

CHAPTER 5

Linear and Nonlinear Approximation

September 9, 2026

When only a limited number of coefficients can be retained, the choice of representation determines which features of a signal survive. Approximation rates quantify this tradeoff and help explain the performance of compression and denoising methods. We compare fixed and adaptive coefficient selection, relate Fourier and wavelet errors to smoothness and edges, and examine triangulations and curvelets for images with smooth contours.

The analysis concerns functions $f \in L^2([0, 1]^d)$ on a continuous domain, with $d = 1, 2$.

5.1 Approximation

5.1.1 Approximation in an Orthonormal Basis

Let $\mathcal{B} = \{\psi_m\}_m$ be an orthonormal basis of $L^2([0, 1]^d)$, with $d = 1$ for signals or $d = 2$ for images. The complete expansion in this basis,

$$f = \sum_{m \in \mathbb{Z}} \langle f, \psi_m \rangle \psi_m$$

reconstructs the signal exactly. Approximation and processing methods modify the coefficients $\langle f, \psi_m \rangle$, introducing a reconstruction error.

The simplest approximation reconstructs the signal from a subset $I_M \subset \mathbb{Z}$ of M coefficients:

$$f_M \stackrel{\text{def.}}{=} \sum_{m \in I_M} \langle f, \psi_m \rangle \psi_m, \quad \text{where } M = |I_M|.$$

The reconstructed signal f_M is the orthogonal projection of f onto the space

$$V_M \stackrel{\text{def.}}{=} \text{Span} \{ \psi_m ; m \in I_M \}.$$

If the selected space V_M depends on f , the overall approximation map $f \mapsto f_M$ may be nonlinear, even though projection onto each fixed space is linear.

Since the basis is orthogonal, the approximation error is

$$\|f - f_M\|^2 = \sum_{m \notin I_M} |\langle f, \psi_m \rangle|^2.$$

The central choice is the retained index set I_M , which may depend on the signal f .

5.1.2 Linear Approximation

Fixing I_M independently of the input gives a linear approximation: the same coefficients are retained for every f . The map $f \mapsto f_M$ is then orthogonal projection onto a fixed space V_M , and it satisfies

$$(f + g)_M = f_M + g_M$$

For the Fourier basis and an even coefficient budget, one usually selects the low-frequency atoms

$$I_M = \{-M/2 + 1, \dots, M/2\}.$$

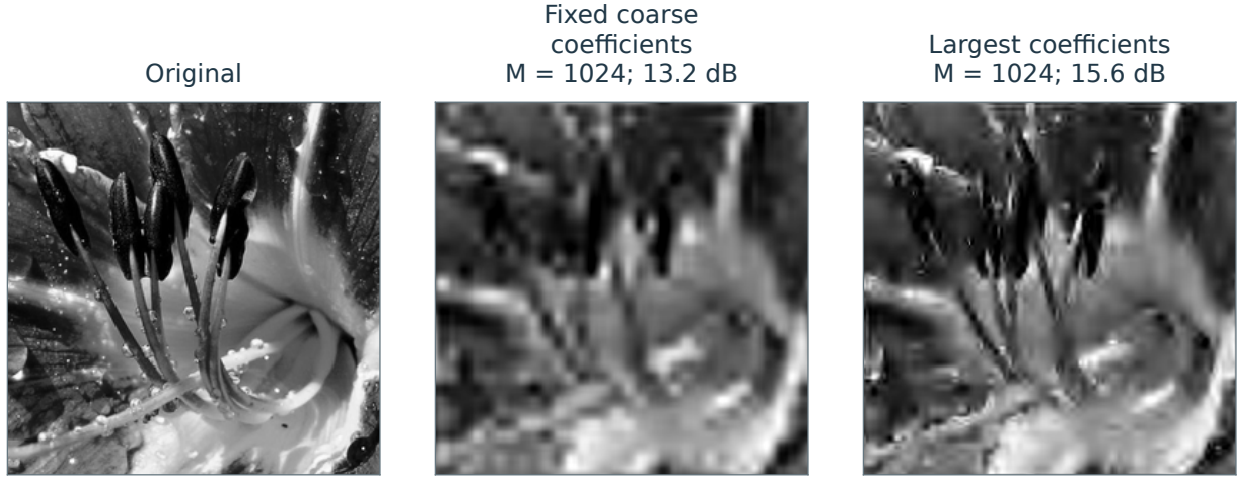


Figure 5.1. Linear versus nonlinear wavelet approximation.

For a 1-D wavelet basis, one usually selects the coarse wavelets

$$I_M = \{\text{coarsest scaling atoms and wavelets } (j, n) \text{ with } j \geq j_0\}$$

where j_0 is selected such that $|I_M| = M$.

Figure 5.1, center, illustrates linear wavelet approximation. Discarding all fine-scale details blurs singularities, where the approximation is least accurate.

5.1.3 Nonlinear Approximation

Allowing I_M to depend on f gives a nonlinear approximation. To choose I_M that minimizes $\|f - f_M\|$, orthogonality reduces the problem to retaining the M coefficients of largest magnitude:

$$I_M = \{M \text{ largest coefficients } |\langle f, \psi_m \rangle|\}.$$

When there are no ties at the cutoff, this can be obtained by thresholding

$$I_M = \{m ; |\langle f, \psi_m \rangle| > T\}$$

where T depends on the number of coefficients M ,

$$M = \#\{m ; |\langle f, \psi_m \rangle| > T\}.$$

Computation of the threshold. Order the coefficient magnitudes in nonincreasing order. For a finite vector, write

$$d_1 \geq \dots \geq d_N \geq 0, \quad \{d_m\}_{m=1}^N = \{|\langle f, \psi_m \rangle|\}_{m=1}^N. \quad (5.1)$$

If $d_M > d_{M+1}$, any threshold $d_{M+1} \leq T < d_M$ retains exactly M coefficients. At a tie, a deterministic rule must select among equal coefficients. Thus the threshold T and coefficient count M do not determine each other uniquely. Figure 5.2 illustrates the relationship.

The following proposition shows that the decay of the ordered coefficients governs the decay of the nonlinear approximation error.

Proposition 5.1. For $\alpha > 0$, with $(d_m)_{m \geq 1}$ in nonincreasing order, one has

$$d_m = O(m^{-\frac{\alpha+1}{2}}) \iff \|f - f_M\|^2 = O(M^{-\alpha}). \quad (5.2)$$

Proof. One has

$$\|f - f_M\|^2 = \sum_{m > M} d_m^2$$

The forward implication follows from $\sum_{m>M} m^{-\alpha-1} \leq \alpha^{-1} M^{-\alpha}$. For the converse, taking M even (integer rounding changes only constants),

$$d_M^2 \leq \frac{2}{M} \sum_{m=M/2+1}^M d_m^2 \leq \frac{2}{M} \sum_{m>M/2} d_m^2 = \frac{2}{M} \|f - f_{M/2}\|^2.$$

□

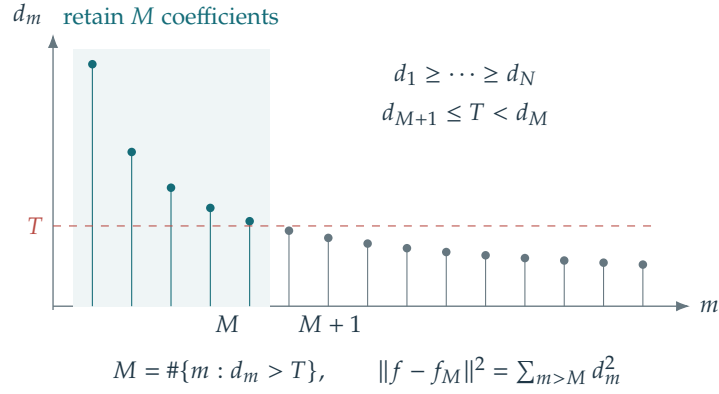


Figure 5.2. Ordered coefficient magnitudes and the threshold used for nonlinear approximation.

Hard thresholding. For the coefficient set selected by a threshold T , the nonlinear approximation can be written as

$$f_M = \sum_{|\langle f, \psi_m \rangle| > T} \langle f, \psi_m \rangle \psi_m = \sum_m S_T(\langle f, \psi_m \rangle) \psi_m, \quad (5.3)$$

where

$$S_T(x) = \begin{cases} x & \text{if } |x| > T \\ 0 & \text{if } |x| \leq T \end{cases} \quad (5.4)$$

is the hard thresholding operator displayed in Figure 5.3.

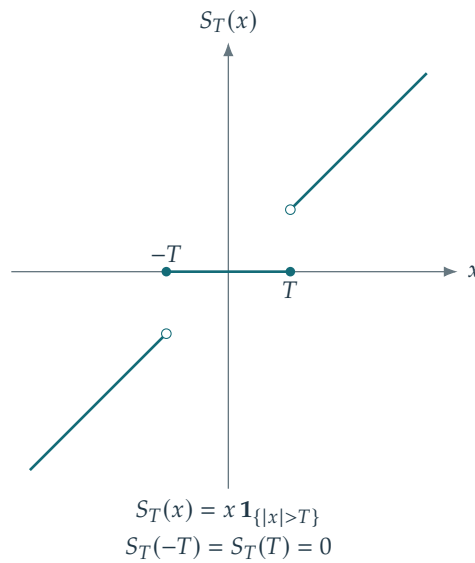


Figure 5.3. Hard thresholding.

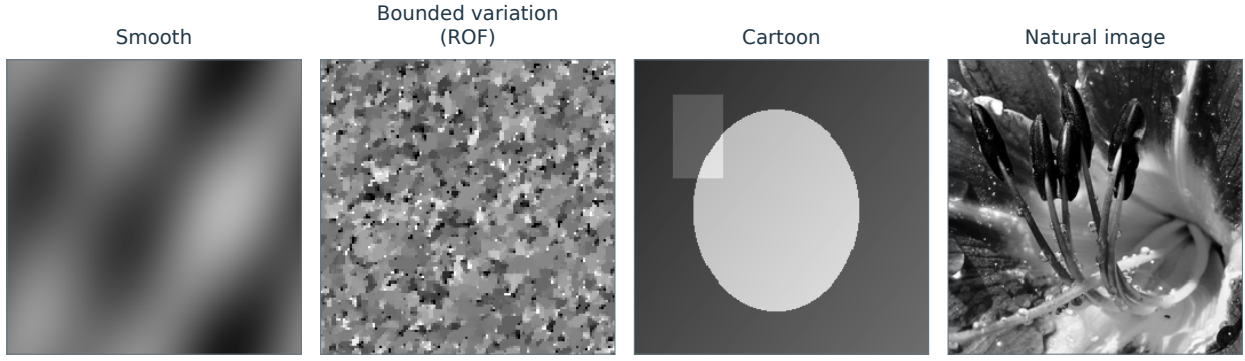


Figure 5.4. Examples of image models. The bounded-variation example is generated by ROF total-variation denoising of white noise.

5.2 Signal and Image Modeling

A signal model specifies a constraint $f \in \Theta$, where $\Theta \subset L^2([0, 1]^d)$ is the class of signals of interest. Figure 5.4 shows different image models, which we describe below.

5.2.1 Uniformly Smooth Signals and Images

Signals with derivatives. The simplest model consists of uniformly smooth signals with bounded derivatives

$$\Theta = \{f \in L^2([0, 1]^d) ; \|f\|_{C^\alpha} \leq C\}, \quad (5.5)$$

where $C > 0$ is fixed. For integer $\alpha \geq 1$, in one dimension

$$\|f\|_{C^\alpha} \stackrel{\text{def.}}{=} \max_{0 \leq k \leq \alpha} \left\| \frac{d^k f}{dt^k} \right\|_\infty.$$

In higher dimensions, take the maximum over all partial derivatives of total order at most α . For noninteger $\alpha = m + \beta$, $0 < \beta < 1$, use the usual Hölder norm: derivatives through order m are bounded, and order- m derivatives are β -Hölder continuous.

Sobolev smooth signals and images. For integer $\alpha \geq 1$, a C^α signal in the model (5.5) has derivatives with bounded energy:

$$\frac{d^\alpha f}{dt^\alpha}(t) = f^{(\alpha)}(t) \in L^2([0, 1]).$$

For integer α and periodic boundary conditions, use the identity

$$\hat{f}_m^{(\alpha)} = (2i\pi m)^\alpha \hat{f}_m$$

Here $\hat{f}$ denotes Fourier coefficients as in (2.2), now on $\mathbb{R}/\mathbb{Z}$ rather than $\mathbb{R}/2\pi\mathbb{Z}$:

$$\hat{f}_n \stackrel{\text{def.}}{=} \int_0^1 e^{-2i\pi n x} f(x) dx,$$

This motivates the homogeneous Sobolev seminorm

$$\|f\|_{\text{Sob}(\alpha)}^2 = \sum_{m \in \mathbb{Z}} |2\pi m|^{2\alpha} |\langle f, e_m \rangle|^2, \quad (5.6)$$

so that $\|f\|_{\text{Sob}(\alpha)} = \|f^{(\alpha)}\|$ for smooth periodic functions. The same seminorm applies when the derivatives exist only in the distributional sense and belong to $L^2([0, 1])$.

This definition extends to distributions and signals $f \in L^2([0, 1]^d)$ of arbitrary dimension d as

$$\|f\|_{\text{Sob}(\alpha)}^2 = \sum_{m \in \mathbb{Z}^d} (2\pi \|m\|)^{2\alpha} |\langle f, e_m \rangle|^2, \quad (5.7)$$

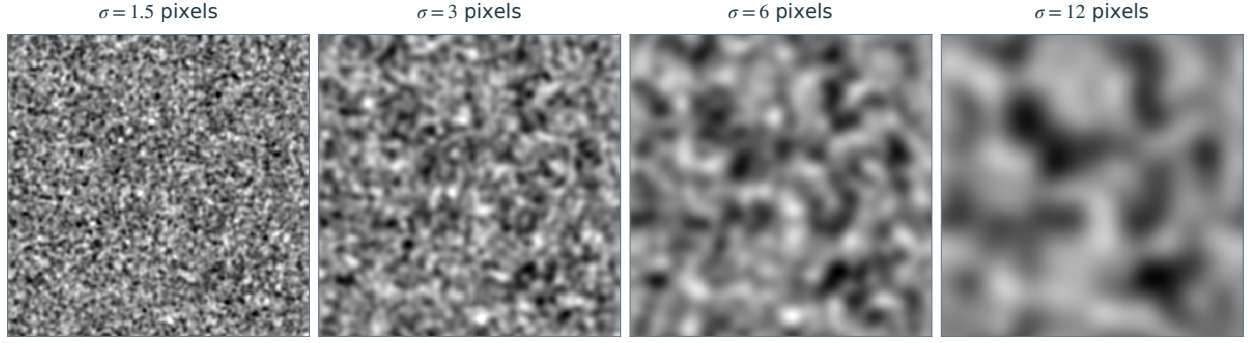


Figure 5.5. Progressive Gaussian filtering of one white-noise image. Smoothing widths increase from left to right; the displayed images have the same mean and variance.

The periodic Sobolev model

$$\Theta = \left\{ f \in L^2([0, 1]^d) ; \|f\|_2^2 + \|f\|_{\text{Sob}(\alpha)}^2 \leq C \right\} \quad (5.8)$$

includes the corresponding periodic C^α model (5.5) for integer α , after adjusting the radius of the ball. For noninteger α , Hölder regularity gives Sobolev regularity of every order $s < \alpha$, but need not give regularity of order α itself.

Figure 5.5 shows progressively smoother images obtained by Gaussian filtering of the same white-noise realization. Each result is rescaled to the same mean and variance, so the comparison emphasizes spatial regularity rather than a loss of contrast.

5.2.2 Piecewise Regular Signals and Images

Piecewise smooth signals. A piecewise smooth 1-D signal $f \in L^2([0, 1])$ is C^α on the intervals separated by at most K interior singular points:

$$\Theta = \left\{ f \in L^2([0, 1]) ; \exists 0 = t_0 < t_1 < \dots < t_{K+1} = 1, \max_{0 \leq i \leq K} \|f|_{(t_i, t_{i+1})}\|_{C^\alpha} \leq C \right\} \quad (5.9)$$

Here $f|_{(t_i, t_{i+1})}$ denotes the restriction of f to the open interval (t_i, t_{i+1}) .

Piecewise smooth images. A piecewise smooth image is a function $f \in L^2([0, 1]^2)$ with C^α regularity away from at most K rectifiable curves:

$$\Theta = \left\{ f \in L^2([0, 1]^2) ; \exists \Gamma = \bigcup_{i=0}^{K-1} \gamma_i, \|f\|_{C^\alpha(\Gamma^c)} \leq C_1 \quad \text{and} \quad |\gamma_i| \leq C_2 \right\} \quad (5.10)$$

Here $|\gamma_i|$ denotes curve length. The C^α bound applies separately within each smooth region, with uniform one-sided regularity up to its edges.

Segmentation methods such as the one proposed by Mumford and Shah [3] implicitly assume such a piecewise smooth image model.

5.2.3 Bounded Variation Signals and Images

Bounded variation provides a more flexible model for signals with edges

$$\Theta = \left\{ f \in L^2([0, 1]^d) ; \|f\|_\infty \leq C_1 \quad \text{and} \quad \|f\|_{\text{TV}} \leq C_2 \right\}. \quad (5.11)$$

For $d = 1$ and $d = 2$, this model includes the piecewise smooth signal and image models (5.9) and (5.10) when $\alpha \geq 1$, after adjusting the constants. Bounds on the derivatives, jump sizes, and total edge length then control the variation.

The total variation of a smooth function is

$$\int \|\nabla f(x)\| dx$$

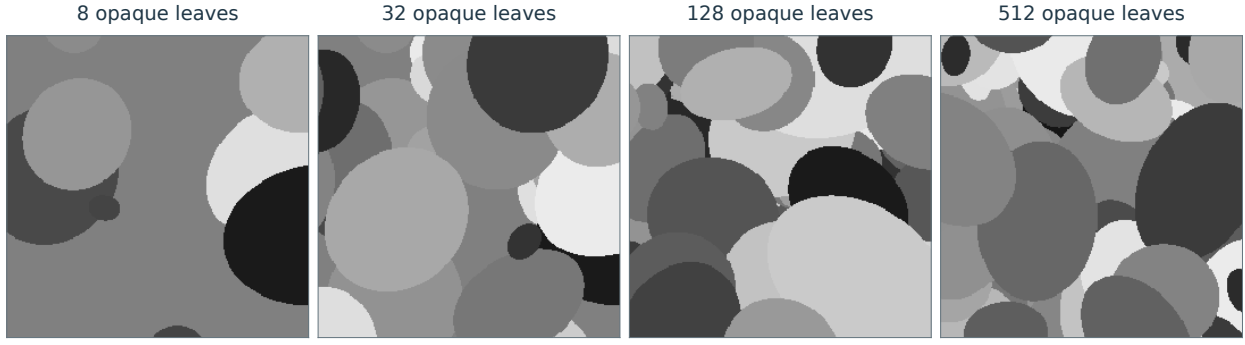


Figure 5.6. Progressive dead-leaves construction: later opaque shapes cover earlier ones, producing a family of cartoon images.

where

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_i} \right)_{i=0}^{d-1} \in \mathbb{R}^d$$

is the gradient at x . Total variation also extends to discontinuous images, including those with jumps across edges. The coarea formula expresses the total variation of a piecewise smooth image as an integral of the perimeters of its superlevel sets:

$$\|f\|_{\text{TV}} = \int_{-\infty}^{\infty} \text{Per}(\{f > t\}; \Omega) dt, \quad \Omega = (0, 1)^d. \quad (5.12)$$

Here the perimeter is relative to the image domain. For smooth functions in two dimensions it agrees, for almost every t , with the length of the level curve. For a bounded set $\Omega \subset \mathbb{R}^2$ with piecewise smooth boundary, taking total variation on all of $\mathbb{R}^2$ gives

$$\|1_{\Omega}\|_{\text{TV}} = |\partial\Omega|.$$

The model of bounded variation was introduced in image processing by Rudin, Osher and Fatemi [4].

5.2.4 Cartoon Images

The bounded-variation model (5.11) controls total perimeter but does not require individual edges to be smooth. For images with smooth contours, incorporating that additional geometry can improve approximation and processing.

The C^α cartoon model consists of 2-D functions that are C^α away from at most K smooth edge curves γ_i :

$$\Theta = \left\{ f \in L^2([0, 1]^2); \exists \Gamma = \bigcup_{i=0}^{K-1} \gamma_i, \|f\|_{C^\alpha(\Gamma^c)} \leq C_1 \quad \text{and} \quad \|\gamma_i\|_{C^\alpha} \leq C_2 \right\} \quad (5.13)$$

Each curve is parameterized by arc length on $[0, A_i]$, with a uniform bound on A_i and its C^α norm. Figure 5.6 illustrates a progressive dead-leaves construction: opaque shapes are placed successively, producing smooth visible contour segments and occlusion boundaries.

Optical diffraction can blur these edges, motivating the blurred cartoon model

$$\tilde{\Theta} = \{ \tilde{f} = f \star h \in L^2([0, 1]^2); f \in \Theta \quad \text{and} \quad h \in \mathcal{H} \} \quad (5.14)$$

where Θ is the sharp-image model (5.13) and $\mathcal{H}$ specifies the admissible blur kernels. For example, $h \geq 0$ may be required to be smooth and localized in both space and frequency. Unknown blur makes it difficult to first detect the edges Γ and then process the smooth regions separately.

Figure 5.7 shows examples of images in Θ and $\tilde{\Theta}$.

5.3 Efficient Approximation

5.3.1 Decay of Approximation Error

To process signals in Θ efficiently, we seek an orthonormal basis in which the nonlinear approximation error $\|f - f_M\|$ tends to 0 as rapidly as possible when M grows.

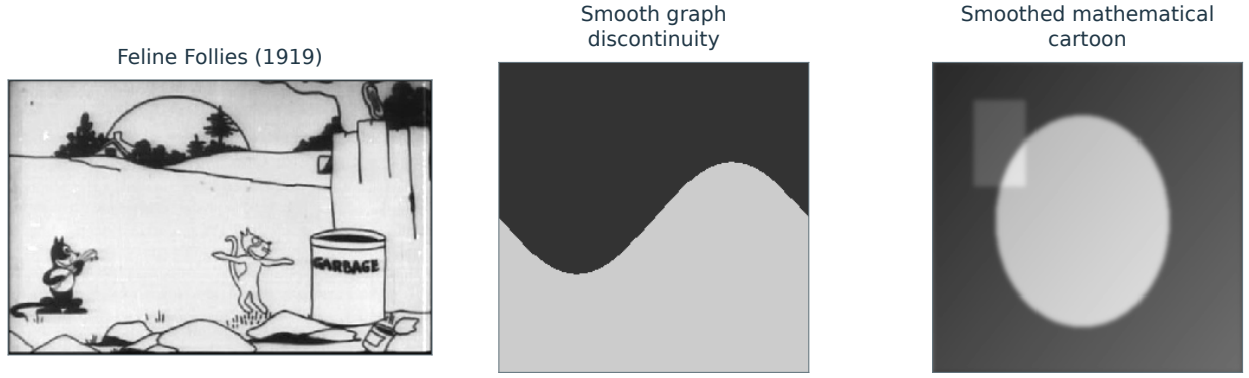


Figure 5.7. A grayscale frame from *Feline Follies* (1919), a mathematical jump along a smooth graph, and a Gaussian-smoothed cartoon. Film frame: Pat Sullivan, via Wikimedia Commons, public domain.

Polynomial error decay. This error decay is measured using a power law

$$\forall f \in \Theta, \forall M, \quad \|f - f_M\|^2 \leq C_f M^{-\alpha} \quad (5.15)$$

Here α is common to all signals in the model; larger values mean faster decay. The exponent depends on the basis and the model Θ , and measures how efficiently the basis represents that class. The constant C_f may depend on the particular signal f .

Relevance for compression, denoising and inverse problems. Approximation rates have practical consequences. Section 6.1 relates compression error to nonlinear approximation error, explaining why an effective approximation basis also supports efficient compression.

Chapter 7 develops a similar connection for thresholding denoisers. Thresholding forms a nonlinear approximation of the noisy image, and the expected denoising error is controlled by how well the basis approximates the clean signal.

Chapter 10 uses compressibility in a chosen basis to address inverse problems such as super-resolution and inpainting. Efficient approximation of the unknown signal helps recover missing information, provided the measurement operator retains enough information about sparse combinations of basis atoms. Recovery performance therefore depends on both the basis and the measurements.

Comparison of signals. With the basis fixed, the decay of $\|f - f_M\|$ compares the approximation complexity of different images. Figure 5.9 illustrates how intricate geometry and textures slow down wavelet approximation.

To reveal approximate power-law behavior, the error curves are displayed on logarithmic axes. An exact power law has the form

$$\log(\|f - f_M\|^2) = \text{cst} - \alpha \log(M)$$

and appears as a straight line with slope $-\alpha$. For sampled data, this behavior is generally studied when $M \ll N$; as $M \approx N$, retaining nearly all coefficients drives the error to zero.

5.3.2 Comparison of bases.

For a fixed image f , the decay of $\|f - f_M\|$ compares the efficiency of different bases. Figure 5.11 makes this comparison for an image containing contours and textures. The Fourier basis of Section 2.5 performs poorly here: periodic boundaries create artifacts, and globally supported atoms do not follow local contours. The cosine basis removes the periodic boundary discontinuity through symmetric extension, but its atoms remain global. Local DCT bases use cosine atoms on small square patches, improving localization. At small coefficient budgets M , however, their block structure produces visible artifacts. The isotropic wavelet basis of Section 4.5.2 performs best in this example by combining localized atoms with a multiresolution organization.

Figure 5.12 summarizes the approximation rates developed in the following sections for different data models.

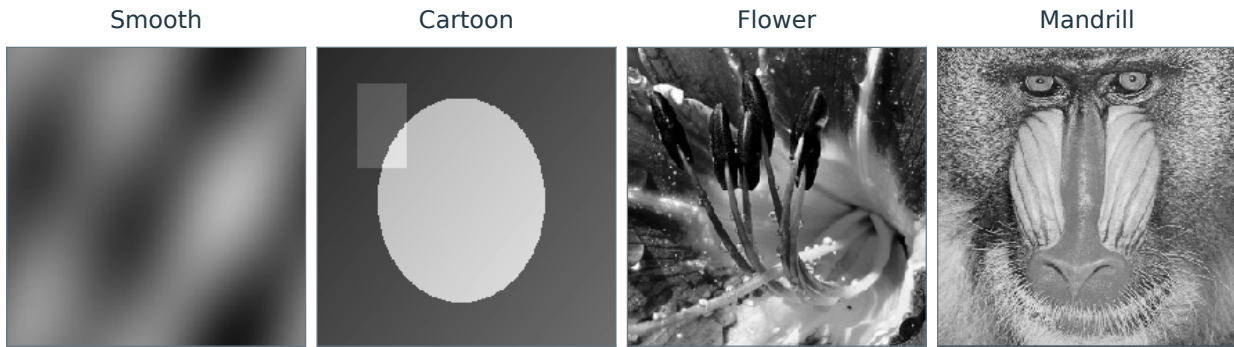


Figure 5.8. Test images with different structures: smooth, cartoon, flower, and mandrill.

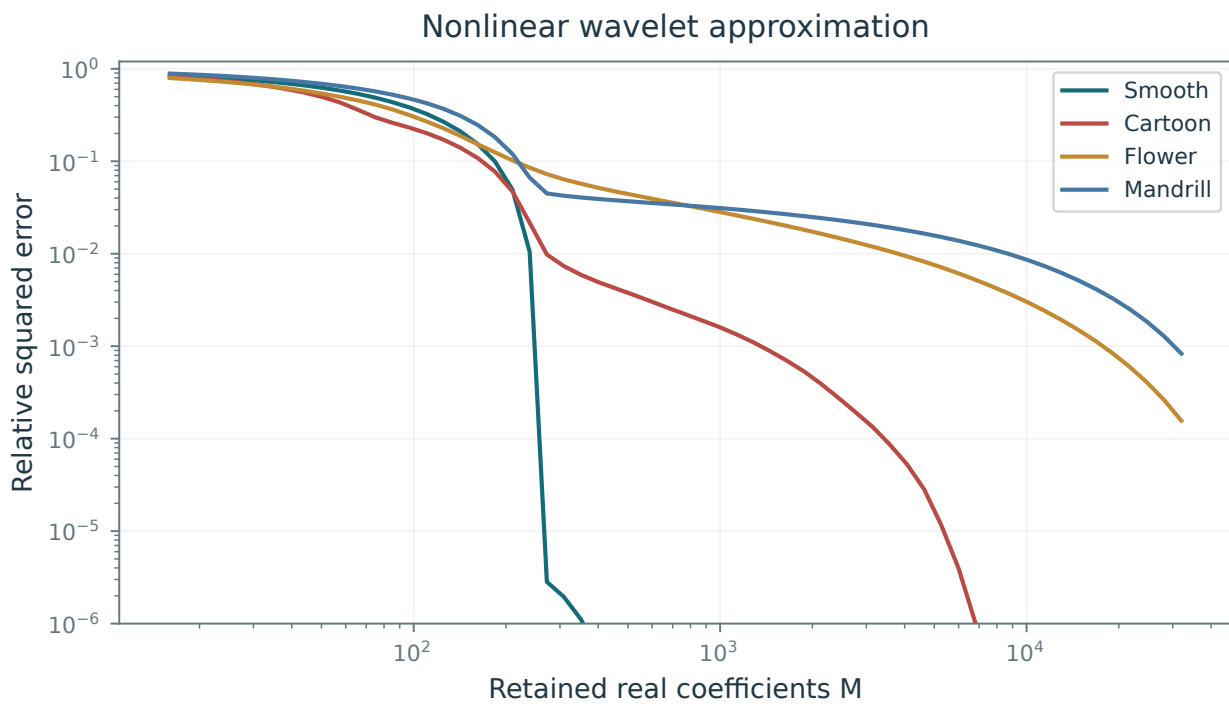


Figure 5.9. Wavelet approximation error for the images in Figure 5.8.

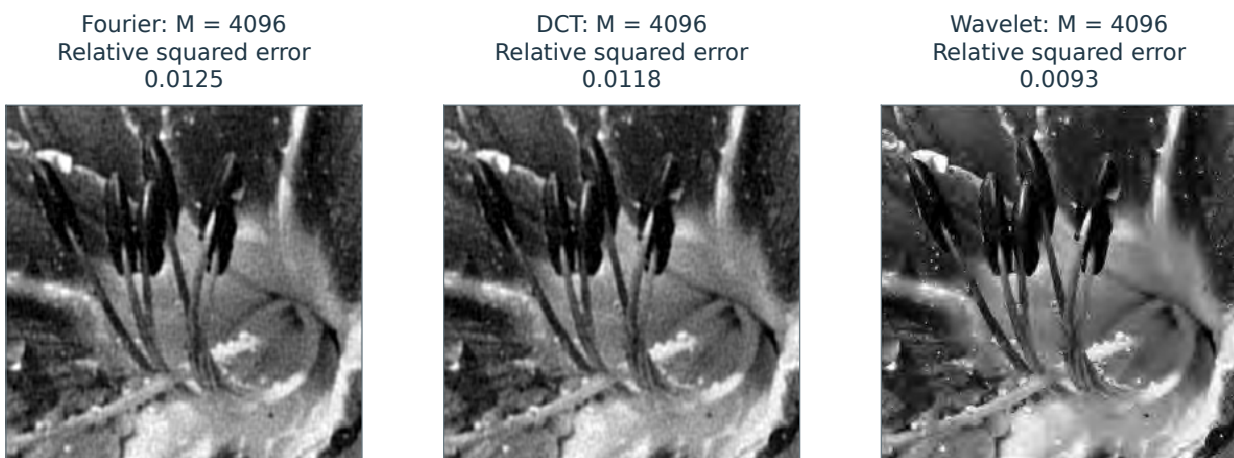


Figure 5.10. Fourier, DCT, and wavelet approximations of the same 256×256 flower image, each retaining $M = 4096$ real orthonormal coefficients.

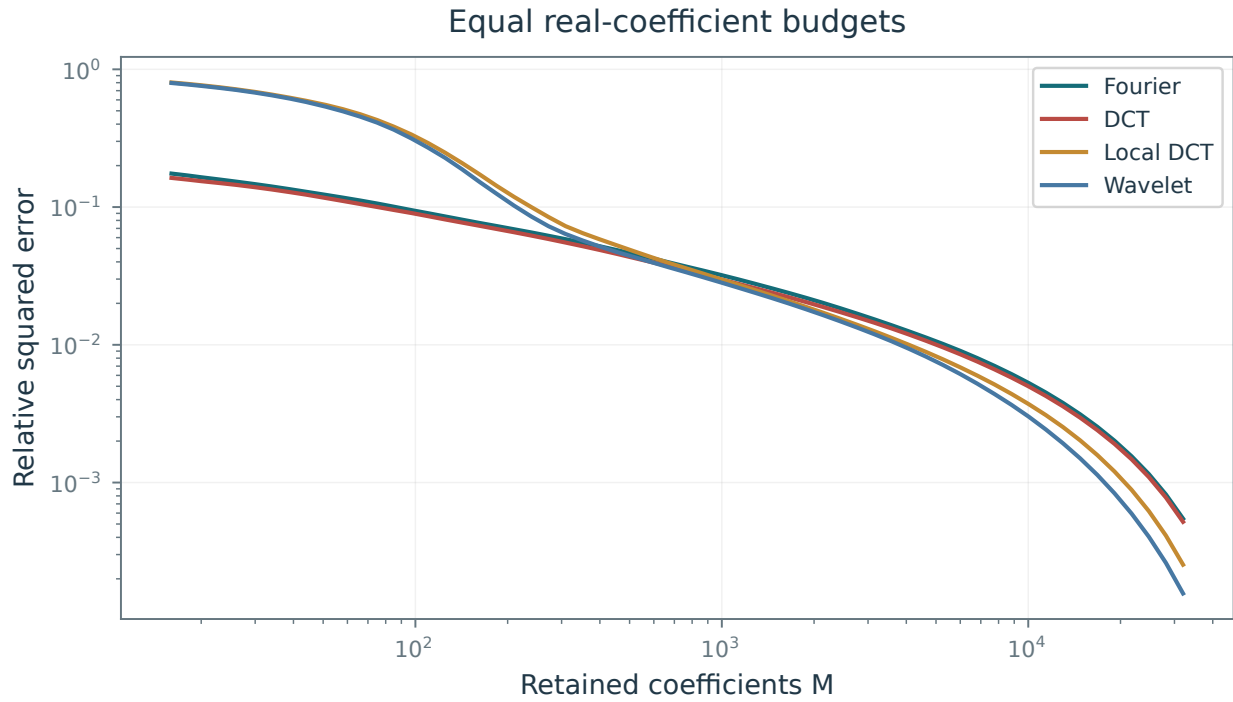


Figure 5.11. Comparison of approximation error decay for different bases.

	Sobolev 1-D	Piecewise smooth	Sobolev 2-D	Bounded variation	Cartoon, C^2
$\ f - f_M\ _2^2 \lesssim$	Sobolev 1-D	piecewise 1-D	Sobolev 2-D	BV 2-D	cartoon 2-D
Fourier, L	$M^{-2\alpha}$	M^{-1}	$M^{-\alpha}$	$M^{-1/2}$	$M^{-1/2}$
Fourier, NL	$M^{-2\alpha}$	M^{-1}	$M^{-\alpha}$	$M^{-1/2}$	$M^{-1/2}$
Wavelets, L	$M^{-2\alpha}$	M^{-1}	$M^{-\alpha}$	$M^{-1/2}$	$M^{-1/2}$
Wavelets, NL	$M^{-2\alpha}$	$M^{-2\alpha}$	$M^{-\alpha}$	M^{-1}	M^{-1}

Cartoon images: adaptive triangulations $O(M^{-2})$; curvelets $O(M^{-2}(\log M)^3)$.

L: linear; NL: nonlinear. Sobolev columns: H^α , $\alpha \geq 1$.

Use compatible wavelet bases and boundary conditions.

The BV class also has the uniform bound assumed in the chapter.

Figure 5.12. Summary of linear and nonlinear approximation rates $\|f - f_M\|_2^2$ for different classes of 1-D signals and images.

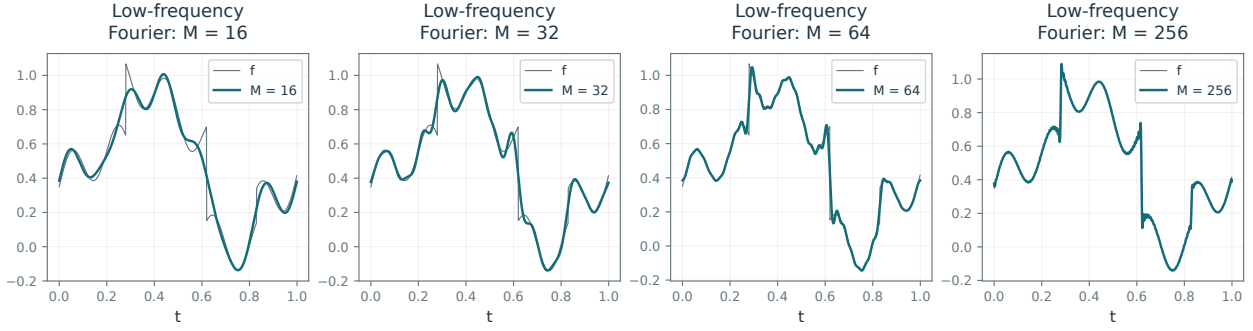


Figure 5.13. Fourier approximation of a signal.

5.4 Fourier Linear Approximation of Smooth Functions

For integer α , the smooth model (5.5) bounds all continuous derivatives through that order. Greater regularity of f , measured by a larger α , permits faster approximation. Figure 5.5 illustrates the related Sobolev measure of smoothness.

5.4.1 1-D Fourier Approximation

A 1-D signal $f \in L^2([0, 1])$ is associated with a 1-periodic function $f(t + 1) = f(t)$ defined for $t \in \mathbb{R}$.

Low pass approximation. For even M , consider the linear Fourier approximation that retains frequencies up to $M/2$:

$$f_M^{\text{lin}} = \sum_{m=-M/2}^{M/2} \langle f, e_m \rangle e_m$$

where we use the 1-D Fourier atoms

$$\forall m \in \mathbb{Z}, \quad e_m(t) \stackrel{\text{def}}{=} e^{2i\pi m t}.$$

This symmetric cutoff retains $M + 1$ Fourier atoms; the extra coefficient does not affect the asymptotic rates below.

Figure 5.13 shows these approximations for increasing M . The jumps in f produce substantial error and ringing near the singularities.

This low-pass approximation is a convolution, since

$$f_M = \sum_{m=-M/2}^{M/2} \langle f, e_m \rangle e_m = f \star h_M \quad \text{where} \quad \hat{h}_M \stackrel{\text{def}}{=} 1_{[-M/2, M/2]}.$$

The kernel h_M is the Dirichlet kernel introduced in the Fourier chapter.

The following elementary bound shows that this approximation error decays for C^α signals.

Proposition 5.2. For integer $\alpha \geq 1$ and $f \in C^\alpha(\mathbb{R}/\mathbb{Z})$

$$\|f - f_M^{\text{lin}}\|^2 = O(M^{-2\alpha+1}).$$

Proof. Using integration by parts, one shows that for a C^α function f in the smooth signal model (5.5), $\mathcal{F}(f^{(\ell)})_m = (2im\pi)^\ell \hat{f}_m$, so that Cauchy–Schwarz gives

$$|\langle f, e_m \rangle| \leq |2\pi m|^{-\alpha} \|f^{(\alpha)}\|_2.$$

This implies

$$\|f - f_M^{\text{lin}}\|^2 = \sum_{|m| > M/2} |\langle f, e_m \rangle|^2 \leq \|f^{(\alpha)}\|_2^2 \sum_{|m| > M/2} |2\pi m|^{-2\alpha} = O(M^{-2\alpha+1}).$$

□

Using the full energy of $f^{(\alpha)}$ gives a sharper bound under the same assumptions. The argument also applies to the larger Sobolev class.

Proposition 5.3. *For $\alpha > 0$ and a periodic signal in the Sobolev model (5.8), one has*

$$\|f - f_M^{\text{lin}}\|^2 \leq C \|f\|_{\text{Sob}(\alpha)}^2 M^{-2\alpha}. \quad (5.16)$$

Proof. One has

$$\|f\|_{\text{Sob}(\alpha)}^2 = \sum_m |2\pi m|^{2\alpha} |\langle f, e_m \rangle|^2 \geq \sum_{|m| > M/2} |2\pi m|^{2\alpha} |\langle f, e_m \rangle|^2 \quad (5.17)$$

$$\geq (\pi M)^{2\alpha} \sum_{|m| > M/2} |\langle f, e_m \rangle|^2 \geq (\pi M)^{2\alpha} \|f - f_M^{\text{lin}}\|^2. \quad (5.18)$$

□

This rate is optimal uniformly over a Sobolev ball. The best M -term Fourier approximation satisfies the same upper bound because it is at least as accurate as the low-pass approximation; individual functions may have faster decay.

For signals with piecewise Lipschitz regularity in the model (5.9), such as the example in Figure 5.13, the general bound for both linear and nonlinear Fourier approximation errors is

$$\|f - f_M\|^2 \leq C_f M^{-1} \quad (5.19)$$

and Fourier atoms are no longer optimal for approximation. For example, the interval indicator $f = 1_{[a,b]}$ has coefficients $\hat{f}_m = (e^{-2\pi i m a} - e^{-2\pi i m b}) / (2\pi i m)$ for $m \neq 0$, and hence $\|f - f_M\|^2 \asymp M^{-1}$ when $0 < b - a < 1$ for both linear and nonlinear approximations.

5.4.2 Sobolev Images

The same argument applies to images and higher-dimensional data through the Sobolev seminorm (5.7) with $d > 1$.

For an α -regular Sobolev image, the linear and nonlinear approximation errors satisfy

$$\|f - f_M\|^2 \leq C \|f\|_{\text{Sob}(\alpha)}^2 M^{-\alpha}.$$

For d -dimensional data $f : [0, 1]^d \rightarrow \mathbb{R}$, the corresponding error bound is $O(M^{-2\alpha/d})$.

For an image with piecewise Lipschitz regularity and finite-length edges in the model (5.10), the linear and nonlinear errors satisfy the slower general bound

$$\|f - f_M\|^2 \leq C_f M^{-1/2}, \quad (5.20)$$

and Fourier atoms are no longer optimal for approximation.

5.5 Wavelet Approximation of Piecewise Smooth Functions

Wavelets improve approximation near singularities because their support is localized.

5.5.1 Decay of Wavelet Coefficients

To approximate smooth regions efficiently, choose wavelets with sufficiently many vanishing moments, denoted by p :

$$\forall k = 0, \dots, p-1, \quad \int \psi(x) x^k dx = 0.$$

Choose $p \geq \alpha$, where α is the regularity of the signal outside singularities (for instance jumps or kinks).

To quantify the approximation error decay for piecewise smooth signals (5.9) and images (5.10), one treats wavelets supported in smooth regions separately from those crossing singularities. Figure 5.15 locates the singular and regular regions of a signal and an image.

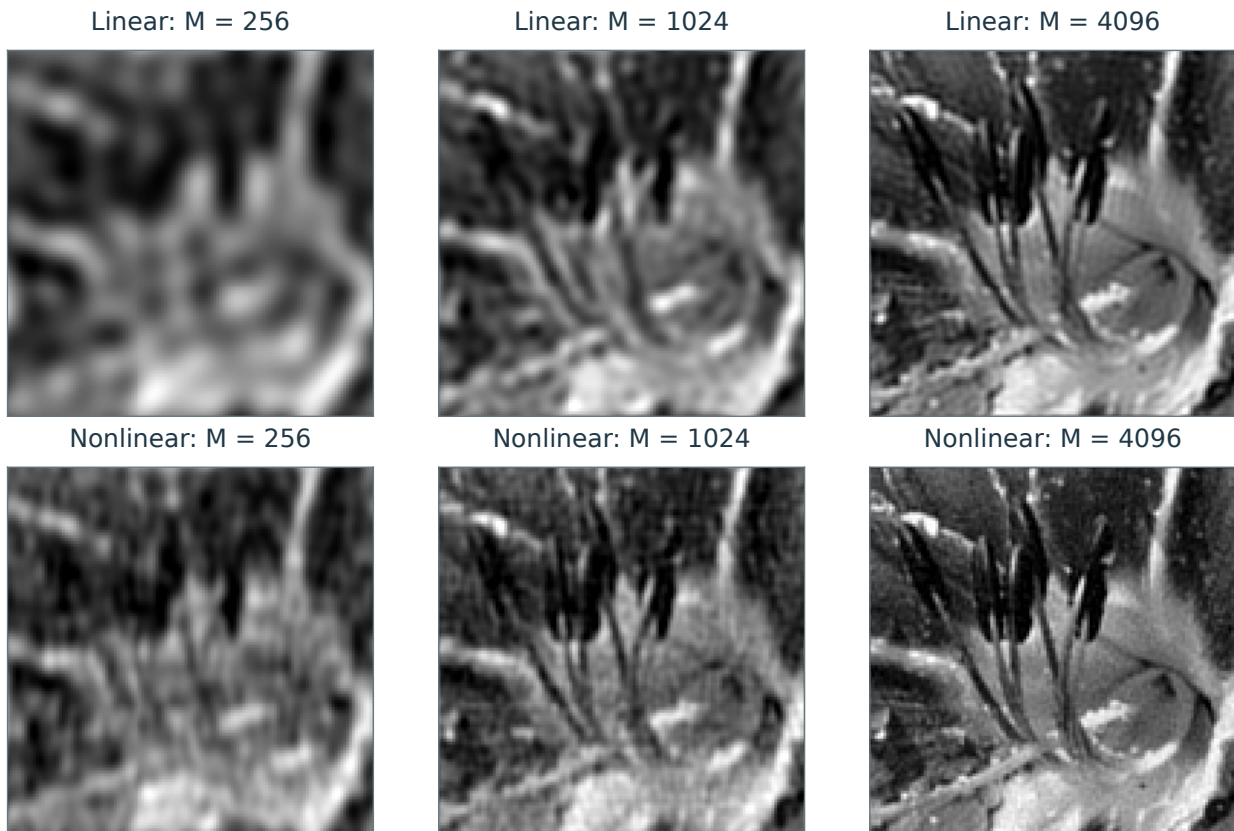


Figure 5.14. Linear (top row) and nonlinear (bottom row) Fourier approximation.

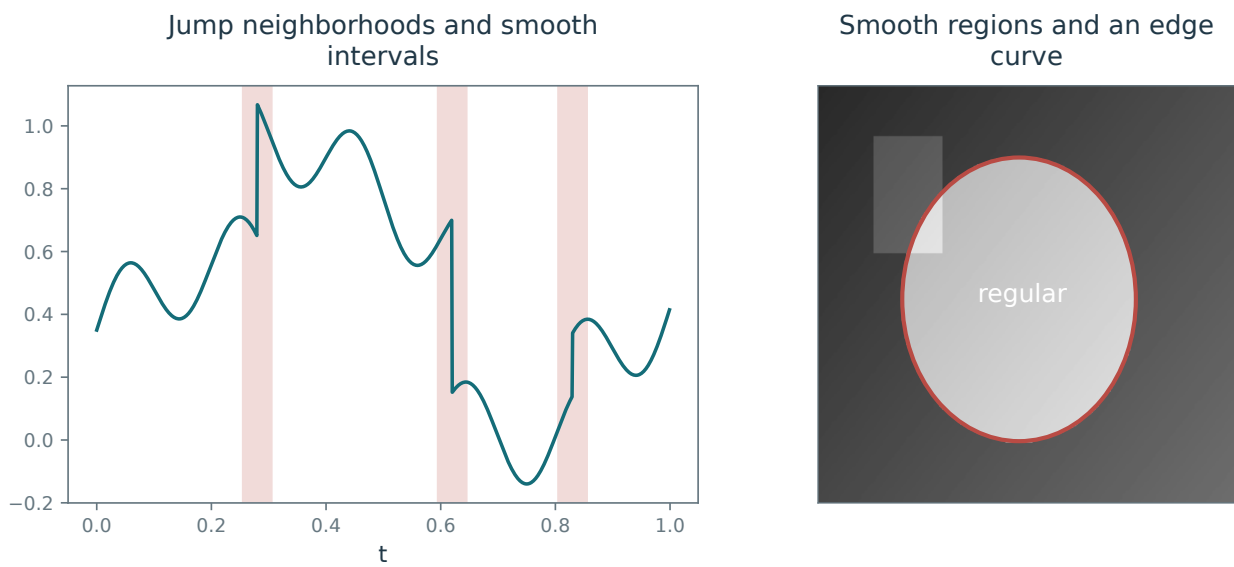


Figure 5.15. Singular and regular regions of a signal (left) and an image (right).

Proposition 5.4. Assume a compactly supported wavelet with $p \geq \alpha$ vanishing moments (all moments of total degree less than p in dimension d). If f admits a C^α extension to a neighborhood of the support, one has

$$|\langle f, \psi_{j,n} \rangle| \leq C_f 2^{j(\alpha+d/2)} \int |t|^\alpha |\psi(t)| dt. \quad (5.21)$$

Without the smoothness assumption, every bounded f satisfies

$$|\langle f, \psi_{j,n} \rangle| \leq \|f\|_\infty \|\psi\|_1 2^{j\frac{d}{2}}. \quad (5.22)$$

Proof. Under these assumptions, one can perform a Taylor expansion of f around the point $2^j n$

$$f(x) = P(x - 2^j n) + R(x - 2^j n) = P(2^j t) + R(2^j t)$$

where $\deg(P) < \alpha$ and

$$|R(x)| \leq C_f \|x\|^\alpha.$$

One then bounds the wavelet coefficient

$$\langle f, \psi_{j,n} \rangle = \frac{1}{2^{j\frac{d}{2}}} \int f(x) \psi\left(\frac{x - 2^j n}{2^j}\right) dx = 2^{j\frac{d}{2}} \int R(2^j t) \psi(t) dt$$

where we substituted $t = \frac{x - 2^j n}{2^j}$. This proves (5.21). The bound (5.22) follows directly by bounding the function in the inner product. $\square$

5.5.2 1-D Piecewise Smooth Approximation

For a piecewise smooth 1-D signal in the model (5.9), the large coefficients $\langle f, \psi_{j,n} \rangle$ concentrate near singularities. Denote the finite set of singular points by $\mathcal{S} \subset [0, 1]$.

Theorem 5.5. For $\alpha > 0$, let f belong to the piecewise C^α model (5.9). With compactly supported orthonormal wavelets having at least α vanishing moments (and a compatible boundary construction), the best M -term error satisfies

$$\|f - f_M\|^2 = O(M^{-2\alpha}). \quad (5.23)$$

Proof. The proof is split into several parts.

Step 1. Separating regular and singular coefficients. At scale 2^j , define the singular index set by the wavelets whose support crosses a singularity:

$$C_j = \{n ; \text{supp}(\psi_{j,n}) \cap \mathcal{S} \neq \emptyset\} \quad (5.24)$$

Compact support and dyadic spacing imply a bound on its cardinality that is uniform in j :

$$|C_j| \leq K|\mathcal{S}| = \text{constant}.$$

Using (5.21) for $d = 1$, the decay of regular coefficients is bounded as

$$\forall n \in C_j^c, \quad |\langle f, \psi_{j,n} \rangle| \leq C 2^{j(\alpha+1/2)}.$$

Using (5.22) for $d = 1$, the decay of singular coefficients is bounded as

$$\forall n \in C_j, \quad |\langle f, \psi_{j,n} \rangle| \leq C 2^{j/2}.$$

For a sufficiently small threshold T , define two cutoff scales, one for regular coefficients and one for singular coefficients, as functions of T :

$$2^{j_1} = (T/C)^{\frac{1}{\alpha+1/2}} \quad \text{and} \quad 2^{j_2} = (T/C)^2.$$

Round these cutoff scales to integers; this changes constants only. All subsequent scale sums are bounded above by the fixed coarsest scale $j_0 = 0$, and the finitely many scaling coefficients are retained. Figure 5.16 shows a schematic segmentation of the set of wavelet coefficients into regular and singular parts, together with the cut-off scales.

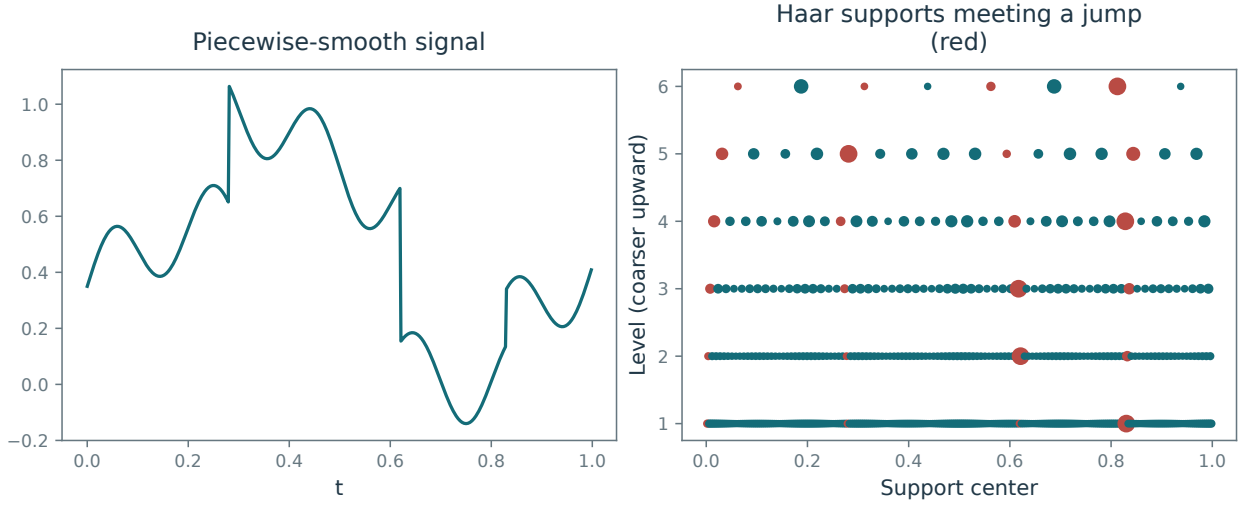


Figure 5.16. Segmentation of the wavelet coefficients into regular and singular parts.

Step 2. Bounding the error. These cut-off scales allow us to define a comparison approximation

$$\tilde{f}_M = P_{V_0}f + \sum_{j_2 \leq j \leq 0} \sum_{n \in C_j} \langle f, \psi_{j,n} \rangle \psi_{j,n} + \sum_{j_1 \leq j \leq 0} \sum_{n \in C_j^c} \langle f, \psi_{j,n} \rangle \psi_{j,n}. \quad (5.25)$$

The error of the comparison M -term approximation $\tilde{f}_M$ bounds the best M -term error from above:

$$\|f - f_M\|^2 \leq \|f - \tilde{f}_M\|^2 \leq \sum_{j < j_2, n \in C_j} |\langle f, \psi_{j,n} \rangle|^2 + \sum_{j < j_1, n \in C_j^c} |\langle f, \psi_{j,n} \rangle|^2 \quad (5.26)$$

$$\leq \sum_{j < j_2} (K|S|) \times C^2 2^j + \sum_{j < j_1} 2^{-j} \times C^2 2^{j(2\alpha+1)} \quad (5.27)$$

$$= O(2^{j_2} + 2^{2\alpha j_1}) = O(T^2 + T^{\frac{2\alpha}{\alpha+1/2}}) = O(T^{\frac{2\alpha}{\alpha+1/2}}). \quad (5.28)$$

Step 3. Counting the Coefficients. The number of coefficients needed to build the approximating signal $\tilde{f}_M$ is

$$M \leq 1 + \sum_{j_2 \leq j \leq 0} |C_j| + \sum_{j_1 \leq j \leq 0} |C_j^c| \leq 1 + \sum_{j_2 \leq j \leq 0} K|S| + \sum_{j_1 \leq j \leq 0} 2^{-j} \quad (5.29)$$

$$= O(|\log(T)| + T^{\frac{-1}{\alpha+1/2}}) = O(T^{\frac{-1}{\alpha+1/2}}). \quad (5.30)$$

Step 4. Putting everything together. Given a coefficient budget M , choose T to be a sufficiently large constant times $M^{-(\alpha+1/2)}$. The coefficient-count bound then uses at most M terms, and the error bound is $O(M^{-2\alpha})$. $\square$

For piecewise Lipschitz signals, the wavelet rate is faster than the jump-limited $O(M^{-1})$ Fourier bound in (5.19). It also matches the smooth-signal rate (5.16): finitely many singularities do not worsen the asymptotic exponent in one dimension. The rate (5.23) is asymptotically optimal for this class.

Figure 5.17 shows examples of wavelet approximation of singular signals.

5.5.3 2-D Piecewise Smooth Approximation

We now give the counterpart of Theorem 5.5 for 2-D functions.

Theorem 5.6. Let f belong to the piecewise smooth image model (5.10), with $\alpha \geq 1$ and finitely many rectifiable edge curves. For compactly supported orthonormal wavelets with at least α vanishing moments and compatible boundary treatment, the best M -term error satisfies

$$\|f - f_M\|^2 = O(M^{-1}). \quad (5.31)$$

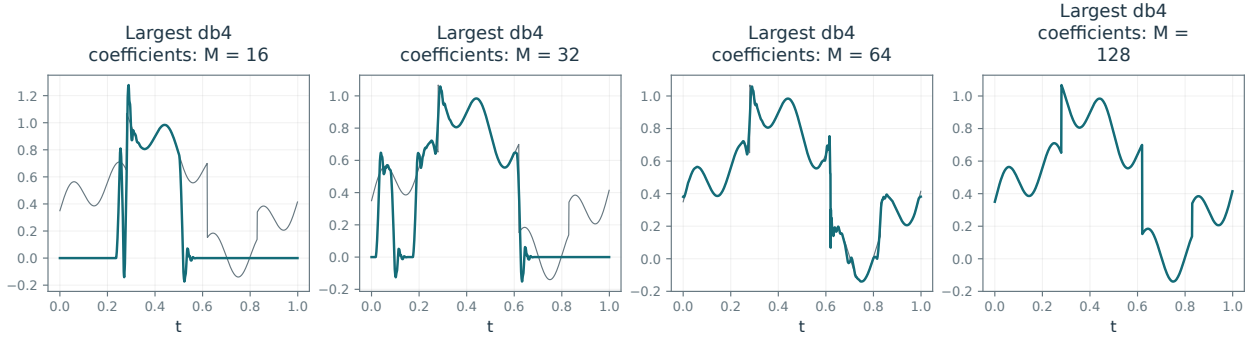


Figure 5.17. 1-D wavelet approximation.

Proof. For an image in the model (5.10), define the singular index set C_j as in (5.24). Unlike the 1-D case, its cardinality grows as the scale 2^j decreases:

$$|C_j^\omega| \leq 2^{-j} K |\mathcal{S}|,$$

where $|\mathcal{S}|$ is the total length of the singular curves $\mathcal{S}$, and $\omega \in \{V, H, D\}$ is the wavelet orientation. Using (5.21) for $d = 2$, the decay of regular coefficients is bounded as

$$\forall n \in (C_j^\omega)^c, \quad |\langle f, \psi_{j,n}^\omega \rangle| \leq C 2^{j(\alpha+1)}.$$

Using (5.22) for $d = 2$, the decay of singular coefficients is bounded as

$$\forall n \in C_j^\omega, \quad |\langle f, \psi_{j,n}^\omega \rangle| \leq C 2^j.$$

After fixing T , the cut-off scales are defined as

$$2^{j_1} = (T/C)^{\frac{1}{\alpha+1}} \quad \text{and} \quad 2^{j_2} = T/C.$$

We define, as in (5.25), a comparison approximation. Similarly to (5.26), we bound the approximation error as

$$\|f - f_M\|^2 \leq \|f - \tilde{f}_M\|^2 = O(2^{j_2} + 2^{2\alpha j_1}) = O(T + T^{\frac{2\alpha}{\alpha+1}}) = O(T)$$

and the number of coefficients as

$$M = O(T^{-1} + T^{-2/(\alpha+1)}) = O(T^{-1}).$$

Since $\alpha \geq 1$, choose T of order M^{-1} with a sufficiently large constant. This meets the coefficient budget and gives the asserted error. For $0 < \alpha < 1$, the same argument gives $O(M^{-\alpha})$ instead. $\square$

The wavelet rate improves on the $O(M^{-1/2})$ Fourier rate (5.20), but it still fails to exploit the full C^α regularity away from the edge curves.

This rate also holds for the broader bounded-variation model (5.11), for which wavelet approximation is asymptotically optimal.

Figure 5.18 shows wavelet approximations of a bounded variation image.

5.6 Approximation of Cartoon Images

The isotropic, compactly supported wavelets considered here occupy squares that do not follow the smooth contours of geometric images (5.13). These contours have more structure than the finite-length level sets allowed by the bounded-variation model (5.11).

5.6.1 Wavelet Approximation of Cartoon Images

The bound (5.31) gives a wavelet approximation rate of $O(M^{-1})$ for cartoon images (5.13). Even the simple indicator $f = 1_\Omega$, where Ω is a disk, can attain this slow rate: isotropic wavelets do not efficiently follow its curved boundary. The smoothed cartoon model (5.14) can behave similarly when the coefficient budget M is too small to resolve the blur.

Figure 5.19 shows that many large coefficients are located near edge curves, and retaining only a small number leads to a poor approximation with visually unpleasant artifacts.

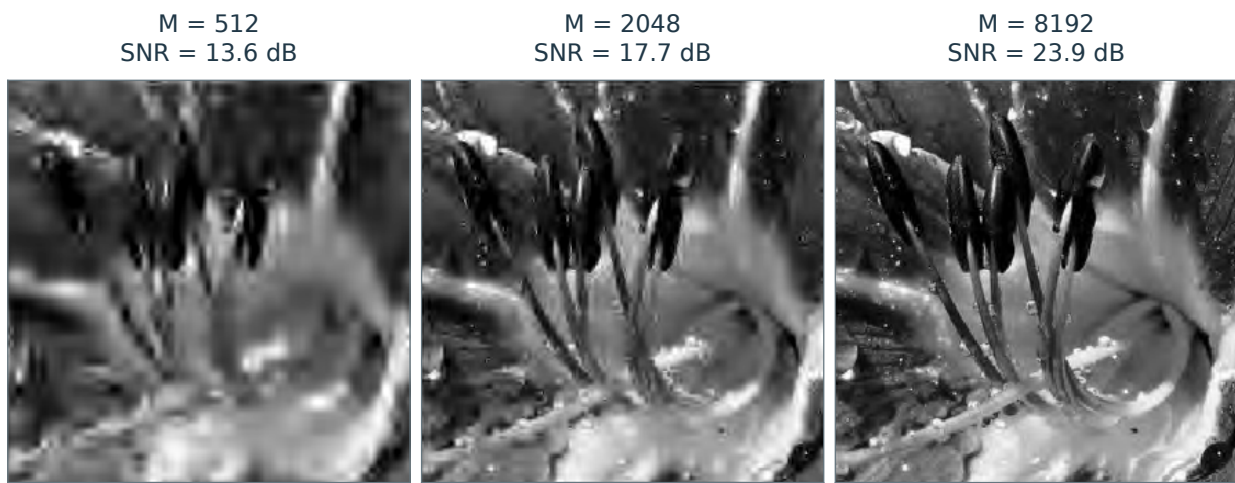


Figure 5.18. 2-D wavelet approximation.



Figure 5.19. Wavelet approximation of a cartoon image.

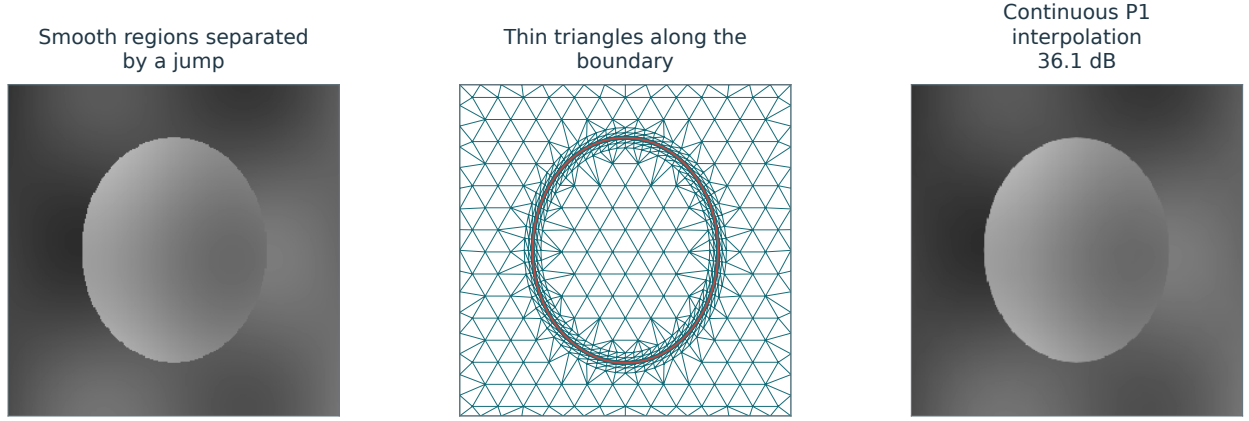


Figure 5.20. A cartoon with smooth variation on both sides of an elliptical jump, a triangulation elongated along that boundary, and the continuous P_1 interpolant.

5.6.2 Finite Element Approximation

An adaptive triangulation can improve the approximation by using elongated triangles aligned with the edges. Figure 5.20 shows an example of such a triangulation.

Select M points in the image domain $[0, 1]^2$ and connect them into a triangulation. Define $\tilde{f}_M$ by piecewise affine interpolation on the resulting triangles.

Figure 5.21 illustrates a construction for C^2 cartoon images (5.13) with $\alpha = 2$. In smooth regions, use $\approx M/2$ nearly equilateral triangles of width $\approx M^{-1/2}$. Near the edges, exploit the C^2 contour regularity by placing $\approx M/2$ elongated triangles of length $\approx M^{-1}$ along the contours and width $\approx M^{-2}$ across them. Such a triangulation yields the error bound

$$\|f - f_M\|^2 = O(M^{-2}), \quad (5.32)$$

which improves over the wavelet approximation error decay (5.31).

Implementing this construction requires estimating the edges. This is particularly difficult for blurred cartoon images with an unknown smoothing kernel h .

Practical methods use heuristics or greedy rules to select sample points and triangulate them from discrete or noisy data. Figure 5.22 uses a known synthetic contour to isolate the effect of mesh geometry. Both regions have smooth, nonaffine intensities. The displayed P_1 image is decoded from an explicit mesh stream whose bit count includes vertex coordinates, values, connectivity, and headers; this example does not assume an optimal mesh code.

5.6.3 Curvelet Approximation

A fixed family of oriented, anisotropic atoms offers an alternative to adaptive triangulations. Candès and Donoho introduced the curvelet frame for this purpose [1].

Curvelets. At fine scales $j \leq 0$, the geometry can be illustrated by applying parabolic scaling to a horizontally oriented mother function c ,

$$c_{2^j}(x_1, x_2) \approx 2^{-3j/4} c(2^{-j/2} x_1, 2^{-j} x_2),$$

followed by translation and rotation,

$$c_{2^j, u}^\theta(x_1, x_2) = c_{2^j}(R_{-\theta}(x - u))$$

where R_θ is the rotation of angle θ .

The atom $c_{2^j, u}^\theta$ is localized near u and oriented at angle θ . Its dimensions obey “width $\approx$ length²”, the same relation used for adaptive triangles near edges. This scaling matches the contour geometry of the cartoon model (5.13) with $\alpha = 2$.

Figure 5.23 illustrates this localization in space and frequency.

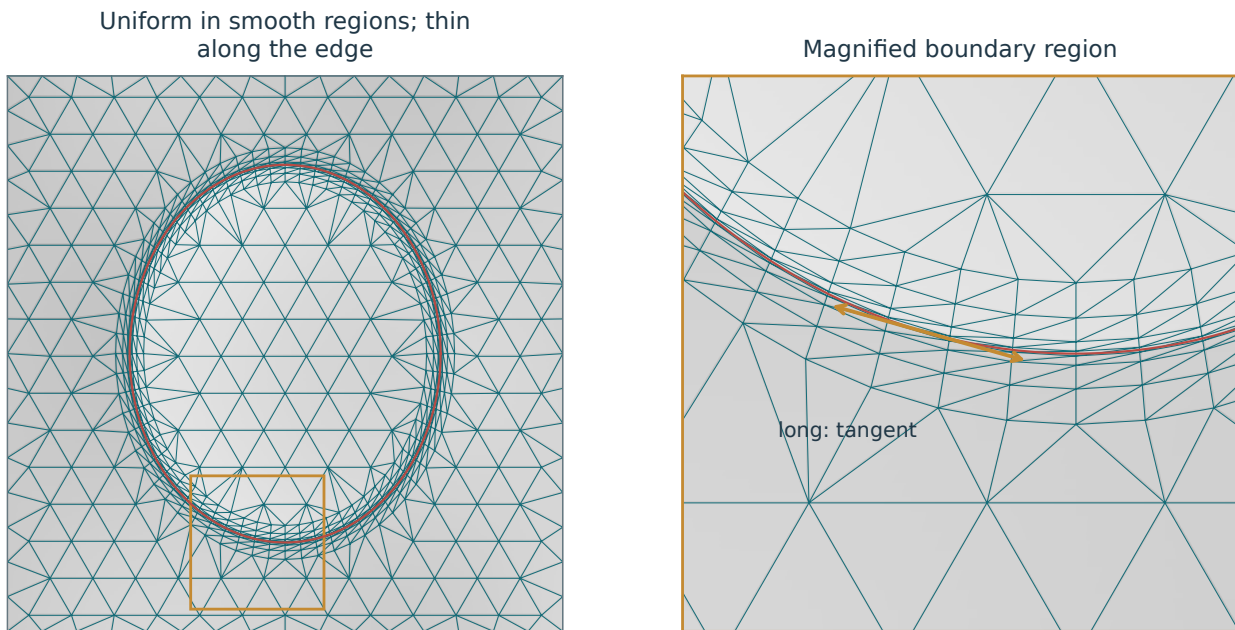


Figure 5.21. The same mesh over a lightened P_1 interpolant (left), with the boxed boundary region enlarged (right). Thin triangles follow the red singularity curve; the arrow indicates its tangent. Smooth regions use nearly equilateral elements.

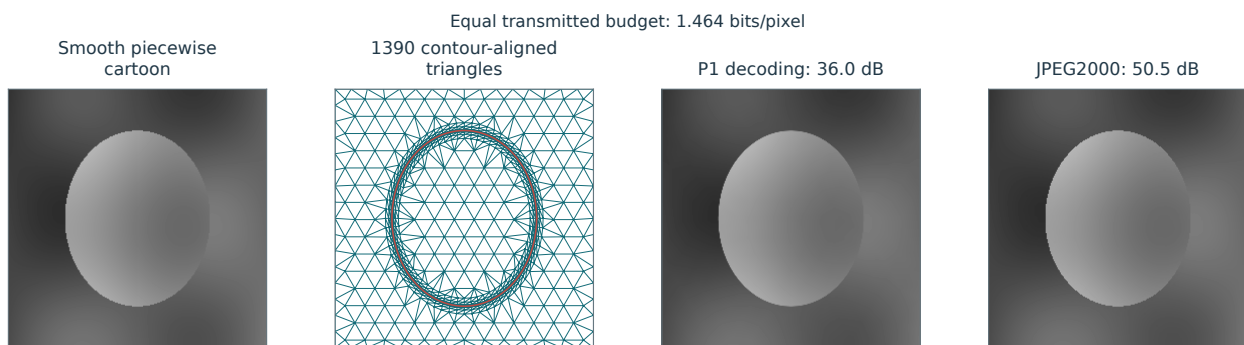


Figure 5.22. Compression of a piecewise-smooth cartoon using a contour-aligned mesh and JPEG-2000 at the same transmitted bit budget. From left to right: original, mesh, decoded P_1 image, and JPEG-2000 image. Mesh costs include coordinates, nodal values, connectivity, and headers.

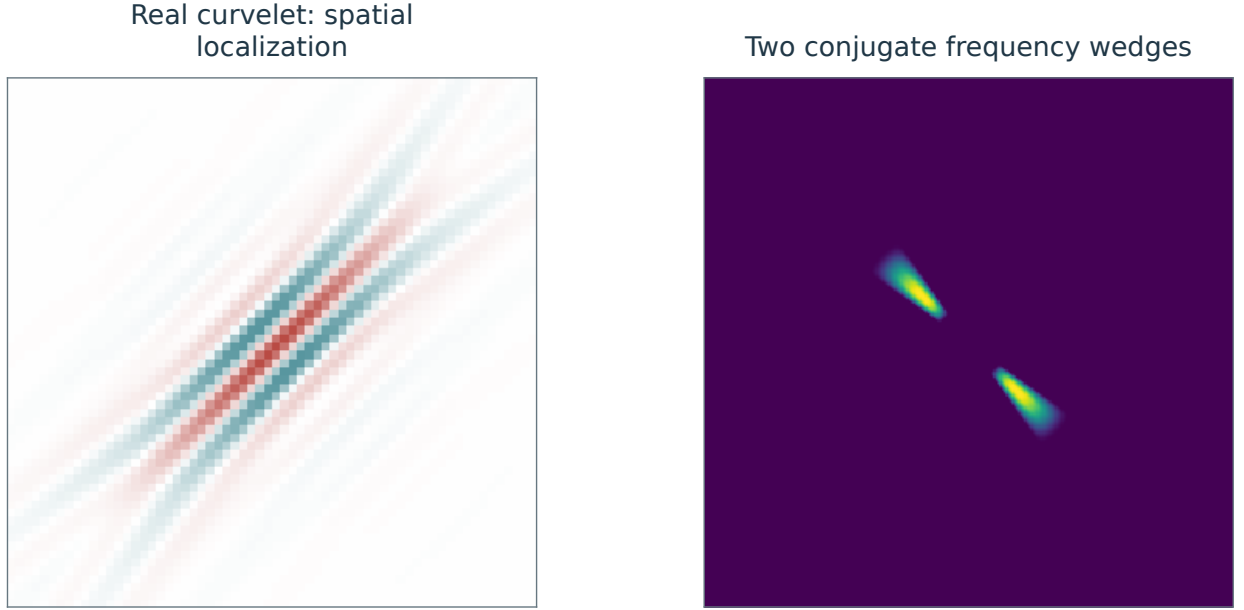


Figure 5.23. A curvelet c_m in space (left) and its localization in the Fourier domain (right).

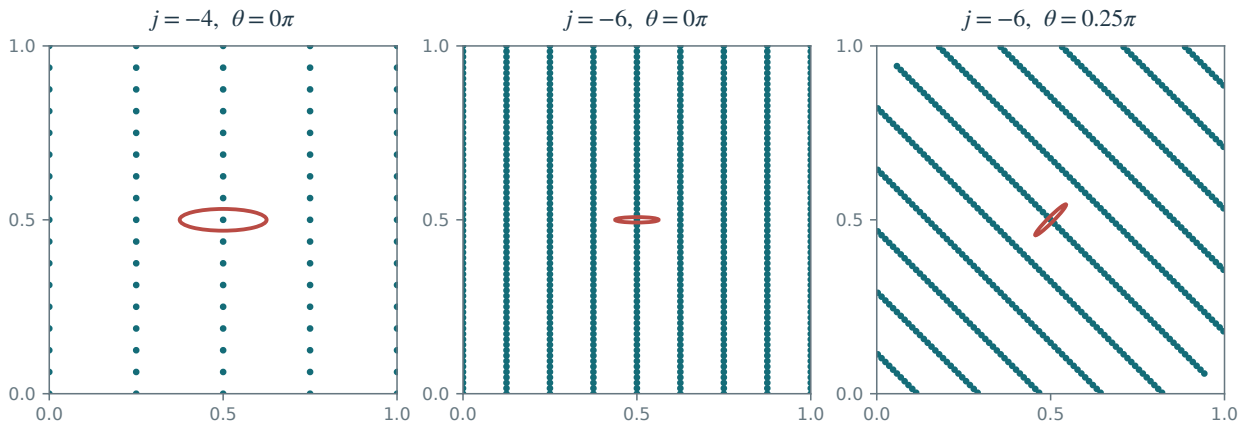


Figure 5.24. Sampling pattern for the curvelet positions.

Parameter discretization. To build an image representation, one needs to sample the position u and orientation θ . The angular sampling interval depends on the scale:

$$\forall 0 \leq k < 2^{-\lceil j/2 \rceil + 2}, \quad \theta_k^{(j)} = k\pi 2^{\lceil j/2 \rceil - 1}$$

and the spatial grid depends on both scale and orientation:

$$\forall m = (m_1, m_2) \in \mathbb{Z}^2, \quad u_m^{(j, \theta)} = R_\theta(2^{j/2}m_1, 2^j m_2).$$

Figure 5.24 shows this sampling pattern.

Curvelet tight frame. The scaling and sampling formulas above describe the geometry of curvelets. An exact tight frame additionally requires carefully chosen frequency windows and a low-frequency family; an arbitrary mother function c does not suffice. With this construction, write the complete family as $(C_\lambda)_{\lambda \in \Lambda}$, where Λ includes the fine-scale indices (j, m, k) and the low-frequency indices. Then

$$\|f\|^2 = \sum_{\lambda \in \Lambda} |\langle f, C_\lambda \rangle|^2, \quad f = \sum_{\lambda \in \Lambda} \langle f, C_\lambda \rangle C_\lambda.$$

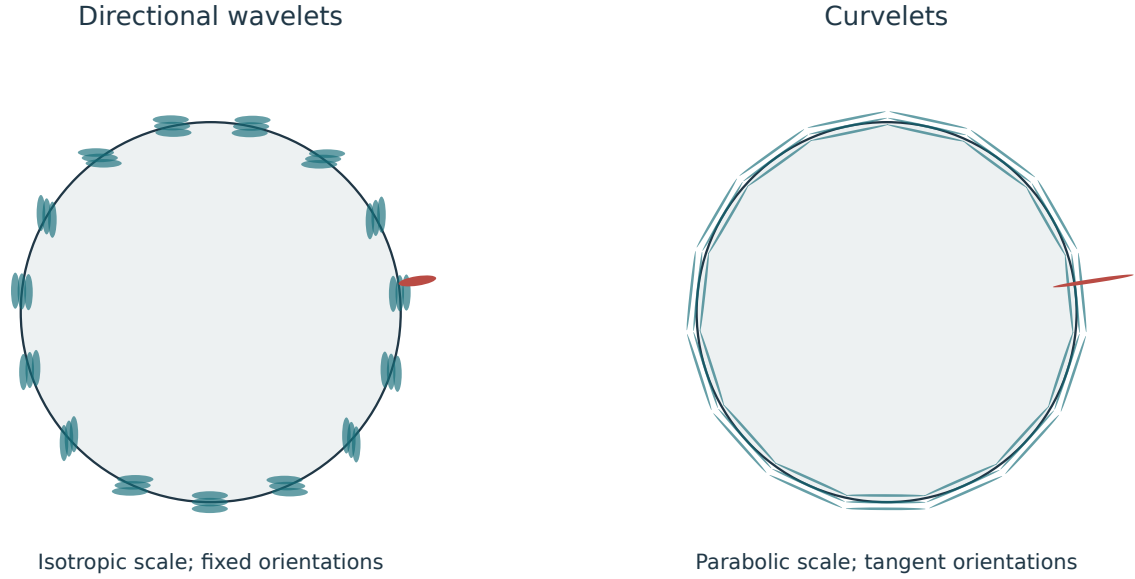


Figure 5.25. Wavelet (left) and curvelet (right) approximation of a cartoon image.

The reconstruction converges in L^2 . These are the Parseval-frame identities; the family is redundant and its atoms are not mutually orthogonal.

One discrete construction for an image with N pixels uses $\approx 5N$ atoms [2].

Curvelet approximation. A nonlinear M -term approximation in curvelets is defined as

$$f_M = \sum_{\lambda \in \Lambda: |\langle f, C_\lambda \rangle| > T} \langle f, C_\lambda \rangle C_\lambda$$

The complete index set Λ includes the low-frequency family, and the threshold T determines the number M of retained coefficients. As before, ties require a selection rule if an exact coefficient count is prescribed. Because the frame is not orthogonal, f_M need not be the best M -term curvelet approximation.

Far from an edge, at locations $u_m^{(j,\theta)}$, cancellation makes the coefficients $\langle f, C_{j,m,k} \rangle$ small. When $u_m^{(j,\theta)}$ lies near an edge with tangent direction $\tilde{\theta}$, the coefficient $\langle f, C_{j,m,k} \rangle$ rapidly decreases as the orientation mismatch $|\theta - \tilde{\theta}|$ grows. Figure 5.25 compares this adaptation to edges with the square supports of directional wavelets.

Combining spatial cancellation, directional selectivity, and scale-dependent sampling gives the approximation bound

$$\|f - f_M\|^2 = O(\log^3(M)M^{-2})$$

for images in the cartoon model (5.13) for $\alpha = 2$. This nearly matches the adaptive-triangulation rate (5.32), while allowing f_M to be computed in $O(N \log(N))$ operations for an image of N pixels.

Redundancy complicates the use of curvelets for compression, but their approximation properties are useful for denoising geometric images and textures. Figure 5.26 compares curvelet and wavelet denoising using the thresholding method of Section 7.3.1.

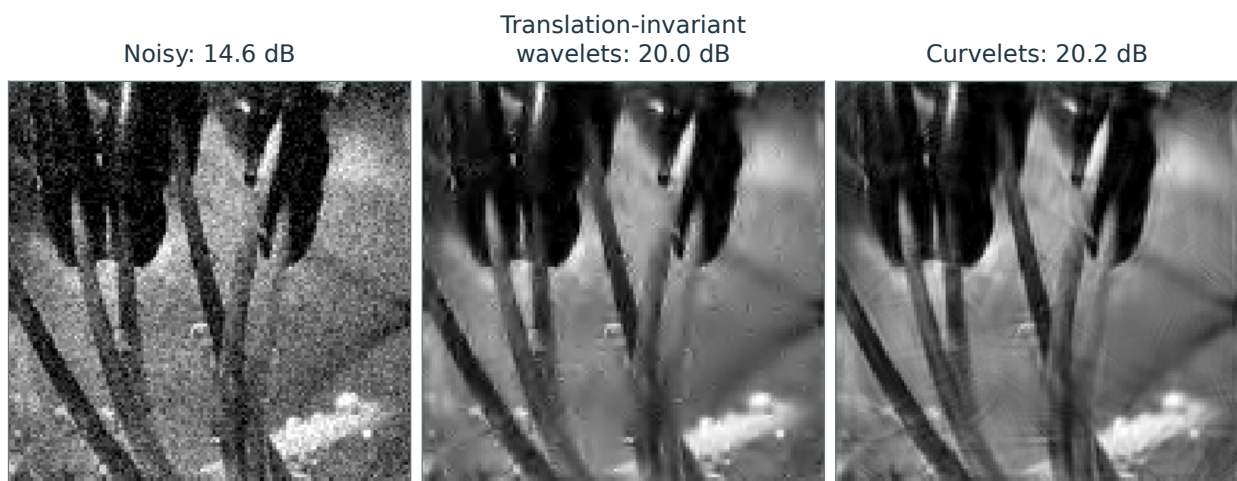


Figure 5.26. Comparison of translation-invariant wavelet denoising and curvelet denoising.

References

- [1] E. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Commun. on Pure and Appl. Math.*, 57(2):219–266, 2004.
- [2] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Modeling and Simulation*, 5:861–899, 2005.
- [3] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. on Pure and Appl. Math.*, 42:577–685, 1989.
- [4] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.