SYSTEM IDENTIFICATION USING WAVELETS

Daniel Coca
Department of Electrical Engineering and Electronics, University of Liverpool, UK

Stephen A. Billings
Department of Automatic Control and Systems Engineering, University of Sheffield, UK

Keywords: basis functions, continuous wavelet transform, differential equation, difference equation, discrete-time system, dynamical system, estimation data set, least squares, model set, model validity tests, parameter estimation, prediction error, structure selection, system identification, wavelet, wavelet multiresolution approximation.

Contents

Summary

Many applications make use of series expansion techniques for characterizing general functions in terms of a predefined set of prototype or basis functions. A well-known example is the Fourier series representation which employs sine and cosine functions of different frequency as the basis functions.

More recently, the wavelet representation has emerged as a powerful alternative to existing series expansion techniques. This method is based on a fundamental concept of representing arbitrary functions in terms of the translations and dilations of a single localized small wave or 'wavelet' function, which decays rapidly towards zero. Unlike the Fourier basis functions, the wavelet basis functions are localized both in space and in frequency so the wavelet analysis can provide time-frequency information about a

function which in many practical situations is more pertinent than the standard Fourier analysis. Wavelets also provide a powerful approximation tool that can be used to synthesize economically, using a minimal number of basis elements, functions which are difficult to approximate by other methods.

This work presents the use of wavelet representations in nonlinear system identification. In particular it is shown how to identify NARMAX models, from data, implemented as a superposition of wavelet basis functions. Two specific implementations one based on radial wavelets and the other based on B-spline wavelet multiresolution approximations are presented.

1. Introduction

The use of linear models in the study of dynamical systems, in the real world is limited by the fact that more often than not the dynamical system of interest turns out to be nonlinear.

Linear models cannot reproduce a wide range of dynamical behavior that results from non-linear interactions. In such situations, it is essential to use nonlinear models to represent such behavior.

In many cases, the only practical way to obtain a model of a nonlinear dynamical system is directly from experimental input/output data recorded from the system. This process is known as system identification.

A fundamental problem in nonlinear system identification is how to infer a good nonlinear model from observations given the fact that the number of possible nonlinear interactions is theoretically infinite. As someone once said, 'a nonlinear function can be nonlinear in so many ways'. How can we decide which nonlinear functions to use when we start to write down the equations that describe a nonlinear process?

The solution is to implement the nonlinear model using specific types of functions, such as polynomials for example, which can approximate arbitrary nonlinear interactions that occur in practical dynamical systems. The approximation properties of each class of functions will determine in practice the range of nonlinear interactions that can be modeled.

In practice, certain types of functions can efficiently approximate only certain nonlinear relationships. In some cases, a given class of functions may not be rich enough to guarantee that a model, that describes the observed nonlinear behavior with good accuracy, can be build using only functions from that class. In other cases the functions chosen to implement the model are more complex than required and can lead to a model that, although it fits the data very well, does not describe the dynamical process.

It follows that in system identification an ideal approximation technique should allow maximum flexibility in adapting the complexity of the model structure to match, as closely as possible, the underlying nonlinear interactions involved in each particular situation. In this context, recent developments in the way arbitrary functions are represented as a superposition of elementary or basis functions provide an excellent framework for implementing such adaptive approximation structures.

Today, many applications of mathematics make use of methods for relating functions or empirically derived variables to a set of basis functions that have important general properties. In the last decade the popular Taylor and Fourier series were complemented by a wealth of alternative series expansions developed in order to represent, analyze and characterize general functions.

Among these new methods none has had such an impact and spurred so much interest as the wavelet representation, which bridged the gap between various fields of research such as approximation theory, signal and image processing.

At the heart of the method is the concept of representing an arbitrary function in terms of dilations and translations of a single wave-like function known as the "mother" wavelet function. In general, the wavelet approximation of a function is very economical, that is, only a small number of wavelets are enough to approximate a function with good accuracy.

Even functions that are notoriously difficult to approximate can be synthesized efficiently using this approach. This makes wavelet functions well suited in nonlinear system identification. It is important however to find among the many wavelet classes available, those that are best suited for this task.

This article focuses on two wavelet based model implementations, which have been successfully employed in practice, namely the wavelet network and the B-spline wavelet multiresolution model. Special emphasis is placed on the model structure selection algorithms. The problem of selecting the wavelet model terms is very important, given the fact that the "atomic" decomposition of the nonlinear function in terms of wavelet basis functions, leads to a very large number of candidate model terms.

In the absence of a proper structure selection strategy, the resulting identified models will have far too many parameters for the model to be practical. In this case, one of the most important features associated with the wavelet representations, that is the ability to describe functions and distributions using a minimal number of wavelet coefficients, is not exploited.

The article highlights the importance of developing efficient algorithms that can be used to assemble piece by piece, like a puzzle, based on a set or library of wavelet basis functions, the simplest model that can describe the system dynamics with good accuracy.

2. Wavelets - A Brief Overview

Although the first appearance of wavelets can be traced back to as early as the beginning of the century it was the work of Morlet and Grossman who initiated the resurgence of wavelet theory in the context of seismic signal processing. They described a function characterized by regularity, localization and an oscillatory nature as a "wavelet". Their work triggered an explosion in the number of publications concerning theoretical and practical treatment of wavelets.

The wavelet transform is founded on the simple notion of approximation of a signal via dilations and translations of a single function $\psi(t)$ referred to as the *mother wavelet* which allows a signal to be viewed at various scales and different positions in time. Since the notion of scale is linked directly to frequency, wavelets represent in fact a time-frequency analysis tool.

Unlike the sinusoidal wave functions used in Fourier theory, the wavelet function is a small wave centered around a given position in time t^* (or space) with a fast decay to zero away from the center. Because of this, wavelets need to be shifted (translated) in order to cover the whole real line. The translation of a wavelet function $\psi(t)$ with an amount b can be written as $\psi(t-b)$.

At the same time, in the frequency domain, $\hat{\psi}(\omega)$, which represents the Fourier transform of ψ, is also centered around a given frequency ω^* with a fast decay to zero away from the center frequency.

The wavelet function is said to generate a window function which is characterized both in the time and frequency domain by its 'center' and 'width' (i.e. the length of the interval over which the function values are significant).

Similar to Fourier theory, where the sinusoidal waves involved have different frequencies, wavelets of different frequencies have to be considered in order to cover the frequency space.

This can be achieved by introducing a dilation parameter a, which controls how fast the wavelet oscillates.

This is equivalent to a translation of $\hat{\psi}(\omega)$ in the frequency space, so that the new center frequency becomes ω^*/a.

Moreover, as shown in Eq. (1), while a translation in time preserves the shape of the time-frequency window, scaling alters it so that for large center-frequency ω^*/a the window narrows in the time-domain and widens in the frequency-domain. The opposite happens at small center-frequencies.

It follows that a single function $\psi(t)$ can be used to generate a whole family of functions parameterized in terms of the dilation (scaling) and translation parameters

$$\psi_{a,b} = |a|^{-1/2}\,\psi\left(\frac{t-b}{a}\right) \qquad (1)$$

which act as time-frequency windows, and can be used to extract spectral information from a signal at specific locations.

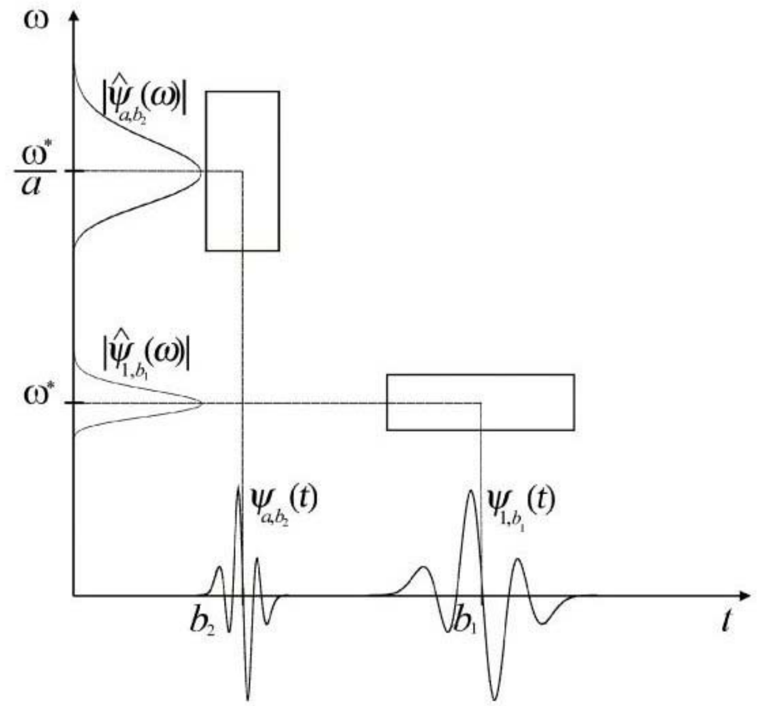

Figure1: Time-frequency windows generated by $\psi_{1,b_1}(t)$ and $\psi_{a,b_2}(t)$ where $0 < a < 1, b_2 < b_1$

The abundance of useful features enjoyed by the wavelet transform has led to its application in a wide range of disciplines such as mathematics, physics, signal processing, approximation theory and numerical analysis. This is also a direct consequence of the fact that the development of wavelet theory follows independent contributions from various fields of research, which eventually led to the establishment of a unified theory.

Similar to Fourier theory, there are two main components of wavelet theory namely the *Continuous Wavelet Transform* and the *Wavelet Series*. Unlike in the Fourier case however, here the two components are closely related. In practice, for a given application, it is often crucial to select the most appropriate of the different forms of the wavelet transforms.

The wavelet transform can be understood as an alternative way of investigating functions, which have previously been studied only by means of Fourier series and integrals. However, the two theories, rather than competing, are complementary since there are applications where wavelet analysis is better suited than Fourier analysis and vice versa. In signal analysis applications, for example, transients are easily recognizable as localized bursts of energy at small scales using the wavelet transform.

In function approximation applications the time-frequency localization property of the wavelet function means that isolated discontinuities can be approximated with good accuracy by refining the wavelet approximation structure locally rather than globally, leading to more economical representations involving fewer parameters.

The ability of wavelets to encode signals efficiently, in terms of a small number of wavelet coefficients, has led to the development of many wavelet-based signal and image compression algorithms.

In estimation applications, a smooth signal can be recovered from noisy measurements by rejecting noise components above the scale at which the underlying signal has significant energy.

2.1 The Continuous Wavelet Transform

The *Continuous Wavelet Transform* (CWT) maps a continuous square integrable function of one variable $f(t)$ to a continuous function of two variables a and b that correspond to scale and location.

A square integrable function is a function for which the following integral has finite value

$$\int_{-\infty}^{\infty} | f(x) |^2 \, dx < \infty \tag{2}$$

The *Continuous Wavelet Transform* (CWT) relative to some basic wavelet called the *mother wavelet* $\psi(x)$ is defined for a square integrable function $f(x)$ as

$$(W_\psi f)(a,b) \overset{\text{def}}{=} \int_{-\infty}^{\infty} f(x) \overline{\psi_{a,b}(x)} \, dx \tag{3}$$

where $a, b \in \mathbb{R}$, $a \neq 0$ and

$$\psi_{a,b}(x) \overset{\text{def}}{=} |a|^{-1/2} \, \psi\left(\frac{x-b}{a}\right) \tag{4}$$

Equation (3) states that $(W_\psi f)$ (a, b) is the correlation of $f(x)$ with a shifted (by b) and scaled (by a) version of ψ. From Figure 1 it is clear that for any given a and b, the CWT gives local information of an analog signal within a time-frequency window whose dimensions and location depend continuously on a and b.

The transform defined in Eq. (3) is invertible subject to a mild restriction imposed on ψ. The restriction is that

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty \tag{5}$$

where $\hat{\psi}$ is the Fourier Transform of ψ. Equation (5) guarantees that $f(x)$ can be recovered from its wavelet transform using the following inversion or synthesis formula

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[(W_\psi f)(b, a) \right] \psi_{b,a}(x) \frac{da}{a^2} db \qquad (6)$$

The expression given in Eq. (6) can be interpreted in two different ways. In one way it shows how to reconstruct *f* from the corresponding integral representation $(W_\psi f)\,(a, b)$. On the other hand, the equation gives a recipe to write any arbitrary *f*(*x*) as a superposition of wavelet functions $\psi_{b,a}(x)$.

Whenever ψ decays sufficiently fast, that is for example when the following integrals are finite

$$\int_{-\infty}^{\infty} x\psi(x)dx < \infty \qquad (7)$$

$$\int_{-\infty}^{\infty} \omega\psi(\omega)d\omega < \infty \qquad (8)$$

the invertibility condition Eq. (5) is equivalent to

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \qquad (9)$$

In this case, the basic wavelet is said to provide a time-frequency window that is, the wavelet function is localized in both the time and the frequency domain. This feature makes the wavelet transform a very suitable time-frequency representation.

There exist other, older and useful time-frequency representations such as the Short-Time Fourier Transform (STFT) or the Gabor transform.

The difference between the time-frequency analysis provided by the wavelet transform and the previous methods is that CWT uses short time-windows (narrower wavelets) at high frequency and long time-windows (wider wavelets) at low frequencies, as opposed to the constant time-frequency window used in STFT.

This makes the wavelet transform particularly suitable for analyzing signals or functions with both very high and very low frequency components.

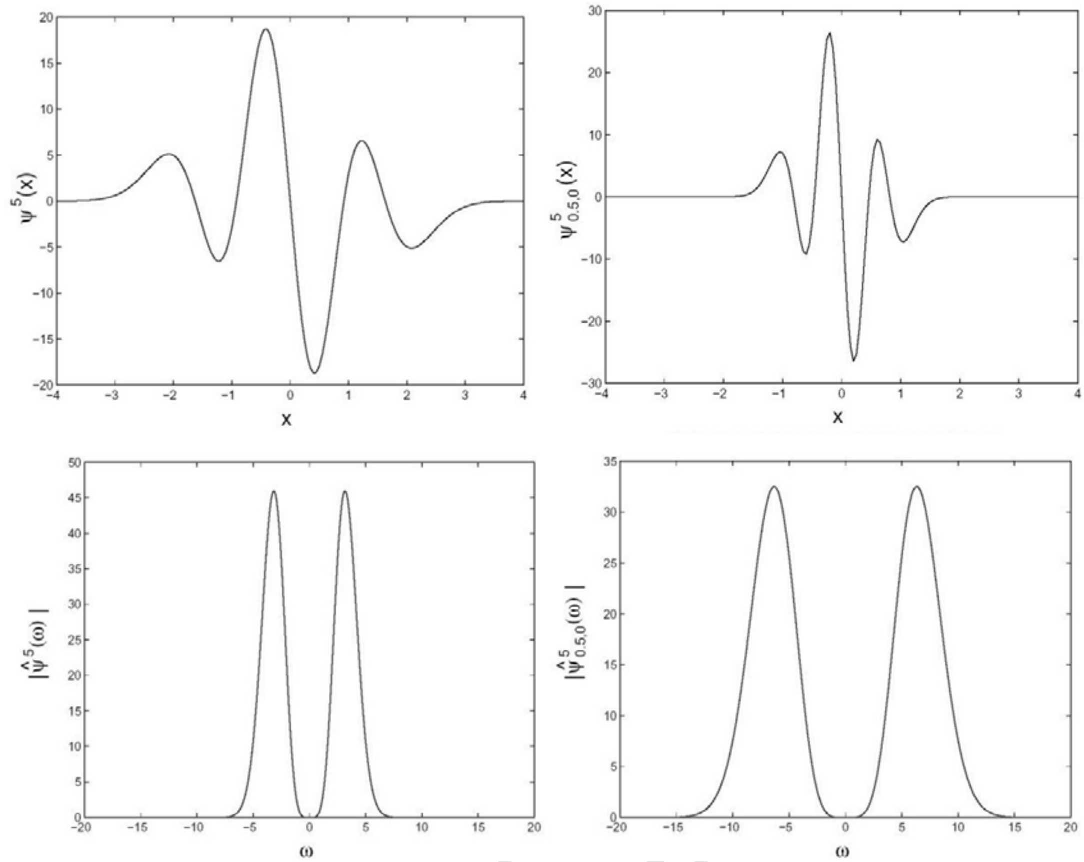

Figure 2: Examples of wavelets: The 5-th derivative of the Gaussian (a) for ***a* = 1** and (c) for ***a* = 0.5**; (b), (d) The corresponding Fourier transforms

An example of a wavelet that is well localized in both the physical and the Fourier domain is the N^{th} derivative of a Gaussian

$$\psi^N(x) = (-1)^N \frac{d^N}{dx^N}\left(e^{\frac{-x^2}{2}}\right) \tag{10}$$

Figures 2a and 2b display the mother wavelet defined in equation Eq. (10) for N = 5, $\psi^5(x)$ along with $\psi^5_{0.5,0}(x)$ and the corresponding Fourier counterparts $\hat{\psi}^5(\omega)$ and $\psi^5_{0.5,0}(\omega)$. It can be seen how the time localization of the wavelet function $\psi^5_{a,b}$ increases with decreasing scale at the expense of the frequency localization of the corresponding Fourier transformed function $\hat{\psi}^5_{a,b}$. The tradeoff between time and frequency localization is a distinguishing feature of the wavelet transform which can be exploited in many applications.

The number of applications of CWT is large, ranging from signal processing and data compression to the analysis of intermittence in turbulence and the physiology of sight. In system identification, the continuous wavelet transform has formed the basis for the development of a neural network type architecture, which will be described in detail in a later section.

2.2 Wavelet Series

In many practical applications, the *Continuous Wavelet Transform* is discretized in the scaling and dilation parameters for computational efficiency. By discretizing both the dilation and translation parameter it is possible to define the *Wavelet Series*. A wavelet series provides an alternative series representation, to the classical Fourier series for example, for square integrable functions.

-
-
-

Bibliography

Billings S. A., Chen S. and Kronenberg M. J. (1989). Identification of MIMO Non-linear Systems Using a Forward-regression Orthonormal Estimator. *International Journal of Control*, **49**, 2157-2189. [Discusses the identification of multi-input, multi-output nonlinear systems using the orthogonal least squares (OLS) algorithm.]

Chen S., Billings S. A. and Luo W. (1989). Orthogonal Least-squares Methods and their Applications to Nonlinear System Identification. *International Journal of Control*, **50**, 1873-1896. [An overview of orthogonal least-squares methods with application in model structure selection for nonlinear systems]

Billings S. A. and Coca D, (1999). Discrete wavelet models for identification and qualitative analysis of chaotic systems *International Journal of Bifurcation and Chaos*, **9**, 1263-1284. [Introduces the B-spline wavelet multiresolution NARMAX model with application in the identification of chaotic systems]

Chui C. K. (1992). *Wavelet Theory and its Applications*. Boston: Academic Press. [An introductory treatise on wavelet analysis with an emphasis on spline wavelets and time-frequency analysis]

Grossman A. and Morlet J. (1984). Decompositions of Hardy Functions Into Square Integrable Wavelets of Constant Shape. *SIAM Journal on Mathematical Analysis*, **15**, 723-736. [A seminal paper on wavelets]

Haber R. and Unbehauen H. (1990). Structure Identification for Nonlinear Dynamical Systems. A Survey on Input-Output Approaches. *Automatica*, **26**, 651-677. [Reviews the identification of certain classes of nonlinear systems.]

Juditski A. , Hjalmarsson H., Beneviste B., Deylon H., Ljung L., Sjóberg, J. and Zhang Q. (1995). Nonlinear Black-Box Models in System Identification: Mathematical Foundations. *Automatica*, **31**, 1725-1750. [Discusses the theoretical foundation of various nonlinear model implementations including wavelets.]

Mallat S. G. (1989). A Theory for Multiresolution Signal Decomposition; The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674-693. [Introduces the concept of wavelet multiresolution decomposition in the context of image processing]

Leontaritis I. and Billings S. A. (1985a). Input-Output Parametric Models for Non-linear Systems - Part I: Deterministic Non-linear Systems. *International Journal of Control*, **41**, 303-328; (1985b) Input-Output Parametric Models for Non-linear Systems - Part II: Stochastic Non-linear Systems. *International Journal of Control*, **41**, 329-344. [Formal introduction to the NARMAX model.]

Zhang Q. G, and Beneviste A. (1992). Wavelet Networks. *IEEE Transactions on Neural Networks*, **3**, 889-898. [Introduces the wavelet network model]

Biographical Sketches

Daniel Coca received the MEng degree in Electrical Engineering from the Transilvania University of Brasov, Romania in 1993 and his Ph.D. in Control Systems Engineering from the University of Sheffield, UK in 1997. He is a Chartered Engineer (CEng) and Member of the IEE (UK). Since 1997 he has worked as a Research Associate in the Department of Automatic Control and Systems Engineering at the University of Sheffield. In August 2002, he was appointed Lecturer in the Department of Electrical Engineering and Electronics at the University of Liverpool. His research interests include nonlinear system identification, wavelets, identification and control of distributed parameter systems, and complex systems.

Stephen A Billings received the BEng degree in Electrical Engineering with first class honors from the University of Liverpool in 1972. He received his Ph.D. in Control Systems Engineering from the University of Sheffield in 1976 and the degree of DEng from the University of Liverpool in 1990. He is a Chartered Engineer (CEng), Chartered Mathematician (CMath), Fellow of the IEE (UK) and Fellow of the Institute of Mathematics and its Applications. He was appointed as Professor in the Department of Automatic Control and Systems Engineering, University of Sheffield, UK in 1990, and leads the Signal Processing and Complex Systems research group. His research interests include system identification and information processing for nonlinear systems, NARMAX methods, model validation, prediction, spectral analysis, adaptive systems, nonlinear systems analysis and design, neural networks, wavelets, fractals, machine vision, cellular automata, spatio-temporal systems, fMRI and optical imagery of the brain.