

Adaptive Control with Multiresolution Bases

Christophe P. Bernard
Centre de Mathématiques Appliquées
Ecole Polytechnique
91128 Palaiseau cedex, FRANCE
bernard@cmapx.polytechnique.fr

Jean-Jacques E. Slotine
Nonlinear Systems Laboratory
Massachusetts Institute of Technology
Cambridge MA02139, USA
jjs@mit.edu

Abstract

This paper presents new results on adaptive nonlinear control using wavelet basis functions. First, it considers how to deal effectively with a potentially infinite number of unknown parameters, rather than using a priori truncations of wavelet expansions. This is done by constructing formally an ideal “infinite” controller, able to manage infinitely many unknown parameters in a convergent fashion, and only then designing a way to approximate its behavior with a finite controller. Besides being theoretically satisfying, the existence of such a consistent underlying infinite controller is easy to guarantee in practice and considerably improves convergence properties in the high frequency range of the unknown function. The paper also shows the advantages of specific multiresolution analysis wavelets over the “Mexican hat”-type wavelet frames now commonly used in control and learning applications. Finally, it discusses several possible constructions in the multidimensional case, and their properties.

Introduction

Adaptive control has been widely developed to cope with nonlinear plants with partially unknown behavior, and standard techniques can provide stable on-line adaptation to a finite number of unknown physical parameters (e.g., inertias or lengths). However, such approaches still require significant prior knowledge of the plant, and so there has been active research aimed at developing control techniques that would directly adapt to the whole system dynamics or to its unknown unparameterized components (see, e.g., [4] for a recent survey)

Wavelet networks are especially well suited for such function learning tasks, and the problem of adapting to an unknown dynamics can be replaced by that of adapting to the coefficients of its expansion in a wavelet base or a wavelet frame (e.g., [1], [8]), although the number of such coefficients can potentially be infinite. This paper derives a number of new technical results in this context.

In section 1, we focus on the design of the wavelets themselves, and show the advantages of particular multiresolu-

tion wavelet systems over functions commonly used “by default” in control applications, such as Mexican hat wavelets. In section 2, we discuss and compare several ways to design multidimensional wavelet representations. Section 3 describes the derivation of an ideal “infinite” adaptive controller, able to manage infinitely many unknown wavelet coefficients in a convergent fashion, and quantifies how its performance can be approximated by finite wavelet expansions.

1 Function approximation framework

Approximation of a function is mostly done by expanding this function over a basis, or more generally over any kind of family. Several such families have been used, like Fourier expansions, radial basis functions, Gaussian functions [7], and most lately wavelet families. Wavelets combine different advantages, like allowing inexpensive high accuracy approximation and low computational costs, and are therefore already massively used for signal and image compression [6].

1.1 Wavelets

Any function with some zero moments¹ can be considered as a possible mother wavelet for a wavelet frame [3]. Given a function $w(x)$ (e.g. the Mexican hat wavelet: $w(x) = d^2/dx^2 e^{-x^2}$, used in [1], [8]), the family $w_{jk}(x) = 2^{j/2}w(2^j x - k\Delta)$ for some step size $\Delta \in \mathbb{R}$ can be proven to be a frame of $L_2(\mathbb{R})$.

We will say that a wavelet basis is just a basis $(w_{jk})_{jk \in \mathbb{Z}}$ that has that dyadic structure

$$w_{jk}(x) = N(j)w(2^j x - k) \quad (1)$$

$N(j)$ is a normalization factor that will depend on the underlying topology.

Advantages of a wavelet structure The only expansion method that can compete with wavelets in terms of completeness and fast error decay is the complex exponential

¹A function f is said to have n zero moments iff $\int f(x)x^k dx = 0$ for $k = 0 \dots n - 1$, or equivalently $\hat{f}^{(k)}(0) = 0, k = 0 \dots n - 1$.

basis². Advantages of the wavelet basis structure are of several kinds:

Compression abilities Compared to a complex exponential basis, a wavelet basis has better compression abilities if the function we approximate is smooth with some local singularities [3][5][6]. In the case of a uniformly smooth function, both bases have similar compression abilities.

Computational efficiency is definitely an advantage of wavelet bases. Estimation of a $f(x)$ for a given x requires $O(\log N)$ computations if f is expanded over a wavelet basis and N computations for a complex exponential basis (where N is the number of coefficients retained after truncation, which is always smaller for wavelets). Moreover, gradient based function estimate updates have similar complexities, and thus also favor the use of wavelets.

Border effects As a last argument towards use of a wavelet structure, wavelets are much better designed to cope with approximation of functions over an interval while complex exponential functions can only well approximate functions over the real line, a torus, and any Cartesian product thereof [3][6].

What kind of wavelets? Usually, if the wavelet is designed so as to have a simple algebraic expression, the resulting family can at best be a frame. Also, such families cannot represent the low-frequency content of a function in a very economical way. For this reason we will use another approach called multiresolution analysis.

1.2 Multiresolutions

Multiresolution analysis [3][5][6] is a powerful tool that can easily provide families with a minimal redundancy: bases. Simple criteria are used to check whether the resulting families are orthogonal or interpolating, and to bound the decay rate of the expansion coefficients, and thus of the truncation error.

We will see that the minimal redundancy of the base is an advantage over a more general frame, since the approximation process in adaptive learning is usually closer to a gradient method (which is of low effectiveness in exploiting redundancy) than to matching pursuit (where redundancy is sought and thus frames are desirable).

1.3 Mother wavelet expansion

Most wavelet families are typically bases of $L_2(\mathbb{R})$. Given a mother wavelet ψ , the wavelet base is $(\psi_{jk})_{j,k \in \mathbb{Z}}$, where $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$. Any function in $L_2(\mathbb{R})$ can be

²Note for instance that Taylor expansions and fixed-grid Spline expansions are not complete representations.

expanded as

$$f = \sum_{j,k} c_{jk}(f) \psi_{jk}$$

where the right hand infinite sum converges in an $L_2(\mathbb{R})$ sense. More precisely,

$$\int \left(f - \sum_{-J..J} \sum_{-K..K} c_{jk} \psi_{jk} \right)^2 dx \rightarrow 0$$

as J and K both go to $+\infty$.

However, since the wavelets have a zero mean, truncations of such expansion approximate nonzero mean functions very poorly. To go around this problem, we will approximate separately the low frequency content of the function f with *father wavelets*.

1.4 Father and mother wavelet expansion

Multiresolution analysis theory provides an extra “father” wavelet [3][5][6] ϕ of non zero integral, that can be used to build hybrid expansions

$$f = \sum_{k=-\infty}^{+\infty} a_{j_0 k}(f) \phi_{j_0 k} + \sum_{j=j_0}^{+\infty} \sum_{k=-\infty}^{+\infty} c_{jk}(f) \psi_{jk}$$

The left hand sum represents the low frequency content of function f and the right hand sum the high frequency content. Such expansions are much sparser as the former ones, and can represent nonzero integral functions in a more natural way.

Multiresolution theory also provides a convenient way to compute dual wavelets $\tilde{\phi}$ and $\tilde{\psi}$ such that³

$$a_{jk}(f) = \langle \tilde{\phi}_{jk}, f \rangle \text{ and } c_{jk}(f) = \langle \tilde{\psi}_{jk}, f \rangle$$

The two following theorems, whose proofs can be found in the appendix, show that such expansion can approximate any continuous function, and give an estimate of the error decay depending on the function smoothness.

Theorem 1 *Let f be a continuous function defined over \mathbb{R} . The sequence of partial sums*

$$f_J = \sum_{k \in \mathbb{Z}} a_{0k} \phi_{0k} + \sum_{0 \leq j \leq J} \sum_{k \in \mathbb{Z}} c_{jk} \psi_{jk} \quad (2)$$

is equal to

$$\sum_{k \in \mathbb{Z}} a_{Jk} \phi_{Jk} \quad (3)$$

and converges to f uniformly on any compact set.

³The dual normalization factor to define $\tilde{\phi}_{jk} = \tilde{N}(j) \tilde{\phi}(2^j x - k)$ and $\tilde{\psi}_{jk} = \tilde{N}(j) \tilde{\psi}(2^j x - k)$ is $\tilde{N}(j) = 2^j / N(j)$. The inner product $\langle f, g \rangle$ is defined as $\langle f, g \rangle = \int_{\mathbb{R}} f(x) \overline{g(x)} dx$.

Theorem 2 *If f is d times differentiable, and if the dual wavelet has at least d zero moments, then there exists a bound M depending on the wavelet shape and on d , such that*

$$\sup |f - f_J| \leq M \sup |f^{(d)}| 2^{-Jd}$$

Usually M is close to 1. Also note that any polynomial of degree less than d will be approximated exactly with a sole father wavelet expansion $\sum_k a_{0k} \phi_{0k}$.

Remark. These theorems can be extended to higher dimensions quite simply. The resulting formulae grow significantly. An example can be given for dimension 2, where truncation error is bounded by

$$|f - f_J| < M \left(\sup \left| \frac{\partial^d f}{\partial x^d} \right| + \sup \left| \frac{\partial^d f}{\partial y^d} \right| \right) 2^{-dJ}$$

for some M depending on the wavelet shape and on the derivation order d .

1.5 Different kinds of multiresolutions

There exists many different kinds of multiresolutions, with two mainly used cases: orthogonal and interpolating multiresolutions. Orthogonality of a wavelet family speaks for itself, while for interpolating multiresolutions, the father wavelet is an interpolation function. The most known orthogonal wavelets are Daubechies wavelets, and the most common interpolating ones are Deslauriers-Dubuc wavelets. Some advantage of the latter are symmetry, higher smoothness, and a quite simple restriction to a finite interval. However, orthogonal wavelets are more natural when quadratic error metrics are used, since they then correspond to L_2 or Sobolev error metrics.

Deslauriers and Dubuc designed a dyadic refinement scheme to build a smooth interpolating function from a sequence of regularly spaced samples. Such interpolation processes are equivalent to building an interpolant $f(x)$ from the samples $f[k]$ as

$$f(x) = \sum_k f[k] \phi(x - k)$$

where ϕ is the Deslauriers-Dubuc interpolating father wavelet. This function is compactly supported and matches a two-scale refinement equation

$$\phi(x) = 2 \sum_{k \in \mathbb{Z}} m[k] \phi(2x - k)$$

and can therefore be used as a starting point to build a multiresolution analysis. Such interpolation scheme and function can have various support size and shape, depending on an integer parameter p .

Donoho suggested to define as a mother wavelet $\psi(x) = \phi(2x - 1)$. Unlike other wavelet families, the mother

wavelet has exactly the same shape as the father wavelet, except for that it is compressed by a factor of two. Moreover, it does not have a zero integral.

For interpolation wavelets, the most natural normalization factor is just $N(j) = 1$ (and thus $\tilde{N}(j) = 2^j$).

Remark. The dual wavelet family is not in $L_2(\mathbb{R})$. The dual father wavelet is

$$\tilde{\phi}(x) = \delta(x)$$

and the dual mother wavelet

$$\tilde{\psi}(x) = 2 \sum_{k \in \mathbb{Z}} m[k] (-1)^k \delta(x - (k + 1)/2).$$

The dual wavelets are compactly supported Radon measures, and inner products $\langle \tilde{\phi}_{jk}, f \rangle$ or $\langle \tilde{\psi}_{jk}, f \rangle$ only make sense if f is continuous. Thanks to theorem 1, this wavelet family can be considered as a basis of the locally convex vector space of continuous functions defined over \mathbb{R} , whose dual vector space happens to be the set of compactly supported Radon measures.

2 Multidimensional wavelets

2.1 Function approximation over a multidimensional domain

Definition 1 *We call a tensor product of n single variable functions f_1, \dots, f_n the n -variable function defined as*

$$F(x_1, \dots, x_n) = f_1(x_1) \times \dots \times f_n(x_n)$$

This tensor product is usually noted $F = f_1 \otimes \dots \otimes f_n$.

We will also note from now the father wavelet $\psi^0 = \phi$ and the mother wavelet $\psi^1 = \psi$, to have simpler higher dimension expressions.

2.1.1 Standard tensor product: Given a basis $\mathfrak{B} = \{b_k : k \in K\}$ for a functional vector space on \mathbb{R} , a straightforward way to build a basis for functions over \mathbb{R}^2 consists in taking tensor products

$$\mathfrak{B}^2 = \mathfrak{B} \otimes \mathfrak{B} = \{b_k \otimes b_{k'} : k, k' \in K\}$$

This construction can be used for any kind of bases, and extension to higher dimensions or number of bases is straightforward.

2.1.2 Homogeneous tensor product: Given a wavelet base $(\psi_{j_0 k}^0)_{k \in \mathbb{Z}} \cup (\psi_{j k}^1)_{j \geq j_0, k \in \mathbb{Z}}$ of a \mathbb{R} function space, we can build a wavelet base for function defined over \mathbb{R}^2 with father wavelets

$$\psi_{j_0 k k'}^{00} = \phi_{j_0 k} \otimes \phi_{j_0 k'}$$

for any integers k, k' and mother wavelets

$$\begin{aligned}\psi_{jkk'}^{10} &= \psi_{jk} \otimes \phi_{jk'} \\ \psi_{jkk'}^{01} &= \phi_{jk} \otimes \psi_{jk'} \\ \psi_{jkk'}^{11} &= \psi_{jk} \otimes \psi_{jk'}\end{aligned}$$

for $j \geq j_0$ and for any integers k, k' . For higher dimensions $n \geq 2$, this construction can be easily extended to define one father wavelet $\psi^{0\dots 0}$ and $2^n - 1$ mother wavelets $\psi^{\varepsilon_1 \dots \varepsilon_n}$ where $\varepsilon_i \in \{0, 1\}$ and $\sum \varepsilon_i \geq 1$.

2.1.3 Comparison: Both constructions have quite different properties. Grouping variables of an unknown function with the latter construction implies assuming similar characteristic scales according to those variables. This can happen for physical variables of the same kind (two lengths, two velocities in x and y). The former construction on the contrary does not imply such an assumption, but it can be used to group variables with different units.

Moreover, both construction can be combined to first group variables of the same kind and then group all variables together to make up the approximation space for our unknown function.

2.2 Vector valued functions

Without any a priori information, learning a vector valued function consists in learning all vector coordinate functions separately. Some more physical approaches can be done, if we suspect that the unknown vector field is a gradient field or can be very well approximated by such a field.

As an example, if we know that the vector field we are trying to learn is a gradient field, we can use as a wavelet family the gradients of a basis of scalar valued wavelets. This will allow to estimate in parallel the underlying potential field.

3 Control and Adaptation

In this section, we consider the standard problem of controlling the dynamical system defined as

$$\dot{x}^{(n)} = f(x, \dots, x^{(n-1)}) + u$$

where f is an unknown function. The measurable state is supposed to be $(x, \dots, x^{(n-1)})$. Let x_d be the desired trajectory, known up to the n -th derivative. Defining the tracking error $\tilde{x} = x - x_d$, the composite tracking error

$$s = \left(\frac{d}{dt} + \lambda \right)^{n-1} \tilde{x}$$

where λ is a strictly positive gain. Ensuring that s converge to 0 guarantees that $\tilde{x}, \dots, \tilde{x}^{(n-1)}$ converge to 0. We first briefly summarize the usual derivation (as in e.g. [7][9]), and then focus on the case of an infinite basis.

3.1 Non adaptive case

If f is known, then a control law is easily derivable. Defining

$$\dot{x}_r^{(n-1)} = x^{(n-1)} - s \quad (4)$$

$$\begin{aligned}&= x_d^{(n-1)} - \sum_{k=1}^{n-1} \binom{n-1}{k} \\ &\times \lambda^k \left(x^{(n-1-k)} - x_d^{(n-1-k)} \right)\end{aligned} \quad (5)$$

we can see from (5) that $x_r^{(n)}$ can be computed from the state $(x, \dot{x}, \dots, x^{(n-1)})$ and the desired trajectory.

Therefore setting $u = x_r^{(n)} - f(x) - k_D s$ ensures, according to (4)

$$\dot{s} = -k_D s$$

thus $s \rightarrow 0$ and $\tilde{x} \rightarrow 0$. Convergence to 0 is exponential.

3.2 Adaptive case

If we assume we can expand our unknown function over a basis $(\psi_K)_{K \in \mathbb{K}}$,

$$f(x) = \sum_{K \in \mathbb{K}} c_K \psi_K(x) = \psi(x)^T c$$

We build an estimate of the unknown function $\hat{f} = \psi(x)^T \hat{c}$. Setting $u = x_r^{(n)} - \hat{f}(x) - k_D s$, the widely used adaptation law

$$\dot{\hat{c}} = \Gamma \psi(x) s$$

guarantees that the positive semidefinite Lyapunov function candidate

$$V = \frac{1}{2} s^2 + \frac{1}{2} \tilde{c}^T \Gamma^{-1} \tilde{c}$$

(where $\tilde{c} = \hat{c} - c$) has a negative semidefinite time derivative. Indeed, $\dot{\tilde{c}} = \dot{\hat{c}}$ and therefore

$$\dot{V} = -k_D s^2 \quad (6)$$

Equation (6) proves uniform global stability of the system, and uniform continuity of s^2 . A standard result is then that uniform continuity of s^2 implies thanks to Barbalat's lemma that s converges to 0. However, no bound on the convergence speed can be given without any further assumptions.

3.3 Case of infinite \mathbb{K} index set

Let us first assume we can deal with infinite sequences and compute infinite sums. We will still have to ensure that those infinite sums converge. To simplify our discussion, we assume that our gain matrices Γ are always diagonal and positive definite. We will note our wavelet indices (j, k) in a condensed way “ K ”.

We first want our Lyapunov function (esp. $\tilde{c}^T \Gamma^{-1} \tilde{c}$) to be finite. This can be achieved whenever both c and \hat{c} are in the Hilbert space $E_\Gamma = \{c : c^T \Gamma^{-1} c < +\infty\}$.

Next, we want expansions $c^T \psi(x)$ to converge reasonably fast to a reasonably smooth function, thus the infinite sum $c^T \psi$ to converge absolutely and uniformly on any finite interval. At last, we need \hat{c} to remain in E_Γ at all times, thus $\dot{\hat{c}}$ to be in E_Γ , which is true if $\Gamma \psi(x) \in E_\Gamma$ for all x .

These last two conditions can be fulfilled if the matrix Γ matches the hereafter defined admissibility condition.

Definition 2 *An infinite diagonal matrix Γ is said to be admissible iff $\psi(x)^T \Gamma \psi(x)$ is an absolutely converging series for any x . Moreover, this absolute convergence has to be uniform with respect to x on any compact set.*

For admissible gain matrices Γ , the corresponding space E_Γ only contains continuous functions, and the sums $c^T \psi(x)$ are uniformly absolutely convergent, since

$$\sum_K |c_K \psi_K(x)| \leq \sqrt{\sum_K \gamma_K^{-1} c_K^2} \sqrt{\sum_K \gamma_K \psi_K(x)^2}$$

Moreover, any $\Gamma \psi(x)$ will be in E_Γ , since

$$(\Gamma \psi(x))^T \Gamma^{-1} (\Gamma \psi(x)) = \psi(x)^T \Gamma \psi(x) < +\infty$$

Note that if there exists a strictly positive sequence (γ_j) such that

$$\Gamma_{(jk), (j', k')} = \delta_{jj'} \delta_{kk'} \gamma_j \quad \forall k \text{ and } \sum_j \gamma_j < +\infty$$

then Γ is admissible.

If we choose $\gamma_j = 2^{-\beta j}$, then the restriction on coefficients c_K that $c^T \Gamma^{-1} c$ be convergent is not very strong, since any compactly supported function that is α -Lipschitz for some $\alpha > (1 + \beta)/2$ has an expansion c in E_Γ . Conversely, any function in E_Γ is at least $\beta/2$ -Lipschitz.

Convergence analysis In a case where the chosen Γ is admissible, the same proof as for a finite number of unknown parameters holds. Namely, s is bounded because the Lyapunov function takes only finite values and is decreasing, and again equation

$$\dot{s} = -k_D s + \tilde{c}^T \Psi(x)$$

leads to

$$|\dot{s}| \leq k_D |s| + \sqrt{\tilde{c}^T \Gamma^{-1} \tilde{c}} \sqrt{\Psi(x)^T \Gamma \Psi(x)}$$

thanks to the Cauchy-Schwarz inequality, and therefore \dot{s} is also bounded. As consequence, the conclusion of Barbalat's lemma also holds in this case.

Finite approximate controller If now f is approximated with finite expansion f^b where c^b is a vector of finite known support, we impose that \hat{c} has the same support. The control law

$$u = x_r^{(n)} - k_D s - \hat{c}^T \psi(x)$$

and the finite adaptation law as in [7][9]

$$\dot{\hat{c}} = \Gamma \psi^b s_\Delta$$

where

$$s_\Delta = \begin{cases} s - \Delta \times \text{sign}(s) & \text{if } |s| > \Delta \\ 0 & \text{else.} \end{cases}$$

ensure that the Lyapunov-like function of the infinite model

$$V = \frac{1}{2} s_\Delta^2 + \frac{1}{2} \tilde{c}^T \Gamma^{-1} \tilde{c},$$

has time derivative

$$\dot{V} = -k_D s s_\Delta + s_\Delta (f - f^b)$$

which can be easily proven to be always less than $-k_D s_\Delta^2$, which in turn shows through Barbalat's lemma that $s_\Delta \rightarrow 0$.

Note that the support of the finite expansions c^b can be time-varying and managed in a similar way to [8] since the value of V is only affected by the set of non-zero values in the vector \hat{c} . Also, Δ can vary accordingly as in [9]. As compared to previous work, existence of this infinite model Lyapunov functions also turns out to enhance significantly the high frequency behavior of the learning algorithm.

Conclusion

As compared to previous work, this paper relies on a model of a controller able to stably handle infinitely many parameters, whose behavior we then approximate with a finite size controller. It also goes further into the specific design of wavelets, and into higher dimension function approximation. Finally, pointwise convergence bounds on the expansion allow us to obtain a stronger convergence result on the tracking error.

Acknowledgments

This paper benefited from interesting discussions with Stéphane Mallat.

Appendix A

Proof of Theorem 1

Equality of formulae 2 and 3 is a consequence of the two-scale refinement formula that ϕ matches.

We therefore only have to prove that as $J \rightarrow +\infty$,

$$\sum_{j=-J}^J \sum_{k=-\infty}^{+\infty} c_{jk} \psi_{jk}(x) \rightarrow f(x)$$

uniformly. This proof relies on the following points:

- f is continuous;
- for any x , $\sum_{k=-\infty}^{+\infty} \phi(x+k) = 1$;
- $\sum_{k=-\infty}^{+\infty} |\phi(x+k)|$ is bounded by B_p , where p is the Deslauriers-Dubuc interpolation parameter;
- ϕ is compactly supported, on $[-2p+1; 2p-1]$.

Let $[a, b]$ be a compact interval, and suppose $x \in [a, b]$. Let $\varepsilon > 0$. According to the Heine theorem, f is uniformly continuous on $[a-1, b+1]$. As a consequence, there exists $\eta > 0$ such that for any x, y in $[a-1, b+1]$, $|x-y| < \eta \Rightarrow |f(x) - f(y)| < \varepsilon$.

If

$$r_j = \sum_{k=-\infty}^{+\infty} f(2^{-j}k) \phi(2^j x - k) - f(x)$$

then

$$r_j = \sum_{k=-\infty}^{+\infty} (f(2^{-j}k) - f(x)) \phi(2^j x - k) - f(x)$$

Since ϕ is supported on $[-2p+1; 2p-1]$, the above sum can be bounded as

$$|r_j| \leq \sum_{k=\lfloor 2^j x - 2p + 1 \rfloor}^{\lfloor 2^j x + 2p - 1 \rfloor} |f(2^{-j}k) - f(x)| \cdot |\phi(2^j x - k)|$$

If $j > \log(2p/\eta)/\log 2$, which implies that $|2^{-j}k - x| \leq \eta$, and if we also have $j > \log 2p/\log 2$ which implies that $2^{-j}k \in [a-1, b+1]$, then

$$|r_j| \leq \varepsilon \sum_{k=-\infty}^{+\infty} |\phi(2^j x - k)| \leq \varepsilon B_p$$

which proves the convergence.

Proof of Theorem 2

Let us suppose f is d times continuously derivable with $d \geq 1$. ψ has $2p$ zero moments, which means that it has a $2p$ order integral that is compactly supported: $\tilde{\psi} = \Psi^{(2p)}$ where Ψ is compactly supported.

Thus, if $d \leq 2p$, we have

$$\int \overline{\tilde{\psi}} f = \int \overline{\Psi^{(2p)}} f = (-1)^d \int \overline{\Psi^{(2p-d)}} f^{(d)}$$

and

$$\begin{aligned} \int \overline{\tilde{\psi}_{jk}} f &= \int \overline{\tilde{\psi}(x-k)} f(2^{-j}x) dx \\ &= (-1)^d 2^{-dj} \int \overline{\Psi^{(2p-d)}(x-k)} f^{(d)}(2^{-j}x) dx \end{aligned}$$

Defining $M_1 = \int |\Psi^{(2p-d)}(x)| dx$, then have the bound

$$|c_{jk}| \leq 2^{-dj} M_1 \sup_{x \in [k-2p+1, k+2p-1]} |f^{(d)}(2^{-j}x)|$$

As a consequence, for a given compact interval $[a, b]$, we have the bound

$$|r_j| \leq \sup_{[a-2p2^{-j}, b+2p2^{-j}]} |f^{(d)}| \times \frac{2^{-dj}}{1-2^{-d}} M_1 B_p$$

For an example $p = 2$ and $d = 4$, we have $B_p < 1.26$, and $\int |\Psi^{(2p-d)}(x)| dx = 3/8$. The bound is therefore

$$|r_j| \leq 0.51 \times 2^{-4j} \sup_{[a-2^{-j+3}, b+2^{-j+3}]} |f^{(4)}|$$

References

- [1] M. Cannon, J-J. Slotine, Space-Frequency Localized Basis Function Networks for Nonlinear System Estimation and Control, *Neurocomputing*, 9(3), 1995.
- [2] A. Cohen, I. Daubechies, B. Jawerth, P. Vial, Multiresolution analysis, wavelets and fast algorithms on an interval, *Comptes Rendus Acad. Sci. Paris*, 316 (série 1), pp. 417-421, 1993.
- [3] I. Daubechies, Ten Lectures on Wavelets, *Society for Industrial and Applied Mathematics, Philadelphia*, 1992.
- [4] F.L. Lewis and Y.H. Kim, "Neural networks for feedback control," in *Encyclopedia of Electrical and Electronics Engineering*, Wiley.
- [5] S.G. Mallat, "A Theory for Multiresolution Signal Decomposition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11:7, pp. 674-693, 1989.
- [6] S.G. Mallat, "A Wavelet Tour of Signal Processing," *Academic Press*, 1997.
- [7] R. Sanner, J-J. Slotine, Gaussian Networks for Direct Adaptive Control, *IEEE Trans. Neural Networks*, November 1992.
- [8] R. Sanner, J-J. Slotine, Structurally Dynamic Wavelet Networks for the Adaptive Control of Uncertain Robotic Systems, *Proc. IEEE Conf. Decision and Control*, New Orleans, 1995.
- [9] J.-J. E. Slotine, J. A. Coetsee, Adaptive Sliding Controller Synthesis for Nonlinear Systems, *Int. Journal of Control*, 10(4) 1986.