Regular Banach Spaces and Large Deviations of Random Sums

Arkadi Nemirovski*

Working paper, version of December 13, 2004

1 Overview

A typical result on large deviations of sums with random terms states that if ξ_t are independent scalar random variables with zero means and such that "ξ_t has as light tails as a Gaussian $\mathcal{N}(0, 4\sigma_t^2)$ random variable", specifically,

$$\mathbf{E}\left\{\exp\{\xi_t^2/\sigma_t^2\}\right\} \leq O(1), \tag{1}$$

and

$$S_N = \sum_{t=1}^N \xi_t,$$

then

$$\text{Prob}\left\{|S_N| > t\sqrt{\sigma_1^2 + ... + \sigma_N^2}\right\} \leq O(1)\exp\{-O(1)t^2\} \tag{2}$$

(from now on, all $O(1)$'s are appropriate positive absolute constants). Our goal is to get similar results for the case when ξ_t are independent random vectors with zero means in a finite-dimensional vector space E equipped with norm $\|\cdot\|$, $S_N = \sum\limits_{t=1}^N \xi_t$ and the "light tail" condition (1) is stated as

$$\mathbf{E}\left\{\exp\{\|\xi_t\|^2/\sigma_t^2\}\right\} \leq \exp\{1\}. \tag{3}$$

Note that a straightforward guess

$$\begin{array}{l}\mathbf{E}\{\xi_t\} = 0 \forall t \ \& \ (3) \ \& \ \{\xi_t\} \text{ are independent} \\ \qquad \Rightarrow \text{Prob}\left\{\|S_N\| > t\sqrt{\sigma_1^2 + ... + \sigma_N^2}\right\} \leq O(1)\exp\{-O(1)t^2\}\end{array} \tag{4}$$

is not true, as it is shown by the following example:

- $E = \mathbf{R}^n$, $\|x\| = \|x\|_1 \equiv \sum\limits_j |x_j|$,
- $(\xi_t)_j = \begin{cases}\epsilon_t, & j = t (\mathrm{mod}\, n) \\ 0, & \text{otherwise}\end{cases}$, where $\epsilon_1, \epsilon_2, ...$ are independent random variables taking values ± 1 with probability 1/2,
- $\sigma_t = 1$, $i \geq 1$.

We have

$$\|S_n\|_1 \equiv n = \sqrt{n}\sqrt{\sigma_1^2 + ... + \sigma_n^2},$$

*Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel, nemirovs@ie.technion.ac.il

so that all we can hope for is a relation of the type

$$\begin{array}{l}\mathbf{E}\{\xi_t\} = 0 \forall t \ \&\ (3)\ \&\ \{\xi_t\} \text{ are independent} \\ \qquad \Rightarrow \text{Prob}\left\{\|S_N\| > t\Theta\sqrt{\sigma_1^2 + ... + \sigma_N^2}\right\} \leq O(1)\exp\{-O(1)t^2\}\end{array} \tag{5}$$

with *depending on* $(E, \|\cdot\|)$ factor Θ which, as our example clearly shows, can be as large as $\sqrt{\dim E}$.

Our major goal is to demonstrate that (5) indeed holds true, provided that Θ^2 is an upper bound on the "constant of regularity" $\kappa(E, \|\cdot\|)$ of $(E, \|\cdot\|)$, with the latter notion defined as follows:

Definition 1.1 *Let $(E, \|\cdot\|)$ be a Banach space, and let $\kappa \geq 1$.*
(i) *Space $(E, \|\cdot\|)$ is called κ-smooth, if the function $p(x) = \|x\|^2$ is continuously differentiable and*

$$\forall x, y \in E : p(x + y) \leq p(x) + Dp(x)[y] + \kappa p(y). \tag{6}$$

(ii) *Space $(E, \|\cdot\|)$ (and the norm $\|\cdot\|$ on E) is called κ-regular, if there exists $\kappa_+ \in [1, \kappa]$ and a norm $\|\cdot\|_+$ on E such that $(E, \|\cdot\|_+)$ is κ_+-smooth and $\|\cdot\|_+$ is κ/κ_+-compatible with $\|\cdot\|$, that is,*

$$\forall x \in E : \|x\|^2 \leq \|x\|_+^2 \leq \frac{\kappa}{\kappa_+}\|x\|^2. \tag{7}$$

The constant $\kappa(E, \|\cdot\|)$ of regularity of $E, \|\cdot\|$ is the infinum of those $\kappa \geq 1$ for which $(E, \|\cdot\|)$ is κ-regular.

In the sequel we

1. Provide a number of interesting examples of normed spaces with "nearly dimension-independent" constants of regularity. Specifically, we demonstrate in Section 2.1 that if $p \geq 2$, then the norm $|X|_p$ on the space of $m \times n$ matrices X:

   $$|X|_p = \|\sigma(X)\|_p \qquad\qquad [\sigma(X) \text{ is the vector of singular values of } X]$$

   is κ-regular with $\kappa = \min[p + 1, (2\ln(\min[m, n]) + 1)\exp\{1\}]$.

2. Demonstrate (Section 3) that (5) indeed is true, provided that $\Theta \geq \sqrt{\kappa(E, \|\cdot\|)}$. In particular,

   (!!) *If ξ_t are independent random $m \times n$ matrices with zero mean such that*

   $$\mathbf{E}\left\{\exp\{|\xi_t|_\infty^2 \sigma_t^{-2}\}\right\} \leq \exp\{1\},$$

   where $|X|_\infty = \|\sigma(X)\|_\infty$ is the usual matrix norm of X, then

   $$\text{Prob}\left\{|\sum_{t=1}^N \xi_t|_\infty \geq t\sqrt{\ln(\min[m+1, n+1])}\sqrt{\sigma_1^2 + ... + \sigma_N^2}\right\} \leq O(1)\exp\{-O(1)t^2\}.$$

Note also that every norm $\|\cdot\|$ on $\mathbf{R}^n$ is n-compatible with Euclidean norm, and the latter is, of course, 1-smooth. Thus, every norm on $\mathbf{R}^n$ is n-regular; in other words, the $(\mathbf{R}^n, \|\cdot\|_1)$-example, where the factor $\Theta = \sqrt{n}$ in (5) is a must, is exactly the worst case.

Major part of the results to follow were announced in [3].

2 Regular Banach spaces

2.1 Basic examples

Example 2.1 *Let $2 \leq p \leq \infty$. The space $(\mathbf{R}^n, \|\cdot\|_p)$ with $n \geq 3$ is $\kappa_p(n)$-regular with*

$$\kappa_p(n) = \min_{\substack{2 \leq \rho \leq p \\ \rho < \infty}} (\rho - 1)n^{\frac{2}{\rho} - \frac{2}{p}} \leq \min[p - 1, 2\ln(n)] \tag{8}$$

Proof. Let $2 \le \rho < \infty$. We claim that in this case the space $(R^n, \|\cdot\|_\rho)$ is $(\rho-1)$-smooth. Indeed, the function $p(x) = \|\cdot\|_\rho^2$ is convex, continuously differentiable everywhere and twice continuously differentiable outside of the origin; for such a function, (6) holds true if and only if

$$D^2 p(x)[h,h] \le 2\kappa_+ p(h) \ \forall (x, h \in E, x \neq 0); \tag{9}$$

since $p(\cdot)$ is homogeneous of degree 2, the validity of (9) for all x, h is equivalent to the validity of the relation for all h and all x normalized by the requirement $p(x) = 1$. Given such an x and h and assuming $\rho > 2$, we have

$$\begin{array}{rcl}
Dp(x)[h] &=& 2\left(\sum_i |x_i|^\rho\right)^{\frac{2}{\rho}-1} \sum_i |x_i|^{\rho-1}\mathrm{sign}(x_i)h_i \\
D^2p(x)[h,h] &=& 2\underbrace{\left(\frac{2}{\rho}-1\right)}_{\le 0}\left(\sum_i |x_i|^\rho\right)^{\frac{2}{\rho}-2}\left(\sum_i |x_i|^{\rho-1}\mathrm{sign}(x_i)h_i\right)^2 \\
&& +2\underbrace{\Big(\sum_i |x_i|^\rho\Big)^{\frac{2}{\rho}-1}}_{=1} \sum_i (\rho-1)|x_i|^{\rho-2}h_i^2 \le 2(\rho-1)\sum_i |x_i|^{\rho-2}h_i^2 \\
&\le& 2(\rho-1)\left(\sum_i (|x_i|^{\rho-2})^{\frac{\rho}{\rho-2}}\right)^{\frac{\rho-2}{\rho}}\left(\sum_i (|h_i|^2)^{\frac{\rho}{2}}\right)^{\frac{2}{\rho}} \\
&=& 2(\rho-1)\|h\|_\rho^2 = 2(\rho-1)p(h)
\end{array}$$

as required in (9) when $\kappa_+ = \rho - 1$. In the case of $\rho = 2$ relation (9) with $\kappa_+ = \rho - 1 = 1$ is evident.

Now, when $\rho \in [2, p]$ and $x \in \mathbf{R}^n$, one has $\|x\|_\rho^2 / \|x\|_p^2 \in [1, n^{\frac{2}{\rho}-\frac{2}{p}}]$, so that $(\mathbf{R}^n, \|\cdot\|_p)$ is κ-regular with $\kappa = (\rho-1)n^{\frac{2}{\rho}-\frac{2}{p}}$, and (8) follows. ∎

Example 2.2 *Let $2 \le p \le \infty$. The norm $|X|_p = \|\sigma(X)\|_p$ on the space $\mathbf{R}^{m\times n}$ of $m \times n$ real matrices, where $\sigma(X)$ is the vector of singular values of X, is $\kappa_p(m,n)$-regular, with*

$$\kappa_p(m,n) = \min_{\substack{2 \le \rho < \infty \\ \rho \le p}} \max[2, \rho-1](\min(m,n))^{\frac{2}{\rho}-\frac{2}{p}} \le \min\left[\max[2, p-1], (2\ln(\min[m,n]+2)-1)\exp\{1\}\right]. \tag{10}$$

Proof. 1^0. We start with the following

Lemma 1 *Let $\rho \ge 2$. Then the space $\mathbf{S}^n$ of symmetric $n \times n$ matrices with the norm $|X|_\rho$ is κ-smooth with*

$$\kappa = \max[2, \rho - 1]. \tag{11}$$

Proof. The statement is evident when $\rho = 2$; thus, from now on we assume that $\rho > 2$.

A. We start with the following fact which is important by its own right (for the proof, see Appendix):

Proposition 2.1 *Let Δ be an open interval on the axis, and f be a C^2 function on Δ such that for certain $\theta_\pm, \mu_\pm \in \mathbf{R}$ one has*

$$\forall (a < b, a, b \in \Delta): \theta_- \frac{f''(a) + f''(b)}{2} + \mu_- \le \frac{f'(b) - f'(a)}{b - a} \le \theta_+ \frac{f''(a) + f''(b)}{2} + \mu_+ \tag{12}$$

Let, further, $\mathcal{X}_n(\Delta)$ be the set of all $n \times n$ symmetric matrices with eigenvalues belonging to Δ. The function

$$F(X) = \mathrm{Tr}(f(X)) : \mathcal{X}_n(\Delta) \to \mathbf{R}$$

is C^2, and for every $X \in \mathcal{X}_n(\Delta)$ and every $H \in \mathbf{S}^n$ one has

$$\theta_- \mathrm{Tr}(Hf''(X)H) + \mu_- \mathrm{Tr}(H^2) \le D^2F(X)[H,H] \le \theta_+ \mathrm{Tr}(Hf''(X)H) + \mu_+ \mathrm{Tr}(H^2). \tag{13}$$

B. Let us apply Proposition 2.1 to $\Delta = \mathbf{R}$, $f(t) = |t|^\rho$ with $\theta_- = \mu_- = 0$, $\mu_+ = 0$ and $\theta_+ = \max\left[\frac{2}{\rho-1}, 1\right]$ (this choice, as it is immediately seen, satisfies (12)). By Proposition, the function $F(X) = |X|_\rho^\rho$ on $\mathbf{S}^n$ is twice continuously differentiable, and

$$\forall X, H : 0 \leq D^2F(X)[H,H] \leq \theta_+ \mathrm{Tr}(f''(x)H^2),\ \theta_+ = \max\left[\frac{2}{\rho-1}, 1\right]. \tag{14}$$

It follows that the function $p(X) = |X|_\rho^2 = (F(X))^{\frac{2}{\rho}}$ is continuously differentiable everywhere and twice continuously differentiable outside of the origin. For $X \neq 0$ we have $Dp(X)[H] = \frac{2}{\rho}(F(X))^{\frac{2}{\rho}-1}DF(X)[H]$, whence

$$\begin{array}{l} X \neq 0 \Rightarrow D^2p(X)[H,H] = \frac{2}{\rho}\underbrace{\left[\frac{2}{\rho}-1\right]}_{<0}(F(X))^{\frac{2}{\rho}-2}(DF(X)[H])^2 + \frac{2}{\rho}(F(X))^{\frac{2}{\rho}-1}D^2F(X)[H,H] \\ \leq \frac{2}{\rho}(F(X))^{\frac{2}{\rho}-1}\theta_+\mathrm{Tr}(f''(x)H^2). \end{array} \tag{15}$$

Setting $Z = \frac{1}{\rho(\rho-1)}(F(X))^{\frac{2}{\rho}-1}f''(X)$, $p = \frac{\rho}{\rho-2}$, it is immediately seen that $|Z|_p = 1$. From (15) we have

$$D^2p(X)[H,H] \leq 2\Theta_+(\rho-1)\mathrm{Tr}(ZH^2) \leq 2\theta_+(\rho-1)|Z|_p|H^2|_{\frac{p}{p-1}} = 2\theta_+(\rho-1)|H^2|_{\frac{\rho}{2}} = 2\theta_+(\rho-1)|H|_\rho^2. \tag{16}$$

Now, if $X, Y \in \mathbf{S}^n$ are such that the segment $[X; X+Y]$ does not contain the origin, then

$$\exists \gamma \in (0,1) : p(X+Y) \leq p(X) + Dp(X)[Y] + \frac{1}{2}D^2p(X+\gamma Y)[Y,Y],$$

and (16) implies that for the outlined X, Y one has

$$p(X+Y) \leq p(X) + Dp(X)[Y] + \theta_+(\rho-1)p(Y).$$

Since p is C^1, the resulting inequality, by continuity, is valid for all X, Y. ∎

2^0. Now we can complete the justification of Example 2.2. W.l.o.g. we may assume that $m \leq n$. Given an $m \times n$ matrix X, let $S(X) = \left[\begin{array}{c|c} & X \\ \hline X^T & \end{array}\right] \in \mathbf{S}^{m+n}$. One clearly has

$$\|\sigma(X)\|_\rho = |X|_\rho = 2^{-1/\rho}|S(X)|_\rho,$$

whence, by Lemma 1 and due to the fact that the mapping $X \mapsto S(X) : \mathbf{R}^{m\times n} \to \mathbf{S}^{m+n}$ is linear, the norm $|\cdot|_\rho$, treated as a norm on $\mathbf{R}^{m\times n}$, is $\max[2, \rho-1]$-smooth whenever $\rho \geq 2$. Since $\sigma(X) \in \mathbf{R}^m$ for $X \in \mathbf{R}^{m\times n}$, for every $\rho \in [2, \infty)$ such that $\rho \leq p$ one has

$$|X|_p^2 \leq |X|_\rho^2 \leq m^{\frac{2}{\rho}-\frac{2}{p}}|X|_p^2.$$

Thus, the space $(\mathbf{R}^{m\times n}, |\cdot|_p)$ is κ-regular with

$$\kappa = \min_{\substack{2\leq\rho<\infty \\ \rho\leq p}} \max[2, \rho-1]m^{\frac{2}{\rho}-\frac{2}{p}},$$

and we arrive at (10). ∎

2.2 "Calculus" of regular spaces

We start with the following well-known fact:

Proposition 2.2 *Let E be finite-dimensional, $\|\cdot\|$ be a norm on E, E^* be the space dual to E. $\|\cdot\|_*$ be the norm on E^* dual to $\|\cdot\|$, $\langle\xi, x\rangle$ be the value of a linear form $x \in E^*$ on a vector $x \in E$. Let also $f(x) = \frac{1}{2}\|x\|^2 : E \to \mathbf{R}$ and $f_*(\xi) = \frac{1}{2}\|\xi\|_*^2 : E^* \to \mathbf{R}$. The following 6 properties are equivalent to each other:*

(i) $(E, \|\cdot\|)$ *is* κ*-smooth;*

(ii) $\partial f(x) = \{f'(x)\}$ *is a singleton for every* x*, and*

$$\langle f'(x) - f'(y), x - y\rangle \leq \kappa\|x-y\|^2 \ \forall x, y \in E; \tag{17}$$

(iii) f *is continuously differentiable, and* $f'(\cdot)$ *is Lipschitz continuous with constant* κ*:*

$$\|f'(x) - f'(y)\|_* \leq \kappa\|x - y\| \ \forall x, y \in E; \tag{18}$$

(iv) One has

$$\forall(\xi, \eta \in E^*, x \in \partial f_*(\xi), y \in \partial f_*(\eta)) : \langle \xi - \eta, x - y\rangle \geq \kappa^{-1}\|\xi - \eta\|_*^2;$$

(v) One has

$$\forall(\xi, \eta \in E^*, x \in \partial f_*(\xi), y \in \partial f_*(\eta)) : \|x - y\| \geq \kappa^{-1}\|\xi - \eta\|_*;$$

(vi) One has

$$\forall(\xi, \eta \in E_*, x \in \partial f_*(\xi)) : f_*(\xi + \eta) \geq f_*(\xi) + \langle \eta, x\rangle + \frac{1}{2\kappa}\|\eta\|_*^2.$$

Proof. (i)⇒(iii): We are in the situation when f is continuously differentiable. Convolving $f(\cdot)$ with smooth nonnegative kernels $\delta_k(\cdot)$ with unit integral and support shrinking to origin as $k \to \infty$, we get a sequence $f_k(\cdot)$ of smooth functions converging to $f(\cdot)$, along with first order derivatives, uniformly on compact sets. We have

$$\begin{array}{rcl} f_k(x+y) & = & \int f(x-z+y)\delta(z)dz \leq \int [f(x-z) + \langle f'(x-z), y\rangle + \kappa f(y)]\delta(z)dz \\ & = & f_k(x) + \langle f_k'(x), y\rangle + \kappa f(y) \end{array}$$

From the resulting inequality combined with smoothness and convexity of f_k it follows that

$$0 \leq D^2 f_k(x)[h, h] \leq \kappa\|h\|^2 \ \forall x, h \in E.$$

Thus, if $\|h\| = \|d\| = 1$, then $4D^2 f_k(x)[h, d] = D^2 f_k(x)[h+d, h+d] - D^2 f_k(x)[h-d, h-d] \leq \kappa\|h+d\|^2 \leq 4\kappa$, whence $D^2 f_k(x)[h, d] \leq \kappa$ whenever $\|h\| = \|d\| = 1$, or, which is the same by homogeneity,

$$|D^2 f_k(x)[h, d]| \leq \kappa\|h\|\|d\| \ \forall x, h, d.$$

Consequently,

$$|\langle f_k'(y) - f_k'(x), h\rangle| = |\int_0^1 D^2 f_k(x + t(y-x))[y - x, h]dt| \leq \int_0^1 \kappa\|y - x\|\|h\|dt \leq \kappa\|y - x\|\|h\|,$$

whence, taking maximum over h with $\|h\| = 1$, $\|f_k'(y) - f_k'(x)\|_* \leq \kappa\|y - x\|$. As $k \to \infty$, $f_k'(x)$ converge to $f'(x)$, and we conclude that $f'(\cdot)$ possesses the required Lipschitz continuity, Q.E.D.
(iii)⇒(ii): evident
(ii)⇒(i): A convex function on $\mathbf{R}^n$ with a singleton differential at every point clearly is continuously differentiable, so that in the case of (ii) f is continuously differentiable. Besides this, in the case of (ii) we have

$$\begin{array}{rcl} f(x+y) & = & f(x) + \langle f'(x), y\rangle + \int_0^1 \langle f'(x+ty) - f'(x), y\rangle dt \\ & \leq & f(x) + \langle f'(x), y\rangle + \int_0^1 \kappa t\|y\|^2 dt = f(x) + \langle f'(x), y\rangle + \kappa f(y), \end{array}$$

which immediately implies (6) (recall that $\|\cdot\|^2 = 2f(\cdot)$), Q.E.D.
(iii)⇔(v): The functions $f(\cdot)$, $f_*(\cdot)$ are the Legendre transforms of each other, so that $x \in \partial f_*(\xi)$ if and only if $\xi \in \partial f(x)$. Now let (iii) be the case, and let $\xi, \eta \in E^*$ and $x \in \partial f_*(\xi)$, $y \in \partial f_*(\eta)$. Then $\xi = f'(x)$,

$\eta = f'(y)$ and therefore, due to (iii), $\|\xi-\eta\|_* \le \kappa\|x-y\|$, so that (v) takes place. Vice versa, let (v) take place, and let $x, y \in E$, $\xi \in \partial f(x)$, $\eta \in \partial f(y)$. Then $x \in \partial f_*(\xi)$, $y \in \partial f_*(y)$, and therefore (v) says that $\|\xi-\eta\|_* \le \kappa\|x-y\|$. We conclude that if $x = y$, then $\xi = \eta$, that is, $\partial f(x)$ always is a singleton, meaning that f is continuously differentiable, and that the inequality in (iii) takes place, that is, (iii) holds true, Q.E.D.

(iv)⇔(iii): Let (iv) take place. If there exists $x \in E$ such that $\partial f(x)$ is not a singleton, then, choosing $\xi, \eta \in \partial f(x)$ with $\xi \neq \eta$, we would have $x \in \partial f_*(\xi)$, $x \in \partial f_*(\eta)$, whence by (iv) we should have $\langle \xi - \eta, x - x\rangle \ge \kappa^{-1}\|\xi-\eta\|_*^2$, which is impossible. Thus, $\partial f(x)$ is a singleton for every x, so that f is continuously differentiable. Besides this, with $x, y \in E$ and $\xi = f'(x)$, $\eta = f'(y)$ we have $x \in \partial f_*(\xi)$, $y \in \partial f_*(\eta)$, whence, by (iv), $\langle \xi-\eta, x-y\rangle \ge \kappa^{-1}\|\xi-\eta\|_*^2$. Since $\langle \xi-\eta, x-y\rangle \le \|\xi-\eta\|_*\|x-y\|$, we get $\|\xi-\eta\|_*\|x-y\| \ge \kappa^{-1}\|\xi-\eta\|_*^2$, whence $\|\xi-\eta\|_* = \|f'(x)-f'(y)\|_* \le \kappa\|x-y\|$, and thus (iii) takes place.

Now let (iii) take place, and let us prove that (iv) takes place as well, or, which is the same in the case of (iii), that $\langle f'(x)-f'(y), x-y\rangle \ge \kappa^{-1}\|f'(x)-f'(y)\|^2$. Setting $g(u) = f(u) - \langle f'(y), u-y\rangle$, we get a continuously differentiable convex function on E such that $\|g'(x)-g'(y)\|_* \le \kappa\|x-y\|$ and $g'(y) = 0$. Due to these relations, $g(y+h) \le g(y) + \frac{\kappa}{2}\|h\|^2$ for all h. Now let $e \in E$ be such that $\langle g'(x), e\rangle = \|g'(x)\|_*$ and $\|e\| = 1$. Due to $\|g'(u)-g'(v)\|_* \le \kappa\|u-v\|$, we have $g(x - \frac{\|g'(x)\|_*}{\kappa}e) \le g(x) - \langle g'(x), \frac{\|g'(x)\|_*}{\kappa}e\rangle + \frac{\kappa}{2}\|\frac{\|g'(x)\|}{\kappa}e\|^2 = g(x) - \frac{\|g'(x)\|_*^2}{\kappa} + \frac{\|g'(x)\|_*^2}{2\kappa} = g(x) - \frac{\|g'(x)\|_*^2}{2\kappa}$. On the other hand, g attains its global minimum at y, so that $g(x) - \frac{\|g'(x)\|_*^2}{2\kappa} \ge g(x - \frac{\|g'(x)\|_*}{\kappa}e) \ge g(y)$. We now have

$$g(y) + \frac{\kappa}{2}\|h\|^2 \ge g(y+h) \ge g(x) + \langle g'(x), y+h-x\rangle \ge g(y) + \frac{\|g'(x)\|_*^2}{2\kappa} + \langle g'(x), y+h-x\rangle,$$

whence

$$\langle g'(x), x-y\rangle \ge \frac{\|g'(x)\|_*^2}{2\kappa} + \langle g'(x), h\rangle - \frac{\kappa}{2}\|h\|^2.$$

This inequality is valid for all h; setting $h = \frac{\|g'(x)\|_*}{\kappa}e$, the right hand side becomes $\frac{\|g'(x)\|_*^2}{\kappa}$. Thus, $\langle f'(x)-f'(y), x-y\rangle = \langle g'(x), x-y\rangle \ge \frac{\|g'(x)\|_*^2}{\kappa} = \frac{\|f'(x)-f'(y)\|_*^2}{\kappa}$, Q.E.D.

(iv) ⇒(vi): Let (iv) take place, let $\xi, \eta \in E^*$ and $x \in \partial f_*(\xi)$. Setting $\xi_t = \xi + t\eta$, $\phi(t) = f_*(\xi_t)$, $0 \le t \le 1$, we get an absolutely continuous function on $[0,1]$ with the derivative which is almost everywhere given by $\phi'(t) = \langle \eta, x_t\rangle$, with $x_t \in \partial f_*(\xi_t)$. We have

$$\begin{array}{l} f_*(\xi+\eta) = \phi(1) = \phi(0) + \int_0^1 \phi'(t)dt = \phi(0) + \int_0^1 \langle \eta, x_t\rangle dt = \phi(0) + \int_0^1 [\langle \eta, x\rangle + \langle \eta, x_t - x\rangle]dt \\ = \phi(0) + \langle \eta, x\rangle + \int_0^1 t^{-1}\langle (\xi+t\eta)-\xi, x_t - x\rangle dt \ge \phi(0) + \langle \eta, x\rangle + \int_0^1 t^{-1}\kappa^{-1}\|[\xi+t\eta]-\xi\|_*^2 dt \\ = \phi(0) + \langle \eta, x\rangle + \frac{1}{2\kappa}\|\eta\|_*^2 = f_*(\xi) + \langle \eta, x\rangle + \frac{1}{2\kappa}\|\eta\|_*^2 \end{array}$$

where the inequality is given by (iv). We end up with the inequality required in (vi), Q.E.D.

(vi)⇒(i): Let (vi) be the case, let $x \in E$ and $\xi \in \partial f(x)$, so that $x \in \partial f_*(\xi)$. We have

$$\begin{array}{l} f(x+y) = \max_{\eta \in E^*} [\langle \xi+\eta, x+y\rangle - f_*(\xi+\eta)] \le \max_{\eta\in E^*} \left[\langle \xi+\eta, x+y\rangle - f_*(\xi) - \langle \eta, x\rangle - \frac{1}{2\kappa}\|\eta\|_*^2\right] \\ = \max_{\eta\in E^*} \left[\langle \xi, x+y\rangle + \langle \eta, y\rangle - f_*(\xi) - \frac{1}{2\kappa}\|\eta\|_*^2\right] \\ = \underbrace{\langle \xi, x\rangle - f_*(\xi)}_{f(x)} + \langle \xi, y\rangle + \max_\eta \left[\langle \eta, y\rangle - \frac{1}{2\kappa}\|\eta\|_*^2\right] = f(x) + \langle \xi, y\rangle + \frac{\kappa}{2}\|y\|^2. \end{array}$$

This relation along with the relation $f(x+y) \ge f(x) + \langle \xi, y\rangle$ implies that ξ is the Frechet derivative of f at x, whence f is convex and differentiable, and thus – continuously differentiable function on E which satisfies the inequality

$$f(x+y) \le f(x) + \langle f'(x), y\rangle + \frac{\kappa}{2}\|y\|^2,$$

Q.E.D.

We have proved that (i)⇔(ii)⇔(iii)⇔(iv)⇔(v) and (iv)⇒(vi)⇒(i), meaning that all 6 properties in question are equivalent to each other. ■

Proposition 2.3 (i) *Let $p \in [2,\infty]$, and let $(E_i, \|\cdot\|_i)$ be finite-dimensional κ-smooth spaces, $i = 1,...,m > 2$. The space $E = E_1 \times ... \times E_m$ equipped with the norm*

$$\|(x^1,...,x^m)\| = \left(\sum_{i=1}^n \|x^i\|_i^p\right)^{1/p}$$

(the right hand side is $\max_i \|x^i\|_i$ when $p=\infty$) is κ^+-regular with

$$\kappa^+ = \min_{2\le\rho\le p}[\kappa+\rho-1]m^{\frac{2}{\rho}-\frac{2}{p}} \le \min[\kappa+p-1, [\kappa+2\ln(m)-1]\exp\{1\}]. \quad (19)$$

(ii) *Let $\|\cdot\|_i$ be κ-smooth norms on E. Then the norm*

$$\|x\| = \sum_{i=1}^m \|x\|_i$$

is $m\kappa$-regular on E.

(iii) *Let $(E,\|\cdot\|)$ be κ-regular, $\alpha>0$, and let L be a linear subspace in E. Then $(L, \alpha\|\cdot\|)$ is κ-regular.*

Proof. (i): To prove (i), let $p_i(x^i) = \|x^i\|_i^2$.

A. Let $\rho \in [2,\infty)$ be such that $\rho \le p$, and let $r = \rho/2$. Our local goal is to prove

Lemma 2 *The norm $\|\cdot\|$ on $E = E_1 \times ... \times E_m$ defined as $\|(x^1,...,x^m)\| = \|(\|x^1\|_1,...,\|x^m\|_m)\|_\rho$ is κ_+-smooth, with*

$$\kappa_+ = \kappa + \rho - 2 \quad (20)$$

Proof. We have

$$p(x^1,...,x^m) \equiv \|(\|x^1\|_1,...,\|x^m\|_m)\|_\rho^2 = \|(p_1(x^1),...,p_m(x^m))\|_r.$$

From this observation it immediately follows that $p(\cdot)$ is continuously differentiable. Indeed, $\rho \ge 2$, whence $r \ge 1$, so that the function $\|y\|_r$ is continuously differentiable everywhere on $\mathbf{R}^m_+$ except for the origin; the functions $p_i(x^i)$ are continuously differentiable by assumption. Consequently, $p(x)$ is continuously differentiable everywhere on $E = E_1 \times ... \times E_m$, except, perhaps, the origin; the fact that p' is continuous at the origin is evident.

Invoking Proposition 2.2, in order to prove Lemma 2 it suffices to verify that

$$\|p'(x) - p'(y)\|_* \le 2\kappa_+\|x-y\| \quad (21)$$

for all x,y. Since p' is continuous, it suffices to prove this relation for a dense in $E\times E$ set of pairs x,y, for example, those for which all blocks $x^i \in E_i$ in x are nonzero. With such x, the segment $[x,y]$ contains finitely many points u such that at least one of the blocks u^i is zero; these points split $[x,y]$ into finitely many consecutive segments, and it suffices to prove that $\|p'(x')-p'(y')\|_* \le 2\kappa_+\|x'-y'\|$ when x',y' are endpoints of such a segment. Since p' is continuous, to prove the latter statement is the same as to prove similar statement for the case when x',y' are interior points of the segment. The bottom line is as follows: in order to prove (21) for all pairs x,y, it suffices to prove the same statement for those pairs x,y for which every segment $[x^i,y^i]$ does not pass through the origin of the corresponding E_i.

Let x,y be such that $[x^i,y^i]$ does not pass through the origin of E_i, $i=1,...,m$. Same as in the item "(i)⇒(iii)" of the proof of Proposition 2.2, for every i there exists a sequence of C^∞ convex functions $\{p_i^t(\cdot) > 0\}_{t=1}^\infty$ on E_i converging to $p_i(\cdot)$ along with first order derivatives uniformly on compact sets and such that

$$|D^2p_i^t(u^i)[h^i,h^i]| \le 2\kappa\|h^i\|_i^2 \;\forall (u^i,h^i \in E_i). \quad (22)$$

Functions $p^t(u) = \|(p_1^t(u^1),...,p_m^t(u^m))\|_r$ clearly are convex, C^∞ (recall that $p_i^t(\cdot)>0$) and converge to $p(\cdot)$, along with their first order derivatives, uniformly on compact sets. It follows that

$$\langle p'(y)-p'(x), h\rangle = \lim_{t\to\infty}\int_0^1 D^2p^t(x+t(y-x))[y-x,h]dt. \quad (23)$$

Setting $F(y_1,...,y_m)=y_1^r+...+y_m^r$, $y\geq 0$, we have $p^t(u)=F^{\frac{1}{r}}(p_1^t(u^1),...,p_m^t(u^m))$. Now let $u\in[x,y]$, and let $v\in E$. We have

$$\begin{array}{rcl}
Dp^t(u)[v] &=& r^{-1}F^{\frac{1}{r}-1}(p_1^t(u^1),...,p_m^t(u^m))\left(\sum_i r(p_i^t(u^i))^{r-1}Dp_i^t(u^i)[v^i]\right)\\
\Rightarrow D^2p^t(u)[v,v] &=& \frac{1}{r}\underbrace{\left(\frac{1}{r}-1\right)}_{\leq 0}F^{\frac{1}{r}-2}(p_1^t(u^1),...,p_m^t(u^m))\left(\sum_i r(p_i^t(u^i))^{r-1}Dp_i^t(u^i)[v^i]\right)^2\\
\multicolumn{3}{l}{+F^{\frac{1}{r}-1}(p_1^t(u^1),...,p_m^t(u^m))\sum_i\left[(r-1)(p_i^t(u^i))^{r-2}(Dp_i^t(u^i)[v_i])^2+(p_i^t(u^i))^{r-1}D^2p_i^t(u^i)[v^i,v^i]\right]}\\
\multicolumn{3}{l}{\leq F^{\frac{1}{r}-1}(p_1^t(u^1),...,p_m^t(u^m))\sum_i\left[(r-1)(p_i^t(u^i))^{r-2}(Dp_i^t(u^i)[v_i])^2+2\kappa(p_i^t(u^i))^{r-1}p_i(v^i)\right]}
\end{array}$$

whence

$$\begin{array}{l}
0\leq D^2p^t(u)[v,v]\\
\leq F^{\frac{1}{r}-1}(p_1^t(u^1),...,p_m^t(u^m))\sum_i\left[(r-1)(p_i^t(u^i))^{r-2}(Dp_i^t(u^i)[v_i])^2+2\kappa(p_i^t(u^i))^{r-1}p_i(v^i)\right].
\end{array}\tag{24}$$

Taking into account that $p_i(\cdot)$ are bounded away from zero on $[x,y]$ and that $p_i^t(\cdot)$ converge, along with first order derivatives, to $p_i(\cdot)$ uniformly on compact sets as $t\to\infty$, the right hand side in bound (24) converges, as $t\to\infty$, uniformly in $u\in[x,y]$ and v, $\|v\|\leq 1$, to

$$\Psi(u,v)=\left(\sum_i\|u^i\|_i^\rho\right)^{\frac{2}{\rho}-1}\sum_i\left[(r-1)\|u^i\|_i^{\rho-4}(Dp_i(u^i)[v_i])^2+2\kappa\|u^i\|_i^{\rho-2}\|v^i\|_i^2\right].$$

By evident reasons, $|Dp_i(u^i)[v_i]|\leq 2\|u^i\|\|v^i\|$, whence

$$\begin{array}{rcl}
\Psi(u,v) &\leq& \left(\sum_i\|u^i\|_i^\rho\right)^{\frac{2}{\rho}-1}\sum_i\left[4(r-1)\|u^i\|_i^{\rho-2}\|v_i\|_i^2+2\kappa\|u^i\|_i^{\rho-2}\|v^i\|_i^2\right]\\
&=& \underbrace{[2\rho+2\kappa-4]}_{2\kappa_+}\left(\sum_i\|u^i\|_i^\rho\right)^{\frac{2}{\rho}-1}\sum_i\|u^i\|_i^{\rho-2}\|v^i\|_i^2
\end{array}\tag{25}$$

When $\rho>2$, we have

$$\begin{array}{rcl}
\sum_i\|u^i\|_i^{\rho-2}\|v^i\|_i^2 &\leq& \left(\sum_i(\|u^i\|_i^{\rho-2})^{\frac{\rho}{\rho-2}}\right)^{\frac{\rho-2}{\rho}}\left(\sum_i(\|v^i\|_i^2)^{\frac{\rho}{2}}\right)^{\frac{2}{\rho}}\\
&=& \left(\sum_i\|u^i\|_i^\rho\right)^{\frac{\rho-2}{\rho}}\left(\sum_i\|v^i\|_i^\rho\right)^{\frac{2}{\rho}},
\end{array}$$

and (25) implies that $\Psi(u,v)\leq 2\kappa_+\|v\|^2$. This inequality clearly is valid for $\rho=2$ as well. Recalling the origin of $\Psi(\cdot,\cdot)$, we conclude that for every $\epsilon>0$ there exists t_ϵ such that

$$t\geq t_\epsilon, u\in[x,y], \|v\|\leq 1\Rightarrow 0\leq D^2p^t(u)[v,v]\leq 2\kappa_+\|v\|^2+\epsilon.$$

The resulting inequality via the same reasoning as in the proof of item "(i)⇒(iii)" of Proposition 2.2 implies that

$$t\geq t_\epsilon, u\in[x,y]\Rightarrow|D^2p^t(u)[v,w]|\leq(2\kappa_++\epsilon)\|v\|\|w\|\ \forall v,w.$$

In view of this bound and (23), we conclude that

$$\langle p'(y)-p'(x),h\rangle\leq(2\kappa_++\epsilon)\|y-x\|\|h\|$$

for all h, whence $\|p'(y)-p'(x)\|_*\leq(2\kappa_++\epsilon)\|y-x\|$. Since $\epsilon>0$ is arbitrary, we arrive at (21).

B. When $\rho \le p$, we have

$$\|(\|x^1\|_1,...,\|x^m\|_m)\|_p^2 \le \|(\|x^1\|_1,...,\|x^m\|_m)\|_\rho^2 \le m^{\frac{2}{\rho}-\frac{2}{p}}\|(\|x^1\|_1,...,\|x^m\|_m)\|_p^2,$$

which combines with Lemma 2 to imply that the norm in (i) is κ-regular with $\kappa = [\rho+\kappa-2]m^{\frac{2}{\rho}-\frac{2}{p}}$, for every $\rho \in [2,p]$, and (i) follows.

(ii): To prove (ii), consider the norm $|(x^1,...,x^m)| = m^{1/2}\sqrt{\|x^1\|_1^2+...+\|x^m\|_m^2}$ on $E\times E\times ...\times E$. As it is immediately seen, this norm is κ-smooth. If, further, $\|(x^1,...,x^m)\|_\dagger = \sum\limits_i \|x^i\|_i$, then

$$\|x\|_\dagger^2 \le |x|^2 \le m\|x\|_\dagger^2 \quad \forall x \in E\times ...\times E,$$

whence $\|\cdot\|_\dagger$ is $m\kappa$-regular. The norm in (ii) is nothing but the restriction of $\|\cdot\|_\dagger$ on the image of E under the embedding $x \mapsto (x,...,x)$ of E into $E\times ...\times E$, and it remains to use (iii).

(iii): Evident. ∎

To proceed, we need the following fact:

Lemma 3 *Let $(E,\|\cdot\|)$ be a finite-dimensional κ-regular space. Then there exists κ-smooth norm $\|\cdot\|_+$ on E such that*

$$\forall (x\in E): \|x\|^2 \le \|x\|_+^2 \le 2\|x\|^2. \tag{26}$$

Proof. By definition, there exists $\kappa_+ \in [1,\kappa]$ and a norm $\pi(\cdot)$ on E which is κ_+-smooth and such that

$$\forall (x\in E): \|x\|^2 \le \pi^2(x) \le \mu\|x\|^2,\ \mu = \kappa/\kappa_+,$$

or, which is the same,

$$\forall \xi \in E^*: \pi_*^2(\xi) \ge \|\xi\|_*^2 \ge \frac{1}{\mu}\pi_*^2(\xi), \tag{27}$$

where E^* is the space dual to E and π_*, $\|\cdot\|_*$ are the norms on E^* conjugate to π, $\|\cdot\|$, respectively.

In the case of $\mu \le 2$, let us take $\|\cdot\|_+ \equiv \pi(\cdot)$, thus getting a κ_+-smooth (and thus – κ-smooth as well) norm on E satisfying (26). Now let $\mu > 2$, so that $\gamma = 1/(\mu-1) \in (0,1)$. Let us set $q_*(\xi) = \sqrt{\gamma\pi_*^2(\xi)+(1-\gamma)\|\xi\|_*^2}$, so that $q_*(\cdot)$ is a norm on E^*. We have

$$\forall \xi \in E^*: q_*^2(\xi) \ge \|\xi\|_*^2 \ge \frac{1}{\gamma\mu+1-\gamma}q_*^2(\xi) = \frac{1}{2}q_*^2(\xi). \tag{28}$$

Further, by Proposition 2.2 we have

$$\forall (\xi,\eta\in E^*, x\in\partial\pi_*^2(\xi)): \pi_*^2(\xi+\eta) \ge \pi_*^2(\xi)+\langle\eta,x\rangle+\frac{1}{\kappa_+}\pi_*^2(\eta),$$

whence, due to $\|\xi+\eta\|_*^2 \ge \|\xi\|_*^2+\langle\eta,y\rangle$ for all ξ,η and every y from the subdifferential $D(\xi)$ of $\|\cdot\|_*^2$ at the point ξ,

$$\forall(\xi,\eta\in E^*, x\in\partial\pi_*^2(\xi), y\in D(\xi)): q_*^2(\xi+\eta) \ge q_*^2(\xi)+\langle\eta,x+y\rangle+\frac{\gamma}{\kappa_+}\pi_*^2(\eta) \ge q_*^2(\xi)+\langle\eta,x+y\rangle+\frac{\gamma}{\kappa_+}q_*^2(\eta)$$

(note that $\pi_*(\cdot) \ge q_*(\cdot)$ by (27)). Since $\partial\pi_*^2(\xi)+D(\xi) = \partial q_*^2(\xi)$ and $\frac{\gamma}{\kappa_+} = \frac{1}{(\mu-1)\kappa_+} \ge \frac{1}{\kappa}$, we get

$$\forall(\xi,\eta\in E^*, z\in\partial q_*^2(\xi)): q_*^2(\xi+\eta) \ge q_*^2(\xi)+\langle\eta,z\rangle+\frac{1}{\kappa}q_*^2(\eta).$$

By the same Proposition 2.2, it follows that the norm $\|\cdot\|_+ \equiv q(\cdot)$ on E such that $q_*(\cdot)$ is the conjugate of $q(\cdot)$ is κ-smooth. At the same time, (28) implies (26). ∎

Lemma 3 allows to prove the following modification of Proposition 2.3.(i,ii):

Proposition 2.4 (i) *Let* $p \in [2,\infty]$, *and let* $(E_i, \|\cdot\|_i)$ *be finite-dimensional* κ*-regular spaces,* $i = 1,...,m > 2$. *The space* $E = E_1 \times ... \times E_m$ *equipped with the norm*

$$\|(x^1,...,x^m)\| = \left(\sum_{i=1}^n \|x^i\|_i^p\right)^{1/p}$$

(the right hand side is $\max_i \|x^i\|_i$ *when* $p = \infty$*) is* κ^{++}*-regular with*

$$\kappa^{++} = 2 \min_{2\leq\rho\leq p}[\kappa+\rho-1]m^{\frac{2}{\rho}-\frac{2}{p}} \leq 2\min[\kappa+p-1,[\kappa+2\ln(m)-1]\exp\{1\}]. \tag{29}$$

(ii) *Let* $\|\cdot\|_i$ *be* κ*-regular norms on a finite-dimensional space* E. *Then the norm*

$$\|x\| = \sum_{i=1}^m \|x\|_i$$

is $2m\kappa$*-regular on* E.

Proof is readily given by Lemma 3 combined with the corresponding items of Proposition 2.3. E.g., to prove (i), note that by Lemma 3 we can find κ-smooth norms $q_i(\cdot)$ on E_i such that $q_i^2(x^i) \leq \|x^i\|_i^2 \leq 2q_i^2(x^i)$ for every i and all $x^i \in E_i$. Applying Proposition 2.3.(i) to the spaces $(E_i, q_i(\cdot))$, we get that the norm $q(x^1,...,x^m) = \left(\sum\limits_{i=1}^n q_i^p(x^i)\right)^{1/p}$ on $E_1 \times ... \times E_m$ is κ^+-regular with κ^+ given by (19). Taking into account the evident relation

$$q^2(x^1,...,x^m) \leq \|(x^1,...,x^m)\|^2 \leq 2q^2(x^1,...,x^m)$$

and recalling the definition of regularity, we conclude that $\|\cdot\|$ is κ^{++}-regular, as required. ■

3 Sums of random vectors in regular spaces

In this Section, we consider the situation as follows. We are given a finite-dimensional κ-regular space $(E, \|\cdot\|)$, a Polish space Ω with Borel probability measure μ and a sequence $\mathcal{F}_0 = \{\emptyset, \Omega\} \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset ...$ of σ-sub-algebras of the Borel σ-algebra of Ω. We denote by $\mathbf{E}_t$, $t = 1,2,...$ the conditional expectation w.r.t. $\mathcal{F}_t$, and by $\mathbf{E} \equiv \mathbf{E}_0$ the expectation w.r.t. μ. The question we are interested in as follows:

> *Given a martingale-difference* $\{\xi_t\}_{t=1}^\infty$ *with values in* E, *so that* ξ_t *is a* $\mathcal{F}_t$*-measurable random vector in* E *such that*
>
> $$\mathbf{E}_{t-1}\{\xi_t\} \equiv 0,\ t = 1,2,...,$$
>
> *what can we say about "typical norms" of the associated sums*
>
> $$S_n = \sum_{t=1}^n \xi_t.$$

In the sequel, we denote by $\|\cdot\|_+$ a κ_+-smooth norm on E which is (κ/κ_+)-compatible with $\|\cdot\|$:

$$\|x\|^2 \leq \|x\|_+^2 \leq (\kappa/\kappa_+)\|x\|^2 \quad \forall x \in E, \tag{30}$$

and set

$$p(x) = \|x\|_+^2.$$

3.1 Bounds on second moments of $\|S_n\|$

Our first observation is nearly tautological:

Proposition 3.1 *Assume that E-valued martingale-difference $\xi = \{\xi_t\}_{t=1}^{\infty}$ is square summable:*

$$\mathbf{E}\left\{\|\xi_t\|^2\right\} \le \sigma_t^2 < \infty.$$

Then

$$\mathbf{E}\left\{\|S_n\|^2\right\} \le \kappa \sum_{t=1}^{n} \sigma_t^2. \tag{31}$$

Proof. Since $\|\cdot\|_+$ is κ_+-smooth, we have

$$p(S_{t+1}) \le p(S_t) + Dp(S_t)[\xi_{t+1}] + \kappa_+ p(\xi_{t+1})$$

whence, taking expectations and making use of the fact that ξ is martingale-difference,

$$\mathbf{E}\left\{p(S_{t+1})\right\} \le \mathbf{E}\left\{p(S_t)\right\} + \kappa_+ \mathbf{E}\left\{p(\xi_{t+1})\right\} \le \mathbf{E}\left\{p(S_t)\right\} + \kappa \mathbf{E}\left\{\|\xi_{t+1}\|^2\right\}$$

(we have used the right inequality in (30)). From this recurrent inequality we get

$$\mathbf{E}\left\{\|S_n\|_+^2\right\} \le \kappa \sum_{t=1}^{n} \mathbf{E}\left\{\|\xi_t\|^2\right\} \le \kappa \sum_{t=1}^{n} \sigma_t^2.$$

The left hand side in this inequality, by (30), is $\ge \mathbf{E}\left\{\|S_n\|^2\right\}$, and (31) follows. ■

3.2 Large deviations for $\|S_n\|$, I

Theorem 3.1 *Let E-valued martingale-difference $\xi = \{\xi_t\}_{t=1}^{\infty}$ and reals $\sigma_t > 0$ be such that*

$$\mathbf{E}_{t-1}\left\{\exp\{\|\xi_t\|^2 \sigma_t^{-2}\}\right\} \le \exp\{1\},\ t = 1, 2, ... \tag{32}$$

Then for all $n \ge 1$ and $\Omega \ge 0$ one has

$$\mathrm{Prob}\left\{\|\sum_{t=1}^{n} \xi_t\| > 15\Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^{n} \sigma_t^2}\right\} \le 3\exp\{-\Omega^2\}. \tag{33}$$

Proof. 1^0. Let us set $\bar\sigma_t = \kappa^{1/2}\sigma_t\kappa_+^{-1/2}$ and $\eta_t = \xi_t\bar\sigma_t^{-1}$, so that $\|\eta_t\|_+^2 \le (\kappa\kappa_+^{-1})\|\eta_t\|^2 = (\kappa\kappa_+^{-1})\|\xi_t\|^2(\kappa^{-1}\kappa_+\sigma_t^{-2}) = \|\xi_t\|^2\sigma_t^{-2}$, whence

$$\xi_t = \bar\sigma_t\eta_t, \quad \mathbf{E}_{t-1}\left\{\|\eta_t\|_+^2\right\} \le \exp\{1\}, \quad \mathbf{E}_{t-1}\left\{\eta_t\right\} = 0, \quad S_n = \sum_{t=1}^{n} \bar\sigma_t\eta_t. \tag{34}$$

Observe that by the moment inequality

$$0 \le \tau \le 1 \Rightarrow \mathbf{E}_{t-1}\left\{\exp\{\tau\|\eta_t\|_+^2\}\right\} \le \exp\{\tau\}. \tag{35}$$

Further, since $\exp\{x\} \ge \frac{x^\ell}{\ell!}$ for all $x \ge 0$, it follows from (35) that

$$\mathbf{E}_{t-1}\left\{\|\eta_t\|_+^{2\ell}\right\} \le e\ell!,\ \ell = 0, 1, ... \tag{36}$$

2^0. Let

$$\omega_n = \left(\sum_{t=1}^{n} \bar\sigma_t^2\right)^{1/2}. \tag{37}$$

Let us prove by induction in $n \geq 1$ that if

$$\epsilon \leq \frac{1}{13 + 2\sqrt{\kappa_+}}, \tag{38}$$

then

$$(P_n): \quad 0 \leq \tau \leq \frac{1}{\omega_n^2} \Rightarrow \mathbf{E}\left\{\exp\{\epsilon^2 \tau p(S_n)\}\right\} \leq \exp\{\omega_n^2 \tau\},$$

Base $n = 1$: evident in view of (35).
Step $n \Rightarrow n+1$: Assume that (P_n) holds true. Denoting by $\|\cdot\|_\dagger$ is the norm on E^* conjugate to $\|\cdot\|_+$ and by $\langle \phi, x\rangle$ the value of a linear functional $\phi \in E^*$ at a vector $x \in E$, we have:

$$\begin{array}{rl} & \mathbf{E}\left\{\exp\{\epsilon^2\tau p(S_{n+1})\}\right\} \\ \leq & \mathbf{E}\left\{\exp\{\epsilon^2\tau\left[p(S_n) + \bar\sigma_{n+1}\langle p'(S_n), \eta_{n+1}\rangle + \kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right]\}\right\} \\ \leq & \mathbf{E}\left\{\exp\{\epsilon^2\tau p(S_n)\}\exp\{\epsilon^2\tau\left[\bar\sigma_{n+1}\langle p'(S_n), \eta_{n+1}\rangle + \kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right]\}\right\} \\ = & \mathbf{E}\left\{\exp\{\epsilon^2\tau p(S_n)\}\mathbf{E}_n\left\{\exp\{\epsilon^2\tau\left[\bar\sigma_{n+1}\langle p'(S_n), \eta_{n+1}\rangle + \kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right]\}\right\}\right\} \end{array} \tag{39}$$

Now let $0 \leq \tau \leq \omega_{n+1}^{-2}$. We have

$$\begin{array}{rl} & \mathbf{E}_n\left\{\exp\{\epsilon^2\tau\left[\bar\sigma_{n+1}\langle p'(S_n), \eta_{n+1}\rangle + \kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right]\}\right\} \\ = & 1 + \epsilon^2\tau\mathbf{E}_n\left\{\bar\sigma_{n+1}\langle p'(S_n), \eta_{n+1}\rangle + \kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}\mathbf{E}_n\left\{\left[\epsilon^2\tau\left[\bar\sigma_{n+1}\langle p'(S_n), \eta_{n+1}\rangle + \kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right]\right]^\ell\right\} \\ \leq & 1 + \epsilon^2\tau\mathbf{E}_n\left\{\kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}\mathbf{E}_n\left\{\left[\epsilon^2\tau\left[\bar\sigma_{n+1}\langle p'(S_n), \eta_{n+1}\rangle + \kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right]\right]^\ell\right\} \\ \leq & 1 + \epsilon^2\tau\mathbf{E}_n\left\{\kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}\mathbf{E}_n\left\{(2\epsilon^2\tau)^\ell\left[\|p'(S_n)\|_\dagger^\ell\bar\sigma_{n+1}^\ell\|\eta_{n+1}\|_+^\ell + \left(\kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right)^\ell\right]\right\} \\ = & 1 + \epsilon^2\tau\mathbf{E}_n\left\{\kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}\mathbf{E}_n\left\{\left(2\epsilon^2\tau\kappa_+\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\right)^\ell\right\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}(2\epsilon^2\tau)^\ell\mathbf{E}_n\left\{\|p'(S_n)\|_\dagger^\ell\bar\sigma_{n+1}^\ell\|\eta_{n+1}\|_+^\ell\right\} \\ \leq & \mathbf{E}_n\left\{\exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\|\eta_{n+1}\|_+^2\}\right\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}(2\epsilon^2\tau)^\ell\mathbf{E}_n\left\{(4p(S_n))^{\frac{1}{2}\ell}\bar\sigma_{n+1}^\ell\|\eta_{n+1}\|_+^\ell\right\} \\ & \qquad\qquad [\text{since } \|p'(u)\|_\dagger \leq \sqrt{4p(u)}] \\ \leq & \exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}(2\epsilon^2\tau)^\ell\mathbf{E}_n\left\{(4p(S_n))^{\frac{1}{2}\ell}\bar\sigma_{n+1}^\ell\|\eta_{n+1}\|_+^\ell\right\} \\ & \qquad\qquad [\text{we have used (35) combined with } 2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2 \leq 1] \\ \leq & \exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \sum\limits_{\ell=2}^\infty \frac{1}{\ell!}(2\epsilon^2\tau\bar\sigma_{n+1})^\ell\left((4p(S_n))^{\frac{1}{2}}\right)^\ell(e\ell!)^{\frac{1}{2}} \\ & \qquad\qquad [\text{we have used (36)}] \\ \leq & \exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \exp\{\frac{1}{2}\}\sum\limits_{\ell=2}^\infty\left(\frac{\left[\overbrace{\left[7\epsilon^2\tau\bar\sigma_{n+1}\right]^2}^{\theta_{n+1}} p(S_n)\right]^\ell}{\ell!}\right)^{\frac{1}{2}} 3^{-\frac{\ell}{2}} \end{array} \tag{40}$$

Further,

$$\begin{array}{l} \exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \exp\{\frac{1}{2}\}\sum\limits_{\ell=2}^\infty\left(\frac{(\theta_{n+1}p(S_n))^\ell}{\ell!}\right)^{\frac{1}{2}} 3^{-\frac{\ell}{2}} \\ \leq \exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \exp\{\frac{1}{2}\}\left(\sum\limits_{\ell=2}^\infty \frac{(\theta_{n+1}p(S_n))^\ell}{\ell!}\right)^{\frac{1}{2}}\left(\sum\limits_{\ell=2}^\infty 3^{-\ell}\right)^{\frac{1}{2}} \\ \leq \exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \left(\sum\limits_{\ell=2}^\infty \frac{(\theta_{n+1}p(S_n))^\ell}{\ell!}\right)^{\frac{1}{2}} \\ \leq \exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \left(\exp\{\theta_{n+1}p(S_n)\} - 1 - \theta_{n+1}p(S_n)\right)^{\frac{1}{2}} \end{array} \tag{41}$$

3^0. We need the following

Lemma 4 *Let $0 \leq \tau$ and $\theta = (7\epsilon^2\tau\bar\sigma_{n+1})^2$. Then*

$$\exp\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2\} + \left(\exp\{\theta p(S_n)\} - 1 - \theta p(S_n)\right)^{\frac{1}{2}} \leq \exp\left\{2\kappa_+\epsilon^2\tau\bar\sigma_{n+1}^2 + 75\epsilon^4\tau^2\bar\sigma_{n+1}^2 p(S_n)\right\} \tag{42}$$

Proof. Observe, first, that for $x \geq 0$ one has

$$\exp\{x\} - 1 - x \leq \frac{x^2}{2}\exp\{x\}, \tag{43}$$

and, second, that for $x, y \geq 0$ one has

$$\exp\{x\}y \leq \exp\{x+y\} - 1. \tag{44}$$

Therefore

$$\begin{array}{l}
\exp\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2\} + (\exp\{\theta p(S_n)\} - 1 - \theta p(S_n))^{\frac{1}{2}} \\
\leq \exp\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2\} + \left[\frac{(\theta p(S_n))^2}{2}\exp\{\theta p(S_n)\}\right]^{\frac{1}{2}} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [\text{by (43)}] \\
= \exp\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2\} + \underbrace{\left(\frac{\theta p(S_n)}{\sqrt{2}}\right)}_{y}\exp\Big\{\underbrace{\frac{1}{2}\theta p(S_n)}_{x}\Big\} \\
\leq \exp\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2\} + \exp\left\{\frac{1}{2}\theta p(S_n) + \frac{\theta p(S_n)}{\sqrt{2}}\right\} - 1 \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [\text{by (44)}] \\
\leq \exp\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2\} + \exp\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2\}\left[\exp\left\{\frac{3}{2}\theta p(S_n)\right\} - 1\right] \\
= \exp\left\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2 + \frac{3}{2}\theta p(S_n)\right\}
\end{array}$$

■

4^0. Combining (39), (40), (41) and (42), we arrive at the relation

$$\begin{array}{l}
0 \leq \tau \leq \omega_{n+1}^{-2} \Rightarrow \\
\mathbf{E}\left\{\exp\{\epsilon^2\tau p(S_{n+1})\}\right\} \leq \mathbf{E}\left\{\exp\left\{\left[\epsilon^2\tau + 75\epsilon^4\tau^2\bar{\sigma}_{n+1}^2\right]p(S_n) + 2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2\right\}\right\}.
\end{array} \tag{45}$$

Now let $0 \leq \tau \leq \omega_{n+1}^{-2}$ and let

$$\mu = \tau + 75\epsilon^2\tau^2\bar{\sigma}_{n+1}^2.$$

Then, setting $\rho = \frac{\bar{\sigma}_{n+1}^2}{\omega_n^2}$,

$$\begin{array}{rcl}
\mu & \leq & \frac{1}{\omega_{n+1}^2} + \frac{75\epsilon^2\bar{\sigma}_{n+1}^2}{\omega_{n+1}^4} = \frac{1}{\omega_n^2+\bar{\sigma}_{n+1}^2} + \frac{75\epsilon^2\bar{\sigma}_{n+1}^2}{(\omega_n^2+\bar{\sigma}_{n+1}^2)^2} = \frac{1}{\omega_n^2}\left[\frac{1}{1+\rho} + \frac{75\epsilon^2\rho}{(1+\rho)^2}\right] \\
& = & \frac{1}{\omega_n^2}\frac{1+\rho(1+75\epsilon^2)}{1+2\rho+\rho^2} < \frac{1}{\omega_n^2}
\end{array}$$

(since $75\epsilon^2 \leq 1$). Consequently, by (P_n) one has

$$\mathbf{E}\left\{\exp\{[\epsilon^2\tau + 75\epsilon^4\tau^2\bar{\sigma}_{n+1}^2]p(S_n)\}\right\} = \mathbf{E}\left\{\exp\{\epsilon^2\mu p(S_n)\}\right\} \leq \exp\{\mu\omega_n^2\},$$

and (45) implies that

$$\begin{array}{rcl}
0 \leq \tau \leq \frac{1}{\omega_{n+1}^2} \Rightarrow \mathbf{E}\left\{\exp\{\epsilon^2\tau p(S_{n+1})\}\right\} & \leq & \exp\left\{2\kappa_+\epsilon^2\tau\bar{\sigma}_{n+1}^2 + \mu\omega_n^2\right\} \\
& = & \exp\left\{\tau\left[2\kappa_+\epsilon^2\bar{\sigma}_{n+1}^2 + [1 + 75\epsilon^2\tau\bar{\sigma}_{n+1}^2]\omega_n^2\right]\right\} \\
& = & \exp\{\tau\omega_{n+1}^2\chi\},
\end{array} \tag{46}$$

where

$$\chi = \frac{2\kappa_+\epsilon^2\rho}{1+\rho} + \frac{1 + 75\epsilon^2(\tau\omega_{n+1}^2)\frac{\rho}{1+\rho}}{1+\rho}.$$

Taking into account that $2\kappa_+\epsilon^2 \leq 1/2$, $\tau\omega_{n+1}^2 \leq 1$ and $75\epsilon^2 \leq 1/2$, we get

$$\chi \leq \frac{\frac{1}{2}\rho}{1+\rho} + \frac{1+\frac{1}{2}\rho}{1+\rho} \leq 1,$$

and (46) implies (P_{n+1}).

5^0. Now we are ready to prove (33). Let us set $\epsilon = \frac{1}{15\sqrt{\kappa_+}}$, so that (38) holds true. Then for $\Omega > 0$ one has

$$\begin{array}{rcl}\mathrm{Prob}\left\{\|S_n\| > 15\Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \sigma_t^2}\right\} & \leq & \mathrm{Prob}\left\{\|S_n\|_+ > 15\Omega\sqrt{\kappa}\sqrt{\kappa_+/\kappa}\sqrt{\sum_{t=1}^n \bar\sigma_t^2}\right\} \\ & & [\text{since } \|\cdot\|_+ \geq \|\cdot\|] \\ & = & \mathrm{Prob}\left\{\|S_n\|_+ > \Omega\epsilon^{-1}\omega_n\right\} \\ & < & \mathbf{E}\left\{\exp\{\epsilon^2\omega_n^{-2}\|S_n\|_+^2\}\right\}\exp\{-\Omega^2\} \\ & \leq & \exp\{1-\Omega^2\}\end{array}$$

(we have used (P_n) with $\tau = \omega_n^{-2}$). ■

Refinements in the case of bounded ξ_t. In this case, constants in Theorem 3.1 can be improved.

Theorem 3.2 *Let $\xi = \{\xi_t\}_{t=1}^\infty$ be a E-valued martingale-difference and $\sigma_t > 0$ be reals such that $\|\xi_t\| \leq \sigma_t$. Then*

$$\mathrm{Prob}\left\{\|S_n\| > \Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \sigma_t^2}\right\} \leq \exp\{1 - \frac{\Omega^2}{4.4}\}. \tag{47}$$

The proof is completely similar to the one of Theorem 3.1.

Note that a result similar to the one of Theorem 3.2 can be easily derived from Talagrand Inequality. Here is this inequality (in slightly extended form presented in [2]):

Theorem 3.3 *Let $(E_t, \|\cdot\|_{E_t})$ be finite-dimensional normed spaces, $t = 1, ..., n$, F be the direct product of $E_1, ..., E_n$ equipped with the norm $\|(x^1, ..., x^n)\|_F = \sqrt{\sum_{t=1}^n \|x^t\|_{E_t}^2}$, μ_t be Borel probability distributions on the unit balls of E_t and μ be the product of these distributions. Given a closed convex set $A \subset F$, let $\mathrm{dist}(x, A) = \min_{y\in A}\|x - y\|_F$. Then*

$$\mathbf{E}_\mu\left\{\exp\{\frac{1}{4}\mathrm{dist}^2(x, A)\}\right\} \leq \frac{1}{\mu(A)}. \tag{48}$$

Talagrand Inequality immediately implies the following result:

Theorem 3.4 *Let $(E, \|\cdot\|)$ be κ-regular space, and let $\xi_1, ..., \xi_n$ be independent random vectors in E with zero means and σ_t be reals such that $\|\xi_t\| \leq \sigma_t$, $t = 1, ..., n$. Then*

$$\Omega \geq 2\sqrt{2\kappa} \Rightarrow \mathrm{Prob}\left\{\|S_n\| > \Omega\sqrt{\sum_{t=1}^n \sigma_t^2}\right\} \leq 2\exp\{-\frac{\Omega^2}{32}\}. \tag{49}$$

Proof. Let F be the direct product of n copies of E equipped with the norm $\|(x^1, ..., x^n)\|_F = \sqrt{\sum_{t=1}^n \|x^t\|^2}$, let $\zeta_t = (2\sigma_t)^{-1}\xi_t$, and let Q be the set of all $x = (x^1, ..., x^n) \in F$ such that $S(x) \equiv 2\sum_{t=1}^n \sigma_t x^t \in E$ satisfies $\|S(x)\| \leq 1$. Note that Q is a closed convex set in F, and that $\|S(x)\| \leq r$ if and only if $x \in rQ$.

Let $\Theta = \sqrt{\sum_{t=1}^n \sigma_t^2}$. Our first observation is that Q contains $\|\cdot\|_F$-ball of the radius $\rho = (2\Theta)^{-1}$ centered at the origin. Indeed, if $\|(x^1, ..., x^n)\|_F \leq \rho$, then $\|S(x)\| \leq \sum_{t=1}^n 2\sigma_t\|x^t\| \leq 2\Theta\sqrt{\sum_{t=1}^n \|x^t\|^2} \leq 2\Theta\rho = 1$.

Next observe that if $\zeta = (\zeta_1, ..., \zeta_n)$, then for every $\gamma > 0$ one has

$$\mathrm{Prob}\left\{\|S(\zeta)\| > \gamma\right\} \equiv \mathrm{Prob}\left\{\zeta \not\in \gamma Q\right\} \le \frac{\kappa\Theta^2}{\gamma^2}. \tag{50}$$

Indeed, we have $S(\zeta) = \sum\limits_{t=1}^n 2\sigma_t\zeta_t = \sum\limits_{t=1}^n \xi_t$; by Proposition 3.1, we have $\mathbf{E}\left\{\|S(\zeta)\|^2\right\} \le \kappa\Theta^2$, and (50) follows from the Tschebyshev inequality.

Let us fix $\gamma > \sqrt{\kappa}\Theta$ and set $A = \gamma Q$; note that A is closed convex set in F symmetric w.r.t. the origin and containing the centered at the origin $\|\cdot\|_F$-ball of radius $\gamma\rho$; besides this,

$$\mathrm{Prob}\left\{\zeta \in A\right\} \ge 1 - \frac{\kappa\Theta^2}{\gamma^2} > 0 \tag{51}$$

by (50). Observe that

$$s > 1,\, x \in F\backslash(sA) \Rightarrow \mathrm{dist}(x, A) > (s-1)\gamma\rho. \tag{52}$$

Indeed, for s, x from the premise of this implication, the set $B = x + (s-1)A$ does not intersect A; since A contains the $\|\cdot\|_F$-ball of radius $\gamma\rho$ centered at the origin, B contains $\|\cdot\|_F$-ball of the radius $(s-1)\gamma\rho$ centered at x. Since $B \cap A = \emptyset$, the conclusion in (52) follows.

Applying (48) to the distribution of ζ, we get

$$\mathbf{E}\left\{\exp\{\frac{1}{4}\mathrm{dist}^2(\zeta, A)\}\right\} \le \frac{1}{\mathrm{Prob}\{\zeta \in A\}} \le \frac{1}{1 - \kappa\Theta^2\gamma^{-2}}$$

(we have used (51)). In view of (52), this bound implies

$$s > 1 \Rightarrow \mathrm{Prob}\left\{\zeta \not\in sA = s\gamma Q\right\} \le \frac{1}{1 - \kappa\Theta^2\gamma^{-2}} \exp\{-\frac{\gamma^2\rho^2(s-1)^2}{4}\} \tag{53}$$

Since $\zeta \not\in \alpha Q$ if and only if $\|\sum\limits_{t=1}^n \xi_t\| > \alpha$, we arrive at

$$\forall(s > 1, \gamma > \sqrt{\kappa}\Theta) : \mathrm{Prob}\left\{\|\sum_{t=1}^n \xi_t\| > \gamma s\right\} \le \frac{1}{1 - \kappa\Theta^2\gamma^{-2}} \exp\{-\frac{\gamma^2(s-1)^2}{8\Theta^2}\}$$

(we have substituted the value of ρ). Given $\Omega \ge 2\sqrt{2\kappa}$ and setting $\gamma = \sqrt{2\kappa}\Theta$ $s = \Omega/\sqrt{2\kappa}$, we arrive at (49). ■

3.3 Large deviations for $\|S_n\|$, II

Theorem 3.5 *Let $\alpha \in (0, 2]$, and let E-valued martingale-difference $\xi = \{\xi_t\}_{t=1}^\infty$ and reals $\sigma_t > 0$ be such that*

$$\mathbf{E}_{t-1}\left\{\exp\{\|\xi_t\|^\alpha\sigma_t^{-\alpha}\right\} \le \exp\{1\},\ t = 1, 2, ... \tag{54}$$

Then for all $n \ge 1$ and $\Omega \ge 0$ one has

$$\mathrm{Prob}\left\{\|\sum_{t=1}^n \xi_t\| > \Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \sigma_t^2}\right\} \le C_\alpha \exp\{-C_\alpha^{-1}\Omega^\alpha\}, \tag{55}$$

where $C_\alpha \ge 2$ depends solely on $\alpha \in (0, 2]$ and is continuous in $\alpha > 0$. In particular, for appropriately chosen $c_\alpha > 0$ depending solely on $\alpha \in (0, 2]$ and continuous in α, one has for all $n \ge 1$:

$$\mathbf{E}\left\{\exp\{\frac{\|\sum\limits_{t=1}^n \xi_t\|^\alpha}{(c_\alpha\sqrt{\kappa}\sqrt{\sum_{t=1}^n \sigma_t^2})^\alpha}\}\right\} \le \exp\{1\}. \tag{56}$$

Proof. 1^0. Let $\rho \geq (2/\alpha)^{1/\alpha}$. Let us set

$$\eta_t = \chi_{\{\|\xi_t\|>\sigma_t\rho\}}\xi_t,\ \zeta_t = \eta_t - \underbrace{\mathbf{E}_{t-1}\left\{\eta_t\right\}}_{\delta_t},\ \omega_t = \xi_t - \zeta_t.$$

Observe that

$$\mathbf{E}_{t-1}\left\{\zeta_t\right\} = 0, \tag{57}$$

whence, due to $\mathbf{E}_{t-1}\left\{\xi_t\right\} = 0$, also

$$\mathbf{E}_{t-1}\left\{\omega_t\right\} = 0. \tag{58}$$

2^0. We have

$$\begin{array}{rcl} \|\delta_t\| & \leq & \mathbf{E}_{t-1}\left\{\chi_{\{\|\xi_t\|>\sigma_t\rho\}}\|\xi_t\|\right\} \\ & = & \mathbf{E}_{t-1}\left\{\exp\{\|\xi_t\|^\alpha\sigma_t^{-\alpha}\}\sigma_t\left[\exp\{-\|\xi_t\|^\alpha\sigma_t^{-\alpha}\}[\|\xi_t\|/\sigma_t]\chi_{\{\|\xi_t\|>\sigma_t\rho\}}\right]\right\} \\ & \leq & \sigma_t\left[\max\limits_{z\geq\rho}\left[z\exp\{-z^\alpha\}\right]\right]\mathbf{E}_{t-1}\left\{\exp\{\|\xi_t\|^\alpha\sigma_t^{-\alpha}\}\right\} \\ & \leq & \sigma_t\exp\{1\}\max\limits_{z\geq\rho}\left[z\exp\{-z^\alpha\}\right] \\ & = & \sigma_t\exp\{1\}\rho\exp\{-\rho^\alpha\} \\ & & \qquad\qquad\qquad [\text{due to } \rho\geq\alpha^{-1/\alpha}] \end{array} \tag{59}$$

Consequently,

$$\begin{array}{rcl} \|\omega_t\| & = & \|\xi_t - [\eta_t - \delta_t]\| \leq \|\xi_t - \eta_t\| + \|\delta_t\| \\ & \leq & \sigma_t\exp\{1\}\rho\exp\{-\rho^\alpha\} + \|\xi_t\|\chi_{\{\|\xi_t\|\leq\sigma_t\rho\}}. \end{array} \tag{60}$$

Setting

$$\widehat{\sigma}_t = 2\sigma_t\rho^{\frac{2-\alpha}{2}},$$

we have from (60):

$$\begin{array}{rl} & \mathbf{E}_{t-1}\left\{\exp\left\{\|\omega_t\|^2\widehat{\sigma}_t^{-2}\right\}\right\} \\ \leq & \mathbf{E}_{t-1}\left\{\exp\left\{0.25\left[\sigma_t\exp\{1\}\rho\exp\{-\rho^\alpha\} + \|\xi_t\|\chi_{\{\|\xi_t\|\leq\sigma_t\rho\}}\right]^2\sigma_t^{-2}\rho^{\alpha-2}\right\}\right\} \\ \leq & \mathbf{E}_{t-1}\left\{\exp\left\{0.5\left[\exp\{2\}\exp\{-2\rho^\alpha\}\rho^\alpha + [\|\xi_t\|^2\sigma_t^{-2}]\rho^{\alpha-2}\chi_{\{\|\xi_t\|\leq\sigma_t\rho\}}\right]\right\}\right\} \\ \leq & \mathbf{E}_{t-1}\left\{\exp\left\{0.5\left[\exp\{2\}\exp\{-2\rho^\alpha\}\rho^\alpha + [\|\xi_t\|^\alpha\sigma_t^{-\alpha}]\chi_{\{\|\xi_t\|\leq\sigma_t\rho\}}\right]\right\}\right\} \\ \leq & \mathbf{E}_{t-1}\left\{\max\left[\exp\left\{\exp\{2\}\exp\{-2\rho^\alpha\}\rho^\alpha\right\}, \exp\left\{[\|\xi_t\|^\alpha\sigma_t^{-\alpha}]\right\}\right]\right\} \\ \leq & \exp\left\{\exp\{2\}\exp\{-2\rho^\alpha\}\rho^\alpha\right\} + \exp\{1\} \\ \leq & \exp\{0.5\exp\{1\}\} + \exp\{1\} \leq \exp\{2\}. \end{array}$$

It follows that with

$$\widetilde{\sigma}_t = 2\sqrt{2}\rho^{\frac{2-\alpha}{2}}\sigma_t \tag{61}$$

one has

$$\mathbf{E}_{t-1}\left\{\exp\{\|\omega_t\|^2\widetilde{\sigma}_t^{-2}\}\right\} \leq \exp\{1\}. \tag{62}$$

3^0. We have $\|\mathbf{E}_{t-1}\left\{\eta_t\right\}\|^2 \leq \mathbf{E}_{t-1}\left\{\|\eta_t\|^2\right\}$, whence

$$\begin{array}{rcl} \mathbf{E}_{t-1}\left\{\|\zeta_t\|^2\right\} & \leq & \mathbf{E}_{t-1}\left\{4\|\eta_t\|^2\right\} = 4\mathbf{E}_{t-1}\left\{\|\xi_t\|^2\chi_{\{\|\xi_t\|>\sigma_t\rho\}}\right\} \\ & = & 4\sigma_t^2\mathbf{E}_{t-1}\left\{\exp\{\|\xi_t\|^\alpha\sigma_t^{-\alpha}\}\left[\|\xi_t\|^2\sigma_t^{-2}\exp\{-\|\xi_t\|^\alpha\sigma_t^{-\alpha}\}\chi_{\{\|\xi_t\|>\sigma_t\rho\}}\right]\right\} \\ & \leq & 4\sigma_t^2\mathbf{E}_{t-1}\left\{\exp\{\|\xi_t\|^\alpha\sigma_t^{-\alpha}\}\max\limits_{z\geq\rho}\left[z^2\exp\{-z^\alpha\}\right]\right\} \\ & = & 4\sigma_t^2\exp\{1\}\rho^2\exp\{-\rho^\alpha\} \\ & & \qquad\qquad\qquad [\text{since } \rho\geq(2/\alpha)^{1/\alpha}] \end{array}$$

Thus,

$$\mathbf{E}_{t-1}\left\{\|\zeta_t\|^2\sigma_t^{-2}\right\} \leq 4\exp\{1\}\rho^2\exp\{-\rho^\alpha\}.$$

Thus, ζ_t is $\mathcal{F}_t$-measurable random vector (by its origin) such that

$$\mathbf{E}_{t-1}\left\{\zeta_t\right\} = 0,\ \mathbf{E}_{t-1}\left\{\|\zeta_t\|^2\sigma_t^{-2}\right\} \leq 4\exp\{1\}\rho^2\exp\{-\rho^\alpha\}, \tag{63}$$

(see (57)). Besides this, ω_t is $\mathcal{F}_t$-measurable random vector (by its origin) such that

$$\mathbf{E}_{t-1}\left\{\omega_t\right\}=0,\ \mathbf{E}_{t-1}\left\{\exp\{\|\omega_t\|^2\widetilde{\sigma}_t^{-2}\}\right\}\le\exp\{1\}, \tag{64}$$

(see (58), (62)).

4^0. Applying Theorem 3.1 to random vectors $\omega_1,...,\omega_n$ and taking into account (64), we get

$$\mathrm{Prob}\left\{\|\sum_{t=1}^n\omega_t\|\ge\rho^{\alpha/2}\sqrt{\kappa}\sqrt{\sum_{t=1}^n\widetilde{\sigma}_t^2}\right\}\le C_1\exp\{-C_2\rho^\alpha\} \tag{65}$$

(from now on, C_i are appropriate positive absolute constants), whence, recalling (61),

$$\mathrm{Prob}\left\{\|\sum_{t=1}^n\omega_t\|\ge 2\sqrt{2}\rho\sqrt{\kappa}\sqrt{\sum_{t=1}^n\sigma_t^2}\right\}\le C_3\exp\{-C_4\rho^\alpha\}. \tag{66}$$

Further, by (63) and Proposition 3.1 as applied to random vectors $\zeta_1,...,\zeta_n$, we have

$$\mathbf{E}\left\{\|\sum_{t=1}^n\zeta_t\|^2\right\}\le C_5\kappa\left(\sum_{t=1}^n\sigma_t^2\right)\rho^2\exp\{-\rho^\alpha\},$$

whence by Tchebyshev inequality

$$\mathrm{Prob}\left\{\|\sum_{t=1}^n\zeta_t\|\ge\rho\sqrt{\kappa}\sqrt{\sum_{t=1}^n\sigma_t^2}\right\}\le C_6\exp\{-C_7\rho^\alpha\}.$$

Combining this inequality with (65) and taking into account that $\xi_t=\omega_t+\zeta_t$, we conclude that

$$\mathrm{Prob}\left\{\|\sum_{t=1}^n\xi_t\|\ge[1+2\sqrt{2}]\rho\sqrt{\kappa}\sqrt{\sum_{t=1}^n\sigma_t^2}\right\}\le C_7\exp\{-C_8\rho^\alpha\}$$

whenever $\rho\ge(2/\alpha)^{1/\alpha}$, and (55) follows.

(56) is an immediate corollary of (55). Indeed, let us fix n, and let $D=\sqrt{\kappa}\sqrt{\sum_{t=1}^n\sigma_t^2}$. For $c>0$, consider the random variable $\theta_c=\|\sum_{t=1}^n\xi_t\|^\alpha/(cD)^\alpha$. By (55), for $t>0$ we have $\psi(t)\equiv\mathrm{Prob}\{\theta_c>t\}=\mathrm{Prob}\{\|\sum_{t=1}^n\xi_t\|>ct^{1/\alpha}D\}\le C_\alpha\exp\{-C_\alpha^{-1}c^\alpha t\}$; setting $c=(2C_\alpha)^{1/\alpha}$, we get $\psi(t)\le C_\alpha\exp\{-2t\}$, whence

$$\mathbf{E}\{\exp\{\theta_c\}\}=-\int_0^\infty\exp\{t\}d\psi(t)=1+\int_0^\infty\exp\{t\}\psi(t)dt\le 1+C_\alpha\int_0^\infty\exp\{-t\}dt=1+C_\alpha.$$

Thus, $\mathbf{E}\left\{\exp\{\frac{\|\sum_{t=1}^n\xi_t\|^\alpha}{2C_\alpha D^\alpha}\}\right\}\le 1+C_\alpha$, whence, by Moment Inequality, (56) holds true with $c_\alpha=(2C_\alpha\ln(1+C_\alpha))^{1/\alpha}$. ■

4 Refinements in Gaussian case

We are about to refine the above results for the case when $\xi=\{\xi_t\}_{t=1}^\infty$ is a sequence of independent Gaussian random vectors with zero mean in a finite-dimensional normed space $(E,\|\cdot\|)$.

4.1 The basic fact

We start with the following fact which seems to be important by its own right.

Let

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int\limits_t^\infty \exp\{-s^2/2\}ds,\quad \phi(r):\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\phi(r)}\exp\{-s^2/2\}ds=r.$$

Proposition 4.1 *Let $\eta\sim\mathcal{N}(0,I_k)$, and let B be a closed convex set in $\mathbf{R}^k$ such that*

$$\mathrm{Prob}\{\eta\in B\}\geq\theta>\frac{1}{2}. \tag{67}$$

Then

$$0<\alpha<1\Rightarrow \mathrm{Prob}\{\alpha\eta\in B\}\geq 1-\exp\left\{-\frac{\phi^2(\theta)}{2\alpha^2}\right\}. \tag{68}$$

Equivalently: for a closed and convex set B and $\zeta\sim\mathcal{N}(0,\Sigma)$ one has

$$\mathrm{Prob}\left\{\zeta\not\in B\right\}\leq\delta<\frac{1}{2}\Rightarrow \mathrm{Prob}\left\{\zeta\not\in\gamma B\right\}\leq\exp\{-\frac{\phi^2(1-\delta)\gamma^2}{2}\}\ \forall\gamma>1. \tag{69}$$

Proof is based on the following fact [1]:

(!) For every $\gamma>0$, $\epsilon\geq0$ and every closed set $X\subset\mathbf{R}^k$ such that $\mathrm{Prob}\{\eta\in X\}\geq\gamma$ one has

$$\mathrm{Prob}\left\{\mathrm{dist}(\eta,X)>\epsilon\right\}\leq\Phi(\phi(\gamma)+\epsilon)$$

where $\mathrm{dist}(a,X)=\min\limits_{x\in X}\|a-x\|_2$.

Now let η,ζ be independent $\mathcal{N}(0,I_k)$ random vectors, and let

$$p(\alpha)=\mathrm{Prob}\{\alpha\eta\not\in B\}.$$

The vector $\alpha\eta+\sqrt{1-\alpha^2}\zeta$ is $\mathcal{N}(0,I_k)$, so that

$$\mathrm{Prob}\{\mathrm{dist}(\alpha\eta+\sqrt{1-\alpha^2}\zeta,B)>t\}\leq\Phi(\phi(\theta)+t) \tag{70}$$

by (!). On the other hand, let $\alpha\eta\not\in B$, and let $e=e(\eta)$ be a unit vector such that $e^T[\alpha\eta]>\max\limits_{x\in B}e^Tx$. If ζ is such that $\sqrt{1-\alpha^2}e^T\zeta>t$, then $\mathrm{dist}(\alpha\eta+\sqrt{1-\alpha^2}\zeta,B)>t$, whence

$$\alpha\eta\not\in B\Rightarrow\mathrm{Prob}\left\{\zeta:\mathrm{dist}(\alpha\eta+\sqrt{1-\alpha^2}\zeta,B)>t\right\}\geq\Phi(t/\sqrt{1-\alpha^2}),$$

whence for all $t\geq0$ such that $\delta(t)\equiv\phi(\theta)+t-t/\sqrt{1-\alpha^2}\geq0$ one has

$$\begin{array}{rcl}
p(\alpha)\Phi(t/\sqrt{1-\alpha^2}) & \leq & \mathrm{Prob}\{\mathrm{dist}(\alpha\eta+\sqrt{1-\alpha^2}\zeta,B)>t\}\leq\Phi(\phi(\theta)+t)\\
\Rightarrow p(\alpha) & \leq & \frac{\Phi(\phi(\theta)+t)}{\Phi(t/\sqrt{1-\alpha^2})}=\frac{\int\limits_{t/\sqrt{1-\alpha^2}}^\infty\exp\{-(s+\delta(t))^2/2\}ds}{\int\limits_{t/\sqrt{1-\alpha^2}}^\infty\exp\{-s^2/2\}ds}\\
 & = & \frac{\int\limits_{t/\sqrt{1-\alpha^2}}^\infty\exp\{-s^2/2-s\delta(t)-\delta^2(t)/2\}ds}{\int\limits_{t/\sqrt{1-\alpha^2}}^\infty\exp\{-s^2/2\}ds}\leq\exp\{-t\delta(t)/\sqrt{1-\alpha^2}-\delta^2(t)/2\}.
\end{array}$$

Setting in the resulting inequality $t=\frac{\phi(\theta)(1-\alpha^2)}{\alpha^2}$, we get

$$p(\alpha)\leq\exp\{-\frac{\phi^2(\theta)}{2\alpha^2}\}.$$

■

4.2 Gaussian version of Theorem 3.1

Theorem 4.1 *Let $(E, \|\cdot\|)$ be κ-regular and let $\xi_1, \xi_2, ...$ be independent Gaussian random vectors in E with zero means. Setting*

$$\delta_t^2 = \mathbf{E}\left\{\|\xi_t\|^2\right\},$$

one has

$$\Omega \geq 3 \Rightarrow \mathrm{Prob}\left\{\|\sum_{t=1}^n \xi_t\| > \Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \delta_t^2}\right\} \leq \exp\{-\frac{\Omega^2}{12.1}\}. \tag{71}$$

Proof. Let U be the unit ball of $\|\cdot\|$. By Proposition 3.1, we have

$$\mathbf{E}\left\{\|S_n\|^2\right\} \leq \underbrace{\kappa\sum_{t=1}^n \delta_t^2}_{\omega_n^2},$$

whence by Tschebyshev inequality for every $\beta > \sqrt{2}$ one has

$$\mathrm{Prob}\left\{\omega_n^{-1}S_n \notin \beta U\right\} \leq \frac{1}{\beta^2} < \frac{1}{2}.$$

Applying (69) (which is legitimate since S_n is Gaussian with zero mean), we arrive at

$$\mathrm{Prob}\left\{\omega_n^{-1}S_n \notin \gamma\beta U\right\} \leq \frac{1}{\beta^2} < \frac{1}{2}\exp\{-\frac{\phi^2(1-\beta^{-2})\gamma^2}{2}\}\ \forall\gamma > 1,$$

or, which is the same, whenever $\Omega > \sqrt{2}$, one has

$$\mathrm{Prob}\left\{\|S_n\| > \Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \delta_t^2}\right\} \leq \inf_{\sqrt{2}<\beta<\Omega} \exp\{-\frac{\phi^2(1-\beta^{-2})\Omega^2}{2\beta^2}\} \tag{72}$$

Maximizing the ratio $\phi^2(1-\beta^{-2})/\beta^2$ in β, we get from (72)

$$\Omega \geq 3 \Rightarrow \mathrm{Prob}\left\{\|S_n\| > \Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \delta_t^2}\right\} \leq \exp\{-\frac{\Omega^2}{12.1}\}. \qquad \blacksquare$$

Discussion. Invoking Proposition 4.1, it easy to demonstrate that if η is a Gaussian vector with zero mean in E, $\delta^2 = \mathbf{E}\left\{\|\eta\|^2\right\}$ and $\sigma = O(1)\delta$ with properly chosen absolute constant $O(1)$, then $\mathbf{E}\left\{\exp\{\|\eta\|^2\sigma^{-2}\}\right\} \leq \exp\{1\}$. On the other hand, if σ is such that the latter inequality holds true, then, by Jensen's inequality, $\exp\{\mathbf{E}\{\|\eta\|^2\sigma^{-2}\}\} \leq \exp\{1\}$, that is, $\sigma \geq \delta$. Thus, in the case when ξ_t are independent Gaussian vectors with zero means, the probabilities of large deviations $\mathrm{Prob}\left\{\|S_n\| \geq \Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \delta_t^2}\right\}$ can be bounded by both Theorem 3.1 and Theorem 4.1. Both bounds are of the same type $O(1)\exp\{-O(1)\Omega^2\}$; however, the absolute constants in the second bound are better than in the first one. Indeed, since the quantities σ_t arising in Theorem 3.1 should be $\geq \delta_t$, bound from Theorem 3.1 is not better than $\mathrm{Prob}\left\{\|S_n\| > \Omega\sqrt{\kappa}\sqrt{\sum_{t=1}^n \delta_t^2}\right\} \leq 3\exp\{-\frac{\Omega^2}{225}\}$, while the right hand side in the bound given by Theorem 4.1 is $\exp\{-\frac{\Omega^2}{12.1}\}$.

5 Extensions to semi-scalar case

In this section we extend the results for Gaussian case to the situation of random sums of the form

$$S_n = \sum_{t=1}^{n} \zeta_t f_t$$

with deterministic vectors $f_t \in E$ and independent random scalars ξ_t which are symmetrically distributed on the axis with "light tail" of the distributions. Since multiplying ξ_t by deterministic positive reals and dividing f_t by the same reals does not affect the situation, in the sequel we normalize ξ_t by the condition

$$\xi_t \sim -\xi_t, \ \mathbf{E}\left\{\exp\{4\xi_t^2\}\right\} \leq \exp\{1\}. \tag{73}$$

5.1 Basic results

We start with results which can be viewed as modifications of Proposition 4.1.

Proposition 5.1 *Let ξ_t be independent and symmetrically distributed random reals such that*

$$\mathbf{E}\{\xi_t^2\} \geq \sigma^2 > 0, \, t = 1, ..., n \tag{74}$$

and either

(i) $|\xi_t| \leq 1/2, t = 1, ..., n,$

or

(ii) $\mathbf{E}\left\{\exp\{4\xi_t^2\}\right\} \leq \exp\{1\}, \, t = 1, ..., n,$

and let $\xi = (\xi_1, ..., \xi_t)$. Let, further, A be a closed convex symmetric w.r.t. the origin set in E such that

$$\text{Prob}\{S_n \in A\} \equiv \mu > \nu \equiv 1 - \frac{\sigma^4}{12\sigma^4 + 1}. \tag{75}$$

Then for every $\vartheta > 1$ one has

in the case of (i):

$$\text{Prob}\left\{\xi \notin \vartheta A\right\} \leq 2 \exp\left\{-\frac{(\vartheta - 1)^2 \sigma^2}{8}\right\} \tag{76}$$

in the case of (ii):

$$\text{Prob}\{\xi \notin \vartheta A\} \leq O(1) \exp\{-O(1) \frac{\sigma^2 (\vartheta - 1)^2}{\sigma(\vartheta - 1) + \ln n}\}, \tag{77}$$

with properly chosen positive absolute constants $O(1)$.

Proof. A. We start with the following

Lemma 5 *Under the premise of (ii) (and thus under the premise of (i) as well), the set A contains the centered at the origin $\|\cdot\|_2$-ball of the radius*

$$\rho = \frac{\sigma}{\sqrt{2}}. \tag{78}$$

Proof. Assume, on the contrary to what should be proved, that there exists $a \notin A$ with $\|a\|_2 = \rho$. Then there exists a vector p, $\|p\|_2 = 1$, such that $p^T x < p^T a \leq \rho$ for all $x \in A$; since A is symmetric w.r.t. the origin, we have

$$\max_{x \in A} |p^T x| < \rho. \tag{79}$$

Consider the random variable $\zeta = |p^T \xi|$, and let $\theta = \sum_i p_i^2 \mathbf{E}\{\xi_i^2\}$, so that $\theta \geq \sigma^2$. We have

$$\mathbf{E}\{\zeta^4\} = 6 \sum_{i<j} p_i^2 p_j^2 \mathbf{E}\{\xi_i^2\} \mathbf{E}\{\xi_j^2\} + \sum_i p_i^4 \mathbf{E}\{\xi_i^4\} \leq \gamma(\theta) \equiv 3\theta^2 + \frac{1}{4}$$

(since $p_i^4\mathbf{E}\{\xi_i^4\} \leq \frac{p_i^2}{8}[\mathbf{E}\{\exp\{4\xi_i^2\}\}-1] \leq \frac{p_i^2}{4}$ due to $t^4 \leq \frac{1}{8}[\exp\{4t^2\}-1]$). Recalling that $\mathrm{Prob}\{\zeta \leq \rho\} \geq \mathrm{Prob}\{\zeta \in A\} > \nu \equiv 1 - \frac{\sigma^4}{12\sigma^4+1}$, we get

$$\theta = \mathbf{E}\{\zeta^2\} \leq \rho^2 \mathrm{Prob}\{0 \leq \zeta \leq \rho\} + \sqrt{\mathbf{E}\{\zeta^4\}}\sqrt{\mathrm{Prob}\{\zeta > \rho\}} < \rho^2 + \sqrt{(1-\nu)\gamma(\theta)},$$

whence

$$\rho^2 > \theta - \sqrt{(1-\nu)\gamma(\theta)} \equiv \phi(\theta). \tag{80}$$

Observe that $\phi'(\theta) \geq 0$ provided that $\sqrt{1-\nu} \leq \frac{\sqrt{3\theta^2+1/4}}{3\theta}$, which definitely is the case when $\nu > \frac{2}{3}$, as guaranteed by the origin of ν (note that $\sigma \leq 1$ due to $\mathbf{E}\{\exp\{4\zeta_t^2\}\} \leq \exp\{1\}$). Consequently, (80) combines with $\theta \geq \sigma^2$ to imply that

$$\rho^2 > \phi(\sigma^2) = \sigma^2 - \sqrt{1-\nu}\sqrt{3\sigma^4+1/4},$$

whence, by (75), $\rho^2 > \sigma^2/2$, which is a contradiction. ■

B. Recall that by Talagrand Inequality, for a sequence of independent random real variables ξ_i, $i = 1,...,n$, taking values in $[-1/2, 1/2]$ and a closed set A in $\mathbf{R}^n$ one has

$$\mathbf{E}\left\{\frac{1}{4}\exp\{\mathrm{dist}^2_{\|\cdot\|_2}(\xi, \mathrm{Conv}(A))\}\right\} \leq \frac{1}{\mathrm{Prob}\{\xi \in A\}}. \tag{81}$$

C. W.l.o.g., let A be a compact set; by Lemma 5, A contains $\|\cdot\|_2$-ball of radius ρ centered at the origin. Let $\vartheta > 1$ and $x \notin \vartheta A$. Consider the norm $\|\cdot\|$ in which A is the unit ball; since $x \notin \vartheta A$, the $\|\cdot\|$-ball B of the radius $\vartheta - 1$ centered at x does not intersect A. Since the $\|\cdot\|_2$-ball of the radius $\rho(\vartheta-1)$, centered at x, is contained in B, this ball does not intersect A as well. Thus,

$$x \notin \vartheta A \Rightarrow \mathrm{dist}_{\|\cdot\|_2}(x, A) \equiv \mathrm{dist}_{\|\cdot\|_2}(x, \mathrm{Conv}(A)) > \rho(\vartheta - 1). \tag{82}$$

D. Assume that (i) is the case. Combining (81) with (82), we arrive at

$$\mathrm{Prob}\{\xi \notin \vartheta A\} \leq \mu^{-1}\exp\{-\rho^2(\vartheta-1)^2/4\}$$

with ρ given by (78), as required in (76).

E. Now assume that (ii) is the case. Let $L > 0$, let Ξ be the event $\{\xi : \|\xi\|_\infty \leq L/2\}$ and p be the probability of the event $\{\xi \notin \Xi\}$. We have

$$p \leq \sum_{i=1}^n \mathrm{Prob}\{|\xi_i| > L/2\} \leq \sum_i \mathbf{E}\left\{\exp\{4\xi_i^2 - L^2\}\right\} \leq n\exp\{1 - L^2\}. \tag{83}$$

Applying (81) to the conditional, by the condition $\xi \in \Xi$, distribution of ξ (which is again a distribution with independent coordinates), we get

$$\begin{array}{rcl}
\mathbf{E}\left\{\exp\{\frac{\mathrm{dist}^2_{\|\cdot\|_2}(\xi,A)}{4L^2}\}\big|\Xi\right\} & \leq & \frac{1}{\mathrm{Prob}\left\{\xi\in A\big|\Xi\right\}} \leq \frac{1-p}{\mu-p} \Rightarrow \\
\mathbf{E}\left\{\exp\{\frac{\mathrm{dist}^2_{\|\cdot\|_2}(\xi,A)}{4L^2}\}\chi_{\xi\in\Xi}\right\} & \leq & \frac{(1-p)^2}{\mu-p} \Rightarrow \\
\mathrm{Prob}\{\xi\in\Xi \;\&\; \xi\notin\vartheta A\} & \leq & \mathrm{Prob}\left\{\xi\in\Xi \;\&\; \frac{\mathrm{dist}^2_{\|\cdot\|_2}(\xi,A)}{4L^2} \geq \frac{\rho^2(\vartheta-1)^2}{4L^2}\right\} \\
 & \leq & \exp\{-\frac{\rho^2(\vartheta-1)^2}{4L^2}\}\mathbf{E}\left\{\exp\{\frac{\mathrm{dist}^2_{\|\cdot\|_2}(\xi,A)}{4L^2}\}\chi_{\xi\in\Xi}\right\} \\
 & \leq & \exp\{-\frac{\rho^2(\vartheta-1)^2}{4L^2}\}\frac{(1-p)^2}{\mu-p} = \exp\{-\frac{\sigma^2(\vartheta-1)^2}{8L^2}\}\frac{(1-p)^2}{\mu-p} \Rightarrow \\
\mathrm{Prob}\{\xi\notin\vartheta A\} & \leq & p + \exp\{-\frac{\sigma^2(\vartheta-1)^2}{8L^2}\}\frac{(1-p)^2}{\mu-p}
\end{array}$$

Assuming $\sigma^2(\vartheta-1)^2 \geq 4$, let $L^2 = \ln n + \sigma(\vartheta - 1)$. With this L, the resulting bound combines with (83) to imply (77). ■

Proposition 5.1 describes a family of probability distributions P on $\mathbf{R}^n$ with the following common property: for every closed convex set A centered at the origin, the "probability mass" $P(\gamma A)$ of γ-enlargement of A rapidly approaches 1 as γ grows, provided that $P(A)$ is not small. A shortcoming of the representation of this phenomenon as given by Proposition 5.1 is that that what is and what is not small depends on the parameter σ; for small value of the parameter, "not small" actually means "close to 1". We are about to present a slightly modified version of Proposition 5.1 in which every fixed positive value of $P(A)$ "is not small".

Proposition 5.2 *Let ξ_t be independent and symmetrically distributed random reals, and let the distributions of ξ_t possess densities $p_t(\cdot)$ such that*

$$p_t(\cdot) \le \frac{1}{3\sqrt{3}\sigma} \tag{84}$$

and either

(i) $|\xi_t| \le 1/2, t = 1, ..., n,$

or

(ii) $\mathbf{E}\left\{\exp\{4\xi_t^2\}\right\} \le \exp\{1\}$, $t = 1, ..., n$,

and let $\xi = (\xi_1, ..., \xi_t)$. Let, further, A be a closed convex symmetric w.r.t. the origin set in E, and let

$$\mu \equiv \text{Prob}\{S_n \in A\} > 0. \tag{85}$$

Then for every $\vartheta > 1$ one has
in the case of (i):

$$\text{Prob}\left\{\xi \notin \vartheta A\right\} \le \frac{1}{\mu} \exp\left\{-\frac{\mu^6\sigma^6(\vartheta-1)^2}{256}\right\} \tag{86}$$

in the case of (ii):

$$\text{Prob}\{\xi \notin \vartheta A\} \le \mu^{-1}\exp\{-O(1)\frac{\mu^6\sigma^6(\vartheta-1)^2}{\mu^3\sigma^3(\vartheta-1)+\ln(n/\mu)}\} \tag{87}$$

with properly chosen positive absolute constant $O(1)$.

Proof. A. We start with the following

Lemma 6 *Let $a \in \mathbf{R}^n$, $\|a\|_2 = 1$, and let $\zeta = a^T\xi$. Then*

$$\rho \in (0, 1/2] \Rightarrow \text{Prob}\left\{|\zeta| \le \rho\right\} \le 2\rho^{1/3}\sigma^{-1}. \tag{88}$$

Proof. Observe, first, that

$$\sigma_t^2 \equiv \mathbf{E}\{\xi_t^2\} \ge \sigma^2. \tag{89}$$

Indeed, for every $\delta \le \frac{3\sqrt{3}\sigma}{2}$, we have $\text{Prob}\{|\xi_t| > \delta\} \ge 1 - \frac{2\delta}{3\sqrt{3}\sigma}$, whence $\sigma_t^2 \ge \delta^2(1 - \frac{2\delta}{3\sqrt{3}\sigma})$. Maximizing over δ, we arrive at (89).

Since $p_t(\cdot)$ are even and $\mathbf{E}\left\{\exp\{4\xi_t^2\}\right\} \le \exp\{1\}$, the generation functions $\phi_t(y) = \mathbf{E}\{\exp\{ixy\}\}$ are real-valued, even and C^∞; besides this, $\phi_t''(0) = -\sigma_t^2 \le -\sigma^2$, $|\phi_t^{(4)}(y)| \le \mathbf{E}\left\{\xi_t^4\right\} \le \frac{1}{8}\mathbf{E}\left\{\exp\{4\xi_t^2\} - 1\right\} \le \frac{e-1}{8}$. Consequently, for $|y| \le 5\sigma$ the remainder in the third order Taylor expansion of $\phi_t(y)$, taken at the origin, does not exceed $\frac{1}{24}\frac{e-1}{8}y^4 \le \sigma^2y^2/4$, whence

$$|y| \le 5\sigma \Rightarrow \phi_t(\sigma) \le 1 - \sigma_t^2y^2/2 + \sigma^2y^2/4 \le 1 - \sigma^2y^2/4 \le \exp\{-\sigma^2y^2/4\}. \tag{90}$$

Now let $\alpha = \max_t |a_t|$ and $\phi(y)$ be the generating function of ζ: $\phi(y) = \prod_{t=1}^n \phi_t(a_ty)$. By (90), we have

$$|y| \le 5\sigma\alpha^{-1} \Rightarrow \phi(y) \le \prod_{t=1}^n \exp\{-a_t^2\sigma^2y^2/4\} \le \exp\{-\sigma^2\|a\|_2^2y^2/4\}. \tag{91}$$

Now let $1/2 \geq \rho > 0$. Consider the function $h(x) = \frac{1}{\sqrt{2\rho}}\chi_{|x|\leq 2\rho}$ along with the function $g = h * h$ ($*$ stands for convolution). Function g clearly is nonnegative and $g(x) \geq 1$ when $|x| \leq \rho$. Observe that the Fourier transform of g is the function $\frac{2\sin^2(\rho y)}{\rho y^2} \in [0,1]$. Denoting by $p(\cdot)$ the density of ζ, we have

$$
\begin{array}{rcl}
\mathrm{Prob}\{|\zeta| \leq \rho\} & \leq & \int p(x)g(x)dx = \frac{1}{2\pi}\int \phi(y)\frac{2\sin^2(\rho y)}{\rho y^2}dy \\
& \leq & \frac{1}{2\pi}\int\limits_{|y|\leq 5\sigma\alpha^{-1}} \phi(y)\frac{2\sin^2(\rho y)}{\rho y^2}dy + \frac{1}{\pi}\int\limits_{5\sigma\alpha^{-1}}^{\infty}\frac{2}{\rho y^2}dy \leq \frac{1}{2\pi}\int \exp\{-\frac{\sigma^2 y^2}{4}\}\frac{2\sin^2(\rho y)}{\rho y^2}dy + \frac{2\alpha}{5\pi\rho\sigma} \\
& \leq & \frac{1}{\pi}\int \exp\{-\frac{\sigma^2 z^2}{4\rho^2}\}\frac{\sin^2(z)}{z^2}dz + \frac{2\alpha}{5\pi\rho\sigma} \leq \frac{1}{\pi}\int \exp\{-\frac{\sigma^2 z^2}{4\rho^2}\}dz + \frac{2\alpha}{5\pi\rho\sigma} = \frac{4\rho}{\sqrt{2\pi}\sigma} + \frac{2\alpha}{5\sigma\rho}
\end{array}
$$

Besides this, the uniform norm of the density of ζ clearly does not exceed the minimum, over t, of the uniform norms of the densities of $a_t\xi_t$, that is, it does not exceed $\frac{1}{3\sqrt{3}\sigma\alpha}$. We conclude that

$$
\mathrm{Prob}\left\{|\zeta| \leq \rho\right\} \leq \min\left[\frac{2\rho}{3\sqrt{3}\sigma\alpha}, \frac{4\rho}{\sqrt{2\pi}\sigma} + \frac{2\alpha}{5\sigma\rho}\right] \tag{92}
$$

In the case of $\alpha \geq \rho^{2/3}$, (92) yields

$$
\mathrm{Prob}\left\{|\zeta| \leq \rho\right\} \leq \frac{2\rho^{1/3}}{3\sqrt{3}\sigma}.
$$

Now let $\alpha < \rho^{2/3}$. Invoking (92), we have

$$
\mathrm{Prob}\{|\zeta| \leq \rho\} \leq \mathrm{Prob}\{|\zeta| \leq \rho^{1/3}\} \leq \frac{1}{\sigma}\left[\frac{4\rho^{1/3}}{\sqrt{2\pi}} + \frac{2\alpha}{5\rho^{1/3}}\right] \leq \frac{\rho^{1/3}}{\sigma}\left[\frac{4}{\sqrt{2\pi}} + \frac{2}{5}\right] \leq 2\rho^{1/3}\sigma^{-1}.
$$

Thus, in all cases $\mathrm{Prob}\left\{|\zeta| \leq \rho\right\} \leq 2\rho^{1/3}\sigma^{-1}$, as claimed. ■

B. We now claim that under the premise of Proposition 5.2, A contains the centered at the origin $\|\cdot\|_2$-ball of radius $\frac{\rho=(\mu\sigma}{2)^3}$. Indeed, otherwise, same as in the proof of Lemma 5, we could find a vector p, $\|p\|_2 = 1$ and $\rho' < \rho$ such that A is contained in the stripe $\{x : |p^T x| < \rho\}$, that is, with $\zeta = p^T\xi$ one has $\mathrm{Prob}\{|\zeta| < \rho'\} \geq \mathrm{Prob}\{\xi \in A\} = \mu$. On the other hand, $\rho' < \rho \leq 1/2$, where the latter inequality follows from the fact that $\sigma \leq 1$ (indeed, $\sigma^2 \leq \sigma_t^2$ by (89), while $\sigma_t^2 = \mathbf{E}\{\xi_t^2\} \leq \mathbf{E}\{\frac{1}{8}[\exp\{4\xi_t^2\} - 1]\} \leq \frac{e-1}{8}$). Applying Lemma 6, we get $\mathrm{Prob}\{|\zeta| < \rho'\} \leq 2(\rho')^{1/2}\sigma < \mu$, which is a contradiction.

C. Now we can complete the proof in exactly the same way as in the case of Proposition 4.1. Specifically, same as in item C of the latter proof, relation (82) with ρ given by B holds true. In the case of (i) this observation combines with Talagrand Inequality (81) to yield the relation

$$
\mathrm{Prob}\left\{\xi \notin \vartheta A\right\} \leq \frac{1}{\mu}\exp\{-\frac{(\vartheta-1)^2\rho^2}{4}\} = \frac{1}{\mu}\exp\{-\frac{(\vartheta-1)^2\rho^2}{4}\} = \frac{1}{\mu}\exp\{-\frac{\mu^6\sigma^6(\vartheta-1)^2}{256}\},
$$

as required in (77). In the case of (ii), the same reasoning as in item E of the proof of Proposition 4.1 results in (87). ■

5.2 Semi-scalar version of Theorem 3.1

Theorem 5.1 *Let f_t be deterministic vectors from a normed finite-dimensional space $(E, \|\cdot\|)$ and ξ_t be independent symmetrically distributed random scalars such that*

$$
\mathbf{E}\{\xi_t^2\} \geq \sigma^2 > 0,\ t = 1, ..., n
$$

and either

(i) $|\xi_t| \leq \frac{1}{2}$, $t = 1, ..., n$,

or

(ii) $\mathbf{E}\left\{\exp\{4\xi_t^2\}\right\} \leq \exp\{1\}$, $t = 1, ..., n$.

Assume that

$$
\mathbf{E}\left\{\|\sum_{t=1}^{n}\xi_t f_t\|^2\right\} < \Theta^2. \tag{93}
$$

Then
in the case of (i):

$$\text{Prob}\left\{\|\sum_{t=1}^{n}\xi_t f_t\| \geq \Omega\Theta\right\} \leq O(1)\exp\{-O(1)\sigma^6\Omega^2\} \tag{94}$$

with appropriate positive absolute constants $O(1)$*;*
in the case of (ii):

$$\text{Prob}\left\{\|\sum_{t=1}^{n}\xi_t f_t\| \geq \Omega\Theta\right\} \leq O(1)\exp\{-O(1)\frac{\sigma^6\Omega^2}{\ln n + \sigma^3\Omega}\} \tag{95}$$

Proof. Let $\xi = (\xi_1, ..., \xi_n)$, $B = \{y \in E : \|y\| \leq r\Theta\}$ and $A = \{s \in \mathbf{R}^n : \sum\limits_{t=1}^{n} s_t f_t \in B\}$. Then A is a convex closed symmetric w.r.t. the origin set such that $\text{Prob}\{\xi \in A\} \geq 1 - r^{-2}$ by (93). Setting $r^2 = \frac{12\sigma^4+1}{\sigma^4}$, we get $\text{Prob}\{\xi \in A\} > \nu \equiv 1 - \frac{\sigma^4}{12\sigma^4+1}$. Applying Proposition 5.1, we get

- In the case of (i):

$$\begin{array}{rcl}\text{Prob}\{\|\sum\limits_{t=1}^{n}\xi_t f_t\| > \Omega\Theta\} &=& \text{Prob}\left\{\xi \notin \Omega r^{-1}A\right\} \leq O(1)\exp\{-O(1)\Omega^2 r^{-2}\sigma^2\} \\ &\leq& O(1)\exp\{-O(1)\sigma^6\Omega^2\}.\end{array}$$

- in the case of (ii):

$$\begin{array}{rcl}\text{Prob}\{\|\sum\limits_{t=1}^{n}\xi_t f_t\| > \Omega\Theta\} &=& \text{Prob}\left\{\xi \notin \Omega r^{-1}A\right\} \leq O(1)\exp\{-O(1)\frac{\Omega^2 r^{-2}\sigma^2}{\ln n + \Omega r^{-1}\sigma}\} \\ &\leq& O(1)\exp\{-O(1)\frac{\sigma^6\Omega^2}{\ln n+\sigma^3\Omega}\}.\end{array}$$

■

References

[1] Borell, C., "The Brunn-Minkowski inequality in Gauss space" – *Inventiones Mathematicae* **30** (2) (1975), 207-216.

[2] Johnson, W.B., Schechtman, G., "Remarks on Talagrand's deviation inequality for Rademacher functions", Banach Archive 2/16/90, *Springer Lecture Notes* 1470 (1991), pp. 72-77.

[3] Nemirovski, A., "On tractable approximations of randomly perturbed convex constraints" – *Proceedings of the 42nd IEEE Conference on Decision and Control Maui, Hawaii USA, December 2003*, 2419-2422.

6 Appendix: Proof of Proposition 2.1

Let $\{f_k(t)\}$ be a sequence of polynomials converging to f, along with the first and the second derivatives, uniformly on every compact subset of Δ. For a polynomial $p(t) = \sum_{j=0}^{N} p_j t^j$ the function $P(X) = \text{Tr}(\sum_j p_j X^j)$ is a polynomial on $\mathbf{S}^n$. Let now $X, H \in \mathbf{S}^n$, let $\lambda_s = \lambda_s(X)$ be the eigenvalues of X, $X = U\text{Diag}\{\lambda\}U^T$ be the eigenvalue decomposition of X, and let $\widehat{H}$ be such that $H = U\widehat{H}U^T$. We have

$$\begin{array}{rcll}P(X) &=& \sum_{s=1}^{n} p(\lambda_s(X)) & (a) \\ DP(X)[H] &=& \text{Tr}(\sum_{j=1}^{N}\sum_{s=0}^{N-1} X^s H X^{N-s-1} = \text{Tr}(p'(X)H) = \sum_{s=1}^{n} p'(\lambda_s(X))\widehat{H}_{ss} & (b)\end{array} \tag{96}$$

Further, let γ be a closed contour in the complex plane encircling all the eigenvalues of X. Then

$$DP(X)[H] = \text{Tr}(p'(X)H) = \frac{1}{2\pi\imath}\oint\limits_{\gamma} p'(z)\text{Tr}((zI - X)^{-1}H)dz$$

$$\Rightarrow D^2P(X)[H,H] = \frac{1}{2\pi\imath}\oint\limits_{\gamma} p'(z)\text{Tr}((zI - X)^{-1}H(zI - X)^{-1}H)dz = \frac{1}{2\pi\imath}\oint\limits_{\gamma}\sum_{s,t=1}^{n}\frac{\widehat{H}_{st}^2 p'(z)}{(z-\lambda_s)(z-\lambda_t)}dz.$$

Computing the residuals, we get

$$D^2P(X)[H,H] = \sum_{s,t} \Gamma_{s,t}[p]\widehat{H}_{st}^2, \quad \Gamma_{s,t}[p] = \left\{ \begin{array}{ll} \frac{p'(\lambda_s)-p'(\lambda_t)}{\lambda_s-\lambda_t}, & \lambda_s \neq \lambda_t \\ p''(\lambda_s), & \lambda_s = \lambda_t \end{array} \right. \tag{97}$$

Substituting $p = f_k$ into $(96.a, b)$ and (97), we see that the sequence of polynomials $F_k(X) = \mathrm{Tr}(f_k(X))$ converges, along with the first and the second order derivatives, uniformly on compact subsets of $\mathcal{X}_n(\Delta)$; by $(96.a)$, the limiting function is exactly $F(X)$. We conclude that $F(X)$ is C^2 on $\mathcal{X}_n(\Delta)$ and that the first and the second derivatives of this function are limits, as $k \to \infty$, of the corresponding derivatives of $F_k(X)$, so that for $X = U\mathrm{Diag}\{\lambda\}U^T \in \mathcal{X}_n(\Delta)$ (where U is orthogonal) and every $H = U\widehat{H}U^T \in \mathbf{S}^n$ we have

$$\begin{array}{rcl} DF(X)[H] & = & \sum_s f'(\lambda_s)\widehat{H}_{ss} = \mathrm{Tr}(f'(X)H) \\ D^2F(X)[H,H] & = & \sum_{s,t} \Gamma_{s,t}[f]\widehat{H}_{st}^2 \end{array} \tag{98}$$

So far, we did not use (12). Invoking the right inequality in (12), we get

$$\begin{array}{l} D^2F(X)[H,H] \leq \sum_{s,t} \left[\theta_+ \frac{f''(\lambda_s)+f''(\lambda_t)}{2} + \mu_+\right] \widehat{H}_{st}^2 = \theta_+ \sum_s f''(\lambda_s) \sum_t \widehat{H}_{st}^2 + \mu_+ \sum_{s,t} \widehat{H}_{st}^2 \\ = \theta_+ \mathrm{Tr}(\mathrm{Diag}\{f''(\lambda_1), ..., f''(\lambda_n)\}\widehat{H}^2) + \mu_+ \mathrm{Tr}(\widehat{H}^2) = \theta_+ \mathrm{Tr}(f''(X)H^2) + \mu_+ \mathrm{Tr}(H^2), \end{array}$$

which is the right inequality in (13). The derivation of the left inequality in (13) is similar. ■