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
email: jan.vybiral@oeaw.ac.at

February 15, 2010

Abstract

We continue our study of the Johnson-Lindenstrauss lemma and its connection to circulant matrices started in [5]. We reduce the bound on k from $k = O(\varepsilon^{-2}\log^3 n)$ proven there to $k = O(\varepsilon^{-2}\log^2 n)$. Our technique differs essentially from the one used in [5]. We employ the discrete Fourier transform and singular value decomposition to deal with the dependency caused by the circulant structure.

AMS Classification: 52C99, 68Q01

Keywords and phrases: 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 $k = O(\varepsilon^{-2}\log n)$, there exists a linear map $f:\mathbb{R}^d \to \mathbb{R}^k$, such that

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

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

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 $k = O(\varepsilon^{-2}\log n)$ means, that there is an absolute constant $C > 0$, such that the statement holds for all natural numbers k with $k \ge C\varepsilon^{-2}\log n$. We shall also always assume, that $k \le d$. Otherwise, the statement becomes trivial.

The original proof of this fact was given by Johnson and Lindenstrauss in [7]. We refer to [4] 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 [6, 1, 2, 3, 9] for details and to [9] 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

$$\mathbb{P}\Big(f \in \mathcal{P} : ||f(x)||_2^2 \geq 1 + \varepsilon\Big) < 1 - \frac{1}{2n}$$

and

$$\mathbb{P}\Big(f \in \mathcal{P} : ||f(x)||_2^2 \leq 1 - \varepsilon\Big) < 1 - \frac{1}{2n},$$

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 best known construction of f (according to the properties mentioned above) was given by Ailon and Chazelle in [2] with an improvement due to Matoušek, cf. [9]. It states, that f may be given as a composition of a sparse matrix, certain random Fourier matrix and a random diagonal matrix. Although it provides a good computational time of $f(x)$ (with high probability $f(x)$ may be computed using $O(d \log d + \min\{d\varepsilon^{-2} \log n, \varepsilon^{-2} \log^3 n\})$ operations), it still needs, that each coordinate of the $k \times d$ matrix is generated independently. In [5], 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 less randomness (only $2d$ compared to kd or $(k+1)d$ used earlier) 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 all the other constructions available in the literature so far is that the components of $f(x)$ are now no longer independent random variables. Decoupling this dependence, we were able to prove in [5] the Johnson-Lindenstrauss lemma for composition of a random circulant matrix and a random diagonal matrix, but only for $k = O(\varepsilon^{-2} \log^3 n)$. It is the main aim of this note to improve this bound to $k = O(\varepsilon^{-2} \log^2 n)$. This comes essentially closer to the standard bound $k = O(\varepsilon^{-2} \log n)$. 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.

To state our main result, we first fix some notation. Let

- $\varepsilon \in (0, \frac{1}{2})$,
- $n \geq d$ be natural numbers,
- $x^1, \dots, x^n \in \mathbb{C}^d$ be n arbitrary points in $\mathbb{C}^d$,
- $k = O(\varepsilon^{-2} \log^2 n)$ be a natural number smaller then d,
- $a = (a_0, \dots, a_{d-1})$ be independent complex Gaussian variables, cf. Definition 2.1,
- $\varkappa = (\varkappa_0, \dots, \varkappa_{d-1})$ be independent Bernoulli variables.

We denote by $M_{a,k}$ and $D_\varkappa$ the partial random circulant matrix and the random diagonal matrix, respectively, cf. Definition 2.2 for details.

Theorem 1.1. *The mapping $f : \mathbb{C}^d \to \mathbb{C}^k$ given by $f(x) = \frac{1}{\sqrt{2k}} M_{a,k} D_\varkappa x$ satisfies*

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

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 1.1, which deals with real Euclidean spaces.

Corollary 1.2. *Let $\varepsilon \in (0, \frac{1}{2})$, $n \ge d$ 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 $k = O(\varepsilon^{-2} \log^2 n)$ 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

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

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 1.1 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

Let us give a brief overview of techniques used in the proof of Theorem 1.1. We shall list only those few properties needed in the sequel.

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

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

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

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

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

2.2 Circulant matrices

Definition 2.1. 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

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

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

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

If $k = d$, we denote by $M_a = M_{a,d}$ the full circulant matrix. This notation extends naturally to the case, when $a = (a_0, \dots, a_{d-1})$ are independent complex Gaussian variables.
(ii) If $\varkappa = (\varkappa_0, \dots, \varkappa_{d-1})$ are independent Bernoulli variables, we put

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

Of course, $D_\varkappa : \mathbb{C}^d \to \mathbb{C}^d$ is also an isomorphism.
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 diagonalised with the use of a discrete Fourier transform, its inverse and a multiple of the discrete Fourier transform of its first row.

2.3 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 $k \le d$. 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

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

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

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

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

A direct calculation shows, that Lemma 2.3 holds also for complex vectors a and $x^1, \dots, x^k$. We present the following formulation of this fact.

Lemma 2.4. *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 1.1

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 [5, Remark 2.5], why this preconditioning may not be omitted.

Lemma 3.1. *Let $n \ge d$ 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 $5/6$*

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

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 $y = (y_0, \dots, y_{d-1}) = \mathcal{F}_d\, D_\varkappa(x)$. Then we may estimate

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

where

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

and

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

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

Using the Markov's inequality and a real parameter $t > 0$, which is at our disposal, we may proceed in a standard way:

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

We have used the inequality $\cosh(v) \le \exp(v^2/2)$, 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 $\exp(-\frac{s^2d}{4})$.

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

$$\mathbb{P}_\varkappa(|y_l| > s) \le 4\exp\Big(-\frac{s^2d}{4}\Big), \quad l = 0,\dots,d-1. \tag{3.3}$$

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

Proof of Theorem 1.1

Let us choose a vector $\varkappa = (\varkappa_0,\dots,\varkappa_{d-1}) \in \{-1,+1\}^d$, such that (3.1) holds. According to the 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 $j = 1,\dots,n$. We show, that there is an absolute constant $c > 0$, such that

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

and

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

holds. From (3.4) and (3.5), Theorem 1.1 follows again by a union bound over all $j = 1,\dots,n$.

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

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

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

Then it holds

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

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,

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

where $\lambda_j, j = 0, \dots, k-1$, 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,

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

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

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

Then

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

We denote

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

The complex version of Lemma 1 from Section 4.1 of [8] (cf. also Lemma 2.2 of [9]) states that

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

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

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

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

$$\mathbb{P}_b(Z \ge 2\varepsilon k) \le \exp\Big(-\frac{ck\varepsilon^2}{\log n}\Big).$$

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

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

which may be again found in Lemma 1, Section 4.1 of [8].

Remark 3.2. The statement and the proof of Theorem 1.1 do not change, if we replace the partial circulant matrix $M_{a,k}$ with any $k \times d$ submatrix of M_a.

Acknowledgement: The author would like to thank Aicke Hinrichs for valuable comments. 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. Syst. Sci.*, 66(4):671-687, 2003.

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

[4] S. Dasgupta and A. Gupta, An elementary proof of a theorem of Johnson and Lindenstrauss. *Random. Struct. Algorithms*, 22:60-65, 2003.

[5] A. Hinrichs and J. Vybíral, Johnson-Lindenstrauss lemma for circulant matrices, submitted, available on `http://arxiv.org/abs/1001.4919`.

[6] P. Indyk and R. Motwani, Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proc. 30th Annual ACM Symposium on Theory of Computing*, pp. 604-613, 1998.

[7] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space. *Contem. Math.*, 26:189-206, 1984.

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

[9] J. Matoušek, On variants of the Johnson-Lindenstrauss lemma, *Random Struct. Algorithms* 33(2):142–156, 2008.