Empirical processes and unconditional log-concave measures

Shahar Mendelson\textsuperscript{a}, Grigoris Paouris\textsuperscript{b}

\textsuperscript{a}Department of Mathematics, Technion, I.I.T, Haifa 32000, Israel.
\textsuperscript{b}Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, U.S.A.

Received ****; accepted after revision +++++

Presented by

Abstract

We show that if \( \mu \) is an unconditional, isotropic, log-concave measure on \( \mathbb{R}^n \), \( n \in \mathbb{N} \) and \( T \) is a symmetric subset of \( \mathbb{R}^n \), then \( \mathbb{E}\|\sum_{i=1}^{N} X_i\|_{T^\circ} \leq C \sqrt{N} E(T) \) and \( \mathbb{E}\sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 - |t|^2 \right| \leq C \left( d_2(T) \frac{E(T)}{\sqrt{N}} + E_2(T) \right) \), where \( C > 0 \) is an absolute constant, \( E(T) = \mathbb{E}\sup_{t \in T}(Z,t) \), \( Z = (z_1, \ldots, z_N) \), \( z_i \) are i.i.d. exponentials with variance 1 and \( d_2(T) = \sup_{t \in T} |t| \) is the Euclidean radius of \( T \).

Résumé

Nous montrons que si \( \mu \) est une mesure log-concave isotrope inconditionnelle sur \( \mathbb{R}^n \), \( n \in \mathbb{N} \) et \( T \) est un sous-ensemble symétrique de \( \mathbb{R}^n \), alors \( \mathbb{E}\|\sum_{i=1}^{N} X_i\|_{T^\circ} \leq C \sqrt{N} E(T) \) et \( \mathbb{E}\sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 - |t|^2 \right| \leq C \left( d_2(T) \frac{E(T)}{\sqrt{N}} + E_2(T) \right) \), où \( C > 0 \) est une constante absolue, \( E(T) = \mathbb{E}\sup_{t \in T}(Z,t) \), \( Z = (z_1, \ldots, z_N) \), \( z_i \) sont des variables aléatoires i.i.d exponentielles de variance 1 et \( d_2(T) = \sup_{t \in T} |t| \) est le rayon Euclidienne de \( T \).

1. Introduction

The goal of this note is to study several natural processes that are generated by independent random vectors distributed according to an isotropic, log-concave probability measure, that is assumed to be unconditional.

Definition 1.1 Let \( \mathbb{R}^n \) be endowed with a fixed, Euclidean structure \( \| \cdot \| \). A measure \( \mu \) on \( \mathbb{R}^n \) is isotropic if it is symmetric and if for every \( t \in \mathbb{R}^n \), \( \mathbb{E}(t, X)^2 = |t|^2 \). It is log-concave if for any two nonempty Borel measurable sets \( A, B \subset \mathbb{R}^n \), and every \( 0 \leq \lambda \leq 1 \), \( \mu(\lambda A + (1 - \lambda)B) \geq \mu^\lambda(A) \cdot \mu^{1-\lambda}(B) \). If \( X = (x_1, \ldots, x_n) \) is distributed according to \( \mu \), then \( X \) is unconditional if it has the same distribution as \( (\varepsilon_1 x_1, ..., \varepsilon_n x_n) \) for every choice of signs \( (\varepsilon_i)_{i=1}^n \).

The study of isotropic, log-concave measures plays a central role in modern asymptotic geometric analysis, and we refer the reader to [4] and references therein for more information on this topic. The question we would like to focus on here is the following:

Question 1.2 Are there “natural” families of functionals \( \Phi_{N,n} \) and \( \Psi_{N,n} \) such that for every isotropic, log-concave measure \( \mu \) on \( \mathbb{R}^n \), every \( T \subset \mathbb{R}^n \), and every independent vectors \( X_1, ..., X_N \) distributed according to \( \mu \),

\[
\mathbb{E}\sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle \right| = \mathbb{E}\left| \sum_{i=1}^{N} X_i \right|_{T^\circ} \leq \Phi_{N,n}(T, \mu) \quad \text{and} \quad \mathbb{E}\sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 - |t|^2 \right| \leq \Psi_{N,n}(T, \mu) \tag{1}
\]

* Shahar Mendelson was partially supported by the Mathematical Sciences Institute – The Australian National University, The European Research Council (under ERC grant agreement n\textdegree\textsuperscript{203314}), the Israel Science Foundation (under grant 900/10) and the Australian Research Council (under grant DP0986563). Grigoris Paouris was partially supported by NSF grant (DMS-0906150).

Email addresses: shahar@tx.technion.ac.il (Shahar Mendelson), grigoris@math.tamu.edu (Grigoris Paouris).

Preprint submitted to Elsevier Science January 15, 2011
Both parts of the question deal with uniform concentration of a sum of i.i.d random variables around their means, but it is clear that the second question is, in general, far more difficult than the first one, because of its non-linear nature and since higher powers of functions accentuate their “large part”, making concentration harder.

To get a clearer picture of what one might hope for in Question 1.2, let us consider the linear case for $N = 1$. In other words, one has to find a natural functional which bounds $E\|X\|_{T^*} = E\sup_{t \in T} \langle X, t \rangle$ for every $T \subset \mathbb{R}^n$. A possible candidate is $E\sup_{t \in T} \langle Z, t \rangle$, if $Z$ is a random vector that weakly-dominates $X$, in the following sense:

**Definition 1.3** A random vector $Z$ on $\mathbb{R}^n$ weakly dominates $X$ if for every $u \geq 1$ and every $t \in \mathbb{R}^n$, $Pr(|\langle X, t \rangle| \geq u|t|) \leq Pr(|\langle Z, t \rangle| \geq u|t|)$.

One natural example in this context is $L$-subgaussian vectors; that is, isotropic vectors $X$ which satisfy that for every $u \geq 1$ and any $t \in \mathbb{R}^n$, $Pr(|\langle X, t \rangle| \geq uL|t|) \leq 2 \exp(-u^2/2)$. Hence, these random variables are dominated by $cLG$, for an isotropic, standard Gaussian vector $G$ and a suitable absolute constant $c$.

The expectation of the supremum of the (isonormal) Gaussian process $E\sup_{t \in T} \langle G, t \rangle$ indexed by a set $T$ has been extensively studied (see, e.g. the books [2,10]) and is determined by the metric structure of $T$ via the $\gamma_2$ functionals.

**Definition 1.4** Let $(T, d)$ be a metric space. A collection of subsets of $T$, $(T_s)_{s=0}^\infty$, is an admissible sequence if $|T_0| = 1$ and $|T_s| \leq 2^s$ for $s \geq 1$. For $\alpha = 1, 2$ let $\gamma_\alpha(T, d) = \inf_{T_s} \sum_{s=0}^\infty 2^\alpha/s \alpha d(t, \pi_s(t))$, where $\pi_s(t)$ is a nearest point to $t$ in $T_s$, and the infimum is taken with respect to all admissible sequences of $T$.

The fundamental property of isotropic Gaussian processes indexed by $T \subset \mathbb{R}^n$ is that $c_1\gamma_2(T, \cdot | \cdot) \leq E\sup_{t \in T} \langle G, t \rangle \leq c_2\gamma_2(T, \cdot | \cdot)$ for absolute constants $c_1$ and $c_2$. The upper bound is due to Fernique [3] while the lower one is Talagrand’s majorizing measures theorem [8]. The proof of both parts can be found in [10]. From here on we will denote $\ell(T) = E\sup_{t \in T} \langle G, t \rangle$.

For a subgaussian vector one has the following:

**Theorem 1.5** There exist absolute constant $c$ and $C$ such that for every $T \subset \mathbb{R}^n$ and any $L$-subgaussian vector $X$, $E\sup_{t \in T} \langle X, t \rangle \leq L\gamma_2(T, \cdot | \cdot) \leq CL\ell(T)$.

In light of the majorizing measures theorem, Theorem 1.5 follows from a chaining argument and since $\gamma_2(T, \psi_2(\mu)) \leq CL\gamma_2(T, \cdot | \cdot)$ (where we identify $T$ with the class of linear functionals $\{\langle t, \cdot \rangle : t \in T\}$, and thus $\|t\|_{\psi_2(\mu)} \sim C_{\psi_2(\mu)}$ as $p \to 2$).

With Theorem 1.5 in mind, it seems natural to conjecture that for a subgaussian vector, $\Phi_{N,n}(T, \mu)$ and $\Psi_{N,n}(T, \mu)$ will depend on $\ell(T)$. And, indeed, under a subgaussian assumption Question 1.2 does have a satisfactory answer, which follows from a general result in empirical processes theory.

**Theorem 1.6** There exist an absolute constant $C$ for which the following holds. Let $X$ be an isotropic vector on $\mathbb{R}^n$. Then for every integer $N$ and every $T \subset \mathbb{R}^n$ which is symmetric (i.e., if $t \in T$ then $-t \in T$), $E\| X \|_{T^*} \leq C\sqrt{N} \gamma_2(T, \psi_2)$ and

$$E\sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 - |t|^2 \right| \leq C \left( d_{\psi_2}(T, \psi_2) \right)^2 + \frac{\gamma_2(T, \psi_2)}{N},$$

where $d_{\psi_2}(T) = \sup_{t \in T} \|\langle X, t \rangle\|_{\psi_2}$. In particular, if $X$ is $L$-subgaussian then one can take $\Phi_{N,n}(T, \mu) \leq CL\sqrt{N} \ell(T)$, and $\Psi_{N,n}(T, \mu) \leq CL^2 \left( d_2(T) \frac{\sqrt{N}}{N} + \frac{\ell(T)}{N} \right)$, where $d_2(T) = \sup_{t \in T} |t|$.

Theorem 1.6 was established in [6] for $T \subset S^{n-1}$, and in [7] for a general subset of $\mathbb{R}^n$. It is optimal in the realm of subgaussian vectors (with the extreme case is when $\mu$ is the standard gaussian vector).

When trying to extend Theorem 1.6 to an arbitrary isotropic, log-concave vector, one encounters several problems. The biggest stumbling block is that if $\mu$ is an isotropic, log-concave measure, the tails of linear functionals $\langle t, \cdot \rangle$ are sub-exponential rather than subgaussian. That is, $\|\langle t, \cdot \rangle\|_{\psi_1(\mu)} \leq c_1\|\langle t, \cdot \rangle\|_{L_2(\mu)}$ for a suitable absolute constant $c$, while $\|\langle t, \cdot \rangle\|_{\psi_2(\mu)}$ might be much larger, or even infinite. Thus, a direct
application of the second part of Theorem 1.6 will often result in a trivial bound, because \( d_{\psi_2}(T) \) is simply too large. Moreover, even if the measure is \( L \)-subgaussian, but \( L \) is very large, the resulting bounds will be clearly loose.

This gap has been partially closed in the main result in [7], where it was shown that for a general, isotropic log-concave measure \( \mu \),

\[
\mathbb{E} \sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 - |t|^2 \right| \leq C \left( \frac{d_{\psi_2}(T) \gamma_2(T, \psi_2(\mu))}{\sqrt{N}} + \frac{\gamma_2^2(T, \psi_2(\mu))}{N} \right),
\]

which leads to a better general upper bound on \( \Psi_{N,n} \).

Although the estimates mentioned above are optimal in the sense that it is impossible to replace \( \gamma_2(T, \psi_2(\mu)) \) by any other \( \gamma_2(T, \psi_2(\mu)) \) for any \( \alpha < 2 \), these bounds are still far from satisfactory. Indeed, except for subgaussian vectors and a few other special cases (e.g. \( T = B_2^n \) under additional assumptions on the measure \( \mu \)), reasonable bounds on \( \gamma_2(T, \psi_2(\mu)) \) are not known. Moreover, because of the possible large gap between \( \| \langle t, \cdot \rangle \|_{\psi_2(\mu)} \sim |t| \), and \( \| \langle t, \cdot \rangle \|_{\psi_2(\mu)} \gamma_2(T, \psi_2(\mu)) \) seems to be “too large” to capture the desired concentration properties.

Our main result is a sharp answer to Question 1.2 if \( \mu \) is unconditional.

The class of unconditional measures has a similar feature to the class of subgaussian measures – it too has a natural “dominating” measure. Indeed, by the Bobkov-Nazarov Theorem [1], if \( X \) is isotropic, log-concave and unconditional, and if we set \( Z = (z_1, \ldots, z_n) \), where \( z_i \) are i.i.d exponentials with variance 1 (i.e., with density \( 2^{-1/2} \exp(-\sqrt{2} |z|) \)), then \( Z \) is dominated by \( c Z \) for a suitable absolute constant \( c \).

Recently, Latała showed in [5] that for every \( T \subset \mathbb{R}^n \), \( \mathbb{E} \sup_{t \in T} \langle X, t \rangle \leq c_1 \mathbb{E} \sup_{t \in T} \langle Z, t \rangle \), which answers Question 1.2 for \( N = 1 \) in this case.

Let us first present a minor extension to Latała’s result (with a completely different proof). We present it solely to give a flavor of the more involved arguments that are needed to resolve Question 1.2 in the unconditional case. To formulate it, let \( Z \) a vector with i.i.d. exponential components as above, and set \( E(T) = \mathbb{E} \sup_{t \in T} \langle Z, t \rangle \).

**Theorem 1.7** There exist absolute constants \( c_1 \) and \( c_2 \) for which the following holds. Let \( X \) be an isotropic, unconditional, log-concave measure on \( \mathbb{R}^n \). Then, for every symmetric \( T \subset \mathbb{R}^n \),

\[
c_1 \mathbb{E} \sup_{t \in T} \langle t, X \rangle \leq \gamma_2(T, \| \cdot \|_1) + \gamma_1(T, \ell_\infty) \leq c_2 E(T).
\]

**Sketch of Proof.** Note that \( Pr(\| \langle X, t \rangle \|e \geq \| \langle X, t \rangle \|_\ell_\infty) \leq \exp(-p) \). Hence, if \( (T_s)_{s \geq 0} \) is an admissible sequence of \( T \) and \( \Delta_s(t) = \pi_{s+1}(t) - \pi_s(t) \), then by a chaining argument, for every \( u > c_1 \), \( \sup_{t \in T} \| \langle X, t - \pi_0(t) \rangle \|_{\ell_\infty} \leq \sum_{s=0}^{\infty} \| \Delta_s(t) \|_{\ell_\infty} \) with probability at least \( 1 - 2 \exp(-cw) \). Next, one has to use the weak domination of \( X \) by \( c_3 Z \) to obtain that for every \( p \geq 1 \) and every \( x \in \mathbb{R}^n \), \( \| \langle X, x \rangle \|_{L_p} \leq c_4 \| \langle Z, x \rangle \|_{L_p} \). Moreover, if \( (x_i^*)_{i=1}^n \) is a monotone ordering of \( (|x_i|)_{i=1}^n \) then for \( p \leq 1 \), \( \| \langle Z, x \rangle \|_{L_p} \sim px_1^* + \sqrt{p} \left( \sum_{i=1}^n (x_i^*)^2 \right)^{1/2} \), and one has a similar control for \( p \geq n \). A correct choice of the admissible sequence gives the first inequality. The second one is a deep result due to Talagrand [9,10].

The main idea of this proof is to use the \( L_p \) norms of increments in the chaining process, which is “local” information, rather than the “global” \( \psi_2 \) structure that has been used in the proof of (2). This simple idea leads to our main result, first by showing that \( \gamma_2(T, \psi_2(\mu)) \) can be replaced by a “local” functional, and then by bounding this functional using \( E(T) \).

**Theorem 1.8** There exist an absolute constant \( C \) for which the following holds. Let \( \mu \) be an unconditional, isotropic, log-concave measure on \( \mathbb{R}^n \). Then for every integer \( N \) and every symmetric \( T \subset \mathbb{R}^n \),

\[ \mathbb{E}\left\| \sum_{i=1}^{N} X_i \right\|_{T^\circ} \leq C \sqrt{N} E(T), \quad \text{and} \quad \mathbb{E} \sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 - |t|^2 \right| \leq C \left( d_2(T) \frac{E(T)}{\sqrt{N}} + \frac{E^2(T)}{N} \right). \]

Moreover, similar bounds hold with high probability.

A part of the proof of Theorem 1.8 is accurate information on the structure of \( P_\mu T = \{(\langle X_i, t \rangle)_{i=1}^{N} : t \in T\} \) which is of independent interest.

**Theorem 1.9** There exist absolute constants \( c_1, c_2 \) and \( c_3 \) for which the following holds. Let \( \mu \) be an unconditional, isotropic, log-concave measure on \( \mathbb{R}^n \) and let \( T \subset \mathbb{R}^n \) be a symmetric set. Then, for every \( u > c_1 \), with probability at least \( 1 - 2 \exp(-c_2 u \log N) \), for every \( I \subset \{1, \ldots, N\} \) and every \( t \in T \),

\[ \left( \sum_{i \in I} \langle t, X_i \rangle^2 \right)^{1/2} \leq c_3 u \left( E(T) + d_2(T) \sqrt{|I| \log(eN/|I|)} \right). \]

Observe that Theorem 1.9 gives the following sharp maximal inequality – that with high probability and in expectation, for every \( I \subset \{1, \ldots, N\} \), \( \| \sum_{i \in I} X_i \|_{T^\circ} \leq E(T) \sqrt{|I|} + d_2(T) |I| \log(eN/|I|) \). Indeed, this follows from Theorem 1.9 by noting that for every \( v \in \mathbb{R}^N \) and every \( I \subset \{1, \ldots, N\} \), \( \sum_{i \in I} |v_i| \leq \sqrt{|I|}(\sum_{i \in I} v_i^2)^{1/2} \).

Finally, let us give two applications (out of many) of Theorem 1.8.

Let \( X \) be an isotropic, unconditional, log concave vector and let \( \Gamma = N^{-1/2} \sum_{i=1}^{N} \langle X_i, \cdot \rangle e_i \) be a random matrix with the independent rows \( (X_i)_{i=1}^{N} \) distributed as \( X \). Then, for every \( N \geq c_1 n \), all the singular values of \( \Gamma \) belong to the interval \([1 - c_2 \sqrt{n/N}, 1 + c_2 \sqrt{n/N}]\). Note that this is the situation for the gaussian ensemble.

The next application is more geometric in nature. If \( T \) is convex and symmetric then with high probability, \( \text{diam}(T \cap \ker \Gamma, \cdot) \leq r_N \), where \( r_N = \inf\{r > 0 : E(T \cap r S^{n-1}) \leq c \sqrt{N} r\} \) for a well chosen absolute constant \( c \). In particular, for any such \( T \), \( \text{diam}(T \cap \ker \Gamma, |\cdot|) \leq c_1 E(T)/\sqrt{N} \), and for \( T = B_1^\circ \), the unit ball of \( \ell_1^\circ \), one has \( \text{diam}(T \cap \ker \Gamma, |\cdot|) \leq c_2 (\log(en/N))/\sqrt{N} \).

**References**


