Some Recent Developments in Spectrum and Harmonic Analysis

David J. Thomson
Bell Laboratories
Whippany, New Jersey 07981

ABSTRACT

In estimating the spectrum of a stationary time series from a finite sample of the process two problems have traditionally been dominant: first, what algorithm should be used so that the resulting estimate is not severely biased; and second, how should one “smooth” the estimate so that the results are consistent and statistically significant.

Within the class of spectrum estimation procedures that have been found successful in the various engineering problems considered, bias control is achieved by iterative model formation and prewhitening combined with robust procedures (the “robust filter”), while “smoothing” is done by an adaptive nonlinear method.

Recently a method has been found which, by using a “local” principal components expansion to estimate the spectrum, provides new solutions to both the bias and smoothing problems and also permits a unification of the differences between windowed and unwindowed philosophies. This estimate, which is an approximate solution of an integral equation, consists of a weighted average of a series of direct spectrum estimates made using discrete prolate spheroidal sequences as orthogonal data windows.

Keywords
Spectrum estimation, prolate spheroidal wave functions, data windows, mixed spectra, harmonic analysis, analysis of variance, smoothing.

1. Introduction

This paper reconsiders the old problem in time series analysis of reliably estimating the spectral density function from a finite sample of an approximately Gaussian stationary process. In particular we consider three related topics: first, What general form should the estimation procedure take? second, Should one use a data window, and if so which one, and third, How should one “smooth” the resulting estimate?

The fact that these questions may be considered distinct is historical; in the following we attempt to summarize some of the developments which led to them and some recent developments towards their unification. While this theory is somewhat orthogonal to the current emphasis on rational spectrum estimates, i.e. autoregressive, moving average, or combined (ARMA) estimates plus their Burg and “maximum entropy” variants, it has significant impact on these problems, in particular on the “super resolution” estimation question.

For several years following the 1946 Royal Statistical Society conference[1] most of the work in the field was directed either towards solving computational problems or to finding better “lag windows” so that the perceived conflict between resolution and variance could be minimized. Advances in computer technology and the discovery of the fast Fourier transform eventually led to more difficult problems being considered but fifteen years ago papers on the “smoothing” problem and the trade between resolution and variance were common and most practitioners used some variation of the Blackman-Tukey method. A few, notably Tukey[1967] and Welch[1967], had advanced to use of direct estimates based on the recently discovered fast Fourier transform. Even though prewhitening had been recommended in Blackman and Tukey in 1957 and earlier, the major bias problems were still largely ignored.

Ten years ago it was beginning to be recognized that the fundamental conflict in spectrum estimation was not between resolution and variance but between resolution and bias. This realization prompted the use of prolate spheroidal wave functions for data windows and led to the use of autoregressive models for prewhitening filters. While the use of these techniques resulted in better estimates, the use of data windows (not to be confused with the older lag windows used for smoothing) also posed a dilemma: first, they clearly could result in vastly superior estimates to unwindowed estimates; second, they effectively gave different weighting to observations which were equally valid. This conflict is difficult to resolve.

Again, five years ago, the spectrum estimation procedures in common use had changed significantly from their predecessors: robust procedures were coming into use, the old smoothers were finally being replaced by adaptive techniques; and the structure of the estimation process was becoming clearer. Details of these procedures are contained in Thomson[1977a,b] and Kleiner et al [1979]. The increasing power of these techniques, however, also made it clear that spectrum estimation, as a subject, was an art consisting of a collection of heuristic methods only loosely bound by theory. Worse, much of the existing theory, particularly the parts depending on asymptotics, was inadequate.

Three years ago Slepian[1978] described discrete prolate spheroidal wave functions and sequences. Using these functions as a basis does much to unify the theory of spectrum estimation: A simple ANOVA-like test for line components makes spectrum estimation and harmonic analysis distinct problems, rather than two names for nearly the same one; The “smoothing” problem disappears; The dilemma between different windowing philosophies can be peacefully resolved; and, not least, the method provides considerable insight into some of the parametric
modelling problems as well.

2. Orthogonal Window Estimates

In this section we present a brief outline of a new and somewhat more unified theory which has a number of interesting features: first, it is an explicit small sample theory with the sample size entering explicitly into the methods and performance bounds; second, it explains the use of data windows and provides a resolution of the dilemma posed earlier; third, smoothing in the conventional manner is unnecessary as the estimate is consistent; fourth, the procedure is data-adaptive; fifth, it provides an analysis of variance test for line components including the process mean (this also distinguishes harmonic analysis from spectrum estimation); and sixth, in the case of multivariate data it results in new classes of estimates. As a particular example of the latter, the technique results in two distinct estimates of coherence, one for line components, one for the continuum.

We assume that the available data consists of $N$ samples, $x_0, x_1, \ldots, x_{N-1}$, representing a stationary, real, ergodic, near Gaussian, time series. $N$ is assumed to be finite. We initially assume that the observations are centered, that is that $E[x_k] = 0$, but later will mention a new estimate of the mean.

We assume that frequency, $f$, and radian frequency, $\omega = 2\pi f$, are defined on their principal domains $(-\frac{1}{2}, \frac{1}{2})$ and $(-\pi, \pi)$ respectively. When estimates are computed on a frequency mesh we assume that the mesh spacing is less than the equivalent Nyquist frequency, $1/2N$.

We begin with the time centered‡ Cramer representation (see Doob[1953])

$$x_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi f(n - \frac{N-1}{2})} dZ(f)$$

In simple cases where there are no line components (which in the line-test procedure becomes the null hypothesis) we have, as usual,

$$E[dZ(f)] = 0$$

but when there are line components we assume the extended representation‡

$$E[dZ(f)] = a \delta(f - f_0)$$

for the simplest case of a single line component. The case for multiple lines is similar but algebraically more complex.

This representation permits a distinction between harmonic and spectrum analysis: harmonic analysis is concerned with the first moments of $dZ(f)$, while spectrum analysis is the problem of estimating the second, and higher, moments of $dZ(f)$, and in particular the second central moment. To quote Parzen[1981]

"Spectral analysis has as its aim the determination of the properties of the function $Z(f)".

For notational simplicity it is convenient to also define the Fourier transform of the observations, $\{x\}$, in time centered form:

$$\hat{x}(f) = \sum_{k=0}^{N-1} e^{-i2\pi f(k - N^{-1})} x_n$$

so that, using the spectral representation for the data in this formula, we have

$$\hat{x}(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\sin N \pi(f - \nu)}{\sin \pi(f - \nu)} dZ(\nu)$$

which is the convolution of the Cramer process, $dZ(\nu)$, with a Dirichlet kernel. The normal interpretation of this equation is that the observed power at frequency $f$ is a complex weighted average of the power at all frequencies with, for $N$ sufficiently large, most of the weight being at frequencies "close" to $f$.

An alternative viewpoint is to consider equation (5) as a linear Fredholm integral equation of the first kind for $dZ(\nu)$. Clearly, since we are dealing with a projection operator, it is impossible to obtain exact or unique solutions: what we desire is a solution which is both statistically and numerically plausible. We thus contemplate "solving" the integral equation in some local interval about $f$, say $(f - W, f + W)$ with the condition that we want the statistics of the "solution" and not the "solution" itself. From this viewpoint the formal solution process for the integral equation can be used as a guide to the statistical approach. Investigation of the solution of an integral equation over an interval naturally suggests an eigenfunction approach and, fortunately, a very detailed description of the eigenfunctions and eigenvalues of the Dirichlet kernel has recently been given by Slepian[1978].

2.1. Diversion: Discrete Prolate Spheroidal Wave Functions and Sequences

In this section we give a short list of formulae and properties of these eigenfunctions from Slepian's paper. The eigenfunctions, denoted by $U_k(N,W;f)$, $k = 0, 1, \ldots, N-1$ are known as discrete prolate spheroidal wave functions and are solutions of the equation:

$$\int_{-W}^{W} \frac{\sin N \pi(f - f')}{\sin \pi(f - f')} U_k(N,W;f') df' = \lambda_k(N,W) U_k(N,W;f)$$

where $W$, $0 < W < \frac{1}{2}$ is the bandwidth defining "local" and is normally of the order $1/N$. The functions are