

# WHAT IS... A RIP MATRIX?

JOEL MOREIRA

ABSTRACT. RIP matrices— shorthand for matrices which satisfy the restricted isometry property— appeared as a byproduct of Compressed Sensing; a method discovered by D. Donoho, E. Candès and T. Tao in 2004 with several applications in computer science. Besides their real world application, RIP matrices are interesting mathematical objects because, on the one hand, a random matrix has a negligible probability of NOT satisfying the RIP, but on the other hand, deterministic constructions of RIP matrices are hard to find. The best known explicit construction, due to J. Bourgain, S. Dilworth, K. Ford, S. Konyagin and D. Kutzarova, is still very far from the best random constructions.

## 1. INTRODUCTION

Let  $K < n$  be natural numbers. A vector  $x \in \mathbb{R}^n$  is  $K$ -*sparse* if it has at most  $K$  non-zero entries.

**Definition 1** (RIP matrices). *Let  $n > m$  be positive integers and let  $\Phi$  be a  $m \times n$  matrix with real entries, let  $\delta > 0$  and let  $K < m$  be an integer. We say that  $\Phi$  is  $(K, \delta)$ -RIP if for every  $K$ -sparse vector  $x \in \mathbb{R}^n$  we have*

$$(1 - \delta)\|x\| \leq \|\Phi x\| \leq (1 + \delta)\|x\|$$

Thus  $\Phi$  acts almost as an isometry when we restrict attention to  $K$ -sparse vectors.

The restricted isometry property is perhaps the ‘simplest’ property of a matrix which makes it ‘good’ for compressed sensing. It was part of the outcome of a series of papers by Donoho, and Candès-Romberg-Tao in 2004-2006 where other slightly different properties were studied [8, 9, 6, 5] (see also the survey [11]). The earliest appearance of the name RIP that I could find was in [7], but even there the exact property is a little different from the modern one (which first appeared, I believe, in [4]).

In the next section I present a very short introduction to the ideas behind compressed sensing and how RIP matrices naturally arise. A more complete exposition of this material is Tao’s blog post [13]. In the final section I address the mathematic interest in RIP matrices, describe some random procedures to construct RIP matrices and mention the main difficulties in finding an explicit deterministic construction.

## 2. DIGITAL SIGNALS, COMPRESSED SENSING AND RIP

A *signal* is a vector  $x \in \mathbb{R}^n$  and can represent a digital sound or image. A *measurement* of  $x$  is the inner product  $\langle x, \psi \rangle$  with some given vector  $\psi \in \mathbb{R}^n$ . If we want to acquire all the entries (think pixels) of  $x$ , we need to make  $n$  measurements, in each taking  $\psi$  to be an element of the basis and obtaining one entry (pixel). In this case, the *measurement matrix*, i.e., the matrix whose lines are the vectors  $\psi$ , is just the identity matrix.

Taking the example when the signal  $x$  is an image, note that the “interesting” images have a large degree of structure. Indeed, a typical (random) image will have each pixel independent of its neighbours, resulting in something like Figure 1. In general, computer scientists realized that, after an orthonormal change of coordinates, an “interesting” signal  $x$  will have very few non-zero coordinates, when compared to the dimension  $n$ . In other words, interesting signals are  $K$ -sparse for a relatively small  $K$ .



FIGURE 1. A typical digital image

This fact has been used for several years in computers to store signals (which are assumed to be sparse). Instead of storing all the entries of  $x$ , one instead stores only the non-zero coordinates of  $x$  together with the positions of those entries. However, during the process of *acquiring* the signal (for instance, taking a photograph) one still needs to capture all the coordinates of  $x$ .

The idea of compressed sensing is to use much less than  $n$  measurements and still recover  $x \in \mathbb{R}^n$ , assuming it is sparse. If one knows which  $K$  coordinates of  $x$  are non-zero one only needs  $K$  measurements to encode  $x$ . However different signals will have the non-zero entries in different places. Thus what is needed is a measurement matrix  $\Phi$  such that  $x$  can be uniquely recovered from  $\Phi x$ , assuming only that  $x$  is  $K$ -sparse.

**Lemma 1.** *Let  $\Phi$  be a  $m \times n$  measurement matrix such that any  $K$  columns of  $\Phi$  are linearly independent. Then any  $K/2$ -sparse vector  $x$  can be uniquely determined from  $\Phi x$ .*

*Proof.* Assume  $\Phi$  is as described and  $\Phi x = \Phi y$  for two  $K/2$ -sparse signals  $x, y \in \mathbb{R}^n$ . Then  $\Phi(x - y) = 0$ . Since  $x - y$  is  $K$ -sparse, this implies that there is a linear combination of  $K$  of the columns of  $\Phi$  equal to 0, and by the hypothesis on  $\Phi$  it follows that  $x - y = 0$ .  $\square$

Thus if any  $K$  columns of a  $K \times n$  matrix  $\Phi$  are linearly independent, one can recover a  $K/2$ -sparse signal  $x \in \mathbb{R}^n$  from only  $b = \Phi x \in \mathbb{R}^K$ . Unfortunately, this theoretical result gives a very slow algorithm:

**Proposition 1.** *Let  $\Phi$  be a  $m \times n$  measurement matrix such that any  $K$  columns of  $\Phi$  are linearly independent. Let  $x \in \mathbb{R}^n$ , let  $b = \Phi x$  and let*

$$x_0 \in \{z \in \mathbb{R}^n : \Phi z = b\} \text{ be the vector with fewest non-zero coordinates}$$

*If  $x$  is  $K/2$ -sparse, then  $x_0 = x$ .*

However, to find  $x_0$  is an NP-hard problem, and thus impractical. An alternative is to find the vector in the hyperplane  $\{z \in \mathbb{R}^n : \Phi z = b\}$  which minimizes the  $\|\cdot\|_1$  norm<sup>1</sup>.

$$(1) \quad \text{Let } x_1 \in \{z \in \mathbb{R}^n : \Phi z = b\} \text{ be the vector which minimizes } \|z\|_1$$

There is, in general, no reason to believe that whenever  $x$  is a sparse vector and  $b = \Phi x$  we have  $x_1 = x$ . However, that holds true for special types of measurement matrices, and in particular when the measurement matrix  $\Phi$  is RIP:

**Theorem 1.**<sup>2</sup> *Let  $K < m < n$  be positive integers, let  $\delta < 1/3$  and let  $\Phi$  be a  $m \times n$  matrix satisfying the  $(K, \delta)$ -RIP. If  $x \in \mathbb{R}^n$  is  $K/4$ -sparse,  $b = \Phi x$  and  $x_1$  is given by (1), then  $x = x_1$ .*

<sup>1</sup>Recall that  $\|x\|_1 = |x_1| + \dots + |x_n|$ .

<sup>2</sup>It is hard to decide who to credit for this theorem. In this form it appeared in [4], but it is only the last evolutionary stage of a long collection of similar results, where the RIP is replaced with a slightly different property, some noise is present, or the conclusions are weakened. A good source for more information is the survey [11].

Thus, if one can easily construct RIP matrices, one can use significantly fewer than  $n$  measurements of a signal  $x \in \mathbb{R}^n$ , and still recover the signal (assuming it is sparse). Moreover, it is computationally feasible to recover  $x$  from the measurement  $b = \Phi x$ .

This is most important in applications where taking measurements is ‘expensive’ such as an MRI scan or an astronomical observation. Another possible application is in digital cameras, which would not be required to process the signal captured and hence save the battery.

### 3. CONSTRUCTING RIP MATRICES

In view of Theorem 1, it is desirable to construct RIP matrices. Specifically, one wants to find a  $(K, \delta)$ -RIP matrix of dimensions  $m \times n$  with the number of measurements,  $m$ , being as small as possible, with respect to the dimension of the signal  $n$  and the level of sparsity,  $K$ . For some reason, this is usually expressed as trying to maximize the level of sparsity  $K$ , depending on the number of measurements  $m$  and the dimension  $n$ .

**3.1. Random constructions.** The constructions of RIP matrices with lowest number of measurements use randomness. Below I present just a few examples:

**Theorem 2.** *Let  $m < n$  be positive integers, let  $\delta > 0$  and let  $K = O\left(\frac{m}{\log^4 n}\right)$ . Let  $\Phi$  be the random matrix defined by one of the following methods.*

- [2] (Normal entries) *Let the entries of  $\Phi$  be i.i.d. with a normal distribution  $N(0, 1/m)$ .*
- [2] (Bernoulli entries) *Let the entries of  $\Phi$  be i.i.d. with a Bernoulli distribution taking the values  $\pm 1/\sqrt{m}$ , each with 50% probability.*
- [12] (Random rows of the Discrete Fourier transform) *Let  $A \subset \{0, \dots, n-1\}$  be a random subset of size  $m$ . Let  $\Phi$  be the matrix obtained from the Discrete Fourier transform matrix (i.e. the matrix  $F$  with entries  $F[\ell, j] = e^{-2\pi i \ell j/n} / \sqrt{n}$  for  $\ell, j \in \{0, \dots, n-1\}$ ) by selecting the rows indexed by  $A$ .*

*Then  $\Phi$  is  $(K, \delta)$ -RIP with probability  $p \approx 1 - e^{-n}$ .*

We are not proving this theorem here, but it is worth noticing that the proof involves an old lemma of Johnson and Lindenstrauss (the father of the fields medalist!) on Lipschitz mappings into Hilbert spaces.

**Lemma 2** (Johnson-Lindenstrauss, [10]). *Let  $k < n$  be positive integers and let  $X \subset \mathbb{R}^n$  be a finite set of cardinality  $k$ . For every  $\epsilon > 0$  there exists a linear map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , with  $m = \frac{8 \log k}{\epsilon^2}$ , such that*

$$(1 - \epsilon)\|x - y\|^2 \leq \|f(x) - f(y)\|^2 \leq (1 + \epsilon)\|x - y\|^2$$

*for all  $x, y \in X$ .*

**3.2. Deterministic constructions.** It is somewhat surprising that, while almost every matrix is RIP, constructing one seems to be a very challenging problem. This situation appears in many other places in mathematics, for instance in explicitly finding normal numbers, or finding Ramsey graphs (graphs in  $n$  vertices without any clique or independent set of size  $\log n$ ). The general problem of constructing deterministic objects with a certain property, for which it is known that a random object will have that property has been called ‘finding hay in a haystack’.

The problem of finding a deterministic family of matrices of increasing dimensions which satisfies the restricted isometry property was explicitly formulated by Tao in his blog [14] (he uses the term UUP instead of RIP, standing for uniform uncertainty principle). The survey [1] contains several different approaches that have been used to make progress in this problem.

One of the most successful techniques to construct deterministic RIP matrices is based on the Gershgorin circle theorem:

**Theorem 3** (Gershgorin). *For each eigenvalue  $\lambda$  of a  $K \times K$  matrix  $A$ , there is an index  $i \in \{1, \dots, K\}$  such that*

$$|\lambda - A[i, i]| \leq \sum_{j \neq i} |A[i, j]|$$

Using this result one can show the following:

**Proposition 2.** *Let  $n, m$  be positive integers and let  $\Phi$  be a  $m \times n$  matrix where each columns has  $\ell^2$  norm 1. Then for any  $K < M$  we have that  $\Phi$  is a  $(K, (K - 1)\mu)$ -RIP, where  $\mu$  is the maximum inner product between two columns of  $\Phi$ .*

This proposition has been used to construct RIP matrices with sparsity  $K = \sqrt{m}$ . However, it has been shown that this is the highest possible level of sparsity one can get from Theorem 3; this is known as the Welch bound [15].

This ‘square root bottleneck’ is present in most other deterministic constructions. The only construction (to date) to beat this bottleneck is due to J. Bourgain, S. Dilworth, K. Ford, S. Konyagin and D. Kutzarova [3], but even there they obtain  $K = m^{1/2+\epsilon}$  for an explicit  $\epsilon$  of size

$$\epsilon \approx 5.5169 \times 10^{-28}$$

## REFERENCES

- [1] Afonso S. Bandeira, Matthew Fickus, Dustin G. Mixon, and Percy Wong. The road to deterministic matrices with the restricted isometry property. *J. Fourier Anal. Appl.*, 19(6):1123–1149, 2013. 3
- [2] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008. 3
- [3] Jean Bourgain, Stephen Dilworth, Kevin Ford, Sergei Konyagin, and Denka Kutzarova. Explicit constructions of RIP matrices and related problems. *Duke Math. J.*, 159(1):145–185, 2011. 4
- [4] Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008. 1, 2
- [5] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006. 1
- [6] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006. 1
- [7] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005. 1
- [8] Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006. 1
- [9] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006. 1
- [10] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984. 3
- [11] Holger Rauhut. Compressive sensing and structured random matrices. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 1–92. Walter de Gruyter, Berlin, 2010. 1, 2
- [12] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008. 3
- [13] T. Tao. Compressed sensing and single-pixel cameras. <http://terrytao.wordpress.com/2007/04/13/compressed-sensing-and-single-pixel-cameras>, April 2007. 1
- [14] T. Tao. Open question: Deterministic uup matrices. <http://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices/>, April 2007. 3
- [15] L.R. Welch. Lower bounds on the maximum cross correlation of signals. *IEEE Trans. Inform. Theory*, 20(3):397–399, 1974. 4