16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

OPTIMAL LOCAL POLYNOMIAL REGRESSION OF NOISY TIME-VARYING SIGNALS

A. Sreenivasa Murthy[1] *and T. V. Sreenivas*[2]

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore, India
Emails: {[1]asmuvce,[2]tvsree}@ece.iisc.ernet.in

ABSTRACT

We address the problem of local-polynomial modeling of smooth time-varying signals with unknown functional form, in the presence of additive noise. The problem formulation is in the time domain and the polynomial coefficients are estimated in the pointwise minimum mean square error (PMMSE) sense. The choice of the window length for local modeling introduces a bias-variance tradeoff, which we solve optimally by using the *intersection-of-confidence-intervals* (ICI) technique. The combination of the local polynomial model and the ICI technique gives rise to an adaptive signal model equipped with a time-varying PMMSE-optimal window length whose performance is superior to that obtained by using a fixed window length. We also evaluate the sensitivity of the ICI technique with respect to the confidence interval width. Simulation results on electrocardiogram (ECG) signals show that at 0 dB signal-to-noise ratio (SNR), one can achieve about 12 dB improvement in SNR. Monte-Carlo performance analysis shows that the performance is comparable to the basic wavelet techniques. For 0 dB SNR, the adaptive window technique yields about 2-3 dB higher SNR than wavelet regression techniques and for SNRs greater than 12 dB, the wavelet techniques yield about 2 dB higher SNR.

1. INTRODUCTION

Naturally occurring signals such as geophysical signals, music, electrocardiogram (ECG), and bioacoustic signals such as speech, cricket calls, bird songs etc. are time-varying in one or more properties. Signals such as speech have approximate generation models [1] whereas signals such as bird songs or music do not have a good and general signal production model. In any case, the signal generation model has to be adaptive because the signal is time-varying. A common property of these signals is that they are produced by mechanical systems which are slowly time varying because, by natural limitation, they cannot vary too fast. These signals are also associated with a temporal structure that is necessary to be preserved by signal processing techniques. Electrocardiogram (ECG) signals, for example contain useful diagnostic information in the temporal structure.

We promote the use of local signal models instead of production-model-based ones since they are more general and applicable to a variety of natural signals. Of particular interest are the local representations using polynomials. The local polynomial model offers a smooth representation using few parameters (the polynomial coefficients) and can be adapted to the local signal properties. Natural signals are also noisy and hence the modeling process must be robust to noise. Thus, there are two principal requirements—adaptability to local signal behaviour and noise-robustness.

Local polynomial modeling (LPM) is a well-established area of research (see [2, 3] for a review) and has found several applications in kernel density estimation, spectral estimation and nonlinear time-series modeling. In almost all LPM problems, there is the need for an optimal window width selection criterion, also known as *bandwidth selection*. In simpler terms, the question that one is faced with is "How much data is too much and how little is too little?" Several criteria have been proposed based on the optimization of a suitable cost function and a good review of these techniques can be found in [2]. In this paper, we consider the pointwise mean square error criterion and show that there is a tradeoff between the adaptability of the model to signal variations and robustness to noise. From a statistical signal processing perspective, this is the bias-variance tradeoff [4, 5]. We show that this tradeoff can be efficiently solved by using a technique known as the *intersection-of-confidence intervals (ICI) technique* [9, 10]. In doing so, we demonstrate that a function model as simple as the polynomial becomes powerful when combined with the notions of local modeling, adaptation, optimality and noise-robustness. The technique can be applied to other kinds of bases as well, but we confine our discussion to polynomials because they are simple.

2. POINTWISE MINIMUM MEAN SQUARE ERROR

Let $\{s[nT], n \in \mathbb{Z}\}$ where T is the sampling period denote the samples of a deterministic time-varying signal $s(t)$ that are corrupted by additive white Gaussian noise samples $w[nT]$ of variance σ_w^2. The noisy observations are given by

$$y[nT] = s[nT] + w[nT]. \tag{1}$$

The objective is to obtain a minimum-mean square error (MMSE) estimate of $s[nT]$ given $y[nT]$. We assume that $s(t)$ is smooth enough to permit a polynomial approximation of a chosen order p, locally over an interval $\left[nT - \frac{L}{2}, nT + \frac{L}{2}\right]$. This approach is motivated by the Weierstrass's theorem [6, 7] which states that a function, continuous in a finite closed interval can be approximated to a desired accuracy by polynomials. With this motivation, the signal can be expressed as:

$$s[nT] = \sum_{k=0}^{p} a_k (nT)^k + \mathcal{O}\{(nT)^{p+1}\}, \tag{2}$$

for $nT \in \left[mT - \frac{L}{2}, mT + \frac{L}{2}\right]$ and where $\mathcal{O}$ is the Landau symbol. The model parameter vector is $\mathbf{a} = [a_0\ a_1\ a_2\ \ldots\ a_p]$. Since $\{y[nT], n \in \mathbb{Z}\}$ are noisy, there is an error in computing the polynomial coefficients and hence we can only obtain an estimate of $s[nT]$ which we denote by $\widehat{s}[nT]$. The estimation error is defined as $\Delta s[nT] = s[nT] - \widehat{s}[nT]$.

Consider the least-squares cost function $\mathcal{C}_m(\mathbf{a})$ at $t = mT$:

$$\mathcal{C}_m(\mathbf{a}) = \sum_{nT=mT-\frac{L}{2}}^{mT+\frac{L}{2}} \left(y[nT] - \sum_{k=0}^{p} a_k (nT)^k \right)^2 h(mT-nT), \tag{3}$$

where $h(t)$ is a window function of length L and symmetric about $t = 0$. The role of $h(t)$ is to localize the model-fitting around the instant mT. We use an L-point rectangular window for simplicity, although other window functions can be used [2]. The optimum coefficient vector $\hat{\mathbf{a}}$ is found as:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \mathcal{C}, \tag{4}$$

which is a local least squares solution. Using these coefficients, one can compute the asymptotic (as $T \to 0$) expressions for the bias and variance of the estimation error $\Delta s(t)$ as follows ([2] contains the derivations in the context of kernel density estimation but they are applicable to our problem too):

$$\mathcal{E}\{\Delta s(t)\} = \begin{cases} \frac{s^{(p+1)}(t)}{(p+2)!} \left(\frac{L}{2}\right)^{p+1} & \text{if } p \text{ is odd and} \\ \frac{s^{(p+2)}(t)}{(p+3)!} \left(\frac{L}{2}\right)^{p+2} & \text{if } p \text{ is even,} \end{cases} \tag{5}$$

$$\mathrm{Var}\{\Delta s(t)\} = \frac{\sigma_w^2}{L}, \tag{6}$$

where Var{.} denotes variance. Since the model uses a p^{th}-order polynomial, the bias contains the higher-order terms starting from order $(p+1)$. Only the dominant term has been retained in the above equation.

We observe that the squared bias $(\mathcal{E}\{\Delta s(t)\})^2$ is directly proportional to L^{2p+2} or L^{2p+4}, according as p is odd or even. The variance is inversely proportional to L. The complementary dependence of bias and variance on L explains the bias-variance tradeoff due to which it is not possible to minimize both squared bias and variance. A reasonable alternative is to minimize the mean-squared error (MSE) which is the sum of squared bias and variance. Without loss of generality, we assume that p is odd and write the MSE as:

$$\mathrm{MSE}\{\widehat{s}(t)\} = \frac{[s^{(p+1)}(t)]^2 L^{2p+2}}{[(p+2)!]^2 2^{2p+2}} + \frac{\sigma_w^2}{L}. \tag{7}$$

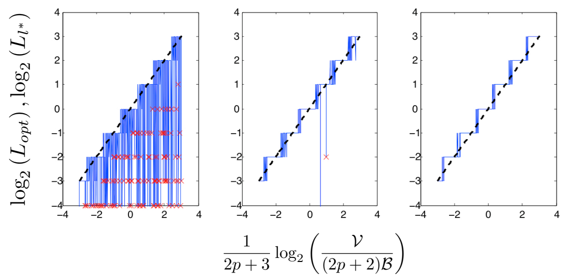

Figure 1: ICI technique reliability as a function of κ for Gaussian noise. The outliers are indicated by a 'x'. The 45^o straight line corresponds to perfect knowledge of the signal derivatives and hence the optimal window length L_{opt}. The plots correspond to $\kappa = 1.5, 2.5$ and 3.5 respectively.

Assuming that p is fixed, we can minimize $\mathrm{MSE}\{\widehat{s}(t)\}$ with respect to L and obtain

$$L_{\text{opt}} = \left(\frac{\sigma_w^2 2^{2p+2} [(p+2)!]^2}{[s^{(p+1)}(t)]^2 (2p+2)} \right)^{\frac{1}{2p+3}}. \tag{8}$$

To emphasize that the MSE is time-varying, we refer to this optimization as the pointwise MMSE (PMMSE) criterion.

Corresponding to $L = L_{\text{opt}}$, we have

$$\mathrm{Var}\{\Delta s(t)\} = \sigma_w^{\frac{2(2p+2)}{2p+3}} \mathcal{B}^{\frac{1}{2p+3}} (2p+2)^{\frac{1}{2p+3}} \text{ and} \tag{9}$$

$$\mathcal{E}\{\Delta s(t)\} = \sqrt{\frac{1}{2p+2}} \sigma_{L_{\text{opt}}}, \tag{10}$$

where

$$\sigma_{L_{\text{opt}}} = \sqrt{\mathrm{Var}\{\Delta s(t)\}} \text{ and} \tag{11}$$

$$\mathcal{B} = \frac{[s^{p+1}(t)]^2}{[(p+2)!]^2 2^{2p+2}}. \tag{12}$$

L_{opt} is a function of the higher-order derivatives of the signal which are not known apriori in practice. Therefore, direct utilization of (8) does not seem to be practically feasible. However, as we show next, a solution close to L_{opt} can be found by probabilistically trading off bias and variance. Henceforth, we drop the time index in $\Delta s(t)$ and $\widehat{s}(t)$. To make the dependence of the estimators on L explicit, we use $\widehat{s}_L$ to denote the estimate of s obtained using the window of length L, and its variance by $\sigma^2(L)$.

3. WINDOW LENGTH OPTIMIZATION BY USING THE ICI TECHNIQUE

3.1 The ICI technique

The ICI technique is an efficient way of finding near-optimal solutions in the PMMSE sense. The technique was applied successfully for the problem of instantaneous frequency estimation [8, 9, 10, 12]. It is based on the notion of a confidence interval. The 2κ-confidence interval of the estimator $\widehat{s}_{L_\ell}$ is defined as

$$D_\ell = [\widehat{s}_{L_\ell} - \kappa\sigma(L_\ell), \widehat{s}_{L_\ell} + \kappa\sigma(L_\ell)]. \quad (13)$$

In the ICI technique, one computes the confidence interval corresponding to an increasing progression of window length values $\{L_\ell, \ell = 1, 2, 3, ...\}$. The progression can be linear/arithmetic: $L_\ell = a + L_{\ell-1}; \ell = 1, 2, ...$ or logarithmic/geometric:$L_\ell = aL_{\ell-1}; \ell = 1, 2, ...$ The initialization in both cases is $L_0 = p+1$ to ensure that there is sufficient data to fit a p^{th}-order polynomial. In our experiments, we choose $a = 2$. The ICI technique defines the optimal window length as that value for which two successive confidence intervals D_ℓ and $D_{\ell+1}$ *just cease to overlap.*

To ensure that the best performance of the ICI technique is achieved, we need to analyze the sensitivity to the confidence interval parameter κ, window length discretization (arithmetic/geometric) and accuracy of noise variance estimation. A sensitivity analysis with respect to confidence interval parameter was carried out in the context of instantaneous frequency estimation using time-frequency representations [10, 13] and sensitivity with respect to window length discretization was carried out for zero-crossing IF estimation in [12]. However, since we are addressing a different problem and the bias-variance expressions are different, we need to perform new experiments in the present context to obtain suitable parameters. Typically, a dyadic set of windows [10] has been used but we show in Sec.3.3 that we can achieve consistently lower MSE by using an arithmetic progression of window lengths. Earlier work on local polynomial modeling has shown that third-order polynomials give rise to smooth curves [2]; therefore, we use $p = 3$ in our experiments.

3.2 Optimal κ

The optimal value of κ is the *smallest* value that keeps the probability of choosing an erroneous window length sufficiently small. Consider the following biased random variable model for the estimation error:

$$s_L - \widehat{s}_L = \Delta s = w\sqrt{\frac{\mathcal{V}}{L}} + \sqrt{\mathcal{B}}L^{p+1}, \quad (14)$$

where

$$\mathcal{B} = \frac{[s^{p+1}(t)]^2}{[(p+2)!]^2 2^{2p+2}} \text{ and} \quad (15)$$

$$\mathcal{V} = \sigma_w^2, \quad (16)$$

and w is a zero-mean, unity variance Gaussian random variable. $\mathcal{B}$ is varied logarithmically such that $\frac{1}{2p+3}\log_2\left(\frac{\mathcal{V}}{(2p+2)\mathcal{B}}\right) \in [-3, 3]$ with a step size of 0.006. For each of the values, we calculate the optimum window length according to (8). The ICI-solution (L_{l^*}) and the actual optimum (L_{opt}) are shown in Fig. 1 on a log scale for $\kappa = 1.5$, 2.5 and 3.5. We have also indicated the number of outliers for each value of κ. For $\kappa = 2.5$, the number of outliers are few (1 in this case). We repeated the experiment an arbitrary 17 times and the number of outliers in each experiment were $\{1, 2, 2, 0, 1, 1, 2, 3, 2, 0, 3, 0, 2, 2, 3, 3, 2\}$. For $\kappa = 3.5$ and above, the number of outliers is zero in this experiment of 1000 realizations. If the number of realizations is increased by a few orders of magnitude, a few outliers will show up for $\kappa = 3.5$ too. Since $\kappa = 2.5$ yields only a few number of outliers which is sufficiently small for our application, we use $\kappa = 2.5$ in our experiments.

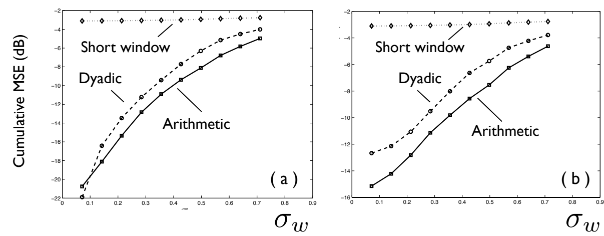

Figure 2: The cumulative MSE vs. σ_w for different window choices: fixed short window, ICI with dyadic window set and ICI with arithmetic window set, corresponding to (a) exact noise variance and (b) estimated noise variance.

3.3 Effects of noise variance estimation and window length discretization

The confidence-interval computation requires the knowledge of the noise variance which is seldom available apriori. However, it can be estimated. To analyze the importance of the accuracy of noise variance estimation, we conduct an experiment with a time-varying signal. We generate a linear frequency-modulated chirp signal according to the definition:

$$s[n] = \sin\left(0.2\frac{n^2}{N} + 0.2n + 0.4\right), \quad 0 \le n \le 255, \quad (17)$$

in the presence of additive white Gaussian noise. The standard deviation of noise σ_w is chosen as α times the empirical standard deviation of $s[n]$. α is varied from 0.1 to 0.9 in steps of 0.1. We used the median estimator for the standard deviation of noise [10] and obtained the performance with both arithmetic as well as dyadic progression of window lengths. In addition, we also obtain the results with a fixed window of 33 samples. To benchmark the performance, we repeated the experiment with exact knowledge of the noise variance. The cumulative (average squared error in the interval $[0, 255]$) MSE is shown in Fig. 2(a) and (b). Note that the adaptive window technique using an arithmetic window set consistently offers the best performance except for very low levels of noise, where the performance is marginally inferior to dyadic window lengths by about 1dB (Fig. 2(a)). The performance with the fixed window is always inferior to the ICI-optimal performance. By comparing Fig. 2(a) and (b), we conclude that improving noise variance estimation can lead to higher accuracy in smoothing.

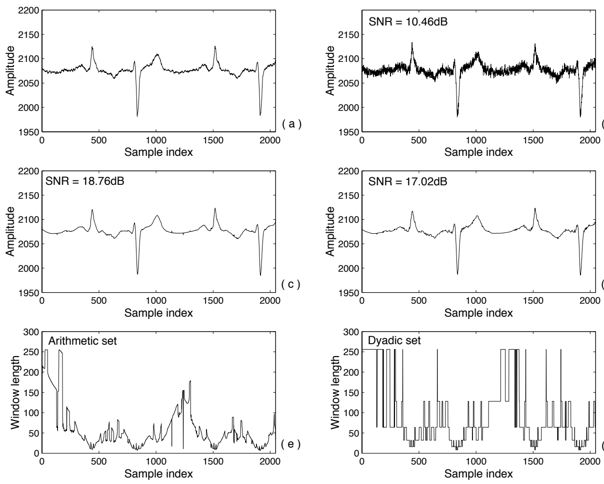

Figure 3: (a) The measured ECG signal (with some inherent measurement noise), (b) noisy ECG signal obtained by adding pseudorandom noise, (c) enhanced ECG signal (arithmetic set of window lengths), (d) enhanced ECG signal (dyadic set of window lengths), (e) instantaneous window lengths (arithmetic set) and (f) instantaneous window lengths (dyadic set)

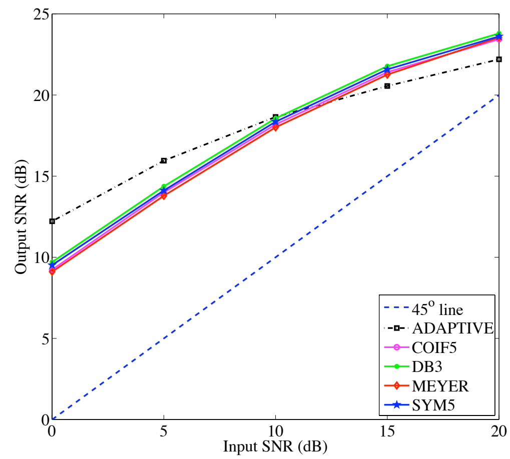

Figure 4: (Color online) Performance comparison of the adaptive local polynomial regression technique (arithmetic set) with wavelet regression techniques and soft-thresholding. (COIF: Coiflet, DB: Daubechies, MEYER: Meyer wavelet, SYM: Symlets).

4. RESULTS ON ECG SIGNALS

To analyze the performance of the technique on real-world signals, we chose the ECG signal because it has a strong temporal structure which needs to be preserved by any kind of processing. The ECG samples are taken from PhysioBank [11], where the signals are recorded at 720 Hz. An example is shown in Fig. 3(a) and its noisy version (additive white Gaussian noise, SNR = 10.46 dB) is shown in Fig. 3(b). The optimum window estimates corresponding to the arithmetic and dyadic windows are shown in Figs. 3(c) and (d) respectively. Notice that there is 8dB improvement in the SNR, which is significant. Also, there is no distortion in the temporal structure of the signal. From Fig. 3(e) and (f), we see that the instantaneous window lengths are optimally chosen by the ICI algorithm in a manner that is noise-robust and that also preserves the temporal structure. Short window lengths are chosen for fast signal variations/high curvature and long window lengths for slow local signal variations.

By using ICI for adaptive regression, we have not only obtained the optimal window lengths and local polynomial model parameters at each instant but also denoised the signal. Therefore, it would be appropriate to compare the denoising performance with the basic wavelet techniques. In Figs. 4, 5 and 6, we compare the performance with respect to standard wavelet-based denoising techniques with soft-thresholding [14]. A three-level decomposition was performed in all the cases. As the figures show, the adaptive regression technique is superior by about 2-3 dB for low SNR and for high SNR, the wavelet techniques have 2-3 dB higher accuracy. These results indicate that the adaptive regression technique performance is quite promising.

Software

The MATLAB software for implementing the adaptive regression technique can be requested from the authors by email. The wavelet denoising techniques are implemented by using the MATLAB Wavelet toolbox.

5. CONCLUSIONS

We addressed the problem of optimal local polynomial regression of noisy time-varying signals. by formulating a pointwise MMSE cost function. In this framework, we applied the ICI technique and analyzed its senstivity to various parameters. The combination of local polynomial regression and ICI techniques yields an *adaptive regression* technique which is ideally suited for processing time-varying signals. We have demonstrated the performance on real-world ECG measurements and showed that the improvement in the SNR is comparable to that of the wavelet techniques.

REFERENCES

[1] J.L. Flanagan, Speech Analysis, Synthesis, and Perception, 2nd ed., *Springer-Verlag*, New York, 1976.

[2] J. Fan and I. Gijbels, Local polynomial modeling and its application, *Chapman and Hall*, London, 1996.

[3] W. Hardle, Applied nonparametric regression, *Cambridge University Press*, Cambridge, 1990.

[4] S. M. Kay, Modern spectral estimation: Theory and applications, *Prentice Hall*, 1988.

[5] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Upper Saddle River, New Jersey, USA: *Prentice-Hall*, 1993.

[6] S. K. Sen and E. V. Krishnamurthy, Numerical Algorithms: Computations in Science and Engineering, *Affiliated East-West Press*, 2001.

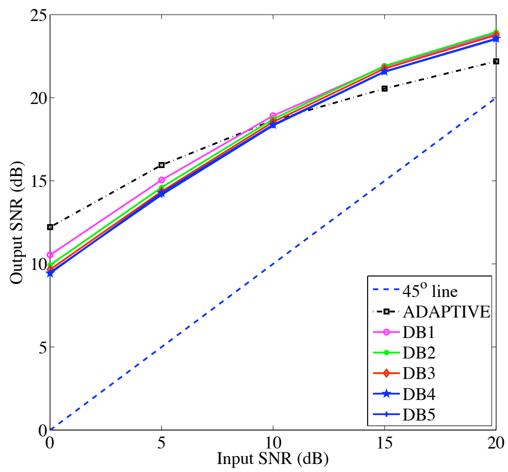

Figure 5: (Color online) Performance comparison of the adaptive local polynomial regression technique (arithmetic set) with wavelet regression techniques and soft-thresholding; Wavelets: DB1 to DB5 (DB: Daubechies).

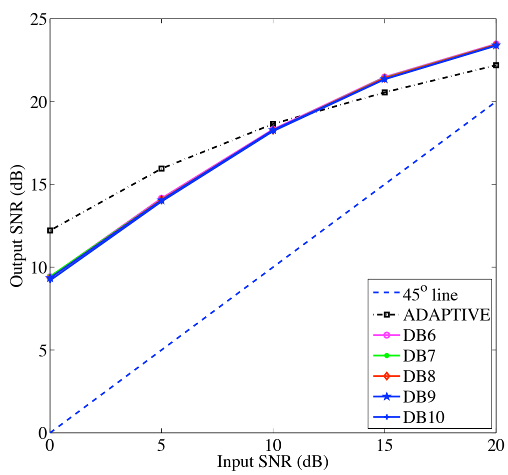

Figure 6: (Color online) Performance comparison of the adaptive local polynomial regression technique (arithmetic set) with wavelet regression techniques and soft-thresholding; Wavelets: DB6 to DB10 (DB: Daubechies).

[7] Weisstein, E. W. "Weierstrass's Theorem," http://mathworld.wolfram.com/WeierstrasssTheorem.html

[8] S. C. Sekhar , T. V. Sreenivas, "Adaptive spectrogram vs. adaptive pseudo-Wigner-Ville distribution for instantaneous frequency estimation", *Signal Proc.*, v.83 n.7, p.1529-1543, July 2003.

[9] V.Katkovnik and L.Stankovic, "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length", *IEEE Trans.on Signal Proc.* vol. 46, pp 2315-2325, Sept 1998.

[10] L. Stankovic and V. Katkovnik, "Algorithm for the instantaneous frequency estimation using time frequency distributions with adaptive window length", *IEEE Sig. Proc. Lett.,*, vol. 5, No.9, pp 224-227, Sept 1998.

[11] http://www.physionet.org/physiobank/database/aami-ec13/.

[12] S. C. Sekhar and T. V. Sreenivas, "Adaptive window zero-crossing-based instantaneous frequency estimation," *EURASIP J. on Applied Sig. Proc.*, vol. 2004, no. 12, pp. 1791-1806, 2004.

[13] B. Boashash, Time-frequency analysis and signal processing - A complete reference, *Elsevier*, 1993.

[14] D. L. Donoho, "Denoising by soft thresholding", *IEEE Trans. Inf. Theory*, Vol. 41, No. 3, pp. 613-627, 1995.