IX

Approximations in Bases

It is time to wonder why are we constructing so many different orthonormal bases. In signal processing, orthogonal bases are of interest because they can provide sparse representations of certain types of signals with few vectors. Compression and denoising are applications studied in Chapters 10 and 11.

Approximation theory studies the error produced by different approximation schemes. Classical sampling theorems are linear approximations that project the analog signal over low-frequency vectors chosen a priori in a basis. The discrete signal representation may be further reduced with a linear projection over the first few vectors of an orthonormal basis. However, better non-linear approximations are obtained by choosing the approximation vectors depending upon the signal. In a wavelet basis, these non-linear approximations adjust locally the approximation resolution to the signal regularity.

Approximation errors depend upon the signal regularity. For uniformly regular signals, linear and non-linear approximations perform similarly, wether in a wavelet or in a Fourier basis. When the signal regularity is not uniform, non-linear approximations in a wavelet basis can considerably reduce the error of linear approximations. This is the case for piecewise regular signals or bounded variation signals and images. Geometric approximations of piecewise regular images with regular edge curves are studied with adaptive triangulations and curvelets.

9.1 Linear Approximations

Analog signals are discretized in Section 3.1.3 with inner products in a basis. The next sections compute the resulting linear approximation error in wavelet and Fourier bases, which depends upon the uniform signal regularity. For signals modeled as realizations of a random vector, Section 9.1.4 proves that the optimal basis is the Karhunen-Loève basis (principal components), which diagonalizes the covariance matrix.

9.1.1 Sampling and Approximation Error

Approximation errors of linear sampling processes are related to the error of linear approximations in an orthogonal basis. These errors are computed from the decay of signal coefficients in this basis.

An analog signal $f(t)$ is discretized with a low-pass filter $\bar{\phi}_s(t)$ and a uniform sampling interval s:

$$f \star \bar{\phi}_s(ns) = \int_{-\infty}^{+\infty} f(u)\, \bar{\phi}_s(ns-u)\, du = \langle f(t), \phi_s(t-ns) \rangle, \tag{9.1}$$

with $\bar{\phi}_s(t) = \phi_s(-t)$. Let us consider an analog signal of compact support, normalized to $[0,1]$. At a resolution N corresponding to $s = N^{-1}$, the discretization is performed over N functions $\{\phi_n(t) = \phi_s(t-ns)\}_{0 \leqslant n < N}$ which are modified at the boundaries to maintain their support in $[0,1]$.

They define a Riesz basis of an approximation space $\mathbf{U}_N \subset \mathbf{L^2}[0,1]$. The best linear approximation of f in $\mathbf{U}_N$ is the orthogonal projection f_N of f in $\mathbf{U}_N$, recovered with the biorthogonal basis $\{\tilde{\phi}_n(t)\}_{1\leqslant n\leqslant N}$:

$$f_N(t) = \sum_{n=0}^{N-1} \langle f, \phi_n\rangle\, \tilde{\phi}_n(t). \tag{9.2}$$

To compute the approximation error $\|f-f_N\|$, we introduce an orthonormal basis $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$ of $\mathbf{L^2}[0,1]$, whose first N vectors $\{g_m\}_{0\leqslant m<N}$ define an orthogonal basis of the same approximation space $\mathbf{U}_N$. Fourier and wavelet bases provide such bases for many classical approximation spaces. The orthogonal projection f_N of f in $\mathbf{U}_N$ can be decomposed on the first N vectors of this basis:

$$f_N = \sum_{m=0}^{N-1} \langle f, g_m\rangle\, g_m.$$

Since $\mathcal{B}$ is an orthonormal basis of $\mathbf{L^2}[0,1]$, $f = \sum_{m=0}^{+\infty} \langle f, g_m\rangle\, g_m$, so

$$f - f_N = \sum_{m=N}^{+\infty} \langle f, g_m\rangle\, g_m,$$

and the resulting approximation error is

$$\varepsilon_l(N,f) = \|f - f_N\|^2 = \sum_{m=N}^{+\infty} |\langle f, g_m\rangle|^2. \tag{9.3}$$

The fact that $\|f\|^2 = \sum_{m=0}^{+\infty} |\langle f, g_m\rangle|^2 < +\infty$ implies that the error decays to zero:

$$\lim_{N\to+\infty} \varepsilon_l(N,f) = 0.$$

However, the decay rate of $\varepsilon_l(N,f)$ as N increases depends on the decay of $|\langle f, g_m\rangle|$ as m increases. The following theorem gives equivalent conditions on the decay of $\varepsilon_l(N,f)$ and $|\langle f, g_m\rangle|$.

Theorem 9.1. *For any $s > 1/2$, there exists $A, B > 0$ such that if $\sum_{m=0}^{+\infty} |m|^{2s}\, |\langle f, g_m\rangle|^2 < +\infty$ then*

$$A \sum_{m=0}^{+\infty} m^{2s}\, |\langle f, g_m\rangle|^2 \leqslant \sum_{N=0}^{+\infty} N^{2s-1}\, \varepsilon_l(N,f) \leqslant B \sum_{m=0}^{+\infty} m^{2s}\, |\langle f, g_m\rangle|^2 \tag{9.4}$$

and hence $\varepsilon_l(N,f) = o(N^{-2s})$.

Proof. By inserting (9.3), we compute

$$\sum_{N=0}^{+\infty} N^{2s-1}\, \varepsilon_l(N,f) = \sum_{N=0}^{+\infty}\sum_{m=N}^{+\infty} N^{2s-1}\, |\langle f, g_m\rangle|^2 = \sum_{m=0}^{+\infty} |\langle f, g_m\rangle|^2 \sum_{N=0}^{m} N^{2s-1}.$$

For any $s > 1/2$

$$\int_0^m x^{2s-1}\, dx \leqslant \sum_{N=0}^{m} N^{2s-1} \leqslant \int_1^{m+1} x^{2s-1}\, dx$$

which implies that $\sum_{N=0}^{m} N^{2s-1} \sim m^{2s}$ and hence proves (9.4).

To verify that $\varepsilon_l(N,f) = o(N^{-2s})$, observe that $\varepsilon_l(m,f) \geqslant \varepsilon_l(N,f)$ for $m \leqslant N$, so

$$\varepsilon_l(N,f) \sum_{m=N/2}^{N-1} m^{2s-1} \leqslant \sum_{m=N/2}^{N-1} m^{2s-1}\, \varepsilon_l(m,f) \leqslant \sum_{m=N/2}^{+\infty} m^{2s-1}\, \varepsilon_l(m,f). \tag{9.5}$$

Since $\sum_{m=1}^{+\infty} m^{2s-1}\varepsilon_l(m,f) < +\infty$ it follows that

$$\lim_{N\to+\infty} \sum_{m=N/2}^{+\infty} m^{2s-1}\, \varepsilon_l(m,f) = 0.$$

Moreover, there exists $C > 0$ such that $\sum_{m=N/2}^{N-1} m^{2s-1} \geqslant C\, N^{2s}$, so (9.5) implies that $\lim_{N\to+\infty} \varepsilon_l(N,f)\, N^{2s} = 0$. ■ ■

This theorem proves that the linear approximation error of f in the basis $\mathcal{B}$ decays faster than N^{-2s} if f belongs to the space

$$\mathbf{W}_{\mathcal{B},s} = \left\{ f \in \mathbf{H} \; : \; \sum_{m=0}^{+\infty} m^{2s} \, |\langle f, g_m \rangle|^2 < +\infty \right\} .$$

One can also prove that this linear approximation is asymptotically optimal over this space [19]. Indeed, there exists no linear or non-linear approximation scheme whose error decays at least like $N^{-\alpha}$, with $\alpha > 2s$, for all $f \in \mathbf{W}_{\mathcal{B},s}$.

The next sections prove that if $\mathcal{B}$ is a Fourier or wavelet basis, then $\mathbf{W}_{\mathcal{B},s}$ is a Sobolev space, and hence that linear approximations of Sobolev functions are optimal in Fourier and wavelet bases. However, we shall also see that for more complex functions, the linear approximation of f from the first N vectors of $\mathcal{B}$ is not always precise because these vectors are not necessarily the best ones to approximate f. Non-linear approximations calculated with vectors chosen adaptively depending upon f are studied in Section 9.2.

9.1.2 Linear Fourier Approximations

The Shannon-Whittaker sampling theorem performs a perfect low-pass filter that keeps the signal low-frequencies. It is thus equivalent to a linear approximation over the lower frequencies of a Fourier basis. Linear Fourier approximation are asymptotically optimal for uniformly regular signals. The approximation error is related to the Sobolev differentiability. It is also calculated for non-uniformly regular signals, such as discontinuous signals having a bounded total variation.

Theorem 3.6 proves (modulo a change of variable) that $\{\mathrm{e}^{i2\pi mt}\}_{m\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{L^2}[0,1]$. We can thus decompose $f \in \mathbf{L^2}[0,1]$ in the Fourier series

$$f(t) = \sum_{m=-\infty}^{+\infty} \langle f(u), \mathrm{e}^{i2\pi mu} \rangle \, \mathrm{e}^{i2\pi mt} \tag{9.6}$$

with

$$\langle f(u), \mathrm{e}^{i2\pi mu} \rangle = \int_0^1 f(u) \, \mathrm{e}^{-i2\pi mu} \, du.$$

The decomposition (9.6) defines a periodic extension of f for all $t \in \mathbb{R}$. The decay of the Fourier coefficients $|\langle f(u), \mathrm{e}^{i2\pi mu} \rangle|$ as m increases depends on the regularity of this periodic extension.

The linear approximation of $f \in \mathbf{L^2}[0,1]$ by the N sinusoïds of lower frequencies is obtained by a linear filtering that sets to zero all higher frequencies:

$$f_N(t) = \sum_{|m| \leqslant N/2} \langle f(u), \mathrm{e}^{i2\pi mu} \rangle \, \mathrm{e}^{i2\pi mt}.$$

It projects f in the space $\mathbf{U}_N$ of functions having Fourier coefficients that are zero above the frequency $N\pi$.

Error Decay versus Sobolev Differentiability The decay of the linear Fourier approximation error depends upon the Sobolev differentiability. The regularity of f can be measured by the number of times it is differentiable. Sobolev differentiability extends derivatives to non-integers, with a Fourier decay condition. We first consider functions $f(t)$ defined for all $t \in \mathbb{R}$, to avoid boundary issues.

Recall that the Fourier transform of the derivative $f'(t)$ is $i\omega\hat{f}(\omega)$. The Plancherel formula proves that $f' \in \mathbf{L^2}(\mathbb{R})$ if

$$\int_{-\infty}^{+\infty} |\omega|^2 \, |\hat{f}(\omega)|^2 \, d\omega = 2\pi \int_{-\infty}^{+\infty} |f'(t)|^2 \, dt < +\infty.$$

This suggests replacing the usual pointwise definition of the derivative by a definition based on the Fourier transform. We say that $f \in \mathbf{L^2}(\mathbb{R})$ is differentiable in the *sense of Sobolev* if

$$\int_{-\infty}^{+\infty} |\omega|^2 \, |\hat{f}(\omega)|^2 \, d\omega < +\infty. \tag{9.7}$$

This integral imposes that $|\hat{f}(\omega)|$ has a sufficiently fast decay when the frequency ω goes to $+\infty$. As in Section 2.3.1, the regularity of f is measured from the asymptotic decay of its Fourier transform.

This definition is generalized for any $s > 0$. The space $\mathbf{W}^s(\mathbb{R})$ of s times differentiable Sobolev functions is the space of functions $f \in \mathbf{L^2}(\mathbb{R})$ whose Fourier transforms satisfy [65]

$$\int_{-\infty}^{+\infty} |\omega|^{2s}\, |\hat{f}(\omega)|^2\, d\omega < +\infty. \tag{9.8}$$

If $s > n + 1/2$, then one can verify (Exercise 9.7) that f is n times continuously differentiable.

Let $\mathbf{W}^s[0,1]$ be the space of functions in $\mathbf{L^2}[0,1]$ that can be extended outside $[0,1]$ into a function $f \in \mathbf{W}^s(\mathbb{R})$. To avoid border problems at $t = 0$ or at $t = 1$, let us consider functions f whose support is strictly included in $(0,1)$. A simple regular extension on $\mathbb{R}$ is obtained by setting its value to 0 outside $[0,1]$, and $f \in \mathbf{W}^s[0,1]$ if this extension is in $\mathbf{W}^s(\mathbb{R})$. In this case, one can prove (not trivial) that the Sobolev integral condition (9.8) reduces to a discrete sum, meaning that $f \in \mathbf{W}^s[0,1]$ if and only if

$$\sum_{m=-\infty}^{+\infty} |m|^{2s}\, |\langle f(u), \mathrm{e}^{i2\pi m u}\rangle|^2 < +\infty. \tag{9.9}$$

For such differentiable functions in the sense of Sobolev, the following theorem computes the approximation error

$$\varepsilon_l(N,f) = \|f - f_N\|^2 = \int_0^1 |f(t) - f_N(t)|^2\, dt = \sum_{|m|>N/2} |\langle f(u), \mathrm{e}^{i2\pi m u}\rangle|^2 \, . \tag{9.10}$$

Theorem 9.2. *Let $f \in \mathbf{L^2}[0,1]$ be a function whose support is included in $(0,1)$. Then $f \in \mathbf{W}^s[0,1]$ if and only if*

$$\sum_{N=1}^{+\infty} N^{2s}\, \frac{\varepsilon_l(N,f)}{N} < +\infty, \tag{9.11}$$

which implies $\varepsilon_l(N,f) = o(N^{-2s})$.

The proof relies on the fact that functions in $\mathbf{W}^s[0,1]$ with a support in $(0,1)$ are characterized by (9.9). This theorem is therefore a consequence of Theorem 9.1. The linear Fourier approximation thus decays quickly if and only if f has a large regularity exponent s in the sense of Sobolev.

Discontinuities and Bounded Variation If f is discontinuous, then $f \notin \mathbf{W}^s[0,1]$ for any $s > 1/2$. Theorem 9.2 thus proves that $\varepsilon_l(N,f)$ can decay like $N^{-\alpha}$ only if $\alpha \leqslant 1$. For bounded variation functions, which are introduced in Section 2.3.3, the following theorem proves that $\varepsilon_l(N,f) = O(N^{-1})$. A function has a bounded variation if

$$\|f\|_V = \int_0^1 |f'(t)|\, dt < +\infty \, .$$

The derivative must be taken in the sense of distributions because f may be discontinuous. If $f = \mathbf{1}_{[0,1/2]}$ then $\|f\|_V = 2$. Recall that $a[N] \sim b[N]$ if $a[N] = O(b[N])$ and $b[N] = O(a[N])$.

Theorem 9.3.
- *If $\|f\|_V < +\infty$ then $\varepsilon_l(N,f) = O(\|f\|_V^2\, N^{-1})$.*
- *If $f = C\,\mathbf{1}_{[0,1/2]}$ then $\varepsilon_l(N,f) \sim \|f\|_V^2\, N^{-1}$.*

Proof. If $\|f\|_V < +\infty$ then

$$\begin{aligned} |\langle f(u), \exp(i2m\pi u)\rangle| &= \left| \int_0^1 f(u)\, \exp(-i2m\pi u)\, du \right| \\ &= \left| \int_0^1 f'(u)\, \frac{\exp(-i2m\pi u)}{-i2m\pi}\, dt \right| \leqslant \frac{\|f\|_V}{2|m|\pi}. \end{aligned}$$

Hence

$$\varepsilon_l(N,f) = \sum_{|m|>N/2} |\langle f(u), \exp(i2m\pi u)\rangle|^2 \leqslant \frac{\|f\|_V^2}{4\pi^2} \sum_{|m|>N/2} \frac{1}{m^2} = O(\|f\|_V^2\, N^{-1}).$$

If $f = C\,\mathbf{1}_{[0,1/2]}$ then $\|f\|_V = 2C$ and

$$|\langle f(u), \exp(i2m\pi u)\rangle| = \begin{cases} 0 & \text{if } m \neq 0 \text{ is even} \\ C/(\pi\,|m|) & \text{if } m \text{ is odd,} \end{cases}$$

so $\varepsilon_l(N,f) \sim C^2\, N^{-1}$. ■ ■

This theorem shows that when f is discontinuous with bounded variations, then $\varepsilon_l(N,f)$ decays typically like N^{-1}. Figure 9.1(b) shows a bounded variation signal approximated by Fourier coefficients of lower frequencies. The approximation error is concentrated in the neighborhood of discontinuities where the removal of high frequencies creates Gibbs oscillations (see Section 2.3.1).

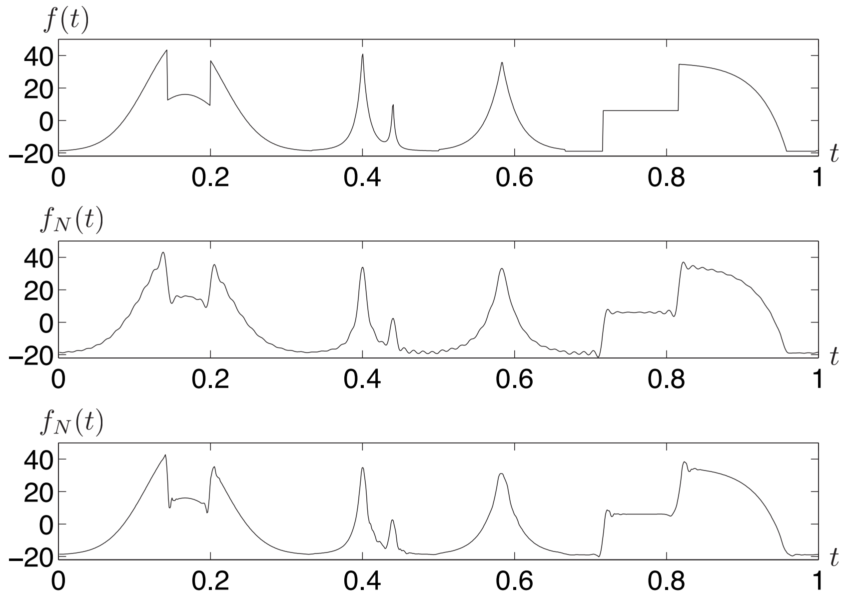


Figure 9.1: Top: Original signal f. Middle: Signal f_N approximated from $N = 128$ lower frequency Fourier coefficients, with $\|f - f_N\|/\|f\| = 8.63\,10^{-2}$. Bottom: Signal f_N approximated from larger scale Daubechies 4 wavelet coefficients, with $N = 128$ and $\|f - f_N\|/\|f\| = 8.58\,10^{-2}$.

Localized Approximations To localize Fourier series approximations over intervals, we multiply f by smooth windows that cover each of these intervals. The Balian-Low Theorem 5.20 proves that one cannot build local Fourier bases with smooth windows of compact support. However, Section 8.4.2 constructs orthonormal bases by replacing complex exponentials by cosine functions. For appropriate windows g_p of compact support $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$, Corollary 8.1 constructs an orthonormal basis of $\mathbf{L^2}(\mathbb{R})$:

$$\left\{ g_{p,k}(t) = g_p(t)\,\sqrt{\frac{2}{l_p}}\,\cos\left[\pi\Big(k+\frac{1}{2}\Big)\frac{t-a_p}{l_p}\right]\right\}_{k\in\mathbb{N},p\in\mathbb{Z}}.$$

Writing f in this local cosine basis is equivalent to segmenting it into several windowed components $f_p(t) = f(t)\,g_p(t)$, which are decomposed in a cosine IV basis. If g_p is $\mathbf{C}^\infty$, the regularity of $g_p(t)\,f(t)$ is the same as the regularity of f over $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$. Section 8.3.2 relates cosine

IV coefficients to Fourier series coefficients. It follows from Theorem 9.2 that if $f_p \in \mathbf{W}^s(\mathbb{R})$, then the approximation

$$f_{p,N} = \sum_{k=0}^{N-1} \langle f, g_{p,k}\rangle\, g_{p,k}$$

yields an error

$$\varepsilon_l(N, f_p) = \|f_p - f_{p,N}\|^2 = o(N^{-2s}).$$

The approximation error in a local cosine basis thus depends on the local regularity of f over each window support.

9.1.3 Multiresolution Approximation Errors with Wavelets

Wavelets are constructed as bases of orthogonal complements of multiresolution approximation spaces. The projection error on multiresolution approximation spaces thus depends upon the decay of wavelet coefficients. Linear approximations with wavelets behave essentially like Fourier approximations, with a better treatment of boundaries. They are thus also asymptotically optimal for uniformly regular signals. The linear error decay is computed for Sobolev differentiable functions and for uniformly Lipschitz α functions.

Uniform Approximation Grid Section 7.5 constructs multiresolution approximation spaces $\mathbf{U}_N = \mathbf{V}_L$ of $\mathbf{L^2}[0,1]$, with their orthonormal basis of $N = 2^{-L}$ scaling functions $\{\phi_{L,n}(t)\}_{0\leqslant n<2^{-l}}$. These scaling functions $\phi_{L,n}(t) = \phi_L(t - 2^L n)$ with $\phi_L(t) = 2^{-L/2}\phi(2^{-L}t)$ are finite elements translated over a uniform grid, modified near 0 and 1 so that their support remains in $[0,1]$. The resulting projection of f in such a space is:

$$f_N = P_{\mathbf{V}_L} f = \sum_{n=0}^{2^{-L}-1} \langle f, \phi_{L,n}\rangle\, \phi_{L,n}. \tag{9.12}$$

and

$$\langle f, \phi_{L,n}\rangle = \int f(t)\, \phi_L(t - 2^L n)\, dt = f \star \bar{\phi}_L(ns) \quad \text{with } \bar{\phi}_L(t) = \phi_L(-t)\ .$$

A different orthogonal basis of $\mathbf{V}_L$ is obtained from wavelets at scales $2^j > 2^L$ and scaling functions at a large scale 2^J:

$$\Big[\{\phi_{J,n}\}_{0\leqslant n<2^{-J}}\ ,\ \{\psi_{j,n}\}_{l<j\leqslant J\,,\,0\leqslant n<2^{-j}}\Big]. \tag{9.13}$$

The approximation (9.12) can thus also be written as a wavelet approximation

$$f_N = P_{\mathbf{V}_L} f = \sum_{j=L+1}^{J} \sum_{n=0}^{2^{-j}-1} \langle f, \psi_{j,n}\rangle\, \psi_{j,n} + \sum_{n=0}^{2^{-J}-1} \langle f, \phi_{J,n}\rangle\, \phi_{J,n}. \tag{9.14}$$

Since wavelets define an orthonormal basis of $\mathbf{L^2}[0,1]$

$$\Big[\{\phi_{J,n}\}_{0\leqslant n<2^{-J}}\ ,\ \{\psi_{j,n}\}_{-\infty<j\leqslant J\,,\,0\leqslant n<2^{-j}}\Big]\ , \tag{9.15}$$

the approximation error is the energy of wavelet coefficients at scales smaller than 2^L:

$$\varepsilon_l(N, f) = \|f - f_N\|^2 = \sum_{j=-\infty}^{L} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^2. \tag{9.16}$$

In the following we suppose that wavelets $\psi_{j,n}$ are $\mathbf{C}^q$ (q times differentiable) and have q vanishing moments. The treatment of boundaries is the key difference with Fourier approximations. Fourier approximations consider that the signal is periodic and if $f(0) \neq f(1)$, then the approximation of f behaves as if f was discontinuous. The periodic orthogonal wavelet bases in Section

7.5.1 do essentially the same. To improve this result, wavelets at the boundaries must keep their q vanishing moments, which requires to modify them near the boundaries so that their supports remain in $[0, 1]$. Section 7.5.3 constructs such wavelet bases. They take advantage of the regularity of f over $[0, 1]$, with no condition on $f(0)$ and $f(1)$. For mathematical analysis, we only use these wavelets, without explicitly writing their shape modifications at the boundaries, to simplify notations. In numerical experiments, the folding boundary solution of Section 7.5.2 is more often used, because it has a simpler algorithmic implementation. Folded wavelets have 1 vanishing moment at the boundary, which is often sufficient in applications.

Approximation Error versus Sobolev Regularity Like a Fourier basis, a wavelet basis provides an efficient approximation of uniformly regular signals. The decay of wavelet linear approximation errors is first related to the differentiability in the sense of Sobolev. Let $\mathbf{W}^s[0,1]$ be the Sobolev space of functions that are restrictions over $[0, 1]$ of s times differentiable Sobolev function $\mathbf{W}^s(\mathbb{R})$ defined over $\mathbb{R}$. If ψ has q vanishing moments then (6.11) proves that the wavelet transform is a multiscale differential operator of order at least q. To test the differentiability of f up to order s, we thus need $q > s$. The following theorem gives a necessary and sufficient condition on wavelet coefficients so that $f \in \mathbf{W}^s[0,1]$.

Theorem 9.4. *Let $0 < s < q$ be a Sobolev exponent. A function $f \in \mathbf{L^2}[0,1]$ is in $\mathbf{W}^s[0,1]$ if and only if*

$$\sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} 2^{-2sj}\, |\langle f, \psi_{j,n}\rangle|^2 < +\infty. \tag{9.17}$$

Proof. We give an intuitive justification but not a proof of this result. To simplify, we suppose that the support of f is included in $(0, 1)$. If we extend f by zeros outside $[0, 1]$ then $f \in \mathbf{W}^s(\mathbb{R})$, which means that

$$\int_{-\infty}^{+\infty} |\omega|^{2s}\, |\hat{f}(\omega)|^2\, d\omega < +\infty. \tag{9.18}$$

The low frequency part of this integral always remains finite because $f \in \mathbf{L^2}(\mathbb{R})$:

$$\int_{|\omega|\leqslant 2^{-J}\pi} |\omega|^{2s}\, |\hat{f}(\omega)|^2\, d\omega \leqslant 2^{-2sJ}\, \pi^{2s} \int_{|\omega|\leqslant\pi} |\hat{f}(\omega)|^2\, d\omega \leqslant 2^{-2sJ}\, \pi^{2s}\, \|f\|^2.$$

The energy of $\hat{\psi}_{j,n}$ is essentially concentrated in the intervals $[-2^{-j}2\pi, -2^{-j}\pi] \cup [2^{-j}\pi, 2^{-j}2\pi]$. As a consequence

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^2 \sim \int_{2^{-j}\pi\leqslant|\omega|\leqslant 2^{-j+1}\pi} |\hat{f}(\omega)|^2\, d\omega.$$

Over this interval $|\omega| \sim 2^{-j}$, so

$$\sum_{n=0}^{2^{-j}-1} 2^{-2sj}\, |\langle f, \psi_{j,n}\rangle|^2 \sim \int_{2^{-j}\pi\leqslant|\omega|\leqslant 2^{-j+1}\pi} |\omega|^{2s}\, |\hat{f}(\omega)|^2\, d\omega.$$

It follows that

$$\sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} 2^{-2sj}\, |\langle f, \psi_{j,n}\rangle|^2 \sim \int_{|\omega|\geqslant 2^{-J}\pi} |\omega|^{2s}\, |\hat{f}(\omega)|^2\, d\omega,$$

which explains why (9.18) is equivalent to (9.17). ■ ■

This theorem proves that the Sobolev regularity of f is equivalent to a fast decay of the wavelet coefficients $|\langle f, \psi_{j,n}\rangle|$ when the scale 2^j decreases. If ψ has q vanishing moments but is not q times continuously differentiable, then $f \in \mathbf{W}^s[0,1]$ implies (9.17), but the opposite implication is not true. The following theorem uses the decay condition (9.17) to characterize the linear approximation error with N wavelets.

Theorem 9.5. *Let $0 < s < q$ be a Sobolev exponent. A function $f \in \mathbf{L^2}[0,1]$ is in $\mathbf{W}^s[0,1]$ if and only if*

$$\sum_{N=1}^{+\infty} N^{2s} \frac{\varepsilon_l(N,f)}{N} < +\infty \ , \tag{9.19}$$

which implies $\varepsilon_l(N,f) = o(N^{-2s})$.

Proof. Let us write the wavelets $\psi_{j,n} = g_m$ with $m = 2^{-j} + n$. One can verify that the Sobolev condition (9.17) is equivalent to

$$\sum_{m=0}^{+\infty} |m|^{2s} \, |\langle f, g_m \rangle|^2 < +\infty.$$

The proof ends by applying Theorem 9.1. ■ ■

If the wavelet has q vanishing moments but is not q times continuously differentiable, then $f \in \mathbf{W}^s[0,1]$ implies (9.19) but the opposite implication is false. Theorem 9.5 proves that $f \in \mathbf{W}^s[0,1]$ if and only if the approximation error $\varepsilon_l(N,f)$ decays slightly faster than N^{-2s}. The wavelet approximation error is of the same order as the Fourier approximation error calculated in (9.11). However, Fourier approximations impose that the support of f is strictly included in $[0,1]$ where as wavelet approximation does not impose this condition or any other boundary condition, because of the finer wavelet treatment of boundaries previously explained.

Lipschitz Regularity A different measure of uniform regularity is provided by Lipschitz exponents, which compute the error of a local polynomial approximation. A function f is uniformly Lipschitz α over $[0,1]$ if there exists $K > 0$, such that for any $v \in [0,1]$ one can find a polynomial p_v of degree $\lfloor \alpha \rfloor$ such that

$$\forall t \in [0,1] \ , \quad |f(t) - p_v(t)| \leqslant K \, |t - v|^\alpha. \tag{9.20}$$

The infimum of the K which satisfy (9.20) is the *homogeneous* Hölder α norm $\|f\|_{\tilde{\mathbf{C}}^\alpha}$. The Hölder α norm of f also imposes that f is bounded:

$$\|f\|_{\mathbf{C}^\alpha} = \|f\|_{\tilde{\mathbf{C}}^\alpha} + \|f\|_\infty \ . \tag{9.21}$$

The space $\mathbf{C}^\alpha[0,1]$ of functions f such that $\|f\|_{\mathbf{C}^\alpha} < +\infty$ is called a Hölder space. The following theorem characterizes the decay of wavelet coefficients.

Theorem 9.6. *There exists $B \geqslant A > 0$ such that*

$$A \, \|f\|_{\tilde{\mathbf{C}}^\alpha} \leqslant \sup_{j \geqslant J, 0 \leqslant n < 2^{-j}} 2^{-j(\alpha+1/2)} \, |\langle f, \psi_{j,n} \rangle| \leqslant B \, \|f\|_{\tilde{\mathbf{C}}^\alpha} \ . \tag{9.22}$$

Proof. The proof of the equivalence between uniform Lipschitz regularity and the coefficient decay of a continuous wavelet transform is given in Theorem 6.3. This theorem gives a nearly equivalent result in the context of orthonormal wavelet coefficients, that correspond to a sampling of a continuous wavelet transform computed with the same mother wavelet. The theorem proof is thus an adaptation of the proof Theorem 6.3. This is illustrated by proving the right inequality of (9.22)

If f is uniformly Lipschitz α on the support of $\psi_{j,n}$, since $\psi_{j,n}$ is orthogonal the polynomial $p_{2^j n}$, approximating f at $v = 2^j n$ yields

$$|\langle f, \psi_{j,n} \rangle| = |\langle f - p_{2^j n}, \psi_{j,n} \rangle| \leqslant \|f\|_{\tilde{\mathbf{C}}^\alpha} \int 2^{-j/2} \, |\psi(2^{-j}(t - 2^j n))| \, |t - 2^j n|^\alpha \, dt \ . \tag{9.23}$$

With a change of variable, we get

$$|\langle f, \psi_{j,n} \rangle| \leqslant \|f\|_{\tilde{\mathbf{C}}^\alpha} \, 2^{j(\alpha+1/2)} \int |\psi(t)| \, |t|^\alpha \, dt,$$

which proves the right inequality of (9.22). Observe that we do not use the wavelet regularity in this proof.

The left inequality of (9.22) is proved by following the steps of continuous wavelet tranfsorm Theorem 6.3, and replacing integrals by discrete sums over the position and scale of orthogonal wavelets. The regularity of wavelets plays an important role as shown by the proof Theorem 6.3. In this case there is no boundary issue because wavelets are adatped to the interval $[0,1]$ and keep their vanishing moments at the boundaries.

Similarly to Theorem 9.4 for Sobolev differentiability, this theorem proves that uniform Lipschitz regularity is characterized by the decay of orthogonal wavelet coefficients when the scale 2^j decreases. Hölder and Sobolev spaces belong to the larger class of Besov spaces, defined in Section 9.2.3.

If ψ has q vanishing moments but is not q times continuously differentiable, then the proof of Theorem 9.6 shows that the right inequality of (9.22) is valid. The following theorem derives the decay of linear approximation errors, for wavelets having q vanishing moments but which are not necessarily $\mathbf{C}^q$.

Theorem 9.7. *If f is uniformly Lipschitz $0 < \alpha \leqslant q$ over $[0,1]$ then $\varepsilon_l(N,f) = O(\|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, N^{-2\alpha})$.*

Proof. Theorem 9.6 proves that

$$|\langle f, \psi_{j,n}\rangle| \leqslant B\, \|f\|_{\tilde{\mathbf{C}}^\alpha}\, 2^{j(\alpha+1/2)}\,. \tag{9.24}$$

There are 2^{-j} wavelet coefficients at a scale 2^j, so there are 2^{-k} wavelet coefficients at scales $2^j > 2^k$. The right inequality (9.22) implies that

$$\varepsilon_l(2^{-k}, f) = \sum_{j=-\infty}^{k} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^2 \leqslant B^2 \sum_{j=-\infty}^{k} 2^{-j}\, 2^{j(2\alpha+1)} = \frac{B^2\, \|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, 2^{2\alpha k}}{1-2^{-2\alpha}}.$$

For $k = -\lfloor \log_2 N \rfloor$, we derive that $\varepsilon_l(N,f) = O(\|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, 2^{2\alpha k}) = O(\|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, N^{-2\alpha})$. ■

Discontinuity and Bounded Variation If f is not uniformly regular then linear wavelet approximations perform poorly. If f has a discontinuity in $(0,1)$ then $f \notin \mathbf{W}^s[0,1]$ for $s > 1/2$ so Theorem 9.5 proves that we cannot have $\varepsilon_l(N,f) = O(N^{-\alpha})$ for $\alpha > 1$.

If f has a bounded total variation norm $\|f\|_V$ then Theorem 9.14 will prove that wavelet approximation error satisfy $\varepsilon_l(N,f) = O(\|f\|_V^2\, N^{-1})$. The same Fourier approximation result was obtained in Theorem 9.3. If $f = C\,\mathbf{1}_{[0,1/2]}$ then one can verify that wavelet approximation gives $\varepsilon_l(N,f) \sim \|f\|_V^2\, N^{-1}$ (Exercise 9.5).

Figure 9.1 gives an example of discontinuous signal with bounded variation, which is approximated by its larger scale wavelet coefficients. The largest amplitude errors are in the neighborhood of singularities, where the scale should be refined. The relative approximation error $\|f - f_N\|/\|f\| = 8.56\, 10^{-2}$ is almost the same as in a Fourier basis.

9.1.4 Karhunen-Loève Approximations

Suppose that signals are modeled as realizations of a random process F. We prove that the basis which minimizes the average linear approximation error is the Karhunen-Loève basis which diagonalizes the covariance operator of F. To avoid the subtleties of diagonalizing infinite dimensional operator, we consider signals of finite dimension P, which means that $F[n]$ is a random vector of size P.

Appendix A.6 reviews the covariance properties of random vectors. If $F[n]$ does not have a zero mean, we subtract the expected value $\mathsf{E}\{F[n]\}$ from $F[n]$ to get a zero mean. The random vector F can be decomposed in an orthogonal basis $\{g_m\}_{0\leqslant m<P}$:

$$F = \sum_{m=0}^{P-1} \langle F, g_m\rangle\, g_m\ .$$

Each coefficient

$$\langle F, g_m\rangle = \sum_{n=0}^{P-1} F[n]\, g_m^*[n]$$

is a random variable (see Appendix A.6). The approximation from the first N vectors of the basis is the orthogonal projection on the space $\mathbf{U}_N$ generated by these vectors:

$$F_N = \sum_{m=0}^{N-1} \langle F, g_m\rangle\, g_m.$$

The resulting mean-square error is

$$\mathsf{E}\{\varepsilon_l(N,F)\} = \mathsf{E}\Big\{\|F - F_N\|^2\Big\} = \sum_{m=N}^{P-1} \mathsf{E}\Big\{|\langle F, g_m\rangle|^2\Big\}.$$

The error is related to the covariance of F defined by

$$R_F[n,m] = \mathsf{E}\{F[n]\,F^*[m]\}.$$

Let K_F be the covariance operator represented by this matrix. It is symmetric and positive and is thus diagonalized in an orthogonal basis called a Karhunen-Loève basis. This basis is not unique if several eigenvalues are equal. The following theorem proves that a Karhunen-Loève basis is optimal for linear approximations.

Theorem 9.8. *For all $N \geqslant 1$, the expected approximation error*

$$\mathsf{E}\{\varepsilon_l(N,F)\} = E\{\|F - F_N\|^2\} = \sum_{m=N}^{P-1} E\{|\langle F, g_m\rangle|^2\}$$

is minimum if and only if $\{g_m\}_{0\leqslant m<P}$ is a Karhunen-Loève basis which diagonalizes the covariance K_F of F, with vectors indexed in decreasing eigenvalue order:

$$\langle K_F g_m, g_m\rangle \geqslant \langle K_F g_{m+1}, g_{m+1}\rangle \quad \textit{for } 0 \leqslant m < P-1.$$

Proof. Let us first observe that

$$\mathsf{E}\{\varepsilon_l(F,N)\} = \sum_{m=N}^{P-1} \langle K_F g_m, g_m\rangle \tag{9.25}$$

because for any vector $z[n]$,

$$\begin{aligned} \mathsf{E}\Big\{|\langle F, z\rangle|^2\Big\} &= \mathsf{E}\left\{\sum_{n=0}^{P-1}\sum_{m=0}^{P-1} F[n]\,F[m]\,z[n]\,z^*[m]\right\} \\ &= \sum_{n=0}^{P-1}\sum_{m=0}^{P-1} R_F[n,m]\,z[n]\,z^*[m] \\ &= \langle K_F z, z\rangle. \end{aligned}$$

We now prove that (9.25) is minimum if the basis diagonalizes K_F. Let us consider an arbitrary orthonormal basis $\{h_m\}_{0\leqslant m<P}$. The trace $\mathrm{tr}(K_F)$ of K_F is independent of the basis:

$$\mathrm{tr}(K_F) = \sum_{m=0}^{P-1} \langle K_F h_m, h_m\rangle.$$

The basis that minimizes $\sum_{m=N}^{P-1} \langle K_F h_m, h_m\rangle$ thus maximizes $\sum_{m=0}^{N-1} \langle K_F h_m, h_m\rangle$.

Let $\{g_m\}_{0\leqslant m<P}$ be a basis that diagonalizes K_F:

$$K_F g_m = \sigma_m^2\, g_m \quad \text{with} \quad \sigma_m^2 \geqslant \sigma_{m+1}^2 \text{ for } 0 \leqslant m < P-1.$$

The theorem is proved by verifying that for all $N \geqslant 0$,

$$\sum_{m=0}^{N-1} \langle K_F h_m, h_m\rangle \leqslant \sum_{m=0}^{N-1} \langle K_F g_m, g_m\rangle = \sum_{m=0}^{N-1} \sigma_m^2.$$

To relate $\langle K_F h_m, h_m\rangle$ to the eigenvalues $\{\sigma_i^2\}_{0\leqslant i<P}$, we expand h_m in the basis $\{g_i\}_{0\leqslant i<P}$:

$$\langle K_F h_m, h_m\rangle = \sum_{i=0}^{P-1} |\langle h_m, g_i\rangle|^2\, \sigma_i^2. \tag{9.26}$$

Hence

$$\sum_{m=0}^{N-1} \langle K_F h_m, h_m \rangle = \sum_{m=0}^{N-1} \sum_{i=0}^{P-1} |\langle h_m, g_i \rangle|^2 \sigma_i^2 = \sum_{i=0}^{P-1} q_i \sigma_i^2$$

with

$$0 \leqslant q_i = \sum_{m=0}^{N-1} |\langle h_m, g_i \rangle|^2 \leqslant 1 \quad \text{and} \quad \sum_{i=0}^{P-1} q_i = N.$$

We evaluate

$$\begin{aligned} \sum_{m=0}^{N-1} \langle K_F h_m, h_m \rangle \quad - & \quad \sum_{i=0}^{N-1} \sigma_i^2 = \sum_{i=0}^{P-1} q_i \sigma_i^2 - \sum_{i=0}^{N-1} \sigma_i^2 \\ = & \quad \sum_{i=0}^{P-1} q_i \sigma_i^2 - \sum_{i=0}^{N-1} \sigma_i^2 + \sigma_{N-1}^2 \Big(N - \sum_{i=0}^{P-1} q_i \Big) \\ = & \quad \sum_{i=0}^{N-1} (\sigma_i^2 - \sigma_{N-1}^2)(q_i - 1) + \sum_{i=N}^{P-1} q_i (\sigma_i^2 - \sigma_{N-1}^2). \end{aligned}$$

Since the eigenvalues are listed in order of decreasing amplitude, it follows that

$$\sum_{m=0}^{N-1} \langle K_F h_m, h_m \rangle - \sum_{m=0}^{N-1} \sigma_m^2 \leqslant 0.$$

Suppose that this last inequality is an equality. We finish the proof by showing that $\{h_m\}_{0 \leqslant m < P}$ must be a Karhunen-Loève basis. If $i < N$, then $\sigma_i^2 \neq \sigma_{N-1}^2$ implies $q_i = 1$. If $i \geqslant N$, then $\sigma_i^2 \neq \sigma_{N-1}^2$ implies $q_i = 0$. This is valid for all $N \geqslant 0$ if $\langle h_m, g_i \rangle \neq 0$ only when $\sigma_i^2 = \sigma_m^2$. This means that the change of basis is performed inside each eigenspace of K_F so $\{h_m\}_{0 \leqslant m < P}$ also diagonalizes K_F. ■ ■

The eigenvectors g_m of the covariance matrix are called the signal *principal components*. Theorem 9.8 proves that a Karhunen-Loève basis yields the smallest expected linear error when approximating a class of signals by their projection on N orthogonal vectors.

Theorem 9.8 has a simple geometrical interpretation. The realizations of F define a cloud of points in $\mathbb{C}^P$. The density of this cloud specifies the probability distribution of F. The vectors g_m of the Karhunen-Loève basis give the directions of the principal axes of the cloud. Large eigenvalues σ_m^2 correspond to directions g_m along which the cloud is highly elongated. Theorem 9.8 proves that projecting the realizations of F on these principal components yields the smallest average error. If F is a Gaussian random vector, the probability density is uniform along ellipsoids whose axes are proportional to σ_m in the direction of g_m. These principal directions are thus truly the preferred directions of the process.

Random Shift Processes If the process is not Gaussian, its probability distribution can have a complex geometry, and a linear approximation along the principal axes may not be efficient. As an example, we consider a random vector $F[n]$ of size P that is a random shift modulo P of a deterministic signal $f[n]$ of zero mean, $\sum_{n=0}^{P-1} f[n] = 0$:

$$F[n] = f[(n - Q) \bmod P]. \tag{9.27}$$

The shift Q is an integer random variable whose probability distribution is uniform on $[0, P-1]$:

$$\Pr(Q = p) = \frac{1}{P} \quad \text{for } 0 \leqslant p < P.$$

This process has a zero mean:

$$E\{F[n]\} = \frac{1}{P} \sum_{p=0}^{P-1} f[(n-p) \bmod P] = 0,$$

and its covariance is

$$\begin{aligned} R_F[n,k] &= \mathsf{E}\{F[n]F[k]\} = \frac{1}{P}\sum_{p=0}^{P-1} f[(n-p) \bmod P]\, f[(k-p) \bmod P] \\ &= \frac{1}{P} f \circledast \bar{f}[n-k] \quad \text{with} \quad \bar{f}[n] = f[-n] \ . \end{aligned} \tag{9.28}$$

Hence $R_F[n,k] = R_F[n-k]$ with

$$R_F[k] = \frac{1}{P} f \circledast \bar{f}[k].$$

Since R_F is P periodic, F is a circular stationary random vector, as defined in Appendix A.6. The covariance operator K_F is a circular convolution with R_F, and is therefore diagonalized in the discrete Fourier Karhunen-Loève basis $\{\frac{1}{\sqrt{P}} \exp\left(\frac{i2\pi mn}{P}\right)\}_{0 \leqslant m < P}$. The eigenvalues are given by the Fourier transform of R_F:

$$\sigma_m^2 = \hat{R}_F[m] = \frac{1}{P}|\hat{f}[m]|^2. \tag{9.29}$$

Theorem 9.8 proves that a linear approximation yields a minimum error in this Fourier basis. To better understand this result, let us consider an extreme case where $f[n] = \delta[n] - \delta[n-1]$. Theorem 9.8 guarantees that the Fourier Karhunen-Loève basis produces a smaller expected approximation error than does a canonical basis of Diracs $\{g_m[n] = \delta[n-m]\}_{0 \leqslant m < P}$. Indeed, we do not know a priori the abscissa of the non-zero coefficients of F, so there is no particular Dirac that is better adapted to perform the approximation. Since the Fourier vectors cover the whole support of F, they always absorb part of the signal energy:

$$\mathsf{E}\left\{\left|\left\langle F[n], \frac{1}{\sqrt{P}} \exp\left(\frac{i2\pi mn}{P}\right)\right\rangle\right|^2\right\} = \hat{R}_F[m] = \frac{4}{P}\sin^2\left(\frac{\pi k}{P}\right).$$

Selecting N higher frequency Fourier coefficients thus yields a better mean-square approximation than choosing a priori N Dirac vectors to perform the approximation.

The linear approximation of F in a Fourier basis is not efficient because all the eigenvalues $\hat{R}_F[m]$ have the same order of magnitude. A simple non-linear algorithm can improve this approximation. In a Dirac basis, F is exactly reproduced by selecting the two Diracs corresponding to the largest amplitude coefficients, whose positions Q and $Q-1$ depend on each realization of F. A non-linear algorithm that selects the largest amplitude coefficient for each realization of F is not efficient in a Fourier basis. Indeed, the realizations of F do not have their energy concentrated over a few large amplitude Fourier coefficients. This example shows that when F is not a Gaussian process, a non-linear approximation may be much more precise than a linear approximation, and the Karhunen-Loève basis is no longer optimal.

9.2 Non-Linear Approximations

Digital images or sounds are signals discretized over spaces of large dimension N, because linear approximation error have a slow decay. Digital camera images have $N \geqslant 10^6$ pixels where as 1 second of a CD recording has $N = 40\,10^3$ samples. Sparse signal representations are obtained by projecting such signals over less vectors selected adaptively in an orthonormal basis of discrete signals in $\mathbb{C}^N$. This is equivalent to perform a non-linear approximation of the input analog signal in a basis of $\mathbf{L^2}[0,1]$.

Th next section analyzes the properties of the resulting non-linear approximation error. Sections 9.2.2 and 9.2.3 proves that non-linear wavelet approximations are equivalent to adaptive grids, and can provide sparse representations of signals including singularities. Approximations of functions in Besov spaces and with bounded variations are studied in Section 9.2.3.

9.2.1 Non-Linear Approximation Error

The discretization of an input analog signal f computes N sample values $\{\langle f, \phi_n\rangle\}_{0\leqslant n<N}$ that specify the projection f_N of f over an approximation space $\mathbf{U}_N$ of dimension N. A non-linear approximation further approximates this projection over a basis providing a sparse representation.

Let $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$ be an orthonormal basis of $\mathbf{L^2}[0,1]$ or $\mathbf{L^2}[0,1]^2$, whose first N vectors is a basis of $\mathbf{U}_N$. The orthogonal projection in $\mathbf{U}_N$ can be written

$$f_N(x) = \sum_{m=0}^{N-1} \langle f, g_m\rangle\, g_m(x),$$

and the linear approximation error is

$$\|f - f_N\|^2 = \sum_{n=N}^{+\infty} |\langle f, g_m\rangle|^2.$$

Let us re-project f_N over over a subset of $M < N$ vectors $\{g_m\}_{m\in\Lambda}$ with $\Lambda \subset [0, N-1]$:

$$f_\Lambda(x) = \sum_{m\in\Lambda} \langle f, g_m\rangle\, g_m(x).$$

The approximation error is the sum of the remaining coefficients:

$$\|f_N - f_\Lambda\|^2 = \sum_{m\notin\Lambda} |\langle f, g_m\rangle|^2. \tag{9.30}$$

The approximation set which minimizes this error is the set Λ_T of M vectors corresponding to the largest inner product amplitude $|\langle f, g_m\rangle|$, and thus above a threshold T that depends on M:

$$\Lambda_T = \{m \;:\; 0 \leqslant m < N \;,\; |\langle f, g_m\rangle| \geqslant T\} \quad \text{with} \quad |\Lambda_T| = M\ . \tag{9.31}$$

The minimum approximation error is the energy of coefficients below T

$$\varepsilon_n(M, f) = \|f_N - f_{\Lambda_T}\|^2 = \sum_{m\notin\Lambda_T} |\langle f, g_m\rangle|^2.$$

In the following we often write $f_M = f_{\Lambda_T}$ this best M-term approximation.

The overall error is the sum of the linear error when projecting f on $\mathbf{U}_N$ and the non-linear approximation error:

$$\varepsilon_n(M, f) = \|f - f_M\|^2 = \|f - f_N\|^2 + \|f_N - f_M\|^2. \tag{9.32}$$

If N is large enough so that all coefficients above T are in the first N:

$$T \geqslant \max_{|m|\geqslant N} |\langle f, g_m\rangle| \;\text{ and hence }\; N > \arg\max_m \{|\langle f, g_m\rangle| \geqslant T\} \tag{9.33}$$

then the M largest signal coefficients are among the first N and the non-linear error (9.32) is the minimum error obtained from M coefficients chosen anywhere in the infinite basis $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$. In this case, the linear approximation space $\mathbf{U}_N$ and f_N do not play any explicit role in the error $\varepsilon_n(M, f)$. If $|\langle f, g_m\rangle| \leqslant C\, m^{-\beta}$ for some $\beta > 0$, then we can choose $N \geqslant C^\beta\, T^\beta$. In the following, this condition condition is supposed to be satisfied.

Discrete Numerical Computations The linear approximation space $\mathbf{U}_N$ is important for discrete computations. A non-linear approximation f_{Λ_T} is computed by calculating the non-linear approximation of the discretized signal $a[n] = \langle f, \phi_n\rangle$ for $0 \leqslant n < N$, and performing a discrete to analog conversion.

Since the discretization family $\{\phi_n\}_{0\leqslant n<N}$ and the approximation basis $\{g_m\}_{0\leqslant n<N}$ are both orthonormal bases of $\mathbf{U}_N$, $\{h_m[n] = \langle g_m, \phi_n\rangle\}_{0\leqslant m<N}$ is an orthonormal basis of $\mathbb{C}^N$. Analog signal inner products in $\mathbf{L^2}[0,1]$ and their discretization in $\mathbb{C}^N$ are then equal

$$\langle a[n], h_m[n]\rangle = \langle f(x), g_m(x)\rangle \quad \text{for } 0 \leqslant m < N\ . \tag{9.34}$$

The non-linear approximation of the signal $a[n]$ in the basis $\{h_m\}_{0\leqslant m<N}$ of $\mathbb{C}^N$ is

$$a_{\Lambda_T}[n] = \sum_{m\in\Lambda_T} \langle a, h_m\rangle\, h_m[n] \quad \text{with} \quad \Lambda_T = \{m \ : \ |\langle a, h_m\rangle| \geqslant T\} \ .$$

It results from (9.34) that the analog conversion of this discrete signal is the non-linear analog approximation:

$$f_{\Lambda_T}(x) = \sum_{n=0}^{N-1} a_{\Lambda_T}[n]\, \phi_n(x) = \sum_{n=0}^{N-1} a_{\Lambda_T}[n]\, \phi_n(x) = \sum_{m\in\Lambda_T} \langle f, g_m\rangle\, g_m(x) \ .$$

The number of operations to compute a_{Λ_T} is dominated by the number of operations to compute the N signal coefficients $\{\langle a, h_m\rangle\}_{0\leqslant n<N}$ which takes $O(N\log_2 N)$ operations in a discrete Fourier basis, and $O(N)$ in a discrete wavelet basis. Reducing N thus decreases the number of operations and does not affect the non-linear approximation error, as long as (9.33) is satisfied. Given this equivalence between discrete and analog non-linear approximations, we now concentrate on analog functions to relate this error to their regularity.

Approximation Error To evaluate the non-linear approximation error $\varepsilon_n(M,f)$, the coefficients $\{|\langle f, g_m\rangle|\}_{m\in\mathbb{N}}$ are sorted in decreasing order. Let $f_{\mathcal{B}}^r[k] = \langle f, g_{m_k}\rangle$ be the coefficient of rank k:

$$|f_{\mathcal{B}}^r[k]| \geqslant |f_{\mathcal{B}}^r[k+1]| \quad \text{with } k>0.$$

The best M-term non-linear approximation computed from the M largest coefficients is:

$$f_M = \sum_{k=1}^{M} f_{\mathcal{B}}^r[k]\, g_{m_k}. \tag{9.35}$$

The resulting error is

$$\varepsilon_n(M,f) = \|f - f_M\|^2 = \sum_{k=M+1}^{+\infty} |f_{\mathcal{B}}^r[k]|^2.$$

The following theorem relates the decay of this approximation error as M increases to the decay of $|f_{\mathcal{B}}^r[k]|$ as k increases.

Theorem 9.9. *Let $s>1/2$. If there exists $C>0$ such that $|f_{\mathcal{B}}^r[k]| \leqslant C\,k^{-s}$ then*

$$\varepsilon_n(M,f) \leqslant \frac{C^2}{2s-1}\, M^{1-2s}. \tag{9.36}$$

Conversely, if $\varepsilon_n(M,f)$ satisfies (9.36) then

$$|f_{\mathcal{B}}^r[k]| \leqslant \left(1 - \frac{1}{2s}\right)^{-s} C\,k^{-s}\ . \tag{9.37}$$

Proof. Since

$$\varepsilon_n(M,f) = \sum_{k=M+1}^{+\infty} |f_{\mathcal{B}}^r[k]|^2 \leqslant C^2 \sum_{k=M+1}^{+\infty} k^{-2s},$$

and

$$\sum_{k=M+1}^{+\infty} k^{-2s} \leqslant \int_M^{+\infty} x^{-2s}\, dx = \frac{M^{1-2s}}{2s-1} \tag{9.38}$$

we derive (9.36).

Conversely, let $\alpha<1$,

$$\varepsilon_n(\alpha M,f) \geqslant \sum_{k=\alpha M+1}^{M} |f_{\mathcal{B}}^r[k]|^2 \geqslant (1-\alpha)\, M\, |f_{\mathcal{B}}^r[M]|^2\ .$$

So if (9.36) is satisfied

$$|f_{\mathcal{B}}^r[M]|^2 \leqslant \frac{\varepsilon_n(\alpha M,f)}{1-\alpha} M^{-1} \leqslant \frac{C^2}{2s-1}\,\frac{\alpha^{1-2s}}{1-\alpha}\, M^{-2s}.$$

For $\alpha = 1 - 1/2s$ we get (9.37) for $k=M$. ■ ■

l^p spaces The following theorem relates the decay of sorted inner products to their $\boldsymbol{\ell}^p$ norm

$$\|f\|_{\mathcal{B},p} = \left(\sum_{m=0}^{+\infty} |\langle f, g_m\rangle|^p\right)^{1/p} .$$

It derives a decay of the error $\varepsilon_n(M,f)$ to

Theorem 9.10. *Let $p<2$. If $\|f\|_{\mathcal{B},p} < +\infty$ then*

$$|f_{\mathcal{B}}^r[k]| \leqslant \|f\|_{\mathcal{B},p}\ k^{-1/p} \tag{9.39}$$

and $\varepsilon_n(M,f) = o(M^{1-2/p})$.

Proof. We prove (9.39) by observing that

$$\|f\|_{\mathcal{B},p}^p = \sum_{n=1}^{+\infty} |f_{\mathcal{B}}^r[n]|^p \geqslant \sum_{n=1}^{k} |f_{\mathcal{B}}^r[n]|^p \geqslant k\,|f_{\mathcal{B}}^r[k]|^p .$$

To show that $\varepsilon_n(M,f) = o(M^{1-2/p})$, we set

$$S[k] = \sum_{n=k}^{2k-1} |f_{\mathcal{B}}^r[n]|^p \geqslant k\,|f_{\mathcal{B}}^r[2k]|^p .$$

Hence

$$\begin{aligned} \varepsilon_n(M,f) &= \sum_{k=M+1}^{+\infty} |f_{\mathcal{B}}^r[k]|^2 \leqslant \sum_{k=M+1}^{+\infty} S[k/2]^{2/p}\,(k/2)^{-2/p} \\ &\leqslant \sup_{k>M/2} |S[k]|^{2/p} \sum_{k=M+1}^{+\infty} (k/2)^{-2/p} . \end{aligned}$$

Since $\|f\|_{\mathcal{B},p}^p = \sum_{n=1}^{+\infty} |f_{\mathcal{B}}^r[n]|^p < +\infty$, it follows that $\lim_{k\to+\infty} \sup_{k>M/2} |S[k]| = 0$. We thus derive from (9.38) that $\varepsilon_n(M,f) = o(M^{1-2/p})$. ■ ■

This theorem specifies spaces of functions that are well approximated by a few vectors of an orthogonal basis $\mathcal{B}$. We denote

$$\mathbf{B}_{\mathcal{B},p} = \Big\{ f \in \mathbf{H} \quad : \quad \|f\|_{\mathcal{B},p} < +\infty \Big\}. \tag{9.40}$$

If $f \in \mathbf{B}_{\mathcal{B},p}$ then Theorem 9.10 proves that $\varepsilon_n(M,f) = o(M^{1-2/p})$. This is called a *Jackson inequality* [19]. Conversely, if $\varepsilon_n(M,f) = O(M^{1-2/p})$ then the *Bernstein inequality* (9.37) for $s = 1/p$ shows that $f \in \mathbf{B}_{\mathcal{B},q}$ for any $q > p$. Section 9.2.3 studies the properties of the spaces $\mathbf{B}_{\mathcal{B},p}$ for wavelet bases.

9.2.2 Wavelet Adaptive Grids

A non-linear approximation in a wavelet orthonormal basis keeps the largest amplitude coefficients. We saw in Section 6.1.3 that these coefficients occur near singularities. A wavelet non-linear approximation thus defines an adaptive grid that refines the approximation scale in the neighborhood of the signal sharp transitions. Such approximations are particularly well adapted to piecewise regular signals. The precision of non-linear wavelet approximation is also studied for bounded variation functions and more general Besov space functions. [208]

We consider a wavelet basis adapted to $\mathbf{L^2}[0,1]$, constructed in Section 7.5.3 with compactly supported wavelets that are $\mathbf{C}^q$ with q vanishing moments:

$$\mathcal{B} = \Big[\{\phi_{J,n}\}_{0\leqslant n<2^{-J}} \ ,\ \{\psi_{j,n}\}_{-\infty<j\leqslant J\,,\,0\leqslant n<2^{-j}} \Big].$$

To simplify notation we write $\phi_{J,n} = \psi_{J+1,n}$.

If the analog signal $f \in \mathbf{L^2}[0,1]$ is approximated at the scale 2^L with $N = 2^{-L}$ samples $\{\langle f, \phi_{L,n}\rangle\}_{0 \leqslant n < 2^{-L}}$ then the corresponding N wavelet coefficients $\{\langle f, \psi_{j,n}\}_{n,j>L}$ are computed with $O(N)$ operations with the fast wavelet transform algorithm of Section 7.3.1. The best non-linear approximation of $f \in \mathbf{L^2}[0,1]$ from M wavelet coefficients above T at scales $2^j > 2^L$ is

$$f_M = \sum_{(j,n)\in\Lambda_T} \langle f, \psi_{j,n}\rangle\, \psi_{j,n} \quad \text{with } \Lambda_T = \{(j,n) \ : \ j > L \ , \ |\langle f, \psi_{j,n}\rangle| \geqslant T\} \ .$$

The approximation error is $\varepsilon_n(M,f) = \sum_{(j,n)\notin\Lambda_T} |\langle f, \psi_{j,n}\rangle|^2$. The following theorem proves that if f is bounded, then for a N sufficiently large the approximation support Λ_T corresponds to all possible wavelet coefficients above T.

Theorem 9.11. *If f is bounded then all wavelets producing coefficients above T are in an approximation space $\mathbf{V}_L$ of dimension $N = 2^{-L} = O(\|f\|_\infty^2\, T^{-2})$.*

Proof. If f is bounded then

$$|\langle f, \psi_{j,n}\rangle| = \Big|\int_0^1 f(t)\, 2^{-j/2}\psi(2^{-j}t - n)\, dt\Big| \leqslant 2^{j/2} \sup_t |f(t)| \int_0^1 |\psi(t)|\, dt = 2^{j/2}\, \|f\|_\infty\, \|\psi\|_1. \tag{9.41}$$

So $|\langle f, \psi_{j,n}\rangle| \ \geqslant \ T$ implies that $2^j \ \geqslant \ T^2\, \|f\|_\infty^{-2}\, \|\psi\|_1^{-2}$, which proves the theorem for $2^L = T^2\, \|f\|_\infty^{-2}\, \|\psi\|_1^{-2}$. ■

This theorem shows that for bounded signals, if the discretization 2^L is sufficiently small then the non-linear approximation error computed from the first $N = 2^{-L}$ wavelet coefficients is equal to the approximation error obtained by selecting the M largest wavelet coefficients in the infinite dimensional wavelet basis. In the following, we suppose that this condition is satisfied, and thus do not have to worry about the discretization scale.

Piecewise Regular Signals Piecewise regular signal define a first simple model where non-linear wavelet approximations considerably outperform linear approximations. We consider signals with a finite number of singularities and which are uniformly regular between singularities. The following theorem characterize the linear and non-linear wavelet approximation error decay for such signals.

Theorem 9.12. *If f has a K discontinuities on $[0,1]$ and is uniformly Lipschitz α between these discontinuities, with $1/2 < \alpha < q$, then*

$$\varepsilon_l(M,f) = O(K\, \|f\|_{\mathbf{C}^\alpha}^2\, M^{-1}) \quad \textit{and} \quad \varepsilon_n(M,f) = O(\|f\|_{\mathbf{C}^\alpha}^2\, M^{-2\alpha}) \ . \tag{9.42}$$

Proof. We distinguish type I wavelets $\psi_{j,n}$ for $n \in I_j$, whose supports include an abscissa where f is discontinuous, from type II wavelets for $n \in II_j$, whose supports are included in a domain where f is uniformly Lipschitz α.

Let C be the support size of ψ. At a scale 2^j, each wavelet $\psi_{j,n}$ has a support of size $C2^j$, translated by $2^j n$. There are thus at most $|I_j| \leqslant C\, K$ type I wavelets $\psi_{j,n}$ whose support includes at least one of the K discontinuities of f. Since $\|f\|_\infty \leqslant \|f\|_{\mathbf{C}^\alpha}$, (9.41) shows for $\alpha = 0$ that there exists B_0 such that $|\langle f, \psi_{j,n}\rangle| \leqslant B_0\, \|f\|_{\mathbf{C}^\alpha}\, 2^{j/2}$.

At fine scales 2^j, there are much more type II wavelets $n \in II_j$, but this number $|II_j|$ is smaller than the total number 2^{-j} of wavelets at this scale. Since f is uniformly Lipschitz α on the support of $\psi_{j,n}$, the right inequality of (9.22) proves that there exists B such that

$$|\langle f, \psi_{j,n}\rangle| \leqslant B\, \|f\|_{\mathbf{C}^\alpha}\, 2^{j(\alpha+1/2)} \, . \tag{9.43}$$

This linear approximation error from $M = 2^{-k}$ wavelets satisfies

$$\begin{aligned}
\varepsilon_l(M,f) &= \sum_{j\leqslant k}\left(\sum_{n\in I_j} |\langle f, \psi_{j,n}\rangle|^2 + \sum_{n\in II_j} |\langle f, \psi_{j,n}\rangle|^2\right) \\
&\leqslant \sum_{j\leqslant k}\Big(C\, K\, B_0^2\, \|f\|_{\mathbf{C}^\alpha}^2\, 2^j + 2^{-j}\, B^2\, \|f\|_{\mathbf{C}^\alpha}^2\, 2^{(2\alpha+1)j}\Big) \\
&\leqslant \|f\|_{\mathbf{C}^\alpha}^2\, 2\, C\, K\, B_0^2\, 2^k + \|f\|_{\mathbf{C}^\alpha}^2\, (1 - 2^{-2\alpha})^{-1}\, B^2\, 2^{2\alpha k} \ .
\end{aligned}$$

This inequality proves that for $\alpha > 1/2$ the error term of type I wavelets dominates, and $\varepsilon_l(M,f) = O(\|f\|^2_{\mathbf{C}^\alpha}\, K\, M^{-1})$.

To compute the non-linear approximaton error $\varepsilon(M,f)$, we evaluate the decay of ordered wavelet coefficients. Let $f^r_{\mathcal{B}}[k] = \langle f, \psi_{j_k,n_k}\rangle$ be the coefficient of rank k: $|f^r_{\mathcal{B}}[k]| \geqslant |f^r_{\mathcal{B}}[k+1]|$ for $k \geqslant 1$. Let $f^r_{\mathcal{B},1}[k]$ and $f^r_{\mathcal{B},II}[k]$ be the values of the wavelet coefficient of rank k among type I and type II wavelets.

For $l > 0$ there are at most $l\,K\,C$ type I coefficients at scales $2^j > 2^{-l}$ and type I wavelet coefficients satisfy $|\langle f, \psi_{j,n}\rangle| \leqslant B_0 \|f\|_{\mathbf{C}^\alpha} 2^{-l/2}$ at scales $2^j \leqslant 2^{-l}$. It results that

$$f^r_{\mathcal{B},I}[lKC] \leqslant B_0\, \|f\|_{\mathbf{C}^\alpha}\, 2^{-l/2}\,,$$

so $f^r_{\mathcal{B},I}[k] = O(\|f\|_{\mathbf{C}^\alpha}\, 2^{-k/(2KC)})$ has an exponential decay.

For $l \geqslant 0$, there are at most 2^l type II wavelet coefficients at scales $2^j > 2^{-l}$, and type II wavelet coefficients satisfy $|\langle f, \psi_{j,n}\rangle| \leqslant B \|f\|_{\mathbf{C}^\alpha} 2^{-l(\alpha+1/2)}$ at scales $2^j \leqslant 2^{-l}$. It results that

$$f^r_{\mathcal{B},II}[2^{-l}] \leqslant B \|f\|_{\mathbf{C}^\alpha}\, 2^{-l(\alpha+1/2)}\,.$$

It follows that $f^r_{\mathcal{B},II}[k] = O(\|f\|_{\mathbf{C}^\alpha}\, k^{-\alpha-1/2})$, for all $k > 0$.

Since type I coefficients have a much faster decay than type II coefficients, putting them together gives $f^r_{\mathcal{B}}[k] = O(\|f\|_{\mathbf{C}^\alpha}\, k^{-\alpha-1/2})$. From the inequality (9.36) of Theorem 9.9 it results that $\varepsilon_n(M,f) = O(\|f\|^2_{\mathbf{C}^\alpha}\, M^{-2\alpha})$. ■

Although there are few large wavelet coefficients created by the potential K discontinuities, the theorem proof shows that the linear approximation error is dominated by these discontinuities. On the contrary, these few wavelet coefficients have a negligible impact on the non-linear approximation error. It thus decays as if there was no such discontinuities and f was uniformly Lipschitz α over its whole support.

Adaptive Grids The approximation f_M calculated from the M largest amplitude wavelet coefficients can be interpreted as an adaptive grid approximation, where the approximation scale is refined in the neighborhood of singularities.

A non-linear approximation keeps all coefficients above a threshold $|\langle f, \psi_{j,n}\rangle| \geqslant T$. In a region where f is uniformly Lipschitz α, since $|\langle f, \psi_{j,n}\rangle| \sim A\, 2^{j(\alpha+1/2)}$ the coefficients above T are typically at scales

$$2^j > 2^l = \left(\frac{T}{A}\right)^{2/(2\alpha+1)}.$$

Setting to zero all wavelet coefficients below the scale 2^l is equivalent to computing a local approximation of f at the scale 2^l. The smaller the local Lipschitz regularity α, the finer the approximation scale 2^l.

Figure 9.2 shows the non-linear wavelet approximation of a piecewise regular signal. Up and down Diracs correspond to positive and negative wavelet coefficients whose amplitude are above T. The largest amplitude wavelet coefficients are in the cone of influence of each singularity. The scale-space approximation support Λ_T specifies the geometry of the signal sharp transitions. Since the approximation scale is refined in the neighborhood of each singularity, they are much better restored than in the fixed scale linear approximation shown in Figure 9.1. The non-linear approximation error in this case is 17 times smaller than the linear approximation error.

Non-linear wavelet approximations are nearly optimal compared to adaptive spline approximations. A spline approximation $\tilde{f}_M$ is calculated by choosing K nodes $t_1 < t_2 < \cdots < t_K$ inside $[0,1]$. Over each interval $[t_k, t_{k+1}]$, f is approximated by the closest polynomial of degree r. This polynomial spline $\tilde{f}_M$ is specified by $M = K(r+2)$ parameters, which are the node locations $\{t_k\}_{1\leqslant k\leqslant K}$ plus the $K(r+1)$ parameters of the K polynomials of degree r. To reduce $\|f - \tilde{f}_M\|$, the nodes must be closely spaced when f is irregular and farther apart when f is smooth. However, finding the M parameters that minimize $\|f - \tilde{f}_M\|$ is a difficult non-linear optimization.

A spline wavelet basis of Battle-Lemarié gives non-linear approximations that are also splines functions, but the nodes t_k are restricted to dyadic locations $2^j n$, with a scale 2^j that is locally adapted to the signal regularity. It is computed with $O(N)$ operations by projecting the signal

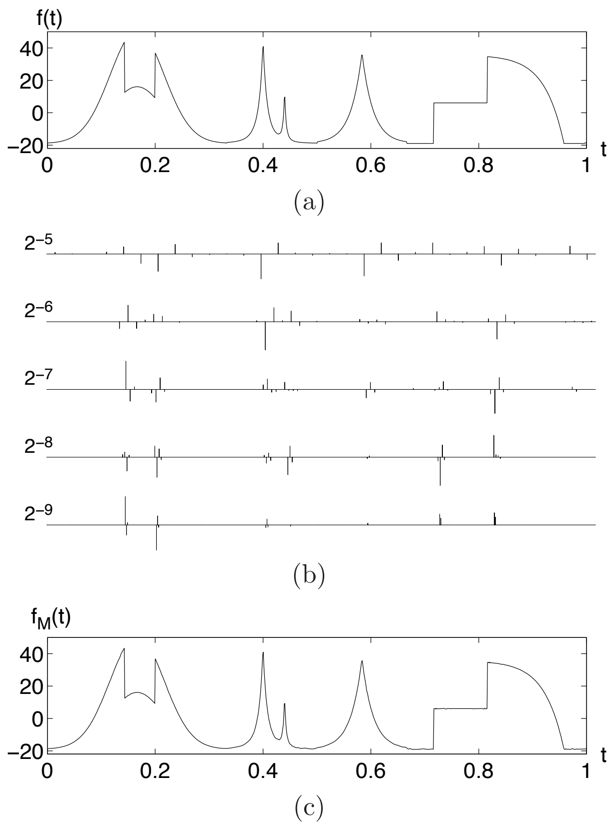


Figure 9.2: (a): Original signal f. (b): Each Dirac corresponds to one of the largest $M = 0.15\,N$ wavelet coefficients, calculated with a Symmlet 4. (c): Non-linear approximation f_M recovered from the M largest wavelet coefficients shown above, $\|f - f_M\|/\|f\| = 5.1\,10^{-3}$.

in an approximation space of dimension $N = O(T^2)$. For large classes of signals, including balls of Besov spaces, the maximum approximation errors with wavelets or with optimized splines have the same decay rate when M increases [209]. The computational overhead of an optimized spline approximation is therefore not worth it.

9.2.3 Approximations in Besov and Bounded Variation Spaces

Studying the performance of non-linear wavelet approximations more precisely requires introducing new spaces. As previously, we write the coarse scale scaling functions $\phi_{J,n} = \psi_{J+1,n}$. The *Besov* space $\mathbf{B}^s_{\beta,\gamma}[0,1]$ is the set of functions $f \in \mathbf{L^2}[0,1]$ such that

$$\|f\|_{s,\beta,\gamma} = \left(\sum_{j=-\infty}^{J+1} \left[2^{-j(s+1/2-1/\beta)} \left(\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^\beta \right)^{1/\beta} \right]^\gamma \right)^{1/\gamma} < +\infty \ . \tag{9.44}$$

Frazier, Jawerth [259] and Meyer [374] proved that $\mathbf{B}^s_{\beta,\gamma}[0,1]$ does not depend on the particular choice of wavelet basis, as long as the wavelets in the basis have $q > s$ vanishing moments and are in $\mathbf{C}^q$. The space $\mathbf{B}^s_{\beta,\gamma}[0,1]$ corresponds typically to functions that have a "derivative of order s" that is in $\mathbf{L}^\beta[0,1]$. The index γ is a fine tuning parameter, which is less important. We need $q > s$ because a wavelet with q vanishing moments can test the differentiability of a signal only up to the order q. When removing from (9.44) the coarsest scale scaling functions $\phi_{J,n} = \psi_{J+1,n}$ the norm $\|f\|_{s,\beta,\gamma}$ is called a homogeneous Besov norm that we shall write $\|f\|^*_{s,\beta,\gamma}$ and the corresponding homogeneous Besov space is $\tilde{\mathbf{B}}^s_{\beta,\gamma}[0,1]$.

If $\beta \geqslant 2$, then functions in $\mathbf{B}^s_{\beta,\gamma}[0,1]$ have a uniform regularity of order s. For $\beta = \gamma = 2$, Theorem 9.4 proves that $\mathbf{B}^s_{2,2}[0,1] = \mathbf{W}^s[0,1]$ is the space of s times differentiable functions in the sense of Sobolev. Theorem 9.5 proves that this space is characterized by the decay of the linear approximation error $\varepsilon_l(N,f)$ and that $\varepsilon_l(N,f) = o(N^{-2s})$. Since $\varepsilon_n(M,f) \leqslant \varepsilon_l(M,f)$ clearly $\varepsilon_n(M,f) = o(M^{-2s})$. One can verify (Exercise 9.11) that non-linear approximation do not improve linear approximations over Sobolev spaces. For $\beta = \gamma = \infty$, Theorem 9.6 proves that the homogeneous Besov norm is a homogeneous Hölder norm

$$\|f\|^*_{s,\infty,\infty} = \sup_{j \geqslant J,n} 2^{-j(\alpha+1/2)} |\langle f, \psi_{j,n}\rangle| \sim \|f\|_{\tilde{\mathbf{C}}^s} \tag{9.45}$$

and the corresponding space $\tilde{\mathbf{B}}^s_{\infty,\infty}[0,1]$ is the homogeneous Hölder space of functions that are uniformly Lipschitz s on $[0,1]$.

For $\beta < 2$, functions in $\mathbf{B}^s_{\beta,\gamma}[0,1]$ are not necessarily uniformly regular. The adaptivity of non-linear approximations then improves significantly the decay rate of the error. In particular, if $p = \beta = \gamma$ and $s = 1/2 + 1/p$, then the Besov norm is a simple $\boldsymbol{\ell}^p$ norm:

$$\|f\|_{s,\beta,\gamma} = \left(\sum_{j=-\infty}^{J+1} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^p \right)^{1/p} .$$

Theorem 9.10 proves that if $f \in \mathbf{B}^s_{\beta,\gamma}[0,1]$, then $\varepsilon_n(M,f) = o(M^{1-2/p})$. The smaller p, the faster the error decay. The proof of Theorem 9.12 shows that although f may be discontinuous, if the number of discontinuities is finite and if f is uniformly Lipschitz α between these discontinuities, then its sorted wavelet coefficients satisfy $|f^r_{\mathcal{B}}[k]| = O(k^{-\alpha-1/2})$, so $f \in \mathbf{B}^s_{\beta,\gamma}[0,1]$ for $1/p < \alpha + 1/2$. This shows that these spaces include functions that are not s times differentiable at all points. The linear approximation error $\varepsilon_l(M,f)$ for $f \in \mathbf{B}^s_{\beta,\gamma}[0,1]$ can decrease arbitrarily slowly because the M wavelet coefficients at the largest scales may be arbitrarily small. A non-linear approximation is much more efficient in these spaces.

Bounded Variation Bounded variation functions are important examples of signals for which a non-linear approximation yields a much smaller error than a linear approximation. The total variation norm is defined in (2.57) by

$$\|f\|_V = \int_0^1 |f'(t)|\, dt \ .$$

The derivative f' must be understood in the sense of distributions, in order to include discontinuous functions. To compute the linear and non-linear wavelet approximation error for bounded variation signals, the following theorem computes an upper and a lower bound of $\|f\|_V$ from the modulus of wavelet coefficients.

Theorem 9.13. *Consider a wavelet basis constructed with ψ such that $\|\psi\|_V < +\infty$. There exist $A, B > 0$ such that for all $f \in \mathbf{L^2}[0,1]$*

$$\|f\|_V \leqslant B \sum_{j=-\infty}^{J+1} \sum_{n=0}^{2^{-j}-1} 2^{-j/2}\, |\langle f, \psi_{j,n}\rangle| = B\, \|f\|_{1,1,1} \ , \tag{9.46}$$

and

$$\|f\|_V \geqslant A \sup_{j \leqslant J} \left(\sum_{n=0}^{2^{-j}-1} 2^{-j/2}\, |\langle f, \psi_{j,n}\rangle| \right) = A\, \|f\|_{1,1,\infty}^{*} \ . \tag{9.47}$$

Proof. By decomposing f in the wavelet basis

$$f = \sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} \langle f, \psi_{j,n}\rangle\, \psi_{j,n} + \sum_{n=0}^{2^{-J}-1} \langle f, \phi_{J,n}\rangle\, \phi_{J,n},$$

we get

$$\|f\|_V \leqslant \sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|\, \|\psi_{j,n}\|_V + \sum_{n=0}^{2^{-J}-1} |\langle f, \phi_{J,n}\rangle|\, \|\phi_{J,n}\|_V \ . \tag{9.48}$$

The wavelet basis includes wavelets whose support are inside $(0,1)$ and border wavelets, which are obtained by dilating and translating a finite number of mother wavelets. To simplify notations we write the basis as if there were a single mother wavelet: $\psi_{j,n}(t) = 2^{-j/2}\psi(2^{-j}t-n)$. Hence, we verify with a change of variable that

$$\|\psi_{j,n}\|_V = \int_0^1 2^{-j/2}\, 2^{-j}\, |\psi'(2^{-j}t-n)|\, dt = 2^{-j/2}\, \|\psi\|_V \ .$$

Since $\phi_{J,n}(t) = 2^{-J/2}\phi(2^{-J}t-n)$ we also prove that $\|\phi_{J,n}\|_V = 2^{-J/2}\, \|\phi\|_V$. The inequality (9.46) is thus derived from (9.48).

Since ψ has at least one vanishing moment, its primitive θ is a function with the same support, which we suppose included in $[-K/2, K/2]$. To prove (9.47), for $j \leqslant J$ we make an integration by parts:

$$\begin{aligned}
\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle| &= \sum_{n=0}^{2^{-j}-1} \left| \int_0^1 f(t)\, 2^{-j/2}\psi(2^{-j}t-n)\, dt \right| \\
&= \sum_{n=0}^{2^{-j}-1} \left| \int_0^1 f'(t)\, 2^{j/2}\theta(2^{-j}t-n)\, dt \right| \\
&\leqslant 2^{j/2} \sum_{n=0}^{2^{-j}-1} \int_0^1 |f'(t)|\, |\theta(2^{-j}t-n)|\, dt \ .
\end{aligned}$$

Since θ has a support in $[-K/2, K/2]$,

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle| \leqslant 2^{j/2}\, K \sup_{t\in\mathbb{R}} |\theta(t)| \int_0^1 |f'(t)|\, dt \leqslant A^{-1} 2^{j/2} \|f\|_V \ . \tag{9.49}$$

This inequality proves (9.47). ■ ■

This theorem shows that the total variation norm is bounded by two Besov norms:

$$A\,\|f\|^*_{1,1,\infty} \leqslant \|f\|_V \leqslant B\,\|f\|_{1,1,1}\ .$$

The lower-bound is a homogeneous norm because the addition of a constant to f does not modify $\|f\|_V$. The space $\mathbf{BV}[0,1]$ of bounded variation functions is therefore embedded in the corresponding Besov spaces:

$$\mathbf{B}^1_{1,1}[0,1] \subset \mathbf{BV}[0,1] \subset \tilde{\mathbf{B}}^1_{1,\infty}[0,1]\ .$$

The following theorem derives linear and non-linear wavelet approximation errors for bounded variation signals.

Theorem 9.14. *For all $f \in \mathbf{BV}[0,1]$ and $M > 2q$*

$$\varepsilon_l(M,f) = O(\|f\|_V^2\, M^{-1})\,, \tag{9.50}$$

and

$$\varepsilon_n(M,f) = O(\|f\|_V^2\, M^{-2})\,. \tag{9.51}$$

Proof. Section 7.5.3 shows a wavelet basis of $\mathbf{L^2}[0,1]$ with boundary wavelets keeping their q vanishing moments has $2q$ different boundary wavelets and scaling functions, so the largest wavelet scale can be $2^J = (2q)^{-1}$ but not smaller.

There are 2^{-j} wavelet coefficients at a scale 2^j, so for any $L \leqslant J$, there are 2^{-L} wavelet and scaling coefficients at scales $2^j > 2^L$. The resulting linear wavelet approximation error is

$$\varepsilon_l(2^{-L},f) = \sum_{j=-\infty}^{L} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^2. \tag{9.52}$$

We showed in (9.47) that

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle| \leqslant A^{-1}\, 2^{j/2}\, \|f\|_V$$

and hence that

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^2 \leqslant A^{-2}\, 2^{j}\, \|f\|_V^2\ .$$

It results from (9.52) that

$$\varepsilon_l(2^{-L},f) \leqslant 2A^{-2}\, 2^L\, \|f\|_V^2\ .$$

Setting $M = 2^{-L}$, we derive (9.50).

Let us now prove the non-linear approximation error bound (9.51). Let $f^r_{\mathcal{B}}[k]$ be the wavelet coefficient of rank k, excluding all the scaling coefficients $\langle f, \phi_{J,n}\rangle$, since we cannot control their value with $\|f\|_V$. We first show that there exists B_0 such that for all $f \in \mathbf{BV}[0,1]$

$$|f^r_{\mathcal{B}}[k]| \leqslant B_0\, \|f\|_V\, k^{-3/2}. \tag{9.53}$$

To take into account the fact that (9.53) does not apply to the 2^J scaling coefficients $\langle f, \phi_{J,n}\rangle$, an upper bound of $\varepsilon_n(M,f)$ is obtained by selecting the 2^J scaling coefficients plus the $M - 2^J$ biggest wavelet coefficients and hence

$$\varepsilon_n(M,f) \leqslant \sum_{k=M-2^J+1}^{+\infty} |f^r_{\mathcal{B}}[k]|^2\ . \tag{9.54}$$

For $M > 2q = 2^{-J}$, inserting (9.53) in (9.54) proves (9.51).

The upper bound (9.53) is proved by computing an upper bound of the number of coefficients larger than an arbitrary threshold T. At scale 2^j, we denote by $f^r_{\mathcal{B}}[j,k]$ the coefficient of rank k among $\{\langle f, \psi_{j,n}\rangle\}_{0\leqslant n\leqslant 2^{-j}}$. The inequality (9.49) proves that for all $j \leqslant J$

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle| \leqslant A^{-1}\, 2^{j/2}\, \|f\|_V.$$

It thus follows from (9.39) that

$$|f^r_{\mathcal{B}}[j,k]| \leqslant A^{-1}\, 2^{j/2}\, \|f\|_V\, k^{-1} = C\, 2^{j/2}\, k^{-1}\ .$$

At a scale 2^j, the number k_j of coefficients larger than T thus satisfies

$$k_j \leqslant \min(2^{-j}\,,\,2^{j/2}\,C\,T^{-1})\,.$$

The total number k of coefficients larger than T is

$$\begin{aligned} k &= \sum_{j=-\infty}^{J} k_j \leqslant \sum_{2^j \geqslant (C^{-1}T)^{2/3}} 2^{-j} + \sum_{2^j > (C^{-1}T)^{2/3}} 2^{j/2} C T^{-1} \\ &\leqslant 6\,(CT^{-1})^{2/3}. \end{aligned}$$

By choosing $T = |f_{\mathcal{B}}^r[k]|$, since $C = A^{-1}\,\|f\|_V$, we get

$$|f_{\mathcal{B}}^r[k]| \leqslant 6^{3/2}\,A^{-1}\,\|f\|_V\,k^{-3/2}\,,$$

which proves (9.53). ■ ■

The asymptotic decay rate of linear and non-linear approximation errors in Theorem 9.14 can not be improved. If $f \in \mathbf{BV}[0,1]$ has discontinuities, then $\varepsilon_l(M,f)$ decays like M^{-1} and $\varepsilon_n(M,f)$ decays like M^{-2}. One can also prove [206] that this error decay rate for all bounded variation functions cannot be improved by any type of non-linear approximation scheme. In this sense, wavelets are optimal for approximating bounded variation functions.

9.3 Sparse Image Representations

Approximation of images is more complex than one-dimensional signals, because singularities often belong to geometrical structures such as edges or textures. Non-linear wavelet approximation define adaptive approximations grids which are numerically highly effective. These approximations are optimal for bounded variation images, but not for images having edges that are geometrically regular. Section 9.3.2 introduces a piecewise regular image model with regular edges, and studies adaptive triangulation approximations. Section 9.3.3 proves that curvelet frames yield asymptotically optimal approximation errors for such piecewise $\mathbf{C}^2$ regular images.

9.3.1 Wavelet Image Approximations

Linear and non-linear approximations of functions in $\mathbf{L^2}[0,1]^d$ can be calculated in separable wavelet bases. We concentrate on the two-dimensional case for image processing, and compute approximation errors for bounded variation images.

Section 7.7.4 constructs a separable wavelet basis of $\mathbf{L^2}[0,1]^2$ from a wavelet basis of $\mathbf{L^2}[0,1]$, with separable products of wavelets and scaling functions. We suppose that all wavelets of the basis of $\mathbf{L^2}[0,1]$ are $\mathbf{C}^q$ with q vanishing moments. The wavelet basis of $\mathbf{L^2}[0,1]^2$ includes three mother wavelets $\{\psi^l\}_{1\leqslant l\leqslant 3}$ that are dilated by 2^j and translated over a square grid of interval 2^j in $[0,1]^2$. Modulo modifications near the borders, these wavelets can be written

$$\psi^l_{j,n}(x) = \frac{1}{2^j}\,\psi^l\left(\frac{x_1 - 2^j n_1}{2^j}\,,\,\frac{x_2 - 2^j n_2}{2^j}\right)\,. \tag{9.55}$$

They have q vanishing moments in the sense that they are orthogonal to two-dimensional polynomials of degree strictly smaller than q. If we limit the scales to $2^j \leqslant 2^J$, we must complete the wavelet family with two-dimensional scaling functions

$$\phi^2_{J,n}(x) = \frac{1}{2^J}\,\phi^2\left(\frac{x_1 - 2^J n_1}{2^J}\,,\,\frac{x_2 - 2^J n_2}{2^J}\right)$$

to obtain orthonormal basis of $\mathbf{L^2}[0,1]^2$

$$\mathcal{B} = \left(\{\phi^2_{J,n}\}_{2^J n\in[0,1)^2}\ \cup\ \{\psi^l_{j,n}\}_{j\leqslant J\,,\,2^j n\in[0,1)^2\,,\,1\leqslant l\leqslant 3}\right)\,.$$

Linear Image Approximation The linear discretization of an analog image in $\mathbf{L^2}[0,1]^2$ can be defined by $N = 2^{-2L}$ samples $\{\langle f, \phi^2_{L,n}\rangle\}_{2^L n\in[0,1]^2}$, which characterize the orthogonal projection of f in the approximation space $\mathbf{V}_L$. The precision of such linear approximations depends upon the uniform image regularity.

Local image regularity can be measured with Lipschitz exponents. A function f is uniformly Lipschitz α over a domain $\Omega \subset \mathbb{R}^2$ if there exists $K > 0$, such that for any $v \in \Omega$ one can find a polynomial p_v of degree $\lfloor \alpha \rfloor$ such that

$$\forall x \in \Omega \ , \quad |f(x) - p_v(x)| \leqslant K\, |x - v|^\alpha \ . \tag{9.56}$$

The infimum of the K which satisfy (9.56) is the *homogeneous* Hölder α norm $\|f\|_{\tilde{\mathbf{C}}^\alpha}$. The Hölder α norm of f also imposes that f is bounded: $\|f\|_{\mathbf{C}^\alpha} = \|f\|_{\tilde{\mathbf{C}}^\alpha} + \|f\|_\infty$. We write $\mathbf{C}^\alpha[0,1]^2$ the Hölder space of functions for which $\|f\|_{\mathbf{C}^\alpha} < +\infty$. Similarly to Theorem 9.15 in one dimension, the following theorem computes the linear approximation error decay of such functions, with $\mathbf{C}^q$ wavelets having q vanishing moments.

Theorem 9.15. *There exists $B \geqslant A > 0$ such that*

$$A\,\|f\|_{\tilde{\mathbf{C}}^\alpha} \leqslant \sup_{1\leqslant l\leqslant 3, j\geqslant J, 0\leqslant n<2^{-j}} 2^{-j(\alpha+1)}\, |\langle f, \psi^l_{j,n}\rangle| \leqslant B\, \|f\|_{\tilde{\mathbf{C}}^\alpha} \ . \tag{9.57}$$

Proof. The proof is essentially the same as the proof of Theorem 9.15 in one dimension. We shall only prove the right inequality. If f is uniformly Lipschitz α on the support of $\psi^l_{j,n}$, since $\psi^l_{j,n}$ is orthogonal the polynomial $p_{2^j n}$ approximating f at $v = 2^j n$, we get

$$\begin{aligned} |\langle f, \psi^l_{j,n}\rangle| &= |\langle f - p_{2^j n}, \psi^l_{j,n}\rangle| \leqslant \|f\|_{\tilde{\mathbf{C}}^\alpha} \iint 2^{-j}\, |\psi^l(2^{-j}(x - 2^j n))|\, |x - 2^j n|^\alpha\, dx \\ &\leqslant \|f\|_{\tilde{\mathbf{C}}^\alpha}\, 2^{(\alpha+1)j} \iint |\psi^l(x)|\, |x|^\alpha\, dx \ , \end{aligned}$$

which proves the right inequality of (9.57). The wavelet regularity is not used to prove this inequality. The left inequality requires that the wavelets are $\mathbf{C}^q$.

The following theorem computes the linear wavelet approximation error decay for images that are uniformly Lipschitz α. It requires that wavelets have q vanishing moments, but no regularity condition is needed.

Theorem 9.16. *If f is uniformly Lipschitz $0 < \alpha \leqslant q$ over $[0,1]^2$ then $\varepsilon_l(N, f) = O(\|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, N^{-\alpha})$.*

Proof. There are $3\,2^{-2j}$ wavelet coefficients at a scale 2^j and 2^{-2k} wavelet coefficients and scaling coefficients at scales $2^j > 2^k$. The right inequality of (9.57) proves that

$$|\langle f, \psi^l_{j,n}\rangle| \leqslant B\, \|f\|_{\tilde{\mathbf{C}}^\alpha}\, 2^{(\alpha+1)j} \ .$$

As a result

$$\varepsilon_l(2^{-2k}, f) = \sum_{j=-\infty}^{k} \sum_{l=1}^{3} \sum_{2^j n\in[0,1]^2} |\langle f, \psi_{j,n}\rangle|^2 \leqslant 3B^2\, \|f\|^2_{\tilde{\mathbf{C}}^\alpha} \sum_{j=-\infty}^{k} 2^{-2j}\, 2^{j(2\alpha+2)} = \frac{3B^2\, \|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, 2^{2\alpha k}}{1 - 2^{-2\alpha}} .$$

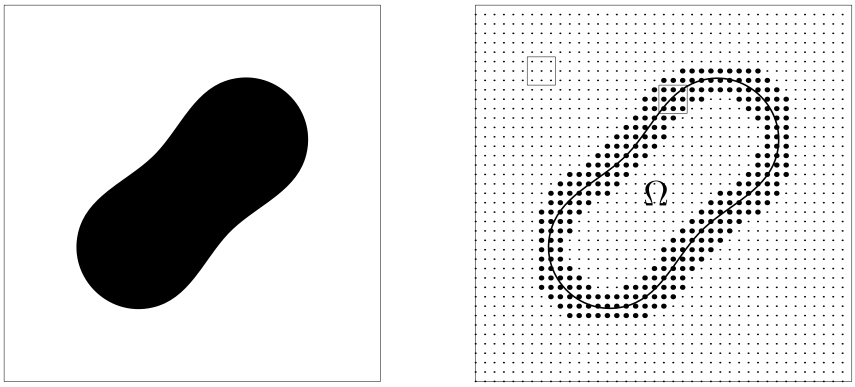
For $2k = -\lfloor \log_2 N \rfloor$, we derive that $\varepsilon_l(N, f) = O(\|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, 2^{2\alpha k}) = O(\|f\|^2_{\tilde{\mathbf{C}}^\alpha}\, N^{-\alpha})$. ■

One can prove that this decay rate is optimal in the sense that no approximation scheme can improve the decay rate $N^{-\alpha}$ over all uniformly Lipschitz α functions [19].

Non-Linear Approximation of Piecewise Regular Images If an image has singularities, then linear wavelet approximations introduce large errors. In one dimension, an isolated discontinuity creates a constant number of large wavelet coefficients at each scales. As a result, non-linear wavelet approximations are marginally influenced by a finite number of isolated singularities. Theorem 9.12 proves that if f is uniformly Lipschitz α between these singularities, then the asymptotic error decay behaves as if there was no singularity. In two dimensions, if f is uniformly Lipschitz α then Theorem

9.16 proves that the linear approximation error from M wavelets satisfies $\varepsilon_l(M, f) = O(M^{-\alpha})$. We can thus hope that piece-wise regular images yield a non-linear approximation error having the same asymptotic decay. Regretfully, this is wrong.

A piecewise regular image has discontinuities along curves of dimension 1, which create a non-negligible number of high amplitude wavelet coefficients. As a result even though the function may be infinitely differentiable between discontinuities, the non-linear approximation error $\varepsilon_n(M, f)$ decays only like M^{-1}. This is illustrated on an example.



(a) (b)

Figure 9.3: (a): Image $f = \mathbf{1}_\Omega$. (b): At the scale 2^j, the wavelets $\psi^l_{j,n}$ have a square support of width proportional to 2^j. This support is translated on a grid of interval 2^j, which is indicated by the smaller dots. The darker dots correspond to wavelets whose support intersects the frontier of Ω, for which $\langle f, \psi^l_{j,n}\rangle \neq 0$.

Suppose that $f = C\,\mathbf{1}_\Omega$ is the indicator function of a set Ω whose border $\partial\Omega$ has a finite length, as shown in Figure 9.3. If the support of $\psi^l_{j,n}$ does not intersect the border $\partial\Omega$, then $\langle f, \psi^l_{j,n}\rangle = 0$ because f is constant over the support of $\psi^l_{j,n}$. The wavelets $\psi^l_{j,n}$ have a square support of size proportional to 2^j, which is translated on a grid of interval 2^j. Since $\partial\Omega$ has a finite length L, there are on the order of $L\,2^{-j}$ wavelets whose support intersects $\partial\Omega$. Figure 9.3(b) illustrates the position of these coefficients.

Since f is bounded, is result from (9.57) (for $\alpha = 0$) that $|\langle f, \psi^l_{j,n}\rangle| = O(C\,2^j)$. Along the border, wavelet coefficients typically have an amplitude $|\langle f, \psi^l_{j,n}\rangle| \sim C\,2^j$. The M largest coefficients are thus typically at scales $2^j \geqslant L/M$. Selecting these M largest coefficients yields an error

$$\varepsilon_n(M, f) \sim \sum_{j=-\infty}^{\log_2(L/M)-1} L 2^{-j}\, C^2\, 2^{2j} = (C\,L)^2\, M^{-1}\ . \tag{9.58}$$

The large number of wavelet coefficients produced by the edges of f thus limit the error decay to M^{-1}.

Two questions then arise. Is the class of image having a wavelet approximation error that decays like M^{-1} sufficiently large to incorporate interesting image models ? The next paragraphs proves that this class includes all bounded variation images. The second question is to understand whether it is possible to find sparse signal representations, that are better than wavelets to approximate images having regular edges. This second question is addressed in Section 9.3.2.

Bounded Variation Images Bounded variation functions provide good models for large class of images which do not have irregular textures. The total variation of f is defined in Section 2.3.3 by

$$\|f\|_V = \int_0^1 \int_0^1 |\vec{\nabla} f(x)|\, dx\ . \tag{9.59}$$

The partial derivatives of $\vec{\nabla} f$ must be taken in the general sense of distributions in order to include discontinuous functions. Let $\partial\Omega_t$ be the level set defined as the boundary of

$$\Omega_t = \{x \in \mathbb{R}^2 \quad : \quad f(x) > t\} \ .$$

Theorem 2.9 proves that the total variation depends on the length $H^1(\partial\Omega_t)$ of level sets:

$$\int_0^1 \int_0^1 |\vec{\nabla} f(x)|\, dx = \int_{-\infty}^{+\infty} H^1(\partial\Omega_t)\, dt. \tag{9.60}$$

It results that if $f = C\, \mathbf{1}_\Omega$ then $\|f\|_V = C\, L$, where L is the length of the boundary of Ω. Indicator function of sets thus have a bounded total variation, which is proportional to the length of their "edges".

Linear and non-linear approximation errors of bounded variation images are computed by evaluating the decay of their wavelet coefficients across scales. We denote by $f^r_{\mathcal{B}}[k]$ the rank k wavelet coefficient of f, without including the 2^{2J} scaling coefficients $\langle f, \phi^2_{J,n}\rangle$. The following theorem gives upper and lower bounds on $\|f\|_V$ from wavelet coefficients. Wavelets are supposed to have a compact support and need only 1 vanishing moment.

Theorem 9.17 (Cohen, DeVore, Pertrushev, Xu)**.** *There exist $A, B_1, B_2 > 0$ such that if $\|f\|_V < +\infty$ then*

$$\sum_{j=-\infty}^{J} \sum_{l=1}^{3} \sum_{2^j n \in [0,1]^2} |\langle f, \psi^l_{j,n}\rangle| + \sum_{2^J n \in [0,1]^2} |\langle f, \phi^2_{J,n}\rangle| \geqslant A\, \|f\|_V \ , \tag{9.61}$$

$$\sup_{\substack{-\infty < j \leqslant J \\ 1 \leqslant l \leqslant 3}} \left(\sum_{2^j n \in [0,1]^2} |\langle f, \psi^l_{j,n}\rangle| \right) \leqslant B_1\, \|f\|_V \tag{9.62}$$

and

$$|f^r_{\mathcal{B}}[k]| \leqslant B_2\, \|f\|_V\ k^{-1} \ . \tag{9.63}$$

Proof. In 2D, a wavelet total variation does not depend upon scale and position. Indeed, with a change of variable $x' = 2^{-j}x - n$, we get

$$\|\psi^l_{j,n}\|_V = \iint |\vec{\nabla}\psi^l_{j,n}(x)|\, dx = \iint |\vec{\nabla}\psi^l(x')|\, dx' = \|\psi^l\|_V \ .$$

Similarly $\|\phi^2_{J,n}\|_V = \|\phi^2\|_V$. The inequalities (9.61) and (9.62) are proved with the same derivation steps as in Theorem 9.13 for one-dimensional bounded variation functions.

The proof of (9.63) is technical and can be found in [174]. The inequality (9.62) proves that wavelet coefficients have a bounded $\mathbf{l^1}$ norm at each scale 2^j. It results from (9.39) in Theorem 9.10 that ranked wavelet coefficients at each scale 2^j have a decay bounded by $B_1\, \|f\|_V\, k^{-1}$. The inequality (9.63) is finer since it applies to the ranking of wavelet coefficients at all scales. ■ ■

Lena is an example of finite resolution approximation of a bounded variation image. Figure 9.4 shows that its sorted wavelet coefficients $\log_2 |f^r_{\mathcal{B}}[k]|$ decays with a slope that reaches -1 as $\log_2 k$ increases, which verifies that $|f^r_{\mathcal{B}}[k]| = O(k^{-1})$. In contrast, the Mandrill image shown in Figure 10.7 does not have a bounded total variation because of the fur texture. As a consequence, $\log_2 |f^r_{\mathcal{B}}[k]|$ decays more slowly, in this case with a slope that reaches -0.65.

A function with finite total variation does not necessarily have a bounded amplitude, but images do have a bounded amplitude. The following theorem incorporates this hypothesis to compute linear approximation errors.

Theorem 9.18. *If $\|f\|_V < +\infty$ and $\|f\|_\infty < +\infty$ then*

$$\varepsilon_l(M, f) = O(\|f\|_V\, \|f\|_\infty\ M^{-1/2}) \tag{9.64}$$

and

$$\varepsilon_n(M, f) = O(\|f\|^2_V\ M^{-1})\, . \tag{9.65}$$

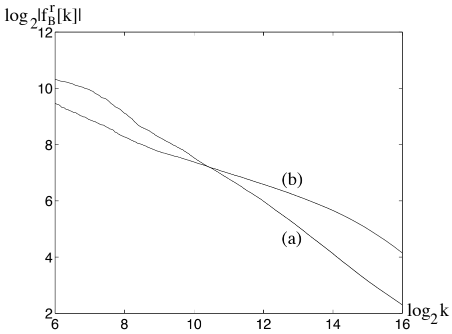
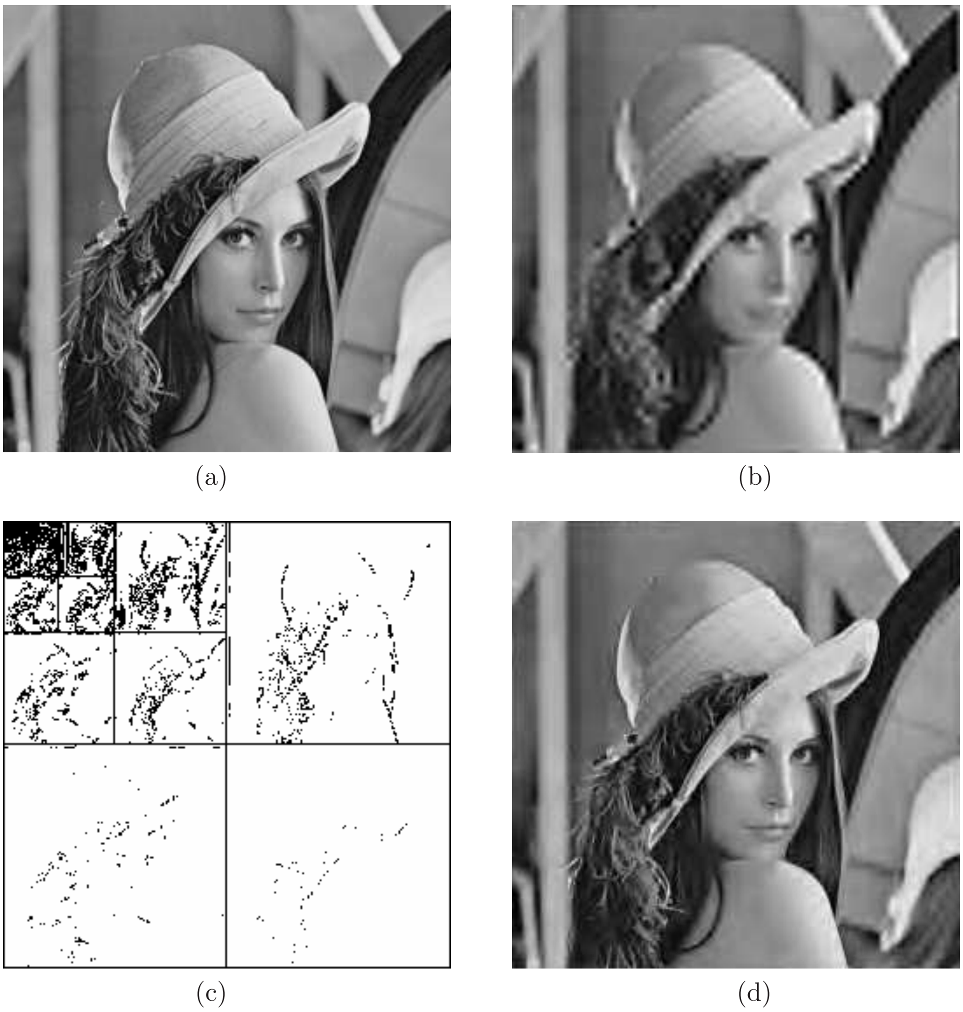


Figure 9.4: Sorted wavelet coefficients $\log_2 |f^r_{\mathcal{B}}[k]|$ as a function of $\log_2 k$ for two images. (a): Lena image shown in Figure 9.5(a). (b): Mandrill image shown in Figure 10.7.



(a) (b) (c) (d)

Figure 9.5: (a): Lena image f of $N = 256^2$ pixels. (b): Linear approximations f_M calculated from the $M = N/16$ Symmlet 4 wavelet coefficients at the largest scales: $\|f - f_M\|/\|f\| = 0.036$. (c): The support of the $M = N/16$ largest amplitude wavelet coefficients are shown in black. (d): Non-linear approximation f_M calculated from the M largest amplitude wavelet coefficients: $\|f - f_M\|/\|f\| = 0.011$.

Proof. The linear approximation error from $M = 2^{-2m}$ wavelets is

$$\varepsilon_l(2^{-2m}, f) = \sum_{j=-\infty}^{m} \sum_{l=1}^{3} \sum_{2^j n \in [0,1]^2} |\langle f, \psi_{j,n}^l \rangle|^2 \ . \tag{9.66}$$

We shall verify that there exists $B > 0$ such that for all j and l

$$\sum_{2^j n \in [0,1]^2} |\langle f, \psi_{j,n}^l \rangle|^2 \leqslant B \, \|f\|_V \, \|f\|_\infty \, 2^j \ . \tag{9.67}$$

Applying this upper bound to the sum (9.66) proves that

$$\varepsilon_l(2^{-2m}, f) \leqslant 6 \, B1 \, \|f\|_V \, \|f\|_\infty \, 2^m \ ,$$

from which (9.64) is derived

The upper bound (9.67) is calculated with (9.62), which shows that there exists $B_2 > 0$ such that for all j and l

$$\sum_{2^j n \in [0,1]^2} |\langle f, \psi_{j,n}^l \rangle| \leqslant B_2 \, \|f\|_V \ . \tag{9.68}$$

The amplitude of a wavelet coefficient can also be bounded:

$$|\langle f, \psi_{j,n}^l \rangle| \leqslant \|f\|_\infty \, \|\psi_{j,n}^l\|_1 = \|f\|_\infty \, 2^j \, \|\psi^l\|_1 \ ,$$

where $\|\psi^l\|_1$ is the $\mathbf{L^1}[0,1]^2$ norm of ψ^l. If $B_3 = \max_{1 \leqslant l \leqslant 3} \|\psi^l\|_1$ this yields

$$|\langle f, \psi_{j,n}^l \rangle| \leqslant B_3 \, 2^j \, \|f\|_\infty \ . \tag{9.69}$$

Since $\sum_n |a_n|^2 \leqslant \sup_n |a_n| \sum_n |a_n|$, we get (9.67) from (9.68) and (9.69).

The non-linear approximation error is a direct consequence of the sorted coefficient decay (9.63). It results from (9.63) and (9.36) that $\varepsilon_n(M, f) = O(\|f\|_V^2 \, M^{-1})$. ■ ■

This theorem shows that non-linear wavelet approximations of bounded variation images can yield much smaller errors than linear approximations. The decay bounds of this theorem are tight in the sense that one can find functions, for which $\varepsilon_l(M, f)$ and $\varepsilon_n(M, f)$ decay respectively like $M^{-1/2}$ and M^{-1}. This is typically the case for bounded variation images including discontinuities along edges, for example $f = C \, \mathbf{1}_\Omega$. Lena in Figure 9.5(a) is another example.

Figure 9.5(b) is a linear approximation calculated with the $M = N/16$ largest scale wavelet coefficients. This approximation produces a uniform blur and creates Gibbs oscillations in the neighborhood of contours. Figure 9.5(c) gives the non-linear approximation support Λ_T of the $M = N/16$ largest wavelet coefficients. Large amplitude coefficients are located where the image intensity varies sharply, in particular along the edges. The resulting non-linear approximation is shown in Figure 9.5(d). The non-linear approximation error is much smaller than the linear approximation error: $\varepsilon_n(M, f) \leqslant \varepsilon_l(M, f)/10$, and the image quality is indeed better. As in one dimension, this non-linear wavelet approximation can be interpreted as an adaptive grid approximation, which refines the approximation resolution near edges and textures by keeping wavelet coefficients at smaller scales.

9.3.2 Geometric Image Models and Adaptive Triangulations

Bounded variation image models correspond to images whose level sets have a finite average length, but it does not imply any geometrical regularity of these level sets. The level sets and "edges" of many images such as Lena are often piecewise regular curves. This geometric regularity can be used to improve the sparsity of image representations.

When an image is uniformly Lipschitz α, Theorem 9.16 proves that wavelet non-linear approximations have an error $\varepsilon_l(M, f) = O(M^{-\alpha})$, which is optimal. However, as soon as the image is discontinuous along an edge, then the error decay rate drops to $\varepsilon_n(M, f) = O(M^{-1})$, because edges create a number of large wavelet coefficients that is proportional to their length. This decay rate is improved by representations taking advantage of edge geometric regularities. We introduce a piecewise regular image model which incorporates the geometric regularity of edges. Approximations of piecewise regular images are studied with adaptive triangulations.

Piecewise $\mathbf{C}^\alpha$ Image Models Piecewise regular image models include edges that are also piecewise regular. These edges are typically occlusion contours of objects in images. The regularity is measured in the sense of uniform Lipschitz regularity, with Hölder norms $\|f\|_{\mathbf{C}^\alpha}$, defined in (9.20) and (9.56) for one and two-dimensional functions. Edges are supposed to be a finite union of curves e_k that are uniformly Lipschitz α in $[0,1]^2$. Between edges, the image is supposed to be uniformly Lipschitz α. To model the blur introduced by the optics or by diffraction phenomena, the image model incorporates a convolution by an unknown regular kernel h_s, whose scale s is a parameter.

Definition 9.1. *A function $f \in \mathbf{L^2}[0,1]^2$ is said to be a piecewise $\mathbf{C}^\alpha$ with a blurring scale $s \geqslant 0$ if $f = \bar{f} \star h_s$ where $\bar{f}$ is uniformly Lipschitz α on $\Omega = [0,1]^2 - \{e_k\}_{1 \leqslant k < K}$. If $s > 0$ then $h_s(x) = s^{-2} h(s^{-1}x)$ where h is a uniformly Lipschitz α kernel with a support in $[-1,1]$, and if $s = 0$ then $h_0 = \delta$. The curves e_k are uniformly Lipschitz α and do not intersect tangentially.*

When the blurring scale $s = 0$ then $f = \bar{f} \star h_0 = \bar{f}$ is typically discontinuous along the edges. When $s > 0$ then $\bar{f} \star h_s$ is blurred and discontinuities along edges are diffused on a neighborhood of size s, as shown in Figure 9.6.

Approximations with Adapted Triangulations Wavelet approximations are inefficient to recover regular edges because it takes many wavelet to cover the edge at each scale, as shown by Figure 9.3. To improve this approximation, it is necessary to use elongated approximation elements. For $\mathbf{C}^2$ piecewise regular images, it is proved that piecewise linear approximations over M adapted triangles can reach the optimal $O(M^{-2})$ error decay, as if the image had no singularities.

For the solution of partial differential equations, the local optimization of anisotropic triangles was introduced by Babuška and A. K. Aziz [88]. Adaptive triangulations are indeed useful in numerical analysis, where shocks or boundary layers require anisotropic refinements of finite elements [75, 409, 433]. A planar triangulation $(\mathcal{V}, \mathcal{T})$ of $[0,1]^2$ is composed of vertices $\mathcal{V} = \{x_i\}_{0 \leqslant i < p}$ and disjoint triangular faces $\mathcal{T} = \{T_k\}_{0 \leqslant k < M}$ that cover the image domain

$$\bigcup_{k=0}^{M-1} T_k = [0,1]^2 \ .$$

Let M be the number of triangles. We consider an image approximation $\tilde{f}_M$ obtained a with linear interpolations of the image values at the vertices. For each $x_i \in \mathcal{V}$, $\tilde{f}_M(x_i) = f(x_i)$, and $\tilde{f}_M(x)$ is linear on each triangular face T_k. A function f is well approximated with M triangles if the shapes of these triangles are optimized to capture the regularity of f.

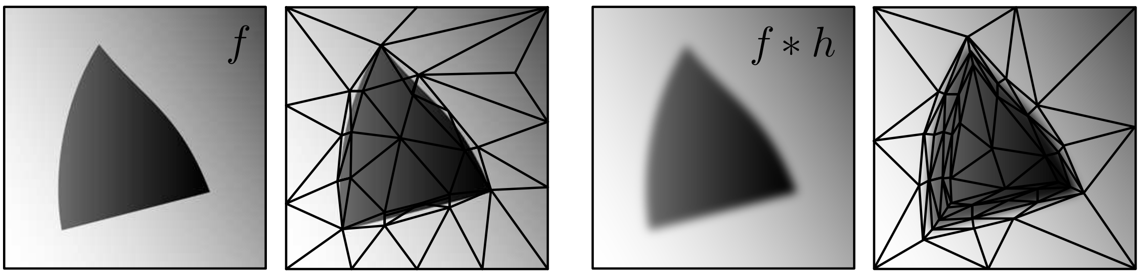


Figure 9.6: Adaptive triangulations for piecewise linear approximations of a piecewise $\mathbf{C}^2$ image, without and with a blurring kernel.

The following theorem proves that adapted triangulations for piecewise $\mathbf{C}^2$ images leads to the optimal decay rate $O(M^{-2})$, by sketching the construction of a well adapted triangulation.

Theorem 9.19. *If f is a piecewise $\mathbf{C}^2$ image, then there exists C such that for any M one can construct a triangulation $(\mathcal{V}, \mathcal{T})$ with M triangles over which the piecewise-linear interpolation $\tilde{f}_M$ satisfies*

$$\|f - \tilde{f}_M\|^2 \leqslant C\, M^{-2}. \tag{9.70}$$

Proof. A sketch of the proof is given. Let us first consider the case where the blurring scale is $s = 0$. Each edge curve e_k is $\mathbf{C}^2$ and can be covered by a band β_k of width ε^2. This band is is a union of straight tubes of width ε^2 and of length ε/C where C is the maximum curvature of the edge curves. Each tube is decomposed in two elongated triangles as illustrated in Figure 9.6. The number of such triangles is $2CL_k\varepsilon^{-1}$ where L_k is length of e_k. Triangle vertices should also be appropriately adjusted at junctions or corners. Let $L = \sum_k L_k$ be the total length of edges and $\beta = \cup_k \beta_k$ be the band covering all edges. This band is divided in $2CL\varepsilon^{-1}$ triangles. Since f is bounded, $\|f - \tilde{f}_M\|^2_{\mathbf{L^2}(\beta)} \leqslant \mathrm{area}(\beta)\,\|f\|^2_\infty \leqslant L\,\varepsilon^2\|f\|^2_\infty$.

The complementary $\beta^c = [0,1]^2 - \beta$ is covered at the center by nearly equilateral triangles of surface of the order of ε, as shown in Figure 9.6. There are $O(\varepsilon^{-1})$ such triangles. Performing this packing requires to use a boundary layer of triangles which connect the large nearly isotropic triangles of width $\varepsilon^{1/2}$ to the anisotropic triangles of size $\sim \varepsilon^2 \times \varepsilon$, along edges. One can verify that such a boundary layer can be constructed with $O(LC\varepsilon^{-1})$ triangles. Since f is $\mathbf{C}^2$ on $\Omega - \beta$, the approximation error of a linear interpolation $\tilde{f}$ over this triangulation is $\|f - \tilde{f}_M\|_{\mathbf{L}^\infty(\beta^c)} = O(\|f\|^2_{\mathbf{C}^2(\beta^c)}\varepsilon^2)$. The total error thus satisfies

$$\|f - \tilde{f}_M\|^2 = O(L\|f\|^2_\infty\,\varepsilon^2 + \|f\|^2_{\mathbf{C}^2}\,\varepsilon^2)$$

with a number of triangles $M = O((C\,L + 1)\,\varepsilon^{-1})$, which verifies (9.70) for $s = 0$.

Suppose now that $s > 0$. According to Definition 9.1, $f = \bar{f} \star h_s$ so edges are diffused into sharp transitions along a tube of width s. Since h is $\mathbf{C}^2$, within this tube f is $\mathbf{C}^2$ but has large amplitude derivatives if s is small. This defines an overall band β of surface $L\,s$ where the derivatives of f are potentially large. The triangulation of the domain $[0,1]^2 - \beta$ can be treated similarly to the case $s = 0$. We thus concentrate on the triangulation of the band β and show that one can find a triangulation with $O(\varepsilon^{-1})$ triangles which yields an error in $O(\varepsilon^2)$ over the band.

The band β has a surface Ls and can thus be covered by $L\,\varepsilon^{-1}$ triangles of surface εs. Let us compute the aspect ratio of these triangles to minimize the resulting error. A blurred piecewise $\mathbf{C}^2$ image $f = \bar{f} \star h_s$ has an anisotropic regularity at a point x close to an edge curve e_k where $\bar{f}$ is discontinuous. Let $\tau_1(x)$ the unit vector which is tangent to e_k at point x, and $\tau_2(x)$ be the perpendicular vector. In the system of coordinates (τ_1, τ_2), for any $u = u_1, \tau_1 + u_2\,\tau_2$ in the neighborhood of x, one can prove that [341, 364]

$$\left|\frac{\partial^{i_1+i_2} f}{\partial u_1^{i_1}\partial u_2^{i_2}}(u)\right| = O(s^{-i_1/2-i_2})\ . \tag{9.71}$$

For s small the derivatives are thus much larger along τ_2 then along τ_1. The error of local linear approximation can be computed with a Taylor decomposition

$$f(x+\Delta) = f(x) + \langle \nabla_x f,\, h\rangle + \frac{1}{2}\langle H_x(f)\,\Delta,\,\Delta\rangle + O(\|\Delta\|^2).$$

where $H_x(f) \in \mathbb{R}^{2\times 2}$ is the symmetric Hessian tensor of second derivatives. Let us decompose $\Delta = \Delta_1\tau_1 + \Delta_2\tau_2$:

$$|f(x+\Delta) - (f(x) + \langle \nabla_x f,\,\Delta\rangle)| = O(s^{-1}|\Delta_1|^2 + s^{-2}|\Delta_2|^2 + s^{-3/2}|\Delta_1|\,|\Delta_2|)\ . \tag{9.72}$$

For a triangle of width and length Δ_1 and Δ_2, we want to minimize the error for a given surface $s\varepsilon \sim \Delta_1\,\Delta_2$, which is obtained with $\Delta_2/\Delta_1 = \sqrt{s}$. It results that $\Delta_1 \sim s^{1/3}\varepsilon^{1/2}$ and $\Delta_2 \sim s^{3/4}\varepsilon^{1/2}$, which yields a linear approximation error (9.72) on this triangle:

$$|f(x+\Delta) - (f(x) + \langle \nabla_x f,\,\Delta\rangle)| = O(s^{-1/2}\varepsilon)\ .$$

This gives

$$\|f - \tilde{f}_M\|^2_{\mathbf{L^2}(\beta)} \leqslant \mathrm{area}(\beta)\,\|f - \tilde{f}_M\|^2_\infty = O(L\varepsilon^2)\ .$$

It proves that the $O(\varepsilon^{-1})$ triangles produce an error of $O(\varepsilon^2)$ on the band β. Since the same result is valid on $[0,1]^2 - \beta$, it yields (9.70). ∎

This theorem proves that an adaptive triangulation can yield an optimal decay rate $O(M^{-2})$ for piecewise $\mathbf{C}^2$ image. The decay rate of the error is independent from the burring scale s, which may be zero or not. However, the triangulation depends upon s and on the edge geometry.

Wherever f is uniformly $\mathbf{C}^2$, it is approximated over large nearly isotropic triangles, of size $O(M^{1/2})$. In the neighborhood of edges, to introduce an error $O(M^{-2})$, triangles must be narrow in the direction of the discontinuity and as long as possible to cover the discontinuity with a

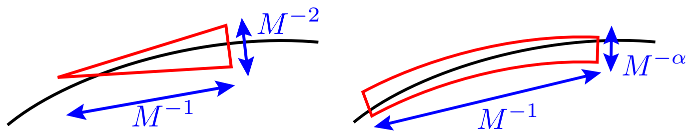


Figure 9.7: Adapted triangle and finite element to approximate a discontinuous function around a singularity curve.

minimum number of triangles. Since edges are $\mathbf{C}^2$, they can be covered with triangles of width and length proportional to M^{-2} and M^{-1}, as illustrated by Figure 9.7.

If the image is blurred at a scale s then the discontinuities are diffused neighborhood of size s where the image has sharp transitions. Theorem 9.19 shows that the tube of width s around each edge should be covered with triangles of width and length proportional to $s^{3/4}\,M^{-1/2}$ and $s^{1/4}\,M^{-1/2}$, as illustrated by Figure 9.8.

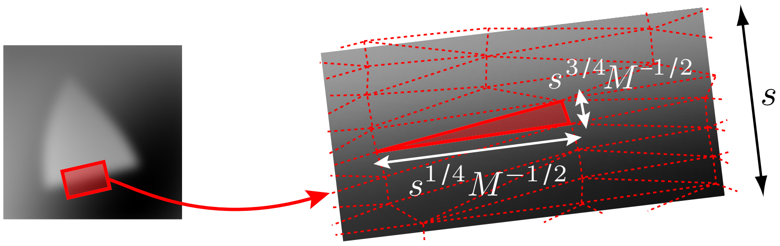


Figure 9.8: Aspect ratio of triangles for the approximation of a blurred contour.

Algorithms to Design Adapted Triangulations It is difficult to adapt the triangulation according to Theorem 9.19 because the geometry of edges and the blurring scale s are unknown. There is currently no adaptive triangulation algorithm that guarantees finding an approximation with an error decay of $O(M^{-2})$ for all piecewise $\mathbf{C}^2$ images. Most algorithms use greedy strategies that iteratively construct a triangulation by progressively reducing the approximation error. Some algorithms progressively increase the number of triangles, while others decimate a dense triangulation.

Delaunay refinement algorithms, introduced by Ruppert [420] and Chew [159], proceed by iteratively inserting points to improve the triangulation precision. Each point is added at a circumcenter of one triangle, and is chosen to reduce as much as possible the approximation error. These algorithms control the shape of triangles and are used to compute isotropic triangulations of images where the size of the triangles varies to match the local image regularity. Extensions of these vertex insertion methods capture anisotropic regularity by locally modifying the notion of circumcenter [336] or with a local optimization of the vertex location [75].

Triangulation thinning algorithms start with a dense triangulation of the domain and progressively remove either a vertex, an edge or a face to increase as slowly as possible the approximation error until M triangles remain [237, 304, 267]. These methods do not control the shape of the resulting triangles, but can create effective anisotropic approximation meshes. They have been used for image compression [204, 238].

These adaptive triangulation are most often used to to mesh the interior of a 2D or a 3D domain having a complex boundary, or to mesh a 2D surface embedded in 3D space.

Approximation of Piecewise $\mathbf{C}^\alpha$ Images Adaptive triangulations could be generalized to higher order approximations of piecewise $\mathbf{C}^\alpha$ images, to obtain an error in $O(M^{-\alpha})$ with M finite elements, for $\alpha > 2$. This would require computing polynomial approximations of order $p = \lceil\alpha\rceil - 1$ on each finite elements, and the support of these finite element should also approximate the geometry of edges at an p. To produce an error $\|f - f_M\|^2 = O(M^{-\alpha})$, it is indeed necessary to cover edge curves with $O(M)$ finite elements of width $O(M^{-\alpha})$, as illustrated in Figure 9.7. However, this gets extremely complicated and has never been implemented numerically. Section 12.2.4 introduce bandlet approximations, which reach the $O(M^{-\alpha})$ error decay for piecewise $\mathbf{C}^\alpha$ images, by choosing a best basis amount a dictionary of orthonormal bases.

9.3.3 Curvelet Approximations

Candès and Donoho [134] showed that a simple thresholding of curvelet coefficients yields nearly optimal approximations of piecewise $\mathbf{C}^2$ images, as opposed to a complex triangulation optimization.

Section 5.5.2 introduces curvelets, defined with a translation, rotation and anisotropic stretching

$$c^\theta_{2^j,u}(x_1,x_2) = c_{2^j}(R_\theta(x-u)) \quad \text{where} \quad c_{2^j}(x_1,x_2) \approx 2^{-3j/4}\, c(2^{-j/2}x_1, 2^{-j}x_2), \tag{9.73}$$

where R_θ is the planar rotation of angle θ. A curvelet $c^\theta_{j,m}$ is elongated in the direction θ, with a *width* proportional to its *length*^{2}. This parabolic scaling corresponds to the scaling of triangles used by Theorem 9.19 to approximate a discontinuous image along $\mathbf{C}^2$ edges.

A tight frame $\mathcal{D} = \{c^\theta_{j,m}\}_{j,m,\theta}$ of curvelets $c^\theta_{j,m}(x) = c^\theta_{2^j}(x - u^{(j,\theta)}_m)$, is obtained with $2^{-\lfloor j/2 \rfloor + 2}$ equispaced angles θ at each scale 2^j, and a translation grid defined by

$$\forall\, m = (m_1, m_2) \in \mathbb{Z}^2 \quad , \quad u^{(j,\theta)}_m = R_\theta(2^{j/2}m_1, 2^j m_2)\ . \tag{9.74}$$

This construction yields a tight frame of $\mathbf{L^2}[0,1]^2$, with periodic boundary conditions [134].

An M-term thresholding curvelet approximation is defined by

$$f_M = \sum_{(\theta,j,m)\in\Lambda_T} \langle f,\, c^\theta_{j,m}\rangle\, c^\theta_{j,m} \quad \text{with} \quad \Lambda_T = \big\{(j,\theta,m)\ :\ |\langle f,\, c^\theta_{j,m}\rangle| > T\big\}$$

where $M = |\Lambda_T|$ is the number of approximation curvelets. Since curvelets define a tight frames with a frame bound $A > 0$

$$\|f - f_M\|^2 \leqslant A^{-1} \sum_{(\theta,j,m)\notin\Lambda_T} |\langle f,\, c^\theta_{j,m}\rangle|^2\ . \tag{9.75}$$

Although the frame is tight, this is not an equality because the thresholded curvelet coefficients of f are not the curvelet coefficients of f_M, due to the frame redundancy. The following theorem shows that a thresholding curvelets approximation of a piecewise $\mathbf{C}^2$ image has an error whose asymptotic decay is nearly optimal.

Theorem 9.20 (Candès, Donoho)**.** *Let f be a piecewise $\mathbf{C}^2$ image. An M-term curvelet approximation satisfies*

$$\|f - f_M\|^2 = O(M^{-2}(\log M)^3). \tag{9.76}$$

Proof. The detailed proof can be found in [134]. We give the main ideas by analyzing how curvelets atoms interact with regular parts and with edges in a piecewise regular image. The burring $f = \bar{f} * h_s$ is nearly equivalent to translating curvelet coefficients from a scale 2^j to a scale $2^j + s$, and thus does not introduce any difficulty in the approximation. In the following, we suppose that $s = 0$.

The approximation error (9.76) is computed with an upper bound on the energy of curvelet coefficients (9.75). This sum is divided in three sets of curvelet coefficients. Type I curvelets have a support mostly concentrated where the image is uniformly $\mathbf{C}^2$. These coefficients are small. Type II curvelets are in a neighborhood of an edge, and are elongated along the direction of the tangent to the edge. These coefficients are large. Type III curvelets are in a neighborhood of an edge, with an angle different from the tangent angle. These coefficients get quickly small as the difference of angle increases. An upper bound of the error is computed by selecting the largest $M/3$ curvelet coefficients for each type of curvelet coefficient and by computing the energy of the left over curvelet coefficients for each type of coefficient.

A type I curvelet is located at a position $u^{(j,\theta)}_m$ at a distance larger than $K\, 2^{j/2}$ from edges. Since curvelets have vanishing moments, keeping type I curvelets at scales larger than 2^j is equivalent to implementing a linear wavelet approximation at a scale 2^j. Theorem 9.16 shows that for linear wavelet approximations of $\mathbf{C}^2$ images, keeping $M/3$ larger scale coefficients yields an error that decreases like $O(M^{-2})$.

For type II curvelets in the neighborhood of an edge, whose angles θ are aligned with the local direction of the tangent to the edge, the sampling interval in this direction is also $2^{j/2}$. There are thus $O(L2^{-j/2})$ type II curvelets along the edge curves of length L in the image. These curvelet coefficient are bounded:

$$|\langle f,\, c^{\theta}_{j,m}\rangle| \leqslant \|f\|_\infty \|c_{2^j}\|_1 = \|f\|_\infty O(2^{3j/4})$$

because $c_{2^j}(x_1,x_2) \approx 2^{-3j/4}\, c(2^{-j}x_1, 2^{-j/2}x_2)$. If we keep the $M/3$ larger scales type II curvelets, since there are $O(L2^{-j/2})$ such curvelets at each scale, it amounts to keep them at scales above $2^l = O(L^2\, M^{-2})$. The left over type II curvelet have an energy $O(L\, 2^{-l/2})\, \|f\|^2_\infty O(2^{3l/2}) = O(M^{-2})$.

Type III curvelet coefficients $|\langle f,\, c^{\theta}_{j,m}\rangle|$ are located in the neighborhood of an edge with an angle θ that deviate from the local orientation θ_0 of edge tangent. The coefficients have a fast decay as $|\theta-\theta_0|$ increases. This is shown in the Fourier domain, by verifying that the Fourier domain where the energy of the localized edge patch is concentrated becomes increasingly disjoint from the Fourier support of $\hat{c}^{\theta}_{j,m}$ as $|\theta-\theta_0|$ increases. The error analysis of these curvelet coefficients is the most technical aspect of the proof. One can prove that selecting the $M/3$ type III largest curvelets yield a an error among type III curvelets that is $O(M^{-2}(\log_2 M)^3)$ [134]. This error dominates the overall approximation error and yields (9.76).

The curvelet approximation rate is optimal up to a $(\log_2 M)^3$ factor. The beauty of this result comes from its simplicity. Unlike an optimal triangulation that must adapt the geometry of each element, curvelets define a fixed frame whose coefficients are selected by a simple thresholding. However, this simplicity has a downside. Curvelets approximation are optimal for piecewise $\mathbf{C}^\alpha$ images for $\alpha = 2$, but they are not optimal if $\alpha > 2$ or if $\alpha < 1$. In particular, curvelets are not optimal for bounded variation functions and their non-linear approximation error does not have the M^{-1} decay of wavelets. Irregular geometric structures and isolated singularities are approximated with more curvelets than wavelets. Section 12.2.4 studies approximations in bandlet dictionaries, which are adapted to the unknown geometric image regularity.

In most images, curvelet frame approximations are not as effective as wavelet orthonormal bases because they have a redundancy factor A which is at least 5, and because most images include some structures that are more irregular than just piecewise $\mathbf{C}^2$ elements. Section 11.3.2 describes curvelet applications to noise removal.

9.4 Exercises

9.1. **2** Suppose that $\{g_m\}_{m\in\mathbb{Z}}$ and $\{\tilde{g}_m\}_{m\in\mathbb{Z}}$ are two dual frames of $\mathbf{L^2}[0,1]$.

(a) Let $f_N = \sum_{m=0}^{N-1} \langle f, g_m\rangle\, \tilde{g}_m$. Prove that the result (9.4) of Theorem 9.1 remains valid.

(b) Let $\langle f, g_{m_k}\rangle$ be the coefficient of rank k: $|\langle f, g_{m_k}\rangle| \geqslant |\langle f, g_{m_{k+1}}\rangle|$. Prove that if $|\langle f, g_{m_k}\rangle| = O(k^{-s})$ with $s > 1/2$ then the best approximation f_M of f with M frame vectors satisfies $\|f - f_M\|^2 = O(M^{1-2s})$.

9.2. **1** *Color images* A color pixel is represented by red, green and blue components (r,g,b), which are considered as orthogonal coordinates in a three dimensional color space. The red $r[n_1,n_2]$, green $g[n_1,n_2]$ and blue $b[n_1,n_2]$ image pixels are modeled as values taken by respectively three random variables R, G and B, that are the three coordinates of a color vector. Estimate numerically the 3 by 3 covariance matrix of this color random vector from several images and compute the Karhunen-Loève basis that diagonalizes it. Compare the color images reconstructed from the two Karhunen-Loève color channels of highest variance with a reconstruction from the red and green channels.

9.3. **2** Suppose $\mathcal{B} = \{g_m\}_{m\in\mathbb{Z}}$ and $\tilde{B} = \{\tilde{g}_m\}_{m\in\mathbb{Z}}$ be two dual frames of $\mathbf{L^2}[0,1]$. Let $f_N = \sum_{m=0}^{N-1} \langle f, g_m\rangle\, \tilde{g}_m$. Prove that the result (9.4) of Theorem 9.1 remains valid eventhough $\mathcal{B}$ is not an orthonormal basis.

9.4. **2** Let $\vec{f} = (f_k)_{0\leqslant k<K}$ be a multichannel signals where each f_k is a signal of size N. We write $\|\vec{f}\|^2_F = \sum_{k=0}^{K-1} \|f_k\|^2$. Let $\vec{f}_M = (f_{k,M})_{0\leqslant k<K}$ be the multichannel signal obtained by projecting all f_k on the same M vectors of an orthonormal basis $\mathcal{B} = \{g_m\}_{0\leqslant m<N}$. Prove that the best M-term approximation $\vec{f}_M$ which minimizes $\|\vec{f} - \vec{f}_M\|^2_F$, is obtained by selecting the M vectors $g_m \in \mathcal{B}$ which maximize $\sum_{k=0}^{K-1} |\langle f_k, g_m\rangle|^2$.

9.5. **1** Verify that for $f = C\,\mathbf{1}_{[0,1/2]}$, a linear approximation with the N largest scale wavelets over $[0,1]$ produces an error which satisfies $\varepsilon_l(N,f) \sim \|f\|^2_V\, N^{-1}$.