ICA for noisy neurobiological data

Shiro Ikeda
PRESTO, JST
BSI, RIKEN, Hirosawa 2–1
Wako, Saitama, 351-0198, Japan
shiro@brain.riken.go.jp

Keisuke Toyama
Shimadzu Technical Lab.
Hikaridai 3–9, Seika
Soraku, Kyoto, 619-0237, Japan
toyama@ext.shimadzu.co.jp

Abstract

ICA (Independent Component Analysis) is a new technique for analyzing multi-variant data. Lots of results are reported in the field of neurobiological data analysis such as EEG (Electroencephalography), MRI (Magnetic Resonance Imaging), and MEG (Magnetoencephalography) using ICA. But there still remain problems. In most of the neurobiological data, there are a large amount of noise, and the number of independent components is unknown which gives difficulties for many ICA algorithms. In this article, we discuss an approach to separate noise-contaminated data without knowing the number of independent components. The idea is to replace PCA (Principal Component Analysis), which is used as the preprocessing of many ICA algorithms, with factor analysis. In the new preprocessing, the number of the sources and the amount of the noise are estimated. After the preprocessing, an ICA algorithm is used to estimate the separation matrix and mixing system. Through the experiments with MEG data, we show this approach is effective.

1 Introduction

The basic ICA problem assumes the following linear relation between the observation $\boldsymbol{x}$ and the source $\boldsymbol{s}$ as,

$$\boldsymbol{x} = A\boldsymbol{s} \qquad \boldsymbol{x} \in R^n, \qquad \boldsymbol{s} \in R^m, \qquad A \in R^{n\times m}. \tag{1}$$

We assume that each component of $\boldsymbol{s}$ is mean 0, mutually independent and drawn from different probability distribution which is not Gaussian expect for at most one. In this article, we also restrict m to be smaller or equal to n for the existence of linear solution.

The goal of ICA is to estimate a separation matrix W which satisfies the following equation,

$$WA = PD \qquad W \in R^{m\times n}, \qquad P \in R^{m\times m}, \qquad D \in R^{m\times m}. \tag{2}$$

P is a permutation matrix which has single entry of one in each row and column, and D is a diagonal matrix. With the separation matrix, we can reconstruct the independent source as $\boldsymbol{y} = W\boldsymbol{x}$. This problem is solved in the framework of semi parametric approach[1], and giving a lot of interesting theoretical and practical results.

When we apply ICA to neurobiological data, the problem defined in eq.(1) does not describe the situation well. In most cases, we cannot avoid the noise and the number of the sources m is unknown. For example, in the case of MEG[4, 11], which we study in this article, the number of the measurements n is large (50∼200), but we believe that the number of the sources is not so large in a macroscopic viewpoint within a short period, and noise are very large.

MEG measures the brain activity through the change of the magnetic field which is extremely small ($\sim 10^{-14}$ T), and we need a special device called SQUID (Super-conducting Quantum Interference Device). The SQUID can detect the brain signals, but they contain a lot of noise. There are two major categories of noise. One is called the artifacts and the other is the quantum mechanical noise. The artifacts include electric power supply, earth magnetism, heart beat, breathing and the brain activity which we are not interested in. They effect on all the sensors simultaneously. The quantum mechanical noise originates from the SQUID itself. SQUID is working under the temperature of liquid Helium, and it cannot avoid quantum mechanical noise which is white and independent to each other. The main technique used so far to lower noise is the averaging over 100 to 200 times. But the averaged data still contains a large amount of noise. After averaging the data, we modeled the data as,

$$\begin{gathered}\boldsymbol{x} = A\boldsymbol{s} + \boldsymbol{\epsilon} \qquad \boldsymbol{x}, \boldsymbol{\epsilon} \in R^n, \quad \boldsymbol{s} \in R^m, \quad A \in R^{n\times m} \\ \boldsymbol{s}\text{: sources and artifacts,} \\ \boldsymbol{\epsilon} \sim N(\mathbf{o}, \Sigma)\text{: quantum mechanical noise,} \quad \Sigma \in R^{n\times n}\text{: diagonal matrix.}\end{gathered} \tag{3}$$

It is difficult to apply most of the ICA algorithms directory to this problem. In this article, we are proposing a new approach to solve this problem. The idea is to use factor analysis for the preprocessing of data. In the following sections, we describe our proposing method, and experimental results.

2 New approach of ICA

2.1 Factor analysis as the preprocessing

From a practical reason, ICA algorithms are often separated into two parts[5, 7, 8]. One is to preprocess the data such that they become uncorrelated. This part is called sphering or whitening. After the preprocessing, the remaining part of the estimation is a rotation matrix and many algorithms are proposed[3, 5, 7, 8]. But for sphering, most of the algorithms are using PCA[7]. Suppose we have a data set as $\{\boldsymbol{x}_t\}$ $(t = 1, \ldots, N)$ and let C be the covariance matrix of observed data $\boldsymbol{x}$, $(C = \sum_t \boldsymbol{x}_t \boldsymbol{x}_t^T / N)$. By defining $P = C^{1/2}$, where $C = PP^T$, PCA transform the observation as, $\boldsymbol{x}' = P^{-1}\boldsymbol{x}$, then $\boldsymbol{x}'$ is uncorrelated because $\sum \boldsymbol{x}' {\boldsymbol{x}'}^T / N = I_n$. But when the data are noisy, P does not make the observation uncorrelated. We can instead use factor analysis.

Factor analysis is one of the well-known techniques for analyzing multi-variant data. In factor analysis, real valued n-dimensional observation $\boldsymbol{x}$ is modeled as,

$$\begin{gathered}\boldsymbol{x} = A\boldsymbol{f} + \boldsymbol{\epsilon}, \qquad \boldsymbol{x}, \boldsymbol{\epsilon} \in R^n, \qquad \boldsymbol{f} \in R^m, \qquad A \in R^{n\times m} \\ \boldsymbol{f} \sim N(\mathbf{o}, I_m), \qquad \boldsymbol{\epsilon} \sim N(\mathbf{o}, \Sigma), \quad \Sigma \in R^{n\times n}\text{: diagonal matrix}\end{gathered} \tag{4}$$

The goal of factor analysis is to estimate m, A (factor loading matrix), and Σ (unique variance matrix) using the second order statistics C. The difference of eq.(3) and eq.(4) is coming from $\boldsymbol{s}$ and $\boldsymbol{f}$, that is $\boldsymbol{f}$ is assumed to be drawn from a normal distribution, but $\boldsymbol{s}$ is not.

Suppose we know A and Σ. Let $Q \in R^{m\times n}$ be the pseudo-inverse of A where $AQA = A$ holds, and we transform the data as, $\boldsymbol{z} = Q\boldsymbol{x}$. Then $\boldsymbol{z}$ becomes the sphered data because,

$$\frac{1}{N} \sum \boldsymbol{z}\boldsymbol{z}^T = I_m + Q\Sigma Q^T.$$

Pseudo-inverse has $(n - m)m$ degree of freedom. And we selected the one which minimizes $\operatorname{tr} Q\Sigma Q^T$. This result shows that we can make the part of observation $\boldsymbol{x}$ due to the sources uncorrelated, and this is the aim of sphering. Therefore, we have to estimate A and m.

For the estimation of m, A (factor loading matrix), and Σ (unique variance matrix), we can use some techniques of factor analysis. There are various estimation methods for A and Σ when m is given. We are going to use the MLE. MLE is defined as,

$$(\hat{A}, \hat{\Sigma})_{\mathrm{MLE}} = \underset{A,\Sigma}{\operatorname{argmax}} \left(-\frac{1}{2} \left\{ \operatorname{tr} \left(C(\Sigma + AA^T)^{-1} \right) + \log(\det(\Sigma + AA^T)) + n \log 2\pi \right\} \right). \tag{5}$$

For solving the equation, we can use the gradient decent algorithm or Gauss-Newton method. Also the EM (Expectation Maximization) algorithm can be applied.

In order to select the number of factors, m, there are also many approaches and we are going to use the model selection approach with an information criteria, MDL (Minimum Description Length). MDL is defined as follows,

$$\mathrm{MDL} = -L(\hat{A}, \hat{\Sigma}) + \frac{\log N}{N} \times \text{the number of free parameters.}$$

The number of the free parameters in factor analysis model is defined as follows. There are $n(m+1)$ parameters in A and Σ. But A has an ambiguity of rotation and $m(m-1)/2$ is the freedom of the ambiguity. Subtracting $m(m-1)/2$ from $n(m+1)$, the number of the free parameters is $n(m+1)-m(m-1)/2$, and,

$$\mathrm{MDL} = -L(\hat{A}, \hat{\Sigma}) + \frac{\log N}{N}\left(n(m+1) - \frac{m(m-1)}{2}\right).$$

For the existence of the estimates, a necessary condition for A to be estimable has been derived[9]. It comes from the fact that $n(n+1)/2 \geq n(m+1) - m(m-1)/2$ has to be satisfied, since C only has $n(n+1)/2$ degrees of freedom. By taking $m < n$ into account, the following bound is obtained

$$m \leq \frac{1}{2}\left\{2n+1-\sqrt{8n+1}\right\}. \tag{6}$$

2.2 Factor Analysis and ICA

After we preprocessed the data $\boldsymbol{x}$ by Q as $\boldsymbol{z} = Q\boldsymbol{x}$. What is left is to determine the rotation matrix. This is also one big problem in factor analysis. In subsection 2.1, we assumed that $\boldsymbol{f}$ and $\boldsymbol{\epsilon}$ are normally distributed. We break a part of the assumption. We still assume that $\boldsymbol{\epsilon}$ is normally distributed, but $\boldsymbol{f}$ is not normally distributed and each component is independent. We can use some ICA algorithm now.

The ICA algorithm we use here, should not be affected by the second order statistics since even if data are preprocessed by factor analysis, $\boldsymbol{z}$ still has second order correlations. Therefore, an algorithm based on higher order correlations is preferable here. We use the JADE algorithm by J.-F. Cardoso, which is based on the 4th order cumulant. The detail of the algorithm is shown in [6]. Suppose a separation matrix $W \in R^{m\times m}$ is estimated by JADE. The separated signal $\boldsymbol{y}$ is obtained as,

$$\boldsymbol{y} = W\boldsymbol{z} = WQ\boldsymbol{x} = WQ(A\boldsymbol{s} + \boldsymbol{\epsilon}). \tag{7}$$

The goal of ICA is to estimate WQA to be $PD : P, D \in R^{m\times m}$ as in eq.(2). And finally, we obtain separation matrix B as $B = WQ$.

Also, we can estimate the mixing system by using W. Let us denote A estimated by factor analysis as A_{FA}, and the new mixing system as A_{ICA} as,

$$A_{\mathrm{ICA}} = A_{\mathrm{FA}} W^T. \tag{8}$$

This A_{ICA} does not have rotation ambiguity, and we can estimate the mixing system. Thus, by using factor analysis and an ICA algorithm, we can estimate the mixing system and the separation matrix.

3 MEG data analysis

3.1 Experiment with phantom data

We applied our algorithm to phantom data. In the phantom, there is a small platinum electrode and we designed the current signal supplied to the electrode to be a 20Hz triangle wave, and averaged data over 100 trials. Fig.1(left) shows 3 signals out of 126 active sensors.

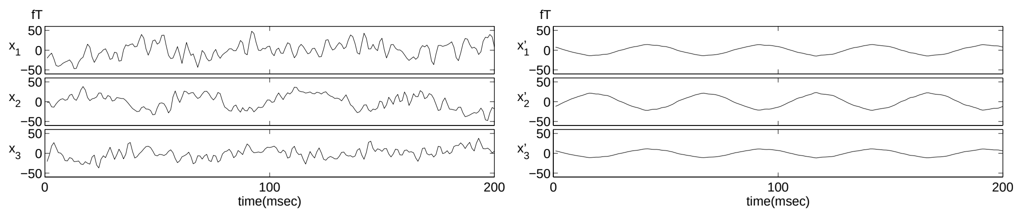

Figure 1: Sensor inputs of phantom data(left), and after removing artifacts(right)

The number of the sources and the separation matrix are estimated by our algorithm. We preprocessed the data with factor analysis and estimated the number of the sources by MDL. In this experiment, the

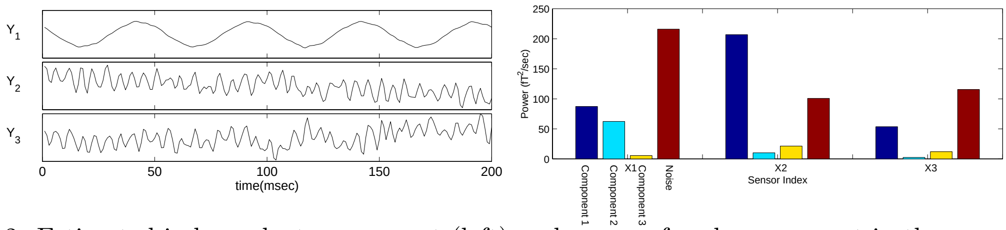

Figure 2: Estimated independent component (left) and power of each component in the sensors(right)

number of the sources is estimated as 3. After rotation matrix is estimated by JADE algorithm, we apply the pseudo-inverse matrix to the preprocessed data. The estimated independent signals are shown in Fig.2(left).

In this experiment, we know the input to the electrode is y_1 in Fig.2(left). After selecting the source, we want to reconstruct the signal. This is possible because the mixing matrix A_{ICA} is estimated as in eq.(8). Each column of $\hat{A}_{\text{ICA}}$ corresponds to the coefficients of each independent component to sensors. Let $\hat{A}_{\text{ICA}} = (\hat{\boldsymbol{a}}_{\text{ICA},1}, \ldots, \hat{\boldsymbol{a}}_{\text{ICA},m})$, where $\hat{\boldsymbol{a}}_{\text{ICA},i}$ is an n dimensional column vector. In this case, we can reconstruct the data with, $\boldsymbol{x}'(t) = \hat{\boldsymbol{a}}_{\text{ICA},1} y_1(t)$. The recovered signals are shown in Fig.1(right). We can see that the noise are reduced by comparing these two figures.

Fig.2(right) shows estimated power of source, artifacts, and noise on the sensors. The powers of source and artifacts are estimated as the components of matrix $\hat{A}_{\text{ICA}}$. Also the power of noise is estimated as the diagonal component of the matrix $\hat{\Sigma}$. From the graph, we can see that some sensors contain much more noise and artifacts than the signal even after averaging over 100 trials.

3.2 Experiment with brain data

We applied our algorithm to the data of brain activity evoked by visual stimulation. The expected results of ICA for MEG data analysis would be summarized as 1) Separating artifacts and brain signals. and 2) The brain activities from different parts to be separated.

We believe that 1) is possible because the artifacts and the brain signals would be independent. But 2) is difficult since it is more natural to think that even though the signals are coming from different parts of the brain, they might be dependent.

First, we show the averaged data in Fig.3(left). A kind of visual stimulations are given to a subject. The data are recorded by 114 sensors in this case (because only 114 of 129 sensors were working correctly). The duration of recording is from 100msec before and to 412msec after the stimulation with 1kHz sampling rate. The same procedure is applied to the subject for 100 times and we averaged the data. Three of the sensors are shown in Fig.3(left). It is observed that there are still a lot of noise after averaging.

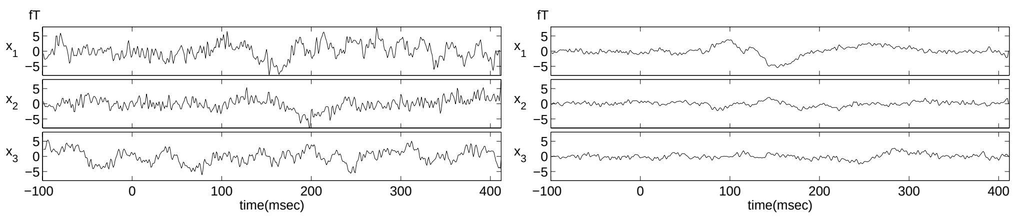

Figure 3: Averaged data(left), and after removing artifacts(right)

We applied our method to the data and 17 independent components were selected by MDL in this experiment. The independent components are shown in Fig.4. We also applied the method to the data from different subjects (4 more), and in all the cases, the selected numbers of sources are roughly the same (from 16 to 19).

Based on the results in Fig.4, we have to separate the sources from the artifacts in which we are not interested. For example, we can see $y_9(t)$ is mainly a high frequency (180Hz) signal which seems to be an artifact. And $y_{10}(t)$ has a very large value at the very end of the record which seems to be some software noise. But for the other 15 components, it is difficult to know if they are brain sources or not. Fortunately, this experiment is a study of evoked response by visual stimuli, and we are not interested in the components

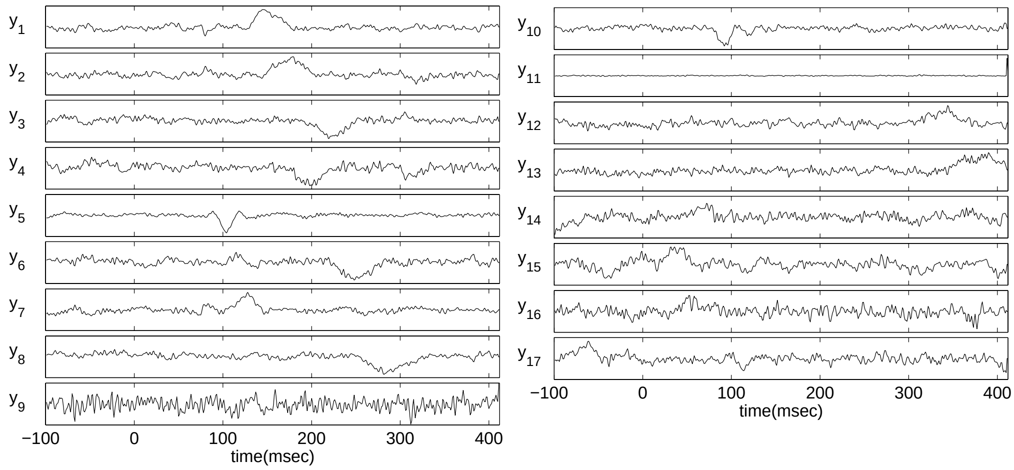

Figure 4: Estimated independent components of visual evoked

which have some power before the stimuli. Therefore, we defined a threshold such that, if a signal has some power before the stimulation, we regarded the signal as an artifact. In this experiment, $y_1(t)$, $y_2(t)$, $y_3(t)$, $y_5(t)$, $y_6(t)$, $y_8(t)$, and $y_{10}(t)$, are selected as the brain signals.

After picking those sources up, we put them back to the orignal sensor signal space, and the result is shown in Fig.3 (right). It looks like that the noise are removed and the data are clear.

We can see the cleaned outputs of the sensors from this result, but we further want to know the relationship between the independent components and brain activities more directly. For the visualization of the result, we implemented SF (spatial filter) technique[10]. SFs are a set of virtual sensors which are located on a hemisphere defined on the brain. We can estimate the current flows on those virtual sensors which describe the MEG observations well. This is an inverse problem and it can be solved in a form of a linear mapping from the MEG sensors to SFs. In our experiment, the part of the brain we are interested in is the visual cortex, and we put 21×21 SFs on a hemisphere, whose center is located at V1.

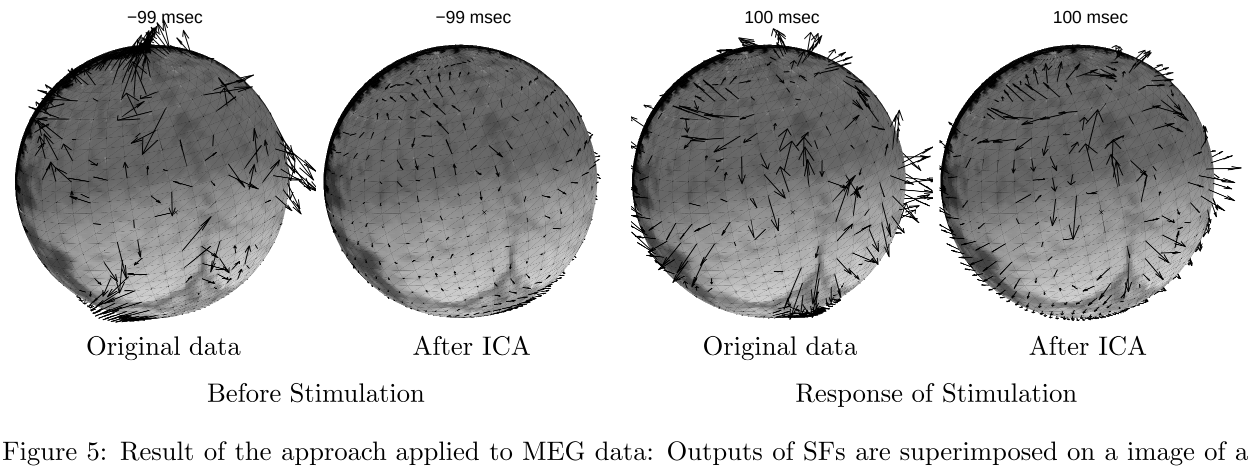

Figure 5: Result of the approach applied to MEG data: Outputs of SFs are superimposed on a image of a brain obtained by MRI

Fig.5 shows the output of the SFs before and after the method. We recorded the response of a subject from 100msec before the visual stimulation to 412msec after the stimulation. The original data includes a lot of noise even before the stimulation and we can remove them very well. The response of the brain is known to be high around 100msec after the stimulation[10], and the characteristics is preserved very well.

4 Discussion

In this article, we proposed a new approach of ICA to noisy data. We applied the algorithm to MEG data, and have shown the approach is effective. We can estimate the number of the sources, and the power of the noise on each sensor which is independent to each other. This is one of the serious problem which has not been well treated in conventional ICA approaches, and this article gives one effective approach. H. Attias[2] has proposed one method to solve this problem. It gives the source distribution by the form of normal mixture and everything is solved by using the parameters in the sense of MLE. But we have proposed a different method based on the semi parametric approach which is one attractive point of ICA.

Although, there still remain a lot of open problems. In the factor analysis, there are a lot of methods to estimate the parameters and the number of the sources, and each of them has each characteristics. We applied MLE for estimation and MDL for estimating the number of the sources. But there are different combinations, and there might be a method which suits better for some particular problems. Also the same thing may be true for the ICA algorithms. We used JADE but there might be a better algorithm. Another problem is the noise distribution. We assumed Gaussian distribution, but if we can have a better model, the algorithm will be improved further.

Our approach also gives a new concept for factor analysis. How to determine the rotation is one of the big problems in factor analysis. And it is not common to use higher order statistics. Therefore our approach gives a new pathway to factor analysis, too.

Acknowlegement

We are grateful to Shun-ichi Amari, Noboru Murata, and Jianting Cao for their comments and suggestions for our research. We also thank Shigeki Kajihara and Shimadzu Inc. for the MEG data.

References

[1] S. Amari and J.-F. Cardoso. Blind source separation – semiparametric statistical approach. *IEEE Trans. Signal Processing*, 45(11):2692–2700, nov 1997.

[2] H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, May 1999.

[3] A. J. Bell and T. J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[4] J. Cao, N. Murata, S. Amari, A. Cichocki, and T. Takeda. ICA approach with pre & post-processing techniques. In *Proceedings of 1998 International Symposium on Nonlinear Theory and its Applications (NOLTA'98)*, volume 1, pages 287–290, September 1998.

[5] J.-F. Cardoso. Higher-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, Janurary 1999.

[6] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian siganals. *IEE-Proceedings-F*, 140(6):362–370, December 1993.

[7] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, apr 1994.

[8] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, July 1997.

[9] W. Ledermann. On the rank of the reduced correlational matrix in multiple-factor analysis. *Psychometrika*, 2:85–93, 1937.

[10] K. Toyama, K. Yoshikawa, Y. Yoshida, Y. Kondo, S. Tomita, Y. Takanashi, Y. Ejima, and S. Yoshizawa. A new method for magnetoencephalography: A three dimensional magnetometer-spatial filter system. *Neuroscience*, 91(2):405–415, 1999.

[11] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in Magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems*, volume 10, pages 229–235. The MIT Press, Cambridge MA, 1998.