DECOMPOSITION OF VIBRATION SIGNALS INTO DETERMINISTIC AND NONDETERMINISTIC COMPONENTS AND ITS CAPABILITIES OF FAULT DETECTION AND IDENTIFICATION

TOMASZ BARSZCZ

Department of Robotics and Mechatronics AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Cracow, Poland e-mail: tbarszcz@agh.edu.pl

The paper investigates the possibility of decomposing vibration signals into deterministic and nondeterministic parts, based on the Wold theorem. A short description of the theory of adaptive filters is presented. When an adaptive filter uses the delayed version of the input signal as the reference signal, it is possible to divide the signal into a deterministic (gear and shaft related) part and a nondeterministic (noise and rolling bearings) part. The idea of the self-adaptive filter (in the literature referred to as SANC or ALE) is presented and its most important features are discussed. The flowchart of the Matlab-based SANC algorithm is also presented. In practice, bearing fault signals are in fact nondeterministic components, due to a little jitter in their fundamental period. This phenomenon is illustrated using a simple example. The paper proposes a simulation of a signal containing deterministic and nondeterministic components. The self-adaptive filter is then applied—first to the simulated data. Next, the filter is applied to a real vibration signal from a wind turbine with an outer race fault. The necessity of resampling the real signal is discussed. The signal from an actual source has a more complex structure and contains a significant noise component, which requires additional demodulation of the decomposed signal. For both types of signals the proposed SANC filter shows a very good ability to decompose the signal.

Keywords: decomposition, vibration, deterministic component, nondeterministic component, rolling bearing.

1. Introduction

According to the Wold theorem, it is possible to decompose each signal into a deterministic and a nondeterministic part. This idea is well developed by statisticians (Widrow *et al.*, 1975) and was successfully applied in telecommunications under the name *Adaptive Line Enhancer* (Zeidler *et al.*, 1978). The decomposition is based on the autocorrelation features of the signal, because the correlation time of the deterministic (i.e., periodic) part is much longer than that of the nondeterministic part. This theory led to the development of adaptive filters, which were a practical application of the method.

The idea of signal decomposition was proposed for machinery diagnostics, as it can be used for the separation of signals emitted from different machinery elements. Such separation can greatly enhance the quality of further signal processing and, finally, improve the ability to detect and identify a fault. In machinery diagnostics, there is a difference in the correlation times of different signals. Longer correlation times exist in devices which maintain

constant frequency, e.g., gears which are phase locked. On the other hand, rolling bearings will always show some degree of slip, even though it may be very small. Therefore, the method could be used to separate the vibration signal into parts generated by gears and by rolling bearings. The method has also great practical importance, as a majority of rotating machinery (except large output machines on sliding bearings) is composed of several shafts coupled by gears, and shafts are almost always also mounted on rolling bearings. What is more, the most common and most important faults are those that occur in shafts, gears and rolling bearings. Classical methods of vibrodiagnostics are based on the investigation of spectral lines, which are generated by a faulty element. When machines are complex and contain several shafts of different rotational speeds, the generated spectra become increasingly complex. Thus, a method to distinguish between spectrum components coming from gears and bearings would be of considerable value. This concept was presented by Chaturvedi and Thomas (1981). It was further developed by Randall and Ho (2000). The main problem was the reference signal, which is necessary for adaptive filters. Randall proposed to use the original signal with a time delay which would be greater than the autocorrelation time of the nondeterministic part of the signal. The concept was called Self Adaptive Noise Cancellation (SANC).

The paper starts with the simulation of a signal similar to that generated by a rotating machine with a faulty rolling bearing. Next, it presents the theoretical basis for SANC-based vibration signal decomposition. First, the Wold theorem is discussed, and then the adaptive filter architecture and theory are presented. The paper then investigates the ability of self-adaptive signal separation to detect rolling bearing faults. An algorithm prepared in Matlab is proposed and described. The performance of the algorithm is verified on the simulated signal. Following this, the algorithm is applied to a real vibration signal from a wind turbine with a faulty outer race rolling bearing. In both cases the algorithm shows a very good ability of signal separation.

2. Simulation

Rolling bearings, unlike gears, are not phase locked to the rotational speed of the machine. Formulas describing characteristic frequencies are well known. For example, the outer ring ball pass frequency is

$$BPFO = S \frac{N_r}{2} \left(1 - \frac{R_d \cos \phi}{P_d} \right), \tag{1}$$

where S is the rotational speed of the shaft, N_r is the number of rolling elements, R_d is the diameter of the roller, P_d is the pitch diameter, and Φ is the angle. A full list of rolling bearing induced frequencies can be found, e.g., in (Zółtowski and Cempel, 2004). Frequencies generated by the rolling bearing are never exactly equal to the presented formula, because the load angle may vary and a slip between rolling elements and races is unavoidable. Therefore, if there is a fault on an element, it generates an impulse which is transmitted through the casing structure and measured by a vibration sensor. We can observe a series of decaying components. The carrier frequency depends on the structure, not the fault. The fault determines the period of decaying impulses. Due to small changes in rolling bearing operation, those impulses will be excited at a slightly varying frequency, close to, e.g., BPFO.

Randall in 2004 proposed frequency-based simulation of such a signal. Here, time-based simulation was performed. The signal was simulated in the Matlab environment. The periodic component was a single sine function at a frequency of 100 Hz, simulating a gear mesh frequency. Bearing induced impulses were simulated as exponentially decaying signals having a carrier frequency of 4 kHz and a period of 11.6 Hz, which is a reasonable value for a BPFO of a bearing on a shaft rotating at 180 rpm (the slow shaft of a gearbox generating the mentioned 100 Hz component). Figure 1 presents the simulated signal of both components (on the left) and of decaying components only (on the right). Plot noise was set to zero for picture clarity. Note that the zero-peak amplitude of impulses was 10 times smaller than that of the sine signal. Three small decaying impulses can be seen in the left part of Fig. 1.

The signal was simulated with and without the jitter of the BPFO frequency. Visually, there was no difference in both signals, but their spectra showed great differences (see Fig. 2). For clarity, only the part of the spectra around the carrier frequency were zoomed. The left panel presents the spectrum of the signal without the jitter in the period of bearing fault components and the right panel with 2% jitter.

It can be clearly seen that the slight variation in the period of impulses causes the smearing of the spectrum and changes its nature from a set of sharp lines to a continuous one. In the next section, theoretical issues concerning signal decomposition will be discussed. The following section will then present the results of the separation of the signal from Fig. 1 back to its source components.

3. Decomposition of signals based on statistical parameters

Vibration signals can be analyzed with methods developed for stochastic signals as they share a large part of their features. Based on the Wold theorem, we can present each random process X(n)—also a vibration signal—as a unique sum of two parts: deterministic p(n) and nondeterministic r(n)

$$X(n) = p(n) + r(n), \tag{2}$$

where r(n) can be presented as

$$r(n) = \sum_{k=1}^{\infty} b_k \varepsilon_{n-k} + \varepsilon_n, \tag{3}$$

 ε being a Gaussian random process and b a vector of prediction coefficients. The component p(n) can then be ideally (i.e., with a zero error) presented as

$$p(n) = \sum_{k=1}^{\infty} a_k p(n-k), \qquad (4)$$

where a is a vector of prediction coefficients.

In (2), p(n) is a deterministic process and r(n) is a stationary random process with zero mean. The process p(n) can be exactly predicted if its previous values are known. The prediction error of p(n) is theoretically zero and does not depend on time. The process r(n) has always a prediction error which will grow with the prediction distance. This feature is the basis for separation, as when two samples of a signal taken with sufficient distance in time



Fig. 1. Simulated signal containing a 100 Hz sine wave with decaying components (left) and a plot of decaying components only, simulating impacts from the rolling bearing fault (right). Note different scales on both plots.



Fig. 2. Fragment of the spectrum of the signal from Fig. 1 without the jitter in the period of bearing fault components (left) and with 2% jitter (right).

will be correlated, the correlation of the nondeterministic part r(n) will be zero and will return the deterministic part p(n). When this idea is applied to machinery vibration, we could decompose each vibration signal into the deterministic and random parts. It has very important practical implications, as we can expect that the deterministic part would be predominantly strictly harmonic excitations (i.e., shafts fundamental frequencies and signals excited by gears). The nondeterministic part would then be noise and excitations from bearings, as they have a significant random factor, as shown in the previous section.

Solving the problem is actually very close to the estimation of adaptive filter parameters. Such a filter is presented in Fig. 3. The goal of the filter is to modify coefficients of the filter H(z), so that the estimated signal x(n) is as close to the reference signal d(n) as possible. The most common goal function is the least mean square.

Adaptive filters need to be stable over as large a span of coefficients as possible. This condition leads to the choice of FIR structures. For the FIR filter, the transfer



Fig. 3. Architecture of the adaptive filter.

function of the filter (given in the z domain) is

j

$$H_n(z) = \sum_{k=0}^{N-1} h_k(n) z^{-k},$$
(5)

where $H_n(z)$ is the transfer function H(z) at the moment $n, h_k(n)$ are filter coefficients at the moment n. Note that $H_n(z)$ contains only the denominator. Thus, the FIR filter

329 amcs

330

output y(n) can be presented as

$$y(n) = \sum_{k=0}^{N-1} h_k(n)x(n-k).$$
 (6)

The least square goal function will then be

$$J(h(n)) = E\left[\left(d(n) - \sum_{k=0}^{N-1} h_k(n)x(n-k)\right)^2\right].$$
 (7)

After calculations, which can be found in (Zieliński, 2007), we obtain the classical Wiener filter equation:

$$h^{\text{opt}} = \left[R_{xx}^{(n)} \right]^{-1} r_{dx}^{(n)},$$
 (8)

where $R_{xx}^{(n)}$ is the autocorrelation matrix of the signal x(n), and $r_{dx}^{(n)}$ is the cross correlation vector of the signals d(n) and x(n-i). In practice, the required order of the filter is very high and Eqn. (8) is not solved directly. Instead, recursive methods are used. The modification of h(n) should be proportional to the gradient of the multidimensional cost function, but with a reversed sign. In general, this formula can be presented as

$$h(n+1) = h(n) + \Delta h(n) = h(n) - \mu \nabla_n,$$
 (9)

where μ is a coefficient deciding about the speed of iteration. If μ is sufficiently small (see (Ho and Randall, 2000)), then Eqn. (8) gives the optimal solution for h(n).

The theory presented above is the basis for the solution of the problem given in (2). In our case we do not have the reference signal. We will rely instead on the length of correlation, as described in the beginning of this section. Such an approach will also modify the architecture of the filter, which will become a self-adaptive filter, as presented in Fig. 4. Now, we will use the signal x(n) as



Fig. 4. Architecture of a self-adaptive filter.

the reference signal, but with a delay, which will cause the autocorrelation of r(n) to be close to zero. Introduction to the theory is presented below. More details can be found in (Antoni and Randall, 2004).

As for (2), we are looking for the predictor of X(n), which can be formulated as the conditional expectation:

$$X(n) = E\{X(n)|X(n-\Delta), \dots, X(n-\Delta-N+1)\},$$
 (10)

where X(n) is predicted from Δ last samples. The parameter Δ should be such that r(n) is not autocorrelated with this distance, i.e.,

$$R_{rr}(m) = 0. \tag{11}$$

Equation (10) should hold for all $|m| > \Delta$. With such preconditions, we can state that

$$\hat{p}(n) = \hat{X}(n), \tag{12}$$

$$\hat{r}(n) = X(n) - \hat{X}(n),$$
 (13)

which are the best estimators (related to the mean square error). We know that X(n) is periodic, so it can be formulated as a sum

$$X(n) = \sum_{k=0}^{N-1} h_k X(n - \Delta - k).$$
 (14)

Thus, we are interested in coefficients h_k which will minimize the mean square prediction error. The solution is again given by the Wiener filter as the solution of the Wiener-Hopf equation:

$$\sum_{k=0}^{N-1} h_k R_{xx}(m - \Delta - k) = R_{xx}(m).$$
(15)

In practice, the filter order is very high, up to several thousands, which makes the set of equations (14) difficult to solve. Recursive solutions are used for the determination of h_i . The method is very similar to (8) and can be presented as

$$h_k^{n+1} = h_k^n + \mu e(n) X(n - \Delta - k).$$
 (16)

The optimal choice of algorithm parameters is very important for the self-adaptive filter. Those parameters include

- filter length: N,
- time delay: Δ ,
- forgetting factor: μ.

The filter length N must be sufficient to extract all important periodic factors from the original signal. In theory, when there is no noise, an order of 2k is required to extract k sinusoidal components. When noise is present, higher orders are required for its rejection. This issue is discussed in detail in (Antoni and Randall, 2004) and, as a conclusion, the following statement can be cited: "So in general, the choice of N faces a compromise between sufficient selectivity of the frequency response on the one hand, and convergence and small estimation bias on the other hand".

The time delay Δ should theoretically be as long as possible. In practice, sampled signals have finite length

331 **AMCS**

and there may be some deviations from the ideal period of deterministic components, due to speed variations (which will be addressed later in the paper) and measurement errors. So, Δ should be slightly larger than the correlation length of the nonperiodic component.

The setting of the forgetting factor μ is theoretically analyzed in (Haykin, 1996). The following formula should be used:

$$0 < \mu < \frac{2}{\lambda_{\max}},\tag{17}$$

where $\lambda_{\max} < \operatorname{trace}(R_{xx}^{(n)})$

In the next sections, a self-adaptive algorithm will be applied first to simulated and then to real data from a wind turbine.

4. Decomposition of simulated data

The described self-adaptive signal decomposition algorithm was developed in Matlab using the Filter Design Toolbox. The algorithm is illustrated in Fig. 5.



Fig. 5. Flowchart of the self-adaptive filter with Matlab commands.

First, the algorithm prepares vectors for the signals x and d (delayed x). It is required that both signals have an equal number of samples, which are afterwards processed simultaneously. Next, the filter object is created. At this point, the object receives the basic parameters, i.e., the block length, the filter length and the forgetting factor. To check the convergence of the filter, the stepmax method is executed. The forgetting factor should be at least two times smaller than the returned parameter *mumax* (Matlab, 2006). Finally, the filter is applied to the data and results are passed to the Matlab workspace. The FD Toolbox contains more filters, some of which are much faster than LMS. If possible, Block LMS or FFT-based block LMS should be used (Shynk, 1992).

The presented algorithm was applied to simulated data, which were described in Section 2 (see Fig. 1). Al-

gorithm parameters were chosen according to the section above and are presented in Table 1.

Table 1. Filter parameters chosen for simulated data and real

data.		
Filter	Value for	Value for
parameter	simulated data	real data
Filter length	64	512
Time delay	12000	12000
Forgetting factor	0.001	0.0005

The filter length was set to 64, as only one sinusoid was to be tracked but sufficient noise rejection was required. The time delay is given in samples (for a signal sampled at 25 kHz), which is more than the distance between exponential components which were added with a random jitter (see Figs. 1 and 2). Figure 6 presents filter outputs. For clarity of the figure, the original signal contained no noise. In the case with noise, the results were very similar.

It can be clearly seen that the self-adaptive filter is capable of decomposing the signal into its original components. The setting time is very short and this does not change with the addition of (relatively small of standard deviation equal to 0.1) white noise.

5. Decomposition of real data

To test the algorithm, it was run on real data from a wind turbine, which experienced an outer ring fault on the generator bearing. Figure 7 presents the setup of the turbine with the location of vibration sensors.



Fig. 7. Mechanical structure of the wind turbine. The location of vibration measurement sensors is shown by "An" symbols.

The main rotor with three blades is supported by the main bearing and transmits the torque to the planetary gear. The planetary gear input is the plate to which the main rotor is connected. The planetary gear has three planets, with their shafts attached to the plate. The planets roll over the stationary ring and transmit the torque to the sun. The sun shaft is the output of the planetary gear. Further, the sun drives the two-stage parallel gear. The paral-



Fig. 6. Filter output and the error signal of the self-adaptive filter working on simulated data.

lel gear has three shafts: the slow shaft connected to the sun shaft, the intermediate shaft and the fast shaft, which drives the generator. The generator produces AC current of slightly varying frequency. This current is converted first into DC power and then into AC power of frequency equal to that of the grid. Electric transformations are performed by the controller at the base of the tower.

332

amcs

The vibration data were acquired by an online vibration monitoring system at a 25 kHz sampling rate. One single record was sampled for 10 s, i.e., contained 250e3 samples. Figure 8 presents an example of a vibration signal and its spectrum. The structure of the signal can be analyzed on its spectrum. One can observe a complex structure, composed of several sharp lines and a few structural resonances. The signal also contains a significant noise component, which will make fault detection more difficult.

The bearing fault occurred in the bearing at the nondriven end of the generator, i.e., data from sensor A6 were investigated. During the sampling session, the generator speed was 1346.5 rpm (22.44 Hz). According to the bearing manufacturer, BPFO (Bearing Pass Frequency Outer ring) is 5.10X, which at this rotational speed gives 114.4 Hz.

To ensure the autocorrelation of periodic components, the vibration signal must be acquired when the machine has exactly equal rotational speed. Even small variations would smear spectral lines and cause the decorrelation of periodic components. Small variations in the rotational speed are common in machinery and in the case of wind turbines rpm changes in a great range (1000 to 1500 rpm in the case of the investigated wind turbine). To make sure that the SANC algorithm can be applied in such cases, the signal must be resampled to the order domain, where there is always a constant number of samples per one rotation of the shaft. An algorithm for such resampling based on software was proposed in (Barszcz, 2004), where its accuracy was also discussed and compared with the hardware-based algorithm. Another possible method for estimating rpm variation is time-frequency analysis, e.g., STFT proposed by Lee and White in 1998.

In the investigated case, the data were tested against the variability of rotational speed. It was found that the variability was around 0.2%, which is much less than the typical jitter in rolling bearing parameters (around 1%). Thus, resampling was not performed in this case. Next, the self-adaptive algorithm was applied to the data. The algorithm parameters are presented in Table 1. To decrease computational complexity, the block LMS algorithm was used. Figure 9 presents filter outputs.

In this case it is harder to evaluate the performance of the filter, although it can be seen that some periodic components were filtered. To gain a better insight, Fig. 10 presents a comparison of spectra of the original and filtered signals. Only a part of the spectrum is presented to keep the figure clear.



Fig. 10. Comparison of spectra of the original signal (grey) and that filtered with the adaptive filter (black).

One can observe that the sharp lines, correspond-



Fig. 8. Waveform and spectrum of the vibration signal from the faulty bearing. The vertical axis is in g. The complex structure of the signal is visible.



Fig. 9. Filter output and the error signal of the self-adaptive filter working on real data.

ing to the periodic components, passed through the filter while the background noise was rejected up to 30 dB. This proves the good performance of the filter. The ability of the filter to extract all the periodic components was not checked, as it was not the goal of the present research. Instead, the error signal should be investigated to see whether it contains impact type components.

Visual inspection of the error signal presented in Fig. 9 (left) does not show a clear pattern of impact type impulses, as one could expect. Such a situation was caused by a large noise component, as mentioned earlier. The analysis of the error signal spectrum (Fig. 11) reveals the presence of a spectral component at a frequency of 115 Hz, but the component is very small and accompanied by several other lines. On the other hand, such behavior is expected, as we investigate exponential impulses excited at 115 Hz. In such a case, only a very small portion of signal energy is within its fundamental frequency. Another method should be used to verify whether the error signal contains the sought fault signature.

The investigation of the spectrum in its full range

showed a strong structural resonance of around 7800 Hz. In such cases, the most efficient method to extract bearing fault induced signals is the demodulation of the signal and the investigation of its envelope. Such a method is described by Ho and Randall (2000). After this approach was applied, envelope signal spectra were calculated for both the filtered and error signals and presented in Fig. 12.

In Fig. 12, for the error signal, the BPFO frequency (115 Hz) is very strong and its second harmonic can be also seen, though it is much smaller. The filtered signal, shown on the left, does not contain any bearing fault related components. Thus, it was shown that the non-deterministic part of the signal, i.e., the error signal from the adaptive filter contains the information about the outer race fault.

6. Conclusion

The paper investigated the possibility of decomposing vibration signals into deterministic and nondeterministic parts, based on the Wold theorem. A short presentation of the theory of adaptive filters was given. The idea of a

333 amcs



Fig. 11. Spectra of the filter error zoomed to 500Hz (left) and in full range (right). A very small component can be seen at the BPFO. Strong structural resonance is visible at around 7800 Hz.



Fig. 12. Comparison of spectra of the filtered signal (left) and the error signal (right), both demodulated around the resonance at 7800 Hz. In the error signal the BPFO (115 Hz) dominates the spectrum. Its second harmonic can be also seen, but is much smaller.

self-adaptive filter (in the literature referred to as SANC or ALE) was presented and its most important features were discussed. The self-adaptive filter was first applied to simulated data and then to a real vibration signal from a wind turbine with an outer race fault. In both cases, the filter showed a very good ability to decompose the signal.

The method should be used on signals resampled to the period of the shaft. If the original vibration signals were to be analyzed, natural variations of rotational speed would in most cases cause the method to be useless. The important contributions of the work include

- implementation of the SANC filter in Matlab,
- proposal of the simulation of a faulty signal,
- application of the method to wind turbine vibration,
- discussion of a real case with a high noise level.

In the future, more research will be aimed at the optimization of the method. This will be based on the

optimization of filter parameters and the algorithm itself. Eventually, it should be investigated whether and how the method can be implemented in commercial monitoring and diagnostic systems. It seems that it could yield promising results for systems monitoring machines, where there are several spectral components masking characteristic frequencies of rolling bearings. Due to long filter lengths, the system should have enough CPU time. Thus, it is unlikely to be used in continuous protection systems, e.g., for large turbosets. On the other hand, systems designed, e.g., for wind turbines sample vibration signals periodically, not spending the majority of time on signal processing, but leaving sufficient CPU time for the proposed advanced analysis.

References

Antoni, J. and Randall, R.B. (2004). Unsupervised noise cancellation for vibration signals: Part I—Evaluation of adaptive algorithms, *Mechanical Systems and Signal Processing* 18(1): 89–101.



- Barszcz, T. (2004). Proposal of new method for mechanical vibration measurement, *Metrology and Measurement Systems* 11(4): 409–421.
- Chaturvedi, G.K. and Thomas, D.W. (1981). Adaptive noise cancelling and condition monitoring, *Journal of Sound and Vibration* **76**(3): 391–405.
- Haykin, S. (1996). *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ.
- Ho, D. and Randall, R.B. (2000). Optimisation of bearing diagnostic techniques using simulated and actual bearing faults, *Mechanical Systems and Signal Processing* 14(5): 763– 788.
- Lee, S.K. and White, P.R. (1998). The enhancement of impulsive noise and vibration signals for fault detection in rotating and reciprocating machinery, *Journal of Sound and Vibration* **217**(3): 485–505.
- Shao, Y. and Nezu, K. (2005). Design of mixture de-noising for detecting faulty bearing signals, *Journal of Sound and Vibration* 282(4): 899–917.
- Shynk, J.J. (1992). Frequency-domain and multirate adaptive filtering, *IEEE Signal Processing Magazine* **9**(1): 14–37.
- The MathWorks (2006). *Filter Design Toolbox*, Matlab documentation, Mathworks V4.2 /R2008a/.
- Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E. and Goodlin, R. C. (1975). Adaptive noise cancelling: Principles and applications, *Proceedings of the IEEE* 63(12): 1692–1716.

- Zeidler, J., Satorius, E., Chabries, D. and Wexler, H. (1978). Adaptive enhancement of multiple sinusoids in uncorrelated noise, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26(3): 240–254.
- Zieliński, T.P. (2007). *Digital Signal Processing. From Theory* to Applications, WKŁ, Warsaw, (in Polish).
- Żółtowski, B. and Cempel, C. (Eds.) (2004). Engineering of Machinery Diagnostics, PTDT ITE PIB Radom, Warsaw/Bydgoszcz/Radom, (in Polish).



Tomasz Barszcz works at the Department of Robotics and Mechatronics of the AGH University of Science and Technology in Cracow, Poland. He is interested in architectures of machinery condition monitoring systems and technical state assessment methods. He is the author of four books and over 60 papers in this field. Technologies developed by Dr. Barszcz have been implemented in various industries including power generation, oil and gas, printing,

railway transport and automotive.

Received: 15 September 2008 Revised: 2 February 2009 Re-revised: 8 February 2009