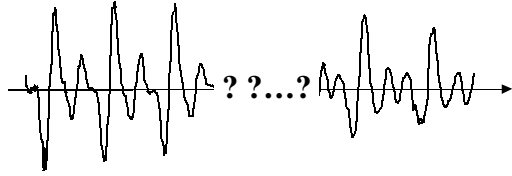


10



INTERPOLATION

- 10.1 Introduction
- 10.2 Polynomial Interpolation
- 10.3 Model-Based Interpolation
- 10.4 Summary

Interpolation is the estimation of the unknown, or the lost, samples of a signal using a weighted average of a number of known samples at the neighbourhood points. Interpolators are used in various forms in most signal processing and decision making systems. Applications of interpolators include conversion of a discrete-time signal to a continuous-time signal, sampling rate conversion in multirate communication systems, low-bit-rate speech coding, up-sampling of a signal for improved graphical representation, and restoration of a sequence of samples irrevocably distorted by transmission errors, impulsive noise, dropouts, etc. This chapter begins with a study of the basic concept of ideal interpolation of a band-limited signal, a simple model for the effects of a number of missing samples, and the factors that affect the interpolation process. The classical approach to interpolation is to construct a polynomial that passes through the known samples. In Section 10.2, a general form of polynomial interpolation and its special forms, Lagrange, Newton, Hermite and cubic spline interpolators, are considered. Optimal interpolators utilise predictive and statistical models of the signal process. In Section 10.3, a number of model-based interpolation methods are considered. These methods include maximum a posteriori interpolation, and least square error interpolation based on an autoregressive model. Finally, we consider time–frequency interpolation, and interpolation through searching an adaptive signal codebook for the best-matching signal.

10.1 Introduction

The objective of interpolation is to obtain a high-fidelity reconstruction of the unknown or the missing samples of a signal. The emphasis in this chapter is on the interpolation of a *sequence* of lost samples. However, first in this section, the theory of ideal interpolation of a band-limited signal is introduced, and its applications in conversion of a discrete-time signal to a continuous-time signal and in conversion of the sampling rate of a digital signal are considered. Then a simple distortion model is used to gain insight on the effects of a sequence of lost samples and on the methods of recovery of the lost samples. The factors that affect interpolation error are also considered in this section.

10.1.1 Interpolation of a Sampled Signal

A common application of interpolation is the reconstruction of a continuous-time signal $x(t)$ from a discrete-time signal $x(m)$. The condition for the recovery of a continuous-time signal from its samples is given by the Nyquist sampling theorem. The Nyquist theorem states that a band-limited signal, with a highest frequency content of F_c (Hz), can be reconstructed from its samples *if* the sampling speed is greater than $2F_c$ samples per second. Consider a band-limited continuous-time signal $x(t)$, sampled at a rate of F_s samples per second. The discrete-time signal $x(m)$ may be expressed as the following product:

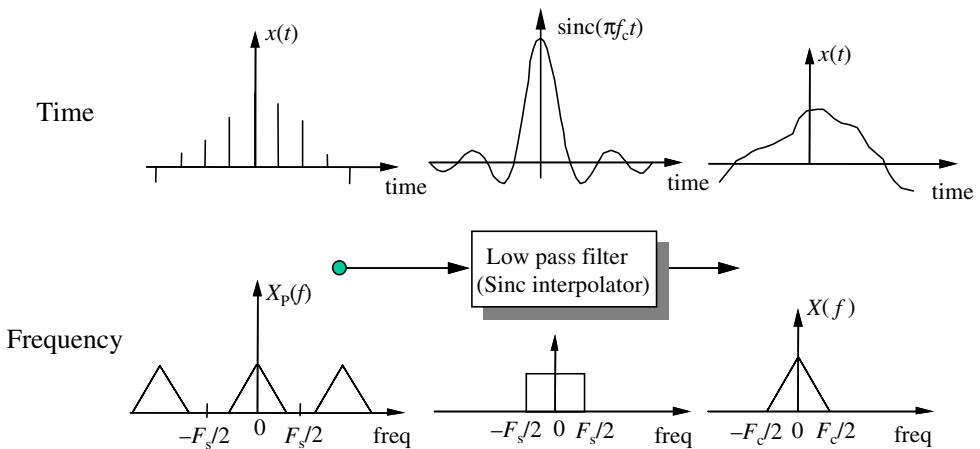


Figure 10.1 Reconstruction of a continuous-time signal from its samples. In frequency domain interpolation is equivalent to low-pass filtering.

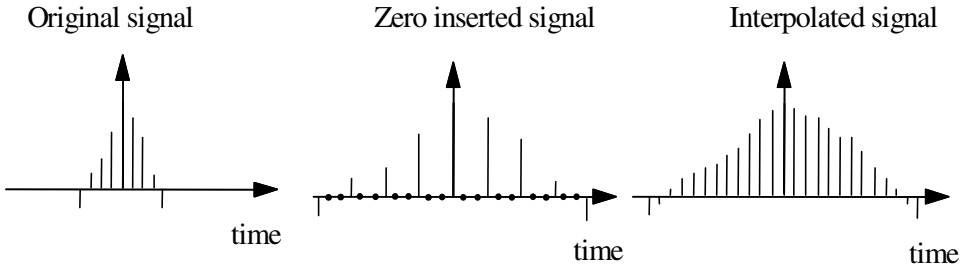


Figure 10.2 Illustration of up-sampling by a factor of 3 using a two-stage process of zero-insertion and digital low-pass filtering.

$$x(m) = x(t) p(t) = \sum_{m=-\infty}^{\infty} x(t) \delta(t - mT_s) \quad (10.1)$$

where $p(t) = \sum \delta(t - mT_s)$ is the sampling function and $T_s = 1/F_s$ is the sampling interval. Taking the Fourier transform of Equation (10.1), it can be shown that the spectrum of the sampled signal is given by

$$X_s(f) = X(f) * P(f) = \sum_{k=-\infty}^{\infty} X(f + kf_s) \quad (10.2)$$

where $X(f)$ and $P(f)$ are the spectra of the signal $x(t)$ and the sampling function $p(t)$ respectively, and $*$ denotes the convolution operation. Equation (10.2), illustrated in Figure 10.1, states that the spectrum of a sampled signal is composed of the original base-band spectrum $X(f)$ and the repetitions or images of $X(f)$ spaced uniformly at frequency intervals of $F_s = 1/T_s$. When the sampling frequency is above the Nyquist rate, the base-band spectrum $X(f)$ is not overlapped by its images $X(f \pm kF_s)$, and the original signal can be recovered by a low-pass filter as shown in Figure 10.1. Hence the ideal interpolator of a band-limited discrete-time signal is an ideal low-pass filter with a sinc impulse response. The recovery of a continuous-time signal through sinc interpolation can be expressed as

$$x(t) = \sum_{m=-\infty}^{\infty} x(m) T_s f_c \text{sinc}[\pi f_c (t - mT_s)] \quad (10.3)$$

In practice, the sampling rate F_s should be sufficiently greater than $2F_c$, say $2.5F_c$, in order to accommodate the transition bandwidth of the interpolating low-pass filter.

10.1.2 Digital Interpolation by a Factor of I

Applications of digital interpolators include sampling rate conversion in multirate communication systems and up-sampling for improved graphical representation. To change a sampling rate by a factor of $V=I/D$ (where I and D are integers), the signal is first interpolated by a factor of I , and then the interpolated signal is decimated by a factor of D .

Consider a band-limited discrete-time signal $x(m)$ with a base-band spectrum $X(f)$ as shown in Figure 10.2. The sampling rate can be increased by a factor of I through interpolation of $I-1$ samples between every two samples of $x(m)$. In the following it is shown that digital interpolation by a factor of I can be achieved through a two-stage process of (a) insertion of $I-1$ zeros in between every two samples and (b) low-pass filtering of the zero-inserted signal by a filter with a cutoff frequency of $F_s/2I$, where F_s is the sampling rate. Consider the zero-inserted signal $x_z(m)$ obtained by inserting $I-1$ zeros between every two samples of $x(m)$ and expressed as

$$x_z(m) = \begin{cases} x\left(\frac{m}{I}\right), & m=0, \pm I, \pm 2I, \dots \\ 0, & \text{otherwise} \end{cases} \quad (10.4)$$

The spectrum of the zero-inserted signal is related to the spectrum of the original discrete-time signal by

$$\begin{aligned} X_z(f) &= \sum_{m=-\infty}^{\infty} x_z(m) e^{-j2\pi f m} \\ &= \sum_{m=-\infty}^{\infty} x(m) e^{-j2\pi f m I} \\ &= X(I.f) \end{aligned} \quad (10.5)$$

Equation (10.5) states that the spectrum of the zero-inserted signal $X_z(f)$ is a frequency-scaled version of the spectrum of the original signal $X(f)$. Figure 10.2 shows that the base-band spectrum of the zero-inserted signal is composed of I repetitions of the based band spectrum of the original signal. The interpolation of the zero-inserted signal is therefore equivalent to filtering out the repetitions of $X(f)$ in the base band of $X_z(f)$, as illustrated in Figure 10.2. Note that to maintain the real-time duration of the signal the

sampling rate of the interpolated signal $x_z(m)$ needs to be increased by a factor of I .

10.1.3 Interpolation of a Sequence of Lost Samples

In this section, we introduce the problem of interpolation of a sequence of M missing samples of a signal given a number of samples on both side of the gap, as illustrated in Figure 10.3. Perfect interpolation is only possible if the missing samples are redundant, in the sense that they carry no more information than that conveyed by the known neighbouring samples. This will be the case if the signal is a perfectly predictable signal such as a sine wave, or in the case of a band-limited random signal if the sampling rate is greater than M times the Nyquist rate. However, in many practical cases, the signal is a realisation of a random process, and the sampling rate is only marginally above the Nyquist rate. In such cases, the lost samples cannot be perfectly recovered, and some interpolation error is inevitable.

A simple distortion model for a signal $y(m)$ with M missing samples, illustrated in Figure 10.3, is given by

$$\begin{aligned} y(m) &= x(m)d(m) \\ &= x(m)[1 - r(m)] \end{aligned} \quad (10.6)$$

where the distortion operator $d(m)$ is defined as

$$d(m) = 1 - r(m) \quad (10.7)$$

and $r(m)$ is a rectangular pulse of duration M samples starting at the sampling time k :

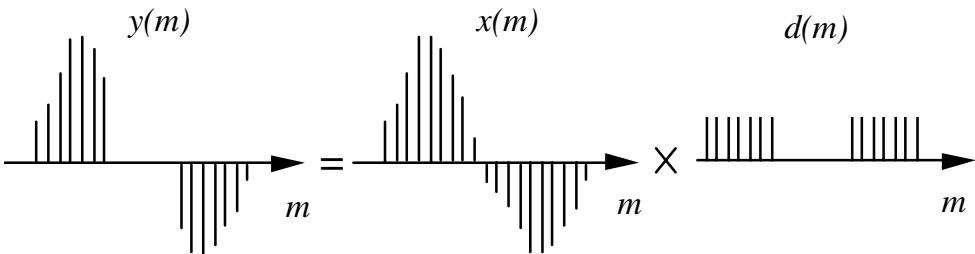


Figure 10.3 Illustration of a distortion model for a signal with a sequence of missing samples.

$$r(m) = \begin{cases} 1, & k \leq m \leq k + M - 1 \\ 0, & \text{otherwise} \end{cases} \quad (10.8)$$

In the frequency domain, Equation (10.6) becomes

$$\begin{aligned} Y(f) &= X(f) * D(f) \\ &= X(f) * [\delta(f) - R(f)] \\ &= X(f) - X(f) * R(f) \end{aligned} \quad (10.9)$$

where $D(f)$ is the spectrum of the distortion $d(m)$, $\delta(f)$ is the Kronecker delta function, and $R(f)$, the frequency spectrum of the rectangular pulse $r(m)$, is given by

$$R(f) = e^{-j2\pi f[k+(M-1)/2]} \frac{\sin(\pi f M)}{\sin(\pi f)} \quad (10.10)$$

In general, the distortion $d(m)$ is a non-invertible, many-to-one transformation, and perfect interpolation with zero error is not possible. However, as discussed in Section 10.3, the interpolation error can be minimised through optimal utilisation of the signal models and the information contained in the neighbouring samples.

Example 10.1 *Interpolation of missing samples of a sinusoidal signal.* Consider a cosine waveform of amplitude A and frequency F_0 with M missing samples, modelled as

$$\begin{aligned} y(m) &= x(m) d(m) \\ &= A(\cos 2\pi f_0 m) [1 - r(m)b] \end{aligned} \quad (10.11)$$

where $r(m)$ is the rectangular pulse defined in Equation (10.7). In the frequency domain, the distorted signal can be expressed as

$$\begin{aligned} Y(f) &= \frac{A}{2} [\delta(f - f_o) + \delta(f + f_o)] * [\delta(f) - R(f)] \\ &= \frac{A}{2} [\delta(f - f_o) + \delta(f + f_o) - R(f - f_o) - R(f + f_o)] \end{aligned} \quad (10.12)$$

where $R(f)$ is the spectrum of the pulse $r(m)$ as in Equation (10.9).

From Equation (10.12), it is evident that, for a cosine signal of frequency F_0 , the distortion in the frequency domain due to the missing samples is manifested in the appearance of sinc functions centred at $\pm F_0$. The distortion can be removed by filtering the signal with a very narrow band-pass filter. Note that for a cosine signal, perfect restoration is possible only because the signal has infinitely narrow bandwidth, or equivalently because the signal is completely predictable. In fact, for this example, the distortion can also be removed using a linear prediction model, which, for a cosine signal, can be regarded as a data-adaptive narrow band-pass filter.

10.1.4 The Factors That Affect Interpolation Accuracy

The interpolation accuracy is affected by a number of factors, the most important of which are as follows:

- (a) The predictability, or correlation structure of the signal: as the correlation of successive samples increases, the predictability of a sample from the neighbouring samples increases. In general, interpolation improves with the increasing correlation structure, or equivalently the decreasing bandwidth, of a signal.
- (b) The sampling rate: as the sampling rate increases, adjacent samples become more correlated, the redundant information increases, and interpolation improves.
- (c) Non-stationary characteristics of the signal: for time-varying signals the available samples some distance in time away from the missing samples may not be relevant because the signal characteristics may have completely changed. This is particularly important in interpolation of a large sequence of samples.
- (d) The length of the missing samples: in general, interpolation quality decreases with increasing length of the missing samples.
- (e) Finally, interpolation depends on the optimal use of the data and the efficiency of the interpolator.

The classical approach to interpolation is to construct a polynomial interpolator function that passes through the known samples. We continue this chapter with a study of the general form of polynomial interpolation, and consider Lagrange, Newton, Hermite and cubic spline interpolators. Polynomial interpolators are *not* optimal or well suited to make efficient use of a relatively large number of known samples, or to interpolate a relatively large segment of missing samples.

In Section 10.3, we study several statistical digital signal processing methods for interpolation of a sequence of missing samples. These include model-based methods, which are well suited for interpolation of small to medium sized gaps of missing samples. We also consider frequency–time interpolation methods, and interpolation through waveform substitution, which have the ability to replace relatively large gaps of missing samples.

10.2 Polynomial Interpolation

The classical approach to interpolation is to construct a polynomial interpolator that passes through the known samples. Polynomial interpolators may be formulated in various forms, such as power series, Lagrange interpolation and Newton interpolation. These various forms are mathematically equivalent and can be transformed from one into another. Suppose the data consists of $N+1$ samples $\{x(t_0), x(t_1), \dots, x(t_N)\}$, where $x(t_n)$ denotes the amplitude of the signal $x(t)$ at time t_n . The polynomial of order N that passes through the $N+1$ known samples is unique (Figure 10.4) and may be written in power series form as

$$\hat{x}(t) = p_N(t) = a_0 + a_1t + a_2t^2 + a_3t^3 + \dots + a_Nt^N \quad (10.13)$$

where $P_N(t)$ is a polynomial of order N , and the a_k are the polynomial coefficients. From Equation (10.13), and a set of $N+1$ known samples, a

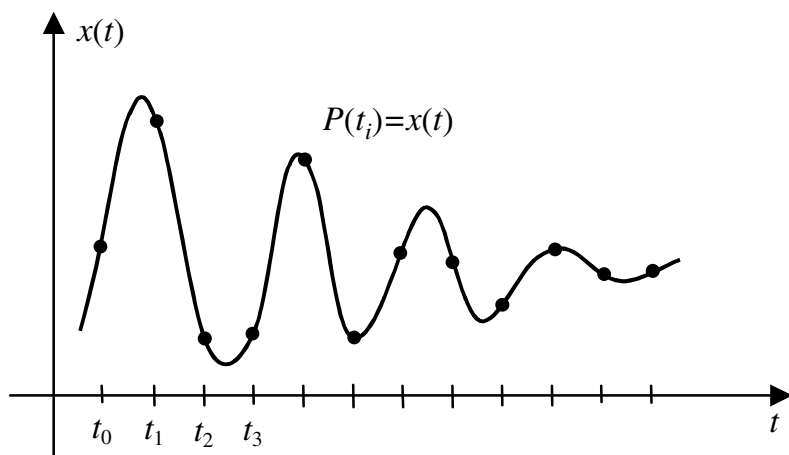


Figure 10.4 Illustration of an Interpolation curve through a number of samples.

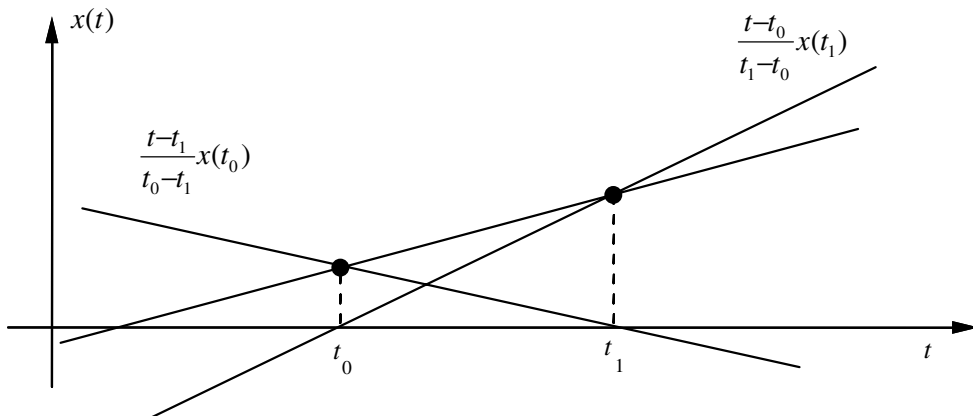


Figure 10.5 The Lagrange line interpolator passing through $x(t_0)$ and $x(t_1)$, described in terms of the combination of two lines: one passing through $(x(t_0), t_0)$ and the other through $(x(t_1), t_1)$.

The line Equation (10.16) may be rearranged and expressed as

$$p_1(t) = \frac{t-t_1}{t_0-t_1} x(t_0) + \frac{t-t_0}{t_1-t_0} x(t_1) \quad (10.17)$$

Equation (10.17) is in the form of a Lagrange polynomial. Note that the Lagrange form of a line interpolator is composed of the weighted combination of two lines, as illustrated in Figure 10.5.

In general, the Lagrange polynomial, of order N , passing through $N+1$ samples $\{x(t_0), x(t_1), \dots, x(t_N)\}$ is given by the polynomial equation

$$P_N(t) = L_0(t)x(t_0) + L_1(t)x(t_1) + \dots + L_N(t)x(t_N) \quad (10.18)$$

where each Lagrange coefficient $L_N(t)$ is itself a polynomial of degree N given by

$$L_i(t) = \frac{(t-t_0) \cdots (t-t_{i-1})(t-t_{i+1}) \cdots (t-t_N)}{(t_i-t_0) \cdots (t_i-t_{i-1})(t_i-t_{i+1}) \cdots (t_i-t_N)} = \prod_{\substack{n=0 \\ n \neq i}}^N \frac{t-t_n}{t_i-t_n} \quad (10.19)$$

Note that the i^{th} Lagrange polynomial coefficient $L_i(t)$ becomes unity at the i^{th} known sample point (i.e. $L_i(t_i)=1$), and zero at every other known sample

(i.e. $L_i(t_j)=0, i \neq j$). Therefore $P_N(t_i)=L_i(t_i)x(t_i)=x(t_i)$, and the polynomial passes through the known data points as required.

The main drawbacks of the Lagrange interpolation method are as follows:

- (a) The computational complexity is large.
- (b) The coefficients of a polynomial of order N cannot be used in the calculations of the coefficients of a higher order polynomial.
- (c) The evaluation of the interpolation error is difficult.

The Newton polynomial, introduced in the next section, overcomes some of these difficulties.

10.2.2 Newton Polynomial Interpolation

Newton polynomials have a recursive structure, such that a polynomial of order N can be constructed by extension of a polynomial of order $N-1$ as follows:

$$p_0(t) = a_0 \tag{d.c. value}$$

$$\begin{aligned} p_1(t) &= a_0 + a_1(t - t_0) \\ &= p_0(t) + a_1(t - t_0) \end{aligned} \tag{ramp}$$

$$\begin{aligned} p_2(t) &= \underbrace{a_0 + a_1(t - t_0)} + a_2(t - t_0)(t - t_1) \\ &= p_1(t) + a_2(t - t_0)(t - t_1) \end{aligned} \tag{quadratic}$$

$$\begin{aligned} p_3(t) &= \underbrace{a_0 + a_1(t - t_0) + a_2(t - t_0)(t - t_1)} + a_3(t - t_0)(t - t_1)(t - t_2) \\ &= p_2(t) + a_3(t - t_0)(t - t_1)(t - t_2) \end{aligned} \tag{cubic} \tag{10.20}$$

and in general the recursive, *order update*, form of a Newton polynomial can be formulated as

$$p_N(t) = p_{N-1}(t) + a_N(t - t_0)(t - t_1) \cdots (t - t_{N-1}) \tag{10.21}$$

For a sequence of $N+1$ samples $\{x(t_0), x(t_1), \dots, x(t_N)\}$, the polynomial coefficients are obtained using the constraint $p_N(t_i)=x(t_i)$ as follows: To solve for the coefficient a_0 , equate the polynomial Equation (10.21) at $t=t_0$ to $x(t_0)$:

$$p_N(t_0)=p_0(t_0)=x(t_0)=a_0 \quad (10.22)$$

To solve for the coefficient a_1 , the first-order polynomial $p_1(t)$ is evaluated at $t=t_1$:

$$p_1(t_1)=x(t_1)=a_0+a_1(t_1-t_0)=x(t_0)+a_1(t_1-t_0) \quad (10.23)$$

from which

$$a_1 = \frac{x(t_1) - x(t_0)}{t_1 - t_0} \quad (10.24)$$

Note that the coefficient a_1 is the slope of the line passing through the points $[x(t_0), x(t_1)]$. To solve for the coefficient a_2 the second-order polynomial $p_2(t)$ is evaluated at $t=t_2$:

$$p_2(t_2)=x(t_2)=a_0+a_1(t_2-t_0)+a_2(t_2-t_0)(t_2-t_1) \quad (10.25)$$

Substituting a_0 and a_1 from Equations (10.22) and (10.24) in Equation (10.25) we obtain

$$a_2 = \left[\frac{x(t_2)-x(t_1)}{t_2-t_1} - \frac{x(t_1)-x(t_0)}{t_1-t_0} \right] / (t_2 - t_0) \quad (10.26)$$

Each term in the square brackets of Equation (10.26) is a slope term, and the coefficient a_2 is the slope of the slope. To formulate a solution for the higher-order coefficients, we need to introduce the concept of *divided differences*. Each of the two ratios in the square brackets of Equation (10.26) is a so-called “divided difference”. The divided difference between two points t_i and t_{i-1} is defined as

$$d_1(t_{i-1}, t_i) = \frac{x(t_i) - x(t_{i-1})}{t_i - t_{i-1}} \quad (10.27)$$

The divided difference between two points may be interpreted as the average difference or the slope of the line passing through the two points. The second-order divided difference (i.e. the divided difference of the divided difference) over three points t_{i-2} , t_{i-1} and t_i is given by

$$d_2(t_{i-2}, t_i) = \frac{d_1(t_{i-1}, t_i) - d_1(t_{i-2}, t_{i-1})}{t_i - t_{i-2}} \quad (10.28)$$

and the third-order divided difference is

$$d_3(t_{i-3}, t_i) = \frac{d_2(t_{i-2}, t_i) - d_2(t_{i-3}, t_{i-1})}{t_i - t_{i-3}} \quad (10.29)$$

and so on. In general the j^{th} order divided difference can be formulated in terms of the divided differences of order $j-1$, in an order-update equation given as

$$d_j(t_{i-j}, t_i) = \frac{d_{j-1}(t_{i-j+1}, t_i) - d_{j-1}(t_{i-j}, t_{i-1})}{t_i - t_{i-j}} \quad (10.30)$$

Note that $a_1 = d_1(t_0, t_1)$, $a_2 = d_2(t_0, t_2)$ and $a_3 = d_3(t_0, t_3)$, and in general the Newton polynomial coefficients are obtained from the divided differences using the relation

$$a_i = d_i(t_0, t_i) \quad (10.31)$$

A main advantage of the Newton polynomial is its computational efficiency, in that a polynomial of order $N-1$ can be easily extended to a higher-order polynomial of order N . This is a useful property in the selection of the best polynomial order for a given set of data.

10.2.3 Hermite Polynomial Interpolation

Hermite polynomials are formulated to fit not only to the signal samples, but also to the derivatives of the signal as well. Suppose the data consists of $N+1$ samples and assume that all the derivatives up to the M^{th} order derivative are available. Let the data set, i.e. the signal samples and the derivatives, be denoted as $[x(t_i), x'(t_i), x''(t_i), \dots, x^{(M)}(t_i), i = 0, \dots, N]$. There

are altogether $K=(N+1)(M+1)$ data points and a polynomial of order $K-1$ can be fitted to the data as

$$p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \dots + a_{K-1} t^{K-1} \quad (10.32)$$

To obtain the polynomial coefficients, we substitute the given samples in the polynomial and its M derivatives as

$$\begin{aligned} p(t_i) &= x(t_i) \\ p'(t_i) &= x'(t_i) \\ p''(t_i) &= x''(t_i) \\ &\vdots \\ p^{(M)}(t_i) &= x^{(M)}(t_i), \quad i=0,1,\dots,N \end{aligned} \quad (10.33)$$

In all, there are $K=(M+1)(N+1)$ equations in (10.33), and these can be used to calculate the coefficients of the polynomial Equation (10.32). In theory, the constraint that the polynomial must also fit the derivatives should result in a better interpolating polynomial that passes through the sampled points and is also consistent with the known underlying dynamics (i.e. the derivatives) of the curve. However, even for moderate values of N and M , the size of Equation (10.33) becomes too large for most practical purposes.

10.2.4 Cubic Spline Interpolation

A polynomial interpolator of order N is constrained to pass through $N+1$ known samples, and can have $N-1$ maxima and minima. In general, the interpolation error increases rapidly with the increasing polynomial order, as the interpolating curve has to wiggle through the $N+1$ samples. When a large number of samples are to be fitted with a smooth curve, it may be better to divide the signal into a number of smaller intervals, and to fit a low order interpolating polynomial to each small interval. Care must be taken to ensure that the polynomial curves are continuous at the endpoints of each interval. In cubic spline interpolation, a cubic polynomial is fitted to each interval between two samples. A cubic polynomial has the form

$$p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \quad (10.34)$$

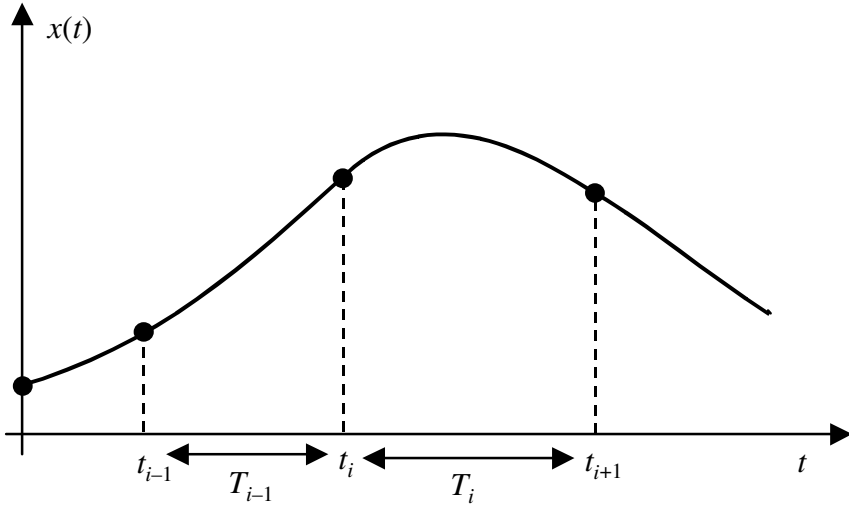


Figure 10.6 Illustration of cubic spline interpolation.

A cubic polynomial has four coefficients, and needs four conditions for the determination of a unique set of coefficients. For each interval, two conditions are set by the samples at the endpoints of the interval. Two further conditions are met by the constraints that the first derivatives of the polynomial should be continuous across each of the two endpoints. Consider an interval $t_i \leq t \leq t_{i+1}$ of length $T_i = t_{i+1} - t_i$ as shown in Figure 10.6. Using a local coordinate $\tau = t - t_i$, the cubic polynomial becomes

$$p(\tau) = a_0 + a_1\tau + a_2\tau^2 + a_3\tau^3 \tag{10.35}$$

At $\tau=0$, we obtain the first coefficient a_0 as

$$a_0 = p(\tau=0) = x(t_i) \tag{10.36}$$

The second derivative of $p(\tau)$ is given by

$$p''(\tau) = 2a_2 + 6a_3\tau \tag{10.37}$$

Evaluation of the second derivative at $\tau=0$ (i.e. $t=t_i$) gives the coefficient a_2

$$a_2 = \frac{p_i''(\tau=0)}{2} = \frac{p_i''}{2} \quad (10.38)$$

Similarly, evaluating the second derivative at the point t_{i+1} (i.e. $\tau=T_i$) yields the fourth coefficient

$$a_3 = \frac{p_{i+1}'' - p_i''}{6T_i} \quad (10.39)$$

Now to obtain the coefficient a_1 , we evaluate $p(\tau)$ at $\tau=T_i$:

$$p(\tau=T_i) = a_0 + a_1T_i + a_2T_i^2 + a_3T_i^3 = x(t_{i+1}) \quad (10.40)$$

and substitute a_0 , a_2 and a_3 from Equations (10.36), (10.38) and (10.39) in (10.40) to obtain

$$a_1 = \frac{x(t_{i+1}) - x(t_i)}{T_i} - \frac{p_{i+1}'' + 2p_i''}{6} T_i \quad (10.41)$$

The cubic polynomial can now be written as

$$p(\tau) = x(t_i) + \left[\frac{x(t_{i+1}) - x(t_i)}{T_i} - \frac{p_{i+1}'' + 2p_i''}{6} T_i \right] \tau + \frac{p_i''}{2} \tau^2 + \frac{p_{i+1}'' - p_i''}{6T_i} \tau^3 \quad (10.42)$$

To determine the coefficients of the polynomial in Equation (10.42), we need the second derivatives and p_{i+1}'' . These are obtained from the constraint that the first derivatives of the curves at the endpoints of each interval must be continuous. From Equation (10.42), the first derivatives of $p(\tau)$ evaluated at the endpoints t_i and t_{i+1} are

$$p'_i = p'(\tau=0) = -\frac{T_i}{6} [p_{i+1}'' + 2p_i''] + \frac{1}{T_i} [x(t_{i+1}) - x(t_i)] \quad (10.43)$$

$$p'_{i+1} = p'(\tau=T_i) = \frac{T_i}{6} [2p_{i+1}'' + p_i''] + \frac{1}{T_i} [x(t_{i+1}) - x(t_i)] \quad (10.44)$$

Similarly, for the preceding interval, $t_{i-1} < t < t_i$, the first derivative of the cubic spline curve evaluated at $\tau = t_i$ is given by

$$p'_i = p'(\tau = t_i) = \frac{T_{i-1}}{6} [2p''_i + p''_{i-1}] + \frac{1}{T_{i-1}} [x(t_i) - x(t_{i-1})] \tag{10.45}$$

For continuity of the first derivative at t_i , p'_i at the end of the interval (t_{i-1}, t_i) must be equal to the p'_i at the start of the interval (t_i, t_{i+1}) . Equating the right-hand sides of Equations (10.43) and (10.45) and repeating this exercise yields

$$T_{i-1} p''_{i-1} + 2(T_{i-1} + T_i) p''_i + T_i p''_{i+1} = 6 \left[\frac{1}{T_{i-1}} x(t_{i-1}) - \left(\frac{1}{T_{i-1}} + \frac{1}{T_i} \right) x(t_i) + \frac{1}{T_i} x(t_{i+1}) \right] \tag{10.46}$$

$i = 1, 2, \dots, N-1$

In Equation (10.46), there are $N-1$ equations in $N+1$ unknowns p''_i . For a unique solution we need to specify the second derivatives at the points t_0 and t_N . This can be done in two ways: (a) setting the second derivatives at the endpoints t_0 and t_N (i.e. p''_0 and p''_N), to zero, or (b) extrapolating the derivatives from the inside data.

10.3 Model-Based Interpolation

The statistical signal processing approach to interpolation of a sequence of lost samples is based on the utilisation of a predictive and/or a probabilistic model of the signal. In this section, we study the maximum a posteriori interpolation, an autoregressive model-based interpolation, a frequency-time interpolation method, and interpolation through searching a signal record for the best replacement.

Figures 10.7 and 10.8 illustrate the problem of interpolation of a sequence of lost samples. It is assumed that we have a signal record of N samples, and that within this record a segment of M samples, starting at time k , $\mathbf{x}_{Uk} = \{x(k), x(k+1), \dots, x(k+M-1)\}$ are missing. The objective is to make an optimal estimate of the missing segment \mathbf{x}_{Uk} , using the remaining $N-k$ samples \mathbf{x}_{Kn} and a model of the signal process. An N -sample signal vector

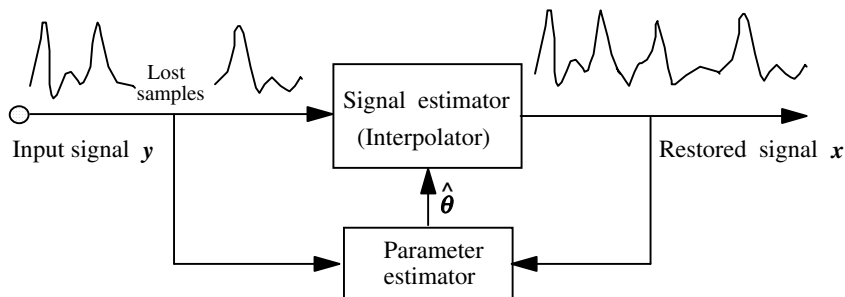


Figure 10.7 Illustration of a model-based iterative signal interpolation system.

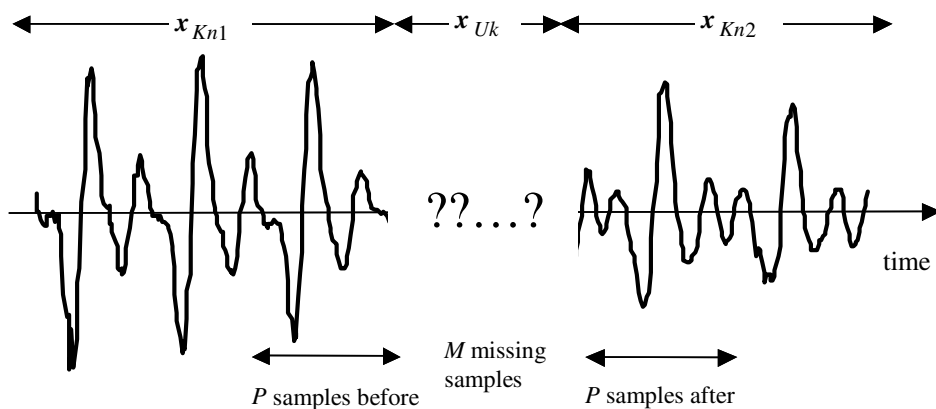


Figure 10.8 A signal with M missing samples and $N-M$ known samples. On each side of the missing segment, P samples are used to interpolate the segment.

\mathbf{x} , composed of M unknown samples and $N-M$ known samples, can be written as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_{Kn_1} \\ \mathbf{x}_U \\ \mathbf{x}_{Kn_2} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{Kn_1} \\ \mathbf{0} \\ \mathbf{x}_{Kn_2} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{x}_{Uk} \\ \mathbf{0} \end{pmatrix} = \mathbf{K} \mathbf{x}_{Kn} + \mathbf{U} \mathbf{x}_{Uk} \quad (10.47)$$

where the vector $\mathbf{x}_{Kn} = [\mathbf{x}_{Kn_1} \ \mathbf{x}_{Kn_2}]^T$ is composed of the known samples, and the vector \mathbf{x}_{Uk} is composed of the unknown samples, as illustrated in Figure 10.8. The matrices \mathbf{K} and \mathbf{U} in Equation (10.47) are rearrangement matrices that assemble the vector \mathbf{x} from \mathbf{x}_{Kn} and \mathbf{x}_{Uk} .

10.3.1 Maximum A Posteriori Interpolation

The posterior pdf of an unknown signal segment \mathbf{x}_{Uk} given a number of neighbouring samples \mathbf{x}_{Kn} can be expressed using Bayes' rule as

$$\begin{aligned}
 f_X(\mathbf{x}_{Uk}|\mathbf{x}_{Kn}) &= \frac{f_X(\mathbf{x}_{Kn}, \mathbf{x}_{Uk})}{f_X(\mathbf{x}_{Kn})} \\
 &= \frac{f_X(\mathbf{x} = \mathbf{K} \mathbf{x}_{Kn} + \mathbf{U} \mathbf{x}_{Uk})}{f_X(\mathbf{x}_{Kn})}
 \end{aligned}
 \tag{10.48}$$

In Equation (10.48), for a given sequence of samples \mathbf{x}_{Kn} , $f_X(\mathbf{x}_{Kn})$ is a constant. Therefore the estimate that maximises the posterior pdf, i.e. the MAP estimate, is given by

$$\hat{\mathbf{x}}_{Uk}^{MAP} = \arg \max_{\mathbf{x}_{Uk}} f_X(\mathbf{K} \mathbf{x}_{Kn} + \mathbf{U} \mathbf{x}_{Uk})
 \tag{10.49}$$

Example 10.2 *MAP interpolation of a Gaussian signal.* Assume that an observation signal $\mathbf{x} = \mathbf{K} \mathbf{x}_{Kn} + \mathbf{U} \mathbf{x}_{Uk}$, from a zero-mean Gaussian process, is composed of a sequence of M missing samples \mathbf{x}_{Uk} and $N - M$ known neighbouring samples as in Equation (10.47). The pdf of the signal \mathbf{x} is given by

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_{xx}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{x}\right)
 \tag{10.50}$$

where $\boldsymbol{\Sigma}_{xx}$ is the covariance matrix of the Gaussian vector process \mathbf{x} . Substitution of Equation (10.50) in Equation (10.48) yields the conditional pdf of the unknown signal \mathbf{x}_{Uk} given a number of samples \mathbf{x}_{Kn} :

$$\begin{aligned}
 f_X(\mathbf{x}_{Uk}|\mathbf{x}_{Kn}) &= \frac{1}{f_X(\mathbf{x}_{Kn})} \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_{xx}|^{1/2}} \times \\
 &\quad \exp\left(-\frac{1}{2} (\mathbf{K} \mathbf{x}_{Kn} + \mathbf{U} \mathbf{x}_{Uk})^T \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{K} \mathbf{x}_{Kn} + \mathbf{U} \mathbf{x}_{Uk})\right)
 \end{aligned}
 \tag{10.51}$$

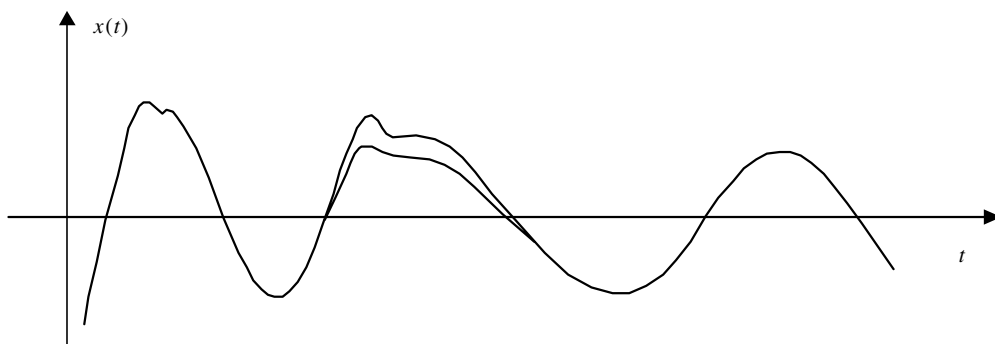


Figure 10.9 Illustration of MAP interpolation of a segment of 20 samples.

The MAP signal estimate, obtained by setting the derivative of the log-likelihood function $\ln f_{\mathbf{x}}(\mathbf{x}|\mathbf{x}_{K_n})$ of Equation (10.51) with respect to \mathbf{x}_{U_k} to zero, is given by

$$\mathbf{x}_{U_k} = -\left(\mathbf{U}^T \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{U}\right)^{-1} \mathbf{U}^T \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{K} \mathbf{x}_{K_n} \quad (10.52)$$

An example of MAP interpolation is shown in Figure 10.9.

10.3.2 Least Square Error Autoregressive Interpolation

In this section, we describe interpolation based on an autoregressive (AR) model of the signal process. The term “autoregressive model” is an alternative terminology for the linear predictive models considered in Chapter 7. In this section, the terms “linear predictive model” and “autoregressive model” are used interchangeably. The AR interpolation algorithm is a two-stage process: in the first stage, the AR model coefficients are estimated from the incomplete signal, and in the second stage the estimates of the model coefficients are used to interpolate the missing samples. For high-quality interpolation, the estimation algorithm should utilise all the correlation structures of the signal process, including periodic or pitch period structures. In Section 10.3.4, the AR interpolation method is extended to include pitch–period correlations.

10.3.3 Interpolation Based on a Short-Term Prediction Model

An autoregressive (AR), or linear predictive, signal $x(m)$ is described as

$$x(m) = \sum_{k=1}^P a_k x(m-k) + e(m) \tag{10.53}$$

where $x(m)$ is the AR signal, a_k are the model coefficients and $e(m)$ is a zero mean excitation signal. The excitation may be a random signal, a quasi-periodic impulse train, or a mixture of the two. The AR coefficients, a_k , model the correlation structure or equivalently the spectral patterns of the signal.

Assume that we have a signal record of N samples and that within this record a segment of M samples, starting from the sample k , $\mathbf{x}_{Uk} = \{x(k), \dots, x(k+M-1)\}$ are missing. The objective is to estimate the missing samples \mathbf{x}_{Uk} , using the remaining $N-k$ samples and an AR model of the signal. Figure 10.8 illustrates the interpolation problem. For this signal record of N samples, the AR equation (10.53) can be expanded to form the following matrix equation:

$$\begin{pmatrix} e(P) \\ e(P+1) \\ \vdots \\ e(k-1) \\ \hline e(k) \\ e(k+1) \\ e(k+2) \\ \vdots \\ e(k+M+P-2) \\ e(k+M+P-1) \\ \hline e(k+M+P) \\ e(k+M+P+1) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} x(P) \\ x(P+1) \\ \vdots \\ x(k-1) \\ \hline x_{Uk}(k) \\ x_{Uk}(k+1) \\ x_{Uk}(k+2) \\ \vdots \\ x(k+M+P-2) \\ x(k+M+P-1) \\ \hline x(k+M+P) \\ x(k+M+P+1) \\ \vdots \\ x(N-1) \end{pmatrix} - \begin{pmatrix} x(P-1) & x(P-2) & \dots & x(0) \\ x(P) & x(P-1) & \dots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(k-2) & x(k-3) & \dots & x(k-P-1) \\ \hline x(k-1) & x(k-2) & \dots & x(k-P) \\ x_{Uk}(k) & x(k-1) & \dots & x(k-P+1) \\ x_{Uk}(k+1) & x_{Uk}(k) & \dots & x(k-P+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(k+M+P-3) & x(k+M+P-2) & \dots & x_{Uk}(k+M-2) \\ x(k+M+P-2) & x(k+M+P-1) & \dots & x_{Uk}(k+M-1) \\ \hline x(k+M+P-1) & x(k+M+P) & \dots & x(k+M) \\ x(k+M+P) & x(k+M+P+1) & \dots & x(k+M+1) \\ \dots & \dots & \ddots & \dots \\ x(N-2) & x(N-3) & \dots & x(N-P-1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{pmatrix} \tag{10.54}$$

where the subscript Uk denotes the unknown samples. Equation (10.54) can be rewritten in compact vector notation as

$$\mathbf{e}(\mathbf{x}_{\text{Uk}}, \mathbf{a}) = \mathbf{x} - \mathbf{X}\mathbf{a} \quad (10.55)$$

where the error vector $\mathbf{e}(\mathbf{x}_{\text{Uk}}, \mathbf{a})$ is expressed as a function of the unknown samples and the unknown model coefficient vector. In this section, the optimality criterion for the estimation of the model coefficient vector \mathbf{a} and the missing samples \mathbf{x}_{Uk} is the minimum mean square error given by the inner vector product

$$\mathbf{e}^T \mathbf{e}(\mathbf{x}_{\text{Uk}}, \mathbf{a}) = \mathbf{x}^T \mathbf{x} + \mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a} - 2\mathbf{a}^T \mathbf{X}^T \mathbf{x} \quad (10.56)$$

The squared error function in Equation (10.56) involves nonlinear unknown terms of fourth order, $\mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a}$, and cubic order, $\mathbf{a}^T \mathbf{X}^T \mathbf{x}$. The least square error formulation, obtained by differentiating $\mathbf{e}^T \mathbf{e}(\mathbf{x}_{\text{Uk}}, \mathbf{a})$, with respect to the vectors \mathbf{a} or \mathbf{x}_{Uk} , results in a set of nonlinear equations of cubic order whose solution is non-trivial. A suboptimal, but practical and mathematically tractable, approach is to solve for the missing samples and the unknown model coefficients in two separate stages. This is an instance of the general estimate-and-maximise (EM) algorithm, and is similar to the linear-predictive model-based restoration considered in Section 6.7. In the first stage of the solution, Equation (10.54) is *linearised* by either assuming that the missing samples have zero values or discarding the set of equations in (10.54), between the two dashed lines, that involve the unknown signal samples. The linearised equations are used to solve for the AR model coefficient vector \mathbf{a} by forming the equation

$$\hat{\mathbf{a}} = \left(\mathbf{X}_{\text{Kn}}^T \mathbf{X}_{\text{Kn}} \right)^{-1} \left(\mathbf{X}_{\text{Kn}}^T \mathbf{x}_{\text{Kn}} \right) \quad (10.57)$$

where the vector $\hat{\mathbf{a}}$ is an estimate of the model coefficients, obtained from the available signal samples.

The second stage of the solution involves the estimation of the unknown signal samples \mathbf{x}_{Uk} . For an AR model of order P , and an unknown signal segment of length M , there are $2M+P$ nonlinear equations in (10.54) that involve the unknown samples; these are

$$\begin{pmatrix} e(k) \\ e(k+1) \\ e(k+2) \\ \vdots \\ e(k+M+P-2) \\ e(k+M+P-1) \end{pmatrix} = \begin{pmatrix} x_{\text{Uk}}(k) \\ x_{\text{Uk}}(k+1) \\ x_{\text{Uk}}(k+2) \\ \vdots \\ x(k+M+P-2) \\ x(k+M+P-1) \end{pmatrix} - \begin{pmatrix} x(k-1) & x(k-2) & \dots & x(k-p) \\ x_{\text{Uk}}(k) & x(k-1) & \dots & x(k-p+1) \\ x_{\text{Uk}}(k+1) & x_{\text{Uk}}(k) & \dots & x(k-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ x_{\text{Uk}}(k+M+P-3) & x_{\text{Uk}}(k+M+P-4) & \dots & x_{\text{Uk}}(k+M-2) \\ x_{\text{Uk}}(k+M+P-2) & x_{\text{Uk}}(k+M+P-3) & \dots & x_{\text{Uk}}(k+M-1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{p-1} \\ a_p \end{pmatrix} \tag{10.58}$$

The estimate of the predictor coefficient vector, obtained from the first stage of the solution, is substituted in Equation (10.58) so that the only remaining unknowns in (10.58) are the missing signal samples. Equation (10.58) may be partitioned and rearranged in vector notation in the following form:

$$\begin{pmatrix} e(k) \\ e(k+1) \\ e(k+2) \\ e(k+3) \\ e(k+4) \\ \vdots \\ e(k+P-1) \\ e(k+P) \\ e(k+P+1) \\ \vdots \\ e(k+M+P-2) \\ e(k+M+P-1) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -a_1 & 1 & 0 & 0 & \dots & 0 \\ -a_2 & -a_1 & 1 & 0 & \dots & 0 \\ -a_3 & -a_2 & -a_1 & 1 & \dots & 0 \\ -a_4 & -a_3 & -a_2 & -a_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_p & -a_{p-1} & -a_{p-2} & -a_{p-3} & \dots & 0 \\ 0 & -a_p & -a_{p-1} & -a_{p-2} & \dots & 0 \\ 0 & 0 & -a_p & -a_{p-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -a_{p-1} \\ 0 & 0 & 0 & 0 & \dots & -a_p \end{pmatrix} \begin{pmatrix} x_{\text{Uk}}(k) \\ x_{\text{Uk}}(k+1) \\ x_{\text{Uk}}(k+2) \\ x_{\text{Uk}}(k+3) \\ \vdots \\ x_{\text{Uk}}(k+M-1) \end{pmatrix} + \begin{pmatrix} -a_p & -a_{p-1} & -a_{p-2} & \dots & -a_1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & -a_p & -a_{p-1} & \dots & -a_2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & -a_p & \dots & -a_3 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -a_p & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & -a_1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & -a_2 & -a_1 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & -a_3 & -a_2 & -a_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & -a_{p-1} & -a_{p-2} & -a_{p-3} & \dots & -a_1 \end{pmatrix} \begin{pmatrix} x(k-p) \\ x(k-p+1) \\ x(k-p+2) \\ \vdots \\ x(k-1) \\ 0 \\ \vdots \\ x(k+M) \\ x(k+M+1) \\ x(k+M+2) \\ \vdots \\ x(k+M+P-1) \end{pmatrix} \tag{10.59}$$

In Equation (10.59), the unknown and known samples are rearranged and grouped into two separate vectors. In a compact vector–matrix notation, Equation (10.58) can be written in the form

$$e = A_1 x_{\text{Uk}} + A_2 x_{\text{Kn}} \tag{10.60}$$

where \mathbf{e} is the error vector, \mathbf{A}_1 is the first coefficient matrix, \mathbf{x}_{Uk} is the unknown signal vector being estimated, \mathbf{A}_2 is the second coefficient matrix and the vector \mathbf{x}_{Kn} consists of the *known* samples in the signal matrix and vectors of Equation (10.58). The total squared error is given by

$$\mathbf{e}^T \mathbf{e} = (\mathbf{A}_1 \mathbf{x}_{Uk} + \mathbf{A}_2 \mathbf{x}_{Kn})^T (\mathbf{A}_1 \mathbf{x}_{Uk} + \mathbf{A}_2 \mathbf{x}_{Kn}) \quad (10.61)$$

The least square AR (LSAR) interpolation is obtained by minimisation of the squared error function with respect to the unknown signal samples \mathbf{x}_{Uk} :

$$\frac{\partial \mathbf{e}^T \mathbf{e}}{\partial \mathbf{x}_{Uk}} = 2\mathbf{A}_1^T \mathbf{A}_1 \mathbf{x}_{Uk} + 2\mathbf{A}_1^T \mathbf{A}_2 \mathbf{x}_{Kn} = 0 \quad (10.62)$$

From Equation (10.62) we have

$$\hat{\mathbf{x}}_{Uk}^{LSAR} = -(\mathbf{A}_1^T \mathbf{A}_1)^{-1} (\mathbf{A}_1^T \mathbf{A}_2) \mathbf{x}_{Kn} \quad (10.63)$$

The solution in Equation (10.62) gives the $\hat{\mathbf{x}}_{Uk}^{LSAR}$ vector which is the least square error estimate of the unknown data vector.

10.3.4 Interpolation Based on Long-Term and Short-term Correlations

For the best results, a model-based interpolation algorithm should utilise all the correlation structures of the signal process, including any periodic structures. For example, the main correlation structures in a voiced speech signal are the short-term correlation due to the resonance of the vocal tract and the long-term correlation due to the quasi-periodic excitation pulses of the glottal cords. For voiced speech, interpolation based on the short-term correlation does not perform well if the missing samples coincide with an underlying quasi-periodic excitation pulse. In this section, the AR interpolation is extended to include both long-term and short-term correlations. For most audio signals, the short-term correlation of each sample with the immediately preceding samples decays exponentially with time, and can be usually modelled with an AR model of order 10–20. In order to include the pitch periodicities in the AR model of Equation (10.53),

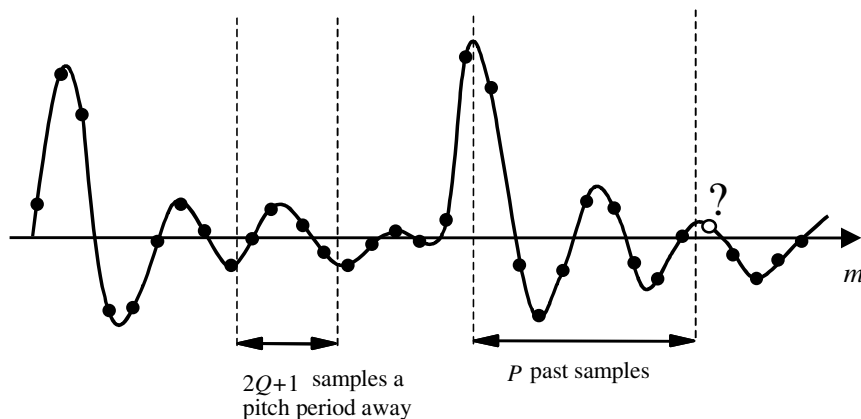


Figure 10.10 A quasiperiodic waveform. The sample marked “ ? ” is predicted using P immediate past samples and $2Q+1$ samples a pitch period away.

the model order must be greater than the pitch period. For speech signals, the pitch period is normally in the range 4–20 milliseconds, equivalent to 40–200 samples at a sampling rate of 10 kHz. Implementation of an AR model of this order is not practical owing to stability problems and computational complexity.

A more practical AR model that includes the effects of the long-term correlations is illustrated in Figure 10.10. This modified AR model may be expressed by the following equation:

$$x(m) = \sum_{k=1}^P a_k x(m-k) + \sum_{k=-Q}^Q p_k x(m-T-k) + e(m) \quad (10.64)$$

The AR model of Equation (10.64) is composed of a *short-term predictor* $\sum a_k x(m-k)$ that models the contribution of the P immediate past samples, and a *long-term predictor* $\sum p_k x(m-T-k)$ that models the contribution of $2Q+1$ samples a pitch period away. The parameter T is the pitch period; it can be estimated from the autocorrelation function of $x(m)$ as the time difference between the peak of the autocorrelation, which is at the correlation lag zero, and the second largest peak, which should happen a pitch period away from the lag zero.

The AR model of Equation (10.64) is specified by the parameter vector $c=[a_1, \dots, a_P, p_{-Q}, \dots, p_Q]$ and the pitch period T . Note that in Figure 10.10

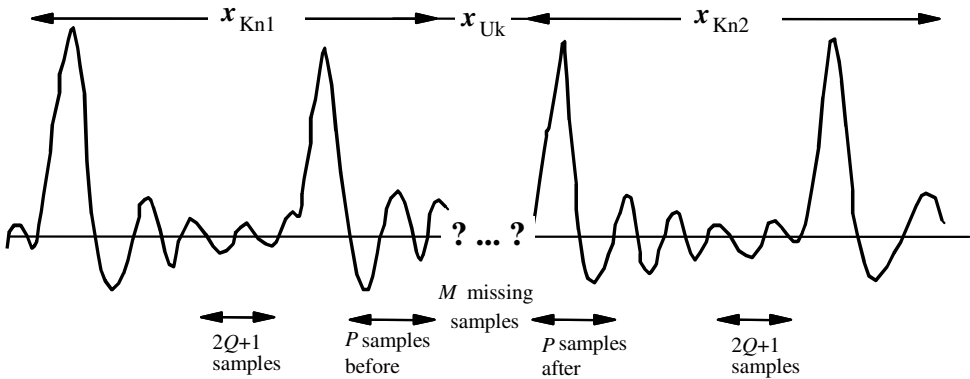


Figure 10.11 A signal with M missing samples. P immediate samples each side of the gap and $2Q+1$ samples a pitch period away are used for interpolation.

the sample marked “?” coincides with the onset of an excitation pulse. This sample is not well predictable from the P past samples, because they do not include a pulse event. The sample is more predictable from the $2Q+1$ samples a pitch period away, since they include the effects of a similar excitation pulse. The predictor coefficients are estimated (see Chapter 7) using the so-called normal equations:

$$c = R_{xx}^{-1} r_{xx} \tag{10.65}$$

where R_{xx} is the autocorrelation matrix of signal x and r_{xx} is the correlation vector. In expanded form, Equation (10.65) can be written as

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \\ p-Q \\ p-Q+1 \\ \vdots \\ p+Q \end{pmatrix} = \begin{pmatrix} r(0) & r(1) & \dots & r(p-1) & r(T+Q-1) & r(T+Q) & \dots & r(T-Q-1) \\ r(1) & r(0) & \dots & r(p-2) & r(T+Q-2) & r(T+Q-1) & \dots & r(T+Q-2) \\ r(2) & r(1) & \dots & r(p-3) & r(T+Q-3) & r(T+Q-2) & \dots & r(T+Q-3) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \dots & r(0) & r(T+Q-p) & r(T+Q-p+1) & \dots & r(T+Q-p) \\ r(T+Q-1) & r(T+Q-2) & \dots & r(T+Q-p) & r(0) & r(1) & \dots & r(2Q) \\ r(T+Q) & r(T+Q-1) & \dots & r(T+Q-p+1) & r(1) & r(0) & \dots & r(2Q-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r(T-Q-1) & r(T-Q-2) & \dots & r(T-Q-p) & r(2Q) & r(2Q-1) & \dots & r(0) \end{pmatrix}^{-1} \begin{pmatrix} r(1) \\ r(2) \\ r(3) \\ \vdots \\ r(p) \\ r(T+Q) \\ r(T+Q-1) \\ \vdots \\ r(T-Q) \end{pmatrix} \tag{10.66}$$

The modified AR model can be used for interpolation in the same way as the conventional AR model described in the previous section. Again, it is assumed that within a data window of N speech samples, a segment of M samples commencing from the sample point k , $x_{Uk} = \{x(k), x(k+1), \dots,$

$x(k+M-1)$ is missing. Figure 10.11 illustrates the interpolation problem. The missing samples are estimated using P samples in the immediate vicinity and $2Q+1$ samples a pitch period away on each side of the missing signal. For the signal record of N samples, the modified AR equation (10.64) can be written in matrix form as

$$\begin{pmatrix} e(T+Q) \\ e(T+Q+1) \\ \vdots \\ e(k-1) \\ e(k) \\ e(k+1) \\ e(k+2) \\ \vdots \\ e(k+M+P-2) \\ e(k+M+P-1) \\ e(k+M+P) \\ e(k+M+P+1) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} x(T+Q) \\ x(T+Q+1) \\ \vdots \\ x(k-1) \\ x_{UK}(k) \\ x_{UK}(k+1) \\ x_{UK}(k+2) \\ \vdots \\ x(k+M+P-2) \\ x(k+M+P-1) \\ x(k+M+P) \\ x(k+M+P+1) \\ \vdots \\ x(N-1) \end{pmatrix} - \begin{pmatrix} x(T+Q-1) & \cdots & x(T+Q-P) & & x(2Q) & \cdots & x(0) \\ x(T+Q) & \cdots & x(T+Q-P+1) & & x(2Q+1) & \cdots & x(1) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ x(k-2) & \cdots & x(k-P-1) & & x(k-T+Q-1) & \cdots & x(k-T-Q-1) \\ x(k-1) & \cdots & x(k-P) & & x(k-T+Q) & \cdots & x(k-T-Q) \\ x_{UK}(k) & \cdots & x(k-P+1) & & x(k-T+Q+1) & \cdots & x(k-T-Q+1) \\ x_{UK}(k+1) & \cdots & x(k-P+2) & & x(k-T+Q+2) & \cdots & x(k-T-Q+2) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ x(k+M+P-3) & \cdots & x_{UK}(k+M-2) & & x(k+M+P-T+Q-2) & \cdots & x(k+M+P-T-Q-2) \\ x(k+M+P-2) & \cdots & x_{UK}(k+M-1) & & x(k+M+P-T+Q-1) & \cdots & x(k+M+P-T-Q-1) \\ x(k+M+P-1) & \cdots & x(k+M) & & x(k+M+P-T+Q) & \cdots & x(k+M+P-T-Q) \\ x(k+M+P) & \cdots & x(k+M+1) & & x(k+M+P-T+Q+1) & \cdots & x(k+M+P-T-Q+1) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ x(N-2) & \cdots & x(N-P-1) & & x(N-T+Q-1) & \cdots & x(N-T-Q-1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \\ \vdots \\ a_{P-Q} \\ \vdots \\ a_{P+Q} \end{pmatrix} \tag{10.67}$$

where the subscript U_k denotes the unknown samples. In compact matrix notation, this set of equation can be written in the form

$$\mathbf{e}(x_{U_k}, \mathbf{c}) = \mathbf{x} + \mathbf{Xc} \tag{10.68}$$

As in Section 10.3.2, the interpolation problem is solved in two stages:

- (a) In the first stage, the known samples on both sides of the missing signal are used to estimate the AR coefficient vector \mathbf{c} .
- (b) In the second stage, the AR coefficient estimates are substituted in Equation (10.68) so that the only unknowns are the data samples.

The solution follows the same steps as those described in Section 10.3.2.

10.3.5 LSAR Interpolation Error

In this section, we discuss the effects of the signal characteristics, the model parameters and the number of unknown samples on the interpolation error. The interpolation error $v(m)$, defined as the difference between the original sample $x(m)$ and the interpolated sample $\hat{x}(m)$, is given by

$$v(m) = x(m) - \hat{x}(m) \quad (10.69)$$

A common measure of signal distortion is the mean square error distance defined as

$$D(\mathbf{c}, M) = \frac{1}{M} \mathcal{E} \left\{ \sum_{m=0}^{M-1} [x(k+m) - \hat{x}(k+m)]^2 \right\} \quad (10.70)$$

where k is the beginning of an M -samples long segment of missing signal, and $\mathcal{E} [\cdot]$ is the expectation operator. In Equation (10.70), the average distortion D is expressed as a function of the number of the unknown samples M , and also the model coefficient vector \mathbf{c} . In general, the quality of interpolation depends on the following factors:

- (a) *The signal correlation structure.* For deterministic signals such as sine waves, the theoretical interpolation error is zero. However information-bearing signals have a degree of randomness that makes perfect interpolation with zero error an impossible objective.
- (b) *The length of the missing segment.* The amount of information lost, and hence the interpolation error, increase with the number of missing samples. Within a sequence of missing samples the error is usually largest for the samples in the middle of the gap. The interpolation Equation (10.63) becomes increasingly ill-conditioned as the length of the missing samples increases.
- (c) *The nature of the excitation underlying the missing samples.* The LSAR interpolation cannot account for any random excitation underlying the missing samples. In particular, the interpolation quality suffers when the missing samples coincide with the onset of an excitation pulse. In general, the least square error criterion causes the interpolator to underestimate the energy of the underlying excitation signal. The inclusion of long-term prediction and the use of quasi-periodic structure of signals improves the ability of the interpolator to restore the missing samples.
- (d) *AR model order and the method used for estimation of the AR coefficients.* The interpolation error depends on the AR model order. Usually a model order of 2–3 times the length of missing data sequence achieves good result.

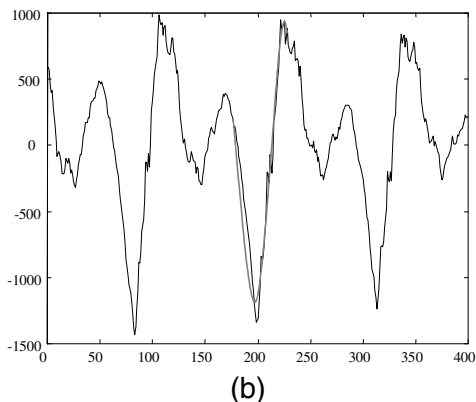
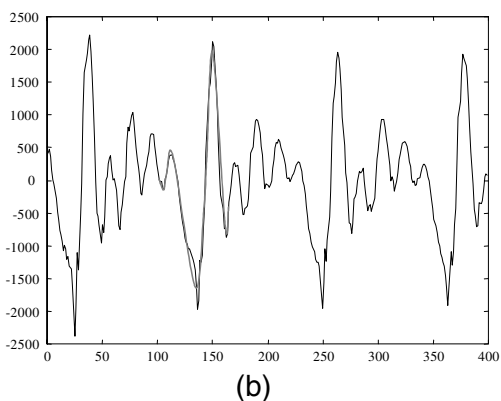
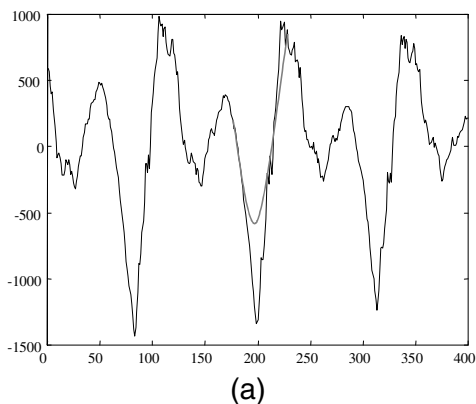
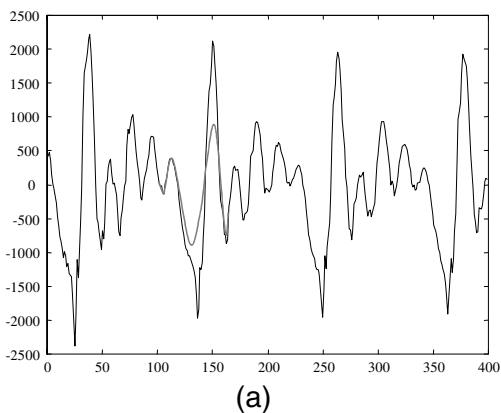


Figure 10.12 (a) A section of speech showing interpolation of 60 samples starting from the sample point 100 (b) Interpolation using short and long-term correlations. Interpolated samples are shown by the light shaded line.

Figure 10.13 (a) A section of speech showing interpolation of 50 samples starting from the sample point 175 (b) Interpolation using short and long-term correlations. Interpolated samples are shown by the light shaded line.

The interpolation error also depends on how well the AR parameters can be estimated from the incomplete data. In Equation (10.54), in the first stage of the solution, where the AR coefficients are estimated, two different approaches may be employed to linearise the system of equations. In the first approach all equations, between the dashed lines, that involve nonlinear terms are discarded. This approach has the advantage that no assumption is made about the missing samples. In fact, from a signal-ensemble point of view, the effect of discarding some equations is

equivalent to that of having a smaller signal record. In the second method, starting from an initial estimate of the unknown vector (such as $\mathbf{x}_{\text{UK}}=\mathbf{0}$), Equation (10.54) is solved to obtain the AR parameters. The AR coefficients are then used in the second stage of the algorithm to estimate the unknown samples. These estimates may be improved in further iterations of the algorithm. The algorithm usually converges after one or two iterations.

Figures 10.12 and 10.13 show the results of application of the least square error AR interpolation method to speech signals. The interpolated speech segments were chosen to coincide with the onset of an excitation pulse. In these experimental cases the original signals are available for comparison. Each signal was interpolated by the AR model of Equation (10.53) and also by the extended AR model of Equation (10.64). The length of the conventional linear predictor model was set to 20. The modified linear AR model of Equation (10.64) has a prediction order of (20,7); that is, the short-term predictor has 20 coefficients and the long-term predictor has 7 coefficients. The figures clearly demonstrate that the modified AR model that includes the long-term as well as the short-term correlation structures outperforms the conventional AR model.

10.3.6 Interpolation in Frequency–Time Domain

Time-domain, AR model-based interpolation methods are effective for the interpolation of a relatively short length of samples (say less than 100 samples at a 20 kHz sampling rate), but suffer severe performance degradations when used for interpolation of large sequence of samples. This is partly due to the numerical problems associated with the inversion of a large matrix, involved in the time-domain interpolation of a large number of samples, Equation (10.58).

Spectral–time representation provides a useful form for the interpolation of a large gap of missing samples. For example, through discrete Fourier transformation (DFT) and spectral–time representation of a signal, the problem of interpolation of a gap of N samples in the time domain can be converted into the problem of interpolation of a gap of one sample, along the time, in each of N discrete frequency bins, as explained next.

Spectral–Time Representation with STFT

A relatively simple and practical method for spectral–time representation of a signal is the short-time Fourier transform (STFT) method. To construct a

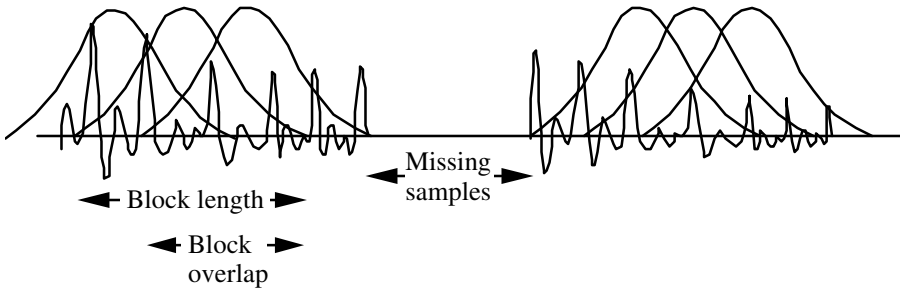


Figure 10.14 Illustration of segmentation of a signal (with a missing gap) for spectral-time representation.

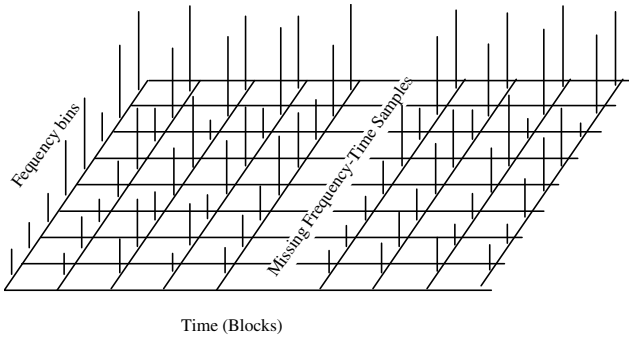


Figure 10.15 Spectral-time representation of a signal with a missing gap.

two-dimensional STFT from a one-dimensional function of time $x(m)$, the input signal is segmented into overlapping blocks of N samples, as illustrated in Figure 10.14. Each block is windowed, prior to discrete Fourier transformation, to reduce the spectral leakage due to the effects of discontinuities at the edges of the block. The frequency spectrum of the m^{th} signal block is given by the discrete Fourier transform as

$$X(k, m) = \sum_{i=0}^{N-1} w(i)x(m(N - D) + i)e^{-j\frac{2\pi}{N}ik}, \quad k = 0, \dots, N-1 \quad (10.71)$$

where $X(k, m)$ is a spectral-time representation with time index m and frequency index k , N is the number of samples in each block, and D is the block overlap. In STFT, it is assumed that the signal frequency composition is time-invariant within the duration of each block, but it may vary across

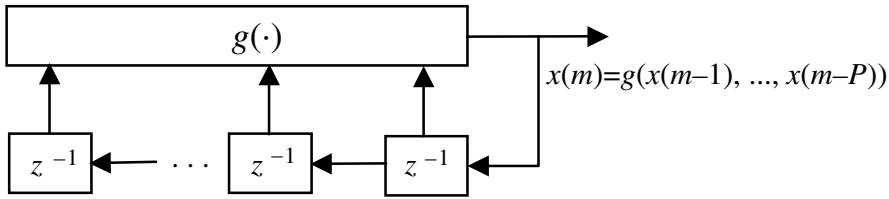


Figure 10.16 Configuration of a digital oscillator.

the blocks. In general, the k^{th} spectral component of a signal has a time-varying character, i.e. it is “born”, evolves for some time, disappears, and then reappears with a different intensity and a different characteristics. Figure 10.15 illustrates a spectralthtime signal with a missing block of samples. The aim of interpolation is to fill in the signal gap such that, at the beginning and at the end of the gap, the continuity of both the magnitude and the phase of each frequency component of the signal is maintained. For most time-varying signals (such as speech), a low-order polynomial interpolator of the magnitude and the phase of the DFT components of the signal, making use of the few adjacent blocks on either side of the gap, would produce satisfactory results.

10.3.7 Interpolation Using Adaptive Code Books

In the LSAR interpolation method, described in Section 10.3.2, the signals are modelled as the output of an AR model excited by a random input. Given enough samples, the AR coefficients can be estimated with reasonable accuracy. However, the instantaneous values of the random excitation during the periods when the signal is missing cannot be recovered. This leads to a consistent underestimation of the amplitude and the energy of the interpolated samples. One solution to this problem is to use a zero-input signal model. Zero-input models are feedback oscillator systems that produce an output signal without requiring an input.

The general form of the equation describing a digital nonlinear oscillator can be expressed as

$$x(m) = g_f(x(m-1), x(m-2), \dots, x(m-P)) \quad (10.72)$$

The mapping function $g_f(\cdot)$ may be a parametric or a non-parametric mapping. The model in Equation (10.72) can be considered as a nonlinear

predictor, and the subscript f denotes forward prediction based on the past samples.

A parametric model of a nonlinear oscillator can be formulated using a Volterra filter model. However, in this section, we consider a non-parametric method for its ease of formulation and stable characteristics. Kubin and Kleijin (1994) have described a non-parametric oscillator based on a codebook model of the signal process.

In this method, each entry in the code book has $P+1$ samples where the $(P+1)^{\text{th}}$ sample is intended as an output. Given P input samples $\mathbf{x}=[x(m-1), \dots, x(m-P)]$, the codebook output is the $(P+1)^{\text{th}}$ sample of the vector in the codebook whose first P samples have a minimum distance from the input signal \mathbf{x} . For a signal record of length N samples, a codebook of size $N-P$ vectors can be constructed by dividing the signal into overlapping segments of $P+1$ samples with the successive segments having an overlap of P samples. Similarly a backward oscillator can be expressed as

$$x_b(m)=g_b(x(m+1),x(m+2),\dots,x(m+P)) \tag{10.73}$$

As in the case of a forward oscillator, the backward oscillator can be designed using a non-parametric method based on an adaptive codebook of the signal process. In this case each entry in the code book has $P+1$ samples where the first sample is intended as an output sample. Given P input samples $\mathbf{x}=[x(m), \dots, x(m+P-1)]$ the codebook output is the first sample of the code book vector whose next P samples have a minimum distance from the input signal \mathbf{x} .

For interpolation of M missing samples, the outputs of the forward and backward nonlinear oscillators may be combined as

$$\hat{x}(k+m)=\left(\frac{M-1-m}{M-1}\right)\hat{x}_f(k+m)+\left(\frac{m}{M-1}\right)\hat{x}_b(k+m) \tag{10.74}$$

where it is assumed that the missing samples start at k .

10.3.8 Interpolation Through Signal Substitution

Audio signals often have a time-varying but quasi-periodic repetitive structure. Therefore most acoustic events in a signal record *reoccur* with some variations. This observation forms the basis for interpolation through

pattern matching, where a missing segment of a signal is substituted by the best match from a signal record. Consider a relatively long signal record of N samples, with a gap of M missing samples at its centre. A section of the signal with the gap in the middle can be used to search for the best-match segment in the record. The missing samples are then substituted by the corresponding section of the best-match signal. This interpolation method is particularly useful when the length of the missing signal segment is large. For a given class of signals, we may be able to construct a library of patterns for use in waveform substitution, Bogner (1989).

10.4 Summary

Interpolators, in their various forms, are used in most signal processing applications. The obvious example is the estimation of a sequence of missing samples. However, the use of an interpolator covers a much wider range of applications, from low-bit-rate speech coding to pattern recognition and decision making systems. We started this chapter with a study of the ideal interpolation of a band-limited signal, and its applications in digital-to-analog conversion and in multirate signal processing. In this chapter, various interpolation methods were categorised and studied in two different sections: one on polynomial interpolation, which is the more traditional numerical computing approach, and the other on statistical interpolation, which is the digital signal processing approach.

The general form of the polynomial interpolator was formulated and its special forms, Lagrange, Newton, Hermite and cubic spline interpolators were considered. The polynomial methods are not equipped to make optimal use of the predictive and statistical structures of the signal, and are impractical for interpolation of a relatively large number of samples. A number of useful statistical interpolators were studied. These include maximum a posteriori interpolation, least square error AR interpolation, frequency-time interpolation, and an adaptive code book interpolator. Model-based interpolation method based on an autoregressive model is satisfactory for most audio applications so long as the length of the missing samples is not too large. For interpolation of a relatively large number of samples the time–frequency interpolation method and the adaptive code book method are more suitable.

Bibliography

- BOGNER R.E. and LI T. (1989) Pattern Search Prediction of Speech. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-89, pp. 180–183, Glasgow.
- COHEN L. (1989) Time-Frequency Distributions-A review. Proc. IEEE, **77(7)**, pp. 941–81.
- CROCHIERE R.E. and RABINER L.R. (1981) Interpolation and Decimation of Digital Signals-A Tutorial review. Proc. IEEE, **69**, pp. 300–331, March.
- GODSILL S.J. (1993) The Restoration of Degraded Audio Signals. Ph.D. Thesis, Cambridge University.
- GODSILL S.J. and Rayner P.J.W. (1993) Frequency domain interpolation of sampled signals. IEEE Int. Conf., Speech and Signal Processing, ICASSP-93, Minneapolis.
- JANSSEN A.J., VELDHUIS R. and VRIES L.B (1984) Adaptive Interpolation of Discrete-Time Signals That Can Be Modelled as Autoregressive Processes. IEEE Trans. Acoustics, Speech and Signal Processing, **ASSP-34**, **2**, pp. 317–330 June.
- KAY S.M. (1983) Some Results in Linear Interpolation Theory. IEEE Trans. Acoustics Speech and Signal Processing, **ASSP-31**, pp. 746–749, June.
- KAY S.M. (1988) Modern Spectral Estimation: Theory and Application. Prentice- Hall, Englewood Cliffs, NJ.
- KOLMOGROV A.N. (1939) Sur l' Interpolation et Extrapolation des Suites Stationnaires, Comptes Rendus de l'Academie des Sciences. **208**, pp. 2043–2045.
- KUBIN G. and KLEIJIN W.B. (1994) Time-Scale Modification of Speech Based On A Nonlinear Oscillator Model. Proc. IEEE Int. Conf., Speech and Signal Processing, ICASSP-94, pp. I453–I456, Adelaide.
- LOCHART G.B. and GOODMAN D.J. (1986) Reconstruction of Missing Speech Packets by Waveform Substitution. Signal Processing 3: Theories and Applications, pp. 357-360.
- MARKS R.J. (1983) Restoring Lost Samples From An Over-Sampled Band-Limited Signal. IEEE Trans. Acoustics, Speech and Signal Processing, **ASSP-31**, **2**. pp. 752–755, June.
- MARKS R.J. (1991) Introduction to Shannon Sampling and Interpolation Theory. Springer Verlag.
- MATHEWS J.H. (1992) Numerical Methods for Mathematics. Science and Engineering, Prentice-Hall, Englewood Cliffs, NJ.

- MUSICUS B.R. (1982) Iterative Algorithms for Optimal Signal Reconstruction and Parameter Identification Given Noisy and Incomplete Data. Ph.D. Thesis, MIT, MA.
- PLATTE H.J. and ROWEDDA V. (1985) A Burst Error Concealment Method for Digital Audio Tape Application. AES Preprint, 2201:1-16.
- PRESS W.H., FLANNERY B.P., TEUKOLSKY S.A. and VETTERELING W.T. (1992) Numerical Recipes in C, 2nd Ed. Cambridge University Press.
- NAKAMURA S. (1991) Applied Numerical Methods with Software. Prentice-Hall, Englewood Cliffs, NJ.
- SCHAFFER, R.W. and RABINER, L.R. (1973) A Digital Signal Processing Approach to Interpolation. Proc. IEEE, **61**, pp. 692-702, June.
- STEELE R. and JAYANT N.S. (1980) Statistical Block Coding for DPCM-AQF Speech. IEEE Trans on Communications. **COM-28**, **11**, pp. 1899-1907, Nov.
- TONG H.(1990) Nonlinear Time Series A Dynamical System Approach. Oxford University Press.
- VASEGHI S.V.(1988) Algorithms for Restoration of Gramophone Records. Ph.D. Thesis, Cambridge University.
- VELDHUIS R. (1990) Restoration of Lost samples in Digital Signals. Prentice-Hall, Englewood, Cliffs NJ.
- VERHELST W. and ROELANDS M. (1993) An Overlap-Add Technique Based on Waveform Similarity (Wsola) for High Quality Time-Scale Modification of Speech. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-93, pp. II-554-II-557, Adelaide.
- WIENER N. (1949) Extrapolation, Interpolation and Smoothing of Stationary Time Series With Engineering Applications. MIT Press, Cambridge, MA.