PROCEEDINGS OF THE ROYAL SOCIETY A | MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

A kepstrum approach to filtering, smoothing and prediction with application to speech enhancement

T. J. Moir and J. F. Barrett

Proc. R. Soc. Lond. A 2003 **459**, 2957-2976
doi: 10.1098/rspa.2003.1137

10.1098/rspa.2003.1137

THE ROYAL SOCIETY

A kepstrum approach to filtering, smoothing and prediction with application to speech enhancement

By T. J. Moir[1] and J. F. Barrett[2]†

[1] *Institute of Information and Mathematical Sciences, Massey University at Albany, Private Bag 102-904, Auckland, New Zealand* (t.j.moir@massey.ac.nz)

[2] *40 Botley Road, Southampton SO19 0NS, UK*

Received 17 April 2002; accepted 16 February 2003; published online 10 September 2003

A kepstrum (or complex-cepstrum) approach to minimum-phase Wiener filtering of stationary scalar processes is proposed and solved for the case of signal plus coloured noise, where the noise possibly includes a white-noise component. A general solution is found in an innovations form. The spectral factorization of the noise model and of the signal-plus-noise model required for the solution are determined from data using the kepstrum technique with the fast Fourier transform. This approach avoids dependence on any form of multidimensional state-space or polynomial-based model and so avoids use of recursive parameter estimation or of Diophantine equations.

Keywords: kepstrum; complex cepstrum; smoothing; prediction; coloured noise

1. Introduction

As is well known, the solution of the minimum mean-square filtering problem was originally found by Kolmogorov (1939, 1941) in discrete time and by Wiener (1949) in continuous time. While these solutions are quite general, they do not result in a convenient closed form for the solution. Instead, they involve awkward spectral factorization and the separation of causal from uncausal terms. It is always possible to work out the solution in any particular case, but for computational ease in the general case and especially for the extension to time-varying systems, new methods have been developed. The first was the Kalman filter in 1960, requiring solution of the matrix-Riccati equation. A later approach was the direct parameter estimation of the innovations model using extended least squares and similar algorithms (Hagander & Wittenmark 1977; Ljung & Söderström 1982). Still another approach was the use of polynomial Diophantine equations (Grimble 1985; Roberts & Newmann 1988; Dabis & Moir 1993). The direct estimation and use of the innovations model has the advantage that algebraic spectral factorization is no longer necessary, the spectral factor being estimated directly from the data. The same method can equally be applied to non-stationary signals by sectioning the signal plus noise into small intervals (assumed to be quasi-stationary). However, when tackling such problems

† Present address: Post Box 6, Limni, Evia 34005, Greece.

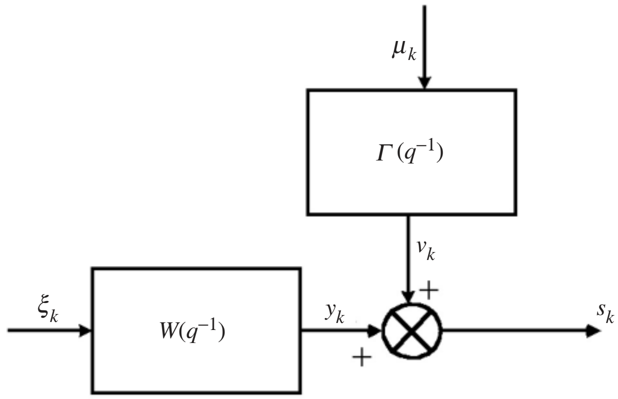

Figure 1. Signal-and-noise generating model.

as speech enhancement (as covered in this paper), the model order rapidly increases, depending on the particular characteristics of the noise. Furthermore, the speech itself requires a high-order model, so that the composite model requires a very large number of parameters, which can be a computational burden for real-time applications. Moreover, algorithms based on the generic recursive least-squares (RLSs) approach do not track very well and require ad hoc approaches such as exponential forgetting factors and so-called 'jacketing software' to avoid possible instability with erroneous data. Similarly, Diophantine equations (which for some problems require simultaneous solution of two such equations (Grimble 1988; Dabis & Moir 1993)) also give a computational problem with higher-order systems. These approaches give rise to the infinite-impulse-response (IIR) adaptive filter or self-tuning filter.

An alternative approach, which has proved more popular in practice, is to use finite-impulse-response (FIR) adaptive filtering based on least-means squares (LMSs) (Widrow & Stearns 1985; Haykin 1986). The tracking ability of LMS-based adaptive filters is generally better than those of RLS and the stability is assured provided the step size is kept below a critical value depending on signal-plus-noise power. Such a method can be thought of as implicit in that, unlike the previous explicit case, the filter is estimated directly via LMS by minimizing the mean-square error rather than computing a filter based on individually estimated component parts (spectral factor, noise variances, etc.). The only slight disadvantage is that the LMS method can also give rise to high-order models (i.e. a large number of weights) in realistic problems and more so in that an FIR model must be larger than an equivalent IIR model, as the latter has both poles and zeros to model the dynamics. This is offset by the fact that the LMS method is by far computationally simpler than the RLS method.

It is also possible to perform optimal estimation using the Kalman (1960) approach. This has been applied to speech processing by Popescu & Zeljkovic (1998), but this also has its drawbacks, since a high-order Riccati equation then needs to be solved iteratively. There are, of course, many other approaches to the problem, e.g. spectral subtraction (Stahl *et al*. 2000; Martinez *et al*. 2001).

The approach used in the present paper is explicit: it does not rely on RLS, but is based around the identification of the spectral factor (and noise variances) using kepstrum (complex-cepstrum) methods and the resulting estimators (filter, smoother and predictor) are innovations based. A distinction is made here between 'kepstrum' (Silvia & Robinson 1978) and 'complex cepstrum' (Oppenheim *et al.* 1968; Oppenheim & Schafer 1975), in that the kepstrum coefficients, as given by the Kolmogorov power series, are theoretical values, while the complex cepstra using the fast Fourier transform (FFT) are estimates of these. The same symbol 'k' is used for both throughout the paper, however, to avoid confusion. For the purposes of this paper, which defines the theoretical framework, it is assumed that the noise can be measured accurately when no signal component is present. The same approach is used as in the LMS method, in that the noise is assumed to be measured separately from the signal plus noise with a second sensor. The FFT approach leads to a computationally sound and stable method for adaptive estimation.

2. Problem statement

The problem we consider is that of filtering a scalar stationary signal from a message process of signal plus additive coloured noise, which may also include white noise. The model is quite general in that it is also valid for the white-noise-only case.

(*a*) *Signal and noise models*

Consider a signal y_k corrupted with additive coloured noise v_k giving a message

$$s_k = y_k + v_k. \tag{2.1}$$

As illustrated in figure 1, the signal and coloured noise are assumed to be generated from stable minimum-phase transfer functions driven by stationary uncorrelated white-noise sources ξ, μ:

$$y_k = W(q^{-1})\xi_k, \tag{2.2}$$

$$v_k = \Gamma(q^{-1})\mu_k. \tag{2.3}$$

For the white-noise case, $\Gamma(\zeta)$ is constant. In practice, there will be usually be a white-noise component to the noise as well as a coloured component, but in this case it will be assumed that the noise spectra have been combined, giving a resulting rational transfer function $\Gamma(\zeta)$ shaping the noise.

The colouring transfer functions $W(\zeta)$, $\Gamma(\zeta)$ may, without loss of generality, be assumed to be such that $W(0)$, $\Gamma(0) = 1$, since any constant multiplier may be absorbed into the variances σ_ξ^2, σ_μ^2 of the white-noise processes ξ_k, μ_k. These processes are then innovations processes.

The signal spectrum transform is

$$g_{\mathrm{ss}}(\zeta) = W(\zeta)W(\zeta^{-1})\sigma_\xi^2, \tag{2.4}$$

which is also the cross-spectrum transform $g_{y\mathrm{s}}(\zeta)$ of message and noise, since signal and noise are uncorrelated. The noise spectrum transform is

$$g_{\mathrm{vv}}(\zeta) = \Gamma(\zeta)\Gamma(\zeta^{-1})\sigma_\mu^2 \tag{2.5}$$

and the spectrum transform of the message (signal plus noise) is

$$g_{\mathrm{ss}}(\zeta) = W(\zeta)W(\zeta^{-1})\sigma_\xi^2 + \Gamma(\zeta)\Gamma(\zeta^{-1})\sigma_\mu^2. \tag{2.6}$$

In the case when signal is absent (e.g. in a so-called noise-only period), the message spectrum transform reduces to the noise spectrum. This gives information about the noise during silence periods of speech or simply if the signal is switched off intentionally for a duration.

The message spectrum transform can be factorized as $\Lambda(\zeta)$,

$$g_{\mathrm{ss}}(\zeta) = \Lambda(\zeta)\Lambda(\zeta^{-1}), \tag{2.7}$$

where here $\Lambda(\zeta)$ will reduce to $\Gamma(\zeta)\sigma_\mu$ when the signal is absent. The corresponding normalized spectral factor $Z(\zeta)$ is defined by $\Lambda(\zeta)/\Lambda(0)$, leading to the noise representation of the message process as

$$s_k = Z(q^{-1})\varepsilon_k, \tag{2.8}$$

which is the innovations model where ε_k is the white innovations sequence for the message process having variance σ_ε^2, where $\Lambda(0) = \sigma_\varepsilon$ and

$$\Lambda(\zeta) = Z(\zeta)\Lambda(0). \tag{2.9}$$

This gives the relation

$$Z(\zeta)Z(\zeta^{-1})\sigma_\varepsilon^2 = W(\zeta)W(\zeta^{-1})\sigma_\xi^2 + \Gamma(\zeta)\Gamma(\zeta^{-1})\sigma_\mu^2 \tag{2.10}$$

between the normalized spectral transforms of message, signal and noise.

(b) *Wiener filter*

The optimal Wiener estimator of a signal y from a message s may be represented as

$$\hat{y}_k = H(q^{-1})s_k, \tag{2.11}$$

where the transfer function $H(\zeta)$, assumed physically realizable, is chosen to minimize the mean-square error

$$E[e_k^2] = E(y_k - \hat{y}_k)^2. \tag{2.12}$$

The solution may be found explicitly as

$$H(\zeta) = \left[\frac{g_{ys}(\zeta)}{\Lambda(\zeta^{-1})}\right]_+ \frac{1}{\Lambda(\zeta)}. \tag{2.13}$$

Here, the notation $[\cdot]_+$ is normally interpreted as meaning that in the expansion of the function inside the brackets into positive and negative powers of ζ, only those terms in ζ^i are kept for $i = 0, 1, 2, \ldots$.

The same solution may be shown to apply when the estimator has the form

$$H(\zeta) = \sum_{i\in\Im} h_i\zeta^i, \tag{2.14}$$

where the estimation takes place over some interval $\Im$ of integers, the bracket $[\cdot]_+$ then meaning that, in the expansion of the function inside the bracket, only those terms in ζ^i are kept for $i \in \Im$ (Barrett & Moir 1987). This leads to a general solution covering lagged filters, prediction and truncation.

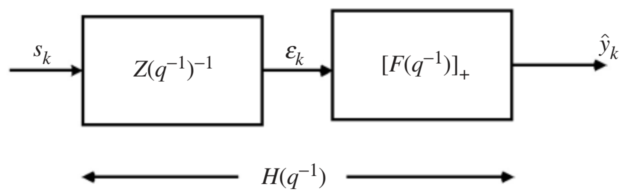

Figure 2. Innovations form of optimal Wiener filter.

On putting

$$F(\zeta) = \frac{g_{ys}(\zeta)}{\Lambda(\zeta^{-1})}, \tag{2.15}$$

the optimal filter may be represented as illustrated in figure 2. Here, the first operation is a whitening filter, giving the innovations process, and the second operation is the spectrum-shaping function $[F(\zeta)]_+$. This interpretation of the optimal filter derives from the approach of Kolmogorov (1941).

(*c*) *Special cases of the Wiener filter*

From the equations

$$g_{ys}(\zeta) = g_{yy}(\zeta) = g_{ss}(\zeta) - g_{vv}(\zeta), \tag{2.16}$$

and using the normalized spectral factors $Z(\zeta)$, $\Lambda(\zeta)$, we can write the optimal filter as

$$H(\zeta) = \left[Z(\zeta) - \frac{\Gamma(\zeta)\Gamma(\zeta^{-1})\sigma_\mu^2}{Z(\zeta^{-1})\sigma_\varepsilon^2}\right]_+ \frac{1}{Z(\zeta)}. \tag{2.17}$$

We now consider the three possible cases.

(i) *Filtering,* $\Im_f = \{0, 1, 2, \dots\}$

This corresponds to instantaneous estimation $\hat{y}_{k/k}$, i.e. information up to and including time k. The estimator becomes

$$H(\zeta) = 1 - \left[\frac{\Gamma(\zeta)\Gamma(\zeta^{-1})}{Z(\zeta^{-1})}\right]_+ \frac{1}{Z(\zeta)} \frac{\sigma_\mu^2}{\sigma_\varepsilon^2}. \tag{2.18}$$

This solution involves the two spectral factors: $Z(\zeta)$ when there is signal and $\Gamma(\zeta)$ when the signal is removed. Since both can be estimated using the kepstrum method, it remains to find a method of simplifying the $[\cdot]_+$ brackets.

(ii) *Fixed-lag smoothing and estimation,* $\Im_s = \{-d, -d+1, -d+2, \dots\}$, $d > 0$

This case gives $\hat{y}_{k/k+d}$, i.e. an estimate of the signal at time k with information up to and including time $k+d$. As this involves future values, it may be reinterpreted as giving $\hat{y}_{k-d/k}$. The estimator becomes

$$H(\zeta) = 1 - \left[\frac{\Gamma(\zeta)\Gamma(\zeta^{-1})}{Z(\zeta^{-1})}\right]_+ \frac{1}{Z(\zeta)} \frac{\sigma_\mu^2}{\sigma_\varepsilon^2}, \tag{2.19}$$

which looks identical to (2.18) except that the observation interval is different. The difference does not show itself in the form of (2.19) above and a further simplification using innovations representations must be sought.

(iii) *Prediction*, $\Im_{\mathrm{p}} = \{d, d+1, d+2, \dots\}$, $d > 0$

This case gives the predicted estimate $\hat{y}_{k/k-d}$, i.e. an estimate of the signal at time k with information up to and including time $k-d$. The estimator becomes

$$H(\zeta) = [Z(\zeta)]_+ \frac{1}{Z(\zeta)} - \left[\frac{\Gamma(\zeta)\Gamma(\zeta^{-1})}{Z(\zeta^{-1})}\right]_+ \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} \frac{1}{Z(\zeta)}. \tag{2.20}$$

From (2.18)–(2.20), the classical Wiener filter is expressed entirely in terms of the two unique spectral factors $Z(\zeta)$ and $\Gamma(\zeta)$, and the problem becomes one of providing innovations models for the three estimators and that of estimating the spectral factors themselves. The innovations models that naturally give rise to the removal of the $[\cdot]_+$ brackets are considered first.

(*d*) *Innovations form for estimators*

The estimators (2.18)–(2.20) in the preceding section are of little computational use, as they require the removal of the $[\cdot]_+$ brackets. It is possible, however, to proceed further by separating the terms within the brackets. Since all three equations have a similar expression, consider the following two power-series expansions

$$\frac{\Gamma(\zeta^{-1})}{Z(\zeta^{-1})} = 1 + p_1\zeta^{-1} + p_2\zeta^{-2} + \cdots \tag{2.21}$$

and

$$\Gamma(\zeta) = 1 + \gamma_1\zeta + \gamma_2\zeta^2 + \cdots. \tag{2.22}$$

Here, we may use p_0 and γ_0 (equal to 1) for the initial constant term. The p_i, $i = 0, 1, 2\dots$, are then impulse-response coefficients of an uncausal sequence, whereas the γ_i, $i = 0, 1, 2\dots$, are impulse-response coefficients of a causal sequence. Multiplication of (2.21) by (2.22) gives

$$\frac{\Gamma(\zeta)\Gamma(\zeta^{-1})}{Z(\zeta^{-1})} = \sum_{k=-\infty}^{\infty} c_k\zeta^k, \tag{2.23}$$

where the coefficients c_k may be found for positive values of k as

$$c_k = \sum_{i=0}^{\infty} p_i\gamma_{k+i}, \quad k = 0, 1, 2, \dots, \tag{2.24}$$

and, for negative values,

$$c_k = \sum_{j=0}^{\infty} p_{-k+j}\gamma_j, \quad k = -1, -2, \dots. \tag{2.25}$$

Having found the coefficients of the Laurent series, the bracket $[\cdot]_+$ may be evaluated for a general interval $\Im$ thus:

$$\left[\frac{\Gamma(\zeta)\Gamma(\zeta^{-1})}{Z(\zeta^{-1})}\right]_+ = \sum_{i\in\Im} c_i\zeta^i. \tag{2.26}$$

This formula includes the effect of truncation. In this case, the interval $\Im$ is of finite length and includes values up to ℓ, where $\ell \gg d$. Typically, for most applications, ℓ should be at least 16, but need not be more than 64.

(*e*) *Formulae using the innovations process*

We now reconsider the three cases using equation (2.8).

(i) *Filtering,* $\Im_{\rm f} = \{0, 1, 2, \dots\}$, $d > 0$

Using (2.18) and (2.26) results in

$$\hat{y}_{k/k} = s_k - \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} \sum_{i=0}^{\ell} c_i \varepsilon_{k-i}. \tag{2.27}$$

(ii) *Fixed-lag smoothing and filtering,* $\Im_{\rm s} = \{-d, -d+1, \dots, \ell\}$, $0 < d < \ell$

Use of (2.19) and (2.26) results in

$$\hat{y}_{k-d/k} = s_{k-d} - \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} \sum_{i=-d}^{\ell} c_i \varepsilon_{k-d-i}. \tag{2.28}$$

(iii) *Prediction,* $\Im_{\rm p} = \{d, d+1, \dots, \ell\}$, $0 < d < \ell$

The coloured-noise predictor has two parts. The first part is similar to the white-noise-predictor case. We use the expansion

$$Z(\zeta) = 1 + a_1\zeta + a_2\zeta^2 + \cdots. \tag{2.29}$$

Then, from (2.20), (2.26) and (2.29) comes the overall result:

$$\hat{y}_{k+d/k} = \sum_{i=d}^{\ell} \left\{ a_i - \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} c_i \right\} \varepsilon_{k+d-i}. \tag{2.30}$$

3. Kepstrum identification

This section describes the kepstrum technique and its application to the analysis and identification of spectral factors. The basic ideas here go back to Kolmogorov's fundamental work; they were later restated and developed by Silvia & Robinson (1978), who coined the word 'kepstrum'. The detailed application of this method in system analysis has been demonstrated in two previous papers (Barrett & Chen 1983; Barrett & Moir 1986).

(*a*) *Definition of kepstrum*

If a general discrete-time transfer function $H(\zeta)$ is both stable and minimum phase, i.e. with no poles or zeros inside or on the unit circle, then it is possible to define the *kepstrum*

$$K(\zeta) = \ln H(\zeta) \tag{3.1}$$

as a regular function within the unit circle. This can also be called *kepstrum generating function*, as it defines the *kepstrum coefficients*, which are the coefficients in the expansion

$$K(\zeta) = k_0 + k_1\zeta + k_2\zeta^2 + \cdots \tag{3.2}$$

valid within the unit circle. The use of the letter k here should cause no confusion with the previous use of subscript k as a discrete-time variable.

The kepstrum has, by virtue of the logarithm, the following properties.

(i) *Additivity for cascading*,

$$\ln\{H_1(\zeta)H_2(\zeta)\} = \ln H_1(\zeta) + \ln H_2(\zeta). \tag{3.3}$$

(ii) *Negation for inversion*,

$$\ln\{1/H(\zeta)\} = -\ln H(\zeta). \tag{3.4}$$

From these it follows that the kepstrum coefficients are also added and negated under cascading and inversion.

(*b*) *Determination of transfer function from kepstrum coefficients*

Suppose that the kepstrum function $K(\zeta)$ corresponds to a minimum-phase transfer function $H(\zeta)$. The kepstrum series for $K(\zeta)$ being known, the series for $H(\zeta)$ may be reconstructed from the equation

$$H(\zeta) = \exp(k_0 + k_1\zeta + k_2\zeta^2 + \cdots). \tag{3.5}$$

It is convenient to take out the constant multiplier and write this in the form

$$H(\zeta) = CZ(\zeta), \tag{3.6}$$

so that

$$C = \exp k_0, \tag{3.7}$$

$$\begin{aligned} Z(\zeta) &= \exp(k_1\zeta + k_2\zeta^2 + k_3\zeta^3 + \cdots) \\ &= a_0 + a_1\zeta + a_2\zeta^2 + a_3\zeta^3 + \cdots, \end{aligned} \tag{3.8}$$

where $a_0 = 1$. From this equation for $Z(\zeta)$, the coefficients are conveniently determined by a recursive method described by Silvia & Robinson (1978). Differentiation gives

$$\frac{\mathrm{d}Z(\zeta)}{\mathrm{d}\zeta} = (k_1 + 2k_2\zeta + 3k_3\zeta^2 + \cdots)Z(\zeta). \tag{3.9}$$

Then, substituting the series representation of $Z(\zeta)$, we get the recursive relations

$$na_n = \sum_{r=1}^{n} rk_r a_{n-r}, \quad n = 1, 2, 3, \ldots. \tag{3.10}$$

(*c*) *Determination of kepstrum coefficients for a spectrum*

The discrete-time spectral density $S(\theta)$ is related to the spectrum transform by

$$S(\theta) = g(\mathrm{e}^{\mathrm{j}\theta}). \tag{3.11}$$

It is a periodic function of the normalized frequency θ. If $g(\zeta)$ has a factorization

$$g(\zeta) = \Lambda(\zeta)\Lambda(\zeta^{-1}), \tag{3.12}$$

with $\Lambda(\zeta)$ being both stable and minimum phase, then

$$S(\theta) = |\Lambda(\mathrm{e}^{\mathrm{j}\theta})|^2. \tag{3.13}$$

From this equation, we have

$$\ln|\Lambda(\mathrm{e}^{\mathrm{j}\theta})| = \tfrac{1}{2}\ln S(\theta). \tag{3.14}$$

Now, using the kepstrum expansion for $\Lambda(\zeta)$, we find

$$\ln \Lambda(\mathrm{e}^{\mathrm{j}\theta}) = k_0 + k_1\mathrm{e}^{-\mathrm{j}\theta} + k_2\mathrm{e}^{-2\mathrm{j}\theta} + \cdots . \tag{3.15}$$

Taking the real part gives

$$\ln|\Lambda(\mathrm{e}^{\mathrm{j}\theta})| = k_0 + k_1\cos\theta + k_2\cos 2\theta + \cdots , \tag{3.16}$$

from which there follow the Fourier formulae for the kepstrum coefficients of $\Lambda(\zeta)$:

$$k_0 = \frac{1}{\pi}\int_0^{2\pi} \ln|\Lambda(\mathrm{e}^{\mathrm{j}\theta})|\,\mathrm{d}\theta, \tag{3.17}$$

$$k_n = \frac{1}{\pi}\int_0^{2\pi} \ln|\Lambda(\mathrm{e}^{\mathrm{j}\theta})|\cos(n\theta)\,\mathrm{d}\theta, \quad n = 1, 2, 3, \ldots . \tag{3.18}$$

Equally, in terms of spectrum, taking into account that it is a symmetrical function:

$$k_0 = \frac{1}{4\pi}\int_0^{2\pi} \ln S(\theta)\,\mathrm{d}\theta, \tag{3.19}$$

$$k_n = \frac{1}{2\pi}\int_0^{2\pi} \ln S(\theta)\cos(n\theta)\,\mathrm{d}\theta, \quad n = 1, 2, 3, \ldots . \tag{3.20}$$

Here, by the last section, k_0 gives the value of $\log\sigma$, σ^2 being the innovations variance. There follows the formula of Kolmogorov,

$$\sigma^2 = \exp\left\{\frac{1}{2\pi}\int_0^{2\pi} \ln S(\theta)\,\mathrm{d}\theta\right\}, \tag{3.21}$$

while the remaining coefficients $k_1, k_2, k_3, \ldots$ determine the expansion of $Z(\zeta)$, the normalized spectral factor $\Lambda(\zeta)/\Lambda(0)$.

4. Determination of the filter from data using the kepstrum

It has been shown that all three estimation problems in coloured noise can be solved, provided the two spectral factors of signal plus noise and noise alone, as well as the ratio $\sigma_\mu^2/\sigma_\varepsilon^2$ of innovations variances, can be determined. In this section we show how these quantities are calculated from observed data.

(*a*) *Estimation of kepstrum coefficients by the cepstrum*

The cepstrum is normally defined as the logarithm of the spectrum estimated from a data sample by using the FFT. This is closely related to the kepstrum, the two ideas being similar but not identical. The cepstrum is an empirical quantity approximating and estimating the theoretical kepstrum.

Suppose that $[x_0, x_1, x_2, \ldots, x_{N-1}]$ is a data vector of a sample of a stationary process. Use of the FFT gives a data vector $[X_0, X_1, X_2, \ldots, X_{N-1}]$, where

$$X_m = \sum_{n=0}^{N-1} x_n w^{mn}, \quad m = 0, 1, 2, \ldots, N-1, \tag{4.1}$$

and w is the so-called twiddle factor $\exp(2\pi\mathrm{j}/N)$.

The periodogram estimate of spectral density S_m at each frequency-bin m is

$$\hat{S}_m = \frac{1}{N}|X_m|^2, \quad m = 0, 1, 2, \ldots, N-1. \tag{4.2}$$

Here, the quantity on the right-hand side is the sum of the squares of the amplitudes of the quadrature components of the data for that frequency. It is known that these components may be considered to be independent Gaussian variables of zero means and equal variances. From this, it may be shown that

$$\ln \hat{S}_m = \ln S_m + f_m, \quad m = 0, 1, 2, \ldots, N-1, \tag{4.3}$$

where f_m are independent random variables having identical distributions with mean $-\gamma$ and variance $\frac{1}{6}\pi^2$, γ (equal to $0.577\,215\ldots$) being Euler's constant (Barrett & Moir 1986). The plot of $\ln \hat{S}_m$ against frequency $2\pi m/N$ shows a downward bias of $-\gamma$, as illustrated in Wahba (1980). This bias shows itself in a corresponding bias in the estimation of k_0, which gives the variance. The higher-order kepstrum coefficients $k_1, k_2, k_3, \ldots$ are unaffected by it. To eliminate the bias, the estimates of the kepstrum coefficients are found by inverse FFT of the vector

$$X_m = \ln \hat{S}_m + \gamma, \quad m = 0, 1, 2, \ldots, N-1, \tag{4.4}$$

these coefficients satisfying $k_i = k_{N-i}$, $i = 0, 1, 2 \ldots, \frac{1}{2}N - 1$.

For smoother estimates and also to take into account time-varying conditions, it is normal to overlap the FFT windows and provide some form of exponential smoothing to the periodogram, as described in Appendix A (Allen *et al.* 1977).

(*b*) *Spectral factor determination from estimated kepstrum coefficients*

The estimated kepstrum expansion of the normalized spectral factor $Z(\zeta)$ for signal plus noise is found from (3.8) as

$$\ln Z(\zeta) = k_1\zeta + k_2\zeta^2 + \cdots, \tag{4.5}$$

and the innovations variance is found from (3.7) and (2.9) with $C = \Lambda(0) = \sigma_\varepsilon$,

$$\sigma_\varepsilon^2 = \exp(2k_0), \tag{4.6}$$

where here the coefficients $k_0, k_1, k_2, \ldots, k_{N-1}$ represent estimated coefficients for the signal. Similarly for noise-only signals,

$$\ln \Gamma(\zeta) = k_1'\zeta + k_2'\zeta^2 + \cdots, \qquad \sigma_\mu^2 = \exp(2k_0'), \tag{4.7}$$

where $k_0, k_1, k_2, \ldots, k_{N-1}$ are the estimated kepstrum coefficients for the noise-only case. The transfer functions for the individual spectral factors can be found by the recursion method of Silvia & Robinson (1978). For the two series, we find the recursions as follows.

(i) *Signal-plus-noise spectral factor*, $Z(\zeta) = a_0 + a_1\zeta + a_2\zeta^2 + \cdots$ $(a_0 = 1)$,

$$na_n = \sum_{r=1}^{n} rk_r a_{n-r}, \quad n = 1, 2, \ldots. \tag{4.8}$$

(ii) *Noise-alone spectral factor*, $\Gamma(\zeta) = \gamma_0 + \gamma_1\zeta + \gamma_2\zeta^2 + \cdots$ $(\gamma_0 = 1)$,

$$n\gamma_n = \sum_{r=1}^{n} rk'_r \gamma_{n-r}, \quad n = 1, 2, \ldots. \tag{4.9}$$

(*c*) *Removal of the* $[\cdot]_+$ *brackets*

A little ingenuity now leads to an efficient way of determining the coefficients in the expansion of $\Gamma(\zeta^{-1})/Z(\zeta^{-1})$. Using the results of (4.8) and (4.9) it would, of course, be possible, with the change of variable from ζ to ζ^{-1}, to divide the two polynomials $\Gamma(\zeta^{-1})$ and $Z(\zeta^{-1})$. However, since division of transfer functions is performed in the kepstrum domain by subtraction of the kepstrum coefficients, it is only necessary to subtract the kepstrum coefficients for $Z(\zeta^{-1})$ from those of $\Gamma(\zeta^{-1})$ and convert them back to the transfer function by using the previous recursive formula. Hence, with $p_0 = 1$, we find

$$np_n = \sum_{r=1}^{n} r(k'_r - k_r)p_{n-r}, \quad n = 1, 2, \ldots. \tag{4.10}$$

Once these coefficients have been found, it remains only to perform the computations to find the coefficients of the Laurent series as given by (2.24) and (2.25).

(*d*) *Innovations estimation*

To estimate the innovations sequence, we use the inverse of the spectral factor $Z(\cdot)$,

$$\varepsilon_k = Z^{-1}(q^{-1})s_k. \tag{4.11}$$

In the expansion

$$Z^{-1}(\zeta) = 1 + \alpha_1\zeta + \alpha_2\zeta^2 + \cdots, \tag{4.12}$$

the coefficients $\alpha_1, \alpha_2, \alpha_3, \ldots$, follow from the recursion

$$n\alpha_n = -\sum_{r=1}^{n} k_r r\alpha_{n-r}, \quad n = 1, 2, 3, \ldots, \tag{4.13}$$

since, using (3.4), inversion changes the sign of the kepstrum coefficients. The innovations estimate now becomes

$$\hat{\varepsilon}_k = \sum_{n=0}^{\ell} \alpha_n s_{k-n}, \tag{4.14}$$

where the coefficients have been truncated at some suitable number ℓ of points.

These results give rise to the algorithms of the next section.

5. Algorithms for adaptive filtering, smoothing and prediction in coloured noise

Here we use the classical Wiener estimators with the kepstrum approach to perform spectral factorization. Innovations forms are used throughout. This is by no means essential, but is computationally convenient.

It is assumed that there is a period when the noise-only signal can be measured, i.e. when the signal is zero (or near zero) between spoken words of speech (Agaiby & Moir 1997) or, alternatively, a separate measurement is available where a sensor is placed near the noise source in a similar manner to that in conventional noise cancellation (Widrow & Stearns 1985).

Algorithm 5.1 (filtering ($d = 0$) and smoothing ($d > 0$)). *For each batch of data* s_k, $k = 0, 1, \dots, N-1$, *and* n_k, $k = 0, 1, \dots, N-1$, *where* N *is the FFT length, we do the following.*

Step 1. *For* $i = 1, 2, \dots, N-1$, *estimate the complex-cepstrum coefficients for the message* k_i *from (4.4) and the complex-cepstrum coefficients* k'_i *of the noise-only signal.*

Step 2. *The coefficients in step 1 are preserved up to some point* $\ell \leqslant \frac{1}{2}N - 1$ *and the rest set to zero (it is assumed that* $\ell \gg d$*). The zeroth coefficients* k_0 *and* k'_0 *are halved. Calculate the variance ratio*

$$\frac{\sigma_\mu^2}{\sigma_\varepsilon^2} = \exp 2(k'_0 - k_0).$$

Step 3. *Estimate the impulse response of the inverse of the signal-plus-noise spectral factor* α_i, $i = 1, 2, \dots, \ell$, *from (4.13).*

Step 4. *Estimate the innovations sequence* $\hat{\varepsilon}_k$ *from (4.14).*

Step 5. *Estimate the impulse-response coefficients* γ_i, $i = 1, 2, \dots, \ell$, *from (4.9) and the impulse-response coefficients* p_i, $i = 1, 2, \dots, \ell$, *from (4.10).*

Step 6. *Find the Laurent series from (2.24) and (2.25), with the summations truncated at some integer* ℓ, *where* $\ell \gg d$. *That is,*

$$c_k = \begin{cases} \displaystyle\sum_{i=0}^{\ell} p_i \gamma_{i+k} & \text{for } k \geqslant 0, \\ \displaystyle\sum_{j=0}^{\ell} p_{-k+j} \gamma_j & \text{for } k < 0, \end{cases}$$

and only the first of these is required in the filtering case when $d = 0$.

Step 7. *The filtered* ($d = 0$) *or d-steps-smoothed estimate of the signal is found from (2.28):*

$$\hat{y}_{k-d/k} = s_{k-d} - \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} \sum_{i=-d}^{\ell} c_i \hat{\varepsilon}_{k-d-i}.$$

Algorithm 5.2 (prediction ($d > 0$)). *For each batch of data* s_k, *with* $k = 0, 1, \dots, N-1$, *where* N *is the FFT length, perform the following computations.*

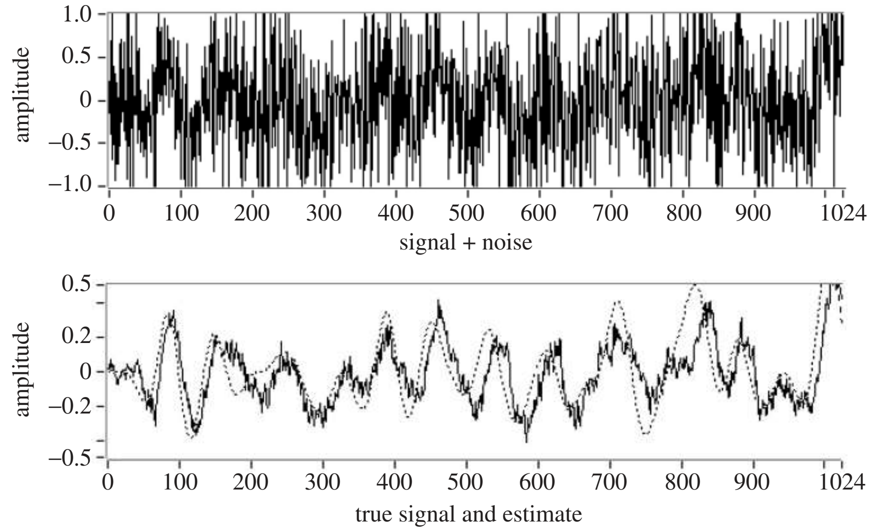

Figure 3. Performance of smoothing filter, $d = 5$ steps. (True signal shown with broken line.)

Steps 1–6. *Identical to that of the filtering-and-smoothing case, except in step 6* $k = d, d+1, d+2, \dots,$ *so only (2.24) is required.*

Step 7. *Estimate the impulse response* a_i, $i = 1, 2, \dots, \ell$, *of the signal-plus-noise spectral factor from (4.8).*

Step 8. *Estimate the* $d > 0$*-steps-ahead predicted estimate from (2.30)*

$$\hat{y}_{k+d/k} = \sum_{i=d}^{\ell} \left\{ a_i - \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} c_i \right\} \hat{\varepsilon}_{k+d-i}.$$

6. Illustrative examples

The overall performance of the kepstrum method can be illustrated with the following three examples, which were chosen to best illustrate the theoretical framework and the practical aspects of this method.

(*a*) *Stationary signal and noise*

First consider the problem of a stationary signal and stationary coloured noise plus white noise. To simulate the signal, unit variance white noise was passed through an eighth-order low-pass IIR Butterworth filter of cut-off frequency 400 Hz. The sampling frequency was 22 050 Hz and 8 bits per sample were used. To this an uncorrelated white-noise sequence of variance 0.2 and a coloured-noise signal were added. The coloured noise was simulated by passing uncorrelated white noise of variance 0.5 through a sixth-order IIR Butterworth bandpass filter. The bandpass filter had lower and upper corner frequencies of 3 and 6 kHz, respectively. An FFT length of $N = 1024$ was used and the signal-to-noise ratio (SNR) was measured as -6.0 dB. Figure 3 shows the signal plus noise, the true signal and the corresponding smoothed estimate for a lag of $d = 5$ steps. A further improvement can be made by averaging

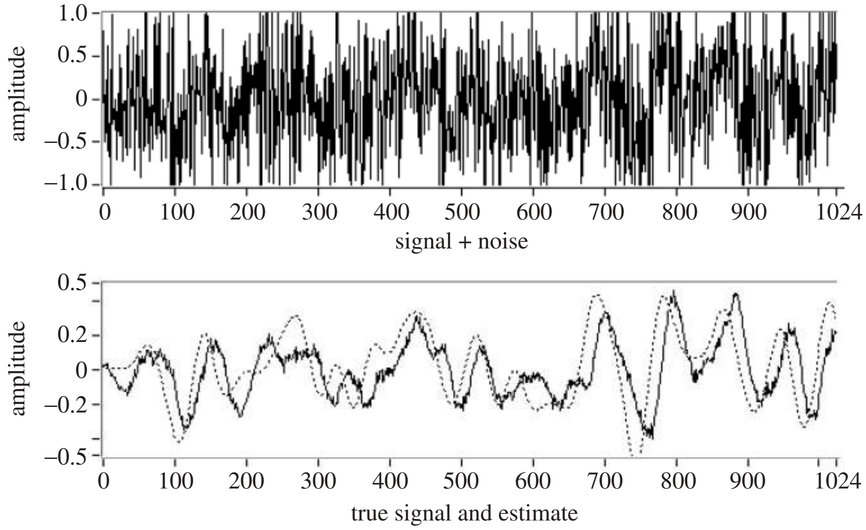

Figure 4. Performance of smoothing filter after averaging and $d = 12$ steps. (True signal shown with broken line.)

Table 1. *SNR improvement for smoothing* (dB)*, stationary example*

$\mathrm{SNR_{in}}$	3	0	−3	−6	−9	−12	−15	−18
$\mathrm{SNR_{out}}$	40	35	30	25	23	15	10.6	4.7

the cepstrum coefficients from overlapping frames (Appendix A). This is shown in figure 4, with $d = 12$ steps, and gives an excellent result.

It is not possible to calculate the SNR improvement directly, as this requires the signal to be of an intermittent nature with periods of 'noise alone'. However, the signal can be manually switched off after convergence and a measurement of SNR improvement made as follows. The SNR at the input to the estimator is defined in dB as

$$\mathrm{SNR_{in}} = 10 \log_{10}\left(\frac{\sigma_y^2}{\sigma_\mathrm{v}^2}\right),$$

where both signal and noise variances are directly measurable in this type of simulation. At the output of the estimator, a measure $\hat{\sigma}_y^2 + \hat{\sigma}_\mathrm{v}^2$ of the variance of the estimated signal plus residual noise can be made. When the signal is manually switched off, only the component representing the residual noise of variance $\hat{\sigma}_\mathrm{v}^2$ remains. Hence some measure, say, P dB, can be made, where

$$P = 10 \log_{10}\left(\frac{\hat{\sigma}_y^2 + \hat{\sigma}_\mathrm{v}^2}{\hat{\sigma}_\mathrm{v}^2}\right),$$

from which the SNR at the output can be calculated as

$$\mathrm{SNR_{out}} = 10 \log_{10}(10^{P/10} - 1),$$

which was estimated as 25 dB for this case. A table of results for this example with various SNRs are shown in table 1.

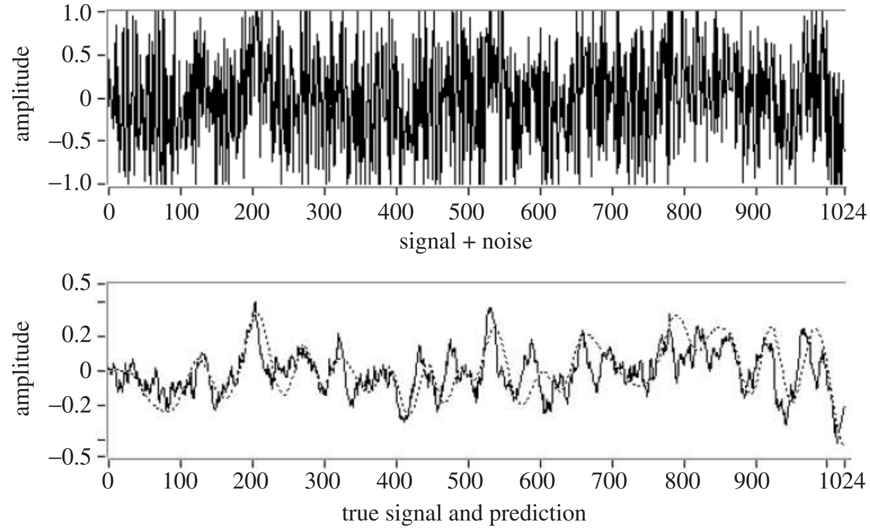

Figure 5. Performance of predictor, $d = 5$ steps. (True signal shown with broken line.)

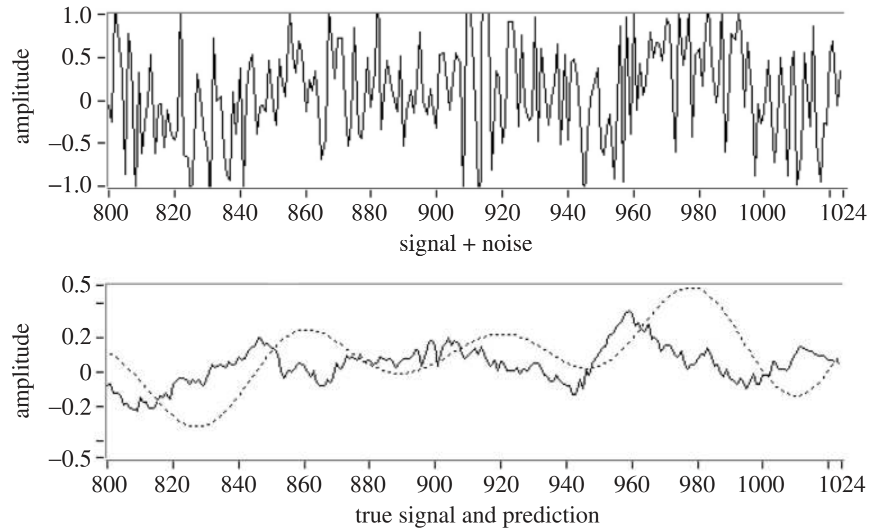

Figure 6. Performance of predictor after averaging, $d = 12$ steps. (True signal shown with broken line.)

For the predictor case, consider the previous signal-and-noise model, but for a five-steps-ahead predictor. The results are shown in figure 5. Although quite good, the results are not as good as those for the smoothing case, since there is less information in the impulse response (i.e. no 'uncausal' information is included).

To illustrate the predictor case further, consider the same problem as the previous example, but with a prediction of $d = 12$ steps. Figure 6 shows the signal plus noise, the true signal and its 12-steps-ahead predicted value. The results are shown a zoomed time-axis to highlight the performance of the predictor.

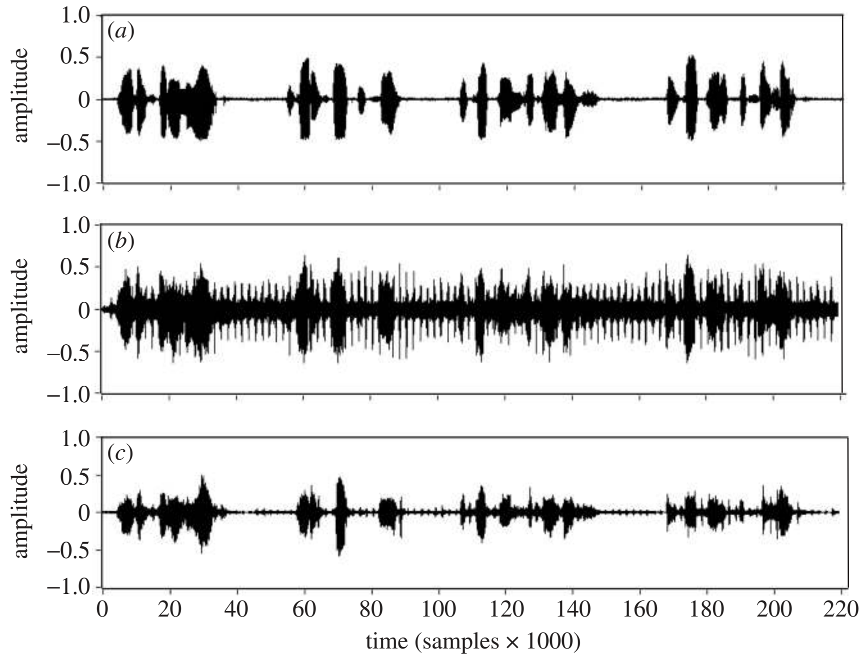

Figure 7. Performance of smoothing filter ($d = 10$ steps) on speech plus helicopter noise. (*a*) Original clean speech. (*b*) Speech plus helicopter noise. (*c*) Smoothed speech signal.

The averaging of the complex-cepstrum coefficients results in a form of convergence of the algorithm, but is not suitable for more realistic environments where the signal and noise are non-stationary. For such problems, the filter and kepstrum coefficients will be time varying for each frame and such an example is considered next with a speech signal and additive helicopter noise.

(*b*) *Speech signal plus helicopter noise*

The clean speech signal shown in figure 7*a* had helicopter noise added. The composite signal plus noise waveform is shown in figure 7*b*. A sampling frequency of 22 050 Hz was used, with an FFT length of 4096 points and 75% overlapping frames. The periodogram was averaged using the method of Appendix A. The SNR before processing was measured as an average −1 dB across the whole waveform. The smoothing algorithm was used to enhance the speech and the result for a lag of 10 steps is shown in figure 7*c*. The helicopter noise was still present, but very much audibly attenuated. The SNR after processing was measured as *ca.* 20 dB across the waveform. This was measured in a similar manner as before by measuring the variance of the signal between utterances (i.e. noise alone) and the variance in the signal plus noise. Knowing this information enables the segmented SNR to be calculated for various utterances and gives an average improvement in SNR of *ca.* 21 dB. However, the above example is somewhat artificial and contrived, as the noise is seldom available in a pure form. A better understanding of the performance of the algorithm is given in the next example.

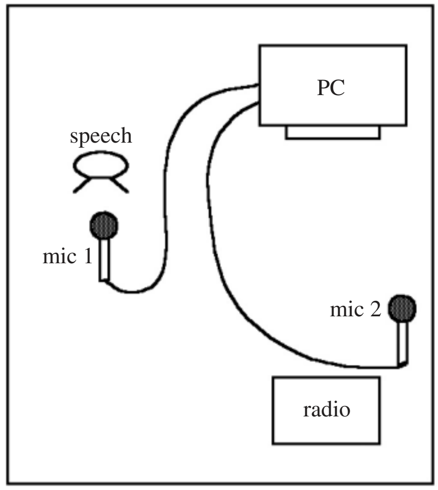

Figure 8. Recording of a speech signal and noise in a real environment.

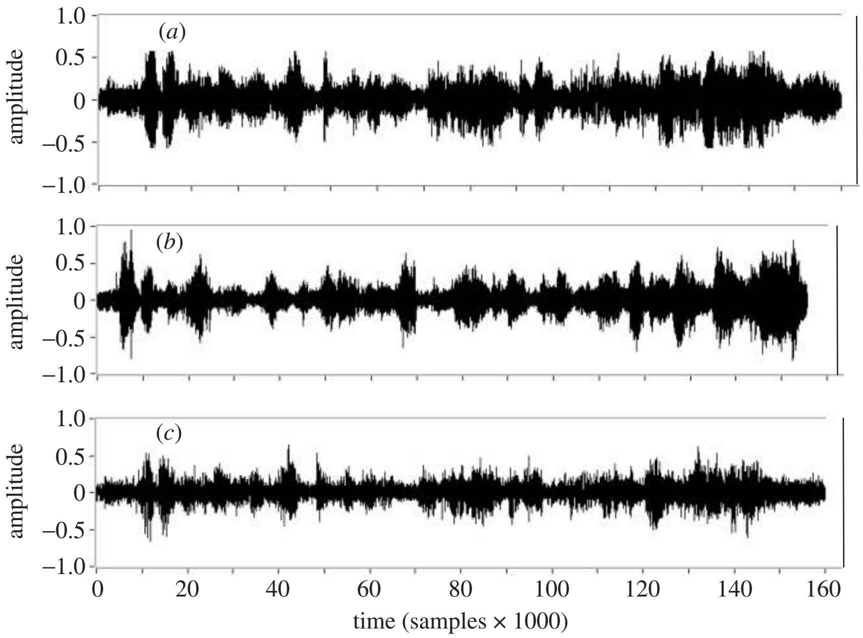

Figure 9. (*a*) Speech plus noise (message), (*b*) kepstrum estimate and (*c*) LMS estimate for real environment.

(*c*) *A noise-cancellation comparison*

Consider the classic noise-cancellation set-up shown in figure 8. Such a configuration has been well known since the early days of noise cancellation and the work of Widrow & Stearns (1985). The room was *ca.* 4 m × 4 m and a typical office environment. The noise source was a radio playing music. A recording was made onto disk and the kepstrum method was compared with that of LMS. The results are shown in figure 9.

A sampling frequency of 22 050 Hz was used, with an FFT length of 4096 samples. The delay was $d = 10$, with $\ell = 64$. The FFT window overlap was 75%. The LMS algorithm was employed using 64 weights so as to make some form of comparison. The average segmented SNR of the signal plus noise was found to be 3.87 dB. The segmented SNR of the kepstrum estimate was found to be 7.08 dB and that of the LMS algorithm 3.52 dB. This indicates an improvement in SNR by using LMSs of 0.35 dB, and 3.2 dB when using the kepstrum method. This result, of course, cannot be said to be definitive, as there are many factors to consider when performing such comparisons.

7. Conclusions

A kepstrum method of spectral factorization has been applied to the general Wiener estimation problem with coloured noise. The algorithms are relatively straight forward, computationally easy to implement and have been shown to give good results on two speech-enhancement examples. Further work needs to be done to compare in more detail these algorithms with existing noise-cancellation and prediction methods and to apply them to real-world problems.

Appendix A. Some comments about averaging

The periodogram is the estimate of spectral density at each frequency-bin m given as

$$\hat{S}_m = \frac{1}{N}|X_m|^2, \quad m = 0, 1, 2, \ldots, N-1, \tag{A 1}$$

but can also be found by the following exponential smoothing method at each adjacent or overlapping FFT frame $j = 1, 2, \ldots,$

$$S_m(j) = \beta S_m(j-1) + (1-\beta) X_m(j) X_m^*(j), \tag{A 2}$$

where $X_m(j)$ is the FFT frequency vector at frame j and $X_m^*(j)$ is its complex conjugate; $0 < \beta < 1$ is a forgetting factor. This method is used here to average the periodogram when the signal and noise are non-stationary. For the special case of stationary signals discussed in § 6, the complex-cepstrum coefficients are averaged across FFT frames according to the recursive mean

$$\bar{k}_i(j) = \bar{k}_i(j-1) + \frac{1}{j}[k_i(j) - \bar{k}_i(j-1)], \quad i = 0, 1, 2, \ldots, N-1, \tag{A 3}$$

where $\bar{k}_i(j)$ is the mean value of the ith coefficient evaluated at frame j. Using (A 3), the complex-cepstrum coefficients will converge to the (theoretical) kepstrum coefficients. Of course, equation (A 3) can only be used for the stationary case, as there is no form of forgetting factor included.

Nomenclature

ζ	transform variable z^{-1}
q^{-1}	backward shift operator
$E[\cdot]$	expectation operator

y_k, s_k	signal and message processes
v_k	coloured-noise process
ξ_k, μ_k	white-noise processes
ε_k	white innovations process
$g_{ys}(\zeta)$	cross-spectral density transform between signal and message
$g_{ss}(\zeta)$	spectral density transform of message process
$\Lambda(\zeta)$, $Z(\zeta)$	spectral factor and normalized spectral factor of signal plus noise
$\Gamma(\zeta)$	normalized noise-only spectral factor

References

Agaiby, H. & Moir, T. J. 1997 Knowing the wheat from the weeds in noisy speech. In *Proc. 5th Eur. Conf. Speech Communication and Technology, Rhodes, Greece, 22–25 September 1997*, vol. 3, pp. 1119–1122.

Allen, J. N., Berkley, D. A. & Blauert, J. 1977 Multi-microphone signal-processing technique to remove room reverberation from speech signals. *J. Acoust. Soc. Am.* **62**, 912–915.

Barrett, J. F. & Chen, X. P. 1983 Numerical spectral factorization using the method of Szegö-Kolmogorov. In *Proc. IASTED Symp. MECO'83, Athens, Greece*.

Barrett, J. F. & Moir, T. J. 1986 The kepstrum method for spectral analysis. *Int. J. Control* **43**, 29–57.

Barrett, J. F. & Moir, T. J. 1987 A unified approach to multivariable discrete-time filtering based on the Wiener theory. *Kybernetica* **23**, 177–196.

Dabis, H. S. & Moir, T. J. 1993 A unified approach to optimal estimation using diophantine equations. *Int. J. Control* **57**, 577–598.

Grimble, M. J. 1985 Polynomial systems approach to optimal linear filtering and prediction. *Int. J. Control* **41**, 1545–1564.

Grimble, M. J. 1988 Single versus double diophantine equation debate: comments on 'a polynomial approach to Wiener filtering'. *Int. J. Control* **48**, 2161–2165.

Hagander, P. & Wittenmark, B. 1977 A self-tuning filter for fixed-lag smoothing. *IEEE Trans. Inform. Theory* **23**, 377–384.

Haykin, S. 1986 *Adaptive filter theory*. Englewood Cliffs, NJ: Prentice Hall.

Kalman, R. E. 1960 A new approach to linear filtering and prediction problems. *J. Basic Engng Trans. ASME* D **82**, 35–45.

Kolmogorov, A. N. 1939 Sur l'interpolation et extrapolation des suites stationnaires. *C. R. Acad. Sci. Paris* **208**, 2043–2045.

Kolmogorov, A. N. 1941 Stationary sequences in Hilbert space. *Bull. Moscow Univ.* **2**, 1–40. (Transl. 1977 *Linear least squares estimation* (ed. T. Kailath), pp. 66–89. Stroudsburg, PA: Dowden, Hutchinson & Ross.)

Ljung, L. & Söderström, T. 1982 *Theory and practice of recursive estimation*. Cambridge, MA: MIT Press.

Martinez, R., Alvarez, A., Gomez, P., Nieto, V. & Rodellar, V. 2001 Combination of adaptive filtering and spectral subtraction for noise removal. In *Proc. 2001 IEEE Int. Symp. Circuits and Systems*, vol. 2, pp. 793–796.

Oppenheim, A. V. & Schafer, R. W. 1975 *Digital signal processing*. Englewood Cliffs, NJ: Prentice Hall.

Oppenheim, A. V., Schafer, R. W. & Stockham, T. G. 1968 Nonlinear filtering of multiplied and convolved signals. *Proc. IEEE* **56**, 1264–1290.

Popescu, D. C. & Zeljkovic, I. 1998 Kalman filtering of coloured noise for speech-enhancement acoustics. In *Proc. 1998 IEEE Int. Conf. Speech and Signal Processing*, vol. 2, pp. 997–1000. Piscataway, NJ: IEEE Press.

Roberts, A. P. & Newmann, M. N. 1988 Polynomial approach to Wiener filtering. *Int. J. Control* **47**, 681–696.

Silvia, M. T. & Robinson, E. A. 1978 Use of the kepstrum in signal analysis. *Geoexploration* **16**, 55–73.

Stahl, V., Fischer, A. & Bippus, R. 2000 Quantile-based noise estimation for spectral subtraction and Wiener filtering. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing 2000 (ICASSP 2000)*, vol. 3, pp. 1875–1878.

Wahba, G. 1980 Automatic smoothing of the log periodogram. *J. Am. Statist. Assoc.* **75**, 122–132.

Widrow, B. & Stearns, S. D. 1985 *Adaptive signal processing*. Englewood Cliffs, NJ: Prentice Hall.

Wiener, N. 1949 *Extrapolation, interpolation and smoothing of stationary time-series with engineering application*. Wiley.