Mathematical Foundations of Data Sciences

Gabriel Peyré
CNRS & DMA, École Normale Supérieure

CHAPTER 10

Sparse Regularization

September 9, 2026

Many signals admit accurate approximations using only a small number of coefficients in a suitable basis or dictionary. Sparse regularization turns this observation into a reconstruction principle for noisy or incomplete measurements. We move from counting nonzero coefficients to a convex penalty, distinguish analysis and synthesis models, derive thresholding and iterative algorithms, and apply them to deconvolution and inpainting.

The main references for this chapter are [5, 7, 6].

10.1 Sparsity Priors

10.1.1 Ideal Sparsity Prior

Chapter 5 shows that an orthonormal basis $\mathcal{B} = \{\psi_m\}_m$ adapted to an image class Θ can approximate each $f \in \Theta$ using relatively few atoms.

The ℓ^0 prior measures representation complexity by counting the nonzero coefficients:

$$J_0(f) \overset{\text{def.}}{=} \#\{m \,;\, \langle f, \psi_m \rangle \neq 0\} \quad \text{where} \quad x_m = \langle f, \psi_m \rangle.$$

This count is also called the ℓ^0 pseudonorm:

$$\|x\|_0 \overset{\text{def.}}{=} J_0(f).$$

It measures sparsity directly, without accounting for the magnitudes of the nonzero coefficients.

Natural images usually have many nonzero coefficients, but can often be approximated by a function f_M with a small count $M = J_0(f_M)$. Retaining coefficients above a threshold gives a best M-term approximation, with M determined by that threshold:

$$f_M = \sum_{|\langle f, \psi_m \rangle| > T} \langle f, \psi_m \rangle \psi_m \quad \text{where} \quad M = \#\{m \,;\, |\langle f, \psi_m \rangle| > T\}.$$

For suitable wavelet bases and the bounded-variation image classes specified in Section 5.2, the error $\|f - f_M\|$ satisfies quantitative decay estimates. These estimates motivate approximating natural images f by functions with small J_0.

Figure 10.1 displays a natural image in the wavelet basis $\psi_m = \psi^{\omega}_{j,n}$, indexed by $m = (j, n, \omega)$. Most coefficients $\langle f, \psi_m \rangle$ have small magnitude, so retaining the few large coefficients captures much of the image.

10.1.2 Convex Relaxation

The ideal sparsity prior J_0 is nonconvex, which makes optimization difficult. For example, if f and g have disjoint nonempty coefficient supports in $\mathcal{B}$, then $J_0((f+g)/2) = J_0(f) + J_0(g)$, violating convexity of J_0.

In a general inverse problem, minimizing J_0 entails a combinatorial search over possible coefficient supports.

To approximate the ideal prior J_0, consider the ℓ^q family with $q > 0$:

$$J_q(f) = \sum_m |\langle f, \psi_m \rangle|^q.$$

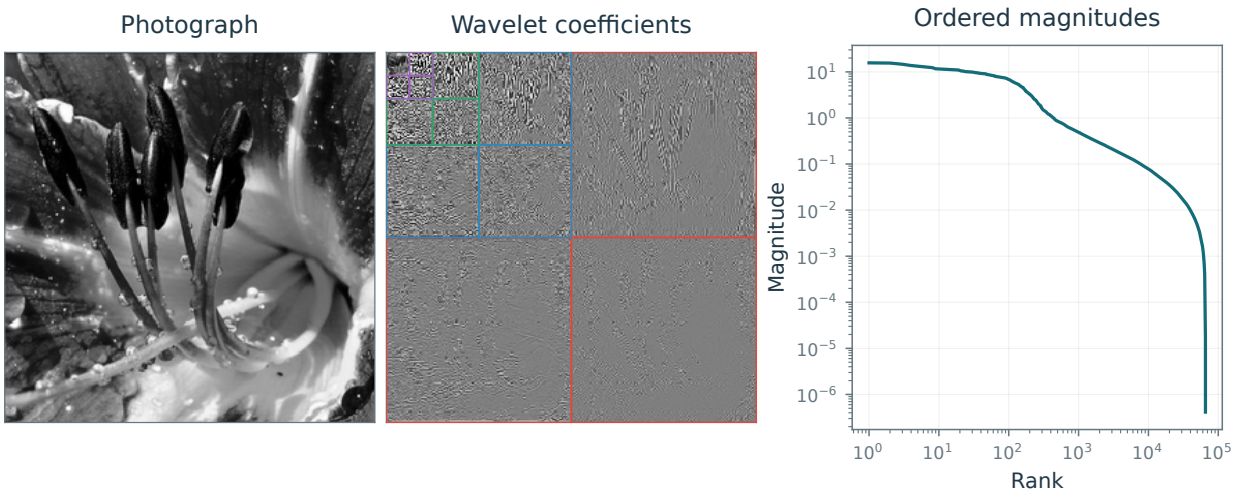

Figure 10.1. Most wavelet coefficients of a natural image have small magnitude.

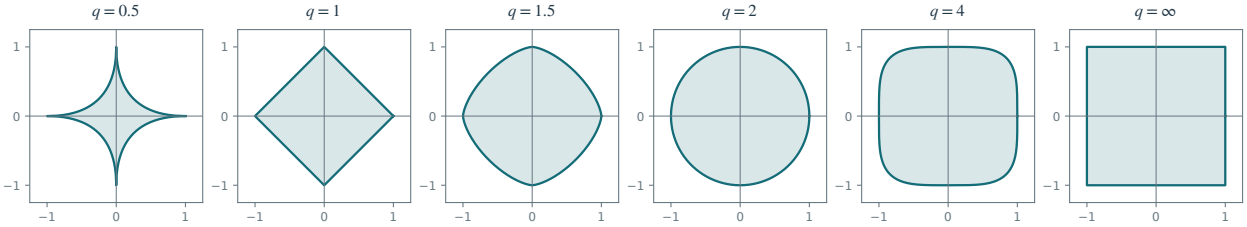

Figure 10.2. ℓ^q balls $\{x \,;\, J_q(x) \leqslant 1\}$ for varying q.

As shown in Figure 10.2, the unit balls in $\mathbb{R}^2$ become concentrated near the coordinate axes as $q \downarrow 0$. For each fixed finite-dimensional vector, $J_q(f) \to J_0(f)$; the limiting shape of these bounded unit balls should not be confused with the unbounded set $\{x : \|x\|_0 \leqslant 1\}$. Small values of q favor sparsity.

The prior J_q is convex exactly when $q \geqslant 1$. The smallest exponent yielding a convex prior, $q = 1$, gives the ℓ^1 prior J_1:

$$J_1(f) = \|(\langle f, \psi_m \rangle)\|_1 = \sum_m |\langle f, \psi_m \rangle|. \tag{10.1}$$

In the following, we consider discrete orthonormal bases $\mathcal{B} = \{\psi_m\}_{m=0}^{N-1}$ of $\mathbb{R}^N$.

10.1.3 Sparse Regularization and Thresholding

Given an orthonormal basis $\{\psi_m\}_m$ of $\mathbb{R}^N$, regularization-based denoising (8.15) can be written using the sparsity priors J_0 and J_1 as

$$f^\star \in \operatorname*{argmin}_{g \in \mathbb{R}^N} \frac{1}{2}\|f - g\|^2 + \lambda J_q(g)$$

for $q = 0$ or $q = 1$, writing $J_0 = J_0$. Orthonormality separates the objective into coefficientwise terms:

$$f^\star = \sum_m x_m^\star \psi_m$$

$$\text{where} \quad x^\star \in \operatorname*{argmin}_{y \in \mathbb{R}^N} \sum_m \frac{1}{2}|x_m - y_m|^2 + \lambda |y_m|^q$$

Here $x_m \overset{\text{def.}}{=} \langle f, \psi_m \rangle$ and $y_m \overset{\text{def.}}{=} \langle g, \psi_m \rangle$. For $q = 0$, we adopt the convention

$$\forall\, u \in \mathbb{R}, \quad |u|^0 = \begin{cases} 0 & \text{if} \quad u = 0, \\ 1 & \text{otherwise.} \end{cases}$$

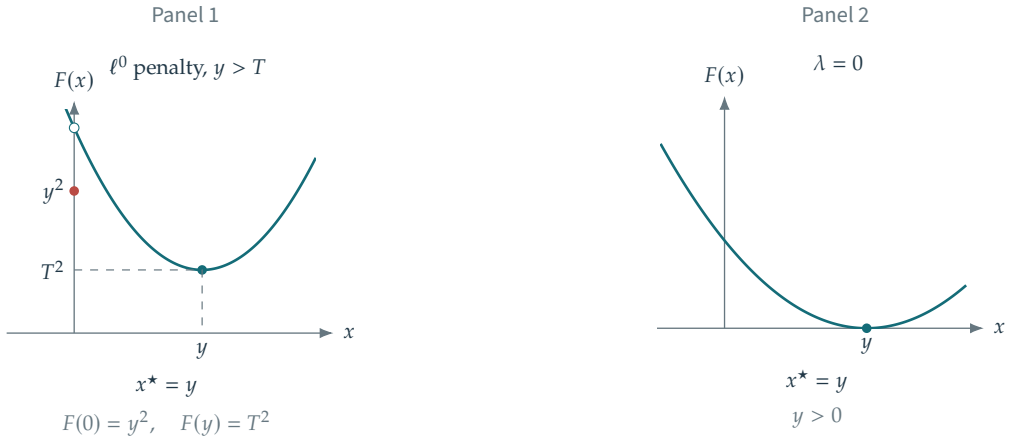

Figure 10.3. Panel 1: the objective $\|\cdot - y\|^2 + T^2\|\cdot\|_0$. Panels 2–5: the scalar objective $F(x) \overset{\text{def.}}{=} \frac{1}{2}|x-y|^2 + \lambda|x|$ for increasing λ.

Each coefficient of the denoised image solves a one-dimensional optimization problem

$$x_m^\star \in \operatorname*{argmin}_{u \in \mathbb{R}} \frac{1}{2}|x_m - u|^2 + \lambda |u|^q \tag{10.2}$$

The following proposition gives a closed-form solution by thresholding.

Proposition 10.1. *One may choose the minimizer*

$$x_m^\star = S_T^q(x_m) \quad \textit{where} \quad \begin{cases} T = \sqrt{2\lambda} & \textit{for} \quad q = 0, \\ T = \lambda & \textit{for} \quad q = 1, \end{cases} \tag{10.3}$$

where

$$\forall\, u \in \mathbb{R}, \quad S_T^0(u) \overset{\text{def.}}{=} \begin{cases} 0 & \textit{if} \quad |u| < T, \\ u & \textit{otherwise} \end{cases} \tag{10.4}$$

is hard thresholding, with the alternative tie convention discussed below, and

$$\forall\, u \in \mathbb{R}, \quad S_T^1(u) \overset{\text{def.}}{=} \operatorname{sign}(u)(|u| - T)_+ \tag{10.5}$$

is the soft thresholding introduced in (7.9). For hard thresholding at $|u| = T$, both 0 and u minimize the objective; the displayed rule selects u. Soft thresholding gives the unique minimizer.

Proof. For $q = 0$, a zero input is minimized at zero. For a nonzero input, the best nonzero candidate is $u = x_m$, with cost λ, whereas $u = 0$ has cost $|x_m|^2/2$. Comparing the two costs gives the threshold $\sqrt{2\lambda}$ and the stated tie. For $q = 1$, the optimality condition is $0 \in u - x_m + \lambda \partial |u|$. If $u > 0$, this gives $u = x_m - \lambda$; if $u < 0$, it gives $u = x_m + \lambda$; and $u = 0$ is optimal exactly when $|x_m| \leqslant \lambda$. These cases give soft thresholding. □

Transforming the thresholded coefficients back to the image domain gives

$$f_{\lambda,q} = \sum_m S_T^q(\langle f, \psi_m \rangle)\psi_m.$$

For Gaussian white noise w of variance σ^2, thresholds can be chosen from the noise level; see Section 7.3. The universal threshold $T = \sigma\sqrt{2 \log N}$ controls the largest noise coefficients and gives asymptotic risk guarantees under the signal assumptions of Section 7.3.3. Common empirical choices are $T \approx 3\sigma$ for hard thresholding (ℓ^0 regularization) and $T \approx 3\sigma/2$ for soft thresholding (ℓ^1 regularization); see Figure 7.13.

10.2 Sparse Regularization of Inverse Problems

Using the ℓ^1 prior in an orthonormal basis $\{\psi_m\}_m$ of $\mathbb{R}^N$, with J_1 defined in (10.1), gives the convex inverse problem

$$f_\lambda \in \operatorname*{argmin}_{f \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi f\|^2 + \lambda \sum_m |\langle f, \psi_m \rangle|. \tag{10.6}$$

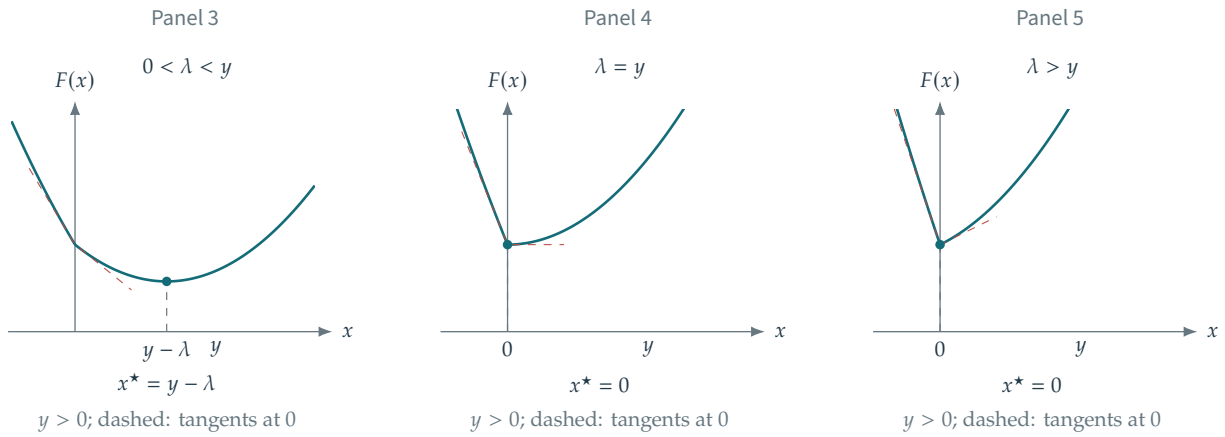

Figure 10.3 (continued). Panel 1: the objective $\|\cdot - y\|^2 + T^2\|\cdot\|_0$. Panels 2–5: the scalar objective $F(x) \overset{\text{def.}}{=} \frac{1}{2}|x - y|^2 + \lambda|x|$ for increasing λ.

Chen, Donoho, and Saunders introduced this basis pursuit denoising formulation in [1]. Section 10.3 derives an iterative thresholding algorithm for (10.6).

Analysis vs. synthesis priors. For a nonorthogonal dictionary $\Psi = \{\psi_m\}_{m=1}^Q$, which is redundant if $Q > N$, there are two distinct extensions of (10.6). Each uses a convex prior J in

$$f_\lambda \in \operatorname*{argmin}_{f \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi f\|^2 + \lambda J(f). \tag{10.7}$$

We also use the dictionary symbol for its synthesis operator; its adjoint is the analysis operator:

$$\Psi : x \in \mathbb{R}^Q \mapsto \Psi x = \sum_m x_m \psi_m \quad \text{and} \quad \Psi^* : f \in \mathbb{R}^N \mapsto (\langle f, \psi_m \rangle)_{m=1}^Q \in \mathbb{R}^Q.$$

The analysis prior penalizes the sum of the magnitudes of correlations with the dictionary atoms:

$$J_1^{\mathrm{A}}(f) \overset{\text{def.}}{=} \sum_m |\langle f, \psi_m \rangle| = \|\Psi^* f\|_1. \tag{10.8}$$

The synthesis prior minimizes the coefficient norm over expansions of f in Ψ:

$$J_1^{\mathrm{S}}(f) \overset{\text{def.}}{=} \min_{x \in \mathbb{R}^Q, \Psi x = f} \|x\|_1. \tag{10.9}$$

The two priors agree for an orthonormal basis, but generally differ for redundant dictionaries. Analysis regularization often requires primal–dual algorithms, as detailed in Chapter 19. Its recovery theory depends on the dictionary and on the structure of the vanishing analysis coefficients. The synthesis formulation leads directly to the coefficient-space algorithms studied below. We assume the dictionary spans $\mathbb{R}^N$; otherwise the synthesis prior is $+\infty$ outside its span.

We now focus on synthesis regularization $J = J_1^{\mathrm{S}}$, and rewrite (10.7) as $f_\lambda = \Psi x_\lambda$ where x_λ is any solution of the following basis pursuit denoising problem

$$x_\lambda \in \operatorname*{argmin}_{x \in \mathbb{R}^Q} \frac{1}{2\lambda}\|y - Ax\|^2 + \|x\|_1 \tag{10.10}$$

where the measurement matrix in coefficient space is

$$A \overset{\text{def.}}{=} \Phi\Psi \in \mathbb{R}^{P \times Q}.$$

For consistent data $y \in \operatorname{Im}(A)$, the limit $\lambda \downarrow 0$ leads to the constrained problem

$$x^\star \in \operatorname*{argmin}_{Ax = y} \|x\|_1 \tag{10.11}$$

and the signal is recovered as $f^\star = \Psi x^\star \in \mathbb{R}^N$.

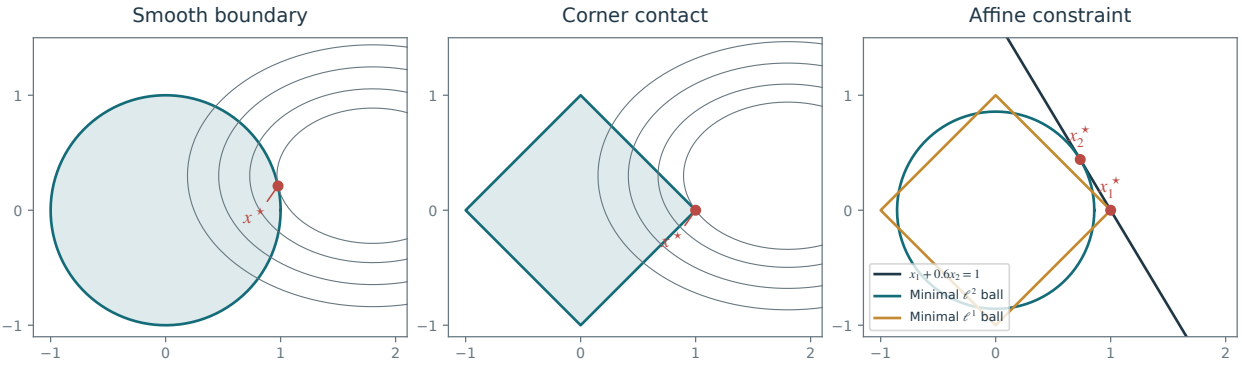

Figure 10.4. Geometry of convex optimization: smooth-boundary and corner contact with objective level sets, followed by the smallest scaled ℓ^2 and ℓ^1 balls touching an affine constraint. Red dots mark the contact points.

10.3 Iterative Soft Thresholding Algorithm

We now derive iterative methods for (10.10), beginning with a comparison to its constrained counterpart.

10.3.1 Noiseless Recovery as a Linear Program

Before treating $\lambda > 0$, note that (10.11) can be written as a linear program. Split $x = x_+ - x_-$ with $(x_+, x_-) \in (\mathbb{R}_+^Q)^2$ and solve

$$(x_+^\star, x_-^\star) \in \operatorname*{argmin}_{(x_+,x_-)\in(\mathbb{R}_+^Q)^2} \left\{ \langle x_+, \mathbb{1}_Q \rangle + \langle x_-, \mathbb{1}_Q \rangle \;;\; y = A(x_+ - x_-) \right\}. \tag{10.12}$$

Then recover $x^\star = x_+^\star - x_-^\star$. For small or moderate Q, the linear program can be solved by simplex or interior-point methods. For large imaging and learning problems, first-order methods are often preferable. One option is Douglas–Rachford splitting, developed in Section 19.5.2. The penalized problem (10.10) admits a particularly simple splitting algorithm because its fidelity term is smooth. The relative computational cost of the two formulations depends on the solver and on A.

10.3.2 Projected Gradient Descent for ℓ^1

We first solve (10.10) by projected gradient descent, which is analyzed in detail in Section 19.1.3. As in (10.12), we rewrite (10.10) as a constrained minimization problem of the form (19.4), with the nonnegative constraint set C and

$$u = (u_+, u_-) \in (\mathbb{R}^Q)^2, \quad C = (\mathbb{R}_+^Q)^2, \quad \text{and} \quad \mathcal{E}(u) = \frac{1}{2}\|A(u_+ - u_-) - y\|^2 + \lambda\langle u_+, \mathbb{1}_Q\rangle + \lambda\langle u_-, \mathbb{1}_Q\rangle.$$

Projection onto C is easy to compute

$$\mathrm{Proj}_{(\mathbb{R}_+^Q)^2}(u_+, u_-) = ((u_+)_\oplus, (u_-)_\oplus) \quad \text{where} \quad (r)_\oplus \overset{\text{def.}}{=} \max(r, 0),$$

and the gradient reads

$$\nabla\mathcal{E}(u_+, u_-) = (\eta + \lambda\mathbb{1}_Q, -\eta + \lambda\mathbb{1}_Q) \quad \text{where} \quad \eta = A^*(A(u_+ - u_-) - y)$$

Denoting $u^{(\ell)} = (u_+^{(\ell)}, u_-^{(\ell)})$ and $x^{(\ell)} \overset{\text{def.}}{=} u_+^{(\ell)} - u_-^{(\ell)}$, the iterates of projected gradient descent (19.5) are

$$u_+^{(\ell+1)} \overset{\text{def.}}{=} \left(u_+^{(\ell)} - \tau_\ell(\eta^{(\ell)} + \lambda)\right)_\oplus \quad \text{and} \quad u_-^{(\ell+1)} \overset{\text{def.}}{=} \left(u_-^{(\ell)} - \tau_\ell(-\eta^{(\ell)} + \lambda)\right)_\oplus$$

where $\eta^{(\ell)} \overset{\text{def.}}{=} A^*(Ax^{(\ell)} - y)$, and adding the scalar λ means adding $\lambda\mathbb{1}_Q$ componentwise.

Theorem 19.2 ensures that $u^{(\ell)} \to u^\star$, a solution of (19.4), if

$$\forall\, \ell, \quad 0 < \tau_{\min} < \tau_\ell < \tau_{\max} < \frac{1}{\|A\|^2},$$

The smooth objective in (u_+, u_-) has gradient Lipschitz constant $2\|A\|^2$, which explains the step-size bound. Consequently, $x^{(\ell)} \to x^\star = u_+^\star - u_-^\star$, a solution of (10.10).

10.3.3 Iterative Soft Thresholding and Forward Backward

Splitting positive and negative parts requires storing $2Q$ coefficients. The iterative soft thresholding algorithm (ISTA) avoids this duplication while retaining comparable convergence guarantees. For a fixed step $0 < \tau < 2/\|A\|^2$, ISTA converges to a minimizer. The surrogate derivation below uses the stricter bound $\tau \leqslant 1/\|A\|^2$, which gives a direct proof of energy decrease.

ISTA was derived by several authors [4, 3]. It is a special case of forward–backward proximal splitting [2], studied in Section 19.4.2.

We derive ISTA for the ℓ^1 penalty by minimizing a sequence of surrogate objectives. Rescale the objective in (10.10) as

$$\mathcal{E}(x) \overset{\text{def.}}{=} \frac{1}{2}\|y - Ax\|^2 + \lambda\|x\|_1$$

For a fixed reference point x', define

$$\mathcal{E}_\tau(x, x') \overset{\text{def.}}{=} \mathcal{E}(x) - \frac{1}{2}\|Ax - Ax'\|^2 + \frac{1}{2\tau}\|x - x'\|^2.$$

We have $\mathcal{E}_\tau(x, x) = \mathcal{E}(x)$, and the difference between the two objectives is

$$K(x, x') \overset{\text{def.}}{=} -\frac{1}{2}\|Ax - Ax'\|^2 + \frac{1}{2\tau}\|x - x'\|^2 = \frac{1}{2}\left\langle \left(\frac{1}{\tau}\mathrm{Id}_Q - A^*A\right)(x - x'),\ x - x'\right\rangle.$$

The correction $K(x, x')$ is nonnegative when $\lambda_{\max}(A^*A) \leqslant 1/\tau$, where the left side is the largest eigenvalue. Equivalently, $\tau \leqslant 1/\|A\|_{\text{op}}^2$, with $\|A\|_{\text{op}} = \sigma_{\max}(A)$ denoting the operator norm. Under this condition, $\mathcal{E}_\tau(x, x')$ is a surrogate that bounds the objective from above and agrees with it at the reference point:

$$\mathcal{E}(x) \leqslant \mathcal{E}_\tau(x, x'), \quad \mathcal{E}_\tau(x', x') = \mathcal{E}(x'), \quad \text{and} \quad \mathcal{E}(\cdot) - \mathcal{E}_\tau(\cdot, x') \text{ is smooth.}$$

Minimizing this surrogate at each iteration gives

$$x^{(\ell+1)} \overset{\text{def.}}{=} \operatorname*{argmin}_x \mathcal{E}_{\tau_\ell}(x, x^{(\ell)}) \tag{10.13}$$

The surrogate upper bound and equality at the current iterate imply

$$\mathcal{E}(x^{(\ell+1)}) \leqslant \mathcal{E}(x^{(\ell)}).$$

Proposition 10.2. *The iterates $x^{(\ell)}$ defined by (10.13) satisfy*

$$x^{(\ell+1)} = \mathcal{S}^1_{\lambda\tau_\ell}\left(x^{(\ell)} - \tau_\ell A^*(Ax^{(\ell)} - y)\right) \tag{10.14}$$

where $\mathcal{S}^1_\lambda(x) = (S^1_\lambda(x_m))_m$ and $S^1_\lambda(r) = \mathrm{sign}(r)(|r| - \lambda)_\oplus$ is the soft thresholding operator defined in (10.5).

Proof. Expanding the quadratic terms and collecting constants independent of the optimization variable gives

$$\begin{aligned}
\mathcal{E}_\tau(x, x') &= \frac{1}{2}\|Ax - y\|^2 - \frac{1}{2}\|Ax - Ax'\|^2 + \frac{1}{2\tau}\|x - x'\|^2 + \lambda\|x\|_1 \\
&= C + \frac{1}{2}\|Ax\|^2 - \frac{1}{2}\|Ax\|^2 + \frac{1}{2\tau}\|x\|^2 - \langle Ax,\ y\rangle + \langle Ax,\ Ax'\rangle - \frac{1}{\tau}\langle x,\ x'\rangle + \lambda\|x\|_1 \\
&= C + \frac{1}{2\tau}\|x\|^2 + \langle x,\ -A^*y + A^*Ax' - \frac{1}{\tau}x'\rangle + \lambda\|x\|_1 \\
&= C' + \frac{1}{\tau}\left(\frac{1}{2}\|x - (x' - \tau A^*(Ax' - y))\|^2 + \tau\lambda\|x\|_1\right)
\end{aligned}$$

Proposition 10.1 therefore identifies the minimizer of $\mathcal{E}_\tau(x, x')$ as $\mathcal{S}^1_{\lambda\tau}(x' - \tau A^*(Ax' - y))$. □

The iterations (10.14) coincide with the forward–backward iterates (19.20). For the coefficient-space objective (10.10), rescaled by λ, use the splitting

$$\mathcal{F} = \frac{1}{2}\|A \cdot - y\|^2 \quad \text{and} \quad \mathcal{G} = \lambda\|\cdot\|_1. \tag{10.15}$$

In the case (10.15), Proposition 10.1 shows that $\mathrm{Prox}_{\rho\mathcal{G}}$ is soft thresholding at level $\rho\lambda$.

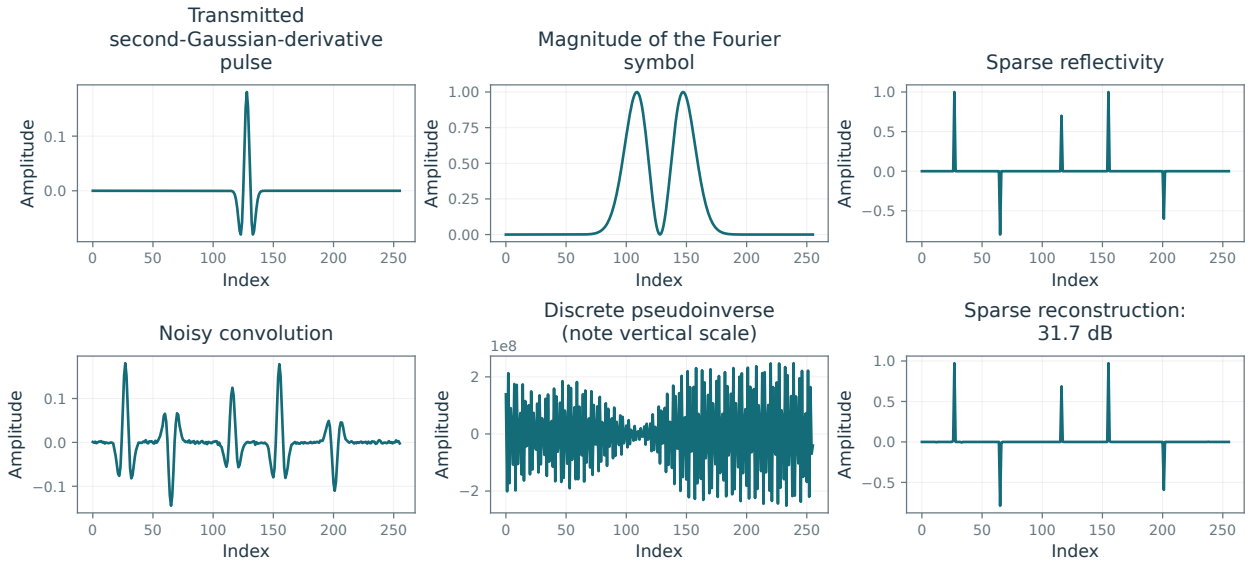

Figure 10.5. Sparse spike recovery by the pseudoinverse and by ℓ^1 regularization.

10.4 Example: Sparse Deconvolution

10.4.1 Sparse Spikes Deconvolution

Sparse spike deconvolution uses sparsity in the spatial domain, which corresponds to the orthonormal basis of Diracs $\psi_m[n] = \delta[n - m]$. In seismic imaging, a simple sparse reflectivity model represents f_0 as a few impulses associated with changes in acoustic impedance.

In a linearized one-dimensional model that neglects multiple reflections, the subsurface reflectivity f_0 produces observations $y = h \star f_0 + w$. Here h is the transmitted pulse, called a seismic wavelet. This use of the word differs from the orthogonal wavelet bases constructed in Chapter 4, although the terminology originated in seismic imaging.

The seismic wavelet h is typically band-pass, balancing spatial and frequency concentration. Figure 10.5 shows a second derivative of a Gaussian and its Fourier transform. The narrow transmitted band illustrates how acquisition suppresses information about f_0.

Using ℓ^1 regularization in the Dirac basis gives

$$f^\star = \operatorname*{argmin}_{f \in \mathbb{R}^N} \frac{1}{2}\|f \star h - y\|^2 + \lambda \sum_m |f_m|.$$

Figure 10.5 shows the result of ℓ^1 minimization with an oracle choice of λ that minimizes the error $\|f^\star - f_0\|$.

ISTA for sparse spike recovery alternates the updates

$$\tilde{f}^{(k)} = f^{(k)} - \tau h^\vee \star (h \star f^{(k)} - y)$$

and

$$f_m^{(k+1)} = S_{\lambda\tau}^1(\tilde{f}_m^{(k)})$$

where $h^\vee[n] = \overline{h[-n]}$ is the adjoint convolution kernel. It equals h for a real symmetric filter. The step size must satisfy

$$0 < \tau < 2/\|\Phi^*\Phi\| = 2/\max_\omega |\hat{h}(\omega)|^2$$

to guarantee convergence. Figure 10.6 shows convergence of both the objective values and the iterates. Objective convergence alone would not ensure convergence of the iterates when minimizers are nonunique; the forward–backward convergence theorem provides the latter guarantee here.

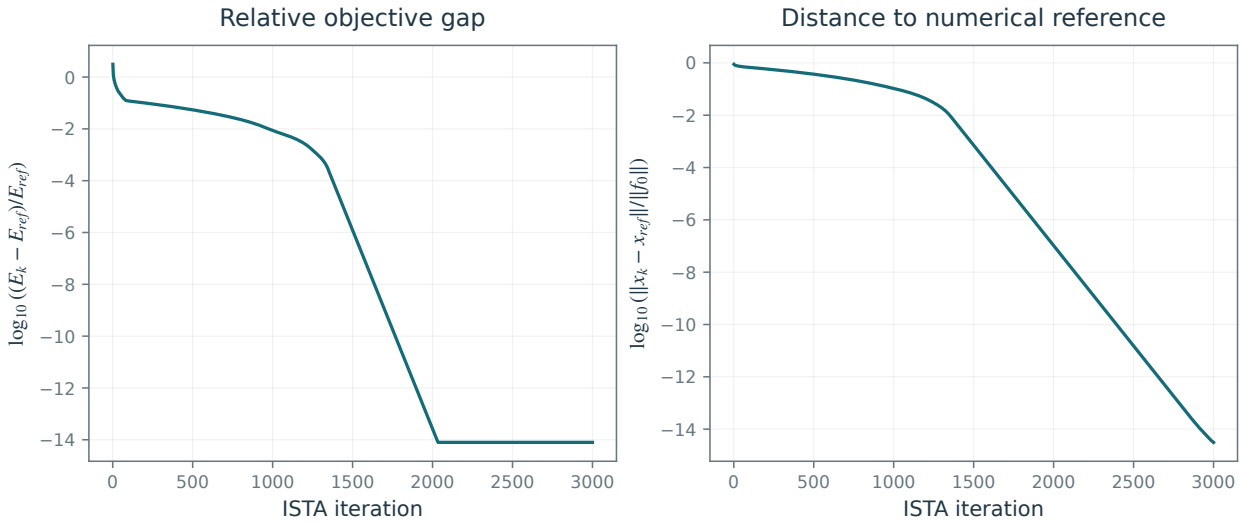

Figure 10.6. Convergence of the energy and iterates under iterative soft thresholding.

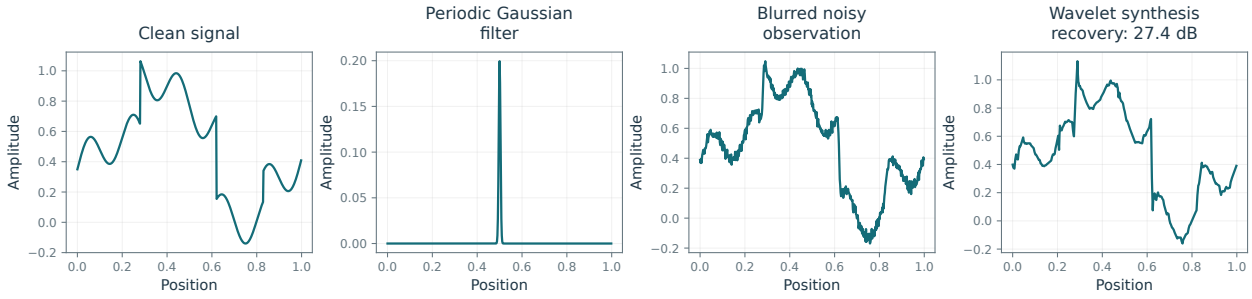

Figure 10.7. One-dimensional deconvolution with sparsity in an orthogonal wavelet basis.

10.4.2 Sparse Wavelets Deconvolution

Camera images can be blurred by defocus, motion during exposure, or diffraction. Assuming spatial invariance, we model the blur operator Φ by convolution:

$$y = f_0 \star h + w.$$

We use a Gaussian filter h of width $\mu > 0$. On a fixed d-dimensional domain, the number of effectively transmitted frequencies scales roughly as μ^{-d}, with a threshold-dependent factor, because higher frequencies are strongly attenuated.

Figures 10.7 and 10.8 show examples of signal and image acquisition with Gaussian blur.

Sobolev regularization (9.23) suppresses high-frequency noise more strongly than the squared-norm penalty (9.10). Both can blur sharp transitions. TV regularization and sparsity in a wavelet basis are better adapted to signals with edges.

Figure 10.7 compares wavelet and Sobolev regularization for a one-dimensional signal; Figure 10.8 shows the corresponding image-deblurring results. Replacing an orthogonal wavelet basis in (10.14) by a translation-invariant tight frame (7.12) can reduce artifacts. For a redundant frame, synthesis ISTA acts in coefficient space with $A = \Phi\Psi$. Simply analyzing, thresholding, and synthesizing is generally not the proximal map of the analysis penalty.

Figure 10.9 plots the SNR against λ. Computing the SNR requires the clean reference image f_0; the reported experiments use the maximizing value of λ.

10.4.3 Sparse Inpainting

This section continues the discussion in Section 9.6.2.

For noiseless inpainting, a small $\lambda > 0$ approximates constrained ℓ^1 minimization. Exact interpolation is obtained in the limit under the assumptions of Proposition 9.5. For an orthonormal basis, the iterative

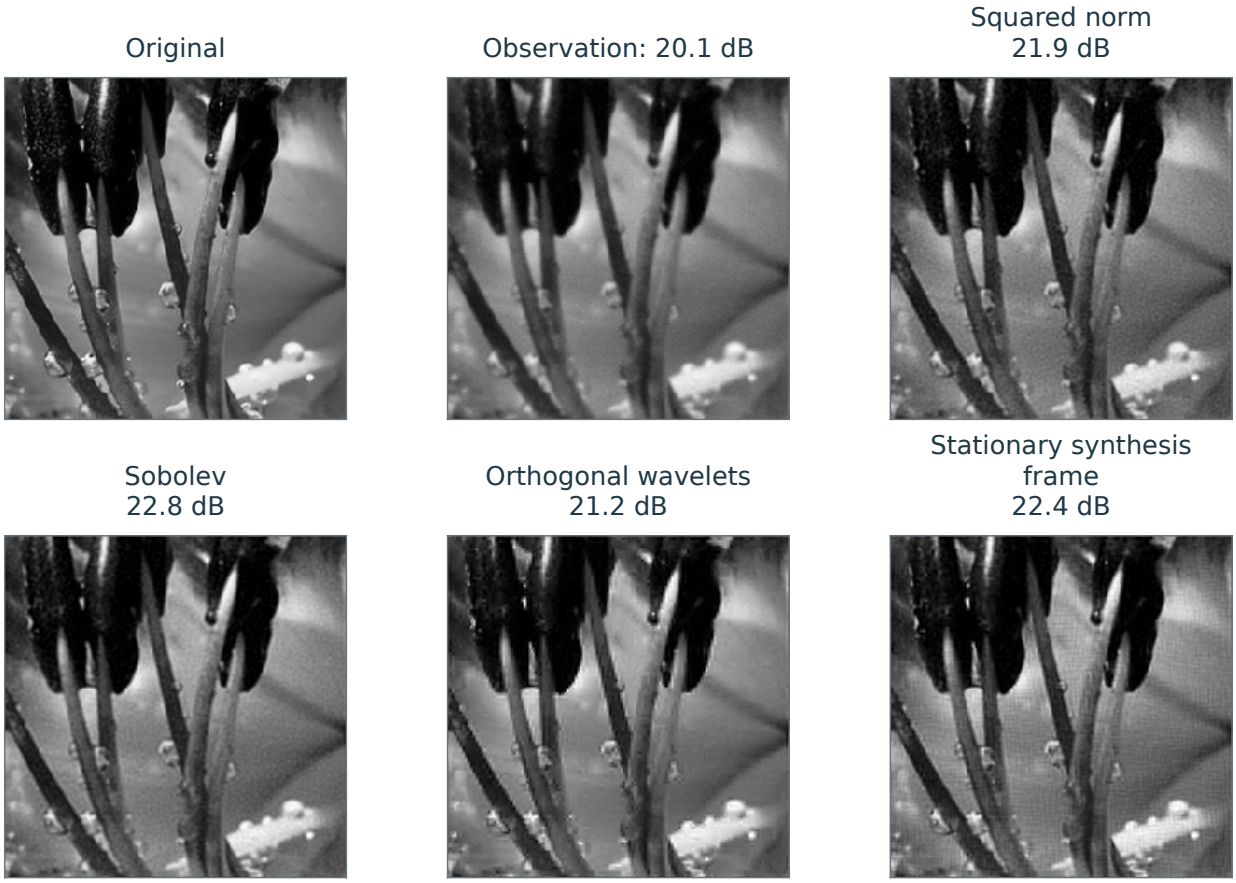

Figure 10.8. Image deconvolution.

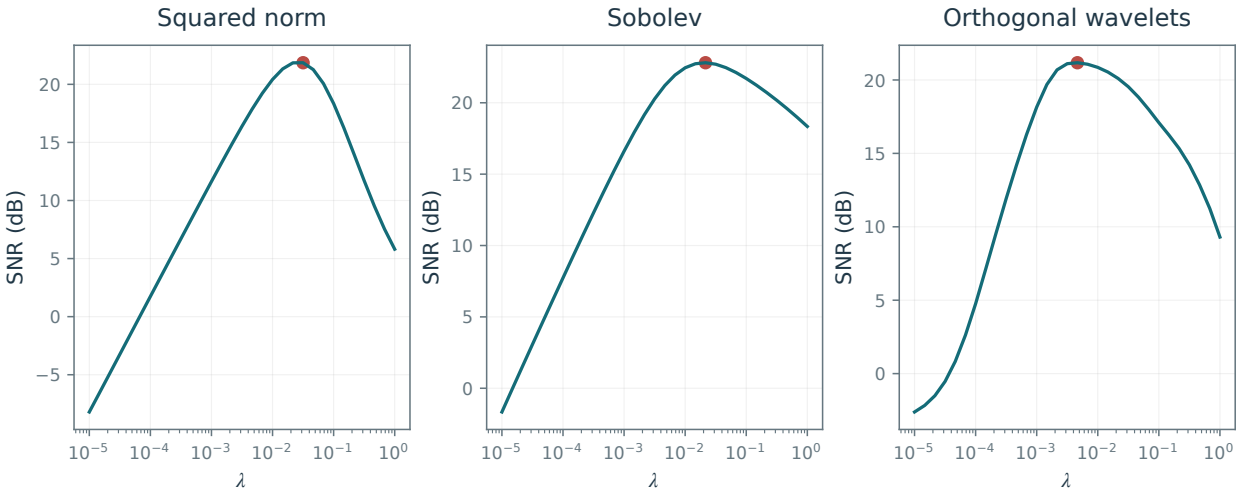

Figure 10.9. Reconstruction SNR as a function of the regularization parameter λ, swept logarithmically from 10^{-5} to 1. Red dots mark the measured maxima.

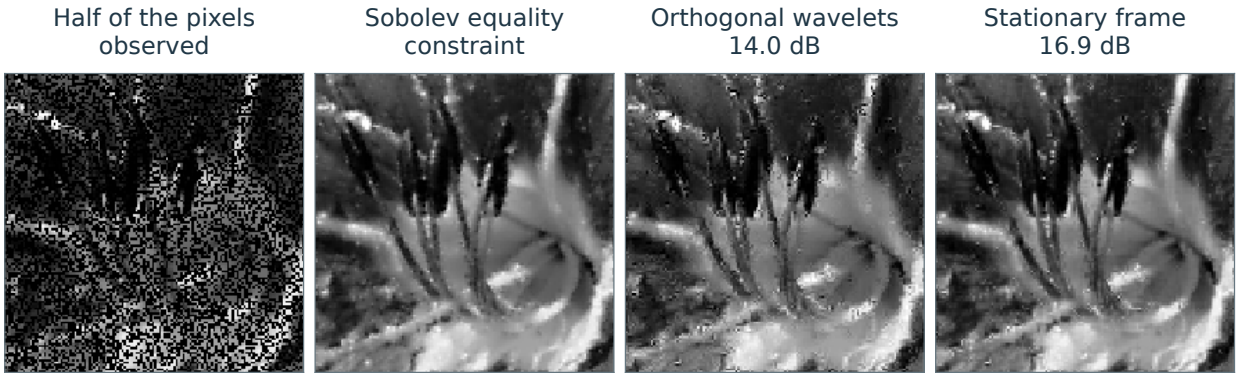

Figure 10.10. Inpainting with Sobolev and wavelet sparsity priors.

thresholding algorithm (10.14) takes the following form for $\tau = 1$:

$$f^{(k+1)} = \sum_m S_\lambda^1(\langle P_y(f^{(k)}), \psi_m \rangle)\psi_m$$

Figure 10.10 compares Sobolev inpainting with wavelet sparsity priors, including a translation-invariant frame.

References

[1] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[2] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Multiscale Modeling and Simulation*, 4(4), 2005.

[3] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. on Pure and Appl. Math.*, 57:1413–1541, 2004.

[4] M. Figueiredo and R. Nowak. An EM Algorithm for Wavelet-Based Image Restoration. *IEEE Trans. Image Proc.*, 12(8):906–916, 2003.

[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.

[7] Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse image and signal processing: Wavelets and related geometric multiscale analysis*. Cambridge university press, 2015.