ウェーブレットの基礎理論とそのニューラルネットへの適用
ー動的環境における学習と適応を中心として（改訂版）

Wavelet basics and their applications to neural networks
— toward learning and adapting under nonstationary environment

大阪教育大学 数理科学 † ・情報科学 * ○芦野 隆一 † ・馬場 則夫 * ・守本 晃 *

R. Ashino†, N. Baba* and A. Morimoto*

Mathematical Sciences†, Information Science*, Osaka Kyoiku University

Abstract A review on several kinds of time-frequency analyses including short-time Fourier transform and wavelet transform is given. Neural networks based on wavelets called wavelet networks or fuzzy wavenets are explained and an application of neural networks to a pre-processing design for multiwavelet filters is given. As a future problem, learning and adapting under nonstationary environment are discussed.

1 はじめに

音楽を考えてみよう．音楽を録音すると関数（信号）$f(x)$ が得られる．この $f(x)$ のすべての値を使えばその音楽が完全に再現されるが，実際には楽器を演奏することによって音楽は作られている．たとえば，ある時刻にあるキーを叩くというように．もちろん関数 $f(x)$ が決まればそのフーリエ変換 $\widehat{f}(\xi)$ は決まり，逆にフーリエ変換 $\widehat{f}(\xi)$ が決まれば元の関数 $f(x)$ が決まるから，音楽を表現する情報としては $f(x)$ と $\widehat{f}(\xi)$ のどちらか一方がわかれば十分である．しかし，通常 $f(x)$ を見ただけでは $\widehat{f}(\xi)$ の挙動はわかりにくいし，逆に $\widehat{f}(\xi)$ を見ただけでは $f(x)$ の挙動はわかりにくい．たとえば，$f(x)$ からはメロディやリズムがわかり，$\widehat{f}(\xi)$ からはどのキーが叩かれているかがわかるのである．

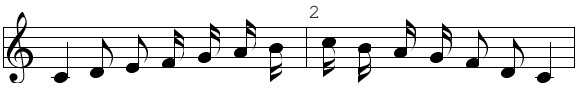

図 1: 音楽の時間周波数解析 — 楽譜

時間周波数解析とは，音楽を関数 $f(x)$（メロディやリズムだけ）あるいは $\widehat{f}(\xi)$（キーだけ）としてではなく，それらを合わせたようなある対象 F（楽譜）として捉えることであるといえるのではないだろうか．関数 $f(x)$ と $\widehat{f}(\xi)$ とは同じ情報を持っているはずであるが，それぞれの関数は対象 F の異なった情報にアクセスしやすいので，ひとつの対象 F の同値ではあるが異なる表現と見なせるのではないだろうか．このひとつの対象 F の 2 つの見方 $f(x)$ と $\widehat{f}(\xi)$ をひとつの関数で表現しようというのが時間周波数解析である．

時間周波数解析では，与えられた関数を**時間周波数原子**と呼ばれる時間と周波数の両方の領域で局在した関数（音楽でいえば音符♪）の線形結合に分解する．さらに，ある時間周波数原子の係数を適当に処理して再構成公式を使って再構成すれば（音楽でいえば音符を変更して演奏者に演奏してもらう），時間周波数領域で関数を処理することが可能となる．たとえば，絶対値がある閾値以下となる時間周波数原子の係数を 0 とすれば，より少ない数の時間周波数原子で与えられた関数を近似できることになり，データ量の圧縮が可能となる．

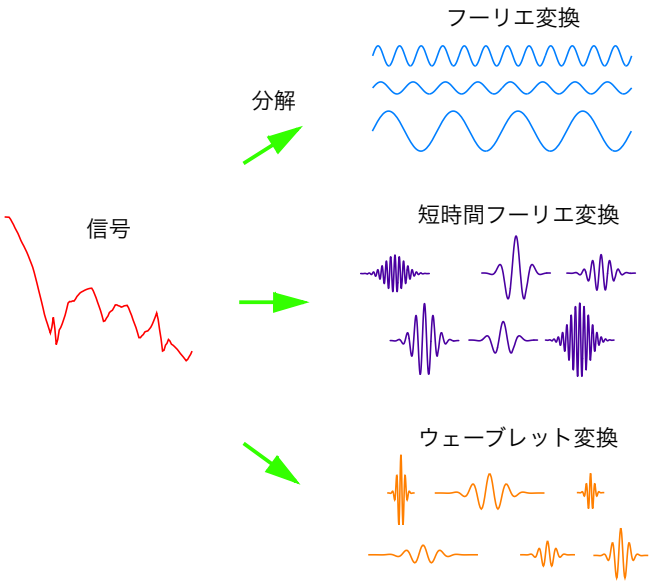

図 2: フーリエ変換と短時間フーリエ変換とウェーブレット変換

図 2 のように，フーリエ変換の場合は正弦波が，短時間フーリエ変換の場合は正弦波にガウシアンのような関数を乗じた波を平行移動した波が，ウェーブレット変換

の場合はひとつの波を伸縮したり平行移動した波が，それぞれ時間周波数原子となる．

2　フーリエ解析からの準備

- ヒルベルト空間 $L^2(\mathbb{R})$

$$L^2(\mathbb{R}) := \Big\{ f \mid \int_{\mathbb{R}} |f(x)|^2\,dx < +\infty \Big\},$$
$$\langle f, g\rangle := \int_{\mathbb{R}} f(x)\overline{g(x)}\,dx, \qquad ||f|| := \sqrt{\langle f, f\rangle}.$$

- 関数 f のフーリエ変換と関数 g の逆フーリエ変換を

$$\mathcal{F}[f(x)](\xi) = \widehat{f}(\xi) := \int e^{-ix\xi}\, f(x)\,dx,$$
$$\mathcal{F}^{-1}[g(\xi)](x) := \frac{1}{2\pi}\int e^{ix\xi}\, g(\xi)\,d\xi.$$

- パーセヴァルの等式

$$\langle f, g\rangle = \frac{1}{2\pi}\langle \widehat{f}, \widehat{g}\rangle, \qquad f, g \in L^2(\mathbb{R}).$$

- 3 つの基本的なユニタリ作用素：

平行移動作用素　$T_y f(x) = f(x-y)$,
変調作用素　$M_\xi f(x) = e^{ix\xi}\, f(x)$,
伸張作用素　$D_\rho f(x) = |\rho|^{-1/2} f(\rho^{-1}x)$, $\rho \in \mathbb{R}\backslash\{0\}$.

- 3 つの基本的な作用素とフーリエ変換の交換：

$\mathcal{F}[T_y f] = M_{-y}\mathcal{F}[f]$,　$\mathcal{F}[M_\omega f] = T_\omega \mathcal{F}[f]$,
$\mathcal{F}[D_\rho f] = D_{1/\rho}\mathcal{F}[f]$.

- $w(x) \in L^2(\mathbb{R})$ が**窓関数** $\overset{\text{def}}{\Longleftrightarrow}$　$xw(x) \in L^2(\mathbb{R})$.

窓関数 w の中心 $\overset{\text{def}}{\Longleftrightarrow}$　$x^* := \dfrac{1}{||w||^2}\displaystyle\int_{\mathbb{R}} x|w(x)|^2\,dx$.

窓関数 w の幅 $\overset{\text{def}}{\Longleftrightarrow}$

$$\Delta_w := \frac{1}{||w||}\Big(\int_{\mathbb{R}} (x-x^*)^2 |w(x)|^2\,dx\Big)^{1/2}.$$

- **フーリエ解析の不確定性原理**：

$$\Delta_w \Delta_{\widehat{w}} \geq 1/2.$$

等号成立は $w = c\,M_a T_b g_\alpha$,　$c \neq 0$,　$a, b \in \mathbb{R}$,

$$g_\alpha(x) := \frac{1}{2\sqrt{\pi\alpha}}\, e^{-x^2/4\alpha}, \quad \alpha > 0$$

の形のときに限る．

- 窓関数 w の時間周波数の窓
$\overset{\text{def}}{\Longleftrightarrow}$　(x,ξ)-平面で考えた長方形領域：

$$[x^* - \Delta_w, x^* + \Delta_w] \times [\xi^* - \Delta_{\widehat{w}}\,, \xi^* + \Delta_{\widehat{w}}\,].$$

3　短時間フーリエ変換

- f の窓関数 $w(x)$ による**短時間フーリエ変換**
$\overset{\text{def}}{\Longleftrightarrow}$　$\mathcal{F}[f(x)\overline{w(x-b)}] = \big\langle f, M_\xi T_b w\big\rangle$.

$$\big\langle f, M_\xi T_b w\big\rangle = \frac{1}{2\pi}\big\langle \widehat{f}, T_\xi M_{-b}\widehat{w}\big\rangle.$$

- 関数 $M_\xi T_b w$ で f から本質的に切り出される部分はどこか？　窓関数 $M_\omega T_b w$ の時間周波数の窓は（図 3）

$$[x^* + b - \Delta_w, x^* + b + \Delta_w] \times [\xi^* + \omega - \Delta_{\widehat{w}}\,, \xi^* + \omega + \Delta_{\widehat{w}}\,].$$

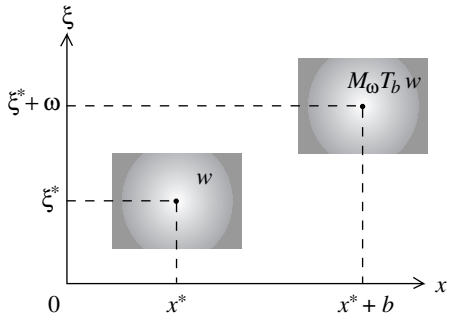

図 3: w の時間周波数の窓と $M_\omega T_b w$ の時間周波数の窓

- 高い振動数のところでは短時間，低い振動数のところでは長時間となるように，運動に応じて時間周波数の窓の形を変えたい．　$\Longrightarrow$　窓関数を $D_a w$ とする．

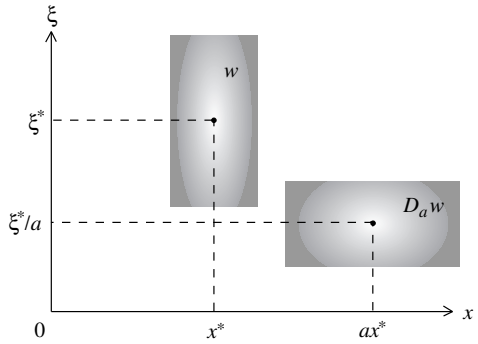

図 4: w の時間周波数の窓と $D_a w$ の時間周波数の窓

4　連続ウェーブレット変換

関数 $D_a w$ による短時間フーリエ変換で a をいろいろ取り替えれば，時間周波数解析に必要な窓が十分多く得られるが，効率的ではない．

$$\begin{aligned} M_\omega T_b D_a w &= e^{i\omega b}\, T_b M_\omega D_a w \\ &= e^{i\omega b}\, T_b D_a M_{a\omega} w = T_b D_a (e^{i\omega b}\, M_{a\omega} w) \end{aligned}$$

$\{e^{i\omega b}\, M_{a\omega} w\}$ のかわりに適当に振動しているひとつの関数 ψ をとれば，関数系 $\{T_b D_a \psi\}$ が十分多くの関数を含む． $\Longrightarrow$ $\langle f, T_b D_a \psi \rangle$ を考える．

- $a, b \in \mathbb{R},\ a \neq 0$（あるいは $a > 0$ ）に対し，

$$(W_\psi f)(b,a) := \langle f, T_b D_a \psi \rangle = \frac{1}{\sqrt{|a|}} \int_{\mathbb{R}} f(x) \overline{\psi\Big(\frac{x-b}{a}\Big)}\, dx$$

を**（連続）ウェーブレット変換**または**ウェーブレット係数**．

実数値関数 $\psi(x)$ とそのフーリエ変換 $\widehat{\psi}(\xi)$ は窓関数であって，

$$\int_{\mathbb{R}} \psi(x)\, dx = 0$$

を満たすならば，**逆ウェーブレット変換**：

$$f(x) = \frac{1}{C_\psi} \int_{\mathbb{R}} \frac{da}{|a|^2} \int_{\mathbb{R}} (W_\psi f)(b,a) \frac{1}{\sqrt{|a|}}\, \psi\left(\frac{x-b}{a}\right) db,$$

$$f(x) = \frac{2}{C_\psi} \int_{\mathbb{R}_+} \frac{da}{a^2} \int_{\mathbb{R}} (W_\psi f)(b,a) \frac{1}{\sqrt{a}}\, \psi\left(\frac{x-b}{a}\right) db$$

が定義できる．C_ψ は ψ から決まる定数．

- b は時間，$\frac{1}{a}$ は周波数，$d\left(\frac{1}{a}\right) = -\frac{1}{a^2}\, da$ だから，積分は時間と周波数に関する積分．ウェーブレット変換の時間周波数の窓はどうなっているのか？　関数 $T_b D_a w$ の時間周波数の窓は（図 5）

$$[ax^* + b - a\Delta_w, ax^* + b + a\Delta_w] \times \Big[\frac{\xi^*}{a} - \frac{\Delta_{\widehat{w}}}{a}, \frac{\xi^*}{a} + \frac{\Delta_{\widehat{w}}}{a}\Big].$$

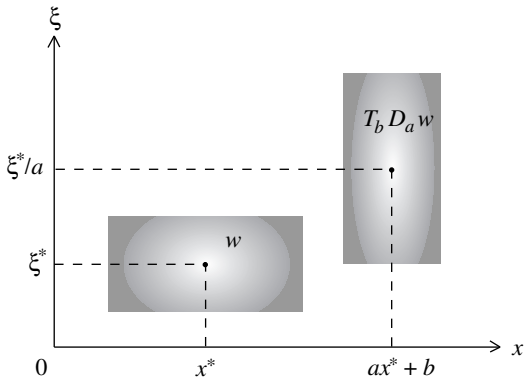

図 5: w の時間周波数の窓と $T_b D_a w$ の時間周波数の窓

短時間フーリエ変換とウェーブレット変換のそれぞれの窓関数による時間周波数平面のタイリングの様子を模式的に表現したのが図 6 である．

- 情報の過剰性と情報の強調
  - ▷ ウェーブレット変換は解析する信号の情報を過剰に含んでいる．

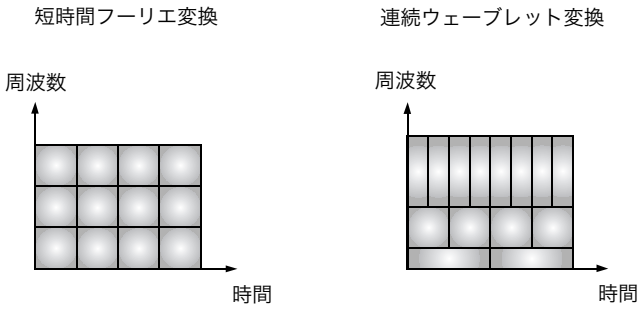

図 6: 短時間フーリエ変換とウェーブレット変換の窓関数によるタイリングの比較

  - ▷ ウェーブレット変換は，情報の繰返しによってある種の性質が強調されてデータが解析しやすくなったり，パターンを認識しやすくなる利点がある．
  - ▷ ウェーブレット変換は元の信号に比べ多くの情報を必要とするので効率的とはいえない．

- f の α 次**モーメント** $\overset{\text{def}}{\Longleftrightarrow}$ $\displaystyle\int_{\mathbb{R}} x^\alpha f(x)\, dx.$

f は α 次の**バニシングモーメント**を持つ $\overset{\text{def}}{\Longleftrightarrow}$ $\displaystyle\int_{\mathbb{R}^n} x^\alpha f(x)\, dx = 0.$

- 導関数の持つ情報とバニシングモーメント

L 次までのバニシングモーメントを持つならば，$N - 1 \leq L$ のとき，テイラー展開：

$$f(x) = \sum_{k=0}^{N-1} \frac{1}{k!}\, f^{(k)}(x_0)(x - x_0)^k + R_N(x)$$

をウェーブレット変換に代入して，

$$\langle f, T_b D_a \psi \rangle = \langle D_{1/a} T_{-b} f, \psi \rangle = \langle R_N, T_b D_a \psi \rangle.$$

  - ▷ ウェーブレット変換が $R_N(x)$ の情報，すなわち $f^{(N)}$ の情報にアクセスできる．
  - ▷ $d^N f/dx^N$ の特異点（急激に変化するところ）の情報を $W_\psi f(b,a)$ の a, b を動かすことにより取り出せる．

5　離散ウェーブレット変換

- 定積分はリーマン和で近似できるから，a, b を a_j, b_k, $j, k \in \mathbb{Z}$ と適当に離散化し，ψ から適当に $\widetilde{\psi}$ を作って

$$f(x) = \sum_{j,k \in \mathbb{Z}} \langle f, T_{b_k} D_{a_j} \psi \rangle\, T_{b_k} D_{a_j} \widetilde{\psi}(x)$$

と表せないか．$\psi_{j,k} := T_{b_k} D_{a_j} \psi$，$\widetilde{\psi}_{j,k} := T_{b_k} D_{a_j} \widetilde{\psi}$ とおけば，**ウェーブレットフレーム**：

$$f(x) = \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle\, \widetilde{\psi}_{j,k}.$$

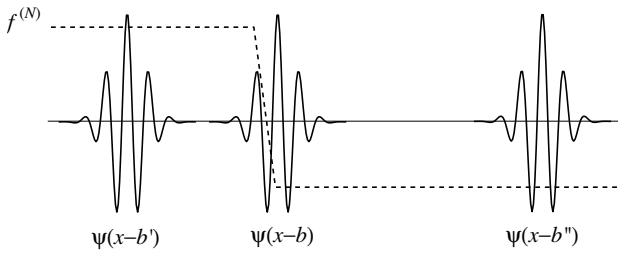

図 7: 特異点情報へのアクセス

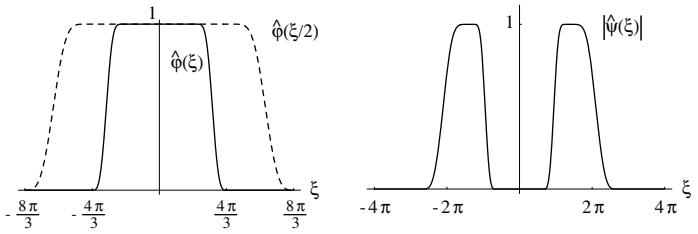

図 8: メイエのスケーリング関数とウェーブレット関数のフーリエ変換

6　ウェーブレット展開

- $f_{jk}(x) := 2^{j/2} f(2^j x - k), \quad j,k \in \mathbb{Z}$
- $\psi \in L^2(\mathbb{R})$ を**（正規直交）ウェーブレット関数** $\overset{\text{def}}{\Longleftrightarrow}$ $\{\psi_{jk}\}_{j,k\in\mathbb{Z}}$ が $L^2(\mathbb{R})$ の正規直交基底.
- $f \in L^2(\mathbb{R})$ の正規直交基底 $\{\psi_{jk}\}_{j,k\in\mathbb{Z}}$ に関する展開

$$f = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{jk}\rangle \, \psi_{jk}$$

を**ウェーブレット展開**.

- **多重解像度解析による方法**：

ウェーブレット関数 ψ は**スケーリング関数**と呼ばれる別の関数 φ から構成する.

- スケーリング関数は次の**伸張方程式**を満たす：

$$\varphi(x) = 2\sum_{k\in\mathbb{Z}} h_k \varphi(2x-k), \tag{1}$$
$$h_k := \langle \varphi(x), \varphi(2x-k)\rangle.$$

▷ $\{h_k\}$ を決めると φ が決まる.

- ウェーブレット関数は次の**ウェーブレット方程式**を満たす：

$$\psi(x) = 2\sum_{k\in\mathbb{Z}} g_k \varphi(2x-k), \tag{2}$$
$$g_k := \langle \psi(x), \varphi(2x-k)\rangle.$$

▷ $\{g_k\}$ は $g_k := (-1)^{1-k}\,\overline{h_{1-k}}$ で与えられる.

▷ スケーリング関数とウェーブレット関数から，それぞれローパスフィルタとハイパスフィルタが作れる.

- $\{\varphi_{j_0k}\}_{k\in\mathbb{Z}} \cup \{\psi_{jk}\}_{j\ge j_0,k\in\mathbb{Z}}$ が $L^2(\mathbb{R})$ の正規直交基底.

▷ 基底の取り替えにより，情報がゆっくり変化する部分（近似）の情報と急激に変化する部分（詳細）の情報に分類・整理される.

- 正規直交基底 $\{\varphi_{j_0k}\}_{k\in\mathbb{Z}} \cup \{\psi_{jk}\}_{j\ge j_0,k\in\mathbb{Z}}$ に関する展開

$$f = \sum\nolimits_{k\in\mathbb{Z}} \langle f, \varphi_{j_0k}\rangle \, \varphi_{j_0k} + \sum\nolimits_{j\ge j_0,k\in\mathbb{Z}} \langle f, \psi_{jk}\rangle \, \psi_{jk}$$

もウェーブレット展開.

- 周波数は $1/a$ に対応．$1/a = 2^j$ とおいたので周波数の対数がスケール j に対応.

$\Longrightarrow$ 厳密には**時間スケール解析**といい，時間周波数解析とは区別.

7　高速ウェーブレット変換アルゴリズム

- 整数 $j,k \in \mathbb{Z}$ に対し，

$$S_{j,k} := 2^{j/2}\langle f, \varphi_{jk}\rangle, \quad D_{j,k} := 2^{j/2}\langle f, \psi_{jk}\rangle.$$

- $\{S_{j,k}\}_{k\in\mathbb{Z}}$ と $\{D_{j,k}\}_{k\in\mathbb{Z}}$ は，$\{S_{j+1,k}\}_{k\in\mathbb{Z}}$ から

$$S_{j,k} = \sum\nolimits_{n\in\mathbb{Z}} \overline{h_{n-2k}}\, S_{j+1,n},$$
$$D_{j,k} = \sum\nolimits_{n\in\mathbb{Z}} \overline{g_{n-2k}}\, S_{j+1,n}.$$

$$\begin{array}{ccc} \{S_{-J+1,k}\} & \xrightarrow{(\downarrow 2)\circ \overline{m}_0} & \{S_{-J,k}\} \\ {\scriptstyle (\downarrow 2)\circ \overline{m}_1}\big\downarrow & & \\ \{D_{-J,k}\} & & \end{array}$$

- $\{S_{j+1,k}\}_{k\in\mathbb{Z}}$ は，$\{S_{j,k}\}_{k\in\mathbb{Z}}$ と $\{D_{j,k}\}_{k\in\mathbb{Z}}$ から

$$S_{j+1,k} = 2\sum\nolimits_{n\in\mathbb{Z}} h_{k-2n}\, S_{j,n} + 2\sum\nolimits_{n\in\mathbb{Z}} g_{k-2n}\, D_{j,n}.$$

$$\begin{array}{ccc} \{S_{-J,k}\} & \xrightarrow{2m_0\circ(\uparrow 2)} & \{S_{-J+1,k}\} \\ & & \big\uparrow{\scriptstyle 2m_1\circ(\uparrow 2)} \\ & & \{D_{-J,k}\} \end{array}$$

▷ $\{S_{j,k}\}_{k\in\mathbb{Z}}$ と $\{D_{j,k}\}_{k\in\mathbb{Z}}$ のデータ数は，それぞれ $\{S_{j+1,k}\}_{k\in\mathbb{Z}}$ のデータ数のほぼ半分.

▷ 多重レベルの分解と再構成を行う.

8　ウェーブレットネットワークス

逆ウェーブレット変換の離散化を考えよう．そのためには，関数 ψ を時間周波数領域で平行移動した N 個の関数 $T_{b_k}D_{a_k}\psi$ の線形結合

$$S_N(\{C_k,b_k,a_k\}) := \sum_{k=1}^{N} C_k T_{b_k} D_{a_k}\psi \tag{3}$$

で与えられた関数 f を近似すればよい．関数 f は離散化されていて，M 個の点の組 $\{(x_i, f(x_i))\}_{i=1,2,\cdots,M}$ として与えられているとする．このとき，2 乗誤差を最小化するように，すなわち，

$$\min \sum_{i=1}^{M} |f(x_i) - S_N(\{C_k,b_k,a_k\})(x_i)|^2 \tag{4}$$

を満たすように $\{C_k,b_k,a_k\}$ を選ぼう．最小化する目的関数を

$$\begin{aligned} &LSP(\{C_k,b_k,a_k\}) \\ &\quad := \sum_{i=1}^{M} |f(x_i) - S_N(\{C_k,b_k,a_k\})(x_i)|^2 \end{aligned} \tag{5}$$

とおく．

式 (3) を図 9 のように書き表し，これを**ウェーブレットネットワークス**という．

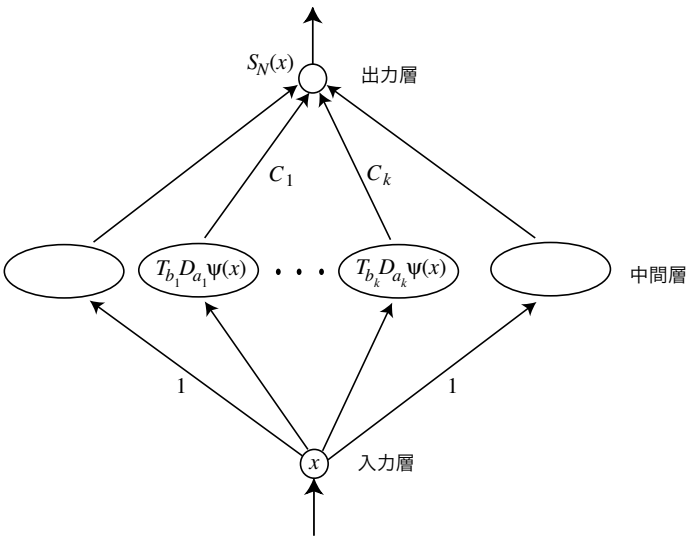

図 9: ウェーブレットネットワークス

特に式 (3) において，$T_{b_k}D_{a_k}\psi$ として

$$\psi_{jk}(x) := D_{1/2^j}T_k\psi = 2^{j/2}\psi(2^j x - k)$$

の形の出力関数を選べば，関数 ψ を時間周波数領域で平行移動した有限個の関数 ψ_{jk} の線形結合

$$S_N(\{C_{jk}\}) := \sum_{j,k} C_{jk}\psi_{jk} \tag{6}$$

を考えることになる．これは，ある種の非線形には有効であろう．さらに，スケーリング関数 φ が存在する場合には，ウェーブレット方程式 (2) を式 (6) の ψ に代入して，

$$S_N(\{C_{jk}\}) = \sqrt{2}\sum_{j,k,\ell} C_{jk} g_\ell \varphi_{j+1,2k+\ell} \tag{7}$$

を考え，係数 C_{jk} のみ最適化することになる．スケーリング関数 φ はラジアル基底 (radial-basis) 関数に近い働きをする．また，ある種のスケーリング関数 φ は，非負であり，対称性を持ち，ただひとつの最大値を持つから，ファジーのメンバーシップ関数に対応させることができる．このため，このウェーブレットネットワークスを**ファジーウェーブレットネットワークス**または**ファジーウェーブネット**と呼ぶ．

式 (4) の最小化問題を適当な初期値 $\{C_k^0,b_k^0,a_k^0\}$ から始めて，各係数で $LSP(\{C_k,b_k,a_k\})$ を偏微分した方向（最急降下法）や，共役勾配の方向に係数たちを少し動かすことを繰り返して，局所的最小値を求めることができる．手順としてまとめておく．

1. 適当な初期値を定める．
2. s 回目の近似値 $\{C_k^s,b_k^s,a_k^s\}$ に対して，$s+1$ 回目の近似値 $\{C_k^{s+1},b_k^{s+1},a_k^{s+1}\}$ を

$$\begin{aligned} C_k^{s+1} &= C_k^s - \eta_{s+1}\frac{\partial LSP}{\partial C_k}\Big|_{\{C_k^s,b_k^s,a_k^s\}} \\ b_k^{s+1} &= b_k^s - \eta_{s+1}\frac{\partial LSP}{\partial b_k}\Big|_{\{C_k^s,b_k^s,a_k^s\}} \\ a_k^{s+1} &= a_k^s - \eta_{s+1}\frac{\partial LSP}{\partial a_k}\Big|_{\{C_k^s,b_k^s,a_k^s\}} \end{aligned}$$

   で定める．ただし，$\eta_{s+1} > 0$ をうまく取る．
3. $LSP(\{C_k^{s+1},b_k^{s+1},a_k^{s+1}\})$ が十分小さくなったら計算を止める．

これがニューラルネットワークのバックプロパゲーション・アルゴリズムに対応しているので，離散逆ウェーブレット変換をニューラルネットワークを使って求めることができる．

どのようなウェーブレットを使うかに関してはいろいろな提案がある．

アルゴリズムの収束に関して重要な問題は次の 2 つである．

(i) ウェーブレットの個数を決めること．

(ii) 初期状態を決めること．

9　ウェーブレットネットワークスの応用

現在では様々なウェーブレットの一般化が試みられている．そのひとつにスケーリング関数を複数個使ったマルチウェーブレットがあり，マルチウェーブレット関数が複数個になるものの自由度が増すため，1 個のスケーリング関数から構成される通常のウェーブレットでは実現できなかった性質を持つ正規直交基底が構成されている．応用上よく使われるマルチスケーリング関数は互いに同じ程度のサポート幅を持つ．したがって，d 個のマルチスケーリング関数から構成されるマルチウェーブレットの場合は，離散データの 1 点に対してマルチでない場合の d 倍のスケーリング関数が重なっていることになる．このようなマルチウェーブレットを数値計算するための高速アルゴリズムには，これらの d 個のマルチスケーリング関数の展開係数に離散データをうまく分配することが必要不可欠となる．この分配問題をマルチウェーブレットの**プレフィルタ設計**という．

このプレフィルタには，d 個のスケーリング関数の取り方に依存してうまく行うことが重要となるが，未だほとんど議論されていない．

この問題に対して論文 [MAV] では，マルチウェーブレットネットワークを用いることを提案した．関数 $f(x)$ の定積分値を有限個の点における $f(x)$ の値の線形結合で近似するクアドラチュア公式などを使ったプレフィルタ設計もあり得るが，クアドラチュア公式は低次の多項式に対して積分値が正確であることを要請するため，滑らかでなかったり何度も振動しているようなスケーリング関数に対しては，あまり良い結果が得られないであろう．また，実際に展開係数を数値積分する方法も可能であるが，ニューラルネットワークを用いる場合には，解析する離散データに適応して分配がある意味で最適化されるため，多い項数で近似する場合などはとりわけ有効であると考えられる．

論文 [MAV] では，我々の提案するマルチウェーブレットネットワークを用いた場合とその他の方法による場合を比較検討して次の結果得られた．

(i) 高解像度な近似（係数の総数とデータ数が同じくらい）の場合には係数を数値積分で求めた場合よりも近似の精度が格段に上がる．

(ii) 低解像度の場合でも近似精度は数値積分で求めた場合よりも 1 桁位は良くなるが，処理時間を考えると逆行列で解く方法が良い．

(iii) 逆行列で偏微分が 0 となる係数を求める方法との比較について，データ数が多くなれば連立 1 次方程式を CG 法（共役勾配法）で解くことになるが，それと比較してマルチウェーブレットネットワークを使った方法は，計算量は同じくらいで使用メモリ量が少なくなるという長所がある．

10　連続線形システムのウェーブレット解析

連続線形システム L は，入力を f とし出力を g とすると，

$$g(x) = L[f(x)]$$

と表せる．シュワルツの核定理 [Tr67] によれば，連続線形システム L は，**核関数**と呼ばれるある超関数 $k(x,y)$ があって，

$$g(x) = \int_{\mathbb{R}} k(x,y) f(y)\, dy$$

と書ける．さらに L が時間不変であるときは，ある超関数 $h(x)$ があって，

$$g(x) = \int_{\mathbb{R}} h(x-y) f(y)\, dy$$

と書ける．つまり，任意の $a \in \mathbb{R}$ に対して，

$$\forall a \in \mathbb{R}:\ \ k(x,y) = k(x-a, y-a)$$

$$\Longleftrightarrow\ \ \exists h(r):\ \ k(x,y) = h(x-y)$$

が超関数の意味で成り立つことが示せる．

離散線形システム L の場合には，入力をベクトルと出力をベクトルをそれぞれ

$$\boldsymbol{f} := (f_1, \ldots, f_n)^T \in \mathbb{C}^n,$$
$$\boldsymbol{g} := (g_1, \ldots, g_n)^T \in \mathbb{C}^n$$

とすると，行列 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$ に対して，

$$\boldsymbol{g} = A\boldsymbol{f}$$

つまり，

$$L[\boldsymbol{f}] = A\boldsymbol{f}$$

となる．このとき，一次独立な n 個の入力 $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ と n 個の出力 $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n$ が既知ならば，

$$(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n) = A(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$$

であるから，行列 $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$ の逆行列 $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)^{-1}$ を右からかけて，

$$A = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n)(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)^{-1}$$

と求めることができる．

一般に，離散線形システム L が時間不変であるならば，添字の範囲が意味を持つような任意の $m \in \mathbb{Z}$ に対して，

$$a_{ij} = a_{i-m,j-m}, \qquad m \in \mathbb{Z}$$

が成り立つことが示せる．

適当な条件 [AMM1] を満たす連続線形システムの核関数 $k(x,y)$ に対して，

$$K(b,a,u,s) := \langle k(x,y), T_b D_a \psi(x) \cdot T_u D_s \overline{\psi(y)} \rangle_{(x,y)}$$

とおけば，

$$W_\psi g(b,a) = \frac{2}{C_\psi} \int_{\mathbb{R}_+} \frac{ds}{s^2} \int_{\mathbb{R}} K(b,a,u,s) W_\psi f(u,s)\, du$$

が成り立つことが示せる．さらに L が時間不変であるときは，

$$W_\psi g(x,a) = L[W_\psi f(x,a)] \tag{8}$$

が成り立つことが示せる．

11　時間不変システムとウェーブレット変換

ここでは (8) の離散化を考えてみよう．通常使われている高速ウェーブレット変換（マラー変換）[芦・山], [Mal] のような離散ウェーブレット変換では (8) のような可換性は成り立たないが，**定常ウェーブレット変換** [NS95] と呼ばれるダウンサンプリングを行わない離散ウェーブレット変換を使えば，数値計算的に L と離散ウェーブレット変換の交換が可能である．

定常ウェーブレット変換

レベル $j=1$ の点列 $S_{1,n}$ に対して，

$$S_{0,k} = \sum \overline{h^0_{n-k}} S_{1,n}, \quad D_{0,k} = \sum \overline{g^0_{n-k}} S_{1,n}$$

を対応させる．ただし，最初のフィルタ h^0 と g^0 は h と g と同じものである．このとき，レベル 0 の $S_{0,k}$ は元の数列 $S_{1,n}$ と同じサイズであり，$S_{0,k}$ をひとつおきに取ればダウンサンプリングを使う離散ウェーブレット変換を行った近似係数列が得られる．

レベル $j+1$ からレベル j を求めるには，フィルタ h^{j+1} と g^{j+1} を アップサンプリングしたフィルタ h^j と g^j を使って，

$$S_{j,k} = \sum \overline{h^j_{n-k}} S_{j+1,n}, \quad D_{j,k} = \sum \overline{g^j_{n-k}} S_{j+1,n}$$

を使う．つまり，レベル -1 を計算するには h^0 と g^0 をアップサンプリングしたフィルタ h^{-1} と g^{-1} のそれぞれと S_0 とのコンボリューションを計算する．レベル -1 の S_{-1} と D_{-1} は，元の数列と同じサイズになり 4 つおきに値をとればダウンサンプリングを使う離散ウェーブレット変換したレベル -1 の近似係数と詳細係数が得られる．このようにして定常ウェーブレット変換が定義される．

12　動的環境における学習と適応を目指して

今のところ，S_N は離散時間 x_i の関数であり，しかも最小 2 乗近似させた範囲の時間内でのみ信頼できる値を出力しないはずであるから，ウェーブレットをどのように動的環境に適応させるのか検討中である．実際，目標とすることは,「この時間範囲では，こういう時間周波数情報が出力されていた」という監視ができて，時間周波数情報が動的に変化したときにその変化を検出できて，またその変化に対応できるようなモデルである．

時間不変性の判定アルゴリズム

離散線形システム L に対する一次独立な n 個の入力 $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ と n 個の出力 $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n$ が既知ならば，離散線形システム L を表現する行列 A が求まることは既に示した．一般に，入力 f や出力 g の連続ウェーブレット変換 $W_\psi f$ や $W_\psi g$ は，入力 f や出力 g に比べ時間周波数領域での局在性がよいので，システムが時間不変であるかどうかを調べるための既知データとしては，より適切であると期待できる．

これらの結果を使って，線形システム L が時間不変であるかどうかの判定アルゴリズムを考えてみよう．

1. 線形システム L が時間不変であると仮定する．
2. 入力 f と出力 g の連続ウェーブレット変換 $W_\psi f$ と $W_\psi g$ を求める．
3. 線形システム L が時間不変性から，入力 $W_\psi f$ に対する出力は $W_\psi g$ である．
4. 離散化する．入力ベクトルを $W_\psi \boldsymbol{f}$ とし出力ベクトルを $W_\psi \boldsymbol{g}$ とする．
5. n 個の一次独立な入力ベクトル $W_\psi \boldsymbol{f}_j$ とそれらの出力ベクトルを $W_\psi \boldsymbol{g}_j$, $j = 1, \ldots, n$ から離散線形システム L を表現する行列 A を求める．
6. 行列 A が，添字の範囲が意味を持つような任意の $m \in \mathbb{Z}$ に対して，

   $$a_{ij} = a_{i-m,j-m}, \qquad m \in \mathbb{Z}$$

   を満たすかどうかを判定する．

7. 行列 A が上の時間不変であるための必要条件を満たさないならば，矛盾するので線形システム L は時間不変でない．(背理法)

数値実験

入出力データにノイズが入っている場合のシステム同定では，(8) に基づいた定常ウェーブレット変換を行う方が今までのシステム同定のアルゴリズムより効果的であることを示す数値実験例を与えよう．詳しくは [AMM1], [AMM2] を参照されたい．

区分的に時間不変システムであるが，途中でシステムが変わってしまうシステムを考える．データ数が 1024 の入力データ列 X_n に対して，前半の 512 個の出力が 2 回差分で与えられ，後半の 512 個の出力が 1 回差分で与えられる次のシステム：

$$Y_n = \begin{cases} X_n/2 - X_{n-1} + X_{n-2}/2, & \text{前半} \\ X_n - X_{n-1}, & \text{後半} \end{cases} \qquad (9)$$

を考える．

ノイズの無い場合

入力データ列を

$$X_n = 5.0 * \cos(n * 3 * \pi/64) + 30 * n/1028. * (1 - n/1028);$$

とした場合の入出力を図 10 に示す．中央付近で出力特性が変わっていることがわかる．

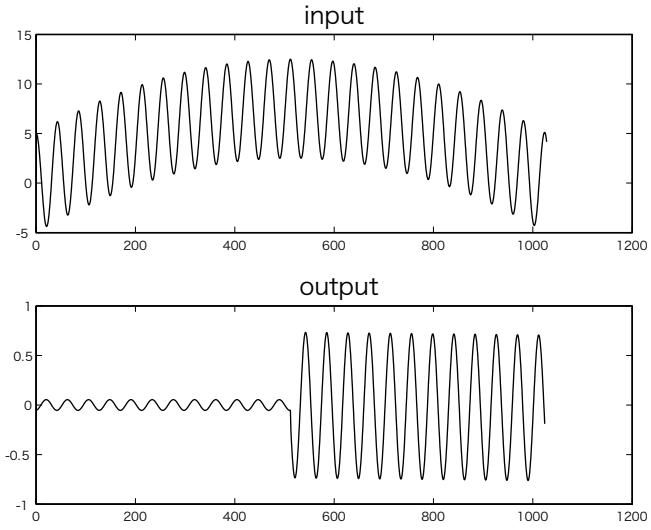

図 10: ノイズの無い場合の入出力

入出力のダウンサンプリングを行わない離散ウェーブレット変換（'db2' で 2 回変換）で近似を求めると，図 11 になる．ただし，入出力の中央部分の 256 点を抜き出した．図の中央の 128 点付近で変化していることがわかる．

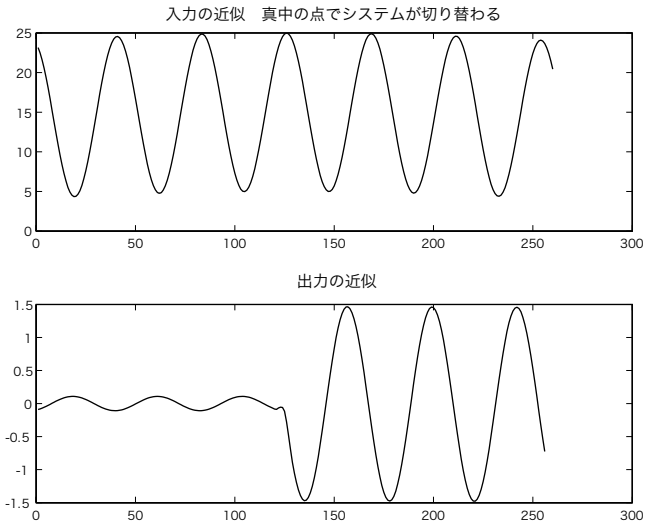

図 11: 入出力の近似：ノイズ無し

また，前半部分のシステムに離散ウェーブレット変換で近似した入力（図 11 の上の図）を入力したときの出力が図 12 の実線であり，出力 Y_n の近似（図 11 の下図）が点線である．この図も中央部分の 256 点を取り出している．これを見ると，前半部分（時間不変な部分）では両者が一致するので，数値計算上は L とダウンサンプリングを行わない離散ウェーブレット変換が交換可能なことがわかる．

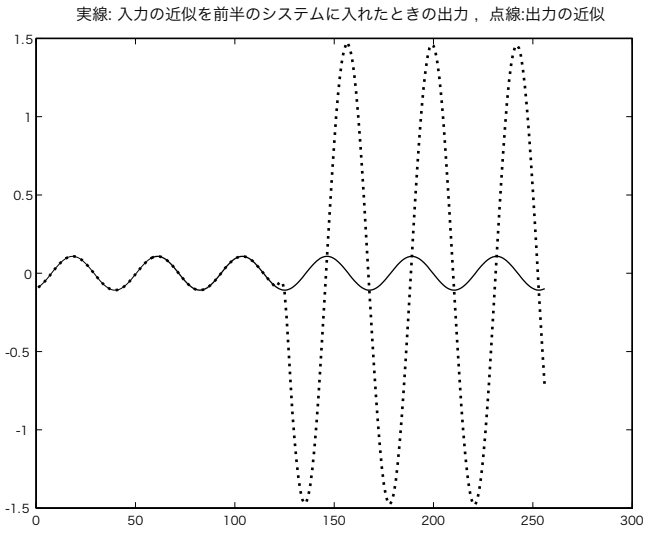

図 12: 近似入力の出力と出力の近似：ノイズ無し

ノイズの入った場合

次に入力 X_n を

$$X_n = 5.0 * \cos(n * 3 * \pi/64) + 30 * n/1028. * (1 - n/1028) + 2 * (\mathtt{rand}(1, 1028) - 0.5);$$

とノイズがある場合を考えよう．ただし，出力側にはノイズは入れない．システムにこの入力を入れたときの出力を図 13 に示す．よく見ると中央付近で出力特性が変わっていることがわかる．

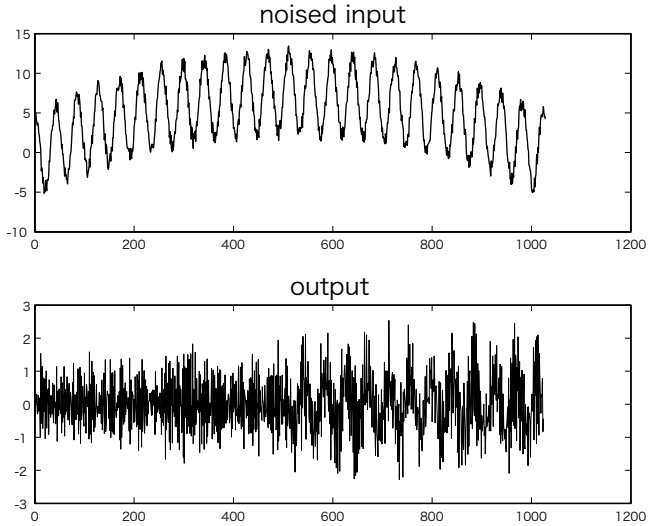

図 13: ノイズのありの入出力

入出力の定常ウェーブレット変換（'db2' 3 回変換）による近似を求めると，図 14 になる．ただし，入出力の中央部の 256 点を抜き出した．128 点付近で変変化しているのがよくわかる

システムにこの近似入力を入力したときの出力が図 15 の実線になり，図 14 の出力の近似が点線である．これを見ると，半分までは（時間不変な部分）両者が一致するので，数値計算上は L と定常離散ウェーブレット変換が交換可能なことがわかる．

ここでは，システムの同定（$g = Af$ の行列 A を求めること，あるいは入力と出力の関係式を求めること）は行っていないが，図 14 の入出力の近似を使うほうが図 13 の実際の入出力を使うより容易であると思われる．なを，ダウンサンプリングを行わない変換では近似のレベルは任意に取れるし，詳細を使っても同様の結果が出ると思われる．

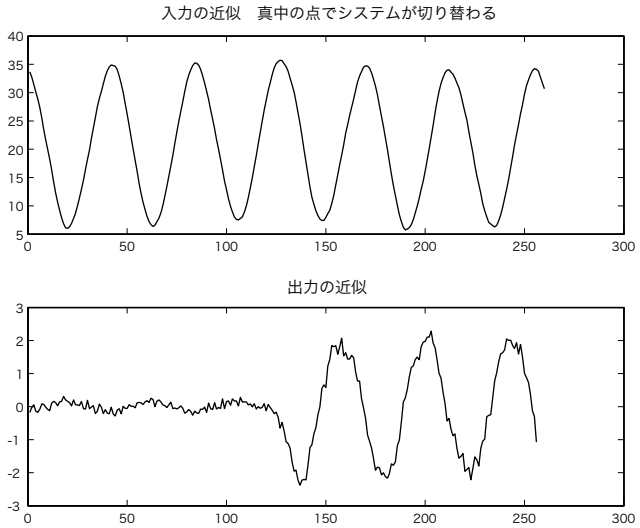

図 14: 入出力の近似：ノイズあり

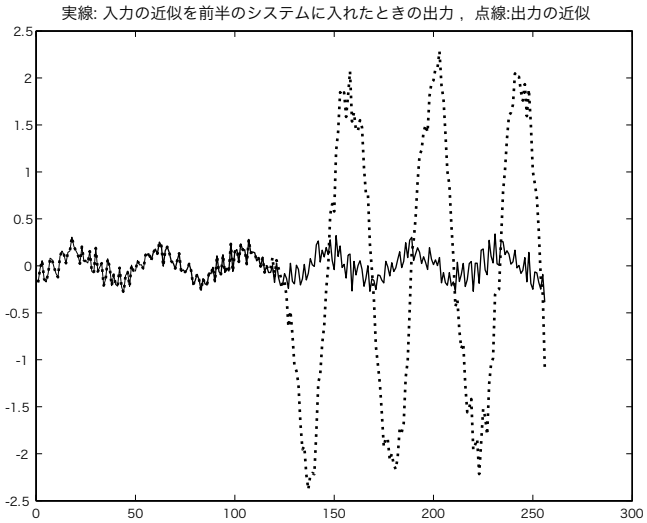

図 15: 近似入力の出力：ノイズあり

参考文献

[芦・山] 芦野 隆一・山本 鎭男，ウェーブレット解析 — 誕生・発展・応用 —，共立出版，1997.

[馬場] 馬場 則夫・小島 史男・小澤 誠一，ニューラルネットの基礎と応用，共立出版，1994.

[ハバード] バーバラ・バーク・ハバード 著　山田 道夫・西野 操 共訳，ウェーブレット入門 — 数学的道具の物語，朝倉書店，2003.

[AMM1] R. Ashino, T. Mandai and A. Morimoto, *System identification based on distribution theory and wavelet transform*, to appear in Applicable Analysis.

[AMM2] R. Ashino, T. Mandai and A. Morimoto, *Application of wavelet transform to system identification*, to appear in Advances in pseudo-differential operators, Birkhauser, Basel.

[Dau] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[GM] A. Grossmann and J. Morlet, *Decomposition of Hardy functions into square integrable wavelets of constant shape*, SIAM J. Math. Anal., **15** (1984), 723–736.

[ICP] S. Iyengar, E. Cho, and V. Phoha, *Foundations of Wavelet Networks and Applications*, Chapman & Hall/CRC, Boca Raton, Florida, 2002.

[Mal] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd edition, Academic Press, San Diego CA, 1999.

[MAV] A. Morimoto, R. Ashino and R. Vaillancourt, *Pre-processing design for multiwavelet filters using neural networks*, to appear in International J. of Wavelets, Multiresolution and Information.

[NS95] G. Nason and B. Silverman, stationary wavelet transform and some statistical applications, *Lecture Notes in Statistics*, Springer-Verlag, New York, NY, **103**, pp. 281–299, 1995.

[Thu] M. Thuillard, *Wavelets in Soft Computing*, World Scientific, Singapore, 2001.

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