

ORDER STATISTICS OF VECTORS WITH DEPENDENT COORDINATES, AND THE KARHUNEN–LOÈVE BASIS

BY ALEXANDER E. LITVAK AND KONSTANTIN TIKHOMIROV

University of Alberta

Let X be an n -dimensional random centered Gaussian vector with independent but not identically distributed coordinates and let T be an orthogonal transformation of \mathbb{R}^n . We show that the random vector $Y = T(X)$ satisfies

$$\mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} X_i^2 \leq C \mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} Y_i^2$$

for all $k \leq n$, where “ j -min” denotes the j th smallest component of the corresponding vector and $C > 0$ is a universal constant. This resolves (up to a multiplicative constant) an old question of S. Mallat and O. Zeitouni regarding optimality of the Karhunen–Loève basis for the nonlinear signal approximation. As a by-product, we obtain some relations for order statistics of random vectors (not only Gaussian) which are of independent interest.

1. Introduction. This work was motivated by the following question raised by S. Mallat and O. Zeitouni in 2000 (it was first posted on Zeitouni’s web page and later in arxiv [15]; see also [19]): Let n be a positive integer, and given $j \leq n$ and a sequence of real numbers a_1, a_2, \dots, a_n , let $j\text{-min}_{i \leq n} a_i$ denote its j th smallest element. Let X be an n -dimensional random Gaussian vector with independent centered coordinates (with possibly different variances). Further, let T be an orthogonal transformation of \mathbb{R}^n and set $Y := T(X)$.

Is it true that for every $k \leq n$, one has

$$(1) \quad \mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} X_i^2 \leq \mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} Y_i^2?$$

This problem has a natural interpretation within the field of signal processing (see [14], Chapter IX). Assume that a signal Y is modeled as an n -dimensional random centered Gaussian vector (for n very large). Our goal is to approximate Y by another vector which allows efficient storage and/or transmission through narrow bandwidth channels (let us note that this setting is distinct from the problem of signal *denoising* [14], Chapter 11, in which the goal is to produce an estimator for the mean of a noncentered signal). Let w_i ($i \leq n$) be a fixed orthonormal basis in

Received December 2016; revised May 2017.

MSC2010 subject classifications. 62G30, 60E15, 60G15, 60G35, 94A08.

Key words and phrases. Order statistics, Karhunen–Loève basis, nonlinear approximation, INID case.

\mathbb{R}^n , so that $Y = \sum_{i=1}^n \langle w_i, Y \rangle w_i$. The standard approach consists in approximating Y with a sparse vector with respect to that basis. The *linear m -term approximation* of Y with respect to the first m basis vectors is given by $\sum_{i=1}^m \langle w_i, Y \rangle w_i$, and the mean square error of the approximation is

$$\mathcal{E}_0(Y, m) = \mathbb{E} \sum_{i=m+1}^n \langle w_i, Y \rangle^2.$$

It is a well-known fact that $\mathcal{E}_0(Y, m)$ is minimized when the basis w_i ($i \leq n$) is the *Karhunen–Loève basis* for Y , that is, when the random variables $\langle w_i, Y \rangle$ ($i \leq n$) are pairwise uncorrelated, and the sequence $(\mathbb{E} \langle w_i, Y \rangle^2)_{i=1}^n$ is nonincreasing (see, e.g., [14], Theorem 9.8). Next, the *nonlinear m -term approximation* is defined as $\sum_{i \in \Lambda} \langle w_i, Y \rangle w_i$, where Λ is the (random) set of indices corresponding to m largest components of $(|\langle w_i, Y \rangle|)_{i \leq n}$. The nonlinear approximation error is given by

$$\mathcal{E}(Y, m) = \mathbb{E} \sum_{i \notin \Lambda} \langle w_i, Y \rangle^2 = \mathbb{E} \sum_{j=1}^{n-m} j \cdot \min_{i \leq n} \langle w_i, Y \rangle^2.$$

Now, observe that the expression on the left-hand side of (1) is the mean square error when approximating a signal X with uncorrelated coordinates with respect to the standard basis e_1, e_2, \dots, e_n using its largest $n - k$ components, and that the basis e_1, e_2, \dots, e_n is the *Karhunen–Loève basis* for X . The right-hand side of (1) corresponds to approximation of X with its $n - k$ largest components with respect to a basis $T^{-1}(e_1), T^{-1}(e_2), \dots, T^{-1}(e_n)$ (for some orthogonal transformation T). Thus, (1) is equivalent to saying that the Karhunen–Loève basis is optimal among all orthonormal bases in \mathbb{R}^n regarding the nonlinear approximation of centered Gaussian vectors. For more information on the signal approximation, we refer to [14], Chapter IX.

Note that the case $k = n$ is trivial. In [15] the authors solved the problem in the special case $k = n - 1$, that is, showed that

$$\mathbb{E} \sum_{j=1}^{n-1} j \cdot \min_{i \leq n} X_i^2 \leq \mathbb{E} \sum_{j=1}^{n-1} j \cdot \min_{i \leq n} Y_i^2.$$

This corresponds to the situation when the signal is approximated by its largest one-dimensional projection. In this paper, we verify (1) up to a multiplicative constant for all $k < n$.

THEOREM 1.1. *Let $1 \leq k < n$. Let X be an n -dimensional centered Gaussian vector with independent coordinates and T be an orthogonal transformation of \mathbb{R}^n . Then, setting $Y := T(X)$, we have*

$$\mathbb{E} \sum_{j=1}^k j \cdot \min_{i \leq n} X_i^2 \leq C \mathbb{E} \sum_{j=1}^k j \cdot \min_{i \leq n} Y_i^2,$$

where $C > 0$ is a universal constant.

The above theorem can be viewed as a relation between sums of *order statistics* of random vectors, in which one vector has independent coordinates and the other admits dependencies. Order statistics of random vectors are well-studied objects, and numerous results regarding their distribution are available. We refer to monography [2] for an account of developments prior to early 2000s. However, in the classical setting order statistics are defined for vectors with i.i.d. coordinates, with some generalizations available in the case of independent but not identically distributed components, as well as special kinds of dependencies (e.g., exchangeable or equicorrelated coordinates; see [2], Chapter 5). In our situation, we are working with coordinates which are *simultaneously* dependent and not equidistributed, making their analysis more problematic. Among recent works dealing with order statistics of vectors with dependent components, let us mention [1, 4, 6–8, 10, 11, 15]. In particular, ideas originated in papers [6, 7, 15] play an essential role in this note.

The proof of Theorem 1.1 can be roughly divided into two (unequal) parts. In the first part, which constitutes the novel element of this paper, we derive a comparison inequality for sums of order statistics of two random vectors, one with independent coordinates and the other with dependencies, under very general assumptions on the distribution of their components. In the second part, which essentially appeared already in [15], we utilize a inequality of A.W. Marshall and F. Proschan [16] (Theorem 2.5 below) to obtain a relation between variances of coordinates of a Gaussian vector and its orthogonal transformation, which, together with the first part, gives the statement of Theorem 1.1. The comparison inequality for order statistics is interesting on its own right, and we state it below as a separate theorem. It holds for a class of distributions satisfying rather mild conditions (see Theorem 5.4 below), however, to avoid technical complications here, we restrict ourselves to vectors with Gaussian components.

THEOREM 1.2. *Let $p > 0$, $1 \leq k \leq n$ and $0 < x_1 \leq \dots \leq x_n$. Let $\xi_i, \eta_i, i \leq n$, be standard Gaussian variables and assume in addition that $\xi_i, i \leq n$, are jointly independent. Then*

$$(2) \quad \mathbb{E} \sum_{j=1}^k j \cdot \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq 6(Cp)^p \mathbb{E} \sum_{j=1}^k j \cdot \min_{1 \leq i \leq n} |x_i \eta_i|^p,$$

where $C > 0$ is an absolute constant.

Note that the dependencies between variables $\eta_i, i \leq n$, can be arbitrary; in particular, we do not require vector (η_1, \dots, η_n) to have multivariate normal distribution.

We would like to mention that in the special case $k = 1$ Theorem 1.2 was previously established in [6, 7]; namely, it was shown that

$$(3) \quad \mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq \Gamma(2 + p) \mathbb{E} \min_{1 \leq i \leq n} |x_i \eta_i|^p.$$

In turn, the last inequality can be viewed as a natural counterpart to the well-known inequality of Šidák (see [3, 18]), asserting that $\mathbb{E} \max_{1 \leq i \leq n} |x_i \xi_i| \geq \mathbb{E} \max_{1 \leq i \leq n} |x_i \eta_i|$. Recently, R. van Handel has provided an example showing that one cannot make the constant multiple on the right-hand side of (3) equal to 1 even in the case $p = 1, n = 3$ [15].

We would also like to note that if j th minima ($1 \leq j \leq k$) in (2) are replaced by corresponding maxima then the expectation of the sum for independent components will be larger (up to a constant multiple), namely

$$(4) \quad C \mathbb{E} \sum_{j=1}^k j\text{-} \max_{1 \leq i \leq n} |x_i \xi_i|^p \geq \mathbb{E} \sum_{j=1}^k j\text{-} \max_{1 \leq i \leq n} |x_i \eta_i|^p,$$

where j -max denotes the j th largest element of corresponding sequences and C is an absolute positive constant. We refer to Theorem 4 in [4] (see also Theorem 2.4 in [5]), where this result was proved in a more general setting involving arbitrary Orlicz norms [note that the sum $\sum_{j=1}^k j\text{-} \max_{i \leq n} |z_i|$ is equivalent to an Orlicz norm of the sequence $(z_i)_{i \leq n}$]. In [17], this result was further extended to an even wider class of norms. We would like to emphasize that although

$$\begin{aligned} \mathbb{E} \sum_{j=1}^n j\text{-} \max_{1 \leq i \leq n} |x_i \eta_i|^p &= \mathbb{E} \sum_{j=1}^n j\text{-} \min_{1 \leq i \leq n} |x_i \eta_i|^p = \mathbb{E} \sum_{j=1}^n |x_i \eta_i|^p \\ &= \left(\sum_{j=1}^n |x_i|^p \right) \mathbb{E} |\eta_1|^p, \end{aligned}$$

the estimates (2) and (4) are incomparable—none of them implies the other one.

One of important ingredients in the proof of Theorem 1.2 is a statement which provides optimal estimates for sums of the smallest order statistics in case of independent components (see Theorem 5.1). The proof is based on using special functionals which were previously employed in papers [6–8].

Another novel element is an argument for working with dependent components (see Theorem 4.5). Absence of such a tool in preceding works [6, 7] was a major obstacle to proving the Mallat–Zeitouni conjecture, even up to a multiple depending on k . The proof of Theorem 4.5 is essentially reduced to considering uniformly bounded dependent variables.

The paper is organized as follows. In Section 2, we fix notation and provide auxiliary statements. Additionally, we introduce several special conditions on distributions which are assumed (in various combinations) in our main statements. Section 3 contains some known results on individual order statistics, which we use later in the paper. For the sake of completeness, we provide the proofs, but we postpone them to Section 8. Section 4 provides new bounds for individual order statistics playing a crucial role in the proof of the main results. The next two sections are devoted to proving Theorems 1.1 and 1.2. In Section 7, we briefly discuss efficiency of the nonlinear approximation based on the largest projections, compared to the linear approximation.

2. Notation and preliminaries. Given a subset $A \subset \mathbb{N}$, we denote its cardinality by $|A|$. Next, for a natural number n and a set $E \subset \{1, 2, \dots, n\}$, we denote by E^c the complement of E inside $\{1, 2, \dots, n\}$. Similarly, for an event A we denote by A^c the complement of the event. Further, we say that a collection of sets $(A_j)_{j \leq k}$ is a partition of $\{1, 2, \dots, n\}$ if each A_j is nonempty, the sets are pairwise disjoint and their union is $\{1, 2, \dots, n\}$. The canonical Euclidean norm and the canonical inner product in \mathbb{R}^n will be denoted by $|\cdot|$ and $\langle \cdot, \cdot \rangle$, respectively. We adopt the conventions $1/0 = \infty$ and $1/\infty = 0$ throughout the text. For a given sequence of real numbers a_1, a_2, \dots, a_n , we denote its k th smallest element by $k\text{-min}_{1 \leq i \leq n} a_i$. In particular, $1\text{-min}_{1 \leq i \leq n} a_i = \min_{1 \leq i \leq n} a_i$, and $(k\text{-min}_{1 \leq i \leq n} a_i)_{k=1}^n$ is the nondecreasing rearrangement of the sequence $(a_i)_{i=1}^n$. As usual, we use the abbreviation c.d.f. for the cumulative distribution function [that is, given a random variable ξ , the c.d.f. of ξ is $F(t) = \mathbb{P}(\xi \leq t)$].

Next, we group together a few combinatorial results which provide basic tools for estimating order statistics in next sections. Let us start with the following simple property of $k\text{-min}_{1 \leq i \leq n} a_i$ which holds for every real sequence $(a_i)_{i=1}^n$: For every partition $(A_j)_{j \leq k}$ of $\{1, 2, \dots, n\}$, one has

$$(5) \quad \sum_{j=1}^k j\text{-min}_{1 \leq i \leq n} a_i \leq \sum_{j=1}^k \min_{i \in A_j} a_i.$$

The next statement is a classical inequality for symmetric means.

THEOREM 2.1 (C. Maclaurin; see [9], Theorem 52). *Let $1 \leq \ell \leq n$ and let a_1, \dots, a_n be nonnegative real numbers. Then*

$$\sum_{\substack{A \subset \{1, 2, \dots, n\} \\ |A| = \ell}} \prod_{i \in A} a_i \leq \binom{n}{\ell} \left(\frac{1}{n} \sum_{i=1}^n a_i \right)^\ell.$$

In [7], it was shown that the above statement, together with Stirling’s formula, implies the following.

COROLLARY 2.2. *Let $1 \leq k \leq n$. Let a_1, \dots, a_n be nonnegative real numbers and assume that*

$$0 < a := \frac{e}{k} \sum_{i=1}^n a_i < 1.$$

Then

$$\sum_{\ell=k}^n \sum_{\substack{A \subset \{1, 2, \dots, n\} \\ |A| = \ell}} \prod_{i \in A} a_i < \frac{1}{\sqrt{2\pi k}} \frac{a^k}{1-a}.$$

The following statement was essentially obtained in [7] (cf. Lemma 4 there). We reproduce the argument in Section 8 for the reader’s convenience.

LEMMA 2.3. *Let $1 \leq k \leq n$ and let $(a_i)_{i=1}^n$ be a nonincreasing sequence of positive real numbers. For each $j \leq n$, set $b_j := \sum_{i=j}^n a_i$ and let $m \leq k$ be the smallest integer such that*

$$a_m \leq \frac{b_m}{k + 1 - m}.$$

Then there exists a partition $(A_j)_{j \leq k}$ of $\{1, 2, \dots, n\}$ such that $A_j = \{j\}$ for $j < m$ and for every $j \geq m$ we have

$$\sum_{i \in A_j} a_i \geq \frac{b_m}{2(k + 1 - m)}.$$

REMARK 2.4. In fact, as one can see from the proof below, the sets A_j can be chosen as intervals, that is, $A_j = \{i \leq n : n_{j-1} < i \leq n_j\}$, $j \leq k$, for some sequence $0 = n_0 < n_1 < \dots < n_k = n$. Moreover, with the partition used in the proof we also have

$$\min_{1 \leq \ell \leq k} \sum_{i \in A_\ell} a_i \geq \frac{1}{2} \min_{1 \leq j \leq k} \frac{1}{k + 1 - j} \sum_{i=j}^n a_i.$$

Next, we introduce several conditions on distributions of random variables. Let $\alpha > 0$ and $\beta > 0$ be parameters. We say that a random variable ξ satisfies the α -condition if

$$(6) \quad \mathbb{P}(|\xi| \leq t) \leq \alpha t \quad \text{for every } t \geq 0$$

and ξ satisfies the β -condition if

$$(7) \quad \mathbb{P}(|\xi| > t) \leq e^{-\beta t} \quad \text{for every } t \geq 0.$$

If both (6) and (7) hold, then we say that ξ satisfies the (α, β) -condition. Note that in this case we necessarily have $\alpha t + e^{-\beta t} \geq 1$ for all $t \geq 0$, which can be true only for $\alpha \geq \beta$. In [7], it was shown that for any $q \geq 1$, a nonnegative random variable ξ with the density function $p(s) = c_q \exp(-s^q)$ ($s \geq 0$), where $c_q := 1/\Gamma(1 + 1/q)$, satisfies (6) and (7) with parameters $\alpha = \beta = c_q$. In particular, for $q = 2$ we get a Gaussian random variable $\mathcal{N}(0, 1/2)$, and $\alpha = \beta = c_2 = 2/\sqrt{\pi}$. This easily implies that the standard Gaussian variable satisfies (6) and (7) with $\alpha = \beta = \sqrt{2/\pi}$. Note also that for $q = 1$ we have an exponentially distributed random variable satisfying the (α, β) -condition with $\alpha = \beta = 1$. Finally, it is not difficult to check that any centered log-concave random variable satisfies the (α, β) -condition for some α and β .

We will employ one more condition on a c.d.f. F of a nonnegative random variable:

(8) *there exist $\delta \in (0, 1)$, $A > 1$ such that $F(t) \geq 2F(t/A)$ whenever $F(t) \leq \delta$.*

Note that the multiple “2” on the right-hand side of (8) can be replaced with any number $a > 1$, at expense of increasing A and decreasing δ .

Finally, we state the following result of Marshall and Proschan, which will be used in Section 6.

THEOREM 2.5 ([16]). *Let ξ_1, \dots, ξ_n be interchangeable random variables (i.e., with the joint distribution invariant under permutations of arguments). Let $(a_i)_{i \leq n}$ and $(b_i)_{i \leq n}$ be nonnegative nonincreasing sequences such that for every $\ell \leq n$:*

$$\sum_{i=1}^{\ell} a_i \geq \sum_{i=1}^{\ell} b_i \quad \text{and} \quad \sum_{i=1}^n a_i = \sum_{i=1}^n b_i.$$

Let φ be a continuous convex function symmetric in its n arguments. Then

$$\mathbb{E}\varphi(a_1\xi_1, \dots, a_n\xi_n) \geq \mathbb{E}\varphi(b_1\xi_1, \dots, b_n\xi_n).$$

3. Known bounds for individual order statistics. In this section, we recall some of results from papers [6, 7] concerning order statistics. For the sake of completeness, we provide their proofs in Section 8.

LEMMA 3.1. *Let $\alpha > 0$ and $p > 0$. Let $0 < x_1 \leq x_2 \leq \dots \leq x_n$ be real numbers and let ξ_1, \dots, ξ_n be (possibly dependent) random variables satisfying the α -condition. Finally, set $b := \sum_{i=1}^n 1/x_i$. Then for every $t > 0$ we have*

$$\mathbb{P}\left\{ \min_{1 \leq i \leq n} |x_i \xi_i| \leq t \right\} \leq \alpha b t.$$

In particular,

$$\text{Med}\left(\min_{1 \leq i \leq n} |x_i \xi_i|^p \right) \geq \frac{1}{2^p \alpha^p b^p} \quad \text{and} \quad \mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p \geq \frac{1}{(1+p)\alpha^p b^p}.$$

LEMMA 3.2. *Let $\beta > 0$ and $p > 0$. Let $0 < x_1 \leq x_2 \leq \dots \leq x_n$ be real numbers and let ξ_1, \dots, ξ_n be independent random variables satisfying the β -condition. Set $b := \sum_{i=1}^n 1/x_i$. Then for every $t > 0$, we have*

$$\mathbb{P}\left\{ \min_{1 \leq i \leq n} |x_i \xi_i| > t \right\} \leq e^{-\beta b t}.$$

In particular,

$$\text{Med}\left(\min_{1 \leq i \leq n} |x_i \xi_i|^p \right) \leq \frac{(\ln 2)^p}{\beta^p b^p} \quad \text{and} \quad \mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq \frac{\Gamma(1+p)}{\beta^p b^p}.$$

An immediate consequence of the above lemmas is the following statement.

COROLLARY 3.3. *Let $p > 0$. Let $(x_i)_{i=1}^n$ be a sequence of real numbers and $f_1, \dots, f_n, \xi_1, \dots, \xi_n$ be random variables satisfying the (α, β) -condition for some $\alpha, \beta > 0$. Assume additionally that the ξ_i 's are jointly independent. Then*

$$\mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq \Gamma(2 + p) \alpha^p \beta^{-p} \mathbb{E} \min_{1 \leq i \leq n} |x_i f_i|^p.$$

In particular, if $f_1, \dots, f_n, \xi_1, \dots, \xi_n$ are $N(0, 1)$ Gaussian random variables, then

$$\mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq \Gamma(2 + p) \mathbb{E} \min_{1 \leq i \leq n} |x_i f_i|^p.$$

The next lemma deals with order statistics other than the smallest one.

LEMMA 3.4. *Let $\alpha > 0, p > 0$ and $1 \leq k \leq n$. Further, let $0 < x_1 \leq x_2 \leq \dots \leq x_n$ be real numbers and let ξ_1, \dots, ξ_n be independent random variables satisfying the α -condition. Set $b := \sum_{i=1}^n 1/x_i, a := \alpha b/k$. Then for every $0 < t < 1/a$ we have*

$$\mathbb{P}\left\{k\text{-} \min_{1 \leq i \leq n} |x_i \xi_i| \leq t\right\} \leq \frac{1}{\sqrt{2\pi k}} \frac{(at)^k}{1 - at}$$

and

$$\frac{1}{2^{1/p} 4\alpha} \max_{1 \leq j \leq k} \frac{k + 1 - j}{\sum_{i=j}^n 1/x_i} \leq \left(\mathbb{E} k\text{-} \min_{1 \leq i \leq n} |x_i \xi_i|^p\right)^{1/p}.$$

REMARK 3.5. Using Lemmas 3.2, 2.3 (with Remark 2.4) and ideas similar to ones used in the proof of Theorem 5.1 below, it was shown in [7] that for variables satisfying the β -condition we have

$$\left(\mathbb{E} k\text{-} \min_{1 \leq i \leq n} |x_i \xi_i|^p\right)^{1/p} \leq C(p, k) \beta^{-1} \max_{1 \leq j \leq k} \frac{k + 1 - j}{\sum_{i=j}^n 1/x_i},$$

where $C(p, k) := C \max\{p, \ln(k + 1)\}$, and C is an absolute positive constant. Moreover, in [8] it was shown that the expectation above is equivalent to some Orlicz norm (up to a factor logarithmic in k).

4. New bounds for individual order statistics. Let ξ be a real-valued random variable and let $F = F_\xi$ be its c.d.f. Let $r \in [0, 1]$. By $q(r) = q_F(r) = q_\xi(r)$, we denote a quantile of order r , that is, a number satisfying

$$\mathbb{P}\{\xi < q(r)\} \leq r \quad \text{and} \quad \mathbb{P}\{\xi \leq q(r)\} \geq r$$

[note that in general $q(r)$ is not uniquely defined]. The following claim provides simple lower bounds on quantiles for a large class of random variables.

CLAIM 4.1. *Let $1 \leq k \leq n$ and $0 < x_1 \leq \dots \leq x_n$. For each $j \leq n$, set $b_j := \sum_{i=j}^n 1/x_i$. Further, let $\xi_i, i \leq n$, be (possibly dependent) random variables satisfying the α -condition for some $\alpha > 0$, and for every $i \leq n$ let F_i be the c.d.f. of $|x_i \xi_i|$. Denote*

$$F := \frac{1}{n} \sum_{i=1}^n F_i \quad \text{and} \quad q := q_F\left(\frac{k-1/2}{n}\right).$$

Then

$$q \geq \frac{1}{2\alpha} \max_{1 \leq j \leq k} \frac{k-j+1}{b_j}.$$

PROOF. By the above definitions, for every $j \leq k$ we have

$$k-1/2 \leq \sum_{i=1}^n F_i(q) \leq j-1 + \sum_{i=j}^n \alpha q/x_i = j-1 + \alpha q b_j,$$

which implies the result. \square

REMARK 4.2. It is not difficult to check that when all ξ_i 's are uniformly distributed on $[0, 1]$, we have

$$q_F\left(\frac{k-1/2}{n}\right) = \max_{1 \leq j \leq k} \frac{k-j+1/2}{b_j}.$$

The next lemma provides lower estimates for order statistics of possibly dependent random variables via quantiles of their truncations.

LEMMA 4.3. *Let $\delta \in (0, 1)$, $A > 1$ and $x_1, \dots, x_n > 0$. Let $\xi_i, i \leq n$, be (possibly dependent) random variables satisfying condition (8) with parameters δ and A . Further, define*

$$t_0 := \min_{i \leq n} \sup\{t > 0 : F_{|\xi_i|}(t) \leq \delta\} \quad \text{and} \quad \eta_i := \min(|\xi_i|, t_0), \quad i \leq n.$$

For every $i \leq n$, we let F_i be the c.d.f. of $x_i \eta_i$. Define

$$F = \frac{1}{n} \sum_{i=1}^n F_i.$$

Then

$$\text{Med}\left(k - \min_{1 \leq i \leq n} |x_i \xi_i|\right) \geq \frac{q_F\left(\frac{k-1/2}{n}\right)}{A}.$$

REMARK 4.4. It may seem natural to obtain a bound for the median in terms of the ‘‘averaged’’ c.d.f. with respect to the original variables ξ_i and not their trun-

cations η_i . The following example (cf. Example 12 in [7]) shows that in fact the truncation is essential. Consider independent standard Gaussian random variables g_1, g_2, \dots, g_n and let $\xi_i := g_1$ for $i \leq n$. Clearly, these random variables satisfy condition (8) with some A and δ . Let $x_1 = \dots = x_k = 1$ and $x_{k+1} = \dots = x_n = n^2$. Then a direct computation shows that for $G := \frac{1}{n} \sum_{i=1}^n F_{|x_i g_i|}$ we have

$$q_G\left(\frac{k-1/2}{n}\right) \approx \text{Med}\left(k - \min_{1 \leq i \leq n} |x_i g_i|\right) \approx \sqrt{\ln k}$$

while

$$\text{Med}\left(k - \min_{1 \leq i \leq n} |x_i \xi_i|\right) \approx \text{const.}$$

PROOF OF LEMMA 4.3. Clearly, $F_{\eta_i}(t) = F_{|\xi_i|}(t) \leq \delta$ for $t < t_0$ and $F_{\eta_i}(t) = 1$ for $t \geq t_0$, and that F_{η_i} also satisfies condition (8) with parameters δ and A . Fix some positive $s < q_F(\frac{k-1/2}{n})/A$ and denote

$$I := \{i \leq n : F_i(As) = 1\}.$$

By the choice of s , we have $\sum_{i \leq n} F_i(As) = nF(As) < k - 1/2$, whence $|I| < k$. Note also that for every $i \notin I$ we have $F_i(As) \leq \delta$. Thus, applying condition (8), we get

$$\begin{aligned} \mathbb{E}|\{i \in I^c : x_i \eta_i < s\}| &= \mathbb{E} \sum_{i \in I^c} \chi_{\{x_i \eta_i < s\}} \leq \sum_{i \in I^c} F_i(s) \\ &\leq \frac{1}{2} \sum_{i \in I^c} F_i(As) = \frac{nF(As) - |I|}{2} < \frac{k - |I|}{2}. \end{aligned}$$

By Markov's inequality, this implies

$$\mathbb{P}(|\{i \in I^c : x_i \eta_i < s\}| \geq k - |I|) \leq \frac{1}{2},$$

whence

$$\mathbb{P}(|\{i \leq n : x_i \eta_i < s\}| \geq k) \leq \frac{1}{2}.$$

Since the event $\{k - \min_{1 \leq i \leq n} x_i \eta_i \geq s\}$ coincides with the event $\{|\{i \leq n : x_i \eta_i < s\}| < k\}$, we obtain

$$\mathbb{P}\left(k - \min_{1 \leq i \leq n} x_i \eta_i \geq s\right) \geq \frac{1}{2},$$

that is,

$$\text{Med}\left(k - \min_{1 \leq i \leq n} |x_i \xi_i|\right) \geq \text{Med}\left(k - \min_{1 \leq i \leq n} |x_i \eta_i|\right) \geq s.$$

Since s was an arbitrary number smaller than $q_F(\frac{k-1/2}{n})/A$, the proof is complete. □

Now, let us formulate a new theorem on order statistics, which essentially states that the lower bound for expectation in Lemma 3.4 does not require independence.

THEOREM 4.5. *Let $\alpha > 0$, $\delta \in (0, 1)$, $A > 1$, $1 \leq k \leq n$ and $0 < x_1 \leq \dots \leq x_n$. For each $j \leq n$, we set $b_j := \sum_{i=j}^n 1/x_i$. Further, let ξ_i , $i \leq n$, be (possibly dependent) random variables satisfying the α -condition and condition (8) with parameters δ and A . Then*

$$\text{Med}\left(k - \min_{1 \leq i \leq n} |x_i \xi_i|\right) \geq \frac{\delta}{2A\alpha} \max_{1 \leq j \leq k} \frac{k - j + 1}{b_j}.$$

PROOF. Let the number t_0 and random variables η_i , $i \leq n$ be defined as in Lemma 4.3. Note that the α -condition on ξ_i 's implies $F_{|\xi_i|}(t) \leq \delta$, $i \leq n$, for $t \leq \delta/\alpha$. Hence, $t_0 \geq \delta/\alpha$. Thus, for every $i \leq n$ we have $F_{\eta_i}(t) \leq \alpha t$ whenever $t < t_0$ and $F_{\eta_i}(t) = 1 \leq (\alpha/\delta)t$ otherwise. In other words, the random variables η_i 's satisfy condition (6) with α/δ replacing α . Combining Lemma 4.3 with Claim 4.1, applied to η_i 's, we obtain the result. \square

In view of Remark 3.5, Theorem 4.5 has the following consequence.

COROLLARY 4.6. *Under the conditions of Theorem 4.5, assuming that independent random variables η_1, \dots, η_n satisfy β -condition with some $\beta > 0$, one has for every $p > 0$,*

$$\left(\mathbb{E}k - \min_{1 \leq i \leq n} |x_i \eta_i|^p\right)^{1/p} \leq \frac{C2^{1/p}A\alpha}{\beta\delta} \max\{p, \ln(k + 1)\} \left(\mathbb{E}k - \min_{1 \leq i \leq n} |x_i \xi_i|^p\right)^{1/p},$$

where C is an absolute positive constant.

REMARK 4.7. The logarithmic factor in Corollary 4.6 cannot be removed as the following example shows. Let ξ be a positive exponential random variable, that is a random variable with the density function $f(t) = e^{-t}$ for $t \geq 0$ and $f(t) = 0$ for $t < 0$. Let η_i , $i \leq n$, be independent copies of ξ and $\xi_1 = \dots = \xi_n = \xi$. Let $x_1 = \dots = x_k = 1$ and $x_{k+1} = \dots = x_n = n^2$. Then

$$\mathbb{E}k - \min_{1 \leq i \leq n} |x_i \eta_i| \approx \mathbb{E} \max_{1 \leq i \leq k} |\eta_i| \approx \ln(k + 1) \quad \text{while} \quad \mathbb{E}k - \min_{1 \leq i \leq n} |x_i \xi_i| = \mathbb{E}|\xi| = 1.$$

5. Bounds for sums of order statistics.

THEOREM 5.1. *Let $p > 0$ and let ξ_i , $i \leq n$, be independent random variables satisfying the (α, β) -condition for some $\alpha, \beta > 0$. Let $0 < x_1 \leq \dots \leq x_n$. For each*

$j \leq n$, set $b_j := \sum_{i=j}^n 1/x_i$. Then for every $k \leq n$ we have

$$\begin{aligned} \frac{1}{2} \left(\frac{1}{16\alpha} \right)^p \sum_{j=1}^k \frac{(k-j+1)^p}{b_j^p} &\leq \mathbb{E} \sum_{j=1}^k j \cdot \min_{1 \leq i \leq n} |x_i \xi_i|^p \\ &\leq W(\beta, p) \sum_{j=1}^k \frac{(k-j+1)^p}{b_j^p}, \end{aligned}$$

where $W(\beta, p) := \beta^{-p} \Gamma(1+p)(1+2 \cdot 4^p)$.

REMARK 5.2. The upper bound can be replaced with a slightly stronger equivalent estimate

$$\beta^{-p} \Gamma(1+p) \sum_{j=1}^{m-1} x_j^p + 2^p \beta^{-p} \Gamma(1+p) \frac{(k-m+1)^{1+p}}{b_m^p},$$

where $m \leq k$ is the smallest positive integer such that

$$\frac{1}{x_m} \leq \frac{b_m}{k+1-m}$$

(see the proof below).

We will need the following calculus lemma.

LEMMA 5.3. Let $p > 0$ and $0 < x_1 \leq \dots \leq x_n$. For $j \leq n$, set $b_j := \sum_{i=j}^n 1/x_i$. Then

$$4^p \sum_{\ell=1}^k \max_{1 \leq j \leq \ell} \frac{(\ell-j+1)^p}{b_j^p} \geq \sum_{j=1}^k \frac{(k-j+1)^p}{b_j^p} \geq 2^{-1-p} \max_{1 \leq j \leq k} \frac{(k-j+1)^{1+p}}{b_j^p}.$$

PROOF. For some fixed $1 \leq s \leq k$, let $\ell := \lceil (k+s)/2 \rceil$. Then

$$\frac{k-s+1}{b_s} \leq 2 \frac{\ell-s+1}{b_s} \leq 2 \max_{1 \leq j \leq \ell} \frac{\ell-j+1}{b_j}.$$

Hence,

$$\sum_{s=1}^k \frac{(k-s+1)^p}{b_s^p} \leq 4^p \sum_{\ell=\lceil k/2 \rceil}^k \max_{1 \leq j \leq \ell} \frac{(\ell-j+1)^p}{b_j^p},$$

which implies the left-hand side inequality.

Now let $s \leq k$ be such that

$$\max_{1 \leq j \leq k} \frac{(k-j+1)^{1+p}}{b_j^p} = \frac{(k-s+1)^{1+p}}{b_s^p}.$$

Let $t := \lfloor (k + s + 1)/2 \rfloor$. Then

$$\begin{aligned} \sum_{j=1}^k \frac{(k - j + 1)^p}{b_j^p} &\geq \sum_{j=s}^t \frac{(k - t + 1)^p}{b_s^p} \geq (t - s + 1) \frac{(k - s + 1)^p}{(2b_s)^p} \\ &\geq \frac{(k - s + 1)^{1+p}}{2^{1+p} b_s^p}, \end{aligned}$$

which completes the proof. \square

PROOF OF THEOREM 5.1. For the lower bound, by Lemma 3.4 we have

$$\mathbb{E} \sum_{j=1}^k j \cdot \min_{1 \leq i \leq n} |x_i \xi_i|^p \geq \frac{1}{2(4\alpha)^p} \sum_{\ell=1}^k \max_{1 \leq j \leq \ell} \frac{(\ell - j + 1)^p}{(\sum_{i=j}^n 1/x_i)^p},$$

and it remains to apply Lemma 5.3 (alternatively, under slightly modified assumptions on random variables we could use Theorem 4.5). Let us prove the upper bound. Let $B_p := \beta^{-p} \Gamma(1 + p)$. Let the integer m and the partition $(A_j)_{j \leq k}$ be given by Lemma 2.3 applied to the sequence $(a_i)_{i \leq n} := (1/x_i)_{i \leq n}$. Using (5) and Lemma 3.2, we get

$$\mathbb{E} \sum_{j=1}^k j \cdot \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq \mathbb{E} \sum_{j=1}^k \min_{i \in A_j} |x_i \xi_i|^p \leq B_p \sum_{j=1}^{m-1} x_j^p + B_p \sum_{j=m}^k \left(\sum_{i \in A_j} 1/x_i \right)^{-p}.$$

Next, note that by the choice of the partition $(A_j)_{j \leq k}$ we have

$$\sum_{j=m}^k \left(\sum_{i \in A_j} 1/x_i \right)^{-p} \leq 2^p \frac{(k - m + 1)^{1+p}}{b_m^p} \leq 2^p \max_{1 \leq j \leq k} \frac{(k - j + 1)^{1+p}}{b_j^p}.$$

Further, applying the definition of m to numbers $x_j, j < m$, we obtain

$$\frac{1}{x_j} > \frac{b_j}{k + 1 - j}, \quad j < m,$$

whence

$$\sum_{j=1}^{m-1} x_j^p \leq \sum_{j=1}^{m-1} \frac{(k - j + 1)^p}{b_j^p}.$$

Combining the estimates and applying the rightmost estimate from Lemma 5.3, we get

$$\begin{aligned} \mathbb{E} \sum_{j=1}^k j \cdot \min_{1 \leq i \leq n} |x_i \xi_i|^p &\leq B_p \sum_{j=1}^{m-1} \frac{(k - j + 1)^p}{b_j^p} + 2^p B_p \max_{1 \leq j \leq k} \frac{(k - j + 1)^{1+p}}{b_j^p} \\ &\leq B_p (1 + 2 \cdot 4^p) \sum_{j=1}^k \frac{(k - j + 1)^p}{b_j^p}, \end{aligned}$$

and the proof is complete. \square

Finally, we formulate the comparison theorem for sums of order statistics (the second part of the theorem below was stated in the [Introduction](#) as Theorem 1.2).

THEOREM 5.4. *Let $p, \alpha, \beta > 0$, $\delta \in (0, 1)$ and $A > 1$. Let $1 \leq k \leq n$ and $0 < x_1 \leq \dots \leq x_n$. Further, let $\xi_i, \eta_i, i \leq n$, be random variables satisfying the (α, β) -condition and condition (8) with parameters δ and A . Assume in addition that $\xi_i, i \leq n$, are jointly independent. Then*

$$\mathbb{E} \sum_{j=1}^k j\text{-} \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq 6 \left(\frac{32A\alpha}{\delta\beta} \right)^p \Gamma(1+p) \mathbb{E} \sum_{j=1}^k j\text{-} \min_{1 \leq i \leq n} |x_i \eta_i|^p.$$

In particular, if ξ_i, η_i are standard Gaussian variables then

$$\mathbb{E} \sum_{j=1}^k j\text{-} \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq 6(Cp)^p \mathbb{E} \sum_{j=1}^k j\text{-} \min_{1 \leq i \leq n} |x_i \eta_i|^p,$$

where $C > 0$ is an absolute constant.

PROOF. In view of Theorem 4.5, we have

$$\begin{aligned} \mathbb{E} \sum_{j=1}^k j\text{-} \min_{1 \leq i \leq n} |x_i \eta_i|^p &\geq \frac{1}{2} \sum_{j=1}^k \left(\text{Med} \left(j\text{-} \min_{1 \leq i \leq n} |x_i \eta_i| \right) \right)^p \\ &\geq \frac{1}{2} \frac{\delta^p}{(2A\alpha)^p} \sum_{\ell=1}^k \max_{1 \leq j \leq \ell} \frac{(\ell - j + 1)^p}{b_j^p}. \end{aligned}$$

Hence, by Lemma 5.3 we get

$$\mathbb{E} \sum_{j=1}^k j\text{-} \min_{1 \leq i \leq n} |x_i \eta_i|^p \geq \frac{\delta^p}{2(8A\alpha)^p} \sum_{j=1}^k \frac{(k - j + 1)^p}{b_j^p}.$$

It remains to apply Theorem 5.1. \square

6. Proof of Theorem 1.1. In [15], it was shown that Theorem 5.4 implies Theorem 1.1. For the sake of completeness we outline the proof here.

Note that for every sequence $(z_i)_{i=1}^n$ and every permutation σ of $\{1, \dots, n\}$ one has

$$(9) \quad \sum_{j=1}^k j\text{-} \min_{i \leq n} z_i = \sum_{j=1}^k j\text{-} \min_{i \leq n} z_{\sigma(i)}.$$

Let $\bar{T} = (\bar{t}_{ij})_{ij}$ be an orthogonal transformation of \mathbb{R}^n , $X = (X_1, \dots, X_n)$ be a centered Gaussian vector with independent components and set $Y = (Y_1, \dots, Y_n) := \bar{T}X$. Fix any $k < n$. For each $i \leq n$, denote the variance of X_i

by \bar{a}_i and the variance of Y_i by \bar{b}_i . By $(a_i)_{i \leq n}$ and $(b_i)_{i \leq n}$ we denote the nonincreasing rearrangements of $(\bar{a}_i)_{i \leq n}$ and $(\bar{b}_i)_{i \leq n}$, and let σ and π be permutations of $\{1, \dots, n\}$ such that $a_i = \bar{a}_{\sigma(i)}$ and $b_i = \bar{b}_{\pi(i)}$ for all $i \leq n$. By (9), we have

$$\mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} X_i^2 = \mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} X_{\sigma(i)}^2 \quad \text{and} \quad \mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} Y_i^2 = \mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} Y_{\pi(i)}^2.$$

For $i, j \leq n$, denote $t_{ij} = \bar{t}_{\pi(i)\sigma(j)}$ and $T = (t_{ij})_{i,j}$, that is, the matrix T is obtained from \bar{T} by multiplying it by permutation matrices corresponding to σ and π . Clearly, T is also orthogonal. Since the coordinates of X are independent, for every $i \leq n$ we have

$$b_i = \sum_{j=1}^n t_{ij}^2 a_j.$$

As T is an orthogonal matrix, $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$. Now we show that for every $\ell < n$ one has

$$\sum_{i=1}^{\ell} a_i \geq \sum_{i=1}^{\ell} b_i.$$

First, note that the case $\ell = 1$ follows by the orthogonality of T and because $(a_i)_{i \leq n}$ is nonincreasing. For $\ell > 1$, again using the orthogonality of T and monotonicity of $(a_i)_i$, we obtain

$$\begin{aligned} \sum_{i=1}^{\ell} b_i &= \sum_{i=1}^{\ell} \sum_{j=1}^n t_{ij}^2 a_j = \sum_{j=1}^{\ell-1} \sum_{i=1}^{\ell} t_{ij}^2 a_j + \sum_{j=\ell}^n \sum_{i=1}^{\ell} t_{ij}^2 a_j \\ &\leq \sum_{j=1}^{\ell-1} \sum_{i=1}^{\ell} t_{ij}^2 a_j + a_{\ell} \sum_{j=\ell}^n \sum_{i=1}^{\ell} t_{ij}^2 \\ &= \sum_{j=1}^{\ell-1} a_j + \sum_{j=1}^{\ell-1} a_j \left(\sum_{i=1}^{\ell} t_{ij}^2 - 1 \right) + a_{\ell} \left(\ell - \sum_{j=1}^{\ell-1} \sum_{i=1}^{\ell} t_{ij}^2 \right) \\ &= \sum_{j=1}^{\ell} a_j + \sum_{j=1}^{\ell-1} (a_j - a_{\ell}) \left(\sum_{i=1}^{\ell} t_{ij}^2 - 1 \right) \\ &\leq \sum_{i=1}^{\ell} a_i. \end{aligned}$$

Note that

$$\|(x_1, \dots, x_n)\| = \left(\sum_{j=1}^{n-k} j\text{-max}_{i \leq n} x_i^2 \right)^{1/2} = \left(\sum_{j=k+1}^n j\text{-min}_{i \leq n} x_i^2 \right)^{1/2},$$

defines a norm on \mathbb{R}^n (recall that j -max is j th maximum of the corresponding sequence). Therefore, the function

$$\varphi(x_1, \dots, x_n) = \sum_{j=k+1}^n j\text{-min}_{i \leq n} x_i^2$$

is convex and thus Theorem 2.5 yields

$$\mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} X_i^2 \leq \mathbb{E} \sum_{j=1}^k j\text{-min}_{i \leq n} (b_i g_i^2),$$

where g_1, \dots, g_n are i.i.d. standard Gaussian variables. Theorem 5.4 completes the proof.

7. Efficiency of the nonlinear approximation. In this section, we briefly discuss the following question: How efficient is the nonlinear approximation based on the largest projections, compared to the linear approximation with respect to the same basis? In what follows, we fix the dimension n . Given a centered random vector X with a well-defined covariance matrix (i.e., each component of X has a bounded variance), denote by $\mathcal{E}(X, m)$ the mean square error of the nonlinear approximation based on m largest projections onto the standard basis vectors, that is,

$$\mathcal{E}(X, m) := \mathbb{E} \sum_{j=1}^{n-m} j\text{-min}_{i \leq n} X_i^2, \quad m < n.$$

Further, we define corresponding error for the linear approximation as

$$\mathcal{E}_0(X, m) := \min_{|J|=n-m} \mathbb{E} \sum_{i \in J} X_i^2 = \sum_{j=1}^{n-m} j\text{-min}_{i \leq n} (\mathbb{E} X_i^2),$$

where the minimum is taken over all subsets of $\{1, 2, \dots, n\}$ of cardinality $n - m$. Obviously, we have

$$(10) \quad \mathcal{E}(X, m) \leq \mathcal{E}_0(X, m)$$

for all $m < n$. Moreover, if for a fixed m we define a random Gaussian vector \tilde{X} with independent components and

$$\mathbb{E} \tilde{X}_i^2 = \begin{cases} 1 & \text{if } i \leq m + 1, \\ 0 & \text{if } i > m + 1, \end{cases}$$

for all $i \leq n$, then it can be checked that $\mathcal{E}_0(\tilde{X}, m) = 1$ whereas $\mathcal{E}(\tilde{X}, m) \approx m^{-2}$. Thus, the nonlinear approximation can in some cases be significantly more efficient than the linear approximation as long as the number of projections is the same. However, as we show below, some kind of a *reverse* inequality for (10) is possible under quite general assumptions on the distribution, if we are allowed to slightly increase the number of projections for the linear approximation:

PROPOSITION 7.1. *Let $u > 0, m < n/2$ and let X be a centered random vector in \mathbb{R}^n with a well defined covariance matrix such that*

$$(11) \quad u\mathbb{E}X_i^2 \leq \int_0^\infty \max\left(\mathbb{P}\{X_i^2 \geq \tau\} - \frac{1}{2}, 0\right) d\tau, \quad i \leq n.$$

Then we have

$$u\mathcal{E}_0(X, 2m) \leq \mathcal{E}(X, m).$$

Before proving the proposition, we would like to remark that condition (11) is invariant with respect to scalar multiplication of X_i 's. Note also that it is satisfied, in particular, for any centered Gaussian random vector with $u = 1/20$. Moreover, this condition holds with $u = \beta^2/(48\alpha^2)$ for random variables satisfying the (α, β) -condition. Indeed, for such a variable ξ , denoting by M the median of ξ^2 , we have by the β -condition

$$\mathbb{E}\xi^2 = 2 \int_0^\infty t\mathbb{P}\{|\xi| > t\} dt \leq 2 \int_0^\infty t \exp(-\beta t) dt = \frac{2}{\beta^2}$$

and, by the α -condition, $M \geq 1/(4\alpha^2)$ and

$$\int_0^M \mathbb{P}\{\xi^2 \geq t\} dt \geq \int_0^{1/(4\alpha^2)} (1 - \alpha\sqrt{t}) dt + \int_{1/(4\alpha^2)}^M \frac{1}{2} dt = \frac{1}{24\alpha^2} + \frac{M}{2},$$

which implies

$$\int_0^\infty \max\left(\mathbb{P}\{\xi^2 \geq t\} - \frac{1}{2}, 0\right) dt = \int_0^M \mathbb{P}\{\xi^2 \geq t\} dt - \frac{M}{2} \geq \frac{1}{24\alpha^2} \geq \frac{\beta^2}{48\alpha^2} \mathbb{E}\xi^2.$$

PROOF OF PROPOSITION 7.1. Let \mathcal{I} be a random subset of $\{1, 2, \dots, n\}$ such that $|\mathcal{I}| = m$ and

$$\sum_{j=1}^{n-m} j \cdot \min_{i \leq n} X_i^2 = \sum_{i \in \mathcal{I}^c} X_i^2$$

everywhere on the probability space. Now, let us distinguish two types of components of X : we set

$$I := \{i \leq n : \mathbb{P}\{i \in \mathcal{I}\} \geq 1/2\}, \quad \text{so that} \quad I^c = \{i \leq n : \mathbb{P}\{i \in \mathcal{I}\} < 1/2\}.$$

Obviously, we have

$$\mathcal{E}(X, m) = \mathbb{E} \sum_{i \in \mathcal{I}^c} X_i^2 = \sum_{i=1}^n \mathbb{E}(X_i^2 \chi_{\{i \in \mathcal{I}^c\}}) \geq \sum_{i \in I^c} \mathbb{E}(X_i^2 \chi_{\{i \in \mathcal{I}^c\}}).$$

Next, observe that for every $\tau > 0$ and every $i \in I^c$,

$$\mathbb{P}\{X_i^2 \chi_{\{i \in \mathcal{I}^c\}} \geq \tau\} = \mathbb{P}\{X_i^2 \geq \tau\} - \mathbb{P}\{X_i^2 \geq \tau \text{ and } i \in \mathcal{I}\} \geq \mathbb{P}\{X_i^2 \geq \tau\} - \frac{1}{2}.$$

Hence, in view of condition (11),

$$\mathcal{E}(X, m) \geq \sum_{i \in I^c} \int_0^\infty \max\left(\mathbb{P}\{X_i^2 \geq \tau\} - \frac{1}{2}, 0\right) d\tau \geq u \sum_{i \in I^c} \mathbb{E}X_i^2 \geq u\mathcal{E}_0(X, |I|).$$

On the other hand,

$$2m = 2\mathbb{E}|I| = 2\mathbb{E} \sum_{i=1}^n \chi_{\{i \in I\}} = 2 \sum_{i=1}^n \mathbb{P}\{i \in I\} \geq |I|,$$

and the proof is complete. \square

REMARK 7.2. In Proposition 7.1, we assumed that m is small compared to n , which is a natural condition in context of signal approximation. For theoretical reasons, it may be interesting to consider the range $m > n/2$. One could ask the following question: Let $m > n/2$ and $k := n - m$. Does there exist an absolute constant $C > 0$ (not depending on k, n) such that $\mathcal{E}_0(X, n - k/2) \leq C\mathcal{E}(X, n - k)$? It turns out that this is not true even in the case of the standard Gaussian random vector. Indeed, a direct computation shows that $\mathcal{E}_0(X, n - k/2) = k/2$ (for even k), while $\mathcal{E}(X, n - k) \approx k^3/n^2$. Thus, the above inequality cannot be true with an absolute constant for $1 \leq k \ll n$.

REMARK 7.3. Note that we were able to obtain a reverse-type inequality for (10) when we agreed to increase the number of one-dimensional projections for the linear approximation, which could be viewed as increasing of the rank (the dimension) of the corresponding projection. The idea to slightly lose on the optimality of dimension in order to gain on other parameters was effectively employed in the study of geometry of high-dimensional convex bodies (see, e.g., [12, 13] and references therein).

8. Proofs of auxiliary results. In this section, we provide the proofs of results from Sections 2 and 3 for the sake of completeness.

PROOF OF LEMMA 2.3. *Case 1:* $m = 1$, so that $a_1 \leq b_1/k$. Let $b := b_1, n_0 := 0$ and, given any $1 \leq \ell \leq k$, let n_ℓ be the largest integer not greater than n such that

$$\sum_{i=1}^{n_\ell} a_i \leq \frac{\ell b}{k}.$$

Since $b/k \geq a_1 \geq a_2 \geq \dots \geq a_n$, we have $0 = n_0 < 1 \leq n_1 < n_2 < \dots < n_k = n$. Define a partition $(A_\ell)_{\ell \leq k}$ of $\{1, 2, \dots, n\}$ as $A_\ell := \{i : n_{\ell-1} < i \leq n_\ell\}$. If $a_i \leq \frac{b}{2k}$ for all i , then we set $t = 0$. Otherwise, let t be the largest number in $\{1, 2, \dots, n\}$

such that $a_t > \frac{b}{2k}$. Then:

- (i) for every $1 \leq \ell \leq k$ such that $n_\ell \leq t$, we have $\sum_{i \in A_\ell} a_i \geq a_{n_\ell} > \frac{b}{2k}$;
- (ii) for every $\ell < k$ such that $n_\ell > t$, we have $\sum_{i \in A_\ell} a_i \geq \frac{b}{2k}$ (otherwise, since $a_{n_\ell+1} \leq \frac{b}{2k}$, we would have

$$\sum_{i=1}^{n_\ell+1} a_i = \sum_{i=1}^{n_\ell-1} a_i + \sum_{i \in A_\ell} a_i + a_{n_\ell+1} < \frac{(\ell-1)b}{k} + \frac{b}{2k} + \frac{b}{2k} = \frac{\ell b}{k},$$

which contradicts the choice of n_ℓ ;

- (iii) for $\ell = k$ we have $\sum_{i \in A_k} a_i = \sum_{i=1}^n a_i - \sum_{i=1}^{n_{k-1}} a_i \geq \frac{b}{k}$.

This completes the proof of the case $m = 1$.

Case 2: $m > 1$. For $1 \leq \ell < m$, choose $A_\ell = \{\ell\}$, and let $(A_\ell)_{\ell=m}^k$ be the partition of $\{m, m+1, \dots, n\}$ into $k+1-m$ sets constructed in the same way as in Case 1. Then, by the above argument, for every $\ell \geq m$ we have

$$\sum_{i \in A_\ell} a_i \geq \frac{b_m}{2(k+1-m)},$$

and the proof is complete. \square

PROOF OF LEMMA 3.1. Denote by $A_k(t)$ the event $\{|x_k \xi_k| > t\} = \{|\xi_k| > t/x_k\}$ and let

$$A(t) := \left\{ \min_{k \leq n} |x_k \xi_k| > t \right\} = \bigcap_{k \leq n} A_k(t), \quad t > 0.$$

By (6), we have $\mathbb{P}(A_k(t)^c) \leq \alpha t/x_k$. Hence,

$$\mathbb{P}(A(t)) \geq 1 - \sum_{k=1}^n \mathbb{P}(A_k(t)^c) \geq 1 - \alpha t \sum_{k=1}^n 1/x_k = 1 - \alpha b t,$$

which proves the first estimate and implies the estimate for the median. The estimate for the expectation follows by the distribution formula

$$\begin{aligned} \mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p &= \int_0^\infty \mathbb{P} \left\{ \min_{1 \leq i \leq n} |x_i \xi_i| > t^{1/p} \right\} dt \geq \int_0^{(\alpha b)^{-p}} (1 - \alpha b t^{1/p}) dt \\ &= \frac{(\alpha b)^{-p}}{1+p}. \end{aligned} \quad \square$$

PROOF OF LEMMA 3.2. As in the last proof, denote $A_k(t) := \{|x_k \xi_k| > t\} = \{|\xi_k| > t/x_k\}$ and let $A(t)$ be the intersection of the events. By (7), we have

$\mathbb{P}(A_k(t)) \leq \exp(-\beta t/x_k)$. Therefore,

$$\mathbb{P}(A(t)) = \prod_{k=1}^n \mathbb{P}(A_k(t)) \leq \exp\left(-\beta \sum_{k=1}^n t/x_k\right) = \exp(-\beta bt),$$

which proves the first estimate and implies the estimate for the median. Again, the bound for the expectation follows by the distribution formula:

$$\begin{aligned} \mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p &= \int_0^\infty \mathbb{P}\left\{\min_{1 \leq i \leq n} |x_i \xi_i| > t^{1/p}\right\} dt \geq \int_0^\infty \exp(-\beta bt^{1/p}) dt \\ &= (\beta b)^{-p} p\Gamma(p). \end{aligned} \quad \square$$

PROOF OF LEMMA 3.4. Denote $B(t) := \mathbb{P}\{k - \min_{1 \leq i \leq n} |x_i \xi_i| \leq t\}$. Clearly, we have

$$\begin{aligned} B(t) &= \mathbb{P}\left\{\exists i_1, i_2, \dots, i_k \leq n : |\xi_{i_j}| \leq \frac{t}{x_{i_j}} \text{ for all } j \leq k\right\} \\ &= \mathbb{P}\left(\bigcup_{\ell=k}^n \bigcup_{\substack{A \subset \{1, \dots, n\} \\ |A|=\ell}} \left\{|\xi_i| \leq \frac{t}{x_i} \text{ for all } i \in A \text{ and } |\xi_i| > \frac{t}{x_i} \text{ for all } i \notin A\right\}\right) \\ &= \sum_{\ell=k}^n \sum_{\substack{A \subset \{1, \dots, n\} \\ |A|=\ell}} \prod_{i \in A} \mathbb{P}\left\{|\xi_i| \leq \frac{t}{x_i}\right\} \prod_{i \notin A} \mathbb{P}\left\{|\xi_i| > \frac{t}{x_i}\right\}. \end{aligned}$$

Hence,

$$B(t) \leq \sum_{\ell=k}^n \sum_{\substack{A \subset \{1, \dots, n\} \\ |A|=\ell}} \prod_{i \in A} \mathbb{P}\left\{|\xi_i| \leq \frac{t}{x_i}\right\} \leq \sum_{\ell=k}^n \sum_{\substack{A \subset \{1, \dots, n\} \\ |A|=\ell}} \prod_{i \in A} \frac{\alpha t}{x_i}.$$

Corollary 2.2 implies the first part of the lemma.

Next, we verify the bound for the expectation. The case $k = 1$ follows by Lemma 3.1, so we assume that $k \geq 2$. Let us start with establishing the bound

$$(12) \quad \frac{k}{2^{1/p} 4\alpha} \left(\sum_{i=1}^n 1/x_i\right)^{-1} \leq \left(\mathbb{E} k - \min_{1 \leq i \leq n} |x_i \xi_i|^p\right)^{1/p}, \quad k \geq 2.$$

Set $\gamma := e/4$. Then

$$\begin{aligned} \mathbb{E} k - \min_{1 \leq i \leq n} |x_i \xi_i|^p &= \int_0^\infty \mathbb{P}\left\{k - \min_{1 \leq i \leq n} |x_i \xi_i| > t^{1/p}\right\} dt \\ &\geq \int_0^{(\gamma/a)^p} \left(1 - \frac{1}{\sqrt{2\pi k}} \frac{a^k t^{k/p}}{1 - at^{1/p}}\right) dt \\ &\geq \left(\frac{\gamma}{a}\right)^p - \frac{1}{\sqrt{2\pi k}} \frac{a^k}{1 - \gamma} \left(\frac{\gamma}{a}\right)^{k+p} \end{aligned}$$

$$\begin{aligned} &\geq \left(\frac{\gamma}{a}\right)^p \left(1 - \frac{1}{2\sqrt{\pi}} \frac{\gamma^2}{1-\gamma}\right) \\ &\geq \frac{1}{2} \left(\frac{\gamma}{a}\right)^p, \end{aligned}$$

which proves (12). Finally, observe that for every sequence $(a_i)_{i=1}^n$ and every $r < k$ one has

$$k\text{-min}(a_i)_{i=1}^n \geq (k-r)\text{-min}(a_i)_{i=r+1}^n;$$

in particular,

$$\begin{aligned} \left(\mathbb{E}k\text{-min}_{1 \leq i \leq n} |x_i \xi_i|^p\right)^{1/p} &\geq \left(\mathbb{E}(k-r)\text{-min}_{r+1 \leq i \leq n} |x_i \xi_i|^p\right)^{1/p} \\ &\geq \frac{k-r}{2^{1/p} 4\alpha} \left(\sum_{i=r+1}^n 1/x_i\right)^{-1}, \end{aligned}$$

where the last inequality is (12) applied to the appropriate “truncated” sequence. The result follows. \square

Acknowledgments. The authors would like to thank Nicole Tomczak-Jaegermann for valuable comments on Section 7. They are also grateful to anonymous referees for helpful remarks and suggestions.

REFERENCES

- [1] ADAMCZAK, R., LATAŁA, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2014). Tail estimates for norms of sums of log-concave random vectors. *Proc. Lond. Math. Soc.* (3) **108** 600–637. [MR3180591](#)
- [2] DAVID, H. A. and NAGARAJA, H. N. (2003). *Order Statistics*, 3rd ed. Wiley-Interscience, Hoboken, NJ. [MR1994955](#)
- [3] GLUSKIN, E. D. (1989). Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces. *Mat. Sb.* **64** 85–96.
- [4] GORDON, Y., LITVAK, A., SCHÜTT, C. and WERNER, E. (2002). Orlicz norms of sequences of random variables. *Ann. Probab.* **30** 1833–1853. [MR1944007](#)
- [5] GORDON, Y., LITVAK, A., SCHÜTT, C. and WERNER, E. (2002). Geometry of spaces between polytopes and related zonotopes. *Bull. Sci. Math.* **126** 733–762. [MR1941083](#)
- [6] GORDON, Y., LITVAK, A., SCHÜTT, C. and WERNER, E. (2005). Minima of sequences of Gaussian random variables. *C. R. Math. Acad. Sci. Paris* **340** 445–448. [MR2135327](#)
- [7] GORDON, Y., LITVAK, A. E., SCHÜTT, C. and WERNER, E. (2006). On the minimum of several random variables. *Proc. Amer. Math. Soc.* **134** 3665–3675. [MR2240681](#)
- [8] GORDON, Y., LITVAK, A. E., SCHÜTT, C. and WERNER, E. (2012). Uniform estimates for order statistics and Orlicz functions. *Positivity* **16** 1–28. [MR2892571](#)
- [9] HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*, 2nd ed. The Univ. Press, Cambridge. [MR0046395](#)
- [10] LATAŁA, R. (2011). Order statistics and concentration of l_r norms for log-concave vectors. *J. Funct. Anal.* **261** 681–696. [MR2799576](#)

- [11] LECHNER, R., PASSENBRUNNER, M. and PROCHNO, J. (2015). Uniform estimates for averages of order statistics of matrices. *Electron. Commun. Probab.* **20** no. 27. [MR3327866](#)
- [12] LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2006). Diameters of sections and coverings of convex bodies. *J. Funct. Anal.* **231** 438–457. [MR2195339](#)
- [13] LITVAK, A. E. and TOMCZAK-JAEGERMANN, N. (2000). Random aspects of high-dimensional convex bodies. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **1745** 169–190. Springer, Berlin. [MR1796719](#)
- [14] MALLAT, S. (2009). *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Elsevier/Academic Press, Amsterdam. [MR2479996](#)
- [15] MALLAT, S. and ZEITOUNI, O. A conjecture concerning optimality of the Karhunen–Loeve basis in nonlinear reconstruction. Available at [arXiv:1109.0489](#).
- [16] MARSHALL, A. W. and PROSCHAN, F. (1965). An inequality for convex functions involving majorization. *J. Math. Anal. Appl.* **12** 87–90. [MR0185062](#)
- [17] MONTGOMERY-SMITH, S. (2002). Rearrangement invariant norms of symmetric sequence norms of independent sequences of random variables. *Israel J. Math.* **131** 51–60. [MR1942301](#)
- [18] ŠIDÁK, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.* **62** 626–633. [MR0216666](#)
- [19] ZEITOUNI, O. (2006). A correlation inequality for nonlinear reconstruction. In *Workshop on the Mathematical Foundations of Learning Theory*, Paris.

DEPARTMENT OF MATHEMATICAL
AND STATISTICAL SCIENCES
UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA T6G 2G1
CANADA
E-MAIL: aelitvak@gmail.com

DEPARTMENT OF MATHEMATICS
PRINCETON UNIVERSITY
FINE HALL
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: ktikhomi@ualberta.ca