

Vidakovic, Brani

Working Paper

Transforms in Statistics

Papers, No. 2004,26

Provided in Cooperation with:

CASE - Center for Applied Statistics and Economics, Humboldt University Berlin

Suggested Citation: Vidakovic, Brani (2004) : Transforms in Statistics, Papers, No. 2004,26, Humboldt-Universität zu Berlin, Center for Applied Statistics and Economics (CASE), Berlin

This Version is available at:

<https://hdl.handle.net/10419/22199>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Transforms in Statistics

Brani Vidakovic

School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive, Atlanta, GA 30332-0205 brani@isye.gatech.edu

It is not an overstatement to say that statistics is based on various transformations of data. Basic statistical summaries such as the sample mean, variance, z-scores, histograms, etc., are all transformed data. Some more advanced summaries, such as principal components, periodograms, empirical characteristic functions, etc., are also examples of transformed data. To give a just coverage of transforms utilized in statistics will take a size of a monograph. In this chapter we will focus only on several important transforms with the emphasis on novel multiscale transforms (wavelet transforms and its relatives).

Transformations in statistics are utilized for several reasons, but unifying arguments are that transformed data

- (i) are easier to report, store, and analyze,
- (ii) comply better with a particular modeling framework, and
- (iii) allow for an additional insight to the phenomenon not available in the domain of non-transformed data.

For example, variance stabilizing transformations, symmetrizing transformations, transformations to additivity, Laplace, Fourier, Wavelet, Gabor, Wigner-Ville, Hugh, Mellin, transforms all satisfy one or more of points listed in (i-iii).

We emphasize that words *transformation* and *transform* are often used interchangeably. However, the semantic meaning of the two words seem to be slightly different. For the word transformation, the synonyms are alteration, evolution, change, reconfiguration. On the other hand, the word transform carries the meaning of a more radical change in which the nature and/or structure of the transformed object are altered. In our context, it is natural that processes which alter the data leaving them unreduced in the same domain should be called transformations (for example Box-Cox transformation) and the processes that radically change the nature, structure, domain, and dimension of data should be called transforms (for example Wigner-Ville transform).

In this Chapter we focus mainly on transforms providing an additional insight on data. After the introduction discussing three examples, several important transforms are overviewed. We selected discrete Fourier, Hilbert, and Wigner-Ville transforms, discussed in Section 2, and given their recent popularity, continuous and discrete wavelet transforms discussed in Sections 3 and 4.

1 Introduction

As an “appetizer” we give two simple examples of use of transformations in statistics, Fisher z and Box-Cox transformations as well as the empirical Fourier-Stieltjes transform.

Example 1. Assume that we are looking for variance transformation $Y = \vartheta(X)$, in the case where $\text{Var } X = \sigma_X^2(\mu)$ is a function of the mean $\mu = \text{E } X$. The first order Taylor expansion of $\vartheta(X)$ about mean μ is

$$\vartheta(X) = \vartheta(\mu) + (X - \mu)\vartheta'(\mu) + O[(X - \mu)^2].$$

Ignoring quadratic and higher order terms we see that

$$\text{E } \vartheta(X) \approx \vartheta(\mu), \quad \text{Var } \vartheta(X) \approx \text{E} [(X - \mu)^2 \vartheta'(\mu)^2] = [\vartheta'(x)]^2 \sigma_X^2(\mu).$$

If $\text{Var}(\vartheta(X))$ is to be c^2 , we obtain

$$[\vartheta'(x)]^2 \sigma_X^2(\mu) = c^2$$

resulting in

$$\vartheta(x) = c \int \frac{dx}{\sigma_X(x)} dx.$$

This is a theoretical basis for the so-called Fisher z -transformation.

Let $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$ be a sample from bivariate normal distribution $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, and $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$, $i = 1, 2$.

The Pearson coefficient of linear correlation

$$r = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{[\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 \cdot \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2]^{1/2}}$$

has a complicated distribution involving special functions, e.g., Anderson (1984), p.113. However, it is well known that the asymptotic distribution for r is normal $N(\rho, \frac{(1-\rho^2)^2}{n})$. Since the variance is a function of mean,

$$\begin{aligned} \vartheta(\rho) &= \int \frac{c\sqrt{n}}{1-\rho^2} d\rho \\ &= \frac{c\sqrt{n}}{2} \int \left(\frac{1}{1-\rho} + \frac{1}{1+\rho} \right) d\rho \\ &= \frac{c\sqrt{n}}{2} \log \left(\frac{1+\rho}{1-\rho} \right) + k \end{aligned}$$

is known as Fisher z -transformation for the correlation coefficient (usually for $c = 1/\sqrt{n}$ and $k = 0$). Assume that r and ρ are mapped to z and ζ as

$$z = \frac{1}{2} \log \left(\frac{1+r}{1-r} \right) = \operatorname{arctanh} r, \quad \zeta = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right) = \operatorname{arctanh} \rho.$$

The distribution of z is approximately normal $N(\zeta, \frac{1}{n-3})$ and this approximation is quite accurate when ρ^2/n^2 is small and when n is as low as 20. The use of Fisher z -transformation is illustrated on finding the confidence intervals for ρ and testing hypotheses about ρ .

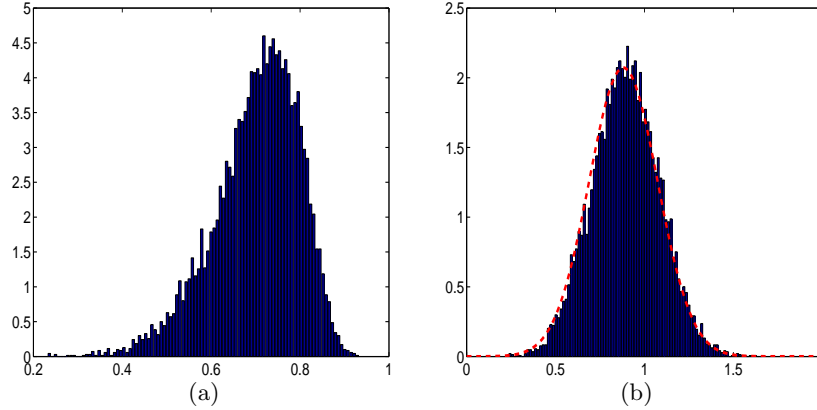


Fig. 1. (a) Simulational run of 10000 r 's from the bivariate population having theoretical $\rho = \sqrt{2}/2$.; (b) The same r 's transformed to z 's with the normal approximation superimposed.

To exemplify the above, we generated $n = 30$ pairs of normally distributed random samples with theoretical correlation $\sqrt{2}/2$. This was done by generating two i.i.d. normal samples a , and b of length 30 and taking the transformation $x_1 = a + b$, $x_2 = b$. The sample correlation coefficient r is found. This was repeated $M = 10000$ times. The histogram of 10000 sample correlation coefficients is shown in Fig. 1(a). The histogram of z -transformed r 's is shown in Fig. 1(b) with superimposed normal approximation $N(\operatorname{arctanh}(\sqrt{2}/2), 1/(30-3))$.

(i) For example, $(1 - \alpha)100\%$ confidence interval for ρ is:

$$\left[\tanh \left(z - \frac{\Phi^{-1}(1 - \alpha/2)}{\sqrt{n-3}} \right), \tanh \left(z + \frac{\Phi^{-1}(1 - \alpha/2)}{\sqrt{n-3}} \right) \right],$$

where $z = \operatorname{arctanh}(r)$ and $\tanh x = (e^x - e^{-x})/(e^x + e^{-x})$ and Φ stands for the standard normal cumulative distribution function.

If $r = -0.5687$ and $n = 28$ $z = -0.6456$, $z_L = -0.6456 - \frac{1.96}{5} = -1.0376$ and $z_U = -0.6456 + \frac{1.96}{5} = -0.2536$. In terms of ρ the 95% confidence interval is $[-0.7769, -0.2483]$.

(ii) Assume that two samples of size n_1 and n_2 , respectively, are obtained from two different bivariate normal populations. We are interested in testing $H_0 : \rho_1 = \rho_2$ against the two sided alternative. After observing r_1 and r_2 and transforming them to z_1 and z_2 , we conclude that the p -value of the test is $2\Phi\left(\frac{-|z_1 - z_2|}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}}\right)$.

Example 2. Box and Cox (1964) introduced a family of transformations, indexed by real parameter λ , applicable to positive data X_1, \dots, X_n ,

$$Y_i = \begin{cases} \frac{X_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log X_i, & \lambda = 0. \end{cases} \quad (1)$$

This transformation is mostly applied to responses in linear models exhibiting non-normality and/or heteroscedasticity. For properly selected λ , transformed data Y_1, \dots, Y_n may look “more normal” and amenable to standard modeling techniques. The parameter λ is selected by maximizing the log-likelihood,

$$(\lambda - 1) \sum_{i=1}^n \log X_i - \frac{n}{2} \log \left[\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 \right], \quad (2)$$

where Y_i are given in (1) and $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.

As an illustration, we apply the Box-Cox transformation to apparently skewed data of CEO salaries.

Forbes magazine published data on the best small firms in 1993. These were firms with annual sales of more than five and less than \$350 million. Firms were ranked by five-year average return on investment. One of the variables extracted is the annual salary of the chief executive officer for the first 60 ranked firms (since one datum is missing, the sample size is 59). Fig. 2(a) shows the histogram of raw data (salaries). The data show moderate skewness to the right. Panel (b) gives the values of likelihood in (2) for different values of λ . Note that (2) is maximized for λ approximately equal to 0.45. Fig. 2(c) gives the transformed data by Box-Cox transformation with $\lambda = 0.45$. The histogram of transformed salaries is notably symmetrized.

Example 3. As an example of transforms utilized in statistics, we provide an application of empirical Fourier-Stieltjes transform (empirical characteristic function) in testing for the independence.

The characteristic function of a probability distribution F is defined as its Fourier-Stieltjes transform,

$$\varphi_X(t) = E \exp(itX), \quad (3)$$

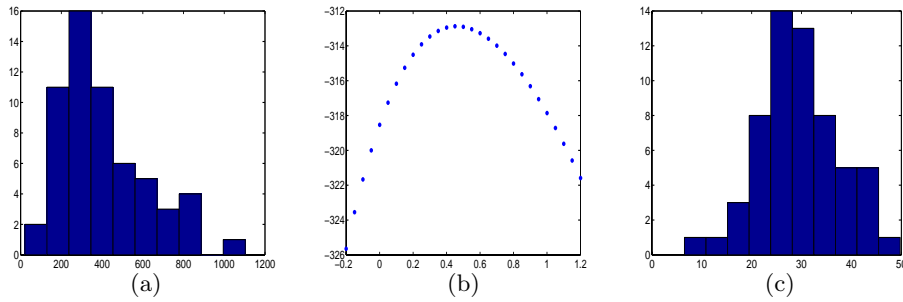


Fig. 2. (a) Histogram of row data [CEO salaries]; (b) Log-likelihood is maximized at $\lambda = 0.45$; and (c) Histogram of Box-Cox-transformed data.

where E is expectation and random variable X has distribution function F . It is well known that the correspondence of characteristic functions and distribution functions is 1-1, and that closeness in the domain of characteristic functions corresponds to closeness in the domain of distribution functions. In addition to uniqueness, characteristic functions are bounded. The same does not hold for moment generating functions which are Laplace transforms of distribution functions.

For a sample X_1, X_2, \dots, X_n one defines empirical characteristic function $\varphi^*(t)$ as

$$\varphi_X^*(t) = \frac{1}{n} \sum_{j=1}^n \exp(itX_j).$$

The result by Feuerverger and Mureika (1977) establishes the large sample properties of the empirical characteristic function

Theorem 1. For any $T < \infty$

$$P \left[\lim_{n \rightarrow \infty} \sup_{|t| \leq T} |\varphi^*(t) - \varphi(t)| = 0 \right] = 1$$

holds. Moreover, when $n \rightarrow \infty$, the stochastic process

$$Y_n(t) = \sqrt{n}(\varphi^*(t) - \varphi(t)), \quad |t| \leq T,$$

converges in distribution to a complex-valued Gaussian zero-mean process $Y(t)$ satisfying $Y(t) = \overline{Y(-t)}$ and

$$E(Y(t)\overline{Y(s)}) = \varphi(t+s) - \varphi(t)\varphi(s),$$

where $\overline{Y(t)}$ denotes complex conjugate of $Y(t)$.

Following Murata (2001) we describe how the empirical characteristic function can be used in testing for the independence of two components in bivariate distributions.

Given the bivariate sample (X_i, Y_i) , $i = 1, \dots, n$, we are interested in testing for independence of the components X and Y . The test can be based on the following bivariate process,

$$Z_n(t, s) = \sqrt{n}(\varphi_{X,Y}^*(t+s) - \varphi_X^*(t)\varphi_Y^*(s)),$$

where $\varphi_{X,Y}^*(t+s) = \frac{1}{n} \sum_{j=1}^n \exp(itX_j + isY_j)$.

Murata (2001) shows that $Z_n(t, s)$ has Gaussian weak limit and that

$$\begin{aligned} \text{Var } Z_n(t, s) &\approx [\varphi_X^*(2t) - (\varphi_X^*(t))^2][\varphi_Y^*(2s) - (\varphi_Y^*(s))^2], \text{ and} \\ \text{Cov}(Z_n(t, s), \overline{Z_n(t, s)}) &\approx (1 - |\varphi_X^*(t)|^2)(1 - |\varphi_Y^*(s)|^2), \end{aligned}$$

The statistics

$$T(t, s) = (\Re Z_n(t, s) \quad \Im Z_n(t, s)) \Sigma^{-1} (\Re Z_n(t, s) \quad \Im Z_n(t, s))'$$

has approximately χ^2 distribution with 2 degrees of freedom for any t and s finite. Symbols \Re and \Im stand for the real and imaginary parts of a complex number. The matrix Σ is 2×2 matrix with entries

$$\begin{aligned} \varsigma_{11} &= \frac{1}{2}(\Re \text{Var}(Z_n(t, s)) + \text{Cov}(Z_n(t, s), \overline{Z_n(t, s)})) \\ \varsigma_{12} = \varsigma_{21} &= \frac{1}{2}\Im \text{Var}(Z_n(t, s)), \text{ and} \\ \varsigma_{22} &= \frac{1}{2}(-\Re \text{Var}(Z_n(t, s)) + \text{Cov}(Z_n(t, s), \overline{Z_n(t, s)})). \end{aligned}$$

Any fixed pair t, s gives a valid test, and in the numerical example we selected $t = 1$ and $s = 1$ for calculational convenience.

We generated two independent components from the Beta(1,2) distribution of size $n = 2000$ and found T statistics and corresponding p -values $M = 2000$ times. Fig. 3(a,b) depicts histograms of T statistics and p values based on 2000 simulations. Since the generated components X and Y are independent, the histogram for T agrees with asymptotic χ^2_2 distribution, and of course, the p -values are uniform on $[0,1]$. In Fig. 3(c) we show the p -values when the components X and Y are not independent. Using two independent Beta(1,2) components X and Y' , the second component Y is constructed as $Y = 0.03X + 0.97Y'$. Notice that for majority of simulational runs the independence hypothesis is rejected, i.e., the p -values cluster around 0.

2 Fourier and Related Transforms

Functional series have a long history that can be traced back to the early nineteenth century. French mathematician (and politician) Jean-Baptiste-Joseph

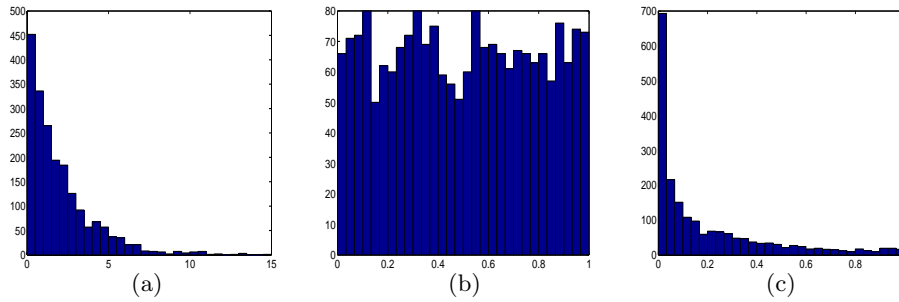


Fig. 3. (a) Histogram of observed T statistics with theoretical χ_2^2 distribution; (b) p -values of the test when components are independent; and (c) p -values if the test when the second component is a mixture of an independent sample and 3% of the first component.

Fourier, decomposed a continuous, periodic on $[-\pi, \pi]$ function $f(x)$ into the series of sines and cosines,

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

where the coefficients a_n and b_n are defined as

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \quad n = 0, 1, 2, \dots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx, \quad n = 1, 2, \dots$$

The sequences $\{a_n, n = 0, 1, \dots\}$ and $\{b_n, n = 1, 2, \dots\}$ can be viewed as a transform of the original function f . It is interesting that at the time of Fourier's discovery the very notion of function was not precisely defined. Fourier methods have long history in statistics especially in the theory of nonparametric function and density estimation and characteristic functions.

There are three types of Fourier transforms: integral, serial, and discrete. Next, we focus on discrete transforms and some modifications of the integral transform.

2.1 Discrete Fourier Transform

The discrete Fourier transform (DFT) of a sequence $\mathbf{f} = \{f_n, n = 0, 1, \dots, N-1\}$ is defined as

$$\mathbf{F} = \left\{ \sum_{n=0}^{N-1} f_n w_N^{nk}, \quad k = 0, \dots, N-1 \right\},$$

where $w_N = e^{-i2\pi/N}$. The inverse is

$$\mathbf{f} = \left\{ \frac{1}{N} \sum_{k=0}^{N-1} F_k w_N^{-nk}, n = 0, \dots, N-1 \right\}.$$

The DFT can be interpreted as the multiplication of the input vector by a matrix; therefore, the discrete Fourier transform is a linear operator. If $Q = \{Q_{nk} = e^{-i2\pi nk}\}_{N \times N}$, then $\mathbf{F} = Q \cdot \mathbf{f}$. The matrix Q is unitary (up to a scale factor), i.e., $Q^*Q = NI$, where I is the identity matrix and Q^* is the conjugate transpose of Q .

There are many uses of discrete Fourier transform in statistics. It turns cyclic convolutions into component-wise multiplication, and the fast version of DFT has a low computational complexity of $O(n \log(n))$, meaning that the number of operations needed to transform an input of size n is proportional to $n \log(n)$. For a theory and various other uses of DFT in various fields reader is directed to Brigham (1988).

We focus on estimation of a spectral density from the observed data, as an important statistical task in a variety of applied fields in which the information about frequency behavior of the phenomena is of interest.

Let $\{X_t, t \in Z\}$ be a real, weakly stationary time series with zero mean and autocovariance function $\gamma(h) = EX(t+h)X(t)$. An absolutely summable complex-valued function $\gamma(\cdot)$ defined on integers is the autocovariance function of X_t if and only if the function

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega} \quad (4)$$

is non-negative for all $\omega \in [-\pi, \pi]$. The function $f(\omega)$ is called the spectral density associated with covariance function $\gamma(h)$, and is in fact a version of discrete Fourier transform of the autocovariance function $\gamma(h)$. The spectral density of a stationary process is a symmetric and non-negative function. Given the spectral density, the autocovariance function can uniquely be recovered via inverse Fourier transform,

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{ih\omega} d\omega, \quad h = 0, \pm 1, \pm 2, \dots$$

A traditional statistic used as an estimator of the spectral density is the *periodogram*. The periodogram $I(\omega)$, based on a sample X_0, \dots, X_{T-1} is defined as

$$I(\omega_j) = \frac{1}{2\pi T} \left| \sum_{t=0}^{T-1} X_t e^{-it\omega_j} \right|^2, \quad (5)$$

where ω_j is the Fourier frequency $\omega_j = \frac{2\pi j}{T}$, $j = [-T/2]+1, \dots, -1, 0, 1, \dots, [T/2]$. By a discrete version of the sampling theorem it holds that $I(\omega)$ is uniquely determined for all $\omega \in [-\pi, \pi]$, given its values at Fourier frequencies.

Calculationally, the periodogram is found by using fast Fourier transform. A simple `matlab` m-function calculating the periodogram is

```
function out = periodogram(ts)
out = abs(fftshift(fft(ts - mean(ts)))).^2/(2*pi*length(ts));
```

An application of spectral and log-spectral estimation involves famous Wolf's sunspot data set. Although in this situation the statistician does not know the "true" signal, the theory developed by solar scientists helps to evaluate performance of the algorithm.

The Sun's activity peaks every 11 years, creating storms on the surface of our star that disrupt the Earth's magnetic field. These "solar hurricanes" can cause severe problems for electricity transmission systems. An example of influence of such periodic activity to everyday life is 1989 power blackout in the American northeast.

Efforts to monitor the amount and variation of the Sun's activity by counting spots on it have a long and rich history. Relatively complete visual estimates of daily activity date back to 1818, monthly averages can be extrapolated back to 1749, and estimates of annual values can be similarly determined back to 1700. Although Galileo made observations of sunspot numbers in the early 17th century, the modern era of sunspot counting began in the mid-1800s with the research of Bern Observatory director Rudolph Wolf, who introduced what he called the *Universal Sunspot Number* as an estimate of the solar activity. The square root of Wolf's yearly sunspot numbers are given in Fig. 4(a), data from Tong (1996), p.471. Because of wavelet data processing we selected a sample of size a power of two, i.e., only 256 observations from 1733 till 1998. The square root transformation was applied to symmetrize and de-trend the Wolf's counts. Panel (b) in Fig. 4 gives a raw periodogram, while the panel (c) in Fig. 4 shows the estimator of log-spectral density (Pensky and Vidakovic, 2003).

The estimator reveals a peak at frequency $\omega^* \approx 0.58$, corresponding to the Schwabe's cycle ranging from 9 to 11.5 (years), with an average of $\frac{2\pi}{0.58} \approx 10.8$ years. The Schwabe cycle is the period between two subsequent maxima or minima the solar activity, although the solar physicists often think in terms of a 22-year magnetic cycle since the sun's magnetic poles reverse direction every 11 years.

2.2 Windowed Fourier Transform

Windowed Fourier Transforms are important in providing simultaneous insight in time and frequency behavior of the functions. Standard Fourier Transforms describing the data in the "Fourier domain" are precise in frequency,

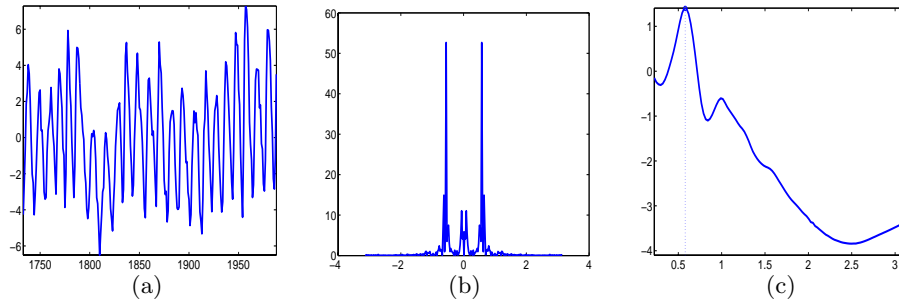


Fig. 4. (a) Square roots of Wolf's yearly sunspot numbers from 1732-1988 (256 observations); (b) Raw periodogram; (c) An estimator of the log-spectra. The frequency $\omega^* \approx 0.58$ corresponds to Schwabe's period of 10.8 (years).

but not in time. Small changes in the signal (data) at one location cause change in the Fourier domain globally. It was of interest to have transformed domains that are simultaneously precise in both time and frequency domains. Unfortunately, the precision of such an insight is limited by the Heisenberg's Uncertainty Principle.

Suppose $f(t)$ is a signal of finite energy. In mathematical terms, the integral of its modulus squared is finite, or shortly, f belongs to $\mathbb{L}_2(\mathbb{R})$ space.

The integral Fourier transform of the signal

$$\mathcal{F}(f)(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}} f(t)e^{-it\xi} dt, \quad (6)$$

describes the allocation of energy content of a signal at different frequencies, but the time-related information is lost.

Windowed Fourier transform (also called *short time Fourier transform*, STFT) was introduced by Gabor (1946), to measure time-localized frequencies of sound. An atom in Gabor's decomposition is defined via:

$$g_{u,\xi}(t) = e^{i\xi t} g(t - u),$$

where g is a real, symmetric, and properly normalized "window" function. [$\|g\| = 1$ so that $\|g_{u,\xi}\| = 1$]

If $f \in \mathbb{L}_2(\mathbb{R})$, then windowed Fourier transform is defined as

$$Sf(u, \xi) = \langle f, g_{u,\xi} \rangle = \int_{\mathbb{R}} f(t)g(t - u)e^{-i\xi t} dt. \quad (7)$$

The chief use of windowed Fourier transforms is to analyze time/frequency distribution of signal energy, via a *spectrogram*.

The spectrogram,

$$P_S f(u, \xi) = |Sf(u, \xi)|^2 = \left| \int_{-\infty}^{\infty} f(t)g(t - u)e^{-i\xi t} dt \right|^2,$$

expresses the energy distribution in the signal f , with respect to time and frequency simultaneously.

The following are some basic properties of STFT. Let $f \in \mathbb{L}_2(\mathbb{R}^2)$. Then

$$\text{[Inverse STFT]} \quad f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} Sf(u, \xi) g(t-u) e^{i\xi t} d\xi du, \quad (8)$$

and

$$\text{[Energy Conservation]} \quad \int_{\mathbb{R}} |f(t)|^2 dt = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} |Sf(u, \xi)|^2 d\xi du. \quad (9)$$

The following is a characterizing property of STFT:

Let $\Phi \in \mathbb{L}_2(\mathbb{R}^2)$. There exist $f \in \mathbb{L}_2(\mathbb{R}^2)$ such that $\Phi(u, \xi) = Sf(u, \xi)$ if and only if

$$\Phi(u_0, \xi_0) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \Phi(u, \xi) \mathbb{K}(u_0, u, \xi_0, \xi) du d\xi, \quad (10)$$

where

$$\mathbb{K}(u_0, u, \xi_0, \xi) = \langle g_{u, \xi}, g_{u_0, \xi_0} \rangle = \int_{\mathbb{R}} g(t-u) g(t-u_0) e^{-i(\xi_0 - \xi)t} dt. \quad (11)$$

2.3 Hilbert Transform

We next describe the Hilbert transform and its use in defining instantaneous frequency, an important measure in statistical analysis of signals.

The Hilbert transform of the function signal) $g(t)$ is defined by

$$H_g(t) = \frac{1}{\pi} (\text{VP}) \int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau} d\tau. \quad (12)$$

Because of the possible singularity at $\tau = t$, the integral is to be considered as a Cauchy principal value, (VP). From equation (12) we see that $H_g(t)$ is a convolution, $\frac{1}{\pi t} * g(t)$.

The spectrum of $H_g(t)$ is related to that of $g(t)$. From the convolution equation,

$$\mathcal{F}(H(t)) = \mathcal{F}\left(\frac{1}{\pi t}\right) \mathcal{F}(g(t)).$$

where \mathcal{F} is the Fourier transform. With a real signal $g(t)$ one can associate a complex function with the real part equal to $g(t)$ and the imaginary part equal to $H(g(t))$, $h(t) = g(t) - iH(g(t))$.

In statistical signal analysis this associated complex function $h(t)$ is known as analytic signal (or causal signal, since $\hat{h}(\xi) = 0$, for $\xi < 0$.) Analytic

signals are important since they possess unique phase $\phi(t)$ which leads to the definition of the instantaneous frequency.

If $h(t)$ is represented as $a(t) \cdot \exp\{i\phi(t)\}$, then the quantity $\frac{d\phi}{dt}$ is instantaneous frequency of the signal $g(t)$, at time t . For more discussion and use of instantaneous frequency, the reader is directed to Flandrin (1992, 1999).

2.4 Wigner-Ville Transforms

Wigner-Ville Transform (or Distribution) is the method to represent data (signals) in the time/frequency domain. In statistics, Wigner-Ville transform provide a tool to define localized spectral density for the nonstationary processes.

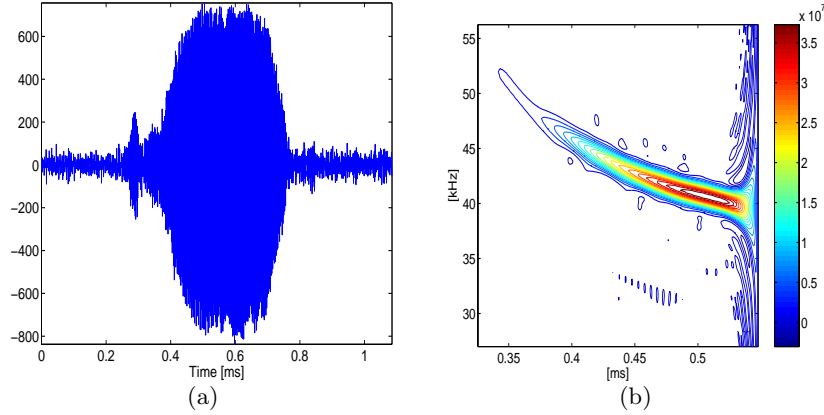


Fig. 5. (a) Sonar signal from flying bat; (b) Its Wigner-Ville transform.

Ville (1948) introduced the quadratic form that measures a local time-frequency energy:

$$P_V f(u, \xi) = \int_{\mathbb{R}} f(u + \frac{\tau}{2}) f^*(u - \frac{\tau}{2}) e^{-i\tau\xi} d\tau,$$

where f^* is conjugate of f .

The Wigner-Ville transform is always real since $f(u + \frac{\tau}{2}) f^*(u - \frac{\tau}{2})$ has Hermitian symmetry in τ .

Time and frequency are symmetric in $P_V f(u, \xi)$, by applying Parseval formula one gets,

$$P_V f(u, \xi) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(\xi + \frac{\gamma}{2}) \hat{f}^*(\xi - \frac{\gamma}{2}) e^{-i\gamma u} d\gamma, \quad (13)$$

For any $f \in \mathbb{L}_2(\mathbb{R})$

$$\int_{\mathbb{R}} P_V f(u, \xi) du = |\hat{f}(\xi)|^2, \quad (14)$$

i.e., the time marginalization reproduces power spectrum, and

$$\int_{\mathbb{R}} P_V f(u, \xi) d\xi = 2\pi |f(u)|^2, \quad (15)$$

i.e., the frequency marginalization is proportional to the squared modulus of the signal.

Integral (13) states that one-dimensional Fourier transform of $g_\xi(u) = P_V f(u, \xi)$, with respect to u is,

$$\hat{g}_\xi(\gamma) = \hat{f}\left(\xi + \frac{\gamma}{2}\right) \hat{f}^*\left(\xi - \frac{\gamma}{2}\right).$$

If $\gamma = 0$, $\hat{g}_\xi(0) = \int_{\mathbb{R}} g_\xi(u) du$, which proves (14). Similarly for (15).

For example,

(i) if $f(t) = \mathbf{1}(-T \leq t \leq T)$, then

$$P_V f(u, \xi) = \frac{2 \sin[2(T - |u|)\xi]}{\xi} \mathbf{1}(-T \leq u \leq T).$$

Plot $P_V f(u, \xi)$.

(ii) if $f(t) = \exp\{i\lambda(t + \alpha t^2/2)\}$, then $P_V(u, \xi) = 2\pi\delta(\xi - \lambda(1 + \alpha u))$.

(iii) a Gaussian $f(t) = (\sigma^2\pi)^{-1/4} \exp(-t^2/(2\sigma^2))$ is transformed into

$$P_V f(u, \xi) = \frac{1}{\pi} \exp\left(-\frac{u^2}{\sigma^2} - \sigma^2 \xi^2\right).$$

In this case, $P_V f(u, \xi) = |f(u)|^2 \cdot |\hat{f}(\xi)|^2$. The Gaussian is the only (up to time and frequency shifts) distribution for which Wigner-Ville transform remains positive. Some basic properties of Wigner-Ville transforms are listed in Table 1.

Table 1. Properties of Wigner-Ville transform

Function	Wigner-Ville
$f(t)$	$P_V f(u, \xi)$
$e^{i\phi} f(t)$	$P_V f(u, \xi)$
$f(t - u_0)$	$P_V f(u - u_0, \xi)$
$e^{i\xi_0 t} f(t)$	$P_V f(u, \xi - \xi_0)$
$e^{iat^2} f(t)$	$P_V f(u, \xi - 2au)$
$\frac{1}{\sqrt{s}} f\left(\frac{t}{s}\right)$	$P_V f\left(\frac{u}{s}, s\xi\right)$

Next we show that expected value of Wigner-Ville transform of a random process can serve as a definition for generalized spectrum of a non-stationary

process. Let $X(t)$ be real-valued zero-mean random process with covariance function

$$EX(t)X(s) = R(t, s) = R\left(u + \frac{\tau}{2}, u - \frac{\tau}{2}\right) = C(u, \tau),$$

after substitution $\tau = t - s$ and $u = \frac{t+s}{2}$.

Now, if the process $X(t)$ is stationary, then $C(u, \tau)$ is a function of τ only and

$$P_X(\xi) = \int_{-\infty}^{\infty} C(\tau)e^{-i\xi\tau} d\tau$$

is its power spectrum.

For arbitrary process Flandrin (1999) defined “power spectrum” as

$$P_X(\xi) = \int_{-\infty}^{\infty} C(u, \tau)e^{-i\xi\tau} d\tau.$$

Thus, $P_X(\xi)$ can be represented as $E P_V X(u, \xi)$, where

$$P_V X(u, \xi) = \int_{-\infty}^{\infty} X\left(u + \frac{\tau}{2}\right)X\left(u - \frac{\tau}{2}\right)e^{-i\xi\tau} d\tau.$$

For more information on Wigner-Ville transforms and their statistical use the reader is directed to Baraniuk (1994), Carmona et al. (1998), Flandrin (1999) and Mallat (1999), among others.

3 Wavelets and Other Multiscale Transforms

Given their recent popularity and clear evidence of wide applicability the most of the space in this Chapter is devoted to Wavelet transforms. Statistical multiscale modeling has, in recent decade, become a well established area in both theoretical and applied statistics, with impact to developments in statistical methodology.

Wavelet-based methods are important in statistics in areas such as regression, density and function estimation, factor analysis, modeling and forecasting in time series analysis, in assessing self-similarity and fractality in data, in spatial statistics.

The attention of the statistical community was attracted in late 1980’s and early 1990’s, when Donoho, Johnstone, and their coauthors demonstrated that wavelet thresholding, a simple denoising procedure, had desirable statistical optimality properties. Since then, wavelets have proved useful in many statistical disciplines, notably in nonparametric statistics and time series analysis. Bayesian concepts and modeling approaches have, more recently, been identified as providing promising contexts for wavelet-based denoising applications.

In addition to replacing traditional orthonormal bases in a variety statistical problems, wavelets brought novel techniques and invigorated some of the existing ones.

3.1 A Case Study

We start first with a statistical application of wavelet transforms. This example emphasizes specificity of wavelet-based denoising not shared by standard state-of-art denoising techniques.

A researcher in geology was interested in predicting earthquakes by the level of water in nearby wells. She had a large ($8192 = 2^{13}$ measurements) data set of water levels taken every hour in a period of time of about one year in a California well. Here is the description of the problem.

The ability of water wells to act as strain meters has been observed for centuries. The Chinese, for example, have records of water flowing from wells prior to earthquakes. Lab studies indicate that a seismic slip occurs along a fault prior to rupture. Recent work has attempted to quantify this response, in an effort to use water wells as sensitive indicators of volumetric strain. If this is possible, water wells could aid in earthquake prediction by sensing precursory earthquake strain.

We have water level records from six wells in southern California, collected over a six year time span. At least 13 moderate size earthquakes (magnitude 4.0 - 6.0) occurred in close proximity to the wells during this time interval. There is a significant amount of noise in the water level record which must first be filtered out. Environmental factors such as earth tides and atmospheric pressure create noise with frequencies ranging from seasonal to semidiurnal. The amount of rainfall also affects the water level, as do surface loading, pumping, recharge (such as an increase in water level due to irrigation), and sonic booms, to name a few. Once the noise is subtracted from the signal, the record can be analyzed for changes in water level, either an increase or a decrease depending upon whether the aquifer is experiencing a tensile or compressional volume strain, just prior to an earthquake.

A plot of the raw data for hourly measurements over one year ($8192 = 2^{13}$ observations) is given in Fig. 6a, with a close-up in panel b. After applying the wavelet transform and further processing the wavelet coefficients (thresholding), we obtained a fairly clean signal with a big jump at the earthquake time. The wavelet-denoised data are given in Fig. 6d. The magnitude of the water level change at the earthquake time did not get distorted in contrast to traditional smoothing techniques. This local adaptivity is a desirable feature of wavelet methods.

For example, Fig. 6c, is denoised signal after applying `supsmo` smoothing procedure. Note that the earthquake jump is smoothed, as well.

3.2 Continuous Wavelet Transform

The first theoretical results in wavelets had been concerned with continuous wavelet decompositions of functions and go back to the early 1980s. Papers of Morlet et al. (1982) and Grossmann and Morlet (1984, 1985) were among the first on this subject.

Let $\psi_{a,b}(x)$, $a \in \mathbb{R} \setminus \{0\}, b \in \mathbb{R}$ be a family of functions defined as translations and re-scales of a single function $\psi(x) \in \mathbb{L}_2(\mathbb{R})$,

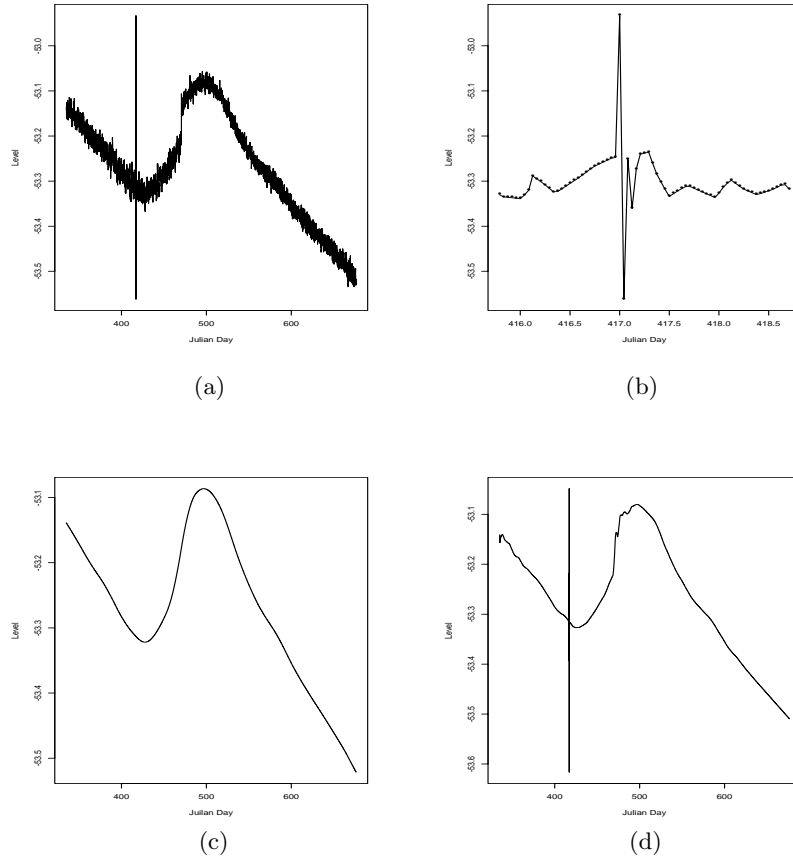


Fig. 6. Panel (a) shows $n = 8192$ hourly measurements of the water level for a well in an earthquake zone. Notice the wide range of water levels at the time of an earthquake around $t = 417$. Panel (b) focusses on the data around the earthquake time. Panel (c) demonstrates action of a standard smoother `supsmo`, and (d) gives a wavelet based reconstruction.

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right). \quad (16)$$

Normalization constant $\frac{1}{\sqrt{|a|}}$ ensures that the norm $\|\psi_{a,b}(x)\|$ is independent of a and b . The function ψ (called *the wavelet function*) is assumed to satisfy the *admissibility condition*,

$$C_\psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \quad (17)$$

where $\Psi(\omega) = \int_{\mathbb{R}} \psi(x)e^{-ix\omega} dx$ is the Fourier transform of $\psi(x)$. The admissibility condition (17) implies

$$0 = \Psi(0) = \int \psi(x)dx.$$

Also, if $\int \psi(x)dx = 0$ and $\int (1 + |x|^\alpha)|\psi(x)|dx < \infty$ for some $\alpha > 0$, then $C_\psi < \infty$.

Wavelet functions are usually normalized to “have unit energy”, i.e., $\|\psi_{a,b}(x)\| = 1$.

For example, the second derivative of the Gaussian function,

$$\psi(x) = \frac{d^2}{dx^2}[-Ce^{-x^2/2}] = C(1 - x^2)e^{-x^2/2}, \quad C = \frac{2}{\sqrt{3}\sqrt{\pi}},$$

is an example of an admissible wavelet, called Mexican Hat or Marr’s wavelet, see Fig. 7.

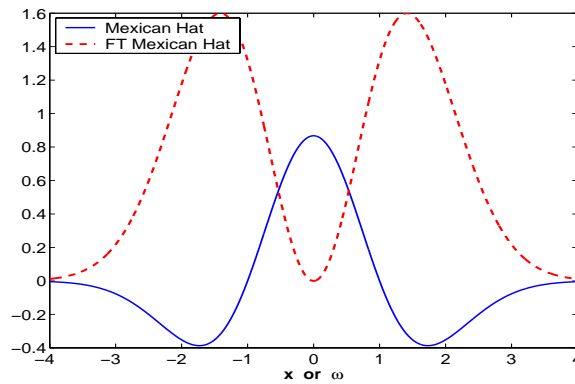


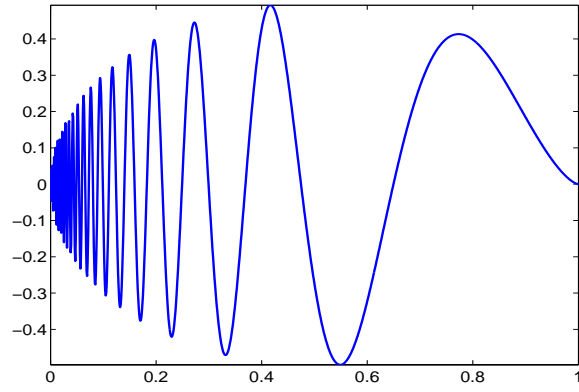
Fig. 7. Mexican hat wavelet (solid line) and its Fourier transform (dashed line)

For any square integrable function $f(x)$, the continuous wavelet transform is defined as a function of two variables

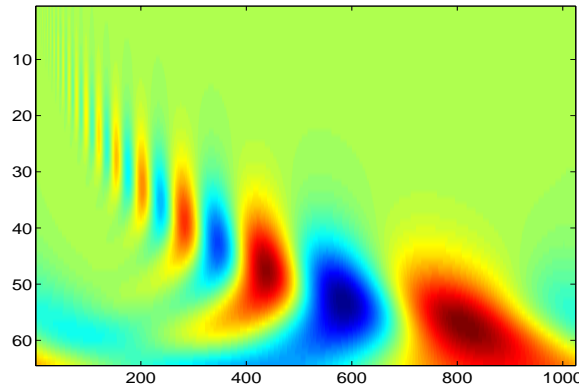
$$CWT_f(a, b) = \langle f, \psi_{a,b} \rangle = \int f(x)\overline{\psi_{a,b}(x)}dx.$$

Here the dilation and translation parameters, a and b , respectively, vary continuously over $\mathbb{R} \setminus \{0\} \times \mathbb{R}$.

Fig. 8 gives the **doppler** test function, $f = \frac{1}{t+0.05}\sqrt{t(1-t)}\sin(2\pi \cdot 1.05t)$, $0 \leq t \leq 1$, and its continuous wavelet transform. The wavelet used was Mexican Hat. Notice the distribution of “energy” in the time/frequency plane in panel (b) of Fig. 8.



(a)



(b)

Fig. 8. (a) Doppler signal; (b) Continuous wavelet transform of doppler signal by the Mexican hat wavelet.

Resolution of Identity. When the admissibility condition is satisfied, i.e., $C_\psi < \infty$, it is possible to find the inverse continuous transform via the relation known as *resolution of identity* or *Calderón's reproducing identity*,

$$f(x) = \frac{1}{C_\psi} \int_{\mathbb{R}^2} \mathcal{CWT}_f(a, b) \psi_{a,b}(x) \frac{da db}{a^2}.$$

The continuous wavelet transform of a function of one variable is a function of two variables. Clearly, the transform is redundant. To “minimize” the transform one can select discrete values of a and b and still have a lossless transform. This is achieved by so called *critical sampling*.

The critical sampling defined by

$$a = 2^{-j}, \quad b = k2^{-j}, \quad j, k \in \mathbb{Z}, \quad (18)$$

will produce the minimal, but complete basis. Any coarser sampling will not produce a unique inverse transform. Moreover under mild conditions on the wavelet function ψ , such sampling produces an orthogonal basis $\{\psi_{jk}(x) = 2^{j/2}\psi(2^jx - k), j, k \in \mathbb{Z}\}$. To formally describe properties of minimal and orthogonal wavelet bases a multiresolution formalism is needed.

3.3 Multiresolution Analysis

Fundamental for construction of critically sampled orthogonal wavelets is a notion of multiresolution analysis introduced by Mallat (1989a,b) A multiresolution analysis (MRA) is a sequence of closed subspaces $V_n, n \in \mathbb{Z}$ in $\mathbb{L}_2(\mathbb{R})$ such that they lie in a containment hierarchy

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots . \quad (19)$$

The nested spaces have an intersection that contains only the zero function and a union that contains all square integrable functions.

$$\cap_n V_j = \{\mathbf{0}\}, \quad \overline{\cup_j V_j} = \mathbb{L}_2(\mathbb{R}).$$

[With \overline{A} we denoted the closure of a set A]. The hierarchy (19) is constructed such that V -spaces are self-similar,

$$f(2^jx) \in V_j \text{ iff } f(x) \in V_0. \quad (20)$$

with the requirement that there exists a *scaling function* $\phi \in V_0$ whose integer-translates span the space V_0 ,

$$V_0 = \left\{ f \in \mathbb{L}_2(\mathbb{R}) \mid f(x) = \sum_k c_k \phi(x - k) \right\},$$

and for which the family $\{\phi(\bullet - k), k \in \mathbb{Z}\}$ is an orthonormal basis. It can be assumed that $\int \phi(x)dx \geq 0$. With this assumption this integral is in fact equal to 1. Because of containment $V_0 \subset V_1$, the function $\phi(x) \in V_0$ can be represented as a linear combination of functions from V_1 , i.e.,

$$\phi(x) = \sum_{k \in \mathbb{Z}} h_k \sqrt{2}\phi(2x - k), \quad (21)$$

for some coefficients $h_k, k \in \mathbb{Z}$. This equation called the *scaling equation* (or two-scale equation) is fundamental in constructing, exploring, and utilizing wavelets.

Theorem 2. *For the scaling function it holds*

$$\int_{\mathbb{R}} \phi(x) dx = 1,$$

or, equivalently,

$$\Phi(0) = 1,$$

where $\Phi(\omega)$ is Fourier transform of ϕ , $\int_{\mathbb{R}} \phi(x) e^{-i\omega x} dx$.

The coefficients h_n in (21) are important in efficient application of wavelet transforms. The (possibly infinite) vector $\mathbf{h} = \{h_n, n \in \mathbb{Z}\}$ will be called a *wavelet filter*. It is a low-pass (averaging) filter as will become clear later by its analysis in the Fourier domain.

To further explore properties of multiresolution analysis subspaces and their bases, we will often work in the Fourier domain.

It will be convenient to use Fourier domain for subsequent analysis of wavelet paradigm. Define the function m_0 as follows:

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} h_k e^{-ik\omega} = \frac{1}{\sqrt{2}} H(\omega). \quad (22)$$

The function in (22) is sometimes called the *transfer function* and it describes the behavior of the associated filter \mathbf{h} in the Fourier domain. Notice that the function m_0 is 2π -periodic and that filter taps $\{h_n, n \in \mathbb{Z}\}$ are in fact the Fourier coefficients in the Fourier series of $H(\omega) = \sqrt{2} m_0(\omega)$.

In the Fourier domain, the relation (21) becomes

$$\Phi(\omega) = m_0\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right), \quad (23)$$

where $\Phi(\omega)$ is the Fourier transform of $\phi(x)$. Indeed,

$$\begin{aligned} \Phi(\omega) &= \int_{-\infty}^{\infty} \phi(x) e^{-i\omega x} dx \\ &= \sum_k \sqrt{2} h_k \int_{-\infty}^{\infty} \phi(2x - k) e^{-i\omega x} dx \\ &= \sum_k \frac{h_k}{\sqrt{2}} e^{-ik\omega/2} \int_{-\infty}^{\infty} \phi(2x - k) e^{-i(2x-k)\omega/2} d(2x - k) \\ &= \sum_k \frac{h_k}{\sqrt{2}} e^{-ik\omega/2} \Phi\left(\frac{\omega}{2}\right) \\ &= m_0\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right). \end{aligned}$$

By iterating (23), one gets

$$\Phi(\omega) = \prod_{n=1}^{\infty} m_0\left(\frac{\omega}{2^n}\right), \quad (24)$$

which is convergent under very mild conditions concerning the rates of decay of the scaling function ϕ .

Next, we prove two important properties of wavelet filters associated with an orthogonal multiresolution analysis, *normalization* and *orthogonality*.

Normalization.

$$\sum_{k \in \mathbb{Z}} h_k = \sqrt{2}. \quad (25)$$

Proof:

$$\begin{aligned} \int \phi(x) dx &= \sqrt{2} \sum_k h_k \int \phi(2x - k) dx \\ &= \sqrt{2} \sum_k h_k \frac{1}{2} \int \phi(2x - k) d(2x - k) \\ &= \frac{\sqrt{2}}{2} \sum_k h_k \int \phi(x) dx. \end{aligned}$$

Since $\int \phi(x) dx \neq 0$ by assumption, (25) follows.

This result also follows from $m_0(0) = 1$.

Orthogonality. For any $l \in \mathbb{Z}$,

$$\sum_k h_k h_{k-2l} = \delta_l. \quad (26)$$

Proof: Notice first that from the scaling equation (21) it follows that

$$\begin{aligned} \phi(x)\phi(x-l) &= \sqrt{2} \sum_k h_k \phi(2x-k)\phi(x-l) \\ &= \sqrt{2} \sum_k h_k \phi(2x-k) \sqrt{2} \sum_m h_m \phi(2(x-l)-m). \end{aligned} \quad (27)$$

By integrating the both sides in (27) we obtain

$$\begin{aligned}
\delta_l &= 2 \sum_k h_k \left[\sum_m h_m \frac{1}{2} \int \phi(2x - k) \phi(2x - 2l - m) d(2x) \right] \\
&= \sum_k \sum_m h_k h_m \delta_{k, 2l+m} \\
&= \sum_k h_k h_{k-2l}.
\end{aligned}$$

The last line is obtained by taking $k = 2l + m$.

An important special case is $l = 0$ for which (26) becomes

$$\sum_k h_k^2 = 1. \quad (28)$$

The fact that the system $\{\phi(\bullet - k), k \in \mathbb{Z}\}$ constitutes an orthonormal basis for V_0 can be expressed in the Fourier domain in terms of either $\Phi(\omega)$ or $m_0(\omega)$.

In terms of $\Phi(\omega)$,

$$\sum_{l=-\infty}^{\infty} |\Phi(\omega + 2\pi l)|^2 = 1. \quad (29)$$

From the Plancherel identity and the 2π -periodicity of $e^{i\omega k}$ it follows

$$\begin{aligned}
\delta_k &= \int_{\mathbb{R}} \phi(x) \overline{\phi(x - k)} dx \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \Phi(\omega) \overline{\Phi(\omega)} e^{i\omega k} d\omega \\
&= \frac{1}{2\pi} \int_0^{2\pi} \sum_{l=-\infty}^{\infty} |\Phi(\omega + 2\pi l)|^2 e^{i\omega k} d\omega.
\end{aligned} \quad (30)$$

The last line in (30) is the Fourier coefficient a_k in the Fourier series decomposition of

$$f(\omega) = \sum_{l=-\infty}^{\infty} |\Phi(\omega + 2\pi l)|^2.$$

Due to the uniqueness of Fourier representation, $f(\omega) = 1$. As a side result, we obtain that $\Phi(2\pi n) = 0, n \neq 0$, and $\sum_n \phi(x - n) = 1$. The last result follows from inspection of coefficients c_k in the Fourier decomposition of $\sum_n \phi(x - n)$, the series $\sum_k c_k e^{2\pi i k x}$. As this function is 1-periodic,

$$c_k = \int_0^1 \left(\sum_n \phi(x - n) \right) e^{-2\pi i k x} dx = \int_{-\infty}^{\infty} \phi(x) e^{-2\pi i k x} dx = \Phi(2\pi k) = \delta_{0,k}.$$

Remark 1. Utilizing the identity (29), any set of independent functions spanning V_0 , $\{\phi(x - k), k \in \mathbb{Z}\}$, can be orthogonalized in the Fourier domain. The orthonormal basis is generated by integer-shifts of the function

$$\mathcal{F}^{-1} \left[\frac{\Phi(\omega)}{\sqrt{\sum_{l=-\infty}^{\infty} |\Phi(\omega + 2\pi l)|^2}} \right]. \quad (31)$$

This normalization in the Fourier domain is used in constructing of some wavelet bases.

(b) In terms of m_0 :

$$|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 = 1. \quad (32)$$

Since $\sum_{l=-\infty}^{\infty} |\Phi(2\omega + 2l\pi)|^2 = 1$, then by (23)

$$\sum_{l=-\infty}^{\infty} |m_0(\omega + l\pi)|^2 |\Phi(\omega + l\pi)|^2 = 1. \quad (33)$$

Now split the sum in (33) into two sums – one with odd and the other with even indices, i.e.,

$$\begin{aligned} 1 &= \sum_{k=-\infty}^{\infty} |m_0(\omega + 2k\pi)|^2 |\Phi(\omega + 2k\pi)|^2 + \\ &\quad \sum_{k=-\infty}^{\infty} |m_0(\omega + (2k+1)\pi)|^2 |\Phi(\omega + (2k+1)\pi)|^2. \end{aligned}$$

To simplify the above expression, we use relation (29) and the 2π -periodicity of $m_0(\omega)$.

$$\begin{aligned} 1 &= |m_0(\omega)|^2 \sum_{k=-\infty}^{\infty} |\Phi(\omega + 2k\pi)|^2 + |m_0(\omega + \pi)|^2 \sum_{k=-\infty}^{\infty} |\Phi((\omega + \pi) + 2k\pi)|^2 \\ &= |m_0(\omega)|^2 + |m_0(\omega + \pi)|^2. \end{aligned}$$

Whenever a sequence of subspaces satisfies MRA properties, there exists (though not unique) an orthonormal basis for $\mathbb{L}_2(\mathbb{R})$,

$$\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), j, k \in \mathbb{Z}\} \quad (34)$$

such that $\{\psi_{jk}(x), j\text{-fixed}, k \in \mathbb{Z}\}$ is an orthonormal basis of the “difference space” $W_j = V_{j+1} \ominus V_j$. The function $\psi(x) = \psi_{00}(x)$ is called a *wavelet function* or informally *the mother wavelet*.

Next, we discuss the derivation of a wavelet function from the scaling function. Since $\psi(x) \in V_1$ (because of the containment $W_0 \subset V_1$), it can be represented as

$$\psi(x) = \sum_{k \in \mathbb{Z}} g_k \sqrt{2}\phi(2x - k), \quad (35)$$

for some coefficients $g_k, k \in \mathbb{Z}$.

Define

$$m_1(\omega) = \frac{1}{\sqrt{2}} \sum_k g_k e^{-ik\omega}. \quad (36)$$

By mimicking what was done with m_0 , we obtain the Fourier counterpart of (35),

$$\Psi(\omega) = m_1\left(\frac{\omega}{2}\right)\Phi\left(\frac{\omega}{2}\right). \quad (37)$$

The spaces W_0 and V_0 are orthogonal by construction. Therefore,

$$\begin{aligned} 0 &= \int \psi(x)\phi(x - k)dx = \frac{1}{2\pi} \int \Psi(\omega)\overline{\Phi(\omega)}e^{i\omega k} d\omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} \sum_{l=-\infty}^{\infty} \Psi(\omega + 2l\pi)\overline{\Phi(\omega + 2l\pi)}e^{i\omega k} d\omega. \end{aligned}$$

By repeating the Fourier series argument, as in (29), we conclude

$$\sum_{l=-\infty}^{\infty} \Psi(\omega + 2l\pi)\overline{\Phi(\omega + 2l\pi)} = 0.$$

By taking into account the definitions of m_0 and m_1 , and by the derivation as in (32), we find

$$m_1(\omega)\overline{m_0(\omega)} + m_1(\omega + \pi)\overline{m_0(\omega + \pi)} = 0. \quad (38)$$

From (38), we conclude that there exists a function $\lambda(\omega)$ such that

$$(m_1(\omega), m_1(\omega + \pi)) = \lambda(\omega) \left(\overline{m_0(\omega + \pi)}, -\overline{m_0(\omega)} \right). \quad (39)$$

By substituting $\xi = \omega + \pi$ and by using the 2π -periodicity of m_0 and m_1 , we conclude that

$$\begin{aligned} \lambda(\omega) &= -\lambda(\omega + \pi), \text{ and} \\ \lambda(\omega) &\text{ is } 2\pi\text{-periodic.} \end{aligned} \quad (40)$$

Any function $\lambda(\omega)$ of the form $e^{\pm i\omega} S(2\omega)$, where S is an $\mathbb{L}_2([0, 2\pi])$, 2π -periodic function, will satisfy (38); however, only the functions for which $|\lambda(\omega)| = 1$ will define an orthogonal basis ψ_{jk} of $\mathbb{L}_2(\mathbb{R})$.

To summarize, we choose $\lambda(\omega)$ such that

- (i) $\lambda(\omega)$ is 2π -periodic,
- (ii) $\lambda(\omega) = -\lambda(\omega + \pi)$, and
- (iii) $|\lambda(\omega)|^2 = 1$.

Standard choices for $\lambda(\omega)$ are $-e^{-i\omega}$, $e^{-i\omega}$, and $e^{i\omega}$; however, any other function satisfying (i)-(iii) will generate a valid m_1 . We choose to define $m_1(\omega)$ as

$$m_1(\omega) = -e^{-i\omega} \overline{m_0(\omega + \pi)}. \quad (41)$$

since it leads to a convenient and standard connection between the filters \mathbf{h} and \mathbf{g} .

The form of m_1 and the equation (29) imply that $\{\psi(\bullet - k), k \in \mathbb{Z}\}$ is an orthonormal basis for W_0 .

Since $|m_1(\omega)| = |m_0(\omega + \pi)|$, the orthogonality condition (32) can be rewritten as

$$|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1. \quad (42)$$

By comparing the definition of m_1 in (36) with

$$\begin{aligned} m_1(\omega) &= -e^{-i\omega} \frac{1}{\sqrt{2}} \sum_k h_k e^{i(\omega + \pi)k} \\ &= \frac{1}{\sqrt{2}} \sum_k (-1)^{1-k} h_k e^{-i\omega(1-k)} \\ &= \frac{1}{\sqrt{2}} \sum_n (-1)^n h_{1-n} e^{-i\omega n}, \end{aligned}$$

we relate g_n and h_n as

$$g_n = (-1)^n h_{1-n}. \quad (43)$$

In signal processing literature, the relation (43) is known as the *quadrature mirror relation* and the filters \mathbf{h} and \mathbf{g} as *quadrature mirror filters*.

Remark 2. Choosing $\lambda(\omega) = e^{i\omega}$ leads to the rarely used high-pass filter $g_n = (-1)^{n-1} h_{-1-n}$. It is convenient to define g_n as $(-1)^n h_{1-n+M}$, where M is a “shift constant.” Such re-indexing of \mathbf{g} affects only the shift-location of the wavelet function.

3.4 Haar Wavelets

In addition to their simplicity and formidable applicability, Haar wavelets have tremendous educational value. Here we illustrate some of the relations discussed in the Section 3.3 using the Haar wavelet. We start with scaling function $\phi(x) = \mathbf{1}(0 \leq x \leq 1)$ and pretend that everything else is unknown. By inspection of simple graphs of two scaled Haar wavelets $\phi(2x)$ and $\phi(2x+1)$ stuck to each other, we conclude that the scaling equation (21) is

$$\begin{aligned} \phi(x) &= \phi(2x) + \phi(2x-1) \\ &= \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x-1), \end{aligned} \quad (44)$$

which yields the wavelet filter coefficients:

$$h_0 = h_1 = \frac{1}{\sqrt{2}}.$$

The transfer functions are

$$m_0(\omega) = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} e^{-i\omega 0} \right) + \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} e^{-i\omega 1} \right) = \frac{1 + e^{-i\omega}}{2}.$$

and

$$m_1(\omega) = -e^{-i\omega} \overline{m_0(\omega + \pi)} = -e^{-i\omega} \left(\frac{1}{2} - \frac{1}{2} e^{i\omega} \right) = \frac{1 - e^{-i\omega}}{2}.$$

Notice that $m_0(\omega) = |m_0(\omega)| e^{i\varphi(\omega)} = \cos \frac{\omega}{2} \cdot e^{-i\omega/2}$ (after $\cos x = \frac{e^{ix} + e^{-ix}}{2}$). Since $\varphi(\omega) = -\frac{\omega}{2}$, the Haar wavelet has *linear phase*, i.e., the scaling function

is symmetric in the time domain. The orthogonality condition $|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1$ is easily verified, as well.

Relation (37) becomes

$$\Psi(\omega) = \frac{1 - e^{-i\omega/2}}{2} \Phi\left(\frac{\omega}{2}\right) = \frac{1}{2} \Phi\left(\frac{\omega}{2}\right) - \frac{1}{2} \Phi\left(\frac{\omega}{2}\right) e^{-i\omega/2},$$

and by applying the inverse Fourier transform we obtain

$$\psi(x) = \phi(2x) - \phi(2x - 1)$$

in the time-domain. Therefore we “have found” the Haar wavelet function ψ . From the expression for m_1 or by inspecting the representation of $\psi(x)$ by $\phi(2x)$ and $\phi(2x - 1)$, we “conclude” that $g_0 = -g_{-1} = \frac{1}{\sqrt{2}}$.

Although the Haar wavelets are well localized in the time domain, in the frequency domain they decay at the slow rate of $O(\frac{1}{n})$ and are not effective in approximating smooth functions.

3.5 Daubechies' Wavelets

The most important family of wavelets was discovered by Ingrid Daubechies and fully described in Daubechies (1992). This family is compactly supported with various degrees of smoothness.

The formal derivation of Daubechies' wavelets goes beyond the scope of this chapter, but the filter coefficients of some of its family members can be found by following considerations.

For example, to derive the filter taps of a wavelet with N vanishing moments, or equivalently, $2N$ filter taps, we use the following equations.

The normalization property of scaling function implies

$$\sum_{i=0}^{2N-1} h_i = \sqrt{2},$$

requirement for vanishing moments for wavelet function ψ leads to

$$\sum_{i=0}^{2N-1} (-1)^{i_j k} h_i = 0, \quad k = 0, 1, \dots, N - 1,$$

and, finally, the orthogonality property can be expressed as

$$\sum_{i=0}^{2N-1} h_i h_{i+2k} = \delta_k \quad k = 0, 1, \dots, N - 1.$$

We obtained $2N + 1$ equations with $2N$ unknowns; however the system is solvable since the equations are not linearly independent.

Example 4. For $N = 2$, we obtain the system:

$$\begin{cases} h_0 + h_1 + h_2 + h_3 = \sqrt{2} \\ h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1 \\ -h_1 + 2h_2 - 3h_3 = 0 \\ h_0 h_2 + h_1 h_3 = 0 \end{cases},$$

which has a solution $h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}$, $h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}$, $h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}$, and $h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$.

For $N = 4$, the system is

$$\begin{cases} h_0 + h_1 + h_2 + h_3 + h_4 + h_5 + h_6 + h_7 = \sqrt{2} \\ h_0^2 + h_1^2 + h_2^2 + h_3^2 + h_4^2 + h_5^2 + h_6^2 + h_7^2 = 1 \\ h_0 - h_1 + h_2 - h_3 + h_4 - h_5 + h_6 - h_7 = 0 \\ h_0 h_2 + h_1 h_3 + h_2 h_4 + h_3 h_5 + h_4 h_6 + h_5 h_7 = 0 \\ h_0 h_4 + h_1 h_5 + h_2 h_6 + h_3 h_7 = 0 \\ h_0 h_6 + h_1 h_7 = 0 \\ 0h_0 - 1h_1 + 2h_2 - 3h_3 + 4h_4 - 5h_5 + 6h_6 - 7h_7 = 0 \\ 0h_0 - 1h_1 + 4h_2 - 9h_3 + 16h_4 - 25h_5 + 36h_6 - 49h_7 = 0 \\ 0h_0 - 1h_1 + 8h_2 - 27h_3 + 64h_4 - 125h_5 + 216h_6 - 343h_7 = 0. \end{cases}$$

Fig. 9 depicts two scaling function and wavelet pairs from the Daubechies family. Panels (a) and (b) depict the pair with two vanishing moments, while panels (c) and (d) depict the pair with four vanishing moments.

4 Discrete Wavelet Transforms

Discrete wavelet transforms (DWT) are applied to discrete data sets and produce discrete outputs. Transforming signals and data vectors by DWT is a process that resembles the fast Fourier transform (FFT), the Fourier method applied to a set of discrete measurements.

Table 2. The analogy between Fourier and wavelet methods

Fourier Methods	Fourier Integrals	Fourier Series	Discrete Fourier Transforms
Wavelet Methods	Continuous Wavelet Transforms	Wavelet Series	Discrete Wavelet Transforms

The analogy between Fourier and wavelet methods is even more complete (Table 2) when we take into account the continuous wavelet transform and wavelet series expansions.

Discrete wavelet transforms map data from the time domain (the original or input data vector) to the wavelet domain. The result is a vector of the same size. Wavelet transforms are linear and they can be defined by matrices

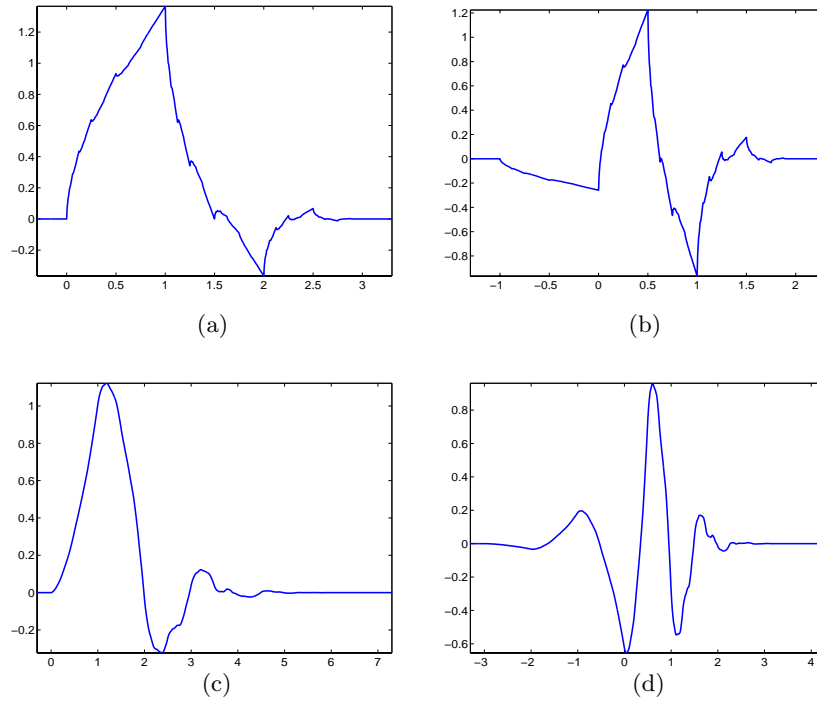


Fig. 9. Wavelet functions from Daubechies family. (a) Daubechies scaling function, 2 vanishing moments, 4 tap filter (b) Wavelet function corresponding to (a),(c) Daubechies scaling function, 4 vanishing moments, 8 tap filter (d) Wavelet function corresponding to (c)

of dimension $n \times n$ if they are applied to inputs of size n . Depending on boundary conditions, such matrices can be either orthogonal or “close” to orthogonal. When the matrix is orthogonal, the corresponding transform is a rotation in \mathbb{R}^n in which the data (a n -tuple) is a point in \mathbb{R}^n . The coordinates of the point in the rotated space comprise the discrete wavelet transform of the original coordinates. Here we provide two toy examples.

Example 5. Let the vector be $(1, 2)$ and let $M(1, 2)$ be the point in \mathbb{R}^2 with coordinates given by the data vector. The rotation of the coordinate axes by an angle of $\frac{\pi}{4}$ can be interpreted as a DWT in the Haar wavelet basis. The rotation matrix is

$$W = \begin{pmatrix} \cos \frac{\pi}{4} & \sin \frac{\pi}{4} \\ \cos \frac{\pi}{4} & -\sin \frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix},$$

and the discrete wavelet transform of $(1, 2)'$ is $W \cdot (1, 2)' = (\frac{3}{\sqrt{2}}, -\frac{1}{\sqrt{2}})'$. Notice that *the energy* (squared distance of the point from the origin) is preserved, $1^2 + 2^2 = (\frac{1}{2})^2 + (\frac{\sqrt{3}}{2})^2$, since W is a rotation.

Example 6. Let $\mathbf{y} = (1, 0, -3, 2, 1, 0, 1, 2)$. The associated function f is given in Fig. 10. The values $f(n) = y_n$, $n = 0, 1, \dots, 7$ are interpolated by a piecewise constant function. We assume that f belongs to Haar's multiresolution space V_0 .

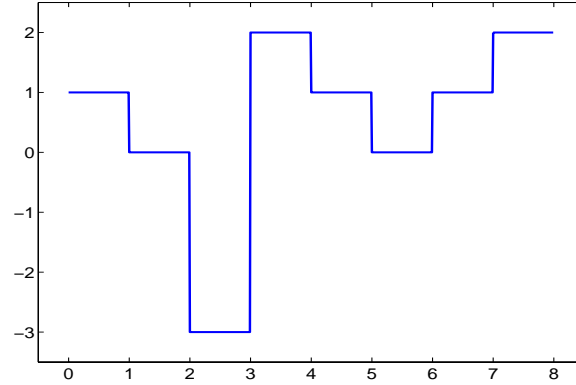


Fig. 10. A function interpolating \mathbf{y} on $[0, 8)$.

The following matrix equation gives the connection between \mathbf{y} and the wavelet coefficients (data in the wavelet domain).

$$\begin{bmatrix} 1 \\ 0 \\ -3 \\ 2 \\ 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} c_{00} \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{bmatrix}.$$

The solution is

$$\begin{bmatrix} c_{00} \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ -\sqrt{2} \\ 1 \\ -1 \\ \frac{1}{\sqrt{2}} \\ -\frac{5}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Thus,

$$\begin{aligned} f &= \sqrt{2}\phi_{-3,0} - \sqrt{2}\psi_{-3,0} + \psi_{-2,0} - \psi_{-2,1} \\ &\quad + \frac{1}{\sqrt{2}}\psi_{-1,0} - \frac{5}{\sqrt{2}}\psi_{-1,1} + \frac{1}{\sqrt{2}}\psi_{-1,2} - \frac{1}{\sqrt{2}}\psi_{-1,3}. \end{aligned} \tag{45}$$

The solution is easy to verify. For example, when $x \in [0, 1)$,

$$f(x) = \sqrt{2} \cdot \frac{1}{2\sqrt{2}} - \sqrt{2} \cdot \frac{1}{2\sqrt{2}} + 1 \cdot \frac{1}{2} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = 1/2 + 1/2 = 1 (= y_0).$$

Applying wavelet transforms by multiplying the input vector with an appropriate orthogonal matrix is conceptually straightforward task, but of limited practical value. Storing and manipulating the transformation matrices for long inputs ($n > 2000$) may not even be feasible.

This obstacle is solved by the link of discrete wavelet transforms with fast filtering algorithms from the field of signal and image processing.

4.1 The Cascade Algorithm

Mallat (1989a,b) was the first to link wavelets, multiresolution analyses and cascade algorithms in a formal way. Mallat’s cascade algorithm gives a constructive and efficient recipe for performing the discrete wavelet transform. It relates the wavelet coefficients from different levels in the transform by filtering with wavelet filter \mathbf{h} and and its mirror counterpart \mathbf{g} .

It is convenient to link the original data with the space V_J , where J is often 0 or $\log n$, where n is a dyadic size of data. Then, coarser smooth and complementing detail spaces are $(V_{J-1}, W_{J-1}), (V_{J-2}, W_{J-2}),$ etc. Decreasing the index in V -spaces is equivalent to coarsening the approximation to the data.

By a straightforward substitution of indices in the scaling equations (21) and (35), one obtains

$$\phi_{j-1,l}(x) = \sum_{k \in \mathbb{Z}} h_{k-2l}\phi_{jk}(x) \quad \text{and} \quad \psi_{j-1,l}(x) = \sum_{k \in \mathbb{Z}} g_{k-2l}\phi_{jk}(x). \tag{46}$$

The relations in (46) are fundamental in developing the cascade algorithm.

In a multiresolution analysis, $\dots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \dots$. Since $V_j = V_{j-1} \oplus W_{j-1}$, any function $v_j \in V_j$ can be represented uniquely as $v_j(x) = v_{j-1}(x) + w_{j-1}(x)$, where $v_{j-1} \in V_{j-1}$ and $w_{j-1} \in W_{j-1}$. It is customary to denote the coefficients associated with $\phi_{jk}(x)$ and $\psi_{jk}(x)$ by c_{jk} and d_{jk} , respectively.

Thus,

$$\begin{aligned} v_j(x) &= \sum_k c_{j,k} \phi_{j,k}(x) \\ &= \sum_l c_{j-1,l} \phi_{j-1,l}(x) + \sum_l d_{j-1,l} \psi_{j-1,l}(x) \\ &= v_{j-1}(x) + w_{j-1}(x). \end{aligned}$$

By using the general scaling equations (46), orthogonality of $w_{j-1}(x)$ and $\phi_{j-1,l}(x)$ for any j and l , and additivity of inner products, we obtain

$$\begin{aligned} c_{j-1,l} &= \langle v_j, \phi_{j-1,l} \rangle \\ &= \langle v_j, \sum_k h_{k-2l} \phi_{j,k} \rangle \\ &= \sum_k h_{k-2l} \langle v_j, \phi_{j,k} \rangle \\ &= \sum_k h_{k-2l} c_{j,k}. \end{aligned} \tag{47}$$

Similarly $d_{j-1,l} = \sum_k g_{k-2l} c_{j,k}$.

The cascade algorithm works in the reverse direction as well. Coefficients in the next finer scale corresponding to V_j can be obtained from the coefficients corresponding to V_{j-1} and W_{j-1} . The relation

$$\begin{aligned} c_{j,k} &= \langle v_j, \phi_{j,k} \rangle \\ &= \sum_l c_{j-1,l} \langle \phi_{j-1,l}, \phi_{j,k} \rangle + \sum_l d_{j-1,l} \langle \psi_{j-1,l}, \phi_{j,k} \rangle \\ &= \sum_l c_{j-1,l} h_{k-2l} + \sum_l d_{j-1,l} g_{k-2l}, \end{aligned} \tag{48}$$

describes a single step in the reconstruction algorithm.

The discrete wavelet transform can be described in terms of operators. Let the operators \mathcal{H} and \mathcal{G} acting on a sequence $a = \{a_n, n \in \mathbb{Z}\}$, satisfy the following coordinate-wise relations:

$$(\mathcal{H}a)_k = \sum_n h_{n-2k} a_n \quad (\mathcal{G}a)_k = \sum_n g_{n-2k} a_n,$$

and their adjoint operators \mathcal{H}^* and \mathcal{G}^* satisfy:

$$(\mathcal{H}^* a)_n = \sum_k h_{n-2k} a_k \quad (\mathcal{G}^* a)_n = \sum_k g_{n-2k} a_k,$$

where $\mathbf{h} = \{h_n\}$ is wavelet filter and $\mathbf{g} = \{g_n\}$ its quadrature-mirror counterpart.

Denote the original signal by $\mathbf{c}^{(J)} = \{c_k^{(J)}\}$. If the signal is of length 2^J , then $\mathbf{c}^{(J)}$ can be interpolated by the function $f(x) = \sum c_k^{(J)} \phi(x - k)$ from V_J . In each step of the wavelet transform, we move to the next coarser approximation (level) $\mathbf{c}^{(j-1)}$ by applying the operator \mathcal{H} , $\mathbf{c}^{(j-1)} = \mathcal{H}\mathbf{c}^{(j)}$. The “detail information,” lost by approximating $\mathbf{c}^{(j)}$ by the “averaged” $\mathbf{c}^{(j-1)}$, is contained in vector $\mathbf{d}^{(j-1)} = \mathcal{G}\mathbf{c}^{(j)}$.

The discrete wavelet transform of a sequence $\mathbf{y} = \mathbf{c}^{(J)}$ of length 2^J can then be represented as

$$(\mathbf{c}^{(J-k)}, \mathbf{d}^{(J-k)}, \mathbf{d}^{(J-k+1)}, \dots, \mathbf{d}^{(J-2)}, \mathbf{d}^{(J-1)}). \quad (49)$$

Notice that the lengths of \mathbf{y} and its transform in (49) coincide. Because of decimation, the length of $\mathbf{c}^{(j)}$ is twice the length of $\mathbf{c}^{(j-1)}$, and $2^J = 2^{J-k} + \sum_{i=1}^k 2^{J-i}$, $1 \leq k \leq J$.

For an illustration of (49), see Fig. 11. By utilizing the operator notation, it is possible to summarize the discrete wavelet transform (curtailed at level k) in a single line:

$$\mathbf{y} \mapsto (\mathcal{H}^k \mathbf{y}, \mathcal{G}\mathcal{H}^{k-1} \mathbf{y}, \dots, \mathcal{G}\mathcal{H}^2 \mathbf{y}, \mathcal{G}\mathcal{H} \mathbf{y}, \mathcal{G}\mathbf{y}).$$

The number k can be any arbitrary integer between 1 and J and it is associated with the coarsest “smooth” space, V_{J-k} , up to which the transform was curtailed. In terms of multiresolution spaces, (49) corresponds to the multiresolution decomposition $V_{J-k} \oplus W_{J-k} \oplus W_{J-k+1} \oplus \dots \oplus W_{J-1}$. When $k = J$ the vector $\mathbf{c}^{(0)}$ contains a single element, $c^{(0)}$.

If the wavelet filter length exceeds 2, one needs to define actions of the filter beyond the boundaries of the sequence to which the filter is applied. Different policies are possible. The most common is a periodic extension of the original signal.

The reconstruction formula is also simple in terms of operators \mathcal{H}^* and \mathcal{G}^* . They are applied on $\mathbf{c}^{(j-1)}$ and $\mathbf{d}^{(j-1)}$, respectively, and the results are added. The vector $\mathbf{c}^{(j)}$ is reconstructed as

$$\mathbf{c}^{(j)} = \mathcal{H}^* \mathbf{c}^{(j-1)} + \mathcal{G}^* \mathbf{d}^{(j-1)}, \quad (50)$$

Recursive application of (50) leads to

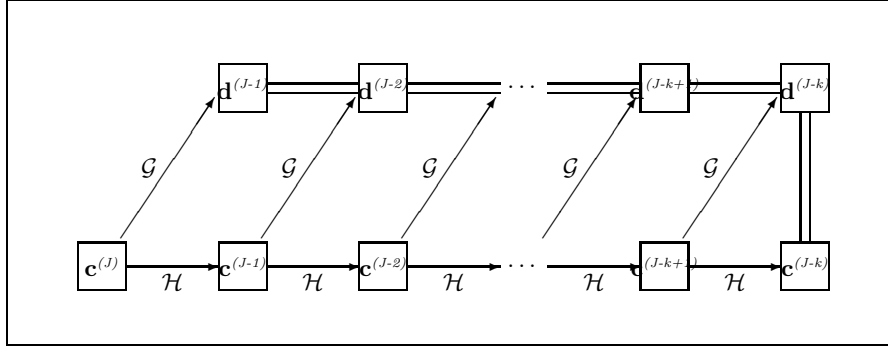


Fig. 11. Forward wavelet transform of depth k (DWT is a vector of coefficients connected by double lines.)

$$\begin{aligned}
 & (\mathcal{H}^k \mathbf{y}, \mathcal{G}\mathcal{H}^{k-1} \mathbf{y}, \dots, \mathcal{G}\mathcal{H}^2 \mathbf{y}, \mathcal{G}\mathcal{H} \mathbf{y}, \mathcal{G} \mathbf{y}) \\
 &= \mathbf{c}^{(J-k)}, \mathbf{d}^{(J-k)}, \mathbf{d}^{(J-k+1)}, \dots, \mathbf{d}^{(J-2)}, \mathbf{d}^{(J-1)} \\
 &\mapsto \sum_{i=1}^{k-1} (\mathcal{H}^*)^{k-1-i} \mathcal{G}^* \mathbf{d}^{(J-k+i)} + (\mathcal{H}^*)^k \mathbf{c}^{(J-k)} = \mathbf{y}.
 \end{aligned}$$

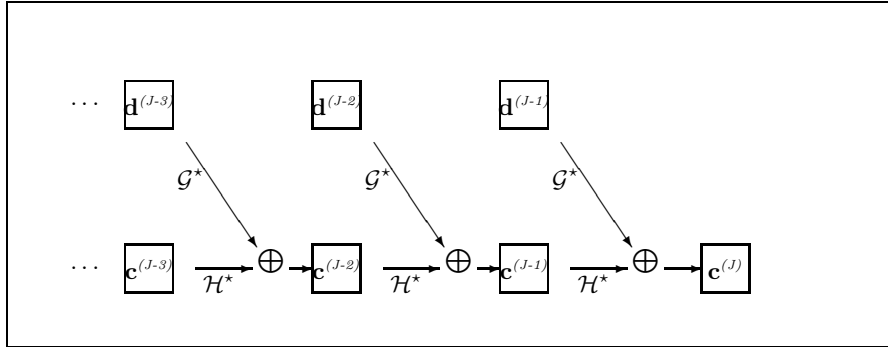


Fig. 12. Inverse Transform

Example 7. Let $\mathbf{y} = (1, 0, -3, 2, 1, 0, 1, 2)$ be an exemplary set we want to transform by Haar's DWT. Let $k = J = 3$, i.e., the coarsest approximation and detail levels will contain a single point each. The decomposition algorithm applied on $\mathbf{y} = (1, 0, -3, 2, 1, 0, 1, 2)$ is given schematically in Fig. 13.

For the Haar wavelet, the operators \mathcal{H} and \mathcal{G} are given by $(\mathcal{H}a)_k = \sum_n h_{n-2k} a_n = \sum_m h_m a_{m+2k} = h_0 a_{2k} + h_1 a_{2k+1} = \frac{a_{2k} + a_{2k+1}}{\sqrt{2}}$. Similarly, $(\mathcal{G}a)_k = \sum_n g_{n-2k} a_n = \sum_m g_m a_{m+2k} = g_0 a_{2k} + g_1 a_{2k+1} = \frac{a_{2k} - a_{2k+1}}{\sqrt{2}}$.

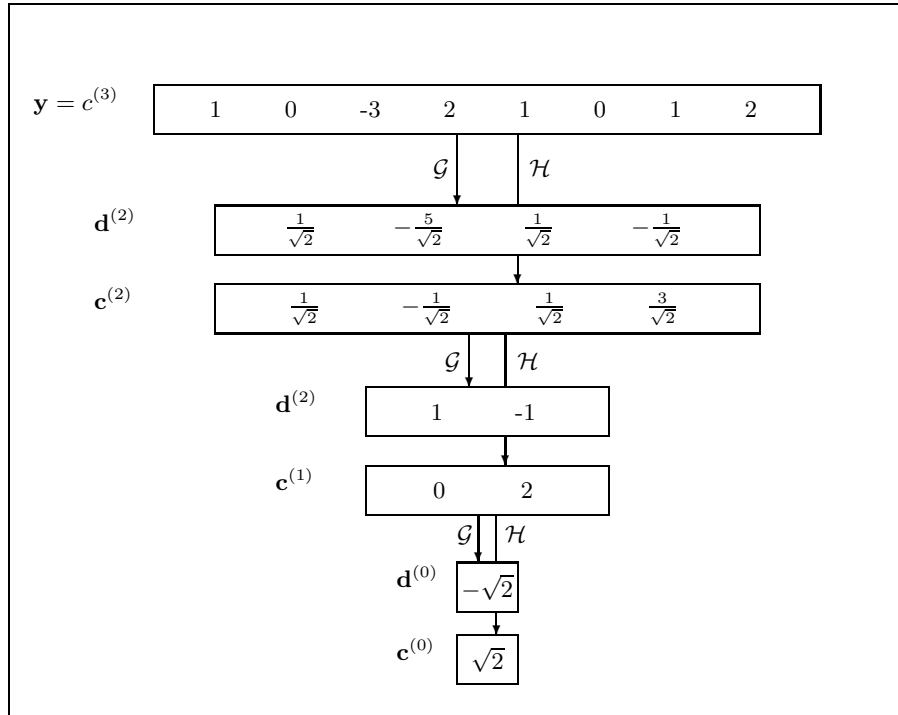


Fig. 13. An illustration of a decomposition procedure.

The reconstruction algorithm is given in Fig. 14. In the process of reconstruction, $(\mathcal{H}^* a)_n = \sum_k h_{n-2k} a_k$, and $(\mathcal{G}^* a)_n = \sum_k g_{n-2k} a_k$. For instance, the first line in Fig. 14 recovers the object $\{1, 1\}$ from $\sqrt{2}$ by applying \mathcal{H}^* . Indeed, $(\mathcal{H}^* \{a_0\})_0 = h_0 \sqrt{2} = 1$ and $(\mathcal{H}^* \{a_0\})_1 = h_1 \sqrt{2} = 1$.

We already mentioned that when the length of the filter exceeds 2, boundary problems occur since the convolution goes outside the range of data.

There are several approaches to resolving the boundary problem. The signal may be continued in a periodic way $(\dots, y_{n-1}, y_n | y_1, y_2, \dots)$, symmetric way $(\dots, y_{n-1}, y_n | y_{n-1}, y_{n-2}, \dots)$, padded by a constant, or extrapolated as a polynomial. Wavelet transforms can be confined to an interval (in the sense of Cohen et al. (1993) and periodic and symmetric extensions can be viewed as special cases. Periodized wavelet transforms are also defined in a simple way.

If the length of the data set is not a power of 2, but of the form $M \cdot 2^K$, for M odd and K a positive integer, then only K steps in the decomposition algorithm can be performed. For precise descriptions of conceptual and calculational hurdles caused by boundaries and data sets whose lengths are not a power of 2, we direct the reader to the monograph by Wickerhauser (1994).

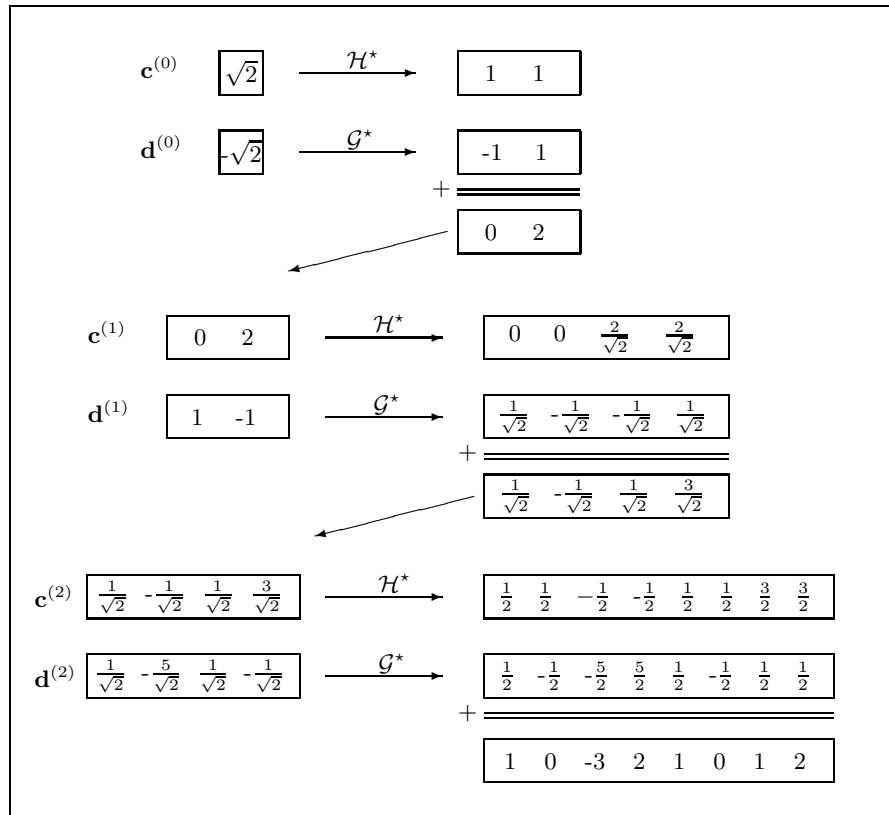


Fig. 14. An illustration of a reconstruction procedure

In this Section we discussed the most basic wavelet transform. Various generalizations include biorthogonal wavelets, multiwavelets, nonseparable multidimensional wavelet transforms, complex wavelets, lazy wavelets, and many more.

For various statistical applications of wavelets (nonparametric regression, density estimation, time series, deconvolutions, etc.) we direct the reader to Antoniadis (1997), Härdle et al. (1998), Vidakovic (1999). An excellent monograph by Walter and Shen (2000) discusses statistical applications of wavelets and various other orthogonal systems.

4.2 Matlab Implementation of Cascade Algorithm

The following two `matlab` m-files implement discrete wavelet transform and its inverse, with periodic handling of boundaries. The data needs to be of dyadic size (power of 2). The programs are didactic, rather than efficient.

For an excellent and comprehensive wavelet package, we direct the reader to wavelab802 module [<http://www-stat.stanford.edu/~wavelab/>] maintained by Donoho and his coauthors.

```
function dwtr = dwtr(data, L, filterh)
% function dwtr = dwt(data, filterh, L);
% Calculates the DWT of periodic data set
% with scaling filter filterh and L scales.
%
% Example of Use:
% data = [1 0 -3 2 1 0 1 2]; filter = [sqrt(2)/2 sqrt(2)/2];
% wt = DWTR(data, 3, filter)
%-----

n = length(filterh);           %Length of wavelet filter
C = data;                      %Data 'live' in V_J
dwtr = [];                     %At the beginning dwtr empty
H =fliplr(filterh);            %Flip because of convolution
G = filterh;                   %Make quadrature mirror
G(1:2:n) = -G(1:2:n);          % counterpart
for j = 1:L                    %Start cascade
    nn = length(C);            %Length needed to
    C = [C(mod((-n-1):-1),nn)+1 C]; % make periodic
    D = conv(C,G);             %Convolve,
    D = D([n:2:(n+nn-2)]+1);   % keep periodic, decimate
    C = conv(C,H);             %Convolve,
    C = C([n:2:(n+nn-2)]+1);   % keep periodic, decimate
    dwtr = [D,dwtr];          %Add detail level to dwtr
end;                            %Back to cascade or end
dwtr = [C, dwtr];              %Add the last 'smooth' part

function data = idwtr(wtr, L, filterh)
% function data = idwt(wtr, L, filterh);
% Calculates the IDWT of wavelet
% transform wtr using wavelet filter
% "filterh" and L scales.
% Example:
%>> max(abs(data - IDWTR(DWTR(data,3,filter), 3,filter)))
%ans = 4.4409e-016
%-----

nn = length(wtr); n = length(filterh); %Lengths
if nargin==2, L = round(log2(nn)); end; %Depth of transform
H = filterh;                    %Wavelet H filter
G =fliplr(H); G(2:2:n) = -G(2:2:n); %Wavelet G filter
LL = nn/(2^L);                  %Number of scaling coeffs
C = wtr(1:LL);                  %Scaling coeffs
for j = 1:L                    %Cascade algorithm
    w = mod(0:n/2-1,LL)+1;      %Make periodic
    D = wtr(LL+1:2*LL);        %Wavelet coeffs
```

```

Cu(1:2:2*LL+n) = [C C(1,w)];           %Upsample & keep periodic
Du(1:2:2*LL+n) = [D D(1,w)];           %Upsample & keep periodic
C = conv(Cu,H) + conv(Du,G);           %Convolve & add
C = C([n:n+2*LL-1]-1);                 %Periodic part
LL = 2*LL;                               %Double the size of level
end;
data = C;                                 %The inverse DWT

```

5 Conclusion

In this Chapter we gave an overview of several transforms useful in computational statistics. We emphasized frequency and scale domain transforms (Fourier and wavelet) since they provide an insight to the phenomena, not available in the domain of untransformed data. Moreover, multiscale transforms are relatively new, and as such deserve more attention. It was pretentious to title this chapter *Transforms in Statistics*, since literally several dozens important transforms are not even mentioned. As it was hinted in the introduction, a just task of overviewing all important transformations used in statistical practice would take a space of a large monograph.

Acknowledgment. Work on this Chapter was supported by DOD/NSA Grant E-24-60R at Georgia Institute of Technology. Editor Jim Gentle read early versions of the chapter and gave many valuable comments. All `matlab` programs that produced figures and simulations are available from the author at request.

References

- Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis, Second Edition, Wiley, New York.
- Antoniadis, A. (1997). Wavelets in statistics: A Review. *J. Ital. Statist. Soc.*, 6: 97–144.
- Baraniuk, R.G. (1994). Wigner–Ville spectrum estimation via wavelet soft–thresholding. In *Proc. IEEE-SP Int. Symp. on Time-Frequency and Time-Scale Analysis*, Philadelphia.
- Box, G.E.P. and Cox, D.R. (1964). An Analysis of Transformations, *Journal of the Royal Statistical Society*, 26: 211–243, discussion 244–252.
- Brigham, E.O. (1988). *The Fast Fourier Transform and Its Applications*, Prentice-Hall, Englewood Cliffs, NJ.
- Carmona, R., Hwang, W-L. and Torr sani, B. (1998). *Practical Time-Frequency Analysis*, volume 9 of *Wavelet Analysis and its Applications*, Academic Press, San Diego.
- Cohen, A, Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, 1(1):54–81.

- Daubechies, I. (1992). *Ten Lectures on Wavelets*, Number 61 in CBMS-NSF Series in Applied Mathematics, SIAM, Philadelphia.
- Feuerverger, A. and Mureika, R. (1977). The empirical characteristic function and its applications, *The Annals of Statistics*, 5: 88–97.
- Flandrin, P. (1992). Time-scale analyses and self-similar stochastic processes. In Byrnes et al. (eds), *Wavelets and Their Applications*, pp. 121 – 142, NATO ASI Series vol. 442.
- Flandrin, P. (1999). *Time-Frequency/Time-scale Analysis*, Academic Press, 386pp.
- Gabor, D. (1946). Theory of communication. *J. IEEE*, 93: 429–457.
- Grossmann, A. and Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math.*, 15: 723–736.
- Grossmann, A. and Morlet, J. (1985). Decomposition of functions into wavelets of constant shape and related transforms. In Streit, L. (ed), *Mathematics and physics, lectures on recent results*, World Scientific, River Edge, NJ.
- Härdle, W., Kerkycharian, G., Pickard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statistics 129. Springer-Verlag, New York.
- Mallat, S.G. (1989a). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315: 69–87.
- Mallat, S.G. (1989b). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Patt. Anal. Mach. Intell.*, 11(7): 674–693.
- Mallat, S.G. (1999). *A Wavelet Tour of Signal Processing*, Second Edition. Academic Press, San Diego.
- Morlet, J., Arens, G., Fourgeau, E. and Giard, D. (1982). Wave propagation and sampling theory. *Geophys.*, 47: 203–236.
- Murata, N. (2001). Properties of the empirical characteristic function and its application to testing for independence. In Lee, Jung, Makeig, and Sejnowski (eds), *Proceedings ICA2001, 3rd International Conference on Independent Component Analysis*, San Diego.
- Pensky, M. and Vidakovic, B. (2003) Bayesian decision theoretic scale-adaptive estimation of log-spectral density. Technical Report 01-2003, ISyE, Georgia Institute of Technology. <http://www.isye.gatech.edu/~brani/isyestat/>
- Tong, H. (1996). *Non-Linear Time Series*, Clarendon Press, Oxford.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, Wiley, NY.
- Ville, J. (1948). Théorie et Applications de la Notion de Signal Analytique, *Cables et Transmission*, 2A: 61–74.
- Walter, G.G. and Shen, X. (2000). *Wavelets and Other Orthogonal Systems*, Second Edition, CRC Press.
- Wickerhauser, M. V. (1994). *Adapted Wavelet Analysis from Theory to Software*, A K Peters, Ltd., Wellesley, MA.

Index

- cascade algorithm, 31
- critical sampling, 18

- decomposition algorithm, 35

- filter
 - high-pass, 26
 - quadrature mirror, 26

- Heisenberg's uncertainty principle, 10

- mirror filter, 31
- moment generating function, 5
- mother wavelet, 24
- MRA, 19
- multiresolution analysis, 19

- normalization property, 21, 27

- orthogonality property, 21, 27

- Parseval formula, 12
- periodogram, 8

- resolution of identity, 18

- scaling equation, 19
- scaling function, 19
- spectral density, 8

- spectrogram, 10

- thresholding, 15

- transform
 - continuous wavelet, 15
 - discrete Fourier, 7
 - discrete wavelet, 28
 - empirical Fourier-Stieltjes, 5
 - fast Fourier, 28
 - Fourier-Stieltjes, 4
 - Hilbert, 11
 - integral Fourier, 10
 - Laplace, 5
 - short time Fourier, 10
 - Wigner-Ville, 12
 - windowed Fourier, 10

- transformation
 - Box-Cox, 4
 - Fisher z , 3

- vanishing moments, 27

- wavelet domain, 28
- wavelets, 14
 - Daubechies, 27
 - Haar, 26
 - Mexican hat, 17
 - periodized, 35
 - sombbrero, 17

