BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 54, Number 1, January 2017, Pages 151–165
http://dx.doi.org/10.1090/bull/1546
Article electronically published on August 25, 2016

A mathematical introduction to compressive sensing, by S. Foucart and H. Rauhut, Applied and Numeric Harmonic Analysis, Birkhäuser/Springer, New York, 2013, xviii+625 pp., ISBN 978-0-8176-4948-7

A mathematical introduction to compressive sensing by Simon Foucart and Holger Rauhut [FR13] is about sparse solutions to systems of random linear equations. To begin, let me describe some striking phenomena that take place in this context. Afterward, I shall try to explain why these facts have captivated so many researchers over the last decade. I shall conclude with some comments on the book.

1. Sparse solutions to random linear systems

Suppose that the vector $\boldsymbol{x}^\natural \in \mathbb{R}^n$ has the property that only s of its coordinates are nonzero. We say that the vector $\boldsymbol{x}^\natural$ is *s-sparse*. More colloquially, if s is somewhat smaller than n, we simply describe the vector as *sparse*.

Next, draw a random Gaussian matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. That is, each entry of $\boldsymbol{A}$ is an independent normal random variable with mean zero and variance one. This is the most fundamental model for a random rectangular matrix.

Consider the random linear system

$$\text{Find } \boldsymbol{x} \text{ such that} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}^\natural. \tag{1.1}$$

The distinguished point $\boldsymbol{x}^\natural$ always solves (1.1). We would like to explore whether (1.1) has another solution as sparse as the planted solution $\boldsymbol{x}^\natural$.

To formalize this question, let us pose an optimization problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \operatorname{card}\{i : x_i \neq 0\} \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}^\natural. \tag{1.2}$$

We may ask when $\boldsymbol{x}^\natural$ is the unique point that achieves the minimum in (1.2). Here is a (nearly) complete answer.

Theorem 1.1. *Suppose that $\boldsymbol{x}^\natural \in \mathbb{R}^n$ has s nonzero entries, where $s < n$, and let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be Gaussian. With probability one,*

1. *if $m > s$, then $\boldsymbol{x}^\natural$ is the unique solution to* (1.2);
2. *if $m = s$, then $\boldsymbol{x}^\natural$ is a nonunique solution to* (1.2);
3. *if $m < s$, then $\boldsymbol{x}^\natural$ is not a solution to* (1.2).

This result follows from a dimension-counting argument [FR13, Thm. 2.16].

Suppose now that we would like to *compute* the sparsest solution to (1.1). Unfortunately, the formulation (1.2) does not lead to practical computer algorithms; see [Nat95] or [FR13, Sec. 2.3]. The intuition behind this fact is that we might have to perform a census of each of the 2^n possible sets of nonzero entries in the variable $\boldsymbol{x}$ to solve (1.2).

2010 *Mathematics Subject Classification.* Primary 46B09, 68P30, 90C90, 94A08, 94A12, 94A20.

Key words and phrases. Image processing, information theory, mathematical programming, probability in Banach spaces, random matrix theory, sampling theory, signal processing.

Instead, we replace (1.2) with a problem that is easier to manage:

$$\text{minimize}_{\boldsymbol{x}} \quad \sum_{i=1}^{n} |x_i| \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}^\natural. \tag{1.3}$$

The ℓ_1 norm in the objective function of (1.3) is a convex proxy for the cardinality in (1.2). We can solve (1.3) efficiently because it is equivalent to a linear programming problem.

We may ask when $\boldsymbol{x}^\natural$ is the unique point that achieves the minimum in (1.3). The following theorem gives a detailed answer.

Theorem 1.2. *Suppose that $\boldsymbol{x}^\natural \in \mathbb{R}^n$ has s nonzero entries, and let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be Gaussian. Parameterize*

$$m \;=\; n \cdot \psi(s/n) \;+\; t \quad \textit{for } t \in \mathbb{R},$$

where $\psi : [0,1] \to [0,1]$ is the following concave function that fixes 0 *and* 1:

$$\psi(\varrho) := \inf_{\tau \geq 0} \left[\varrho(1+\tau^2) + (1-\varrho)\sqrt{\frac{2}{\pi}} \int_\tau^\infty (u-\tau)^2 \mathrm{e}^{-u^2/2}\,\mathrm{d}u \right].$$

With probability increasing rapidly to one as $|t|$ increases,

(1) *if $t > 0$, then $\boldsymbol{x}^\natural$ is the* unique *solution to* (1.3);
(2) *if $t < 0$, then $\boldsymbol{x}^\natural$ is* not *a solution to* (1.3).

Bounds for the probability are available.

Theorem 1.2 is a deep result that can be established using methods from integral geometry [DT09b, ALMT14] or statistical physics [BLM15] or Gaussian process theory [Sto13, TOH14]. See [ALMT14, Sec. 10] for some history.

The easiest way to understand Theorem 1.2 is to look at the computer experiment in Figure 1.1. In words, there is a sharp change, or *phase transition*, in the behavior of (1.3) when $m = n\psi(s/n)$. It is possible to bound the phase transition curve using elementary functions. The condition $m > 2s(1 + \log(n/s))$ is sufficient to ensure that $\boldsymbol{x}^\natural$ is the unique solution to (1.3) with high probability; this bound is sharp for small s.

Here is one interpretation of these results. Theorem 1.1 states that we need precisely $m = s+1$ random linear equations to determine an s-sparse vector uniquely. Theorem 1.2 implies that $m = 2s(1 + \log(n/s))$ random linear equations are sufficient if we want to find the s-sparse vector by means of the computationally tractable formulation (1.3). In sharp contrast, we need fully $m = n$ linear equations to describe a general vector in $\mathbb{R}^n$. These results suggest that the number of degrees of freedom in a sparse vector is roughly proportional to the sparsity level, not the ambient dimension.

2. Stability

Theorem 1.2 demonstrates that the ℓ_1 minimization method (1.3) is a powerful approach for finding a sparse solution to a random linear system. We can also investigate how this type of formulation behaves when we relax our assumptions.

Let $\boldsymbol{x}^\natural \in \mathbb{R}^n$ and $\boldsymbol{e} \in \mathbb{R}^m$ be arbitrary vectors, and draw a Gaussian matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Consider the perturbed ℓ_1 minimization problem

$$\text{minimize}_{\boldsymbol{x}} \quad \sum\nolimits_{i=1}^{n} |x_i| \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}^\natural + \boldsymbol{e}. \tag{2.1}$$

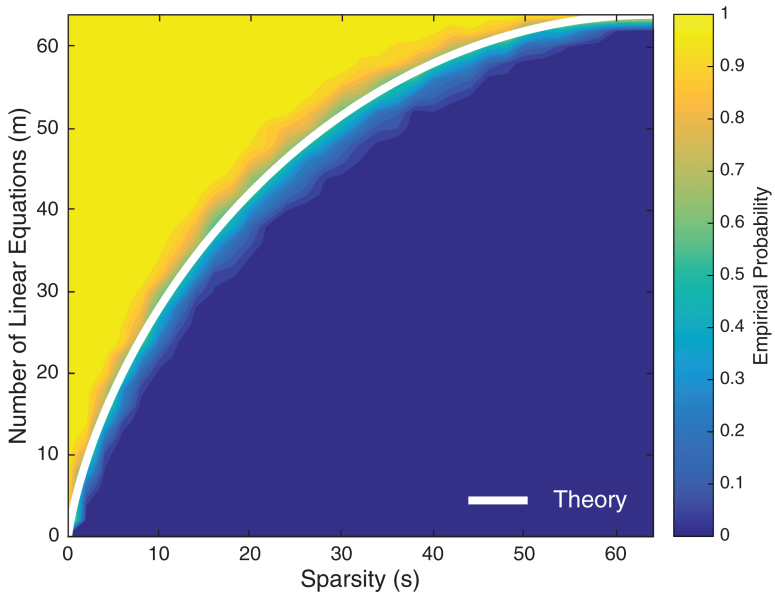

FIGURE 1.1. Computationally efficient recovery of a sparse vector from random linear equations. For $n = 64$, the heatmap displays the empirical probability that an n-dimensional vector $\boldsymbol{x}^\natural$ with s nonzero entries is the unique solution to (1.3) for an $m \times n$ Gaussian matrix $\boldsymbol{A}$. The white curve is the theoretical phase transition $m = n\psi(s/n)$.

We may study how well a solution to the optimization problem (2.1) approximates the distinguished point $\boldsymbol{x}^\natural$. Here is a satisfactory result.

Theorem 2.1. *Fix arbitrary vectors $\boldsymbol{x}^\natural \in \mathbb{R}^n$ and $\boldsymbol{e} \in \mathbb{R}^m$, and let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be Gaussian. Select a positive integer s for which*

$$m > \text{Const} \cdot s \log(n/s). \tag{2.2}$$

With probability increasing rapidly to one as m increases, each solution $\boldsymbol{x}^\star$ to the optimization problem (2.1) *satisfies*

$$\|\boldsymbol{x}^\natural - \boldsymbol{x}^\star\|^2 \leq \text{Const} \cdot \left[\min_{s\text{-sparse } \boldsymbol{y}} \|\boldsymbol{x}^\natural - \boldsymbol{y}\|^2 + \frac{1}{m} \|\boldsymbol{e}\|^2 \right].$$

We have written $\|\cdot\|$ for the ℓ_2 norm, and the constants are absolute.

The analysis requires methods from approximation theory and geometry of Banach spaces. See [FR13, Thm. 11.23] for the proof and background.

When $\boldsymbol{x}^\natural$ is sparse and $\boldsymbol{e} = \boldsymbol{0}$, Theorem 2.1 gives the same qualitative prediction for (2.1) as Theorem 1.2. As a consequence, we cannot improve the bound (2.2) on the sparsity level s beyond the precise value of the constant. When $\boldsymbol{x}^\natural$ is not sparse, Theorem 2.1 demonstrates that (2.1) always delivers an approximation to $\boldsymbol{x}^\natural$ that is comparable with the best s-sparse approximation.

One may wonder whether the behavior of the optimization problems (1.3) and (2.1) persists even when the matrix $\boldsymbol{A}$ is not Gaussian. In fact, there is extensive numerical [DT09a] and theoretical [OT15] evidence that the detailed result of Theorem 1.2 remains valid for some other types of random matrices. Moreover, it is possible to prove qualitative results, such as Theorem 2.1, for a wide class of examples that also includes random matrices with additional structure [FR13].

3. Compressed sensing

The field of compressed sensing is founded on a claim about the technological implications of results like Theorems 1.2 and 2.1. Compressed sensing interprets this theory as evidence that we can design sensing technologies that acquire some types of data using far fewer measurements than classical systems allow. This claim is surprising, persuasive, and inspirational. But the jury has not yet reached a verdict on the practical impact; see Section 5.

To provide some background, let me say a few words about sensing technologies. We use the word *signal* to describe an information-bearing function, such as an image, a sound, or a radio wave. For simplicity, we only consider signals that are represented as finite-dimensional vectors, rather than continuous functions.

Sensors are mechanisms for acquiring data about a target signal out in the world. Sensors include antennas, microphones, microscopes, digital cameras, X-ray machines, magnetic resonance imagers (MRI), and far more. Given the measured data from one of these systems, we wish to perform *signal processing* tasks, such as classifying the signal or producing a digital reconstruction.

The main goal of compressed sensing is to develop sensors that allow us to acquire and reconstruct a target signal using as few measurements as possible. The approach is based on a family of principles about how to model sensing technologies and perform signal reconstructions:

Sparsity. The first principle is that we should look for problem domains where signals of interest are *sparse*. More precisely, after an appropriate change of coordinates, the target signal $\boldsymbol{x}^\natural \in \mathbb{R}^n$ should be well approximated by a sparse vector. For example, audio and radar signals can often be approximated using a small number of frequency components. Images can be expressed efficiently using a basis of functions called *wavelets*; see Figure 3.1.

Linear measurements. We can abstract the action of a sensor as a linear transformation that maps a target signal to a vector of measurements. If we write $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ for the measurement map, then the measurements of a signal $\boldsymbol{x}^\natural$ take the form $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}^\natural + \boldsymbol{e}$, where the vector $\boldsymbol{e}$ models measurement errors. Linearity gives a good description of many practical systems. For example, a digital camera returns the intensity of light at locations in the image plane. An X-ray device collects line integrals of the density of the target object.

Randomness. Some sensing technologies have enough design flexibility that we can modify the type of measurements they collect. Compressed sensing promotes the idea that the measurement map $\boldsymbol{A}$ should be "as random as possible" to make the measurements more informative. For example, in an imaging system, we might introduce randomness by placing a series of patterned screens between the image and the sensor.

FIGURE 3.1. Images are approximately sparse in a wavelet basis. [left] A 1024×1024 pixel grayscale image. [right] An approximation performed by selecting the largest 1% of the coefficients when the image is expressed in the orthonormal basis of Daubechies-4 tensor product wavelets. The reconstruction has dimension $n = 1.05 \cdot 10^6$ and sparsity level $s = 1.05 \cdot 10^4$.

Efficient algorithms. Once we have collected noisy data $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}^\natural + \boldsymbol{e}$ about a sparse signal, we wish to perform signal processing tasks, such as digital reconstruction. To do so, we should use efficient algorithms that can exploit sparsity. For example, we may solve the optimization problem (2.1) using linear programming software.

Minimal sampling. The focus of compressed sensing is to acquire a target signal using the minimal number of measurements. If $\boldsymbol{x}^\natural$ is well approximated by an s-sparse vector, then Theorem 2.1 predicts that $m = \text{Const} \cdot s \log(n/s)$ random measurements are sufficient for us to approximate the target signal by solving the tractable formulation (2.1).

Summary. In short, compressed sensing proposes that we should identify domains where signals are very sparse. We should then design sensors that collect random measurements. Random measurements allow us to determine a sparse signal from fewer numbers than its nominal dimension. Last, we should use efficient computational methods to reconstruct these signals digitally.

This package of ideas is usually attributed to a pair of papers that were released in 2004. The first paper [CRT06], written by Candès, Romberg, and Tao (CRT), is called "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." The second paper [Don06], written by Donoho, is called "Compressed sensing."

4. History and context

There is a persistent myth that the subject of compressed sensing was born in 2004, without a history and fully grown, like Athena springing from the forehead of Zeus. In fact, most of the conceptual elements outlined above had already been established firmly before the appearance of the CRT and Donoho papers. It is

worth noting a few of the precedents, although a complete account of the history would require a separate essay.

4.1. **Algorithms for sparse approximation.** A *sparse approximation* problem requests the sparsest vector that almost solves a linear system:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \text{card}\{i : x_i \neq 0\} \quad \text{subject to} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\| \leq \varepsilon. \tag{4.1}$$

Here, the matrix $\boldsymbol{A}$ might be deterministic or random. It is NP-hard to solve (4.1) for some classes of problem data [Nat95].

Sparse approximation arose in the statistics literature in the 1950s as a model for selecting a small set of informative variables for linear regression [Mil02]. To solve (4.1), statisticians developed heuristic greedy algorithms that build up a solution iteratively, adding one nonzero entry at a time to reduce the error. During the 1970s, researchers in geophysics proposed techniques based on ℓ_1 minimization, similar to (2.1), to solve inverse problems that arise in geological imaging [CM73, TBM79]. In spite of the long history, there was little rigorous theory on sparse approximation before the 1990s, aside from a few scattered papers [Log65, SS86, DS89].

We can trace contemporary interest in sparse approximation to research in applied harmonic analysis in the 1970s and 1980s. This community demonstrated that there is a correspondence between the regularity of a function and the complexity of representing the function. In particular, a piecewise smooth function admits a sparse approximation as a linear combination of wavelets. This idea plays an important role in image and video compression because many real-world signals are piecewise smooth; see Figure 3.1 for an illustration. More background appears in [DVDD98, Mal09].

By the early 1990s, a number of researchers were studying procedures for producing sparse approximations of functions with respect to wavelets and other systems of basis functions. *Inter alia*, see [CW92, Bar93, DMZ94, DJ94, DT96, CDS98]. Around the same time, statisticians began to direct fresh attention to variable selection problems as they sought to improve the accuracy and interpretability of predictions in large-scale regression models [Tib96]. We should also mention a third line of work on total-variation minimization that emerged in imaging science; the connection endures because the total variation of a signal is roughly the ℓ_1 norm of its gradient [ROF92].

After 2000, researchers began to develop a systematic theory of sparse approximation. This effort led to geometric assumptions on the matrix $\boldsymbol{A}$ which ensure that tractable sparse approximation methods, such as (2.1), solve the problem (4.1). By 2004, a family of sharp results was available for deterministic sparse approximation problems [DH01, GMS03, DE03, Tro04, Fuc05, DET06, Tro06]. At this point, the arrival of compressed sensing instigated a careful study of random sparse approximation problems. See Foucart and Rauhut [FR13] for more discussion.

4.2. **Randomness and minimal sampling.** There are several strands of research that support the notion that random measurements offer an efficient mechanism for acquiring information about a sparse vector.

In the early 1990s, the theoretical computer science literature developed the idea that we can reconstruct a function with a sparse approximation from the information in random point evaluations [KM93, Man95]. In the early 2000s, Gilbert

et al. introduced efficient computer algorithms that use a small number of random point evaluations to solve sparse approximation problems with respect to wavelets [GGI+02b] or discrete Fourier modes [GGIMS02]. The primary goal of these papers is to develop fast algorithms for specific computational problems, rather than to design new sensors.

Donoho [Don06] has also identified a mathematical prototype for compressed sensing in classical results from approximation theory. In a Euclidean space, the Gel'fand m-width of a centrally symmetric set is the diameter of the smallest intersection between the set and a subspace of codimension m. In other words, the Gel'fand width measures how accurately we can approximate the points in the set using the best fixed choice of m linear functionals. Kashin [Kas77] and Garnaev and Gluskin [GG84] showed that a *random* subspace achieves the Gel'fand width of the ℓ_1^n ball, up to a constant factor, with high probability. Vectors in the ℓ_1^n ball are well approximated by sparse vectors, per Markov's inequality, and we can find a point in the ℓ_1^n ball that satisfies linear constraints by means of linear programming software. Connecting the dots, we discover a sketch of the compressed sensing idea.

4.3. **Computational sensing.** Compressed sensing can also be classified as a species in a larger research taxon called *computational sensing*, which dates at least as far back as the 1970s. In this field, the basic approach is to use modeling and computation to extract more information from sensing technologies. Let us mention a few examples.

Most imaging systems can only collect the amplitude of a light field. In the early 1970s, researchers in optics began to apply computer algorithms to reconstruct the missing phase information [GS72, Fie82].

Around the same time, optics researchers developed imaging systems that use a series of patterned screens to modulate a light field. These systems are designed to reduce the intensity of the illumination on the sample, to reduce the number of measurements required, and to increase the signal-to-noise ratio. They also use computer algorithms to perform image reconstruction [HS79].

Geophysicists rely on reflection seismograms to map geological features under the Earth's surface. The measurements consist of the convolution of a known waveform with an unknown sparse vector that encodes changes in the subsurface density. Around 1980, geophysicists began using ℓ_1 minimization methods to improve the quality of these geological maps [TBM79, LF81, OSL83].

4.4. **Compressed sensing.** We have seen that the research literature already describes many of the principles that underlie compressed sensing. So, why did the CRT and Donoho papers have such an outsize impact?

First, these works contain *theorems* that improve qualitatively on prior art. Both papers demonstrate that we can reconstruct an s-sparse vector in $\mathbb{R}^n$ from about $m \approx s \log n$ random linear equations by solving a tractable optimization problem like (1.3). The best previous results in the applied harmonic analysis literature require $m \approx s^2$ equations [DE03, Tro04].

Second, the CRT and Donoho papers raise the prospect of technological applications of these new theoretical results. As a motivating example, the CRT paper discusses an idealization of MRI. The Donoho paper considers examples in imaging and spectroscopy. In short, these two works explicitly announce the compressed sensing project.

The CRT and Donoho papers ignited an explosion of research. This body of work includes everything from theoretical analyses of structured random matrices all the way to concrete attempts to design and build technologies based on compressed sensing principles. After a decade, some of the smoke has cleared, and we are in a better position to assess the successes and failures of the compressed sensing program.

5. Applications?

Compressed sensing was not developed in response to the exigencies of a particular application. Rather, it is a mathematical theory that includes claims about its own significance. As we seek to understand the technological value of this approach, we must delineate conditions under which the theory is relevant and specify applications in which these conditions prevail.

Compressed sensing has the following prerequisites:

(1) The signal $\boldsymbol{x}^\natural$ must be well approximated by a sparse vector.
(2) The signal-to-noise ratio (SNR) must be high. The SNR is roughly the relative size of $\boldsymbol{x}^\natural$ and the measurement error $\boldsymbol{e}$.
(3) The sensing technology must be able to acquire random measurements.
(4) The "unit cost" per measurement must be high.
(5) We can afford the computational cost of the reconstruction algorithm.

There are a few applications where the planets align, but the requirements are stringent enough that many attempts to design compressed sensing technologies have failed.

The most prominent successful application is the "Sparse MRI" system, which uses compressed sensing principles to accelerate signal acquisition in MRI [LDP07]. This approach works because medical images are relatively sparse, the highly engineered equipment achieves good SNR, and MRI measurements can be modified to insert some randomness. The measurement "cost" reflects the amount of time a patient spends in the MRI scanner, so it is acceptable to trade the computational cost of reconstruction for reduced scanning time. Sparse MRI can trim the image acquisition time by a factor between two and six, which is a meaningful clinical improvement [VMA+11].

The literature also describes some less successful attempts to invoke the compressed sensing paradigm. For instance, there was an enthusiastic effort to use these principles to design wideband analog-to-digital converters (ADCs), which record the voltage of an electrical signal at a high rate. One can make a theoretical case for the value of this type of technology [TLD+10]. But projects to build practical systems foundered because of inadequate sparsity and SNR, the difficulty of performing random sampling in hardware, and also the expense of reconstruction algorithms. See [DLTB12] for a discussion of some of these issues.

Overall, the greatest obstacles to compressed sensing are the requirements of high sparsity and high SNR. Many signals do admit sparse approximations, but the sparsity level s may not be small enough to support a substantial reduction in the number m of measurements. At the same time, when the measurements are corrupted by random noise, every reduction in the number m of samples by a factor of two also decreases the quality of the digital reconstruction by a factor of two [DLTB12, CD13]; see also Theorem 2.1. As a consequence, reducing the number of measurements—the very goal of compressed sensing—can be self-defeating.

The theory of compressed sensing promises an exponential decrease in the number of measurements needed to acquire signals. At this point, I am not aware of settings where the reduction has been more dramatic than a small, but useful, constant factor. It is hard to predict what applications the future will bring, but compressed sensing has not had the technological impact that its strongest proponents anticipated.

6. The halo effect

Although compressed sensing has not yet achieved its goals fully, the coherent research effort generated ancillary benefits. First, researchers are better educated about established insights from applied harmonic analysis, convex optimization, and random matrix theory. Second, the community has made substantial progress on basic theoretical and algorithmic methods that will have wider applications.

As we have discussed, compressed sensing depends on foundational research in applied harmonic analysis. In particular, sparse signal models and algorithms for sparse approximation play a central role. These powerful ideas have become more widely known over the last decade. In fact, when people refer to "compressed sensing", they often mean "sparse approximation".

Because of research in compressed sensing, our knowledge of sparse signal models has increased significantly. We now have a keener understanding of how to find other types of "structured" solutions to systems of linear equations. For instance, we can design and analyze optimization templates, similar to (1.3), for finding a low-rank matrix that satisfies linear constraints [Faz02, RFP10]. This is just one example from a wide spectrum of problems. See [CRPW12, ALMT14, TOH15] for some others.

Compressed sensing also makes a vigorous argument for the importance of convex optimization. Indeed, many approaches, such as (1.3), for finding structured solutions to linear systems depend on convex optimization. Although we can solve these problems efficiently in principle, they can be quite challenging in practice. Over the last decade, there has been a flood of work on algorithms for large-scale convex optimization. We now have significantly better computational methods than we did a decade ago. For example, see [TW10, BCG11]. These algorithmic advances will have an impact far beyond compressed sensing.

The theory of compressed sensing also depends heavily on results about the geometric and spectral properties of random matrices. To model realistic measurement systems more accurately, we must study random matrices with more structure than a Gaussian matrix. The tools for this inquiry derive from random matrix theory, stochastic processes, convex geometry, operator theory, and more. These methods have become a part of the standard toolkit for an entire generation of students.

Theoretical problems in compressed sensing have also motivated the introduction of novel probabilistic techniques. For instance, we have developed an entire suite of new inequalities for random matrices [Tro12]. We have also achieved a deeper understanding of random processes with quadratic structure [KMR14].

Finally, enthusiasm about compressed sensing has inspired engineers to be creative about designing new types of sensors. For example, Baraniuk's group has developed a lensless digital camera that is very thin [AAS+15]. Another group of researchers has been working on biological assays that can efficiently identify the

constituents of a sample [EGN+15]. Both of these systems rely on clever measurements and sophisticated algorithms.

Although most problem domains are not amenable to minimal sampling, some of the other principles that support compressed sensing do remain valuable. Indeed, the field of *computational sensing* is alive and well.

7. What's in it for mathematicians?

For mathematicians, compressed sensing might seem like a spectator sport. This is hardly the case. The field draws on an astonishing range of mathematical techniques; it has led to a deeper understanding of parts of mathematics; and it even has some new ramifications. This section offers a brief tour of the connections with theoretical mathematics.

Random matrix theory. Rudelson and Vershynin used ideas from compressed sensing to develop quantitative bounds for the minimum singular value of a square random matrix [RV08]. Related constructions led to universality laws for the action of a random matrix on a convex set [OT15].

Approximation theory. Proof techniques from compressed sensing allowed researchers to compute the Gel'fand widths of ℓ_p balls in the range $0 < p < 1$ [FPRU10].

Combinatorial geometry. The behavior of the optimization problem (1.3) is closely related to the facial structure of a random image of the ℓ_1 ball. This connection has tendered new results on the existence of "neighborly" polytopes [DT09b].

Integral geometry. Investigations in compressed sensing led to a deeper understanding of the intrinsic volumes of convex cones [ALMT14, MT14]. This work also revealed new phase transition phenomena that take place in classical problems on the intersection properties of randomly oriented cones [MT13].

Harmonic analysis. There is a long-standing connection between sparsity and uncertainty principles [DS89, DH01, Tao05, CRT06]. In addition, the properties of random measurement matrices are closely connected with the Λ_1 set problem; see [FR13, Sec. 12.7].

Number theory. There is also a long-standing connection between sparsity and the analytic principle of the large sieve [DL92]. More recently, authors have used properties of the Legendre symbol to construct nonrandom measurement matrices that mimic the properties of random measurements [BFMM16].

Additive combinatorics. Researchers have also used deep facts about the existence of sum sets inside product sets to construct nonrandom measurement matrices [BDF+11].

Graph theory. Expander graphs can be applied to design measurement matrices that have few nonzero entries [GI10].

This dizzying assortment of topics indicates why so many researchers have been drawn to compressed sensing. We expect that some of the research challenges that have arisen from compressed sensing will only be resolved by the development of new mathematics.

8. About the book

The book by S. Foucart and H. Rauhut is the first textbook on the subject of compressed sensing, although several compendia [BCEM10, EK12] of expository

research articles appeared earlier. The book is intended as a guide for students and researchers who are interested in the mathematical underpinnings of the field. The book does many things well, and it is the best reference I know for some of its content. On the other hand, it is too exhaustive to be truly introductory, and it was written before some of the most striking mathematical developments took place.

Foucart and Rauhut have focused their attention on the years 2000–2012, which was the period of greatest activity in sparse approximation and compressed sensing. Almost all of the significant research from this era appears within the book. The authors discuss the properties of sparse solutions to linear systems, both deterministic and random. They present and analyze several classes of algorithms for solving these problems. They explain techniques for studying the properties of random matrices *qua* compressed sensing. They draw deep connections with several areas of mathematics and computer science. The notes provide a thorough and honest account of the history, with even-handed discussions about competing claims of novelty. There is a selection of exercises with varying degrees of difficulty. The authors, wisely, do not discuss specific applications after the introduction.

In my view, the strongest aspect of the book is the comprehensive treatment of methods for studying structured random matrices. This is a beautiful and important subject that used to be quite difficult to learn. The authors have produced the first accessible treatment that proceeds from first principles. I expect that their presentation of this material will serve as a lasting reference.

The book often feels more like a research monograph than an invitation to students because of the breadth of topics and the level of detail. I think this is because Foucart and Rauhut have eschewed judgments about the relative value of different parts of the compressed sensing project. Researchers who have contributed to compressed sensing will appreciate this ecumenical approach. But some readers may wish that the authors had curated the material more aggressively.

In spite of its breadth, the book does omit a few subjects that a mathematically oriented reader might appreciate. For example, Foucart and Rauhut motivate the idea of sparse models in image processing using numerical examples. There is limited discussion about how harmonic analysis justifies this approach; readers will want Mallat's book [Mal09] for remediation.

Our theoretical understanding of compressed sensing and related problems has continued to mature since the appearance of Foucart and Rauhut's book. We now recognize that ordinary sparsity is part of a spectrum of models that have relatively few degrees of freedom. We have machinery for developing optimization problems that can enforce various types of structure. We have sharp results, such as Theorem 1.2, that describe the number of random measurements we need to determine a signal from a general model. We also have a precise analysis of the impact of noise in the measurements. See [CRPW12, ALMT14, TOH15, OT15, Thr16] for presentations of this material.

Compressed sensing has provided an opportunity for electrical engineers to learn new mathematics, and it has given mathematicians some challenging new problems to consider. Foucart and Rauhut have written a comprehensive survey of the ideas and methods from this field. Their book will engage the interest of many researchers, both theoretical and applied.

Acknowledgments

The author gratefully acknowledges support from ONR Award N00014-11-1002 and the Gordon & Betty Moore Foundation. Many thanks to Richard Baraniuk, John Bruer, Mark Davenport, Anna Gilbert, Michael McCoy, and Madeleine Udell for their contributions.

References

[ALMT14] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, *Living on the edge: phase transitions in convex programs with random data*, Inf. Inference **3** (2014), no. 3, 224–294. MR3311453

[AAS+15] M. S. Asif, A. Ayremlou, A. C. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, *Flatcam: Thin, bare-sensor cameras using coded aperture and computation*, 2015. Available at `http://arXiv.org/abs/1509.00116`.

[BFMM16] A. S. Bandeira, M. Fickus, D. G. Mixon, and J. Moreira, *Derandomizing Restricted Isometries via the Legendre Symbol*, Constr. Approx. **43** (2016), no. 3, 409–424. MR3493967

[BCEM10] R. G. Baraniuk, E. Candès, M. Elad, and Y. Ma, *Applications of sparse representation and compressive sensing*, Proceedings of the IEEE **98** (2010June), no. 6. Special issue.

[Bar93] A. R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory **39** (1993), no. 3, 930–945. MR1237720

[BLM15] M. Bayati, M. Lelarge, and A. Montanari, *Universality in polytope phase transitions and message passing algorithms*, Ann. Appl. Probab. **25** (2015), no. 2, 753–822. MR3313755

[BCG11] S. R. Becker, E. J. Candès, and M. C. Grant, *Templates for convex cone problems with applications to sparse signal recovery*, Math. Program. Comput. **3** (2011), no. 3, 165–218. MR2833262

[BDF+11] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova, *Explicit constructions of RIP matrices and related problems*, Duke Math. J. **159** (2011), no. 1, 145–185. MR2817651

[CD13] E. J. Candès and M. A. Davenport, *How well can we estimate a sparse vector?*, Appl. Comput. Harmonic Anal. **34** (2013Mar.), no. 2, 317–323.

[CRT06] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52** (2006), no. 2, 489–509. MR2236170

[CRPW12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, *The convex geometry of linear inverse problems*, Found. Comput. Math. **12** (2012), no. 6, 805–849. MR2989474

[CDS98] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. **20** (1998), no. 1, 33–61. MR1639094

[CM73] J. F. Claerbout and F. Muir, *Robust modeling with erratic data*, Geophysics **38** (1973), no. 5, 826–844, available at `http://geophysics.geoscienceworld.org/content/38/5/826.full.pdf`.

[CW92] R. R. Coifman and M. V. Wickerhauser, *Entropy-based algorithms for best basis selection*, IEEE Transactions on Information Theory **38** (1992March), no. 2, 713–718.

[DLTB12] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk, *The pros and cons of compressive sensing for wideband signal acquisition: noise folding versus dynamic range*, IEEE Trans. Signal Process. **60** (2012), no. 9, 4628–4642. MR2960550

[DMZ94] G. Davis, S. Mallat, and Z. Zhang, *Adaptive time-frequency approximations with matching pursuits*, Wavelets: theory, algorithms, and applications (Taormina, 1993), 1994, pp. 271–293. MR1321432

[DT96] R. A. DeVore and V. N. Temlyakov, *Some remarks on greedy algorithms*, Adv. Comput. Math. **5** (1996), no. 2-3, 173–187. MR1399379

[DL92] D. L. Donoho and B. F. Logan, *Signal recovery and the large sieve*, SIAM J. Appl. Math. **52** (1992), no. 2, 577–591. MR1154788

[DT09a] D. Donoho and J. Tanner, *Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **367** (2009), no. 1906, 4273–4293. With electronic supplementary materials available online. MR2546388

[Don06] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52** (2006), no. 4, 1289–1306, DOI 10.1109/TIT.2006.871582. MR2241189

[DE03] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization*, Proc. Natl. Acad. Sci. USA **100** (2003), no. 5, 2197–2202 (electronic). MR1963681

[DET06] D. L. Donoho, M. Elad, and V. N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory **52** (2006), no. 1, 6–18. MR2237332

[DH01] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory **47** (2001), no. 7, 2845–2862. MR1872845

[DJ94] D. L. Donoho and I. M. Johnstone, *Minimax risk over l_p-balls for l_q-error*, Probab. Theory Related Fields **99** (1994), no. 2, 277–303. MR1278886

[DS89] D. L. Donoho and P. B. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. **49** (1989), no. 3, 906–931. MR997928

[DT09b] D. L. Donoho and J. Tanner, *Counting faces of randomly projected polytopes when the projection radically lowers dimension*, J. Amer. Math. Soc. **22** (2009), no. 1, 1–53, DOI 10.1090/S0894-0347-08-00600-0. MR2449053

[DVDD98] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, *Data compression and harmonic analysis*, IEEE Trans. Inform. Theory **44** (1998), no. 6, 2435–2476, DOI 10.1109/18.720544. Information theory: 1948–1998. MR1658775

[EK12] Y. C. Eldar and G. Kutyniok (eds.), *Compressed sensing. Theory and applications*, Cambridge University Press, Cambridge, 2012. MR2961961

[EGN+15] Y. Erlich, A. Gilbert, H. Ngo, A. Rudra, N. Thierry-Mieg, M. Wootters, D. Zielinski, and O. Zuk, *Biological screens from linear codes: theory and tools*, 2015. Unpublished.

[Faz02] M. Fazel, *Matrix rank minimization with applications*, Ph.D. Thesis, Palo Alto, 2002.

[Fie82] J. R. Fienup, *Phase retrieval algorithms: a comparison*, Appl. Opt. **21** (1982Aug), no. 15, 2758–2769.

[FPRU10] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich, *The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$*, J. Complexity **26** (2010), no. 6, 629–640. MR2735423

[FR13] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, New York, 2013. MR3100033

[Fuc05] J. J. Fuchs, *Recovery of exact sparse representations in the presence of bounded noise*, IEEE Trans. Inform. Theory **51** (2005), no. 10, 3601–3608. MR2237526

[GG84] A. Yu. Garnaev and E. D. Gluskin, *The widths of a Euclidean ball*, Dokl. Akad. Nauk SSSR **277** (1984), no. 5, 1048–1052. MR759962

[GS72] R. W. Gerchberg and W. O. Saxton, *A practical algorithm for the determination of the phase from image and diffraction plane pictures*, Optik **35** (1972), no. 2, 237–246.

[GI10] A. Gilbert and P. Indyk, *Sparse recovery using sparse matrices*, Proceedings of the IEEE **98** (2010June), no. 6, 937–947.

[GGIMS02] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, *Near-optimal sparse Fourier representations via sampling*, Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, 2002, pp. 152–161. MR2121138

[GGI+02b] A. C. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. J. Strauss, *Fast, small-space algorithms for approximate histogram maintenance*, Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, 2002, pp. 389–398. MR2121164

[GMS03] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, *Approximation of functions over redundant dictionaries using coherence*, Proceedings of the fourteenth annual acm-siam symposium on discrete algorithms, 2003, pp. 243–252.

[HS79] M. Harwit and N. J. A. Sloane, *Hadamard transform optics*, Academic Press, 1979.

[Kas77] B. S. Kašin, *The widths of certain finite-dimensional sets and classes of smooth functions*, Izv. Akad. Nauk SSSR Ser. Mat. **41** (1977), no. 2, 334–351, 478. MR0481792

[KMR14] F. Krahmer, S. Mendelson, and H. Rauhut, *Suprema of chaos processes and the restricted isometry property*, Comm. Pure Appl. Math. **67** (2014), no. 11, 1877–1904. MR3263672

[KM93] E. Kushilevitz and Y. Mansour, *Learning decision trees using the Fourier spectrum*, SIAM J. Comput. **22** (1993), no. 6, 1331–1348. MR1247193

[LF81] S. Levy and P. K. Fullagar, *Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution*, Geophysics **46** (1981), no. 9, 1235–1243, available at `http://geophysics.geoscienceworld.org/content/46/9/1235.full.pdf`.

[Log65] B. F. Logan, *Properties of high-pass signals*, Ph.D. Thesis, New York, 1965.

[LDP07] M. Lustig, D. Donoho, and J. M. Pauly, *Sparse mri: The application of compressed sensing for rapid mr imaging*, Magnetic Resonance in Medicine **58** (2007), no. 6, 1182–1195.

[Mal09] S. Mallat, *A wavelet tour of signal processing, The sparse way, 3rd ed.*, Elsevier/Academic Press, Amsterdam, 2009. MR2479996

[Man95] Y. Mansour, *Randomized interpolation and approximation of sparse polynomials*, SIAM J. Comput. **24** (1995), no. 2, 357–368. MR1320215

[MT13] M. McCoy and J. A. Tropp, *The achievable performance of convex demixing*, 2013. Available at `http://arXiv.org/abs/1309.7478`.

[MT14] M. B. McCoy and J. A. Tropp, *From Steiner formulas for cones to concentration of intrinsic volumes*, Discrete Comput. Geom. **51** (2014), no. 4, 926–963. MR3216671

[Mil02] A. Miller, *Subset selection in regression*, Second, Monographs on Statistics and Applied Probability, vol. 95, Chapman & Hall/CRC, Boca Raton, FL, 2002. MR2001193

[Nat95] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput. **24** (1995), no. 2, 227–234. MR1320206

[OSL83] D. W. Oldenburg, T. Scheuer, and S. Levy, *Recovery of the acoustic impedance from reflection seismograms*, GEOPHYSICS **48** (1983), no. 10, 1318–1337, available at `http://dx.doi.org/10.1190/1.1441413`.

[OT15] S. Oymak and J. A. Tropp, *Universality laws for randomized dimension reduction, with applications*, 2015. Available at `http://arXiv.org/abs/1511.09433`.

[RFP10] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev. **52** (2010), no. 3, 471–501. MR2680543

[RV08] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633. MR2407948

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), no. 1-4, 259–268. Experimental mathematics: computational issues in nonlinear science, (Los Alamos, NM, 1991). MR3363401

[SS86] F. Santosa and W. W. Symes, *Linear inversion of band-limited reflection seismograms*, SIAM J. Sci. Statist. Comput. **7** (1986), no. 4, 1307–1330. MR857796

[Sto13] M. Stojnic, *Regularly random duality*, 2013. Available at `http://arXiv.org/abs/1303.7295`.

[Tao05] T. Tao, *An uncertainty principle for cyclic groups of prime order*, Math. Res. Lett. **12** (2005), no. 1, 121–127. MR2122735

[TBM79] H. L. Taylor, S. C. Banks, and J. F. McCoy, *Deconvolution with the l 1 norm*, Geophysics **44** (1979), no. 1, 39–52, available at `http://geophysics.geoscienceworld.org/content/44/1/39.full.pdf`.

[TOH14] C. Thrampoulidis, S. Oymak, and B. Hassibi, *The Gaussian min-max theorem in the presence of convexity*, 2014. Available at `http://arXiv.org/abs/1408.4837`.

[Thr16] C. Thrampoulidis, *Recovering structured signals in high dimensions via non-smooth convex optimization: Precise performance analysis.*, Ph.D. Thesis, Pasadena, 2016.

[TOH15] C. Thrampoulidis, S. Oymak, and B. Hassibi, *Recovering structured signals in noise: least-squares meets compressed sensing*, Compressed sensing and its applications, 2015, pp. 97–141. MR3382104

[Tib96] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B **58** (1996), no. 1, 267–288. MR1379242

[TW10] J. A. Tropp and S. J. Wright, *Computational methods for sparse solution of linear inverse problems*, Proceedings of the IEEE **98** (2010June), no. 6, 948–958.

[Tro04] J. A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory **50** (2004), no. 10, 2231–2242. MR2097044

[Tro06] J. A. Tropp, *Just relax: convex programming methods for identifying sparse signals in noise*, IEEE Trans. Inform. Theory **52** (2006), no. 3, 1030–1051. MR2238069

[Tro12] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math. **12** (2012), no. 4, 389–434. MR2946459

[TLD+10] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, *Beyond Nyquist: efficient sampling of sparse bandlimited signals*, IEEE Trans. Inform. Theory **56** (2010), no. 1, 520–544. MR2589462

[VMA+11] S. Vasanawala, M. Murphy, M. Alley, P. Lai, K. Keutzer, J. Pauly, and M. Lustig, *Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients*, 2011 ieee international symposium on biomedical imaging: From nano to macro, 2011March, pp. 1039–1043.

Joel A. Tropp
California Institute of Technology, Pasadena, California 91125-5000
E-mail address: jtropp@cms.caltech.edu