数理解析研究所講究録 1385 巻 2004 年 120-135

Wavelet analysis for system identification*

大阪教育大学・数理科学 芦野 隆一 (Ryuichi Ashino)
Mathematical Sciences, Osaka Kyoiku University
大阪電気通信大学・数理科学研究センター 萬代 武史 (Takeshi Mandai)
Research Center for Physics and Mathematics,
Osaka Electro-Communication University
大阪教育大学・情報科学 守本 晃 (Akira Morimoto)
Information Science, Osaka Kyoiku University

Abstract

By the Schwartz kernel theorem, to every continuous linear system there corresponds a unique distribution, called *kernel distribution*. Formulae using wavelet transform to access time-frequency information of kernel distributions are deduced. A new wavelet based system identification method for health monitoring systems is proposed as an application of a discretized formula.

1 System Identification

A *system* L illustrated in Figure 1 is an object in which variables of different kinds interact and produce observable signals. The observable signals g that are of interest to us are called *outputs*. The system is also affected by external stimuli. External signals f that can be manipulated by the observer are called *inputs*. Others are called *disturbances* and can be divided into those, denoted by w, that are directly measured, and those, denoted by v, that are only observed through their influence on the output.

Mathematically, a system can be regarded as a mapping which relates inputs to outputs. Hereafter, we will use the terminology "system" rather

*This research was partially supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Scientific Research (C), 15540170 (2003–2004), and by the Japan Society for the Promotion of Science, Japan-U.S. Cooperative Science Program (2003–2004).

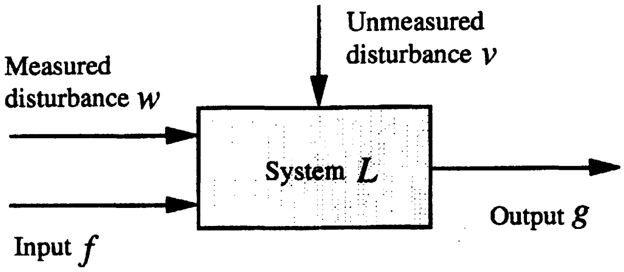


Figure 1: System L.

than "mapping". A system L is said to be *linear* if

$$L[\alpha_1 f_1 + \alpha_2 f_2] = \alpha_1 L[f_1] + \alpha_2 L[f_2],$$

for arbitrary constants α_1, α_2 and arbitrary inputs f_1, f_2. To choose a model set for linear systems, the most general setting should be given by using the Schwartz kernel theorem [Tr67].

Let us denote by $\mathscr{D}(\mathbb{R}^n)$, the space of compactly supported C^∞ functions with the canonical LF-topology [Tr67]. A *distribution* T in $\mathbb{R}^n$ is a continuous linear form on $\mathscr{D}(\mathbb{R}^n)$. The set of all the distributions in $\mathbb{R}^n$ is denoted by $\mathscr{D}'(\mathbb{R}^n)$ and the duality of distributions is denoted by $\langle \cdot, \cdot \rangle^*$, that is, the duality between $T \in \mathscr{D}'$ and $\phi \in \mathscr{D}$ is written as

$$T(\phi) = \langle T, \phi \rangle^* = \langle T(x), \phi(x) \rangle_x^*.$$

The space of C^∞ functions in $\mathbb{R}^n$ with the canonical Fréchet-topology is denoted by $\mathscr{E}(\mathbb{R}^n)$ and its topological dual space is denoted by $\mathscr{E}'(\mathbb{R}^n)$, which is the space of compactly supported distributions in $\mathbb{R}^n$. On the other hand, as the inner product of L^2 is denoted by $\langle \cdot, \cdot \rangle$, we have

$$\langle f, g \rangle^* = \langle f, \overline{g} \rangle, \qquad f, g \in L^2.$$

The space of C^∞ functions in $\mathbb{R}^n$ rapidly decreasing at infinity is called *Schwartz space* and denoted by $\mathscr{S}(\mathbb{R}^n)$. The topological dual space of $\mathscr{S}(\mathbb{R}^n)$ is denoted by $\mathscr{S}'(\mathbb{R}^n)$ and an element of $\mathscr{S}'(\mathbb{R}^n)$ is called *tempered distribution.*

Theorem 1 (The Schwartz kernel theorem) *Let $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ be a continuous linear system. Then, there is a unique distribution $k \in \mathscr{D}'(\mathbb{R}^{2n})$ such that*

$$L[f](x) = \langle k(x, y), f(y) \rangle_y^*, \qquad f \in \mathscr{D}(\mathbb{R}^n).$$

The distribution k is called kernel distribution of L. If $L : \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}'(\mathbb{R}^n)$, then $k \in \mathscr{S}'(\mathbb{R}^{2n})$.

Invariance under fundamental operators

Let us define three fundamental operators for time-frequency analysis. They are unitary operators when they act on $L^2(\mathbb{R}^n)$. Define the *translation operator* T_a by

$$T_a f(x) = f(x-a), \qquad a \in \mathbb{R}^n,$$

the *modulation operator* M_ξ by

$$M_\xi f(x) = e^{ix\xi} f(x), \qquad \xi \in \mathbb{R}^n,$$

and the *dilation operator* D_ρ by

$$D_\rho f(x) = \rho^{-n/2} f(\rho^{-1}x), \qquad \rho \in \mathbb{R}_+ := \{x \in \mathbb{R}\,;\, x > 0\}.$$

For simplicity, we consider D_ρ only for $\rho > 0$ although we can consider D_ρ for $\rho \in \mathbb{R}\backslash\{0\}$.

A continuous linear system $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ is said to be *translation-invariant* if $T_a L[f] = L[T_a f]$ for every $f \in \mathscr{D}(\mathbb{R}^n)$ and every $a \in \mathbb{R}^n$. A continuous linear system $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ is said to be *modulation-invariant* if $M_\xi L[f] = L[M_\xi f]$ for every $f \in \mathscr{D}(\mathbb{R}^n)$ and every $\xi \in \mathbb{R}^n$. A continuous linear system $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ is said to be *dilation-invariant* if $D_\rho L[f] = L[D_\rho f]$ for every $f \in \mathscr{D}(\mathbb{R}^n)$ and every $\rho \in \mathbb{R}_+$.

In this section, we will give necessary and sufficient conditions on the kernel distribution $k(x, y)$ corresponding to a continuous linear system $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ for invariance under the three fundamental operators. The proofs can be found in [AMM2].

Proposition 1 *Let $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ be a continuous linear system and $k(x, y)$ be its kernel distribution. The system L is translation-invariant if and only if there exists a unique $h \in \mathscr{D}'(\mathbb{R}^n)$ such that $k(x, y) = h(x-y)$, that is, $L[f] = h * f$. As a result, we have $L(\mathscr{D}(\mathbb{R}^n)) \subset \mathscr{E}(\mathbb{R}^n)$. The distribution h is called the impulse response of L.*

If L is continuous from $\mathscr{S}(\mathbb{R}^n)$ to $\mathscr{S}'(\mathbb{R}^n)$, then $h \in \mathscr{S}'(\mathbb{R}^n)$, and hence we have $L(\mathscr{S}(\mathbb{R}^n)) \subset \mathscr{O}_M(\mathbb{R}^n)$, where $\mathscr{O}_M$ is the space of slowly increasing C^∞ functions.

Proposition 2 *Let $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ be a continuous linear system and $k(x, y)$ be its kernel distribution. The system L is modulation-invariant if and only if there exists a unique $g \in \mathscr{D}'(\mathbb{R}^n)$ such that $L[f] = gf$ for every $f \in \mathscr{D}(\mathbb{R}^n)$.*

Proposition 3 *Let $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ be a continuous linear system and $k(x, y)$ be its kernel distribution. The system L is dilation-invariant if and only if $k(\rho x, \rho y) = \dfrac{1}{\rho^n} k(x, y)$ for every $\rho \in \mathbb{R}_+$.*

Stability

Let $L : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ be a continuous linear system, and k be its kernel distribution. L is said to be *L^p-stable* $(1 \le p \le \infty)$ if there exists a constant C such that $\|L[f]\|_{L^p} \le C\|f\|_{L^p}$ for every $f \in \mathscr{D}(\mathbb{R}^n)$. If $p < \infty$ and L is L^p-stable, then L can be extended to a bounded linear operator from $L^p(\mathbb{R}^n)$ to $L^p(\mathbb{R}^n)$. Here, we concentrate on the case when $p = 2$.

When the kernel distribution k is locally integrable, the following is well-known. (For example, [La02], Theorem 2 and Theorem 3, §16.1.)

Proposition 4 (1) *If $k \in L^2(\mathbb{R}^{2n})$, then L is L^2-stable, and*

$$\|L[f]\|_{L^2(\mathbb{R}^n)} \le \|k\|_{L^2(\mathbb{R}^{2n})}\|f\|_{L^2(\mathbb{R}^n)} \qquad \textit{for every } f \in L^2(\mathbb{R}^n). \tag{1.1}$$

(2) *Assume that there exist constants M_1, M_2 such that*

$$\int_{\mathbb{R}^n} |k(x,y)|\,dx \le M_1, \quad \int_{\mathbb{R}^n} |k(x,y)|\,dy \le M_2 \quad \textit{for a.e. } x \in \mathbb{R}^n. \tag{1.2}$$

Then, L is L^2-stable, and

$$\|L[f]\|_{L^2(\mathbb{R}^n)} \le \sqrt{M_1 M_2}\|f\|_{L^2(\mathbb{R}^n)} \qquad \textit{for every } f \in L^2(\mathbb{R}^n).$$

Let L be translation-invariant and h be its impulse response. We have the following ([St70], IV, §3.1; [Hö60]).

Proposition 5 *L is L^2-stable $\iff$ $\hat{h} \in L^\infty(\mathbb{R}^n)$, where $\hat{h}$ is the Fourier transform of h.*

Causality

Causality is natural for a physical system in which the variable is time. It means that the response at time t depends only on what has happened before and at t. In particular, a system does not respond before there is an input. Thus causality is a necessary condition for a system to be physically realizable.

Let L be a continuous linear system $\mathscr{D}(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$, and $k \in \mathscr{D}'(\mathbb{R}^{2n})$ be its kernel distribution. A continuous linear system L is said to be *causal* if

$$\operatorname{supp} L[f] \subset \operatorname{supp} f + \overline{\mathbb{R}_+}^n$$

for every $f \in \mathscr{D}(\mathbb{R}^n)$. Here, $A + B := \{\, a + b \;;\; a \in A,\, b \in B \,\}$ and $\overline{\mathbb{R}_+} := [0, \infty)$. We simply write $a + B$ for $\{a\} + B$. A distribution $f \in \mathscr{D}'(\mathbb{R}^n)$ is said to be *causal* if $\operatorname{supp} f \subset \overline{\mathbb{R}_+}^n$. When L is translation-invariant, we have the following lemma.

Lemma 1 *Let $h \in \mathscr{D}'(\mathbb{R}^n)$. If $L[f] = h * f$ for $f \in \mathscr{D}(\mathbb{R}^n)$, then the following three conditions are equivalent.*

(a) *L is causal.*

(b) *If f is causal then $L[f]$ is causal.*

(c) *h is causal.*

2 Wavelet Analysis of Linear Systems

Assume that $\psi \in L^2(\mathbb{R}^n)$ satisfy

$$0 < C_\psi := \int_0^\infty \frac{|\widehat{\psi}(s\xi)|^2}{s} \, ds < \infty, \tag{2.1}$$

where C_ψ is independent of ξ. Note that this condition is satisfied if $\psi(\neq 0)$ is a radially symmetric continuous function with compact support, and $\int_{\mathbb{R}^n} \psi(x) \, dx = 0$. For simplicity we further assume that $\|\psi\|_{L^2(\mathbb{R}^n)} = 1$.

The *continuous wavelet transform* of $f \in L^2(\mathbb{R}^n)$ with respect to ψ is defined by

$$W_\psi f(b,a) : = |a|^{-n/2} \int_{\mathbb{R}^n} f(x) \, \overline{\psi\Big(\frac{x-b}{a}\Big)} \, dx = \langle f, T_b D_a \psi \rangle_{L^2(\mathbb{R}^n)}, \quad (b,a) \in \mathbb{H}_n := \mathbb{R}^n \times \mathbb{R}_+. \tag{2.2}$$

It is well-known that $W_\psi f \in \mathscr{H}_1 := L^2(\mathbb{H}_n; dbda/a^{n+1})$ and

$$\langle W_\psi f, W_\psi g \rangle_{\mathscr{H}_1} = C_\psi \langle f, g \rangle_{L^2(\mathbb{R}^n)} \quad \text{for every } f, g \in L^2(\mathbb{R}^n). \tag{2.3}$$

(See, for example, [Gr01]. As for a group representation theoretic approach to wavelet transforms, see [Wo02].)

Let L be a bounded linear operator from $L^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$. We state theorems concerning the interaction between L and W_ψ. The proofs can be found in [AMM2].

Wavelet analysis of kernels

For $(b,a), (u,s) \in \mathbb{H}_n$, set

$$K_\psi(b,a;u,s) := C_\psi{}^{-1} \langle L[T_u D_s \psi], T_b D_a \psi \rangle_{L^2(\mathbb{R}^n)}. \tag{2.4}$$

Theorem 2 (1) *K_ψ is a bounded continuous function on $(\mathbb{H}_n)^2$, and*

$$\sup_{(b,a;u,s)\in(\mathbb{H}_n)^2} |K_\psi(b,a;u,s)| \le C_\psi{}^{-1}\|L\|_{op}, \tag{2.5}$$

where $\|L\|_{op}$ denotes the operator norm of L.
(2) *For every fixed $(u,s)\in\mathbb{H}_n$, we have*

$$K_\psi(\cdot,\cdot;u,s) = C_\psi{}^{-1}\, W_\psi(L[T_uD_s\psi]) \in \mathscr{H}_1, \tag{2.6}$$
$$\|K_\psi(\cdot,\cdot;u,s)\|_{\mathscr{H}_1} \le C_\psi{}^{-1/2}\|L\|_{op}. \tag{2.7}$$

Similarly, for every fixed $(b,a)\in\mathbb{H}_n$, we have

$$K_\psi(b,a;\cdot,\cdot) = C_\psi{}^{-1}\, \overline{W_\psi(L^*[T_bD_a\psi])} \in \mathscr{H}_1, \tag{2.8}$$
$$\|K_\psi(b,a;\cdot,\cdot)\|_{\mathscr{H}_1} \le C_\psi{}^{-1/2}\|L\|_{op}, \tag{2.9}$$

where $*$ denotes the adjoint operator.
(3) *For every $F\in\mathscr{H}_1$ and $(b,a)\in\mathbb{H}_n$, set*

$$L_\psi[F](b,a) := \int_{\mathbb{H}_n} K_\psi(b,a;u,s)\,F(u,s)\,du\frac{ds}{s^{n+1}} = \langle K_\psi(b,a;\cdot,\cdot),\overline{F}\rangle_{\mathscr{H}_1}.$$

Then, we have $L_\psi[F]\in\mathscr{H}_1$ for every $F\in\mathscr{H}_1$, and $\|L_\psi\|_{op}\le\|L\|_{op}$. Further, we have $W_\psi\, L = L_\psi\, W_\psi$, that is,

$$W_\psi(L[f]) = L_\psi[W_\psi f] \quad \textit{for every } f\in L^2(\mathbb{R}^n). \tag{2.10}$$

(4) *We have an inversion formula for L from L_ψ:*

$$L[f](x) = C_\psi{}^{-1}\int_{(\mathbb{H}_n)^2} K_\psi(b,a;u,s)\,W_\psi f(u,s) \times T_bD_a\psi(x)\,du\frac{ds}{s^{n+1}}db\frac{da}{a^{n+1}} \tag{2.11}$$

for every $f\in L^2(\mathbb{R}^n)$. Here, the L^2-valued integral can be considered, for example, in the weak sense:

$$\langle L[f],g\rangle = C_\psi{}^{-1}\int_{\mathbb{H}_n}\left(\int_{\mathbb{H}_n} K_\psi(b,a;u,s)\,W_\psi f(u,s)\,du\frac{ds}{s^{n+1}}\right) \times \langle T_bD_a\psi,g\rangle\,db\frac{da}{a^{n+1}} \tag{2.12}$$

or

$$\langle L[f], g\rangle = C_\psi{}^{-1} \lim_{M\to\infty} \int_{([-M,M]^n\times[1/M,M])^2} K_\psi(b,a;u,s) \times W_\psi f(u,s)\,\langle T_b D_a \psi, g\rangle\, du \frac{ds}{s^{n+1}} db \frac{da}{a^{n+1}} \quad (2.13)$$

for every $f, g \in L^2(\mathbb{R}^n)$. Note that $W_\psi f$ and $\langle T_b D_a\psi, g\rangle = \overline{W_\psi g}$ belong to $\mathscr{H}_1$.

Equality (2.10) can be written as

$$W_\psi(L[f])(b,a) = \int_{\mathbb{H}_n} K_\psi(b,a;u,s)\, W_\psi f(u,s)\, du \frac{ds}{s^{n+1}} \quad \text{for every } f \in L^2(\mathbb{R}^n). \quad (2.14)$$

Formula (2.14) enables us to access to information of K_ψ by wavelet transforms $W_\psi f$ and $W_\psi(L[f])$, which are computable from the observed input f and output $L[f]$.

When L is the identity operator, the kernel K_ψ is the reproducing kernel of the reproducing kernel Hilbert space Range(W_ψ) ([Da92], [Ma99]).

Theorem 2 and other contents of this section are deeply connected with [Wo02]. Roughly speaking, the localization operator in [Wo02] corresponds to the operator $W_\psi^{-1}(F\times)W_\psi$, where $F = F(b,a)$ is a bounded function of (b,a), while we are interested in $W_\psi L W_\psi^{-1}$. Since W_ψ is not injective, W_ψ^{-1} should be treated with care. We will choose the operator V_ψ such that $V_\psi W_\psi = I$, where I is the identity operator, and $W_\psi V_\psi$ has good properties. (See (6.22).)

Wavelet analysis of Hilbert–Schmidt kernels

When L is a Hilbert–Schmidt operator, that is, the kernel distribution k belongs to $L^2(\mathbb{R}^{2n})$, then we have the following additional result.

Theorem 3 *Let $k \in L^2(\mathbb{R}^{2n})$. Then,*

$$K_\psi(b,a;u,s) = C_\psi{}^{-1} \langle k, T_b D_a \psi \otimes \overline{T_u D_s \psi}\rangle_{L^2(\mathbb{R}^{2n})}, \quad (2.15)$$

where $(f\otimes g)(x,y) := f(x)g(y)$, and

$$K_\psi \in \mathscr{H}_2 := L^2\Big((\mathbb{H}_n)^2_{(b,a,u,s)}; db\frac{da}{a^{n+1}} du \frac{ds}{s^{n+1}}\Big),$$
$$\|K_\psi\|_{\mathscr{H}_2} = \|k\|_{L^2(\mathbb{R}^{2n})}.$$

Thus, L_ψ is also a Hilbert–Schmidt operator.

We also have an inversion formula for k:

$$k(x,y) = C_\psi^{-1} \int_{(\mathbb{H}_n)^2} K_\psi(b,a;u,s) \times T_b D_a \psi(x)\, \overline{T_u D_s \psi(y)}\, db \frac{da}{a^{n+1}} du \frac{ds}{s^{n+1}}, \quad (2.16)$$

where the integral is, for example, in the weak sense:

$$\langle k, h\rangle_{L^2(\mathbb{R}^{2n})} = C_\psi^{-1} \int_{(\mathbb{H}_n)^2} K_\psi(b,a;u,s) \times \langle T_b D_a \psi \otimes \overline{T_u D_s \psi}, h\rangle\, db \frac{da}{a^{n+1}} du \frac{ds}{s^{n+1}} \quad (2.17)$$

$$\text{for every } h \in L^2(\mathbb{R}^{2n}).$$

Note that $\langle T_b D_a \psi \otimes \overline{T_u D_s \psi}, h\rangle \in \mathscr{H}_2$, just like K_ψ.

We can access to time-frequency information of k by a similar way to the ordinary wavelet analysis, because (2.15) means that K_ψ is a kind of wavelet transform of k and (2.16) is a kind of inverse wavelet transform.

Next, assume condition (1.2) for k.

Theorem 4 *Let k satisfy (1.2). Then, for every $(b,a;u,s) \in (\mathbb{H}_n)^2$, we have $k(x,y) T_u D_s \psi(y) \overline{T_b D_a \psi(x)} \in L^1(\mathbb{R}^{2n})$ and*

$$K_\psi(b,a;u,s) = C_\psi{}^{-1} \int_{\mathbb{R}^{2n}} k(x,y)\, T_u D_s \psi(y)\, \overline{T_b D_a \psi(x)}\, dxdy. \quad (2.18)$$

We also have the inversion formula (2.16) for k, for example, in the following sense, which is weaker than (2.17).

$$\begin{aligned} &\int_{\mathbb{R}^{2n}} k(x,y)\, \overline{\phi_1(x)\phi_2(y)}\, dxdy \quad (2.19) \\ &= C_\psi^{-1} \int_{\mathbb{H}_n} \left(\int_{\mathbb{H}_n} K_\psi(b,a;u,s) \langle \overline{T_u D_s \psi}, \phi_2\rangle\, du \frac{ds}{s^{n+1}} \right) \times \langle T_b D_a \psi, \phi_1\rangle\, db \frac{da}{a^{n+1}} \\ &= C_\psi^{-1} \lim_{M\to\infty} \int_{([-M,M]^n \times [1/M,M])^2} K_\psi(b,a;u,s) \times \langle T_b D_a \psi, \phi_1\rangle\, \langle \overline{T_u D_s \psi}, \phi_2\rangle\, db \frac{da}{a^{n+1}} du \frac{ds}{s^{n+1}} \end{aligned}$$

for every $\phi_1, \phi_2 \in L^2(\mathbb{R}^n)$.

Wavelet analysis of translation-invariant systems

When the system L is translation-invariant, we have the following result, a discrete version of which will be used in subsequent numerical experiments.

Theorem 5 *Let L be a continuous translation-invariant linear system from $L^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$. Then, we have*

$$L_\psi[F] = P_\psi(G), \quad G(b,a) := L[F(\cdot,a)](b) \tag{2.20}$$

for every $F \in \mathscr{H}_1$, where P_ψ is the projection in $\mathscr{H}_1$ onto the range of W_ψ. In particular,

$$(W_\psi(L[f]))(\cdot,a) = L[W_\psi f(\cdot,a)] \tag{2.21}$$

for every $f \in L^2(\mathbb{R}^n)$ and $a \in \mathbb{R}_+$.

Let us discuss about system identification of a translation-invariant system L. In the real world, there are disturbances. After observing the input f, there could be an unmeasured disturbance ν^{in}. Then the real input to L is $f+\nu^{\text{in}}$. Similarly, before observing the output, there could be an unmeasured disturbance ν^{out}. Then the observed output from L is $L[f] + L[\nu^{\text{in}}] + \nu^{\text{out}}$. In this case, the observed input-output pair is $\{f,\ L[f] + L[\nu^{\text{in}}] + \nu^{\text{out}}\}$, which could cause a bad identification. Applying the continuous wavelet transform W_ψ to this observed input-output pair and using (2.21), then we have $\{W_\psi f,\ L[W_\psi f] + L[W_\psi \nu^{\text{in}}] + W_\psi \nu^{\text{out}}\}$. As the denoising property of the continuous wavelet transform reduces certain kinds of disturbances ν^{in} and ν^{out}, we may have an input-output pair close to $\{W_\psi f,\ L[W_\psi f]\}$, which causes a better identification. In another disturbed case, we may observe not f but $f + \nu^{in}$ as the input. By applying the continuous wavelet transform to the observed input-output pair $\{f + \nu^{in}, L[f + \nu^{in}] + \nu^{out}\}$, we could also have a better identification.

A function ψ is called *wavelet function for causality* if $W_\psi f(b,a)$ is causal with respect to b for every causal function f. If we define the *involution* $\mathcal{I}$ by

$$\mathcal{I}[g(x)] := \overline{g(-x)},$$

then wavelet transform $W_\psi f(b,a)$ can be represented as

$$W_\psi f(b,a) = (f * D_a \mathcal{I}\psi)(b). \tag{2.22}$$

Corollary 1 stated below follows easily from Lemma 1 and (2.22).

Corollary 1 *Let $a > 0$. Then, the following two conditions are equivalent.*

(i) *f is causal $\implies$ $W_\psi f(b,a)$ is causal with respect to b.*

(ii) $\mathcal{I}[\psi]$ *is causal.*

In the following section, we will use a discretized version of the wavelet function for causality constructed as follows. Take a continuous orthonormal wavelet function ψ with compact support such as Daubechies' orthonormal wavelet functions. Then, there exists $b \in \mathbb{R}$ such that $\operatorname{supp} T_b\mathcal{I}\psi \subset \overline{\mathbb{R}_+}$. Since $T_b\mathcal{I}\psi = \mathcal{I}T_{-b}\psi$, the function $\mathcal{I}T_{-b}\psi$ is causal. Hence $T_{-b}\psi$ is a wavelet function for causality, because the function $T_{-b}\psi$ is a continuous function with compact support satisfying

$$\int_{\mathbb{R}} T_{-b}\psi(x)\,dx = 0.$$

3 System Identification of Discrete Systems

Assume that the discrete system to be identified has the following form:

$$y_n = \sum_{\ell=0}^{m-1} \alpha_\ell \, x_{n-\ell}, \tag{3.1}$$

where $\{x_n\}$ is the input and $\{y_n\}$ is the output. The assumption means that we are considering a translation-invariant causal system of finite filter length. Denote the filter coefficients to be identified by

$$A = [\alpha_{m-1}, \alpha_{m-2}, \cdots, \alpha_0]^T.$$

There are various types of observable input-output pairs. We need to modify system identification methods according to the type. In this paper, we will deal with only one input-output pair of long length. This type of input-output pairs are observed, for example, in health monitoring systems for structures.

Health monitoring systems

Structures under dynamic load, such as buildings, bridges, and so on, store cumulative damages on their structural members. The main concern of health monitoring systems is to have an efficient identification method of the structural parameters and to find when those parameters have been changed. Although these damages are generally estimated by continuous observation of several measurements, such as acceleration, velocity and displacement at several observing points, health monitoring systems based on these measurements could be expensive. Therefore, approaches to health monitoring systems utilizing only one measurement are growing in importance. Such health monitoring systems have only one input-output pair of long length. In this case, we subdivide the input-output pair into an enough number of input-output pairs of short length.

Conventional method

First, let us explain the conventional method of system identification. Take N successive elements starting with the index k from the output $\{y_n\}$ and denote the N-dimensional column vector by

$$Y = [y_k, y_{k+1}, y_{k+2}, \cdots, y_{k+N-1}]^T.$$

Take a successive $N+m-1$ elements starting with the index $k-m+1$ from the input $\{x_n\}$ and construct an $N \times m$ matrix X defined by

$$X = \begin{bmatrix} x_{k-m+1} & x_{k-m+2} & \cdots & x_k \\ x_{k-m+2} & x_{k-m+3} & \cdots & x_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k+N-m} & x_{k+N-m+1} & \cdots & x_{k+N-1} \end{bmatrix}.$$

The conventional method solves $Y = XA$ by the least square method.

Wavelet method

Next, we propose our wavelet method of system identification. We use a wavelet function for causality. Then, the time-invariant discrete wavelet transform called **stationary wavelet transform** is represented as

$$S_{j+1,k} = \sum_{n=-\infty}^{k} \overline{h_{j,n-k}} S_{j,n}, \qquad D_{j+1,k} = \sum_{n=-\infty}^{k} \overline{g_{j,n-k}} S_{j,n}. \tag{3.2}$$

Let a pair of input $\{x_n\}$ and output $\{y_n\}$ be given. Put

$$S_{0,n}^{\text{in}} = x_n, \qquad S_{0,n}^{\text{out}} = y_n.$$

Applying the stationary wavelet transform (3.2) for causal systems, calculate inductively the approximation $S_{j,k}^{\text{in}}$ and the detail $D_{j,k}^{\text{in}}$ of level j for the input and the approximation $S_{j,k}^{\text{out}}$ and the detail $D_{j,k}^{\text{out}}$ of level j for the output. Then, as a discrete version of Theorem 5, we have the following Theorem 6.

Theorem 6 *Let a pair of input $\{x_n\}$ and output $\{y_n\}$ be given. Assume that the system to be identified has the form* (3.1). *Then,*

$$S^{\text{out}}_{j,k} = \sum_{\ell=0}^{m-1} \alpha_\ell \, S^{\text{in}}_{j,k-\ell}, \quad D^{\text{out}}_{j,k} = \sum_{\ell=0}^{m-1} \alpha_\ell \, D^{\text{in}}_{j,k-\ell}.$$

Choosing enough approximation pairs $\{(j_i, k_i)\}_{i \in I_a}$, $I_a = \{1, 2, \ldots, N_a\}$ and detail pairs $\{(j_i, k_i)\}_{i \in I_d}$, $I_d = \{1, 2, \ldots, N_d\}$, we have the following system to solve for A:

$$\begin{cases} S^{\text{out}}_{j_i,k_i} = \displaystyle\sum_{\ell=0}^{m-1} \alpha_\ell \, S^{\text{in}}_{j_i,k_i-\ell}, & i \in I_a, \\ D^{\text{out}}_{j_i,k_i} = \displaystyle\sum_{\ell=0}^{m-1} \alpha_\ell \, D^{\text{in}}_{j_i,k_i-\ell}, & i \in I_d. \end{cases} \tag{3.3}$$

The wavelet method solves (3.3) by the least square method.

4 Numerical Experiment

The aim of the following numerical experiment is to compare the conventional method with the wavelet method. Here we will deal with a prototypal mathematical model of simplified health monitoring systems. The model to be identified is not a translation-invariant system. Let us use MATLAB's colon operator. The expression $J : K$ is the same as the row vector $[J, J+1, \ldots, K]$, where J, $K \in \mathbb{Z}$ and $J \leq K$. For an input x_n, $n = 1 : 1028$, the output y_n is given by

$$\begin{cases} y_n = x_n - x_{n-1}, & n = 2 : 514, \\ y_n = x_n/2 - x_{n-1} + x_{n-2}/2, & n = 515 : 1028, \end{cases} \tag{4.1}$$

where $n = 515$ is the critical moment. This model changes its structural parameters at a moment, that is, the filter coefficients of the system are not constants but step functions.

In our paper [AMM1], we propose a system identification method based on wavelet analysis for a different kind of model called *ARX model* and give a numerical experiment on a simple model of vehicle suspension systems.

Outline of numerical experiment

Using the filter coefficients identified from the first part of input-output pair, compute an output, which will be called *predicted output,* from the real input.

Compare the predicted output and the real output to detect the critical moment for input-output pairs without noise and with white noise. We will explain only for the case with white noise. The input-output pair $(\widetilde{x}_n, \widetilde{y}_n)$ with white noise is illustrated in Figure 2.

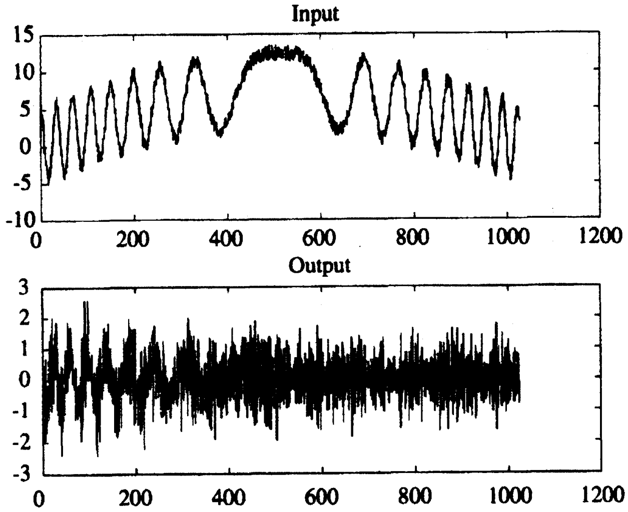


Figure 2: Input-output pair with white noise.

Input-output pair with white noise

Conventional method : Replacing the input-output pair (x_n, y_n) with the noised input-output pair $(\widetilde{x}_n, \widetilde{y}_n)$, do the same experiment as in the case without noise. Comparison of the predicted output $p\widetilde{y}_n$ with the real output $\widetilde{y}_n$ is illustrated with Figure 3 (a). It is impossible to detect the time when the system changed.

Wavelet method : Replacing the input-output pairs:

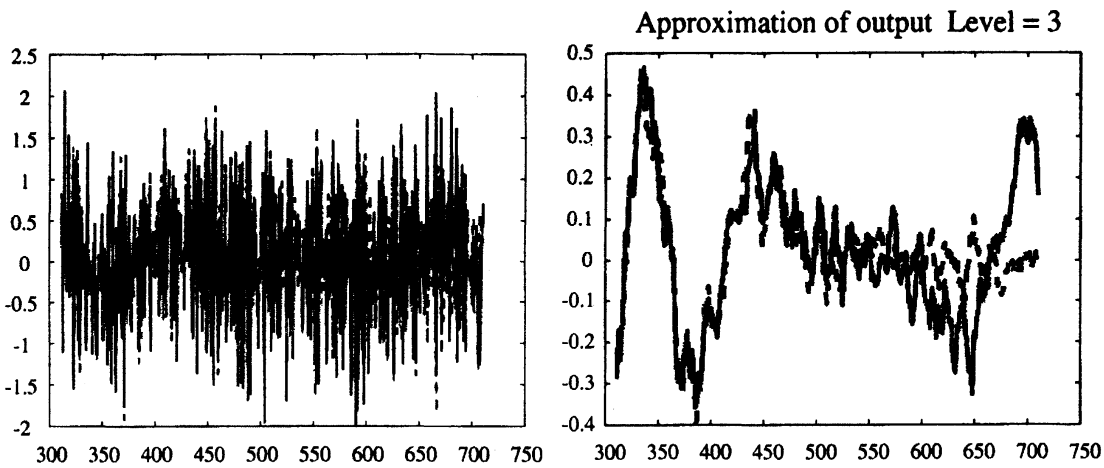
$$(D^{\text{in}}_{1,n}, D^{\text{out}}_{1,n}), \quad (D^{\text{in}}_{2,n}, D^{\text{out}}_{2,n}), \quad (S^{\text{in}}_{3,n}, S^{\text{out}}_{3,n}), \quad (D^{\text{in}}_{3,n}, D^{\text{out}}_{3,n})$$

with the noised input-output pairs:

$$(\widetilde{D}^{\text{in}}_{1,n}, \widetilde{D}^{\text{out}}_{1,n}), \quad (\widetilde{D}^{\text{in}}_{2,n}, \widetilde{D}^{\text{out}}_{2,n}), \quad (\widetilde{S}^{\text{in}}_{3,n}, \widetilde{S}^{\text{out}}_{3,n}), \quad (\widetilde{D}^{\text{in}}_{3,n}, \widetilde{D}^{\text{out}}_{3,n}),$$

respectively, do the same experiment as in the case without noise.

Comparison of the predicted output $P\widetilde{S}^{\text{out}}_{3,n}$ with the real output $\widetilde{S}^{\text{out}}_{3,n}$ is illustrated with Figure 3 (b). It is easy to detect the time when the system changed. The difference between $P\widetilde{S}^{\text{out}}_{3,n}$ and $\widetilde{S}^{\text{out}}_{3,n}$ is illustrated with Figure 4



(a) Conventional Method (b) Wavelet Method

Figure 3: Comparisons of the predicted outputs with the real outputs.

and those between $P\widetilde{D}^{\rm out}_{j,n}$ and $\widetilde{D}^{\rm out}_{j,n}$, $j = 1, 2, 3$ are illustrated with Figure 5. For each level $j = 1, 2, 3$, it is not so hard to detect the time when the system changed.

Conclusion of numerical experiment

For an input-output pair without noise, both the conventional and the wavelet methods can detect the critical moment. On the contrary, for an input-output pair with white noise, the conventional method cannot detect the critical moment but the wavelet method can do. This is because the wavelet method can filter out the white noise.

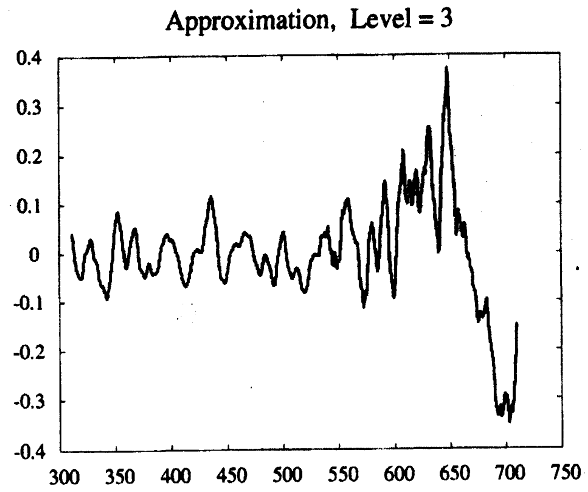


Figure 4: Difference between $P\widetilde{S}^{\text{out}}_{3,n}$ and $\widetilde{S}^{\text{out}}_{3,n}$.

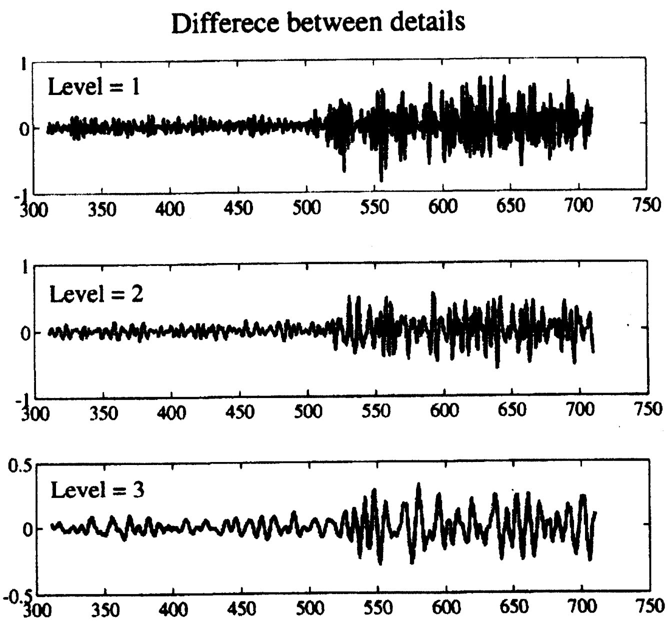


Figure 5: Wavelet, with white noise. Differences between predicted outputs $P\widetilde{D}^{\text{out}}_{j,n}$ and real outputs $\widetilde{D}^{\text{out}}_{j,n}$, $j = 1, 2, 3$.

References

[AMM1] Ashino, R., Mandai, T., Morimoto, A., Applications of wavelet transform to system identification, to appear in *Advances in pseudo-differential operators*, Birkhäuser, Basel.

[AMM2] Ashino, R., Mandai, T., Morimoto, A., System identification based on distribution theory and wavelet transform, to appear in Applicable Analysis.

[Da92] Daubechies, I., *Ten lectures on wavelets*, Regional Conference Series in Applied Math., SIAM, Philadelphia, PA, 1992.

[Gr01] Gröchenig, K., *Foundations of time-frequency analysis*, Birkhäuser Boston, Cambridge, MA, 2001.

[Hö60] Hörmander, L., Estimates for translation invariant operators in L^p spaces, *Acta Math.*, **104**(1960), 93–140.

[La02] Lax, P. D., *Functional analysis*, John Wiley & Sons, New York, NY, 2002.

[Ma99] Mallat, S., *A wavelet tour of signal processing*, Second Edition, Academic Press, San Diego, CA, 1999.

[Sc66] Schwartz, L., *Théorie des distributions*, Nouvelle édition, Hermann, Paris, 1966.

[St70] Stein, E. M., *Singular integrals and differentiability properties of functions*, Princeton University Press, Princeton, NJ, 1970.

[Tr67] Treves, F., *Topological vector spaces, distributions and kernels*, Academic Press, New York, NY, 1967.

[Wo02] Wong, M. W., *Wavelet transforms and localization operators*, Operator Theory: Advances and Applications **136**, Birkhäuser, Basel, 2002.