Compressive Sensing, Structured Random Matrices and Recovery of Functions in High Dimensions

Holger Rauhut

Compressive sensing predicts that sparse vectors can be recovered efficiently from highly undersampled measurements. While it is well-understood by now that Gaussian random matrices provide optimal measurement matrices in this context, such "highly" random matrices suffer from certain drawbacks: applications require more structure arising from physical or other constraints, and recovery algorithms such as greedy methods or algorithms for ℓ_1-minimization demand fast matrix vector multiplies in order to make them feasible for large scale problems. In order to meet such desiderata, we study two types of structured random measurement matrices: partial random circulant matrices, and random sampling matrices associated to bounded orthonormal systems (e.g. random Fourier type matrices). The latter maybe used to study reconstruction problems in high spatial dimensions.

Compressive Sensing. A vector $x \in \mathbb{C}^N$ is called s-sparse if $\|x\|_0 := \#\{\ell, x_\ell \neq 0\} \leq s$. The ℓ_p-norm is defined as usual, $\|x\|_p := (\sum_{\ell=1}^N |x_\ell|^p)^{1/p}$, $0 < p < \infty$. The best s-term approximation error of an arbitrary vector in ℓ_p is defined as

$$\sigma_s(x)_p = \inf_{\|z\|_0 \leq s} \|x - z\|_p.$$

Informally, x is called compressible if $\sigma_s(x)_p$ decays quickly in s. An estimate originally due to Stechkin states that $\sigma_s(x)_p \leq s^{1/p-1/q}\|x\|_q$ for $q < p$ so that $B_q^N = \{x \in \mathbb{C}^N, \|x\|_q \leq 1\}$ is a good model for compressible vectors if $q \in (0,1]$ is chosen small.

The task of compressive sensing is to recover a sparse or compressible vector $x \in \mathbb{C}^N$ from undersampled measurements

$$y = Ax \ \in \mathbb{C}^m,$$

where $A \in \mathbb{C}^{m \times N}$ is a suitable measurement matrix and $m \ll N$. The first approach for recovering x that probably comes to mind consists in solving the ℓ_0-minimization problem

$$\min_{z \in \mathbb{C}^N} \|z\|_0 \quad \text{subject to } Az = y.$$

Unfortunately, this combinatorial optimization problem is NP hard in general. For this reason, several tractable alternatives have been introduced, most notably ℓ_1-minimization, which consists in solving the convex optimization problem

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \quad \text{subject to } Az = y.$$

A very useful concept for analyzing ℓ_1-minimization are the restricted isometry constants. For $s < N$ they are defined as the smallest constant δ_s such that

$$(1-\delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_s)\|x\|_2^2 \quad \text{for all } x \in \mathbb{C}^N, \|x\|_0 \leq s.$$

If $\delta_{2s} < 0.46$ then ℓ_1-minimization reconstructs all s-sparse vectors x exactly from $y = Ax$, and compressible vectors approximately, see [1, 4] for precise statements.

It is an open problem to construct deterministic (explicit) measurement matrices that have small restricted isometry constants for small m (or large s, respectively). So far, all good constructions use randomness. A matrix with independent standard normal distributed entries is called a Gaussian random matrix. It is by now well-known [2] that a rescaled Gaussian matrix $\frac{1}{\sqrt{m}}A \in \mathbb{R}^{m\times N}$ satisfies $\delta_s \le \delta$ with probability at least $1-\varepsilon$ provided

$$m \ge C\delta^{-2}(s\log(N/s) + \log(\varepsilon^{-1})),$$

where $C > 0$ is a universal constant. In particular, exact recovery of s-sparse vectors via ℓ_1-minimization is possible if $m \asymp s\log(N/s)$. This estimate for the minimal number m of measurements is optimal as follows from lower bounds of Gelfand widths of B_q^N [5].

Partial random circulant matrices. While Gaussian random matrices are optimal for compressive sensing, they are not structured at all, which poses severe limitations for practical applications as mentioned above. Therefore, we consider instead the following structured random matrix. For a vector $b \in \mathbb{C}^N$ we define its associated circulant matrix $\Phi = \Phi(b) \in \mathbb{C}^{N\times N}$ with entries

$$\Phi_{k,j} = b_{j-k \mod N}, \quad k,j = 1,\dots,N.$$

For an abritrary subset $\Theta \subset \{1,\dots,N\}$ we define the restriction operator $R_\Theta : \mathbb{C}^N \to \mathbb{C}^\Theta$ as $(R_\Theta x)_\ell = x_\ell$, $\ell \in \Theta$. Then the partial circulant matrix $\Phi^\Theta = \Phi^\Theta(b) = R_\Theta\Phi(b)$ consists of the rows of $\Phi = \Phi(b)$ indexed by the set Θ. An application of Φ^Θ to a vector x corresponds to convolution with b followed by subsampling on Θ. Since a circulant matrix can be diagonalized by the Fourier matrix, the FFT can be used for fast matrix vector multiplies. For the purpose of compressive sensing, the vector b is chosen at random, more precisely, as Rademacher sequence, that is, all entries are independent, and take the value $+1$ or -1 with equal probability. This turns the matrix $\Phi^\Theta = \Phi^\Theta(b)$ into a partial random circulant matrix. In [6] the following nonuniform recovery result for Φ^Θ has been shown.

Theorem. *Let $\Theta \subset \{1,\dots,N\}$ be an arbitrary (deterministic) set of cardinality m. Let $x \in \mathbb{C}^N$ be s-sparse such that the signs of its non-zero entries form a Rademacher or Steinhaus sequence. Choose $b \in \{-1,+1\}^N$ to be a Rademacher sequence. Let $y = \Phi^\Theta(b)x \in \mathbb{C}^m$. If*

$$m \ge 57s\ln^2(17N^2/\varepsilon)$$

then x can be recovered from y via ℓ_1-minimization with probability at least $1-\varepsilon$.

Unfortunately, this result does not imply the existence of a single matrix $\Phi^\Theta(b)$ that guarantees recovery of all s-sparse vectors simultaneously. Such type of statement is implied by the next theorem on the restricted isometry constants shown in [7].

Theorem. *Let $\Theta \subset \{1,\dots,N\}$ be an arbitrary (deterministic) set of cardinality m. Choose $b \in \mathbb{R}^N$ to be a Rademacher sequence. Assume that*

$$m \ge C\delta^{-1}s^{3/2}\log^{3/2}(N), \tag{1}$$

and, for $\varepsilon \in (0,1)$, $m \ge C\delta^{-2}s\log^2(s)\log^2(N)\log(\varepsilon^{-1})$. Then with probability at least $1-\varepsilon$ the restricted isometry constants of $\frac{1}{\sqrt{m}}\Phi^\Theta(b)$ satisfy $\delta_s \le \delta$.

The exponent $3/2$ in (1) does not seem to be optimal. Unfortunately, the proof technique in [7] is likely not powerful enough in order to obtain the expected exponent 1.

Random Sampling in Bounded Orthonormal Systems. Let $\Omega \subset \mathbb{R}^d$ be endowed with a probability measure ν, and $\phi_1, \ldots, \phi_N$ be a system of orthonormal functions, i.e., $\int_\Omega \phi_j(t)\overline{\phi_k(t)}d\nu(t) = \delta_{j,k}$. We further assume that the function system is bounded in the sense that

$$\sup_{j=1,\ldots,N} \|\phi_j\|_\infty \leq K$$

for some constant $K \geq 1$. A function of the form

$$f(t) = \sum_{j=1}^{N} x_j \phi_j(t)$$

is called s-sparse if $\|x\|_0 \leq s$. Our goal is to reconstruct sparse (or compressible) functions from sample values $f(t_1), \ldots, f(t_m)$ with $t_1, \ldots, t_m \in \Omega$. Introducing the sampling matrix $A \in \mathbb{C}^{m \times N}$ with entries $A_{k,j} = \phi_j(t_k)$ yields $y = (f(t_1), \ldots, f(t_m))^T = Ax$. Therefore, we are interested in the restricted isometry constants of the sampling matrix. We choose the points $t_1, \ldots, t_m$ independent and distributed according to ν. This makes A a structured random matrix. The most important example consists in choosing $\phi_j(t) = e^{2\pi i t \cdot j}, j \in \mathbb{Z}^d, t \in [0,1]^d$, $\Omega = [0,1]^d$ and ν to be the Lebesgue measure. The resulting sampling matrix is then a non-equispaced Fourier matrix for which fast (approximate) matrix vector multiplies are available. In [6] the following estimate for the restricted isometry constants has been derived, generalizing and improving slightly on [2, 9].

Theorem. *Let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated to a bounded orthonormal system with constant $K \geq 1$. If*

$$\frac{m}{\ln(m)} \geq CK^2\delta^{-2}s\ln^2(s)\ln(N).$$

then with probability at least $1 - N^{-\gamma \ln^2(s)\ln(m)}$ the restricted isometry constant of $\frac{1}{\sqrt{m}}A$ satisfies $\delta_s \leq \delta$. The constants $C, \gamma > 0$ are universal.

This result can be used to extend reconstruction from sample values via compressive sensing to infinite dimensional function spaces, and in particular, to suitable spaces of functions of many variables. We refer to [3, 8] for details.

References

[1] E. Candès, The restricted isometry property and its implications for compressed sensing, *C. R. Acad. Sci. Paris S'er. I Math.* 346, 589–592, 2008.

[2] E. Candès and T. Tao, Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory*, 52, 5406–5425, 2006.

[3] A. Cohen, R. DeVore, S. Foucart, H. Rauhut, Recovery of functions in high dimensions via compressive sensing, in preparation.

[4] S. Foucart, A note on guaranteed sparse recovery via ℓ_1-minimization, *Appl. Comput. Harmon. Anal.*, to appear.

[5] S. Foucart, A. Pajor, H. Rauhut, T. Ullrich, The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$, *J. Complexity*, to appear.

[6] H. Rauhut, Compressive sensing and structured random matrices, In: M. Fornasier (ed.), *Theoretical Foundations and Numerical Methods for Sparse Recovery*, 1–92, deGruyter, 2010.

[7] H. Rauhut, J. Romberg, J. Tropp, Restricted isometries for partial random circulant matrices, in preparation.

[8] H. Rauhut, R. Ward, Sparse Legendre expansions via ℓ_1-minimization, preprint, 2010.

[9] M. Rudelson and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, *Comm. Pure Appl. Math.* 61, 1025–1045, 2008.