Gabriel Peyré
CNRS & DMA, École Normale Supérieure, PSL

CHAPTER 3

Sparsity, inverse problems and compressed sensing

September 9, 2026

How can a small number of measurements describe a complicated signal? Sparse representations provide a link between compression and the recovery of incomplete or noisy data. This chapter develops that link, explains the role of convex optimization, and states the assumptions under which compressed sensing gives reliable reconstructions.

3.1 From continuous signals to samples

An acquisition device turns a continuous signal $\tilde{f}$ into a finite vector $f = (f_q)_{q=1}^{Q} \in \mathbb{R}^Q$. For sound, these entries represent measurements at successive times. For a grayscale image, they represent light collected over small sensor regions. An idealized pixel measurement is

$$f_q = \int_{c_q} \tilde{f}(s)\,\mathrm{d}s,$$

where c_q is the area of the sensor element. Dividing by its area gives an average intensity; exposure time and calibration introduce further scale factors. Color and multichannel signals can be handled by recording several values at each location.

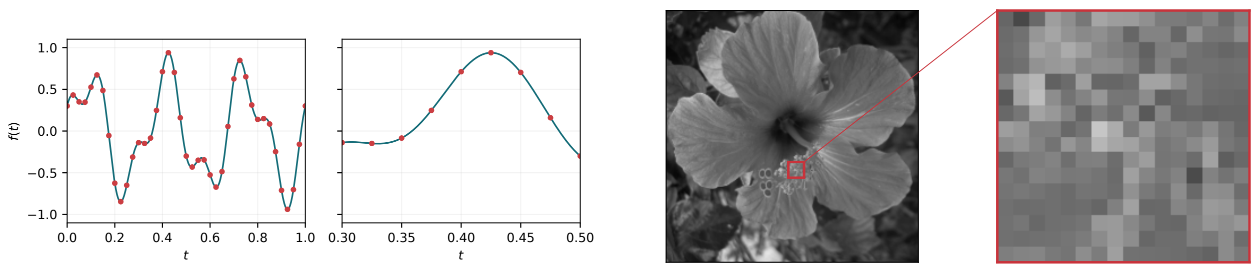


Figure 3.1. Sampling a one-dimensional signal and a two-dimensional image. The dots mark time samples; the red frames identify the image region enlarged on the right.

Sampling preserves information only under suitable assumptions. For example, let $\tilde{f} \in L^2(\mathbb{R})$ have Fourier transform supported in $[-B, B]$, with frequency measured in cycles per unit time. If the sampling interval satisfies $0 < T < 1/(2B)$, the sampling theorem [15] gives

$$\tilde{f}(t) = \sum_{k\in\mathbb{Z}} \tilde{f}(kT)\,\operatorname{sinc}\left(\frac{t}{T} - k\right), \qquad \operatorname{sinc}(u) = \begin{cases} \sin(\pi u)/(\pi u), & u \neq 0, \\ 1, & u = 0. \end{cases}$$

For the continuous representative of this band-limited function, the series converges in L^2 and uniformly. The theorem uses samples over the entire time axis. A finite recording, quantization, and measurement noise introduce additional errors. Frequencies above the acquisition band must be attenuated before sampling to avoid aliasing; increasing numerical precision afterwards cannot undo aliasing.

This is a different question from lossless coding: sampling concerns the representation of a continuous signal by measurements, whereas coding concerns the binary representation of those measurements. Sparsity provides another useful assumption on the signal, and leads to a different acquisition strategy.

3.2 Nonlinear approximation and compression

3.2.1 Sparse representations

A dictionary $\Psi = (\psi_n)_{n=1}^N$ consists of vectors $\psi_n \in \mathbb{R}^Q$. We seek an approximation

$$f \approx \Psi x = \sum_{n=1}^{N} x_n \psi_n$$

using few nonzero coefficients. Sinusoids, local cosine functions, and wavelets are familiar examples of atoms. Transform coding uses such representations to concentrate much of a signal's energy in a small number of coefficients.

Given an integer budget $0 \le M \le N$, a best M-term approximation solves

$$x^\star \in \arg \min_{x \in \mathbb{R}^N,\, \|x\|_0 \le M} \|f - \Psi x\|_2^2, \qquad \|x\|_0 = \#\{n : x_n \neq 0\}. \tag{3.1}$$

Here $\|v\|_2^2 = \sum_n |v_n|^2$. The notation $\|x\|_0$ is conventional: this count is not a norm, since multiplying a nonzero vector by a nonzero scalar leaves the count unchanged. For a general dictionary, sparse approximation is NP-hard [12]. Exhaustively testing the possible supports is usually prohibitive; this worst-case difficulty does not rule out efficient algorithms for special dictionaries or particular data.

3.2.2 The orthonormal case

Suppose $N = Q$ and Ψ is an orthonormal basis, so that $\langle \psi_n, \psi_m \rangle = \delta_{nm}$. With $a_n = \langle f, \psi_n \rangle$, Parseval's identity gives

$$f = \sum_{n=1}^{N} a_n \psi_n, \qquad \|f - \Psi x\|_2^2 = \sum_{n=1}^{N} |a_n - x_n|^2. \tag{3.2}$$

Once a support is chosen, the best coefficients are the corresponding a_n. To minimize the discarded energy, keep the M largest magnitudes. More precisely, choose an ordering $|a_{n_1}| \ge \cdots \ge |a_{n_N}|$. Then

$$x_n^\star = \begin{cases} a_n, & n \in \{n_1, \ldots, n_M\}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

Ties at the threshold can produce several minimizers. If fewer than M coefficients are nonzero, the approximation uses fewer than M terms.

f $\quad$ $M = N/4$ $\quad$ $M = N/8$ $\quad$ $M = N/16$

Figure 3.2. Best M-term wavelet approximations of a flower image with $N = 256^2$ pixels. The orthonormal transform uses Daubechies wavelets with four vanishing moments, five levels, and periodic boundary conditions. Display intensities are clipped to $[0, 1]$.

The reconstruction $x \mapsto \Psi x$ is linear, but selecting the largest coefficients depends on the data. The approximation $f \mapsto \Psi x^\star$ is therefore nonlinear. For many piecewise smooth signals, wavelets give small errors with relatively few coefficients; the regularity of the signal controls the decay of this error [10, Chapter 9]. Textures and fine oscillations can require substantially more terms.

The fraction M/N counts retained coefficients; it is not a bit rate. A complete compressor must also quantize the retained values, describe their locations, and encode the resulting symbols. Entropy coding, introduced in the chapter on information theory, completes this step. Wavelet transforms also admit fast algorithms whose cost is proportional to the number of samples [10, Chapter 7].

3.3 Inverse problems and regularization

3.3.1 The observation model

An inverse problem asks us to recover an unknown signal $f \in \mathbb{R}^Q$ from measurements

$$y = \Phi f + w \in \mathbb{R}^P. \tag{3.4}$$

The known linear operator $\Phi \in \mathbb{R}^{P\times Q}$ describes acquisition, and w represents unknown errors. Usually one applies Φ by an algorithm rather than storing this potentially enormous matrix.

f $\quad$ Φf: blur $\quad$ Φf: mask

Figure 3.3. Two noiseless observation operators applied to the same flower image: Gaussian blurring and a mask retaining about 10% of the pixels. Missing entries are displayed in black.

Several standard problems fit this model:

- Denoising: $\Phi = \mathrm{Id}$ and $P = Q$.
- Deblurring: $\Phi f = \varphi \star f$, with a filter φ describing the blur and a specified boundary convention.
- Missing data: Φ selects observed coordinates. Equivalently, one may keep the full array and use the diagonal mask $\Phi = \mathrm{diag}(\mu_q)$, where $\mu_q \in \{0, 1\}$.
- Tomography: Φ models line integrals of the object, as in a discretized Radon transform.

If Φ has a nontrivial kernel, the measurements alone do not determine f. Even when a square Φ is invertible, its small singular values can amplify errors in $\Phi^{-1}y = f + \Phi^{-1}w$. Regularization introduces additional information to select a useful, stable reconstruction.

3.3.2 Sparse regularization

Assuming $f \approx \Psi x$ with few active coefficients leads to

$$x^\star \in \arg\min_{\|x\|_0 \le M} \|y - \Phi\Psi x\|_2^2, \qquad f^\star = \Psi x^\star. \tag{3.5}$$

For denoising in an orthonormal basis, this is hard thresholding of the coefficients of the *observed signal* y. It gives the best sparse fit to y for the chosen budget; it does not by itself guarantee the smallest error relative to the unknown clean signal f. For a general acquisition operator, $\Phi\Psi$ is not orthogonal, and the simple thresholding formula no longer solves the problem.

Greedy algorithms and convex relaxation provide two ways to approach this difficulty [11, 17]. To motivate convex relaxation, consider

$$\|x\|_\alpha^\alpha = \sum_{n=1}^{N} |x_n|^\alpha, \qquad \alpha > 0.$$

For each fixed vector x, this quantity tends to $\|x\|_0$ as $\alpha \downarrow 0$. The sets $B_\alpha = \{x : \sum_n |x_n|^\alpha \le 1\}$ lie in $[-1,1]^N$ and converge, in Hausdorff distance, to

$$B_0 = \{x \in [-1,1]^N : \|x\|_0 \le 1\}.$$

This limiting set consists of the coordinate-axis segments, including the origin. The set $\{x : \|x\|_0 \le 1\}$ without the box restriction instead contains the entire, unbounded coordinate axes.

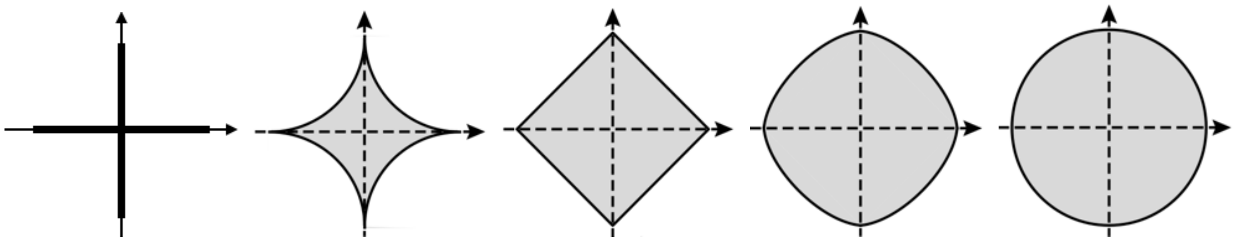


Figure 3.4. The sets B_α in two dimensions. The first panel shows the bounded limiting set B_0.

Within this family, $\|\cdot\|_\alpha$ is a norm and B_α is convex precisely when $\alpha \ge 1$. The choice $\alpha = 1$ balances convexity with a geometry that favors zero coefficients. This leads to the constrained Lasso problem

$$x^\star \in \arg\min_{\|x\|_1 \le \tau} \frac{1}{2}\|Ax - y\|_2^2, \qquad A = \Phi\Psi. \tag{3.6}$$

Increasing τ allows a closer fit to the measurements, but can also fit their errors. The parameter must be chosen with the noise level, operator, and signal model in mind. Sending τ to infinity removes the regularization and generally leaves many possible reconstructions when $P < N$.

Two closely related formulations are useful:

$$\min_x \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_1, \qquad \lambda > 0, \tag{3.7}$$

$$\min_x \|x\|_1 \quad \text{subject to } \|Ax - y\|_2 \le \varepsilon, \qquad \varepsilon \ge 0. \tag{3.8}$$

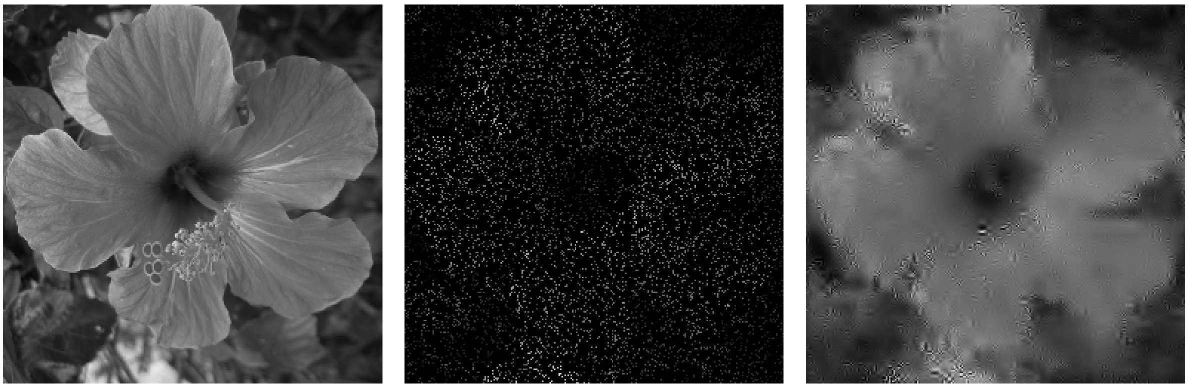
The penalized formulation trades data fit against sparsity. The second formulation, basis pursuit denoising, is convenient when an error bound ε is known. For $\varepsilon = 0$, it becomes basis pursuit, $\min_{Ax=y}\|x\|_1$. These formulations are connected through suitable Lagrange multipliers; their parameters are not interchangeable numerical values. Early developments include seismic inversion, basis pursuit, and the Lasso [5, 14, 16].

Convexity makes global optimization possible, but the large dimension and the nondifferentiability of $\|x\|_1$ still require suitable algorithms. Proximal methods exploit the explicit soft-thresholding operation associated with this penalty [6]. A numerical solver reaches a prescribed accuracy, rather than an exact real-valued solution after a guaranteed small number of operations.

3.3.3 Why the geometry matters

In two dimensions, an ℓ^1 ball has corners on the coordinate axes. A level set of the data-fitting term can first meet the ball at one of these corners, producing a zero coefficient. The round ℓ^2 ball typically gives two nonzero coefficients in the same configuration (Figure 3.6). This is an explanation of a tendency, not a guarantee for every operator or every datum: a contact along a face can give several minimizers, and an ℓ^1 solution need not be sparse enough to recover the truth.

Recovery depends on how the measurement operator acts on sparse vectors. Increasing the ambient dimension alone does not improve it. Precise conditions are available for many random and structured operators; convolution and Fourier acquisition require assumptions adapted to their geometry [1, 3, 4, 9].



f $\qquad$ y $\qquad$ $f^\star = \Psi x^\star$

Figure 3.5. Reconstruction from 6,554 observed pixels out of 256^2 (about 10%). The penalized problem (3.7) is solved numerically with $\lambda = 0.005$, intensities in $[0, 1]$, and the same orthonormal wavelets as in Figure 3.2. The remaining texture errors illustrate the limits of the model.

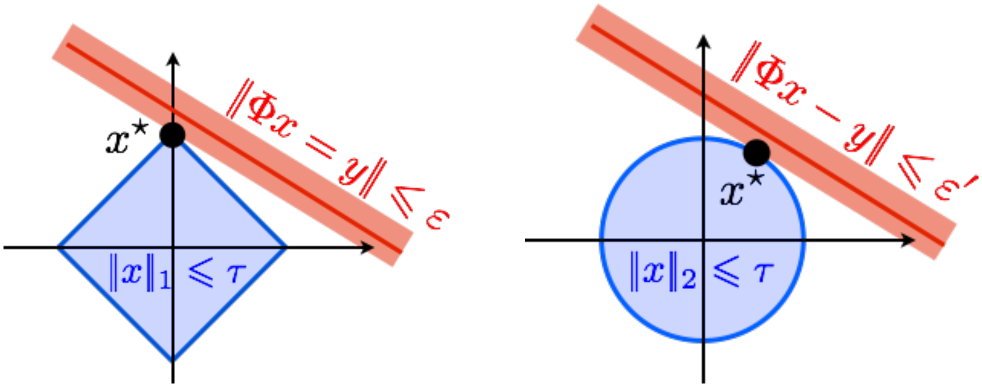


Figure 3.6. Constrained data fitting with an ℓ^1 ball (left) and an ℓ^2 ball (right). Starting from $\varepsilon = 0$, enlarge the strip $\{x : \|Ax - y\|_2 \le \varepsilon\}$ until it first meets the constraint ball. The contact point minimizes the residual over that ball.

3.4 Compressed sensing

3.4.1 Measuring combinations of pixels

Conventional transform coding first acquires Q samples and then keeps a much smaller representation. Compressed sensing asks whether we can instead acquire $P \ll Q$ carefully chosen linear measurements and reconstruct a signal that is sparse, or approximately sparse, in a known basis.

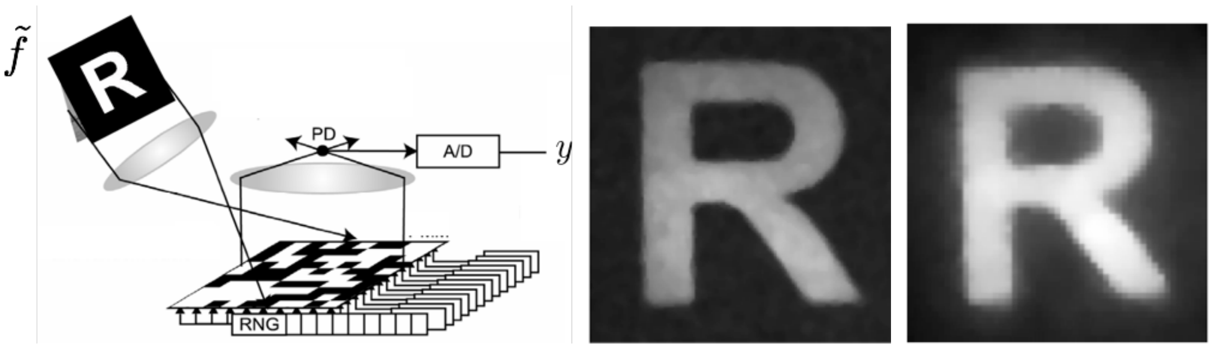
The single-pixel camera developed at Rice University illustrates this idea [8]. A grid of Q micromirrors directs selected portions of the light towards one detector; other mirrors direct light away from it. For each pattern p, the detector measures the sum

$$y_p = \sum_{q=1}^{Q} \Phi_{p,q} \int_{c_q} \tilde{f}(s)\,\mathrm{d}s + w_p = \sum_{q=1}^{Q} \Phi_{p,q} f_q + w_p, \qquad \Phi_{p,q} \in \{0, 1\}.$$

The mirrors do not record individual pixels. Changing their configuration produces the vector $y \in \mathbb{R}^P$ directly. The scene must remain sufficiently stable during these measurements. Complementary patterns, or a calibrated measurement of total light, can convert 0/1 measurements into centered signed measurements.

3.4.2 A recovery guarantee

Here is a standard form of the compressed-sensing guarantee [2, 7]. Suppose Ψ is an orthonormal basis of $\mathbb{R}^N$, $f = \Psi x_0$, and $A = \Phi\Psi$ has independent Gaussian entries with mean zero and variance $1/P$. Equivalently, Φ may be chosen with that Gaussian distribution, since it is invariant under multiplication by a fixed orthogonal matrix.



Acquisition $\quad$ f $\quad$ $f^\star$

Figure 3.7. Single-pixel acquisition and the original reconstruction example, reproduced from the material accompanying [8]. The stated measurement reduction factor is $Q/P = 6$, so that $P/Q \approx 1/6$.

There are universal positive constants C, c, C_0, C_1 such that, for $1 \le M \le N$ and

$$P \ge CM \log(eN/M), \tag{3.9}$$

with probability at least $1 - 2e^{-cP}$ over the choice of A, the following bound holds simultaneously for all x_0 and all errors with $\|w\|_2 \le \varepsilon$. Any solution $x^\star$ of (3.8), with $y = Ax_0 + w$, satisfies

$$\|x^\star - x_0\|_2 \le C_0 \frac{\sigma_M(x_0)_1}{\sqrt{M}} + C_1 \varepsilon, \qquad \sigma_M(x_0)_1 = \min_{\|z\|_0 \le M} \|x_0 - z\|_1. \tag{3.10}$$

The logarithm in (3.9) is natural; changing its base only changes the constant. The interesting compressed regime is one in which the required P is smaller than N.

The quantity $\sigma_M(x_0)_1$ is the sum of the magnitudes of the discarded coefficients after keeping the M largest. If x_0 is exactly M-sparse and the noiseless problem uses $\varepsilon = 0$ (hence $w = 0$), both terms in (3.10) vanish: basis pursuit recovers x_0 exactly. For approximately sparse signals and noisy measurements, the two terms separate the approximation error from the measurement error. Orthonormality gives the same Euclidean error for the reconstructed signal, $\|\Psi x^\star - f\|_2 = \|x^\star - x_0\|_2$.

One route to this theorem is the restricted isometry property: A approximately preserves Euclidean lengths of vectors supported on sufficiently small sets. Random matrices satisfy this property when the number of measurements has the order in (3.9). Appropriate centered subgaussian designs also admit guarantees, with constants depending on their distribution. A raw 0/1 sensing matrix requires the centering and normalization to be accounted for; it is not the Gaussian model stated above.

The theorem concerns a number of real measurements, not a number of bits. Their precision, quantization, and storage must still be considered. The sensing matrix must be known to the reconstruction algorithm, for instance through a shared random seed. Keeping that matrix secret is not, by itself, a cryptographic security guarantee. Likewise, savings in acquisition time or energy depend on the physical apparatus and the reconstruction cost.

Structured acquisition, including magnetic resonance imaging, imposes additional constraints: one cannot generally choose arbitrary independent Gaussian measurements. Sampling trajectories and sparsity models must therefore be analyzed together [1, 4]. The guiding principle remains to design acquisition and reconstruction jointly, using a model that reflects the data.

Further reading. The interplay between approximation, inverse problems, and optimization extends beyond sparsity. Other penalties encode smoothness, low rank, or geometric structure. The review [17] develops these ideas, and the Numerical Tours [13] provide computational examples.

Acknowledgements. I thank Charles Dossal, Jalal Fadili, Samuel Vaiter, Stéphane Seuret, and the anonymous reviewer of the original text for their help.

References

[1] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing. *CoRR*, abs/1302.0561, 2013. URL http://arxiv.org/abs/1302.0561.

[2] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[3] E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014. ISSN 1097-0312. doi: 10.1002/cpa.21455. URL http://dx.doi.org/10.1002/cpa.21455.

[4] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss. Variable density sampling with continuous trajectories. *SIAM Journal on Imaging Sciences*, 7(4):1962–1992, 2014.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998. doi: 10.1137/S1064827596304010.

[6] Patrick L. Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011. doi: 10.1007/978-1-4419-9569-8_10.

[7] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[8] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, March 2008. ISSN 1053-5888.

[9] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015. doi: 10.1007/s10208-014-9228-6. URL http://hal.archives-ouvertes.fr/hal-00839635/.

[10] S. G. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Elsevier/Academic Press, Amsterdam, third edition, 2009.

[11] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[12] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2): 227–234, 1995. doi: 10.1137/S0097539792240406.

[13] Gabriel Peyré. The numerical tours of signal processing. *Computing in Science & Engineering*, 13(4):94–97, 2011. doi: 10.1109/MCSE.2011.71. URL https://www.numerical-tours.com/.

[14] F. Santosa and W.W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330, 1986. doi: 10.1137/0907087.

[15] C. E. Shannon. Communication in the presence of noise. *Proc. Institute of Radio Engineers*, 37(1):10–21, 1949.

[16] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.

[17] Samuel Vaiter, Gabriel Peyré, and Jalal Fadili. Low complexity regularization of linear inverse problems. In Götz E. Pfander, editor, *Sampling Theory, a Renaissance*, pages 103–153. Birkhäuser, 2015. URL https://hal.science/hal-01018927.