SIGNALS IN COLOURED NOISE: JOINT NON-PARAMETRIC ESTIMATION OF SIGNAL AND OF NOISE SPECTRUM

Victor Solo

School of Electrical Engineering
University of New South Wales
Sydney, NSW 2052, AUSTRALIA
email: vsolo@syscon.ee.unsw.edu.au

ABSTRACT

There is a considerable literature on non-parametric estimation of signals buried in coloured noise of known or finitely parameterised spectrum. And a considerable literature on nonparamtric estimation of the spectrum of an observed coloured noise. Here we consider the bivariate problem of jointly, non-parametrically estimating a signal in coloured noise as well as the coloured noise spectrum. It is straightforward enough to develop estimators for each infinite dimensional parameter but what is much less clear is how jointly choose the two required tuning parameters. We develop, apparently for the first time a criterion that achieves this.

1. INTRODUCTION

The problem of non-parametric estimation of a signal (or function) in noise is a central and so much studied ill-conditioned inverse problem e.g. [1],[2]. A crucial part of the solution process is choosing a tuning parameter such as a window-width or a Tikhonov penalty parameter [3]. Similarly much studied, is the problem of non-parametric estimation of a spectral density either directly e.g. by windowing the auto-covariance [4] or indirectly e.g. by auto-regressive modeling with an empirically chosen order [5]. It is the conjunction of the two that we are interested in here and we call this a bivariate ill-conditioned inverse problem .

It is known in fact in the statistics literature that non-parametric signal estimation in the presence of coloured noise requires a non-trivial adjustment be made to procedures for automatic selection of tuning parameters [6],[7],[8]. Indeed each of the references discusses methods for doing that which generally involve crude estimation of the correlation structure. But the problem we consider here namely joint non-parametric estimation of signal and noise spectral density is not discussed in these works and indeed there seems to have been no work at all on this topic. While it may be easy enough to think up estimators it is much less obvious how to do the joint tuning parameter selection. Our aim indeed is to show how to do this.

In the next section we describe the problem setup as well as some estimators. In section 3 the model selection criterion is developed. In section 4 are results of a simple simulation. Conclusions are given in section 5.

This work was partly supported by the Australian Research Council

2. SIGNAL IN COLOURED NOISE

Consider the problem of estimating an unknown function or signal $f(t)$ buried in coloured noise of spectrum $F(\omega)$ from data (scaled to lie on $[0,1]$)

$$\begin{aligned} y_i &= f(\frac{i}{n}) + \epsilon_i,\ i = 1, \cdots, n \\ cov(\epsilon_r, \epsilon_s) &= \gamma_{r-s} = \int_{-\pi}^{\pi} e^{j\omega(r-s)} F(\omega) \frac{d\omega}{2\pi} \end{aligned}$$

A standard estimator for $f(t)$ is the kernel (or two-sided filter) estimator

$$\hat{f}(t) = \frac{1}{nh} \Sigma_1^n K(\frac{t - \frac{i}{n}}{h}) y_i \tag{2.1}$$

Here $K(\cdot)$ is a hump shaped kernel e.g. a triangle or a Gaussian and h is a window-width or feature size which controls the bias and variance of the estimator. The properties of this kind of estimator are well understood [1] and automatic methods of selecting the window-width are well known [1], [3]. If we take Fourier transforms we find

$$\begin{aligned} \tilde{f}(\omega) &= \tilde{K}(\omega h) \tilde{y}(\omega) \\ \tilde{y}(\omega) &= \frac{1}{n} \Sigma_1^n y_i e^{-j\omega \frac{i}{n}} \end{aligned}$$

We see the estimator provides a low pass filtering of y_i and that $\frac{1}{h}$ is a bandwidth for the filter. (Unfortunately in the statistics literature h itself is usually called a bandwidth; we prefer the term window-width to avoid confusion).

Similarly in the absence of the signal then a common method to estimate $F(\omega)$ is to fit an auto-regressive model and calculate the corresponding spectrum. Automatic methods of selecting the order p are again well studied [5],[9].

Now consider simultaneous estimation of $f(t), F(\omega)$. We can still use the kernel estimator (2.1) and it is even known how to adjust the automatic selector for the presence of autocorrelation [6]. The adjustment depends on empirical estimation of the autocorrelation function as well as another window-width for which however no empirical selection method has been given. Further the autocorrelation function estimator is not shown to lead to a spectrum estimate. Our aim here is to directly deal with both estimation of $f(t), F(\omega)$ as well as automatic selection of the two needed tuning parameters; h and p.

3. TUNING PARAMETER SELECTION WITH NURE

Our development follows earlier work of the author [10],[11] which while independently developed can be seen as an extension of the methods of [9]. We denote the parameters by $\theta = (\beta, \alpha)$; these consist of e.g. β: Fourier coefficients in the Fourier series expansion of $f(t)$ on $[0, 1]$; α: auto-regressive parameters.

To develop a tuning paramter selection criterion we need two items: (i) a data fidelity criterion $J_\theta(y)$ such as a log likelihood, used for parameter estimation. (ii) A joint measure of the quality of the estimators of $f(t), F(\omega)$. This is a discrepancy or 'risk' measure, call it $D(\theta)$ - below we will use Kullback-Liebler information. We also need an unbiassed estimator of $D(\theta)$ say $D_\theta(y)$; usuallly fairly easily constructed.

If we could calculate $D(\theta)$ we could choose h, p jointly to minimise it. Since we cannot in general calculate $D(\theta)$ the idea is to construct a nearly unbiassed estimator of it (NURE) and minimise that instead. The natural candidate is $D_{\hat\theta}(y)$ but it is biassed. However in [10] it is shown how to develop a bias correction and this leads to the selection criterion

$$\begin{aligned} NURE &= D_{\hat\theta}(y) + trace(J_2^{-1}W) \\ J_2 &= E(\frac{d^2 J_\theta(y)}{d\theta d\theta^T})|_{\theta=\theta_e} \\ W &= E(\frac{dJ_\theta(y)}{d\theta}\frac{dD_\theta(y)}{d\theta^T})|_{\theta=\theta_e} \\ \theta_e &= arg.min.E(J_\theta(y)) \end{aligned}$$

In practice θ_e is replaced by $\hat\theta$.

This NURE is very general and extends the approach in [9] where it is implicitly assumed $D(\theta) = E(J_\theta(y))$. Note that the criterion does not assume the data generating process comes from a model in the model class being fitted as e.g. AIC [9] does. To implement the criterion we have to specify then $J_\theta(y), D(\theta)$. It is far from straightforward to construct a discrepancy measure that incorporates both signal and noise components and the most easily accessible candidate is the Kullback-Liebler information KL(θ).

When the noise is Gaussian it is shown in [11] and is anyway easy to check that an unbiassed estimator of $KL(\theta)$ is the frequency domain likelihood type expression [5]

$$KL_\theta(y) = \frac{1}{2}\Sigma lnF_k(\alpha) + \frac{1}{2}\Sigma\frac{I_k(\beta)}{F_k(\alpha)}$$

where $I_k(\beta) = |\frac{\tilde y_k}{n} - \tilde\beta_k|^2$ and where e.g. $\tilde y_k$ is the DFT of y_i and $F_k(\alpha) \equiv F(\frac{2\pi k}{n})$ is the noise spectrum.

We use the following data fidelity criterion which enables simple construction of estimators. The log-likehood is of course a natural alternative but is computationally more demanding and will be pursued elsewhere.

$$\begin{aligned} J_\theta(y) &= S_\beta(y) + N_\alpha(y) \\ S_\beta(y) &= \frac{1}{2}\Sigma|\frac{\tilde y_k}{n} - \tilde\beta_k|^2 + \frac{1}{2}\Sigma h^2(2\pi k)^2|\tilde\beta_k|^2 \\ N_\alpha(y) &= -\frac{1}{2}\Sigma lnF_k(\alpha) - \frac{1}{2}\Sigma\frac{I_k}{F_k(\alpha)} \\ I_k &= |\frac{\tilde y_k}{n} - \tilde K(\frac{2\pi k}{n}h)\frac{\tilde y_k}{n}|^2, \tilde K(\omega) = \frac{1}{1+\omega^2} \end{aligned}$$

Optimising the signal criterion $S_\beta(y)$ leads easily to a smoothing spline [12] which is also a kernel estimator

$$\tilde\beta_k = \frac{1}{1+(hk)^2}\frac{\tilde y_k}{n} \Rightarrow \tilde K(\omega h) = \frac{1}{1+(\omega h)^2}$$

Carrying through the computations indicated above for NURE and assuming the data generating process belongs to the model class we find

$$NURE = nln\hat\sigma^2 + \frac{2}{h} + 2p$$

where $\hat\sigma^2$ is the residual mean squred error. This is a very unusual criterion since it involves both a window-width and a model order. If there is no function to be estimated the second term is missing and it reduces to AIC for selecting the order of an auto-regression. If the noise is white the third term is missing and it is very close to cross-validation methods for choosing h [1].

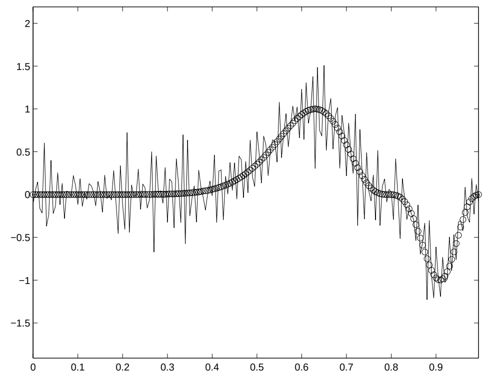

Figure 1: Function in Coloured Noise

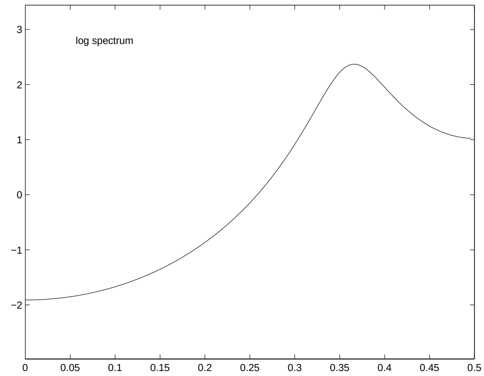

Figure 2: Noise Spectrum

4. RESULTS

In Fig.1 we show a simulation of $n = 200$ data values of the function $f(t) = sin^3(2\pi x^3)$ in coloured AR(2) noise

$$\epsilon_i = \nu_i - \phi_1 \epsilon_{i-1} - \phi_2 \epsilon_{i-2}$$

where ν_i is a zero mean white noise of variance σ_ν^2. In the simulation $\phi_1, \phi_2, \sigma_\nu^2 = -1, -.6, .025$. The noise spectrum is shown in Fig.2.

In Fig.3 we show the NURE surface (plotted against h and $-p$ i.e. in reverse order on the p axis so that its shape can be easily seen). The minimum is in the vicinity of $p = 3, h = .04$. The plot in Fig.4. enables this to be pinpointed. The estimated function and spectrum according to these values are shown in Fig 5 and Fig.6.

The signal reconstruction is a little noisy in the flat region but otherwise reasonably good. Clearly the spectrum estimate is biased at low frequency but the location of the peak is well obtained. The bias is not surprising since there is lack of identifiability at low frequency between signal and noise. We have made no attempt to deal with this. But some low pass filtering before spectrum estimation would ameliorate it. This issue needs further investigation, perhaps along the lines of [13].

5. CONCLUSIONS

In this paper we have developed, apparently for the first time a criterion that allows selection of two tuning parameters in a bivariate ill-conditioned inverse problem . The approach is quite general and can handle e.g. computed tomography reconstruction in coloured noise as well as choice of any number of tuning parameters. The criterion is unusual in the application discussed here in that the penalty term involves both

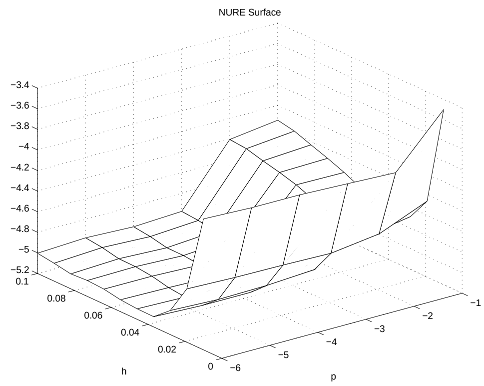

Figure 3: NURE Surface

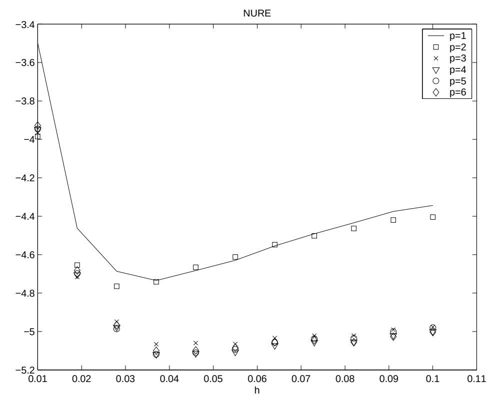

Figure 4: NURE as a function of h for each p

a model dimension and a window-width. Compared to other potential approaches such as Bayesian methods the computational load here is trivial. Problems for future work will involve application to other reconstruction problems as well as development of statistical properties of the estimators.

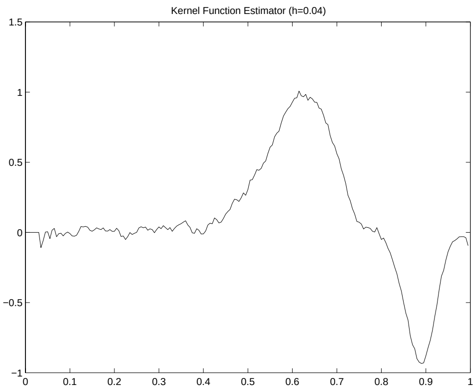

Figure 5: Estimated Signal ($h = .3$)

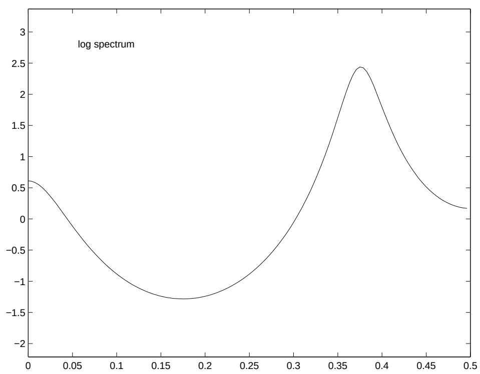

Figure 6: Estimated Noise Spectrum ($p = 3$)

REFERENCES

[1] M.P. Wand and M.C. Jones, *Kernel Smoothing*, Chapman and Hall, London, 1995.

[2] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, New York, 1989.

[3] J. Fan and I Gijbels, *Local Polynomial Modelling and its applications*, Chapman and Hall, London, 1996.

[4] G. Jenkins and D. Watts, *Spectral analysis and its applications*, Holden-Day, San Francisco, 1969.

[5] E.J. Hannan and M. Deistler, *Statistical theory of Linear Systems*, J. Wiley, New York, 1988.

[6] N.S. Altman, “Kernel smoothing of data with correlated errors,” *Jl. Amer. Stat. Assoc*, vol. 85, pp. 749–759, 1990.

[7] I.M. Johnstone and B. Silverman, “Wavelet threshold estimators for data with correlated noise,” *Jl. Royal. Stat. Soc. Ser. B*, vol. 59, pp. 319–352, 1997.

[8] V. Solo, “Total variation denoising in coloured noise,” in *Proc ICASSP2000*, Istanbul, Turkey, 2000, IEEE.

[9] H. Linhart and W. Zucchini, *Model selection*, J. Wiley, New York, 1986.

[10] V. Solo, “Transfer function estimation with a H_∞ criterion,” in *Proc 37th IEEE CDC*, Tampa, FL, 1998, IEEE.

[11] V Solo, P Purdon, R Weisskoff, and E Brown, “A signal estimation approach to functional mri,” *IEEE Trans Med Imaging*, vol. 20, pp. 26–35, 2001.

[12] G. Wahba, *Spline models for observational data. CBMS-NSF, Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1990.

[13] S.T. Chiu, “Bandwidth selection for kernel estimate with correlated noise,” *Stat.Prob.Lett.*, vol. 8, pp. 347–354, 1989.