Available online at www.sciencedirect.com

ScienceDirect

Journal of Functional Analysis 260 (2011) 1096–1105

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

A variant of the Johnson–Lindenstrauss lemma for circulant matrices

Jan Vybíral

Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria

Received 15 February 2010; accepted 26 November 2010

Available online 4 December 2010

Communicated by K. Ball

Abstract

We continue our study of the Johnson–Lindenstrauss lemma and its connection to circulant matrices started in Hinrichs and Vybíral (in press) [7]. We reduce the bound on k from $k = \Omega(\varepsilon^{-2}\log^3 n)$ proven there to $k = \Omega(\varepsilon^{-2}\log^2 n)$. Our technique differs essentially from the one used in Hinrichs and Vybíral (in press) [7]. We employ the discrete Fourier transform and singular value decomposition to deal with the dependency caused by the circulant structure.

Keywords: Johnson–Lindenstrauss lemma; Circulant matrix; Discrete Fourier transform; Singular value decomposition

1. Introduction

Let $x^1,\dots,x^n \in \mathbb{R}^d$ be n points in the d-dimensional Euclidean space $\mathbb{R}^d$. The classical Johnson–Lindenstrauss lemma tells that, for a given $\varepsilon \in (0,\frac{1}{2})$ and a natural number $k = \Omega(\varepsilon^{-2}\log n)$, there exists a linear map $f:\mathbb{R}^d \to \mathbb{R}^k$, such that

$$(1-\varepsilon)\|x^j\|_2^2 \leqslant \|f(x^j)\|_2^2 \leqslant (1+\varepsilon)\|x^j\|_2^2$$

for all $j \in \{1,\dots,n\}$.

E-mail address: jan.vybiral@oeaw.ac.at.

0022-1236/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jfa.2010.11.014

Here $\|\cdot\|_2$ stands for the Euclidean norm in $\mathbb{R}^d$ or $\mathbb{R}^k$, respectively. Furthermore, here and any time later, the condition $k = \Omega(\varepsilon^{-2}\log n)$ means, that there is an absolute constant $C > 0$, such that the statement holds for all natural numbers k with $k \geqslant C\varepsilon^{-2}\log n$. We shall also always assume, that $k \leqslant d$. Otherwise, the statement becomes trivial.

The original proof of this fact was given by Johnson and Lindenstrauss in [9]. We refer to [6] for a beautiful and self-contained proof. Since then, it has found many applications for example in algorithm design. These applications inspired numerous variants and improvements of the Johnson–Lindenstrauss lemma, which try to minimize the computational costs of $f(x)$, the memory used, the number of random bits used and to simplify the algorithm to allow an easy implementation. We refer to [8,1–3,12] for details and to [12] for a nice description of the history and the actual "state of the art".

All the known proofs of the Johnson–Lindenstrauss lemma work with random matrices and proceed more or less in the following way. One considers a probability measure $\mathbb{P}$ on a some subset $\mathcal{P}$ of all $k \times d$ matrices (i.e. all linear mappings $\mathbb{R}^d \to \mathbb{R}^k$). The proof of the Johnson–Lindenstrauss lemma then emerges by some variant of the following two estimates

$$\mathbb{P}\big(f \in \mathcal{P}\colon\ \big\|f(x)\big\|_2^2 \geqslant 1+\varepsilon\big) < 1 - \frac{1}{2n}$$

and

$$\mathbb{P}\big(f \in \mathcal{P}\colon\ \big\|f(x)\big\|_2^2 \leqslant 1-\varepsilon\big) < 1 - \frac{1}{2n},$$

which have to be proven for all unit vectors $x \in \mathbb{R}^d$, and a simple union bound over all points $x^j/\|x^j\|_2$, $j = 1,\dots,n$. Here and later on we assume, without loss of generality, that $x^j \neq 0$ for all $j = 1,\dots,n$.

The biggest breakthrough in the attempts to minimize the running time of f was achieved by Ailon and Chazelle in [2] (with improvements by Matoušek [12] and Ailon and Liberty [4]). The mapping f is given in [2] as the composition of a sparse matrix, a certain random Fourier matrix and a random diagonal matrix. The value $f(x)$ can be computed with high probability very efficiently, i.e. using $O(d\log d + \min\{d\varepsilon^{-2}\log n, \varepsilon^{-2}\log^3 n\})$ operations. This was later further improved by Ailon and Liberty to $O(d\log k)$ for $k = O(d^{1/2-\delta})$, for any arbitrary small fixed $\delta > 0$.

In [7], we studied a different construction of f, namely the possibility of a composition of a random circulant matrix with a random diagonal matrix. As a multiple of a circulant matrix may be implemented with the help of a discrete Fourier transform, it provides the running time of $O(d\log d)$, requires very few random bits (only $2d$ random bits in the case of Bernoulli variables) and allows a very simple implementation, as the Fast Fourier Transform is a part of every standard mathematical software package.

The main difference between this approach and the usual constructions available in the literature is that the components of $f(x)$ are now no longer independent random variables. Decoupling this dependence, we were able to prove in [7] the Johnson–Lindenstrauss lemma for composition of a random circulant matrix and a random diagonal matrix, but only for $k = \Omega(\varepsilon^{-2}\log^3 n)$. It is the main aim of this note to improve this bound to $k = \Omega(\varepsilon^{-2}\log^2 n)$. This comes essentially closer to the standard bound $k = \Omega(\varepsilon^{-2}\log n)$. Reaching this optimal bound (and keeping the control of the constants involved) remains an open problem and a subject of a challenging research.

We use a completely different technique here. We use the discrete Fourier transform and the singular value decomposition of circulant matrices. That is the reason, why we found it more instructive to state and prove our variant of Johnson–Lindenstrauss lemma for complex vectors and Gaussian random variables. As a corollary, we obtain of course a corresponding real version.

Before we state our main result, we give the necessary definitions.

Definition 1.1. Let α and β be independent real Gaussian random variables with

$$\mathbb{E}\alpha = \mathbb{E}\beta = 0 \quad \text{and} \quad \mathbb{E}|\alpha|^2 = \mathbb{E}|\beta|^2 = 1.$$

Then we call

$$a = \alpha + i\beta$$

a complex Gaussian variable.

Let us note, that if a is a complex Gaussian variable, then

$$\mathbb{E}a = \mathbb{E}\alpha + i\mathbb{E}\beta = 0 \quad \text{and} \quad \mathbb{E}|a|^2 = \mathbb{E}\alpha^2 + \mathbb{E}\beta^2 = 2.$$

Definition 1.2. (i) Let $k \leqslant d$ be natural numbers. Let $a = (a_0, \ldots, a_{d-1}) \in \mathbb{C}^d$ be a fixed complex vector. We denote by $M_{a,k}$ the partial circulant matrix

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

If $k = d$, we denote by $M_a = M_{a,d}$ the full circulant matrix. This notation extends naturally to the case, when $a = (a_0, \ldots, a_{d-1})$ are independent complex Gaussian variables.

(ii) If $\varkappa = (\varkappa_0, \ldots, \varkappa_{d-1})$ are independent Bernoulli variables, we put

$$D_\varkappa = \operatorname{diag}(\varkappa) := \begin{pmatrix} \varkappa_0 & 0 & \cdots & 0 \\ 0 & \varkappa_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varkappa_{d-1} \end{pmatrix} \in \mathbb{R}^{d\times d}.$$

Of course, $D_\varkappa : \mathbb{C}^d \to \mathbb{C}^d$ is an isomorphism.

Theorem 1.3. *Let $\varepsilon \in (0, \frac{1}{2})$, $n \geqslant d$ be natural numbers, and let $x^1, \ldots, x^n \in \mathbb{C}^d$ be n arbitrary points in $\mathbb{C}^d$. Let $a = (a_0, \ldots, a_{d-1})$ be d independent complex Gaussian variables and let $\varkappa = (\varkappa_0, \ldots, \varkappa_{d-1})$ be independent Bernoulli variables.*

If $k = \Omega(\varepsilon^{-2} \log^2 n)$ is a natural number, then the mapping $f : \mathbb{C}^d \to \mathbb{C}^d$ given by $f(x) = \frac{1}{\sqrt{2k}} M_{a,k} D_\varkappa x$ satisfies

$$(1-\varepsilon)\|x^j\|_2^2 \leqslant \|f(x^j)\|_2^2 \leqslant (1+\varepsilon)\|x^j\|_2^2$$

for all $j \in \{1,\dots,n\}$ with probability at least $2/3$. Here $\|\cdot\|_2$ stands for the ℓ_2-norm in $\mathbb{C}^d$ or $\mathbb{C}^k$, respectively.

For reader's convenience, we formulate also a variant of Theorem 1.3, which deals with real Euclidean spaces.

Corollary 1.4. *Let $\varepsilon \in (0, \frac{1}{2})$, $n \geqslant d$ be natural numbers, and let $x^1,\dots,x^n \in \mathbb{R}^{2d}$ be n arbitrary points in $\mathbb{R}^{2d}$. Let $\alpha_0,\dots,\alpha_{d-1},\beta_0,\dots,\beta_{d-1}$ be $2d$ independent real Gaussian variables and let $\varkappa = (\varkappa_0,\dots,\varkappa_{d-1})$ be independent Bernoulli variables.*

If $k = \Omega(\varepsilon^{-2}\log^2 n)$ is a natural number, then the mapping $f:\mathbb{R}^{2d}\to\mathbb{R}^{2k}$ given by

$$f(x) = \frac{1}{\sqrt{2k}}\begin{pmatrix} M_{\alpha,k} & -M_{\beta,k} \\ M_{\beta,k} & M_{\alpha,k}\end{pmatrix}\begin{pmatrix} D_\varkappa & 0 \\ 0 & D_\varkappa\end{pmatrix} x$$

satisfies

$$(1-\varepsilon)\|x^j\|_2^2 \leqslant \|f(x^j)\|_2^2 \leqslant (1+\varepsilon)\|x^j\|_2^2$$

for all $j \in \{1,\dots,n\}$ with probability at least $2/3$. Here $\|\cdot\|_2$ stands for the ℓ_2-norm in $\mathbb{R}^{2d}$ or $\mathbb{R}^{2k}$, respectively.

The proof follows trivially from Theorem 1.3 by considering complex Gaussian variables $a = (\alpha_0+i\beta_0,\dots,\alpha_{d-1}+i\beta_{d-1})$ and complex vectors $y^j = (x_0^j+ix_d^j,\dots,x_{d-1}^j+ix_{2d-1}^j) \in \mathbb{C}^d$, $j=1,\dots,n$.

2. Used techniques

We give an overview of the techniques used in the proof of Theorem 1.3.

2.1. Discrete Fourier transform

Our main tool in this note is the discrete Fourier transform. If d is a natural number, then the discrete Fourier transform $\mathcal{F}_d:\mathbb{C}^d\to\mathbb{C}^d$ is defined by

$$(\mathcal{F}_d x)(\xi) = \frac{1}{\sqrt{d}}\sum_{u=0}^{d-1} x_u \exp\left(-\frac{2\pi i u\xi}{d}\right).$$

With this normalization, $\mathcal{F}_d$ is an isomorphism of $\mathbb{C}^d$ onto itself. The inverse discrete Fourier transform is given by

$$(\mathcal{F}_d^{-1}x)(\xi) = \frac{1}{\sqrt{d}}\sum_{u=0}^{d-1} x_u \exp\left(\frac{2\pi i u\xi}{d}\right).$$

Observe, that the matrix representation of $\mathcal{F}_d^{-1}$ is the conjugate transpose of the matrix representation of $\mathcal{F}_d$, i.e. $\mathcal{F}_d^{-1}=\mathcal{F}_d^*$.

The fundamental connection between discrete Fourier transform and circulant matrices is given by

$$M_a=\mathcal{F}_d\,\operatorname{diag}(\sqrt{d}\mathcal{F}_d a)\mathcal{F}_d^{-1}, \tag{2.1}$$

which may be verified by direct calculation. Hence every circulant matrix may be diagonalized with the use of a discrete Fourier transform, its inverse and a multiple of the discrete Fourier transform of its first row.

2.2. Singular value decomposition

The last tool needed in the proof is the singular value decomposition. Let $M:\mathbb{C}^d\to\mathbb{C}^k$ be a $k\times d$ complex matrix with $k\leqslant d$. Then there exists a decomposition

$$M=U\Sigma V^*,$$

where U is a $k\times k$ unitary complex matrix, Σ is a $k\times k$ diagonal matrix with nonnegative entries on the diagonal, V is a $d\times k$ complex matrix with k orthonormal columns and V^* denotes the conjugate transpose of V. Hence V^* has k orthonormal rows. The entries of Σ are the singular values of M, namely the square roots of the eigenvalues of MM^*.

If $a=(a_0,\dots,a_{d-1})\in\mathbb{C}^d$ is a complex vector and M_a is the corresponding circulant matrix, then its singular values may be calculated using (2.1). We obtain

$$\begin{aligned} M_aM_a^* &= \mathcal{F}_d\operatorname{diag}(\sqrt{d}\mathcal{F}_d a)\mathcal{F}_d^{-1}\big[\mathcal{F}_d\operatorname{diag}(\sqrt{d}\mathcal{F}_d a)\mathcal{F}_d^{-1}\big]^* \\ &= \mathcal{F}_d\operatorname{diag}(\sqrt{d}\mathcal{F}_d a)\operatorname{diag}(\overline{\sqrt{d}\mathcal{F}_d a})\mathcal{F}_d^{-1} \\ &= \mathcal{F}_d\operatorname{diag}\big(d|\mathcal{F}_d a|^2\big)\mathcal{F}_d^{-1}. \end{aligned}$$

Hence, the singular values of M_a are $\{\sqrt{d}|(\mathcal{F}_d a)(\xi)|\}_{\xi=0}^{d-1}$.

The action of an arbitrary projection onto a vector of independent real Gaussian variables is very well known. It may be described as follows.

Lemma 2.1. *Let $a=(a_0,\dots,a_{d-1})$ be independent real Gaussian variables. Let $k\leqslant d$ be a natural number and let $x^1,\dots,x^k$ be mutually orthogonal unit vectors in $\mathbb{R}^d$. Then*

$$\big\{\langle a,x^j\rangle\big\}_{j=1}^k$$

is equidistributed with a k-dimensional vector of independent real Gaussian variables.

A direct calculation shows, that Lemma 2.1 holds also for complex vectors a and $x^1,\dots,x^k$. We present the following formulation of this fact.

Lemma 2.2. *Let $a=(a_0,\dots,a_{d-1})$ be independent complex Gaussian variables. Let W be a $k\times d$ matrix with k orthonormal rows. Then Wa is equidistributed with a k-dimensional vector of independent complex Gaussian variables.*

3. Proof of Theorem 1.3

We shall need the following statement, which describes the preconditioning role of the diagonal matrix $D_\varkappa$. A similar fact has been used also in [2]. Nevertheless, using discrete Fourier transform instead of a Hadamard matrix does not pose any restrictions on the underlying dimension d. Without repeating the details, we point out, that we discussed briefly in [7, Remark 2.5], why this preconditioning may not be omitted.

Lemma 3.1. *Let $n \geqslant d$ be natural numbers and let $x^1, \dots, x^n \in \mathbb{C}^d$ be complex vectors. Let $\varkappa = (\varkappa_0, \dots, \varkappa_{d-1})$ be independent Bernoulli variables. Then there is an absolute constant $C > 0$, such that with probability at least* $5/6$,

$$\left\|\mathcal{F}_d D_\varkappa\left(x^j\right)\right\|_\infty \leqslant \frac{C\sqrt{\log n}}{\sqrt{d}} \cdot \left\|x^j\right\|_2 \tag{3.1}$$

holds for all $j \in \{1, \dots, n\}$.

Proof. Let $x = \alpha + i\beta$ be a unit complex vector in $\mathbb{C}^d$. We put $y = (y_0, \dots, y_{d-1}) = \mathcal{F}_d D_\varkappa(x)$. Combining the inclusion

$$\bigl\{z \in \mathbb{C}\colon |z| > s\bigr\} = \bigl\{z \in \mathbb{C}\colon (\Re z)^2 + (\Im z)^2 > s^2\bigr\} \subset \left\{z \in \mathbb{C}\colon |\Re z| > \frac{s}{\sqrt{2}}\right\}$$
$$\cup \left\{z \in \mathbb{C}\colon |\Im z| > \frac{s}{\sqrt{2}}\right\}$$

with

$$\mathbb{P}_\varkappa\left(|\Re y_l| > \frac{s}{\sqrt{2}}\right) = 2\mathbb{P}_\varkappa\left(\Re y_l > \frac{s}{\sqrt{2}}\right),$$

we may estimate

$$\mathbb{P}_\varkappa\bigl(|y_l| > s\bigr) \leqslant 2\mathbb{P}_\varkappa\left(\Re y_l > \frac{s}{\sqrt{2}}\right) + 2\mathbb{P}_\varkappa\left(\Im y_l > \frac{s}{\sqrt{2}}\right), \quad l = 0, \dots, d-1, \tag{3.2}$$

where

$$\Re y_l = \frac{1}{\sqrt{d}} \sum_{u=0}^{d-1} \varkappa_u\bigl[\alpha_u \cos(2\pi l u/d) + \beta_u \sin(2\pi l u/d)\bigr]$$

and

$$\Im y_l = \frac{1}{\sqrt{d}} \sum_{u=0}^{d-1} \varkappa_u\bigl[\beta_u \cos(2\pi l u/d) - \alpha_u \sin(2\pi l u/d)\bigr]$$

are the real and the imaginary part of y_l, respectively.

Let $t > 0$ be a real parameter to be chosen later. Using Markov's inequality we may proceed in a standard way:

$$\begin{aligned}
\mathbb{P}_\varkappa\left(\Re y_l > \frac{s}{\sqrt{2}}\right) &= \mathbb{P}_\varkappa\left(\exp\left(t\Re y_l - \frac{st}{\sqrt{2}}\right) > 1\right)\\
&\leqslant \exp\left(-\frac{st}{\sqrt{2}}\right)\mathbb{E}_\varkappa \exp(t\Re y_l)\\
&= \exp\left(-\frac{st}{\sqrt{2}}\right)\prod_{u=0}^{d-1}\cosh\left[\frac{t}{\sqrt{d}}\left[\alpha_u\cos(2\pi lu/d) + \beta_u\sin(2\pi lu/d)\right]\right]\\
&\leqslant \exp\left(-\frac{st}{\sqrt{2}}\right)\prod_{u=0}^{d-1}\exp\left(\frac{t^2}{2d}\left[\alpha_u\cos(2\pi lu/d) + \beta_u\sin(2\pi lu/d)\right]^2\right)\\
&\leqslant \exp\left(-\frac{st}{\sqrt{2}}\right)\prod_{u=0}^{d-1}\exp\left(\frac{t^2}{2d}\left[\alpha_u^2 + \beta_u^2\right]\right) = \exp\left(-\frac{st}{\sqrt{2}} + \frac{t^2}{2d}\right).
\end{aligned}$$

We have used the inequality $\cosh(v) \leqslant \exp(v^2/2)$, which holds for all $v \in \mathbb{R}$, and the inequality between geometric and quadratic means. For the optimal $t = \frac{sd}{\sqrt{2}}$, this is equal to $\exp(-\frac{s^2 d}{4})$.

As the second summand in (3.2) may be estimated in the same way, we obtain

$$\mathbb{P}_\varkappa\big(|y_l| > s\big) \leqslant 4\exp\left(-\frac{s^2 d}{4}\right), \quad l = 0, \ldots, d-1. \tag{3.3}$$

Choosing $s = \Omega(d^{-1/2}\sqrt{\log n})$ and applying the union bound over all $nd \leqslant n^2$ components of $\{\mathcal{F}_d D_\varkappa(x^j/\|x^j\|_2)\}_{j=1}^n$, we obtain the result. □

Proof of Theorem 1.3. Let us choose a vector $\varkappa = (\varkappa_0, \ldots, \varkappa_{d-1}) \in \{-1, +1\}^d$, such that (3.1) holds. According to Lemma 3.1 this happens with probability at least $5/6$.

Let us take $\tilde{x} = \frac{x^j}{\|x^j\|_2}$ for any fixed $j = 1, \ldots, n$. We show, that there is an absolute constant $c > 0$, such that

$$\mathbb{P}_a\big(\|M_{a,k}D_\varkappa\tilde{x}\|_2^2 \geqslant 2(1+\varepsilon)k\big) \leqslant \exp\left(-\frac{ck\varepsilon^2}{\log n}\right) \tag{3.4}$$

and

$$\mathbb{P}_a\big(\|M_{a,k}D_\varkappa\tilde{x}\|_2^2 \leqslant 2(1-\varepsilon)k\big) \leqslant \exp\left(-\frac{ck\varepsilon^2}{\log n}\right) \tag{3.5}$$

hold. From (3.4) and (3.5), Theorem 1.3 follows again by a union bound over all $j = 1, \ldots, n$.

Let $y^j = S^j(D_\varkappa\tilde{x}) \in \mathbb{C}^d$, $j = 0, \ldots, k-1$, where S is the shift operator defined by

$$S : \mathbb{C}^d \to \mathbb{C}^d, \qquad S(z_0, \ldots, z_{d-1}) = (z_1, \ldots, z_{d-1}, z_0).$$

We denote by Y the $k \times d$ matrix with rows $y^0, \ldots, y^{k-1}$.

Then it holds

$$\|M_{a,k}D_\varkappa\tilde{x}\|_2^2=\sum_{j=0}^{k-1}\left|\sum_{u=0}^{d-1}a_{(u-j)\bmod d}\varkappa_u\tilde{x}_u\right|^2=\sum_{j=0}^{k-1}\left|\sum_{u=0}^{d-1}y_u^j a_u\right|^2=\|Ya\|_2^2.$$

Let $Y=U\Sigma V^*$ be the singular value decomposition of Y. As mentioned above, $b:=V^*a$ is a k-dimensional vector of independent complex Gaussian variables. Hence,

$$\begin{aligned}\mathbb{P}_a\big(\|Ya\|_2^2>\tau\big)&=\mathbb{P}_a\big(\big\|U\Sigma V^*a\big\|_2^2>\tau\big)=\mathbb{P}_b\big(\|U\Sigma b\|_2^2>\tau\big)\\&=\mathbb{P}_b\big(\|\Sigma b\|_2^2>\tau\big)=\mathbb{P}_b\left(\sum_{j=0}^{k-1}\lambda_j^2|b_j|^2>\tau\right),\end{aligned}$$

holds for every $\tau>0$. Here, λ_j, $j=0,\dots,k-1$, are the singular values of Y. Let us denote $\mu_j=\lambda_j^2$. Then

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

where $\|Y\|_F$ is the Frobenius norm of Y.

Moreover,

$$\begin{aligned}\|\mu\|_\infty=\|\lambda\|_\infty^2&=\sup_{z\in\mathbb{C}^d,\|z\|_2\leqslant 1}\|Yz\|_2^2\\&\leqslant\sup_{z\in\mathbb{C}^d,\|z\|_2\leqslant 1}\|M_{D_\varkappa\tilde{x}}z\|_2^2=d\big\|\mathcal{F}_dD_\varkappa(\tilde{x})\big\|_\infty^2\leqslant C^2\log n,\end{aligned}\tag{3.6}$$

where $M_{D_\varkappa\tilde{x}}$ stands for the $d\times d$ complex circulant matrix with the first row equal to $D_\varkappa\tilde{x}$.

This leads finally also to

$$\|\mu\|_2\leqslant\sqrt{\|\mu\|_1\cdot\|\mu\|_\infty}\leqslant C\sqrt{k\log n}.\tag{3.7}$$

Then

$$\mathbb{P}_a\big(\|Ya\|_2^2>2(1+\varepsilon)k\big)=\mathbb{P}_b\left(\sum_{j=0}^{k-1}\mu_j\big(|b_j|^2-2\big)>2\varepsilon k\right).$$

We denote

$$Z:=\sum_{j=0}^{k-1}\mu_j\big(|b_j|^2-2\big).$$

The complex version of Lemma 1 from Section 4.1 of [11] (cf. also Lemma 2.2 of [12]) states that

$$\mathbb{P}_b\big(Z \geqslant 2\sqrt{2}\|\mu\|_2\sqrt{t} + 2\|\mu\|_\infty t\big) \leqslant \exp(-t). \tag{3.8}$$

Using (3.6) and (3.7), we arrive at

$$\mathbb{P}_b\big(Z \geqslant 2\sqrt{2}C\sqrt{tk\log n} + 2C^2 t\log n\big) \leqslant \exp(-t).$$

Choosing $t = \frac{c'k\varepsilon^2}{C^2\log n}$ for $c' > 0$ small enough, we get

$$\mathbb{P}_b(Z \geqslant 2\varepsilon k) \leqslant \exp\left(-\frac{ck\varepsilon^2}{\log n}\right).$$

This finishes the proof of (3.4). Let us note, that (3.5) follows in the same manner with (3.8) replaced by

$$\mathbb{P}_b\big(Z \leqslant -2\sqrt{2}\|\mu\|_2\sqrt{t}\big) \leqslant \exp(-t),$$

which may be again found in Lemma 1, Section 4.1 of [11]. □

Remark 3.2. The statement and the proof of Theorem 1.3 do not change, if we replace the partial circulant matrix $M_{a,k}$ with any $k \times d$ submatrix of M_a.

Note added in proof

Interesting new work of Ailon and Liberty [5] appeared during the review process of this paper. Their transformation is the composition of a random sign matrix with a random selection of a suitable number k of rows from a Fourier matrix. Their bound on k, namely $k = \Omega(\varepsilon^{-4}\log n \cdot \operatorname{polylog} d)$, is optimal up to the $\operatorname{polylog} d$ factor. Depending on d and n, this may be better than our bound.

In another very recent preprint [10], Krahmer and Ward applied the RIP bounds of [13] to prove that partial circulant matrices satisfy the Johnson–Lindenstrauss lemma if

$$k = \Omega\big(\max\big(\varepsilon^{-1}\log^{3/2} n \cdot \log^{3/2} d, \varepsilon^{-2}\log n \cdot \log^4 d\big)\big).$$

Acknowledgments

The author would like to thank Aicke Hinrichs for valuable comments and the anonymous referee for the careful reading of the manuscript. The author also acknowledges the financial support provided by the FWF project Y 432-N15 START-Preis “Sparse Approximation and Optimization in High Dimensions”.

References

[1] D. Achlioptas, Database-friendly random projections: Johnson–Lindenstrauss with binary coins, J. Comput. System Sci. 66 (4) (2003) 671–687.

[2] N. Ailon, B. Chazelle, Approximate nearest neighbors and the fast Johnson–Lindenstrauss transform, in: Proc. 38th Annual ACM Symposium on Theory of Computing, 2006.

[3] N. Ailon, B. Chazelle, The fast Johnson–Lindenstrauss transform and approximate nearest neighbors, SIAM J. Comput. 39 (1) (2009) 302–322.

[4] N. Ailon, E. Liberty, Fast dimension reduction using Rademacher series on dual BCH codes, Discrete Comput. Geom. 42 (4) (2009) 615–630.

[5] N. Ailon, E. Liberty, Almost optimal unrestricted fast Johnson–Lindenstrauss transform, http://arxiv.org/abs/1005.5513.

[6] S. Dasgupta, A. Gupta, An elementary proof of a theorem of Johnson and Lindenstrauss, Random Structures Algorithms 22 (2003) 60–65.

[7] A. Hinrichs, J. Vybíral, Johnson–Lindenstrauss lemma for circulant matrices, Random Structures Algorithms, in press.

[8] P. Indyk, R. Motwani, Approximate nearest neighbors: Towards removing the curse of dimensionality, in: Proc. 30th Annual ACM Symposium on Theory of Computing, 1998, pp. 604–613.

[9] W.B. Johnson, J. Lindenstrauss, Extensions of Lipschitz Mappings into a Hilbert Space, Contemp. Math., vol. 26, 1984, pp. 189–206.

[10] F. Krahmer, R. Ward, New and improved Johnson–Lindenstrauss embeddings via the Restricted Isometry Property, http://arxiv.org/abs/1009.0744.

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

[12] J. Matoušek, On variants of the Johnson–Lindenstrauss lemma, Random Structures Algorithms 33 (2) (2008) 142–156.

[13] H. Rauhut, J. Romberg, J. Tropp, Restricted isometries for partial random circulant matrices, http://arxiv.org/abs/1010.1847.