Chapter 5
Wavelet Packet Transform

While discrete wavelet transform provides flexible time–frequency resolution, it suffers from a relatively low resolution in the high-frequency region. This deficiency leads to difficulty in differentiating high-frequency transient components. The wavelet packet transform (WPT), in comparison, further decomposes the *detailed* information of the signal in the high-frequency region, thereby overcoming this limitation. Figure 5.1 schematically illustrates a WPT-based signal decomposition process, where a four-level WPT produces a total of 16 subbands, with each subband covering one-sixteenth of the signal frequency spectrum (Gao and Yan 2006). The enhanced signal decomposition capability makes WPT an attractive tool for detecting and differentiating transient elements with high-frequency characteristics.

In this chapter, we introduce the theoretical basis of a wavelet packet and algorithms to realize the WPT. Representative applications of the WPT are then introduced to illustrate this computational technique.

5.1 Theoretical Basis of Wavelet Packet

5.1.1 Definition

The wavelet packet is defined by the following equation (Wickerhauser 1991):

$$\begin{cases} u_{2n}^{(j)}(t) = \sqrt{2}\sum_k h(k)u_n^{(j)}(2t-k) \\ u_{2n+1}^{(j)}(t) = \sqrt{2}\sum_k g(k)u_n^{(j)}(2t-k). \end{cases} \quad \text{with } n=0,1,2,\ldots \text{ and } k=0,1,\ldots,m \tag{5.1}$$

with $u_0^{(0)}(t)$ being the scaling function $\phi(t)$, that is, $u_0^{(0)}(t) = \phi(t)$, and $u_1^{(0)}(t)$ being the base wavelet function $\psi(t)$, that is, $u_1^{(0)}(t) = \psi(t)$ (Wickerhauser 1991). The superscript (j) in (5.1) denotes the jth level wavelet packet basis, and there will be 2^j wavelet packet bases at the jth level.

R.X. Gao and R. Yan, *Wavelets: Theory and Applications for Manufacturing*,
DOI 10.1007/978-1-4419-1545-0_5, © Springer Science+Business Media, LLC 2011

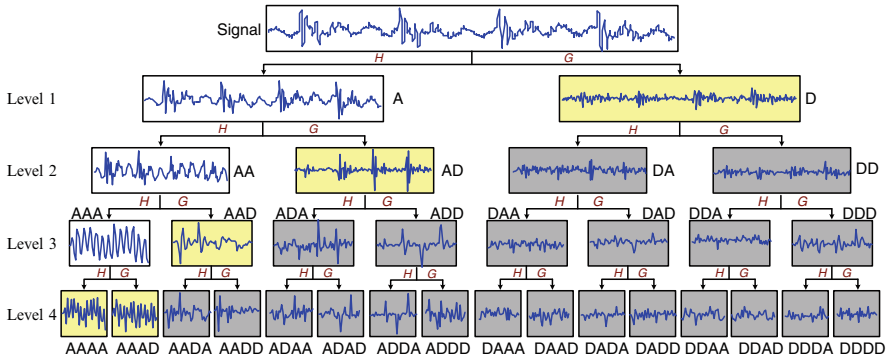

Fig. 5.1 Procedure for signal decomposition using wavelet packet transform. Note: *A* approximate information, *D* detailed information, *H* low-pass filter, *G* high-pass filter

To illustrate the derivation process of wavelet packet basis, the Haar wavelet (Haar 1910) is used here as an example. The coefficients $h(k)$ and $g(k)$ for Haar wavelet are defined as (Daubechies 1992)

$$\begin{cases} h(0) = h(1) = \dfrac{1}{\sqrt{2}},\ h(k) = 0 \ \text{ when } k = 2, 3, \ldots, m \\ g(0) = g(1) = -\dfrac{1}{\sqrt{2}},\ g(k) = 0 \ \text{ when } k = 2, 3, \ldots, m \end{cases} \tag{5.2}$$

From (5.1) and (5.2), the first level of the Haar wavelet packet basis, indicated by the superscript (1) is obtained as

$$\begin{cases} u_0^{(1)}(t) = u_0^{(0)}(2t) = \phi(2t) \\ u_1^{(1)}(t) = \sqrt{2}\dfrac{1}{\sqrt{2}}[u_0^{(1)}(2t) - u_0^{(1)}(2t-1)] = u_0^{(1)}(2t) - u_0^{(1)}(2t-1) \end{cases} \tag{5.3}$$

Similarly, the second and third levels of the Haar wavelet packet basis can be derived using (5.4) and (5.5), respectively:

$$\begin{cases} u_0^{(2)}(t) = \phi(4t) \\ u_1^{(2)}(t) = u_0^{(2)}(2t) - u_0^{(2)}(2t-1) \\ u_2^{(2)}(t) = u_1^{(2)}(2t) + u_1^{(2)}(2t-1) \\ u_3^{(2)}(t) = u_1^{(2)}(2t) - u_1^{(2)}(2t-1) \end{cases} \tag{5.4}$$

$$\begin{cases} u_0^{(3)}(t) = \phi(8t) \\ u_{2n}^{(3)}(t) = u_n^{(3)}(2t) + u_n^{(3)}(2t-1), \quad n = 1,2,3 \\ u_{2n+1}^{(3)}(t) = u_n^{(3)}(2t) - u_n^{(3)}(2t-1), \quad n = 0,1,2,3 \end{cases} \tag{5.5}$$

Figure 5.2a–c illustrates the waveforms of the Haar wavelet packet bases at levels 1 through 3 that are derived from the scaling function. Using the same approach, the Haar wavelet packet bases at all other levels can be obtained.

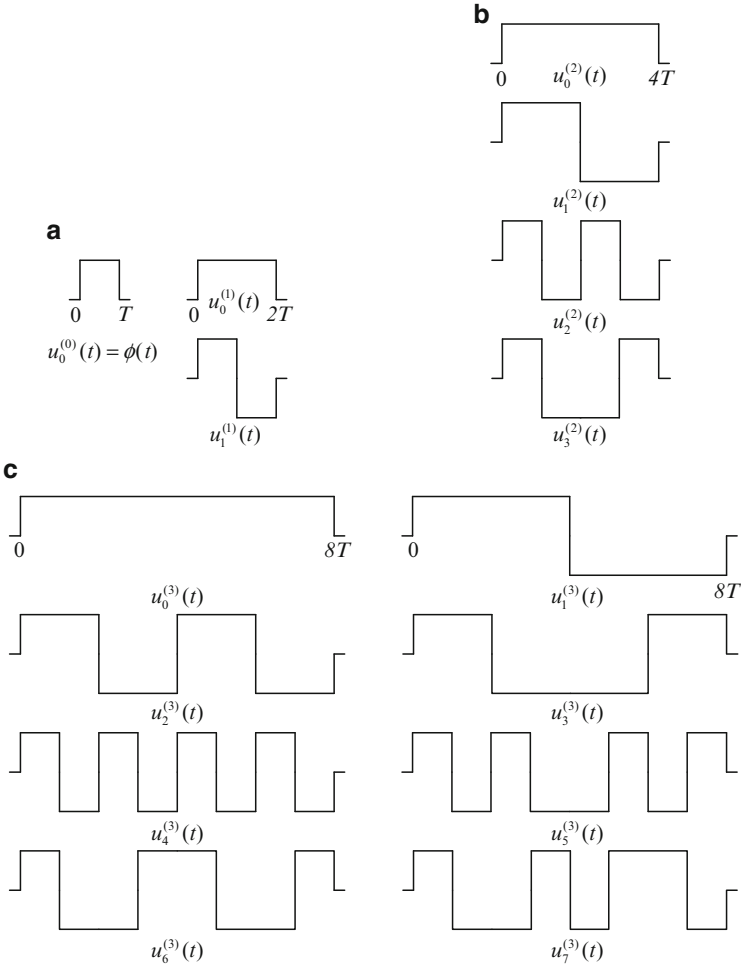

Fig. 5.2 Wavelet packet bases for the Haar wavelet: (**a**) level 1; (**b**) level 2; and (**c**) level 3

5.1.2 Wavelet Packet Property

Equation (5.1) indicates that the wavelet packet has the following properties (Wickerhauser 1991; Coifman et al. 1992).

5.1.2.1 Shift Orthogonality

If $\{u_n^{(j)}(t)\}_{n\in Z}$ is the set of wavelet packet bases obtained from the scaling function $u_0^{(0)}(t) = \phi(t)$ of an orthogonal base wavelet, then these bases hold the property of shift orthogonality:

$$\langle u_n^j(t),\ u_n^j(t-k)\rangle = \delta_k, \quad k \in z \tag{5.6}$$

where $\langle\cdot\rangle$ denotes inner product operation. The symbol δ_k represents a Dirac function.

Proof: When n = 0, $u_0^{(j)}(t)$, and $u_1^{(j)}(t)$ are the scaled versions of $\phi(t)$ and $\psi(t)$, respectively. By definition of the scaling function and base wavelet function, they are orthogonal (Daubechies 1992).

When $n \neq 0$, as $u_2^{(j)}(t)$ and $u_3^{(j)}(t)$ are both a linear combination of $u_1^{(j)}(t)$ as seen in (5.3) and (5.4), and $u_1^{(j)}(t)$ is a scaled version of the wavelet function $\psi(t)$, which is orthogonal and normalized (Daubechies 1992), $u_2^{(j)}(t)$ and $u_3^{(j)}(t)$ are orthogonal.

As an example, if we have

$$\begin{cases} u_2^{(j)}(t) = \sqrt{2}\sum_{k'} h_{k'} u_1^{(j)}(2t-k') \\ u_2^{(j)}(t-k) = \sqrt{2}\sum_{k''} h_{k''} u_1^{(j)}(2t-2k-k'') \end{cases} \tag{5.7}$$

where $k' = 0, 1, \ldots, m$ and $k'' = 0, 1, \ldots, m$.

Then

$$\langle u_2^{(j)}(t),\ u_2^{(j)}(t-k)\rangle = 2\sum_{k'}\sum_{k''} h_{k'}h_{k''}\langle u_1^{(j)}(2t-k'),\ u_1^{(j)}(2t-2k-k'')\rangle \tag{5.8}$$

The inner product on the right-hand side of (5.8) is equal to 1/2 when $k' = 2k + k''$; otherwise, it is equal to zero. Therefore,

$$\langle u_2^{(j)}(t),\ u_2^{(j)}(t-k)\rangle = \sum_{k'} h_{k''}h_{2k+k''} = \delta_k \tag{5.9}$$

Similarly, $u_4^{(j)}(t)$ and $u_5^{(j)}(t)$ are both linear combinations of $u_2^{(j)}(t)$, and they are also orthogonal. Using the same approach, wavelet packet bases of higher levels can be derived.

5.1.2.2 Orthogonal Relationship between $u_{2n}^{(j)}(t)$ and $u_{2n+1}^{(j)}(t)$

$$\langle u_{2n}^{(j)}(t),\ u_{2n+1}^{(j)}(t)\rangle = 0 \tag{5.10}$$

Proof From (5.1), we have

$$\begin{aligned}\langle u_{2n}^{(j)}(t),\ u_{2n+1}^{(j)}(t)\rangle &= 2\int \sum_{k'}\sum_{k''} h_{k'}g_{k''}u_n^{(j)}(2t-2k-k')u_n^{(j)}(2t-k'')\,dt \\ &= 2\sum_{k'}\sum_{k''} h_{k'}g_{k''}\int u_n^{(j)}(2t-2k-k')u_n^{(j)}(2t-k'')\,dt\end{aligned} \tag{5.11}$$

The result of integral part in (5.11) is equal to zero except when $k'' = 2k + k'$. Therefore,

$$\langle u_{2n}^{(j)}(t),\ u_{2n+1}^{(j)}(t)\rangle = \sum_{k''} h_{k'}g_{k''} = 0 \tag{5.12}$$

5.2 Recursive Algorithm

Once the wavelet packet basis is defined using (5.1), a recursive algorithm can be designed to implement WPT for signal decomposition. The result of the decomposition is given by (Mallat 1999):

$$\begin{cases} d_{j+1,2n} = \sum_m h(m-2k)d_{j,n} \\ d_{j+1,2n+1} = \sum_m g(m-2k)d_{j,n} \end{cases} \tag{5.13}$$

where $d_{j,n}$ denotes the wavelet coefficients at the jth level, nth subband, $d_{j+1,2n}$, and $d_{j+1,2n+1}$ denote the wavelet coefficients at the $(j+1)$th level, $2n$th, and $(2n+1)$th subbands, respectively, and m is the number of the wavelet coefficients.

Theoretically, there are multiple ways ($>2^L$) to analyze a signal using an L-level decomposition (Mallat 1999). This makes it possible to optimize the signal decomposition process and improve the effectiveness. Various criteria, such as $l_p(p \le 2)$ norm, logarithmic entropy, and Shannon entropy, can be utilized as the cost function to facilitate the optimization process. A widely applied criterion for optimal WPT-based signal representation is the Shannon entropy (Coifman and Wickerhauser 1992). For wavelet coefficients at the nth subfrequency band within the level j, $d_{j,n} = \{d_{j,n} : n = 1, 2, \ldots, 2^j\}$, the Shannon entropy is defined as

$$Entropy\,(d_{j,n}) = -\sum_i p_i \cdot \log(p_i) \tag{5.14}$$

where p_i is the probability distribution of the energy contained in the wavelet coefficients at the nth subfrequency band within the level j. The probability distribution function is defined as

$$p_i = \left|d_{j,n}(i)\right|^2 / \left\|d_{j,n}\right\|^2 \tag{5.15}$$

with $\sum_{i=1}^{m} p_i = 1$, and $p_i \cdot \log_2 p_i = 0$ if $p_i = 0$. The upper limit m represents the number of wavelet coefficients at the nth subfrequency band within the level j.

Equations (5.13) and (5.14) indicate that the entropy of the wavelet coefficients is bounded by

$$0 \leq E_{\text{entropy}}(d_{j,n}) \leq \log_2 m \tag{5.16}$$

From (5.16), we see that the Shannon entropy will have a large value if the energy content is spread out across the constituent wavelet coefficients within the subfrequency band. Conversely, it assumes a small value if the energy is concentrated on a few dominant components. As we want the signal information to be concentrated within as few coefficients as possible, the minimum Shannon entropy should be contained in the wavelet coefficients as a result of the signal decomposition. Mathematically, such a process involves comparing the entropy of the lower level (e.g., in the subbands *DAAA* and *DAAD*, Fig. 5.1) of the tree structure with the entropy of the higher source level (e.g., subband *DAA*), starting from the bottom of the decomposition (e.g., level 4). If the higher level has returned smaller entropy than the sum of the entropies from the lower level, then the higher level subfrequency band will be retained. Otherwise, it will be replaced by the two subfrequency bands at the lower level. Such a process is executed until it reaches the top level of the decomposition.

5.3 FFT-Based Harmonic Wavelet Packet Transform

Besides the recursive algorithm introduced in the previous section, another algorithm for WPT, based on the Fourier transform, has been shown to be effective when realizing the harmonic wavelet packet transform (HWPT) (Samuel et al. 2000; Yan and Gao 2005).

5.3.1 Harmonic Wavelet Transform

The mathematical expression of the harmonic wavelet is defined in Chap. 3 as

$$\Psi_{m,n}(f) = \begin{cases} \dfrac{1}{(n-m)} & m \leq f \leq n \\ 0 & \text{elsewhere} \end{cases} \tag{5.17}$$

Accordingly, its corresponding time domain expression is obtained by taking the inverse Fourier transform as (Yan and Gao 2005)

$$\psi_{m,n}(t) = \frac{e^{jn2\pi t} - e^{jm2\pi t}}{j2\pi(n-m)t} \tag{5.18}$$

If the harmonic wavelet is translated by a step $k/(m - n)$, in which k is the translation parameter, a generalized expression that is centered at $t = k/(n - m)$ with a bandwidth of $(n - m)$ can be obtained as (Newland 1994)

$$\psi_{m,n}\left(t - \frac{k}{n-m}\right) = \frac{\left(e^{jn2\pi\left(t-\frac{k}{n-m}\right)} - e^{jm2\pi\left(t-\frac{k}{n-m}\right)}\right)}{j2\pi(n-m)\left(t-\frac{k}{n-m}\right)} \tag{5.19}$$

On the basis of the generalized expression, the harmonic wavelet transform of a signal $x(t)$ can be performed as

$$hwt(m,n,k) = (n-m)\int_{-\infty}^{\infty} x(\tau)\psi^*_{m,n}\left(\tau - \frac{k}{n-m}\right)d\tau \tag{5.20}$$

where $hwt(m, n, k)$ is the harmonic wavelet coefficient.

By taking the Fourier transform of (5.20), an equivalent expression of the harmonic wavelet transform in the frequency domain can be expressed as

$$HWT(m,n,f) = X(f) \cdot \Psi^*((n-m)f) \tag{5.21}$$

where $X(f)$ is the Fourier transform of the signal $x(t)$, and $\Psi^*((n - m)f)$ is the conjugate of $\Psi((n - m)f)$, which is the Fourier transform of the harmonic wavelet at the scale (m, n). As the harmonic wavelet has compact frequency expression as shown in (5.17), the harmonic wavelet transform can be readily obtained through a pair of Fourier transform and inverse Fourier transform operations (Newland 1993).

As shown in Fig. 5.3, after taking the Fourier transform of a signal $x(t)$ to obtain its frequency domain expression $X(f)$, the inner product $HWT(m, n, f)$ of $X(f)$ and the conjugate of the harmonic wavelet $\Psi^*((n - m)f)$ at the scale (m, n) are calculated. Finally, the harmonic wavelet transform of the signal $x(t)$, denoted as $hwt(m, n, k)$, is obtained by taking the inverse Fourier transform of the inner product $HWT(m, n, f)$.

5.3.2 Harmonic Wavelet Packet Algorithm

The scale parameters m and n determine the bandwidth that the harmonic wavelet covers. Shown in Fig. 5.4a–d are the real and imaginary parts of the generalized harmonic wavelet under two exemplary sets of scale parameters, $m = 0$, $n = 16$ and

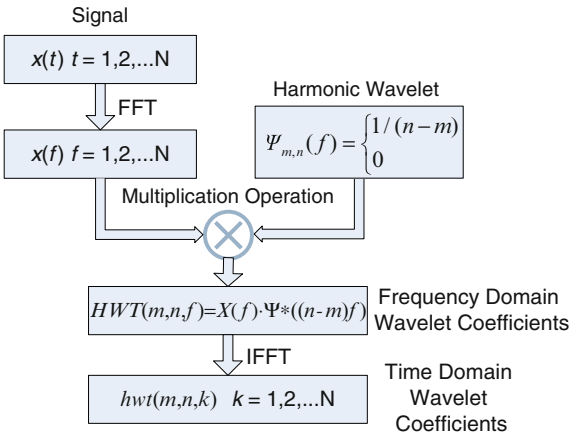

Fig. 5.3 Algorithm for implementing the harmonic wavelet transform

$m = 16$, $n = 32$, while the translation parameter $k = 8$ remains the same. We can see that, through appropriate variation of these two scale parameters, the harmonic wavelet can be scaled to match the signal within different frequency regions associated with the same bandwidth of $(n - m)$ (16 in this example), as shown in Fig. 5.4e–f. As a result, the HWPT operation is realized.

Similar to the WPT, the number of frequency subbands for the HWPT has to be s powers of two, in which s corresponds to the decomposition level of WPT. As a result, the signal can be decomposed into 2^s frequency subbands, with the bandwidth expressed in Hertz for each subband that is defined by

$$f_{\text{band}} = \frac{f_{\text{h}}}{2^s} \tag{5.22}$$

In (5.20), f_{h} is the highest frequency component of the signal to be analyzed. As the bandwidth of the harmonic wavelet is $(n - m)$, selection of the values for m and n of the HWPT has to satisfy the following condition:

$$m - n = f_{\text{band}} \tag{5.23}$$

Thus, the harmonic wavelet packet coefficients $hwpt(s, i, k)$ can be obtained as

$$hwpt(s, i, k) = hwt(m, n, k) \tag{5.24}$$

where s is the decomposition level, i is the index of the subband, and k is the index of the coefficient. In addition, the parameters m and n need to satisfy the following condition:

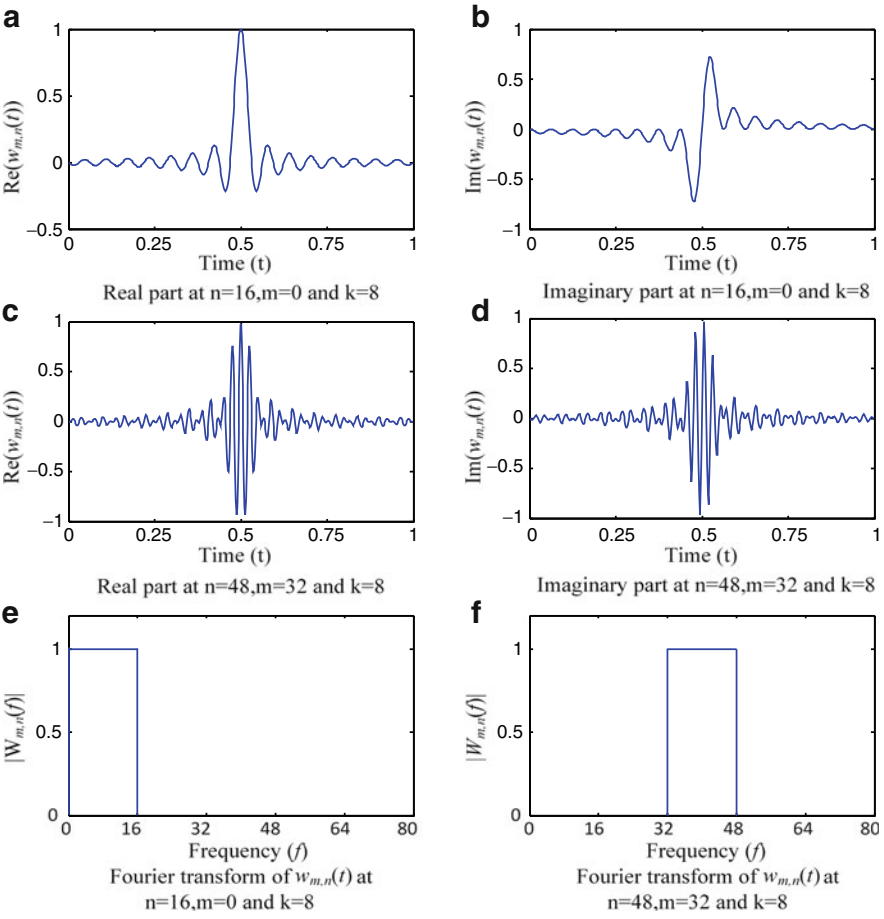

Fig. 5.4 Waveforms of the harmonic wavelet with their Fourier transforms under different scale parameters

$$\begin{cases} m = i \times f_{\text{band}} = i \times \dfrac{f_{\text{h}}}{2^s} \\ n = (i+1) \times f_{\text{band}} = (i+1) \times \dfrac{f_{\text{h}}}{2^s} \end{cases}, \quad i = 0, 1, \ldots, 2^s - 1 \tag{5.25}$$

As a result, by selecting the appropriate parameter pairs (m, n) based on (5.25), the FFT-based HWPT algorithm can be realized through the computational process as illustrated in Fig. 5.3.

5.4 Application of Wavelet Packet Transform

Using the WPT, we can determine a signal's time–frequency composition, thereby having a good understanding of what is contained within the signal. Furthermore, the WPT can be applied to remove noise contained in the signal. In the following, we demonstrate two examples of these applications.

5.4.1 Time-Frequency Analysis

Figure 5.5 shows a vibration signal measured on a ball bearing during a run-to-failure test. Physically, when a localized defect is initiated in a rolling element bearing, for example, due to spalling on the surface of the bearing raceway, impact will be generated every time when a rolling element rolls over the defect. Such impacts subsequently excite the intrinsic modes of the bearing system, giving rise to transient vibrations at the mode-related resonant frequencies. As the defect size increases, different intrinsic modes of the bearing system will be excited, leading to frequency shifts of the impact-induced transient vibrations. Therefore, by evaluating the time–frequency distributions of the vibration signal, degradation of the bearing's health condition can be monitored.

Applying the WPT to the vibration data, we have seen in Fig. 5.6 that not only all the major transient elements are identified, but the corresponding frequency shifts are also clearly seen. The result also shows the increased number of frequency components after the 45-ms time point, reflecting the defect size propagation.

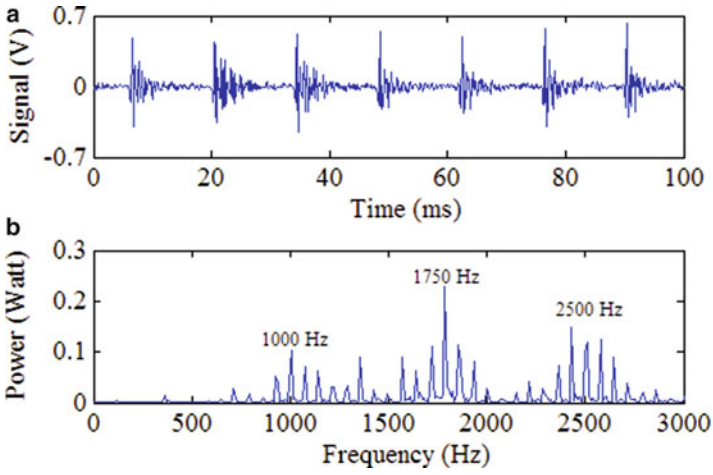

Fig. 5.5 Vibration signal from a ball bearing

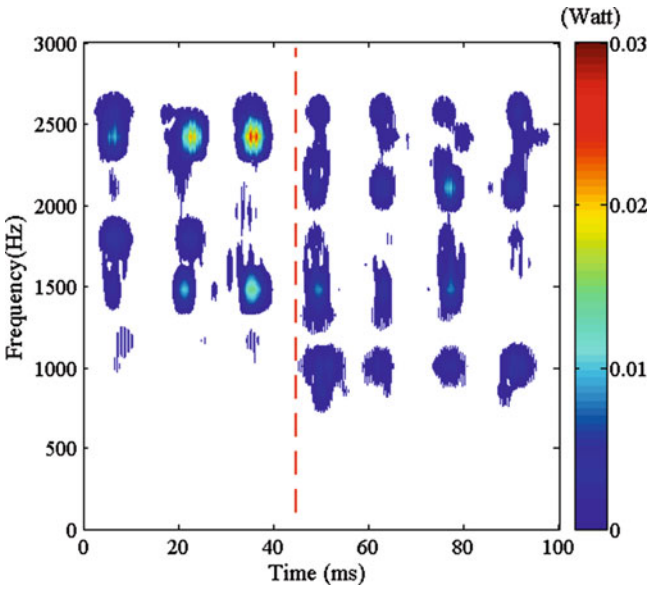

Fig. 5.6 Wavelet packet transform of the bearing vibration signal

5.4.2 Wavelet Packet for Denoising

Figure 5.7a shows a noisy chirp signal, where Gaussian noise is added to the signal, leading to a signal-to-noise ratio of seven. The denoising idea illustrated here is in principle identical to that developed in the wavelet framework in Chap. 4. The only difference is that the WPT provides better flexibility because of a more complete analysis of the signal. In this example, the Stein's unbiased estimate of risk (SURE) criterion threshold is used to construct the wavelet coefficients (Donoho 1995; Donoho and Johnstone 1995). For the purpose of comparison, the signal is processed using both the wavelet packets-based denoising and wavelet-based denoising techniques, and the results are shown in Fig. 5.7b, c, respectively. It can be seen that the performance of the wavelet packets-based denoising approach is better than that of the wavelet-based approach.

5.5 Summary

This chapter begins with the introduction of a theoretical basis of a wavelet packet, where the definition of the wavelet packet and its related properties are presented. Two approaches for implementing the WPT are then discussed. Applications of the WPT on time-frequency analysis and denoising are illustrated in Sect. 5.4.

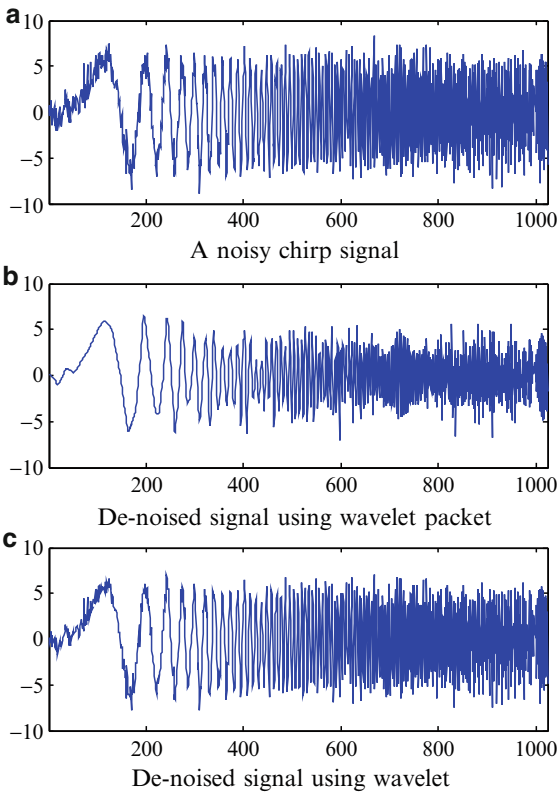

Fig. 5.7 Example of wavelet packet for denoising. (**a**) A noisy chirp signal, (**b**) denoised signal using wavelet packet, and (**c**) denoised signal using wavelet

5.6 References

Coifman RD et al (1992) Wavelet and signal processing. In: Ruskai (ed) Wavelet and their application. Jones and Bartlett Publishers, Boston, MA

Coifman RR, Wickerhauser MV (1992) Entropy based algorithms for best basis selection. IEEE Trans Inform Theory 38(2):713–718

Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia, PA

Donoho DL (1995) De-noising by soft-thresholding. IEEE Trans Inform Theory 41(3):613–627

Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 90(432):1200–1244

Gao R, Yan RQ (2006) Non-stationary signal processing for bearing health monitoring. Int J Manuf Res 1(1):18–40

Haar A (1910) Zur theorie der orthgonalen funktionensysteme. Math Annalen 69:331–371

Mallat SG (1998) A wavelet tour of signal processing. Academic, San Diego, CA

Newland DE (1993) Random vibrations, spectral and wavelet analysis. 3rd edn. Addison Wesley Longman, Boston, MA

Newland DE (1994) Wavelet analysis of vibration part I: theory; part II: wavelet maps. J Vib Acous 116(4):409–425

Samuel PD, Pines DJ, Lewicki DG (2000) A comparison of stationary and non-stationary metrics for detecting faults in helicopter gearboxes. J Am Helicopter Soc 45:125–136

Wickerhauser MV (1991) INRIA lectures on wavelet packet transform

Yan R, Gao R (2005) An efficient approach to machine health evaluation based on harmonic wavelet packet transform. Robot Comput Integrated Manuf 21:291–301