Concentration of quadratic forms under a Bernstein moment assumption

Pierre C. Bellec
Rutgers University, ENSAE and UMR CNRS 9194

Abstract: A concentration result for quadratic form of independent subgaussian random variables is derived. If the moments of the random variables satisfy a “Bernstein condition”, then the variance term of the Hanson-Wright inequality can be improved.

1. Concentration of a quadratic form of subgaussian random variables

Throughout this note, \( A \in \mathbb{R}^{n \times n} \) is a real matrix, and \( \xi = (\xi_1, \ldots, \xi_n)^T \) is a centered random vector with independent components. We are interested in the concentration behavior of the random variable

\[
\xi^T A \xi - \mathbb{E}[\xi^T A \xi],
\]

Let \( \sigma_i^2 = \mathbb{E}[\xi_i^2] \) for all \( i = 1, \ldots, n \) and define \( D_\sigma = \text{diag}(\sigma_1, \ldots, \sigma_n) \). If the random variables \( \xi_1, \ldots, \xi_n \) are Gaussian, we have the following concentration inequality.

**Proposition 1** (Gaussian chaos of order 2). Let \( \xi_1, \ldots, \xi_n \) be independent zero-mean normal random variables with for all \( i = 1, \ldots, n \), \( \mathbb{E}[\xi_i^2] = \sigma_i^2 \). Let \( A \) be any \( n \times n \) real matrix. Then for any \( x > 0 \),

\[
P \left( \xi^T A \xi - \mathbb{E}[\xi^T A \xi] > t \right) \leq \exp \left( -c \min \left( \frac{t^2}{K^4 \|A\|_F^2}, \frac{t}{K^2 \|A\|_2^2} \right) \right)
\]

A proof of this concentration result can be found in [3, Example 2.12]. We will refer to the term \( 2 \|D_\sigma A D_\sigma\|_F \sqrt{x} + 2 \|D_\sigma A D_\sigma\|_2 x \) as the variance term, since if \( A \) is diagonal-free, the random variable \( \xi^T A \xi \) is centered with variance \( \|D_\sigma A D_\sigma\|_F^2 \).

A similar concentration result is available for subgaussian random variables. It is known as the Hanson-Wright inequality and is given in Proposition 2 below. First versions of this inequality can be found in Hanson and Wright [5] and Wright [9], although with a weaker statement than Proposition 2 below since these results involve \( \|(|a_{ij}|)\|_2 \) instead of \( \|A\|_2 \). Recent proofs of this concentration inequality with \( \|A\|_2 \) instead of \( \|(|a_{ij}|)\|_2 \) can be found in Rudelson and Vershynin [6] or Barthe and Milman [2, Theorem A.5].

**Proposition 2** (Hanson-Wright inequality [6]). There exist an absolute constant \( c > 0 \) such that the following holds. Let \( n \geq 1 \) and \( \xi_1, \ldots, \xi_n \) be independent zero-mean subgaussian random variables with \( \max_{i=1,\ldots,n} \|\xi_i\|_{\psi_2} \leq K \) for some real number \( K > 0 \). Let \( A \) be any \( n \times n \) real matrix. Then for all \( t > 0 \),

\[
P \left( \xi^T A \xi - \mathbb{E}[\xi^T A \xi] > t \right) \leq \exp \left( -c \min \left( \frac{t^2}{K^4 \|A\|_F^2}, \frac{t}{K^2 \|A\|_2^2} \right) \right)
\]
where $\xi = (\xi_1, \ldots, \xi_n)^T$. Furthermore, for any $x > 0$, with probability greater than $1 - \exp(-x)$,

$$\xi^T A \xi - \mathbb{E}[\xi^T A \xi] \leq cK^2 \|A\|_2^2 x + cK^2 \|A\|_F \sqrt{x}.$$ 

For some random variables $\xi_1, \ldots, \xi_n$, the “variance term” $K^2 \|A\|_F \sqrt{x}$ is far from the variance of the random variable $\xi^T A \xi$. The goal of the present paper is to show that under a mild assumption on the moments of $\xi_1, \ldots, \xi_n$, it is possible to substantially reduce the variance term. This assumption is the following.

**Assumption 1** (Bernstein condition on $\xi_1^2, \ldots, \xi_n^2$). Let $K > 0$ and assume that $\xi_1, \ldots, \xi_n$ are independent and satisfy

$$\forall p \geq 1, \quad \mathbb{E}[|\xi_i|^p] \leq \frac{1}{2} p! \sigma_i^2 K^{2(p-1)}.$$  (2)

**Example 1.** Centered variables almost surely bounded by $K$ and zero-mean Gaussian random variables with variance smaller than $K^2$ satisfy (2).

**Example 2** (Log-concave random variables). In [7], the authors consider a slightly stronger condition [7, Definition 1.1]. They consider random variables $Z$ satisfying for any integer $p \geq 1$ and some constant $K$:

$$\mathbb{E}[|Z|^p] \leq pK \mathbb{E}[|Z|^{p-1}],$$  (3)

and they showed in [7, Section 7] that any distribution that is log-concave satisfies (3). Thus, if $X$ is log-concave then our assumption (2) holds. See [1, Section 6] for a comprehensive list of the common log-concave distributions.

The next theorem provides a concentration inequality for quadratic forms of independent random variables satisfying the moment assumption (2). It is sharper than the Hanson-Wright inequality given in Proposition 2.

**Theorem 3.** Assume that the random variable $\xi = (\xi_1, \ldots, \xi_n)^T$ satisfies Assumption 1 for some $K > 0$. Let $A$ be any $n \times n$ real matrix. Then for all $t > 0$,

$$\mathbb{P}\left(\xi^T A \xi - \mathbb{E}[\xi^T A \xi] > t\right) \leq \exp\left(-\min\left(\frac{t^2}{192K^2 \|A\|_F^2}, \frac{t^2}{256K^2 \|A\|_2^2}\right)\right),$$  (4)

where $D_\sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$. Furthermore, for any $x > 0$, with probability greater than $1 - \exp(-x)$,

$$\xi^T A \xi - \mathbb{E}[\xi^T A \xi] \leq 256K^2 \|A\|_2^2 x + 8\sqrt{3}K \|AD_\sigma\|_F \sqrt{x}.$$  (5)

The proof of this result relies on the decoupling inequality for quadratic forms [8][4, Theorem 8.11].

If $t$ is small, the right hand side of (4) becomes

$$\exp\left(-\frac{t^2}{192K^2 \|A D_\sigma\|_F^2}\right),$$

whereas the right hand side of the Hanson-Wright inequality (1) becomes

$$\exp\left(-\frac{c-1}{K^4 \|A\|_F^2}\right),$$

for some absolute constant $c > 0$. The element of the diagonal matrix $D_\sigma$ are bounded from above by $K$, so Theorem 3 gives a sharper bound than the Hanson-Wright inequality in this regime.
2. Proof of Theorem 3

The goal of this section is to prove Theorem 3. We start with preliminary calculations that will be useful in the proof. Let $A$ be any $n \times n$ real matrix. Let $\lambda > 0$ satisfy
\begin{equation}
128\|A\|_2 K^2 \lambda \leq 1, \tag{6}
\end{equation}
and define
\begin{equation}
\eta = 32K^2 \lambda^2. \tag{7}
\end{equation}
The inequality (6) can be rewritten in terms of $\eta$:
\begin{equation}
512K^2 \|A\|_2^2 \eta \leq 1. \tag{8}
\end{equation}
Let $A_0$ be the matrix $A$ with the diagonal entries set to 0. Then, using the triangle inequality with $A_0 = A - \text{diag}(a_{11},...,a_{nn})$ and $|a_{ii}| \leq \|A\|_2$ for all $i = 1,...,n$, we obtain
\begin{equation}
\|A_0\|_2 \leq 2\|A\|_2. \tag{9}
\end{equation}
Let $B = A_0^2 A_0 = (b_{ij})_{i,j=1,...,n}$ and let $B_0$ be the matrix $B$ with the diagonal entries set to 0. Then
\begin{equation}
\forall i = 1,...,n, \quad 0 \leq b_{ii} = \sum_{j \neq i} a_{ji}^2 \leq \|A\|^2. \tag{10}
\end{equation}
By using the decomposition $B_0 = B - \text{diag}(b_{11},...,b_{nn})$ and the inequality $\|v + v'\|_2^2 \leq 2\|v\|_2^2 + 2\|v'\|_2^2$, (10) and (9), we have:
\begin{equation}
\|B_0 \xi\|_2^2 \leq 2\|B \xi\|_2^2 + 2 \sum_{i=1}^n b_{ii}^2 \xi_i^2,
\end{equation}
\begin{equation}
\leq 2\|A_0\|_2^2\|A_0 \xi\|_2^2 + 2\|A\|_2^2 \sum_{i=1}^n b_{ii} \xi_i^2,
\end{equation}
\begin{equation}
\leq 8\|A\|_2^2\|A_0 \xi\|_2^2 + 2\|A\|_2^2 \sum_{i=1}^n b_{ii} \xi_i^2.
\end{equation}
Combining the previous display with (8), we obtain for any $K > 0$:
\begin{equation}
16K^2 \eta^2 \|B_0 \xi\|_2^2 \leq (512K^2 \|A\|_2^2 \eta) \left(\frac{\eta}{4} \|A_0 \xi\|_2^2 + \frac{\eta}{16} \sum_{i=1}^n b_{ii} \xi_i^2\right),
\end{equation}
\begin{equation}
\leq \frac{\eta}{4} \|A_0 \xi\|_2^2 + \frac{\eta}{16} \sum_{i=1}^n b_{ii} \xi_i^2. \tag{11}
\end{equation}

Proof of Theorem 3. Throughout the proof, let $\lambda > 0$ satisfy (6). The value of $\lambda$ will be specified later.

First we treat the diagonal terms by bounding the moment generating function of
\begin{equation}
S_{\text{diag}} := \sum_{i=1}^n a_{ii} \xi_i^2 - \sum_{i=1}^n a_{ii} \sigma_i^2.
\end{equation}
Using the independence of $\xi_1, \ldots, \xi_n$ and (17) with $s = a_{ii}\lambda$ with each $i = 1, \ldots, n$:

$$\mathbb{E}\exp(\lambda S_{\text{diag}}) \leq \exp \left( \lambda^2 \sum_{i=1}^{n} a_{ii}^2 \sigma_i^2 K^2 \right), \quad (12)$$

provided that for all $i = 1, \ldots, n$, $2|a_{ii}| \lambda K^2 \leq 1$ which is satisfied as (6) holds and $|a_{ii}| \leq \|A\|_2$.

Now we bound the moment generating function of the off-diagonal terms. Let $S_{\text{off-diag}} := \sum_{i,j=1,\ldots,n: i \neq j} a_{ij}\xi_i\xi_j$.

Let the random vector $\xi' = (\xi'_1, \ldots, \xi'_n)^T$ be independent of $\xi$ with the same distribution as $\xi$. We apply the decoupling inequality [8] (see also [4, Theorem 8.11]) to the convex function $s \to \exp(\lambda s)$:

$$\mathbb{E}\exp(\lambda S_{\text{off-diag}}) \leq \mathbb{E}\exp \left( 4\lambda \sum_{i,j=1,\ldots,n: i \neq j} a_{ij}\xi'_i\xi'_j \right).$$

Conditionally on $\xi_1, \ldots, \xi_n$, for each $i = 1, \ldots, n$, we use the independence of $\xi'_1, \ldots, \xi'_n$ and (16) applied to $\xi'_i$ with $s = 4\sum_{j=1,\ldots,n: j \neq i} a_{ij}\xi'_j$:

$$\mathbb{E}\exp \left( 4\lambda \sum_{i \neq j} a_{ij}\xi'_i\xi'_j \right) \leq \mathbb{E}\exp \left( 16K^2\lambda^2 \sum_{i=1,\ldots,n} \left( \sum_{j=1,\ldots,n: j \neq i} a_{ij}\xi'_j \right)^2 \right) = \mathbb{E}\exp \left( 16K^2\lambda^2 \|A_0\xi\|_2^2 \right) = \mathbb{E}\exp \left( \frac{\eta}{2}\|A_0\xi\|_2^2 \right),$$

where $\eta$ is defined in (7) and $A_0$ is the matrix $A$ with the diagonal entries set to 0. Let $B = A_0^T A_0 = (b_{ij})_{i,j=1,\ldots,n}$. Then $\|A_0\xi\|_2^2 = \sum_{i=1}^{n} b_{ii}\xi_i^2 + \sum_{i\neq j} b_{ij}\xi_i\xi_j$.

We use the Cauchy-Schwarz inequality to separate the diagonal terms from the off-diagonal ones:

$$\left( \mathbb{E}\exp \left( \frac{\eta}{2}\|A_0\xi\|_2^2 \right) \right)^2 \leq \mathbb{E}\exp \left( \eta \sum_{i=1}^{n} b_{ii}\xi_i^2 \right) \mathbb{E}\exp \left( \eta \sum_{i \neq j} b_{ij}\xi_i\xi_j \right). \quad (13)$$

For the off-diagonal terms of (13), using the decoupling inequality [8] (see also [4, Theorem 8.11]) we have:

$$\mathbb{E}\exp \left( \eta \sum_{i \neq j} b_{ij}\xi_i\xi_j \right) \leq \mathbb{E}\exp \left( 4\eta \sum_{i \neq j} b_{ij}\xi'_i\xi'_j \right).$$

Again, conditionally on $\xi_1, \ldots, \xi_n$, for each $j = 1, \ldots, n$, we use (16) applied to $\xi'_j$ and the
independence of $\xi_1', \ldots, \xi_n$:

$$E \exp \left( 4\eta \sum_{i \neq j} b_{ij} \xi_i \xi_j \right) \leq E \exp \left( 16K^2 \eta^2 \left( \sum_{i=1}^{n} \left( \sum_{j=1, \ldots, n: j \neq i} b_{ij} \xi_j \right)^2 \right) \right),$$

$$= E \exp \left( 16K^2 \eta^2 \| B_0 \xi \|_2^2 \right),$$

$$\leq E \exp \left( \frac{\eta}{4} \| A_0 \xi \|_2^2 + \frac{\eta}{16} \sum_{i=1}^{n} b_{ii} \xi_i^2 \right),$$

where we used the preliminary calculation (11) for the last display. Finally, the Cauchy-Schwarz inequality yields

$$E \exp \left( 4\eta \sum_{i \neq j} b_{ij} \xi_i \xi_j \right) \leq \sqrt{E \exp \left( \frac{\eta}{2} \| A_0 \xi \|_2^2 \right)} \sqrt{E \exp \left( \frac{\eta}{8} \sum_{i=1}^{n} b_{ii} \xi_i^2 \right)}.$$  

We plug this upper bound back into (13). After rearranging, we find

$$\left( E \exp \left( \frac{\eta}{2} \| A_0 \xi \|_2^2 \right) \right)^{3/2} \leq E \exp \left( \eta \sum_{i=1}^{n} b_{ii} \xi_i^2 \right) \sqrt{E \exp \left( \frac{\eta}{8} \sum_{i=1}^{n} b_{ii} \xi_i^2 \right)}.$$  

As $b_{ii} \geq 0$, this implies:

$$E \exp(\frac{\eta}{2} \| A_0 \xi \|_2^2) \leq E \exp \left( \eta \sum_{i=1}^{n} b_{ii} \xi_i^2 \right).$$

For each $i = 1, \ldots, n$, we apply (18) to the variable $\xi_i$ with $s = b_{ii} \eta \geq 0$. Using the independence of $\xi_1', \ldots, \xi_n'$, we obtain:

$$E \exp \left( \eta \sum_{i=1}^{n} b_{ii} \xi_i^2 \right) = \prod_{i=1}^{n} E \exp(\eta b_{ii} \xi_i^2),$$

$$\leq \exp \left( \frac{3}{2} \eta \sum_{i=1}^{n} b_{ii} \sigma_i^2 \right) = \exp \left( \frac{3}{2} \eta \| A_0 \sigma \|_F^2 \right),$$

provided that for all $i = 1, \ldots, n$, $2K^2 b_{ii} \eta \leq 1$ which is satisfied thanks to (6) and (10).

We remove $\eta$ from the above displays using its definition (7):

$$E \exp(\lambda S_{\text{off-diag}}) \leq \exp \left( 48\lambda^2 K^2 \| A_0 \sigma \|_F^2 \right),$$

where $A_0$ is the matrix $A$ with the diagonal entries set to 0.

Now we combine the bound on the moment generating function of $S_{\text{diag}}$ and $S_{\text{off-diag}}$, given respectively in (12) and (14). Using the Chernoff bound and the Cauchy-Schwarz inequality: we have that for all $\lambda$ satisfying (6),

$$\mathbb{P} \left( S_{\text{diag}} + S_{\text{off-diag}} > t \right) \leq \exp(-\lambda t) E[\exp(\lambda S_{\text{diag}}) \exp(\lambda S_{\text{off-diag}})],$$

$$\leq \exp \left( -\lambda t \right) \sqrt{E[\exp(2\lambda S_{\text{diag}})]} \sqrt{E[\exp(2\lambda S_{\text{off-diag}})]},$$

$$\leq \exp \left( -\lambda t + \lambda^2 K^2 \left( \sum_{i=1}^{n} \sigma_i^2 a_{ii}^2 + 48 \| A_0 \sigma \|_F^2 \right) \right),$$

$$\leq \exp \left( -\lambda t + 48\lambda^2 K^2 \| AD \sigma \|_F^2 \right),$$

(15)
where for the last display we used the equality
\[ \|AD\sigma\|_F^2 = \sum_{i,j=1,\ldots,n} a_i^2 \sigma_i^2 = \|A_0D\sigma\|_F^2 + \sum_{i=1}^n a_i^2 \sigma_i^2. \]

It now remains to choose the parameter \( \lambda \). The unconstrained minimum of (15) is attained at \( \bar{\lambda} = t/(96K^2 \|AD\sigma\|_F^2) \). If \( \bar{\lambda} \) satisfies the constraint (6), then
\[ \mathbb{P}(S_{\text{diag}} + S_{\text{off-diag}} > t) \leq \exp\left(-\frac{t^2}{192K^2 \|AD\sigma\|_F^2}\right). \]

On the other hand, if \( \bar{\lambda} \) does not satisfy (6), then the constraint (6) is binding and the minimum of (15) is attained at \( \lambda_b = \frac{1}{128 \|A\|_2^2} \). In this case,
\[ -t\lambda_b + \lambda_b^2 48K^2 \|AD\sigma\|_F^2 = -t\lambda_b + \frac{t}{2} \lambda_b = -\frac{t}{256K^2 \|A\|_2^2}. \]

Combining the two regimes, we obtain
\[ \mathbb{P}(S_{\text{diag}} + S_{\text{off-diag}} > t) \leq \exp\left(-\min\left(\frac{t^2}{192K^2 \|AD\sigma\|_F^2}, \frac{t}{256K^2 \|A\|_2^2}\right)\right). \]

The proof of (4) is complete.

Now we prove (5). The function
\[ t \rightarrow x(t) = \min\left(\frac{t^2}{192K^2 \|AD\sigma\|_F^2}, \frac{t}{256K^2 \|A\|_2^2}\right) \]

is increasing and bijective from the set of positive real numbers to itself. Furthermore, for all \( t > 0 \),
\[ t \leq 8\sqrt{3}K \|AD\sigma\|_F \sqrt{x(t)} + 256K^2 \|A\|_2 x(t), \]

so the variable change \( x = x(t) \) completes the proof of (5).

## 3. Technical lemmas: bounds on moment generating functions

The condition (2) leads to the following bounds on the moment generating functions of \( X \) and \( X^2 \), which are crucial to prove Theorem 3.

**Proposition 4.** Let \( K > 0 \) and let \( \xi_i \) be a random variable satisfying (2) with \( \sigma_i^2 = \mathbb{E}[\xi_i^2] \). Then for all \( s \in \mathbb{R} \):
\[ \mathbb{E}\exp(s\xi_i) \leq \exp(s^2 K^2), \quad (16) \]
Furthermore, if \( 0 \leq 2sK^2 \leq 1 \), then
\[ \mathbb{E}\exp(s\xi_i^2 - s\sigma_i^2) \leq \exp(s^2 \sigma_i^2 K^2), \quad (17) \]
\[ \mathbb{E}\exp(s\xi_i^2) \leq \exp\left(\frac{3}{2} s\sigma_i^2\right). \quad (18) \]

Inequality (16) shows that a random variable \( X \) satisfying the moment assumption (2) is subgaussian and its \( \psi_2 \) norm is bounded by \( K \) up to a multiplicative absolute constant. The proof of Proposition 4 is based on Taylor expansions and some algebra.
Proof of Proposition 4. To simplify the notation, let $X = \xi$ and $\sigma = \sigma_i$. We first prove (17). We apply the assumption on the even moments of $X$:

$$
E \exp(sX^2) = 1 + s\sigma^2 + \sum_{p \geq 2} \frac{s^p E X^{2p}}{p!},
$$

$$
\leq 1 + s\sigma^2 + \frac{\sigma^2}{2} \sum_{k=1}^{\infty} (sK^2)^k = 1 + s\sigma^2 + \frac{\sigma^2 K^2 s^2}{2(1 - sK^2)},
$$

and using the inequality $0 < 2sK^2 \leq 1$, we obtain:

$$
E \exp(sX^2) \leq 1 + s\sigma^2 + \sigma^2 s^2 K^2 \leq \exp(s\sigma^2 + s^2 \sigma^2 K^2),
$$

which completes the proof of (17). Inequality (18) is a direct consequence of (17) after applying again the inequality $2sK^2 \leq 1$.

We now prove (16). Using the Cauchy-Schwarz inequality and the assumption on the moments for $p = 2$, we get $\sigma^4 \leq E[\xi^4] \leq \sigma^2 K^2$, so $\sigma \leq K$. Let $p \geq 1$. For the even terms of the expansion of $E \exp(sX)$, we get:

$$
\frac{s^{2p} E X^{2p}}{(2p)!} \leq \frac{1}{2} (sK)^{2p} \frac{p!}{(2p)!} \leq \frac{1}{2} \frac{(sK)^{2p}}{p!},
$$

where for the last inequality we used $(p!)^2 \leq (2p)!$. For the odd terms, by using the Jensen inequality for $p \geq 1$:

$$
\frac{s^{2p+1} E X^{2p+1}}{(2p+1)!} \leq \frac{1}{2} |sK|^{2p+1} \frac{(p+1)!}{(2p+1)!},
$$

If $|sK| > 1$, we use the inequality $(p+1)!^2 \leq (2p+1)!$ to obtain

$$
\frac{s^{2p+1} E X^{2p+1}}{(2p+1)!} \leq \frac{|sK|^{2(p+1)}}{2((p+1)!)},
$$

and by combining the inequality for the even and the odd terms:

$$
E \exp(sX) = 1 + \sum_{p \geq 1} \frac{s^{2p} E X^{2p}}{(2p)!} + \frac{s^{2p+1} E X^{2p+1}}{(2p+1)!},
$$

$$
\leq 1 + \frac{1}{4} \sum_{p \geq 1} \frac{(sK)^{2p}}{p!} + \frac{|sK|^{2(p+1)}}{(2p+1)!},
$$

$$
\leq 1 + \sum_{p \geq 1} \frac{(sK)^{2p}}{p!} = \exp(s^2 K^2).
$$

If $|sK| \leq 1$, we use the inequality $(p+1)!p! \leq (2p+1)!$ to obtain

$$
\frac{s^{2p+1} E X^{2p+1}}{(2p+1)!} \leq \frac{(sK)^{2p}}{2(p!)},
$$

$$
E \exp(sX) = 1 + \sum_{p \geq 1} \frac{s^{2p} E X^{2p}}{(2p)!} + \frac{s^{2p+1} E X^{2p+1}}{(2p+1)!},
$$

$$
\leq 1 + \frac{1}{4} \sum_{p \geq 1} \frac{(sK)^{2p}}{p!} + \frac{(sK)^{2(p+1)}}{(2p+1)!},
$$

$$
\leq 1 + \sum_{p \geq 1} \frac{(sK)^{2p}}{p!} = \exp(s^2 K^2).
$$
and by combining the inequality for the even and the odd terms:

\[
\mathbb{E}\exp(sX) = 1 + \sum_{p \geq 1} \frac{s^{2p+1} \mathbb{E}X^{2p+1}}{(2p+1)!} + \frac{s^{2p} \mathbb{E}X^{2p}}{(2p)!} \\
\leq 1 + \frac{1}{2} \sum_{p \geq 1} \frac{(sK)^{2p}}{p!} + \frac{(sK)^{2p}}{p!} = 1 + \sum_{p \geq 1} \frac{(sK)^{2p}}{p!} = \exp(s^2K^2).
\]

\[\square\]

References

[1] Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic theory*, 26(2):445–469, 2005.
[2] Franck Barthe and Emanuel Milman. Transference principles for log-Sobolev and spectral-gap with applications to conservative spin systems. *Comm. Math. Phys.*, 323(2):575–625, 2013. URL http://dx.doi.org/10.1007/s00220-013-1782-2.
[3] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.
[4] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.
[5] David Lee Hanson and Farroll Tim Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.
[6] Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18:no. 82, 1–9, 2013. URL http://ecp.ejpecp.org/article/view/2865.
[7] Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 437–446. SIAM, 2012.
[8] Roman Vershynin. A simple decoupling inequality in probability theory. *preprint*, 2011. URL http://www-personal.umich.edu/~romanv/papers/decoupling-simple.pdf.
[9] Farroll Tim Wright. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *The Annals of Probability*, 1(6):1068–1070, 1973.