Statistica Sinica **14**(2004), 377-393

CONVERGENCE PROPERTIES OF WAVELET ESTIMATORS WITH MULTIPLE SAMPLING RATES

Peter Hall, Gérard Kerkyacharian and Dominique Picard

Australian National University, Université de Paris X and Université de Paris VII

Abstract: Adaptive sampling schemes with multiple sampling rates have the potential to significantly improve the efficiency and effectiveness of methods for signal analysis. For example, in the case of equipment which transmits data continuously, multi-rate methods can reduce the cost of transmission. For equipment which transmits data only periodically they can reduce the costs of both storage and transmission. When multiple sampling rates are used in connection with wavelet estimators, the most natural algorithms for rate-switching are arguably those based on threshold-crossings by wavelet coefficients. In this paper we study the performance of such algorithms, and show that even simple threshold-crossing rules can achieve near-optimal convergence rates. A new mathematical model is suggested for assessing performance, combining the simplicity and familiarity of global approaches with an account of the local variation to which multi-rate sampling responds.

Key words and phrases: Besov space, convergence rates, curve estimation, Hölder space, minimax, online estimation, optimality, rate switching, signal analysis, threshold.

1. Introduction

In conventional problems of signal analysis, noisy observations are made of a signal at regular time intervals. The sampling rates can vary widely; they are as low as 8kHz for digital telephony, 44.1kHz for conventional CDs, 96kHz or 192kHz for DVD audio, and several mHz for new multi-channel systems. However, the rate for a specific device generally does not vary in time. While this constraint makes for simpler technology, it prevents users from accessing the advantages of a more adaptive, and hence potentially more effective, system.

Those advantages can arise in several ways. First, in the case of data storage, using a constant sampling rate means that unnecessarily large amounts of storage space are taken up when the signal is essentially flat, or uninteresting. If the sampling rate could be reduced in such circumstances then this space could be saved. Not being able to do so can be a major drawback for remote, automated equipment which spends most of its time off-line, perhaps transmitting its stored data only occasionally when a satellite passes overhead. On the other hand,

using a low sampling rate at all times may cause the device to miss important detail in high-frequency events which occur only infrequently. Likewise, multi-rate sampling is potentially beneficial in the context of data transmission; sending masses of data containing little information can be cost-ineffective.

For these reasons, and others, there is growing interest in technologies where sampling rates vary with time, in an adaptive fashion determined by signal complexity. See, for example, work of Liu (1996), Liu and Walter (1996), Jetter and Stöckler (1997), Aldroubi and Feichtinger (1998), Chen, Itoh and Shiki (1998) and Aldroubi and Gröchenig (2001). Of course, wavelet methods offer an exceptionally flexible approach to signal analysis, and there a natural technique for determining (and responding to) signal complexity is to threshold the wavelet coefficients. Online, quasi real-time versions of this approach have been suggested and discussed be Hall and Penev (2002), and shown both theoretically and numerically to have the sorts of properties desired of them. In particular, they allow the sampling rate to increase when an empirical assessment of the signal shows that it has become relatively complex, and to decrease again when complexity appears to be returning to more conventional, lower levels.

While the methods discussed by Hall and Penev (2002) are reasonably close to those that might be used in practice, the results are developed only for a single signal, not uniformly in a large class of signals. An advantage of the former approach is that it permits detailed analysis of local properties, but a disadvantage is that it precludes access to traditional minimax optimality arguments, based on global performance on intervals. To overcome this drawback, in the present paper we introduce a global measure of performance which nevertheless encompasses local features of multi-rate estimators, and which reveals the ability of multi-rate sampling schemes to accommodate, in an essentially optimal way, functions whose regularity may by highly variable on an interval.

In practice, a multiple sampling rate technology would probably be based on only a small number of rates, perhaps two, three or four. In a theoretical treatment, addressing more than two rates is often neither more revealing nor more difficult than considering only two, and so we confine most of our attention to the latter case. It has the advantage of being simpler to discuss. We briefly consider multi-rate generalisations in Sections 2.4 and 3.1.

In the two-rate case we treat a threshold-based rate-switching rule which is designed to use the lower sampling rate at timepoints where a signal is judged to be of high regularity, and to increase the rate when the regularity appears to decrease below a threshold. High regularity is defined in terms of the true signal, g say, being in a conventional Hölder space Λ^s, and low regularity as $g \in \Lambda^t \setminus \Lambda^s$, where $t < s$. Rate-switching decisions are based on empirical wavelet coefficients, using a new level of threshold which is designed to distinguish

between Λ^s and $\Lambda^t \setminus \Lambda^s$. It is shown that this approach can achieve virtually minimax-optimal convergence rates uniformly over the larger, lower-regularity class Λ^t, and at the same time ensure that the lower (and less expensive) sampling rate is employed in the higher-regularity class Λ^s, except when functions are near the boundary that separates Λ^s from $g \in \Lambda^t \setminus \Lambda^s$.

In a dual-rate scheme applied in cases where the signal is usually relatively mundane, but has occasional interesting, high-frequency bursts, the sampler would spend most of its time at the lower rate, increasing the rate only for relatively short periods when adaptive data analysis indicated that doing so was warranted. Our theory is motivated by this context and, in Section 3.1 we discuss our results in this setting. The diametrically opposite alternative, where most sampling is done at the high rate but drops to the low rate occasionally, is usually hard to justify in practical terms, since it leads to only minor savings in sampling, storage and transmission costs.

Finally let us observe that we express regularity in term of Hölder spaces. However our results can be generalised to weak Besov spaces. In this regard, note particularly the definition of $m_q(j_0, \mathcal{J})$ and Lemma 1.

2. Models and Methodology

2.1. Models for data and sampling

Suppose data Y_u are generated at "observation times" u according to the model,

$$Y_u = g(u) + \epsilon_u\,, \quad u \in \mathcal{I} \equiv [0,1]\,, \tag{2.1}$$

where the function g represents a signal, ϵ_u (denoting noise) is a random variable with zero mean and finite variance σ^2, and the timepoints u are chosen discretely. The experimenter has some latitude in selecting the observation times. In particular, we suppose that u comes from a design that may be chosen on one of two grids, either "slow" with $u = k/v_n$ for $1 \le k \le v_n$, or "fast" with $u = k/n$ for $1 \le k \le n$, where n will be permitted to increase without bound.

Motivation for dual-rate sampling is perhaps better appreciated if we assume that

$$\begin{gathered} n \ge v_n, \text{ both } n \text{ and } v_n \text{ are integer powers of 2,} \\ \text{and } v_n \to \infty \text{ as } n \to \infty, \text{ but with } v_n/n \to 0\,. \end{gathered} \tag{2.2}$$

The relationship $n \ge v_n$ implies that the fast grid corresponds to a higher sampling rate than does the slow grid. Our algorithm will prescribe fast-grid sampling when there is empirical evidence, local to a point in time, that the signal g is relatively complex. Our primary example of v_n is $v_n \asymp n^\xi$ where $0 < \xi < 1$, and the notation $a_n \asymp b_n$, for positive constants a_n and b_n, means that a_n/b_n is bounded away from zero and infinity. The assumption, in (2.2), that n and v_n

are both powers of 2 is inessential but conventional, since it simplifies notation for, and computation of, wavelet estimators. On the present occasion it also implies that the slow grid is a subset of the fast grid. This means that if we draw a slow-grid sample but then decide (after an analysis of those data) that we really need a fast-grid sample, the slow-grid sample can still be used; we need only adjoin extra data to it.

2.2. Wavelet expansion and estimators

To introduce the expansion, denote by ψ and ϕ the mother and father wavelet functions, assumed bounded and compactly supported. Let $p = 2^{j_0}$ be the (inverse of the) "maximal bandwidth," or the "primary resolution level," these nomenclatures being drawn from Donoho, Johnstone, Kerkyacharian and Picard (1995) and Hall and Patil (1995), respectively. Put $p_j = 2^j p$ for $j \geq 0$, and define $\phi_j(u) = p^{1/2}\,\phi(pu - j)$ and $\psi_{jk}(u) = p_j^{1/2}\,\psi(p_j u - k)$. A wavelet expansion of g is

$$g = \sum_\ell \alpha_\ell \phi_\ell + \sum_{j=j_0}^{\infty} \sum_k \beta_{jk}\,\psi_{jk}\,, \tag{2.3}$$

where $\alpha_\ell = \int g\,\phi_\ell$ and $\beta_{jk} = \int g\,\phi_{jk}$.

We assume g is in the Hölder class $\Lambda^s = B_{s\infty\infty}$, consisting of functions for which, with the above definitions of α_ℓ and β_{jk},

$$\sup_\ell |\alpha_\ell| + \sup_{j \geq j_0} 2^{j\{s+(1/2)\}} \sup_k |\beta_{jk}| < \infty\,. \tag{2.4}$$

Of course the series here depends on j_0, but it is finite for some j_0 if and only if it is finite for all j_0, in particular for $j_0 = 0$. The value of the left-hand side of (2.4) in the latter case is the norm $\|g\|_{s\infty\infty}$. Given a constant $\gamma > 0$, let $\Lambda^s(\gamma)$ denote the set of g for which $\|g\|_{s\infty\infty} \leq \gamma$. We usually suppress the notation γ.

In Section 2.3 we discuss a wavelet estimator with dual sampling rates, where the slow and fast rates are designed primarily to estimate g on intervals for which its restriction is in Λ^s, or in $\Lambda^t \setminus \Lambda^s$, respectively, with $1/2 < t < s < r$. Here, the positive number r denotes the greatest "regularity" that we envisage the functions might enjoy, and t the least.

Next we define empirical wavelet coefficients and thresholds. In our main results we take the primary resolution level to satisfy

$$p \asymp v_n^{1/(2r+1)}\,. \tag{2.5}$$

With superscripts S and F denoting "slow" and "fast," respectively, the empirical wavelet coefficients computed on the slow and fast grids are

$$\hat\alpha_\ell^{\mathrm{S}} = v_n^{-1} \sum_{i=1}^{v_n} \phi_\ell(i/v_n)\,Y_{i/v_n}\,, \quad \hat\beta_{jk}^{\mathrm{S}} = v_n^{-1} \sum_{i=1}^{v_n} \psi_{jk}(i/v_n)\,Y_{i/v_n}\,,$$

$$\hat{\alpha}_\ell^{\mathrm{F}} = n^{-1} \sum_{i=1}^{n} \phi_\ell(i/n)\, Y_{i/n}\,, \qquad \hat{\beta}_{jk}^{\mathrm{F}} = n^{-1} \sum_{i=1}^{n} \psi_{jk}(i/n)\, Y_{i/n}\,.$$

Choose integers J^{S} and J^{F} such that

$$2^{J^{\mathrm{S}}} \asymp v_n/\log v_n\,, \quad 2^{J^{\mathrm{F}}} \asymp n/\log n\,, \tag{2.6}$$

and define the thresholds

$$\delta_n^{\mathrm{S}} = K^{\mathrm{S}}\,(v_n^{-1}\,\log v_n)^{1/2}\,, \quad \delta_n^{\mathrm{F}} = K^{\mathrm{F}}\,(n^{-1}\,\log n)^{1/2}\,, \tag{2.7}$$

where K^{S} and K^{F} are constants.

In this notation, the conventional wavelet estimator computed on the slow grid is

$$\hat{g}^{\mathrm{S}}(u) = \sum_\ell \hat{\alpha}_\ell^{\mathrm{S}}\, \phi_\ell(u) + \sum_{j=j_0}^{J^{\mathrm{S}}} \sum_k \hat{\beta}_{jk}^{\mathrm{S}}\, I(|\hat{\beta}_{jk}^{\mathrm{S}}| > \delta_n^{\mathrm{S}})\, \psi_{jk}(u)\,. \tag{2.8}$$

Its fast-grid version, $\hat{g}^{\mathrm{F}}$, is obtained by changing S to F throughout.

2.3. Dual-rate estimator

The dual-rate estimator $\hat{g}^*$ is constructed in a block-wise fashion, taking the intervals $\mathcal{I}_{j_0\ell}$ as the blocks. These may be addressed in arbitrary order, but practice is better reflected if we imagine that time increases as we pass from the left- to the right-hand side of $\mathcal{I}$. Thus, having computed $\hat{g}^*$ on $[0, (\ell-1)/p]$, the next step is to calculate it on $\mathcal{I}_{j_0\ell} = ((\ell-1)/p, \ell/p]\,.$

If the wavelets are from the Haar sequence then all the wavelet functions that influence $\hat{g}^{\mathrm{S}}$ or $\hat{g}^{\mathrm{F}}$ on $\mathcal{I}_{j_0\ell}$ are supported in that interval. More generally, however, ψ and ϕ are supported on intervals that are of greater than unit length, and consequently there is some spillage of influence from either side of $\mathcal{I}_{j_0\ell}$. Therefore it is helpful to describe the "domain of influence" of $\mathcal{I}_{j_0\ell}$.

Let $\mathcal{S}_{jk}$ and $\mathcal{T}_k$ denote the supports of ψ_{jk} and ϕ_k, respectively. Define $\mathcal{I}(\ell)$, the domain of influence of $\mathcal{I}_{j_0\ell}$, to be the union of $\mathcal{S}_{jk}$ over all pairs (j,k) such that $j \ge j_0$ and $\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \ne \emptyset$, together with the union of $\mathcal{T}_k$ over all k such that $\mathcal{T}_k \cap \mathcal{I}_{j_0\ell} \ne \emptyset$. If $\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \ne \emptyset$, then we might fairly say that the corresponding wavelet coefficient β_{jk} "falls within the domain of influence of g on $\mathcal{I}_{j_0\ell}$."

Excepting small effects at the endpoints 0 and 1 of $\mathcal{I}$, we may interpret $\mathcal{I}(\ell)$ as the smallest subset of $\mathcal{I}$ such that data drawn at any timepoint $u \in \mathcal{I}(\ell)$ might affect the value of either $\hat{g}^{\mathrm{S}}$ or $\hat{g}^{\mathrm{F}}$. Recall that the length of $\mathcal{I}_{j_0\ell}$ equals p^{-1}; the length of each $\mathcal{I}(\ell)$ equals $O(p^{-1})$, uniformly in ℓ.

The first step in computing $\hat{g}^*$ on $\mathcal{I}_{j_0\ell}$ is to calculate all the slow-grid empirical wavelet coefficients $\hat{\beta}_{jk}^{\mathrm{S}}$, for $j_0 \le j \le J^{\mathrm{S}}$ and such that $\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \ne \emptyset$,

and check whether any of them exceed (in absolute value) a pre-determined, and relatively large, "rate-switching" threshold Δ_n, say. If one argues that the slow grid is essentially designed for estimating functions whose regularity is not less than s, in the sense that they are elements of Λ^s, and the fast grid is primarily for estimating functions of lower regularity in $\Lambda^t \setminus \Lambda^s$, where $t \in (1/2, s)$, then one should choose Δ_n towards the upper extremity of the absolute values of coefficients β_{jk} in the case of functions from Λ^s. A guide to the size of Δ_n is given by (2.4): consider taking Δ_n to be of smaller order than $2^{-j_0\{s+(1/2)\}}$.

Note particularly that the rate-switching threshold has a completely different purpose, and consequently a very different size, from the conventional thresholds δ_n^{S} and δ_n^{F}. The latter serve to distinguish an empirical wavelet coefficient from fluctuations due to noise. On the other hand, Δ_n separates (in an approximate way) wavelet coefficients whose size reflects functions in Λ^s, from coefficients whose size reflects functions in $\Lambda^t \setminus \Lambda^s$ where $1/2 < t < s$.

Define the event

$$\mathcal{E}_\ell^{\mathrm{S}} \ : \ \text{"for all } (j,k) \text{ with } j_0 \le j \le J^{\mathrm{S}} \text{ and } \mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \ne \emptyset,\ |\hat\beta_{jk}^{\mathrm{S}}| \le \Delta_n \text{ ,"} \tag{2.9}$$

and let

$$\hat g^* = \begin{cases} \hat g^{\mathrm{S}} & \text{on } \mathcal{I}_{j_0\ell} \text{ if } \mathcal{E}_\ell^{\mathrm{S}} \text{ obtains,} \\ \hat g^{\mathrm{F}} & \text{on } \mathcal{I}_{j_0\ell} \text{ otherwise.} \end{cases} \tag{2.10}$$

Therefore, our algorithm requires slow-grid sampling to be undertaken throughout $\mathcal{I}$, albeit in a block-wise manner. The overlapping influence of neighbouring blocks means that, except in the case of Haar wavelets, some data on the slow grid will be used more than once − indeed, once for each interval $\mathcal{I}(\ell)$ into which the corresponding timepoints fall. This causes no difficulty however.

For simplicity we assume that when it is necessary to sample at the fast rate on $\mathcal{I}_{j_0\ell}$, the necessary data are drawn independently of the slow-rate data. This assumption is made explicit at (3.7) below. However, since the ratio of the sizes of the fast- and slow-grid samples diverges to infinity as $n \to \infty$, making the fast-grid data completely independent of their slow-grid counterparts involves only negligible loss of efficiency. Nevertheless, with a little extra analysis one may treat the case where the fast-rate sample is drawn only at fast-grid points where no slow-grid data were previously available. (This possibility is permitted by (2.2).) That is, when we decide to use $\hat g^{\mathrm{F}}$ instead of $\hat g^{\mathrm{S}}$ on $\mathcal{I}_{j_0\ell}$ we simply adjoin some extra data to the slow-grid sample, to convert it to a fast-grid sample.

The sampling rule described above is an abstraction of that suggested by Hall and Penev (2002). It captures the essential features of the latter, but is more amenable to mathematical study of its performance over function classes. It can be converted to an online algorithm by, for example, increasing the sampling rate

in the next interval $\mathcal{I}_{j_0,\ell-1}$, rather than the current interval $\mathcal{I}_{j_0\ell}$, if the threshold Δ_n is exceeded in $\mathcal{I}_{j_0\ell}$, and likewise delaying a return to the slower sampling rate. Theoretical properties of this lagged approach, for functions where high-frequency episodes last for more than one block, are in many respects similar to those of the slightly simpler technique on which it is based.

2.4. Multi-rate estimators

In the case of Q sampling rates, and Q function classes $\Lambda^{s_1} \subseteq \ldots \subseteq \Lambda^{s_Q}$ with regularities satisfying $r > s_1 > \ldots > s_Q > 1/2$, we determine thresholds $0 < \Delta_n^{(1)} \leq \ldots \leq \Delta_n^{(Q)} = \infty$ in such a way that $\Delta_n^{(q)}$ is larger than the wavelet-coefficient bound for functions in $\Lambda^{s_{q-1}}$, but less than the bound for functions in Λ^{s_q}, for each q. For example, we might take $\Delta_n^{(q)} = K_q\, 2^{-j_0\{u_q+(1/2)\}}$, where $u_1 > s_1 > u_2 > s_2 > \ldots > s_{Q-1} > 0 > u_Q = -\infty$, and the constants satisfied $K_q > 0$. There are likewise Q sampling rates, $v_{nq} = n^{\xi_q}$ for $1 \leq q \leq Q$, where $\xi_1 < \ldots < \xi_Q$. One algorithm for switching rates is the following: work primarily with the lowest rate, v_{n0}, which we continue to call the slow rate and for which the empirical wavelet coefficients still have the notation $\hat{\beta}_{jk}^{\mathrm{S}}$. (We assume $p = 2^{j_0} \asymp v_{n0}^{1/(2r+1)}$.) Within interval $\mathcal{I}_{j_0\ell}$, switch to rate v_{nq} if q is the largest integer such that $|\hat{\beta}_{jk}^{\mathrm{S}}| > \Delta_n^{(q)}$ for some (j,k) with $j_0 \leq j \leq J^{\mathrm{S}}$ and $\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \neq \emptyset$; and use the slow rate if each $|\hat{\beta}_{jk}^{\mathrm{S}}|$ is less than $\Delta_n^{(1)}$. There are of course variants of this scheme, some of which involve making thresholding decisions using sampling rates that are faster than the slowest.

3. Theoretical Properties

3.1. Main result, and discussion

Our main result is derived in Section 4. Recall from Section 2 that our aim is to use the slow and fast rates to estimate g when it is in Λ^s and in $\Lambda^t \setminus \Lambda^s$, respectively, where $1/2 < t < s$. To assess the potential performance of our empirical rate-switching rule, consider the deterministic event

$$\mathcal{A}_\ell(g) \;:\; \text{“}|\beta_{jk}| \leq (1+\eta)\,\Delta_n \text{ for all } (j,k) \text{ with } j_0 \leq j \leq J^{\mathrm{S}} \text{ and } \mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \neq \emptyset\,\text{,”} \tag{3.1}$$

and let $\mathcal{A}_{\ell 0} = \mathcal{A}_{\ell 0}(g)$ denote the version of $\mathcal{A}_\ell$ that arises when $\eta = 0$. In a sense, an “ideal rate-switching rule” is one that results in either $\hat{g}^* = \hat{g}^{\mathrm{S}}$ or $\hat{g}^* = \hat{g}^{\mathrm{F}}$ on $\mathcal{I}_{j_0\ell}$, according as $\mathcal{A}_{\ell 0}$ is true or false, respectively. Thus, we would ideally like to use the fast grid only to estimate g when its wavelet coefficients are relatively large among functions in Λ^s.

Consider employing this rule to estimate $g \in \Lambda^t$. Functions of this type which satisfy $\mathcal{A}_{\ell 0}$ look like functions in Λ^s when constrained to $\mathcal{I}_{j_0\ell}$, and so

when estimated there using $\hat{g}^{\mathrm{S}}$, their integrated L^2 convergence rates should be bounded above by $p^{-1}\,(\delta_n^{\mathrm{S}})^{4s/(2s+1)}$. (To appreciate why, observe that the optimal uniform convergence rate for estimating functions in Λ^s on the unit interval is $O\{(\delta_n^{\mathrm{S}})^{4s/(2s+1)}\}$, modulo a logarithmic factor. If we confine attention to the subinterval $\mathcal{I}_{j_0\ell}$ of length p^{-1} then this should be multiplied by p^{-1}.) Using the ideal rule, we take $\hat{g}^* = \hat{g}^{\mathrm{F}}$ when $\mathcal{A}_{\ell 0}$ fails; there, the optimal rate of our estimator is $O\{p^{-1}\,(\delta_n^{\mathrm{F}})^{4t/(2t+1)}\}$, uniformly in $g \in \Lambda^t$.

Therefore computing $\hat{g}^*$ using the ideal rule, rather than the empirical rule suggested in Section 2, we aspire to the following L^2 convergence rate on $\mathcal{I}_{j_0\ell}$:

$$p^{-1}\left(I\{\mathcal{A}_{\ell 0}(g)\}\,(\delta_n^{\mathrm{S}})^{4s/(2s+1)} + [1 - I\{\mathcal{A}_{\ell 0}(g)\}]\,(\delta_n^{\mathrm{F}})^{4t/(2t+1)}\right), \tag{3.2}$$

uniformly in $g \in \Lambda^t$. Adding over $1 \le \ell \le p$ we see that the sought-after L^2 convergence rate on $\mathcal{I}$ is

$$\int_{\mathcal{I}} E|\hat{g}^* - g|^2 = O\Big[\,\rho(g)\,(\delta_n^{\mathrm{S}})^{4s/(2s+1)} + \{1 - \rho(g)\}\,(\delta_n^{\mathrm{F}})^{4t/(2t+1)}\,\Big], \tag{3.3}$$

uniformly in $g \in \Lambda^t$, where $\rho(g) = p^{-1}\sum_\ell I\{\mathcal{A}_{\ell 0}(g)\}$ denotes the proportion of indices ℓ for which the ideal rate-switching rule classifies the restriction of g to $\mathcal{I}_{j_0\ell}$ as being in the more regular function class Λ^s, rather than simply in Λ^t.

Result (3.3) implies that

$$\sup_{g\in\Lambda^t} \int_{\mathcal{I}} E|\hat{g}^* - g|^2 = O\{n^{-a}\,(\log n)^{-b}\}\,, \tag{3.4}$$

where

$$a = \min\left(\frac{2\xi s}{2s+1}\,,\ \frac{2t}{2t+1}\right), \quad b = \frac{2s}{2s+1}\,.$$

Provided the exponent, ξ, of the slow sampling rate is not less than $(2s+1)t/(2t+1)s$, which is always strictly less than 1 since $t < s$, (3.4) shows that dual-rate sampling allows us to maintain the optimal convergence rate at the value $(\delta_n^{\mathrm{S}})^{4s/(2s+1)}$ it would assume under slow-grid sampling for $g \in \Lambda^s$, despite the fact that we are actually addressing all functions in the less regular class Λ^t. Of course, the price to be paid is that we must sample more often (specifically, at rate n rather than rate n^ξ per unit time) when estimating functions that the rate-switching rule assesses to be in Λ^t but not in Λ^s.

However, if the functions g we are likely to observe will for the most part be in Λ^s rather than in $\Lambda^t \setminus \Lambda^s$, and if even those functions in $\Lambda^t \setminus \Lambda^s$ will look like Λ^s functions for most of the length of $\mathcal{I}$, then the price will rarely have to be paid, and more seldom still will it have to be paid in full. As argued in Section 1, dual-rate sampling is often intended to address this context, where relatively complex

functions (or relatively complex parts of functions) are present relatively rarely in data, and the higher sampling rate is needed only occasionally, to accommodate such cases. It is in these terms that (3.3) and (3.4), and likewise Theorem 1 below, provide theoretical justification, in terms of uniform convergence rates over Hölder classes, for dual-rate sampling.

Of course, the majority of this discussion applies to the ideal rule, not the empirical rule suggested in Section 2.3. However, (3.9) of Theorem 1 shows that convergence rates for the empirical rule have almost exactly the same property; the only change is that the slow- and fast-grid sampling fractions, ρ and $1-\rho$, are slightly altered. Similarly, (3.10) of the theorem shows that the empirical sampling fraction differs little from its ideal-rule version.

Next we give regularity conditions for Theorem 1. Assume $\Delta_n > 0$ satisfies

$$\Delta_n = o(p^{-\{s+(1/2)\}}), \quad (v_n/\log v_n)^{1/2}\,\Delta_n \to \infty \tag{3.5}$$

as $n \to \infty$. The first part of this condition ensures that Δ_n is of smaller order than the upper bound, $|\beta_{jk}| \le C\,p^{-\{s+(1/2)\}}$, which applies uniformly in $j \ge j_0$ and $g \in \Lambda^s$; see (2.4). Therefore, functions g in Λ^t that satisfy $\mathcal{A}_{\ell 0}(g)$ look like functions in Λ^s when viewed on $\mathcal{I}_{j_0\ell}$.

The second part of (3.5) asks that Δ_n be of strictly larger order than the threshold, δ_n^{S}, for the slow-rate estimator. One choice of Δ_n that satisfies both parts of (3.5) is $\Delta_n = K n^{-\xi(2u+1)/2(2r+1)}$, where $K > 0$ is arbitrary and $s < u < r$.

Suppose the mesh, v_n, of the slow grid satisfies

$$v_n \asymp n^{\xi}, \quad \text{where} \quad 0 < \xi < 1. \tag{3.6}$$

Assume of the errors in the model at (2.1) that:

> the ϵ_u's for the slow grid (placed throughout $\mathcal{I}$) are independent and identically distributed as normal $\mathrm{N}(0,\sigma^2)$, and, conditional on the slow-grid data, the ϵ_u's for the fast grid (placed at places in $\mathcal{I}$ that are determined by the slow-grid data) are independent and identically normal $\mathrm{N}(0,\sigma^2)$. (3.7)

Suppose too that the threshold constants K^{S} and K^{F} satisfy

$$\min(K^{\mathrm{S}}, K^{\mathrm{F}}) \ge 2^{3/2}\,\sigma. \tag{3.8}$$

Write $\mathcal{B}_\ell = \mathcal{B}_\ell(g)$ for the event corresponding to $\mathcal{A}_\ell(g)$, defined at (3.1), when $1+\eta$ is replaced by $1-\eta$. Let $\hat\rho$ denote the proportion of indices ℓ, $1 \le \ell \le p$, such that $\hat g^* = \hat g^{\mathrm{S}}$ on $\mathcal{I}_{j_0\ell}$, and write $\rho_{\mathcal{A}}(g)$ for the proportion of indices ℓ for which the event $\mathcal{A}_\ell(g)$ holds for the function g: $\rho_{\mathcal{A}}(g) = p^{-1}\sum_{\ell=1}^{p} I\{\mathcal{A}_\ell(g)\}$. Let $\rho_{\mathcal{B}}(g)$ be its counterpart when $\mathcal{A}_\ell$ is replaced by $\mathcal{B}_\ell$ on the right-hand side.

Theorem 1. *Let $1/2 < t < s < r$. Assume conditions (2.5) – (2.7) and (3.5) – (3.8). Recall the definition of $\hat{g}^*$ at (2.10). Then, for each $\eta > 0$ in the definitions of $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{B}}$, we have as $n \to \infty$:*

$$\sup_{g \in \Lambda^t} \int_{\mathcal{I}} E|\hat{g}^* - g|^2 = O\Big[\rho_{\mathcal{A}}(g)\,(\delta_n^{\mathrm{S}})^{4s/(2s+1)} + \{1 - \rho_{\mathcal{B}}(g)\}\,(\delta_n^{\mathrm{F}})^{4t/(2t+1)}\Big]\,, \quad (3.9)$$

$$\inf_{g \in \Lambda^t} P_g\{\rho_{\mathcal{B}}(g) \le \hat{\rho} \le \rho_{\mathcal{A}}(g)\} \to 1\,. \quad (3.10)$$

The full force of the second part of (3.5) is needed only to prove (3.10). It may be replaced by the assumption that $\liminf_{n\to\infty} (v_n/\log v_n)^{1/2}\,\Delta_n > 0$ if (3.9) is the objective.

The methods used to derive Theorem 1 also lead to results analogous to (3.2), where performance is assessed on a local, block-wise basis. Such formulae have the advantage of showing more explicitly that the benefits of dual-rate sampling are available at a local level. However, the mean integrated squared error view taken at (3.9) is more standard, and therefore more easily interpreted. It reflects the local result through the sampling fractions $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{B}}$.

Theorem 1 has close analogues in the case of multi-rate sampling schemes. For example, when the scheme is that described in Section 2.4, (3.9) changes to

$$\sup_{g \in \Lambda^{s_1}} \int_{\mathcal{I}} E|\hat{g}^* - g|^2 = O\Big\{ \sum_{q=1}^{Q} \rho_q(g)\,(\delta_n^{(q)})^{4s_q/(2s_q+1)} \Big\}\,,$$

where $\delta_n^{(q)} = (n^{-\xi_q} \log n)^{1/2}$ and $\rho_q(g)$ denotes the proportion of indices ℓ for which q is the largest integer such that (a) $|\beta_{jk}| > (1-\eta)\Delta_n^{(q-1)}$ for some (j,k) with $j_0 \le j \le J^{\mathrm{S}}$ and $\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \ne \emptyset$, and (b) $|\beta_{jk}| \le (1+\eta)\Delta_n^{(q)}$ for all such (j,k). (We take $\Delta_n^{(Q)} = \infty$.)

The next two sections develop properties of wavelet estimators on small intervals, and are used in our proof of the theorem.

3.2. Index of local sparsity

We introduce an index, m_q, of the sparsity of wavelet coefficients of the signal, g. The index, which is especially simple to work with, bounds the number of large wavelet coefficients which exceed an arbitrary level. Since less regular signals have more large coefficients, then smaller values of m_q correspond to greater regularity. A global index has been discussed by Kerkyacharian and Picard (2000) for another purpose; the present index is local.

Let $\mathcal{J}$ denote a subinterval of $\mathcal{I} = [0,1]$, with length $|\mathcal{J}|$ not less than 2^{-j_0}. Let g be a function admitting the wavelet expansion at (2.3), and put

$$m_q(j_0, \mathcal{J}) = \sup_{\lambda > 0} \mathrm{Card}\,\Big\{ (j,k) \,:\, j \ge j_0\,,\ |\beta_{jk}| > \lambda\,,\ \mathcal{S}_{jk} \cap \mathcal{J} \ne \emptyset \Big\}\, \lambda^q\,.$$

The lemma below shows that if $m_q(j_0, \mathcal{J})$ is finite then the number of wavelet coefficients β_{jk} exceeding λ in absolute value, with their indices k in the interval $\mathcal{J}$ and their levels $j \geq j_0$, is always less than a constant multiple of λ^q.

Recall that $p = p(j_0) = 2^{j_0}$. Fix $s > 0$, and put $q = (s + 1/2)^{-1}$.

Lemma 1. *If $g \in \Lambda^s(\gamma)$ then for any interval $\mathcal{J}$ of length $|\mathcal{J}|$ satisfying $|\mathcal{J}| \geq p^{-1}$, $m_q(j_0, \mathcal{J}) \leq C\,|\mathcal{J}|$, where $C > 0$ depends only on s, γ and the support of the wavelet.*

To derive the lemma, write $\log_2$ to indicate logarithms to base 2, let $C_1 > 0$ and let j_λ denote the integer part of $(s+1/2)^{-1}\,|\log_2 \lambda| + C_1$. Since $g \in \Lambda^s$ then, for any $j \geq 0$, $\sup_k |\beta_{jk}| \leq C_2\, 2^{-j\{s+(1/2)\}}$, where $C_2 = \|g\|_{s\infty\infty}$. Therefore, if C_1 is sufficiently large and $j > j_\lambda$ then $|\beta_{jk}| \leq \lambda$. Hence,

$$\begin{aligned}\text{Card}\Big\{ (j,k) \,:\, j \geq j_0\,,\ |\beta_{jk}| > \lambda\,,\ \mathcal{S}_{jk} \cap \mathcal{J} \neq \emptyset \Big\} \\ \leq C_3 \sum_{j=j_0}^{j_\lambda} 2^j\,|\mathcal{J}| \leq 2\,C_3\,2^{j_\lambda}\,|\mathcal{J}| \leq C_4\,|\mathcal{J}|\,\lambda^{-q}\,,\end{aligned}$$

uniformly in $g \in \Lambda^s$, where $C_3, C_4 > 0$ depend only on s, $\|g\|_{s\infty\infty}$ and the support of the wavelet.

3.3. Local performance of thresholding algorithms

Let

$$\hat{g}(u) = \sum_\ell \hat{\alpha}_\ell\,\phi_\ell(u) + \sum_{j=j_0}^{J} \sum_k \hat{\beta}_{jk}\,I(|\hat{\beta}_{jk}| > K\,c_n)\,\psi_{jk}(u) \tag{3.11}$$

denote a general wavelet estimator, in which the wavelet coefficients $\hat{\alpha}_\ell$ and $\hat{\beta}_{jk}$ are calculated in an as-yet unspecified way and the threshold is a constant multiple, K, of a sequence of constants c_n converging to zero. We choose J and p, the latter in the definitions of ϕ_ℓ and ψ_{jk}, such that

$$2^J \asymp c_n^{-2}\,, \quad p \leq C\,c_n^{-2/(2r+1)}\,, \tag{3.12}$$

where, here and below, $C > 0$ is a constant not depending on n. We assume $r > s$, and give a bound for the integral of $E(\hat{g} - g)^2$ over small intervals $\mathcal{J}$. See also Cohen, DeVore, Kerkyacharian and Picard (2001) and Kerkyacharian and Picard (2000).

Recall that $\mathcal{S}_{jk}$ denotes the support of ψ_{jk}, and let $\mathcal{T}_\ell$ denote the support of ϕ_ℓ. We assume that the coefficients $\hat{\alpha}_\ell$ and $\hat{\beta}_{jk}$ satisfy:

$$\text{for all } \ell \text{ such that } \mathcal{T}_\ell \cap \mathcal{J} \neq \emptyset\,, \quad E|\hat{\alpha}_\ell - \alpha_\ell|^2 \leq C\,c_n^2\,, \tag{3.13}$$

for all (j,k) such that $j_0 \le j \le J$ and $\mathcal{S}_{jk} \cap \mathcal{J} \neq \emptyset$, $E|\hat\beta_{jk} - \beta_{jk}|^4 \le C c_n^4$, (3.14)

for all (j,k) such that $j_0 \le j \le J$ and $\mathcal{S}_{jk} \cap \mathcal{J} \neq \emptyset$,

$$P(|\hat\beta_{jk} - \beta_{jk}| \ge \frac{1}{2} K c_n) \le C c_n^4, \quad (3.15)$$

where α_ℓ and β_{jk} are the wavelet coefficients in the expansion at (2.3).

As in Lemma 1, put $q = (s + 1/2)^{-1}$. Let $\mathcal{C}$ denote the class of intervals $\mathcal{J} \subseteq \mathcal{I}$ such that $|\mathcal{J}| \ge C p^{-1}$, where $C > 0$ is arbitrary.

Proposition 1. *Assume $r > s$ and that (3.12)–(3.15) hold. Let $\hat g$ be the estimator defined at (3.11). Then, as $n \to \infty$,*

$$\int_{\mathcal{J}} E|\hat g - g|^2 = O\{c_n^{4s/(2s+1)}\, m_q(j_0, \mathcal{J})\} + o(c_n^{4s/(2s+1)}\, |\mathcal{J}|) + \sum_{j=J+1}^{\infty} \sum_{k\,:\,\mathcal{S}_{jk} \cap \mathcal{J} \neq \emptyset} \beta_{jk}^2, \quad (3.16)$$

uniformly in $\mathcal{J} \in \mathcal{C}$ and functions $g \in \Lambda^s$.

Of course, the proposition continues to hold if we replace s by t throughout. It is proved in Section 4, and implies a variety of bounds to convergence rates, on various spaces. See, for example, Donoho and Johnstone (1995). We apply it here to the case where $\hat g = \hat g^{\mathrm{S}}$ or $\hat g^{\mathrm{F}}$, to show how local convergence rates depend on respective sampling rates. Recall that the thresholds δ_n^{S} and δ_n^{F} involve the constants K^{S} and K^{F}, respectively.

Proposition 2. *Let $\hat g^{\mathrm{S}}$ denote the slow-grid estimator defined at (2.8), and let $\hat g^{\mathrm{F}}$ be its fast-grid counterpart, obtained by changing* S *to* F *throughout (2.8). Assume conditions (2.5)–(2.7) on the parameters used in the estimator, that $1/2 < s < t < r$ and $\min(K^{\mathrm{S}}, K^{\mathrm{F}}) \ge 2^{3/2}\, \sigma$, and that the errors ϵ_u at (2.1) are independent and normal $N(0, \sigma^2)$. Then, as $n \to \infty$,*

$$\sup_{g \in \Lambda^s} \int_{\mathcal{J}} E|\hat g^{\mathrm{S}} - g|^2 = O\Big\{|\mathcal{J}|\, (\delta_n^{\mathrm{S}})^{4s/(2s+1)}\Big\}, \quad (3.17)$$

$$\sup_{g \in \Lambda^t} \int_{\mathcal{J}} E|\hat g^{\mathrm{F}} - g|^2 = O\Big\{|\mathcal{J}|\, (\delta_n^{\mathrm{F}})^{4t/(2t+1)}\Big\}, \quad (3.18)$$

uniformly in $\mathcal{J} \in \mathcal{C}$.

4. Proofs of Propositions and Theorem

4.1. Proof of Proposition 1

For $u \in \mathcal{J}$ we may write

$$\hat g(u) - g(u) = \sum_{\ell\,:\,\mathcal{T}_\ell \cap \mathcal{J} \neq \emptyset} (\hat\alpha_\ell - \alpha_\ell)\, \phi_\ell(u)$$

$$+\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}\{\hat\beta_{jk}\,I(|\hat\beta_{jk}|>K\,c_n)-\beta_{jk}\}\,\psi_{jk}(u)$$
$$-\sum_{j=J+1}^{\infty}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}\beta_{jk}\,\psi_{jk}(u)\,.$$

The integral of the square of the left-hand side over the interval $\mathcal{J}$ is bounded above by the integral of the square of the right-hand side over the whole real line. Hence, by orthogonality,

$$\int_{\mathcal{J}}|\hat g-g|^2\le S_1+S_2+S_3\,, \tag{4.1}$$

where

$$S_1=\sum_{\ell\,:\,\mathcal{T}_\ell\cap\mathcal{J}\neq\emptyset}(\hat\alpha_\ell-\alpha_\ell)^2\,,\quad S_2=\sum_{j=J+1}^{\infty}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}\beta_{jk}^2\,,$$
$$S_3=\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}\{\hat\beta_{jk}\,I(|\hat\beta_{jk}|>K\,c_n)-\beta_{jk}\}^2\,.$$

In view of (3.12) and (3.13),

$$E(S_1)\le C_1\,c_n^2\,\mathrm{Card}\,\{\ell\,:\,\mathcal{T}_\ell\cap\mathcal{J}\neq\emptyset\}\le C_1\,p\,c_n^2\,|\mathcal{J}|\le C_2\,c_n^{-4r/(2r+1)}\,|\mathcal{J}|\,, \tag{4.2}$$

where, here and below, $C_1, C_2, \ldots$ denote constants not depending on n or $\mathcal{J}$. Furthermore, $1/2\,E(S_3)\le s_{31}+s_{32}$ and $s_{3j}\le s_{3j1}+s_{3j2}$, where

$$s_{31}=\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}E\{(\hat\beta_{jk}-\beta_{jk})^2\,I(|\hat\beta_{jk}|>K\,c_n)\}\,,$$
$$s_{32}=\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}\beta_{jk}^2\,P(|\hat\beta_{jk}|\le K\,c_n)\,,$$
$$s_{311}=\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}E\{(\hat\beta_{jk}-\beta_{jk})^2\,I(|\hat\beta_{jk}-\beta_{jk}|\ge\frac{1}{2}\,K\,c_n)\}\,,$$
$$s_{312}=\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}E(\hat\beta_{jk}-\beta_{jk})^2\,I(|\beta_{jk}|\ge\frac{1}{2}\,K\,c_n)\,,$$
$$s_{321}=\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}\beta_{jk}^2\,P(|\hat\beta_{jk}-\beta_{jk}|>K\,c_n)\,,$$
$$s_{322}=\sum_{j=j_0}^{J}\sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset}\beta_{jk}^2\,I(|\beta_{jk}|\le 2\,K\,c_n)\,.$$

Define $\mathcal{J}' = \bigcup_{j=J+1}^{\infty} \bigcup_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset} \mathcal{S}_{jk}$. Using (3.12), (3.14), (3.15) and the Cauchy-Schwartz inequality, we may prove that

$$
\begin{aligned}
s_{311} &\le C\,c_n^4 \sum_{j=j_0}^{J} \sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset} 1 \le C_3\,c_n^4\,2^J\,|\mathcal{J}'| \le C_4\,c_n^2\,|\mathcal{J}'| \le C_5\,c_n^2\,|\mathcal{J}|\,,\\
s_{312} &\le C\,c_n^2\,\mathrm{Card}\Big\{\,(j,k)\;:\;j_0\le j\le J\,,\;|\beta_{jk}|\ge\frac{1}{2}\,K\,c_n\,,\;\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset\,\Big\}\\
&\le C_6\,c_n^2\,m_q(j_0,\mathcal{J})\,(K\,c_n/2)^{-q} \le C_7\,c_n^{4s/(2s+1)}\,m_q(j_0,\mathcal{J})\,,\\
s_{321} &\le C\,c_n^4 \sum_{j=j_0}^{J} \sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset} \beta_{jk}^2 \le C_8\,c_n^4\,|\mathcal{J}|\,,\\
s_{322} &\le \int_0^{2Kc_n} 2x\,\mathrm{Card}\Big\{\,(j,k)\;:\;j_0\le j\le J\,,\;|\beta_{jk}|\ge x\,,\;\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset\,\Big\}\,dx\\
&\le C_9\,m_q(j_0,\mathcal{J})\int_0^{2Kc_n} x\cdot x^{-q}\,dx \le C_{10}\,m_q(j_0,\mathcal{J})\,c_n^{2-q}\\
&= C_{10}\,c_n^{4s/(2s+1)}\,m_q(j_0,\mathcal{J})\,.
\end{aligned}
$$

Proposition 1 follows on combining (4.1) with the bounds from (4.2) down.

4.2. Proof of Proposition 2

(We drop the superscripts S and F, working on each occasion with either grid.) First we establish the regularity conditions for Proposition 1. There, $c_n = (v_n^{-1}\log v_n)^{1/2}$ or $c_n = (n^{-1}\log n)^{1/2}$ in the slow- or fast-grid cases, respectively. Since $\min(s,t) > 1/2$, standard methods show that for both types of grid, $(E\hat\alpha_\ell - \alpha_\ell)^2 = O(c_n^2)$ uniformly in ℓ such that $\mathcal{T}_\ell\cap\mathcal{J}\neq\emptyset$, and $(E\hat\beta_{jk}-\beta_{jk})^2 = O(c_n^2)$ uniformly in (j,k) such that $j_0\le j\le J$ and $\mathcal{S}_{jk}\cap\mathcal{I}\neq\emptyset$. Results (3.13) and (3.14) now follow from standard moment bounds, and (3.15) from a standard moderate deviation bound. The first part of (3.12) follows from (2.6), and the second from (2.5).

It remains to show that the bound at (3.16) implies those at (3.17) and (3.18). By Lemma 1, $m_q(j_0,\mathcal{J}) \le C\,|\mathcal{J}|$. Since $g\in\Lambda^s$ then it follows from (2.4) that $|\beta_{jk}| \le C_1\,2^{-j(2s+1)/2}$. Therefore,

$$
\begin{aligned}
\sum_{j=J+1}^{\infty} \sum_{k\,:\,\mathcal{S}_{jk}\cap\mathcal{J}\neq\emptyset} \beta_{jk}^2 &\le C_2\,|\mathcal{J}| \sum_{j=J+1}^{\infty} 2^{-j(2s+1)} \le C_3\,|\mathcal{J}|\,2^{-2Js}\\
&\le C_3\,|\mathcal{J}|\,2^{-2Js/(2s+1)} \le C_4\,|\mathcal{J}|\,\delta_n^{4s/(2s+1)}\,.
\end{aligned}
$$

4.3. Proof of (3.9)

Recall the definition of $\mathcal{E}_\ell^{\mathrm{S}}$ at (2.9). Write $\mathcal{E}_\ell^{\mathrm{F}}$ for the complement of $\mathcal{E}_\ell^{\mathrm{S}}$, and put

$$
a_\ell^{\mathrm{S}} = \int_{\mathcal{I}_{j_0\ell}} E\Big\{|\hat g^{\mathrm{S}}(t) - g(t)|^2\,I(\mathcal{E}_\ell^{\mathrm{S}})\Big\}\,dt\,.
$$

Define a_ℓ^{F} to be the same quantity but with S replaced by F throughout. Then,

$$a_\ell \equiv \int_{\mathcal{I}_{j_0\ell}} E|\hat{g}^* - g|^2 = a_\ell^{\mathrm{S}} + a_\ell^{\mathrm{F}}\,. \tag{4.3}$$

For arbitrary $\eta > 0$ put $\mathcal{A}_\ell^{\mathrm{S}} = \mathcal{A}_\ell = \mathcal{A}_\ell(g)$, the latter defined at (3.1), and let $\mathcal{A}_\ell^{\mathrm{F}}$ denote the complement of $\mathcal{A}_\ell^{\mathrm{S}}$. Put

$$a_\ell^{\mathrm{SS}} = \int_{\mathcal{I}_{j_0\ell}} E\Big\{|\hat{g}^{\mathrm{S}}(t) - g(t)|^2\, I(\mathcal{A}_\ell^{\mathrm{S}} \cap \mathcal{E}_\ell^{\mathrm{S}})\Big\}\, dt\,,$$

and let a^{SF} be the same quantity but with $\mathcal{A}_\ell^{\mathrm{S}}$ replaced by $\mathcal{A}_\ell^{\mathrm{F}}$. In this notation,

$$a_\ell^{\mathrm{S}} = a_\ell^{\mathrm{SS}} + a_\ell^{\mathrm{SF}}\,. \tag{4.4}$$

Let $\mathcal{B}_\ell^{\mathrm{S}} = \mathcal{B}_\ell$ be as defined in Section 3, and write $\mathcal{B}_\ell^{\mathrm{F}}$ for the complement of $\mathcal{B}_\ell^{\mathrm{S}}$. Put

$$a_\ell^{\mathrm{FS}} = \int_{\mathcal{I}_{j_0\ell}} E\Big\{|\hat{g}^{\mathrm{F}}(t) - g(t)|^2\, I(\mathcal{B}_\ell^{\mathrm{S}} \cap \mathcal{E}_\ell^{\mathrm{F}})\Big\}\, dt\,,$$

and let a_ℓ^{FF} be the same quantity but with $\mathcal{B}_\ell^{\mathrm{S}}$ replaced by $\mathcal{B}_\ell^{\mathrm{F}}$. Then,

$$a_\ell^{\mathrm{F}} = a_\ell^{\mathrm{FS}} + a_\ell^{\mathrm{FF}}\,. \tag{4.5}$$

Combining (4.3)–(4.5) we deduce that

$$a_\ell \le I(\mathcal{A}_\ell) \int_{\mathcal{I}_{j_0\ell}} E\{|\hat{g}^{\mathrm{S}}(t) – g(t)|^2\}dt + \{1 – I(\mathcal{B}_\ell)\} \int_{\mathcal{I}_{j_0\ell}} E\{|\hat{g}^{\mathrm{F}}(t) – g(t)|^2\}dt + a_\ell^{\mathrm{SF}} + a_\ell^{\mathrm{FS}}. \tag{4.6}$$

Let $\mathcal{F}_\ell$ denote the event that $|\hat{\beta}_{jk}^{\mathrm{S}} - \beta_{jk}| > \eta\Delta_n$ for some (j,k) with $j_0 \le j \le J^{\mathrm{S}}$ and $\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \ne \emptyset$. In this notation, $\mathcal{A}_\ell^{\mathrm{F}} \cap \mathcal{E}_\ell^{\mathrm{S}} \subseteq \mathcal{F}_\ell$ and

$$a_\ell^{\mathrm{SF}} + a_\ell^{\mathrm{FS}} \le \int_{\mathcal{I}_{j_0\ell}} \{b^{\mathrm{S}}(t) + b^{\mathrm{F}}(t)\}\, dt\,, \tag{4.7}$$

where $b^{\mathrm{S}}(t) = E\{|\hat{g}^{\mathrm{S}}(t) - g(t)|^2\, I(\mathcal{F}_\ell)\}$ and b^{F} denotes the same quantity but with $\hat{g}^{\mathrm{F}}$ on the right-hand side. We prove that

$$\sup_{g \in \Lambda^t} \sum_{\ell=1}^{p} \int_{\mathcal{I}_{j_0\ell}} b^{\mathrm{S}}(t)\, dt = O\Big[\min\Big\{(\delta_n^{\mathrm{S}})^{4s/(2s+1)}, (\delta_n^{\mathrm{F}})^{4t/(2t+1)}\Big\}\Big]\,. \tag{4.8}$$

An identical bound may be derived for the integral of b^{F}.

Using the argument leading to (4.3), and taking $\mathcal{J}$ there to equal $\mathcal{I}_{j_0\ell}$, we deduce that

$$\int_{\mathcal{I}_{j_0\ell}} b^{\mathrm{S}}(t)\, dt \le b_1(\ell) + b_2(\ell) + b_3(\ell) \tag{4.9}$$

where, by the Cauchy-Schwarz inequality and the bound $J^{\mathrm{S}} \le C_1 \log n$,

$$
\begin{aligned}
b_1(\ell) &= \sum_{k\,:\,\mathcal{T}_k \cap \mathcal{I}_{j_0\ell} \neq \emptyset} E\{(\hat\alpha_k^{\mathrm{S}} - \alpha_k)^2\, I(\mathcal{F}_\ell)\} \\
&\le P(\mathcal{F}_\ell)^{1/2} \sum_{k\,:\,\mathcal{T}_k \cap \mathcal{I}_{j_0\ell} \neq \emptyset} \{E(\hat\alpha_k^{\mathrm{S}} - \alpha_k)^4\}^{1/2} \le C_2\, P(\mathcal{F}_\ell)^{1/2}\, (\delta_n^{\mathrm{S}})^2\,, \\
b_2(\ell) &= \sum_{j=j_0}^{J^{\mathrm{S}}} \sum_{k\,:\,\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \neq \emptyset} E\Big[\{\hat\beta_{jk}^{\mathrm{S}}\, I(|\hat\beta_{jk}^{\mathrm{S}}| > K^{\mathrm{S}}\, c_n) - \beta_{jk}\}^2\, I(\mathcal{F}_\ell)\Big] \\
&\le C_2 \Big[\{E(\hat\beta_{jk} - \beta_{jk})^2\}^{1/2}\, P(\mathcal{F}_\ell)^{1/2} + P(\mathcal{F}_\ell)\Big]\, \log n \\
&\le C_3 \Big\{P(\mathcal{F}_\ell)^{1/2}\, (\delta_n^{\mathrm{S}})^2 + P(\mathcal{F}_\ell)\Big\}\, \log n\,, \\
b_3(\ell) &= P(\mathcal{F}_\ell) \sum_{j=J^{\mathrm{S}}+1}^{\infty} \sum_{k\,:\,\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \neq \emptyset} \beta_{jk}^2 \le C_2\, P(\mathcal{F}_\ell)\,,
\end{aligned}
$$

uniformly in $g \in \Lambda^t$. Since $\Delta_n/\delta_n^{\mathrm{S}} \to \infty$ (see the second part of (3.5)) then, using a bound for a moderate deviations of a normal random variable, $P(\mathcal{F}_\ell) = O(n^{-B})$ for all $B > 0$, the order relation holding uniformly in $g \in \Lambda^t$ and $1 \le \ell \le p$. Combining the results from (4.9) down we deduce that $\sum_{\ell=1}^p \int_{\mathcal{I}_{j_0\ell}} b^{\mathrm{S}}(t)\, dt = O(n^{-1})$, uniformly in $g \in \Lambda^t$. This implies (4.8).

Combining (4.3), (4.6), (4.7), (4.8) and the analogue of the latter for b^{F} rather than b^{S}, we deduce that

$$
\begin{aligned}
\int_{\mathcal{I}} E|\hat g^* - g|^2 = O\Big[&\sum_{\ell=1}^p I(\mathcal{A}_\ell) \int_{\mathcal{I}_{j_0\ell}} E|\hat g^{\mathrm{S}} - g|^2 + \sum_{\ell=1}^p \{1 - I(\mathcal{B}_\ell)\} \int_{\mathcal{I}_{j_0\ell}} |\hat g^{\mathrm{F}} - g|^2 \\
&+ \min\Big\{(\delta_n^{\mathrm{S}})^{4s/(2s+1)}, (\delta_n^{\mathrm{F}})^{4t/(2t+1)}\Big\}\Big]\,,
\end{aligned}
$$

uniformly in $g \in \Lambda^t$. Result (3.9) follows from this formula and Proposition 2. Note that when $\mathcal{A}_\ell = \mathcal{A}_\ell(g)$ holds and $g \in \Lambda^t$, we may equivalently view g as the restriction to $\mathcal{I}_{j_0\ell}$ of a function in Λ^s.

4.4. Proof of (3.10)

It suffices to show that

$$
\sup_{g \in \Lambda^t} \sum_{\ell=1}^p \sum_{j=j_0}^{J^{\mathrm{S}}} \sum_{k\,:\,\mathcal{S}_{jk} \cap \mathcal{I}_{j_0\ell} \neq \emptyset} \sup_{g \in \Lambda^t} \pi_{jk}(g) \to 0\,, \tag{4.10}
$$

where $\pi_{jk}(g) = P_g(|\hat\beta_{jk}^{\mathrm{S}} - \beta_{jk}| > 1/2\eta\Delta_n)$. For this it is adequate to prove that (4.10) holds with

$$
\pi_{jk}(g) = P_g(|\hat\beta_{jk}^{\mathrm{S}} - E\hat\beta_{jk}^{\mathrm{S}}| > \frac{1}{2}\,\eta\,\Delta_n) \tag{4.11}
$$

and with $\pi_{jk}(g) = I(|E\hat{\beta}^{\mathrm{S}}_{jk} - \beta_{jk}| > 1/2\eta\Delta_n)$. However, $|E\hat{\beta}^{\mathrm{S}}_{jk} - \beta_{jk}| = O(\delta^{\mathrm{S}}_n)$ uniformly in $g \in \Lambda^t$, since $t > 1/2$. Moreover, in view of the second part of (3.A1), $\delta^{\mathrm{S}}_n/\Delta_n \to 0$. Therefore it suffices to treat the case where π_{jk} is given by (4.11).

There we have, using a moderate deviation bound for the normal distribution, $\sup_{g\in\Lambda^t} \pi_{jk}(g) = O(n^{-B})$ for all $B > 0$. The number of indices k in the series at (4.10) is uniformly bounded, and so the number of indices (j, k, ℓ) there equals $O(pJ^{\mathrm{S}}) = O(n^{\xi/(2r+1)} \log n)$, uniformly in g. Therefore (4.10) holds.

References

Aldroubi, A. and Feichtinger, H. (1998). Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: the L_p-theory. *Proc. Amer. Math. Soc.* **126**, 2677-2686.

Aldroubi A. and Gröchenig, K. (2001). Nonuniform sampling and reconstruction in shift-invariant spaces. *SIAM Rev.* **43**, 585-620.

Chen, W., Itoh, S. and Shiki, J. (1998). Irregular sampling theorems for wavelet subspaces. *IEEE Trans. Inform. Theory* **44**, 1131-1142.

Cohen, A., DeVore, R., Kerkyacharian, G. and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.* **11**, 167-191.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 1200-1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (With discussion.) *J. Roy. Statist. Soc. Ser. B* **57**, 301-369.

Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **23**, 905-928.

Hall, P. and Penev, S. (2002). Wavelet-based signal analysis with multiple sampling rates. Manuscript.

Jetter, K. and Stöckler, J. (1997). Topics in scattered data interpolation and non-uniform sampling. In *Surface Fitting and Multiresolution Methods* (Edited by A. Le Méhautée, C. Rabut and L. L. Schumaker), 191-208. Vanderbilt Univ. Press, Nashville, Tennessee.

Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms and well-concentrated bases. *Test* **9**, 283-344.

Liu, Y. (1996). Irregular sampling for spline wavelet subspaces. *IEEE Trans. Inform. Theory* **42**, 623-627.

Liu, Y. and Walter, G. G. (1996). Irregular sampling in wavelet subspaces. *J. Fourier Anal. Appl.* **2**, 181-189.

Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia.

E-mail: peter.hall@maths.anu.edu.au

Département de Mathématiques, Université de Paris X, 92001 Nanterre, France.

E-mail: kerk@math.jussieu.fr

Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris VII, 75252 Paris, France

E-mail: picard@math.jussieu.fr

(Received December 2002; accepted June 2003)