# Random Projections Preserve Linearity in Sparse Spaces

**Mahdi Milani Fard** and **Yuri Grinberg** and **Joelle Pineau** and **Doina Precup**
School of Computer Science
McGill University, Montreal, Canada
{mmilan1, ygrinb, jpineau, dprecup}@cs.mcgill.ca

## Introduction

This paper includes a tightening of the bound presented in Theorem 2 of Fard et al. (2012). The new bound is simpler and easier to prove.

## Notations and Sparsity Assumption

Throughout this paper, column vectors are represented by lower case bold letters, and matrices are represented by bold capital letters. $|.|$ denotes the size of a set, and $\|.\|_0$ is Donoho's zero "norm" indicating the number of non-zero elements in a vector. $\|.\|$ denotes the $L^2$ norm for vectors and the operator norm for matrices: $\|\mathbf{M}\| = \sup_{\mathbf{v}} \|\mathbf{M}\mathbf{v}\|/\|\mathbf{v}\|$. Also, we denote the Moore-Penrose pseudo-inverse of a matrix $\mathbf{M}$ with $\mathbf{M}^\dagger$ and the smallest singular value of $\mathbf{M}$ by $\sigma_{\min}^{(M)}$.

We will be working in sparse input spaces. Our input is represented by a vector $\mathbf{x} \in \mathcal{X}$ of $D$ features, having $\|\mathbf{x}\| \leq 1$. We assume that $\mathbf{x}$ is $k$-sparse in some known or unknown basis $\mathbf{\Psi}$, implying that $\mathcal{X} \triangleq \{\mathbf{\Psi}\mathbf{z},\ \text{s.t.}\ \|\mathbf{z}\|_0 \leq k \text{ and } \|\mathbf{z}\| \leq 1\}$. For a concrete example, the signals can be natural images and $\Psi$ can represent these signals in the frequency domain (e.g., see Olshausen, Sallee, and Lewicki (2001)).

## Random Projections and Inner Product

In this work, we assume that each entry in a projection $\mathbf{\Phi}^{D\times d}$ is an i.i.d. sample from a Gaussian [1]:

$$\phi_{i,j} = \mathcal{N}(0, 1/d). \tag{1}$$

We build our work on the following (based on theorem 4.1 from Davenport, Wakin, and Baraniuk (2006)), which shows that for a finite set of points, inner product with a fixed vector is almost preserved after a random projection.

**Theorem 1.** *(Davenport, Wakin, and Baraniuk (2006)) Let $\mathbf{\Phi}^{D\times d}$ be a random projection according to Eqn 1. Let $S$ be a finite set of points in $\mathbb{R}^D$. Then for any fixed $\mathbf{w} \in \mathbb{R}^D$ and $\epsilon > 0$:*

$$\forall \mathbf{s} \in S : \left|\langle \mathbf{\Phi}^T\mathbf{w}, \mathbf{\Phi}^T\mathbf{s}\rangle - \langle \mathbf{w}, \mathbf{s}\rangle\right| \leq \epsilon\|\mathbf{w}\|\|\mathbf{s}\|, \tag{2}$$

*fails with probability less than* $(4|S| + 2)e^{-d\epsilon^2/48}$.

[1] The elements of the projection are typically taken to be distributed with $\mathcal{N}(0, 1/D)$, but we scale them by $\sqrt{D/d}$, so that we avoid scaling the projected values (see e.g. Davenport, Wakin, and Baraniuk (2006)).

We derive the corresponding theorem for sparse feature spaces.

**Theorem 2.** *Let $\Phi^{D\times d}$ be a random projection according to Eqn 1. Let $\mathcal{X}$ be a $D$-dimensional $k$-sparse space. Then for any fixed $\mathbf{w}$ and $\epsilon > 0$ :*

$$\forall \mathbf{x} \in \mathcal{X} : \left|\langle \mathbf{\Phi}^T\mathbf{w}, \mathbf{\Phi}^T\mathbf{x}\rangle - \langle \mathbf{w}, \mathbf{x}\rangle\right| \leq \epsilon\|\mathbf{w}\|\|\mathbf{x}\|, \tag{3}$$

*fails with probability less than* $(4D + 2)e^{-d\epsilon^2/48k}$.

## Proof of Theorem 2

*Proof of Theorem 2.* Let $e_i$ be the $i$th column of $\Psi$. Then $S = \{\mathbf{e}_i\}_{1\leq i\leq D}$ is an orthonormal basis under which the signal $\mathbf{x}$ is sparse and all $\mathbf{e}_i$'s are in $\mathcal{X}$. Using Theorem 1, with probability no less than $(4D + 2)e^{-d\epsilon^2/48k}$:

$$\forall i : \left|\langle \mathbf{\Phi}^T\mathbf{w}, \mathbf{\Phi}^T\mathbf{e}_i\rangle - \langle \mathbf{w}, \mathbf{e}_i\rangle\right| \leq \frac{\epsilon}{\sqrt{k}}\|\mathbf{w}\|. \tag{4}$$

For any $\mathbf{x} = \sum_i \alpha_i\mathbf{e}_i$:

$$\begin{aligned}
&\left|\langle \mathbf{\Phi}^T\mathbf{w}, \mathbf{\Phi}^T\mathbf{x}\rangle - \langle \mathbf{w}, \mathbf{x}\rangle\right| \\
&\quad\leq \sum_i \left|\alpha_i(\langle \mathbf{\Phi}^T\mathbf{w}, \mathbf{\Phi}^T\mathbf{e}_i\rangle - \langle \mathbf{w}, \mathbf{e}_i\rangle)\right| \\
&\quad\leq \frac{\epsilon}{\sqrt{k}}\|\mathbf{w}\| \sum_i |\alpha_i|\,.
\end{aligned}$$

As $\mathbf{x}$ is $k$-sparse in $\{\mathbf{e}_i\}$, we have that $\sum_i |\alpha_i| \leq \sqrt{k}\|\mathbf{x}\|$, which completes the proof.

□

## References

Davenport, M.; Wakin, M.; and Baraniuk, R. 2006. Detection and estimation with compressive measurements. *Dept. of ECE, Rice University, Tech. Rep.*

Fard, M.; Grinberg, Y.; Pineau, J.; and Precup, D. 2012. Compressed least-squares regression on sparse spaces. In *AAAI*.

Olshausen, B.; Sallee, P.; and Lewicki, M. 2001. Learning sparse image codes using a wavelet pyramid architecture. In *Proceedings of Advances in neural information processing systems*.