A NON-PARAMETERIC ML ESTIMATOR WITH UNKNOWN CHANNEL ORDER

Yu Gong[⋆], *Vimal Bhatia*[+], *Bernard Mulgrew*[+], *and Colin F. N. Cowan*[⋆]

[⋆]Institute of Electronics, Communications and Information Technology, Queen's University of Belfast
Northern Ireland Science Park, Queen's Road,
Belfast BT3 9DT, United Kingdom
email: y.gong, c.f.n.cowan@ee.qub.ac.uk

[+]IDCOM, School of Engineering and Electronics,
The University of Edinburgh
Kings Buildings, Mayfield Road
Edinburgh, EH9 3JL, United Kingdom
email: v.bhatia, b.mulgrew@ee.ed.ac.uk

ABSTRACT

A recently proposed non-parameteric maximum likelihood (NPML) channel estimator shows superior performance to the least square (LS) estimator in presence of non-Gaussian noise. The derivation of the NPML estimator assumed perfect knowledge of the channel order, which, however, does not comply with most applications. In this paper, first we study the effects of the inaccurate order assumption on the NPML estimator, and then show that the traditional order selection criteria like the AIC are unreliable to apply for the NPML estimator. Finally we propose a simple method to trace the channel order where the order selection and channel estimation are carried out simultaneously.

1. INTRODUCTION

A typical block diagram of a channel estimator is shown in Fig. 1, where $x(n)$ is the channel input signal, $h(n)$ is the unknown channel, (n) is the channel noise, $\hat{h}(n)$ is the channel estimator and $e(n)$ is the error signal. In a recent paper [1], Bhatia (et al.) proposed a non-parameteric maximum likelihood (NPML) channel estimator with significant superior performance to the least square (LS) estimator in presence of non-Gaussian noise. In deriving the NPML estimator, however, the channel order was assumed known which is generally not the case for most applications. It is noted that in this paper we take the presence of co-channel interference in Gaussian noise as combined non-Gaussian noise [1].

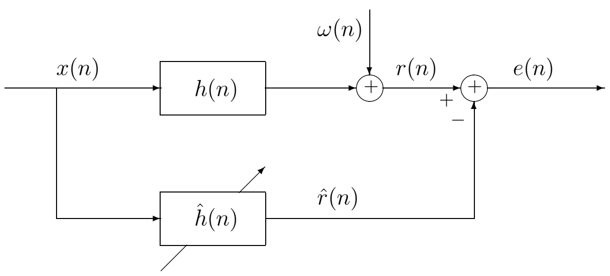


Figure 1: Block diagram of the channel estimator.

Although an old topic, the order estimation remains an incompletely solved problem [2]. The most widely used order estimation criterion are the Akaike's Information Criterion (AIC) [3], the Final Prediction Error (FPE) [3] and the Minimum Description Length (MDL) [4], all of which, however, are unreliable to apply when the noise is not Gaussian as will be assumed in this paper. All other order estimation algorithm are evaluated against the above three popular criteria. Usually the criterion indices for the possible orders are calculated, before making the final order selection. This *brute-force* approach demands high computation, impeding its application to on-line systems. Although some approaches (e.g. [5]) can carry out the order selection and channel estimation simultaneously, they are limit to specific applications and hard to be applied to the NPML estimator.

For a channel estimator, the order "under-estimate" is more serious a problem than the order "over-estimate" in terms of performance. Thus in practice, it is usually not necessary, if not impossible, to have a precise order estimate as long as the order is not underestimated, thereby making it possible to use simpler methods to estimate the channel order. Recently, Gong (et al.) proposed a novel variable tap-length adaptive algorithm which can be used to track the channel order on-line [6]. However, based on the symbol-based adaptive algorithm such as the LMS algorithm, the proposed algorithm cannot be used for the NPML estimator which is block-based.

In this paper, we will first investigate the influences of the inaccurate order assumption on the NPML channel estimator. Then, after showing that the classic AIC criterion is unreliable to apply in presence of non-Gaussian noise, we will propose a simple method to search for the channel order where the order selection and channel estimation can be carried out simultaneously. Simulation results are presented at the end.

2. NON-PARAMETERIC ML CHANNEL ESTIMATOR

According to Fig. 1, and assuming N as the total number of samples and P_o as the true channel order, the channel output vector can be expressed as:

$$\mathbf{r} = \mathbf{X}_{P_o}\mathbf{h}_{P_o} + \mathbf{w}, \tag{1}$$

where $\mathbf{h}_{P_o}$ is the channel vector, $\mathbf{w}$ is the noise vector, and $\mathbf{X}_{P_o}$ is the channel input matrix which is given by:

$$\mathbf{X}_{P_o} = \begin{bmatrix} x(1) & 0 & 0\cdots & 0 \\ x(2) & x(1) & 0\cdots & 0 \\ \vdots & \vdots & & \vdots \\ x(N) & x(N-1) & \cdots & x(N-P_o+1) \end{bmatrix} \tag{2}$$

This research was sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/S00217/01.

p	3	4	5	6	7	8	9	10	11
NMSE	0.0961	0.0215	**0.0018**	0.0023	0.0029	0.0034	0.0038	0.0047	0.0055

Table 1: NMSE for different assumed channel order.

The ML estimator maximizes the log-likelihood function

$$\mathscr{L}(\mathbf{r}|\hat{\mathbf{h}}_p) = \log f(\mathbf{r}|\hat{\mathbf{h}}_p) = \quad_{n=1}^{N} \log f \quad (e(n)) \qquad (3)$$

with respect to the channel estimator vector $\hat{\mathbf{h}}_p$, where the assumed channel order is p and $f \ (.)$ is the scalar probability density function (pdf) of the channel noise $\ (n)$.

As has been shown in [1], the ML estimator can be obtained by the gradient ascent search as:

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \quad \frac{\mathscr{L}(\mathbf{r}|\mathbf{h})}{\mathbf{h}}\Big|_{\mathbf{h}=\hat{\mathbf{h}}(k)}. \qquad (4)$$

Since, in our system model, it is assumed that the noise distribution is "unknown" (in (3)), a kernel density estimator is used to estimate this density as,

$$\hat{f} \ (e) = \frac{1}{N} \quad_{n=1}^{N} \quad (e - e(n)), \qquad (5)$$

where $\ (.)$ is the Gaussian kernel [1]. Then from (3) and (5), and with some manipulations, we have:

$$\frac{\mathscr{L}(\mathbf{r}|\mathbf{h})}{\mathbf{h}}\Big|_{\mathbf{h}=\hat{\mathbf{h}}(k)} = -\frac{1}{N} \quad_{n=1}^{N} \frac{\ _{i=1}^{N}(e(n)-e(i))(\mathbf{x}(n)-\mathbf{x}(i)) \ (e(n)-e(i))}{\ _{j=1}^{N} \ (e(n)-e(j))}. \qquad (6)$$

Finally substituting (6) into (4) gives the NPML estimator.

3. CHANNEL ORDER MIS-ESTIMATION

In general, if the channel order is assumed inaccurately, the estimation error comes from two parts: the coefficient-estimation error in the assumed model space and the space-estimation error between the true model and assumed model spaces [7]. As the assumed order increases, the coefficient-estimation error always increases, while the space-estimation error decreases until the assumed order is equal to, or large than, the true channel order.

To be specific, if the channel order is under estimated, i.e. $p < P_o$, only the first p coefficients of the channel can be effectively estimated, and the received signal can be expressed as:

$$r(n) = \quad_{i=1}^{p} h(i)x(n-i) + \ '(n), \qquad n = 1, \cdots, N, \qquad (7)$$

where $\ '(n) = h(p+1)x(n-p-1) + \cdots + h(P_o)x(n-P_o) + \ (k)$. Then the problem reduces to estimating the first p channel coefficients with the *equivalent channel noise* of $\ '(n)$. Hence when $p < P_o$, beside that there are $P_o - p$ taps "missing", even the estimation errors corresponding to the first p coefficients is larger than those when $p = P_o$ since $\ ^2_{\ '(n)} > \ ^2_{\ (n)}$. Therefore the order under-estimate results in significantly performance loss.

It is interesting to observe that $\ '(n)$ basically forms a Gaussian mixture. Thus under rare circumstance, can $\ '(n)$ be Gaussian. Further noting that NPML estimator demonstrates significantly superior performance to the LS estimator [1] in presence of non-Gaussian noise, we conclude that the NPML estimator is always better than, or more robust to, the LS estimator when the channel order is under-estimated.

On another front, if the channel order is over-estimated (i.e. $p > P_o$), the "space-estimation error" disappears and only the "coefficient-estimation error" remains. Then the estimator vector can be expressed as:

$$\hat{\mathbf{h}} = [\mathbf{h}_{P_o}^{\mathrm{T}} \ \mathbf{0}^{\mathrm{T}}]^{\mathrm{T}} + \quad_{\hat{\mathbf{h}}}, \qquad (8)$$

where $\ _{\hat{\mathbf{h}}}$ can be regarded as a perturbation to the ideal estimate. In general, the larger the N is, the smaller the perturbation term is. Particularly, it can be easily verified that, if $x(n)$ and $w(n)$ are independent to each other and either of them has zero mean, we have

$$\lim_{N\to} \quad_{\hat{\mathbf{h}}_{\mathrm{LS}}} = \mathbf{0}. \qquad (9)$$

Thus if the data number is large enough, the last $p - P_o$ coefficients of the estimator are very small.

As an example, we consider a system with presence of co-channel interference (CCI) where SNR=20dB, $\mathbf{h}_{P_o} = [1 \ \ 0.8 \ \ 0.6 \ \ 0.4 \ \ 0.2]^{\mathrm{T}}$ with $P_o = 5$, the interfering channel has signal-to-interference-ratio (SIR) of 10dB, and the total number of samples is 100. Table 1 shows the normalized mean-square-error (NMSE) of the NPML estimator when the assumed channel order varies from 3 to 11 respectively. The NMSE is a performance index to measure the "goodness" of an estimator and is defined as

$$\mathrm{NMSE} = \frac{\mathrm{E}[\ \ _{n=1}(h(n)-\hat{h}(n))^2]}{\mathbf{h}_{P_o}^{\mathrm{T}}\mathbf{h}_{P_o}}. \qquad (10)$$

It is clearly shown in Table 1 that the NMSE reaches the minimum at $p = P_o$. But when $p \geqslant P_o$, the NMSE are within a narrow range, all significantly below those for $p < P_o$. This indicates that the order under-estimation is more serious a problem than the order over-estimate in terms of performance, though the latter imposes more complexity.

4. NPML ESTIMATOR WITH ORDER ESTIMATION

In this section, first we will show that the traditional AIC is unreliable to apply for case of estimated and non-Gaussian pdf, and then propose a simple method to estimate the channel order.

p	SNR=20dB, N=100 No CCI		SNR=20dB, N=50 No CCI		SNR=40dB, N=50 SIR = 10dB	
	AIC$_{\hat{\mathscr{L}}}$	AIC$_{\hat{}^2}$	AIC$_{\hat{\mathscr{L}}}$	AIC$_{\hat{}^2}$	AIC$_{\hat{\mathscr{L}}}$	AIC$_{\hat{}^2}$
3	-85.39	-144.77	-35.50	-64.05	-33.59	-57.75
4	-179.63	-286.81	-85.21	-64.04	-54.92	-100.15
5	-464.65	**-455.82**	-193.96	**-211.84**	-57.72	-106.89
6	-465.80	-454.01	-192.40	-210.10	-55.93	-105.25
7	**-470.73**	-453.39	-201.67	-208.40	-54.21	-103.78
8	-468.73	-451.38	-199.69	-206.71	-55.64	-107.21
9	-467.33	-449.36	-198.09	-205.06	-53.89	-105.70
10	-466.22	-449.22	**-205.55**	-207.00	**-57.97**	**-111.96**
11	-468.58	-451.17	-203.81	-205.54	-56.78	-110.22

Table 2: AIC for different scenarios

4.1 Order estimation based on AIC

AIC is the most widely used order selection criterion which is defined as [3]:

$$\text{AIC}_{\hat{\mathscr{L}}} = -2\mathscr{L}(\mathbf{r}|\hat{\mathbf{h}}_p) + 2p, \quad (11)$$

where $\mathscr{L}(\mathbf{r}|\hat{\mathbf{h}}_p)$ is defined in (3). When the noise is Gaussian, (11) can be simplified to:

$$\text{AIC}_{\hat{}^2} = N \log \hat{}^2 + 2p, \quad (12)$$

where $\hat{}^2 = (1/N) \;\; {}_{n=1}^{N} e^2(n)$.

Unfortunately, neither AIC$_{\hat{\mathscr{L}}}$ nor AIC$_{\hat{}^2}$ is reliable to estimate the channel order for the NPML estimator: first, although the kernel density estimation (5) can be used to estimate the likelihood, it is not accurate enough to calculate the AIC$_{\hat{\mathscr{L}}}$; second, AIC$_{\hat{}^2}$ is only limited to Gaussian cases.

For illustration, we calculate the AIC for the same channel as that used in the previous section, and show the results in Table 2, where the minimum values are highlighted in bold. Recall the true channel order P_o is 5. In the first case, we have a pure Gaussian channel, where SNR=20dB, the number of sample $N = 100$ and no CCI. It is clear that AIC$_{\hat{}^2}$ has its minimum at $p = 5$ but AIC$_{\hat{\mathscr{L}}}$ at $p = 7$ which is biased away from P_o. In the second case, we have the same channel but N is decreased to 50. We observe that AIC$_{\hat{}^2}$ still finds the true order, but AIC$_{\hat{\mathscr{L}}}$ has the minimum which is further away from P_o. This is not surprising because, as N decreases, the density estimation becomes poorer and so does estimated AIC$_{\hat{\mathscr{L}}}$. In the last case, the interfering channel is introduced where SIR=10dB, the SNR is increased to 40dB, by which the channel becomes totally different from Gaussian. Under such scenario, neither AIC$_{\hat{\mathscr{L}}}$ nor AIC$_{\hat{}^2}$ estimates the channel order well. In conclusion, the AIC is unreliable for use with NPML estimator for order selection.

4.2 A simple order estimation method for the NPML estimator

Below we describe a simple method to estimate the channel order. The idea is based on the previous observation that, when the channel order is over-estimated, the extra taps are usually small compared to the others.

To be specific, at every iteration of the NPML estimation, the summation of squares for the last M coefficients of the estimator is measured. If it is smaller than M times of a pre-set threshold , then the order is decreased by 1; otherwise, if the summation of squares for the last $M-1$ taps is larger than $(M-1)$, the order is increased by 1, or the order remains unchanged. In summary, we have the following procedure combining the order selection and the NPML channel estimation together:

For every iteration , $k = 1, 2, 3, \cdots$
  Do the kernel density estimation based on (5).
  Update the estimator according to (4).
  if ${}_{i=p(k)-M+1}^{p(k)} \hat{\mathbf{h}}^2_{p(k)} < M \cdot$
    $p(k+1) = p(k) - 1$
  else if ${}_{i=p(k)-M+2}^{p(k)} \hat{\mathbf{h}}^2_{p(k)} < (M-1) \cdot$
    $p(k+1) = p(k)$
    else
    $p(k+1) = p(k) + 1$
    end
  end

In the above procedure, $p(k)$ is the tap-length at kth iteration and M is an integer no less than 1. M has two effects: first, to create a "guard margin" so that the estimation is based on M, rather than 1, coefficient values; second, to make the search escape from the local minima which are the zero coefficients within the range of the channel spread. Then if the threshold value is properly chosen, $p(k)$ will converge to within the range of $[P_o, P_o + M - 1]$. Obviously this method tends to over-estimate the order.

The threshold depends on both the channel specifics and the channel estimator. When the sample number N is large enough, the extra taps are normally very small, leaving us big room to choose . When N is small, the NPML estimator significantly outperforms the LS estimator as the former can explore the "local statistics" much better than LS. In fact, with a fixed $= 0.01$, we have tried extensive simulations under different scenarios such as different channel, SNR, N and etc. The results show that the proposed method always works well as long as N is reasonable large (e.g. $N \geqslant 30$).

Alternatively, we may use dynamic threshold, i.e. varies at each iteration. It has been shown in Section 3 that the channel estimation consists of the true channel plus a perturbation term. It is obviously that, the larger the N is, or the smaller the $\hat{}^2$ is, the smaller the perturbation is and then the smaller the we should have. Inspired by this observation, we can have a dynamic threshold as:

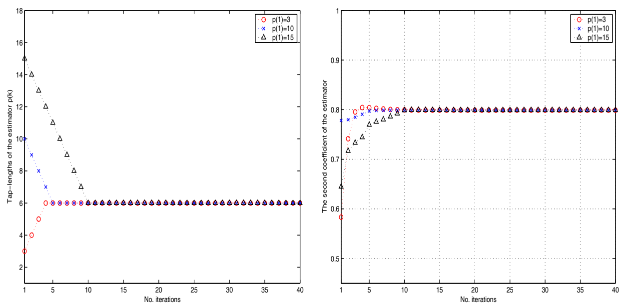
$$(k) = \frac{C\cdot\hat{\ }^{2}(k)}{N}, \qquad (13)$$

where C is a constant. To make the algorithm robust, we ensure that $_{min} < (k) < _{max}$, where $_{max}$ and $_{min}$ are maximum and minimum values for the threshold.

5. NUMERICAL SIMULATIONS

For the simulations in this section, the channel is the same as that for the previous examples in this paper, $M = 3$, the dynamic threshold based on (13) is used where $C = 10$, $_{max} = 0.05$ and $_{min} = 0.005$. All results are based on one typical run. The learning curves of the tap-length and the second tap coefficient of the estimator are shown in (a) and (b) respectively for each figure.

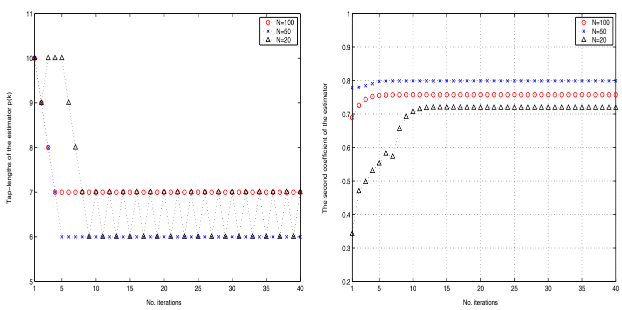
Fig. 2 investigates the proposed algorithm for different initialization of the estimator's tap-length, where SNR= 20dB, SIR= 10dB and $N = 100$. It is clear that, for all initializations, the individual tap-lengths converge to '6' which is in the range of $[P_o, P_o + M - 1]$ as expected.



(a) Tap-length learning curve. (b) 2^{nd} tap learning curve.

Figure 2: Learning curves for different tap-length initialization.

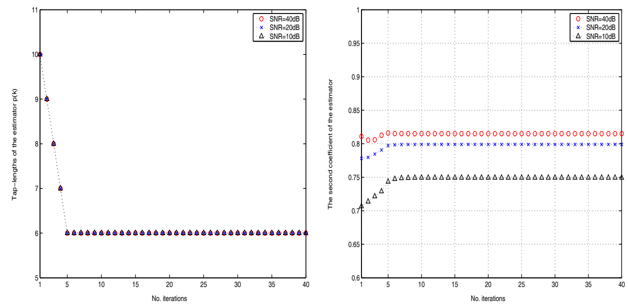
Fig. 3 compares the results for different sample number N, where SNR=20dB and SIR=10dB. We observe that even if N is as low as '20', the algorithm can still track the order, although it oscillates between '6' and '7' as shown in Fig. 3 (a). Accordingly, we observe slower coefficient convergence for $N = 20$ in Fig. 3 (b).



(a) Tap-length learning curve. (b) 2^{nd} tap learning curve.

Figure 3: Learning curves for different number of samples.

Fig. 4 shows the results for different SNR-s, where the SIR= 10dB and $N = 50$. It is obvious that the proposed algorithm works well for all these SNR-s. From Fig. 4 (b), it is interesting to note that NPML algorithm performs better for SNR=20dB than for 40dB, as the former converges closer to the true 2^{nd} coefficient (which is 0.8) of the channel. This is because that, in presence of CCI, the channel with SNR at 40dB is further "away" from Gaussian than that with SNR at 20dB, resulting in less accuracy for the kernel density estimation.



(a) Tap-length learning curve. (b) 2^{nd} tap learning curve.

Figure 4: Learning curves for different SNR.

6. CONCLUSION

In this paper, we propose a simple method to estimate the channel order and the channel coefficients for an unknown additive noise. It was shown that traditional AIC based techniques fail to estimate the channel order in non-Gaussian senarios for NPML estimator. The simulation results confirm that a better order estimate can be obtained by using proposed algorithm even in the presence of severe CCI.

REFERENCES

[1] V. Bhatia and B. Mulgrew, "A EM-kernel density method for channel estimtion in non-gaussian noise," in *IEEE 60th Vehicular Technology Conference*, Los Angeles, CA, USA, Sept. 2004.

[2] I. A. Rezek and S. J. Roberts, "Parametric model order estimation: a brief review," in *IEE Colloquium on the Use of Model Based Digital Signal Processing Techniques in the Analysis of Biomedical Signals*, London, UK, April 1997, pp. 3/1–3/6.

[3] H. Akaike, "A new look at the statistical model identification," *IEEE Trans Automat. Contr*, vol. AC-19, pp. 716 – 723, Dec. 1974.

[4] J. Rissanen, "Modeling by shortest data description," *Automatica*, 1978.

[5] S. K. Katsikas, S. D. Likothanassis, and D. G. Lainiotis, "AR model identification with unknown process order," *IEEE Transactions on Acoustics Speech & Signal Processing*, 1990.

[6] Y. Gong and C. F. N. Cowan, "Structure adaptation of linear MMSE adaptive filters," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 151, no. 4, pp. 271 – 277, Aug. 2004.

[7] T. Matsuoka and T. J. Ulrych, "Information theoy measures with application to model identification," *IEEE Trans ASSP*, vol. ASSP-34, no. 3, pp. 511 – 517, June 1986.