ON FREQUENCY WEIGHTING IN AUTOREGRESSIVE SPECTRAL ESTIMATION

Anders Blomqvist

Department of Mathematics
Royal Institute of Technology
SE-100 44 Stockholm, Sweden
anders.blomqvist@math.kth.se

Bo Wahlberg

Department of Signal, Sensors and Systems
Royal Institute of Technology
SE-100 44 Stockholm, Sweden
bo.wahlberg@s3.kth.se

ABSTRACT

This paper treats the problem of approximating a complex stochastic process in a given frequency region by an estimated autoregressive (AR) model. Two frequency domain approaches are discussed: a weighted frequency domain maximum likelihood method and a prefiltered covariance extension method based on the theory of Lindquist and coworkers. It is shown that these two approaches are very closely related and can both be formulated as convex optimization problems. An examples illustrating the methods and the effect of prefiltering/weighting is provided. The results show that these methods are capable of tuning the AR model fit to a specified frequency region.

1. INTRODUCTION

Suppose that $\{y(t),\ t = \ldots -1, 0, 1, \ldots\}$ is a stationary linear regular random process generated by feeding a sequence of independent identically distributed random variables $\{e(t),\ t = \ldots -1, 0, 1, \ldots\}$ with zero means and with variances σ_0^2 through a shaping, or noise, filter $H_0(q) := 1 + \sum_{k=1}^{\infty} h_k^0 q^{-k}$ where q^{-1} is the delay operator. The power spectral density equals $\Phi_0(e^{i\omega}) = \sigma_0^2 |H_0(e^{i\omega})|^2$. For an autoregressive moving average (ARMA) process the noise filter equals

$$H_0(q) = \frac{C_0(q)}{A_0(q)} =: \frac{1 + c_1^0 q^{-1} + \ldots c_{n_c^0}^0 q^{-n_c^0}}{1 + a_1^0 q^{-1} + \ldots a_{n_a^0}^0 q^{-n_a^0}}.$$

The poles and zeros of $H_0(q)$ are restricted to be in the unit disc. The AR process corresponds to the special case $C_0(q) \equiv 1$.

Given observations $y(1) \ldots y(N)$ the task is to estimate the power spectral density or the noise filter and innovation variance. This problem is classical and many methods are available. The books [1, 2] provide very good introductions to this field.

Let $H(q, \theta)$ be a noise filter model set parameterized by the d-dimensional vector θ. For an ARMA model $\theta = [a_1 \ldots a_{n_a}, c_1 \ldots c_{n_c}]^T$. The corresponding prediction error equals $\epsilon(t, \theta) = [H(q, \theta)]^{-1} y(t)$, and the prediction error identification method (PEM) estimate $\hat{\theta}_{PEM}$ is obtained by minimizing the sum of squared prediction errors under the constraints that the poles and zeros of $H(q, \theta)$ should be inside the unit circle. For an AR model this simplifies to a linear least square problem.

If the true process is ARMA and in the model set and the $\{e(t)\}$ are Gaussian distributed, the PEM estimate then coincides with the conditional maximum likelihood (ML) estimate. Furthermore, in large samples the ML method attains the ultimate statistical performance, i.e., it is consistent and achieves the Cramér-Rao lower bound, and is thus asymptotically statistically efficient.

However, if the true process is more complex than the model less is known – in particular when the attention is given to models which are good in a certain frequency region. Another open issue is reliable online implementations of the PEM estimator which needs monitoring to guarantee feasible estimates and to avoid local minima etc.

The objective of this contribution is to study how to focus the estimation into a certain frequency range when the true process is more complex than the model!

2. WEIGHTING IN THE FREQUENCY DOMAIN

Weighting is often more easily done in the frequency than in the time domain. The ML method can also be formulated using frequency domain data, see e.g., [2, 3]. Let $Y_N(\omega) := \frac{1}{\sqrt{N}} \sum_{t=1}^{N} y(t) e^{-i\omega t}$ be the discrete Fourier transform of the sequence $\{y(t),\ t = 1 \ldots N\}$. By also including the noise variance σ^2 as a parameter, the frequency domain ML (FDML) cost function equals, see [2, p. 230],

$$V_{FDML}(\theta, \sigma^2) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|Y_N(\omega)|^2}{\sigma^2 |H(e^{i\omega}, \theta)|^2} d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(\sigma^2 |H(e^{i\omega}, \theta)|^2\right) d\omega. \quad (1)$$

This is the discrete time version of the classical Whittle likelihood, and can interpreted as the likelihood of the asymptotic distribution of the periodogram, see e.g., [4, p. 347]. Here we have used an integral formulation, instead of the more common summation over the frequencies $\omega_k = 2\pi k/N$, $k = 1\dots N$. The difference is negligible for large N. In [5] this frequency domain approach is further refined by also taking the initial state of the noise filter into account to obtain a leakage free spectral representation.

To obtain a frequency weighted estimate, FDML(f), we can assign a non-negative weight, $W(\omega) \geq 0$, to each frequency in (1):

$$V_{FDML(f)}(\theta,\sigma^2) := \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)\frac{|Y_N(\omega)|^2}{\sigma^2|H(e^{i\omega},\theta)|^2}d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)\log\left(\sigma^2|H(e^{i\omega},\theta)|^2\right)d\omega. \quad (2)$$

One possible frequency weighting is to give a constant weight to all frequencies in a desired region and zero weight to all others. In the discrete case the cost function then is of the form

$$\tilde{V}_{FDML(f)}(\theta,\sigma^2) = \frac{1}{M}\sum_{\omega_k\in\Omega}\frac{|Y_N(\omega_k)|^2}{\sigma^2|H(e^{i\omega_k},\theta)|^2} + \frac{1}{M}\sum_{\omega_k\in\Omega}\log\left(\sigma^2|H(e^{i\omega_k},\theta)|^2\right), \quad (3)$$

where Ω is a specified set of M important frequencies.

It is possible to perform analytic minimization of the cost-functions with respect to σ^2 as shown in [2, p. 230]. For $V_{FDML(f)}(\theta,\sigma^2)$ this leads to the estimates

$$\hat{\theta}_{FDML(f)} := \arg\min_{\theta}\left[\bar{W}\log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}W(\omega)\frac{|Y_N(\omega)|^2}{|H(e^{i\omega},\theta)|^2}d\omega\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi}W(\omega)\log|H(e^{i\omega},\theta)|^2 d\omega\right], \quad (4)$$

where $\bar{W} := \frac{1}{2\pi}\int_{-\pi}^{\pi}W(\omega)d\omega$. For any monic, stable and inversely stable transfer function $H(e^{i\omega},\theta)$ we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log|H(e^{i\omega},\theta)|^2 d\omega = 0, \quad (5)$$

and hence the second term in (4) disappears if $W(\omega) = 1$ (no weighting). This term is, however, most important for the weighted case to regularize the optimization problem, and can be viewed as a kind of barrier function in constrained optimization.

3. WEIGHTED COVARIANCE EXTENSION

Next we discuss a weighted covariance extension approach. Let $r(\tau)$ be the covariances of $y(t)$. The *rational covariance extension problem* amounts to, for a given covariance function, to find a noise filter with a matching covariance function up to a given lag, i.e., solve

$$r(\tau) - \frac{1}{2\pi}\int_{-\pi}^{\pi}e^{i\omega\tau}\Phi(e^{i\omega})d\omega = 0, \ \tau = 0\dots n$$
$$\Phi(z) = \sigma^2\frac{C(z)C(z^{-1})}{A(z)A(z^{-1})},$$

with respect to $[a_1\dots a_n]^T$ and σ^2. Since only $n+1$ covariances are exactly matched, the $C(z)$-polynomial can be taken as a free design variable (with roots inside the unit disc). This gives different feasible covariance extensions $r(\tau), \ \tau > n$ of the covariance sequence, hence the name. For example $C \equiv 1$, i.e., an AR model, gives the maximum entropy extension. See [6] for the state-of-art of the rational covariance extension problem.

The trick to solve this problem is to use the coefficients of $Q(z) = A(z)A(z^{-1})/\sigma^2 = \sum_{k=0}^{n} q_k(z^k + z^{-k})/2$ as parameters, and then solve

$$r(\tau) - \frac{1}{2\pi}\int_{-\pi}^{\pi}e^{i\omega\tau}\frac{|C(e^{i\omega})|^2}{Q(e^{i\omega})}d\omega = 0, \quad \tau = 0\dots n. \quad (6)$$

The set of feasible $\bar{\theta} := [q_0\dots q_n]^T$,

$$\mathcal{Q}_+ := \left\{\bar{\theta}\in\mathbb{R}^{n+1} : \ Q(e^{i\omega}) > 0, \ \omega\in[-\pi,\pi]\right\},$$

is easily seen to be a convex set. The left hand side of the equations (6) are just the derivatives with respect to q_τ, $\tau = 1\dots n$ of the functional

$$V_{CE}(\bar{\theta}) := [r(0)\dots r(n)]\bar{\theta} - \frac{1}{2\pi}\int_{-\pi}^{\pi}|C(e^{i\omega})|^2\log Q(e^{i\omega})d\omega.$$

The first term of $V_{CE}(\bar{\theta})$ equals the normalized variance of the prediction errors for an AR model. For the special case $C(z) \equiv 1$, the second term of $V_{CE}(\bar{\theta})$ equals $\log\sigma^2$. Hence we have then derived the asymptotic prediction error cost functional for an AR model, which after analytic optimization with respect to σ^2 simplifies to a quadratic least squares optimization problem (solving the normal equations). The second term is however most important for an ARMA model, $C(z) \neq 1$, and makes the optimization problem non-quadratic. In [6] it is, however, shown that $V_{CE}(\bar{\theta})$ is a convex functional and that the problem has a unique interior minimizer, and hence the ARMA covariance extension problem can be solved using convex optimization methods.

Next we generalize this approach to the weighted/prefiltered case. Let $y_f(t) := L_f(q)y(t)$, $r_f(\tau)$ be the filtered covariances, and

$$\hat{r}_f(\tau) := \frac{1}{N}\sum_{t=\tau+1}^{N}y_f(t)y_f(t-\tau).$$

The key idea is to apply the rational covariance extension method to the prefiltered covariances. This leads to the following matching problem, for $\tau = 0, \dots n$:

$$r_f(\tau) - \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega\tau}|L_f(e^{i\omega})|^2 \frac{|C(e^{i\omega})|^2}{Q(e^{i\omega})}d\omega = 0$$

which integrates to the cost function

$$\begin{aligned} V_{CE(f)}(\bar\theta) &:= [r_f(0)\dots r_f(n)]\bar\theta \\ &-\frac{1}{2\pi}\int_{-\pi}^{\pi} |C(e^{i\omega})|^2|L_f(e^{i\omega})|^2 \log Q(e^{i\omega})d\omega. \quad (7)\end{aligned}$$

This is also a convex function of $\bar\theta$ as can be shown using the arguments in [7]. The optimization problem then becomes

$$\min_{\bar\theta\in\mathcal{Q}_+} V_{CE(f)}(\bar\theta). \qquad (8)$$

In [7] it is shown to have a unique interior minimizer.

4. ON THE RELATION BETWEEN THE WEIGHTING IN THE FREQUENCY DOMAIN AND COVARIANCE EXTENSION

Here we will study the connection between the FDML(f) functional and the CE(f) functional. Under a particular choice of weighting function and prefilter, they will in fact be the same.

The cost function $V_{CE(f)}(\bar\theta)$ in (7) with

$$C = 1,\ H(q,\theta) = 1/A(q),\ 1/Q(q) = \sigma^2 H(q,\theta)H(1/q,\theta)$$

can be expressed as

$$\begin{aligned}
&\hat V_{CE(f)}(\bar\theta) \\
&= [\hat r_f(0)\dots\hat r_f(n)]\bar\theta - \frac{1}{2\pi}\int_{-\pi}^{\pi}|L_f(e^{i\omega})|^2\log Q(e^{i\omega})d\omega \\
&= \sum_{k=0}^{n} q_k \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{ki\omega}+e^{-ki\omega}}{2}\Big[\hat r_f(0) \\
&\quad + \sum_{l=1}^{n}\hat r_f(l)(e^{li\omega}+e^{-li\omega}) + S(e^{i\omega})\Big]d\omega \\
&\quad + \frac{1}{2\pi}\int_{-\pi}^{\pi}|L_f(e^{i\omega})|^2\log\left(\sigma^2|H(e^{i\omega},\theta)|^2\right)d\omega, \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}|L_f(e^{i\omega})|^2\frac{\hat\Phi(e^{i\omega})}{\sigma^2|H(e^{i\omega},\theta)|^2}d\omega \\
&\quad + \frac{1}{2\pi}\int_{-\pi}^{\pi}|L_f(e^{i\omega})|^2\log\left(\sigma^2|H(e^{i\omega},\theta)|^2\right)d\omega,
\end{aligned} \qquad (9)$$

with

$$\begin{aligned}\hat\Phi(\omega) \quad := \quad & \frac{1}{|L_f(e^{i\omega})|^2}\left(\hat r_f(0) + \hat r_f(1)(e^{i\omega}+e^{-i\omega}) + \right. \\ & \left. \cdots + \hat r_f(n)(e^{ni\omega}+e^{-ni\omega}) + S(e^{i\omega})\right),\end{aligned}$$

where $S(z)$ is *any* function of z^k for all $k > n$ and $k < -n$. Theorems 2.1 and 2.2 in [2] imply that $\hat\Phi(\omega) = |Y_N(\omega)|^2 + \mathcal{O}(1/\sqrt{N})$ if we choose $S(z) = \sum_{k=n+1}^{N}\hat r_f(k)(z^k + z^{-k})$. The functional (9) is then equal to the FDML(f) functional in (2) if we take $\hat\Phi(\omega) = |Y_N(\omega)|^2$ and $W(\omega) = |L_f(e^{i\omega})|^2$. An ideal filter

$$|L_f(e^{i\omega})| = \begin{cases} 1, & \omega_k \le \omega \le \omega_l, \quad \omega_k,\omega_l \in \Omega \\ 0, & \text{otherwise}\end{cases}$$

would give that (9) is close to (3). However, this choice is not feasible since we assume that $0 < |L_f(e^{i\omega})| < \infty$ for the covariance extension problem.

Inspired by this we can re-parameterize the FDML functional $\tilde V_{FDML}(\theta,\sigma^2)$ in terms of the q_k parameters. This also yields a convex problem. However, in contrast to the integral formulation of the covariance extension approach there is no guarantee that there exists an interior optimizer. This never happens in the integral formulation as shown in [6, 7]. However, for reasonable data this should not be a major problem in the finite approximation.

5. A SIMULATION EXAMPLE

Next we shall consider an example illustrating the proposed procedure while comparing it to the FDML(f) and PEM estimators.

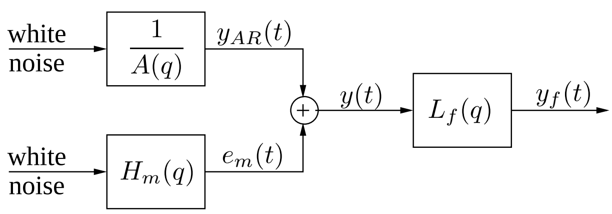

Fig. 1. The system setup

Example 1 (AR(2) in colored noise) This example amounts to identifying an AR(2) process in colored noise. The signal $y(t)$ is generated as in Figure 1. Let the roots of $A(q)$ be in $0.98e^{\pm 0.2\pi i}$ and let the driving noise be Gaussian white with unit variance. Let the added noise be color by driving Gaussian white with unit variance through a sixth order shaping filter $H_m(z)$. This noise has most of its power in the higher frequency region. The signal-to-noise ratio of the signal $y(t)$ will then be high around the peak located at the frequency 0.2π while it will be very small at high frequencies. Assuming that we possess this *a priori* knowledge of the process, we, for instance, use the low-pass prefilter

$$L_f(z) = \frac{(z-0.6e^{2i})^3(z-0.6e^{-2i})^3}{(z-0.6e^{0.8i})^3(z-0.6e^{-0.8i})^3}, \qquad (10)$$

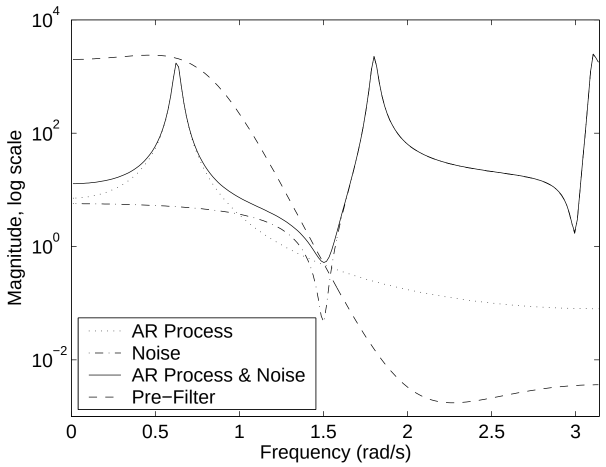

Fig. 2. The spectrum of the AR(2) process and the noise together with the spectral density of the prefilter.

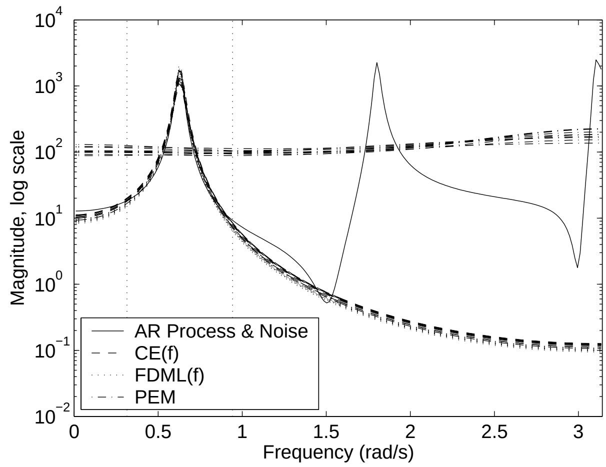

Fig. 3. Spectral densities for ten estimations of Example 1.

in the method. The spectral representation of $y_{AR}(t)$, $e_m(t)$, and the prefilter are plotted in Figure 2.

Now we apply the proposed CE(f) method. We compare our estimator to the PEM estimator of [2] and the FDML(f) estimator. In the latter we take

$$\Omega = \left\{ \omega_k = \frac{2\pi k}{M} : \omega_k \in [0.1\pi, 0.3\pi] \right\}.$$

In a sense this correspond to an ideal band-pass filter and thus differs significantly from $L_f(z)$ in (10). To avoid transient error we use the large sample size $N = 5000$.

In Figure 3 the frequency responses for ten different realizations of the noises $\{e(t)\}$ and $\{e_m(t)\}$ are given. We clearly see that the FDML(f) and CE(f) estimators are better for the lower frequencies than the PEM estimator. However, this is of course at the expense of a worse match for higher frequencies. That the FDML(f) and CE(f) estimators give approximately the same result despite the fact that the FDML(f) uses an ideal band-pass filter whereas CE(f) uses H_m indicates a low sensitivity with respect to the choice of prefilter/weight.

6. CONCLUSIONS

Models are always approximations of true data generating processes. The quality of a model depends heavily on its intended use. If the objective is prediction, the prediction error approach is optimal. If the interest is in spectral properties in certain frequency band the answer is more complicated. In this paper we have proposed a prefiltered covariance extension approach to introduce frequency weighting in AR estimation and shown that it is closely related to frequency weighted maximum likelihood estimation. An example indicates that this is a promising way to affect the model error distribution. This paper is an abbreviated version of [8].

7. REFERENCES

[1] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

[2] L. Ljung, *System Identification, Theory for the User*, Prentice Hall, 1999.

[3] R. Pintelon and J. Schoukens, *System Identification, A Frequency Domain Approach*, IEEE Press, 2001.

[4] P.J Brockwell and R.J. Davis, *Time Series: Theory and Methods*, Springer, New York, 1991.

[5] R. Pintelon and J. Schoukens, “Time series analysis in the frequency domain,” *IEEE Trans. Signal Processing*, vol. 47, no. 1, pp. 206–210, 1999.

[6] C. I. Byrnes, S. V. Gusev, and A. Lindquist, “From finite covariance windows to modeling filters: A convex optimization approach,” *SIAM Review*, vol. 43, no. 4, pp. 645–675, 2001.

[7] T. T. Georgiou and A. Lindquist, “Kullback-Leibler approximation of spectral density functions,” *IEEE Trans. Information Theory*, vol. 49, no. 11, pp. 2910–2917, November 2003.

[8] A. Blomqvist and B. Wahlberg, “On affecting the frequency domain error distribution in autoregressive spectral estimation using prefiltering,” Submitted for publication, 2004.