Fast Random Projections using Lean Walsh Transforms
Yale University Technical report #1390

Edo Liberty* Nir Ailon† Amit Singer‡

Abstract

We present a $k \times d$ random projection matrix that is applicable to vectors $x \in \mathbb{R}^d$ in $O(d)$ operations if $d \geq k^{2+\delta'}$. Here, k is the minimal Johnson Lindenstrauss dimension and δ' is arbitrarily small. The projection succeeds, with probability $1 - 1/n$, in preserving vector lengths, up to distortion ε, for all vectors such that $\|x\|_\infty \leq \|x\|_2 k^{-1/2} d^{-\delta}$ (for arbitrary small δ). Sampling based approaches are either not applicable in linear time or require a bound on $\|x\|_\infty$ that is strongly dependant on d. Our method overcomes these shortcomings by rapidly applying dense tensor power matrices to incoming vectors.

1 Introduction

The application of various random matrices has become a common method for accelerating algorithms both in theory and in practice. These procedures are commonly referred to as *random projections*. The critical property of a $k \times d$ random projection matrix, Φ, is that the mapping $x \mapsto \Phi x$ not only reduces the dimension from d to k, but also preserves lengths, up to distortion ε, with probability at least $1-1/n$ for some small ε and large n. The name *random projections* was coined after the first construction of Johnson and Lindenstrauss [1] in 1984, who showed that such mappings exist for $k \geq O(\log(n)/\varepsilon^2)$. Other constructions of random projection matrices have been discovered since [2, 3, 4, 5, 6]. Their properties make random projections a key player in rank-k approximation algorithms [7, 8, 9, 10, 11, 12, 13, 14], other algorithms in numerical linear algebra [15, 16, 17], compressed sensing [18], and various other applications, e.g, [19, 20].

Considering the usefulness of random projections it is natural to ask the following question: what should be the structure of a random projection matrix, Φ, such that mapping $x \mapsto \Phi x$ would require the least amount of computation? A naïve construction of a $k \times d$ unstructured matrix Φ would result in an $O(dk)$ application cost. This is prohibitive even for moderate values of k and d.

*Yale University, Department of Computer Science, Supported by NGA and AFOSR.
†Google Research
‡Yale University, Department of Mathematics, Program in Applied Mathematics.
[0]EL and AS thank the Institute for Pure and Applied Mathematics (IPAM) and its director Mark Green for their warm hospitality during the fall semester of 2007.

In [21], Ailon and Chazelle proposed the first Fast Johnson Lindenstrauss Transform. Their matrix is a composition of a sparse sampling matrix and a discrete Fourier matrix. This achieves a running time of $O(d\log(d) + k^3)$. Recently, Ailon and Liberty [22] further improved this to $O(d\log(k))$[1] by composing a deterministic code matrix and a randomized block diagonal matrix. The idea behind both fast constructions is similar: they start with applying a randomized isometric $d \times d$ matrix Ψ, which maps all vectors in R^d (w.h.p) into a set $\chi \subset \mathbb{R}^d$, and then use a $k \times d$ matrix A to project all vectors from χ to $\mathbb{R}^k$. There seems to be a tradeoff between the possible computational efficiency of applying A and the size of χ: the smaller χ is, the faster A can be applied. This, however, might require a time costly preprocessing application of Ψ.

In the present work we examine the connection between A and χ for any matrix A (Section 2). We propose in Section 3 a new type of fast applicable matrices and in Section 4 explore their χ. These matrices are constructed using tensor products and can be applied to any vector in $\mathbb{R}^d$ in linear time, i.e, in $O(d)$. Due to the similarity in their construction to Walsh-Hadamard matrices and their rectangular shape we term them *Lean Walsh Matrices*[2].

	The rectangular $k \times d$ matrix A	Application time	$x \in \chi$ if $\|x\|_2 = 1$ and:
Johnson, Lindenstrauss [1]	k rows of a random unitary matrix	$O(kd)$	
Various Authors [2, 4, 5, 6]	i.i.d random entries	$O(dk)$	
Ailon, Chazelle [21]	Sparse Gaussian entrees	$O(k^3)$	$\|x\|_\infty \leq O((d/k)^{-1/2})$
Ailon, Liberty [22]	4-wise independent Code matrix	$O(d\log k)$	$\|x\|_4 \leq O(d^{-1/4})$
This work	Any deterministic matrix	?	$\|x\|_A \leq O(k^{-1/2})$
This work	Lean Walsh Transform	$O(d)$	$\|x\|_\infty \leq O(k^{-1/2}d^{-\delta})$

Table 1: Types of $k \times d$ matrices and the subsets χ of $\mathbb{R}^d$ for which they constitute a random projection. The norm $\|\cdot\|_A$ is defined below.

Due to their construction the Lean Walsh matrices are of size $\tilde{d} \times d$ where $\tilde{d} = d^\alpha$ for some $0 < \alpha < 1$. In order to reduce the dimension to $k \leq \tilde{d}$, $k = O(\log(n)/\varepsilon^2))$, we compose the lean Walsh matrix, A, with a known Johnson Lindenstrauss matrix construction R. Applying R in $O(d)$ requires some relation between d, k and α as explained in subsection 4.1.

2 Norm concentration and $\chi(A, \varepsilon, n)$

We compose an arbitrary deterministic $\tilde{d} \times d$ matrix A and random sign diagonal matrix D_s and study the behavior of such matrices as random projections. In order for AD_s to exhibit the property of a random

[1] Their method applies to cases for which $k \leq d^{1/2-\delta}$ for some arbitrary small δ.

[2] The terms *Lean Walsh Transform* or simply *Lean Walsh* are also used interchangeably.

projection it is enough for it to preserve the length of any single *unit* vector $x \in R^d$ with very high probability:

$$\Pr\left[|\ \|AD_s x\|_2^2 - 1\ | \geq \varepsilon) \right] < 1/n \tag{1}$$

Here D_s is a diagonal matrix such that $D_s(i,i)$ are random signs (i.i.d ± 1 w.p $1/2$ each) and n is chosen according to a desired success probability, usually polynomial in the intended number of projected vectors.

Note that we can replace the term $AD_s x$ with $AD_x s$ where D_x is a diagonal matrix holding on the diagonal the values of x, i.e $D_x(i,i) = x(i)$ and similarly $s(i) = D_s(i,i)$. Denoting $M = AD_x$, we view the term $\|Ms\|_2$ as a function over the product space $\{1,-1\}^d$ from which the variable s is uniformly chosen. This function is convex over $[-1,1]^d$ and Lipschitz bounded. In his book, Talagrand [23] describes a strong concentration result for such functions.

Lemma 2.1 (Talagrand [23])**.** *Given a matrix M and a random vector s ($s(i)$ are i.i.d ± 1 w.p $1/2$) define the random variable $Y = \|Ms\|_2$. Denote by μ the median of Y, and by $\sigma = \|M\|_{2\to 2}$ the spectral norm of M. Then*

$$\Pr\left[\, |Y - \mu| > t \right] \leq 4e^{-t^2/8\sigma^2} \tag{2}$$

Lemma 2.1 asserts that $\|AD_x s\|$ is distributed like a (sub) Gaussian around its median, with standard deviation 2σ.

First, in order to have $E[Y^2] = 1$ it is necessary and sufficient for the columns of A to be normalized to 1 (or normalized in expectancy). To estimate the median, μ, we substitute $t^2 \to t'$ and compute:

$$\begin{aligned} E[(Y-\mu)^2] &= \int_0^\infty \Pr[(Y-\mu)^2] > t']dt' \\ &\leq \int_0^\infty 4e^{-t'/(8\sigma^2)} dt' = 32\sigma^2 \end{aligned}$$

Furthermore, $(E[Y])^2 \leq E[Y^2] = 1$, and so $E[(Y-\mu)^2] = E[Y^2] - 2\mu E[Y] + \mu^2 \geq 1 - 2\mu + \mu^2 = (1-\mu)^2$. Combining, $|1-\mu| \leq \sqrt{32}\sigma$. We set $\varepsilon = t + |1-\mu|$:

$$\Pr[|Y-1| > \varepsilon] \leq 4e^{-\varepsilon^2/32\sigma^2} \quad \text{,for } \ \varepsilon > 2|1-\mu| \tag{3}$$

If we set $k = 33\log(n)/\varepsilon^2$ (assuming $\log(n)$ is larger than some constant) the requirement of equation 1 is met for $\sigma \leq k^{-1/2}$. Moreover $\varepsilon > 2|1-\mu|$. We see that a condition on $\sigma = \|AD_x\|_{2\to 2}$ is sufficient for the projection to succeed w.h.p. This naturally defines χ.

Definition 2.1. *For a given matrix A we define the vector pseudonorm of x with respect to A as $\|x\|_A \equiv \|AD_x\|_{2\to 2}$ where D_x is a diagonal matrix such that $D_x(i,i) = x(i)$.*

Definition 2.2. *We define $\chi(A,\varepsilon,n)$ as the intersection of the Euclidian unit sphere and a ball of radius $k^{-1/2}$ in the norm $\|\cdot\|_A$*

$$\chi(A,\varepsilon,n) = \left\{ x \in \mathbb{R}^d \mid \|x\|_2 = 1, \|x\|_A \leq k^{-1/2} \right\} \tag{4}$$

for $k = 33\log(n)/\varepsilon^2$.

Lemma 2.2. *For any column normalized matrix, A, and an i.i.d random ± 1 diagonal matrix, D_s, the following holds:*

$$\forall x \in \chi(A, \varepsilon, n) \quad \Pr\left[\left| \|AD_s x\|_2^2 - 1 \right| \geq \varepsilon \right] \leq 1/n \tag{5}$$

Proof. For any $x \in \chi$, by definition 2.2, $\|x\|_A = \|AD_x\|_{2\to 2} = \sigma \leq k^{-1/2}$. The lemma follows from substituting the value of σ into equation 3. □

It is convenient to think about χ as the "good" set of vectors for which AD_s is length preserving with high probability. En route to explore $\chi(A, \varepsilon, n)$ for lean Walsh matrices we first turn to formally defining them.

3 Lean Walsh transforms

The *Lean* Walsh Transform, similar to the Walsh Transform, is a recursive tensor product matrix. It is initialized by a constant seed matrix, A_1, and constructed recursively by using Kronecker products $A_{\ell'} = A_1 \otimes A_{\ell'-1}$. The main difference is that the Lean Walsh seeds have fewer rows than columns. We formally define them as follows:

Definition 3.1. *A_1 is a Lean Walsh seed (or simply 'seed') if i) A_1 is a rectangular matrix $A_1 \in \mathbb{C}^{r\times c}$, such that $r < c$; ii) A_1 is absolute valued $1/\sqrt{r}$ entree-wise, i.e, $|A_1(i,j)| = r^{-1/2}$; iii) the rows of A_1 are orthogonal; and iv) all inner products between its different columns are equal in absolute value to a constant $\rho \leq 1/\sqrt{(c-1)}$. ρ is called the Coherence of A_1.*

Definition 3.2. *A_ℓ is a Lean Walsh transform, of order ℓ, if for all $\ell' \leq \ell$ we have $A'_\ell = A_1 \otimes A_{\ell'-1}$, where $\otimes$ stands for the Kronecker product and A_1 is a seed according to definition 3.1.*

The following are examples of seed matrices:

$$\begin{array}{cc} A'_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} & A''_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \end{pmatrix} \\ r' = 3,\ c' = 4,\ \rho' = 1/3 & r'' = 2,\ c'' = 3,\ \rho'' = 1/2 \end{array} \tag{6}$$

These examples are a part of a large family of possible seeds. This family includes, amongst other constructions, sub-Hadamard matrices (like A'_1) or sub-Fourier matrices (like A''_1). A simple construction is given for possible larger seeds.

Fact 3.1. *Let F be the $c \times c$ Discrete Fourier matrix such that $F(i,j) = e^{2\pi\sqrt{-1}ij/c}$. Define A_1 to be the matrix consisting of the first $r = c-1$ rows of F normalized by $1/\sqrt{r}$. A_1 is a lean Walsh seed with coherence $1/r$.*

Proof. The facts that $|A_1(i,j)| = 1/\sqrt{r}$ and that the rows of A_1 are orthogonal are trivial. Moreover, due to the orthogonality of the columns of F, the inner product of two different columns of A_1 must equal $\rho = 1/r$ in absolute value.

$$\left|\langle A_1^{(j_1)}, A_1^{(j_2)}\rangle\right| = \frac{1}{r}\left|\sum_i^r \bar{F}(i,j_1)F(i,j_2)\right| = \frac{1}{r}\left|-\bar{F}(c,j_1)F(c,j_2)\right| = \frac{1}{r} \tag{7}$$

here $\bar{F}(\cdot,\cdot)$ stands for the complex conjugate of $F(\cdot,\cdot)$. □

We use elementary properties of Kronecker products to characterize A_ℓ in terms of the number of rows, r, the number of columns, c, and the coherence, ρ, of A_1. The following facts hold true for A_ℓ:

Fact 3.2. *i) A_ℓ is of size*[3] *$d^\alpha \times d$, where $\alpha = \log(r)/\log(c) < 1$ is the skewness of A_1 ii) for all i and j, $A_\ell(i,j) \in \pm\tilde{d}^{-1/2}$ which means that A_ℓ is column normalized; and iii) the rows of A_ℓ are orthogonal.*

Fact 3.3. *The time complexity of applying A_ℓ to any vector $z \in \mathbb{R}^d$ is $O(d)$.*

Proof. Let $z = [z_1; \ldots; z_c]$ where z_i are sections of length d/c of the vector z. Using the recursive decomposition for A_ℓ we compute $A_\ell z$ by first summing over the different z_i according to the values of A_1 and applying to each sum the matrix $A_{\ell-1}$. Denoting by $T(d)$ the time to apply A_ℓ to $z \in \mathbb{R}^d$ we get that $T(d) = rT(d/c) + rd$. Due to the Master Theorem, and the fact that $r < c$ we have that $T(d) = O(d)$. More precisely, $T(d) \le dcr/(c-r)$. □

For clarity, we demonstrate Fact 3.3 for A'_1 (equation 6):

$$A'_\ell z = A'_\ell \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \frac{1}{\sqrt{3}} \begin{pmatrix} A'_{\ell-1}(z_1 + z_2 - z_3 - z_4) \\ A'_{\ell-1}(z_1 - z_2 + z_3 - z_4) \\ A'_{\ell-1}(z_1 - z_2 - z_3 + z_4) \end{pmatrix} \tag{8}$$

In what follows we characterize $\chi(A, \varepsilon, n)$ for a general Lean Walsh transform by the parameters of its seed, r, c and ρ. The omitted notation, A, stands for A_ℓ of the right size to be applied to x, i.e, $\ell = \log(d)/\log(c)$. Moreover, we freely use α to denote the skewness $\log(r)/\log(c)$ of the seed at hand.

[3]The size of A_ℓ is $r^\ell \times c^\ell$. Since the running time is linear, we can always pad vectors to be of length c^ℓ without effecting the asymptotic running time. From this point on we assume w.l.o.g $d = c^\ell$ for some integer ℓ

4 An ℓ_p bound on $\|\cdot\|_A$

After describing the lean Walsh transforms we turn our attention to exploring their "good" sets χ .We remind the reader that $\|x\|_A \leq k^{-1/2}$ entails $x \in \chi$:

$$\|x\|_A^2 = \|AD_x\|_{2\to 2}^2 = \max_{y,\|y\|_2=1} \|y^T AD_x\|_2^2 \tag{9}$$

$$= \max_{y,\|y\|_2=1} \sum_{i=1}^{d} x^2(i)(y^T A^{(i)})^2 \tag{10}$$

$$\leq \left(\sum_{i=1}^{d} x^{2p}(i)\right)^{1/p} \left(\max_{y,\|y\|_2=1} \sum_{i=1}^{d} (y^T A^{(i)})^{2q}\right)^{1/q} \tag{11}$$

$$= \|x\|_{2p}^2 \|A^T\|_{2\to 2q}^2 \tag{12}$$

The transition from the second to the third line follows from Hölder's inequality for dual norms p and q, satisfying $1/p + 1/q = 1$. We are now faced with the computing $\|A^T\|_{2\to 2q}$ in order to obtain the constraint on $\|x\|_{2p}$.

Theorem 4.1. [Riesz-Thorin] *For an arbitrary matrix B, assume $\|B\|_{p_1\to r_1} \leq C_1$ and $\|B\|_{p_2\to r_2} \leq C_2$ for some norm indices p_1, r_1, p_2, r_2 such that $p_1 \leq r_1$ and $p_2 \leq r_2$. Let λ be a real number in the interval $[0, 1]$, and let p, r be such that $1/p = \lambda(1/p_1) + (1 - \lambda)(1/p_2)$ and $1/r = \lambda(1/r_1) + (1 - \lambda)(1/r_2)$. Then $\|B\|_{p\to r} \leq C_1^{\lambda} C_2^{1-\lambda}$.*

In order to use the theorem, let us compute $\|A^T\|_{2\to 2}$ and $\|A^T\|_{2\to\infty}$. From $\|A^T\|_{2\to 2} = \|A\|_{2\to 2}$ and the orthogonality of the rows of A we get that $\|A^T\|_{2\to 2} = \sqrt{d/\tilde{d}} = d^{(1-\alpha)/2}$. From the normalization of the columns of A we get that $\|A^T\|_{2\to\infty} = 1$. Using the theorem for $\lambda = 1/q$, for any $q \geq 1$, we obtain $\|A^T\|_{2\to 2q} \leq d^{(1-\alpha)/2q}$. It is worth noting that $\|A^T\|_{2\to 2q}$ might actually be significantly lower then the given bound. For a specific seed, A_1, one should calculate $\|A_1^T\|_{2\to 2q}$ and use $\|A_\ell^T\|_{2\to 2q} = \|A_1^T\|_{2\to 2q}^{\ell}$ to achieve a possibly lower value for $\|A^T\|_{2\to 2q}$.

Lemma 4.1. *For a lean Walsh transform, A, we have that for any $p > 1$ the following holds:*

$$\{x \in \mathbb{R}^d \mid \|x\|_2 = 1, \|x\|_{2p} \leq k^{-1/2} d^{-\frac{1-\alpha}{2}(1-\frac{1}{p})}\} \subset \chi(A, \varepsilon, n) \tag{13}$$

where $k = O(\log(n)/\varepsilon^2)$, $\alpha = \log(r)/\log(c)$, r is the number of rows, and c is the number of columns in the seed of A.

Proof. We combine the above and use the duality of p and q:

$$\|x\|_A \leq \|x\|_{2p} \|A^T\|_{2\to 2q} \tag{14}$$

$$\leq \|x\|_{2p} d^{\frac{1-\alpha}{2q}} \tag{15}$$

$$\leq \|x\|_{2p} d^{\frac{1-\alpha}{2}(1-\frac{1}{p})} \tag{16}$$

The desired property, $\|x\|_A \leq k^{-1/2}$, is achieved if $\|x\|_{2p} \leq k^{-1/2} d^{-\frac{1-\alpha}{2}(1-\frac{1}{p})}$ for any $p > 1$. ∎

4.1 Controlling α and choosing R

We see that increasing α is beneficial from the theoretical stand point since it weakens the constraint on $\|x\|_p$. However, the application oriented reader should keep in mind that this requires the use of a larger seed, which subsequently increases the constant hiding in the big O notation of the running time.

Consider the seed constructions described in Fact 3.1 for which $r = c - 1$. Their skewness $\alpha = \log(r)/\log(c)$ approaches 1 as their size increases. Namely, for any positive constant δ there exists a constant size seed such that $1 - 2\delta \leq \alpha \leq 1$.

Lemma 4.2. *For any positive constant $\delta > 0$ there exists a Lean Walsh matrix, A, such that:*

$$\{x \in \mathbb{R}^d \mid \|x\|_2 = 1\ ,\ \|x\|_\infty \leq k^{-1/2} d^{-\delta}\} \subset \chi(A, \varepsilon, n) \tag{17}$$

Proof. Generate A from a seed such that its skewness $\alpha = \log(r)/\log(c) \geq 1 - 2\delta$ and substitute $p = \infty$ into the statement of Lemma 4.1. □

The constant α also determines the minimal dimension d (relative to k) for which the projection can be completed in $O(d)$ operations, the reason being that the vectors $z = AD_s x$ must be mapped from dimension $\tilde{d}$ ($\tilde{d} = d^\alpha$) to dimension k in $O(d)$ operations. This is done using the Ailon and Liberty [22] construction serving as the random projection matrix R. R is a $k \times \tilde{d}$ Johnson Lindenstrauss projection matrix which can be applied in $\tilde{d}\log(k)$ operations if $\tilde{d} = d^\alpha \geq k^{2+\delta''}$ for arbitrary small δ''. For the same choice of a seed as in lemma 4.2, the condition becomes $d \geq k^{2+\delta''+2\delta}$ which can be achieved by $d \geq k^{2+\delta'}$ for arbitrary small δ' depending on δ and δ''. Therefore for such values of d the matrix R exists and requires $O(d^\alpha \log(k)) = O(d)$ operations to apply.

5 Conclusion and work in progress

We have shown that any $k \times d$ (column normalized) matrix, A, can be composed with a random diagonal matrix to constitute a random projection matrix for some part of the Euclidian space, χ. Moreover, we have given sufficient conditions, on $x \in \mathbb{R}^d$, for belonging to χ depending on different $\ell_2 \to \ell_p$ operator norms of A^T and l_p norms of x. We have also seen that lean Walsh matrices exhibit both a "large" χ and a linear time computation scheme. These properties make them good building blocks for the purpose of random projections.

However, as explained in the introduction, in order for the projection to be complete, one must design a linear time preprocessing matrix Ψ which maps all vectors in $\mathbb{R}^d$ into χ (w.h.p). Achieving such Ψ would be extremely interesting from both the theoretical and practical stand point. Possible choices for Ψ may include random permutations, various wavelet/wavelet-like transforms, or any other sparse orthogonal transformation.

The authors would like to thank Steven Zucker, Daniel Spielman, and Yair Bartal for their insightful ideas and suggestions.

References

[1] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[2] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory Series A*, 44:355–362, 1987.

[3] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC)*, pages 604–613, 1998.

[4] S. DasGupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. *Technical Report, UC Berkeley*, 99-006, 1999.

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

[6] J. Matousek. On variants of the Johnson-Lindenstrauss lemma. *Preprint*, 2006.

[7] P. Drineas, M. W. Mahoney, and S.M. Muthukrishnan. Sampling algorithms for ℓ_2 regression and applications. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Miami, Florida, United States, 2006.

[8] T. Sarlós. Improved approximation algorithms for large matrices via random projections. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, Berkeley, CA, 2006.

[9] A. M. Frieze, R. Kannan, and S. Vempala. Fast monte-carlo algorithms for finding low-rank approximations. In *IEEE Symposium on Foundations of Computer Science*, pages 370–378, 1998.

[10] S. Har-Peled. A replacement for Voronoi diagrams of near linear size. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 94–103, Las Vegas, Nevada, USA, 2001.

[11] D. Achlioptas and F. McSherry. Fast computation of low rank matrix approximations. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 2001.

[12] P. Drineas and R. Kannan. Fast monte-carlo algorithms for approximate matrix multiplication. In *IEEE Symposium on Foundations of Computer Science*, pages 452–459, 2001.

[13] P.G. Martinsson, V. Rokhlin, and M. Tygert. A randomized algorithm for the approximation of matrices. *In review. Yale CS research report YALEU/DCS/RR-1361.*

[14] E. Liberty, F. Woolfe, P. G. Martinsson, V. Rokhlin, and M. Tygert. Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences (PNAS)*, Dec 2007.

[15] A. Dasgupta, P. Drineas, B. Harb, R. Kumar, and M. W. Mahoney. Sampling algorithms and coresets for ℓ_p regression. *Proc. of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008.

[16] P. Drineas, M. W. Mahoney, S. Muthukrishnan, and T. Sarlos. Faster least squares approximation. *TR arXiv:0710.1435*, submitted for publication, 2007.

[17] P. Drineas, M.W. Mahoney, and S. Muthukrishnan. Relative-error cur matrix decompositions. *TR arXiv:0708.3696*, submitted for publication, 2007.

[18] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, Dec. 2006.

[19] P. Paschou, E. Ziv, E. Burchard, S. Choudhry, W. Rodriguez-Cintron, M. W. Mahoney, and P. Drineas. Pca-correlated snps for structure identification in worldwide human populations. *PLOS Genetics, 3, pp. 1672-1686*, 2007.

[20] P. Paschou, M. W. Mahoney, J. Kidd, A. Pakstis, K. Kidd S. Gu, and P. Drineas. Intra- and inter-population genotype reconstruction from tagging snps. *Genome Research, 17(1), pp. 96-107*, 2007.

[21] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *Proceedings of the 38st Annual Symposium on the Theory of Compututing (STOC)*, pages 557–563, Seattle, WA, 2006.

[22] N. Ailon and E. Liberty. Fast dimension reduction using Rademacher series on dual BCH codes. In *Symposium on Discrete Algorithms (SODA), accepted*, 2008.

[23] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes.* Springer-Verlag, 1991.