Proceedings of Machine Learning Research 260:-, 2024 ACML 2024

Non-Oblivious Performance of Random Projections

Maciej Skorski MACIEJ.SKORSKI@GMAIL.COM *University of Warsaw* and **Alessandro Temperoni** *University of Luxembourg*

Editors: Vu Nguyen and Hsuan-Tien Lin

Abstract

Random projections are a cornerstone of high-dimensional computations. However, their analysis has proven both difficult and inadequate in capturing the empirically observed accuracy. To bridge this gap, this paper studies random projections from a novel perspective, focusing on data-dependent, that is, *non-oblivious*, performance.

The key contribution is the precise and data-dependent accuracy analysis of Rademacher random projections, achieved through elegant geometric methods of independent interest, namely, *Schur-concavity*. The result formally states the following property: the less spread-out the data is, the better the accuracy. This leads to notable improvements in accuracy guarantees for data characterized by sparsity or distributed with a small spread.

The key tool is a novel algebraic framework for proving Schur-concavity properties, which offers an alternative to derivative-based criteria commonly used in related studies. We demonstrate its value by providing an alternative proof for the extension of the celebrated Khintchine inequality.

Keywords: Johnson-Lindenstrauss Lemma; Rademacher Chaos; Random Projections; Schur-convexity; Concentration Inequalities.

1. Introduction

1.1. Background and Motivation

The seminal result of Johnson and Lindenstrauss (1984) set a milestone in high-dimensional data analysis, demonstrating that *random linear projections* are perfect embeddings: they *compress almost isometrically*, mapping data into a much smaller dimension while guaranteeing low distortion of distances. Remarkably, this feature comes with *precise probability guarantees*, which enabled applications to a range of problems and areas, including nearest-neighbour search Kleinberg (1997), clustering Dasgupta (1999); Cohen et al. (2015); Makarychev et al. (2022), outlier detection Aouf and Park (2012), ensemble learning Cannings and Samworth (2017), adversarial machine learning Vinh et al. (2016), feature hashing in machine learning Jagadeesan (2019), numerical linear algebra Sarlos (2006); Cohen et al. (2015); Clarkson and Woodruff (2017), convex optimization Zhang et al. (2013), differential privacy Blocki et al. (2012), neuroscience Ganguli and Sompolinsky (2012) and numerous others; for a comprehensive literature review we refer the reader to the recent survey Freksen (2021). The long-line of research Frankl and Maehara (1988); Indyk and Motwani (1998); Achlioptas (2003); Ailon and Chazelle (2006); Matoušek (2008); Dasgupta et al. (2010); Kane and Nelson (2014); Jagadeesan (2019); Skorski (2022, 2021); Skorski et al. (2022) has established various accuracy guarantees in the form of accuracy-confidence bounds

$$\|\Phi x\| \approx \|x\| \quad \text{with high probability,} \tag{1}$$

where an appropriately sampled random matrix $\Phi \in \mathbb{R}^{m \times n}$ represents the projection of an n-dimensional vector x to m dimensions ($m \ll n$ is desired), and the relative approximation error in (1) is referred to as *distortion*.

All of the mentioned prior bounds were *input-oblivious*, that is independent of the data structure. However, empirical study in Venkatasubramanian and Wang (2011) has shown that existing accuracy bounds, although proven nearly optimal in worst-case scenarios, do not fully explain the performance observed on real data. While non-oblivious bounds are very useful, it is necessary to link the performance to actual data to bridge the aforementioned gap between theory and practice. Therefore, it is both interesting and scientifically important to address the following problem:

Give *precise and non-oblivious* accuracy analysis of random projections.

As for related work, little is known about non-oblivious accuracy of random projections. Extensive experiments Venkatasubramanian and Wang (2011); Fedoruk et al. (2018) highlight the gap, but didn't provide a pathway towards understanding the root cause. Some theoretical studies have suggested that theoretical overestimation may be partly attributed to neglecting the relative positions of data points, particularly the angles between vectors Kaban (2015), but assumed arbitrary structure of individual points. Very recently, Skorski (2022) found that the theoretical performance of sparse random projections improves a bit when inputs are described in non-euclidean, Renyi-entropy related, distances; these results are however of rather qualitative nature, as they were established asymptotically with crude constants and for statistically suboptimal variants of random projections.

1.2. Contributions

This paper addresses the posed problem by providing a detailed analysis of the statistical accuracy of *Rademacher random projections*. These projections, which utilize scaled ± 1 random matrices, were established as *dimension-optimal in a strict sense* by Burr et al. (2018). Notably, other popular variants of random projections have been shown to be significantly less accurate[1]. The paper makes the following novel contributions:

Non-oblivious and numerically optimal bounds. The presented bounds, for the first time, relate the accuracy to input structure, specifically to the *spreadness*. The more spread-out the input data is, the heavier the distribution of distortion becomes, and the worse the statistical guarantees are. As a particular case, we *characterize the best performance on sparse data*. These bounds are discussed in Section 3.1, and formally establish that stochastic moments of the projection distortion are Schur-concave with respect to the input.

Geometric insights via Schur-concavity framework. The novelty of the proposed approach, in the context of prior work on random projections, lies in its utilization of Schur-concavity—a tool for optimizing objectives by relating changes in input spreadness. This technique is discussed in detail in Section 3.2. Furthermore, as an application of this framework, the paper provides an elegant proof of the celebrated Khintchine inequality.

1. For instance, the known guarantees for the dimension of sparse projections are a constant factor away from this optimal result, as demonstrated in Kane and Nelson (2014).

Supplementary experiments. The project repository at GitHub[2] hosts the empirical results and code that reproduces the analysis.

1.3. Organization

The remainder of the paper is organized as follows: Section 2 introduces notation and technical notions used throughout the paper; then Section 3 overviews novel results of this paper, Section 4 presents numerical evaluation, Section 5 presents proofs, and Section 6 concludes the work. The paper is completed by the supplementary material available from the repository. The full version of this work is also available on arXiv[3].

2. Preliminaries

We start by recalling some basic concepts from probability theory, algebra, combinatorics and optimization.

Throughout the paper we work with standard probability distributions: normal, binomial, and Rademacher; sampling from these distributions is denoted as usual by $\sim \mathsf{Norm}(\mu, \sigma^2)$, $\sim \mathsf{Binom}(n, k)$ and $\sim \{-1, 1\}$ respectively.

Vector norms used in this paper are the Euclidean norm denoted by $\|\cdot\|$ and the norm $\|\cdot\|_0$ that counts the number of non-zero components; we also say that x is ℓ-sparse when $\ell = \|x\|_0$, for example, $x = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0\right)$ is 3-sparse.

Partitions represent a positive integer q as a sum of positive integers. A non-increasing and non-negative sequence λ is a partition of q, denoted as $\lambda \vdash q$, when $\sum_i \lambda_i = q$; in the *frequency notation* distinct parts are assigned frequencies, so that $\lambda = 1^{f_1} \ldots q^{f_q}$ where $\sum_i i f_i = q$. For example, we have $\lambda = 1^2 2 = (2, 1, 1) \vdash 4$.

Monomial symmetric functions for a given partition $\lambda = (\lambda_1, \ldots, \lambda_k) \vdash q$ are defined as the sum of all distinct monomials with exponent λ, that is $\mathbf{m}_\lambda(x) = \sum_{i_1,\ldots,i_k} x_{i_1}^{\lambda_1} \cdots x_{i_k}^{\lambda_k}$ where $i_1, \ldots, i_k$ and monomials $x_{i_1}^{\lambda_1} \cdots x_{i_k}^{\lambda_k}$ are distinct. For example, $\mathbf{m}_{(2,1,1,1)}(x_1, x_2, x_3, x_4) = x_1^2 x_2 x_3 x_4 + x_1 x_2^2 x_3 x_4 + x_1 x_2 x_3^2 x_4 + x_1 x_2 x_3 x_4^2$.

Elementary symmetric polynomials are defined as $\mathbf{e}_k(x) = \sum_{i_1 < \ldots < i_k} x_{i_1} x_{i_2} \cdots x_{i_k}$. Both monomial and elementary symmetric polynomials form a basis of symmetric polynomials. For more details, we refer to the monographs Alexandersson (2020).

The *majorization order* on n-dimensional vectors is defined as follows: we say that x dominates y, denoted by $y \prec x$, if for their non-increasing rearrangements $(x_i^{\downarrow})$ and $(y_i^{\downarrow})$ we have the inequality $\sum_{i=1}^k x_i^{\downarrow} \geqslant \sum_{i=1}^k y_i^{\downarrow}$ for $k = 1, \ldots, n$ with equality when $k = n$; equivalently, y can be produced from x by a finite sequence of *Robin-Hood operations* which replace $x_i > x_j$ by $x_i \leftarrow x_i - \epsilon, x_j \leftarrow x_j + \epsilon$ for $\epsilon \in \left(0, \frac{x_i - x_j}{2}\right)$. Intuitively, x dominating y means that x is less spread-out, or more dispersed, compared to y. For example, we have $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \prec \left(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3}\right)$ (the transformation takes 3 Robin-Hood transfers).

The *Schur-convexity* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the following property: $x \prec y$ implies $f(x) \leqslant f(y)$; we speak of Schur-concavity when the inequality is reversed. Schur-convex or Schur-concave functions are necessarily symmetric; a symmetric function is Schur-convex if

2. https://github.com/maciejskorski/NonobliviousRademacherProjections
3. https://arxiv.org/abs/2303.11774

$\left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}\right)(x_i - x_j) \geqslant 0$ (Schur-Ostrowski criterion). For example, power sums $\sum_i x_i^q$ for $q \geqslant 1$ are Schur-convex. For more on *majorization* and *Schur-convexity* we refer to Arnold and Sarabia (2018); Shi and Technology (2019).

We define the *moment domination* of a random variable Y over X, denoted as $X \prec_m Y$, by requiring $\mathbf{E}X^q \leqslant \mathbf{E}Y^q$ for all positive integers q. In particular, it implies that MGF of Y dominates the MGF of X, the majorization in the Lorentz stochastic order Arnold and Sarabia (2018).

Rademacher chaos is understood as a quadratic form in Rademacher random variables, as usually in probability theory Boucheron et al. (2003).

For assessing statistical accuracy, variants of Johson-Lindenstrauss Lemma are developed for fixed input. Technically, this is referred to as the *Distributional Johnson-Lindenstrauss Lemma* Freksen (2021); performance on many points $\{x^i, \ldots, x^n)$ can be reduced to study the performance on pairwise differences $x = x^i - x^j$.

3. Results

3.1. Main Result

In Theorem 1 below, we provide the promised numerically sharp, non-oblivious and geometrically insightful bounds for Rademacher random projections, defined by the matrix

$$\Phi_{k,i} = \frac{1}{\sqrt{m}} r_{k,i}, \quad r_{k,i} \sim^{IID} \{-1, +1\}. \tag{2}$$

In the particularly interesting case of sparse inputs, we obtain more explicit formulas involving binomial distributions.

Theorem 1 (Sharp Moment Bounds for Rademacher Random Projections) *Let Φ be sampled according to the Rademacher scheme (2), and define the distortion as*

$$E(x) \triangleq \frac{\|\Phi x\|^2}{\|x\|^2} - 1. \tag{3}$$

Then the following holds true:

(a) *The moments of $E(x)$ are Schur-concave symmetric polynomials in (x_i^2)*

(b) *$E(x)$ is moment-dominated by E_* defined as*

$$E_* = \frac{1}{m}\sum_{i=1}^{m}(Z_i^2 - 1) \tag{4}$$

where Z_i are standardized binomial r.v.s.:

$$Z_i \sim^{IID} \frac{B - \mathbf{E}B}{\sqrt{\mathbf{Var}[B]}}, \; B \sim \mathsf{Binom}\left(\|x\|_0, \frac{1}{2}\right). \tag{5}$$

Equivalently,

$$\mathbf{E}E(x)^q \leqslant \mathbf{E}E_*^q \tag{6}$$

holds for $q = 2, 3, \ldots$ with equality when all components of the input x are equal.

Remark 2 (Distortion variants) *Consistently with the literature, we upper-bound* $E(x) = \|\Phi x\|^2/\|x\|^2 - 1$ *which is tractable and also yields an upper bound on* $\|\Phi x\|/\|x\| - 1$*, because of the inequality* $|e-1| \leqslant \sqrt{e^2-1}$ *valid for non-negative* e*.*

Remark 3 (Distortion concentration) *Majorization by* E_**, which is a sum of independent zero-mean r.vs, implies that* $E(x) \approx 0$*, equivalently* $\|\Phi x\| \approx \|x\|$*, with high probability.*

The first part of Theorem 1 establishes that the more spread-out the input vector, the higher the moments, resulting in a heavier distortion tail and consequently worse accuracy in (3); Schur-concavity is used to precisely capture how much the input is spread-out / dispersed. While this allows for comparing the performance between different points, the second result captures the worst-case performance by input sparsity (the higher sparsity, the better); bounds are stated in terms of moment majorization by explicit distributions, and therefore can be converted into concrete probability tails with Markov's inequality.

Author	Result
Burr et al. (2018)	$\max_x \mathbf{P}[\|E(x)\| > \epsilon] \geqslant 2\exp\left(-\frac{m\epsilon^2(1+o(1))}{4})\right)$ when $m \gg \epsilon^{-2}, n \gg 1$
Achlioptas (2001)	$\mathbf{P}[\|E(x)\| > \epsilon] \leqslant 2\exp\left(-\frac{m\epsilon^2}{4}\left(1-\frac{2}{3}\epsilon\right)\right)$
this paper	$E(x) \prec_m \frac{\sum_{i=1}^m Z_i^2 - 1}{m}$, $Z_i \sim^{IID} \frac{B - \mathbf{E}B}{\sqrt{\mathbf{Var}[B}}$, $B \sim \mathsf{Binom}\left(\|x\|_0, \frac{1}{2}\right)$

Table 1: Bounds from Theorem 1 compared with the best prior bounds Achlioptas (2001) and the sharp no-go results Burr et al. (2018). Our bounds imply those from prior work by standard "normal majorization" arguments.

The presented result is *numerically optimal* and *captures input sparsity*. It should be compared against the bounds from Achlioptas (2001) and the no-go result from Burr et al. (2018), as illustrated in Table 1. To see that our bounds are better than those in Achlioptas (2001), it suffices to use the Gaussian majorization argument to obtain a weaker bound $E(x) \prec_m \frac{\sum_{i=1}^m (N_i^2 - 1)}{m}$ where N_i are independent standard normal random variables, and use known sub-gamma tail bounds for chi-square distributions (for example, those developed in the monograph on concentration inequalities by Boucheron et al. (2003)). On sparse inputs, our bounds improve confidence by a constant factor, as shown by the empirical evaluation discussed in Section 4.

We now briefly overview the proof of Theorem 1 (see Figure 1): it starts by a reduction to the non-normalized distortion and the dimension $m = 1$, and writing the distortion as a Rademacher chaos of order 2; we then find extreme values of its moments geometrically, by means of *Schur optimization*. Finally, these extreme values can be found explicitly and efficiently by linking them to binomial moments.

3.2. Techniques: Proving Schur Convexity

We present a useful framework for proving Schur convexity properties. It makes repeated use of few auxiliary facts to eventually reduce the task to a 2-dimensional problem. This is often easier than the classical approach of evaluating derivative tests.

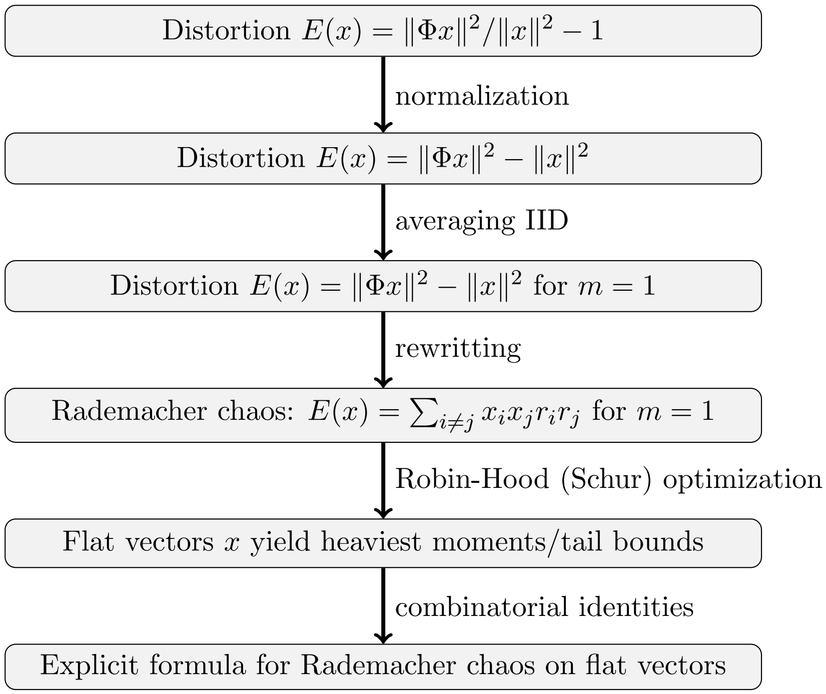


Figure 1: The proof roadmap for Theorem 1.

Theorem 4 Non-negative *Schur-convex (or concave) functions form a semi-ring.*

Lemma 5 *A multivariate function is Schur-convex (respectively, Schur-concave) if and only if it is symmetric and Schur-convex (respectively, Schur-concave) with respect to each pair of variables.*

To demonstrate the usefulness of these facts, we sketch an alternative proof of a refined version of celebrated Khintchine's Inequality, due to Efron. This refinement plays an important role in statistics, namely in proving properties of the popular Student-t tests.

Corollary 6 (Refined Khintchine Inequality Efron (1968)) *The mapping*

$$x \to \mathbf{E}\left(\sum_i x_i r_i\right)^q$$

is a Schur-concave function of (x_i^2). Consequently, for $\sigma = \|x\|_2$ we have

$$\mathbf{E}\left(\sum_{i=1}^n x_i r_i\right)^q \leqslant \mathbf{E}\left(\frac{\sigma}{\sqrt{n}}\sum_{i=1}^n r_i\right)^q \leqslant \mathbf{E}\mathsf{Norm}(0,\sigma^2)^q.$$

Proof The symmetry with respect to (x_i) is obvious. Applying the multinomial expansion to $(\sum_i x_i r_i)^q$, taking the expectation and using the symmetry of Rademacher random variables, we conclude that $\mathbf{E}(\sum_i x_i r_i)^q$ is a symmetric polynomial in variables

(x_i^2) with non-negative coefficients. By Lemma 5, it suffices to prove the Schur-concavity property for x_1^2, x_2^2. By the binomial formula and the independence of r_1, r_2 from $(r_i)_{i>2}$, we see that $\mathbf{E}(\sum_i x_i r_i)^q = \sum_k \binom{q}{k} \mathbf{E}(\sum_{i>2} x_i r_i)^{q-k} \cdot \mathbf{E}(x_1 r_1 + x_2 r_2)^k$ is a combination of expressions $\mathbf{E}(x_1 r_1 + x_2 r_2)^k$ with coefficients $c_k = \binom{q}{k}\mathbf{E}(\sum_{i>2} x_i r_i)^{q-k}$ that are independent of x_1, x_2 and non-negative due to the symmetry of r_i. By Theorem 4, it suffices to prove that $F_k \triangleq \mathbf{E}(x_1 r_1 + x_2 r_2)^k$ is a Schur-concave function of x_1^2, x_2^2. Define $G_k \triangleq \mathbf{E}(x_1 r_1 + x_2 r_2)^k x_1 x_2 r_1 r_2$. By $(x_1 r_1 + x_2 r_2)^k = (x_1 r_1 + x_2 r_2)^{k-2}(x_1^2 + x_2^2 + 2x_1 x_2 r_1 r_2)$ we have $F_k = (x_1^2 + x_2^2) F_{k-2} + 2G_{k-2}$ and $G_k = (x_1^2 + x_2^2) G_{k-2} + 2x_1^2 x_2^2 F_{k-2}$. Since $x_1^2 + x_2^2$ and $x_1^2 x_2^2$ are both Schur-concave in x_1^2, x_2^2, the Schur-concativity property of F_k, G_k is proven when it is proven for $k := k-2$. By mathematical induction, it suffices to realize that $F_0 = 1, F_1 = 0, G_1 = 1, G_2 = x_1^2 x_2^2$ are Schur-concave in x_1^2, x_2^2.

Let $\mathbf{1}_n$ be the vector of n ones. The first inequality follows then by $\frac{\sum_{i=1}^n x_i^2}{n} \cdot \mathbf{1}_n \prec (x_1^2, \ldots, x_n^2)$, and is clearly sharp. Since $\frac{1}{n+1}\mathbf{1}_{n+1} \prec \frac{1}{n}\mathbf{1}_n 0$, the Schur-concavity implies that $\mathbf{E}(\sum_{i=1}^n r_i/\sqrt{n})^q$ increases with n; the second inequality follows by the Central Limit Theorem: first we see that $\sum_{i=1}^n r_i/\sqrt{n}$ converges to $\mathsf{Norm}(0,1)$ in distribution, then we apply the extended CLT due to Bernstein (see for instance Hall (1978)) which guarantees also the convergence of moments, as $\max_i \mathbf{E} r_i^q < +\infty$. ■

3.3. Techniques: Rademacher Chaoses

Of independent interests are the techniques used in this work. The first result analyses the quadratic Rademacher chaos geometrically. It is similar in the spirit of the results of Efron (1968) and Eaton (1970), which however concern only a first-order Rademacher chaos.

Theorem 7 (Schur-concavity of Rademacher Chaoses) *Let (r_i) be a sequence of independent Rademacher random variables. Then the Rademacher chaos moment*

$$\mathbf{R}_q(x) \triangleq \mathbf{E}\left(\sum_{i \neq j} x_i x_j r_i r_j\right)^q \tag{7}$$

is a Schur-concave function of (x_i^2) for every positive integer q.

The second result is a recipe for explicitly computing the extreme moment values:

Theorem 8 (Extreme moments of Rademacher Chaos) *For any x and $K = \|x\|_0$ the following holds:*

$$\mathbf{R}_q(x) \leqslant \mathbf{R}_q(x^*), \quad x^* = \underbrace{\left(\frac{\|x\|_2}{\sqrt{K}}, \ldots, \frac{\|x\|_2}{\sqrt{K}}\right)}_{K \text{ times}}, \tag{8}$$

and furthermore the explicit value of this bound equals

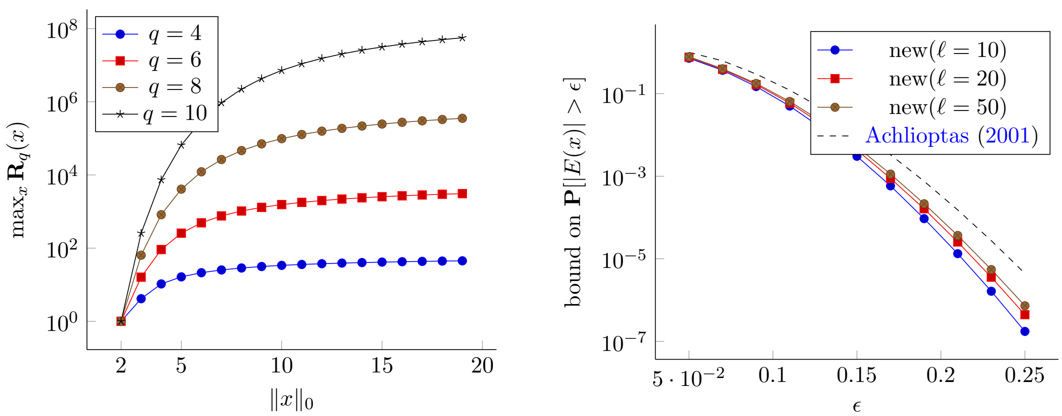
$$\mathbf{R}_q(x^*) = \|x\|_2^{2q} \cdot \mathbf{E}_{\bar{B}}(\bar{B}^2 - 1)^q, \tag{9}$$

where $\bar{B} = \frac{B - K/2}{\sqrt{K/4}}$ standardizes the symmetric binomial distribution with K trials B.

4. Numerical Comparison

To validate our findings, we performed experiments on synthetic and real-world data, as detailed below. More results, along with the code, are available from the GitHub repository.

Figure 2(a) and Figure 2(b) demonstrate numerical improvements, highlighting that the input dispersion is important: random projections are seen *less distorted when input data are less dispersed / more sparse*. This is further illustrated on Hamlet quotes in Figure 3.



(a) Distortion moments vs data dispersion

(b) Distortion tail

Figure 2: (a) The more spread-out the input (controlled by sparsity $\|x\|_0$), the more distorted the projected output (captured by $\mathbf{R}_q(x)$, the Rademacher chaos moment). Utilizing the input dispersion improves probability bounds by orders of magnitude. Note: for normalization, we assume $\|x\|_2 = 1$. (b) Capturing input-sparsity ($\ell = \|x\|_0$) improves the bounds on Rademacher random projections, as demonstrated by distortion probability tails.

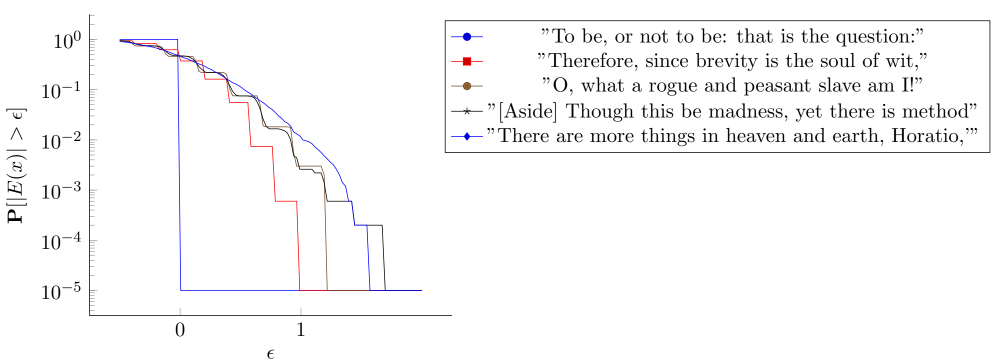


Figure 3: Performance of random embeddings of dimension 10, measured by distortion tails, applied to the famous quotations from Hamlet. The corpus was preprocessed with TF-IDF.

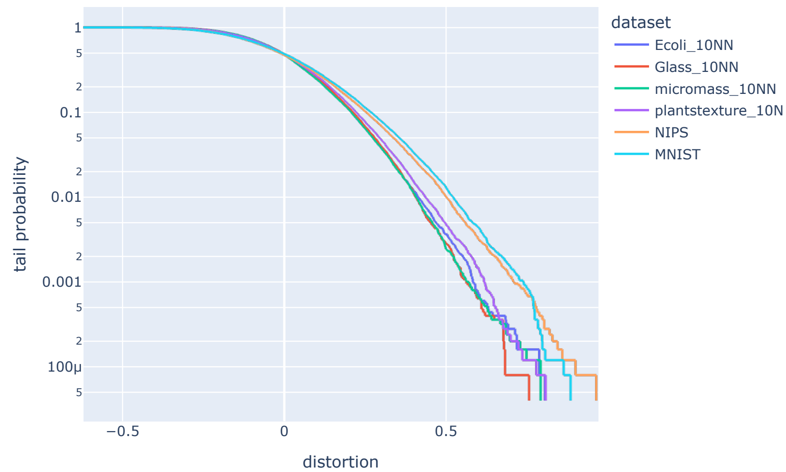


Figure 4: Distortion, shown as probability tails, measured on various datasets.

5. Proofs

5.1. Proof of Theorem 4

Consider two non-negative functions f, g and inputs $x \prec y$. Consider the identity

$$f(y)g(y) - f(x)g(x) = (f(y) - f(x)) \cdot g(y) + f(x) \cdot (g(y) - g(x)). \tag{10}$$

If f, g are Schur-convex then $f(y) - f(x) \geqslant 0$ and $g(y) - g(x) \geqslant 0$ and the whole expression is non-negative when f, g are non-negative. This shows that $f \cdot g$ is also Schur-convex. The claim for Schur-concave functions follows analogously (the expression is then non-positive).

5.2. Proof of Lemma 5

The proof follows from the fact that x is dominated by y if and only if x can be produced from y by a sequence of *Robin-Hood operations*, and the fact that Robin-Hood operations change only two fixed components of vectors.

5.3. Proof of Theorem 7

Proof Note that $\mathbf{R}_q$ is a polynomial in x_i^2 with integer coefficients, and thus a well-defined function of (x_i^2). This follows by applying the multinomial expansion and noticing that monomials with odd exponents have expectation zero due to the symmetry of Rademacher distribution. $\mathbf{R}_q$ is obviously symmetric. By Lemma 5 it now suffices to validate the Schur-concavity for x_1^2, x_2^2 and any fixed choice of $(x_i)_{j>2}$. Define the following expressions

$$P = \sum_{i \not\in \{1,2\}} x_i r_i, \quad R = \sum_{i,j \not\in \{1,2\}} x_i x_j r_i r_j, \tag{11}$$

then our task is to prove the Schur-concavity of the function

$$\mathbf{R}_q \triangleq \mathbf{E}\left(P(x_1r_1+x_2r_2)+x_1x_2r_1r_2+R\right)^q, \tag{12}$$

with respect to x_1^2, x_2^2.

By the multinomial expansion we find that

$$\mathbf{R}_q \triangleq \sum_{q_1+q_2+q_3=q} \binom{q}{q_1,q_2,q_3} \left[\mathbf{E}\left[P^{q_1}R^{q_3}\right] \quad \mathbf{E}\left[(x_1r_1+x_2r_2)^{q_1}(x_1x_2r_1r_2)^{q_2}\right]\right], \tag{13}$$

where we used the independence of r_1, r_2 on $(r_i)_{i>2}$ and thus also on P, R. Observe that $P^{q_1}R^{q_3}$ is, by definition and our assumption $x_i \geqslant 0$, a polynomial in symmetric random variables r_i with non-negative coefficients. This observation shows that

$$\mathbf{E}\left[P^{q_1}R^{q_3}\right] \geqslant 0, \tag{14}$$

and by Theorem 4 it suffices to prove that

$$F \triangleq \mathbf{E}\left[(x_1r_1+x_2r_2)^{q_1}(x_1x_2r_1r_2)^{q_2}\right] \tag{15}$$

is Schur-concave as a function of x_1^2, x_2^2 for any non-negative integers q_1, q_2.

To see that F is indeed a well-defined function of x_1^2, x_2^2, note that it equals the expectation of a polynomial in the symmetric random variables $y_i = x_i r_i$; thus only monomials with even-degrees contribute, and the result is a polynomial in $y_i^2 = x_i^2$. In fact, F equals the sum of even-degree monomials in the expanded polynomial $(x_1+x_2)^{q_1}(x_1x_2)^{q_2}$.

We next observe that

$$F = \begin{cases} (x_1x_2)^{q_2}\mathbf{E}\left[(x_1r_1+x_2r_2)^{q_1}\right] & q_2 \text{ even} \\ (x_1x_2)^{q_2-1}\mathbf{E}\left[(x_1r_1+x_2r_2)^{q_1}x_1x_2r_1r_2\right] & q_2 \text{ odd.} \end{cases} \tag{16}$$

Note that x_1x_2 is Schur-concave in non-negative x_1, x_2; indeed, the identity $(x_1+\epsilon)(x_2-\epsilon) = x_1x_2+\epsilon(x_2-x_1-\epsilon)$ shows that Robin-Hood transfers increase the value. By Theorem 4 we conclude that $(x_1x_2)^k$ is Schur concave in x_1^2, x_2^2 for non-negative even k. Thus, by Equation (16) and Theorem 4 we conclude that it suffices to consider the case $q_2 = 1$, that is, to prove the Schur-concavity of the following two functions:

$$G_k \triangleq \mathbf{E}\left[(x_1r_1+x_2r_2)^k\right] \tag{17}$$

$$H_k \triangleq \mathbf{E}\left[(x_1r_1+x_2r_2)^k x_1x_2r_1r_2\right], \tag{18}$$

with respect to x_1^2, x_2^2, for any non-negative integer k.

Using the identity $(x_1r_1+x_2r_2)^k = (x_1r_1+x_2r_2)^{k-2}(x_1^2+x_2^2+2x_1x_2r_1r_2)$, which follows as $r_1^2 = r_2^2 = 1$, we obtain the following recurrence relations

$$G_k = (x_1^2+x_2^2)G_{k-2} + 2H_{k-2} \tag{19}$$

$$H_k = 2x_1^2x_2^2G_{k-2} + (x_1^2+x_2^2)H_{k-1}, \tag{20}$$

valid for $k \geqslant 2$. Since $x_1^2 + x_2^2$ and $x_1^2 x_2^2$ are Schur-concave as functions of x_1^2, x_2^2, by Theorem 4 the concavity property proven for $k-2$ implies that it is valid also for k. By induction, it suffices to verify the case $k=0$ and $k=1$. But we see that

$$\begin{aligned} G_0 &= 1 \\ G_1 &= 0 \\ H_0 &= 1 \\ H_1 &= 2x_1^2 x_2^2 \end{aligned} \tag{21}$$

are all Schur-concave as functions of x_1^2, x_2^2. This completes the proof. ■

5.4. Proof of Theorem 8

Without loss of generality, we assume that $\|x\|_2 = 1$. From Theorem 7 and the fact that (x_i^2) majorizes (x_i^{*2}) we obtain

$$\begin{aligned} \max_{\|x\|_0 \leqslant K} \mathbf{E}\left(\sum_{i<j} x_i x_i r_i r_j\right)^q &= \mathbf{E}\left(\sum_{i<j} x_i^* x_i^* r_i r_j\right)^q \\ &= \mathbf{E}\left(\frac{1}{K}\sum_{1\leqslant i<j\leqslant K} r_i r_j\right)^q, \end{aligned} \tag{22}$$

Observe that $r_i = 1 - 2b_i$ where (b_i) is a sequence of independent Bernoulli random variables with parameter $\frac{1}{2}$. Therefore,

$$\begin{aligned} \mathbf{E}\left(\sum_{i=1}^K r_i\right)^q &=^{(a)} \sum_{k\in\mathbb{Z}} k^q \cdot \mathbf{P}\left\{\sum_{i=1}^K r_i = k\right\} \\ &=^{(b)} \sum_{k\in\mathbb{Z}} k^q \cdot \mathbf{P}\left\{\sum_{i=1}^K b_i = \frac{K-k}{2}\right\} \\ &=^{(c)} \sum_{i=0}^K (K-2i)^q \cdot \mathbf{P}\left\{\mathsf{Binom}\left(K, \frac{1}{2}\right) = i\right\} \\ &=^{(d)} \frac{1}{2^K}\sum_{i=0}^K \binom{K}{i}(K-2i)^q \\ &=^{(e)} \frac{1}{2^K}\sum_{i} \binom{K}{i}(-K+2i)^q, \end{aligned} \tag{23}$$

where in (a) we use the fact that $\sum_i r_i$ takes integer values, (b) follows by the identity $r_i = 1 - 2b_i$, (c) follows by $\mathsf{Binom}(K, 1/2) \sim \sum_{i=1}^K b_i$, (d) uses the explicit formula on the binomial probability mass function, and finally in (e) we substitute $i := K - i$ and use the symmetry of binomial coefficients $\binom{K}{i} = \binom{K}{K-i}$.

Using the above formula, we further calculate

$$
\begin{aligned}
\mathbf{E}\left(\sum_{1\leqslant i\neq j\leqslant K} r_i r_j\right)^q &=^{(a)} \mathbf{E}\left(\left(\sum_{i=1}^{K} r_i\right)^2 - \sum_{i=1}^{K} r_i^2\right)^q \\
&=^{(b)} \sum_j \binom{q}{j}(-K)^{q-j}\mathbf{E}\left(\sum_{i=1}^{K} r_i\right)^{2j} \\
&=^{(c)} \frac{1}{2^K}\sum_{i,j}\binom{q}{j}\binom{K}{i}(-K+2i)^{2j}(-K)^{q-j} \\
&=^{(d)} \frac{(-K)^q}{2^K}\sum_i \binom{K}{i}\left(1-\frac{(-K+2i)^2}{K}\right)^q,
\end{aligned} \tag{24}
$$

where (a) follows by the square sum completion, (b) follows by the binomial formula and $r_i^2 = 1$, (c) follows by Equation (23), and (d) is obtained by algebraic rearrangements.

Inserting Equation (23) into Equation (22), we arrive at

$$
\max_{x:\|x\|_0\leqslant K} \mathbf{E}\left(\sum_{1\leqslant i\neq j\leqslant K} x_i x_j r_i r_j\right)^q = \frac{1}{2^K}\sum_{i=0}^{K}\binom{K}{i}\left(\frac{(-K+2i)^2}{K}-1\right)^q. \tag{25}
$$

To simplify further, let $Z \sim \frac{\mathsf{Binom}\left(K,\frac{1}{2}\right)-\frac{K}{2}}{\sqrt{\frac{K}{4}}}$ be the standardization of the symmetric binomial distribution. Denoting $i \sim \mathsf{Binom}\left(K,\frac{1}{2}\right)$ we have $Z^2 \sim \frac{\left(i-\frac{K}{2}\right)^2}{\frac{K}{4}} = \frac{(-K+2i)^2}{K}$, and we can rewrite Equation (25) as follows:

$$
\max_{x:\|x\|_0\leqslant K} \mathbf{E}\left(\sum_{1\leqslant i\neq j\leqslant K} x_i x_j r_i r_j\right)^q = \mathbf{E}_Z\left(Z^2-1\right)^q, \tag{26}
$$

which finishes the proof.

5.5. Proof of Theorem 1

We have to prove that for the distortion $E(\cdot)$ defined as in Equation (3) the following inequality holds true:

$$
E(x) \leqslant E(y), \quad (y_i^2) \prec (x_i^2). \tag{27}
$$

The proof goes through several reduction steps until Schur-concavity of few simple functions.

We first observe that it suffices to prove that the moments of the expression

$$
x \to \|\Phi x\|^2 - \|x\|^2, \tag{28}
$$

are Schur-concavity with respect to (x_i^2). Indeed, since $(y_i^2) \prec (x_i^2)$ implies $\|x\|^2 = \sum_i x_i^2 = \sum_i y_i^2 = \|y\|^2$, we have $\mathbf{E}E(x)^q \leqslant \mathbf{E}E(y)^q$ if and only if $\mathbf{E}(\|\Phi x\|^2 - \|x\|^2)^q \leqslant \mathbf{E}(\|\Phi y\|^2 - \|y\|^2)^q$, by the definition of E.

We first prove that the distortion of m-dimensional projections is the average of m IID distortions of 1-D projections. Observe that

$$\|\Phi x\|^2 - \|x\|^2 = \sum_{k=1}^{m} \left((\Phi_k x)^2 - \mathbf{E}(\Phi_k x)^2\right), \tag{29}$$

where Φ_k is the k-th row of Φ; this follows by $\mathbf{E}(\Phi_k x)^2 = \sum_i x_i^2 \mathbf{Var}[\Phi_{k,i}] = \frac{1}{m}\|x\|^2$. Furthermore, the summands in (29) are independent and identically distributed:

$$(\Phi_k x)^2 - \mathbf{E}(\Phi_k x)^2 \sim \frac{1}{m}\sum_{i\neq j} x_i x_j r_i r_j. \tag{30}$$

Then we note that the Schur-concativity test can be done on the 1-D case. This follows because, due to the multinomial expansion applied to Equation (30), the q-th moment of m-dimensional distortion is a multivariate polynomial in 1-D distortion moments of order $k \leqslant q$, with non-negative coefficients; the distortion moments are themselves non-negative, and by Theorem 4 and Theorem 7 we obtain the first part of the theorem.

Finally, applying Theorem 8 proves the second part.

6. Conclusions

This work revisits the performance analysis of Rademacher random projections, connecting the statistical guarantees with the input structure as quantified by spreadness and, as a special case, sparsity. The main result of this paper proves Schur-concavity property for distortion moments, which makes the bounds numerically sharp and data-aware (non-obliviuos), while giving a geometric perspective to the performance of the projections. We benchmarked our bounds both theoretically and empirically by measuring the distortion of the projected vectors against the original input data. As a result, dense projections are preferred, and they work incredibly well with sparse input data. We believe that our findings are of broader interest for a variety of statistical-inference applications.

Acknowledgments

The authors thank the anonymous referees of the conference ACML'24 for detailed and constructive feedback. The first author acknowledges the financial support from the University of Warsaw, under the program "Excellence Initiative – Research University (2020-2026)".

References

Dimitris Achlioptas. Database-friendly random projections. In *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 274–281, 2001.

Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of computer and System Sciences*, 66(4):671–687, 2003.

Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 557–563, 2006.

Per Alexandersson. The symmetric functions catalog. 2020. URL https://www.symmetricfunctions.com.

Mazin Aouf and Laurence AF Park. Approximate document outlier detection using random spectral projection. In *Australasian Joint Conference on Artificial Intelligence*, pages 579–590. Springer, 2012.

B.C. Arnold and J.M. Sarabia. *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics*. Statistics for Social and Behavioral Sciences. Springer International Publishing, 2018. ISBN 9783319937731. URL https://books.google.pl/books?id=ibhmDwAAQBAJ.

Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss transform itself preserves differential privacy. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 410–419. IEEE, 2012.

Stéphane Boucheron, Gábor Lugosi, and Olivier Bousquet. Concentration inequalities. In *Summer school on machine learning*, pages 208–240. Springer, 2003.

Michael Burr, Shuhong Gao, and Fiona Knoll. Optimal bounds for Johnson-Lindenstrauss transformations. *The Journal of Machine Learning Research*, 19(1):2920–2941, 2018.

Timothy I Cannings and Richard J Samworth. Random-projection ensemble classification. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):959–1035, 2017.

Kenneth L Clarkson and David P Woodruff. Low-rank approximation and regression in input sparsity time. *Journal of the ACM (JACM)*, 63(6):1–45, 2017.

Michael B Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. Dimensionality reduction for k-means clustering and low rank approximation. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 163–172, 2015.

Anirban Dasgupta, Ravi Kumar, and Tamás Sarlós. A sparse Johnson-Lindenstrauss transform. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 341–350, 2010.

Sanjoy Dasgupta. Learning mixtures of Gaussians. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 634–644. IEEE, 1999.

Morris L Eaton. A note on symmetric bernoulli random variables. *The annals of mathematical statistics*, 41(4):1223–1226, 1970.

Bradley Efron. Student's t-test under non-normal conditions. Technical report, HARVARD UNIV CAMBRIDGE MA DEPT OF STATISTICS, 1968.

John Fedoruk, Byron Schmuland, Julia Johnson, and Giseon Heo. Dimensionality reduction via the Johnson–Lindenstrauss Lemma: theoretical and empirical bounds on embedding dimension. *The Journal of Supercomputing*, 74(8):3933–3949, 2018.

Peter Frankl and Hiroshi Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory, Series B*, 44(3):355–362, 1988.

Casper Benjamin Freksen. An introduction to johnson-lindenstrauss transforms. *arXiv preprint arXiv:2103.00564*, 2021.

Surya Ganguli and Haim Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual review of neuroscience*, 35: 485–508, 2012.

Peter Hall. On the rate of convergence of moments in the central limit theorem. *Journal of the Australian Mathematical Society*, 25(2):250–256, 1978.

Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 604–613, 1998.

Meena Jagadeesan. Understanding sparse JL for feature hashing. In *NeurIPS*, pages 15177–15187, 2019. (Full version: https://arxiv.org/pdf/1903.03605.pdf).

William B Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

Ata Kaban. Improved bounds on the dot product under random projection and random sign projection. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 487–496, 2015.

Daniel M Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss transforms. *Journal of the ACM (JACM)*, 61(1):1–23, 2014.

Jon M Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 599–608, 1997.

Konstantin Makarychev, Yury Makarychev, and Ilya Razenshteyn. Performance of Johnson–Lindenstrauss transform for k-means and k-medians clustering. *SIAM Journal on Computing*, (0):STOC19–269, 2022.

Jiří Matoušek. On variants of the Johnson–Lindenstrauss lemma. *Random Structures & Algorithms*, 33(2):142–156, 2008.

Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *2006 47th annual IEEE symposium on foundations of computer science (FOCS'06)*, pages 143–152. IEEE, 2006.

H. Shi and H.I. Technology. *Schur-Convex Functions and Inequalities: Volume 1: Concepts, Properties, and Applications in Symmetric Function Inequalities*. De Gruyter, 2019. ISBN 9783110607840. URL https://books.google.at/books?id=VrztDwAAQBAJ.

Maciej Skorski. Johnson-Lindenstrauss Transforms with best confidence. In *Conference on Learning Theory*, pages 3989–4007. PMLR, 2021.

Maciej Skorski. Entropy Matters: Understanding Performance of Sparse Random Embeddings. In *33rd International Symposium on Algorithms and Computation (ISAAC 2022)*, volume 248 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:18, 2022. ISBN 978-3-95977-258-7. doi: 10.4230/LIPIcs.ISAAC.2022.18. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ISAAC.2022.18.

Maciej Skorski, Temperoni Alessandro, and Theobald Martin. Robust and provable guarantees for sparse random embeddings. In *Advances in Knowledge Discovery and Data Mining*, pages 211–223, 2022.

Suresh Venkatasubramanian and Qiushi Wang. The Johnson-Lindenstrauss transform: an empirical study. In *2011 Proceedings of the Thirteenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 164–173. SIAM, 2011.

Nguyen Xuan Vinh, Sarah Erfani, Sakrapee Paisitkriangkrai, James Bailey, Christopher Leckie, and Kotagiri Ramamohanarao. Training robust models using random projection. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 531–536. IEEE, 2016.

Lijun Zhang, Mehrdad Mahdavi, Rong Jin, Tianbao Yang, and Shenghuo Zhu. Recovering the optimal solution by dual random projection. In *Conference on Learning Theory*, pages 135–157. PMLR, 2013.