Euro. Jnl of Applied Mathematics (1998), *vol.* 9, *pp.* 55–79. Printed in the United Kingdom

Singular systems analysis as a moving-window spectral method

A. C. FOWLER[1] and G. KEMBER[2]

[1]*Mathematical Institute, Oxford University, 24–29 St. Giles', Oxford OX1 3LB, UK*
[2]*Department of Applied Mathematics, Technical University of Nova Scotia, Halifax, Nova Scotia, Canada*

(*Received 15 January 1997; in revised form 17 October 1997*)

Singular Systems Analysis (SSA), or time domain Principal Component Analysis (PCA), is most appropriately analysed in terms of local, moving-window spectral analysis. The behaviour of Empirical Orthogonal Functions (EOF) of this theory are examined, for continuously sampled data, in the limits of large and small window length, and for centre or end projection. Filters obtained by projecting on to these EOFs are shown to approximate local, *linear* band pass filters, where the EOFs depend upon the correlation structure (or the power spectral density) of the signal and the window length. Power in the spectra is not generally conserved, and projection to the endpoints of a window may not converge to the underlying signal in the absence of noise. The filters are independent of the phase of the Fourier transform, and are therefore unable to distinguish dynamically between a signal and a surrogate (phase-randomized) transform of it. Iteration of such local filters using a prediction error-based stopping criterion can and does lead to improved results, but the choice of window length must be made *a priori*. Hence, we introduce an iterative local filter with the window length being determined as part of the filtering procedure. This involves the determination of the predictability of the projected time series, and hence allows SSA to be used in a genuinely nonlinear way.

1 Moving spectral filters

It is often the case in time series analysis that one wants to monitor the existence of local periods of oscillation. A geophysical example can be found in Dehant *et al.* (1993), where certain oscillatory frequencies in the earth's gravity field are thought to be due to the motions of the earth's core, and these are buried in a noisy background time series. A medical application requiring a similar analysis is the extraction of the respiratory signal taken during REM sleep (Pilgram *et al.*, 1995) from a noisy background. Elsewhere, in neonatal respiratory monitoring (Fleming *et al.*, 1988), it is of interest when monitoring abnormalities to chart the existence of periodic breathing, indicated by a slow (12–15 second period) oscillation in the basic rhythm. In these examples, one wants a filter to select these oscillations, and moreover that it be local, insofar as the amplitude of the signal may be locally varying.

A different application is to the assessment of baroreceptor sensitivity (Robbe *et al.*, 1987). Here one measures heart rate and blood pressure, which interact in the cardiovascular system via the baroreceptor reflex, and the object is to determine the

strength of this coupling through analysis of the two time series. Since the frequency of the signals emitted by the baroreceptors is known, this can be done spectrally, if the coupling is linear. However, that is not in fact the case, and the coupling acts both ways. Therefore, one needs to separate the part of the heart rate signal which is exclusively due to forcing by variations in blood pressure. State-of-the-art methods use moving segment Fourier spectra, but it has been suggested that more flexibility is available through SSA (Vautard & Ghil, 1989) where the choice of basis functions is adaptive. SSA has been applied to the decomposition of short, noisy time series (Vautard *et al.*, 1992) into dynamical components. A medical example of such an application of SSA is to the separation of EEG potentials into trends, alpha and beta components (Mineva & Popivanov, 1996).

Thus, the main aim in this paper is to establish the equivalence between linear, moving-window spectral methods and SSA. This link is important, since it will establish the extent to which SSA improves the Fourier approach, and expose the limitations of a linear approach such as SSA for decomposing signals into dynamical components.

Let $X(t)$ be a real, continuous (for convenience) time series, given for all t, and which in any particular realization exists on the finite interval $[-T, T]$, where T is large. Assume that the series is normalized to have mean zero and variance one, and the subsequent discussion assumes that this normalization has been carried out. Given such a realization, X is extended to be zero outside $[-T, T]$ (thus X depends implicitly upon T), so that the Fourier transform $\hat{X}(f)$ is defined as

$$\hat{X}(f) = \int_{-T}^{T} X(t) e^{2\pi i f t} \mathrm{d}t. \tag{1.1}$$

Note that $\overline{\hat{X}(f)} = \hat{X}(-f)$, if $X(t)$ is real. The (one-sided) *power spectral density* $P_X(f)$ is then given by Press *et al.* (1989):

$$P_X(f) = \lim_{T\to\infty} \frac{1}{T} \left|\hat{X}(f)\right|^2, \tag{1.2}$$

supposing this exists.

The autocorrelation function is defined by

$$C_X(t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(s) X(s-t)\, ds, \tag{1.3}$$

and is a smooth, even function of t. The Wiener–Khintchin theorem can then be written in the form

$$\hat{C}_X(f) = \frac{1}{2} P_X(f). \tag{1.4}$$

For each t, a *centred window* is defined as the interval $(t-\tau, t+\tau)$ and $2\tau = \tau_w$ is the window length. Each value $X(t)$ of the time series is associated with a local window $X(t-s)$, where $s \in (-\tau, \tau)$. When X is sampled discretely, $X(t-s)$ is a matrix, called the trajectory matrix, and in the continuous case is called the *trajectory kernel.*

The aim of a moving spectral window is to choose an orthogonal basis of functions in a suitable function space on $[-\tau, \tau]$, and expand $X(t-s)$ in terms of this. An appropriate

space is $L_2[-\tau,\tau]$ of square-integrable functions with inner product

$$(f,g) = \frac{1}{2\tau}\int_{-\tau}^{\tau} f\bar{g}\ \mathrm{d}t. \tag{1.5}$$

If $\{\rho_k(s)\}$ is a complete, countable basis for $L_2[-\tau,\tau]$, then for each t in $\mathbf{R}$,

$$X(t-s) = \sum_k X_k(t)\rho_k(s), \tag{1.6}$$

and if $\{\rho_k\}$ is an orthonormal basis, then the components are given by

$$X_k(t) = \frac{1}{2\tau}\int_{-\tau}^{\tau} X(t-s)\bar{\rho}_k(s)ds, \tag{1.7}$$

where $\bar{\rho}_k$ denotes the complex conjugate of ρ_k.

A *filter* is then a projection on to one or more values of k. *Centre* projection occurs if $s = 0$, thus

$$X_F(t) = \sum_{\{k\}} \rho_k(0)X_k(t), \tag{1.8}$$

where $\{k\}$ denotes a subset of the integers k which label the basis. *End* projection occurs if $s = \pm\tau$, and for example back projection is taken to be $s = -\tau$,

$$X_F(t+\tau) = \sum_{\{k\}} \rho_k(-\tau)X_k(t). \tag{1.9}$$

The difference between centre and end projection lies in the way the (discrete) time series is embedded in the phase space. Associated with the sequence $\{X_n\}_{n=1}^N$, a d_E-dimensional embedding can be chosen variously as

$$\begin{aligned}
\mathbf{x}_n &= (X_n, X_{n-1}, X_{n-2}, \ldots X_{n-(d_E-1)}),\\
\mathbf{x}_n &= (X_n, X_{n+1}, X_{n+2}, \ldots X_{n+(d_E-1)}),\\
\mathbf{x}_n &= (X_{n-[d_E/2]}, \ldots, X_{n+[(d_E-1)/2]}).
\end{aligned} \tag{1.10}$$

When SVD and projection on to the singular vectors is carried out, then the embeddings above lead respectively to backward, forward and centre projection. While there would be appear to be little to distinguish these, we shall see that (at least in the continuous limit) there are differences concerning the ability of the various projections to converge to the data set at the window end-points.

The aim of a filter is to select $\{\rho_k\}$ which can optimally capture certain parts of the signal. For example, consider a moving Fourier window, $\rho_k = e^{\pi iks/\tau}$, where $k \in \mathbf{Z}$. The filtered signal determined by centre projection is

$$X_F(t) = \sum_{\{k\}} X_k(t), \tag{1.11}$$

and the components X_k are just the Fourier components of X over the window. The difference between centre and end projection is immediately apparent. For each t, the Fourier representation of $X(t-s)$, $s \in [-\tau,\tau]$, will converge to $X(t-s)$ for all $s \in (-\tau,\tau)$ but not in general at the end points, where the Fourier series converges to $\frac{1}{2}[X(t-\tau)+X(t+\tau)]$.

The question now arises as to what basis to choose for ρ_k. SSA answers this as follows.

By Hilbert-Schmidt theory, the positive, symmetric integral operator

$$A\rho = \frac{1}{2\tau}\int_{-\tau}^{\tau} C_X(t-s)\rho(s)\mathrm{d}s. \tag{1.12}$$

has a denumerable sequence of eigenfunctions ρ_k with positive eigenvalues $\lambda_k = \sigma_k^2$, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant 0$. These eigenfunctions form a complete orthonormal basis for $L_2[-\tau,\tau]$ if A is non-degenerate and their form is discussed in the following section.

It is perhaps easiest to understand in what sense the eigenfunctions of (1.12) might be 'better' than (say) a Fourier basis in terms of the practical implementation of the methodology above. Just as the discrete Fourier transform is the practical way to do Fourier analysis, so singular value decomposition (SVD) is the discrete method corresponding to SSA (Broomhead & King, 1986; Vautard & Ghil, 1989; Vautard *et al.*, 1992). The singular values correspond to σ_k, while the singular vectors correspond to ρ_k. However, in this implementation, one can interpret the window vectors as points in phase space, and the singular vectors then represent the principal axes of the trajectory explored in the phase space. The point is that the singular vectors sequentially select the directions with most power (in a least squares sense); thus projection on to them is most efficient insofar as representing the data with the fewest possible components is concerned. The same observation applies to the continuous system.

To put it another way, SSA is 'better' at representing data since, even where the EOFs are indeed sinusoidal (as they are when the window is large – see below), their period is not constrained to be an integer fraction of the window width, as would be the case for the base period of a Fourier basis. On the other hand, the fact that the EOFs are themselves orthogonal is a limitation, and indicates that other, 'better' bases ought to be available.

2 Eigenfunctions

In the preceding section, centre and end projection were shown to lead to different results if a Fourier basis is chosen. To examine whether this difference occurs in SSA, the behaviour of the eigenfunctions ρ_k for centred windows (§ 1), is examined for the limits where $\tau \to \infty$ and $\tau \to 0$.

The limit $\tau \to \infty$

Consider the eigenvalue problem $A\rho = \lambda\rho$, A given by (1.12), in the limit where $\tau \to \infty$. When $T \gg \tau \gg 1$, the local transform, defined as

$$\hat{X}(w,f) = \int_{w-\tau}^{w+\tau} X(t)e^{2\pi i f t}\mathrm{d}t, \tag{2.1}$$

is independent of w when τ is large. Essentially, this means that the power spectrum of X is stationary. If $\rho = e^{2\pi i f t}$, then

$$\begin{aligned} A\rho &\approx \frac{1}{2T}\cdot\frac{1}{2\tau}\int_{-T}^{T}\int_{-\tau}^{\tau} X(u)X(u-t+s)e^{2\pi i f s}\mathrm{d}u\mathrm{d}s \\ &= \left[\frac{1}{2T}\int_{-T}^{T} X(u)e^{-2\pi i f u}\left\{\frac{1}{2\tau}\int_{u-t-\tau}^{u-t+\tau} X(v)e^{2\pi i f v}\mathrm{d}v\right\}\mathrm{d}u\right]\rho. \end{aligned} \tag{2.2}$$

By assumption, the expression in curly brackets, which is $\hat{X}(u-t,f)$, is independent of $u-t$, and in this case (2.2) may be approximately written as

$$A\rho = \lambda\rho, \tag{2.3}$$

where

$$\lambda = \frac{1}{2\tau}\cdot\frac{1}{2T}\int_{-T}^{T} X(u)e^{-2\pi ifu}du\int_{-\tau}^{\tau} X(v)e^{2\pi ifv}\mathrm{d}v. \tag{2.4}$$

Using (1.1), (1.2) and (1.4), we deduce that for large values of τ, the pair

$$\rho = e^{2\pi ift},\quad \lambda = \frac{1}{2\tau}\hat{C}_X(f), \tag{2.5}$$

form an approximate eigensolution for (2.3). In this case, for centred windows, large window lengths are equivalent to Fourier analysis. The practical effect of projection to such a basis is examined in §3. For completeness, the behaviour of eigenfunctions for windows which are not centred is examined in the Appendix, but the lesson is the same; projection to either endpoint of a large window does not guarantee convergence to the underlying signal in the absence of noise.

The limit $\tau \to 0$

Gibson *et al.* (1992) analysed the discrete formulation of this eigenvalue problem (i.e. the singular value decomposition of $X(t-s)$), and showed that in the limit $\tau \to 0$, the singular vectors could be approximated by discrete Legendre polynomials. The continuous version is analysed here, and it is shown that the appropriate eigenfunctions, as expected, are approximated by Legendre polynomials. Our intention is not to suggest that such short windows should be used in practice, because we expect an optimal window length to be comparable to the 'recurrence time' (Kember & Fowler, 1993), but rather to gain some analytic insight into the sort of structure the eigenfunctions can be expected to have at moderate window lengths.

Recall that

$$A\rho = \frac{1}{2\tau}\int_{-\tau}^{\tau} C_X(t-s)\rho(s)\mathrm{d}s = \lambda\rho. \tag{2.6}$$

If $t = \tau\xi$, $\rho(t) = \phi(\xi)$, and C_X is assumed to be analytic, then

$$\begin{aligned}\lambda\phi &= \frac{1}{2}\int_{-1}^{1} C_X[\tau(\xi-u)]\phi(u)\mathrm{d}u\\ &= \frac{1}{2}\sum_j \frac{\tau^j}{j!}c_j\int_{-1}^{1}(\xi-u)^j\phi(u)du,\end{aligned} \tag{2.7}$$

where $c_j = C_X^{(j)}(0)$. The assumption of analyticity is necessary in writing this, and is crucial to the result that the eigenfunctions are approximately Legendre polynomials. For example, a red noise process with $C_X(t) = e^{-a|t|}$ has eigenfunctions which are sinusoidal in $(-\tau,\tau)$ (for all τ) (Vautard & Ghil, 1989).

The orthonormal polynomials over $L_2[-1,1]$ are the Legendre polynomials $P_n(\xi)$, and

we write the nth (starting from $n = 0$) eigenfunction as

$$\phi_n = \sum_0^\infty a_k^{(n)} P_k(\xi). \tag{2.8}$$

The superscript n is omitted where convenient. Substituting this in, and using the orthogonality property of the Legendre polynomials over $[-1, 1]$,

$$\lambda \sum_0^\infty a_k P_k(\xi) = \frac{1}{2} \sum_{j\geq 0} \frac{\tau^j}{j!} c_j \sum_{l\geq 0} a_l \int_{-1}^1 (\xi - u)^j P_l(u) \mathrm{d}u, \tag{2.9}$$

whence the coefficients a_k satisfy the equations

$$\frac{2\lambda}{2k+1} a_k = \frac{1}{2} \sum_{j\geq 0} \frac{\tau^j}{j!} c_j \sum_{l\geq 0} a_l \int_{-1}^1 \int_{-1}^1 (\xi - u)^j P_l(u) P_k(\xi) \mathrm{d}u \, \mathrm{d}\xi. \tag{2.10}$$

Since C_X is an even function, $c_j = 0$ if j is odd, and integration by parts gives

$$c_{2s} = (-1)^s \kappa_s, \tag{2.11}$$

where

$$\kappa_n = \langle \{X^{(n)}\}^2 \rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^T \{X^{(n)}\}^2 \mathrm{d}t. \tag{2.12}$$

Note that if

$$b_{ij} = \int_{-1}^1 u^i P_j(u) \mathrm{d}u, \tag{2.13}$$

then $b_{ij} = 0$ if $i < j$. By expanding the double integral in (2.10), that equation can be written in the form

$$\frac{2\lambda}{2k+1} a_k = \sum_{s\geq k/2} \sum_{l=0}^{2s-k} \gamma_{ls}^{(k)} a_l \tau^{2s}, \tag{2.14}$$

where

$$\gamma_{ls}^{(k)} = \sum_{r=k}^{2s-l} \frac{(-1)^{s+r} \kappa_s b_{r,k} b_{2s-r,l}}{r!(2s-r)!}, \tag{2.15}$$

and the sum limits in (2.14) may be restricted, since $\gamma_{ls}^{(k)} = 0$ if $k + l > 2s$.

Next, it is easy to show that if $\phi(\xi)$ satisfies (2.7) for some λ, then so does $\phi(-\xi)$. Hence, the eigenfunctions can be considered as being either even or odd, and in particular can be written respectively as

$$\phi_n = \sum_{r\geq 0} A_r P_{2r}(\xi) \tag{2.16}$$

(n even, and $A_r = a_{2r}^{(n)}$), with

$$\frac{2\lambda}{4r+1} A_r = \sum_{s\geq r} \sum_{m=0}^{s-r} \gamma_{2m,s}^{(2r)} A_m \tau^{2s}, \tag{2.17}$$

and

$$\phi_n = \sum_{r\geq 0} B_r P_{2r+1}(\xi) \tag{2.18}$$

(n odd, and $B_r = a^{(n)}_{2r+1}$), with

$$\frac{2\lambda}{4r+3}B_r = \sum_{s\geqslant r}\sum_{m=0}^{s-r}\gamma^{(2r+1)}_{2m+1,s+1}B_m\tau^{2s+2}. \tag{2.19}$$

The solutions are power series in τ^2 as $\tau \to 0$. For example, at leading order, (2.17) implies

$$2\lambda A_0 = \gamma^{(0)}_{0,0}A_0 + O(\tau^2), \quad \lambda A_r = O(\tau^2) \text{ for } r > 0, \tag{2.20}$$

suggesting $\lambda_0 \approx \gamma^{(0)}_{0,0}/2$, $\phi_0 \approx P_0(\xi)$ for the 0th eigenfunction. Since the eigenfunctions are orthogonal, the next one must have A_0 small ($O(\tau^2)$) so that $\lambda_1 = O(\tau^2)$. Proceeding in this manner suggests $\lambda_n \propto \tau^{2n}$, $\phi_n \propto P_n(\xi)$, with A_n and B_n being $O(1)$, and the rest smaller, in fact $A_{n\pm j}$ and $B_{n\pm j}$ are $O(\tau^{2j})$. This can be formalized by redefining λ, A_m and B_m in (2.17) and (2.19) as follows. For the *even* eigenfunctions, put $n = 2N$, and

$$\lambda = \mu\tau^{4N}, \quad A_m = \tau^{2|N-m|}C_m; \tag{2.21}$$

for the odd eigenfunctions, put $n = 2N+1$, and

$$\lambda = \mu\tau^{4N+2}, \quad B_m = \tau^{2|N-m|}D_m. \tag{2.22}$$

It is reasonable to anticipate that eigenfunctions and eigenvalues can be found in which, $\mu, A_m, B_m = O(1)$ as $\tau \to 0$. Rearranging the summations, the two expressions (2.17) and (2.19) can be simplified as

$$\begin{aligned}\frac{2\mu}{4r+1}\tau^{2[N-r+|N-r|]}C_r = &\sum_{j=0}^{\infty}\sum_{m=0}^{N}\gamma^{(2r)}_{2m,m+j+r}C_m\tau^{2j} \\ &+\sum_{q=0}^{\infty}\sum_{p=0}^{[q/2]}\gamma^{(2r)}_{2(N+1+p),q+N+1-p+r}C_{N+1+p}\tau^{4+2q},\end{aligned} \tag{2.23}$$

and

$$\begin{aligned}\frac{2\mu}{4r+3}\tau^{2[N-r+|N-r|]}D_r = &\sum_{j=0}^{\infty}\sum_{m=0}^{N}\gamma^{(2r+1)}_{2m+1,m+j+r+1}D_m\tau^{2j} \\ &+\sum_{q=0}^{\infty}\sum_{p=0}^{[q/2]}\gamma^{(2r+1)}_{2(N+1+p)+1,q+2+N-p+r}D_{N+1+p}\tau^{4+2q}.\end{aligned} \tag{2.24}$$

The eigenfunctions and eigenvalues are then determined at leading order as follows, considering for example the $2N$th eigenfunction. C_N can be taken as given with no loss of generality (it simply normalizes ϕ_{2N}). Then for $r < N$, $\{N - r + |N - r|\} > 0$, so to leading order

$$\sum_{m=0}^{N}\gamma^{(2r)}_{2m,m+r}C_m \approx 0, \quad r = 0, 1, \ldots N-1, \tag{2.25}$$

giving N equations for the N coefficients $C_0, \ldots C_{N-1}$. When $r = N$,

$$\frac{2\mu}{4N+1}C_N \approx \sum_{m=0}^{N}\gamma^{(2N)}_{2m,m+N}C_m \tag{2.26}$$

then determines μ, while for $r > N$

$$\frac{2\mu}{4r+1}C_r \approx \sum_{m=0}^{N} \gamma_{2m,m+r}^{(2r)} C_m \tag{2.27}$$

gives the leading order expressions for C_r. Evidently the procedure can be carried to higher powers of τ^2, and the process is identical for the odd eigenfunctions.

Legendre functions as $n \to \infty$

Thus the eigenfunctions are sinusoidal for large τ, but more like Legendre polynomials when $\tau \to 0$. These two behaviours can be reconciled by examining the behaviour of the Legendre polynomials when n is large. The Mehler–Dirichlet integrals for P_n are

$$P_n(\cos\phi) = \frac{2}{\pi}\int_0^{\phi} \frac{\cos[(n+\frac{1}{2})\theta]\mathrm{d}\theta}{[2\cos\theta - 2\cos\phi]^{1/2}}. \tag{2.28}$$

With $x = \cos\phi$, $a = e^{i\phi}$, $z = e^{i\theta}$,

$$P_n(x) = Re\left\{\frac{2}{\pi i}\int_0^{a} \frac{z^n \mathrm{d}z}{[z^2 - 2xz + 1]^{1/2}}\right\}. \tag{2.29}$$

The roots of the quadratic are at $\theta = \pm\phi$, and a branch cut is taken along the arc of the circle joining $\bar{a}$ to a. Applying the method of steepest descents as $n \to \infty$ leads to

$$P_n(\cos\phi) \sim \sqrt{\frac{2}{\pi n}} \frac{\sin\left\{\left(n+\frac{1}{2}\right)\phi + \frac{\pi}{4}\right\}}{\sin^{1/2}\phi}, \tag{2.30}$$

which is locally sinusoidal for large n.

This expression is invalid near the end points, e.g. $\phi \to 0$, and specifically when $\phi \sim 1/n$. Putting $n\phi = O(1)$ as $n \to \infty$, we have

$$P_n(\cos\phi) \sim J_0(n\phi) \text{ as } n \to \infty \ (n\phi \sim O(1)), \tag{2.31}$$

(see Abramowitz & Stegun, 1964, p. 787). Hence

$$P_n(x) \sim J_0\left[n(2(1-x))^{1/2}\right] \tag{2.32}$$

as $x \to 1$, with even or odd extension as $x \to -1$ if n is even or odd. As $1 - x$ becomes $O(1)$, the Bessel function asymptotes back to (2.30).

3 Filters

Now consider the question of local filtering. Recall that if $\{\rho_k(s)\}$ is a complete orthonormal sequence of functions in $L_2[-\tau, \tau]$, then projection of $X(t-s)$, $s \in [-\tau, \tau]$ on to a set of k-values is, from (1.6),

$$X(t-s) = \sum_{\{k\}} X_k(t)\rho_k(s), \tag{3.1}$$

and the resulting filter is, for centre projection,

$$X_F(t) = \sum_{\{k\}} \rho_k(0) X_k(t). \tag{3.2}$$

The spectral transfer function of such a filter is of interest and is defined by

$$\Phi(f) = P_{X_F}(f)/P_X(f), \tag{3.3}$$

i.e. the ratio of the filtered power spectrum to that of the original signal, and is also given in terms of the basis functions as follows. The Fourier transform of X_k is

$$\hat{X}_k(f) = \tilde{\rho}_k(f)\hat{X}(f), \tag{3.4}$$

where

$$\tilde{\rho}_k(f) = \frac{1}{2\tau}\int_{-\tau}^{\tau} \rho_k(s)e^{2\pi i f s}ds \tag{3.5}$$

is the reduced transform of ρ_k. Thus the Fourier transform of X_F (found by centre projection)

$$\hat{X}_F(f) = \sum_{\{k\}} \rho_k(0)\hat{X}_k(f), \tag{3.6}$$

can be written as

$$\hat{X}_F(f) = \sum_{\{k\}} \rho_k(0)\tilde{\rho}_k(f)\hat{X}(f), \tag{3.7}$$

so that

$$P_{X_F}(f) = |\sum_{\{k\}} \rho_k(0)\tilde{\rho}_k(f)|^2 P_X(f) \tag{3.8}$$

and

$$\Phi(f) = \frac{P_{X_F}(f)}{P_X(f)} = |\sum_{\{k\}} \rho_k(0)\tilde{\rho}_k(f)|^2. \tag{3.9}$$

The filtering effect of SSA is determined by the transfer function $\Phi(f)$. $\Phi(f)$ is determined by the eigenfunctions ρ_k of the integral operator (1.12), which depends upon the correlation structure of the signal, $C_X(t)$, and the window length $\tau_w = 2\tau$. Thus the effect of such a filter is independent of the phase of the signal and has the same spectral effect on a deterministic signal and a surrogate signal which has had its phase randomised (Provenzale *et al.*, 1992): *this filter has an effect that is independent of whether the signal is random or not.* The apparent inference is that SSA is unable to selectively filter oscillatory dynamical signals from an irregular time series, as has been claimed (Ghil & Vautard 1991). The transfer function is now constructed for the limits of large and small centred windows.

The limit $\tau \to \infty$

In this limit, the eigenfunctions are $e^{2\pi i f t}$, corresponding to eigenvalues $\lambda = \hat{C}_X(f)/2\tau$ (see (2.5)); a pair $f = \pm f^*$ with corresponding eigenfunctions $\rho_\pm = e^{\pm 2\pi i f^* t}$ is chosen, and the corresponding reduced transforms are

$$\tilde{\rho}_\pm = \frac{\sin[2\pi(f \mp f^*)\tau]}{2\pi(f \mp f^*)\tau}. \tag{3.10}$$

For large τ, $\tilde{\rho}_+ \approx (2\tau)^{-1}\delta(f - f^*)$, where δ is the delta function; thus the eigenfunctions approximately select the frequency f^*, much as an ordinary Fourier transform does. Evaluating $\Phi(f)$, the spectral transfer function corresponding to centre projection on to

$f = \pm f^*$ gives

$$\Phi(f) \approx \left(\frac{\sin[2\pi\tau(f-f^*)]}{2\pi\tau(f-f^*)} + \frac{\sin[2\pi\tau(f+f^*)]}{2\pi\tau(f+f^*)} \right)^2 \tag{3.11}$$

whereas for back projection (as in (1.9)),

$$\Phi(f) \approx \cos^2(2\pi f^*\tau) \left(\frac{\sin[2\pi\tau(f-f^*)]}{2\pi\tau(f-f^*)} + \frac{\sin[2\pi\tau(f+f^*)]}{2\pi\tau(f+f^*)} \right)^2 + \sin^2(2\pi f^*\tau) \left(\frac{\sin[2\pi\tau(f-f^*)]}{2\pi\tau(f-f^*)} - \frac{\sin[2\pi\tau(f+f^*)]}{2\pi\tau(f+f^*)} \right)^2. \tag{3.12}$$

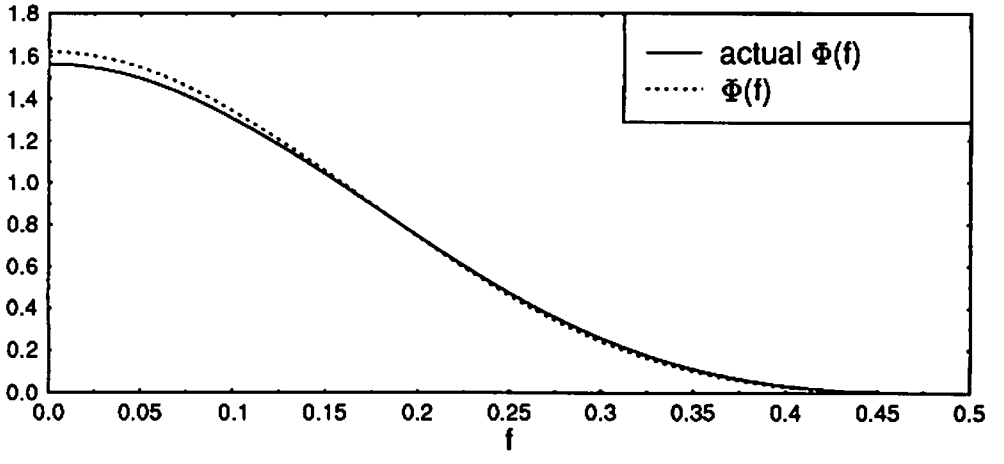
Thus, for centre projection (3.11), and back projection (3.12), the function $\Phi(f)$ selects a weighted band of frequencies near $f = f^*$, and when $\tau \to \infty$ both filters approach, near $f = f^*$,

$$\Phi(f) \approx \frac{\sin^2[2\pi\tau(f-f^*)]}{[2\pi\tau(f-f^*)]^2}. \tag{3.13}$$

If the eigenfunctions approximate a Fourier basis when $\tau = O(1)$, which is not unusual, then the corresponding transfer functions are different; for example when $\tau = 1/4f^*$, centre projection (3.11) has $\Phi(0) > 1$ and spectral power is fed to the filtered signal, whereas back projection (3.12) has $\Phi(0) = 0$.

As an example, consider the Rössler equations: $\dot{x} = -z - x$, $\dot{y} = x + ay$, $\dot{z} = b + z(x-c)$, $a = 0.15$, $b = 0.2$, $c = 10$, integrated numerically with a timestep $\Delta t = \pi/100$ (with a fourth order Runge Kutta solver), and outputted at a time step of $\Delta t = \pi/10$. The Rössler equations possess a recurrence time $T \approx 2\pi$, and a centred embedding (the discrete version of a centred window) is constructed from $y(t)$ with $\tau_w = 2\tau = T/2 = \pi$, where $d_E = 11$. In Figure 1, a comparison is made of $\Phi(f)$, computed from (3.11) with $f^* = 1/T = 1/2\pi$, $\tau = \pi/2$, and the same function found directly from the singular system analysis of the embedding of $y(t)$ by projecting onto the eigen-pair ($d_P = 2$) associated with the largest singular values of the trajectory matrix. This filter selects a band of frequencies near $f^* = 1/2\pi$. Good agreement is observed between (3.11) and $\Phi(f)$ computed for the embedding of $y(t)$, since $y(t)$ has a strong periodic component of period $T \approx 2\pi$ and in fact the first pair of eigenfunctions are approximately $\rho_0 = \sqrt{2}\cos t$, and $\rho_1 = \sqrt{2}\sin t$.

The Rössler equations are an example of a system for which the embedded trajectory is approximately spanned by a few Fourier modes, despite the fact that the half-window width τ is $O(1)$. As shown in Figure 2, this is due to the strong periodicity of the signal, as exhibited in the power spectrum $P_y(f)$ of $y(t)$. In this figure, the result of projection on to the first two singular vectors ($d_P = 2$) for both centre projection (as in Figure 1) and end projection is also shown. The transfer functions are quite different. In particular, the results in Figure 2 suggest that if Gaussian noise is added to $y(t)$, a *better* filter might be back projection, even though back projection onto the entire basis (§2) may not recover the original signal in the absence of noise. The results may be worse for a red-noise process with an algebraically decaying power spectrum added to $y(t)$, since centre projection to an eigen-pair is likely to enhance power spuriously at lower frequencies.



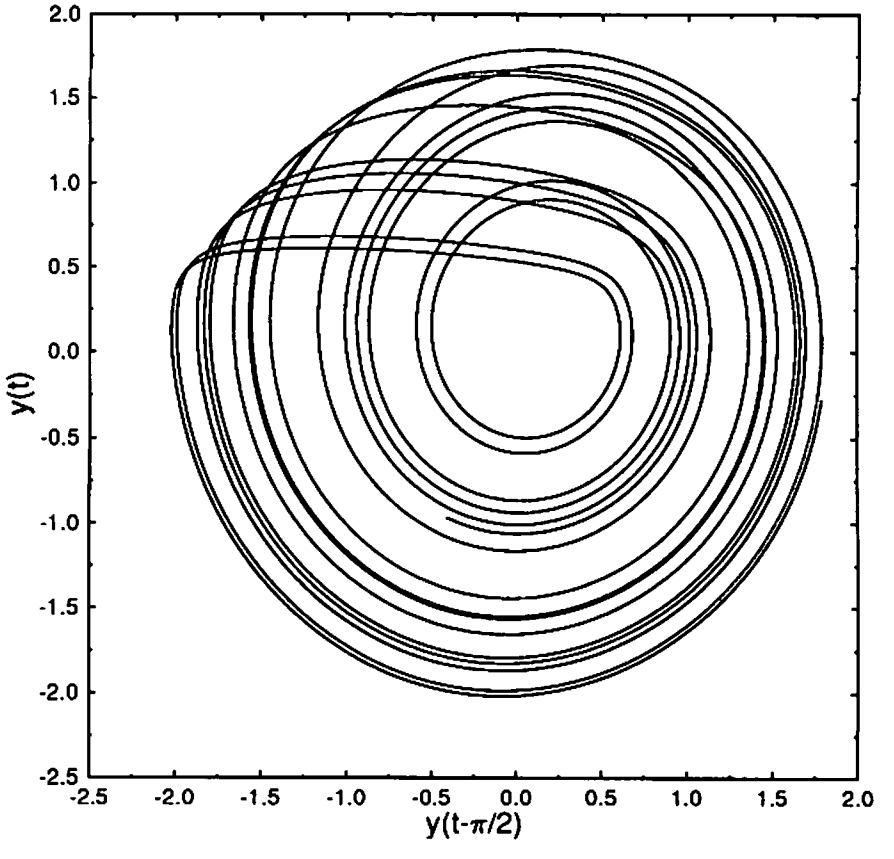


FIGURE 1. Phase portrait of numerical solution of Rössler equations (see text), together with analytic (dotted) and numerically (or 'actual' in the legend) computed filter $\Phi(f)$. The extent to which the $\Phi(f)$ are different is due to the lack of resolution of the eigenfunctions provided with a $d_E = 11$ dimensional embedding.

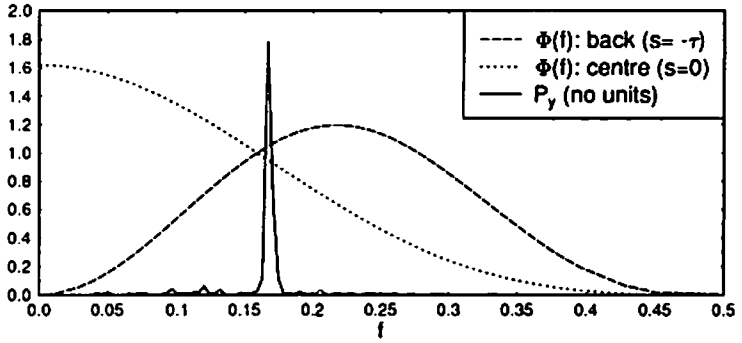


FIGURE 2. Power spectrum of numerical solution of the Rössler equations shown with spectral transfer functions obtained by centre projection ($s = 0$ in (1.6)) and back projection ($s = -\tau$) in equation (1.6) when the eigenfunctions ρ_k approximate a Fourier basis. The power spectrum (no units are shown) is computed with a square window using 4096 points taken from the numerical solution (see text) of the Rössler equations and the Nyquist frequency is approximately $f = 16$. The spectral transfer function is strongly dependent upon the choice of projection location s.

The limit $\tau \to 0$

The limit $\tau \to 0$ is also of interest. Recall from (2.8) that the nth ($n \geqslant 0$) eigenfunction is given by

$$\rho_n(\tau\xi) = \phi_n(\xi) = \sum_{k=0}^{\infty} a_k^{(n)} P_k(\xi), \tag{3.14}$$

where $\{a_k^{(n)}\}$ and the corresponding eigenvalues λ_n are given by (2.17) and (2.19), or (2.21)–(2.24). Thus the reduced transforms are

$$\begin{aligned}\tilde{\rho}_n(f) &= \frac{1}{2\tau}\int_{-\tau}^{\tau} \tilde{\rho}_n(s) e^{2\pi i f s} \mathrm{d}s \\ &= \frac{1}{2}\int_{-1}^{1} \phi_n(\xi) e^{2\pi i f \tau \xi} \mathrm{d}\xi.\end{aligned} \tag{3.15}$$

Consider the case of centre projection. In this case, since $\rho_k(0) = \phi_k(0) = 0$ for odd k, projection eliminates odd components. For even n ($= 2N$)

$$\tilde{\rho}_{2N}(f) = \frac{1}{2}\int_{-1}^{1} \phi_{2N}(\xi)\cos(2\pi f \tau \xi)\mathrm{d}\xi. \tag{3.16}$$

Now, if projection is on to the $2N$-th eigenfunction (or the pair $2N$ and $2N + 1$), then (from (3.9)) the corresponding transfer function

$$\Phi_{2N}(f) = |\rho_{2N}(0)\tilde{\rho}_{2N}(f)|^2. \tag{3.17}$$

To compute the spectral transfer function, $\rho_{2N}(0)$ and $\tilde{\rho}_{2N}(f)$ are required. From § 2, (2.16), (2.21) and (3.14),

$$\rho_{2N}(0) = C_N P_{2N}(0) + O(\tau^2) \tag{3.18}$$

where $P_{2k}(0) = (-1)^k(2k)!/\{2^{2k}(k!)^2\}$, thus $P_0(0) = 1$, $P_2(0) = -1/2$, $P_4(0) = 3/8$, etc. C_N is chosen to normalise ϕ_{2N} to a unit norm. It follows from (2.16), (2.21), and the fact that

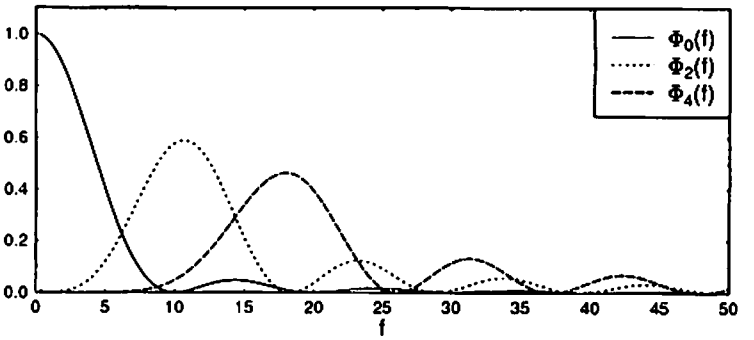


FIGURE 3. Small window approximation to $\Phi_{2N}(f)$ $N = 0, 1, 2$, from (3.17), (3.20) and (3.21), where $\tau = 0.05$.

$\frac{1}{2}\int_{-1}^{1} P_n^2(\xi)d\xi = 1/(2n+1)$, that

$$C_N = (4N+1)^{1/2} + O(\tau^4), \tag{3.19}$$

hence,

$$\rho_{2N}(0) = (4N+1)^{1/2}\frac{(-1)^n(2n)!}{2^{2n}(n!)^2} + O(\tau^2). \tag{3.20}$$

To find $\tilde{p}_{2N}(f)$, ϕ_{2N} is expanded to leading order in terms of the Legendre polynomials. It is clear from (3.15) that $\tilde{p}_{2N}$ is in fact a function of $f\tau$; moreover in practice one *chooses* τ inversely proportional to the recurrence frequency of interest (Kember & Fowler, 1993). Therefore $f\tau$ is assumed to be of $O(1)$. Using elementary properties of the Legendre polynomials and some algebra, eventually we find

$$\tilde{p}_{2N}(f) \approx C_N \left[\left(\sum_{m=0}^{N} \frac{(-1)^m 1.3.5\ldots(4m-1)\binom{2N+2m}{2N-2m}}{(2\pi f\tau)^{2m}} \right) \frac{\sin 2\pi f\tau}{2\pi f\tau} + \left(\sum_{\substack{m=0 \\ (N\geqslant 1)}}^{N-1} \frac{(-1)^m 1.3.5\ldots(4m+1)\binom{2N+2m+1}{2N-2m-1}}{(2\pi f\tau)^{2m+1}} \right) \frac{\cos 2\pi f\tau}{2\pi f\tau} \right] \tag{3.21}$$

and $\Phi_{2N}(0) = 0$ for $N \geqslant 1$. The transfer function to leading order, for projection on to ϕ_0 only, is

$$\Phi_0(f) = \frac{\sin^2(2\pi f\tau)}{(2\pi f\tau)^2}, \tag{3.22}$$

equivalent to the large window result (3.11) when $f^* \ll 1$, and, for projection on to ϕ_2 only,

$$\Phi_2(f) = \frac{25}{4}\left\{ \frac{\sin(2\pi f\tau)}{2\pi f\tau} - 3\frac{\sin(2\pi f\tau)}{(2\pi f\tau)^3} + 3\frac{\cos(2\pi f\tau)}{(2\pi f\tau)^2} \right\}^2. \tag{3.23}$$

As N increases, Φ_{2N} selects an approximately constant bandwidth (about $1/2\tau$ for $N = 0$ and $1/\tau$ for $N \geqslant 1$), centred at larger frequency but with a smaller peak amplitude. Φ_{2N}, $N = 0, 1, 2$ is depicted in figure 3 for $\tau = 0.05$. Approximating Φ_{2N} at large N, away from

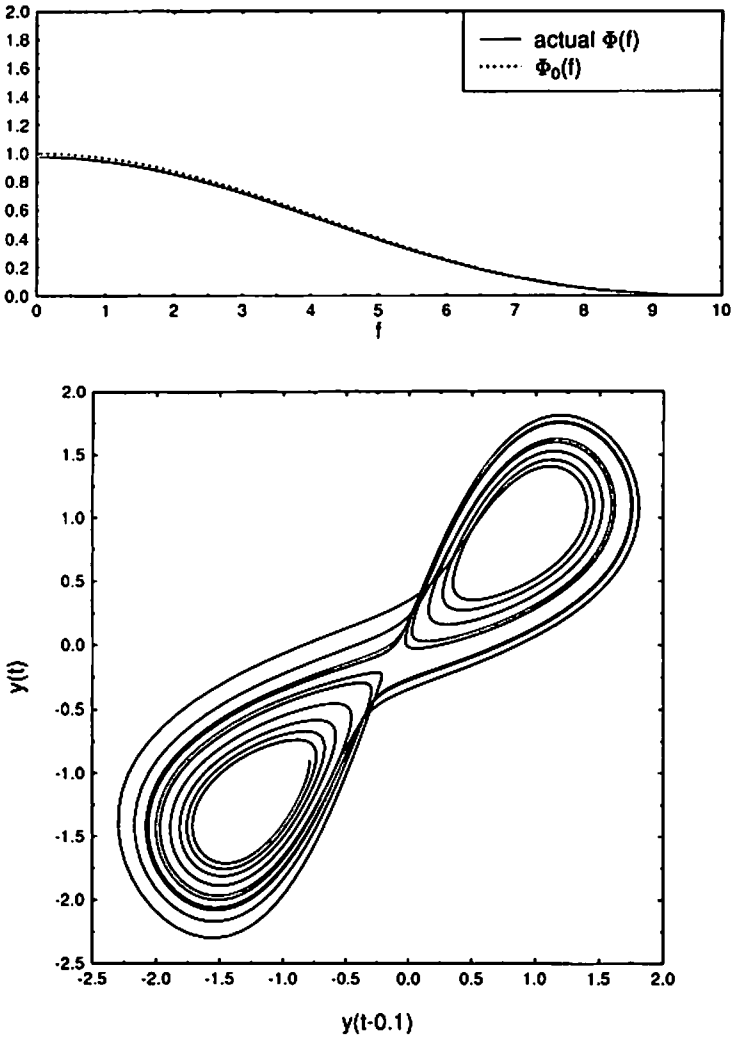


FIGURE 4. Phase portrait of the Lorenz equations (see text), together with analytic (dotted) and numerically (or 'actual' in the legend) computed filter $\Phi_0(f)$. The extent to which the transfer functions are different is due to the lack of resolution of the eigenfunctions provided with a $d_E = 11$ dimensional embedding.

$f = 0$ (use (2.32), and the method of stationary phase (see, for example, Bender & Orszag, 1978, Chapter 6)) leads to

$$\Phi_{2N}(f) = \frac{M}{N^2}\cos^2(2\pi f\tau), \tag{3.24}$$

for $f\tau = O(1)$, $M = O(1)$. It can readily be shown that the transfer function $\cos^2(2\pi f\tau)$ corresponds to a moving average filter of the form $X_F(t) = \frac{1}{2}(X(t-\tau)+X(t+\tau))$, thus as $N \to \infty$, projection on to each pair of eigenfunctions indexed by $2N$, $2N-1$ corresponds to a two-point moving average weighted by $1/2N$. To illustrate some aspects of the small window limit, SSA is applied to the Lorenz system; $\dot{x} = \sigma(y-x)$, $\dot{y} = rx - y - xz$, $\dot{z} = -bz + xy$, $\sigma = 10$, $b = 8/3$, $r = 28$, and we work with $y(t)$. These equations are

integrated numerically (with a fourth order Runge Kutta method), using a time step $\Delta t = 0.01$ and with the solution outputted at the same time step; $y(t)$ is centre-embedded over a window $\tau_w = 2\tau = 0.1$, with $d_E = 11$. In figure 4, the low frequency cut-off filter $\Phi_0(f)$ (3.22), is compared to the same computed directly from the SSA of $y(t)$, with projection to the eigenfunctions ρ_k, associated with the two largest singular values. As before with the Rössler equations, there is close agreement between Eq. (3.22) and $\Phi(f)$ computed for the embedding of $y(t)$, since the eigenfunctions approximate the Legendre polynomials.

Iterative filtering

The transfer function $\Phi(f)$ depends on $C_X(\tau)$, and so it may be of interest to iterate the filtering process by filtering the filtered series, etc.. Indeed, it has been found (Fowler *et al.*, 1994) that such iterations can apparently lead to convergent sequences, and it is part of the aim of this paper to examine what such sequences are convergent to. One hope is that they may improve noise reduction. To examine this, the approximate forms for ρ_k derived in §2 are used to compute approximations for $\Phi(f)$ in the limits of large and small window width.

The power spectrum of the filtered signal is determined by the map

$$P_{X_F}(f) = \Phi[f; P_X(f)]P_X(f), \tag{3.25}$$

where it is understood that at each iterate, the filtered series X_F is renormalised to have zero mean and unit variance. An iterated sequence of power spectra can be defined where $P_X^{(n)}(f)$ represents the nth form in the sequence. Ideally, the sequence would converge to the spectrum of the true signal, but in fact continued iteration of the filter reduces the signal substantially. This can be illustrated in the small window limit, where projection on to the first eigenfunction is given by $\Phi_0(f)$ in (3.22) so that each iterated transfer function is approximately $\Phi_0(f)$, and thus

$$P_X^{(n)}(f) \approx K_n \frac{\sin^{2n}(2\pi f\tau)}{(2\pi f\tau)^{2n}} P_X^{(0)}(f), \tag{3.26}$$

where K_n is constant. Continued iteration of the filter picks out the 0-th frequency, and leads (in this approximation) to the substantial reduction of the signal. More generally, the maxima of Φ_{2k} will lead to the filter picking out a nonzero frequency. Nevertheless, a few iterations can produce successive improvements in the filtered series. An example of this is for the filtering of Gaussian noise added to a single dynamical system. For example, the Lorenz equations possess a decreasing and continuous power spectrum, and essentially require the specification of a low pass filter.

To illustrate this, consider $y(t)$ taken from a numerical solution of the Lorenz equations (described above). Gaussian noise of zero mean and variance 16 is added to each outputted data point, and this corresponds to a signal to noise ratio (the ratio of standard deviation of signal to that of the noise) of about two. The noisy $y(t)$ is centre-embedded, over an optimal window (the methodology of optimization is explained by Kember & Fowler (1993)), with $\tau_w = 2\tau = 0.36$, and $d_E = 37$. Figure 5 shows the effect of iterating the filter using the same embedding definition and projecting on to the $d_P = 3$ singular

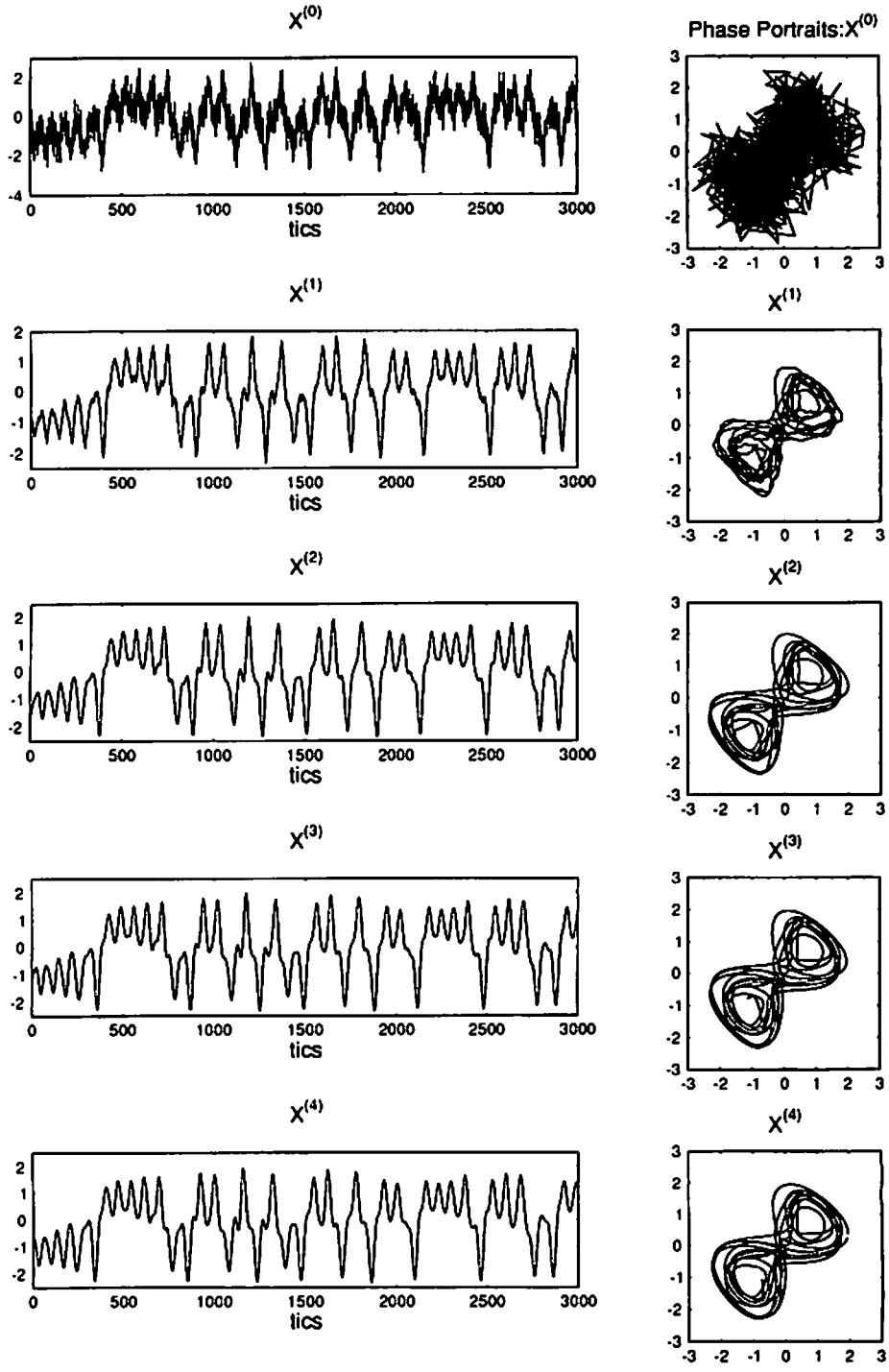


FIGURE 5. Lorenz time series (see text) plus added white noise of variance 16 (about a quarter of the variance of the clean signal), normalised to have variance one and mean zero; labelled $X^{(0)}$. The diagram illustrates the original noisy series $X^{(0)}$, and the results of successive iterates $X^{(j)}$, $j = 1, 2, 3, 4$, along with their phase portraits (shown beside each series), constructed with a lag of 0.18.

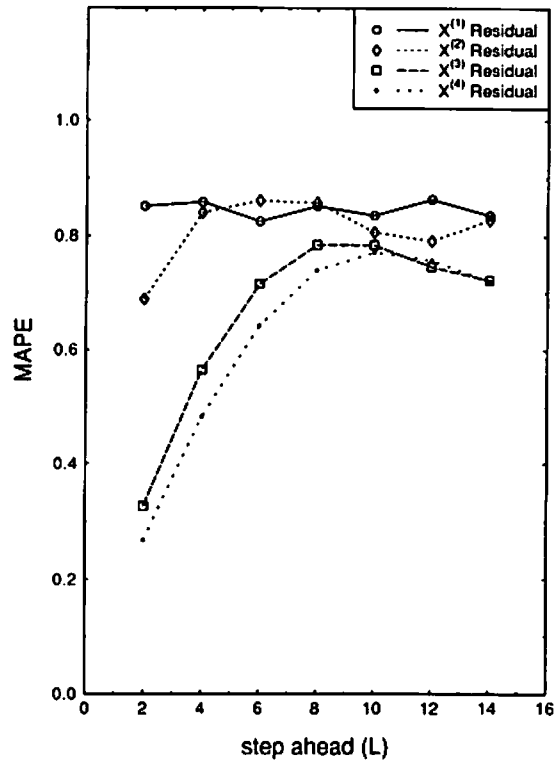


FIGURE 6. Prediction errors of the residual series of the first four passes of the iterative procedure of Figure 5, using a nearest neighbour linear predictor (see text), as a function of prediction time step ahead. The evolution of slope at pass three indicates predictability, and hence the loss of signal in the filter.

directions associated with singular values above the noise floor of the singular spectrum (the methodology is described in Broomhead & King (1986)). It is clear that the iterative procedure provides an apparently convergent sequence of filtered series, which seem to improve and this is particularly apparent in the phase portraits. As mentioned above, the repeated application eventually starts stripping more signal than noise. To evaluate this more objectively, Figure 6 shows the prediction error of the residual series at each iterate. The prediction error is evaluated as a function of the prediction step ahead within the embedded phase space, by using a nearest neighbour method (Sugihara & May, 1990). Specifically, an L step-ahead predictor of a predictee point in the embedded phase space is formed by first locating $d_E + 1$ nearest (in the Euclidean sense) predictee neighbours taken from a training set. The predictor is then set equal to the simple average of the predictee neighbours taken L steps into the future. The mean, absolute, prediction error (MAPE) at L steps ahead, is the mean, absolute difference between the predictors and the observed predictees L steps ahead (about 1000 predictions were made for each value of L). The result in Figure 6 indicates that the second iterate removes noise but that the third iterate removes signal, as the residual on the third iterate has some predictability. This suggests halting the procedure after two iterates.

A practical difficulty with the filtering procedure followed above is the specification

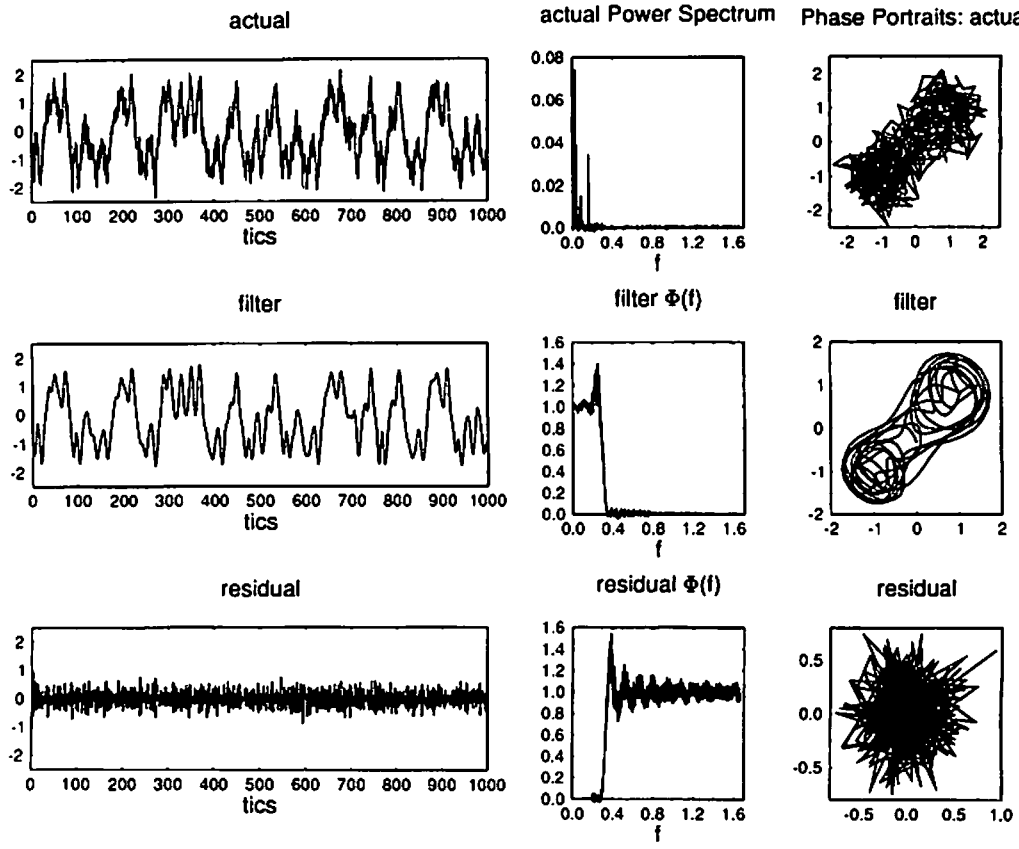


FIGURE 7. Time series of the solution of the forced Duffing equation, its power spectrum, and embedded phase portrait ($x(t)$ is plotted vertically against $x(t-1.5)$) (top row). Below are the plots of the filtered and residual time series, their spectral transfer functions and their phase portraits, also using a lag of 0.15. (See text for details.)

of an optimal window (or equivalently of a recurrence time) and a robust means to find this. The literature contains various methods to establish optimal windows based on the analysis of singular values (Broomhead & King, 1986), computation of mutual information (Fraser & Swinney, 1986), and measures of 'spreadedness' of the attractor in the embedding phase space (Buzug & Pfister, 1992; Liebert & Schuster, 1991; Kember & Fowler, 1993). Two different possible strategies for selecting the window are illustrated, and to test the algorithm further, the forced Duffing equation $\ddot{x} + a\dot{x} + x^3 - x = b\sin t$, $a = b = 0.4$, which has a more complicated power spectrum than the Lorenz system, is analysed. This equation is numerically integrated with a time step $\Delta t = 0.3$ and outputted at the same time step. Gaussian noise with zero mean and variance $\sigma^2 = 0.0625$ is added to $x(t)$ which yields a noisy time series with a signal to noise ratio of about four to one (top series in Figure 7). The Duffing equation possesses a recurrence time $T \approx 30$. $x(t)$ is centre embedded over an optimal window $\tau_w = 2\tau = T/2 \approx 15$, with $d_E = 50$. Figure 7 shows the effect of projecting on to the first ten singular directions ($d_P = 10$) that correspond to the ten largest singular values, all of which are above the noise floor of the singular spectrum. The spectral transfer functions of the filtered and residual series in

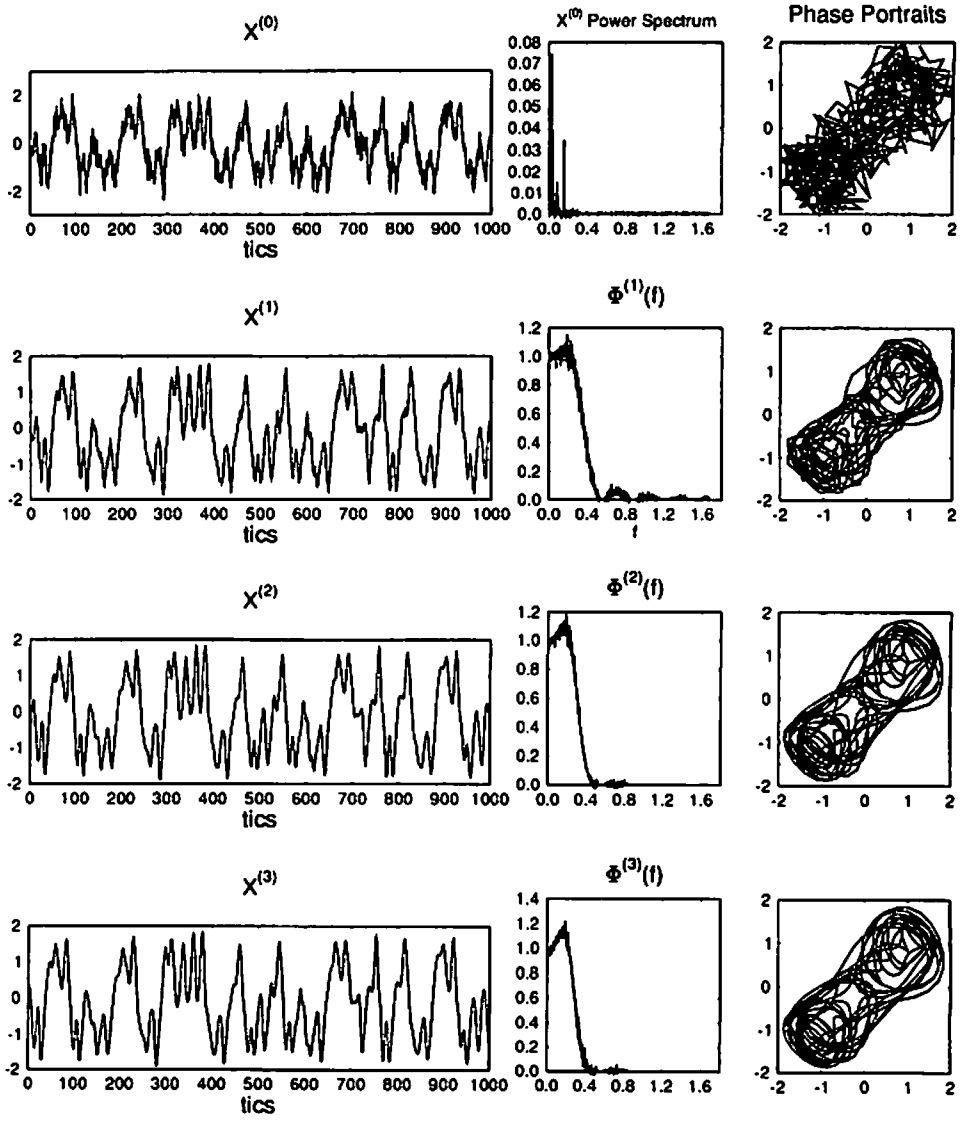


FIGURE 8. The result of iterating the filter, as described in the text, for the forced Duffing equation with a sub-optimal window. The top row shows the original series, spectrum and phase plot, and the ones below show successively the time series, accumulated spectral transfer function and phase plots of the first, second and third iterates. The accumulated spectral transfer functions are found by forming $P_X^{(n)}(f)/P_X^{(0)}(f)$ ($P_X^{(j)}(f)$ is the power spectrum of $X^{(j)}$).

Figure 7 (found by dividing the filtered and residual power spectra, respectively, by the original power spectrum) indicate a low pass filter with a cut-off at about $f = 0.35$. In this case, further iteration (not shown) does not lead to any improvement since the choice of low pass filter is optimal; and this is precisely because the window was selected to cater for the basic recurrence of the signal.

A more robust and useful method is to build the appropriate choice of window length into the filtering approach. The basic idea is to iterate a filter taken from a sub-optimal

(shorter) window where the effect of each linear filter is evaluated using predictability (as described in the Lorenz example above) of the dynamics of the residual series after filtering. Specifically, the procedure, in two parts, is as follows. Step one: a shorter window low pass filter $\Phi(f)$ is found. To do this, the noisy Duffing series considered above is centre embedded over a window $\tau_w = 3$, with $d_E = 11$. Inspection of the first few singular vectors indicates that they approximate the Legendre polynomials (naturally, since the window $\tau_w = 3$ is much smaller than half the recurrence time $T/2 = 15$) and hence this is a shorter window. To determine the filter to be iterated, the projection dimension d_P is reduced until the residual series begins to show improved predictability, and this occurs, in this example, at $d_P = 4$. Hence, a $d_P = 4$ projection is used to define a short window, low pass filter $\Phi(f)$. Step two: having derived $\Phi(f)$, iterate this filter (as described in the Lorenz example) until the residual series at each iterate begins to show some predictability (not shown), and this occurs at the third iterate. The accumulated transfer function of the filtered series at each iterate, $\Phi^{(n)}(f) \equiv P_X^{(n)}(f)/P_X^{(0)}(f)$, is shown in Figure 8, and these indicate a low pass filter with a cut-off at $f \approx 0.4$, and this approximates the same in Figure 7. The series (Figure 8), and in particular the phase portraits, also subjectively indicate that similar results to those in Figure 7 have been achieved after the second iterate.

4 Conclusions

The analysis of SSA presented here suggests that it is equivalent to a linear spectral filter, but with the important qualification that local information is maintained due to projection to a moving window. Although the filter is data-adaptive, the basis functions which specify the spectral dependence of the filter are determined by the correlation structure of the time series and the length of the moving window. When such filters are iterated, it has been found that the resulting sequence of filtered series can show some initial improvement in the quality of the filter, as measured by the lack of predictability of the successive residuals, thus apparently giving an improved characterisation of the uncontaminated signal, at least when white noise is added. However, continued iteration eventually leads to decimation of the signal, and moreover the results are very dependent on the choice of window used, since effectively the filters act as local band pass filters. Thus, for example, pure white noise can be filtered to produce apparently cyclic signals and this iterative filter is therefore best used when there is (at least one) clearly identifiable recurrence time.

It has also been shown, both theoretically and in practical demonstration, that widely different results can be obtained by end and centre projection. More generally, the choice of embedding style, embedding dimension, projection location, and projection dimension have a significant effect on the nature of the resulting filter.

Much has been made recently of the applicability of SSA to the analysis of data sets of climatic, geophysical significance, with the particular aim of identifying oscillatory trends in the data (Elsner & Tsonis, 1991; Plaut *et al.*, 1995), and this work has been critically appraised by Allen & Smith (1996). In a similar vein, Pilgram *et al.* (1995) and Mineva & Popivanov (1996) seek to dynamically characterise EEG signals in spite of the qualitative criticisms of SSA by Palus & Dvorak (1992) which we have quantified here. Our results

reaffirm that applications of SSA should be treated with some degree of caution, since for example it is possible to enhance low frequency components in signal filtering, and this could lead to the apparent detection of spurious low frequency dynamics (or trends) in what may simply be a red noise process.

As a recommendation, we consider that purely data-driven analyses of the type discussed here require a 'control' or 'benchmark' system, against which hypotheses can be tested. Typically, this will be some form of model, and we therefore term the combined process *nonlinear modelling*.

Appendix

Instead of a centred window for each t, a *backward window* can be defined as the interval $(t-2\tau, t)$ and $2\tau = \tau_w$ is the window length. Each value $X(t)$ of the time series is associated with a local window $X(t-s)$, where $s \in (0, 2\tau)$. To define the eigenvalue problem in term of the centred window formalism, set $\rho(t) = \phi(t+\tau)$, so that

$$A\rho = \lambda\rho \tag{A 1}$$

can be written, with $t = -\tau + \xi$,

$$\lambda\phi(\xi) = \frac{1}{2\tau}\int_0^{2\tau} C_X(\xi - u)\phi(u)\mathrm{d}u, \tag{A 2}$$

and $\xi = 0$ and $\xi = 2\tau$ correspond to the endpoints of the backward window. Writing $C_X/2\tau = c$, the limit as $\tau \to \infty$ of the eigenvalue problem is

$$\lambda\phi = \int_0^{\infty} c(\xi - u)\phi(u)\mathrm{d}u, \tag{A 3}$$

and this problem can be solved using the Wiener–Hopf technique (Carrier *et al.*, 1966).

$\phi(\xi)$ is defined for $\xi > 0$; its definition is extended by putting $\phi = 0$ for $\xi < 0$, thus

$$h(\xi) + \lambda\phi(\xi) = \int_{-\infty}^{\infty} c(\xi - u)\phi(u)\mathrm{d}u, \tag{A 4}$$

where $h = 0$ for $\xi > 0$, but h is to be determined for $\xi < 0$. Taking Fourier transforms of (A.3) gives (with $\hat{\phi}(f) = \int_{-\infty}^{\infty} \phi(\xi)e^{2\pi i f\xi}d\xi$, etc.)

$$(\hat{c} - \lambda)\hat{\phi}_+ = \hat{h}_-. \tag{A 5}$$

The transforms (and this equation (A 5)) are defined for real values of f, but they can be extended to complex f. Since $\phi = 0$ for $\xi < 0$, the transform $\hat{\phi}$ will be analytic in an upper half plane Im $f > 0$, and the transform notation is written as $\hat{\phi}_+$ to indicate this. Similarly, the transform of h is $\hat{h}_-$ and is analytic in Im $f < 0$. However, $\hat{\phi}$ may have (isolated) singularities on the real axis.

The aim is now to choose functions $G_\pm(z)$ (the notation indicating G_+ and G_- are respectively analytic in Im $z > 0$ and < 0), so that

$$\lambda - \hat{c}(f) = G_+(f)/G_-(f) \text{ for } f \in \mathbf{R}, \tag{A 6}$$

since then $G_+\hat{\phi}_+ = -G_-\hat{h}_-$, and, presuming there is a common strip of analyticity, $G_+\hat{\phi}_+$ then extends to an entire function which is determined by conditions at infinity. Now

note that $\hat{c}$ is real and positive for real f, by the Wiener–Khintchin theorem. It is also symmetric in f. Also $\hat{c} \to 0$ as $f \to \infty$.

Suppose firstly that $\lambda > \hat{c}$ for all real f; then $\lambda > 0$ and we define functions $L_\pm$ via

$$\ln L_+ - \ln L_- = \ln[(\lambda - \hat{c}(f))/\lambda], \tag{A 7}$$

for which a solution is

$$L(z) = \exp\left[\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\ln[(\lambda - \hat{c}(f))/\lambda]df}{f - z}\right] \tag{A 8}$$

(the integral exists, since $\hat{c}(f) \to 0$ at ∞), and $L(\infty) = 1$. Thus $\lambda - \hat{c} = \lambda L_+/L_-$, so that $\lambda L_+\hat{\phi}_+ = -L_-\hat{h}_-$. Now it is expected that $\hat{\phi}_+$, $\hat{h}_- \to 0$ as $z \to \infty$ (being Fourier transforms). Therefore, since $L_\pm \to 1$, the analytic continuation of $L_+\hat{\phi}_+$ is zero everywhere, and thus $\hat{\phi}_+ = 0$. It follows that if $\lambda > \hat{c}(f)$ everywhere, then $\phi = 0$. We therefore suppose that $\hat{c}(f) = \lambda$ at real values $\pm f_1, \pm f_2, \ldots, \pm f_n$.

Consider for example the case that $\hat{c}(0) > \lambda$, which will typically apply if n is odd. Define the polynomial

$$p(f) = \prod_{i=1}^{n}(f^2 - f_i^2), \tag{A 9}$$

then $g(f) = [\lambda - \hat{c}(f)]/p(f)$ is of one sign (positive) for all real f. Therefore, $\ln[g(f)]$ is well-defined for all real f. However, the integral corresponding to (A 8) does not exist. Therefore,

$$L(z) = \exp\left[\frac{1}{2\pi i}\int_{-\infty}^{\infty} \ln\left\{\frac{[\lambda - \hat{c}(f)](f - z_0)^n(f - \bar{z}_0)^n}{\lambda p(f)}\right\}\frac{\mathrm{d}f}{f - z}\right], \tag{A 10}$$

is defined where z_0 is an arbitrary complex number with $\mathrm{Im}\, z_0 > 0$. If $l - i\gamma$, $l + i\gamma$, $\gamma > 0$, are defined to be the branches of $\ln(f - z_0)$ and $\ln(f - \bar{z}_0)$ then the logarithm in (A 10) tends to zero as $f \to \pm\infty$, and the integral exists. Moreover, $L(\infty) = 1$, and

$$L_+/L_- = \frac{[\lambda - \hat{c}(f)](f - z_0)^n(f - \bar{z}_0)^n}{\lambda p(f)}; \tag{A 11}$$

it follows that

$$\lambda p L_+\hat{\phi}_+\left[\frac{1}{(f - \bar{z}_0)^n}\right]_+ = -L_-\hat{h}_-(f - z_0)^n, \tag{A 12}$$

and thus either side defines an entire function, a polynomial by Liouville's theorem, and hence (since $L \to 1$, $\hat{h} \to 0$ at ∞) of degree $n - 1$. It therefore follows that

$$\hat{\phi}_+(f) = \frac{(f - \bar{z}_0)^n p_{n-1}(f)}{\lambda p(f) L_+(f)}, \tag{A 13}$$

where p_{n-1} is an arbitrary polynomial of degree $n - 1$.

The inversion contour must be indented above the zeroes of p (since $\hat{\phi}$ is analytic in $\mathrm{Im}\, f > 0$), thus

$$\phi(\xi) = \int_C \frac{(f - \bar{z}_0)^n p_{n-1}(f) e^{-2\pi i f\xi}\mathrm{d}f}{\lambda p(f) L_+(f)}, \tag{A 14}$$

where C is the real axis indented upwards at $\pm f_i$. To compute the integral, the contour is completed in the lower half-plane, and to do this, (A 11) is used to continue the integrand

there, and thus

$$\phi(\xi) = \int_{\bar{C}} \frac{p_{n-1}(f)e^{-2\pi i f\xi}\mathrm{d}f}{[\lambda - \hat{c}(f)](f - z_0)^n L_-(f)}, \tag{A 15}$$

where $\bar{C}$ is the union of the upwardly-indented real axis and the limit of a large semi-circle $|f| = R$, $\mathrm{Im}\, f < 0$, as $R \to \infty$.

The evaluation of (A 15) depends on the singularities of $[\lambda - \hat{c}(f)]^{-1}$. On the real axis, these are $f = \pm f_i$, whence

$$\phi = \sum_{\substack{k=-n \\ k \neq 0}}^{n} A_k e^{-2\pi i f_k \xi}, \tag{A 16}$$

where $A_{-k} = \bar{A}_k$ if ϕ is real (as can be assumed). Note that, from (A.10), $L_- = (z - z_0)^{-n}\psi_-$, $L_+ = (z - \bar{z}_0)^n \psi_+$, where $\psi_\pm$ are analytic in $\mathrm{Im}\, z \neq 0$, so that (A 15) is independent of z_0 and $\bar{z}_0$ (as it must be). There are n independent solutions, corresponding to the coefficients of $p_{n-1}(f)$. If $\hat{c} = \lambda$ for some complex f in $\mathrm{Im} f < 0$, then for each such $f = f^*$, ϕ has a corresponding term proportional to $\exp[-2\pi i f^* \xi] = \exp[-2\pi f_I \xi]\exp[2\pi i f_R \xi]$, if $f^* = f_R - i f_I$. Thus, in general, the (real) eigenvalues are given by

$$\lambda = \hat{c}(f) = \frac{1}{2\tau}\hat{C}_X(f) \tag{A 17}$$

(just as in (2.5)), and the corresponding eigenfunctions are

$$\phi = \sin(2\pi f \xi + v) + \exp[-O(\xi)], \tag{A 18}$$

where the phase is fixed, and the exponential term is an end correction when $\xi = 0$ and is exponentially small when $\xi = 2\tau \gg 1$. If $\lambda - \hat{c} \neq 0$, for all complex f, then there is no end correction.

As an example, consider the autocorrelation function for a red-noise process, with variance one:

$$C_X(t) = e^{-a|t|}, \tag{A 19}$$

where $a > 0$. Clearly

$$\hat{c}(f) = \frac{a}{\tau(a^2 + 4\pi^2 f^2)}, \tag{A 20}$$

and there is an eigenvalue λ for all $0 < \lambda < 1/\tau a$, and since the only two roots of $\hat{c} = \lambda$ are the real values

$$f = \pm\frac{a}{2\pi}\left[\frac{1}{\lambda\tau a} - 1\right]^{1/2} = \pm f_1 \tag{A 21}$$

there is no end-correction, $n = 1$ (so $p_{n-1} = A =$ constant), and the eigenfunctions are given by

$$\phi = A\int_{\bar{C}} \frac{e^{-2\pi i f\xi}\mathrm{d}f}{[\lambda - \hat{c}(f)](f - z_0)L_-(f)}. \tag{A 22}$$

In this case, from (A 22) and (A 21),

$$\lambda - \hat{c}(f) = \frac{\lambda p}{f^2 + f_0^2}, \tag{A 23}$$

where $p = f^2 - f_1^2$ and $f_0 = a/2\pi$, so that L satisfies

$$\frac{L_+}{L_-} = \frac{(f - z_0)(f - \bar{z}_0)}{(f - if_0)(f + if_0)}. \tag{A 24}$$

It is easiest to take $z_0 = if_0$, so that $L = 1$ and (A 22) is

$$\phi = A \int_{\mathcal{C}} \frac{(f + if_0)e^{-2\pi if\xi} df}{\lambda(f^2 - f_1^2)}. \tag{A 25}$$

Calculating the residues,

$$\begin{aligned}\phi &= -\frac{A}{\lambda}\left[\frac{1}{2}\left(1 + \frac{if_0}{f_1}\right)e^{-2\pi if_1\xi} + \frac{1}{2}\left(1 - \frac{if_0}{f_1}\right)e^{-2\pi if_1\xi}\right] \\ &= \tilde{A}\left[\cos(2\pi f_1\xi) + \left\{\frac{1}{\lambda\tau a} - 1\right\}^{-1/2}\sin(2\pi f_1\xi)\right],\end{aligned} \tag{A 26}$$

where $\tilde{A} = -A/\lambda$. See also Carrier *et al.* (1966, p. 388f).

The expression (A 26) can be loosely compared with the results of Vautard & Ghil (1989). Given that the result above applies for $\tau \to \infty$, we see that the eigenfunction is a linear combination of the even and odd eigensolutions (2.11a) and (2.12a) of Vautard and Ghil. However, the point is that their analysis uses a centred window, which allows both solutions corresponding to the same eigenvalue, whereas the use (as here) of a back projection gives only *one* eigensolution for each eigenvalue. This distinction is clear in the continuous limit we study, although it will inevitably be less so in the practical, discrete implementation of SVD.

In summary, the eigensolution structure for end projection differs from that of centred windows in the phase specification (and the possible exponential end correction at $\xi = 0$). These differences allow for the possibility that the eigenfunction expansion (1.6) converges at the end points, i.e. for $s = \pm\tau$, although this point is not pursued here.

References

[1] Abramowitz, M. & Stegun, I. (eds.) 1964 *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, DC.

[2] Allen, M. R. & Smith, L. A. 1996 Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise. *J. Climate* **9**, 3373–3404.

[3] Bender, C. M. & Orsag, S. A. (1978) *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, New York.

[4] Broomhead, D. S. & King, G. P. 1986 Extracting qualitative dynamics from experimental data. *Physica D* **20**, 217–236.

[5] Buzug, Th. & Pfister, G. 1992 Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behavior of strange attractors. *Phys. Rev. A* **45**, 7073–7084.

[6] Carrier, G. F., Krook, M. & Pearson, C. E. 1966 *Functions of a Complex Variable*, McGraw-Hill, New York.

[7] Dehant, V., Ducarme, B. & Defraigne, P. D. 1993 New analysis of the superconducting gravimeter data of Brussels. In: J.-L. Le Mouël, D. E. Smylie and T. Herring (eds.), *Dynamics of Earth's Deep Interior and Earth Rotation, Geophys. Monog.* **72**, 35–44. AGU, Washington, DC.

[8] Elsner, J. B. & Tsonis, A. A. 1991 Do bidecadal oscillations exist in the global temperature record? *Nature* **353**, 551–553.

[9] Fleming, P. J., Levine, M. R., Long, A. M. & and Cleave, J. P. 1988 Postneonatal development of respiratory oscillations. *Ann. N.Y. Acad. Sci.* **533**, 305–313.

[10] Fowler, A. C., Kember, G., Johnson, J., Walter, S. J., Fleming, P. & Clements, M. 1994 A method for filtering respiratory oscillations. *J. Theor. Biol.* **170**, 273–281.

[11] Fraser, A. M. & Swinney, H. L. 1986 Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**, 1134–1140.

[12] Ghil, M. & Vautard, R. 1991 Interdecadal oscillations and the warming trend in global temperature time series. *Nature* **350**, 324–327.

[13] Gibson, J. F., Farmer, J. D., Casdagli, M. & Eubank, S. 1992 An analytic approach to practical state space reconstruction. *Physica* **57**, 1–30.

[14] Kember, G. & Fowler, A. C. 1993 A correlation function for choosing time delays in phase portrait reconstructions. *Phys. Letts. A* **179**, 72–80.

[15] Liebert, W. & Schuster, H. G. 1989 Proper choice of the time delay for the analysis of chaotic time series. *Phys. Letts. A* **142**, 107–111.

[16] Mineva A. & Popivanov, D. 1996 Method for single-trial readiness potential identification, based on singular spectrum analysis. *J. Neurosci. Math.* **68**, 91–99.

[17] Palus, M. & Dvorak, I. 1992 Singular-value decomposition in attractor reconstruction: pitfalls and precautions. *Physica D* **55**, 221–234.

[18] Pilgram, B., Schappacher, W., Löscher, W. N. & Pfurtscheller, G. 1995 Application of the correlation integral to respiratory data of infants during REM sleep. *Biol. Cybern.* **72**, 543–551.

[19] Plaut, G., Ghil, M. & Vautard, R. 1995 Interannual and interdecadal variability from a long temperature time series. *Science* **268**, 710–713.

[20] Press, W. H., Flannery, B. P., Teukolsky, S. A. & and Vetterling, W. T. 1989 *Numerical Recipes*, Cambridge University Press, Cambridge.

[21] Provenzale, A., Smith, L. A., Vio, R. & Murante, G. 1992 Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D* **58**, 31–49.

[22] Robbe, H. W. J., Mulder, L. J. M., Rüddel, H., Langewitz, W. A., Veldman, J. B. P. & Mulder, G. 1987 Assessment of baroreceptor reflex sensitivity by means of spectral analysis. *Hypertension* **10**, 538–543.

[23] Sugihara, G. & May, R. M. 1990 Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**, 734–741.

[24] Vautard, R.& Ghil, M. 1989 Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D* **35**, 395–424.

[25] Vautard, R., Yiou, P. & Ghil, M. 1992 Singular spectrum analysis: a toolkit for short, noisy chaotic signals. *Physica D* **58**, 95–126.