HAL open science

Large Deviations of Vector-valued Martingales in 2-Smooth Normed Spaces

Anatoli B. Juditsky, Arkadii S. Nemirovski

To cite this version:

Anatoli B. Juditsky, Arkadii S. Nemirovski. Large Deviations of Vector-valued Martingales in 2-Smooth Normed Spaces. 2008. hal-00318071

HAL Id: hal-00318071

https://hal.science/hal-00318071v1

Preprint submitted on 3 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Submitted to the Annals of Applied Probability

LARGE DEVIATIONS OF VECTOR-VALUED MARTINGALES IN 2-SMOOTH NORMED SPACES

By Anatoli B. Juditsky and Arkadi S.Nemirovski*

Université Grenoble I and Georgia Institute of Technology

In this paper, we derive exponential bounds on probabilities of large deviations for "light tail" martingales taking values in finite-dimensional normed spaces. Our primary emphasis is on the case where the bounds are dimension-independent or nearly so. We demonstrate that this is the case when the norm on the space can be approximated, within an absolute constant factor, by a norm which is differentiable on the unit sphere with a Lipschitz continuous gradient. We also present various examples of spaces possessing the latter property.

1. Introduction. It is well-known that for a sequence of independent zero mean random reals $\{\xi_i\}_{i=1}^\infty$ with light tail distributions (e.g., such that $\mathbf{E}\left\{\exp\{|\xi_i|^\alpha\sigma_i^{-\alpha}\}\right\} \le \exp\{1\}$ for certain $\alpha \in [1,2]$ and deterministic $\sigma_t > 0$), a "typical magnitude" of the sum $S_t = \sum_{i=1}^t \xi_i$ is "at most of order of $\sqrt{\sum_{i=1}^t \sigma_i^2}$ ", meaning that

$$\mathrm{Prob}\left\{|S_t| > [1+\gamma]\sqrt{\sum_{i=1}^t \sigma_i^2}\right\} \le O(1)\exp\{-O(1)\gamma^\alpha\}$$

for all $\gamma \ge 0$; here in what follows, all $O(1)$ are positive absolute constants. The question we focus on in this paper is to which extent the above large deviation bound is preserved when passing from scalar random variables to independent zero mean random variables taking values in a normed space $(E, \|\cdot\|)$ of (possibly, large) dimension $n < \infty$. Now our "light tail" condition reads

$$\mathbf{E}\left\{\exp\{\|\xi_i\|^\alpha\sigma_i^{-\alpha}\}\right\} \le \exp\{1\} \tag{1}$$

* Research of this author was partly supported by the NSF award DMI-0619977.

AMS 2000 subject classifications: Primary 60F10; secondary 60B12, 90C15

Keywords and phrases: large deviations, matrix-valued martingales, deviations of vector norms

for some $\alpha \in [1,2]$, and what we want to get is a bound of the form

$$\forall \gamma \geq 0 : \mathrm{Prob}\left\{ \|\sum_{i=1}^{t} \xi_i\| > [\theta+\gamma]\sqrt{\sum_{i=1}^{t} \sigma_i^2} \right\} \leq O(1)\exp\{-O(1)\gamma^\alpha\} \qquad (*)$$

with a "moderate" value of the constant θ. It is immediately seen that our goal is not always attainable. For instance, let $(E, \|\cdot\|)$ be ℓ_1^n (i.e., $\mathbf{R}^n$ equipped with the norm $\|x\|_1 = \sum_{i=1}^n |x_i|$), and let ξ_i take values $\pm e_i$ with probability 1/2, $1 \leq i \leq n$, where e_i are the standard basic orths in $\mathbf{R}^n$. Then (1) holds true with $\sigma_i = 1$, while $\|S_k\|_1 \equiv k$ whenever $k \leq n$. We see that in order for $(*)$ to be true, θ should be as large as $O(1)\sqrt{n}$. On the other hand, with $\theta = O(1)\sqrt{\dim E}$, $(*)$ indeed is true independently of the norm $\|\cdot\|$ in question (see Example 3.1 in Section 3.1). Our major goal in this paper is to show that a sufficient condition for $(*)$ to be valid with certain θ is θ^2-*regularity* of the space $(E, \|\cdot\|)$. The latter means, essentially, that $\|\cdot\|$ can be approximated within an absolute constant factor by a norm $p(\cdot)$ which is continuously differentiable outside of the origin and possesses Lipschitz continuous, with the Lipschitz constant θ^2, derivative on its unit sphere:

$$p(x) = p(y) = 1 \Rightarrow p_*(p'(x) - p'(y)) \leq \theta^2 p(x-y) \qquad (2)$$

(here p_* is the norm on the dual space E^*, which is dual to p). Examples of κ-regular norms with "moderate" κ include the spaces $(\mathbf{R}^n, \|\cdot\|_p)$ (L_p on an n-point set with unit masses of points) and the spaces $(\mathbf{R}^{m\times n}, |\cdot|_p)$, $2 \leq p \leq \infty$, of $m \times n$ matrices with the Shatten norms $|X|_p = \|\sigma(X)\|_p$, $\sigma(X)$ being the vector of singular values of a matrix X; in both cases, $p \in [2,\infty]$. The spaces of the first series are κ-regular with $\kappa = O(1)\min[p, \ln(n+1)]$, while the spaces of the second series are κ-regular with $\kappa = O(1)\min[p, \ln(m+1), \ln(n+1)]$.

Norms $p(\cdot)$ satisfying (2) play important role in the theory of Banach spaces (where they are called norms with smoothness modulus of power 2). In particular, a number of results on the properties of martingales taking values in Banach spaces with smooth norms (see, e.g., [3, 4]) are available. However, we were unable to locate in the literature a result equivalent to Theorem 2.1 which establishes the validity of (somehow refined) bound $(*)$ in the case of a θ^2-regular space $(E, \|\cdot\|)$. Thus, the main result of this paper, to the best of our (perhaps incomplete) knowledge, is new. The preliminary and slightly less accurate, version of Theorem 2.1 was announced in [10] and proved in the preprint [11].

While the question we address seems to be important by its own right, our interest in it stems mainly from various applications of (somehow rudimentary) bounds of type $(*)$ we have encountered over the years. These applications include investigating performance of Euclidean and non-Euclidean stochastic approximation [7, 5], nonparametric statistics [8, 5, 9], optimization under uncertainty [10], investigating quality of semidefinite relaxations of some difficult combinatorial problems [12], etc.

Our paper is organized as follows: the main result on large deviations (Theorem 2.1) is formulated in Section 2. Section 3.1 contains instructive examples and characterizations of κ-regular spaces, along with a kind of "calculus" of these spaces. All proofs are placed in the appendix.

In what follows, if not explicitly stated otherwise, we suppose all the relations between random variables to hold a.s..

2. Main result.

2.1. *Regular spaces.* We start with the following

Definition 2.1. *Let $(E, \|\cdot\|)$ be a finite-dimensional normed space and let $\kappa \geq 1$.*
(i) *The function $p(x) = \|x\|^2$ called κ-smooth if it is continuously differentiable and*

$$\forall x, y \in E : p(x+y) \leq p(x) + Dp(x)[y] + \kappa p(y). \tag{3}$$

(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,*

$$\forall x \in E : \|x\|^2 \leq \|x\|_+^2 \leq \frac{\kappa}{\kappa_+}\|x\|^2. \tag{4}$$

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

As an immediate example, an Euclidean space $(\mathbf{R}^n, \|\cdot\|_2)$ is 1-smooth and thus 1-regular.

2.2. *Main result.* Assume that we are given

- a finite-dimensional 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 $\mathbf{E}_i$, $i = 1, 2, ...$ the conditional expectation w.r.t. $\mathcal{F}_i$, and by $\mathbf{E} \equiv \mathbf{E}_0$ the expectation w.r.t. μ.

We further assume that we are given an E-valued martingale-difference sequence $\xi^\infty = \{\xi_i\}_{i=1}^\infty$ of Borel E-valued functions on Ω such that ξ_i is $\mathcal{F}_i$-measurable and

$$\mathbf{E}_{i-1}\{\xi_i\} \equiv 0,\ i = 1, 2, ...$$

An immediate consequence of Definition 2.1 of the regular norm is as follows: assume that an E-valued martingale-difference $\xi = \{\xi_t\}_{t=1}^\infty$ is square-integrable: $\mathbf{E}\{\|\xi_t\|^2\} \le \sigma_t^2 < \infty$. Then

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

Indeed, $\|\cdot\|_+$ 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 a martingale-difference,

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

by the right inequality of (4). Then, by the left inequality of (4),

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

Our primary objective is to establish exponential bounds on the probabilities of large deviations for an E-valued martingale difference $\{\xi_i\}$. To this end, we impose on $\{\xi_i\}$ a "light tail" assumption as follows. Let $\alpha \in [1, 2]$ and a sequence $\sigma^\infty = \{\sigma_i > 0\}_{i=1}^\infty$ of (deterministic) positive reals be given. We introduce the following condition on the sequence ξ^∞:

$$\forall i \ge 1 : \mathbf{E}_{i-1}\left\{\exp\{\|\xi_i\|^\alpha \sigma_i^{-\alpha}\}\right\} \le \exp\{1\} \text{ almost surely} \qquad (\mathcal{C}_\alpha[\sigma^\infty])$$

Our main result is the large deviation bound for $S_N = \sum_{i=1}^N \xi_i$ as follows:

THEOREM 2.1. *Let $(E, \|\cdot\|)$ be κ-regular, let E-valued martingale-difference ξ^∞ satisfy $(\mathcal{C}_\alpha[\sigma^\infty])$, and let $S_N = \sum_{i=1}^N \xi_i$, $\sigma^N = [\sigma_1; ...; \sigma_N]$. Then*

(i) *for* $1 \le \alpha \le 2$*, one has for all* $N \ge 1$ *and* $\gamma \ge 0$*:*

(5)
$$\mathrm{Prob}\left\{\|S_N\| \ge \left[\sqrt{2e\kappa}+\sqrt{2}\gamma\right]\sqrt{\sum_{i=1}^N \sigma_i^2}\right\} \le 2\exp\left\{-\frac{1}{64}\min\left[\gamma^2;\gamma_*^{2-\alpha}\gamma^\alpha\right]\right\},$$

where

(6)
$$\gamma_* \equiv \gamma_*(\alpha,\nu^N) = \begin{cases} 32\left[\frac{8\alpha_*}{2^{\alpha_*}}\right]^{\frac{\alpha-1}{2-\alpha}}\left[\frac{\|\nu^N\|_2}{\|\nu^N\|_{\alpha_*}}\right]^{\frac{\alpha}{2-\alpha}} \ge 16\left[\frac{\|\nu^N\|_2}{\|\nu^N\|_{\alpha_*}}\right]^{\frac{\alpha}{2-\alpha}} \ge 16, & 1<\alpha<2,\\ \left[\alpha_* = \frac{\alpha}{\alpha-1}, \nu^N = [\nu_1;...;\nu_N]\right], & \\ \lim_{\alpha\to 1+0}\gamma_*(\alpha,\nu^N) = 16\frac{\|\nu^N\|_2}{\|\nu^N\|_\infty}, & \alpha=1,\\ \lim_{\alpha\to 2-0}\gamma_*(\alpha,\nu^N) = +\infty, & \alpha=2.\end{cases}$$

(ii) *When* $\alpha = 2$*, the bound* (42) *improves to*

(7)
$$(\forall N\ge 1,\gamma\ge 0): \mathrm{Prob}\left\{\|S_N\| \ge \left[\sqrt{2\kappa}+\sqrt{2}\gamma\right]\sqrt{\sum_{i=1}^N \sigma_i^2}\right\} \le \exp\{-\gamma^2/3\}.$$

(iii) *When the condition* $\mathbf{E}_{i-1}\left\{\exp\{\|\xi_i\|^2\sigma_i^{-2}\}\right\} \le \exp\{1\}$ *in* $(\mathcal{C}_2[\sigma^\infty])$ *is strengthened to* $\|\xi_i\| \le \sigma_i$ *almost surely,* $i = 1,2,...$*, the bound* (42) *improves to*

(8)
$$(\forall N\ge 1,\gamma\ge 0): \mathrm{Prob}\left\{\|S_N\| \ge \left[\sqrt{2\kappa}+\sqrt{2}\gamma\right]\sqrt{\sum_{i=1}^N \sigma_i^2}\right\} \le \exp\left\{-\gamma^2/2\right\}.$$

3. Regular spaces. To make Theorem 2.1 meaningful, we need to point out a spectrum of interesting κ-smooth/regular spaces, and this is the issue we consider in this Section.

3.1. *Basic examples.* Let E be an n-dimensional linear space, and let $\|\cdot\|$ be a norm on E. It is well known [2] that there exists an ellipsoid Q centered at the origin such that $Q \subset \{x \in E : \|x\| \le 1\} \subset \sqrt{n}Q$, or, equivalently, there exists a Euclidean norm $\|\cdot\|_+$ on E such that $\|x\|^2 \le \|x\|_+^2 \le n\|x\|^2$. Since the Euclidean space $(E, \|\cdot\|_+)$ is 1-smooth, we conclude that

Example 3.1. . *Every finite-dimensional normed space* $(E, \|\cdot\|)$ *is* $(\dim E)$*-regular.*

We are about to present a number of less trivial examples, those where the regularity parameter κ is dimension-independent (or nearly so).

EXAMPLE 3.2. *Let $2 \le p \le \infty$. The space $(\mathbf{R}^n, \|\cdot\|_p)$ with $n \ge 3$ is $\kappa_p(n)$-regular with*

$$\kappa_p(n) = \min_{\substack{2\le\rho\le p\\ \rho<\infty}} (\rho-1)n^{\frac{2}{\rho}-\frac{2}{p}} \le \min[p-1, 2\ln(n)] \tag{9}$$

EXAMPLE 3.3. *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*

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

The proof of the bound (10) is based upon the fact which is important by its own right:

PROPOSITION 3.1. *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*

$$\forall(a<b, a,b\in\Delta): \theta_-\frac{f''(a)+f''(b)}{2}+\mu_- \le \frac{f'(b)-f'(a)}{b-a} \le \theta_+\frac{f''(a)+f''(b)}{2}+\mu_+ \tag{11}$$

Let, further, $\mathcal{X}_n(\Delta)$ be the set of all $n\times n$ symmetric matrices with eigenvalues belonging to Δ. Then $\mathcal{X}_n(\Delta)$ is an open convex set in the space $\mathbf{S}^n$ of $n\times n$ symmetric matrices, 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

$$\theta_-\,\mathrm{Tr}(Hf''(X)H)+\mu_-\,\mathrm{Tr}(H^2) \le D^2F(X)[H,H] \le \theta_+\,\mathrm{Tr}(Hf''(X)H)+\mu_+\,\mathrm{Tr}(H^2). \tag{12}$$

3.2. *Dual characterization of smoothness and regularity.* The following well-known fact can be seen as dual characterization of κ-smoothness:

PROPOSITION 3.2. *Let $(E, \|\cdot\|)$ be a finite-dimensional normed space, E^* be the space dual to E, $\|\cdot\|_*$ be the norm on E^* dual to $\|\cdot\|$; and let $\langle\xi, x\rangle$ stand for the value of a linear form $\xi \in E^*$ 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 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*

$$\langle f'(x) - f'(y), x - y\rangle \le \kappa\|x-y\|^2 \ \ \forall x, y \in E; \tag{13}$$

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

$$\|f'(x) - f'(y)\|_* \le \kappa\|x-y\| \ \ \forall x, y \in E; \tag{14}$$

(iv) *One has*

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

(v) *One has*

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

(vi) *One has*

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

Another characterization of regular spaces is as follows:

Proposition 3.3. *Let* $(E, \|\cdot\|)$ *be a finite-dimensional normed space,* E^* *be the space dual to* E, $\|\cdot\|_*$ *be the norm on* E^* *dual to* $\|\cdot\|$, *and let* $\langle \xi, x\rangle$ *stand for the value of a linear form* $\xi \in E^*$ *on a vector* $x \in E$. *Let also* B_* *be the unit* $\|\cdot\|_*$*-ball of* E^*.

(i) *If* $(E, \|\cdot\|)$ *is* κ*-regular, then the exists a continuous function* $V : B_* \to \mathbf{R}$ *which is strongly convex, with coefficient* 1 *w.r.t.* $\|\cdot\|_*$, *on* B_*, *that is, possesses the following equivalent to each other properties:*

(15)

(*a*) $\forall(\xi,\ \eta \in \mathrm{int}B_*,\ x \in \partial v(\xi), y \in \partial v(\eta)) : \ \ \langle \xi - \eta, x - y\rangle \ge \|\xi - \eta\|_*^2,$

(*b*) $\forall(\xi, \eta :\ \xi, \xi + \eta \in \mathrm{int}B_*,\ x \in \partial v(\xi)) : \ \ v(\xi + \eta) \ge v(\xi) + \langle \eta, x\rangle + \frac{1}{2}\|\eta\|_*^2;$

and, in addition, is such that

$$\max_{B_*} v - \min_{B_*} v \le \frac{\kappa}{2} \tag{16}$$

(ii) *Assume that the unit ball* B_* *of* $(E^*, \|\cdot\|_*)$ *admits a function* v *satisfying* (15), (16). *Then* $(E, \|\cdot\|)$ *is* $O(1)\kappa$*-regular with an appropriately chosen absolute constant* $O(1)$.

3.3. *"Calculus" of smooth and regular spaces.*

PROPOSITION 3.4. *Let $(E, \|\cdot\|_E)$ be a finite-dimensional normed space, L be a linear subspace of E, and $F = E/L$ be the factor-space of E equipped with the factor-norm $\|\bar{f}\|_F = \min_{f\in\bar{f}} \|f\|_E$. If $(E, \|\cdot\|_E)$ is κ-smooth (κ-regular), then $(L, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ also are κ-smooth, respectively, κ-regular.*

PROPOSITION 3.5. (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*

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

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

$$\kappa^+ = \min_{2\le\rho\le p} [\kappa + \rho - 1] m^{\frac{2}{\rho}-\frac{2}{p}} \le \min[\kappa + p - 1, [\kappa + 2\ln(m) - 1]\exp\{1\}]. \tag{17}$$

(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.

PROPOSITION 3.6. (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*

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

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

$$\kappa^{++} = 2 \min_{2\le\rho\le p} [\kappa + \rho - 1] m^{\frac{2}{\rho}-\frac{2}{p}} \le 2\min[\kappa + p - 1, [\kappa + 2\ln(m) - 1]\exp\{1\}]. \tag{18}$$

(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.

4. Appendix: Proofs.

4.1. *Proofs for Section 3.1.*

4.1.1. *Justifying the Examples.*

Example 3.2:. 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, (3) holds true if and only if

$$D^2p(x)[h,h] \le 2\kappa_+ p(h) \ \forall (x, h \in E, x \ne 0); \tag{19}$$

since $p(\cdot)$ is homogeneous of degree 2, the validity of (19) 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

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

as required in (19) when $\kappa_+ = \rho - 1$. In the case of $\rho = 2$ relation (19) 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 (9) follows. ∎

Example 3.3:. 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*

$$\kappa = \max[2, \rho - 1]. \tag{20}$$

Proof. The statement is evident when $\rho = 2$; thus, from now on we assume that $\rho > 2$. Let us apply Proposition 3.1 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 (11)). By Proposition, the function $F(X) = |X|_\rho^\rho$ on $\mathbf{S}^n$ is twice continuously differentiable, and

(21)
$$\forall X, H : 0 \le D^2F(X)[H,H] \le \theta_+ \operatorname{Tr}(f''(x)H^2),\ \theta_+ = \max\left[\frac{2}{\rho-1}, 1\right].$$

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

(22)
$$X \neq 0 \Rightarrow D^2p(X)[H,H] = \frac{2}{\rho}\underbrace{\left[\frac{2}{\rho} - 1\right]}_{<0}(F(X))^{\frac{2}{\rho}-2}(DF(X)[H])^2 + \frac{2}{\rho}(F(X))^{\frac{2}{\rho}-1}D^2F(X)[H,H]$$
$$\le \frac{2}{\rho}(F(X))^{\frac{2}{\rho}-1}\theta_+ \operatorname{Tr}(f''(x)H^2).$$

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 (22) we have

(23)
$$D^2p(X)[H,H] \le 2\Theta_+(\rho-1)\operatorname{Tr}(ZH^2) \le 2\theta_+(\rho-1)|Z|_p|H^2|_{\frac{p}{p-1}} = 2\theta_+(\rho-1)|H^2|_{\frac{\rho}{2}}$$
$$= 2\theta_+(\rho-1)|H|_\rho^2.$$

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

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

and (23) implies that for the outlined X, Y one has

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

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 3.3. W.l.o.g. we may assume that $m \le n$. 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 $\kappa = \min\limits_{\substack{2\leq\rho<\infty\\ \rho\leq p}} \max[2,\rho-1]m^{\frac{2}{\rho}-\frac{2}{p}}$, and we arrive at (10). ∎

4.1.2. *Proof of Proposition 3.1.* 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 $P(X) = \mathrm{Tr}(\sum_j p_j X^j)$ 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, $X = U\,\mathrm{Diag}\{\lambda\}U^T$ be the eigenvalue decomposition of X, and let $\widehat{H}$ be such that $H = U\widehat{H}U^T$. We have

(24)
$$\begin{array}{rcll} P(X) & = & \sum_{s=1}^n p(\lambda_s(X)) & (a)\\ DP(X)[H] & = & \mathrm{Tr}(\sum_{j=1}^N\sum_{s=0}^{N-1} X^s H X^{N-s-1} = \mathrm{Tr}(p'(X)H) = \sum_{s=1}^n p'(\lambda_s(X))\widehat{H}_{ss} & (b)\end{array}$$

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

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

Computing the residuals, we get

$$D^2P(X)[H,H] = \sum_{s,t}\Gamma_{s,t}[p]\widehat{H}_{st}^2, \quad \Gamma_{s,t}[p] = \begin{cases} \frac{p'(\lambda_s)-p'(\lambda_t)}{\lambda_s-\lambda_t}, & \lambda_s \neq \lambda_t\\ p''(\lambda_s), & \lambda_s = \lambda_t\end{cases} \tag{25}$$

Substituting $p = f_k$ into (24.a,b) and (25), 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 (24.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

$$\begin{array}{rcl} DF(X)[H] & = & \sum_s f'(\lambda_s)\widehat{H}_{ss} = \mathrm{Tr}(f'(X)H)\\ D^2F(X)[H,H] & = & \sum_{s,t}\Gamma_{s,t}[f]\widehat{H}_{st}^2\end{array} \tag{26}$$

So far, we did not use (11). Invoking the right inequality in (11), we get

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

which is the right inequality in (12). The derivation of the left inequality in (12) is similar. ■

4.1.3. *Proof of Proposition 3.2.*

(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

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

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

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

Thus, if $\|h\|=\|d\|=1$, then

$$4D^2f_k(x)[h,d] = D^2f_k(x)[h+d,h+d]-D^2f_k(x)[h-d,h-d] \le \kappa\|h+d\|^2 \le 4\kappa$$

. Whence $D^2f_k(x)[h,d]\le\kappa$ whenever $\|h\|=\|d\|=1$, or, which is the same by homogeneity,

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

Consequently,

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

whence, taking maximum over h with $\|h\|=1$,

$$\|f_k'(y)-f_k'(x)\|_* \le \kappa\|y-x\|$$

. As $k\to\infty$, $f_k'(x)$ converge to $f'(x)$, and we conclude that $f'(\cdot)$ possesses the required Lipschitz continuity. ■

(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

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

which immediately implies (3) (recall that $\|\cdot\|^2 = 2f(\cdot)$). ■

(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),

$$\|\xi - \eta\|_* \leq \kappa \|x - y\|,$$

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

$$\|\xi - \eta\|_* \leq \kappa \|x - y\|.$$

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. ■

(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

$$\langle \xi - \eta, x - x\rangle \geq \kappa^{-1} \|\xi - \eta\|_*^2,$$

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),

$$\langle \xi - \eta, x - y\rangle \geq \kappa^{-1} \|\xi - \eta\|_*^2.$$

Since

$$\langle \xi - \eta, x - y\rangle \leq \|\xi - \eta\|_* \|x - y\|,$$

we get

$$\|\xi-\eta\|_*\|x-y\| \geq \kappa^{-1}\|\xi-\eta\|_*^2,$$

whence

$$\|\xi-\eta\|_* = \|f'(x)-f'(y)\|_* \leq \kappa\|x-y\|,$$

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 $\langle f'(x)-f'(y), x-y\rangle \geq \kappa^{-1}\|f'(x)-f'(y)\|^2$. Setting

$$g(u) = f(u) - \langle f'(y), u-y\rangle,$$

we get a continuously differentiable convex function on E such that

$$\|g'(x)-g'(y)\|_* \leq \kappa\|x-y\|$$

and $g'(y)=0$. Due to these relations,

$$g(y+h) \leq g(y) + \frac{\kappa}{2}\|h\|^2$$

for all h. Now let $e \in E$ be such that $\langle g'(x), e\rangle = \|g'(x)\|_*$ and $\|e\|=1$. Due to

$$\|g'(u)-g'(v)\|_* \leq \kappa\|u-v\|,$$

we have

$$\begin{aligned} g(x-\frac{\|g'(x)\|_*}{\kappa}e) &\leq g(x) - \langle g'(x), \frac{\|g'(x)\|_*}{\kappa}e\rangle + \frac{\kappa}{2}\|\frac{\|g'(x)\|}{\kappa}e\|^2 \\ &= g(x) - \frac{\|g'(x)\|_*^2}{\kappa} + \frac{\|g'(x)\|_*^2}{2\kappa} = g(x) - \frac{\|g'(x)\|_*^2}{2\kappa}. \end{aligned}$$

On the other hand, g attains its global minimum at y, so that

$$g(x) - \frac{\|g'(x)\|_*^2}{2\kappa} \geq g(x-\frac{\|g'(x)\|_*}{\kappa}e) \geq g(y).$$

We now have

$$\begin{aligned} g(y) + \frac{\kappa}{2}\|h\|^2 &\geq g(y+h) \geq g(x) + \langle g'(x), y+h-x\rangle \\ &\geq g(y) + \frac{\|g'(x)\|_*^2}{2\kappa} + \langle g'(x), y+h-x\rangle, \end{aligned}$$

whence

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

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,

$$\langle f'(x)-f'(y), x-y\rangle = \langle g'(x), x-y\rangle \geq \frac{\|g'(x)\|_*^2}{\kappa} = \frac{\|f'(x)-f'(y)\|_*^2}{\kappa}. \quad \blacksquare$$

(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

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

where the inequality is given by (iv). We end up with the inequality required in (vi). ■

(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

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

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

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

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. ■

4.1.4. *Proof of Proposition 3.3.*

(i):. Let $(E, \|\cdot\|)$ be κ-regular, and let $\kappa_+ \in [1, \kappa]$ and $\|\cdot\|_+$ be such that $(E, \|\cdot\|_+)$ is κ-smooth and (4) holds true, and let $\|\cdot\|_{+,*}$ be the norm on E^* dual to $\|\cdot\|_+$; note that

$$\frac{\kappa_+}{\kappa}\|\cdot\|_*^2 \leq \|\cdot\|_{+,*}^2 \leq \|\cdot\|_*^2 \tag{27}$$

due to (4). Invoking Proposition 3.2, the function $v(\xi) = \frac{\kappa}{2}\|\xi\|_{*,+}^2 : B_* \to \mathbf{R}$ satisfies

$$\forall(\xi, \eta \in \mathrm{int}B_*,\ x \in \partial v(\xi),\ y \in \partial v(\eta)) :\ \langle \xi - \eta, x - y\rangle \geq \frac{\kappa}{\kappa_+}\|\xi - \eta\|_{+,*}^2,$$

and thus satisfies (15.a) due to (27). At the same time,

$$\max_{B_*} v - \min_{B_*} v = \frac{\kappa}{2}\max_{\|\xi\|_* \leq 1}\|\xi\|_{+,*}^2 \leq \frac{\kappa}{2},$$

where the concluding inequality is due to (27). (i) is proved.

(ii):. Let $v(\cdot)$ satisfy (15) and (16); clearly, the function $\frac{1}{2}[v(\xi) + v(-\xi)] - v(0)$ also satisfy these relations; thus, we can assume w.l.o.g. that $v(\xi) = v(-\xi)$ and $v(0) = 0$. Let V be the Legendre transform of $v(\cdot)\Big|_{B_*}$, that is,

$$V(x) = \max_{\|\xi\|_* \leq 1}\left[\langle \xi, x\rangle - v(x)\right].$$

By the standard properties of the Legendre transform, (15) implies that V is a continuously differentiable convex function on E such that

$$V'(x) = \operatorname*{argmin}_{\xi \in B_*}\left[\langle \xi, x\rangle - v(\xi)\right] \in B_* \text{ and } \|V'(x) - V'(y)\|_* \leq \|x - y\|\ \forall x, y.$$

In addition, we clearly have $V(x) = V(-x)$ and $\|x\| - \frac{\kappa}{2} \leq V(x) \leq \|x\|$ for all x by (16). Convolving V with a smooth symmetric w.r.t. the origin nonnegative kernel with unit integral and small support and subtracting a constant to make function vanish at the origin, we see that for every $\epsilon > 0$ there exists a C^∞ convex function $W = W_\epsilon$ on E such that for all $x \in E$ one has

$$\begin{array}{ll} (a) & W_\epsilon(x) = W_\epsilon(-x),\ W_\epsilon(0) = 0; \\ (b) & \|x\| - \frac{\kappa}{2} - \epsilon \leq W_\epsilon(x) \leq \|x\| + \epsilon \\ (c) & \|W'(x)\|_* \leq 1 \\ (d) & 0 \leq \langle W''(x)dx, dx\rangle \leq \|dx\|^2\ \forall dx \in E. \end{array} \tag{28}$$

Assuming $\epsilon \leq \kappa/10$, let us set $B = \{x : W(x) \leq \kappa\}$. Then B is a closed convex set symmetric w.r.t. the origin and such that

$$\{x : \|x\| \leq \frac{9}{10}\kappa\} \subset B \subset \{x : \|x\| \leq \frac{5}{2}\kappa\} \tag{29}$$

due to (28.b). B is the unit ball of certain norm $r(x)$ on E; by (29) we have

$$\frac{2}{5}\|x\| \leq \kappa r(x) \leq \frac{10}{9}\|x\|. \tag{30}$$

Setting $L(x) = p^2(x)$, observe that the function L is given by the equation

$$V(x/\sqrt{L(x)}) = \kappa.$$

It follows immediately from the Implicit Function Theorem that L is C^∞ outside of the origin, and since this function is the square of a norm, it is therefore C^1 on the entire space. Let us compute the second order differential of L at a point $x \neq 0$. Differentiating twice the equation specifying L, we get

$$\begin{aligned} DL(x)[dx] &= 2L\frac{\langle W', dx\rangle}{\langle W', x\rangle}, \\ D^2L(x)[dx, dx] &= 2L\left[\frac{\langle W', dx\rangle}{\langle W', x\rangle}\right]^2 + \frac{2L^{1/2}}{\langle W', x\rangle}\langle W''\left[dx - \frac{\langle W', dx\rangle}{\langle W', x\rangle}x\right], \left[dx - \frac{\langle W', dx\rangle}{\langle W', x\rangle}x\right]\rangle, \\ \text{where} \quad & L = L(x), W' = W'(L^{-1/2}x), W'' = W''(L^{-1/2}x). \end{aligned}$$

We claim that

$$x \neq 0 \Rightarrow 0 \leq D^2L(x)[dx, dx] \leq \frac{27}{\kappa}\|dx\|^2. \tag{31}$$

Indeed, $D^2L(x)[dx, dx]$ is homogeneous of degree 0 in x, so that it suffices to verify the required relation when $L(x) = 1$, i.e., when $W(x) = \kappa$. In this case, the required bound is readily given by the expression for D^2L combined with (28.c, d) and the following observations: (1) for x in question, we have $\langle W', x\rangle \geq W(x) - W(0) = \kappa$, and (2) $\|x\| \leq \frac{5}{2}\kappa$ by (29).

Setting $\|x\|_+ = \frac{5}{2}\kappa r(x)$ and invoking (29), we have

$$\|\cdot\|^2 \leq \|\cdot\|_+^2 \leq O(1)\|\cdot\|^2, \tag{32}$$

while from (31) it follows that the function $f(x) = \|x\|_+^2$ satisfies

$$\|f'(x) - f'(y)\|_* \leq O(1)\kappa\|x - y\|,$$

which combines with (32) to imply that

$$\|f'(x) - f'(y)\|_{+,*} \leq O(1)\kappa\|x - y\|_*.$$

Thus, $(E, \|\cdot\|)$ is $O(1)\kappa$-smooth, whence, by (32), $(E, \|\cdot\|)$ is $O(1)\kappa$-regular. ■

4.1.5. *Proof of Proposition 3.4.* The fact that a subspace of a κ-smooth/regular space equipped with the induced norm is κ-smooth/regular is evident. As about the factor-space $F = E/L$, note that the space dual to $(F, \|\cdot\|_F)$ is nothing but the subspace $L^\perp = \{\xi : \langle\xi, x\rangle = 0 \forall x \in L\}$ in E^* equipped by the norm induced by $\|\cdot\|_*$. Now assume that $(E, \|\cdot\|_E)$ is κ-smooth. By Proposition 3.2, it follows that $\|\cdot\|_*$ possesses property (iv) and therefore its restriction on $L^\perp$ possesses the same property. Applying Proposition 3.2 again, we conclude that $(F, \|\cdot\|_F)$ is κ-smooth. We see that passing to a factor-space preserves κ-smoothness, and since this transformation preserves also relations like (4), it preserves κ-regularity as well. ■

4.1.6. *Proof of Proposition 3.5.*

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

A.. Let $\rho \in [2, \infty)$ be such that $\rho \leq p$, 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

$$\kappa_+ = \kappa + \rho - 2 \tag{33}$$

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, $\rho \geq 2$, whence $r \geq 1$, 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 3.2, in order to prove Lemma 2 it suffices to verify that

$$\|p'(x) - p'(y)\|_* \leq 2\kappa_+\|x - y\| \tag{34}$$

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

$$\|p'(x') - p'(y')\|_* \leq 2\kappa_+\|x' - y'\|$$

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 (34) 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 3.2, 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

$$|D^2 p_i^t(u^i)[h^i, h^i]| \leq 2\kappa\|h^i\|_i^2 \ \forall (u^i, h^i \in E_i). \tag{35}$$

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

$$\langle p'(y) - p'(x), h\rangle = \lim_{t\to\infty} \int_0^1 D^2 p^t(x + t(y - x))[y - x, h]dt. \tag{36}$$

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

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

whence
(37)
$$\begin{aligned}&0 \le D^2p^t(u)[v,v]\\ &\le F^{\frac{1}{r}-1}(p_1^t(u^1),...,p_m^t(u^m))\sum_i\left[(r-1)(p_i^t(u^i))^{r-2}(Dp_i^t(u^i)[v_i])^2+2\kappa(p_i^t(u^i))^{r-1}p_i(v^i)\right].\end{aligned}$$

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 (37) converges, as $t\to\infty$, uniformly in $u\in[x,y]$ and v, $\|v\|\le 1$, 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, $|Dp_i(u^i)[v_i]|\le 2\|u^i\|\|v^i\|$, whence
(38)
$$\begin{aligned}\Psi(u,v) &\le \left(\sum_i\|u^i\|_i^\rho\right)^{\frac{2}{\rho}-1}\sum_i\left[4(r-1)\|u^i\|_i^{\rho-2}\|v_i\|_i^2+2\kappa\|u^i\|_i^{\rho-2}\|v^i\|_i^2\right]\\ &= \underbrace{[2\rho+2\kappa-4]}_{2\kappa_+}\left(\sum_i\|u^i\|_i^\rho\right)^{\frac{2}{\rho}-1}\sum_i\|u^i\|_i^{\rho-2}\|v^i\|_i^2\end{aligned}$$

When $\rho>2$, we have

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

and (38) implies that $\Psi(u,v)\le 2\kappa_+\|v\|^2$. 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

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

The resulting inequality via the same reasoning as in the proof of item "(i)⇒(iii)" of Proposition 3.2 implies that

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

In view of this bound and (36), we conclude that

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

for all h, whence $\|p'(y)-p'(x)\|_*\le(2\kappa_++\epsilon)\|y-x\|$. Since $\epsilon>0$ is arbitrary, we arrive at (34).

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, $\|(x^1, ..., x^m)\|_\dagger = \sum_i \|x^i\|_i$, 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 Proposition 3.4. ■

4.1.7. *Proof of Proposition 3.6.*

A useful lemma.. We start with 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*

$$\forall (x \in E) : \|x\|^2 \le \|x\|_+^2 \le 2\|x\|^2. \tag{39}$$

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,

$$\forall \xi \in E^* : \pi_*^2(\xi) \ge \|\xi\|_*^2 \ge \frac{1}{\mu}\pi_*^2(\xi), \tag{40}$$

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 (39). 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

$$\forall \xi \in E^* : q_*^2(\xi) \ge \|\xi\|_*^2 \ge \frac{1}{\gamma\mu + 1 - \gamma}q_*^2(\xi) = \frac{1}{2}q_*^2(\xi). \tag{41}$$

Further, by Proposition 3.2 we have

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

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

$$q_*^2(\xi+\eta)\geq q_*^2(\xi)+\langle\eta,x+y\rangle+\frac{\gamma}{\kappa_+}\pi_*^2(\eta)\geq q_*^2(\xi)+\langle\eta,x+y\rangle+\frac{\gamma}{\kappa_+}q_*^2(\eta)$$

(note that $\pi_*(\cdot)\geq q_*(\cdot)$ by (40)). Since $\partial\pi_*^2(\xi)+D(\xi)=\partial q_*^2(\xi)$ and $\frac{\gamma}{\kappa_+}=\frac{1}{(\mu-1)\kappa_+}\geq\frac{1}{\kappa}$, we get

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

By the same Proposition 3.2, 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, (41) implies (39). ■

Proof of Proposition 3.6. is readily given by Lemma 3 combined with the corresponding items of Proposition 3.5. 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 3.5.(i) to the spaces $(E_i,q_i(\cdot))$, we get that the norm $q(x^1,...,x^m)=\left(\sum\limits_{i=1}^m q_i^p(x^i)\right)^{1/p}$ on $E_1\times...\times E_m$ is κ^+-regular with κ^+ given by (17). 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. ■

4.2. *Proof of Theorem 2.1.*

4.2.1. *Reduction to the case of a smooth norm.* We intend to reduce the situation to the one where $(E,\|\cdot\|)$ is κ-smooth rather than κ-regular. Specifically, we are about to prove the following fact:

THEOREM 4.1. *Let $(E,\|\cdot\|)$ be κ-smooth, let E-valued martingale-difference ξ^∞ satisfy $(\mathcal{C}_\alpha[\sigma^\infty])$, and let $S_N=\sum_{i=1}^N\xi_i$, $\sigma^N=[\sigma_1;...;\sigma_N]$. Then*

(i) *When* $1 \le \alpha \le 2$*, one has for all* $N \ge 1$ *and* $\gamma \ge 0$*:*

(42)
$$\mathrm{Prob}\left\{\|S_N\| \ge \left[\sqrt{\exp\{1\}\kappa}+\gamma\right]\sqrt{\sum_{i=1}^N \sigma_i^2}\right\} \le 2\exp\left\{-\frac{1}{64}\min\left[\gamma^2;\gamma_*^{2-\alpha}\gamma^\alpha\right]\right\},$$

where

(43)
$$\gamma_* \equiv \gamma_*(\alpha,\sigma^N) \quad = \quad \begin{cases} 32\left[\frac{8\alpha_*}{2^{\alpha_*}}\right]^{\frac{\alpha-1}{2-\alpha}}\left[\frac{\|\sigma^N\|_2}{\|\sigma^N\|_{\alpha_*}}\right]^{\frac{\alpha}{2-\alpha}} \ge 16\left[\frac{\|\sigma^N\|_2}{\|\sigma^N\|_{\alpha_*}}\right]^{\frac{\alpha}{2-\alpha}} \ge 16, & 1<\alpha<2,\\ \alpha_* = \frac{\alpha}{\alpha-1}, & \\ \lim_{\alpha\to 1+0}\gamma_*(\alpha,\sigma^N) = 16\frac{\|\sigma^N\|_2}{\|\sigma^N\|_\infty}, & \alpha=1,\\ \lim_{\alpha\to 2-0}\gamma_*(\alpha,\sigma^N) = +\infty, & \alpha=2.\end{cases}$$

(ii) *When* $\alpha = 2$*, the bound* (42) *improves to*

$$(\forall N\ge 1,\gamma\ge 0): \mathrm{Prob}\left\{\|S_N\| \ge \left[\sqrt{\kappa}+\gamma\right]\sqrt{\sum_{i=1}^N \sigma_i^2}\right\} \le \exp\{-\gamma^2/3\}. \tag{44}$$

(iii) *When the condition* $\mathbf{E}_{i-1}\left\{\exp\{\|\xi_i\|^2\sigma_i^{-2}\}\right\} \le \exp\{1\}$ *in* $(\mathcal{C}_2[\sigma^\infty])$ *is strengthened to* $\|\xi_i\| \le \sigma_i$ *almost surely,* $i = 1,2,...$*, the bound* (42) *improves to*

$$(\forall N\ge 1,\gamma\ge 0): \mathrm{Prob}\left\{\|S_N\| \ge \left[\sqrt{\kappa}+\gamma\right]\sqrt{\sum_{i=1}^N \sigma_i^2}\right\} \le \exp\{-\gamma^2/2\}. \tag{45}$$

It is immediately seen that Theorem 4.1 implies Theorem 2.1. Indeed, if $(E,\|\cdot\|)$ is κ-regular, by Lemma 3 there exists a norm $\|\cdot\|_+$ on E such that $(E,\|\cdot\|_+)$ is κ-smooth and (39) holds true. Setting $\widehat{\sigma}_i = \sqrt{2}\sigma_i$, observe that (39) combines with $(\mathcal{C}_\alpha[\sigma^\infty])$ to imply that $\mathbf{E}_{i-1}\left\{\exp\{\|\xi_i\|_+^2\widehat{\sigma}_i^{-2}\}\right\} \le \exp\{1\}$. Applying Theorem 4.1.(i) to the κ-smooth space $(E,\|\cdot\|_+)$ and $\widehat{\sigma}_i$ in the role of σ_i and taking into account that $\|S_N\| \le \|S_N\|_+$, we see that Theorem 2.1.(i) is an immediate corollary of Theorem 4.1.(i), and similarly for Theorem 2.1.(ii-iii).

4.2.2. *Proof of Theorem 4.1: preliminaries.* In the sequel, we need the following (essentially, well-known) fact.

Proposition 4.1. *Let* ψ_i*,* $i = 1,...,N$*, be Borel functions on* Ω *such that* ψ_i *is* $\mathcal{F}_i$*-measurable, let* $\alpha \in [1,2]$*, and let* μ_i*,* $\nu_i > 0$ *be deterministic reals. Assume that almost surely one has*

$$\mathbf{E}_{i-1}\{\psi_i\} \le \mu_i, \mathbf{E}_{i-1}\left\{\exp\{|\psi_i|^\alpha/\nu_i^\alpha\}\right\} \le \exp\{1\},\ 1\le i\le N. \tag{46}$$

Then for every $\gamma \geq 0$ one has

$$\text{(47)}\quad \mathrm{Prob}\left\{\sum_{i=1}^N \psi_i > \sum_{i=1}^N \mu_i + \gamma\sqrt{\sum_{i=1}^N \nu_i^2}\right\} \leq 2\exp\{-\frac{1}{64}\min\left[\gamma^2, \gamma_*^{2-\alpha}\gamma^\alpha\right]\},$$

where

(48)

$$\gamma_* \equiv \gamma_*(\alpha,\nu^N) = \begin{cases} 32\left[\frac{8\alpha_*}{2^{\alpha_*}}\right]^{\frac{\alpha-1}{2-\alpha}}\left[\frac{\|\nu^N\|_2}{\|\nu^N\|_{\alpha_*}}\right]^{\frac{\alpha}{2-\alpha}} \geq 16\left[\frac{\|\nu^N\|_2}{\|\nu^N\|_{\alpha_*}}\right]^{\frac{\alpha}{2-\alpha}} \geq 16, & 1<\alpha<2,\\ \left[\alpha_* = \frac{\alpha}{\alpha-1}, \nu^N = [\nu_1;...;\nu_N]\right], & \\ \lim_{\alpha\to 1+0}\gamma_*(\alpha,\nu^N) = 16\frac{\|\nu^N\|_2}{\|\nu^N\|_\infty}, & \alpha=1,\\ \lim_{\alpha\to 2-0}\gamma_*(\alpha,\nu^N) = +\infty, & \alpha=2.\end{cases}$$

To make the text self-contained, here is the proof.

0^0.. Till item **4**^{0} of the proof, we restrict ourselves with the case when $1<\alpha<2$. Besides this, by evident homogeneity reasons we may assume w.l.o.g. that $\nu \equiv \sum_{i=1}^N \nu_i^2 = 1$.

1^0.. We start with the following

Lemma 4. *Let $\alpha \in (1,2)$, $\nu > 0$ and ψ be a real-valued random variable such that*

$$\text{(49)}\qquad \mathbf{E}\{\exp\{|\psi/\nu|^\alpha\}\} \leq \exp\{1\}.$$

Then

(50)

$$t \geq 0 \Rightarrow \ln\left(\mathbf{E}\{\exp\{t\psi\}\}\right) \leq t\mathbf{E}\{\psi\} + 8(t\nu)^2 + 2^{\alpha_*}\alpha_*^{-1}|t\nu|^{\alpha_*}, \quad \alpha_* = \frac{\alpha}{\alpha-1}.$$

Proof. 1) Let $t \geq 0$ be fixed. W.l.o.g. we can assume that $\nu = 1$. By Young inequality, we have

$$t\psi = (2t)(\psi/2) \leq \frac{|\psi/2|^\alpha}{\alpha} + \frac{(2t)^{\alpha_*}}{\alpha_*};$$

since $\alpha^{-1}(1/2)^\alpha < 1$ and $\nu = 1$, we have $\mathbf{E}\{\exp\{\alpha^{-1}|\psi/2|^\alpha\}\} \leq \exp\{\alpha^{-1}(1/2)^\alpha\}$, whence

$$\mathbf{E}\{\exp\{t\psi\}\} \leq \mathbf{E}\{\exp\{\alpha^{-1}|\psi/2|^\alpha + \alpha_*^{-1}(2t)^{\alpha_*}\}\} \leq \exp\{\alpha^{-1}(1/2)^\alpha + \alpha_*^{-1}(2t)^{\alpha_*}\}.$$

2) Let $f(t) = \mathbf{E}\{\exp\{t\psi\}\}$. Since $\alpha > 1$, f is a C^∞ function on the axis such that $f(0) = 1$, $f'(0) = \mathbf{E}\{\psi\}$ and

$$f''(t) = \mathbf{E}\left\{\exp\{t\psi\}\psi^2\right\}$$

It is easily seen that

$$0 \le t \le 1/4 \Rightarrow \exp\{t|s|\}s^2 \le \exp\{|s|^\alpha\}\,\forall s,$$

whence under the premise of Lemma 4 one has

$$0 \le t \le 1/4 \Rightarrow f''(t) \le \exp\{1\}$$

(recall that $\nu = 1$). It follows that

$$0 \le t \le 1/4 \Rightarrow f(t) \le 1 + t\mathbf{E}\{\psi\} + \frac{\exp\{1\}}{2}t^2 \le \exp\{t\mathbf{E}\{\psi\} + \frac{\exp\{1\}}{2}t^2\}.$$

Thus, one has

$$(51)\qquad \begin{array}{lrcl} (a) & 0 \le t \le 1/4 \Rightarrow \ln f(t) & \le & t\mathbf{E}\{\psi\} + \frac{\exp\{1\}}{2}t^2, \\ (b) & t \ge 0 \Rightarrow \ln f(t) & \le & \alpha^{-1}(1/2)^\alpha + \alpha_*^{-1}(2t)^{\alpha_*}. \end{array}$$

Since $8t^2 \ge \frac{\exp\{1\}}{2}t^2$ and $8t^2 \ge \alpha^{-1}(1/2)^\alpha$ when $t \ge 1/4$, (51) implies (50).

2^0.. Since $\alpha > 1$, we have for all $t \ge 0$

$$\begin{array}{l} \mathbf{E}\left\{\exp\{t\sum_{i=1}^n \psi_j\}\right\} = \mathbf{E}\left\{\exp\{t\sum_{i=1}^{n-1}\psi_j\}\mathbf{E}_{n-1}\{\exp\{t\psi_n\}\}\right\} \\ \le \mathbf{E}\left\{\exp\{\exp\{t\sum_{i=1}^{n-1}\psi_j\}\right\}\exp\{\mu_n t + 8(t\nu_n)^2 + \alpha_*^{-1}2^{\alpha_*}(t\nu_n)^{\alpha_*}\}, \end{array}$$

whence

$$\begin{array}{l} \ln\left(\mathbf{E}\{t\sum_{i=1}^N \psi_i\}\right\} \le A_N t + B_N t^2 + C_N t^{\frac{\alpha}{\alpha-1}}, \\ A_N = \sum_{i=1}^N \mu_i,\ B_N = 8\sum_{i=1}^N \nu_i^2,\ C_N = \alpha_*^{-1}2^{\alpha_*}\sum_{i=1}^N \nu_i^{\alpha_*}. \end{array}$$

3^0.. Recall that we are in the situation $\sum_{i=1}^N \nu_i^2 = 1$. We have for all $t > 0$:

$$\begin{array}{rcl} \mathrm{Prob}\{\Psi_N > A_N + \gamma\nu\} & = & \mathrm{Prob}\{\exp\{t\Psi_N\} > \exp\{tA_N + t\gamma\}\} \\ & \le & \mathbf{E}\{\exp\{t\Psi_N\}\}\exp\{-tA_N - t\gamma\} \le \exp\{B_N t^2 + C_N t^{\frac{\alpha}{\alpha-1}} - t\gamma\}, \end{array}$$

whence

$$\mathrm{Prob}\{\Psi_N > A_N + \gamma\} \le \inf_{t>0}\exp\{B_N t^2 + C_N t^{\frac{\alpha}{\alpha-1}} - t\gamma\}.$$

whence also

$$\ln\left(\mathrm{Prob}\{\Psi_N > A_N + \gamma\}\right) \le \ln(2) + \inf_{t>0}\big[\underbrace{\max[2B_N t^2, 2C_N t^{\frac{\alpha}{\alpha-1}}]}_{\phi(t)} - \gamma t\big] \equiv \ln(2) - \phi_*(\gamma),$$

where ϕ_* is the Legendre transform of ϕ, $\mathrm{Dom}\phi = [0,\infty)$. Let $t_* = t_*(\alpha)$ be the unique positive root of the equation $B_N t^2 = C_N t^{\alpha_*}$, that is,

$$t_* = (B_N/C_N)^{\frac{\alpha-1}{2-\alpha}}.$$

The function $\phi(t)$ is strongly convex on $[0,\infty)$, equals $2B_N t^2$ to the left of t_* and equals $2C_N t^{\alpha_*}$ to the right of t_*. Let $\gamma_- = \gamma_-(\alpha)$ be the left, and $\gamma_+ = \gamma_+(\alpha)$ be the right derivative of ϕ at t_*, so that

$$4B_N t_* = \gamma_- \le \gamma_+ = 2C_N \alpha_* t_*^{\frac{1}{\alpha-1}}.$$

The function $\phi_*(\gamma)$ is as follows: since ϕ is strongly convex on $[0,\infty)$, $\phi'(0) = 0$ and $\phi(t)/t \to \infty$ as $t \to \infty$, ϕ_* is continuously differentiable and convex on $[0,\infty)$; when $0 \le \gamma \le \gamma_-$, ϕ_* coincides with the Legendre transform $\phi_{*,-}(\gamma) = \frac{1}{8B_N}\gamma^2$ of the function $2B_N t^2$ on the axis; when $\gamma \ge \gamma_+$, ϕ_* coincides with the Legendre transform $\phi_{*,+}(\gamma) = \frac{(2C_N)^{1-\alpha}}{\alpha}\gamma^\alpha$ of the function $2C_N|t|^{\alpha_*}$ on the axis. In the segment $[\gamma_-,\gamma_+]$ ϕ_* is linear with the slope $\phi'_{*,-}(\gamma_-) = \phi'_{*,+}(\gamma_+) = t_*$. Now let $\theta = \phi_{*,-}(\gamma_-)/\phi_{*,+}(\gamma_-)$, and let $\omega(\gamma) = \theta\phi_{*,+}(\gamma)$. Observe that $\omega(\gamma) \le \phi_*(\gamma)$ when $\gamma \ge \gamma_-$.

Indeed, at the point γ_+ the functions $\phi_{*,+}$ and ϕ_* have equal values and equal derivatives, and since ϕ_* is linear in $\Delta = [\gamma_-,\gamma_+]$, we conclude from convexity of $\phi_{*,+}(\cdot)$ that $\phi_{*,+}(\gamma) \ge \phi_*(\gamma)$ on Δ, while $0 \le \phi'_{*,+}(\gamma) \le \phi'_*(\gamma) \equiv \phi'_{*,+}(\gamma_+)$ on Δ. Therefore $\theta \le 1$, and since ϕ'_* is nondecreasing, we have $\omega'(\gamma) \le \phi'_*(\gamma)$ on Δ. Since $\omega(\gamma_-) = \phi_*(\gamma_-)$, we conclude that $\omega \le \phi_*$ everywhere on Δ. Since $\theta < 1$ and $\phi_{*,+}$ is positive, when $\gamma \ge \gamma_+$ we have $\omega(\gamma) \le \phi_{*,+}(\gamma) = \phi_*(\gamma)$.

The bottom line is that

$$\phi_*(\gamma) \ge \begin{cases} \frac{1}{8B_N}\gamma^2 & ,0 \le \gamma \le \gamma_- \\ D_N\gamma^\alpha & ,\gamma \ge \gamma_- \end{cases}, \; D_N = \frac{\phi_{*,-}(\gamma_-)}{\gamma_-^\alpha}$$

Recalling the definition of A_N, B_N. C_N, we arrive at (47) – (48).

4^0.. We have proved the assertion of Proposition in the case of $1 < \alpha < 2$. This combines with the standard approximation arguments to yield the assertion in the cases of $\alpha = 1$ and $\alpha = 2$.

4.2.3. *Completing the proof of Theorem 4.1.*

1^0*: Preparations..* Given κ-smooth space $(E, \|\cdot\|)$, let us set

$$V(\xi) = \begin{cases} \frac{1}{2}\|\xi\|^2 & ,\|\xi\| \le 1 \\ \|\xi\| - \frac{1}{2} & ,\|\xi\| \ge 1 \end{cases}, \quad V_\beta(\xi) = \beta V(\xi/\beta) \quad [\beta > 0], \quad v(x) = \frac{1}{2}\|x\|_*^2.$$

Observe that

1. $V_\beta(\cdot)$ is the Legendre transform of the restriction of $\beta v(\cdot)$ on the $\|\cdot\|_*$-unit ball, whence $\|V'_\beta(\xi)\|_* \leq 1$ for all $\beta > 0$ and all ξ, and

$$\|x\|_* \leq 1 \Rightarrow \langle x, \xi\rangle \leq \beta v(x) + V_\beta(\xi) \leq \frac{\beta}{2} + V_\beta(\xi)\,\forall \xi. \tag{52}$$

2. $V(\cdot)$ is continuously differentiable with $\|V'(\xi) - V'(\eta)\|_* \leq \kappa\|\xi - \eta\|$ and is Lipschitz continuous, with constant 1, w.r.t. $\|\cdot\|$;

   The second claim is evident. To prove the first, note that the function $v(\cdot)$ on the entire $\mathbf{R}^n$ is strongly convex w.r.t. $\|\cdot\|_*$ with parameter κ^{-1}, whence, of course, so is the function $\hat{v}$ which is equal to v in the unit ball and is $+\infty$ outside of this ball. Given ξ, η and setting $x = V'(\xi)$, $y = V'(\eta)$, we have $\xi \in \partial\hat{v}(\xi)$, $\eta \in \partial\hat{v}(y)$, whence

$$\|\xi - \eta\|\|x - y\|_* \geq \langle x - y, \xi - \eta\rangle \geq \kappa^{-1}\|x - y\|_*^2,$$

   so that

$$\|V'(\xi) - V'(y)\|_* = \|x - y\|_* \leq \kappa\|\xi - \eta\|.$$

3. One has

$$\begin{array}{ll} (a) & |V_\beta(\xi + \eta) - V_\beta(\xi)| \leq \|\eta\| \\ (b) & V_\beta(\xi + \eta) - V_\beta(\xi) \leq \langle V'_\beta(\xi), \eta\rangle + \frac{\kappa}{2\beta}\|\eta\|^2. \end{array} \tag{53}$$

   It clearly suffices to consider the case of $\beta = 1$, that is, $V_\beta \equiv V$. By the second claim in item 2, V is Lipschitz continuous with constant 1 w.r.t. the norm $\|\cdot\|$, which implies (53.a). Relation (53.b) is readily given by the Lipschitz continuity of V', see the first claim in item 2.

2^0: Proof of Theorem 4.1.(i).. Let us fix $\beta > 0$ and set

$$S_n = \sum_{i=1}^n \xi_i,\ a_n = V'_\beta(S_{n-1}),\ \psi_n = V_\beta(S_n) - V_\beta(S_{n-1}),$$

so that a_n is $\mathcal{F}_{n-1}$-measurable, and ψ_n is $\mathcal{F}_n$-measurable. By (53.a) we have $|\psi_n| \leq \|\xi_n\|$, whence

$$\mathbf{E}_{n-1}\left\{\exp\{|\psi_n|^\alpha/\sigma_n^\alpha\}\right\} \leq \exp\{1\}, \tag{54}$$

while by (53.b) we have

$$\begin{array}{l} \mathbf{E}_{n-1}\{\psi_n\} \leq \mathbf{E}_{n-1}\left\{\langle a_n, \xi_n\rangle + \frac{\kappa}{2\beta}\|\xi_n\|^2\right\} = \mathbf{E}_{n-1}\left\{\langle a_n, \xi_n\rangle + \frac{\kappa}{2\beta}\|\xi_n\|^2\right\} \\ = \mathbf{E}_{n-1}\left\{\frac{\kappa}{2\beta}\|\xi_n\|^2\right\} \quad [\text{since } a_n \text{ is } \mathcal{F}_{n-1}\text{-measurable and } \mathbf{E}_{n-1}\{\xi_n\} = 0] \\ \leq \frac{\kappa}{2\beta}\sigma_n^2 \exp\{1\}. \end{array}$$

The concluding inequality above can be justified as follows: setting $\zeta_n = \|\xi_n\|/\sigma_n$, we have $\mathbf{E}_{n-1}\{\exp\{\zeta_n^\alpha\}\} \leq \exp\{1\}$. At the same time, it is immediately seen that

$$s^2 \leq (\alpha \exp\{1\}/2)^{-2/\alpha} \exp\{|s|^\alpha\}$$

for all s, and since $(\alpha \exp\{1\}/2)^{-2/\alpha} \leq 1$ when $1 \leq \alpha \leq 2$, we get $\mathbf{E}_{n-1}\{\zeta_n^2\} \leq \mathbf{E}_{n-1}\{\exp\{|\zeta_n|^\alpha\}\}$. Thus, we arrive at

$$\mathbf{E}_{n-1}\{\psi_n\} \leq \mu_n := \exp\{1\}\sigma_n^2. \tag{55}$$

Invoking (52), we get

$$\|S_N\| \leq \frac{\beta}{2} + V_\beta(S_N) = \frac{\beta}{2} + \sum_{i=1}^N \psi_i.$$

Taking into account (54), (55) and applying Proposition 4.1, we arrive at

$$\forall \gamma \geq 0 : \mathrm{Prob}\left\{\|S_N\| \geq \left[\frac{\beta}{2} + \frac{\kappa \exp\{1\}\sum_{i=1}^N \sigma_i^2}{2\beta}\right] + \gamma\sqrt{\sum_{i=1}^N \sigma_i^2}\right\} \leq 2\exp\{-\tfrac{1}{64}\min[\gamma^2, \gamma_*^{2-\alpha}\gamma^\alpha]\},$$

with $\gamma_* = \gamma_*(\alpha, \sigma^N)$ given by (48). Optimizing this bound in $\beta > 0$, we arrive at (42). Theorem 4.1.(i) is proved.

3^0: Proof of Theorem 4.1.(ii-iii).. These results are given by exactly the same reasoning as above, with the role of Proposition 4.1 played by the following statement:

PROPOSITION 4.2. *Let ψ_i, $i = 1, ..., N$, be Borel functions on Ω such that ψ_i is $\mathcal{F}_i$-measurable, and let $\mu_i \geq 0$, $\nu_i > 0$ be deterministic reals. Assume that almost surely one has*

$$\forall i : \mathbf{E}_{i-1}\{\psi_i\} \leq \mu_i,$$

and either

$$\forall i : \mathbf{E}_{i-1}\left\{\exp\{\psi_i^2/\nu_i^2\}\right\} \leq \exp\{1\}, \tag{56}$$

or

$$\forall i : |\psi_i| \leq \nu_i. \tag{57}$$

Then for every $\gamma \geq 0$ one has

$$\mathrm{Prob}\left\{\sum_{i=1}^N \psi_i > \sum_{i=1}^N \mu_i + \gamma\sqrt{\sum_{i=1}^N \nu_i^2}\right\} \leq \begin{cases} \exp\{-\gamma^2/3\}, & \text{case of (56)} \\ \exp\{-\gamma^2/2\}, & \text{case of (57)} \end{cases}. \tag{58}$$

Proof. Let (56) be the case. It is immediately seen that $\exp\{s\} \leq s + \exp\{9s^2/16\}$ for all s. We conclude that if $0 \leq t \leq \frac{4}{3\nu_i}$, then

$$\begin{aligned} \mathbf{E}_{i-1}\left\{\exp\{t\psi_i\}\right\} &\leq t\mu_i + \mathbf{E}_{i-1}\left\{\exp\{9t^2\psi_i^2/16\}\right\} \\ &\leq t\mu_i + \exp\{9t^2\nu_i^2/16\} \leq \exp\{t\mu_i + 9t^2\nu_i^2/16\}. \end{aligned} \tag{59}$$

Besides this, we have $tx \leq \frac{3t^2\nu_i^2}{8} + \frac{2x^2}{3\nu_i^2}$, so that

$$\mathbf{E}_{i-1}\left\{\exp\{t\psi_i\}\right\} \leq \exp\left\{\frac{3t^2\nu_i^2}{8} + \frac{2}{3}\right\},$$

and the latter quantity is $\leq \exp(\frac{3t^2\nu_i^2}{4})$ when $t \geq \frac{4}{3\nu_i}$. Invoking (59), we arrive at

$$t \geq 0 \Rightarrow \mathbf{E}_{n-1}\left\{\exp\{t\phi_n\}\right\} \leq \exp\{t\mu_i + 3t^2\nu_n^2/4\}. \tag{60}$$

It follows that

$$\mathbf{E}\exp\left\{t\sum_{i=1}^n \psi_i\right\} = \mathbf{E}\left\{\mathbf{E}_{n-1}\left\{\exp\left\{t\sum_{i=1}^n \psi_i\right\}\right\}\right\} \leq \mathbf{E}\left\{\exp\left\{t\sum_{i=1}^{n-1} \psi_i\right\}\right\}\exp(t\mu_n + 3t^2\nu_n^2/4),$$

whence

$$t \geq 0 \Rightarrow \mathbf{E}\left\{\exp\{t\textstyle\sum_{i=1}^N \psi_i\}\right\} \leq \exp\left\{t\textstyle\sum_{i=1}^N \mu_i + \frac{3t^2}{4}\sum_{i=1}^N \nu_i^2\right\}.$$

Therefore for $\gamma \geq 0$ we get

$$\begin{aligned} &\mathrm{Prob}\left\{\textstyle\sum_{i=1}^N \psi_i > \sum_{i=1}^N \mu_i + \gamma\sqrt{\sum_{i=1}^N \nu_i^2}\right\} \\ &\leq \min_{t>0}\left[\mathbf{E}\left\{\exp\{t\textstyle\sum_{i=1}^N \psi_i\}\right\}\exp\{-t\sum_{i=1}^N \mu_i - t\gamma\sqrt{\sum_{i=1}^N \nu_i^2}\}\right] \\ &\leq \min_{t>0}\exp\{t\textstyle\sum_{i=1}^N \mu_i + \frac{3t^2}{4}\sum_{i=1}^N \nu_i^2 - t\sum_{i=1}^N \mu_i - t\gamma\sqrt{\sum_{i=1}^N \nu_i^2}\} = \exp\{-\gamma^2/3\} \end{aligned},$$

as required in the first bound in (58). In the case of (57), by Azuma-Hoeffding's inequality [1], we have

$$\forall t \geq 0 : \mathbf{E}_{n-1}\left\{\exp\{t\phi_i\}\right\} \leq \exp\{t\mu_i + \sigma_i^2/2\};$$

with this relation in the role of (60), the above reasoning results in the second bound in (58). ■

REFERENCES

[1] Azuma, K. Weighted sums of certain dependent random variables. *Tōkuku Math. J.*, **19** (1967), 357-367.

[2] John, F., "Extremum problems with inequalities as subsidiary conditions" – in: *Studies and Essays presented to R. Courant on his 60th Birthday, January 8, 1948.* Intersciense Piblichers Inc., NY 1948, 187-204.

[3] Garling, D.J.H., "Functional Central Limit Theorems in Banach Spaces" – *Annals of Probability* **4** (1976), 600–911.

[4] Garling, D.J.H., "Convexity, smoothness and martingale inequalities" – *Israel Journal of Mathematics* **28** (1978), 189-198.

[5] Juditsky, A., and Nemirovski, A., "Functional aggregation for nonparametric regression" – *Annals of Stat.* **28:3** (2000), 681-712.

[6] Juditsky, A., Lan, G., Nemirovski, A., Shapiro, A., "Stochastic Approximation approach to Stochastic Programming" – To appear in *SIAM J. Optim.*.

[7] Nemirovski, A., and Yudin, D. *Problem complexity and method efficiency in optimization.* - Nauka Publishers, Moscow, 1978 (in Russian); English translation: John Wiley & Sons, 1983.

[8] Nemirovskii, A., Polyak, B., and Tsybakov, A., "Rate of convergence for nonparametric maximum likelihood estimates." (in Russian) - *Problemy Peredachi Informacii*, **21:4** (1985) (English translation: *Problems of Information Transmission*)

[9] Nemirovskii, A. "Nonparametric estimation of functions satisfying differential inequalities." - in: R. Khasminskii, Ed. *Advances in Soviet Mathematics, v. 12: "Selected Topics in Nonparametric Statistics"* - American Mathematical Society, Providence, 1992.

[10] 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.

[11] Nemirovski, A. (2004), *Regular Banach spaces and large deviations of random sums.* – Working Paper,
E-print: http://www2.isye.gatech.edu/∼nemirovs/LargeDev2004.pdf

[12] Nemirovski, A., "Sums of random symmetric matrices and quadratic optimization under orthogonality constraints" – *Math. Progr.*, **109:2-3** (2007), 283–317.

Anatoli Juditsky
Laboratoire Jean Kuntzmann
51 rue des Mathématiques
BP 53
38041 Grenoble Cedex 9 France
E-mail: E-mail: juditsky@imag.fr

Arkadi Nemirovski
School of Industrial and Systems Engineering
Georgia Institute of Technology
765 Ferst Drive
Atlanta GA 30332-0205 USA
E-mail: E-mail: nemirovs@isye.gatech.edu