IMS Collections
From Probability to Statistics and Back: High-Dimensional Models and Processes
Vol. 9 (2013) 254–265

DOI: 10.1214/12-IMSCOLL918

Around Nemirovski's inequality

Pascal Massart and Raphaël Rossignol

Univ. Paris-Sud 11

Abstract: Nemirovski's inequality states that given independent and centered at expectation random vectors $X_1, \ldots, X_n$ with values in $\ell^p(\mathbb{R}^d)$, there exists some constant $C(p,d)$ such that

$$\mathbb{E}\|S_n\|_p^2 \leq C(p,d) \sum_{i=1}^n \mathbb{E}\|X_i\|_p^2.$$

Furthermore $C(p,d)$ can be taken as $\kappa(p \wedge \log(d))$. Two cases were studied further in [*Am. Math. Mon.* **117**(2) (2010) 138–160]: general finite-dimensional Banach spaces and the special case $\ell^\infty(\mathbb{R}^d)$. We show that in these two cases, it is possible to replace the quantity $\sum_{i=1}^n \mathbb{E}\|X_i\|_p^2$ by a smaller one without changing the order of magnitude of the constant when d becomes large. In the spirit of [*Am. Math. Mon.* **117**(2) (2010) 138–160], our approach is probabilistic. The derivation of our version of Nemirovski's inequality indeed relies on concentration inequalities.

1. Introduction

Let $\mathbb{B}$ be some separable Banach space and let $X_1, \ldots, X_n$ be independent and centered at expectation $\mathbb{B}$-valued random variables. Let us denote as usual:

$$S_n = \sum_{i=1}^n X_i.$$

In [1], the validity of an inequality like

$$\mathbb{E}\left\|S_n\right\|^2 \leq C \sum_{i=1}^n \mathbb{E}\left\|X_i\right\|^2 \tag{1.1}$$

is discussed and the structure of the constant C is investigated. Of course, when $\mathbb{B}$ is a Hilbert space, (1.1) holds with the optimal constant $C = 1$, but what about the non-Hilbertian case ? Whenever $\mathbb{B} = \ell^p(\mathbb{R}^d)$, Nemirovski's inequality (as stated in [4]) ensures that for every $p \geq 2$

$$\mathbb{E}\left\|S_n\right\|_p^2 \leq C\left(p,d\right) \sum_{i=1}^n \mathbb{E}\left\|X_i\right\|_p^2, \tag{1.2}$$

where the constant $C(p,d)$ can be taken as $C(p,d) = \kappa(p \wedge \ln(d))$ for some universal constant κ.

Département de Mathématiques, Université Paris-Sud, 91405 Orsay, France, e-mail: pascal.massart@math.u-psud.fr; raphael.rossignol@ujf-grenoble.fr

Keywords and phrases: Efron-Stein's inequality, high dimensional Banach space, Nemirovski's inequality, concentration inequalities, maximal inequalities

In [1] two different questions are addressed. The optimal value of the constant $C(p, d)$ is investigated, especially when $p = \infty$ and d is large. It is in particular proved that

$$\liminf_{d\to\infty} \frac{C(\infty, d)}{2\ln(d)} \geq 1,$$

and

$$\limsup_{d\to\infty} \frac{C(\infty, d)}{2\ln(d)} \leq e.$$

In another direction, some extensions of Nemirovski's inequality are considered. For instance, it is shown that in any Banach space of dimension d, (1.1) is true with $C = d$. Infinite dimensional Banach spaces such as $\mathbb{L}_p$ spaces are also considered. In contrast to Nemirovski's original approach which is based on analytic arguments, the methods introduced in [1] are probabilistic. In some sense we would like to go further in this direction by using some concentration tools.

The main purpose of the present note is to emphasize the benefits one gets from thinking in terms of concentration inequalities when tackling those questions. The first benefit is to reduce easily the problem of bounding $\mathbb{E}(\|S_n\|^2)$ to the problem of bounding $(\mathbb{E}\|S_n\|)^2$. The second advantage is that the concentration approach will naturally lead to some stronger version of Nemirovki's inequality without much effort. To be more precise, let us notice that one can write the norm of any vector x in $\mathbb{B}$ as

$$\|x\| = \sup_{\|t\|\leq 1} \langle t, x\rangle, \tag{1.3}$$

where the elements t belong to the dual space of $\mathbb{B}$. From this perspective the right-hand side of inequality (1.2) may be rewritten as

$$V = \sum_{i=1}^{n} \mathbb{E}\|X_i\|^2 = \mathbb{E}\left[\sum_{i=1}^{n} \sup_{\|t\|\leq 1} \langle t, X_i\rangle^2\right].$$

Now let us introduce the following "weak variance" (as opposed to the "strong variance" V) :

$$v = \mathbb{E}\left[\sup_{\|t\|\leq 1} \sum_{i=1}^{n} \langle t, X_i\rangle^2\right].$$

Those definitions look very similar. Passing from the definition of V to that of v just means commuting the supremum and the summation which of course leads to $v \leq V$. The second benefit of the concentration approach is to lead naturally to bounds involving v rather than V, with essentially the same constants. We shall demonstrate this in the case of finite dimensional Banach spaces. Since v can be substantially smaller than V, this may be a strong improvement upon Nemirovski's inequality. In the case $\mathbb{B} = \ell^\infty(\mathbb{R}^d)$, such a result was in fact implicitly present in [1], although their authors did not emphasize it. To demonstrate the importance of trading V for v, we shall investigate this in details in the case $\mathbb{B} = \ell^\infty(\mathbb{R}^d)$, which was announced in [1] as the most interesting (at least from the point of view of statistical applications).

The paper is organised as follows. The crucial tools from concentration inequalities are presented in Section 2. Notably, Efron-Stein inequality is used in Section 3 to reduce, in any Banach space, the control of $\mathbb{E}\|S_n\|^2$ to that of $(\mathbb{E}\|S_n\|)^2$.

Then, Section 4 focuses on the finite dimensional case, showing that the inequality $\mathbb{E}\|S_n\|^2 \leq C_{\mathbb{B}} v$ holds with a constant $C_{\mathbb{B}}$ proportional to the dimension. In Section 5, we investigate in details the case $\mathbb{B} = \ell^\infty(\mathbb{R}^d)$. We show that the inequality $\mathbb{E}\|S_n\|^2 \leq 8\log(2d)v$ may be considerably stronger than $\mathbb{E}\|S_n\|^2 \leq 8\log(2d)V$. Then, that inequality is refined in Section 5.3 when dealing with bounded individual variables $X_{i,j}$. Finally some perspectives are provided in Section 6.

2. Some concentration tools

Throughout the paper, we shall use some tools from concentration inequalities. The first one is the so-called Efron-Stein inequality, cf. [2], Section 2.5.4.

Theorem 2.1 (Efron-Stein inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables and let $Z = f(X)$ be a square-integrable function of $X = (X_1, \ldots, X_n)$. Then,*

$$\mathsf{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}\left[(Z - \mathbb{E}^{(i)} Z)^2\right],$$

where $\mathbb{E}^{(i)}$ denotes the expectation operator conditioned on all the variables except X_i.

Next we present some material which will find useful in the finite dimensional cases, cf. Sections 4 and 5. The following lemma allows to transfer a common bound on the Laplace transform into a maximal inequality. For a real random variable Z, we shall denote by $\psi_Z(\lambda)$ its log-Laplace transform:

$$\forall \lambda \in \mathbb{R}, \quad \psi_Z(\lambda) = \ln \mathbb{E} \exp(\lambda Z).$$

Lemma 2.2. *Let $\{Z(t), t \in T\}$ be a finite family of real valued random variables. Assume that for some nonnegative constants s and c, for every $\lambda \in (0, 1/c)$ (or $\lambda \in (0, +\infty)$ if $c = 0$) and $t \in T$,*

$$\psi_{Z(t)}(\lambda) \leq \frac{\lambda^2 s}{2(1 - c\lambda)}.$$

Then,

$$\mathbb{E}\left[\sup_{t \in T} Z(t)\right] \leq \sqrt{2s \ln(|T|)} + c \ln(|T|). \tag{2.1}$$

The case $c = 0$ corresponds to the well-known sub-gaussian setting. On the other hand, Lemma 2.2 applies if $\{X(t), t \in T\}$ is a finite family of random variables, each one following the chi-square distribution with p degrees of freedom. One obtains

$$\mathbb{E}\left[\sup_{t \in T} X(t) - p\right] \leq 2\sqrt{p \ln(|T|)} + 2 \ln(|T|). \tag{2.2}$$

Although there is nothing new here, by sake of completeness we give a proof of Lemma 2.2 in the appendix where we shall also recall the statements of Hoeffding and Bernstein's inequalities that we shall use in conjunction with Lemma 2.2.

When T is infinite, one may classically use the inequalities above in conjunction with a *chaining technique* and metric entropy computations. For instance, one may derive the following result, which will be useful in Section 4.

Proposition 2.3. *Let B be the unit ball of a d-dimensional Banach space. Let $(Z_t)_{t\in\mathbb{B}}$ be a stochastic process indexed by B such that:*

$$\forall t,t' \in B,\ \forall \lambda \in \mathbb{R}^+, \quad \psi_{Z(t)-Z(t')}(\lambda) \leq \frac{\lambda^2 s\|t-t'\|^2}{2},$$

then, for any $t_0 \in B$,

$$\mathbb{E}\left[\sup_{t\in T} Z(t) - Z(t_0)\right] \leq 12 \int_0^1 \sqrt{sd\log\left(1+\frac{2}{\varepsilon}\right)}\, d\varepsilon.$$

For a proof, one may combine for isntance Theorem 3.17 page 72 of [2], Lemma 4.16 page 56 and inequality (5.7) page 63 of [5].

3. The square does not matter

In the rest of this paper, we keep the notations of the introduction, i.e $\mathbb{B}$ is some separable Banach space and $X_1, \dots, X_n$ are independent and centered at expectation $\mathbb{B}$-valued random variables. Furthermore, we denote:

$$S_n = \sum_{i=1}^n X_i.$$

As noted in the introduction, inequality (1.1) is an equality in Hilbert spaces, with $C = 1$. In other Banach spaces, the optimal constant should reflect the geometry of the norm. The idea of this section is to use a concentration argument in order to reduce the control of $\mathbb{E}\|S_n\|^2$ to that of $(\mathbb{E}\|S_n\|)^2$, showing in the meantime that the geometric meaning of the constant C in inequality (1.1) is entirely contained in the relation between $(\mathbb{E}\|S_n\|)^2$ and V.

Proposition 3.1.

$$\mathbb{E}[\|S_n\|^2] \leq [\mathbb{E}\|S_n\|]^2 + V.$$

Proof. Using the notations of Theorem 2.1, notice that for any $X^{(i)}$-measurable and square-integrable random variable Z_i,

$$\mathbb{E}\big[(Z-\mathbb{E}^{(i)}Z)^2\big] \leq \mathbb{E}[(Z-Z_i)^2].$$

Thus, Theorem 2.1 and the triangle inequality give:

$$\mathsf{Var}(\|S_n\|) \leq \sum_{i=1}^n \mathbb{E}(\|S_n\| - \|S_n - X_i\|)^2 \leq V.$$

□

Now, let the weak variance enter the stage just by refining the preceeding argument.

Proposition 3.2.

$$\mathbb{E}[\|S_n\|^2] \leq [\mathbb{E}\|S_n\|]^2 + 2v.$$

Proof. Let $X' = (X'_1, \dots, X'_n)$ be an independent copy of X. Using the notations of Theorem 2.1, let $Z'_i = f(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)$. Then,

$$\mathbb{E}\big[(Z-\mathbb{E}^{(i)}Z)^2\big] = \mathbb{E}[(Z-Z'_i)_+^2].$$

Now, let $Z = \|S_n\|$ and use the representation (1.3):

$$Z = \sup_{\|t\| \leq 1} \sum_{i=1}^{n} \langle t, X_i \rangle.$$

Suppose that the supremum is attained at some t_X. Then,

$$(Z - Z_i')_+ \leq |\langle t_X, X_i \rangle - \langle t_X, X_i' \rangle|.$$

Since X_i' is centered and independent of X_i, taking the expectation with respect to X_i' leads to:

$$\mathbb{E}[(Z - Z_i')_+^2 | X] \leq \langle t_X, X_i \rangle^2 + \mathbb{E}[\langle t_X, X_i' \rangle^2 | X] = \langle t_X, X_i \rangle^2 + \mathbb{E}[\langle t_X, X_i' \rangle | X]^2.$$

Thus, Theorem 2.1 gives:

$$\begin{aligned} \mathsf{Var}\|S_n\| &\leq \sum_{i=1}^{n} \mathbb{E}[\langle t_X, X_i \rangle^2] + \sum_{i=1}^{n} \mathbb{E}[\langle t_X, X_i' \rangle^2] \\ &\leq 2v. \end{aligned}$$

□

Apart from the absolute constant 2, we see that one does not loose much in terms of constants when bounding the variance by the weak quantity v rather than the strong one V. As we shall see in Section 5.2, it may lead to substantial gains since v and V can be of very different order of magnitude, notably in $\ell^\infty(\mathbb{R}^d)$ when d becomes large.

4. Finite-dimensional Banach spaces

It is proved in [1] Corollary 2.10 that when $X_1, \ldots, X_n$ are centered independent random variables in a Banach space of dimension d:

$$\mathbb{E}\|S_n\|^2 \leq d \sum_{i=1}^{n} \mathbb{E}\|X_i\|^2. \tag{4.1}$$

Below, we shall show that one may replace $V = \sum_{i=1}^{n} \mathbb{E}\|X_i\|^2$ by v at the price of losing some universal, constant factor.

Theorem 4.1. *Let $\mathbb{B}$ be a Banach space of dimension d. There exists a universal constant κ such that if $X_1, \ldots, X_n$ are centered independent random variables in $\mathbb{B}$,*

$$\mathbb{E}\left[\|S_n\|_{\mathbb{B}}^2\right] \leq \kappa d v.$$

The constant κ may be chosen as 1154.

Proof. Let us denote by $\mathcal{D}$ the unit ball of the dual space $\mathbb{B}^*$, so that

$$\|S_n\| = \sup_{t \in \mathcal{D}} \sum_{i=1}^{n} \langle t, X_i \rangle.$$

Assume first the random variables X_i to be symmetric and let $(\varepsilon_i)_{1 \leq i \leq n}$ be i.i.d. Rademacher variables which are independent from the variables X_i's. Then,

$$\mathbb{E}\|S_n\| = \mathbb{E}\left[\sup_{t \in \mathcal{D}} \sum_{i=1}^{n} \varepsilon_i \langle t, X_i \rangle\right].$$

Now working conditionally on $(X_i)_{i\leq n}$, $\sup_{t\in\mathcal{D}}\sum_{i=1}^n \epsilon_i\langle t, X_i\rangle$ is the supremum of a process indexed by $\mathcal{D}$, with sub-Gaussian increments:

$$\begin{aligned}\mathsf{Var}\left(\sum_{i=1}^n \epsilon_i\langle t-t', X_i\rangle\right) &= \sum_{i=1}^n \langle t-t', X_i\rangle^2 \\ &\leq \|t-t'\|^2 \sup_{s\in\mathcal{D}}\sum_{i=1}^n\langle s, X_i\rangle^2\,,\end{aligned}$$

where $\|t-t'\|$ is the norm of $t-t'$ in $\mathbb{B}^*$. Thus, Lemma 6.1 and Proposition 2.3 gives:

$$\begin{aligned}&\mathbb{E}\left[\sup_{t\in\mathcal{D}}\sum_{i=1}^n \epsilon_i\langle t, X_i\rangle \mid X_1,\ldots,X_n\right] \\ &\quad\leq \left(\sup_{s\in\mathcal{D}}\sum_{i=1}^n\langle s, X_i\rangle^2\right)^{1/2} 12\int_0^1 \sqrt{d\log\left(1+\frac{2}{\varepsilon}\right)}\,d\varepsilon.\end{aligned}$$

By Jensen's inequality,

$$\begin{aligned}\left(\mathbb{E}\left[\|S_n\|\right]\right)^2 &\leq d\times\mathbb{E}\left[\sup_{s\in\mathcal{D}}\sum_{i=1}^n\langle s, X_i\rangle^2\right]\left(12\int_0^1\sqrt{\log\left(1+\frac{2}{\varepsilon}\right)}\,d\varepsilon\right)^2 \\ &\leq 144d\times v\times\int_0^1 \log\left(1+\frac{2}{\varepsilon}\right)\,d\varepsilon \\ &\leq 144vd(3\log 3-2\log 2)\leq 2\times 144vd.\end{aligned}$$

When the X_i's are not symmetric, a standard symmetrization argument gives:

$$\left(\mathbb{E}\left[\|S_n\|\right]\right)^2 \leq 8\times 144\times vd.$$

Combining this upper bound with Proposition 3.1,

$$\mathbb{E}\left[\|S_n\|^2\right] \leq (2+1152d)v \leq 1154dv. \qquad \square$$

Notice that in the proof above, Proposition 3.1 is particularly important if one want to be able to use the chaining result of Proposition 2.3.

5. The case $\mathbb{B}=\ell^\infty(\mathbb{R}^d)$

In this section, we focus on the key case where $\mathbb{B}=\ell_\infty(\mathbb{R}^d)$. It is proved in [1] Corollary 2.3 and Corollary 3.5 that when $X_1,\ldots,X_n$ are centered independent random variables in $\ell^\infty(\mathbb{R}^d)$:

$$\mathbb{E}\|S_n\|^2 \leq C(\infty,d)\sum_{i=1}^n \mathbb{E}\|X_i\|^2 \tag{5.1}$$

with:

$$C(\infty,d)=\begin{cases}2e\ln d-e & \text{if } d\geq 3,\\ d & \text{if } d\leq 2.\end{cases}$$

In addition, when $X_1, \ldots, X_n$ are symetric, one may take $C(\infty, d) = 2\ln(2d)$. Looking at the proof of Corollary 3.5 in [1], one realizes that Wellner et al. actually show that when $X_1, \ldots, X_n$ are symmetric:

$$\mathbb{E}\left\|S_n\right\|^2 \leq 2v\log(2d),$$

and when $X_1, \ldots, X_n$ are centered:

$$\mathbb{E}\left\|S_n\right\|^2 \leq 8v\log(2d).$$

In this section, we shall first derive (a slightly weaker version of) those results using our techniques, and then describe some cases where v is dramatically smaller than $V = \sum_{i=1}^n \mathbb{E}\left\|X_i\right\|^2$.

5.1. Improving Nemirovski's inequality

First, notice that

$$v = \mathbb{E}\sup_{\|t\|_1 \leq 1}\sum_{i=1}^n \langle t, X_i\rangle^2 = \mathbb{E}\max_{j\leq d}\sum_{i=1}^n X_{i,j}^2. \tag{5.2}$$

Indeed, $t \mapsto \sum_{i=1}^n \langle t, X_i\rangle^2$ is a convex function and the unit ℓ^1-unit ball is the convex hull of the set $\{\pm b_1, \ldots, \pm b_d\}$ where $b_1, \ldots, b_d$ is the standard basis of $\mathbb{R}^d$. Now, assume first the random variables X_i to be symmetric. Then, if $(\varepsilon_i)_{1\leq i\leq n}$ are i.i.d. Rademacher variables which are independent from the variables X_i's

$$\mathbb{E}\left\|S_n\right\| = \mathbb{E}\max_{j\leq d}\left|\sum_{i=1}^n \varepsilon_i X_{i,j}\right|.$$

Working conditionally to the variables X_i's, note that Lemma 6.1 implies that the assumptions of Lemma 2.2 are satisfied with $s = \max_{j\leq d}\sum_{i=1}^n X_{i,j}^2$ and $c = 0$. Therefore

$$\mathbb{E}\left[\max_{j\leq d}\left|\sum_{i=1}^n \varepsilon_i X_{i,j}\right| \mid X_1, \ldots, X_n\right] \leq \sqrt{2\ln(2d)\max_{j\leq d}\sum_{i=1}^n X_{i,j}^2},$$

where we used the fact that:

$$\max_{j\leq d}\left|\sum_{i=1}^n \varepsilon_i X_{i,j}\right| = \max_{j\leq 2d}\sum_{i=1}^n \varepsilon_i X_{i,j}$$

if we define $X_{i,j} = -X_{i,j-d}$ for $d+1 \leq j \leq 2d$. Taking expectation on both sides of this inequality we derive via Jensen's inequality that

$$\mathbb{E}\left[\max_{j\leq d}\left|\sum_{i=1}^n \varepsilon_i X_{i,j}\right|\right] \leq \sqrt{2\ln(2d)\,\mathbb{E}\max_{j\leq d}\sum_{i=1}^n X_{i,j}^2}.$$

Combining this inequality with (5.2) and Proposition 3.2, we finally get

$$\mathbb{E}\left\|S_n\right\|_\infty^2 \leq 2\left(1+\ln(2d)\right)v. \tag{5.3}$$

In the general case, by using a standard symmetrization argument, we loose some factor 2 in the control of $\mathbb{E}\|S_n\|$ and the inequality becomes

$$\mathbb{E}\left\|S_n\right\|_\infty^2 \leq 2\left(1+4\ln\left(2d\right)\right)v. \tag{5.4}$$

We recover thus (a slightly weaker version of) the result of Wellner et al.

Interestingly, the constant $2(1+\ln(2d))$ has the right order of magnitude when d goes to infinity and the X_i are symmetric. This is a consequence of the counterexample given in [1] Example 1.2 and we shall by the way provide some alternative counterexample below, in Section 5.2. Thus in the symmetric case we were able to produce a stronger version of Nemirovski's inequality (we mean involving v instead of V) with an asymptotically optimal constant.

5.2. Comparison of strong and weak variances

Let us come back a moment to the general Banach framework. Since

$$v \geq \mathbb{E}\left[\max_{1\leq i\leq n}\sup_{\|t\|\leq 1}\langle t, X_i\rangle^2\right]$$

we derive that

$$v \leq V \leq nv.$$

Now we focus on the finite dimensional case and denote by v_p, respectively V_p, the values of the quantities v and V respectively whenever $\mathbb{B}=\ell_p(\mathbb{R}^d)$. Noticing that

$$v_\infty = \mathbb{E}\left[\max_{j\leq d}\sum_{i=1}^n X_{i,j}^2\right]$$

we derive that

$$V_\infty \leq V_2 = \mathbb{E}\left[\sum_{j=1}^d\sum_{i=1}^n X_{i,j}^2\right] \leq dv_\infty$$

and therefore

$$v_\infty \leq V_\infty \leq (n\wedge d)\, v_\infty.$$

Our purpose now is to show that the ratio V_∞/v_∞ can indeed be of order $n\wedge d$ in two very different (and in some sense opposite) situations described below.

5.2.1. The non i.i.d. case

In this case, it is very easy to exhibit a situation where the ratio V_∞/v_∞ is maximal. Let us indeed consider the following highly non i.i.d. scheme. Let $X_{i,j}=\varepsilon_i a_{i,j}$ where $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d. Rademacher random variables. Then

$$v_\infty = \max_{j\leq d}\sum_{i=1}^n a_{i,j}^2 \qquad \text{while } V_\infty = \sum_{i=1}^n \max_{j\leq d} a_{i,j}^2.$$

Choosing $a_{j,j}=1$ for every $j\leq n\wedge d$ and $a_{i,j}=0$ otherwise, we readily see that $v_\infty = 1$ while $V_\infty = n\wedge d$. This shows that the ratio V_∞/v_∞ can indeed achieve the maximal value $n\wedge d$.

5.2.2. The i.i.d. case

It is even more interesting to investigate the value of the ratio V_∞/v_∞ in the case where the variables $X_{i,j}$ are i.i.d. standard normal random variables. In this case, considering $\xi_1,\dots,\xi_d$ i.i.d. standard normal random variables, one has $V_\infty = n\mathbb{E}[\max_{j\le d}\xi_j^2]$. Moreover, Nemirovski's inequality writes

$$\mathbb{E}\max_{j\le d}\left|\sum_{i=1}^n X_{i,j}\right|^2 \le C(\infty,d)\,V_\infty$$

and is equivalent to

$$V_\infty \le C(\infty,d)\,V_\infty.$$

Since $C(\infty,d)$ is of order $\ln(d)$ as d goes to infinity, this means that Nemirovski's inequality misses the target from a $\ln(d)$ factor. This is completely due to the use of the strong variance V_∞ instead of the weak one v_∞. Indeed since each of the random variables $\sum_{i=1}^n X_{i,j}^2$ has a chi-square distribution with n degrees of freedom, it comes from (2.2) below that

$$\mathbb{E}\left[\max_{j\le d}\left(\sum_{i=1}^n X_{i,j}^2\right)-n\right] \le 2\sqrt{n\ln(d)}+2\ln(d)$$

i.e.

$$v_\infty \le n+2\left(\sqrt{n\ln(d)}+\ln(d)\right).$$

Since obviously $v_\infty \ge n$, we derive the following inequalities for the ratio V_∞/v_∞

$$\frac{\mathbb{E}\left[\max_{j\le d}\xi_j^2\right]}{1+2\sqrt{\ln(d)/n}+2(\ln(d)/n)} \le \frac{V_\infty}{v_\infty} \le \mathbb{E}\left[\max_{j\le d}\xi_j^2\right]. \tag{5.5}$$

Let us now consider three different asymptotic regimes of dependency of $d=d_n$ w.r.t. n as n goes to infinity.

1. $\ln(d_n)/n$ tends to 0 as n goes to infinity
2. $\ln(d_n)/n$ tends to $\alpha>0$ as n goes to infinity
3. $\ln(d_n)/n$ tends to infinity as n goes to infinity.

Recalling that $\mathbb{E}[\max_{j\le d}\xi_j^2]/2\ln(d)$ tends to 1 as d goes to infinity, in the sub-exponential asymptotic regime 1., inequality (5.5) implies that the ratio V_∞/v_∞ is equivalent to $2\ln(d_n)$ as n goes to infinity. Since inequality (5.3) can be rewritten as

$$V_\infty \le 2(1+\ln(2d_n))\,v_\infty$$

this proves on the one hand the asymptotic optimality of this inequality: i.e. the $2\ln(d_n)$ factor cannot be avoided. On the other hand, since the convergence to 0 of $\ln(d_n)/n$ can be arbitrary slow, the ratio V_∞/v_∞ is close to its maximal value n. In the exponential regime 2., our inequality (5.3) is no longer optimal, but it misses the target from a factor at most $1+\sqrt{2\alpha}+2\alpha$. This time the ratio V_∞/v_∞ is actually of order n (up to some constant) in the sense that

$$\liminf_{n\to\infty}\frac{V_\infty}{nv_\infty} \ge \frac{2\alpha}{1+\sqrt{2\alpha}+2\alpha}.$$

Finally, for the over-exponential regime 3., our inequality misses the target from a factor bounded by $\ln(d_n)/n$ and in the meantime the ratio V_∞/v_∞ is asymptotically of the order of its maximal value n, in the sense that V_∞/nv_∞ tends to 1 as n goes to infinity.

5.3. A refined inequality for bounded variables

In order to produce some ready to be used inequalities for applications, it would be more convenient to put moment hypotheses on the individual variables $X_{i,j}$ rather than hypotheses on $\|X_i\|$. In this case, neither the original version of Nemirovski's inequality which involves the strong variance nor our version involving with the weak variance give an ultimate control on $\mathbb{E}(\|S_n\|^2)$. We do not want here to produce a whole bunch of results in this direction but just show that it is possible to do it at least in the easiest case where the variables are uniformly bounded. Indeed, using Proposition 3.2, Hoeffding and Bernstein's inequalities allows to derive the following control for bounded individual variables.

Proposition 5.1. *Let $X_1, \dots, X_n$ be centered independent random variables in $\ell^\infty(\mathbb{R}^d)$. Suppose that:*

$$\sup_{i,j} \|X_{i,j}\|_\infty \le L \quad \textit{and} \quad \sup_j \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_{i,j}^2) \le \sigma^2.$$

Then,

$$\mathbb{E}(\|S_n\|_\infty^2) \le 2n\sigma^2 \left\{ \ln(2d) \left(1 + \frac{L}{3\sigma} \sqrt{\frac{\ln(2d)}{2n}} \right)^2 + 1 + \frac{L^2}{\sigma^2} \sqrt{\frac{\ln(2d)}{2n}} \right\}.$$

Proof. Recall that

$$\mathbb{E}(\|S_n\|_\infty^2) \le 2v + (\mathbb{E}\|S_n\|_\infty)^2.$$

We may use Bernstein's inequality to bound $\mathbb{E}(\|S_n\|_\infty)$. Indeed, from Lemmas 2.2 and 6.2, we derive

$$\mathbb{E}(\|S_n\|_\infty) \le \sqrt{2\sigma^2 \ln(2d)} + L\ln(2d)/3.$$

Now, we use Hoeffding's inequality to bound v. From Lemmas 2.2 and 6.1, we get

$$v = \mathbb{E}(\max_j \sum_{i=1}^n X_{i,j}^2) \le \max_j \mathbb{E}(\sum_{i=1}^n X_{i,j}^2) + L^2 \sqrt{\frac{n\ln(2d)}{2}}.$$

Gathering the two bounds ends the proof. □

Notice that when d is moderately large compared to n, in the sense that $(\ln d)/n$ goes to zero, then Proposition 5.1 leads to a bound equivalent to:

$$\mathbb{E}(\|S_n\|_\infty^2) \le 2n\sigma^2 \ln(d),$$

which is optimal.

6. Perspective

When focusing on the case $\ell^\infty(\mathbb{R}^d)$, we have shown that trading the strong variance with the weak one improved significantly Nemirovski's inequality. On the other hand, in $\ell^2(\mathbb{R}^d)$, one cannot beat Nemirovski's inequality since it is an equality, with an optimal constant $C = 1$. Thus, one may naturally wonder what happens in the case $\mathbb{B} = \ell^p(\mathbb{R}^d)$ with $p \in (2, \infty)$. A trivial step is done by noticing that for any $x \in \mathbb{R}^d$,

$$\|x\|_\infty \le \|x\|_p \le d^{1/p}\|x\|_\infty,$$

which gives for centered independent variables X_i in $\ell^p(\mathbb{R}^d)$:

$$\mathbb{E}(\|S_n\|_p^2) \le d^{1/p} 2(1 + 4\ln(2d))v_p,$$

with:

$$v_p = \mathbb{E}\left[\sup_{\|t\|_{p'} \le 1} \sum_{i=1}^n \langle t, X_i\rangle^2\right],$$

and $\frac{1}{p'} + \frac{1}{p} = 1$. Notably, when p is at least of the order of $\ln(d)$, we recover a constant comparable to that of Nemirovski's inequality, $C(p, d) = \kappa(p \wedge \ln(d))$, but with the weak variance v_p instead of the strong one. Probably there should be some value of p below which the weak variance does not help and it would be interesting to understand this phenomenon.

Besides, in the spirit of Section 5.3, it would be interesting to go beyond the case of bounded variables. This would require a tool to replace Bernstein's inequality in that context and is in fact a delicate question. Recent progress in this direction can be found in [3].

Appendix

First we give a proof of Lemma 2.2. Setting $x = \mathbb{E}\left[\sup_{t\in T} Z(t)\right]$, we have by Jensen's inequality

$$\exp(\lambda x) \le \mathbb{E}\left[\exp\left(\lambda \sup_{t\in T} Z(t)\right)\right] = \mathbb{E}\left[\sup_{t\in T} \exp(\lambda Z(t))\right],$$

for any $\lambda \in (0, 1/c)$. Hence, recalling that $\psi_{Z(t)}(\lambda) = \ln \mathbb{E}\left[\exp(\lambda Z(t))\right]$,

$$\exp(\lambda x) \le \sum_{t\in T} \mathbb{E}\left[\exp(\lambda Z(t))\right] \le |T| \exp(\psi(\lambda)).$$

Therefore for any $\lambda \in (0, 1/c)$ we have

$$\lambda x - \psi(\lambda) \le \ln(|T|),$$

which means that

$$x \le \inf_{\lambda\in(0,1/c)} \left[\frac{\ln(|T|) + \psi(\lambda)}{\lambda}\right].$$

Simple calculus shows that the right-hand side is optimized at $\lambda = 1/(c + \sqrt{s/(2\ln|T|)})$, which gives the value announced in Lemma 2.2.

Finally, below we give Hoeffding and Bernstein inequalities in their exponential form, i.e the control on the Laplace transform which implies the usual forms of these well-known inequalities.

Lemma 6.1 (Hoeffding inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables with values in $[a_i, b_i]$ and define:*

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

Then,

$$\psi_S(\lambda) \leq \frac{\lambda^2}{8} \sum_{i=1}^{n} (b_i - a_i)^2.$$

Lemma 6.2 (Bernstein inequality)**.** *Let* $X_1, \ldots, X_n$ *be independent random variables such that* $X_i \leq L$*, and define:*

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

Then, for every $\lambda \in (0, 1/c)$

$$\psi_S(\lambda) \leq \frac{\lambda^2 s}{2\,(1 - c\lambda)}$$

with $s = \sum_{i=1}^{n} \mathbb{E}(X_i^2)$ *and* $c = L/3$*.*

References

[1] Dümbgen, L., van de Geer, S. A., Veraar, M. C. and Wellner, J. A. (2010). Nemirovski's inequalities revisited. *Am. Math. Mon.* **117**(2) 138–160.

[2] Massart, P. (2007). *Concentration Inequalities and Model Selection. Ecole d'Eté de Probabilités de Saint-Flour XXXIII – 2003.* Lecture Notes in Mathematics 1896. Springer, Berlin, xiv, 337 p.

[3] Mendelson, S. (2008). On weakly bounded empirical processes. *Math. Ann.* **340**(2) 293–314.

[4] Nemirovski, A. (2000). *Topics in Non-Parametric Statistics*, pp. 85–277. *Lecture Notes in Mathematics* **1738**. Springer, Berlin.

[5] Pisier, G. (1989). *The Volume of Convex Bodies and Banach Space Geometry. Cambridge Tracts in Mathematics* **94**. Cambridge University Press, Cambridge.