January 3, 2009

Compressed Sensing: Basic results and self contained proofs

Shai Shalev-Shwartz

Abstract

Compressed sensing is a linear dimensionality reduction technique which utilizes a prior assumption that the original vector is (approximately) sparse in some basis. In this note we summarize some of the known results and provide self contained, easy to follow, proofs.

1 Motivation

Consider a vector $\mathbf{x} \in \mathbb{R}^d$ that has at most s non-zero elements. That is,

$$\|\mathbf{x}\|_0 \stackrel{\text{def}}{=} |\{i : x_i \neq 0\}| \leq s \,.$$

Clearly, we can compress $\mathbf{x}$ by representing it using s (index,value) pairs. Furthermore, this compression is lossless – we can reconstruct $\mathbf{x}$ exactly from the s (index,value) pairs. Now, lets take one step forward and assume that $\mathbf{x} = U\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a sparse vector, $\|\boldsymbol{\alpha}\|_0 \leq s$, and U is a fixed orthonormal matrix. That is, $\mathbf{x}$ has a sparse representation in another basis. It turns out that many natural vectors are (at least approximately) sparse in some representation. In fact, this assumption underlies many modern compression schemes. For example, the JPEG-2000 format for image compression relies on the fact that natural images are approximately sparse in a wavelet basis.

Can we still compress $\mathbf{x}$ into roughly s numbers? Well, one simple way to do this is to multiply $\mathbf{x}$ by U^T, which yields the sparse vector $\boldsymbol{\alpha}$, and then represent $\boldsymbol{\alpha}$ by its s (index,value) pairs. However, this requires to first 'sense' $\mathbf{x}$, to store it, and then to multiply it by U^T. This raises a very natural question: Why go to so much effort to acquire all the data when most of what we get will be thrown away? Can't we just directly measure the part that won't end up being thrown away?

Compressed sensing is a technique that simultaneously acquire and compress the data. The key result is that a random linear transformation can compress $\mathbf{x}$ without loosing information. The number of measurements needed is order of $s \log(d)$. That is, we roughly acquire only the important information about the signal. As we will see later, the price we pay is a slower reconstruction phase. In some situations, it makes sense to save time in compression even at the price of a slower reconstruction. For example, a security camera should sense and compress a large amount of images while most of the time we do not need to decode the compressed data at all. Furthermore, in many practical applications, compression by a linear transformation is advantageous because it can be performed efficiently in hardware. For example, a team led by Baraniuk and Kelly have proposed a camera architecture that employs a digital micromirror array to perform optical calculations of a linear transformation of an image. In this case, obtaining each compressed measurement is as easy as obtaining a single raw measurement. Another important application of compressed sensing is medical imaging, in which requiring less measurements translates to less radiation for the patient.

2 Main results

Informally, the main results are the following three "surprising" results:

1. It is possible to fully reconstruct any sparse signal if it was compressed by $\mathbf{x} \mapsto W\mathbf{x}$, where W is a matrix which satisfies a condition so-called Restricted Isoperimetric Property (RIP). A matrix that satisfies this property is guaranteed to have a low distortion of the norm of any sparse representable vector.

2. The reconstruction can be calculated in polynomial time by solving a linear program.

3. A random $n \times d$ matrix is likely to satisfies the RIP condition provided that n is greater than order of $s \log(d)$.

Formally,

Definition 1 (RIP) *A matrix $W \in \mathbb{R}^{n,d}$ is (ϵ, s)-RIP if for all $\mathbf{x} \neq 0$ s.t. $\|\mathbf{x}\|_0 \leq s$ we have*

$$\left| \frac{\|W\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} - 1 \right| \leq \epsilon \ .$$

The first theorem establishes that RIP matrices yield a lossless compression scheme for sparse vectors. It also provides a (non-efficient) reconstruction scheme.

Theorem 1 *Let $\epsilon < 1$ and let W be a $(\epsilon, 2s)$-RIP matrix. Let $\mathbf{x}$ be a vector s.t. $\|\mathbf{x}\|_0 \leq s$, let $\mathbf{y} = W\mathbf{x}$ be the compression of $\mathbf{x}$, and let*

$$\tilde{\mathbf{x}} \in \operatorname*{argmin}_{\mathbf{v}: W\mathbf{v}=\mathbf{y}} \|\mathbf{v}\|_0$$

be a reconstructed vector. Then, $\tilde{\mathbf{x}} = \mathbf{x}$.

Proof We prove the theorem by assuming the contrary, namely assuming that $\tilde{\mathbf{x}} \neq \mathbf{x}$. Since $\mathbf{x}$ satisfies the constraints in the optimization problem that defines $\tilde{\mathbf{x}}$ we clearly have that $\|\tilde{\mathbf{x}}\|_0 \leq \|\mathbf{x}\|_0 \leq s$. Therefore, $\|\mathbf{x} - \tilde{\mathbf{x}}\|_0 \leq 2s$ and we can apply the RIP inequality on the vector $\mathbf{x} - \tilde{\mathbf{x}}$. But, since $W(\mathbf{x} - \tilde{\mathbf{x}}) = \mathbf{0}$ we get that $|0 - 1| \leq \epsilon$, which leads to a contradiction. ■

The reconstruction scheme given in Theorem 1 seems to be non-efficient because we need to minimize a combinatorial objective (the sparsity of $\mathbf{v}$). Quite surprisingly, it turns out that we can replace the combinatorial objective, $\|\mathbf{v}\|_0$, with a convex objective, $\|\mathbf{v}\|_1$, which leads to a linear programming problem that can be solved efficiently. This is stated formally in the following theorem.

Theorem 2 *Assume that the conditions of Theorem 1 holds and that $\epsilon < \frac{1}{1+\sqrt{2}}$. Then,*

$$\mathbf{x} \ = \ \operatorname*{argmin}_{\mathbf{v}: W\mathbf{v}=\mathbf{y}} \|\mathbf{v}\|_0 \ = \ \operatorname*{argmin}_{\mathbf{v}: W\mathbf{v}=\mathbf{y}} \|\mathbf{v}\|_1 \ .$$

In fact, we will prove an even stronger result, which holds even if $\mathbf{x}$ is not a sparse vector.

Theorem 3 *Let $\epsilon < \frac{1}{1+\sqrt{2}}$ and let W be a $(\epsilon, 2s)$-RIP matrix. Let $\mathbf{x}$ be an arbitrary vector and denote*

$$\mathbf{x}_s \in \operatorname*{argmin}_{\mathbf{v}: \|\mathbf{v}\|_0 \leq s} \|\mathbf{x} - \mathbf{v}\|_1 \ .$$

That is, $\mathbf{x}_s$ is the vector which equals $\mathbf{x}$ on the s largest elements of $\mathbf{x}$ and equals 0 elsewhere. Let $\mathbf{y} = W\mathbf{x}$ be the compression of $\mathbf{x}$ and let

$$\mathbf{x}^\star \in \operatorname*{argmin}_{\mathbf{v}: W\mathbf{v}=\mathbf{y}} \|\mathbf{v}\|_1$$

be the reconstructed vector. Then,

$$\|\mathbf{x}^\star - \mathbf{x}\|_2 \leq 2(1-\rho)^{-1} s^{-1/2} \|\mathbf{x} - \mathbf{x}_s\|_1 \ ,$$

where $\rho = \sqrt{2}\epsilon/(1-\epsilon)$.

Note that in the special case that $\mathbf{x} = \mathbf{x}_s$ we get an exact recovery, $\mathbf{x}^\star = \mathbf{x}$, so Theorem 2 is a special case of Theorem 3. The proof of Theorem 3 is given in Section 3.1.

Finally, the last theorem tells us that random matrices with $n \geq \Omega(n \log(d))$ are likely to be RIP. In fact, the theorem shows that multiplying a random matrix by an orthonormal matrix also provides an RIP matrix. This is important for compressing signals of the form $\mathbf{x} = U\boldsymbol{\alpha}$ where $\mathbf{x}$ is not sparse but $\boldsymbol{\alpha}$ is sparse. In that case, if W is a random matrix and we compress using $\mathbf{y} = W\mathbf{x}$ then this is the same as compressing $\boldsymbol{\alpha}$ by $\mathbf{y} = (WU)\boldsymbol{\alpha}$ and since WU is also RIP we can reconstruct $\boldsymbol{\alpha}$ (and thus also $\mathbf{x}$) from $\mathbf{y}$.

Theorem 4 *Let U be an arbitrary fixed $d \times d$ orthonormal matrix, let ϵ, δ be scalars in $(0,1)$, let s be an integer in $[d]$, and let n be an integer that satisfies*

$$n \geq 100 \, \frac{s \ln(40d/(\delta\,\epsilon))}{\epsilon^2} \, .$$

Let $W \in \mathbb{R}^{n,d}$ be a matrix s.t. each element of W is distributed normally with zero mean and variance of $1/n$. Then, with proabability of at least $1-\delta$ over the choice of W, the matrix WU is (ϵ, s)-RIP.

The proof of Theorem 4 is given in Section 3.2.

3 Proofs

3.1 Proof of Theorem 3

We follow a proof due to [Candes, "The restricted isometry property and its implications for compressed sensing"].

Notation : Given a vector $\mathbf{v}$ and a set of indices I we denote by $\mathbf{v}_I$ the vector whose ith element is v_i if $i \in I$ and otherwise its ith element is 0. Let $\mathbf{h} = \mathbf{x}^\star - \mathbf{x}$.

The first trick we use is to partition the set of indices $[d] = \{1, \ldots, d\}$ into disjoint sets of size s. That is, we will write $[d] = T_0 \cup T_1 \cup T_2 \ldots T_{d/s-1}$ where for all i, $|T_i| = s$, and we assume for simplicity that d/s is an integer. We define the partition as follows. In T_0 we put the s indices corresponding to the s largest elements in absolute values of $\mathbf{x}$ (ties are braked arbitrarily). Let $T_0^c = [d] \setminus T_0$. Next, T_1 will be the s indices corresponding to the s largest elements in absolute value of $\mathbf{h}_{T_0^c}$. Let $T_{0,1} = T_0 \cup T_1$ and $T_{0,1}^c = [d] \setminus T_{0,1}$. Next, T_2 will correspond to the s largest elements in absolute value of $\mathbf{h}_{T_{0,1}^c}$. And, we will construct $T_3, T_4, \ldots$ using the same way.

To prove the theorem we first need the following lemma which shows that RIP also implies approximate orthogonality.

Lemma 1 *Let W be an (ϵ, s)-RIP matrix. Then, for any two disjoint sets I, J, both of size at most s, and for any vector $\mathbf{u}$ we have that $\langle W\mathbf{u}_I, W\mathbf{u}_J \rangle \leq \epsilon \|\mathbf{u}_I\| \, \|\mathbf{u}_J\|$.*

Proof W.l.o.g. assume $\|\mathbf{u}_I\| = \|\mathbf{u}_J\| = 1$.

$$\langle W\mathbf{u}_I, W_J\mathbf{u} \rangle = \frac{\|W\mathbf{u}_I + W\mathbf{u}_J\|^2 - \|W\mathbf{u}_I - W\mathbf{u}_J\|^2}{4} \, .$$

But, since $|J \cup I| \leq 2s$ we get from the RIP condition that $\|W\mathbf{u}_I + W\mathbf{u}_J\|^2 \leq (1+\epsilon)(\|\mathbf{u}_I\|^2 + \|\mathbf{u}_J\|^2) = 2(1+\epsilon)$ and that $-\|W\mathbf{u}_I - W\mathbf{u}_J\|^2 \leq -(1-\epsilon)(\|\mathbf{u}_I\|^2 + \|\mathbf{u}_J\|^2) = -2(1-\epsilon)$, which concludes our proof. ■

We are now ready to proving the theorem. Clearly,

$$\|\mathbf{h}\|_2 = \|\mathbf{h}_{T_{0,1}} + \mathbf{h}_{T_{0,1}^c}\|_2 \leq \|\mathbf{h}_{T_{0,1}}\|_2 + \|\mathbf{h}_{T_{0,1}^c}\|_2 \, . \tag{1}$$

To prove the theorem we will show the following two claims:

Claim 1: $\|\mathbf{h}_{T_{0,1}^c}\|_2 \leq \|\mathbf{h}_{T_0}\|_2 + 2s^{-1/2}\|\mathbf{x} - \mathbf{x}_s\|_1$.

Claim 2: $\|\mathbf{h}_{T_{0,1}}\|_2 \leq \frac{\rho}{1-\rho} s^{-1/2}\|\mathbf{x} - \mathbf{x}_s\|_1$

Combining these two claims with Eq. (1) we get that

$$\begin{aligned}\|\mathbf{h}\|_2 &\leq \|\mathbf{h}_{T_{0,1}}\|_2 + \|\mathbf{h}_{T_{0,1}^c}\|_2 \leq 2\|\mathbf{h}_{T_{0,1}}\|_2 + 2s^{-1/2}\|\mathbf{x} - \mathbf{x}_s\|_1 \\ &\leq 2\left(\tfrac{\rho}{1-\rho} + 1\right) s^{-1/2}\|\mathbf{x} - \mathbf{x}_s\|_1 \\ &= 2(1-\rho)^{-1} s^{-1/2}\|\mathbf{x} - \mathbf{x}_s\|_1 \ ,\end{aligned}$$

and this will conclude our proof.

Proving claim 1: To prove this claim we do not use the RIP condition at all but only use the fact that $\mathbf{x}^\star$ minimizes the ℓ_1 norm. Take $j > 1$. For each $i \in T_j$ and $i' \in T_{j-1}$ we have that $|h_i| \leq |h_{i'}|$. Therefore, $\|\mathbf{h}_{T_j}\|_\infty \leq \|\mathbf{h}_{T_{j-1}}\|_1 / s$. Thus,

$$\|\mathbf{h}_{T_j}\|_2 \leq s^{1/2}\|\mathbf{h}_{T_j}\|_\infty \leq s^{-1/2}\|\mathbf{h}_{T_{j-1}}\|_1 \ .$$

Summing the above over $j = 2, 3, \ldots$ and using the triangle inequality we obtain that

$$\|\mathbf{h}_{T_{0,1}^c}\|_2 \leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2 \leq s^{-1/2}\|\mathbf{h}_{T_0^c}\|_1 \ . \tag{2}$$

Next, we show that $\|\mathbf{h}_{T_0^c}\|_1$ cannot be large. Indeed, since $\mathbf{x}^\star = \mathbf{x} + \mathbf{h}$ has minimal ℓ_1 norm and since $\mathbf{x}$ satisfies the constraint in the definition of $\mathbf{x}^\star$ we have that $\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{h}\|_1$. Thus, using the triangle inequality we obtain that

$$\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{h}\|_1 = \sum_{i \in T_0} |x_i + h_i| + \sum_{i \in T_0^c} |x_i + h_i| \geq \|\mathbf{x}_{T_0}\|_1 - \|\mathbf{h}_{T_0}\|_1 + \|\mathbf{h}_{T_0^c}\|_1 - \|\mathbf{x}_{T_0^c}\|_1 \ . \tag{3}$$

and since $\|\mathbf{x}_{T_0^c}\|_1 = \|\mathbf{x} - \mathbf{x}_s\|_1 = \|\mathbf{x}\|_1 - \|\mathbf{x}_{T_0}\|_1$ we get that

$$\|\mathbf{h}_{T_0^c}\|_1 \leq \|\mathbf{h}_{T_0}\|_1 + 2\|\mathbf{x}_{T_0^c}\|_1 \ . \tag{4}$$

Combining the above with Eq. (2) we get that

$$\|\mathbf{h}_{T_{0,1}^c}\|_2 \leq s^{-1/2}\left(\|\mathbf{h}_{T_0}\|_1 + 2\|\mathbf{x}_{T_0^c}\|_1\right) \leq \|\mathbf{h}_{T_0}\|_2 + 2s^{-1/2}\|\mathbf{x}_{T_0^c}\|_1 \ ,$$

which concludes the proof of claim 1.

Proving claim 2: For the second claim we use the RIP condition to get that

$$(1-\epsilon)\|\mathbf{h}_{T_{0,1}}\|_2^2 \leq \|W\mathbf{h}_{T_{0,1}}\|_2^2 \ . \tag{5}$$

Since $W\mathbf{h}_{T_{0,1}} = W\mathbf{h} - \sum_{j \geq 2} W\mathbf{h}_{T_j} = -\sum_{j \geq 2} W\mathbf{h}_{T_j}$ we have that

$$\|W\mathbf{h}_{T_{0,1}}\|_2^2 = -\sum_{j \geq 2} \langle W\mathbf{h}_{T_{0,1}}, W\mathbf{h}_{T_j}\rangle = -\sum_{j \geq 2} \langle W\mathbf{h}_{T_0} + W\mathbf{h}_{T_1}, W\mathbf{h}_{T_j}\rangle \ .$$

From the RIP condition on inner products we obtain that for all $i \in \{1, 2\}$ and $j \geq 2$ we have

$$|\langle W\mathbf{h}_{T_i}, W\mathbf{h}_{T_j}\rangle| \leq \epsilon\|\mathbf{h}_{T_i}\|_2\|\mathbf{h}_{T_j}\|_2 \ .$$

Since $\|\mathbf{h}_{T_0}\|_2 + \|\mathbf{h}_{T_1}\|_2 \leq \sqrt{2}\|\mathbf{h}_{T_{0,1}}\|_2$ we therefore get that

$$\|W\mathbf{h}_{T_{0,1}}\|_2^2 \;\leq\; \sqrt{2}\epsilon\|\mathbf{h}_{T_{0,1}}\|_2 \sum_{j\geq 2}\|\mathbf{h}_{T_j}\|_2 \;.$$

Combining the above with Eq. (2) and Eq. (5) we obtain

$$(1-\epsilon)\|\mathbf{h}_{T_{0,1}}\|_2^2 \leq \sqrt{2}\epsilon\|\mathbf{h}_{T_{0,1}}\|_2 s^{-1/2}\|\mathbf{h}_{T_0^c}\|_1 \;.$$

Rearranging the above gives

$$\|\mathbf{h}_{T_{0,1}}\|_2 \leq \frac{\sqrt{2}\epsilon}{1-\epsilon} s^{-1/2}\|\mathbf{h}_{T_0^c}\|_1 \;.$$

Finally, using Eq. (4) we get that

$$\|\mathbf{h}_{T_{0,1}}\|_2 \leq \rho s^{-1/2}\left(\|\mathbf{h}_{T_0}\|_1 + 2\|\mathbf{x}_{T_0^c}\|_1\right) \leq \rho\|\mathbf{h}_{T_0}\|_2 + 2\rho s^{-1/2}\|\mathbf{x}_{T_0^c}\|_1 \;,$$

but since $\|\mathbf{h}_{T_0}\|_2 \leq \|\mathbf{h}_{T_{0,1}}\|_2$ this implies

$$\|\mathbf{h}_{T_{0,1}}\|_2 \leq \frac{\rho}{1-\rho}\, s^{-1/2}\|\mathbf{x}_{T_0^c}\|_1 \;,$$

which concludes the proof of the second claim.

3.2 Proof of Theorem 4

To prove the theorem we follow the approach of Baraniuk, Davenport, DeVore, and Wakin, "A simple proof of the RIP for random matrices". The idea is to combine Johnson-Lindenstrauss (JL) lemma with a simple covering argument. For completeness, we provide JL lemma and its proof in Section 3.3 below.

We start with a covering property of the unit ball.

Lemma 2 *Let $\epsilon \in (0,1)$. There exists a finite set $Q \subset \mathbb{R}^d$ of size $|Q| \leq \left(\frac{5}{\epsilon}\right)^d$ such that*

$$\sup_{\mathbf{x}:\|\mathbf{x}\|\leq 1} \min_{\mathbf{v}\in Q} \|\mathbf{x}-\mathbf{v}\| \;\leq\; \epsilon \;.$$

Proof Let k be an integer and let

$$Q' = \{\mathbf{x}\in\mathbb{R}^d : \forall j, \exists i \in \{-k, -k+1, \ldots, k\} \text{ s.t. } x_j = \tfrac{i}{k}\} \;.$$

Clearly, $|Q'| = (2k+1)^d$. We shall set $Q = Q' \cap B_2(1)$, where $B_2(1)$ is the unit L_2 ball of $\mathbb{R}^d$. Since the points in Q' are distributed evenly on the unit cube, the size of Q is the size of Q' times the ratio between the volumes of the unit L_2 ball and the unit cube. The volume of the unit cube is 1 and the volume of $B_2(1)$ is

$$\frac{\pi^{d/2}}{\Gamma(1+d/2)} \;.$$

For simplicity, assume that d is even and therefore

$$\Gamma(1+d/2) \;=\; (d/2)! \;\geq\; \left(\frac{d/2}{e}\right)^{d/2} \;,$$

where in the last inequality we used Stirling's approximation. Overall we obtained that

$$|Q| \leq (2k+1)^d\,(\pi/e)^{d/2}\,(d/2)^{-d/2} \;. \tag{6}$$

Now lets specify k. For each $\mathbf{x} \in B_2(1)$ let $\mathbf{v} \in Q$ be the vector whose ith element is $\text{sign}(x_i)\, \lfloor |x_i|\, k \rfloor$. Then, for each element we have that $|x_i - v_i| \leq 1/k$ and thus

$$\|\mathbf{x} - \mathbf{v}\| \leq \frac{\sqrt{d}}{k} \,.$$

To ensure that the right-hand side of the above will be at most ϵ we shall set $k = \lceil \sqrt{d}/\epsilon \rceil$. Plugging this value into Eq. (6) we conclude that

$$|Q| \leq (3\sqrt{d}/\epsilon)^d \, (\pi/e)^{d/2} \, (d/2)^{-d/2} = \left(\tfrac{3}{\epsilon} \sqrt{\tfrac{2\pi}{e}} \right)^d \leq \left(\tfrac{5}{\epsilon} \right)^d \,.$$

■

Let $\mathbf{x}$ be a vector that can be written as $\mathbf{x} = U\boldsymbol{\alpha}$ with U being some orthonormal matrix and $\|\boldsymbol{\alpha}\|_0 \leq s$. Combining the covering property above and the JL lemma (Lemma 5) enables us to show that a random W will not distort any such $\mathbf{x}$.

Lemma 3 *Let U be an orthonormal $d \times d$ matrix and let $I \subset [d]$ be a set of indices of size $|I| = s$. Let S be the span of $\{U_i : i \in I\}$, where U_i is the ith column of U. Let $\delta \in (0,1)$, $\epsilon \in (0,1)$, and n be an integer such that*

$$n \geq 24 \, \frac{\ln(2/\delta) + s \, \ln(20/\epsilon)}{\epsilon^2} \,.$$

Then, with probability of at least $1 - \delta$ over a choice of a random matrix $W \in \mathbb{R}^{n,d}$ such that each element of W is independently distributed according to $N(0, 1/n)$ we have

$$\sup_{\mathbf{x} \in S} \left| \frac{\|W\mathbf{x}\|}{\|\mathbf{x}\|} - 1 \right| < \epsilon \,.$$

Proof It suffices to prove the lemma for all $\mathbf{x} \in S$ of unit norm. We can write $\mathbf{x} = U_I \boldsymbol{\alpha}$ where $\boldsymbol{\alpha} \in \mathbb{R}^s$, $\|\boldsymbol{\alpha}\|_2 = 1$, and U_I is the matrix whose columns are $\{U_i : i \in I\}$. Using Lemma 2 we know that there exists a set Q of size $|Q| \leq (20/\epsilon)^s$ such that

$$\sup_{\boldsymbol{\alpha} : \|\boldsymbol{\alpha}\| = 1} \min_{\mathbf{v} \in Q} \|\boldsymbol{\alpha} - \mathbf{v}\| \leq (\epsilon/4) \,.$$

But, since U is orthogonal we also have that

$$\sup_{\boldsymbol{\alpha} : \|\boldsymbol{\alpha}\| = 1} \min_{\mathbf{v} \in Q} \|U_I\boldsymbol{\alpha} - U_I\mathbf{v}\| \leq (\epsilon/4) \,.$$

Applying Lemma 5 on the set $\{U_I\mathbf{v} : \mathbf{v} \in Q\}$ we obtain that for n satisfying the condition given in the lemma, the following holds with probability of at least $1 - \delta$:

$$\sup_{\mathbf{v} \in Q} \left| \frac{\|WU_I\mathbf{v}\|^2}{\|U_I\mathbf{v}\|^2} - 1 \right| \leq \epsilon/2 \,,$$

This also implies that

$$\sup_{\mathbf{v} \in Q} \left| \frac{\|WU_I\mathbf{v}\|}{\|U_I\mathbf{v}\|} - 1 \right| \leq \epsilon/2 \,.$$

Let a be the smallest number such that

$$\forall \mathbf{x} \in S, \;\; \frac{\|W\mathbf{x}\|}{\|\mathbf{x}\|} \leq 1 + a \,.$$

Clearly $a < \infty$. Our goal is to show that $a \leq \epsilon$. This follows from the fact that for any $\mathbf{x} \in S$ of unit norm there exists $\mathbf{v} \in Q$ such that $\|\mathbf{x} - U_I \mathbf{v}\| \leq \epsilon/4$ and therefore

$$\|W\mathbf{x}\| \leq \|WU_I\mathbf{v}\| + \|W(\mathbf{x} - U_I\mathbf{v})\| \leq 1 + \epsilon/2 + (1+a)\epsilon/4 \,.$$

Thus,

$$\forall \mathbf{x} \in S, \quad \frac{\|W\mathbf{x}\|}{\|\mathbf{x}\|} \leq 1 + (\epsilon/2 + (1+a)\epsilon/4) \;.$$

But, the definition of a implies that

$$a \leq \epsilon/2 + (1+a)\epsilon/4 \;\;\Rightarrow\;\; a \leq \frac{\epsilon/2 + \epsilon/4}{1 - \epsilon/4} \leq \epsilon \,.$$

This proves that for all $\mathbf{x} \in S$ we have $\frac{\|W\mathbf{x}\|}{\|\mathbf{x}\|} - 1 \leq \epsilon$. The other side follows from this as well since

$$\|W\mathbf{x}\| \geq \|WU_I\mathbf{v}\| - \|W(\mathbf{x} - U_I\mathbf{v})\| \geq 1 - \epsilon/2 - (1+\epsilon)\epsilon/4 \geq 1 - \epsilon \,.$$

■

The above lemma tells us that for $\mathbf{x} \in S$ of unit norm we have

$$(1 - \epsilon) \leq \|W\mathbf{x}\| \leq (1 + \epsilon) \,,$$

which implies that

$$(1 - 2\,\epsilon) \leq \|W\mathbf{x}\|^2 \leq (1 + 2\,\epsilon) \,.$$

The proof of Theorem 4 follows from the above by a union bound over all choices of I.

3.3 Random Projections and Johnson-Lindenstrauss lemma

We provide a variant of a famous lemma due to Johnson and Lindenstrauss, showing that random projections do not distort Euclidean distances too much.We start with analyzing the distortion caused by applying a random projection on a single vector.

Lemma 4 *Fix some $\mathbf{x} \in \mathbb{R}^d$. Let $W \in \mathbb{R}^{n,d}$ be a random matrix such that each $W_{i,j}$ is an independent normal random variable. Then, for any $\epsilon \in (0,3)$ we have*

$$\mathbb{P}\left[\left|\frac{\|(1/\sqrt{n})W\mathbf{x}\|^2}{\|\mathbf{x}\|^2} - 1\right| > \epsilon\right] \;\leq\; 2\,e^{-\epsilon^2 n/6} \,.$$

Proof Without loss of generality we can assume that $\|\mathbf{x}\|^2 = 1$. Therefore, an equivalent inequality is

$$\mathbb{P}\left[(1-\epsilon)n \leq \|W\mathbf{x}\|^2 \leq (1+\epsilon)n\right] \geq 1 - 2e^{-\epsilon^2 n/6} \,.$$

Let $\mathbf{z}_i$ be the ith row of W. The random variable $\langle \mathbf{z}_i, \mathbf{x}\rangle$ is a weighted sum of d independent normal random variables and therefore it is normally distributed with zero mean and variance $\sum_j x_j^2 = \|\mathbf{x}\|^2 = 1$. Therefore, the random variable $\|W\mathbf{x}\|^2 = \sum_{i=1}^n (\langle \mathbf{z}_i, \mathbf{x}\rangle)^2$ has a χ^2_n distribution. The claim now follows directly from a measure concentration property of χ^2 random variables stated in Lemma 6 in Section 3.3.1 below. ■

The Johnson-Lindenstrauss lemma follows from the above using a simple union bound argument.

Lemma 5 (Johnson-Lindenstrauss lemma) *Let Q be a finite set of vectors in $\mathbb{R}^d$. Let $\delta \in (0,1)$ and n be an integer such that*

$$\epsilon = \sqrt{\frac{6\,\ln(2|Q|/\delta)}{n}} \leq 3\,.$$

Then, with probability of at least $1-\delta$ over a choice of a random matrix $W \in \mathbb{R}^{n,d}$ such that each element of W is independently distributed according to $N(0,1/n)$ we have

$$\sup_{\mathbf{x}\in Q}\left|\frac{\|W\mathbf{x}\|^2}{\|\mathbf{x}\|^2}-1\right| < \epsilon\,.$$

Proof Using Lemma 4 and a union bound we have that for all $\epsilon \in (0,3)$:

$$\mathbb{P}\left[\sup_{\mathbf{x}\in Q}\left|\frac{\|W\mathbf{x}\|^2}{\|\mathbf{x}\|^2}-1\right| > \epsilon\right] \leq 2\,|Q|\,e^{-\epsilon^2 n/6}\,.$$

Let δ denote the right-hand side of the above and solve for ϵ we obtain that $\epsilon = \sqrt{\frac{6\,\ln(2|Q|/\delta)}{n}}$. ■

3.3.1 Concentration of χ^2 variables

Let $X_1,\ldots,X_k$ be k independent normally distributed random variables. That is, for all i, $X_i \sim N(0,1)$. The distribution of the random variable X_i^2 is called χ^2 (chi square) and the distribution of the random variable $Z = X_1^2+\ldots+X_k^2$ is called χ_k^2 (chi square with k degrees of freedom). Clearly, $\mathbb{E}[X_i^2]=1$ and $\mathbb{E}[Z]=k$. The following lemma states that X_k^2 is concentrated around its mean.

Lemma 6 *Let $Z \sim \chi_k^2$. Then, for all $\epsilon > 0$ we have*

$$\mathbb{P}[Z \leq (1-\epsilon)k)] \leq e^{-\epsilon^2 k/6}\,,$$

and for all $\epsilon \in (0,3)$ we have

$$\mathbb{P}[Z \geq (1+\epsilon)k)] \leq e^{-\epsilon^2 k/6}\,.$$

Finally, for all $\epsilon \in (0,3)$,

$$\mathbb{P}\left[(1-\epsilon)k \leq Z \leq (1+\epsilon)k\right] \geq 1-2e^{-\epsilon^2 k/6}\,.$$

Proof Let us write $Z=\sum_{i=1}^k X_i^2$ where $X_i \sim N(0,1)$. To prove both bounds we use Chernoff's bounding method. For the first inequality, we first bound $\mathbb{E}[e^{-\lambda X_1^2}]$, where $\lambda > 0$ will be specified later. Since $e^{-a} \leq 1-a+\frac{a^2}{2}$ for all $a \geq 0$ we have that

$$\mathbb{E}[e^{-\lambda X_1^2}] \;\leq\; 1-\lambda\,\mathbb{E}[X_1^2]+\frac{\lambda^2}{2}\,\mathbb{E}[X_1^4]\,.$$

Using the well known equalities, $\mathbb{E}[X_1^2]=1$ and $\mathbb{E}[X_1^4] \;=\; 3$, and the fact that $1-a \leq e^{-a}$ we obtain that

$$\mathbb{E}[e^{-\lambda X_1^2}] \;\leq\; 1-\lambda+\frac{3}{2}\lambda^2 \leq e^{-\lambda+\frac{3}{2}\lambda^2}\,.$$

Now, applying Chernoff's bounding method we get that

$$\begin{aligned}
\mathbb{P}[-Z \geq -(1-\epsilon)k)] &= \mathbb{P}\left[e^{-\lambda Z} \geq e^{-(1-\epsilon)k\lambda}\right] && (7)\\
&\leq e^{(1-\epsilon)k\lambda}\;\mathbb{E}\left[e^{-\lambda Z}\right] && (8)\\
&= e^{(1-\epsilon)k\lambda}\;\left(\mathbb{E}\left[e^{-\lambda X_1^2}\right]\right)^k && (9)\\
&\leq e^{(1-\epsilon)k\lambda}\,e^{-\lambda k+\frac{3}{2}\lambda^2 k} && (10)\\
&= e^{-\epsilon k\lambda+\frac{3}{2}k\lambda^2}\,. && (11)
\end{aligned}$$

Choose $\lambda = \epsilon/3$ we obtain the first inequality stated in the lemma.

For the second inequality, we use a known closed form expression for the moment generating function of a χ^2_k distributed random variable:

$$\forall \lambda < \tfrac{1}{2}, \;\; \mathbb{E}\left[e^{\lambda Z^2}\right] \;=\; (1-2\lambda)^{-k/2}\,. \tag{12}$$

Based on the above and using Chernoff's bounding method we have:

$$\begin{aligned}
\mathbb{P}[Z \geq (1+\epsilon)k)] &= \mathbb{P}\left[e^{\lambda Z} \geq e^{(1+\epsilon)k\lambda}\right] && (13)\\
&\leq e^{-(1+\epsilon)k\lambda}\,\mathbb{E}\left[e^{\lambda Z}\right] && (14)\\
&= e^{-(1+\epsilon)k\lambda}\,(1-2\lambda)^{-k/2} && (15)\\
&\leq e^{-(1+\epsilon)k\lambda}\,e^{k\lambda} = e^{-\epsilon k\lambda}\,, && (16)
\end{aligned}$$

where the last inequality is because $(1-a) \leq e^{-a}$. Setting $\lambda = \epsilon/6$ (which is in $(0, 1/2)$ by our assumption) we obtain the second inequality stated in the lemma.

Finally, the last inequality follows from the first two inequalities and the union bound. ■