Mathematical Methods for Machine Learning and Signal Processing SS 2019

Lectures 14 – 17: Compressed Sensing – Randomized sampling

V. Morgenshtern (compiled from notes by H. Bölcskei)

Agenda:

1. Sampling at information rate
2. How to choose the measurement matrix?
3. The Restricted Isometry Property (RIP) and signal recovery
4. The Johnson-Lindenstrauss Lemma
5. Verifying the RIP from concentration inequalities

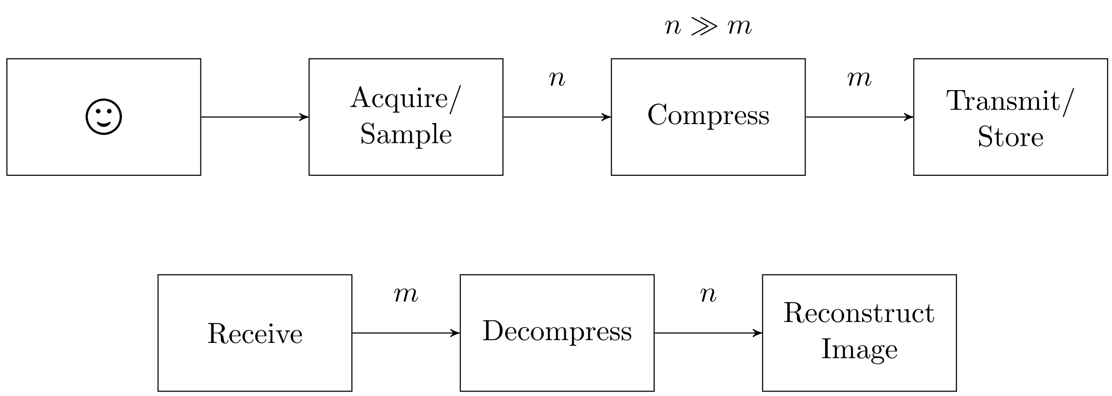
1 Sampling at information rate

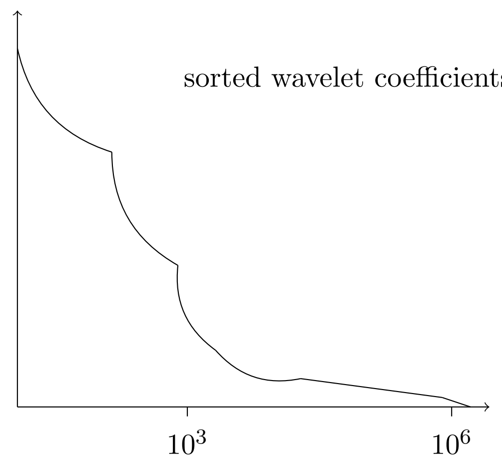
In signal separation problems we did not have control over designing the dictionary $\mathbf{D}$. The dictionary was determined by the nature of separation problem under consideration.

In super-resolution problems we also did not have control over the dictionary $\mathbf{D}$. The dictionary was the low-frequency part of the Fourier matrix.

As we will see, in Compressed Sensing we have a lot of control on how we design $\mathbf{D}$. This will allow us to break the square-root bottleneck and also the non-negativity constraints as in super-resolution will not be necessary.

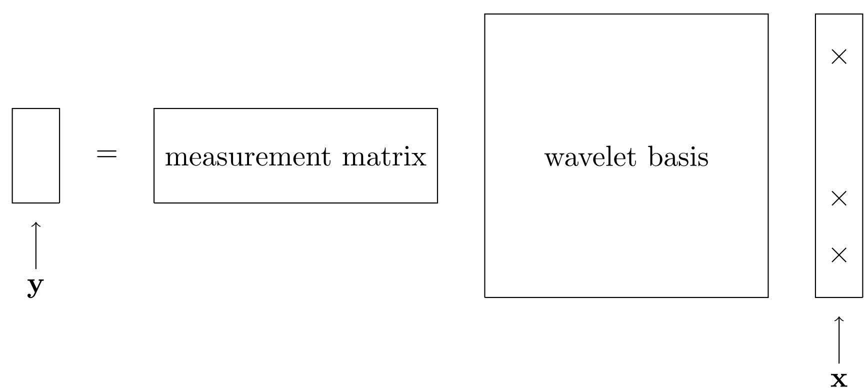
Consider a typical sensing pipeline:



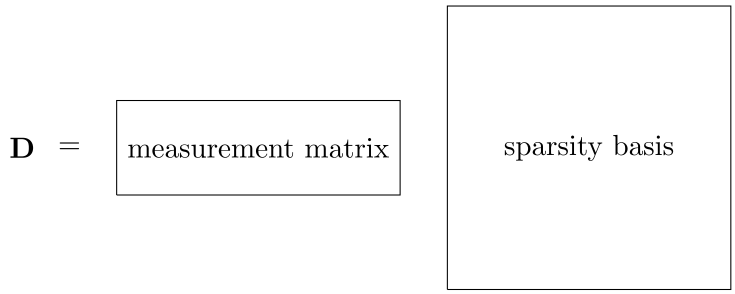


Typically, we obtain a graph similar to the figure above when plotting the amplitude of the sorted wavelet coefficients when we sample natural images. Hence, $10^3 - 10^6$ of the costly acquired wavelet coefficients are thrown away in the process of compression. It is therefore important to ask whether we cannot just acquire the information that will not end up being thrown away.

In many problems of practical importance we can perform *generalized measurement*. One generalized measurement is simply an inner product between the signal we try to recover and a (known) measurement vector. If we stack the measurement vectors into rows of a matrix, we obtain the *measurement matrix*. We would like to take fewer measurements than the dimensionality of the signal. For example, if the signal is sparse in the wavelet domain, we would like the number of measurements to be proportional to the sparsity level of the signal. In other words the generalized measurement matrix is fat. The generalized measurement pipeline is structured as follows:



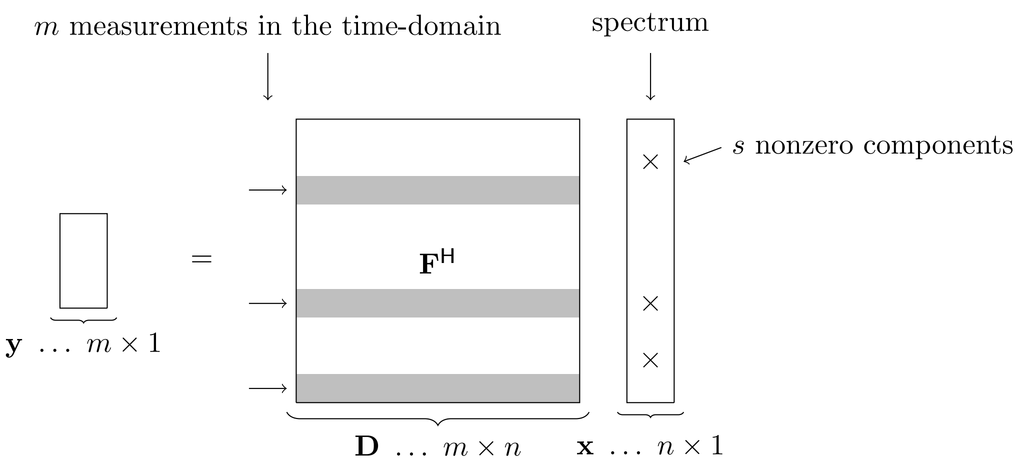
To reduce the generalized measurement pipeline to the setup we studied so far, define:



For a given sparsity basis (e.g., wavelets), we want to find a sampling basis such that the resulting matrix $\mathbf{D}$ admits successful recovery of all s-sparse vectors.

2 How to choose the measurement matrix?

Consider the example of spectrally sparse signals.

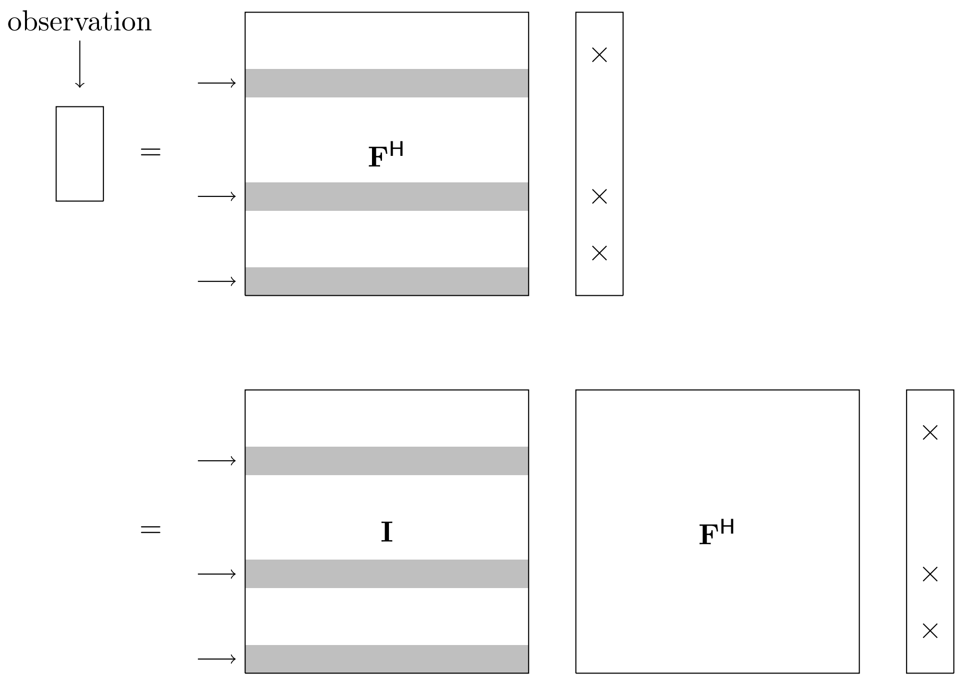


In the Figure above

$$\mathbf{F} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)^2} \end{bmatrix}$$

is the $n \times n$ DFT matrix and $\omega_n = e^{-2\pi \mathrm{i}/n}$.

In terms of the generalized measurement pipeline, this example can be rewritten as follows:

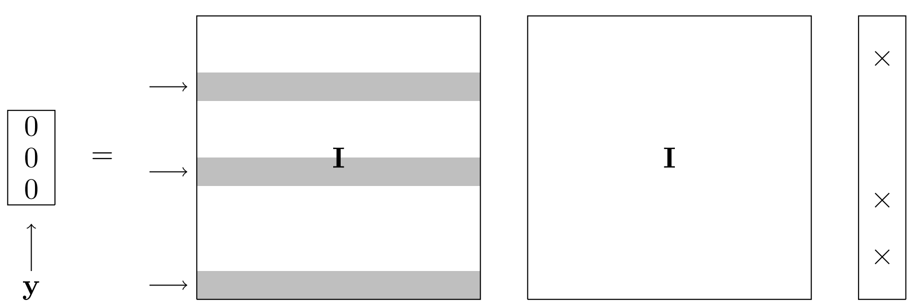


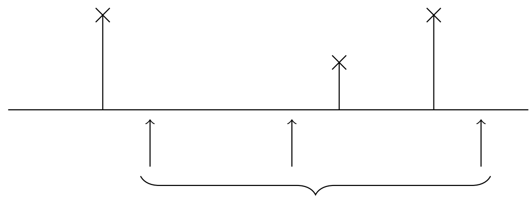
Above, the matrix $\mathbf{F}^{\mathsf{H}}$ is the *sparsity* basis. In this basis the signal has a sparse representation. The identity matrix with removed columns is the *measurement* matrix.

We can apply the following universal sampling pattern. Choose the first $m = 2s$ rows of $\mathbf{F}$. In principle, $\mathbf{x}$ is now uniquely determined by measurements $\mathbf{y}$. To see this, assume that there are two s-sparse vectors $\mathbf{x}_1 \neq \mathbf{x}_2$ that satisfy $\mathbf{D}\mathbf{x}_1 = \mathbf{D}\mathbf{x}_2$. Therefore, $\|\mathbf{D}\left(\mathbf{x}_1 - \mathbf{x}_2\right)\|^2 = 0$. Let $\mathcal{S}_i$ denote the support of $\mathbf{x}_i$, $i = 1, 2$ and $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$. The matrix $\mathbf{D}_{\mathcal{S}}$ is a square matrix Vandermonde matrix. Therefore it is fullrank with rank $2s$. Thus, it is invertible and we conclude that $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}$. So that $\mathbf{x}_1 = \mathbf{x}_2$. Note that we did not provide a recovery algorithm, but we proved that the recovery is possible in principle.

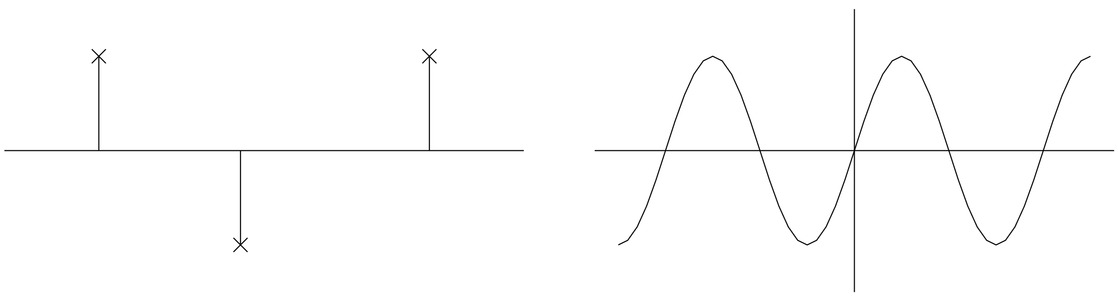
Would this work for every choice of sparsity basis and sampling matrix?

Since we require the compressive sensing scheme to be *universal*, recovery must be possible independently of the s-sparse signal vector $\mathbf{x}$. In the example depicted below, this is clearly not the case.





Spikes and sinusoids are *incoherent*. Spikes and spikes are not *incoherent*.



We conclude that if the measurement matrix and the sparsity basis are sufficiently incoherent, the resulting matrix $\mathbf{D}$ might have low coherence, $\mu(\mathbf{D})$. Hence, we might be able to use the theory we derived earlier to get a recovery guarantee for all s-sparse signals. However, as we discussed, the coherence-based deterministic bounds can only provide recovery guarantee up to the sparsity level $s \lesssim \sqrt{m}$, which is very pessimistic. In the special case of super-resolution of positive signals we have seen that recovery is possible for $s \propto m$. Next we study the signal recovery theory based on the resticted isometry property (RIP). This theory allows us to obtain recovery guarantees for $s \propto m$ for a large class of dictionaries, $\mathbf{D}$, in which the sampling basis is random or partially random.

3 The RIP and signal recovery

We present the theory developed by E. J. Candès [1]. Assume that we observe $\mathbf{y} = \mathbf{D}\mathbf{x} \in \mathbb{R}^m$, where $\mathbf{x} \in \mathbb{R}^n$ is a signal, unknown to us and that we want to reconstruct, and $\mathbf{D} \in \mathbb{C}^{m \times n}$ is a known measurement matrix. Here, we consider the underdetermined case with fewer equations than unknowns, i.e., $m < n$.

Definition 1. For each integer $s = 1, 2, \ldots$, define the isometry constant δ_s of a matrix $\mathbf{D}$ as the smallest number such that

$$(1 - \delta_s) \|\mathbf{x}\|_2^2 \leq \|\mathbf{D}\mathbf{x}\|_2^2 \leq (1 + \delta_s) \|\mathbf{x}\|_2^2$$

holds for all s-sparse vectors $\mathbf{x}$. A vector is said to be s-sparse if it has at most s nonzero entries.

We denote the best s-sparse approximation to an arbitrary vector $\mathbf{x}$ by $\mathbf{x}_s$, i.e. $\mathbf{x}_s$ is the vector $\mathbf{x}$ with all but the s-larges entries (in absolute value) set to zero.

Theorem 1 (Noiseless recovery)**.** Assume that $\delta_{2s} < \sqrt{2} - 1$. Then, the solution $\mathbf{x}^*$ to

$$\text{find } \arg\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{D}\hat{\mathbf{x}} = \mathbf{y}$$

obeys

$$\|\mathbf{x}^* - \mathbf{x}\|_1 \leq C_0 \|\mathbf{x} - \mathbf{x}_s\|_1$$

and

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0 s^{-1/2} \|\mathbf{x} - \mathbf{x}_s\|_1$$

for some constant C_0 given explicitly below. In particular, if $\mathbf{x}$ is s-sparse, recovery is exact.

Theorem 2 (Noisy recovery)**.** Assume that $\delta_{2s} < \sqrt{2} - 1$ and $\|\mathbf{n}\|_2 \leq \varepsilon$. Then, the solution $\mathbf{x}^*$ to

$$\text{find } \arg \min_{\hat{\mathbf{x}} \in \mathbb{R}^n} \|\hat{\mathbf{x}}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{D}\hat{\mathbf{x}}\|_2 \leq \varepsilon$$

obeys

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0 s^{-1/2} \|\mathbf{x} - \mathbf{x}_s\|_1 + C_1 \varepsilon$$

with the same constant C_0 as before and some constant C_1 given explicitly below.

Proofs

Lemma 3. We have

$$\left|\langle \mathbf{D}\mathbf{v}, \mathbf{D}\mathbf{v}' \rangle\right| \leq \delta_{s+s'} \|\mathbf{v}\|_2 \|\mathbf{v}'\|_2$$

for all $\mathbf{v}$, $\mathbf{v}'$ supported on disjoint subsets $\mathcal{T}, \mathcal{T}' \subseteq \{1, \ldots, n\}$ with $|\mathcal{T}| \leq s$ and $|\mathcal{T}'| \leq s'$.

Proof. Without loss of generality, let us assume that $\|\mathbf{v}\|_2 = \|\mathbf{v}'\|_2 = 1$. By definition of the restricted isometry constant $\delta_{s+s'}$, it holds that

$$\|\mathbf{v} \pm \mathbf{v}'\|_2^2 (1 - \delta_{s+s'}) \leq \|\mathbf{D}(\mathbf{v} \pm \mathbf{v}')\|_2^2 \leq \|\mathbf{v} \pm \mathbf{v}'\|_2^2 (1 + \delta_{s+s'}).$$

Since $\mathbf{v}$ and $\mathbf{v}'$ are disjointly supported and $\|\mathbf{v}\|_2 = \|\mathbf{v}'\|_2 = 1$ by assumption, we have

$$\|\mathbf{v} \pm \mathbf{v}'\|_2^2 = \|\mathbf{v}\|_2^2 + \|\mathbf{v}'\|_2^2 = 2.$$

It follows that

$$2(1 - \delta_{s+s'}) \leq \|\mathbf{D}(\mathbf{v} \pm \mathbf{v}')\|_2^2 \leq 2(1 + \delta_{s+s'}).$$

Applying the polarization identity

$$\langle \mathbf{u}, \mathbf{u}' \rangle = \frac{1}{4} \left(\|\mathbf{u} + \mathbf{u}'\|_2^2 - \|\mathbf{u} - \mathbf{u}'\|_2^2 \right)$$

to $\mathbf{u} = \mathbf{D}\mathbf{v}$ and $\mathbf{u}' = \mathbf{D}\mathbf{v}'$, we obtain

$$\left|\langle \mathbf{D}\mathbf{v}, \mathbf{D}\mathbf{v}' \rangle\right| = \frac{1}{4} \left| \|\mathbf{D}\mathbf{v} + \mathbf{D}\mathbf{v}'\|_2^2 - \|\mathbf{D}\mathbf{v} - \mathbf{D}\mathbf{v}'\|_2^2 \right|.$$

Resolving $|\cdot|$ such that arg > 0 yields

$$\begin{aligned}\left|\langle \mathbf{D}\mathbf{v}, \mathbf{D}\mathbf{v}'\rangle\right| &= \frac{1}{4}\left(\left\|\mathbf{D}\mathbf{v} + \mathbf{D}\mathbf{v}'\right\|_2^2 - \left\|\mathbf{D}\mathbf{v} - \mathbf{D}\mathbf{v}'\right\|_2^2\right) \\ &\leq \frac{1}{4}\cdot 2(1+\delta_{s+s'}) - \frac{1}{4}\cdot 2(1-\delta_{s+s'})) \\ &= \frac{1}{2}(1+\delta_{s+s'} - 1 + \delta_{s+s'}) = \delta_{s+s'}.\end{aligned}$$

Resolving $|\cdot|$ such that arg < 0 yields

$$\begin{aligned}\left|\langle \mathbf{D}\mathbf{v}, \mathbf{D}\mathbf{v}'\rangle\right| &= \frac{1}{4}\left(\left\|\mathbf{D}\mathbf{v} - \mathbf{D}\mathbf{v}'\right\|_2^2 - \left\|\mathbf{D}\mathbf{v} + \mathbf{D}\mathbf{v}'\right\|_2^2\right) \\ &\leq \frac{1}{4}\cdot 2(1+\delta_{s+s'}) - \frac{1}{4}\cdot 2(1-\delta_{s+s'})) \\ &= \frac{1}{2}(1+\delta_{s+s'} - 1 + \delta_{s+s'}) = \delta_{s+s'}.\end{aligned}$$

□

Let us denote by $\mathbf{x}_{\mathcal{T}}$ the vector equal to $\mathbf{x}$ on the index set $\mathcal{T}$ and zero elsewhere. Let us first prove the noisy case. We start with the basic observation:

$$\|\mathbf{D}(\mathbf{x}^* - \mathbf{x})\|_2 \leq \underbrace{\|\mathbf{D}\mathbf{x}^* - \mathbf{y}\|_2}_{\leq\varepsilon(\text{as } \mathbf{x}^* \text{ is feasible})} + \underbrace{\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2}_{=\|\mathbf{n}\|_2\leq\varepsilon} \leq 2\varepsilon.$$

Write $\mathbf{x}^*$ as $\mathbf{x}^* = \mathbf{x} + \mathbf{h}$, and decompose $\mathbf{h}$ into a sum of vectors $\mathbf{h}_{\mathcal{T}_0}, \mathbf{h}_{\mathcal{T}_1}, \ldots$, each of sparsity at most s. $\mathcal{T}_0$ corresponds to the locations of the s largest coefficients of $\mathbf{x}$, $\mathcal{T}_1$ to the locations of the s largest (in absolute value) coefficients of $\mathbf{h}_{\mathcal{T}_0^c}$ and so on. The proof proceeds in two steps:

1. the size of $\mathbf{h}$ outside $\mathcal{T}_0 \cup \mathcal{T}_1$ is essentially bounded by that of $\mathbf{h}$ on $\mathcal{T}_0 \cup \mathcal{T}_1$,
2. $\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2$ is approximately small.

For the first step, we note that for each $j \geq 2$, we have

$$\left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2 \leq s^{1/2}\left\|\mathbf{h}_{\mathcal{T}_j}\right\|_\infty \leq s^{-1/2}\left\|\mathbf{h}_{\mathcal{T}_{j-1}}\right\|_1$$

because $s\left\|\mathbf{h}_{\mathcal{T}_j}\right\|_\infty \leq \left\|\mathbf{h}_{\mathcal{T}_{j-1}}\right\|_1$. We therefore get

$$\begin{aligned}\sum_{j\geq 2}\left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2 &\leq s^{-1/2}\left(\|\mathbf{h}_{\mathcal{T}_1}\|_1 + \|\mathbf{h}_{\mathcal{T}_2}\|_1 + \ldots\right) \\ &\leq s^{-1/2}\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1.\end{aligned}$$

This gives the useful estimate

$$\left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)^c}\right\|_2 = \left\|\sum_{j\geq 2}\mathbf{h}_{\mathcal{T}_j}\right\|_2 \leq \sum_{j\geq 2}\left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2 \leq s^{-1/2}\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1. \tag{1}$$

The key point is that $\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1$ cannot be very large as $\|\mathbf{x}+\mathbf{h}\|_1 = \|\mathbf{x}^*\|_1$ is minimum. By applying the reverse triangle inequality twice, we obtain

$$\begin{aligned}\|\mathbf{x}\|_1 \geq \|\mathbf{x}+\mathbf{h}\|_1 &= \sum_{j\in\mathcal{T}_0} |x_j+h_j| + \sum_{j\in\mathcal{T}_0^c} |x_j+h_j| \\ &\geq \|\mathbf{x}_{\mathcal{T}_0}\|_1 - \|\mathbf{h}_{\mathcal{T}_0}\|_1 + \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 - \left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 .\end{aligned}$$

This yields the following chain of inequalities

$$\begin{aligned}\|\mathbf{x}\|_1 - \|\mathbf{x}_{\mathcal{T}_0}\|_1 + \left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 &\geq -\|\mathbf{h}_{\mathcal{T}_0}\|_1 + \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 \\ \|\mathbf{x}_{\mathcal{T}_0}\|_1 + \left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 - \|\mathbf{x}_{\mathcal{T}_0}\|_1 + \left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 &\geq -\|\mathbf{h}_{\mathcal{T}_0}\|_1 + \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 \\ 2\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 + \|\mathbf{h}_{\mathcal{T}_0}\|_1 &\geq \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 \\ 2\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 + \|\mathbf{h}_{\mathcal{T}_0}\|_1 &\geq s^{1/2}\left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)^c}\right\|_2 ,\end{aligned}$$

where the last inequality follows directly from (1). Using the fact that $\|\mathbf{h}_{\mathcal{T}_0}\|_1 \leq s^{1/2}\|\mathbf{h}_{\mathcal{T}_0}\|_2$, this becomes

$$\begin{aligned}2\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 + s^{1/2}\|\mathbf{h}_{\mathcal{T}_0}\|_2 &\geq s^{1/2}\left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)^c}\right\|_2 \\ 2s^{-1/2}\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 + \|\mathbf{h}_{\mathcal{T}_0}\|_2 &\geq \left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)^c}\right\|_2 .\end{aligned}$$

By definition, $\mathbf{x}_{\mathcal{T}_0^c} = \mathbf{x} - \mathbf{x}_s$. Therefore,

$$2\underbrace{s^{-1/2}\|\mathbf{x}-\mathbf{x}_s\|_1}_{=e_0} + \|\mathbf{h}_{\mathcal{T}_0}\|_2 \geq \left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)^c}\right\|_2 .$$

Next, we bound $\left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)^c}\right\|_2$ from above. We have the following

$$\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1} = \mathbf{D}\left(\mathbf{h} - \sum_{j\geq 2}\mathbf{h}_{\mathcal{T}_j}\right) = \mathbf{D}\mathbf{h} - \sum_{j\geq 2}\mathbf{D}\mathbf{h}_{\mathcal{T}_j},$$

which implies

$$\|\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2^2 = \langle\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}, \mathbf{D}\mathbf{h}\rangle - \left\langle\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}, \sum_{j\geq 2}\mathbf{D}\mathbf{h}_{\mathcal{T}_j}\right\rangle. \tag{2}$$

The Cauchy-Schwarz inequality gives

$$|\langle\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}, \mathbf{D}\mathbf{h}\rangle| \leq \|\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 \|\mathbf{D}\mathbf{h}\|_2 .$$

Moreover, it holds that

$$\|\mathbf{D}\underbrace{(\mathbf{x}^*-\mathbf{x})}_{=\mathbf{h}}\|_2 \leq \|\mathbf{D}\mathbf{x}^*-\mathbf{y}\|_2 + \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2 \leq 2\varepsilon,$$

which, combined with the definition of the restricted isometry constant, gives

$$\begin{aligned}|\langle\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}, \mathbf{D}\mathbf{h}\rangle| &\leq \|\mathbf{D}\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 \cdot 2\varepsilon \\ &\leq 2\varepsilon\sqrt{1+\delta_{2s}}\,\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 .\end{aligned} \tag{3}$$

It follows from Lemma 3 that for all j, we have

$$\left|\left\langle \mathbf{D}\mathbf{h}_{\mathcal{T}_0}, \mathbf{D}\mathbf{h}_{\mathcal{T}_j} \right\rangle\right| \le \delta_{2s} \left\|\mathbf{h}_{\mathcal{T}_0}\right\|_2 \left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2. \tag{4}$$

The sets $\mathcal{T}_0$ and $\mathcal{T}_1$ are disjoint, and therefore,

$$\left\|\mathbf{h}_{\mathcal{T}_0}\right\|_2 + \left\|\mathbf{h}_{\mathcal{T}_1}\right\|_2 \le \sqrt{2} \left\|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2. \tag{5}$$

This can be seen as follows:

$$\left\|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2^2 = \underbrace{\left\|\mathbf{h}_{\mathcal{T}_0}\right\|_2^2}_{=a^2} + \underbrace{\left\|\mathbf{h}_{\mathcal{T}_1}\right\|_2^2}_{=b^2},$$

and we have

$$\begin{aligned}
\sqrt{2(a^2+b^2)} &\ge a+b \\
2(a^2+b^2) &\ge a^2+b^2+2ab \\
a^2+b^2-2ab &\ge 0 \\
(a-b)^2 &\ge 0.
\end{aligned}$$

Using the triangle inequality, (4), and (5), we obtain

$$\begin{aligned}
\left|\left\langle \mathbf{D}\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}, \sum_{j\ge 2} \mathbf{D}\mathbf{h}_{\mathcal{T}_j} \right\rangle\right| &\le \left|\left\langle \mathbf{D}\mathbf{h}_{\mathcal{T}_0}, \sum_{j\ge 2} \mathbf{D}\mathbf{h}_{\mathcal{T}_j} \right\rangle\right| + \left|\left\langle \mathbf{D}\mathbf{h}_{\mathcal{T}_1}, \sum_{j\ge 2} \mathbf{D}\mathbf{h}_{\mathcal{T}_j} \right\rangle\right| \\
&\le \sum_{j\ge 2} \delta_{2s} \left(\left\|\mathbf{h}_{\mathcal{T}_0}\right\|_2 + \left\|\mathbf{h}_{\mathcal{T}_1}\right\|_2\right) \left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2 \\
&\le \sqrt{2}\delta_{2s} \left\|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2 \sum_{j\ge 2} \left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2. \qquad (6)
\end{aligned}$$

Combining (2), (3), (6) and using again the definition of the restricted isometry constant, we get

$$(1-\delta_{2s}) \left\|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2^2 \le \left\|\mathbf{D}\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2^2 \le \left\|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2 \left(2\varepsilon\sqrt{1+\delta_{2s}} + \sqrt{2}\delta_{2s} \sum_{j\ge 2} \left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2\right)$$

Furthermore, we have

$$\sum_{j\ge 2} \left\|\mathbf{h}_{\mathcal{T}_j}\right\|_2 \le s^{-1/2} \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1$$

which implies

$$\left\|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2 \le \underbrace{\frac{2\sqrt{1+\delta_{2s}}}{1-\delta_{2s}}}_{=\alpha} \varepsilon + \underbrace{\frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}}_{=\rho} s^{-1/2} \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 = \alpha\varepsilon + \rho s^{-1/2} \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1.$$

Note that we can divide by $1-\delta_{2s}$ because $\delta_{2s} < 1$.

$$\left\|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\right\|_2 \le \alpha\varepsilon + \rho s^{-1/2} \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1$$

Using

$$\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 \le \left\|\mathbf{h}_{\mathcal{T}_0}\right\|_1 + \underbrace{2\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1}_{=2\|\mathbf{x}-\mathbf{x}_s\|_1},$$

this becomes

$$\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 \leq \alpha\varepsilon + \rho \underbrace{s^{-1/2}\|\mathbf{h}_{\mathcal{T}_0}\|_1}_{\leq\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_1} + \underbrace{\rho s^{-1/2} 2\|\mathbf{x}-\mathbf{x}_s\|_1}_{2\rho\mathbf{e}_0}.$$

This gives

$$\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 \leq \alpha\varepsilon + \rho\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 + 2\rho e_0,$$

and therefore, since we assumed that $\delta_{2s} < \sqrt{2}-1$, it holds that $1/(1-\rho) > 0$, and we have

$$\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 \leq \frac{\alpha\varepsilon + 2\rho e_0}{1-\rho}.$$

And finally

$$\|\mathbf{h}\|_2^2 = \|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2^2 + \left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)}\right\|_2^2$$

$$\begin{aligned}
\|\mathbf{h}\|_2 &\leq \|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 + \left\|\mathbf{h}_{(\mathcal{T}_0\cup\mathcal{T}_1)^c}\right\|_2 \\
&\leq \|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 + \|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 + 2e_0 \\
&= 2\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 + 2e_0 \leq 2\frac{\alpha\varepsilon + 2\rho e_0}{1-\rho} + 2e_0 \\
&= 2\frac{\alpha\varepsilon + 2\rho e_0 + e_0 - e_0\rho}{1-\rho} = 2\frac{\alpha\varepsilon + e_0\rho + e_0}{1-\rho} \\
&= 2\frac{\alpha\varepsilon + e_0(1+\rho)}{1-\rho} = \underbrace{2\frac{\alpha\varepsilon}{1-\rho}}_{=C_1\varepsilon} + \underbrace{2\frac{1+\rho}{1-\rho}}_{=C_0} s^{-1/2}\|\mathbf{x}-\mathbf{x}_{\mathcal{S}}\|_1
\end{aligned}$$

Lemma 4. Let $\mathbf{h}$ be any vector in the null space of $\mathbf{D}$ and let $\mathcal{T}_0$ be any set of cardinality s. Then,

$$\|\mathbf{h}_{\mathcal{T}_0}\|_1 \leq \rho\left\|\mathbf{h}_{\lambda Q_0^c}\right\|_1$$

with $\rho = \sqrt{2}\delta_{2s}(1-\delta_{2s}^{-1})$.

Proof. We have

$$\begin{aligned}
\|\mathbf{h}_{\mathcal{T}_0}\|_1 \leq s^{-1/2}\|\mathbf{h}_{\mathcal{T}_0}\|_2 &\leq s^{1/2}\|\mathbf{h}_{\mathcal{T}_0\cup\mathcal{T}_1}\|_2 \\
&\leq s^{-1/2}(\rho s^{-1/2}\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1) \\
&= \rho\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1
\end{aligned}$$

$$\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 \leq \|\mathbf{h}_{\mathcal{T}_0}\|_1 + 2\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 \leq \frac{2}{1-\rho}\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1$$

Therefore, in the noiseless case, we have

$$\begin{aligned}
\|\mathbf{h}\|_1 = \|\mathbf{h}_{\mathcal{T}_0}\|_1 + \left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 &\leq (\rho+1)\left\|\mathbf{h}_{\mathcal{T}_0^c}\right\|_1 \\
&\leq 2\frac{1+\rho}{1-\rho}\left\|\mathbf{x}_{\mathcal{T}_0^c}\right\|_1 \\
&= 2\frac{1+\rho}{1-\rho}\|\mathbf{x}-\mathbf{x}_s\|_1.
\end{aligned}$$

□

4 The Johnson-Lindenstrauss Lemma

Suppose we are given a set $\mathcal{U}$ of m points in $\mathbb{R}^n$. We would like to embed these points into a lower dimensional Euclidean space (i.e., in $\mathbb{R}^k$ with $k < n$), while approximately preserving the distances between the points in $\mathcal{U}$. The Johnson-Lindenstrauss (JL) Lemma, stated below, shows that any set of m points can be embedded in $k = O(\log m/\epsilon^2)$ dimensions while the distances between any two points change by at most a factor of $1 \pm \epsilon$. The JL Lemma, in particular, the concentration of measure inequality from which the JL Lemma follows (as shown later), will turn out to be an essential ingredient in proving the restricted isometry property (RIP) for random matrices considered in class. As a reference for these notes, see [2, 3].

Lemma 5 (Johnson-Lindenstrauss Lemma)**.** Choose ϵ with $0 < \epsilon < 1$ and suppose k satisfies

$$k \geq \frac{8}{\epsilon^2 - \epsilon^3} \log(2m). \tag{7}$$

Then, for every set $\mathcal{U}$ of m points, there is a (linear) map $f\colon \mathbb{R}^n \to \mathbb{R}^k$ such that for all $\mathbf{u}, \mathbf{u}' \in \mathcal{U}$,

$$(1-\epsilon) \left\|\mathbf{u} - \mathbf{u}'\right\|^2 \leq \left\|f(\mathbf{u}) - f(\mathbf{u}')\right\|^2 \leq (1+\epsilon) \left\|\mathbf{u} - \mathbf{u}'\right\|^2. \tag{8}$$

The JL Lemma is essentially tight according to [4, Thm. 9.3].

For concreteness, set $\epsilon = 0.5$, i.e., the distances between points may be reduced by no more than 50 percent, then k must be larger than $64\log(2m)$.

The original proof of the JL Lemma, as well as the proof discussed here, is based on random projections. Essentially, it is shown that projecting an arbitrary m-point subset into a random subspace only changes the inter-point distances by a factor of $1 \pm \epsilon$ with positive probability.

The JL Lemma will follow directly from the following concentration inequality. This concentration inequality will be an essential ingredient for verifying the RIP for random matrices considered in class.

Lemma 6. Let $\mathbf{A} \in \mathbb{R}^{k \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1/k)$ entries. Then, for ϵ with $0 < \epsilon < 1$ and a fixed $\mathbf{u} \in \mathbb{R}^n$,

$$\mathbb{P}\left(\left|\|\mathbf{A}\mathbf{u}\|^2 - \mathbb{E}\left[\|\mathbf{A}\mathbf{u}\|^2\right]\right| \geq \epsilon \|\mathbf{u}\|^2\right) < 2e^{-k\frac{\epsilon^2 - \epsilon^3}{4}}. \tag{9}$$

Also:

$$\mathbb{E}\left[\|\mathbf{A}\mathbf{u}\|^2\right] = \|\mathbf{u}\|^2. \tag{10}$$

In words, Lemma 6 states that the random variable $\|\mathbf{A}\mathbf{u}\|^2$ is concentrated around its expectation. An equation of the form (9) is called "concentration of measure inequality" or simply "concentration inequality" in the literature. Lemma 6 is not restricted to Gaussian random matrices, but generalizes to other random matrices. Essentially the same inequality holds if each entry $a_{i,j}$ of $\mathbf{A}$ is i.i.d. sub-Gaussian, i.e., its tail probability satisfies $\mathbb{P}(|a_{i,j}| > t) \leq c_1 e^{-c_2 t^2}$ for constants c_1, c_2.

Before proving Lemma 6, we will show how it implies the JL Lemma.

Proof of the JL Lemma. We will show that the (linear) map $f(\mathbf{u}) = \mathbf{A}\mathbf{u}$ with $\mathbf{A} \in \mathbb{R}^{k\times n}$ a random matrix with i.i.d. $\mathcal{N}(0, 1/k)$ entries, satisfies (8) for all $\mathbf{u}, \mathbf{u}' \in \mathcal{U}$ with non-zero probability.

Applying the union bound over all $m(m-1)/2 < m^2$ pairs of points in $\mathcal{U}$, it follows from Lemma 6 that (8) is violated for any pair of points $(\mathbf{u}, \mathbf{u}')$ with $\mathbf{u}, \mathbf{u}' \in \mathcal{U}$ with probability less than $m^2 2e^{-k\frac{\epsilon^2-\epsilon^3}{4}}$. If we take k as in (7), we obtain

$$k \geq \frac{4}{\epsilon^2-\epsilon^3} 2\log(2m) \Leftrightarrow -k\frac{\epsilon^2-\epsilon^3}{4} \leq \log(1/(4m^2)) \Leftrightarrow m^2 2e^{-k\frac{\epsilon^2-\epsilon^3}{4}} \leq 1/2.$$

This ensures $f(\mathbf{u}) = \mathbf{A}\mathbf{u}$ satisfies (8) with probability at least 1/2. □

Observe: If we make k grow faster than $\frac{4}{\epsilon^2-\epsilon^3} 2\log(2m)$ as a function of m, by the same logic we can demonstrate that $f(\mathbf{u}) = \mathbf{A}\mathbf{u}$ satisfies (8) with probability arbitrary close on one for large m.

Proof of Lemma 6. First observe that

$$\mathbb{E}\left[\|\mathbf{A}\mathbf{u}\|^2\right] = \mathbb{E}\left[\mathbf{u}^\intercal\mathbf{A}^\intercal\mathbf{A}\mathbf{u}\right] = \mathbf{u}^\intercal\mathbb{E}\left[\mathbf{A}^\intercal\mathbf{A}\right]\mathbf{u} = \mathbf{u}^\intercal\mathbf{I}\mathbf{u} = \|\mathbf{u}\|^2$$

which proves (10).

Next, let $\mathbf{a}_j^\intercal$ be the j-th row of $\mathbf{A}$, and set $X_j = \frac{\sqrt{k}}{\|\mathbf{u}\|}\mathbf{a}_j^\intercal\mathbf{u}$. Note that $\mathbf{a}_j^\intercal\mathbf{u}$ is the sum of independent Gaussians and is therefore $\|\mathbf{u}\|\mathcal{N}(0, 1/k)$ distributed. It follows that the X_j are i.i.d. $\mathcal{N}(0, 1)$ distributed. Next set $X = \sum_{j=1}^{k} X_j^2$. With this notation, we have

$$X = \sum_{j=1}^{k} X_j^2 = \frac{k}{\|\mathbf{u}\|^2}\sum_{j=1}^{k}\left|\mathbf{a}_j^\intercal\mathbf{u}\right|^2 = \frac{k}{\|\mathbf{u}\|^2}\|\mathbf{A}\mathbf{u}\|^2.$$

Thus, for $\lambda \geq 0$,

$$\begin{aligned}
\mathbb{P}\left(\|\mathbf{A}\mathbf{u}\|^2 \geq (1+\epsilon)\|\mathbf{u}\|^2\right) &= \mathbb{P}\left(X \geq (1+\epsilon)k\right) \\
&= \mathbb{P}\left(e^{\lambda X} \geq e^{\lambda(1+\epsilon)k}\right) \\
&\leq \frac{1}{e^{(1+\epsilon)k\lambda}}\mathbb{E}\left[e^{\lambda X}\right] && (11) \\
&= \frac{1}{e^{(1+\epsilon)k\lambda}}\prod_{j=1}^{k}\mathbb{E}\left[e^{\lambda X_j^2}\right] && (12) \\
&= \frac{1}{e^{(1+\epsilon)k\lambda}}\left(\mathbb{E}\left[e^{\lambda X_1^2}\right]\right)^k && (13)
\end{aligned}$$

where we used Markov's inequality[1] for a nonnegative random variable in (11), independence of the X_j for (12) and that all X_j have the same distribution for (13).

[1]For a nonnegative random variable X, $\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$.

It remains to evaluate the moment generating function $\mathbb{E}\left[e^{\lambda X_1^2}\right]$. Since X_1 is $\mathcal{N}(0,1)$ distributed,

$$\begin{aligned}\mathbb{E}\left[e^{\lambda X_1^2}\right] &= \int_{-\infty}^{\infty} e^{\lambda x^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\ &= \frac{1}{\sqrt{1-2\lambda}} \int_{-\infty}^{\infty} \frac{\sqrt{1-2\lambda}}{\sqrt{2\pi}} e^{-\frac{x^2}{2}(1-2\lambda)} dx \\ &= \frac{1}{\sqrt{1-2\lambda}} \end{aligned} \tag{14}$$

where we used that the integrand is the normal density with standard deviation $\frac{1}{\sqrt{1-2\lambda}}$. The conclusion above holds for any $\lambda < 1/2$. Using (14) in (13) yields

$$\mathbb{P}\left(\|\mathbf{A}\mathbf{u}\|^2 \geq (1+\epsilon)\|\mathbf{u}\|^2\right) \leq \left(\frac{e^{-2(1+\epsilon)\lambda}}{1-2\lambda}\right)^{\frac{k}{2}}. \tag{15}$$

We next minimize the right hand side (RHS) of (15). To this end, we choose λ such that the term $\frac{e^{-2(1+\epsilon)\lambda}}{1-2\lambda}$ is minimal. It is easily verified (by setting the derivative to zero) that the optimal choice is $\lambda = \frac{\epsilon}{2(1+\epsilon)}$. With this choice,

$$\mathbb{P}\left(\|\mathbf{A}\mathbf{u}\|^2 \geq (1+\epsilon)\|\mathbf{u}\|^2\right) \leq \left((1+\epsilon)e^{-\epsilon}\right)^{\frac{k}{2}} < e^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \tag{16}$$

where for the last inequality we used

$$1+\epsilon < e^{\epsilon - \frac{\epsilon^2-\epsilon^3}{2}}.$$

To prove this inequality it is sufficient to show that

$$\log(1+\epsilon) \leq \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{2}.$$

To show this, recall that the Taylor approximation to the logarithm is:

$$\log(1+\epsilon) = \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3} - \ldots.$$

so that for $0 \leq \epsilon \leq 1$:

$$\log(1+\epsilon) \leq \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3} \leq \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{2}.$$

Similarly, we obtain

$$\mathbb{P}\left(\|\mathbf{A}\mathbf{u}\|^2 \leq (1-\epsilon)\|\mathbf{u}\|^2\right) < e^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \tag{17}$$

Combining (16) and (17) via the union bound concludes the proof. □

5 Verifying the RIP from concentration inequalities

We show how to prove the RIP for random matrices.

Given $\mathcal{S}$ with $|\mathcal{S}| \leq s$, denote

$$\mathcal{X}_{\mathcal{S}} = \{\mathbf{x} \in \mathbb{R}^n : x_i = 0 \text{ for } i \in \mathcal{S}^c\}.$$

This is an s-dimensional linear subspace.

Our approach is to construct finite nets of points in each s-dimensional subspace, $\mathcal{X}_{\mathcal{S}}$, then apply the concentration inequality to all these points using the union bound, and then extend the result from the finite net of points to all possible s-dimensional signals.

Lemma 7. Let $\mathbf{D} \in \mathbb{R}^{m\times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1/m)$ entries. Then, for all $\mathcal{S}$ with $|\mathcal{S}| = s < m$ and all $0 < \delta < 1$,

$$(1-\delta)\left\|\mathbf{x}\right\|_2 \leq \left\|\mathbf{D}\mathbf{x}\right\|_2 \leq (1+\delta)\left\|\mathbf{x}\right\|_2 \tag{18}$$

is satisfied for all $\mathbf{x} \in \mathcal{X}_{\mathcal{S}}$ simultaneously with probability at least

$$1 - 2(12/\delta)^s e^{-c_0(\delta/2)m},$$

where $c_0(w) = \frac{1}{4}(w^2 - w^3)$.

Proof. Since $\mathbf{D}\mathbf{x}$ is a linear map, it suffices to prove (18) for $\left\|\mathbf{x}\right\|_2 = 1$. Choose a finite set of points $\mathcal{T}_{\mathcal{S}}$ such that $\mathcal{T}_{\mathcal{S}} \subseteq \mathcal{X}_{\mathcal{S}}$, $\left\|\mathbf{q}\right\|_2 = 1$ for all $\mathbf{q} \in \mathcal{T}_{\mathcal{S}}$, and for all $\mathbf{x} \in \mathcal{X}_{\mathcal{S}}$ with $\left\|\mathbf{x}\right\|_2 = 1$, we have

$$\min_{\mathbf{q}\in\mathcal{T}_{\mathcal{S}}} \left\|\mathbf{x}-\mathbf{q}\right\|_2 \leq \delta/4.$$

From the theory of covering numbers, it is known that we can choose such a set $\mathcal{T}_{\mathcal{S}}$ with $|\mathcal{T}_{\mathcal{S}}| \leq (12/\delta)^s$ elements. The finite set $\mathcal{T}_{\mathcal{S}}$ is called $\delta/4$-net for the infinite set $\mathcal{X}_{\mathcal{S}}$.

We use the union bound to apply Lemma 6 to this set of points with $\varepsilon = \delta/2$, which yields that

$$(1-\delta/2)\left\|\mathbf{q}\right\|_2^2 \leq \left\|\mathbf{D}\mathbf{q}\right\|_2^2 \leq (1+\delta/2)\left\|\mathbf{q}\right\|_2^2$$

holds for all $\mathbf{q} \in \mathcal{T}_{\mathcal{S}}$ *simultaneously* with probability at least

$$1 - 2(12/\delta)^s e^{-c_0(\delta/2)m},$$

where

$$c_0(w) = \frac{w^2 - w^3}{4}.$$

We next define A as the smallest number such that

$$\left\|\mathbf{D}\mathbf{x}\right\|_2 \leq (1+A)\left\|\mathbf{x}\right\|_2, \text{ for all } \mathbf{x} \in \mathcal{X}_{\mathcal{S}} \text{ with } \left\|\mathbf{x}\right\|_2 = 1.$$

Our goal is to show that $A \leq \delta$. To this end, recall that for every $\mathbf{x} \in \mathcal{X}_{\mathcal{S}}$ with $\left\|\mathbf{x}\right\|_2 = 1$, we can find a $\mathbf{q} \in \mathcal{T}_{\mathcal{S}}$ such that $\left\|\mathbf{x}-\mathbf{q}\right\|_2 \leq \delta/4$. Hence, we have

$$\left\|\mathbf{D}\mathbf{x}\right\|_2 \leq \left\|\mathbf{D}\mathbf{q}\right\|_2 + \left\|\mathbf{D}(\mathbf{x}-\mathbf{q})\right\|_2 \leq 1 + \delta/2 + (1+A)\delta/4.$$

Since by definition, A is the smallest number for which $\|\mathbf{D}\mathbf{x}\|_2 \leq (1+A)\|\mathbf{x}\|_2$, we have

$$A \leq \delta/2 + (1+A)\delta/4$$
$$A(1-\delta/4) \leq \delta/2 + \delta/4$$
$$A \leq \frac{\delta/2+\delta/4}{1-\delta/4} = \frac{2\delta+\delta}{4-\delta} \leq \frac{3\delta}{3} = \delta$$

as desired. We therefore proved that

$$\|\mathbf{D}\mathbf{x}\|_2 \leq (1+\delta)\|\mathbf{x}\|_2 .$$

The inequality $\|\mathbf{D}\mathbf{x}\|_2 \geq (1-\delta)\|\mathbf{x}\|_2$ follows since

$$\begin{aligned}\|\mathbf{D}\mathbf{x}\|_2 \geq \|\mathbf{D}\mathbf{q}\|_2 - \|\mathbf{D}(\mathbf{x}-\mathbf{q})\|_2 &\geq (1-\delta/2) - (1+\delta)\delta/4 \\ &= 1 - \delta/2 - \delta/4 - \delta^2/4 \\ &\geq 1 - \delta/2 - \delta/4 - \delta/4 = 1-\delta,\end{aligned}$$

which completes the proof. □

Theorem 8. Fix m, n, and $0 < \delta < 1$. If the pdf generating $\mathbf{D}$ satisfies the concentration inequality in Lemma 6, then there exist constants $c_1, c_2 > 0$ depending only on δ such that with probability at least $\geq 1 - 2e^{-c_2 m}$, δ is the restricted isometry constant of $\mathbf{D}$ for every sparsity level $s \leq c_1 m/\log(n/s)$.

Proof. By Lemma 7, we know that for each of the s-dimensional spaces $\mathcal{X}_{\mathcal{S}}$, the matrix $\mathbf{D}$ will fail to satisfy

$$(1-\delta)\|\mathbf{x}\|_2 \leq \|\mathbf{D}\mathbf{x}\|_2 \leq (1+\delta)\|\mathbf{x}\|_2 , \text{ for all } \mathbf{x} \in \mathcal{X}_{\mathcal{S}} \tag{19}$$

with probability no larger than

$$2(12/\delta)^s e^{-c_0(\delta/2)m}.$$

There are $\binom{n}{s} \leq (en/s)^s$ such subspaces. Hence, by the union bound, δ will fail to be the restricted isometry constant of $\mathbf{D}$ for sparsity level s with probability no larger than

$$2(en/s)^s(12/\delta)^s e^{-c_0(\delta/2)m} = 2e^{-c_0(\delta/2)m+s[\log(en/s)+\log(12/\delta)]}. \tag{20}$$

Thus, for every $c_1 > 0$, whenever

$$s \leq \frac{c_1 m}{\log(n/s)},$$

the exponent in (20) is smaller than $-c_2 m$, provided that

$$c_2 \leq c_0(\delta/2) - c_1\left(1 + \frac{1+\log(12/\delta)}{\log(n/s)}\right).$$

This is true because

$$e^{-c_0(\delta/2)m+s[\log(en/s)+\log(12/\delta)]} \leq e^{-c_2 m}$$

follows from

$$s \leq \frac{c_1 m}{\log(n/s)}$$
$$e^{-m\left(c_0(\delta/2)-c_1\frac{\log(en/s)+\log(12/\delta)}{\log(n/s)}\right)} \leq e^{-c_2 m}$$
$$c_2 \leq c_0(\delta/2) - c_1\left(1 + \frac{1+\log(12/\delta)}{\log(n/s)}\right).$$

Hence, we can always choose $c_1 > 0$ sufficiently small to ensure that $c_2 > 0$. This proves that with probability at least $1 - 2e^{-c_2 m}$, δ is the restricted isometry constant of $\mathbf{D}$ for sparsity level s. □

References

[1] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences*, vol. 346, pp. 589–592, May 2008.

[2] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of johnson and lindenstrauss," *Random Structures and Algorithms*, vol. 22, pp. 60–65, January 2003.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[4] N. Alon, "Problems and results in extremal combinatorics-i," *Discrete Math.*, vol. 273, pp. 31–53, Dec. 2003.