**Mathematical Foundations of Data Sciences**

Gabriel Peyré
CNRS & DMA, École Normale Supérieure

**CHAPTER 12**

# Compressed Sensing

September 9, 2026

Compressed sensing aims to recover sparse signals from far fewer measurements than their ambient dimension. This can reduce acquisition costs when measurements are expensive, but success depends on how the sensing operator interacts with the signal representation. Motivated by the single-pixel camera, we analyze recovery from random measurements, distinguish guarantees for fixed signals from uniform guarantees, and extend the discussion to structured Fourier sampling.

Hardware constraints often prevent direct implementation of an ideal random matrix. The theory nevertheless guides the design of structured acquisition schemes, whose physical assumptions and computational costs must be considered separately.

The main references for this chapter are [5, 4, 6].

## 12.1 Motivation and Potential Applications

### 12.1.1 Single-Pixel Camera

The "single-pixel camera" developed at Rice University [3] motivates compressed acquisition. Figure 12.1 illustrates the reconstruction principle with simulated measurements of a flower image. A central idea is to combine acquisition and compression in a single device. Conventional acquisition first samples the analog scene $\tilde{f}_0$ at high resolution to obtain an image $f_0 \in \mathbb{R}^Q$, then compresses it to $M$ significant coefficients using (5.3). Compressed acquisition instead records a vector $y \in \mathbb{R}^P$ directly. The aim is to keep the measurement count $P$ close to the sparsity level $M$, typically with a logarithmic overhead, while retaining enough information for accurate reconstruction.

The scene $\tilde{f}_0$ is modeled as a light-intensity function. Its value $\tilde{f}_0(s)$ gives the intensity at position $s \in \mathbb{R}^2$ in the camera's focal plane. Light is focused onto an array of $Q$ micromirrors in the focal plane. Each mirror directs light either toward the sensor or away from it; the mirrors themselves record no information. The device cycles rapidly through $P$ mirror configurations. For $p = 1, \dots, P$, the entry $\Phi_{p,q} \in \{0,1\}$ indicates whether mirror $q$ is off or on during exposure $p$. A single sensor integrates the light directed toward it during each exposure, producing one measurement $y_p \in \mathbb{R}$. Let $c_q$ be the region occupied by mirror $q$ in the focal plane, and define its integrated intensity by $f_{0,q} = \int_{c_q} \tilde{f}_0(s)\,\mathrm{d}s$. The discrete image $f_0 \in \mathbb{R}^Q$ and the measurements $y \in \mathbb{R}^P$ are then related by

$$\forall\, p = 1, \dots, P, \quad y_p \approx \sum_{q=1}^{Q} \Phi_{p,q} \int_{c_q} \tilde{f}_0(s)\mathrm{d}s = (\Phi f_0)_p,$$

The symbol $\approx$ accounts for acquisition noise. We therefore use the usual inverse-problem model

$$y = \Phi f_0 + w \in \mathbb{R}^P$$

where $w$ is the noise vector. The intermediate image $f_0$ is never recorded: the device measures $y$ directly from the analog scene $\tilde{f}_0$. Binary mirror patterns are not centered random variables. Centered signed measurements can be formed by subtracting an appropriate total-intensity measurement or by differencing complementary

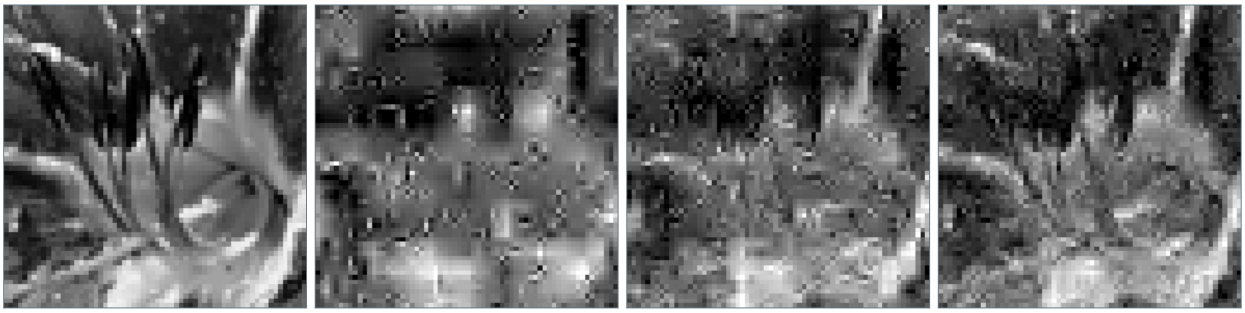


**Figure 12.1.** Simulated compressed sensing of the flower image, with $Q = 4096$ pixels. The three wavelet-$\ell^1$ reconstructions use $P = 512$, 1024, and 2048 orthonormal signed Walsh–Hadamard measurements, respectively.

exposures, followed by normalization. This distinction matters when applying the random-matrix assumptions below.

Recovering the ideal discrete image $f_0$ from $y$ requires inverting the acquisition model while accounting for the noise $w$. Figure 12.1 uses nested subsets of signed Walsh–Hadamard patterns with a fixed random pixel permutation. Each reconstruction minimizes the same wavelet-$\ell^1$ objective with its own measurements. The examples are noiseless simulations; the signed measurements correspond to centered patterns, rather than individual binary exposures.

### 12.1.2 Sparse Recovery

Following Section 10.2, write $f_0 = \Psi x_0$, where $\Psi \in \mathbb{R}^{Q\times N}$ is a dictionary and $x_0$ is sparse. Denoting $A \overset{\text{def.}}{=} \Phi\Psi \in \mathbb{R}^{P\times N}$, this leads us to consider the usual $\ell^1$ regularized problem (10.10)

$$x_\lambda \in \underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2\lambda}\|y - Ax\|^2 + \|x\|_1, \tag{$\mathcal{P}_\lambda(y)$}$$

so that the reconstructed image is $f_\lambda = \Psi x_\lambda$. We also sometimes consider the constrained problem

$$x_\varepsilon \in \underset{\|Ax-y\|\leqslant\varepsilon}{\operatorname{argmin}} \|x\|_1, \tag{$\mathcal{P}^\varepsilon(y)$}$$

Assume that a noise bound $\|w\| \leqslant \varepsilon$ is known, so that the true coefficient vector is feasible. Convex optimality conditions relate the two formulations through suitable Lagrange multipliers. The parameter correspondence depends on $y$ and need not be one-to-one; a range of parameter values may produce the same solution. The endpoint $\varepsilon = 0$ may require a limiting parameter value.

The distribution of $A = \Phi\Psi$ depends on both the acquisition matrix and the dictionary. If $\Psi$ is square and orthogonal and $\Phi$ has independent centered Gaussian entries with a common variance, right orthogonal invariance gives $\Phi\Psi$ the same distribution as $\Phi$. For a general dictionary, this simplification fails. The results below therefore impose distributional assumptions directly on $A$.

## 12.2 Dual Certificate Theory and Non-Uniform Guarantees

### 12.2.1 Random Projection of Polytopes

For noiseless data, $w = 0$, we can use the geometric characterization of $(\mathcal{P}^0(Ax_0)) = (\mathcal{P}_0(Ax_0))$ from Section 11.1.2. Proposition 11.1 relates identifiable $s$-sparse vectors to $(s-1)$-dimensional faces of the $\ell^1$ ball $B_1$ that survive projection onto $AB_1$. Recovery thus becomes a problem of counting faces of random polytopes. For Gaussian projections, the asymptotic analysis of Donoho and Tanner distinguishes two thresholds, described by functions $C_A$ and $C_M$. In the proportional-growth regime, with probability tending to one as the dimensions increase,

– Every vector $x_0$ with $\|x_0\|_0 \leqslant C_A(P/N)P$ is identifiable.
– A typical support/sign pattern with sparsity at most $C_M(P/N)P$ is identifiable.

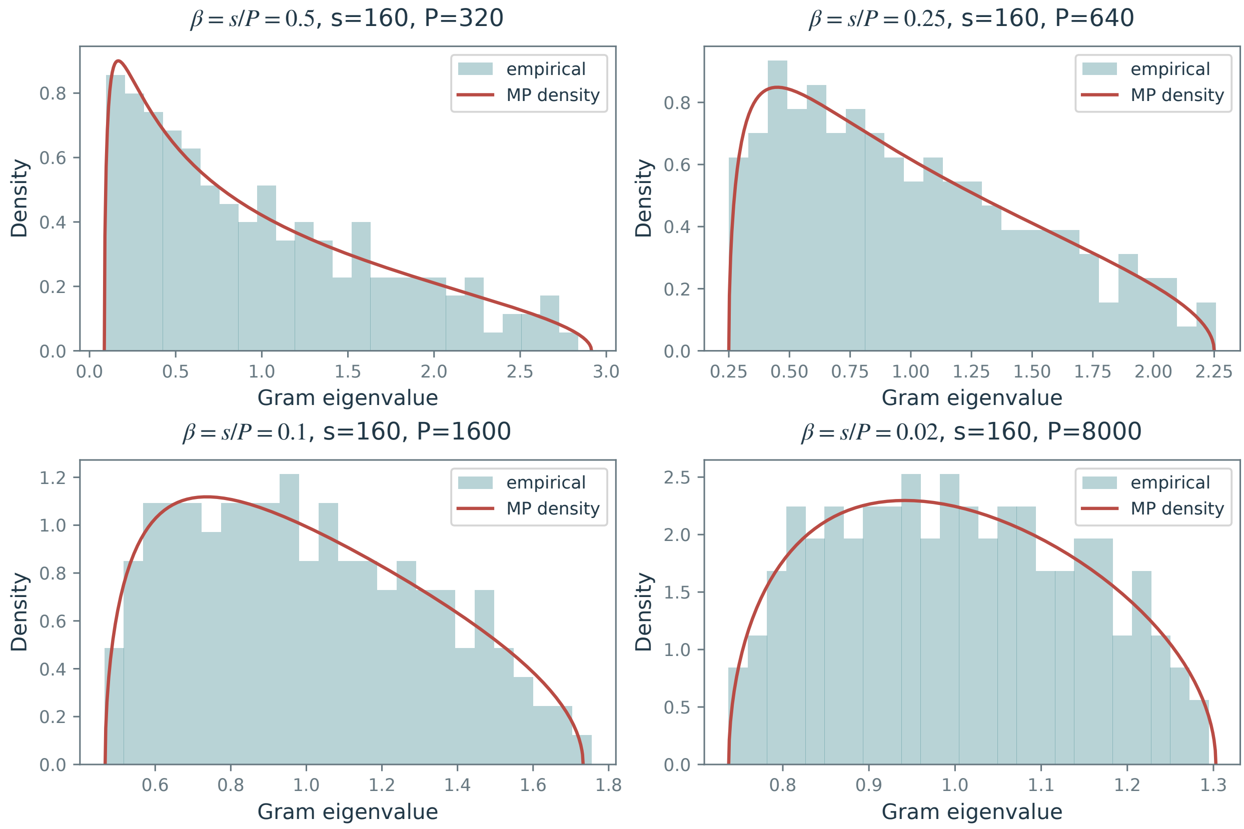


**Figure 12.2.** Empirical eigenvalue distributions approaching the Marchenko–Pastur law.

For illustration, at $P/N = 1/4$, the weak threshold is roughly one quarter of $P$. Strong thresholds require recovery for every support/sign pattern; weak thresholds concern a typical pattern. Their precise values depend on the asymptotic regime. Figure 12.5 illustrates the two transitions numerically. Above the weak threshold $C_M(P/N)P$, a typical support/sign pattern fails to be identifiable with high probability. The sharp transition between these regimes is called a phase transition. The function $C_M$ can be computed numerically and has inverse-logarithmic decay, $C_M(r) \asymp 1/\log(1/r)$, as $r \to 0$. Thus, in the high-compression regime, typical sparse vectors remain recoverable when their sparsity is proportional to the number of measurements, up to a logarithmic factor.

### 12.2.2 Random Matrices

To prove recovery guarantees, we need quantitative control of the singular values of random column submatrices. The Gaussian ensemble provides a useful starting point.

Let $I$ contain $s$ column indices and set $B = A_I \in \mathbb{R}^{P \times s}$. Throughout this subsection, the entries of $A$ are independent $\mathcal{N}(0, 1/P)$ variables. The Gram matrix $B^*B$ is a scaled Wishart matrix; when it is invertible, its inverse has the corresponding inverse-Wishart distribution. We need to control both matrices. For fixed $s$, the law of large numbers gives $B^*B \to \mathrm{Id}_s$ almost surely as $P \to \infty$. Compressed sensing also requires estimates when $s$ grows with $P$.

**Proportional growth.** Suppose $s, P \to \infty$ with $s/P \to \beta \in (0, 1)$. The empirical eigenvalue distribution of $B^*B$ converges almost surely to the Marchenko–Pastur law:

$$\frac{1}{s}\sum_{j=1}^{s} \delta_{\lambda_j(B^*B)} \rightharpoonup f_\beta(\lambda)\,\mathrm{d}\lambda, \qquad f_\beta(\lambda) = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\beta\lambda}\,1_{[\lambda_-,\lambda_+]}(\lambda),$$

where $\lambda_\pm = (1 \pm \sqrt{\beta})^2$. For Gaussian matrices, the extreme eigenvalues also converge to these endpoints. When $\beta < 1$ there is no atom at zero; such an atom appears when $s > P$. Figure 12.2 illustrates this convergence.

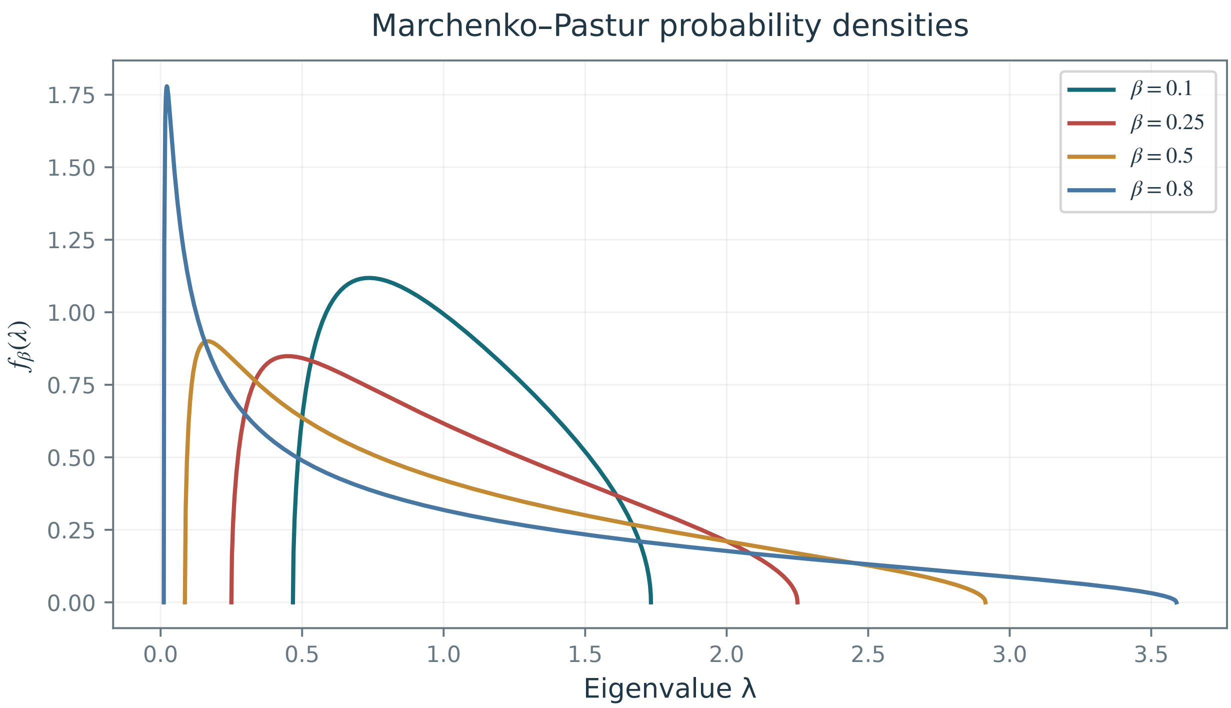


**Figure 12.3.** Marchenko–Pastur densities $f_\beta$ for several aspect ratios $\beta$.

**Finite-sample concentration.** For a fixed support $I$ of size $s$, Gaussian singular-value concentration gives, for $t > 0$, probability at least $1 - 2e^{-Pt^2/2}$ that

$$1 - \sqrt{s/P} - t \leqslant \sigma_{\min}(A_I) \leqslant \sigma_{\max}(A_I) \leqslant 1 + \sqrt{s/P} + t. \tag{12.1}$$

In particular, writing $a = \sqrt{s/P} + t$, on this event

$$\|A_I^* A_I - \mathrm{Id}_s\|_{\mathrm{op}} \leqslant 2a + a^2.$$

Thus the singular-value bounds also control the deviation of the Gram matrix from the identity. They show how increasing the measurement count $P$ relative to the support size $s$ improves conditioning.

### 12.2.3 Dual Certificates

We seek both a small $\ell^2$ reconstruction error and exact support recovery at low noise levels. Section 11.2.3 gives a sufficient condition: the Fuchs precertificate (11.16) must be nondegenerate, meaning

$$\|\eta_{F,I^c}\|_\infty < 1 \quad \text{where} \quad \eta_F = A^* A_I (A_I^* A_I)^{-1} \operatorname{sign}(x_{0,I}). \tag{12.2}$$

Figure 11.9 suggests nondegeneracy when $P$ is sufficiently large relative to $\|x_0\|_0$. We now quantify this observation.

**Coherence-based analysis.** We begin with a bound based on the coherence of $A = (a_j)_{j=1}^N$. Assume that each column $a_j \in \mathbb{R}^P$ has unit norm. The coherence is

$$\mu \stackrel{\text{def.}}{=} \max_{i \neq j} |\langle a_i, a_j \rangle| \tag{12.3}$$

This is an entrywise maximum, not the induced $\ell^\infty$ matrix norm used below. An orthogonal matrix has coherence 0. In general, coherence is at most 1, so $\mu \in [0, 1]$. Small coherence controls the conditioning of submatrices with few columns and gives a sufficient condition for nondegeneracy of $\eta_F$.

**Proposition 12.1.** *Let $I = \operatorname{supp}(x_0)$ and $s = |I| \geqslant 1$. If $(s-1)\mu < 1$, then*

$$\|\eta_{F,I^c}\|_\infty \leqslant \frac{s\mu}{1-(s-1)\mu} \quad \textit{and} \quad \|p_F\|^2 \leqslant \frac{s}{1-(s-1)\mu}. \tag{12.4}$$

*In particular, if $\mu > 0$ and $s < \frac{1}{2}\left(1+\frac{1}{\mu}\right)$, then $\|\eta_{F,I^c}\|_\infty < 1$, so Theorem 11.15 applies. If $\mu = 0$, the columns are orthonormal and the conclusion holds directly.*

*Proof.* We recall that the $\ell^\infty$ operator norm (see Remark 11.14) is

$$\|B\|_\infty = \max_i \sum_j |B_{i,j}|.$$

Write $s_I = \operatorname{sign}(x_{0,I})$ and $C = A^*A$. Then

$$\|A_{I^c}^* A_I\|_\infty = \max_{j\in I^c} \sum_{i\in I} |C_{i,j}| \leqslant s\mu \quad \text{and} \quad \|\mathrm{Id}_s - A_I^* A_I\|_\infty = \max_{j\in I} \sum_{i\in I, i\neq j} |C_{i,j}| \leqslant (s-1)\mu$$

One also has

$$\begin{aligned}\|(A_I^*A_I)^{-1}\|_\infty &= \|(\mathrm{Id}_s - (\mathrm{Id}_s - A_I^*A_I))^{-1}\|_\infty = \|\sum_{k\geqslant 0} (\mathrm{Id}_s - A_I^*A_I)^k\|_\infty \\ &\leqslant \sum_{k\geqslant 0} \|\mathrm{Id}_s - A_I^*A_I\|_\infty^k \leqslant \sum_{k\geqslant 0} ((s-1)\mu)^k = \frac{1}{1-(s-1)\mu}\end{aligned}$$

The Neumann series converges because $(s-1)\mu < 1$. Using these two bounds, one has

$$\|\eta_{F,I^c}\|_\infty = \|A_{I^c}^* A_I (A_I^*A_I)^{-1} \operatorname{sign}(x_{0,I})\|_\infty \leqslant \|A_{I^c}^* A_I\|_\infty \|(A_I^*A_I)^{-1}\|_\infty \|\operatorname{sign}(x_{0,I})\|_\infty \leqslant (s\mu) \times \frac{1}{1-(s-1)\mu} \times 1.$$

One has

$$\frac{s\mu}{1-(s-1)\mu} < 1 \quad \Longleftrightarrow \quad 2s\mu < 1+\mu$$

which gives the last statement. One also has

$$\|p_F\|^2 = \langle (A_I^*A_I)^{-1} s_I, s_I \rangle \leqslant \|(A_I^*A_I)^{-1}\|_\infty \|s_I\|_1 \leqslant \frac{s}{1-(s-1)\mu}$$

□

The proof applies to every sign pattern on every support satisfying the coherence condition. It also yields support containment for arbitrary noise levels when $\lambda$ satisfies the noise-dependent ERC bound from the preceding chapter. An arbitrary choice of $\lambda$ need not preserve the support.

For $N > P$, the Welch bound gives

$$\mu \geqslant \sqrt{\frac{N-P}{P(N-1)}} \tag{12.5}$$

which is equivalent to $1/\sqrt{P}$ for $N \gg P$. For Gaussian matrices with normalized columns $A \in \mathbb{R}^{P\times N}$, a union bound gives, with high probability when $\log N$ is small relative to $P$,

$$\mu = O\left(\sqrt{\log N/P}\right)$$

Thus Gaussian matrices attain the Welch scale up to a factor of order $\sqrt{\log N}$ when $N \gg P$. Ignoring logarithmic factors, Proposition 12.1 guarantees support stability for sparsities of order $\sqrt{P}$. This coherence bound is conservative: a direct probabilistic analysis permits sparsities of order $P$, again up to logarithmic factors.

**Randomized analysis of the Fuchs certificate.** For the probabilistic analysis, we specify independent entries with a common variance and a uniform Gaussian-type tail bound.

**Definition 12.2** (sub-Gaussian random matrix)**.** A random matrix $A \in \mathbb{R}^{P\times N}$ is called normalized sub-Gaussian if its entries are independent, $\mathbb{E}(A_{i,j}) = 0$, $\mathbb{E}(A_{i,j}^2) = 1/P$, and, for some constants $\beta, \kappa > 0$ and every $t > 0$,

$$\mathbb{P}(|\sqrt{P}A_{i,j}| \geqslant t) \leqslant \beta e^{-\kappa t^2}. \tag{12.6}$$

The entries may have different distributions, but the tail constants $(\beta, \kappa)$ must be uniform over $(i, j)$. The normalization by $\sqrt{P}$ is consistent with unit expected squared column norm, $\mathbb{E}(\|a_j\|^2) = 1$, where $a_j$ denotes a column of $A$.

Examples include Gaussian matrices with entry variance $1/P$ and matrices with independent entries taking the values $\pm 1/\sqrt{P}$ with equal probability.

**Theorem 12.3.** *Fix a nonzero $x_0 \in \mathbb{R}^N$, independently of $A$, and write $s = \|x_0\|_0$. For a normalized sub-Gaussian matrix, there is a constant $C$, depending only on the sub-Gaussian parameters, such that*

$$P \geqslant Cs\log(2N/\varepsilon), \qquad 0 < \varepsilon < 1,$$

*implies, with probability at least $1-\varepsilon$, that $A_I$ is injective and $\|\eta_{F,I^c}\|_\infty < 1$. The support-stability conclusion of Theorem 11.15 then holds for sufficiently small $\lambda$ and sufficiently small $\|A^*w\|_\infty/\lambda$.*

*Proof.* We outline the concentration argument. With probability at least $1-\varepsilon/2$, taking $P$ as in the theorem with a sufficiently large constant gives $\sigma_{\min}(A_I) \geqslant 1/2$, hence

$$\|p_F\|^2 = \langle (A_I^*A_I)^{-1}s_I, s_I\rangle \leqslant 4s, \qquad s_I = \mathrm{sign}(x_{0,I}). \tag{12.7}$$

Conditioning on $A_I$ fixes $p_F$, while the off-support columns $a_j$ remain independent of this vector. The probability that $|\langle a_j, p_F\rangle|$ reaches or exceeds 1 is bounded by $2\exp(-cP/\|p_F\|^2)$. A union bound over at most $N$ columns gives the stated sample complexity and the strict off-support bound, by excluding $|\langle a_j, p_F\rangle| \geqslant 1$.

For Gaussian entries, conditional on $A_I$, these inner products are Gaussian with variance $\|p_F\|^2/P$. When $s/P$ is small, $\|p_F\|^2$ is close to $s$, so the largest of the $N-s$ absolute correlations is approximately $\sqrt{2s\log(N-s)/P}$. This gives the sharper leading-order Gaussian threshold $P \approx 2s\log N$.

Rotational invariance makes this statement more precise:

$$\frac{Ps}{\|p_F\|^2} \sim \chi^2_{P-s+1}.$$

To see this, rotate the $s$ columns so that $s_I/\sqrt{s}$ becomes the first coordinate vector. The inverse of the corresponding diagonal entry of $(G^*G)^{-1}$, for $G = \sqrt{P}A_I$, is the squared distance of the first column of $G$ to the span of the remaining columns. This is a chi-square variable with $P-s+1$ degrees of freedom. Thus $\|p_F\| \leqslant (1+\delta)\sqrt{s}$ with high probability when $s/P$ is sufficiently small and $P$ is large; the factor must exceed one. □

In a Gaussian asymptotic regime with $s/P \to 0$, the leading-order scaling is

$$P \gtrsim 2s\log(2N/\varepsilon), \tag{12.8}$$

with slack needed for a finite-sample probability bound. The unspecified constant in Theorem 12.3 applies to general sub-Gaussian ensembles; it should not be identified with this Gaussian leading constant.

This is a nonuniform guarantee: the vector $x_0$ is fixed before the matrix $A$ is drawn. The RIP theory in Section 12.3 gives a single event on which recovery holds for all sparse vectors simultaneously.

## 12.3 RIP Theory for Uniform Guarantees

### 12.3.1 Restricted Isometry Constants

For a positive integer $s$, the restricted isometry constant $\delta_s$ of $A \in \mathbb{R}^{P\times N}$ is the smallest nonnegative number such that

$$\forall z \in \mathbb{R}^N, \quad \|z\|_0 \leqslant s \quad \Longrightarrow \quad (1-\delta_s)\|z\|^2 \leqslant \|Az\|^2 \leqslant (1+\delta_s)\|z\|^2, \tag{12.9}$$

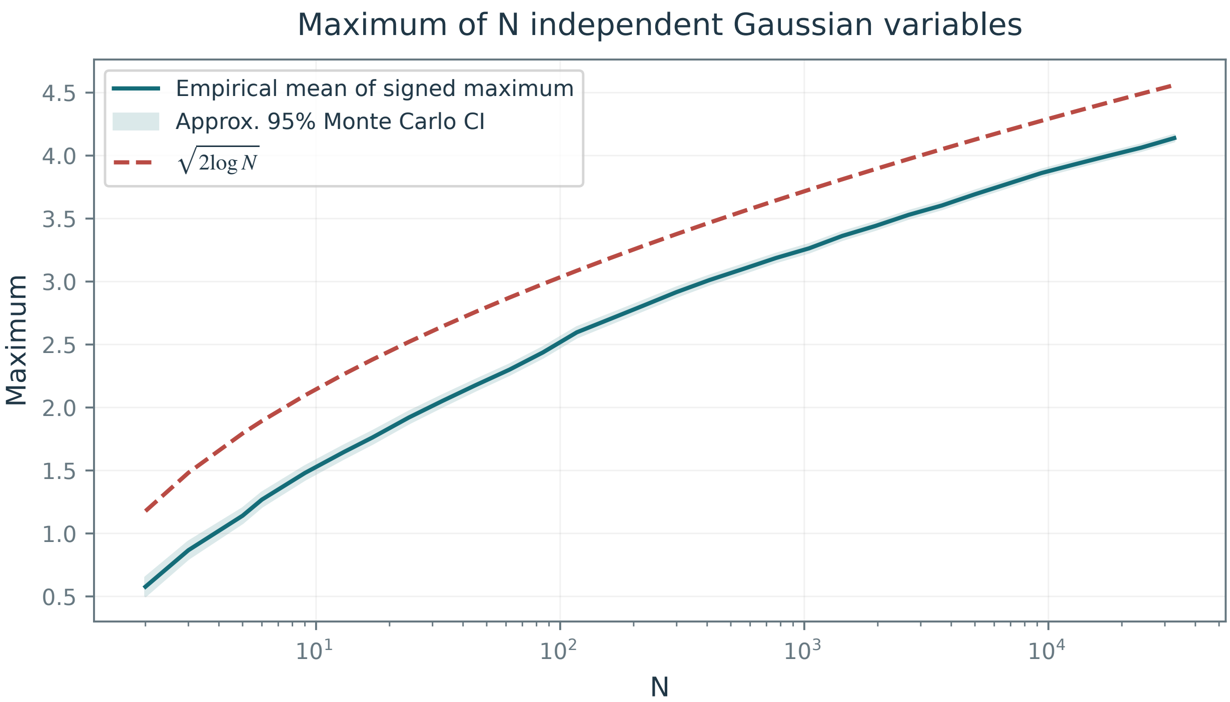


**Figure 12.4.** Maximum of $N$ independent standard Gaussian variables, compared with the dashed curve $\sqrt{2\log(N)}$.

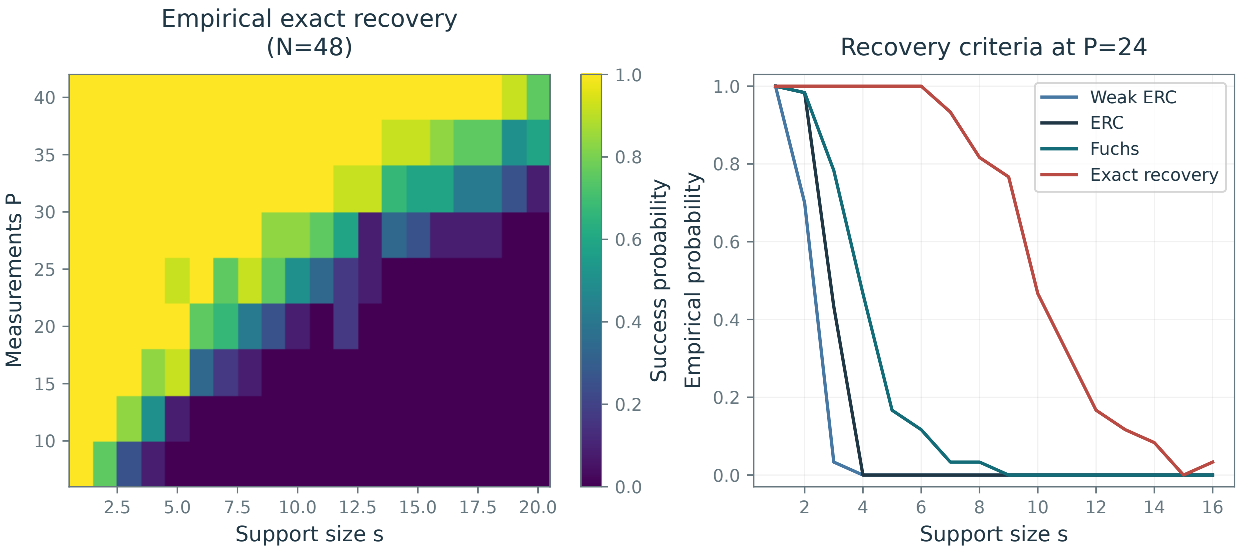


**Figure 12.5.** Phase transitions. Right: empirical probabilities that recovery criteria hold as sparsity varies. Blue: weak ERC; black: ERC; green: $\|\eta_F\|_\infty \leqslant 1$; red: exact recovery.

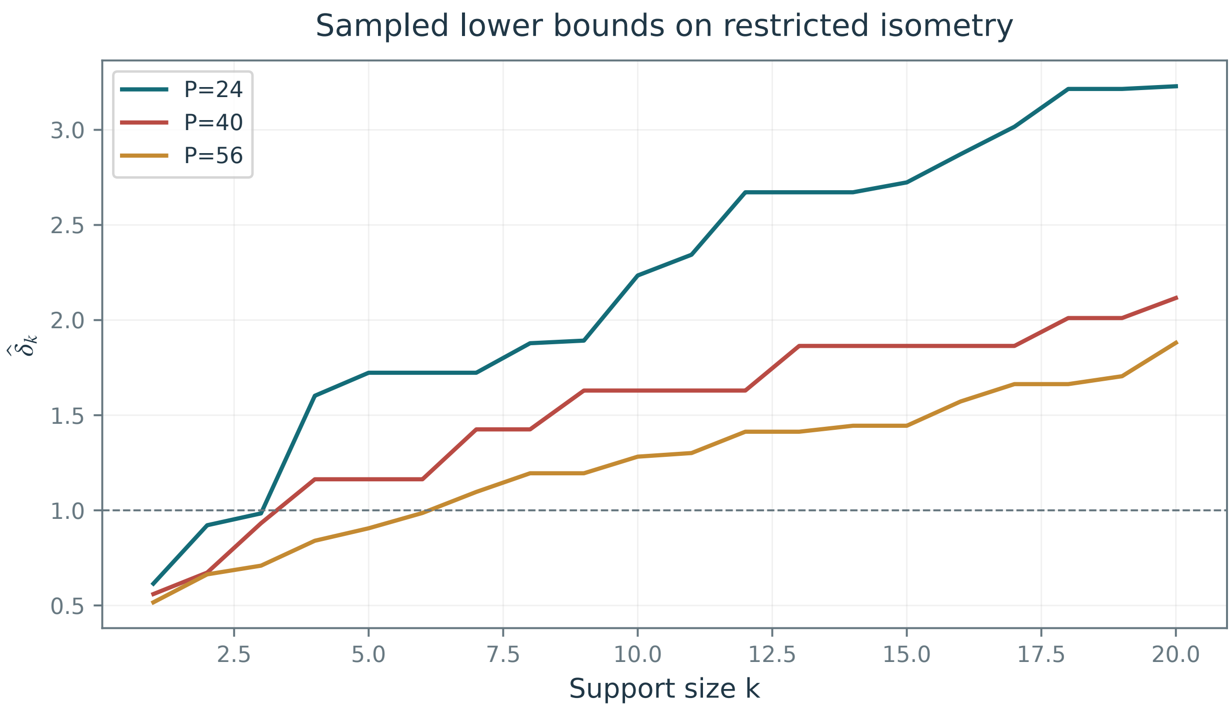


**Figure 12.6.** Evolution of lower bounds $\hat{\delta}_k$ on the RIP constant.

A small $\delta_s$ therefore means that $A$ approximately preserves the Euclidean norm on every coordinate subspace of dimension at most $s$.

A related quantity is the restricted orthogonality constant $\theta_{s,s'}$: the smallest nonnegative number such that, whenever $\|x\|_0 \leqslant s$, $\|x'\|_0 \leqslant s'$ and the supports are disjoint,

$$|\langle Ax, Ax'\rangle| \leqslant \theta_{s,s'}\|x\|\|x'\|$$

The next lemma relates restricted isometry and restricted orthogonality.

**Lemma 12.4.** *One has*

$$\theta_{s,s'} \leqslant \delta_{s+s'} \leqslant \theta_{s,s'} + \max(\delta_s, \delta_{s'}).$$

*Proof.* For the first inequality, normalize nonzero disjointly supported vectors so that $\|z\| = \|z'\| = 1$. RIP applied to $z + z'$ and $z - z'$ and polarization give

$$|\langle Az, Az'\rangle| = \frac{1}{4}\left|\|Az + Az'\|^2 - \|Az - Az'\|^2\right| \leqslant \delta_{s+s'}.$$

Rescaling gives the general claim. For the second inequality, split any vector supported on at most $s + s'$ indices as $x + x'$ with disjoint supports of sizes at most $s, s'$. Then

$$\left|\|A(x + x')\|^2 - \|x + x'\|^2\right| \leqslant \max(\delta_s, \delta_{s'})(\|x\|^2 + \|x'\|^2) + 2\theta_{s,s'}\|x\|\|x'\|,$$

and use $2\|x\|\|x'\| \leqslant \|x\|^2 + \|x'\|^2$. □

The next theorem bounds the number of measurements needed to keep restricted isometry constants small at sparsity level $s$.

**Theorem 12.5.** *For $0 < \delta < 1$ and $1 \leqslant s \leqslant N$, if $A$ is a normalized sub-Gaussian matrix and*

$$P \geqslant C\delta^{-2}s\log(eN/s) \tag{12.10}$$

*it satisfies $\delta_s \leqslant \delta$ with probability at least $1 - 2e^{-\delta^2\frac{P}{2C}}$, where $C$ depends only on the sub-Gaussian parameters in (12.6).*

The proof combines concentration on each support with a union bound over supports. The RIP condition (12.9) is equivalent to the spectral bound $\mathrm{eig}(A_I^* A_I) \subset [1-\delta_s, 1+\delta_s]$ for every column submatrix $A_I$ with $|I| \leqslant s$. Thus the proof reduces to controlling singular values uniformly over supports. For Gaussian matrices, Section 12.2.2 provides these estimates for $B^*B \in \mathbb{R}^{s\times s}$, where $B \in \mathbb{R}^{P\times s}$ has independent $\mathcal{N}(0, 1/P)$ entries. The concentration bound (12.1) controls $\delta_s$ on each support. Analogous nonasymptotic estimates hold under the stated sub-Gaussian assumptions; finite covariance alone would not suffice. For a fixed support, the smallest eigenvalue is close to $(1-\sqrt{s/P})^2$, a deficit of $2\sqrt{s/P} - s/P$ from 1. Thus a fixed tolerance requires $P$ to scale with $s$. A union bound over all supports adds the logarithmic factor. This step requires nonasymptotic concentration, which the Marchenko–Pastur limit alone does not provide.

### 12.3.2 RIP Implies Dual Certificates

The following theorem shows that a sufficiently small restricted isometry constant guarantees a valid dual certificate. Theorem 11.11 then yields stable recovery of sparse signals.

**Theorem 12.6.** *Assume $\delta_{2s} + \theta_{s,s} + \theta_{s,2s} < 1$. For every $x_0$ with $\|x_0\|_0 \leqslant s$, there is a certificate $\eta = A^*p$ satisfying $\eta_I = \mathrm{sign}(x_{0,I})$, where $I = \mathrm{supp}(x_0)$, and*

$$\|\eta_{I^c}\|_\infty \leqslant \frac{\theta_{s,s}}{1-\delta_{2s}-\theta_{s,2s}} < 1.$$

*Moreover, if $Q = \theta_{s,2s}/(1-\delta_{2s})$, one may choose*

$$\|p\| \leqslant \frac{\sqrt{s}}{\sqrt{1-\delta_s}} + \frac{\theta_{s,s}\sqrt{s}}{(1-\delta_s)\sqrt{1-\delta_{2s}}(1-Q)}.$$

Lemma 12.4 gives

$$\delta_{2s} + \theta_{s,s} + \theta_{s,2s} \leqslant \delta_{2s} + \delta_{2s} + \delta_{3s} \leqslant 3\delta_{3s}$$

so that condition $\delta_{2s} + \theta_{s,s} + \theta_{s,2s} < 1$ is implied by $\delta_{3s} < 1/3$. Sharper recovery conditions are available, for example $\delta_{2s} < \sqrt{2} - 1$ for robust recovery [1]. These sufficient bounds on restricted isometry constants are commonly referred to as restricted isometry properties (RIP). The numerical constants in elementary RIP bounds can be conservative, so the resulting sufficient sample sizes may exceed the observed phase-transition thresholds.

The iterative construction below follows [2], under a slightly stronger hypothesis than the sharpest certificate results. Rather than requiring the Fuchs precertificate $\eta_F$ in (12.2) to be feasible, we correct its large off-support entries. At each step, an interpolant cancels the largest $s$ entries outside $I$ while preserving the values on the target support. To construct these corrections to $\eta_F$, we first bound minimum-norm interpolants using restricted isometry and restricted orthogonality.

**Lemma 12.7.** *We assume $\delta_s < 1$. Let $c \in \mathbb{R}^N$ with $\|c\|_0 \leqslant s$ and $\mathrm{supp}(c) = J$, and let $s' \geqslant 1$ be an integer. Let $p$ be the minimum-norm solution of $A_J^* p = c_J$ and set $\eta = A^*p$. Explicitly,*

$$p \overset{\text{def.}}{=} A_J^{*,+} c_J = A_J (A_J^* A_J)^{-1} c_J$$

*(note that $A_J$ is full rank because $\delta_s < 1$). Then, letting $K \subset J^c$ index the $\min(s', |J^c|)$ largest entries in magnitude of $\eta_{J^c}$, one has*

$$\|\eta_{(J\cup K)^c}\|_\infty \leqslant \frac{\theta_{s,s'}\|c\|}{(1-\delta_s)\sqrt{s'}} \quad \text{and} \quad \|\eta_K\| \leqslant \frac{\theta_{s,s'}\|c\|}{1-\delta_s} \quad \text{and} \quad \|p\| \leqslant \frac{\|c\|}{\sqrt{1-\delta_s}}$$

*Proof.* Since $|J| \leqslant s$, the smallest eigenvalue satisfies $\lambda_{\min}(A_J^* A_J) \geqslant 1 - \delta_s$. Hence

$$\|(A_J^* A_J)^{-1}\| \leqslant \frac{1}{1-\delta_s}$$

One has

$$\|p\|^2 = \langle A_J(A_J^*A_J)^{-1}c_J,\, A_J(A_J^*A_J)^{-1}c_J\rangle = \langle (A_J^*A_J)^{-1}c_J,\, c_J\rangle \leqslant \frac{\|c\|^2}{1-\delta_s}.$$

Let $L$ be any set with $L \cap J = \emptyset$ and $|L| \leqslant s'$. One has

$$\|\eta_L\|^2 = |\langle \eta_L, \eta_L\rangle| = |\langle A_J(A_J^*A_J)^{-1}c_J,\, A_L\eta_L\rangle| \leqslant \theta_{s,s'}\|(A_J^*A_J)^{-1}c_J\|\|\eta_L\| \leqslant \frac{\theta_{s,s'}}{1-\delta_s}\|c_J\|\|\eta_L\|.$$

If $\eta_L \neq 0$, divide by $\|\eta_L\|$; if $\eta_L = 0$, the bound is immediate. In either case,

$$\|\eta_L\| \leqslant \frac{\theta_{s,s'}}{1-\delta_s}\|c_J\|. \tag{12.11}$$

Let us denote $\bar{K} = \{k \in J^c \ ; \ |\eta_k| > T\}$ where $T \overset{\text{def.}}{=} \frac{\theta_{s,s'}}{(1-\delta_s)\sqrt{s'}}\|c_J\|$. One necessarily has $|\bar{K}| \leqslant s'$, otherwise one would have, taking $K \subset \bar{K}$ the $s'$ largest entries (in fact any $s'$ entries)

$$\|\eta_K\| > \sqrt{s'T^2} = \frac{\theta_{s,s'}}{1-\delta_s}\|c_J\|$$

This contradicts (12.11). Consequently, all entries of $\eta_{J^c}$ beyond rank $s'$ have magnitude at most $T$. □

We can now prove Theorem 12.6.

*Proof.* For $x_0 = 0$, take $p = 0$. Otherwise, let $I_0 = \operatorname{supp}(x_0)$ and apply Lemma 12.7 with $J = I_0$, $c_{I_0} = \operatorname{sign}(x_{0,I_0})$, and $s' = s$. It gives $\eta_1 = A^* p_1$ and a set $I_1 \subset I_0^c$, $|I_1| \leqslant s$, such that

$$\eta_{1,I_0} = \operatorname{sign}(x_{0,I_0}), \quad \|\eta_{1,I_1}\| \leqslant b\sqrt{s}, \quad \|\eta_{1,(I_0\cup I_1)^c}\|_\infty \leqslant b, \qquad b = \frac{\theta_{s,s}}{1-\delta_s}.$$

Also $\|p_1\| \leqslant \sqrt{s}/\sqrt{1-\delta_s}$. Given $\eta_n$ and $I_n$, apply the lemma on $J = I_0 \cup I_n$ to the interpolation values $(0, \eta_{n,I_n})$, using $(2s, s)$ in place of $(s, s')$. This produces $\eta_{n+1} = A^* p_{n+1}$ and $I_{n+1}$ disjoint from $I_0 \cup I_n$, with

$$\eta_{n+1,I_0} = 0, \quad \eta_{n+1,I_n} = \eta_{n,I_n}, \quad \|\eta_{n+1,I_{n+1}}\| \leqslant b\sqrt{s}\,Q^n,$$

$$\|\eta_{n+1,(I_0\cup I_n\cup I_{n+1})^c}\|_\infty \leqslant bQ^n, \qquad Q = \frac{\theta_{s,2s}}{1-\delta_{2s}} < 1. \tag{12.12}$$

The denominator involves $\delta_{2s}$ because the interpolation support can have $2s$ indices. Furthermore,

$$\|p_{n+1}\| \leqslant \frac{b\sqrt{s}}{\sqrt{1-\delta_{2s}}}Q^{n-1}.$$

Thus $p = \sum_{n\geqslant 1}(-1)^{n-1}p_n$ converges and $\eta = A^* p$ interpolates the signs on $I_0$.

Fix $j \notin I_0$. Whenever $j \in I_n$, the two contributions $\eta_{n,j}$ and $\eta_{n+1,j}$ cancel. Since consecutive $I_n$ are disjoint, these cancellations do not overlap. Every remaining term obeys the pointwise geometric bound, including the initial bound for $n = 1$. Hence

$$|\eta_j| \leqslant \sum_{n\geqslant 1} bQ^{n-1} = \frac{b}{1-Q} \leqslant \frac{\theta_{s,s}}{1-\delta_{2s}-\theta_{s,2s}} < 1.$$

Summing the norm bounds on $p_1$ and the remaining $p_n$ gives the stated bound on $\|p\|$. □

### 12.3.3 RIP Implies Stable Recovery

For uniform stability, it is convenient to use the noise-constrained formulation ($\mathcal{P}^\varepsilon(y)$). The standard robust RIP estimate [1], together with Theorem 12.5, gives the following statement.

**Theorem 12.8.** *Let $1 \leqslant s \leqslant N$ and let $A$ be a normalized sub-Gaussian matrix. There are constants $C, c, C_0, C_1 > 0$, depending only on its sub-Gaussian parameters, such that if*

$$P \geqslant Cs\log(eN/s), \tag{12.13}$$

*then, with probability at least $1 - 2e^{-cP}$, the following holds simultaneously for every $x_0 \in \mathbb{R}^N$, every $\varepsilon \geqslant 0$, every noise vector $w$ with $\|w\| \leqslant \varepsilon$, and every minimizer $x_\varepsilon$ of ($\mathcal{P}^\varepsilon(y)$) with $y = Ax_0 + w$:*

$$\|x_\varepsilon - x_0\| \leqslant C_0\varepsilon + C_1\frac{\|x_0 - x_{0,s}\|_1}{\sqrt{s}}.$$

*Here $x_{0,s}$ is a best $s$-term approximation. In particular, exactly $s$-sparse vectors are recovered with error at most $C_0\varepsilon$, and uniquely when $\varepsilon = 0$.*

The penalized formulation also admits linear noise-error bounds: combine Theorem 12.6 with Theorem 11.11 and choose $\lambda$ accordingly. The bounds hold for every minimizer, but do not assert uniqueness of noisy Lasso solutions. For approximately sparse vectors, the constrained estimate measures the $\ell^1$ approximation tail divided by $\sqrt{s}$, without an additional factor $\|A\|$.

The order of the quantifiers distinguishes uniform from nonuniform recovery. In the noiseless case, this theorem bounds the probability of a single event on which every $s$-sparse vector is recovered. Theorem 12.3 instead bounds the recovery probability separately for each vector fixed independently of $A$.

For uniform norm recovery, RIP analysis replaces $\log N$ by $\log(eN/s)$. The Fuchs analysis using $\eta_F$ addresses an additional question: stability of the support. The guarantees also differ in their constants. Elementary RIP estimates can be conservative, while Gaussian fixed-signal calculations give a sharp leading constant in their asymptotic regime.

The uniform certificate constructed above need not coincide with the Fuchs precertificate $\eta_F$.

### 12.3.4 RIP for Fourier Sampling

Fully independent random measurements can be difficult to implement. A practical alternative is to sample a random subset of coefficients in an orthonormal basis $\Xi = (\xi_\omega)_{\omega=1}^N$ of $\mathbb{R}^N$ or $\mathbb{C}^N$:

$$Ax \overset{\text{def.}}{=} \sqrt{N/P}\,(\langle x,\, \xi_\omega\rangle)_{\omega\in\Omega} \tag{12.14}$$

The factor $\sqrt{N/P}$ makes the expected Gram matrix the identity. For a complex basis, the measurements are complex. The index set $\Omega \subset \{1,\dots,N\}$ is sampled uniformly without replacement, with $|\Omega| = P$. The following theorem gives an RIP guarantee when the basis vectors are sufficiently spread out, as quantified by

$$\rho(\Xi) \overset{\text{def.}}{=} \sqrt{N} \max_{1\leqslant\omega\leqslant N} \|\xi_\omega\|_\infty.$$

**Theorem 12.9** (Rudelson–Vershynin)**.** *For every fixed $0 < c < 1$ and failure exponent $\gamma > 0$, there is a constant $C = C(c,\gamma)$ such that, provided*

$$P \geqslant C\rho(\Xi)^2 s \log(2N)^4 \tag{12.15}$$

*the normalized operator (12.14) satisfies $\delta_{2s} \leqslant c$ with probability at least $1 - N^{-\gamma}$.*

One always has $1 \leqslant \rho(\Xi)^2 \leqslant N$. The worst case is the Dirac basis, $\Xi = \mathrm{Id}_N$, for which $\rho(\Xi)^2 = N$. Indeed, uniform recovery of even one-sparse vectors then requires observing every coordinate, so $P = N$. Fourier atoms $\xi_\omega = (N^{-1/2} e^{\frac{2i\pi}{N}\omega n})_{n=1}^N \in \mathbb{C}^N$ attain the smallest possible value, $\rho(\Xi) = 1$; Hadamard bases do so as well. Up to logarithmic factors, the sample count (12.15) then matches the sub-Gaussian scaling (12.13).

Theorem 12.9 assumes that $x_0$ is sparse in the coordinate basis. If instead $f_0$ is sparse in an orthonormal basis $\Psi = (\psi_m)_{m=1}^N$, use the coordinates $x = \Psi^* f$, so that $x$ contains the coefficients of $f$ in $\Psi$. The factor $\rho(\Xi)$ in (12.15) is then replaced by the mutual coherence of the sampling and sparsity bases:

$$\rho(\Psi^*\Xi) \overset{\text{def.}}{=} \sqrt{N} \max_{1\leqslant\omega,m\leqslant N} |\langle \psi_m,\, \xi_\omega\rangle|.$$

Small mutual coherence therefore yields stronger measurement bounds. Fourier and Dirac bases are maximally incoherent. Fourier and wavelet bases can have large correlations, whereas specially constructed “noiselet” bases have low coherence with suitable wavelets. Unlike the Gaussian ensemble, these structured measurements are not universal: their recovery guarantees depend on the sparsity basis $\Psi$.

# References


[1] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9–10):589–592, 2008.

[2] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[3] Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.

[4] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Birkhäuser Basel, 2013.

[5] Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[6] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, Frank Lenzen, and L Sirovich. *Variational methods in imaging*. Springer, 2009.