Electron. Commun. Probab. **0** (2020), article no. 0, 1–11.
https://doi.org/10.1214/YY-TN
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A noncommutative Freedman inequality*

Andrzej Małek† Adam Osękowski‡

Abstract

The paper contains the extension of Freedman's inequality to the context of self-adjoint martingales on general finite von Neumann algebras.

Keywords: Martingale ; Freedman's inequality ; operator ; exponential ; self-adjoint.
MSC2020 subject classifications: 60G42 ; 46L53.
Submitted to ECP on July 24, 2023, final version accepted on July 12, 2024.

1 Introduction

As evidenced in numerous works, exponential inequalities form a powerful tool to control tail probabilities for random variables satisfying various boundedness conditions. Such a tight control is of significant importance for many areas, including statistics, learning theory, discrete mathematics, statistical mechanics, information theory and convex geometry. The primary goal of this paper is to establish a yet another result in this important direction: we will be concerned with an extension of Freedman's martingale inequality to the more general, operator-algebraic context.

To formulate the statements and relate them to those existing in the literature, let us say a few words about the historical background. Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a fixed probability space. A celebrated result of Bernstein asserts that if X_0, X_1, X_2, $\ldots$ are independent, centered random variables satisfying the uniform one-sided bound $X_n \leq M$, $n = 0, 1, 2, \ldots$, then for any positive number t we have

$$\mathbb{P}\left(\sum_{k=0}^{n} X_k \geq t\right) \leq \exp\left(-\frac{t^2/2}{\sum_{k=0}^{n} \mathbb{E}X_k^2 + Mt/3}\right), \qquad n = 0, 1, 2, \ldots. \tag{1.1}$$

This statement can be successfully generalized to the martingale setting. Suppose that the space $(\Omega, \mathcal{F}, \mathbb{P})$ is filtered by a nondecreasing sequence $(\mathcal{F}_n)_{n\geq 0}$ of sub-σ-algebras of $\mathcal{F}$. Let $f = (f_n)_{n\geq 0}$ be an adapted, real-valued martingale starting from zero, with the associated difference $df = (df_n)_{n\geq 0}$ determined uniquely by the requirement

$$f_n = \sum_{k=0}^{n} df_k \qquad \text{for all } n = 0, 1, 2, \ldots.$$

*A. Osękowski are supported by Narodowe Centrum Nauki (Poland), grant no. 2018/30/Q/ST1/00072.
†University of Warsaw, Poland. E-mail: am406230@students.mimuw.edu.pl
‡University of Warsaw, Poland. E-mail: A.Osekowski@mimuw.edu.pl

Then $s(f)=(s_n(f))_{n\geq 0}$, the conditional square function of f, is given by

$$s_n(f)=\left(\sum_{k=0}^{n}\mathbb{E}(df_k^2|\mathcal{F}_{k-1})\right)^{1/2}, \qquad n=0,\,1,\,2,\,\ldots, \tag{1.2}$$

with the convention $\mathcal{F}_{-1}=\mathcal{F}_0$. Using an appropriate stopping-time argument, Freedman [3] established the following extension of (1.1).

Theorem 1.1. *Suppose that $f=(f_n)_{n\geq 0}$ is a martingale satisfying $f_0=0$, whose difference sequence $df=(df_n)_{n\geq 0}$ satisfies the uniform one-sided bound*

$$df_n\leq M \quad \textit{almost surely for all } n=1,\,2,\,\ldots.$$

Then for any $t>0$ and any $\sigma^2>0$ we have

$$\mathbb{P}\big(\exists_{n\geq 0}\,:\, f_n\geq t \textit{ and } s_n(f)\leq\sigma\big)\leq \exp\left(-\frac{t^2}{2(\sigma^2+Mt/3)}\right). \tag{1.3}$$

We will be interested in the further generalization of the above into the operator context. In [9], Oliveira established a version of (1.3) for matrix-valued martingales and applied it in the study of properties of random graphs. A slight improvement of Oliveira's results can be found in Tropp's paper [19]. To describe these extensions of (1.3), suppose that the martingale f takes values in the class of self-adjoint matrices of dimension $d\times d$. Here the martingale structure is considered entry-wise: that is, for each $i,\,j\in\{1,2,\ldots,d\}$ and any $n=0,\,1,\,2,\ldots$, we have $\mathbb{E}(f_{n+1}^{ij}|\mathcal{F}_n)=f_n^{ij}$ almost surely, where

$$f_n=\begin{pmatrix} f_n^{11} & f_n^{12} & f_n^{13} & \ldots & f_n^{1d}\\ f_n^{21} & f_n^{22} & f_n^{23} & \ldots & f_n^{2d}\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ f_n^{d1} & f_n^{d2} & f_n^{13} & \ldots & f_n^{dd}\end{pmatrix}.$$

Then the associated difference sequence and conditional square function are defined with the same formulas as in the scalar setting. In the statement below, for a given self-adjoint matrix A, $\lambda_{\max}(A)$ stands for its largest eigenvalue.

Theorem 1.2. *Consider a matrix martingale f, whose values are self-adjoint matrices of dimension $d\times d$. Assume that $f_0=0$ and the associated difference sequence df satisfies*

$$\lambda_{\max}(df_n)\leq M \quad \textit{almost surely for } n=1,\,2,\,\ldots.$$

Then for any $t>0$ we have

$$\mathbb{P}\big(\exists_{n\geq 0}\,:\lambda_{\max}(f_n)\geq t \textit{ and } \lambda_{\max}(s_n(f))\leq\sigma\big)\leq d\cdot\exp\left(-\frac{t^2}{2(\sigma^2+Mt/3)}\right). \tag{1.4}$$

This is a perfect extension of Freedman's inequality: if $d=1$, then the estimates (1.3) and (1.4) coincide. Comparing the right-hand sides, we see that the only change in the matrix setting is the appearance of the factor d. It is easy to construct examples showing that this dependence on dimension cannot be removed.

We would like to extend Theorem 1.2 in the following three directions.

(D1) There is a question whether the estimate (1.4) can be improved to capture the matrix structure of martingales; from this viewpoint, Theorem 1.2 is a bit inefficient. To explain this, consider the following two simple examples. Pick an arbitrary real-valued martingale $f = (f_n)_{n\geq 0}$ satisfying $f_0 = 0$, $df_n \leq M$ and $s_n(f) \leq \sigma$ almost surely for each n. Letting

$$f_n^{(1)} = \begin{pmatrix} f_n & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \qquad f_n^{(2)} = \begin{pmatrix} f_n & 0 & 0 & \dots & 0 \\ 0 & f_n & 0 & \dots & 0 \\ 0 & 0 & f_n & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & f_n \end{pmatrix},$$

we easily see that $f^{(1)} = (f_n^{(1)})_{n\geq 0}$ and $f^{(2)} = (f_n^{(2)})_{n\geq 0}$ are matrix martingales satisfying $\lambda_{\max}(df_n^{(j)}) \leq M$ and $\lambda_{\max}(s_n(f^{(j)})) \leq \sigma$ almost surely for all n and $j = 1, 2$. However, for both these processes the estimate (1.4) is weak: indeed, for $j = 1, 2$ we have

$$\begin{aligned} &\mathbb{P}\big(\exists_{n\geq 0} : \lambda_{\max}(f_n^{(j)}) \geq t \text{ and } \lambda_{\max}(s_n(f^{(j)})) \leq \sigma\big) \\ &= \mathbb{P}\big(\exists_{n\geq 0} : f_n \geq t \text{ and } s_n(f) \leq \sigma\big) \leq \exp\left(-\frac{t^2}{2(\sigma^2 + Mt/3)}\right), \end{aligned}$$

where the last passage is due to (1.3). Thus the factor d appearing in (1.4) is not needed for these examples. In a sense, the reason for this flaw is that the estimate refers to the maximal eigenvalue of f only, and does not capture the behavior of the remaining ones (which for $f^{(1)}$ and $f^{(2)}$ is quite different; in the first case, all the remaining eigenvalues are zero; for $f^{(2)}$, all eigenvalues coincide). This flaw is easily removed for $f^{(1)}$: this martingale actually takes values in the class of matrices of dimension 1×1, hence we may apply (1.4) with $d = 1$. Our contribution will enable to fix the problem for $f^{(2)}$: we will replace the probability on the left of (1.4) by a different, bigger functional invoking the full structure of eigenvalues.

(D2) Theorem 1.2 concerns matrix martingales; a natural idea is to generalize the statement to the context of operators acting on separable Hilbert spaces. This is not possible by a simple limiting argument, since the right-hand side of (1.4) depends on the dimension.

(D3) We will also enrich the martingale structures allowed in the exponential estimates. Theorem 1.2 concerns the case in which the martingales are relative to the probabilistic filtration $(\mathcal{F}_n)_{n\geq 0}$; more precisely, it studies matrices whose all entries are scalar-valued martingales. We will extend this context to the case in which conditional expectations refer to specific sub-σ-algebras of operators.

Before we proceed, let us mention that some results in the directions (D2) and (D3) have been obtained in a beautiful paper [6] by Junge and Zeng, and in the more recent work [15] by Randrianantoanina. Our main result can be regarded as a significant refinement of the estimates obtained in those papers. See Section 3 for the discussion.

The remaining part of the paper is organized as follows. The next section contains the description of the background needed for our further investigation. The last part is devoted to the formulation and the proof of our main result.

2 Preliminaries

Let us recall some basic facts from operator theory, which will be needed for our further investigation. For the detailed exposition of the subject we refer the reader to [7, 8, 17]. Throughout, $\mathcal{M}$ will denote a von Neumann algebra with a normal, faithful and finite trace τ. We will restrict ourselves to the probabilistic context and assume that $\tau(I) = 1$, where I is the identity element of $\mathcal{M}$. We may treat $\mathcal{M}$ as a subalgebra of the algebra of all bounded operators on a certain Hilbert space H. Let $\mathcal{P}$ be the lattice of projections on $\mathcal{M}$. given a sequence $(e_n)_{n\geq 1} \subset \mathcal{P}$, we define its infimum by $\bigwedge_{n\geq 1} e_n$: it is a projection onto the subspace $\bigcap_{n\geq 1} e_n(H)$. Two projections e and f are said to be equivalent, if there exists a partial isometry $u \in \mathcal{M}$ such that $uu^* = e$ and $u^*u = f$. Then if $e \wedge f = 0$ (i.e., the intersection $e(H) \cap f(H)$ is a null space), then e is equivalent to a subprojection of $I - f$ and in particular $\tau(e) \leq \tau(I - f)$.

Denote by $|x| = (x^*x)^{1/2}$ the left modulus of $x \in \mathcal{M}$. In addition, for any Borel function f on $\mathbb{R}$ and any self-adjoint operator a admitting the spectral decomposition $a = \int_{-\infty}^{\infty} \lambda de_\lambda$, the operator $f(a)$ is defined spectrally by $f(a) = \int_{-\infty}^{\infty} f(\lambda) de_\lambda$; this in particular allows to consider spectral projections $I_B(a)$ for any Borel subset B of $\mathbb{R}$. For two self-adjoint operators $x, y \in \mathcal{M}$, we will write $x \leq y$ if the difference $y - x$ is positive, i.e., we have $\langle (y-x)\xi, \xi\rangle \geq 0$ for all $\xi \in H$. For $1 \leq p < \infty$, we define the noncommutative L^p-space as the closure of $\mathcal{M}$ with respect to the norm $\|x\|_p = (\tau(|x|^p))^{1/p}$. For $p = \infty$, the space $L^p(\mathcal{M}, \tau)$ coincides with $\mathcal{M}$ with its usual operator norm. Then $(L^p(\mathcal{M}))^* = L^{p'}(\mathcal{M})$ for $1 \leq p < \infty$, where $p' = p/(p-1)$ stands for the conjugate exponent.

We turn our attention to the description of the general setup for noncommutative martingales. The starting point is the appropriate extension of the notion of a conditional expectation. Suppose that $\mathcal{N}$ is a von Neumann subalgebra of $\mathcal{M}$. Then the restriction of τ to $\mathcal{N}$ is a normal and faithful trace on $\mathcal{N}$, and one can isometrically identify $L^1(\mathcal{N}) = L_1(\mathcal{N}, \tau|_{\mathcal{N}})$ as a subspace of $L^1(\mathcal{M})$. Let $\iota : L^1(\mathcal{N}) \to L_1(\mathcal{M})$ be the natural inclusion and let $\mathcal{E} = \iota^* : \mathcal{M} \to \mathcal{N}$ be its adjoint (here we use the duality $(L^1(\mathcal{M}))^* = \mathcal{M}$ and $(L^1(\mathcal{N}))^* = \mathcal{N}$). It can be checked that $\mathcal{E}$ is positive, contractive, normal and satisfies

$$\mathcal{E}(axb) = a\mathcal{E}(x)b \qquad \text{for all } a,\, b \in \mathcal{N} \text{ and all } x \in \mathcal{M}.$$

In addition, it enjoys the trace-preserving property $\tau(\mathcal{E}(x)) = \tau(x)$ for all $x \in \mathcal{M}$. We will call $\mathcal{E}$ the conditional expectation of $\mathcal{M}$ with respect to $\mathcal{N}$.

Next, suppose that $(\mathcal{M}_n)_{n\geq 0}$ is a noncommutative filtration, i.e., a nondecreasing sequence of von Neumann subalgebras of $\mathcal{M}$. Let $(\mathcal{E}_n)_{n\geq 0}$ be the associated sequence of conditional expectations. A sequence $f = (f_n)_{n\geq 0}$ in $L^1(\mathcal{M})$ is called a (noncommutative) martingale with respect to the filtration $(\mathcal{M}_n)_{n\geq 0}$, if for any $n \geq 0$ we have

$$\mathcal{E}_n(f_{n+1}) = f_n.$$

Throughout the text, we will assume that f is self-adjoint, i.e., for any n the operator f_n satisfies $f_n^* = f_n$. The difference sequence of f is defined by $df = (df_n)_{n\geq 0}$, where $df_0 = f_0$ and $df_n = f_n - f_{n-1}$ for $n \geq 1$. Finally, we define the associated conditional square function by

$$s_n(f) = \left(\sum_{k=0}^{n} \mathcal{E}_{k-1}(df_k^2) \right)^{1/2}, \qquad n = 0,\, 1,\, 2,\, \ldots,$$

with the convention $\mathcal{E}_{-1} = \mathcal{E}_0$.

We conclude this short section with the discussion on the operator-valued version of Golden-Thompson inequality. In [16], Sutter et al. proved that if $H_1, H_2, \ldots, H_n$ is a sequence of Hermitian matrices of the same dimension, then we have

$$\left\| \exp\left(\sum_{k=1}^{n} H_k \right) \right\|_{L^p} \leq \sup_{u\in\mathbb{R}} \left\| \prod_{k=1}^{n} \exp\Big((1+iu)H_k \Big) \right\|_{L^p}, \tag{2.1}$$

where $\|\cdot\|_{L^p}$ is the Schatten L^p norm: $\|A\|_{L^p} = (\mathrm{Tr}(|A|^p))^{1/p}$. This estimate extends to random Hermitian matrices, either by adapting the proof, or by combining Fubini's theorem and Jensen's inequality with the following stronger result from [16]:

$$\left\| \exp\left(\sum_{k=1}^{n} H_k \right) \right\|_{L^p} \leq \exp\left\{ \int_{\mathbb{R}} \ln \left\| \prod_{k=1}^{n} \exp((1+iu)H_k) \right\|_{L^p} \mathrm{d}\beta(u) \right\},$$

where β is the probability measure $\beta(u) = \frac{\pi}{2}(\cosh(\pi u)+1)^{-1}$. For the general version of (2.1) for arbitrary self-adjoint operators H_1, H_2, ..., H_n, with the Schatten norm replaced by the L^p norm $\|\cdot\|_{L^p(\mathcal{M})}$, see [5, Theorem 1.2].

3 Noncommutative Freedman inequality

3.1 The main result and related statements in the literature

We are ready to formulate the noncommutative version of Freedman's inequality. Recall that we work in the probabilistic setup: we have $\tau(I) = 1$.

Theorem 3.1. *Suppose that $f = (f_n)_{n\geq 0}$ is a noncommutative self-adjoint martingale with $f_0 = 0$, whose differences satisfy $df_n \leq MI$ for all $n = 0, 1, 2, \ldots$. Then for any $t > 0$ and $\sigma^2 > 0$ there is a projection e such that*

(i) if a nonzero vector ξ lies in $e(H)$, then for any $n = 0, 1, 2, \ldots$,

$$\langle f_n\xi, \xi\rangle < t|\xi|^2 \qquad \text{or} \qquad \langle s_n^2(f)\xi, \xi\rangle > \sigma^2|\xi|^2. \tag{3.1}$$

(ii) we have

$$\tau(I-e) \leq \exp\left(-\frac{t^2}{2(\sigma^2 + Mt/3)}\right). \tag{3.2}$$

Let us discuss the relation of this statement to other results mentioned in the introductory section. First, we immediately see that Theorem 3.1 is a perfect extension of the classical Freedman's inequality, with the same constant. Indeed, when restricted to the commutative setting, the above result asserts the existence of an event E satisfying

$$\mathbb{P}(\Omega\setminus E) \leq \exp\left(-\frac{t^2}{2(\sigma^2 + Mt/3)}\right)$$

on which we have $f_n < t$ or $s_n^2(f) > \sigma^2$ for any n. This yields (1.3), since the event $\{\exists_{n\geq 0} : f_n \geq t \text{ and } s_n(f) \leq \sigma\}$ must be contained in the complement of E. Next, Theorem 3.1 is an extension of Tropp's result. Indeed, consider the von Neumann algebra $\mathcal{M} = M_{d\times d}\otimes L^\infty(\Omega,\mathcal{F},\mathbb{P})$ equipped with the normalized tensor trace $\tau = \mathrm{Tr}\otimes\mathbb{E}/d$ and the filtration $\mathcal{M}_n = M_{d\times d}\otimes L^\infty(\Omega,\mathcal{F}_n,\mathbb{P})$, $n = 0, 1, 2, \ldots$. When specified to this context, Theorem 3.1 implies, for any $t > 0$ and $\sigma^2 > 0$, the existence of a random matrix E, taking values in projections, satisfying

$$\mathbb{E}(\mathrm{Tr}(I-E)) \leq d\exp\left(-\frac{t^2}{2(\sigma^2 + Mt/3)}\right). \tag{3.3}$$

Furthermore, for almost any ω, any nonzero vector ξ lying in the range of $E(\omega)$ and any n we have $\langle f_n(\omega)\xi, \xi\rangle < t|\xi|^2$ or $\langle s_n^2(f)(\omega)\xi,\xi\rangle > \sigma^2|\xi|^2$. On the other hand, suppose that for a given $\omega\in\Omega$, if there is an integer n such that $\lambda_{\max}(f_n(\omega)) \geq t$ and $\lambda_{\max}(s_n(f)(\omega)) \leq \sigma$. This implies the existence of a nonzero vector $\xi = \xi(\omega)$ for which $\langle f_n(\omega)\xi,\xi\rangle \geq t|\xi|^2$ and $\langle s_n^2(f)(\omega)\xi,\xi\rangle \leq \sigma^2|\xi|^2$. Thus such a vector cannot lie in the range of $E(\omega)$, and hence in particular the dimension of $I - E(\omega)$ is at least one. Consequently, we obtain

$$\mathbb{P}\big(\exists_{n\geq 0} : \lambda_{\max}(f_n) \geq t \text{ and } \lambda_{\max}(s_n(f)) \leq \sigma\big) \leq \mathbb{P}(\dim(I-E)\geq 1) \leq \mathbb{E}\big(\mathrm{Tr}(I-E)\big)$$

and it remains to apply (3.3) to get (1.4). Actually, the fact that our approach allows to control $\mathbb{E}\big(\mathrm{Tr}(I-E)\big)$, not only $\mathbb{P}(\dim(I-E)\geq 1)$, is closely related to the direction (D1) discussed in the introductory section. Indeed, recall the martingale $f^{(2)}$ discussed there:

put

$$f_n^{(2)} = \begin{pmatrix} f_n & 0 & 0 & \dots & 0 \\ 0 & f_n & 0 & \dots & 0 \\ 0 & 0 & f_n & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & f_n \end{pmatrix},$$

where $(f_n)_{n\geq 0}$ is a real-valued martingale satisfying $f_0 = 0$, $df_n \leq M$ and $s_n(f) \leq \sigma$ almost surely for each n. Now, for given t and σ^2, if E is the projection guaranteed by Theorem 3.1, then we have $\mathrm{Tr}(I-E) = 0$ or $\mathrm{Tr}(I-E) = d$ almost surely, directly from the form of the martingale $f^{(2)}$. Consequently, we have $1_{\{\dim(I-E)\geq 1\}} = \mathrm{Tr}(I-E)/d$ and hence (3.2) yields

$$\begin{aligned} &\mathbb{P}\big(\exists_{n\geq 0} : \lambda_{\max}(f_n) \geq t \text{ and } \lambda_{\max}(s_n(f)) \leq \sigma\big) \\ &\leq \mathbb{P}(\dim(I-E)\geq 1) = \frac{1}{d}\mathbb{E}\big(\mathrm{Tr}(I-E)\big) \leq \exp\left(-\frac{t^2}{2(\sigma^2+Mt/3)}\right). \end{aligned}$$

Thus our approach fixes the problem indicated in (D1). Next, let us relate Theorem 3.1 to the following result of Junge and Zeng [6].

Theorem 3.2. *Suppose that $(f_n)_{n\geq 0}$ is a noncommutative self-adjoint martingale with $f_0 = 0$, whose differences satisfy $|df_n| \leq MI$ for all $n = 0, 1, 2, \dots$. If the conditional square function satisfies $s_n^2(f) \leq \sigma^2 I$ for all $n = 0, 1, 2, \dots$, then for any n and any $t > 0$,*

$$\tau(I_{[t,\infty)}(f_n)) \leq \exp\left(-\frac{t^2}{8\sigma^2 + 4tM}\right). \tag{3.4}$$

This statement is weaker than our main result. It concerns martingales whose differences satisfy the *two-sided* boundedness condition and assumes that the conditional square function is bounded. Furthermore, it does not capture any interplay between the martingale and its square function on the left-hand side. We should also point out that (3.4) follows from Theorem 3.1, actually with a better constant $\exp\left(-t^2/(2\sigma^2+2Mt/3)\right)$. Indeed, let us apply our result: then for any martingale satisfying the assumptions of Theorem 3.2 there is a projection e enjoying conditions (i) and (ii). Since $s_n^2(f) \leq \sigma^2 I$ for all n, we see that for any $\xi \in e(H)$ the first inequality in (3.1) must hold. This implies $I_{[t,\infty)}(f)(\xi) = 0$, so $e \wedge I_{[t,\infty)}(f) = 0$ and hence $I_{[t,\infty)}(f)$ is equivalent to a subprojection of $I - e$ (see the beginning of Section 2). This gives $\tau(I_{[t,\infty)}(f)) \leq \tau(I-e)$ and the application of (ii) gives the claim.

There is a closely related result of Randrianantoanina, obtained recently in [15].

Theorem 3.3. *Suppose that $(f_n)_{n\geq 0}$ is a noncommutative self-adjoint martingale with $f_0 = 0$, whose differences satisfy $|df_n| \leq MI$ for all $n = 0, 1, 2, \dots$. If the conditional square function satisfies $s_n^2(f) \leq \sigma^2 I$ for all $n = 0, 1, 2, \dots$, then for any n and any $t > 0$,*

$$\tau(I_{[t,\infty)}(f_n)) \leq \exp\left(-\frac{t}{2M}\operatorname{arcsinh}\left(\frac{tM}{2\sigma^2}\right)\right). \tag{3.5}$$

The result improves on (3.4) (it yields better constants), but it has similar drawbacks: the martingale differences are still assumed to satisfy the two-sided boundedness condition, the conditional square function is bounded, and the left-hand side of (3.5) does not involve any interactions between f and $s^2(f)$.

Finally, let us mention another recent work [18] by Talebi et al. The authors study the extension of Freedman's inequality in which the left-hand side is of the weaker form

$$\sup_{n\geq 1} \frac{\tau\big(I_{[t,\infty)}(f_n) \wedge I_{[0,\sigma^2]}(s_n^2(f))\big)}{2^{n-1}},$$

and work under an additional boundedness assumption expressed in terms of f and $s^2(f)$.

3.2 Idea of proof: a glimpse at the classical case

To explain the idea behind the proof of Theorem 3.1, it is convenient to inspect the argument leading to Freedman's inequality in the commutative setting. So, pick a classical martingale $f = (f_n)_{n\geq 0}$ started at zero and satisfying the appropriate boundedness requirement. Fix a positive integer N and two auxiliary parameters α, $\beta > 0$ which will be specified in a moment. We obviously have

$$\mathbb{P}\Big(\exists_{0\leq n<N} \,:\, f_n \geq t \text{ and } s_n^2(f) \leq \sigma^2\Big) \leq \mathbb{P}\Big(\exists_{0\leq n<N} \,:\, \alpha f_n - \beta s_n^2(f) \geq \alpha t - \beta\sigma^2\Big).$$

Now, for any $n = 0, 1, 2, \ldots, N$, denote $x_n = \alpha f_n - \beta s_n^2(f)$ and introduce the stopping time $\tau = \inf\{n \,:\, x_n \geq \alpha t - \beta\sigma^2\}$, with the convention $\inf\emptyset = N$. Then we have

$$\begin{aligned}\mathbb{P}\Big(\exists_{0\leq n<N} \,:\, \alpha f_n - \beta s_n^2(f) \geq \alpha t - \beta\sigma^2\Big) &= \mathbb{P}\Big(\exists_{0\leq n<N} \,:\, x_n \geq \alpha t - \beta\sigma^2\Big) \\ &\leq \mathbb{P}(x_\tau \geq \alpha t - \beta\sigma^2) \leq e^{-\alpha t+\beta\sigma^2}\mathbb{E}e^{x_\tau},\end{aligned}$$

where the last passage is due to Chebyshev's inequality. Now, one can show (see Proposition 3.5 below for the more general fact in the noncommutative setting) that if we set β to be equal to $(e^{\alpha M} - 1 - \alpha M)/(\alpha^2 M^2)$, then $(e^{x_n})_{0\leq n\leq N}$ is a supermartingale and hence $\mathbb{E}e^{x_\tau} \leq \mathbb{E}e^{x_0} \leq 1$. Plugging this above and optimizing over α one gets

$$\mathbb{P}(\exists_{0\leq n<N} \,:\, f_n \geq t \text{ and } s_n^2(f) \leq \sigma^2) \leq \exp\left(-\frac{t^2/2}{\sigma^2 + Mt/3}\right).$$

It remains to let $N \to \infty$ to get Freedman's inequality.

The above proof cannot be directly carried over to the noncommutative setting, since the notion of stopping time does not make any sense in the operator setting. However, what actually matters, is not the stopping time τ itself, but rather the stopped process x_τ and the key estimate for $\mathbb{E}e^{x_\tau}$. In our considerations below, we will construct an appropriate noncommutative version of this stopped process; as a by-product, this will also lead us to the desired projection e appearing in the assertion of Theorem 3.1. The construction rests on the following simple, yet crucial observation: the (commutative) stopped process x_τ is given by

$$x_\tau = \sum_{k=1}^N (x_k - x_{k-1})1_{\{\tau\geq k\}} = \sum_{k=1}^N (\alpha df_k - \beta\mathcal{E}_{k-1}(df_k^2))1_{\{\tau\geq k\}}.$$

3.3 Noncommutative case

We fix a martingale $(f_n)_{n\geq 0}$ as in the statement of Theorem 3.1 and, as above, set $x_n = \alpha f_n - \beta s_n^2(f)$, where $\alpha > 0$ and $\beta = (e^{\alpha M} - 1 - \alpha M)/(\alpha^2 M^2)$. The key point is that while τ does not seem to be transferable to this setting, the indicator functions $1_{\{\tau\geq k\}}$ can be extended with the use of approach originating in the work of Cuculescu [2]. The argument, generally speaking, rests on the construction of a certain class of projections, and its various modifications have been successfully exploited in many contexts of noncommutative probability and harmonic analysis (cf. [4, 10, 11, 12, 13, 14], for example). In our case, we introduce the sequence $(R_n)_{n=0}^N$ of projections and the sequence $(X_n)_{n=0}^N$ of self-adjoint operators by the following recursive procedure. We set

$R_0 = I$, $X_0 = 0$ and

$$X_n = \sum_{k=1}^{n} R_{k-1}\Big(\alpha df_k - \beta\mathcal{E}_{k-1}(df_k^2)\Big)R_{k-1},$$
$$R_n = R_{n-1}I_{(-\infty,\alpha t-\beta\sigma^2)}(R_{n-1}X_nR_{n-1})$$

for $n = 1, 2, \ldots, N$. It is easy to check that in the commutative case, the operator X_n is precisely the stopped variable $x_{\tau\wedge n}$, while R_n becomes the indicator function $1_{\{\tau>n\}}$. Thus, in particular, the operator X_N is an extension of the stopped variable x_τ; it has all the necessary properties, as we will check later. The desired projection e will be equal to the intersection $\bigwedge_{N\geq 1} R_N$.

Let us record some useful properties of the projections $(R_n)_{n=0}^N$. The proof is immediate, we omit the straightforward details.

Lemma 3.4. *The sequence $(R_n)_{n=0}^N$ enjoys the following.*

(i) It is nonincreasing: we have $R_n \leq R_{n-1}$ for any $n = 1, 2, \ldots, N$.

(ii) For any $n = 1, 2, \ldots, N$, the projection R_n commutes with $R_{n-1}X_nR_{n-1}$.

(iii) For any $n = 1, 2, \ldots, N$ we have

$$(R_{n-1} - R_n)X_n(R_{n-1} - R_n) \geq (\alpha t - \beta\sigma^2)(R_{n-1} - R_n).$$

Furthermore, if ξ is a nonzero element of $R_n(H)$, then

$$\langle R_nX_nR_n\xi, \xi\rangle < (\alpha t - \beta\sigma^2)|\xi|^2.$$

(iv) We have $R_n \in \mathcal{M}_n$ for any $n = 0, 1, 2, \ldots, N$.

We will also need the following supermartingale property.

Lemma 3.5. *For any $n = 1, 2, \ldots, N$ we have $\tau(e^{X_n}) \leq \tau(e^{X_{n-1}})$.*

Proof. We start with the observation that for any k we have

$$\begin{aligned}\mathcal{E}_{k-1}\big(e^{\alpha R_{k-1}df_kR_{k-1}}\big) &= R_{k-1} + R_{k-1}\mathcal{E}_{k-1}\Big(e^{\alpha R_{k-1}df_kR_{k-1}} - I - \alpha R_{k-1}df_kR_{k-1}\Big)R_{k-1}\\ &\leq I + \frac{e^{\alpha M} - 1 - \alpha M}{M^2}R_{k-1}\mathcal{E}_{k-1}\big(R_{k-1}df_kR_{k-1}df_kR_{k-1}\big)R_{k-1}\\ &\leq I + \beta R_{k-1}\mathcal{E}_{k-1}\big(df_k^2\big)R_{k-1}.\end{aligned}$$

Here in the first passage we have used the martingale property $\mathcal{E}_{k-1}(df_k) = 0$, while the first inequality is due to the assumption $df_k \leq MI$ and the pointwise estimate

$$e^x - 1 - x \leq \frac{e^{\alpha M} - 1 - \alpha M}{\alpha^2M^2}x^2 \qquad \text{for } x \leq \alpha M.$$

Since $1 + x \leq e^x$ for all $x \in \mathbb{R}$, we obtain

$$\mathcal{E}_{k-1}\big(e^{\alpha R_{k-1}df_kR_{k-1}}\big) \leq \exp\big(\beta R_{k-1}\mathcal{E}_{k-1}(df_k^2)R_{k-1}\big)$$

and hence for any real number u, we have

$$\begin{aligned}&\exp\left(-(1-iu)\cdot\frac{\beta}{2}R_{k-1}\mathcal{E}_{k-1}(df_k^2)R_{k-1}\right)\\ &\qquad\mathcal{E}_{k-1}\big(e^{\alpha R_{k-1}df_kR_{k-1}}\big)\exp\left(-(1+iu)\cdot\frac{\beta}{2}R_{k-1}\mathcal{E}_{k-1}(df_k^2)R_{k-1}\right) \leq I.\end{aligned} \tag{3.6}$$

Now, we apply (2.1), with $p=2$, $H_1=\frac{\alpha}{2}R_{k-1}df_kR_{k-1}$, $H_2=-\frac{\beta}{2}R_{k-1}\mathcal{E}_{k-1}(df_k^2)R_{k-1}$ and $H_3=\frac{1}{2}X_{k-1}$. The operators H_1, H_2, H_3 are self-adjoint, so we get

$$\tau(e^{X_k})=\tau(e^{2(H_1+H_2+H_3)})=\left\|e^{H_1+H_2+H_3}\right\|_2^2\leq\sup_{u\in\mathbb{R}}\left\|e^{(1+iu)H_1}e^{(1+iu)H_2}e^{(1+iu)H_3}\right\|_2^2.$$

But the latter expression is equal to

$$\sup_{u\in\mathbb{R}}\tau\Big(e^{(1-iu)H_3}e^{(1-iu)H_2}e^{(1-iu)H_1}e^{(1+iu)H_1}e^{(1+iu)H_2}e^{(1+iu)H_3}\Big)$$
$$=\sup_{u\in\mathbb{R}}\tau\Big(e^{(1-iu)H_3}e^{(1-iu)H_2}e^{2H_1}e^{(1+iu)H_2}e^{(1+iu)H_3}\Big).$$

Note that H_2 and H_3 are measurable with respect to $\mathcal{M}_{k-1}$. Consequently, inserting the conditional expectation with respect to $\mathcal{E}_{k-1}$, we see that the above quantity equals

$$\sup_{u\in\mathbb{R}}\tau\Big(e^{(1-iu)H_3}e^{(1-iu)H_2}\mathcal{E}_{k-1}(e^{2H_1})e^{(1+iu)H_2}e^{(1+iu)H_3}\Big).$$

But by (3.6), we have $e^{(1-iu)H_2}\mathcal{E}_{k-1}(e^{2H_1})e^{(1+iu)H_2}\leq I$. This gives

$$\tau(e^{X_k})\leq\sup_{u\in\mathbb{R}}\tau\Big(e^{(1-iu)H_3}e^{(1+iu)H_3}\Big)=\tau(e^{2H_3})=\tau(e^{X_{k-1}})$$

and completes the proof. □

We are ready for the proof of our main result.

Proof of Theorem 3.1. Suppose that a nonzero vector ξ belongs to $R_N(H)$. Then by the monotonicity of $(R_n)_{n=0}^N$, we have $\xi\in R_n(H)$ for any $n=0,1,2,\ldots,N$ and hence

$$\langle x_n\xi,\xi\rangle=\langle x_nR_n\xi,R_n\xi\rangle=\langle R_nx_nR_n\xi,\xi\rangle$$

(recall that $x_n=\alpha f_n-\beta s_n^2(f)$). However, by the very definition of X_n and the monotonicity of R_k's, we have $R_nx_nR_n=R_nX_nR_n$, so by Lemma 3.4 (iii),

$$\langle x_n\xi,\xi\rangle=\langle R_nX_nR_n\xi,\xi\rangle<(\alpha t-\beta\sigma^2)|\xi|^2. \tag{3.7}$$

Next, observe that by the monotonicity again, the difference $R_{n-1}-R_n$ is again a projection. We will show that for any vectors $\xi\in(R_{n-1}-R_n)(H)$ and $\eta\in(R_{m-1}-R_m)(H)$, for some $m\neq n$, we have $\langle X_N\xi,\eta\rangle=0$ (at the operator level, this is equivalent to saying that X_N commutes with $R_{m-1}-R_m$). Since X_N is self-adjoint, we may assume that $n<m$. Then $R_{m-1}-R_m\leq R_{m-1}\leq R_{n-1}$ and we may write

$$\langle X_N\xi,\eta\rangle=\langle X_N(R_{n-1}-R_n)\xi,(R_{m-1}-R_m)\eta\rangle=\langle R_{n-1}X_N(R_{n-1}-R_n)\xi,(R_{m-1}-R_m)\eta\rangle.$$

Now we have $R_{n-1}X_N(R_{n-1}-R_n)=R_{n-1}X_n(R_{n-1}-R_n)=R_{n-1}X_nR_{n-1}(R_{n-1}-R_n)$, since the projection $R_{n-1}-R_n$ annihilates the appropriate summands in the definition of X_N: precisely, for any $k>n$ we have

$$R_{k-1}\Big(\alpha df_k-\beta\mathcal{E}_{k-1}(df_k^2)\Big)R_{k-1}(R_{n-1}-R_n)=0.$$

By Lemma 3.4 (ii), R_n commutes with $R_{n-1}X_nR_{n-1}$ and therefore $R_{n-1}-R_n$ also has this property. This gives $R_{n-1}X_n(R_{n-1}-R_n)=(R_{n-1}-R_n)X_nR_{n-1}$ and hence

$$\langle X_N\xi,\eta\rangle=\langle(R_{n-1}-R_n)X_nR_{n-1}\xi,(R_{m-1}-R_m)\xi\rangle=0,$$

as desired. Here we have used the fact that the ranges of $R_{n-1}-R_n$ and $R_{m-1}-R_m$ are orthogonal (which again is due to the monotonicity of $R_k's$). Consequently, if $\xi \in (I-R_N)(H)$ and $\xi=\xi_1+\xi_2+\ldots+\xi_N$ is the unique decomposition with $\xi_n \in (R_{n-1}-R_n)(H)$ for all n, then

$$\langle X_N\xi,\xi\rangle=\sum_{n=1}^N\langle X_N\xi_n,\xi_n\rangle=\sum_{n=1}^N\big\langle (R_{n-1}-R_n)X_n(R_{n-1}-R_n)\xi,\xi\big\rangle\geq(\alpha t-\beta\sigma^2)|\xi|^2,$$

where the last passage is due to Lemma 3.4 (iii). The above estimate implies that $I-R_N$ is equivalent to a subprojection of $I_{[\alpha t-\beta\sigma^2,\infty)}(X_N)$: indeed, the intersection $(I-R_N)(H)\cap I_{(-\infty,\alpha t-\beta\sigma^2)}(X_N)(H)$ is the trivial, null subspace of H. Consequently, Czebyshev's inequality yields

$$\tau(I-R_N)\leq\tau\big(I_{[\alpha t-\beta\sigma^2,\infty)}(X_N)\big)\leq e^{-\alpha t+\beta\sigma^2}\tau(e^{X_N})\leq e^{-\alpha t+\beta\sigma^2}, \tag{3.8}$$

where in the last line we have exploited Lemma 3.5.

Now we perform the limiting procedure as $N\to\infty$, by considering the projection $e=\bigwedge_{N=1}^\infty R_N$. By (3.7), for any $n=0,\,1,\,2,\,\ldots$ and any nonzero vector $\xi\in e(H)$ we have $\langle x_n\xi,\xi\rangle<(\alpha t-\beta\sigma^2)|\xi|^2$. But $x_n=\alpha f_n-\beta s_n^2(f)$, so the latter estimate implies that (3.1) holds. As for (3.2), we write

$$\tau(I-e)=\lim_{N\to\infty}\tau(I-R_N)\leq e^{-\alpha t+\beta\sigma^2}.$$

Optimizing over α, we get the claim. □

Acknowledgment

The authors would like to thank anonymous Referees for the careful reading of the first version of the paper and several helpful comments and suggestions.

References

[1] Boucheron S., Lugosi G., Massart, P.: Concentration Inequalities. A Nonasymptotic Theory of Independence. *Oxford University Press*, Oxford, 2013. MR3185193

[2] Cuculescu, I.: Martingales on von Neumann algebras. *J. Multivariate Anal.* **1**, (1971), 17–27. MR0295398

[3] Freedman, D. A.: On tail probabilities for martingales. *Ann. Probab.* **3**, (1975), 100–118. MR0380971

[4] Jiao, Y., Osękowski, A. and Wu, L.: Strong differential subordinates for noncommutative submartingales. *Ann. Probab.* **47**, 2019, no. 5, 3108–3142.

[5] Junge, M. and LaRacuente, N.: Multivariate trace inequalities, p-fidelity, and universal recovery beyond tracial settings. *J. Math. Phys.* **63**, (2022), no. 12, Paper No. 122204. MR4520657

[6] Junge M. and Zeng, Q.: Noncommutative martingale deviation and Poincaré type inequalities with applications. *Probab. Theory Related Fields* **161**, (2015), 449–507. MR3334274

[7] Kadison, K. V. and Ringrose, J. R.: Fundamentals of the Theory of Operator Algebras. Vol. I, Elementary Theory, *Academic Press*, New York, 1983. MR0719020

[8] Kadison, R. V. and Ringrose, J. R.: Fundamentals of the theory of operator algebras. Vol. II, Advanced Theory, *Academic Press*, Orlando, FL, 1986. MR0859186

[9] Oliveira, R. I.: Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges, arXiv:0911.0600.

[10] Osękowski, A.: Weak type inequality for noncommutative differentially subordinated martingales. *Probab. Theory Related Fields* **140**, 2008, no. 3-4, 553–568.

[11] Parcet, J.: Pseudo-localization of singular integrals and noncommutative Calderón-Zygmund theory. *J. Funct. Anal.* **256**, 2009, no.2, 509–593.

[12] Parcet, J. and Randrianantoanina, N.: Gundy's decomposition for non-commutative martingales and applications. *Proc. London Math. Soc.* **93**, 2006, no. 1, 227–252.

[13] Randrianantoanina, N.: Non-commutative martingale transforms. *J. Funct. Anal.* **194**, 2002, no. 1, 181–212.

[14] Randrianantoanina, N.: Conditioned square functions for noncommutative martingales. *Ann. Probab.* **35**, 2007, no. 3, 1039–1070.

[15] Randrianantoanina, N.: Triple operator version of the Golden-Thompson inequality for traces on von Neumann algebras. *Ann. Inst. Fourier (Grenoble)* **74**, 2024, no. 1, 193–233.

[16] Sutter, D., Berta M. and Tomamichel, M.: Multivariate trace inequalities. *Comm. Math. Phys.* **352**, 2017, 37–58. MR3623253

[17] Takesaki, M.: Theory of Operator Algebras. I, *Springer-Verlag*, New York, 1979. MR0548728

[18] Talebi, A., Sadeghi, Gh. and Moslehian, M. S.: Freedman inequality in noncommutative probability spaces. *Complex Anal. Oper. Theory* **16**, 2022, no. 2, Paper No. 22, 18 pp.

[19] Tropp, J. A.: Freedman's inequality for matrix martingales. *Elect. Comm. in Probab.* **16**, (2011), 262–270. MR2802042