Wavelets

(notes by Pavel Chigansky)

Local irregularities of a function may drastically affect the behavior of its Fourier coefficients with respect to the trigonometric basis. For example, if a smooth function is modified by a local jump discontinuity, the Fourier coefficients would decrease much slower than in the original expansion. This increases the bias term in the risk of projection estimators. To avoid such an effect it makes sense to consider bases of functions localized in space. This chapter introduces *wavelets*, a particular kind of bases which are capable of adapting to spatially varying smoothness.

1. The Haar basis

Let $\phi(x) = \mathbf{1}_{\{x\in(0,1]\}}$, $x \in \mathbb{R}$ and define the shifts $\phi_{0m}(x) = \phi(x-m)$, $m \in \mathbb{Z}$. Having disjoint supports, the functions ϕ_{0m} are orthonormal with respect to the usual scalar product in $L^2(\mathbb{R})$. Let V_0 be the linear subspace spanned by $\{\phi_{0m}\}$. It consists of functions constant on the intervals $(m, m+1]$. Next, define the set

$$V_1 = \big\{ f(2x) : f \in V_0 \big\}.$$

This set consists of all functions in $L^2(\mathbb{R})$ constant on the intervals $(\frac{m}{2}, \frac{m+1}{2}]$, $m \in \mathbb{Z}$. It is a linear subspace spanned by the orthonormal basis, which consists of the functions

$$\phi_{1m} = \sqrt{2}\phi(2x-m), \quad m \in \mathbb{Z}.$$

Since ϕ_{0m}'s are linear combinations of ϕ_{1m}'s the inclusion $V_0 \subset V_1$ holds. By the same pattern, define an increasing sequence of linear subspaces $V_j \subset V_{j+1}$,

$$V_j = \big\{ f(2^j x) : f \in V_0 \big\}, \quad j \in \mathbb{N}. \tag{1.1}$$

The subspace V_j consists of all functions, constant on the intervals $(\frac{m-1}{2^j}, \frac{m}{2^j}]$, $m \in \mathbb{Z}$ and is spanned by the orthonormal basis $\{\phi_{jm}\}$ with

$$\phi_{jm}(x) = 2^{j/2}\phi(2^j x - m), \quad m \in \mathbb{Z}.$$

LEMMA 1.1. *$\lim_{j\to\infty} V_j = \bigcup_{j=0}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$.*

PROOF. (Problem 1). □

---

lecture notes for "Advanced Statistical Models B" course.

The system of functions $\big\{\{\phi_{0m}\},\{\phi_{1m}\},...\big\}$ is obviously not orthogonal, but can be used to construct a complete orthonormal basis in $L^2(\mathbb{R})$. To this end let $W_0 = V_1 \ominus V_0$, the orthogonal complement of V_0 in V_1,

$$W_0 = \{f \in V_1 : f \perp V_0\}.$$

LEMMA 1.2. *The functions* $\psi_{0k}(x) = \psi(x-k)$, $k \in \mathbb{Z}$ *where*

$$\psi(x) = \begin{cases} 1 & x \in (0, \frac{1}{2}] \\ -1 & x \in (\frac{1}{2}, 1] \end{cases}$$

form an orthonormal basis in W_0.

PROOF. Having disjoint supports, $\{\psi_{0k}\}$ are orthogonal. Denote their span by U. To prove that $U = W_0$ it suffices to show that $V_1 = U \oplus V_0$ (orthogonal sum). Any ψ_{0k} is orthogonal to any ϕ_{0m} with $m \neq k$, since they have nonintersecting supports, and is also orthogonal to ϕ_{0k} by direct calculation. Hence $U \perp V_0$. Note that

$$\phi_{10}(x) = \frac{1}{\sqrt{2}}\psi_{01}(x) + \frac{1}{\sqrt{2}}\phi_{01}(x)$$
$$\phi_{11}(x) = \frac{1}{\sqrt{2}}\psi_{01}(x) - \frac{1}{\sqrt{2}}\phi_{01}(x).$$

Analogously, any ϕ_{1m} is a linear combination of $\psi_{0[m/2]}$ and $\phi_{0[m/2]}$. Thus any $f \in V_1$ decomposes into $f = h + g$ with $h \in U$ and $g \in V_0$, i.e. $V_1 = U \oplus V_0$. $\square$

We can proceed in the same manner to define $W_j = V_{j+1} \ominus V_j$ for all $j \in \mathbb{N}$. The functions

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad k \in \mathbb{Z},$$

form an orthonormal basis in W_j and

$$V_j = V_0 \oplus W_0 \oplus ... \oplus W_{j-1}.$$

Since $\cup_{j=0}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$, so is the space $V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$, i.e., any $f \in L^2(\mathbb{R})$ admits the series representation

$$f = \sum_k \alpha_{0k}\phi_{0k} + \sum_{j=0}^{\infty}\sum_k \beta_{jk}\psi_{jk}, \tag{1.2}$$

where α_{0k} and β_{jk} are the scalar products of f with ϕ_{0k} and ψ_{jk}. To recap, we obtained the following result.

COROLLARY 1.3. *The system of functions* $\big\{\{\phi_{0k}\},\{\psi_{0k}\},\{\psi_{1k}\},...\big\}$ *is an orthonormal basis in* $L^2(\mathbb{R})$.

This system is called the *Haar basis*. Like the Fourier series, the Haar expansion is somewhat localized in frequency: if f consists of pure harmonics, coefficients other than those corresponding to the present frequencies will have smaller (but nonzero!) magnitudes. However it also enjoys the localization property in space: if a function f is smooth everywhere except for a small vicinity, only the coefficients corresponding to basis functions, which are nonzero over that vicinity will be affected. This differs drastically from the Fourier basis.

2. Multiresolution analysis

The approximation accuracy by truncated Haar expansion can improve quite slowly with the number of terms in the truncated series (1.2) even for smooth functions (see Problem 2). This stems from the fact that the Haar basis functions are discontinuous. The basic properties of the Haar basis can be mimicked to define wavelets in a much greater generality.

Suppose $\phi(x)$ is a function such that $\phi_{0k}(x) = \phi(x-k)$, $k \in \mathbb{Z}$ are orthonormal on $\mathbb{R}$ (not necessarily with disjoint or even finite supports). For each $j \in \mathbb{N} \cup \{0\}$, define

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k), \quad k \in \mathbb{Z}, \tag{2.1}$$

and let V_j be the linear subspace of $L^2(\mathbb{R})$ spanned by ϕ_{jk}'s,

$$V_j = \Big\{ f = \sum_k c_k \phi_{jk} : \sum_k c_k^2 < \infty \Big\}.$$

DEFINITION 2.1. *The system of subspaces $\{V_j, j \in \mathbb{N} \cup \{0\}\}$ is called Multiresolution Analysis (MRA) generated by a father wavelet ϕ if*

(1) *the spaces V_j are nested, i.e. $V_j \subset V_{j+1}$, and*

(2) *$\bigcup_{j=0}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$.*

By this definition Haar basis is an MRA. Proceeding as before, let $W_j = V_{j+1} \ominus V_j$ so that $V_j = V_0 \oplus \bigoplus_{i=0}^{j-1} W_i$, and by (2) above,

$$\bigcup_{j=0}^{\infty} V_j = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j \text{ is dense in } L^2(\mathbb{R}).$$

Thus any function $f \in L^2(\mathbb{R})$ admits the representation

$$f = \sum_k \alpha_{0k}\phi_{0k} + \sum_{j=0}^{\infty}\sum_k \beta_{jk}\psi_{jk}, \tag{2.2}$$

where $\{\psi_{jk}\}$ is a basis in W_j and the coefficients are scalar products of f with the corresponding basis functions. This is the *multiresolution* expansion of f. The spaces W_j or/and their bases are called *resolution* levels.

Note that at this point it is not at all clear that ψ_{jk}'s can be chosen so that

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad k \in \mathbb{Z},\ j \in \mathbb{N} \cup \{0\}, \tag{2.3}$$

for some function ψ, called a *mother* wavelet. If this turns out to be the case, the series (2.2) is called the *wavelet* expansion. Hence the Haar series is also a wavelet expansion. Another important example is the Shannon wavelet (Problem 6).

3. Construction of wavelets

In view of the above definitions, construction of a wavelet MRA amounts to finding a father wavelet ϕ such that

(i) its translates $\phi_{0k}(x) = \phi(x-k)$ are orthogonal;

(ii) the spaces V_j are nested;

(iii) the union $\bigcup_{j=0}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$;

(iv) bases of the resolution level subspaces $W_j = V_{j+1} \ominus V_j$ satisfy wavelet property (2.3) for some mother wavelet ψ;

(v) mother wavelet has properties, relevant to statistical purposes, such as smoothness, compact support, etc.

Realization of this program turns out to be quite involved, especially in view of additional requirements in (v). Let us sketch a few first steps, which address some properties in the list, referring the reader to a comprehensive account in [1] for more details. The following lemma characterizes functions which satisfy (i).

LEMMA 3.1. *Let $\phi \in L^2(\mathbb{R})$. The functions $\{\phi_{0k}\}$ are orthonormal if and only if the Fourier transform of ϕ satisfies*

$$\sum_k \left|\widehat{\phi}(\lambda + 2\pi k)\right|^2 = 1, \quad a.e. \tag{3.1}$$

PROOF. By the Poisson summation formula from Lemma A.3 and Remark A.4, identity (3.1) holds if and only if the Fourier transform of $g(x) := |\widehat{\phi}(x)|^2$ satisfies

$$\widehat{g}(k) = 2\pi\delta_{k0}, \quad k \in \mathbb{Z}. \tag{3.2}$$

By the correlation property,

$$\widehat{g}(k) = \int_{\mathbb{R}} g(x)e^{-ikx}dx = 2\pi\left(\mathcal{F}^{-1} \circ g\right)(-k) = 2\pi\int_{\mathbb{R}} \phi(x)\phi(x-k)dx,$$

and hence (3.2) and in turn (3.1) is equivalent to

$$\int_{\mathbb{R}} \phi_{0k}(x)\phi_{0m}dx = \int_{\mathbb{R}} \phi(x)\phi(x+(k-m))dx = \delta_{km},$$

as claimed. $\square$

The next lemma gives a characterization for property (ii).

LEMMA 3.2. *Assume that (i) holds. Then $V_j \subset V_{j+1}$ if and only if there exists a 2π-periodic function $m_0 \in L^2([0,2\pi])$ such that*

$$\widehat{\phi}(\lambda) = m_0(\lambda/2)\widehat{\phi}(\lambda/2), \quad a.e. \tag{3.3}$$

PROOF.

1) *Necessity.* Let us assume that the subspaces are nested and construct m_0 as above. If the assumption holds, then, in particular, $V_0 \subset V_1$. By definition V_1 is spanned by the functions $\phi_{1k}(x) = \sqrt{2}\phi(2x-k)$, $k \in \mathbb{Z}$. Then by assumption, ϕ itself, which belongs to V_0, must also belong to V_1,

$$\phi(x) = \sum_k \langle\phi, \phi_{1k}\rangle\phi_{1k}(x) = \sum_k h_k\sqrt{2}\phi(2x-k),$$

where the coefficients $h_k := \sqrt{2}\int_{\mathbb{R}} \phi(u)\phi(2u-k)du$ are square summable. Taking the Fourier transform of both sides and using the scaling and the shift properties,

$$\widehat{\phi}(\lambda) = \sqrt{2}\frac{1}{2}\widehat{\phi}(\lambda/2)\sum_k e^{-i\lambda k/2}h_k =: \widehat{\phi}(\lambda/2)m_0(\lambda/2),$$

where the function

$$m_0(\lambda) = \frac{1}{\sqrt{2}}\sum_k h_k e^{-i\lambda k}$$

is 2π-periodic and square integrable on $[0,2\pi]$.

2) *Sufficiency.* Let us assume that property (3.3) is satisfied for some $m_0 \in L^2([0,2\pi])$ and show that $V_j \subset V_{j+1}$. The proof is similar for all j's and we will check $V_0 \subset V_1$. Note that $f = \sum_k c_k\phi_{0k} \in V_0$ satisfies

$$\widehat{f}(\lambda) = \int_{\mathbb{R}}\Big(\sum_k c_k\phi(x-k)\Big)e^{-i\lambda x}dx = \sum_k c_k e^{-i\lambda k}\widehat{\phi}(\lambda) =: m(\lambda)\widehat{\phi}(\lambda),$$

where $m \in L^2([0,2\pi])$ and is 2π-periodic. Hence the space of Fourier transforms of functions from V_0 is

$$\widehat{V}_0 := \Big\{m(\lambda)\widehat{\phi}(\lambda) : m \in L^2([0,2\pi]) \text{ is } 2\pi\text{-periodic}\Big\}.$$

Similarly, the Fourier transforms of functions from V_1 are

$$\widehat{V}_1 = \Big\{m(\lambda/2)\widehat{\phi}(\lambda/2) : m \in L^2([0,2\pi]) \text{ is } 2\pi\text{-periodic}\Big\}.$$

By invertibility of the Fourier transform, the desired property $V_0 \subset V_1$ holds if we show that $\widehat{V}_0 \subset \widehat{V}_1$.

In view of (3.3), a function $\widehat{f} \in V_0$ admits the representation

$$\widehat{f}(\lambda) = m(\lambda)\widehat{\phi}(\lambda) = m(\lambda)m_0(\lambda/2)\widehat{\phi}(\lambda/2).$$

Note that the function $m(2\lambda)m_0(\lambda)$ is 2π-periodic. Hence to prove that $\widehat{V}_0 \subset \widehat{V}_1$, it remains to show that it belongs to $L^2([0,2\pi])$. To this end it suffices to show that m_0, which satisfies (3.3), is, in fact, a bounded function. We will argue that m_0 satisfies

$$|m_0(\lambda)|^2 + |m_0(\lambda+\pi)|^2 = 1, \quad a.e.,$$

and therefore is bounded by 1. Indeed, if ϕ satisfies (i), then by Lemma 3.1, for almost all λ,

$$\begin{aligned}
&1 = \sum_k \big|\widehat{\phi}(2\lambda+2\pi k)\big|^2 \overset{(3.3)}{=} \sum_k \big|m_0(\lambda+\pi k)\widehat{\phi}(\lambda+\pi k)\big|^2 = \\
&\sum_k \big|m_0(\lambda+2\pi k)\widehat{\phi}(\lambda+2\pi k)\big|^2 + \sum_k \big|m_0(\lambda+2\pi k+\pi)\widehat{\phi}(\lambda+2\pi k+\pi)\big|^2 \overset{\dagger}{=} \\
&|m_0(\lambda)|^2\sum_k \big|\widehat{\phi}(\lambda+2\pi k)\big|^2 + |m_0(\lambda+\pi)|^2\sum_k \big|\widehat{\phi}(\lambda+2\pi k+\pi)\big|^2 = \\
&|m_0(\lambda)|^2 + |m_0(\lambda+\pi)|^2,
\end{aligned}$$

where $\dagger$ holds, since m_0 is periodic.

$\square$

Property (iii) can be shown to hold when ϕ is a father wavelet, satisfying (3.1) and (3.3), which also satisfies an additional mild integrability condition. The following result settles (iv).

LEMMA 3.3. *Let ϕ be a father wavelet which generates an MRA of $L^2(\mathbb{R})$ and let m_0 be a solution of (3.3). Then the inverse Fourier transform ψ of*

$$\widehat{\psi}(\lambda) = m_1(\lambda/2)\widehat{\phi}(\lambda/2), \tag{3.4}$$

where $m_1(\lambda) = \overline{m_0(\lambda+\pi)}e^{-i\lambda}$, is a mother wavelet.

PROOF. (see [1]) $\square$

The formula (3.4) can be used to construct a mother wavelet from the farther, which in turn can be constructed by iterations of (3.3),

$$\widehat{\phi}(\lambda) = \prod_{j=1}^{\infty} m_0(\lambda/2^j), \tag{3.5}$$

if the infinite product is convergent and nontrivial. Therefore, a particular wavelet basis is defined by the choice of m_0, which must satisfy the properties stated in the lemmas above and, moreover, guarantee additional features of the obtained wavelet, mentioned in (v).

Many useful wavelet families have been constructed over the years. The Daubechies wavelets are generated by $m_0(\lambda)$ satisfying

$$|m_0(\lambda)|^2 = c_N \int_{\lambda}^{\pi} (\sin x)^{2N-1} dx, \tag{3.6}$$

where c_N is the constant chosen so that $m_0(0) = 1$. For $N = 1$ it coincides with the Haar basis (Problem 7), which turns out to be the only wavelet having a *symmetric* compactly supported father wavelet ϕ. For other N's no closed form formulas for ϕ or ψ exist, but they can be approximated numerically (see Figure 1). Here are some important properties of the Daubechies wavelets.

(1) Both father and mother wavelets are compactly supported,

$$\operatorname{supp}(\phi) \subseteq [0, 2N-1] \quad \text{and} \quad \operatorname{supp}(\psi) \subseteq [-N+1, N].$$

(2) The first $N-1$ moments of the mother wavelet vanish

$$\int_{\mathbb{R}} \psi(x) x^{\ell} dx = 0, \quad \ell = 0, \dots, N-1.$$

The second property plays the central role in nonparametric estimation, since it allows to control the bias terms in the MSE decomposition on appropriate spaces. Daubechies wavelets have vanishing moments for the mother wavelet, but not for the father. This is taken care of by another family of wavelets called *coiflets*, [1].

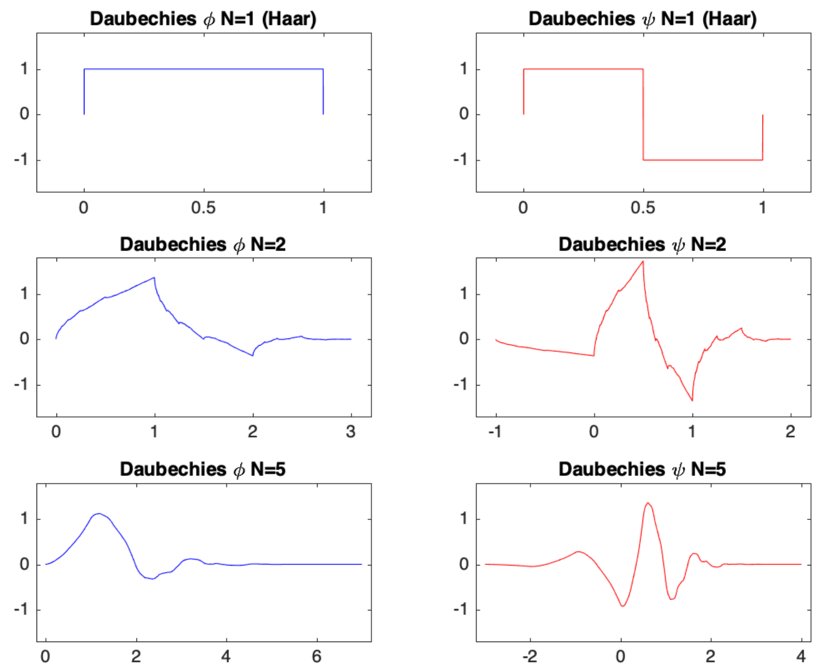

FIGURE 1. Father and mother Daubechies wavelets

4. Asymptotic minimax

Consider the standard regression model

$$Y_j = f(X_j) + \varepsilon_j, \quad j = 1, \dots, n,$$

where ε_j's are i.i.d. $N(0,1)$ random variables, the design is uniform, $X_j = j/n$, and f is the unknown function to be estimated given the data $(X_1, Y_1), \dots, (X_n, Y_n)$.

4.1. Linear and nonlinear wavelet estimators. The usual projection wavelet estimator is obtained by replacing the coefficients in (2.2) with their empirical counterparts

$$\widehat{\alpha}_{0,k} = \frac{1}{n} \sum_{m=1}^{n} Y_m \phi_k(m/n) \quad \text{and} \quad \widehat{\beta}_{j,k} = \frac{1}{n} \sum_{m=1}^{n} Y_m \psi_{jk}(m/n),$$

and truncating the series to a finite number of terms or, more generally, multiplying the terms by some weights. Such an estimator is *linear* with respect to the samples.

One can also consider nonlinear estimators by applying some nonlinear function to the coefficients estimates prior to plugging them into (2.2). A simple choice is the *hard limiter* function $\delta_\lambda^H(u) = |u| \mathbf{1}_{\{|u| \geq \lambda\}}$ with a threshold $\lambda \in \mathbb{R}_+$. The nonlinear wavelet estimator with thresholds λ_k and λ_{jk} is then

$$\widehat{f}_{n,\lambda}(x) = \sum_k \delta^H_{\lambda_k}\big(\widehat{\alpha}_{0k}\big)\phi_{0k}(x) + \sum_{j=0}^{\infty}\sum_k \delta^H_{\lambda_{jk}}\big(\widehat{\beta}_{jk}\big)\psi_{jk}(x). \tag{4.1}$$

The coefficients tend to decrease and hence the summations have only finitely many terms. Another simple transformation is the *soft limiter*

$$\delta^S_\lambda(u) = \mathrm{sign}(u)(|u|-\lambda)_+, \quad u \in \mathbb{R}.$$

The thresholds λ_k and λ_{jk} are design parameters and various practical rules for their data-driven choice are available in the literature.

4.2. Asymptotic optimality. Asymptotic minimax theory of wavelet estimators is technically involved and is way beyond the scope of these notes. Let us briefly sketch some representative results, extracted from [3].

DEFINITION 4.1. *Let $P = (x_0, \dots, x_n)$ be a partition of an interval $[a,b]$ with points*

$$a = x_0 < \dots < x_n = b.$$

Variation of a function $f : [a,b] \mapsto \mathbb{R}$ on the partition P is

$$\bigvee_P f = \sum_j^{|P|} |f(x_j) - f(x_{j-1})|.$$

The quantity

$$\bigvee_a^b f = \sup_{P\in\mathcal{P}} \bigvee_P f,$$

where $\mathcal{P}$ is the set of all finite partitions, is called the total variation of f on $[a,b]$. A function is said to have bounded (total) variation if $\bigvee_a^b f < \infty$.

It is not hard to see that a function f with continuous derivative f' has bounded variation

$$\bigvee_a^b f = \int_a^b |f'(x)|dx < \infty.$$

More generally, if f is piecewise continuously differentiable on $[a,b]\setminus\{x_1, x_2, \dots\}$ with derivative f' and has absolutely summable jumps at x_j's then

$$\bigvee_a^b f = \int_a^b |f'(x)|dx + \sum_j |f(x_j) - f(x_{j-1})| < \infty.$$

Thus in addition to smooth functions the total variation body ("ball")

$$TV(C) = \Big\{ f : \bigvee_a^b f \le C \Big\}$$

contains, both smooth and non-smooth functions, e.g. with jumps.

For an estimator $\widehat{f}_n$, based on the sample of size n, define the MISE risk

$$R(f, \widehat{f}_n) = \mathbb{E}_f \|f - \widehat{f}_n\|_2^2.$$

The theory developed in [3] provides the following insights.

(1) The optimal minimax risk on $TV(C)$ satisfies [1]

$$\inf_{\widehat{f}_n} \sup_{f\in TV(C)} R(f,\widehat{f}_n) \asymp n^{-2/3}.$$

(2) The optimal minimax risk on $TV(C)$ of linear estimators satisfies

$$\inf_{\widehat{f}_n\in\mathcal{F}_{\mathrm{lin}}} \sup_{f\in TV(C)} R(f,\widehat{f}_n) \asymp n^{-1/2}.$$

Here $\mathcal{F}_{\mathrm{lin}}$ stands for the set of *all* estimators, linear with respect to Y_j's (including linear wavelet estimators).

(3) It is possible to choose the thresholds λ_k and λ_{jk} (depending on n and C) so that the nonlinear wavelet estimator in (4.1) achieves the minimax rate $n^{-2/3}$.

The optimal minimax asymptotics of nonlinear wavelet estimators is the manifestation of their adaptivity to spatial irregularity. More generally, properly tuned wavelet estimators are asymptotically optimal on Besov spaces. These spaces are defined in terms of certain smoothness characteristics of functions, such as moduli of continuity, etc. Ellipsoids in such function spaces can be also defined in terms of appropriate norms of sequences of their wavelet expansions. It is this fact which makes Besov spaces suitable for analysis of wavelet estimators ([3], [1]).

Computer assignment

(1) Implement projection estimator with the Haar basis, which returns the estimated function at the design points [2]. Generate the regression data for the function

$$f(x) = x\sin(2\pi/x), \quad x\in[0,1],$$

on the uniform design $X_m = m/n$, $m\in\{1,\dots,n\}$ with $N(0,1)$ noise. Experiment with the sample sizes and the choice of the maximal resolution level $J_{\max}$ at which the series is truncated. Plot estimates for several values of $J_{\max}$ versus the true function and comment on the results.

(2) Approximate the discrete MISE risk of your estimator

$$\mathrm{MISE}(J_{\max}) = \mathbb{E}_f \frac{1}{n}\sum_{m=1}^n \big(f(X_m)-\widehat{f}(X_m)\big)^2$$

for $J_{\max}\in\{1,\dots,5\}$ by averaging over $M = 10{,}000$ MC trials. Find the minimal approximated MISE and the corresponding optimal "oracle" level $J^*_{\max}$.

[1] for positive real sequences a_n and b_n, $a_n \asymp b_n$ means that $\inf_n a_n/b_n > 0$ and $\sup_n a_n/b_n < \infty$.

[2] think how to use matrix multiplication to make your code as efficient as possible

(3) Repeat (1)-(2) for the projection estimator with Fourier basis. Compare its optimal "oracle" MISE risk to that obtained in the previous question.

(4) Plot the optimal "oracle" estimates with Haar and Fourier bases versus the noisy data and the true function.

Exercises

PROBLEM 1. Prove Lemma 1.1, following the steps below.

(1) Argue that no generality is lost, if $f(x) \ge 0$ is assumed.

Hint: consider the decomposition $f(x) = f^+(x) - f^-(x)$ where f^+ and f^- are the positive and negative parts of f.

(2) Show that a nonnegative function $f \in L^2(\mathbb{R})$ can be approximated by a bounded function. More precisely, for any $\varepsilon > 0$, there exists a function $\widetilde{f}$ such that $\widetilde{f}(x) \le M$ for some M and $\|f - \widetilde{f}\|_2 \le \varepsilon$. Thus without loss of generality the claim of Lemma 1.1 can be proved for bounded nonnegative functions, in particular $0 \le f \le 1$ can be assumed.

Hint: try $\widetilde{f}_M(x) = f(x)\mathbf{1}_{\{f(x)\le M\}}$ with a suitable M.

(3) Show that a function $f \in L^2(\mathbb{R})$ with $0 \le f(x) \le 1$ can be approximated by simple functions, which take only finite number of values. More precisely, for any $\varepsilon > 0$ there exist m subsets $A_1, \dots, A_m$ and constants $c_1, \dots, c_m$ such that $\|f - \widetilde{f}_m\|_2 \le \varepsilon$ where

$$\widetilde{f}_m(x) = \sum_{j=1}^m c_j \mathbf{1}_{\{x \in A_j\}}, \quad x \in \mathbb{R}. \tag{4.2}$$

Hint: try the above approximations with $m = 2^n$, $n \in \mathbb{N}$, coefficients $c_j = \frac{j-1}{2^n}$ and subsets $A_j = \{x : f(x) \in [\frac{j-1}{2^n}, \frac{j}{2^n})\}$, that is,

$$\widetilde{f}_n(x) = \sum_{j=1}^{2^n} \frac{j-1}{2^n} \mathbf{1}_{\{f(x)\in[\frac{j-1}{2^n}, \frac{j}{2^n})\}}.$$

(4) Complete the proof by arguing that any function of the form (4.2) can be approximated by a function from V_k from (1.1) for sufficiently large k.

Hint: it is known that a bounded Borel subset A of a real line can be approximated in the Lebesgue measure by dyadic sets of sufficiently fine resolution.

PROBLEM 2. This problem demonstrates that even for very smooth functions the accuracy of the Haar approximation can improve quite slowly with the number of terms included into the truncated series. To this end let us consider the function

$$f(x) = \sin(2\pi x)\mathbf{1}_{\{x\in[0,1]\}}. \tag{4.3}$$

(1) Find the coefficients β_{jk} in the Haar expansion (1.2) for the function (4.3).

(2) Check that the series (1.2) truncated at $j := J$ contains $2^{J+1} - 3$ nonzero summands.

(3) Argue that the $L^2(\mathbb{R})$ approximation accuracy by the Haar expansion improves as $O(n^{-1})$ where n is the total number of summands in the truncated series, while the Fourier expansion is trivially finite and exact.

PROBLEM 3. Prove the properties (i)-(iv) of the Fourier transform, formulated in Appendix A.

PROBLEM 4. Find the Fourier transforms of the following functions.

(1) $f(x) = \mathbf{1}_{\{|x|\le 1\}}$

(2) $f(x) = (1-|x|)\mathbf{1}_{\{|x|\le 1\}}$

(3) $f(x) = e^{-|x|}$

PROBLEM 5. The goal of this problem is to exercise the conditions of Lemmas 3.1-3.2 for the Haar basis.

(1) Check that Haar's father wavelet satisfies the condition of Lemma 3.1.

Hint: use the formula

$$\sum_{k\in\mathbb{Z}} \frac{1}{(t+\pi k)^2} = \frac{1}{\sin^2(t)}.$$

(2) Find $m_0(\lambda)$ in Lemma 3.2 and check that $m_0 \in L^2([0,2\pi])$.

(3) Find Haar's mother wavelet using Lemma 3.3.

PROBLEM 6 (Shannon's wavelet). One of the central results in Signal Processing is the sampling theorem, which asserts that a function f whose Fourier transform

$$\widehat{f}(\lambda) = \int_{\mathbb{R}} f(x) e^{-i\lambda x} dx,$$

vanishes outside the interval $[-\pi,\pi]$, is defined by its values at integers and satisfies

$$f(x) = \sum_k f(k) \frac{\sin \pi(x-k)}{\pi(x-k)}, \quad x \in \mathbb{R}. \tag{4.4}$$

The goal of this problem is to show that the function (extended by continuity to 0),

$$\phi(x) = \frac{\sin(\pi x)}{\pi x}, \quad x \in \mathbb{R}, \tag{4.5}$$

is a valid father wavelet.

(1) Show that the function (4.5) is the inverse Fourier transform of $\mathbf{1}_{\{|\lambda|\le\pi\}}$, that is,

$$\frac{\sin(\pi x)}{\pi x} = \frac{1}{2\pi}\int_{\mathbb{R}} \mathbf{1}_{\{|\lambda|\le\pi\}} e^{i\lambda x} d\lambda.$$

(2) Show that $\{\phi_{0k}\}$ are orthonormal on $\mathbb{R}$.

Hint: use the convolution property of the Fourier transform

(3) Show that $V_j \subset V_{j+1}$.

Hint: use the scaling and translation properties of the Fourier transform and representation (4.4).

(4) Show that $\bigcup_{j=0}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$, thus concluding that $\big\{V_j, j \in \mathbb{N} \cup \{0\}\big\}$ is MRA generated by the father wavelet (4.5).

Hint: use Plancherel's theorem, which asserts that for $f \in L^2(\mathbb{R})$,

$$\|f\|_2 = \|\widehat{f}\|_2.$$

(5) Find the mother wavelet for the Shannon basis using Lemma 3.3.

PROBLEM 7. Show that Daubechies wavelets for $N = 1$ coincide with the Haar basis.

Appendix A. A brief on the Fourier transform

A.1. Definitions. The *Fourier transform* of a function $f \in L^1(\mathbb{R})$ is defined by the integral, called the *Fourier operator*,

$$\widehat{f}(\lambda) = (\mathcal{F}f)(\lambda) = \int_{\mathbb{R}} f(x)e^{-i\lambda x}dx, \quad \lambda \in \mathbb{R}.$$

By the dominated convergence theorem it is a continuous function and the following results asserts that it has vanishing tails.

THEOREM A.1 (Riemann-Lebesgue lemma). *For any $f \in L^1(\mathbb{R})$,*

$$\lim_{|\lambda|\to\infty} \widehat{f}(\lambda) = 0.$$

If $\widehat{f} \in L^1(\mathbb{R})$ then it can be shown that

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\lambda)e^{i\lambda x}d\lambda, \quad a.e.$$

This integral is called the *inverse Fourier transform*. Thus the action of the Fourier operator is invertible at least on the subspace of continuous integrable functions.

The domain of the Fourier operator can be extended to spaces other than $L^1(\mathbb{R})$. An important case is the Hilbert space $L^2(\mathbb{R})$ and the following lemma is instrumental.

LEMMA A.2 (Plancherel formula). *If $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then*

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\lambda)\overline{\widehat{g}(\lambda)}d\lambda.$$

By this lemma the Fourier operator is an *isometry* from $L^1(\mathbb{R})\cap L^2(\mathbb{R})$ to $L^2(\mathbb{R})$, i.e., it is a linear map which preserves the norm. The subspace $L^1(\mathbb{R})\cap L^2(\mathbb{R})$ can be shown dense in $L^2(\mathbb{R})$. This means that for any $f\in L^2(\mathbb{R})$ there is a sequence $f_n\in L^1(\mathbb{R})\cap L^2(\mathbb{R})$ such that $\|f-f_n\|^2\to 0$. By the isometry property

$$\frac{1}{2\pi}\|\widehat{f}_m-\widehat{f}_n\|^2=\|f_m-f_n\|^2\xrightarrow[m,n\to\infty]{}0$$

which implies that $\widehat{f}_n$ is a Cauchy sequence in $L^2(\mathbb{R})$. By completeness there exists a function $\widehat{f}\in L^2(\mathbb{R})$ such that $\|\widehat{f}_n-\widehat{f}\|\to 0$. The limit does not depend on the choice of the approximating sequence f_n: indeed, if g_n is another such sequence, i.e., $\|g_n-f\|\to 0$ and $\widehat{g}_n\to\widehat{g}$, then

$$\|\widehat{f}-\widehat{g}\|\le\|\widehat{f}-\widehat{f}_n\|+\|\widehat{f}_n-\widehat{g}_n\|+\|\widehat{g}-\widehat{g}_n\|\to 0,$$

where the second terms vanishes by the isometry property. This limiting procedure extends the definition of the Fourier operator to $L^2(\mathbb{R})$.

A.2. Basic properties. The Fourier transform has many useful properties which makes it a powerful tool in a variety of applications. Below is a very partial list[3].

(i) Translation property:

$$f(x-a)\quad\Longleftrightarrow\quad\widehat{f}(\lambda)e^{-ia\lambda}.$$

(ii) Scale property:

$$f(ax)\quad\Longleftrightarrow\quad|a|^{-1}\widehat{f}(a^{-1}\lambda),\quad a\neq 0.$$

(iii) Convolution property:

$$f_1*f_2\quad\Longleftrightarrow\quad\widehat{f}_1(\lambda)\widehat{f}_2(\lambda),$$

where $*$ stands for the convolution integral

$$(f_1*f_2)(x)=\int_{\mathbb{R}}f_1(u)f_2(x-u)du.$$

In particular, convolution between $f(x)$ and $f(-x)$ is mapped to $|\widehat{f}(\lambda)|^2$.

(iv) If $\widehat{f}$ has finite p-th absolute moment, then f is p times differentiable and

$$\frac{d^p}{dx^p}f(x)\quad\Longleftrightarrow\quad(i\lambda)^p\widehat{f}(\lambda).$$

A.3. Poisson summation formula. The following formula relates the Fourier transform of a function to the Fourier series of the periodic functions generated by its translates.

[3] $\Longleftrightarrow$ stands for correspondence between functions and their Fourier transforms

LEMMA A.3. *Let $f \in L^1(\mathbb{R})$. Then the series*

$$S(x) = \sum_k f(x + 2\pi k)$$

converges in $L^1([0,2\pi])$ to a periodic function with Fourier series coefficients

$$\widehat{s}_k = \frac{1}{2\pi}\widehat{f}(k).$$

PROOF. Define the partial sums $S_n(x) = \sum_{|k|\le n} f(x+2\pi k)$, $n \in \mathbb{N}$. For $n \ge m$,

$$\begin{aligned}
\|S_n - S_m\|_1 \le \sum_{m<|k|\le n} \int_0^{2\pi} |f(x+2\pi k)|dx = \\
\sum_{k=-n}^{-m} \int_{2\pi k}^{2\pi(1+k)} |f(y)|dy + \sum_{k=m}^{n} \int_{2\pi k}^{2\pi(1+k)} |f(y)|dy = \\
\int_{-2\pi n}^{2\pi(1-m)} |f(y)|dy + \int_{2\pi m}^{2\pi(1+n)} |f(y)|dy \le \\
\int_{\mathbb{R}} |f(y)| \mathbf{1}_{\{|y|\ge 2\pi(m-1)\}} dy \xrightarrow[m\to\infty]{} 0,
\end{aligned}$$

where the convergence holds since $f \in L^1(\mathbb{R})$. This shows that S_n is a Cauchy sequence in $L^1([0,2\pi])$ and therefore it converges to a function $S \in L^1([0,2\pi])$ by completeness. Its Fourier coefficients are given by

$$\widehat{s}_m = \frac{1}{2\pi}\sum_k \int_0^{2\pi} f(x+2\pi k)e^{-imx}dx = \frac{1}{2\pi}\int_{\mathbb{R}} f(x)e^{-imx}dx = \frac{1}{2\pi}\widehat{f}(m).$$

□

REMARK A.4. In particular, $S(x)$ equals 1 a.e. if and only if $\widehat{f}(k) = 2\pi\delta_{k0}$.

References

[1] Härdle, W., Kerkyacharian, G., Picard, D., & Tsybakov, A. (2012). Wavelets, approximation, and statistical applications (Vol. 129). Springer Science & Business Media.

[2] Pinsky, Mark A. Introduction to Fourier analysis and wavelets. Vol. 102. American Mathematical Soc., 2008.

[3] Donoho, David L., and Iain M. Johnstone. ”Minimax estimation via wavelet shrinkage.” The annals of Statistics 26.3 (1998): 879-921.

DEPARTMENT OF STATISTICS, THE HEBREW UNIVERSITY, MOUNT SCOPUS, JERUSALEM 91905, ISRAEL

E-mail address: Pavel.Chigansky@mail.huji.ac.il