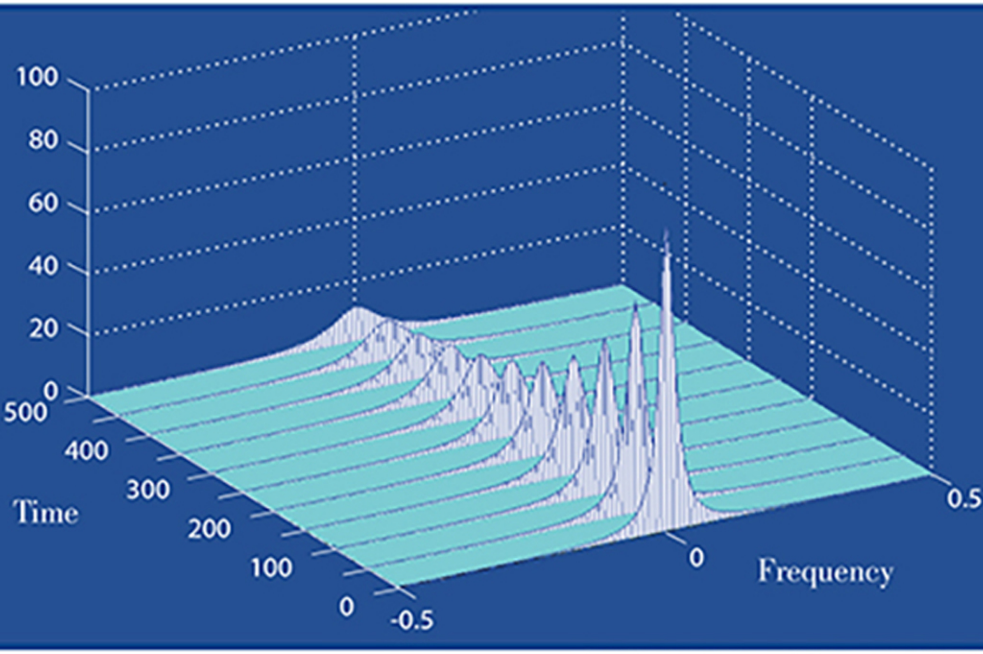
Volume III

FUNDAMENTALS OF STATISTICAL SIGNAL PROCESSING

PRACTICAL ALGORITHM DEVELOPMENT



STEVEN M. KAY

CD-ROM INCLUDED

Fundamentals of Statistical Signal Processing, Volume III

In summary, signal model selection proceeds as follows.

1. Choose a general signal model *type*, such as the sinusoid model given by

$$s[n] = \sum_{i=1}^{p} A_i \cos(2\pi f_i n + \phi_i) \quad n = 0, 1, \ldots, N-1.$$

2. Choose the model order p, i.e., the number of sinusoids.

3. Choose the values of the parameters $\{A_1, \ldots, A_p, f_1, \ldots, f_p, \phi_1, \ldots, \phi_p\}$ either as fixed constants to be used in a fixed algorithm or to be determined on-line in a data-adaptive algorithm.

In general, the model accuracy will depend upon its complexity. Using more parameters leads to a more detailed and potentially more accurate model. However, if those model parameters need to be estimated from data and thus are subject to statistical error, then a less accurate model may result if too many parameters are included. Hence, model order selection is an especially important task and presents itself as somewhat of a "balancing act". An analogy might be the goal of modeling the exterior of an automobile for visual identification. As "data" one is shown a *particular* automobile. An *adequate* model would be one that incorporates four wheels, some windows, and a box-like shape, much as what a child would draw. A more *precise* model would also include the chrome wheels, the spoiler and other adornments. But the latter could hardly be counted upon to identify a general class of automobiles. The latter more detailed model has too many "parameters" to allow it to be a good *general* model.

The selection of a signal model is aided by first utilizing knowledge of any physical constraints, and second, by analysis of field data. The first consideration is to guide the model selection by the physics of the signal generation process. For example, consider the signal modeling of the acoustic emissions of a helicopter. The manufacturer may have physical knowledge of the acoustic harmonic frequencies of a helicopter blade operating at a given speed, which has been established from previous testing and development studies. As already mentioned, for best algorithm performance any prior knowledge that can be *reliably* assumed should be incorporated into the algorithm as a *constraint*. The second consideration, analysis of field data, is a critical one in that it lends insight into the task at hand and can provide test data for preliminary performance analysis of the algorithm. Note that once again the same type of analysis of the field data that is used to formulate a model, may in fact be latter implemented in the actual algorithm if it is to be data-adaptive.

5.2 Signal Modeling

We next provide an overall approach to signal modeling, which is guided by a "roadmap".

5.2.1 A Roadmap

Table 3.2 given in Chapter 3 listed some useful deterministic signal models. We now describe an explicit procedure for choosing a signal model and then illustrate it with an example. The overall approach is summarized by the "roadmap" given in Figure 5.1. The selection process proceeds as follows.

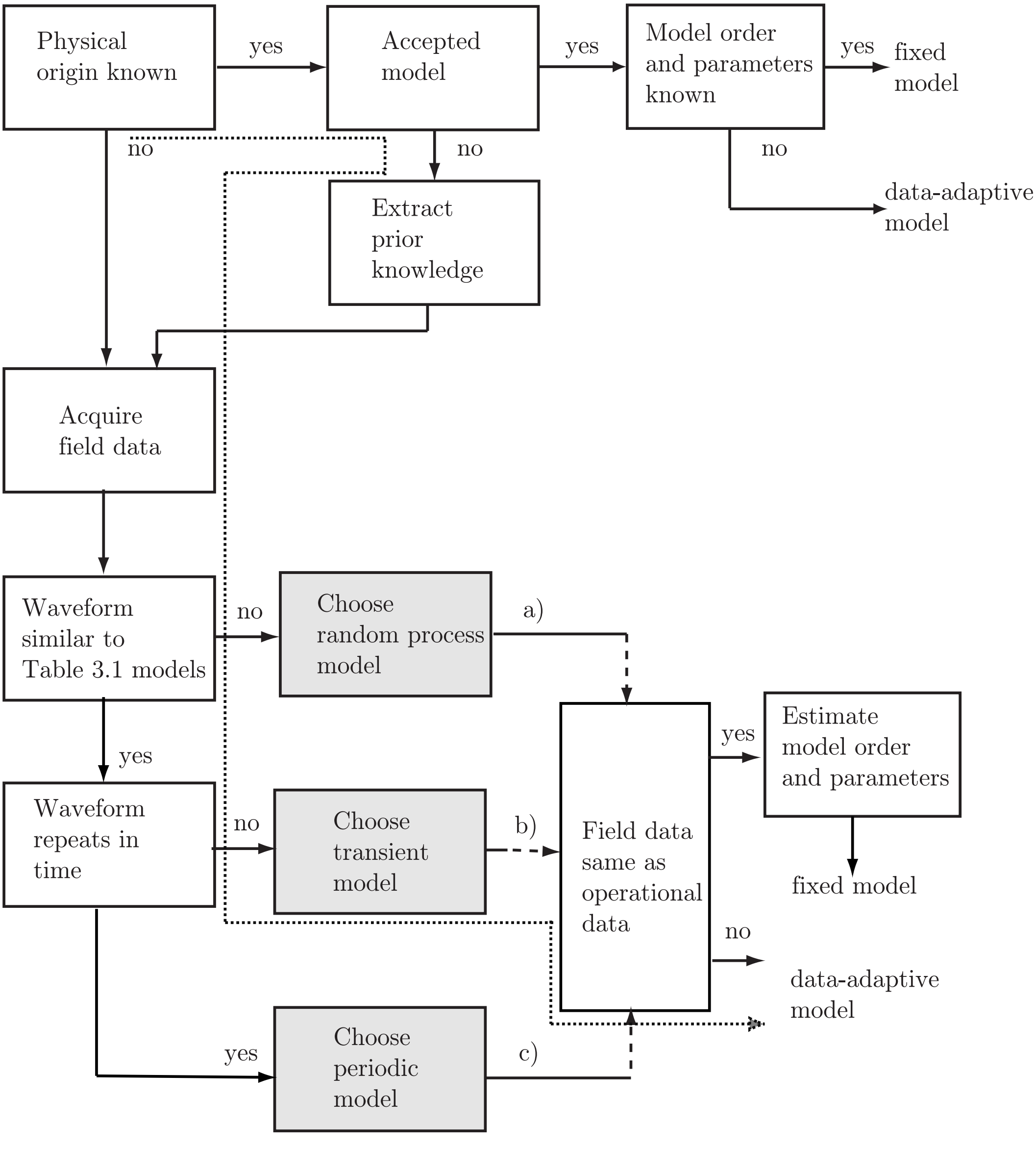
We first determine if the physical origin of the signal is known and if there is an accepted model. If this is the case, then we use the model as is, with known values for the model order and parameters. This produces a *fixed* signal model. If, however, the model order and/or the parameters are unknown and may vary from data set to data set, then we will need to estimate them on-line, leading to a *data-adaptive* model.

If the physical origin of the signal is unknown, or if it is known but the model is not sufficiently accurate, we next obtain some field data for study and analysis. Note that even if an acceptable model is not available, we can usually still extract information from the problem physics and use it later on to help in model selection. Having acquired field data, we ask if the waveform appears to be of a deterministic nature (a readily recognizable time series pattern) or more random in nature. If the latter, we choose a random process model from Table 4.1. If the random process model parameters are known or can be estimated and assumed not to vary from data set to data set, then we are done. This produces a fixed model. If not, then we will have to estimate them on-line to produce a data-adaptive signal model.

Much the same procedure is used when the signal waveform has a definite pattern, i.e., similar to the mathematical models given in Table 3.1. The two cases of interest are waveforms that repeat in time and those that do not. If the latter, we choose either damped exponentials, damped sinusoids, a phase modulated signal, or a polynomial. If it repeats with a stable pattern, i.e., is periodic in time, we choose either a set of harmonically related sinusoids (a Fourier series representation model) or a temporal signal with known or unknown signal samples within the basic period. As before, we use a fixed model if the model order and parameters are known, or are estimated and do not change from the field data used to estimate them to the operational data. Otherwise, we use a data-adaptive model.

5.3 An Example

Consider the problem of detecting a fault in the bearing of a precision machine [Liu 2008]. We describe the decision-making path, which is shown in Figure 5.1 as the dashed line. The physical origin of the fault is generally known, i.e., it can be due to inadequate lubrication, metal fatigue, etc., but translating that knowledge into a suitable model for the signal at the output of an accelerometer is not possible. This is mainly due to the many possible operating conditions of the machine. For example, the waveform will be dependent upon the speed of the motor which itself depends



a) See Table 4.1
b) damped exponentials, damped sinusoids, phase modulated signal, or polynomial
c) harmonic sinusoids or periodic time signal

Figure 5.1: Roadmap for signal model selection.

upon a load that is changing with time. Nonetheless, from past experience it has been observed that the fault waveform, when it appears, has a spiky nature that is relatively constant in shape but whose amplitude varies with each occurrence. The spikes are nearly periodic but this assumption is violated when the machine changes speed, which is again load dependent, and cannot be predicted. An example might be as shown in Figure 5.2.

It is seen that the spike amplitudes are different and the spacing appears to increase with time, indicating a decrease in bearing speed. Hence, the waveform does not repeat, prompting the choice of a *transient* model. If we examine a single spike occurring at time $t = 0.125$ sec, we observe the sampled waveform shown in Figure 5.3.

It has been sampled at 10,000 samples/sec so that a time interval of 0.01 seconds yields $N = 100$ samples as shown. The waveform now appears to be a damped sinusoid or sum of damped sinusoids with some additive noise. In fact, by plotting its periodogram, which is defined as

$$I(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} s[n]\exp(-j2\pi f n)\right|^2 \tag{5.1}$$

and is the normalized and magnitude-squared discrete-time Fourier transform, we observe in Figure 5.4 two resonances and therefore, choose a damped sinusoid model with $p = 2$ as

$$s[n] = \sum_{i=1}^{2} A_i r_i^n \cos(2\pi f_i n + \phi_i) \quad n_0 \le n \le n_0 + N - 1 \tag{5.2}$$

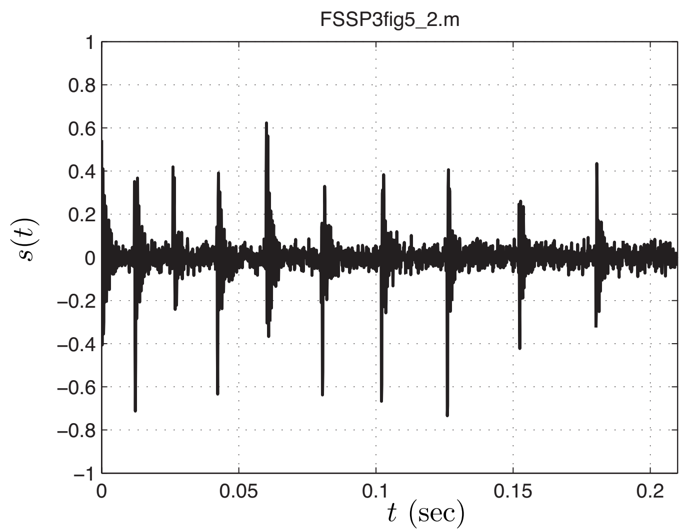


Figure 5.2: Example of accelerometer output waveform indicating a bearing fault.

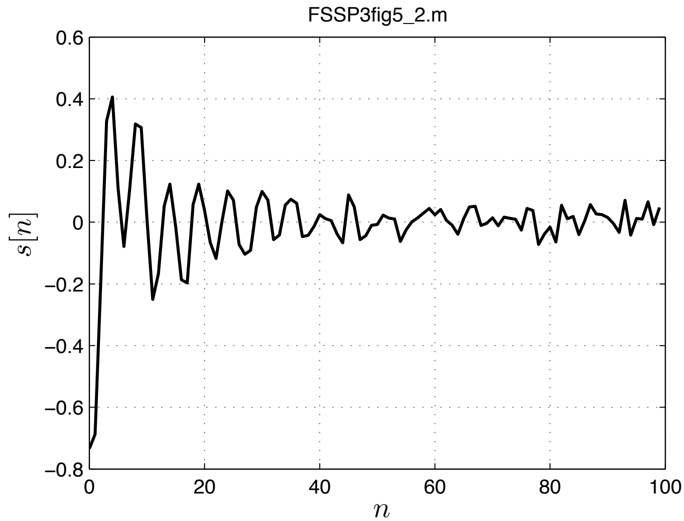


Figure 5.3: Expanded version of sampled accelerometer output waveform showing a single spike starting at $t = 0.125$ sec. The sample points have been connected for easier viewing.

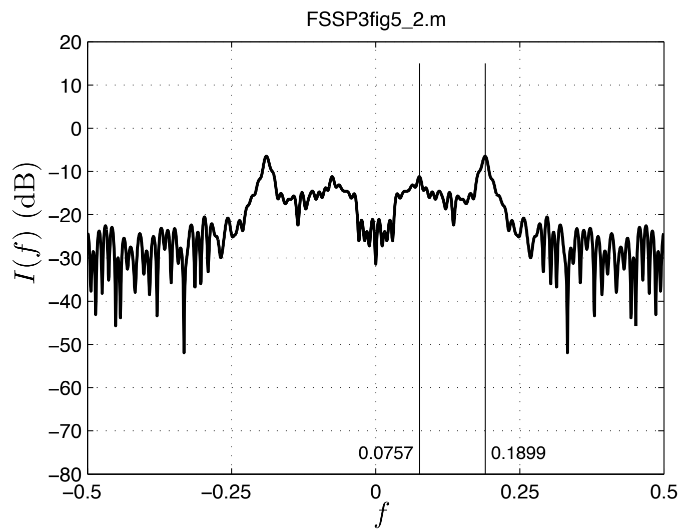


Figure 5.4: Periodogram of time-sampled spike waveform shown in Figure 5.3. The two local maxima, termed resonances, are indicated by the vertical lines. Their locations are also indicated.

where n_0 is the index of the first signal sample of the chosen spike, and N is the length of the signal.

Exercise 5.1 – Length of signal
For a single damped sinusoid $s[n] = Ar^n \cos(2\pi f_0 n + \phi)$ with $A = 2$, $r = 0.95$, $f_0 = 0.05$, and $\phi = 0$, plot the signal $s[n]$ versus n for $n = 0, 1, \ldots, 199$. From your plot determine N, which is the essential duration of the signal. Assuming the phase is zero, a rule of thumb is that the signal is essentially zero if

$$\left| \frac{s[N]}{s[0]} \right| < 0.001.$$

Thus, the signal length N can be found. Show that for a single damped sinusoid N is given explicitly as

$$N = \frac{\ln 0.001}{\ln r}.$$

How does this compare with your visual determination of the signal length? How could you extend this rule of thumb to the case when $\phi \neq 0$?

•

Exercise 5.2 – Starting time of signal
A simple method for determining the starting time of a signal, based on noise-corrupted data $x[n]$, is to implement a *sliding window power detector* (see also Section 4.5). One chooses the signal start time as the time when the average power

$$T[n] = \frac{1}{L+1} \sum_{k=n-L/2}^{n+L/2} x^2[n]$$

exceeds a threshold γ. Load the sampled data of the waveform plotted in Figure 5.2 by using the MATLAB command `load FSSP3exer5_2`. The sampling rate is 10,000 samples/sec. Implement $T[n]$ using a window length of $L + 1 = 11$ and plot $T[n]$ versus n for $5 \leq n \leq 2094$. Can you determine the starting times of the spikes?

•

Since the amplitudes A_i's and phases ϕ_i's and time interval between spikes cannot be determined in advance (note how they change from spike to spike in Figure 5.2), we will use (5.2) as a signal model for a data-adaptive algorithm. The unknown parameters will need to be estimated within the algorithm. From prior experience the damping factors r_1, r_2 and resonance frequencies f_1, f_2 are known not to vary so that we can estimate them from the data given in Figure 5.3 and then fix their values in the model. The amplitudes and phases, however, will need to be estimated in a data-adaptive algorithm. Using this data the parameter estimates obtained (the exact method to be described shortly) are

$$r_1 = 0.86 \quad f_1 = 0.08 \quad r_2 = 0.91 \quad f_2 = 0.18. \tag{5.3}$$