Commun. Math. Sci.
Vol. 12, No. 4, pp. 695–705

NEW BOUNDS FOR CIRCULANT JOHNSON-LINDENSTRAUSS EMBEDDINGS[*]

HUI ZHANG[†] AND LIZHI CHENG[‡]

Abstract. This paper analyzes circulant Johnson-Lindenstrauss (JL) embeddings which, as an important class of structured random JL embeddings, are formed by randomizing the column signs of a circulant matrix generated by a random vector. With the help of recent decoupling techniques and matrix-valued Bernstein inequalities, we obtain a new bound $k=O(\epsilon^{-2}\log^{(1+\delta)}(n))$ for Gaussian circulant JL embeddings. Moreover, by using the Laplace transform technique (also called Bernstein's trick), we extend the result to the subgaussian case. The bounds in this paper offer a small improvement over the current best bounds for Gaussian circulant JL embeddings for certain parameter regimes and are derived using more direct methods.

Key words. Johnson-Lindenstrauss embedding, circulant matrix, Laplace transform, decoupling technique, matrix-valued Bernstein inequality.

AMS subject classifications. 52C99, 68Q01.

1. Introduction

The Johnson-Lindenstrauss (JL) lemma [6] is by now a standard technique in high dimensional data processing. The lemma shows the existence, with high probability, of JL embeddings, or linear maps $A\in\mathbb{R}^d\to\mathbb{R}^k$ (with $k<d$) which embed a fixed set of n points $\{x_1\cdots x_n\}\subset\mathbb{R}^d$ into $\mathbb{R}^k$ with distortion at most ϵ. The best known embedding dimension k, as is achieved by, e.g., Gaussian random matrices, is $k=O(\epsilon^{-2}\log(n))$. Recently, there is growing interest in analyzing structured random JL embeddings which, unlike Gaussian random matrices, have fast matrix-vector multiplication routines. In this paper, we focus on circulant JL embeddings which, as an important class of such structured random JL embeddings, are formed by randomizing the column signs of a circulant matrix generated by a random vector. The first result for circulant JL embeddings might be formulated as follows.

Theorem 1.1. *([4]) Let $x^1,x^2,\cdots,x^n$ be n points in the d-dimensional Euclidean space $\mathbb{R}^d$. Let $\epsilon\in(0,\frac{1}{2})$ and let $k=O(\epsilon^{-2}\log^3(n))$ be a natural number. Assume that f is a composition of a $k\times d$ random circulant matrix $M_{a,k}$ with a $d\times d$ random diagonal matrix $D_\varkappa$, i.e., $f(x)=\frac{1}{\sqrt{k}}M_{a,k}D_\varkappa x$. Then with probability at least 2/3 the following holds:*

$$(1-\epsilon)\|x^i-x^j\|_2^2\le\|f(x^i)-f(x^j)\|_2^2\le(1+\epsilon)\|x^i-x^j\|_2^2,\quad i,j,=1,\cdots,n. \tag{1.1}$$

Here, the random circulant matrix $M_{a,k}$ is defined by a random vector $a=(a_0,\cdots,a_{d-1})$ whose entries are independent Bernoulli variables or independent nor-

[*] Received: December 6, 2012; accepted (in revised form): May 29, 2013. Communicated by Amit Singer.

[†] Department of Mathematics and Systems Science, College of Science, National University of Defense Technology, Changsha, Hunan, 410073, P.R. China (hhuuii.zhang@gmail.com).

[‡] Department of Mathematics and Systems Science, College of Science, National University of Defense Technology, Changsha, Hunan, 410073, P.R. China (clzcheng@nudt.edu.cn).

mally distributed variables. Concretely,

$$M_{a,k}=\begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{d-1} \\ a_{d-1} & a_0 & a_1 & \cdots & a_{d-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{d-k+1} & a_{d-k+2} & a_{d-k+3} & \cdots & a_{d-k} \end{pmatrix} \in \mathbb{R}^{k\times d}.$$

The random diagonal matrix $D_\varkappa$ is

$$D_\varkappa=\begin{pmatrix} \varkappa_0 & 0 & 0 & \cdots & 0 \\ 0 & \varkappa_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \varkappa_{d-1} \end{pmatrix} \in \mathbb{R}^{d\times d},$$

where $\varkappa=(\varkappa_0,\varkappa_1,\cdots,\varkappa_{d-1})$ is a Bernoulli sequence, i.e., each entry of $\varkappa$ takes the values $+1$ or -1 with probability $1/2$. Here and thereafter, we will call the mapping f a (sub)gaussian circulant JL embedding when the random vector a is set as a (sub)gaussian random vector.

Compared with the standard bound $k=O(\epsilon^{-2}\log(n))$, Theorem 1.1 only established a worse bound $k=O(\epsilon^{-2}\log^3(n))$. Later on, Vybíral [12] improved the bound to $k=O(\epsilon^{-2}\log^2(n))$ by employing the discrete Fourier transform and singular value decomposition to deal with the dependence caused by the circulant structure. Recently, by randomizing the column signs of matrices that have the Restricted Isometry Property (RIP) [2], Krahmer and Ward [3] further improved the bound to $k=O(\epsilon^{-2}\log(n)\log^4(d))$ which is better than another recent bound $k=O(\epsilon^{-4}\log(n)\log^4(d))$ by Ailon and Liberty [1]. Most recently, Krahmer, Mendelson, and Rauhut [7] derived new bounds for the RIP of partial circulant matrices. By combining these bounds with the connection between RIP and JL in [3], the current best bound for Gaussian circulant JL embedding reads

$$k=O(\epsilon^{-2}\log(n)(\log d)^2(\log\log d)^2). \tag{1.2}$$

We summarize these JL bounds in table 1.1.

work	JL bound
[4]	$k=O(\epsilon^{-2}\log^3(n))$
[12]	$k=O(\epsilon^{-2}\log^2(n))$
[1]	$k=O(\epsilon^{-4}\log(n)\log^4(d))$
[3]	$k=O(\epsilon^{-2}\log(n)\log^4(d))$
[3], [7]	$k=O(\epsilon^{-2}\log(n)(\log d)^2(\log\log d)^2)$

Table 1.1. *Bounds for Gaussian circulant JL embeddings*

1.1. Main results.

In this study, we combine the decoupling technique in [12] with the matrix-value Bernstein inequality in [11] to derive a new and improved bound $k=O(\epsilon^{-2}\log^{(1+\delta)}(n))$ for Gaussian circulant JL embeddings.

Traditionally, the key step in the JL lemma is to estimate the probability bounds of $\mathbb{P}(\|f(x)\|_2^2\geq(1+\epsilon)k)$ and $\mathbb{P}(\|f(x)\|_2^2\leq(1-\epsilon)k)$. The authors in [4] obtained the following estimations:

$$\mathbb{P}(\|f(x)\|_2^2 \geq (1+\epsilon)k) \leq \exp(-c(k\epsilon^2)^{1/3}) \tag{1.3}$$

and

$$\mathbb{P}(\|f(x)\|_2^2 \leq (1-\epsilon)k) \leq \exp(-c(k\epsilon^2)^{1/3}), \tag{1.4}$$

where c is an absolute constant. One can see that it is just the power 1/3 making the bound to be $k\epsilon^2 \sim \log^3(n)$, i.e., $k = O(\epsilon^{-2}\log^3(n))$. Vybíral [12] improved the right-hand side of inequalities (1.3) and (1.4) to $\exp(-\frac{ck\epsilon^2}{\log n})$, and hence directly derived a better bound $k = O(\epsilon^{-2}\log^2(n))$. Our main result, stated in Theorem 1.2, is more general and can recover the result in [12] under a strictly weaker constraint on the number n if $d > 12$. Also, the bound for Gaussian circulant JL embeddings, derived in Corollary 1.3, offers an improvement over existing bounds.

Theorem 1.2 (Main result). *Let $k \leq d$ be natural numbers and let $\epsilon \in (0, \frac{1}{2})$. Let $x \in \mathbb{R}^d$ be a unit vector, $a = (a_0, a_1, \cdots, a_{d-1}) \sim N_d(0, I_d)$. Assume that f is a composition of a $k \times d$ Gaussian circulant matrix $M_{a,k}$ with a $d \times d$ random diagonal matrix $D_\varkappa$, i.e., $f(x) = M_{a,k}D_\varkappa x$. Then with probability at least $1-(d+k)e^{-\frac{\tau \log^\delta n}{2}}$, it holds that*

$$\mathbb{P}(\|f(x)\|_2^2 \geq (1+\epsilon)k) \leq \exp\left(-\frac{c(\tau)k\epsilon^2}{\log^\delta n}\right) \tag{1.5}$$

and

$$\mathbb{P}(\|f(x)\|_2^2 \leq (1-\epsilon)k) \leq \exp\left(-\frac{c(\tau)k\epsilon^2}{\log^\delta n}\right), \tag{1.6}$$

where $c(\tau) = \frac{1}{8\tau}$, δ, and τ are positive parameters.

Remark 1.1. Setting $\tau = 2$ and $\delta = 1$ in Theorem 1.2 gives $c(\tau) = \frac{1}{16}$ and

$$1-(d+k)e^{-\frac{\tau \log^\delta n}{2}} = 1 - \frac{d+k}{n}.$$

Thus, letting $1 - \frac{d+k}{n} \geq \frac{5}{6}$, i.e., $n \geq \sqrt{6(d+k)}$, Theorem 1.2 rederives the inequalities (3.4) and (3.5) in [12] with an explicit value $c = \frac{1}{16}$, and the condition on n is strictly relaxed from $n \geq d$ to $n \geq \sqrt{6(d+k)}$ when $d > 12$, since $n \geq d > \sqrt{6(d+k)}$ for any $k < d$; for more details please refer to Lemma 3.1 and the proof of Theorem 1.3 in [12].

The following corollary follows by a union bound over C_n^2 pairs of points; note that we have set $\tau = 2$ for simplicity.

Corollary 1.3. *Let $x^1, x^2, \cdots, x^n$ be n points in the d-dimensional Euclidean space $\mathbb{R}^d$. Let $\epsilon \in (0, \frac{1}{2})$ and let $k = O(\epsilon^{-2}\log^{(1+\delta)}(n))$ be a natural number, where $\delta > 0$. Assume that f is a composition of a $k \times d$ Gaussian circulant matrix $M_{a,k}$ with a $d \times d$ random diagonal matrix $D_\varkappa$, i.e., $f(x) = \frac{1}{\sqrt{k}}M_{a,k}D_\varkappa x$. Then with probability at least $\frac{2}{3}\left(1-(d+k)e^{-\log^\delta n}\right)$, the following holds:*

$$(1-\epsilon)\|x^i - x^j\|_2^2 \leq \|f(x^i) - f(x^j)\|_2^2 \leq (1+\epsilon)\|x^i - x^j\|_2^2, \quad i,j, = 1, \cdots, n. \tag{1.7}$$

Remark 1.2. Compared with the current best bound (1.2), Corollary 1.3 only offers an improved bound for a relatively small non asymptotic range of n. In fact, in order for the stated probability to be positive, we derive that $\log(d+k)<\log^{\delta}(n)$ and hence $\log(d)<\log^{\delta}(n)$. On the other hand, we need $\log^{\delta}(n)\le\log^2(d)$ to have an improved estimate over (1.2). Therefore, for the parameter regimes satisfying

$$\log(d)<\log^{\delta}(n)\le\log^2(d),$$

Corollary 1.3 indeed offers an improved bound $k=O(\epsilon^{-2}\log^{(1+\delta)}(n))$ for Gaussian circulant JL embedding. However, once n is sufficiently large that $\log^2(d)\le\log^{\delta}(n)$, the derived bound in Corollary 1.3 becomes increasingly worse than (1.2). In other words, the bound (1.2) is asymptotically stronger than that in Corollary 1.3.

1.2. Extension. We generalize the main result to the case of subgaussian circulant JL embedding by borrowing the Laplace transform technique (also called Bernstein's trick). Here X is a subgaussian random variable with constant η referring to $E[\exp(tX)]\le\exp(\eta t^2)$ for some $\eta>0$. We only discuss the case of $\eta\le 1/2$, which includes many types of random circulant matrices we are interested in. An important type is the Bernoulli circulant matrix. In fact, if X is a Bernoulli random variable, then $E[\exp(tX)]=\frac{1}{2}\exp(t)+\frac{1}{2}\exp(-t)=\cosh(t)\le\exp(\frac{1}{2}t^2)$. So the Bernoulli random variable X is subgaussian with $\eta=\frac{1}{2}$. For the subgaussian case, we have the following results.

Theorem 1.4. *Let $k\le d$ be natural numbers and let $\epsilon\in(0,\frac{1}{2})$. Let $x\in\mathbb{R}^d$ be a unit vector, and choose a subgaussian vector $a=(a_0,a_1,\cdots,a_{d-1})$ having a uniform subgaussian constant $\eta>0$. Assume that f is a composition of a $k\times d$ subgaussian circulant matrix $M_{a,k}$ with a $d\times d$ random diagonal matrix $D_{\varkappa}$, i.e., $f(x)=M_{a,k}D_{\varkappa}x$. Then with probability at least $1-(d+k)e^{-\frac{\tau\log^{\delta}n}{2}}$ it holds that*

$$\mathbb{P}(\|f(x)\|_2^2\ge(1+\epsilon)k)\le\exp\left(-\frac{c(\theta,\eta,\tau)k\epsilon^2}{\log^{2\delta}n}\right) \tag{1.8}$$

and

$$\mathbb{P}(\|f(x)\|_2^2\le(1-\epsilon)k)\le\exp\left(-\frac{c(\theta,\eta,\tau)k\epsilon^2}{\log^{2\delta}n}\right). \tag{1.9}$$

Here, $c(\theta,\eta,\tau)=\theta(\frac{1}{2\tau\eta}-4\theta)$ is some absolute constant, where $0<\theta<\min\{1,\frac{1}{8\eta\tau}\}$ and $\delta,\tau>0$ are fixed parameters, the number n needs to be set big enough such that $\frac{2\theta\epsilon}{\log^{\delta}n}<\frac{1}{2}$, and the subgaussian constant η obeys $\frac{1}{2}\frac{1-\beta^2}{1+\beta^2}\le\eta\le\frac{1}{2}$ with $\beta=\frac{\theta\epsilon}{\tau\log^{2\delta}n}<\frac{1}{2}$.

Again, the following corollary follows by a union bound over C_n^2 pairs of points and setting $\tau=2$.

Corollary 1.5. *Let $x^1,x^2,\cdots,x^n$ be n points in the d-dimensional Euclidean space $\mathbb{R}^d$. Let $\epsilon\in(0,\frac{1}{2})$ and let $k=O(\epsilon^{-2}\log^{1+2\delta}n)$ be a natural number, where δ is a fixed positive parameter. Assume that f is a composition of a $k\times d$ subgaussian circulant matrix $M_{a,k}$ with a $d\times d$ random diagonal matrix $D_{\varkappa}$, i.e., $f(x)=\frac{1}{\sqrt{k}}M_{a,k}D_{\varkappa}x$. Assume that the subgaussian constant η obeys $\frac{1}{2}\frac{1-\beta^2}{1+\beta^2}\le\eta\le\frac{1}{2}$, where $\beta=\frac{\theta\epsilon}{\log^{2\delta}n}<1$, $0<\theta<\min\{1,\frac{1}{16\eta}\}$, and n is big enough such that $\frac{2\theta\epsilon}{\log^{\delta}n}<\frac{1}{2}$. Then with probability at least $\frac{2}{3}\left(1-(d+k)e^{-\log^{\delta}(n)}\right)$ the following holds:*

$$(1-\epsilon)\|x^i-x^j\|_2^2\le\|f(x^i)-f(x^j)\|_2^2\le(1+\epsilon)\|x^i-x^j\|_2^2,\quad i,j,=1,\cdots,n. \tag{1.10}$$

Remark 1.3. The bound $k = O(\epsilon^{-2}\log^{1+2\delta} n)$ is independent of the parameters η, β, and θ. These parameters are used to bound the subgaussian constant. In other words, the conclusion in Corollary 1.5 only applies to some special subgaussian cases.

Remark 1.4. Although our main result can be extended to the subgaussian case, we have to admit that the bound $k = O(\epsilon^{-2}\log^{1+2\delta}(n))$ in Corollary 1.5 is weaker than (1.2) due to the factor $\log^{2\delta}(n)$ and the implicit requirement $\log(d) < \log^{\delta}(n)$ in the probability bound. However, our analysis is more direct than that in [7] and our bound is comparable to (1.2) when the number of points n is approximately the same as the ambient dimension d.

2. Proof of Theorem 1.2

In this section, we will prove Theorem 1.2 by showing that for any fixed unit vector x, $f(x) = M_{a,k}D_{\varkappa}x$ has the concentration property. We divide the proof of Theorem 1.2 into three steps. Since the random matrix $M_{a,k}D_{\varkappa}$ couples the random vectors a and $\varkappa$ together, the first step decouples these two random vectors so that we can apply some existing concentration results to them separately. The second step estimates the spectral norm of random matrix Y whose randomness is from the Bernoulli random vector $\varkappa$. By using the special structure of the random matrix Y, we deduce a tighter and more general estimate than that from [12]. Our derivation relies on the matrix-valued Berstein inequality in [11]. The last step is a direct application of the concentration of quadratic function to the Gaussian random vector a.

Step 1: Decoupling. We define the matrix

$$Y = \begin{pmatrix} x_0\varkappa_0 & x_1\varkappa_1 & x_2\varkappa_2 & \cdots & x_{d-1}\varkappa_{d-1} \\ x_1\varkappa_1 & x_2\varkappa_2 & x_3\varkappa_3 & \cdots & x_0\varkappa_0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{k-1}\varkappa_{k-1} & x_k\varkappa_k & x_{k+1}\varkappa_{k+1} & \cdots & x_{k-2}\varkappa_{k-2} \end{pmatrix} \in \mathbb{R}^{k\times d}.$$

Then it holds that

$$\|f(x)\|_2^2 = \|M_{a,k}D_{\varkappa}x\|_2^2 = \|Ya\|_2^2.$$

Let $Y = U\Sigma V^T$ be the singular value decomposition of Y. Since $Y \in \mathbb{R}^{k\times d}$, we take matrices $U \in \mathbb{R}^{k\times k}, V \in \mathbb{R}^{d\times k}$ to be real orthogonal matrices [5]. Thus $b = V^T a$ is a $k-$dimensional vector of independent Gaussian variables. Hence,

$$\|Ya\|_2^2 = \|U\Sigma V^T a\|_2^2 = \|U\Sigma b\|_2^2 = \|\Sigma b\|_2^2 = \sum_{j=0}^{k-1} |\lambda_j|^2 b_j^2,$$

where $\lambda_j, j = 0, 1, \cdots, k-1$ are the singular values of Y, and $b_j = \sum_{i=0}^{d-1} V_{ij}a_i$. Let $\mu_j = |\lambda_j|^2$. Then

$$\|\mu\|_1 = \sum_{j=0}^{k-1} |u_j| = \sum_{j=0}^{k-1} |\lambda_j|^2 = \|Y\|_F^2 = k, \tag{2.1}$$

where $\|Y\|_F$ is the Frobenius norm of Y, and the last identity is due to the fact that $x \in \mathbb{R}^d$ is a unit vector.

Step 2: Spectral estimate. While the analysis of the decoupling process in the first step closely follows from Vybíral [12], the estimate of the spectral norm of Y is quite different. We begin with the following lemma [11].

Lemma 2.1 (Matrix-valued Bernstein inequality). *Consider a finite sequence $\{B_i\}$ of fixed matrices with dimension $d_1 \times d_2$, and let $\{\xi_i\}$ be a finite sequence of independent standard normal variables or symmetrical Bernoulli variables. Then, for all $t \geq 0$,*

$$\mathbb{P}\{\|\sum_i \xi_i B_i\| \geq t\} \leq (d_1 + d_2) e^{-t^2/2\sigma^2}, \tag{2.2}$$

where

$$\sigma^2 := \max\{\|\sum_i B_i B_i^T\|, \|\sum_i B_i^T B_i\|\}.$$

To apply this lemma to our case, we define two $d \times d$ permutation matrices

$$P = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{d \times d} \quad \text{and} \quad C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 & 0 \end{pmatrix} \in \mathbb{R}^{d \times d}.$$

Let $\mathrm{S} = (I_k \quad 0_{k \times (d-k)})$. By multiplying matrix S at the left-hand side of an arbitrary matrix, one obtains its first k rows as a new rectangular matrix with dimension $k \times d$; thus the matrix Y can be written in the form

$$Y = \sum_{i=0}^{d-1} \varkappa_i x_i \mathrm{S} P^i C \triangleq \sum_{i=0}^{d-1} \varkappa_i B_i, \tag{2.3}$$

where the random matrix $B_i = x_i \mathrm{S} P^i C$. Now, we estimate the spectral norm of random matrix Y by using Lemma 2.1.

Lemma 2.2. *Let* Y *be defined as before. Then it holds that*

$$\mathbb{P}\{\|\mathrm{Y}\| \geq t\} \leq (d + k) e^{-t^2/2}. \tag{2.4}$$

Proof. By Lemma 2.1, we only need to show that

$$\max\left\{\left\|\sum_{i=0}^{d-1} B_i B_i^T\right\|, \left\|\sum_{i=0}^{d-1} B_i^T B_i\right\|\right\} = 1,$$

where $B_i = x_i \mathrm{S} P^i C$. In fact, on one hand,

$$\sum_{i=0}^{d-1} B_i B_i^T = \sum_{i=0}^{d-1} x_i^2 \mathrm{S} \mathrm{P}^i \mathrm{C} \mathrm{C}^T (\mathrm{P}^i)^T \mathrm{S}^T = \sum_{i=0}^{d-1} x_i^2 I_k, \tag{2.5}$$

where we have employed the property $Q^T = Q^{-1}$ for every permutation Q, $\mathrm{S}\mathrm{S}^T = I_k$, $\mathrm{C}\mathrm{C}^T = I_d$, and $\mathrm{P}^i (\mathrm{P}^i)^T = I_d$. Since x is a unit vector, we get $\sum_{i=0}^{d-1} B_i B_i^T = I_k$, which implies $\|\sum_{i=0}^{d-1} B_i B_i^T\| = 1$. On the other hand,

$$\sum_{i=0}^{d-1} B_i^T B_i = \sum_{i=0}^{d-1} x_i^2 \mathrm{C}^T (\mathrm{P}^i)^T \mathrm{S}^T \mathrm{S} \mathrm{P}^i \mathrm{C} \tag{2.6a}$$

$$=\sum_{i=0}^{d-1} x_i^2 \mathrm{C}^T(\mathrm{P}^i)^T \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \mathrm{P}^i \mathrm{C} \tag{2.6b}$$

$$\preceq \sum_{i=0}^{d-1} x_i^2 I_d = I_d. \tag{2.6c}$$

Thus, $\|\sum_{i=0}^{d-1} B_i^T B_i\| \le 1$. This completes the proof. □

Taking $t=\sqrt{\tau}\log^{\delta/2} n$ with δ,τ being positive parameters in the probability inequality (2.4), we have the estimation

$$\|\mu\|_\infty = \|\lambda\|_\infty^2 = \|\mathbf{Y}\|^2 \le \tau \log^\delta n \tag{2.7}$$

with probability at least $1-(d+k)e^{-\frac{\tau\log^\delta n}{2}}$. From (2.1) and (2.7),

$$\|\mu\|_2 \le \sqrt{\|\mu\|_1 \|\mu\|_\infty} \le \sqrt{\tau k \log^\delta n} \tag{2.8}$$

holds with probability at least $1-(d+k)e^{-\frac{\tau\log^\delta n}{2}}$.

Step 3: Concentration. To finish the proof, we need the following concentration result [8], which is also the main tool employed in [4, 12].

LEMMA 2.3. *Let $Z=\sum_{i=1}^s \alpha_i(a_i^2-1)$, where the a_i are independent identically distributed (i.i.d.) normal variables and α_i are nonnegative numbers. Then for any $t>0$,*

$$\mathbb{P}(Z \ge 2\|\alpha\|_2\sqrt{t} + 2\|\alpha\|_\infty t) \le \exp(-t), \tag{2.9}$$

$$\mathbb{P}(Z \le -2\|\alpha\|_2\sqrt{t}) \le \exp(-t). \tag{2.10}$$

Now, let us complete the proof. First, we have

$$\mathbb{P}(\|Ya\|_2^2 \ge (1+\epsilon)k) = \mathbb{P}\Big(\sum_{j=0}^{k-1} \mu_j(b_j^2-1) \ge \epsilon k\Big). \tag{2.11}$$

Denote $Z=\sum_{j=0}^{k-1}\mu_j(b_j^2-1)$; then we need to estimate $\mathbb{P}(Z\ge k\epsilon)$. By the estimation (2.9) in Lemma 2.3, we get

$$\mathbb{P}(Z \ge 2\|\mu\|_2\sqrt{t} + 2\|\mu\|_\infty t) \le \exp(-t). \tag{2.12}$$

Using (2.7) and (2.8), we derive

$$\mathbb{P}(Z \ge 2\sqrt{\tau k t \log^\delta n} + 2\tau t \log^\delta n) \le \exp(-t). \tag{2.13}$$

Setting $t=\frac{c(\tau)k\epsilon^2}{\log^\delta n}$ with $c(\tau)=\frac{1}{8\tau}$, we have

$$2\sqrt{\tau k t \log^\delta n} + 2\tau t \log^\delta n = \Big(\frac{\sqrt{2}}{2} + \frac{\epsilon}{4}\Big)k\epsilon \le k\epsilon. \tag{2.14}$$

Thus, we finally get

$$\mathbb{P}(Z \ge k\epsilon) \le \exp\Big(-\frac{c(\tau)k\epsilon^2}{\log^\delta n}\Big), \tag{2.15}$$

which shows (1.5). The inequality (1.6) can be proved in the same manner by invoking the estimation (2.10) in Lemma 2.3.

3. Proof of Theorem 1.4

For the subgaussian case, we provide a direct proof by using the Laplace transform technique. First, we need the following lemma.

Lemma 3.1. *If X is subgaussian with constant $\eta>0$ and $X_i\sim X$ are i.i.d., then*

$$E[\exp(\lambda W^2)]\leq \frac{1}{\sqrt{1-4\eta\lambda}}, \tag{3.1}$$

where $W=\sum_{i=0}^{k-1}X_i\beta_i$ with $\beta_i\in\mathbb{R}$ satisfying $\sum_{i=0}^{k-1}\beta_i^2=1$. Moreover, define $\varphi(\lambda)=\log E[\exp(\lambda(W^2-1))]$. Then it holds that

$$\varphi(\lambda)\leq \frac{8\eta^2\lambda^2}{1-4\eta\lambda}, \quad \textit{for} \quad \lambda<\frac{1}{4\eta} \quad \textit{and} \quad \eta\leq\frac{1}{2}. \tag{3.2}$$

Proof. For the proof of the first part, see [9]. Here we only show the second part. Using the estimate of bound $E[\exp(\lambda W^2)]$ in (3.1) and the conditions $\lambda<\frac{1}{4\eta}$ and $\eta\leq\frac{1}{2}$, we calculate

$$\begin{aligned}
\varphi(\lambda)&\leq -\frac{1}{2}\log(1-4\eta\lambda)-\lambda && \text{(3.3a)}\\
&=2\eta\lambda-\lambda+\sum_{m=2}^{\infty}2^{m-1}\frac{(2\eta\lambda)^m}{m} && \text{(3.3b)}\\
&\leq\sum_{m=2}^{\infty}2^{m-1}\frac{(2\eta\lambda)^m}{m}=\sum_{m=2}^{\infty}\frac{8\eta^2\lambda^2(4\eta\lambda)^{m-2}}{m} && \text{(3.3c)}\\
&\leq\sum_{m=2}^{\infty}8\eta^2\lambda^2(4\eta\lambda)^{m-2}=\sum_{m=0}^{\infty}8\eta^2\lambda^2(4\eta\lambda)^m && \text{(3.3d)}\\
&=\frac{8\eta^2\lambda^2}{1-4\eta\lambda}, \quad \text{for} \quad \lambda<\frac{1}{4\eta} \quad \text{and} \quad \eta\leq\frac{1}{2}, && \text{(3.3e)}
\end{aligned}$$

which completes the proof. □

We divide the proof of Theorem 1.4 into two parts.

Part A: Proof of probability inequality (1.8). Similar to the argument in the proof of Theorem 1.2, we need to estimate

$$\mathbb{P}\Big(\sum_{j=0}^{k-1}\mu_j(b_j^2-1)\geq\epsilon k\Big), \tag{3.4}$$

where $b_j=\sum_{i=0}^{d-1}V_{ij}a_i$ is not a Gaussian variable but a linear combination of subgaussian variables, i.e., each b_j has the form of W in Lemma 3.1. So we cannot directly invoke Lemma 2.2. Here we use the Laplace transform technique to complete the proof. We derive that

$$\begin{aligned}
\mathbb{P}\Big(\sum_{j=0}^{k-1}\mu_j(b_j^2-1)\geq\epsilon k\Big)&=\mathbb{P}(\exp\Big(\sum_{j=0}^{k-1}\lambda\mu_j(b_j^2-1)\Big)\geq\exp(\lambda\epsilon k)), \quad \textit{for} \quad \lambda>0 && \text{(3.5a)}\\
&\leq\inf_{\lambda>0}\frac{E\Big[\exp\Big(\sum_{j=0}^{k-1}\lambda\mu_j(b_j^2-1)\Big)\Big]}{\exp(\lambda\epsilon k)} && \text{(3.5b)}
\end{aligned}$$

$$= \inf_{\lambda>0} \frac{\prod_{j=0}^{k-1} E[\exp(\lambda\mu_j(b_j^2-1))]}{\exp(\lambda\epsilon k)} \tag{3.5c}$$

$$\le \inf_{0<\lambda\mu_j<1/4\eta} \frac{\prod_{j=0}^{k-1} \exp(\varphi(\lambda\mu_j))}{\exp(\lambda\epsilon k)}, \tag{3.5d}$$

where (3.5b) follows from the Markov inequality, (3.5c) follows from the independence of b_j, and (3.5d) is due to the additional restriction of λ and the expression of $\varphi(\cdot)$. Denote $f(\lambda)=\frac{8\eta^2\lambda^2}{1-4\eta\lambda}$; then it is a monotonically increasing function since its derivative is positive. Moreover $\|\mu\|_\infty \le \tau\log^\delta n$ with probability at least $1-(d+k)e^{-\frac{\tau\log^\delta n}{2}}$ from (2.8). Thus, together with Lemma 3.1 we have

$$\varphi(\lambda\mu_j)\le f(\lambda\mu_j)\le \frac{8\eta^2\lambda^2\tau^2\log^{2\delta}n}{1-4\eta\lambda\tau\log^\delta n}, \quad for \quad j=0,\cdots,k-1. \tag{3.6}$$

With this uniform bound and a tighter restriction of λ, we continue to estimate the probability inequality (3.5d) and get

$$\mathbb{P}\Big(\sum_{j=0}^{k-1}\mu_j(b_j^2-1)\ge\epsilon k\Big)\le \inf_{0<\lambda<1/4\eta\tau\log^\delta n}\exp\left(\frac{8k\eta^2\lambda^2\tau^2\log^{2\delta}n}{1-4\eta\lambda\tau\log^\delta n}-\lambda\epsilon k\right). \tag{3.7}$$

Take $\lambda=\frac{\theta\epsilon}{2\eta\tau\log^{2\delta}n}$, where θ is a positive parameter and ϵ obeys $0<\epsilon<1/2$. In order to satisfy the constraint $0<\lambda<(4\eta\tau\log^\delta n)^{-1}$, one needs to require that $0<\theta<1$. Now, using this special choice of λ, we get an upper bound

$$\mathbb{P}\Big(\sum_{j=0}^{k-1}\mu_j(b_j^2-1)\ge\epsilon k\Big)\le\exp\left(-\frac{k\epsilon^2}{\log^{2\delta}n}\left(\frac{\theta}{2\eta\tau}-\frac{2\theta^2}{1-\frac{2\theta\epsilon}{\log^\delta n}}\right)\right). \tag{3.8}$$

For any fixed parameter δ, let n be big enough such that $\frac{2\theta\epsilon}{\log^\delta n}<\frac12$. Then the upper bound can be relaxed to

$$\mathbb{P}\Big(\sum_{j=0}^{k-1}\mu_j(b_j^2-1)\ge\epsilon k\Big)\le\exp\Big(-\frac{k\epsilon^2}{\log^{2\delta}n}\Big(\frac{\theta}{2\eta\tau}-4\theta^2\Big)\Big). \tag{3.9}$$

Let $c(\theta,\eta,\tau)=\theta(\frac{1}{2\eta\tau}-4\theta)$; then it is a positive constant depending on parameters θ,η,τ if $\theta<\frac{1}{8\eta\tau}$. Thus, the probability inequality (1.8) holds.

Part B: Proof of probability inequality (1.9). In the following, we will show that the inequality (1.9) can be obtained in the same manner under the additional parameters constraint $\frac12\frac{1-\beta^2}{1+\beta^2}\le\eta\le\frac12$ where $\beta=\frac{\theta\epsilon}{\tau\log^{2\delta}n}<\frac12$. Our aim is to estimate

$$\mathbb{P}\Big(\sum_{j=0}^{k-1}\mu_j(1-b_j^2)\ge\epsilon k\Big). \tag{3.10}$$

Define a new function

$$\phi(\lambda)=\log E[\exp(\lambda(1-W^2))], \tag{3.11}$$

where the random variable W is defined as in Lemma 3.1. Applying the Laplace transform technique and using the new function above, we get

$$\mathbb{P}\Big(\sum_{j=0}^{k-1}\mu_j(1-b_j^2)\geq \epsilon k\Big)\leq \inf_{\lambda>0}\frac{E[\exp(\sum_{j=0}^{k-1}\lambda\mu_j(1-b_j^2))]}{\exp(\lambda\epsilon k)} \tag{3.12a}$$

$$\leq \inf_{0<\lambda\mu_j<1/4\eta}\frac{\prod_{j=0}^{k-1}\exp(\phi(\lambda\mu_j))}{\exp(\lambda\epsilon k)}. \tag{3.12b}$$

If we could prove the inequality

$$\phi(\lambda)\leq \frac{8\eta^2\lambda^2}{1-4\eta\lambda}, \quad \text{when} \quad 2\lambda\eta=\beta<\frac{1}{2} \quad \text{and} \quad \frac{1}{2}\frac{1-\beta^2}{1+\beta^2}\leq \eta \leq \frac{1}{2}, \tag{3.13}$$

where $\beta=\frac{\theta\epsilon}{\tau\log^{2\delta}n}$, then we can prove (1.9) as **Part A** because setting $\lambda=\frac{\theta\epsilon}{2\tau\eta\log^{2\delta}n}$, $\varphi(\lambda)$ and $\phi(\lambda)$ take the same upper bound. Now, let us show inequality (3.13) as follows:

$$\phi(\lambda)\leq -\frac{1}{2}\log(1+4\eta\lambda)+\lambda \tag{3.14a}$$

$$=-\frac{1}{2}\sum_{m=1}^{\infty}(-1)^{m-1}\frac{(4\eta\lambda)^m}{m}+\lambda \tag{3.14b}$$

$$=\frac{1}{2}\sum_{m=1}^{\infty}\frac{(4\eta\lambda)^m}{m}-\sum_{m\equiv 1(mod 2)}^{\infty}\frac{(4\eta\lambda)^m}{m}+\lambda \tag{3.14c}$$

$$=\sum_{m=2}^{\infty}2^{m-1}\frac{(2\eta\lambda)^m}{m}+2\eta\lambda-\sum_{l=1}^{\infty}\frac{(4\eta\lambda)^{2l-1}}{2l-1}+\lambda, \tag{3.14d}$$

where (3.14a) follows from the first part of Lemma 3.1. From (3.3c) to (3.3e), it holds under the condition $\lambda\eta<\frac{1}{4}$ that

$$\sum_{m=2}^{\infty}2^{m-1}\frac{(2\eta\lambda)^m}{m}\leq\frac{8\eta^2\lambda^2}{1-4\eta\lambda}. \tag{3.15}$$

Denote $g(\lambda)=\sum_{l=1}^{\infty}\frac{(4\eta\lambda)^{2l-1}}{2l-1}$. Then

$$g(\lambda)=\sum_{l=1}^{\infty}2\frac{2^{2l-2}}{2l-1}(2\eta\lambda)^{2l-1}\geq 2\sum_{l=1}^{\infty}(2\eta\lambda)^{2l-1}=\frac{4\eta\lambda}{1-4\eta^2\lambda^2}, \tag{3.16}$$

where the inequality follows from that $\frac{2^{2l-2}}{2l-1}\geq 1$ for every positive number l. Hence, it suffices to show $2\eta\lambda-\frac{4\eta\lambda}{1-4\eta^2\lambda^2}+\lambda\leq 0$, or equivalently to show $\beta-\frac{2\beta}{1-\beta^2}+\frac{\beta}{2\eta}\leq 0$ since $2\lambda\eta=\beta$. After a simple calculation, one needs $\frac{1}{2}\frac{1-\beta^2}{1+\beta^2}\leq\eta$, which is just the assumed condition. Thus, the inequality (3.13) holds and hence the estimate (1.9) follows.

Remark 3.1. The condition on the subgaussian constant $\frac{1}{2}\frac{1-\beta^2}{1+\beta^2}\leq\eta$ is only required in estimating (1.9). Such a requirement can guarantee that inequality (3.13) holds and hence gives us a uniform probability estimates. If one gives up the uniform expressions in (1.8) and (1.9), then the condition may be relaxed. We leave the possible improvements of the lower bound on the subgaussian constant open for further investigations.

Acknowledgment. The authors thank the anonymous reviewers for their valuable comments and referring us to [7], which greatly improved the quality of the paper, and thank Profs. Jan Vybíral and Wotao Yin for their helpful discussion, and also thank Dr. Ming Yan for a careful reading of the manuscript. The work of H. Zhang is supported by China Scholarship Council during his visit to Rice University, and in part by the Graduate School of NUDT under Funding of Innovation B110202 and by Hunan Provincial Innovation Foundation For Postgraduate CX2011B008. The work of L.Z. Cheng is supported by the National Science Foundation of China under Grants No. 61271014 and No.61072118. H. Zhang thanks Rice University, CAAM Department, for hosting him.

REFERENCES

[1] N. Ailon and E. Liberty, *Almost optimal unrestricted fast Johnson-Lindenstrauss transform*, Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, 185–191, 2011.

[2] E.J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52(2), 489–509, 2006.

[3] F. Krahmer and R. Ward, *New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property*, SIAM J. Math. Anal., 43(3), 1269–1281, 2011.

[4] A. Hinrichs and J. Vybíral, *Johnson-Lindenstrauss lemma for circulant matrices*, Random. Struct. Algor., 39(3), 391–398, 2011.

[5] R.A. Horn and C.R. Johnson, *Topics in Matrices Analysis*, Posts and Telecom. Press, 2007.

[6] W.B. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Contemp. Math., 26, 189–206, 1984.

[7] F. Krahmer, S. Mendelson, and H. Rauhut, *Suprema of Chaos Processes and the Restricted Isometry Property*, Comm. Pure Appl. Math., accepted, 2013. arXiv:1207.0235

[8] B. Laurent and P. Massart, *Adaptive estimation of a quadratic functional by model selection*, Ann. Stat., 28(5), 1302–1338, 2000.

[9] J.R. Lee, *Randomized Algorithms and Probabilistic Analysis*, Lecture 10: Feb 11, CSE 525, 2008.

[10] J. Matoušk, *On variants of the Johnson-Lindenstrauss lemma*, Random. Struct. Algor., 33(2), 142–156, 2008.

[11] J.A. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math., 12(4), 389–434, 2012.

[12] J. Vybíral, *A variant of the Johnson-Lindenstrauss lemma for circulant matrices*, J. Funct. Anal., 260, 1096–1105, 2011.