OPTIMAL MAP ESTIMATION OF BILINEAR SYSTEMS VIA THE EM ALGORITHM

Vikram Krishnamurthy, Leigh Johnston and Andrew Logothetis

Department of Electrical and Electronic Engineering
University of Melbourne
Parkville, Victoria 3052, Australia
{v.krishnamurthy, l.johnston}@ee.mu.oz.au

ABSTRACT

In this paper we present a finite dimensional iterative algorithm for optimal *maximum a posteriori* (MAP) state estimation of bilinear systems. Bilinear models are appealing in their ability to represent or approximate a broad class of nonlinear systems. We show that several bilinear models previously considered in the literature are special cases of the general bilinear model we propose. Our iterative algorithm for state estimation is based on the Expectation–Maximization (EM) algorithm and outperforms the widely used Extended Kalman filter (EKF). Unlike the EKF, our algorithm is an optimal (in the MAP sense) finite–dimensional solution to the state sequence estimation problem for bilinear models.

1. INTRODUCTION

Bilinear models [12] are widely used to model nonlinear processes in signal processing and communication systems. Nonlinear systems to which bilinear models can be applied arise in channel equalization [2], echo cancellation [1], nonlinear tracking [6], multiplicative disturbance tracking [11], and many other areas of engineering, socioeconomics and biology [4]. Further details are given in Section 2.

Due to the widespread use of bilinear models, there is strong motivation to develop estimation algorithms for the state of such systems given noisy observations. Unfortunately, the optimal filter for reconstructing conditional mean estimates of the state of a partially observed bilinear system is infinite dimensional. Thus practical filtering algorithms for bilinear systems are necessarily suboptimal. For example, the extended Kalman filter (EKF) is an approximate filter that linearizes around conditional mean state estimates at each time instant.

Rather than attempting to compute approximations to the conditional mean estimates, in this paper we adopt a different criterion for state estimation – maximum a posteriori (MAP) state sequence estimation. *We present an iterative finite dimensional algorithm for computing the optimal MAP state sequence estimate for a bilinear system*. In particular, we use the Expectation Maximization (EM) algorithm to compute the MAP state sequence estimate. Note that EM is traditionally used for maximum likelihood parameter estimation; the novel idea here is to use EM for MAP state sequence estimation.

Partially supported by the Co-operative Research Centre for Sensor Signal and Information Processing (CSSIP)

Fnaiech and Ljung's paper [5] discusses methods for parameter identification of bilinear systems. These methods are directly transferred from linear system identification methods, such as least squares and recursive prediction error methods. A Conjugate Gradient method for identification of bilinear systems is developed by Bose and Chen in [3]. Uses of bilinear models are not restricted to system identification. Hazarika, Tsoi and Sergejew use bilinear models to improve the modeling of EEG signals, analysis of which is carried out via an artificial neural network [7]. Lewis *et.al.* use Walsh functions to estimate the state of bilinear systems [10].

The signal model considered in this paper is based on a much broader class of bilinear processes than in [5],[12] where the input is a known deterministic sequence or white noise. We generalize [5],[12] by assuming the input signal to be an unknown autoregressive (AR) process. We formulate the bilinear processes in vector state-space form, allowing for the possibility of bilinearity in the state and observation equations.

We call our EM algorithm the KSEM algorithm, as it is an expectation–maximization (EM) algorithm that combines two Kalman smoothers. Highlights of the KSEM include:

(i) Each iteration of the KSEM has a finite dimensional analytical solution. The solution is somewhat remarkable: it involves the exact separation of the expectation and maximization steps, each carried out via a Kalman smoother (see Section 3).

(ii) Proven convergence, based on convergence of the EM algorithm (see [13] for proof), albeit to a local stationary point. Most approximate nonlinear filtering methods are heuristic with no convergence proof [7].

(iii) Application to more general models than have been previously considered in the literature [12],[5].

(iv) Simulations show that the KSEM algorithm outperforms the extended Kalman filter (EKF), in its ability to estimate single and multiple autoregressive (AR) processes (see Section 4).

(v) Potential for online recursive implementation. (This will be explored in future work [9])

2. BILINEAR SYSTEM MODELS

The general T-point, discrete time, bilinear model we will consider is

$$\mathbf{x}_{k+1} = F(\mathrm{s}_k)\mathbf{x}_k + D\mathrm{s}_k + \mathbf{w}_k \quad (1)$$

$$\mathbf{s}_{k+1} = B\mathbf{s}_k + \mathbf{v}_k \quad (2)$$
$$y_k = H(\mathbf{s}_k)'\mathbf{x}_k + e_k \quad (3)$$

where $F(\mathbf{s}_k)$ is a matrix–valued function and $H(\mathbf{s}_k)$ is a vector–valued function, both linear in $\mathbf{s}_k$. This system model is bilinear in both state and observation equations.

Practical Examples: We relate (1)–(3) to models considered in [11] and [7].

Example 1: The multiplicative disturbance model proposed in [11] is a case of the above system model, for which

$$F(\mathbf{s}_k) = \begin{bmatrix} s_{1_k} & \dots & s_{(p-1)_k} & s_{p_k} \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}$$
$$B = I_{p\times p}\ ,\ D = 0$$
$$\mathbf{w}_k = [w_{1_k}, w_{2_k}, \dots, w_{p_k}]'$$
$$\mathbf{v}_k = [v_k, 0, \dots, 0]'$$

Example 2: The bilinear processes considered by Hazarika *et.al.* to model EEG signals [7] are of the form

$$x_k + \sum_{i=1}^{p} a_i x_{k-i} = \sum_{i=1}^{m}\sum_{j=1}^{r} b_{ij} X_{k-i} v_{k-j} + v_k \quad (4)$$

This is another case of our general bilinear system model, for which

$$s_k = v_k\ \forall k$$
$$F(\mathbf{s}_k) = \left[A + [\mathbf{r}_{k-1}\ \mathbf{r}_{k-2}\ \dots\ \mathbf{r}_{k-r}] * M'\right]$$
$$\mathbf{r}_{k-i} = [s_{k-i}\ 0\ \dots\ 0]'$$
$$M = \{b_{ij}\}$$
$$A = \begin{bmatrix} -a_1 & \dots & -a_{p-1} & -a_p \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}$$
$$B = 0\ ,\ D = [1\ 0\ \dots\ 0]$$
$$\sigma_e^2 = 0\ ,\ \sigma_w^2 = 0$$

Similarly, one can express several other bilinear models as special cases of the general bilinear model of (1)-(3). For the purposes of this paper, $H(\mathbf{s}_k) = H$ (no observation process bilinearity) and $D = 0$ (no linear s_k term in the $\mathbf{x}_k$ process). The KSEM algorithm is derived according to these conditions on the system matrices. Other algorithms follow from other models, which are to be discussed in [9].

Aim: Let $S_T = [\mathbf{s}_1\ \dots \mathbf{s}_T]$ and $Y_T = [y_1\ \dots y_T]$. Given the signal model (1)-(3), our objective is to compute the MAP state sequence estimate

$$S_T^{MAP} = \arg\max_{S_T} f(S_T, Y_T)\ . \quad (5)$$

Here $f(S_T, Y_T)$ denotes the joint density of the state sequence S_T and observation sequence Y_T.

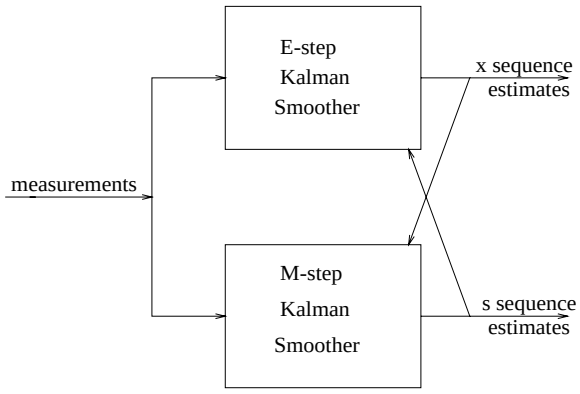

Figure 1: Structure of the KSEM algorithm

3. MAP ESTIMATION ALGORITHM

In this section we present an EM algorithm, called the KSEM, to compute the MAP state sequence estimate defined in (5). Each iteration of the algorithm has the following intuitive interpretation: It cross-couples two Kalman smoothers, one for the E-step which estimates $\{\mathbf{x}_k\}$ and the other for the M-step which computes the the MAP $\{\mathbf{s}_k\}$ state sequence estimate. If either the $\{\mathbf{x}_k\}$ or $\{\mathbf{s}_k\}$ sequences are known, the bilinear signal model (1)-(3) reduces to a linear, Gaussian system, the optimal estimate of which is obtained by a Kalman smoother. The remarkable result we present here is that when both AR processes are unknown in the bilinear system model, optimal MAP estimates of $\mathbf{s}_k$ can be achieved in a clean, precise manner by cross-coupling two Kalman smoothers (one for the E-step and the other for the M-step). The structure of the KSEM algorithm is shown in Figure 1.

The EM algorithm described below obtains S_T^{MAP} by iteratively generating new state sequence estimates such that the state sequence likelihood $f(S_T, Y_T)$ monotonically increases until the MAP sequence S_T^{MAP} is obtained. Given mild regularity conditions, it can be proven [13] that the EM algorithm converges to a local stationary point on the state sequence likelihood $f(S_T, Y_T)$. Details appear in [9].

KSEM Algorithm: For iterations $l = 1, 2, \dots$

E step: Evaluate the conditional expectation of the log likelihood

$$\mathcal{Q}(S_T, S_T^{(l)}) \triangleq E\{\ln f(S_T, X_T, Y_T)|Y_T, S_T^{(l)}\} \quad (6)$$

to obtain estimates of $\widehat{\mathbf{x}_k\mathbf{x}_k'}$ and $\widehat{\mathbf{x}_k\mathbf{x}_{k-1}'}$.

M step: Compute $S_T^{(l+1)}$:

$$S_T^{(l+1)} = \arg\max_{S_T} \mathcal{Q}(S_T, S_T^{(l)}) \quad (7)$$

For the bilinear system model (1)-(3) these steps have compact structures as a result of the Gaussian nature of the distributions.

$$\mathcal{Q}(S_T, S_T^{(l)}) = -\frac{1}{2}\sum_{k=1}^{T}(y_k - H\mathbf{x}_k)'\Sigma_e^{-1}(y_k - H\mathbf{x}_k)$$
$$-\frac{1}{2}\sum_{k=1}^{T}(\mathbf{s}_k - B\mathbf{s}_{k-1})'\Sigma_v^{-1}(\mathbf{s}_k - B\mathbf{s}_{k-1})$$

$$-\frac{1}{2}\sum_{k=1}^{T}(\mathbf{x}_k - F(\mathbf{s}_{k-1})\mathbf{x}_{k-1})'\Sigma_w^{-1}(\mathbf{x}_k - F(\mathbf{s}_{k-1})\mathbf{x}_{k-1}) \quad (8)$$

Computational Requirements: Each iteration of the KSEM algorithm requires $O(p^3T)$ for a T-point data sequence, p being the dimension of $\mathbf{s}_k$. This is the cost associated with a Kalman smoother.

3.1. Computing the E step

Estimates of $\widehat{\mathbf{x}_k\mathbf{x}_k'}$ and $\widehat{\mathbf{x}_k\mathbf{x}_{k-1}'}$ are required in the M step (see (19), (20)). This latter term is of dimension $p+1$, where p is the dimension of the $\mathbf{s}_k$ and $\mathbf{x}_k$ vectors. Therefore, ignoring terms in (8) independent of $\mathbf{x}_k$ and assuming knowledge of the sequence $S_T^{(l)}$, $\mathcal{Q}(S_T, S_T^{(l)})$ is the density of a linear, Gaussian system, the state sequence $\{\mathbf{x}_k\}$ of which can be estimated via a Kalman smoother on the augmented system equations

$$\begin{bmatrix} \mathbf{x}_{k+1} \\ x_{k-p} \end{bmatrix} = \begin{bmatrix} F(\mathbf{s}_k^{(l)}) & 0 \\ 0_{1xp} & 1 \end{bmatrix}\begin{bmatrix} \mathbf{x}_k \\ x_{k-p-1} \end{bmatrix} + \begin{bmatrix} I_{pxp} & 0 \\ 0_{1xp} & 1 \end{bmatrix}\mathbf{w}_k \quad (9)$$

$$y_k = \begin{bmatrix} H & 0 \end{bmatrix}\begin{bmatrix} \mathbf{x}_k \\ x_{k-p-1} \end{bmatrix} + e_k \quad (10)$$

3.2. Computing the M step

We seek to maximize $\mathcal{Q}(S_T, S_T^{(l)})$ to obtain an updated state sequence estimate, $S_T^{(l+1)}$. The surprising result is that this is achievable by application of a Kalman smoother on a modified linear, Gaussian system. The proof of this fact follows.

Theorem 1 A recursive solution to the maximization

$$S_T^{(l+1)} = \arg\max_{S_T} \mathcal{Q}(S_T, S_T^{(l)}) \quad (11)$$

can be achieved by applying a Kalman smoother to the linear, Gaussian system

$$\mathbf{s}_{k+1} = B\mathbf{s}_k + \mathbf{v}_k \quad (12)$$

$$\tilde{\mathbf{y}}_k = \tilde{H}\mathbf{s}_k + \mathbf{e}_k \quad (13)$$

where

$$\tilde{H}\tilde{H}' = G_2(\mathbf{x}\mathbf{x}')\big|_{\widehat{\mathbf{x}\mathbf{x}'}} \quad (14)$$

$$\tilde{\mathbf{y}} = \tilde{H}^{-'}G_1(\mathbf{x}_k\mathbf{x}_{k-1}')\big|_{\widehat{\mathbf{x}_k\mathbf{x}_{k-1}'}} \quad (15)$$

Remark: Theorem 1 states that the M-step, i.e. the maximization of (7), can be explicitly carried out via a Kalman smoother. This Kalman smoother operates on an averaged state space system (averaged over the states of the $\hat{X}_T$ sequence).

Proof of Theorem 1: We aim to write $\mathcal{Q}(S_T, S_T^{(l)})$ as a likelihood function for the system (12)-(13). Firstly, when maximizing $\mathcal{Q}(S_T, S_T^{(l)})$ of (8), all terms that are independent of S_T can be ignored, as they remain constant in the maximization procedure. Assuming that $\Sigma_v = \sigma_v^2 I$ and $\Sigma_w = \sigma_w^2 I$, expanding $\mathcal{Q}(S_T, S_T^{(l)})$ gives

$$\begin{aligned}\mathcal{Q}(S_T, S_T^{(l)}) = \\ -\tfrac{1}{2\sigma_v^2}\sum_{k=1}^{T}[\mathbf{s}_k'\mathbf{s}_k - 2\mathbf{s}_k'B\mathbf{s}_{k-1} + \mathbf{s}_{k-1}'B'B\mathbf{s}_{k-1} \\ + \mathbf{x}_{k-1}'F(\mathbf{s}_{k-1})'F(\mathbf{s}_{k-1})\mathbf{x}_{k-1}] \\ + \text{terms independent of } S_T\end{aligned} \quad (16)$$

As $F(\mathbf{s}_k)$ is linear in s_k, rewriting $F(\mathbf{s}_k)\mathbf{x}_k$ in terms of a function of $\mathbf{x}_k$ gives

$$F(\mathbf{s}_k)\mathbf{x}_k \triangleq G(\mathbf{x}_k)\mathbf{s}_k. \quad (17)$$

$$-2\mathbf{x}_k'F(\mathbf{s}_{k-1})\mathbf{x}_{k-1} \triangleq -2\mathbf{x}_k'G(\mathbf{x}_{k-1})\mathbf{s}_{k-1} \quad (18)$$

Estimates of $\widehat{\mathbf{x}_k'\mathbf{x}_{k-1}}$ are available from the augmented Kalman smoother output of the E-step. Make the substitution:

$$-2\mathbf{x}_k'G(\mathbf{x}_{k-1})\mathbf{s}_{k-1} = -2G_1(\mathbf{x}_k\mathbf{x}_{k-1}')\big|_{\widehat{\mathbf{x}_k\mathbf{x}_{k-1}'}}\mathbf{s}_{k-1} \quad (19)$$

As $G(\mathbf{x}_k)$ is linear, $G(\mathbf{x}_k)'G(\mathbf{x}_k)$ is quadratic. Estimates of $\widehat{\mathbf{x}\mathbf{x}'}$ are also a product of the Kalman smoother in the E-step, hence make a further substitution:

$$G(\mathbf{x}_k)'G(\mathbf{x}_k) = G_2(\mathbf{x}\mathbf{x}')\big|_{\widehat{\mathbf{x}\mathbf{x}'}} \quad (20)$$

By making these substitutions in (16) and further removing terms independent of S_T, one can complete the square to obtain:

$$\begin{aligned}\mathcal{Q}(S_T, S_T^{(l)}) = -\tfrac{1}{2\sigma_v^2}\sum_{k=1}^{T}(\mathbf{s}_k - B\mathbf{s}_{k-1})'(\mathbf{s}_k - B\mathbf{s}_{k-1}) \\ -\tfrac{1}{2}\sum_{k=1}^{T}(\tilde{\mathbf{y}} - \tilde{H}\mathbf{s}_k)'(\tilde{\mathbf{y}} - \tilde{H}\mathbf{s}_k) \\ + \text{terms independent of } S_T\end{aligned} \quad (21)$$

The signal model (12)–(13) follows directly from the expression for $\mathcal{Q}(S_T, S_T^{(l)})$ in (21).

Due to the Gaussianity and independence of v_k, e_k and s_0, the following expression for the joint density of S_T and $\tilde{Y}_T$ is easily derived:

$$\ln f(S_T, \tilde{Y}_T) = \mathcal{Q}(S_T, S_T^{(l)}) + [\text{terms independent of } s_k] \quad (22)$$

where $\mathcal{Q}(S_T, S_T^{(l)})$ is defined in (21). The MAP estimate of any linear Gaussian system e.g. (12) and (13), is known to be given by a fixed-interval Kalman smoother [8]. That is,

$$(s_{1|T}, \ldots, s_{T|T}) = \arg\max_{S_T} f(S_T, \tilde{Y}_T). \quad (23)$$

From (22),

$$\arg\max_{S_T} f(S_T, \tilde{Y}_T) = \arg\max_{S_T} \mathcal{Q}(S_T, S_T^{(l)}). \quad (24)$$

This is precisely the quantity that the EM algorithm computes.

$$S_T^{(l+1)} = \arg\max_{S_T} \mathcal{Q}(S_T, S_T^{(l)}) \quad (25)$$

It follows then that maximizing $\mathcal{Q}(S_T, S_T^{(l)})$ is equivalent to applying a Kalman smoother to the linear Gaussian model in (12) and (13). □

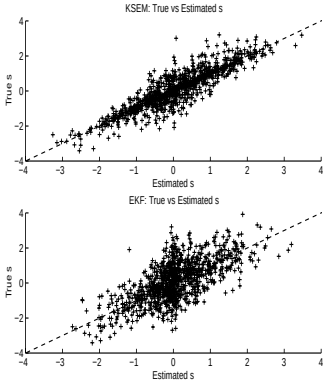

Figure 2: Scatterplots of true s_k sequence vs estimated s_k sequence for KSEM and EKF

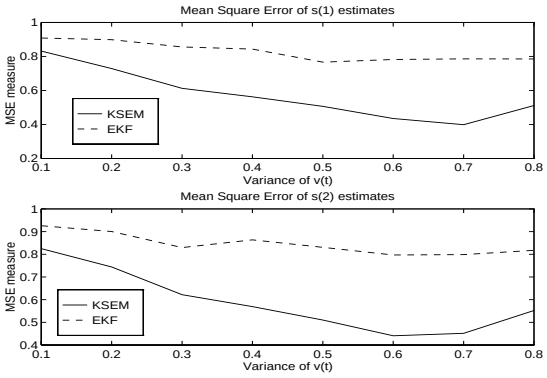

Figure 3: Proportional Mean Square Error of KSEM and EKF

4. SIMULATIONS

The bilinear system model of Section 2 is unstable for some parameter and sequence values. The s_k process cannot be too large, as it drives the time-varying coefficients of the x_k process. Over the noise variance ranges for which the bilinear system is stable, the KSEM algorithm outperforms the EKF. While both algorithms estimate the x_k states well, the EKF fails to estimate the s_k states with the same degree of accuracy as the KSEM.

Consider the particular system with observation noise variance $\sigma_e^2 = 1$, x state noise variance $\sigma_w^2 = 1$, and s state noise variance $\sigma_v^2 = 0.5$. The KSEM algorithm is iterated 10 times. Figure 2 shows plots of *true* $\{\mathrm{s}_k\}$ vs *estimated* $\{\mathrm{s}_k\}$ for the KSEM and EKF algorithms. Ideally the plot should be a linear relationship as shown by the broken line. The cloud obtained from the KSEM algorithm is more accurate than that obtained from the EKF. These plots are typical of algorithm performance.

To compare performance over a range of noise values, Figure 3 shows how the proportional Mean Square Error varies against s_k process noise variance, σ_v^2. Both σ_e^2 and $\sigma_w^2 = 1$ and the dimension of s_k is 2,

$$B = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.8 \end{bmatrix} \quad , \quad F(\mathrm{s}_k) = \begin{bmatrix} s_{1_k} & 0 \\ 0 & s_{2_k} \end{bmatrix} \tag{26}$$

and the KSEM algorithm is iterated 10 times.

The proportional Mean Square Error (MSE) metric we use is

$$\text{proportional MSE} = \frac{\sum_{k=1}^{T} (s_k - \hat{s}_k)^2}{\sum_{k=1}^{T} s_k^2}. \tag{27}$$

Dividing by the signal power ensures that the MSE of the x_k process is a proportion of its size, not an absolute measure. Figure 3 clearly shows the superiority of the KSEM algorithm over the EKF. Values around 1 indicate that signal estimate errors occur with the same power as the original signal. While not evident in this plot alone, the KSEM can estimate multiple $\{\mathrm{s}_k\}$ sequences, while the EKF only estimates one with any degree of success. Such facts will be explored further in future work [9].

5. REFERENCES

[1] O. Agazzi and D.G. Messerschmitt. *Nonlinear echo cancellation of data signals*, IEEE Transactions on Communications, Vol.30, pp. 2421-2433, November 1982.

[2] S. Benedetto and E. Biglieri, *Nonlinear equalization of digital satellite channels*, IEEE Journal of Selected Areas in Communications, Vol.1, pp. 57-62, January 1983.

[3] T. Bose and M. Chen, *Conjugate Gradient method in adaptive bilinear filtering*, IEEE Transactions on Signal Processing, Vol.43, No.6, pp. 1503-1508, 1995.

[4] C. Bruni, G. DiPillo, and G. Koch, *Bilinear systems: an appealing class of 'nearly linear systems' in theory and applications*, IEEE Transactions on Automatic Control, Vol.19, pp. 334-348, 1974.

[5] F. Fnaiech and L. Ljung *Recursive identification of bilinear systems* International Journal of Control, Vol.45, No.2. pp. 453-470, 1987.

[6] T.U. Halawani, R.R. Mohler, and W.J. Kolodziej *A two-step bilinear filtering algorithm* IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.32, No.2, pp. 344-352, 1984.

[7] N. Hazarika, A.C. Tsoi, and A.A. Sergejew, *Nonlinear considerations in EEG signal classification*, IEEE Transactions on Signal Processing, Vol.45, No.4, pp. 829-836, 1997.

[8] A.H. Jazwinski, *Stochastic processes and filtering theory*, Academic Press, 1970.

[9] L. Johnston and V. Krishnamurthy *Iterative algorithms for optimal MAP sstimation of bilinear systems*, In preparation.

[10] F.L. Lewis, V.G. Mertzios, G. Vachtsevanos, and M.A. Christodoulou, *Analysis of bilinear systems using Walsh functions*, IEEE Transactions on Automatic Control, Vol.35, No.1. pp. 119-123, 1990.

[11] M. Niedzwiecki and K. Cisowski, *Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals*, IEEE Transactions on Signal Processing, Vol.44, No.3, pp. 528-537, 1996.

[12] M.B. Priestley, *Non-linear and non-stationary time series analysis*, Academic Press, 1991.

[13] C.F.J. Wu *On the convergence properties of the EM algorithm*, J. Roy. Stat. Soc, Ser. B, Vol.39, No.1. pp. 1-38, 1983.