Biometrika (1998), **85**, 1, *pp.* 201–213
Printed in Great Britain

Matching sequences: Cross-spectral analysis of categorical time series

BY DAVID S. STOFFER
Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, U.S.A.
stoffer@stat.pitt.edu

AND DAVID E. TYLER
Department of Statistics, Rutgers University, New Brunswick, New Jersey 08903, U.S.A.
dtyler@stat.rutgers.edu

SUMMARY

We consider the problem of quantifying the degree to which two stationary categorical time series are coherent. The goal is to discover whether or not the sequences contain similar patterns. The problem is motivated by the problem of matching two DNA sequences. Following the ideas used in defining the spectral envelope for a qualitative-valued time series, the methods we present here focus on the problem of obtaining coherency envelopes for measuring the similarity between two categorical time series. Estimation is based on the fast Fourier transform so that the methods are computationally simple and fast, and can be applied to long sequences.

Some key words: Canonical analysis; Coherency; DNA sequences; Numerical maximisation.

1. INTRODUCTION

Let $X_1(t)$ and $X_2(t)$ $(t=0, \pm 1, \pm 2, \ldots)$ be stationary categorical time series taking values in the finite sets $\mathscr{S}_1 = \{c_{11}, \ldots, c_{1,k_1+1}\}$ and $\mathscr{S}_2 = \{c_{21}, \ldots, c_{2,k_2+1}\}$, respectively, such that

$$p_{ij} = \text{pr}\{X_i(t) = c_{ij}\} > 0 \quad (j = 1, \ldots, k_i + 1;\ i = 1, 2).$$

The problem we consider is that of quantifying the degree to which the sequences $X_1(t)$ and $X_2(t)$ match via a measure of coherency. The general problem is motivated by, but not restricted to, the matching of DNA sequences; for a recent review, see Waterman & Vingron (1994). The focus is on obtaining computationally fast and simple methods that can easily be applied to, for example, large genomic sequences. Hence we examine the problem via nonparametric spectral methods based on the fast Fourier transform. Our approach builds on the ideas underlying the spectral envelope for categorical time series as described in Stoffer, Tyler & McDougall (1993).

In particular, let $g\{X_1(t)\}$ and $h\{X_2(t)\}$ be the real-valued time series obtained via nonconstant transformations g and h of the categorical sequences $X_1(t)$ and $X_2(t)$, such that $g\{X_1(t)\}$ has continuous spectral density $f_{gg}(\omega)$, and $h\{X_2(t)\}$ has continuous spectral density $f_{hh}(\omega)$, with $f_{gh}(\omega)$ being the complex-valued cross-spectral density. A measure of the degree of similarity between the sequences $g\{X_1(t)\}$ and $h\{X_2(t)\}$ at frequency

$\omega \in (-\pi, \pi]$ is the squared coherency

$$K^2_{gh}(\omega) = \frac{|f_{gh}(\omega)|^2}{f_{gg}(\omega) f_{hh}(\omega)}.$$

The value of $K^2_{gh}(\omega)$ will, of course, depend on the choices of the transformations g and h. If $X_1(t)$ and $X_2(t)$ are independent, then so are $g\{X_1(t)\}$ and $h\{X_2(t)\}$ for any nonconstant transformations g and h, and hence $K^2_{gh}(\omega) = 0$ for any such g and h. Our goal is to find g and h to maximise, under various constraints, the squared coherency $K^2_{gh}(\omega)$. If the maximised value of $K^2_{gh}(\omega)$ is small we can say that the two sequences $X_1(t)$ and $X_2(t)$ do not match at frequency ω. If the maximised value of $K^2_{gh}(\omega)$ is large, then the resulting transformations g and h can help in understanding the nature of the similarity between the two sequences. It should be noted that, if g' and h' are location and scale changes of g and h, respectively, then $K^2_{g'h'}(\omega) = K^2_{gh}(\omega)$.

We identify the categorical sequences $X_i(t)$ with the $k_i \times 1$ point processes $Y_i(t)$ defined by

$$Y_i(t) = \begin{cases} e_{ij} & \text{if } X_i(t) = c_{ij} \text{ for } j = 1, \ldots, k_i, \\ 0_i & \text{if } X_i(t) = c_{ik_i+1}, \end{cases}$$

for $i = 1, 2$, where e_{ij} is a $k_i \times 1$ vector with a one in the jth position and zeros elsewhere, and 0_i is a $k_i \times 1$ vector of zeros. Note that there is a one-to-one correspondence between $X_1(t)$ and $Y_1(t)$, and between $X_2(t)$ and $Y_2(t)$. We assume throughout the existence of the $k_i \times k_i$ $(i = 1, 2)$, nonsingular spectral density matrices $f_{11}(\omega)$ and $f_{22}(\omega)$ of $Y_1(t)$ and $Y_2(t)$, respectively, and denote the $k_1 \times k_2$ cross-spectral matrix between $Y_1(t)$ and $Y_2(t)$ by $f_{12}(\omega)$.

2. CANONICAL VARIATE SERIES

Brillinger (1975, Ch. 10) discusses a time series extension of canonical correlation analysis that could be used here as a special case. Briefly, consider real constants μ_i and $k_i \times 1$ linear filters $\{\beta_i(j)\}$ such that $\sum \|\beta_i(j)\| < \infty$ $(i = 1, 2)$. The real-valued univariate series

$$Z_1(t) = \mu_1 + \sum_{j=-\infty}^{\infty} \beta_1'(t-j) Y_1(j), \quad Z_2(t) = \mu_2 + \sum_{j=-\infty}^{\infty} \beta_2'(t-j) Y_2(j),$$

having maximum squared-coherency $K^2_{12}(\omega)$ at each ω subject to $b_i^*(\omega) f_{ii}(\omega) b_i(\omega) = 1$, for $i = 1, 2$, where $b_i(\omega)$ is the Fourier transform of $\{\beta_i(j)\}$, and * denotes complex conjugate transpose, are given by finding the largest scalar $\lambda(\omega)$ such that, with $f_{21}(\omega) = f_{12}^*(\omega)$,

$$f_{22}(\omega)^{-\frac{1}{2}} f_{21}(\omega) f_{11}(\omega)^{-1} f_{12}(\omega) f_{22}(\omega)^{-\frac{1}{2}} \gamma(\omega) = \lambda(\omega)\gamma(\omega).$$

The maximum squared-coherency achieved between $Z_1(t)$ and $Z_2(t)$ is $\lambda(\omega)$, and $b_1(\omega)$ and $b_2(\omega)$ are taken proportional to $f_{11}(\omega)^{-1} f_{12}(\omega) f_{22}(\omega)^{-\frac{1}{2}} \gamma(\omega)$ and $f_{22}(\omega)^{-\frac{1}{2}} \gamma(\omega)$, respectively.

An interpretation of the maximal squared coherency can be given that is consistent with the notion of scaling and the spectral envelope established in Stoffer, Tyler & McDougall (1993). Specifically, we may regard $\lambda(\omega)$ as a coherency envelope, in the following sense. Let b_i be complex-valued $k_i \times 1$ vectors, $i = 1, 2$, and consider the scaled complex-valued processes $Z_1(t, b_1) = b_1^* Y_1(t)$ and $Z_2(t, b_2) = b_2^* Y_2(t)$ having squared coherency $K^2_{12}(\omega, b_1, b_2)$. If we find b_1 and b_2 so that the squared coherency between $Z_1(t, b_1)$ and $Z_2(t, b_2)$ is maximised at a particular frequency $\omega = \omega_0$, then the maximum

squared coherency at frequency ω_0 is $\lambda(\omega_0)$, and the complex-valued scalings for which the maximum squared coherency is achieved are proportional to

$$b_1(\omega_0) = f_{11}(\omega_0)^{-1} f_{12}(\omega_0) f_{22}(\omega_0)^{-\frac{1}{2}} \gamma(\omega_0), \quad b_2(\omega_0) = f_{22}(\omega_0)^{-\frac{1}{2}} \gamma(\omega_0).$$

Thus, for any nonzero complex vectors b_1 and b_2, $K_{12}^2(\omega, b_1, b_2) \leqslant \lambda(\omega)$, with equality when b_1 is proportional to $b_1(\omega)$ and b_2 is proportional to $b_2(\omega)$.

Although this method finds the maximal coherency, it is perhaps too brutal to be used in applications. A drawback is that one arrives at complex-valued scalings that are different for each sequence and interpretation may be difficult. In addition, the real-valued canonical series $g\{X_1(t)\} = Z_1(t)$ and $h\{X_2(t)\} = Z_2(t)$, which are obtained by linear filters, may be difficult to interpret because of the complexity of the transformations. Also, for moderate sample sizes, the estimated maximal squared coherency might be close to one even if the two series are independent. That is, the significance levels, the null hypothesis being that the two series are not coherent, may be too large to be useful. Instead, to enhance interpretability, we concentrate on the problem of scaling where we require the scales to be real valued. Nevertheless, the maximal coherency gives an upper bound that can be used as a benchmark for other techniques.

3. Models and applications

3·1. *Detecting a common signal: Local alignment*

In local alignment, we assume that the two sequences under investigation are in phase, and that each sequence contains at most one common signal or pattern. They may be subsequences of larger sequences. The case of global alignment, where we do not assume that the sequences are in phase but may share several different common signals, will be considered in § 3·2. Waterman & Vingron (1994) provide a discussion of local and global alignments as they pertain to DNA sequences. A general theory for the real-valued scaling problem is developed in the Appendix. In this section we assume that the sequences are defined on the same state-space and we set $k_1 = k_2 = k$.

Stoffer (1987) proposes the local model

$$Y_i(t) = p_i + S(t) + e_i(t), \tag{3·1}$$

where $p_i = (p_{i1}, \ldots, p_{ik})'$, and $S(t)$ is a zero-mean realisation of a stationary $k \times 1$ vector-valued time series that is uncorrelated with the zero-mean, stationary $k \times 1$ vector-valued series $e_i(t)$ $(i = 1, 2)$. There may, however, be some dependence structure between $S(t)$ and $e_i(t)$; see Stoffer (1987) for details. Furthermore, $S(t)$ has a $k \times k$ spectral density matrix $f_{ss}(\omega)$, and $e_i(t)$ $(i = 1, 2)$ have common $k \times k$ spectra denoted by $f_{ee}(\omega)$. Here, $S(t)$ is a stochastic signal that is common to both time series $X_1(t)$ and $X_2(t)$, or, equivalently, $Y_1(t)$ and $Y_2(t)$.

Let $\beta = (\beta_1, \ldots, \beta_k)' \in \mathbb{R}^k$ $(\beta \neq 0)$ be a vector of common scalings associated with the categories $\{c_1, \ldots, c_k\}$ of both sequences $X_i(t)$ $(i = 1, 2)$. Define the real-valued series $X_i(t, \beta) = \beta' Y_i(t)$ for $i = 1, 2$. We will show in the Appendix, Proposition A·2, that, under the model conditions (3·1), the optimal strategy is to select common scales for both sequences $X_1(t)$ and $X_2(t)$. The model can now be rewritten, for $i = 1, 2$, as

$$X_i(t; \beta) = \beta' p_i + \beta' S(t) + \beta' e_i(t).$$

Let $f_{11}(\omega; \beta)$ be the spectrum of scaled process $X_1(t; \beta)$; similarly, let $f_{22}(\omega; \beta)$ denote the spectrum of $X_2(t; \beta)$ and let $f_{12}(\omega; \beta)$ denote the cross-spectrum between $X_1(t; \beta)$ and

$X_2(t;\beta)$. The following identities hold:

$$f_{ii}(\omega;\beta)=\beta' f_{ss}(\omega)\beta+\beta' f_{ee}(\omega)\beta=\beta' f_{ss}^{re}(\omega)\beta+\beta' f_{ee}^{re}(\omega)\beta \quad (i=1,2),$$

$$f_{12}(\omega;\beta)=\beta' f_{ss}(\omega)\beta=\beta' f_{ss}^{re}(\omega)\beta.$$

Both $f_{ss}^{im}(\omega)$ and $f_{ee}^{im}(\omega)$ are skew symmetric, where the supersripts 're' and 'im' denote the real part and the imaginary part, respectively, of the complex matrix. The coherence between $X_1(t;\beta)$ and $X_2(t;\beta)$ is

$$K_{12}(\omega;\beta)=\frac{\beta' f_{ss}^{re}(\omega)\beta}{\beta' f_{ss}^{re}(\omega)\beta+\beta' f_{ee}^{re}(\omega)\beta}. \tag{3·2}$$

Setting $b=\{f_{ss}^{re}(\omega)+f_{ee}^{re}(\omega)\}^{\frac{1}{2}}\beta$, subject to the standardisation $b'b=1$, we write (3·2) as

$$K_{12}(\omega;b)=b'\{f_{ss}^{re}(\omega)+f_{ee}^{re}(\omega)\}^{-\frac{1}{2}} f_{ss}^{re}(\omega)\{f_{ss}^{re}(\omega)+f_{ee}^{re}(\omega)\}^{-\frac{1}{2}}b. \tag{3·3}$$

It follows that the maximum value of (3·3) is the largest scalar $\lambda(\omega)$ such that

$$\{f_{ss}^{re}(\omega)+f_{ee}^{re}(\omega)\}^{-\frac{1}{2}} f_{ss}^{re}(\omega)\{f_{ss}^{re}(\omega)+f_{ee}^{re}(\omega)\}^{-\frac{1}{2}}b(\omega)=\lambda(\omega)b(\omega). \tag{3·4}$$

The optimal scaling, $\beta(\omega)$, is taken proportional to $\{f_{ss}^{re}(\omega)+f_{ee}^{re}(\omega)\}^{-\frac{1}{2}}b(\omega)$. This value of $\beta(\omega)$ will maximise the coherency at frequency ω between the two sequences, with the maximum value being $\lambda(\omega)$. That is, $K_{12}(\omega;\beta)\leqslant K_{12}\{\omega;\beta(\omega)\}=\lambda(\omega)$, with equality when β is proportional to $\beta(\omega)$.

If $\hat{f}_{ij}(\omega)$ denotes consistent estimators of the corresponding spectra and cross-spectra, $f_{ij}(\omega)$, of the $Y_i(t)$ processes, for $i,j=1,2$, consistent estimators of $f_{ss}^{re}(\omega)$ and $f_{ss}^{re}(\omega)+f_{ee}^{re}(\omega)$, under the model conditions (3·1), are given by

$$\hat{f}_{ss}^{re}(\omega)=\{\hat{f}_{12}^{re}(\omega)+\hat{f}_{21}^{re}(\omega)\}/2=\{\hat{f}_{12}^{re}(\omega)+\hat{f}_{12}^{re}(\omega)'\}/2, \tag{3·5}$$

$$\hat{f}_{ss}^{re}(\omega)+\hat{f}_{ee}^{re}(\omega)=\{\hat{f}_{11}^{re}(\omega)+\hat{f}_{22}^{re}(\omega)\}/2. \tag{3·6}$$

These lead to consistent estimators of the maximal coherency and the corresponding optimal scalings via (3·4). If $\lambda(\omega)$ is a unique root, from which it follows that $\lambda(\omega)>0$, then the asymptotic distributions of the estimator of $\lambda(\omega)$ and the corresponding estimator of $\beta(\omega)$ based on (3·5) and (3·6) are normal (Stoffer, Tyler & McDougall, 1993).

An alternative approach is to test the null hypothesis that there is no signal common to both sequences, that is, $f_{ss}(\omega)=0$. Note that, if the model (3·1) is correct, this hypothesis is equivalent to $K_{12}(\omega;\beta)=0$ for any scaling β. For a fixed value of β, a frequency-based F-test for testing for a common signal in the scaled sequences $X_1(t;\beta)$ and $X_2(t;\beta)$ was described in Stoffer (1987) by extending the work of Brillinger (1980) to the discrete-valued time series case. We now extend that approach by selecting $\beta(\omega)\in\mathbb{R}^k$, $\beta(\omega)\neq 0$, to maximise the F-statistic at each Fourier frequency.

Given data $\{Y_i(t), t=1,\ldots,n\}$, for $i=1,2$, let

$$d_i(\omega)=(2\pi n)^{-\frac{1}{2}}\sum_{t=1}^{n} Y_i(t)\exp(-i\omega t)$$

be the corresponding finite Fourier transform. Let $d_i(\omega;\beta)$ denote the Fourier transform of the scaled series $X_i(t;\beta)$ $(i=1,2)$ and note that $d_i(\omega;\beta)=\beta' d_i(\omega)$ $(i=1,2)$. The average transform of the scaled series is

$$d_+(\omega;\beta)=\{d_1(\omega;\beta)+d_2(\omega;\beta)\}/2=\beta' d_+(\omega),$$

where $d_+(\omega)=\{d_1(\omega)+d_2(\omega)\}/2$.

Let ω_j be a frequency of interest and consider the collection of $(2M+1)$, for M fixed, Fourier frequencies centred at ω_j, $\{\omega_{j+m} = \omega_j + 2\pi m/n;\ m = -M, \ldots, M\}$, but excluding $\omega = 0, \pi$. Define

$$H(\omega_j;\ \beta) = \sum_{m=-M}^{M} |d_+(\omega_{j+m};\ \beta)|^2,$$

$$E(\omega_j;\ \beta) = \sum_{i=1}^{2} \sum_{m=-M}^{M} |d_i(\omega_{j+m};\ \beta) - d_+(\omega_{j+m};\ \beta)|^2.$$

For fixed β, it can be shown that, assuming the $Y_i(t)$ processes are mixing, as $n \to \infty$,

$$H(\omega_j;\ \beta) \to \{\beta' f_{ss}^{re}(\omega_j)\beta + \tfrac{1}{2}\beta' f_{ee}^{re}(\omega_j)\beta\}\chi_\nu^2/2, \tag{3·7}$$

$$E(\omega_j;\ \beta) \to \{\beta' f_{ee}^{re}(\omega_j)\beta\}\chi_\nu^2/2 \tag{3·8}$$

in distribution, where the χ^2 variates are independent, each with $\nu = 2(2M+1)$ degrees of freedom (Brillinger, 1980). From (3·7) and (3·8) it is seen that a frequency and scale dependent F-statistic that can be used for testing the presence of a common signal is

$$F(\omega_j;\ \beta) = \frac{2H(\omega_j;\ \beta)}{E(\omega_j;\ \beta)},$$

which, under our null hypothesis, has an asymptotic $F(\nu, \nu)$ distribution.

We seek the scaling $\beta(\omega_j) \neq 0$ that maximises $F(\omega_j;\ \beta)$ at frequency ω_j, for each ω_j of interest. First define the matrices $H(\omega_j)$ and $E(\omega_j)$ as

$$H(\omega_j) = \sum_{m=-M}^{M} d_+(\omega_{j+m}) d_+^*(\omega_{j+m}),$$

$$E(\omega_j) = \sum_{i=1}^{2} \sum_{m=-M}^{M} \{d_i(\omega_{j+m}) - d_+(\omega_{j+m})\}\{d_i(\omega_{j+m}) - d_+(\omega_{j+m})\}^*.$$

We can then write

$$F(\omega_j;\ \beta) = \frac{2\beta' H(\omega_j)\beta}{\beta' E(\omega_j)\beta} = \frac{2\beta' H^{re}(\omega_j)\beta}{\beta' E^{re}(\omega_j)\beta},$$

so that the maximum value of $F(\omega_j;\ \beta)$ is the largest scalar, $\lambda_F(\omega_j)$, satisfying

$$2H^{re}(\omega_j)b(\omega_j) = \lambda_F(\omega_j)E^{re}(\omega_j)b(\omega_j);$$

the optimal scaling, $\beta(\omega_j)$, is taken proportional to $E^{re}(\omega_j)^{-\frac{1}{2}}b(\omega_j)$. This value of $\beta(\omega_j)$ will maximise the F-statistic at frequency ω_j, with the maximum value of the F-statistic being $\lambda_F(\omega_j)$.

Under the assumption that $Y_1(t)$ and $Y_2(t)$ are mixing, the asymptotic $(n \to \infty)$ null distribution of $\lambda_F(\omega_j)$ is that of Roy's largest root, the distribution of which can be found in Muirhead (1982, pp. 481–4) along with some approximations; for some tables see Anderson (1984, Appendix B). Finite sample null distributions are discussed in § 4·2.

There is a link between the hypothesis testing approach and consistent estimation as described in (3·5) and (3·6). Note that the averaged periodograms

$$\bar{f}_{ss}^{re}(\omega_j) = (2M+1)^{-1} \sum_{m=-M}^{M} \frac{1}{2}\{I_{12}^{re}(\omega_{j+m}) + I_{21}^{re}(\omega_{j+m})\}, \tag{3·9}$$

$$\bar{f}_{ss}^{re}(\omega_j) + \bar{f}_{ee}^{re}(\omega_j) = (2M+1)^{-1} \sum_{m=-M}^{M} \frac{1}{2}\{I_{11}^{re}(\omega_{j+m}) + I_{22}^{re}(\omega_{j+m})\}, \tag{3·10}$$

where $I_{ij}(\omega) = d_i(\omega)d_j^*(\omega)$ $(i, j = 1, 2)$, can be written in terms of $H^{re}(\omega_j)$ and $E^{re}(\omega_j)$. In particular, (3·9) and (3·10) can be written as

$$\bar{f}_{ss}^{re}(\omega_j) = (2M+1)^{-1}\{H^{re}(\omega_j) - \tfrac{1}{2}E^{re}(\omega_j)\}, \quad \bar{f}_{ee}^{re}(\omega_j) = (2M+1)^{-1}E^{re}(\omega_j).$$

Using (3·9) and (3·10), we may write

$$\bar{K}_{12}(\omega_j; \beta) = \frac{\beta'\bar{f}_{ss}^{re}(\omega_j)\beta}{\beta'\bar{f}_{ss}^{re}(\omega_j)\beta + \beta'\bar{f}_{ee}^{re}(\omega_j)\beta} = \frac{H(\omega_j; \beta) - \frac{1}{2}E(\omega_j; \beta)}{H(\omega_j; \beta) + \frac{1}{2}E(\omega_j; \beta)} = \frac{F(\omega_j; \beta) - 1}{F(\omega_j; \beta) + 1},$$

provided that $F(\omega_j; \beta) \geqslant 1$; otherwise, set $\bar{K}_{12}(\omega_j; \beta) = 0$. It is clear that, for any fixed M, the scaling $\beta(\omega_j)$ that maximises the F-statistic, $F(\omega_j; \beta)$, also maximises $\bar{K}_{12}(\omega_j; \beta)$.

3·2. *Detecting common signals: Global alignment*

We now extend the model of § 3·1 to include the possibility that there is more than one signal common to each sequence, and that the sequences are not necessarily aligned. Here, we assume the global model

$$Y_1(t) = p_1 + \sum_{j=1}^{q} S_j(t) + e_1(t), \quad Y_2(t) = p_2 + \sum_{j=1}^{q} S_j(t - \tau_j) + e_2(t),$$

where $S_j(t)$ $(j = 1, \ldots, q)$ are zero-mean realisations of stationary $k \times 1$ vector-valued time series that are mutually uncorrelated, and in addition are uncorrelated with the zero-mean, stationary $k \times 1$ vector-valued series $e_1(t)$ and $e_2(t)$. Furthermore, $S_j(t)$ has $k \times k$ spectral density matrix $g_{jj}(\omega)$ $(j = 1, \ldots, q)$ and $e_i(t)$ $(i = 1, 2)$ are uncorrelated and have common $k \times k$ spectra denoted by $f_{ee}(\omega)$. It is hypothesised that the processes $S_j(t)$ are stochastic signals that are common to both time series $X_1(t)$ and $X_2(t)$, or, equivalently, $Y_1(t)$ and $Y_2(t)$. We do not specify $\tau_1, \ldots, \tau_q$, or the integer $q \geqslant 0$; although the problem of their estimation is interesting, we do not consider that problem here. Of particular interest is whether or not $q = 0$.

Using the notation of the previous sections, we note the following conditions:

$$f_{11}(\omega) = f_{22}(\omega) = \sum_{j=1}^{q} g_{jj}(\omega) + f_{ee}(\omega), \quad f_{12}(\omega) = \sum_{j=1}^{q} g_{jj}(\omega)\exp(i\omega\tau_j).$$

Let $\beta = (\beta_1, \ldots, \beta_k)' \in \mathbb{R}^k$ $(\beta \neq 0)$ be a vector of common scalings, and write the scaled processes as $X_i(t, \beta) = \beta' Y_i(t)$ $(i = 1, 2)$. Then the squared coherency between $X_1(t, \beta)$ and $X_2(t, \beta)$ is

$$K_{12}^2(\omega; \beta) = \frac{|\sum_{j=1}^{q} \beta' g_{jj}^{re}(\omega)\beta \exp(i\omega\tau_j)|^2}{|\beta' f^{re}(\omega)\beta|^2}, \tag{3·11}$$

where we have written $f(\omega) = f_{11}(\omega) = f_{22}(\omega)$. We will show in the Appendix, Proposition A·2, that the optimal strategy is to select common scalings.

Setting $b = f^{re}(\omega)^{\frac{1}{2}}\beta$, with the constraint $b'b = 1$, we write (3·11) as

$$K_{12}^2(\omega; b) = \left| b' \left\{ \sum_{j=1}^{q} f^{re}(\omega)^{-\frac{1}{2}} g_{jj}^{re}(\omega) f^{re}(\omega)^{-\frac{1}{2}} \exp(i\omega\tau_j) \right\} b \right|^2. \tag{3·12}$$

Define the complex-valued matrix

$$Q(\omega) = \sum_{j=1}^{q} f^{re}(\omega)^{-\frac{1}{2}} g_{jj}^{re}(\omega) f^{re}(\omega)^{-\frac{1}{2}} \exp(i\omega\tau_j) = Q^{re}(\omega) + iQ^{im}(\omega),$$

and note that $Q^{re}(\omega)$ and $Q^{im}(\omega)$ are symmetric, but not necessarily positive definite, matrices. Now, write (3·12) as

$$K_{12}^2(\omega; b) = \{b'Q^{re}(\omega)b\}^2 + \{b'Q^{im}(\omega)b\}^2. \tag{3·13}$$

The squared coherency, (3·13), can be maximised via Proposition A·2. In particular, the recursion

$$b_j = \mathscr{E}_1\{Q^{re}(\omega)b_{j-1}b_{j-1}'Q^{re}(\omega) + Q^{im}(\omega)b_{j-1}b_{j-1}'Q^{im}(\omega)\} \quad (j = 1, 2, \ldots) \tag{3·14}$$

can be employed. In (3·14), the notation $\mathscr{E}_1(A)$ means the eigenvector associated with the largest eigenvalue of A. The algorithm is initialised by setting b_0 to either $\mathscr{E}_1\{Q^{re}(\omega)^2\}$ or $\mathscr{E}_1\{Q^{im}(\omega)^2\}$ depending on which vector produces the larger value of $K_{12}^2(\omega; b)$. From Propositions A·1 and A·2 we know that $K_{12}^2(\omega; b_j) \geqslant K_{12}^2(\omega; b_{j-1})$. Maximisation can also be achieved iteratively via Newton–Raphson; the details are given in the Appendix. The optimal scaling vector at any particular frequency, $\beta(\omega)$, is taken proportional to $f^{re}(\omega)^{-\frac{1}{2}}b(\omega)$, where $b(\omega)$ is the maximising vector.

Given consistent estimators $\hat{f}_{11}(\omega)$, $\hat{f}_{22}(\omega)$ and $\hat{f}_{12}(\omega)$, of the corresponding spectral matrices, we estimate $f(\omega)$ by

$$\hat{f}(\omega) = \tfrac{1}{2}\{\hat{f}_{11}(\omega) + \hat{f}_{22}(\omega)\}.$$

Also, since

$$\{f_{12}^{re}(\omega) + f_{12}^{re}(\omega)'\}/2 = \{f_{12}^{re}(\omega) + f_{21}^{re}(\omega)\}/2 = \sum_{j=1}^{q} g_{jj}^{re}(\omega)\cos(\omega\tau_j),$$

$$\{f_{12}^{im}(\omega) + f_{12}^{im}(\omega)'\}/2 = \{f_{12}^{im}(\omega) - f_{21}^{im}(\omega)\}/2 = \sum_{j=1}^{q} g_{jj}^{re}(\omega)\sin(\omega\tau_j),$$

consistent estimators of $Q^{re}(\omega)$ and $Q^{im}(\omega)$ are given by

$$\hat{Q}^{re}(\omega) = \{\hat{f}_{11}^{re}(\omega) + \hat{f}_{22}^{re}(\omega)\}^{-\frac{1}{2}}\{\hat{f}_{12}^{re}(\omega) + \hat{f}_{12}^{re}(\omega)'\}\{\hat{f}_{11}^{re}(\omega) + \hat{f}_{22}^{re}(\omega)\}^{-\frac{1}{2}},$$

$$\hat{Q}^{im}(\omega) = \{\hat{f}_{11}^{re}(\omega) + \hat{f}_{22}^{re}(\omega)\}^{-\frac{1}{2}}\{\hat{f}_{12}^{im}(\omega) + \hat{f}_{12}^{im}(\omega)'\}\{\hat{f}_{11}^{re}(\omega) + \hat{f}_{22}^{re}(\omega)\}^{-\frac{1}{2}},$$

respectively.

It is more difficult to obtain asymptotic null distributions for this case than in local alignment and we do not address this problem here. Finite sample null distributions are discussed in § 4·2.

4. Discussion and examples

4·1. *Numerical behaviour*

We first focus on the numerical methods of § 3·2. Fix ω and drop it from the notation. Since $K_{12}^2(b)$ given in (3·13) is a quartic, numerical maximisation should present no problem when there is a unique maximum and a good starting value. Proposition A·1 in the Appendix provides such a starting value, namely b_0 described after (3·14). Our experience is that $K_{12}^2(b_0)$ is frequently very close to the maximum and the difference between the maximum value, $K_{12}^2(\tilde{b})$, say, and $K_{12}^2(b_0)$ rarely exceeds 5%. It is often enough to carry out a few iterations of (3·14). If a high degree of accuracy is desired, then one can run the cheaper (3·14) a small number of times and then run the more costly Newton–Raphson to accelerate the convergence. We have found this combination to be extremely fast and to be practicable even with large sequences. Some computation times are reported in § 4·3 and convergence of the algorithm is discussed further in the Appendix.

On a rare occasion, we found a difference of about 12% between $K_{12}^2(\tilde{b})$ and $K_{12}^2(b_0)$. The Newton–Raphson procedure, however, found the maximum in four iterations using fourth place accuracy. In this example, $k=3$ and the vector of scalings, $\beta=(\beta_1, \beta_2, \beta_3)'$ is scaled so that β_3 is held fixed at 1. Figure 1(a) shows the surface of $K_{12}^2(\beta)$, defined in (3·11), as β_1 and β_2 vary. In this particular example,

$$Q^{\text{re}} = \begin{pmatrix} 0{\cdot}41 & 0{\cdot}15 & -0{\cdot}06 \\ 0{\cdot}15 & -0{\cdot}20 & 0{\cdot}25 \\ -0{\cdot}06 & 0{\cdot}25 & 0{\cdot}49 \end{pmatrix}, \quad Q^{\text{im}} = \begin{pmatrix} -0{\cdot}13 & -0{\cdot}47 & -0{\cdot}08 \\ -0{\cdot}47 & 0{\cdot}16 & -0{\cdot}12 \\ -0{\cdot}08 & -0{\cdot}12 & 0{\cdot}06 \end{pmatrix}. \tag{4·1a}$$

Note that, although there is a unique maximum, the surface contains ridges of near maximum height. The global maximum is relatively small, corresponding to a squared coherency of less than 45%.

Figure 1(b) shows an example of a favourable situation where the maximal coherency is near one. In this case convergence, using fourth place accuracy, was attained after one iteration of (3·14); in this example,

$$Q^{\text{re}} = \begin{pmatrix} -0{\cdot}07 & 0{\cdot}07 & -0{\cdot}35 \\ 0{\cdot}07 & -0{\cdot}09 & 0{\cdot}67 \\ -0{\cdot}35 & 0{\cdot}67 & -0{\cdot}18 \end{pmatrix}, \quad Q^{\text{im}} = \begin{pmatrix} 0{\cdot}14 & -0{\cdot}01 & -0{\cdot}03 \\ -0{\cdot}01 & 0{\cdot}05 & -0{\cdot}15 \\ -0{\cdot}03 & -0{\cdot}15 & -0{\cdot}05 \end{pmatrix}. \tag{4·1b}$$

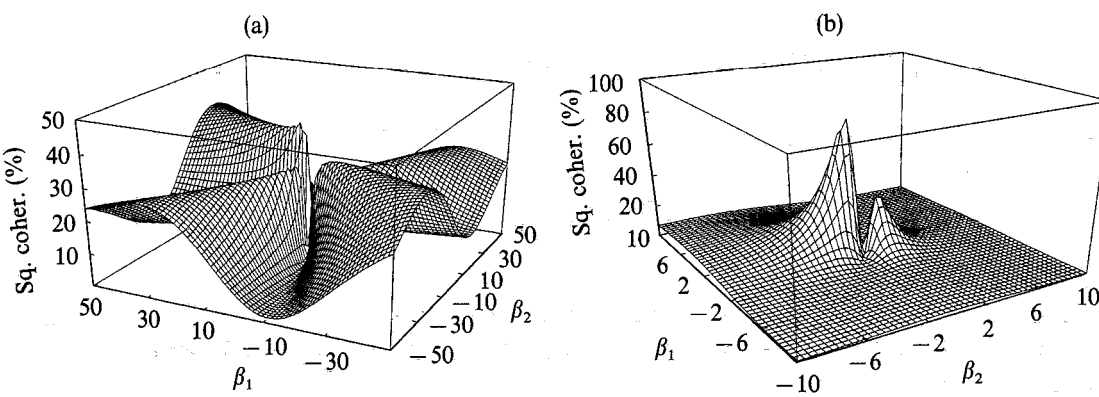

Fig. 1. Surface of $K_{12}^2(\beta)$ (a) based on (4·1a), (b) based on (4·1b) as described in §4·1.

4·2. *Simulation results*

We briefly examine the finite sample null behaviour of the maximum F-statistic, $\lambda_F(\omega)$, described in § 3·1, and the global alignment maximal squared coherency, $\hat{\lambda}_{\text{GA}}(\omega)$, say, described in § 3·2. Here, and henceforth, frequency ω is measured in cycles per unit time. In the first example we simulated two independent sequences of $n=1000$ observations. The first series is a white noise sequence on four categories with $p_{11}=0{\cdot}18$, $p_{12}=0{\cdot}31$, $p_{13}=0{\cdot}29$, $p_{14}=0{\cdot}22$; the second series is also discrete white noise with $p_{21}=0{\cdot}30$, $p_{22}=0{\cdot}21$, $p_{23}=0{\cdot}19$, $p_{24}=0{\cdot}30$. These proportions and sample lengths pertain to the second example of § 4·3 and we will use the results there. The spectral estimates were based on averaged periodograms with $M=5$. The 5% and 1% critical values for $\lambda_F(\omega)$ based on 2500 repeated simulations were 3·68 and 4·62, respectively. Referring to the

discussion in the final paragraph of § 3·1, we note that 3·68 corresponds to a squared coherency of $\{(3{\cdot}68-1)/(3{\cdot}68+1)\}^2 = 32{\cdot}8\%$, and 4·62 corresponds to a squared coherency of 41·5%. In a similar simulation, the 5% and 1% critical values of $\hat{\lambda}_{GA}(\omega)$ were 47·5% and 56·2%, respectively.

To examine further the finite sample null distribution of $\lambda_F(\omega)$ when signals are present, we repeated the simulation described above, except that $p_{ij} = \frac{1}{4}$, and then the first series, $X_1(t)$, has an additional 80% chance that every third observation is c_1, while the second series, $X_2(t)$, has an additional 70% chance that every fourth observation is c_1. Thus, each series has its own individual signal, and both series have a common, but weak, signal at $\omega = \frac{1}{12}$ and its harmonics. In this case, the 5% and 1% critical values were 3·71 and 4·94, respectively. These critical values apply to any frequency except $\omega = \frac{1}{12}$, its harmonics and the endpoints.

Finally, in a simulation of 10 000 observations of Roy's largest root statistic with 22 numerator and denominator degrees of freedom, we found 3·59 and 4·76 to be the approximate asymptotic 5% and 1% critical values of $\lambda_F(\omega)$.

4·3. *Applications to DNA sequences*

First, we compare two DNA sequences each of approximately 4000 base pairs (bp) long. One is a gene of the Epstein–Barr virus (EBV) and the other is a gene of the herpes virus saimiri (HVS). Both genes have been labelled BNRF1 since similarities between the two have been noted by molecular biologists. The spectral envelopes, not shown here, indicate that each gene contains a strong signal at a cycle of approximately $\frac{1}{3}$. The EBV BNRF1 gene was studied extensively in Stoffer, Tyler, McDougall & Schachtel (1993).

Figure 2 compares three methods of computing maximal squared coherency. Figure 2(a) shows the canonical variates approach of § 2, $\hat{\lambda}_{CV}(\omega)$, say; Fig. 2(b) shows the local alignment approach first mentioned in § 3·1, equations (3·3)–(3·6), $\hat{\lambda}^2_{LA}(\omega)$, say; and Fig. 2(c) shows the global alignment approach of § 3·2, $\hat{\lambda}_{GA}(\omega)$, say. In each case we used a triangular set of $2M+1$, $M = 15$, weights to do the smoothing. Each method appears to give similar results in that the two genes show a strong match at a frequency of one cycle every three bp, but the evidence is much stronger in $\hat{\lambda}_{CV}(\omega)$ and $\hat{\lambda}_{GA}(\omega)$. The fact that $\hat{\lambda}^2_{LA}(\frac{1}{3})$ is about half of the value of $\hat{\lambda}_{CV}(\frac{1}{3})$ and $\hat{\lambda}_{GA}(\frac{1}{3})$ is most likely because the genes are not homogeneous. We also note that, at frequency $\frac{1}{3}$, the global alignment method attains nearly the same squared coherency, 0·82, as the canonical variates method, 0·86. Thus, while enhancing the interpretation, little information is lost in considering common real-valued scales.

With the global alignment model, the scales at the one-third frequency were $A = 3{\cdot}6$, $C = 6{\cdot}8$, $G = 7{\cdot}1$, $T = 0$, which suggests that the match is due to strong bonding, that is $C = G$, and that the nucleotides A and T are different and separate from $C = G$. Quite a different outcome occurs when we consider the local alignment model, although this model would not be considered entirely appropriate for this situation. In this case the scales were $A = 9{\cdot}3$, $C = 1{\cdot}2$, $G = 8{\cdot}8$, $T = 0$, which suggests that the match occurs with the pairing of the purines, $A = G$, and the pyrimidines, $C = T$. As we shall see in the next example, this result could be true for part of the BNRF1 genes.

The entire calculation of $\hat{\lambda}_{GV}(\omega)$ and $\hat{\lambda}^2_{LA}(\omega)$, across all Fourier frequencies, took about 25 seconds each on a Pentium 100 PC, with 16 MB of RAM, using the Gauss programming language. The computation of $\hat{\lambda}_{GA}(\omega)$, to third place accuracy, took about 35 seconds; the average number of iterations was 2, and the maximum number of iterations was 4.

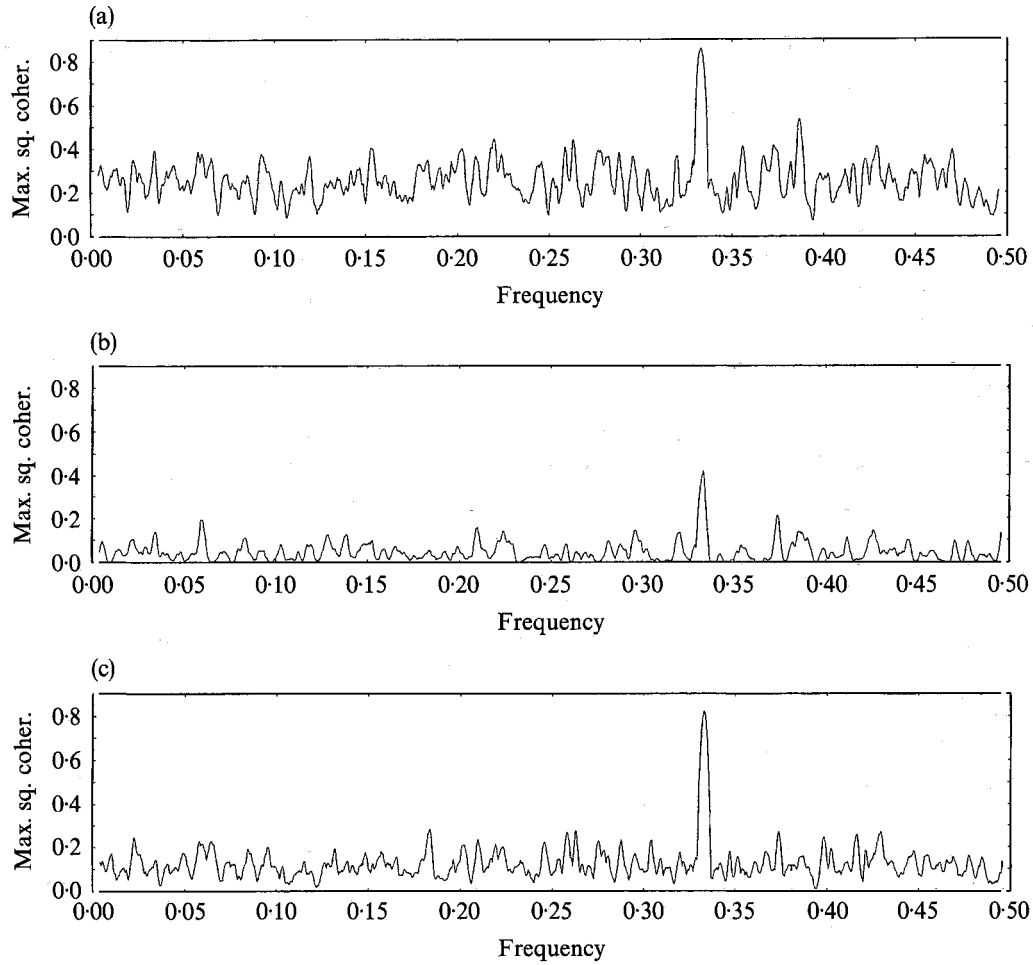

Fig. 2. Comparison of the estimated maximal squared coherency between the gene BNRF1 of EBV and of HVS using three methods as described in § 4·3: (a) canonical variates, $\hat{\lambda}_{CV}(\omega)$; (b) local alignment, $\hat{\lambda}^2_{LA}(\omega)$; (c) global alignment, $\hat{\lambda}_{GA}(\omega)$.

In Stoffer, Tyler, McDougall & Schachtel (1993), we reported a rather strange result about the gene BNRF1 of EBV. There, it was found that, although a cycle of $\frac{1}{3}$ could be found in most of the gene, the last 1000 bp appeared not to have any cyclic behaviour. Figure 3(a) shows the spectral envelope of the final 1000 bp of BNRF1 of EBV superimposed on that of HVS. Note that there is considerable power at frequency $\frac{1}{3}$ for the HVS whereas nothing is there for the EBV gene.

Nevertheless, Fig. 3(b) shows a startling result. To see if the two genes are coherent in the final 1000 bp, we used the maximum F-statistic approach described in the second part of § 3·1, with $M = 5$, which gives $\nu = 22$. The proportions of A, C, G, T in each series are given in § 4·2, and, based on that simulation, the 0·01 significance threshold is 4·62. In Fig. 3(b), it is seen that $\lambda_F(\frac{1}{3}) = 7$, which we can safely consider to be significant. Hence, while no signal was evident in this part of EBV-BNRF1 using the spectral envelope, there is evidence that it is coherent with a similar gene in HVS.

The scales at $\omega = \frac{1}{3}$ were $A = 4{\cdot}8$, $C = -1{\cdot}5$, $G = 8{\cdot}7$, $T = 0$, which suggests, to a certain degree, the pairing of the purines, $A = G$, and the pyrimidines, $C = T$. Recall that this

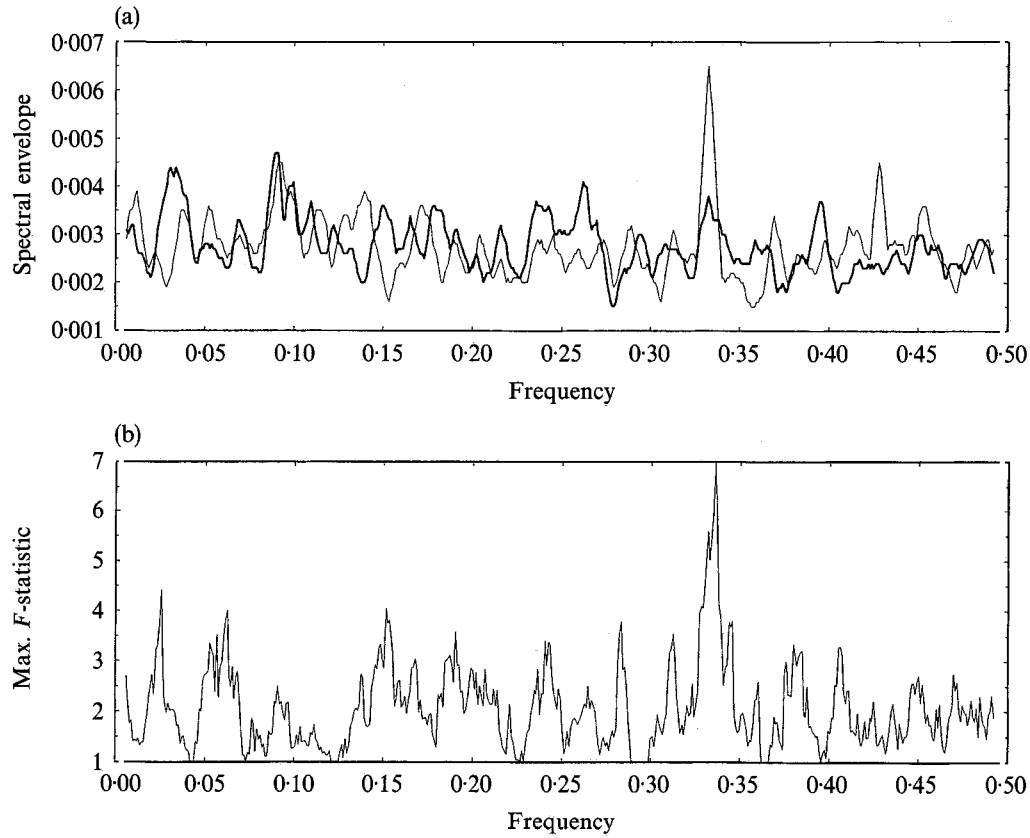

Fig. 3. (a) The spectral envelope of the final 1000 bp of BNRF1 of EBV (dark curve) and of HVS. (b) Maximum F-statistic method (§ 3·1) applied to the final 1000 bp of BNRF1 of EBV and of HVS, as described in § 4·3.

possibility was suggested in the previous example. The global alignment method applied to these subsequences yielded approximately the same results.

Acknowledgement

The work was supported in part by grants from the National Science Foundation.

Appendix

General theory and algorithms

First, consider the problem of obtaining maximal coherency with real-valued scales. That is, let $\alpha = (\alpha_1, \ldots, \alpha_{k_1})' \in \mathbb{R}^{k_1}$ $(\alpha \neq 0)$ be a vector of scalings associated with the categories of the first series, $X_1(t)$, and let $\beta = (\beta_1, \ldots, \beta_{k_2})' \in \mathbb{R}^{k_2}$ $(\beta \neq 0)$ be a vector of scalings associated with the categories of the second series, $X_2(t)$. The scaled series are $X_1(t, \alpha) = \alpha' Y_1(t)$, and $X_2(t, \beta) = \beta' Y_2(t)$. The squared coherency between $X_1(t, \alpha)$ and $X_2(t, \beta)$ can be written as

$$K_{12}^2(\omega; \alpha, \beta) = \frac{|\alpha' f_{12}(\omega)\beta|^2}{\{\alpha' f_{11}^{\mathrm{re}}(\omega)\alpha\}\{\beta' f_{22}^{\mathrm{re}}(\omega)\beta\}}. \tag{A·1}$$

Setting $a = f_{11}^{re}(\omega)^{\frac{1}{2}}\alpha$ and $b = f_{22}^{re}(\omega)^{\frac{1}{2}}\beta$, subject to the standardisations $a'a = 1$ and $b'b = 1$, and writing

$$Q(\omega) = f_{11}^{re}(\omega)^{-\frac{1}{2}} f_{12}(\omega) f_{22}^{re}(\omega)^{-\frac{1}{2}} = Q^{re}(\omega) + iQ^{im}(\omega),$$

we may write (A·1) as

$$K_{12}^2(\omega; a, b) = \{a'Q^{re}(\omega)b\}^2 + \{a'Q^{im}(\omega)b\}^2. \tag{A·2}$$

The goal is to find a and b to maximise (A·2) for each ω of interest. The maximisation can be accomplished iteratively via an algorithm based on the following proposition.

PROPOSITION A·1. *Let A and B be real-valued $k_1 \times k_2$ matrices, and let x and y be real-valued unit length vectors of dimensions k_1 and k_2, respectively. Define*

$$u(x, y) = (x'Ay)^2 + (x'By)^2,$$

and let y_0 be an arbitrary real-valued $k_2 \times 1$ unit length vector. Define the sequence of vectors x_j to be the eigenvector corresponding to the largest root of the at most rank 2, nonnegative definite matrix

$$Ay_{j-1}y_{j-1}'A' + By_{j-1}y_{j-1}'B',$$

and the sequence y_j to be the eigenvector corresponding to the largest root of the at most rank 2, nonnegative definite matrix

$$A'x_jx_j'A + B'x_jx_j'B,$$

for $j = 1, 2, \ldots$. Let $\mathscr{L}_1(.)$ and $\mathscr{E}_1(.)$ denote the largest eigenvalue function and corresponding eigenvector function, respectively, and define

$$u_{\min} = \max[\mathscr{L}_1(A'A) + \{\mathscr{E}_1'(AA')B\mathscr{E}_1(A'A)\}^2, \mathscr{L}_1(B'B) + \{\mathscr{E}_1'(BB')A\mathscr{E}_1(B'B)\}^2],$$

$$u_{\max} = \mathscr{L}_1(A'A) + \mathscr{L}_1(B'B).$$

Then,

$$u(x_{j+1}, y_{j+1}) \geqslant u(x_j, y_j) \quad (j = 1, 2, \ldots), \quad u_{\min} \leqslant u(\tilde{x}, \tilde{y}) \leqslant u_{\max},$$

where $\tilde{x}$ and $\tilde{y}$ maximise $u(x, y)$.

Proof. First note that

$$u(x, y) = x'(Ayy'A' + Byy'B')x = y'(A'xx'A + B'xx'B)y.$$

From the first relationship it follows that $u(x_{j+1}, y_j) \geqslant u(x, y_j)$, for any x of unit length, and from the second relationship it follows that $u(x_{j+1}, y_{j+1}) \geqslant u(x_{j+1}, y)$, for any y of unit length. Hence $u(x_{j+1}, y_{j+1}) \geqslant u(x_{j+1}, y_j) \geqslant u(x_j, y_j)$. The upper bound $u_{\max}$ is easily established and the lower bound is established using the singular value decompositions of A and B. □

If we set $A = Q^{re}(\omega)$ and $B = Q^{im}(\omega)$ for a given ω, the algorithm given in Proposition A·1 can be used to find the scalings, $a(\omega)$ and $b(\omega)$, say, that maximise (A·2). The algorithm is initialised by setting $b_0(\omega)$ equal to either $\mathscr{E}_1\{Q^{re}(\omega)'Q^{re}(\omega)\}$ or $\mathscr{E}_1\{Q^{im}(\omega)'Q^{im}(\omega)\}$, depending on which vector produces the lower bound $u_{\min}(\omega)$. In turn, $\alpha(\omega)$ and $\beta(\omega)$ are taken proportional to $f_{11}^{re}(\omega)^{-\frac{1}{2}}a(\omega)$ and $f_{22}^{re}(\omega)^{-\frac{1}{2}}b(\omega)$, respectively. Note that the algorithm requires only the computation of latent roots and vectors of at most rank 2, nonnegative definite matrices, regardless of the dimension of the state-spaces. Moreover, by Proposition A·1 we know that the objective function increases with each step. Since the objective function is bounded the algorithm must converge. However, the algorithm is local, and convergence to the global maximum is not guaranteed. As discussed in § 4·1, it is our experience that $u_{\min}(\omega)$ is typically close to the global maximum and that the algorithm usually converges. The maximum value of $K_{12}^2(\omega; a, b)$ is also bounded above by $\lambda_{CV}(\omega)$, which corresponds to allowing a and b to be complex as described in § 2. Hence, a plot of $u_{\min}(\omega)$ and $\min\{u_{\max}(\omega), \lambda_{CV}(\omega)\}$ across the frequencies ω can be useful in detecting potential problems. Typically, these bounds are very tight.

In § 3·2, the problem was to maximise a function of the form $u(b) = (b'Ab)^2 + (b'Bb)^2$, where A and B are $k \times k$ symmetric matrices. The maximisation can still be accomplished iteratively via Proposition A·1 in conjunction with the following result.

PROPOSITION A·2. *If, in addition to the conditions stated in Proposition* A·1, $k_1 = k_2 = k$ *and the matrices* A *and* B *are symmetric, the maximum value of* $u(x, y)$ *is attained over* (x, y) *when* $x = y$.

Proof. Let $\tilde{x}, \tilde{y}$ maximise $u(x, y)$ over all x, y of unit length. Since A and B are symmetric,

$$u(x, y) = x'(Ayy'A + Byy'B)x = y'(Axx'A + Bxx'B)y,$$

so that $\tilde{x}$ and $\tilde{y}$ satisfy $\tilde{x} = \mathscr{E}_1(A\tilde{y}\tilde{y}'A + B\tilde{y}\tilde{y}'B)$ and $\tilde{y} = \mathscr{E}_1(A\tilde{x}\tilde{x}'A + B\tilde{x}\tilde{x}'B)$. Hence there are constants λ_1 and λ_2 such that $\lambda_1\tilde{x} = (A\tilde{y}\tilde{y}'A + B\tilde{y}\tilde{y}'B)\tilde{x} = (\gamma_1 A + \gamma_2 B)\tilde{y} \equiv C\tilde{y}$, where $\gamma_1 = \tilde{x}'A\tilde{y}$ and $\gamma_2 = \tilde{x}'B\tilde{y}$, and, similarly, $\lambda_2\tilde{y} = (A\tilde{x}\tilde{x}'A + B\tilde{x}\tilde{x}'B)\tilde{y} = (\gamma_1 A + \gamma_2 B)\tilde{x} \equiv C\tilde{x}$. It follows that $\lambda_1 = \tilde{x}'C\tilde{y}$ and $\lambda_2 = \tilde{y}'C\tilde{x}$, and thus $\lambda_1 = \lambda_2 = u(\tilde{x}, \tilde{y})$. Since $\lambda_1 = \lambda_2 \equiv \lambda$, we have $\lambda^2\tilde{x} = C\lambda\tilde{y} = C^2\tilde{x}$ and $\lambda^2\tilde{y} = C\lambda\tilde{x} = C^2\tilde{y}$; that is, λ is the largest eigenvalue of C. Hence we can choose $\tilde{x} = \tilde{y} = \mathscr{E}_1(C)$. □

Propositions A·1 and A·2 lead to the recursion given in (3·14). Another important consequence of Proposition A·2 is that it gives sufficient conditions under which choosing common scales is not only parsimonious but optimal. Specifically, if $k_1 = k_2 = k$ and $Q^{\mathrm{re}}(\omega)$ and $Q^{\mathrm{im}}(\omega)$ are both symmetric, then the maximum of $K_{12}^2(\omega, a, b)$, as defined in (A·2), is achieved when $a = b$; note that the models of § 3 would produce this form.

As noted in § 3·2, (3·13) can be maximised iteratively via Newton–Raphson. For ease of discussion, we fix ω and drop it from the notation. To solve the constrained maximisation problem using Lagrange's method, we set $V(b) = K_{12}^2(b) - \eta(b'b - 1)$. Then, setting the $k \times 1$ vector $V^{(1)}(b) = \partial V(b)/\partial b = 0$, and using the fact that Q^{re} and Q^{im} are symmetric, we find $\eta = 2(b'Q^{\mathrm{re}}b) + 2(b'Q^{\mathrm{im}}b)$ so that

$$V^{(1)}(b) = (b'Q^{\mathrm{re}}b)\{Q^{\mathrm{re}} - (b'Q^{\mathrm{re}}b)I\}b + (b'Q^{\mathrm{im}}b)\{Q^{\mathrm{im}} - (b'Q^{\mathrm{im}}b)I\}b,$$

where I is the $k \times k$ identity matrix. One can then proceed by solving $V^{(1)}(b) = 0$ via Newton–Raphson.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.

BRILLINGER, D. R. (1980). Analysis of variance problems under time series models. *Handbook of Statistics*, **1**, Ed. P. R. Krishnaiah, pp. 237–78. Amsterdam: North Holland.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.

STOFFER, D. S. (1987). Walsh-Fourier analysis of discrete-valued time series. *J. Time Ser. Anal.* **8**, 449–67.

STOFFER, D. S., TYLER, D. E. & MCDOUGALL, A. J. (1993). Spectral analysis for categorical time series: Scaling and the spectral envelope. *Biometrika* **80**, 611–22.

STOFFER, D. S., TYLER, D. E., MCDOUGALL, A. J. & SCHACHTEL, G. A. (1993). Spectral analysis of DNA sequences (with Discussion). *Bull. Int. Statist. Inst.*, Bk 1, 345–61, Bk 4, 63–9.

WATERMAN, M. S. & VINGRON, M. (1994). Sequence comparison significance and Poisson approximation. *Statist. Sci.* **9**, 367–81.

[*Received September* 1995. *Revised July* 1997]