A Shift Invariance-Based Order-Selection Technique for Exponential Data Modelling

Jean-Michel Papy, Lieven De Lathauwer, and Sabine Van Huffel

***Abstract*—This paper presents a new subspace-based technique for automatic detection of the number of exponentially damped sinusoids. It consists in studying the shift-invariance of the dominant subspace of the Hankel data matrix. No threshold setting and no penalization terms are necessary. This model-based method, easy to implement, can be plugged into most subspace-based harmonic retrieval algorithms.**

***Index Terms*—Exponentially damped sinusoids, harmonic retrieval, order selection, singular value decomposition, subspace, total least squares.**

I. Introduction

THE exponentially damped sinusoids (EDS) model is a very basic model which is ubiquitous in signal processing. It has, among others, applications in nuclear magnetic resonance (NMR) [1], [2], material engineering [3], and audio processing [4]. Generally, the single-channel noise-free signal model is written in a complex form

$$x_n = \sum_{k=1}^{K} c_k z_k^n \tag{1}$$

where z_k is the kth pole of the signal, c_k denotes the kth associated complex amplitude, K is the model order, and n is the sample time index. More information about the model can be found in [1]. The popularity of subspace-based techniques for the estimation of the model parameters relies on the fact that this model lends itself to subspace representation. The data x_n in (1) are first mapped to an $L \times M$ Hankel matrix $\boldsymbol{H}$, as follows:

$$\boldsymbol{H} = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{M-1} \\ x_1 & x_2 & x_3 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{L-1} & \cdots & \cdots & x_{N-2} & x_{N-1} \end{pmatrix} \tag{2}$$

with $\{L, M\}$ chosen such that $N = L + M - 1$, $L > K$, and $M \geqslant K$. This is the starting point of popular subspace-based parameter estimation algorithms such as Tufts and Kumaresan's linear prediction method [16] or the HSVD algorithm [15]. The total least squares (TLS) version of the latter algorithm called the HTLS algorithm [1], [2] is often referred to as a special case of the ESPRIT algorithm [5] and also a TLS variant of Kung *et al.*'s algorithm [6]. In the noise-free case, $\boldsymbol{H}$ has a *Vandermonde decomposition* (VDMD)

$$\boldsymbol{H} = \boldsymbol{S}_{L,K}\, \boldsymbol{C}\, \boldsymbol{S}_{M,K}^T \tag{3}$$

where

$$\boldsymbol{S}_{L,K} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_K \\ \vdots & \vdots & \cdots & \vdots \\ z_1^{L-1} & z_2^{L-1} & \cdots & z_K^{L-1} \end{pmatrix} \tag{4}$$

$$\boldsymbol{C} = \operatorname{diag}\{c_1, c_2, \ldots, c_K\} \tag{5}$$

and ^{T} denotes the transpose. In the remainder, we will refer to the Vandermonde matrix $\boldsymbol{S}_{L,K}$ as $\boldsymbol{S}$. Notice two important properties:

- $\boldsymbol{H}$ is rank-K and its column space is exactly spanned by the set of K Vandermonde vectors $\boldsymbol{S}$;
- $\boldsymbol{S}$ has the following shift-invariance property:

$$\boldsymbol{S}^{\uparrow} = \boldsymbol{S}_{\downarrow} \boldsymbol{D} \tag{6}$$

  where $^{\uparrow}$ (respectively, $_{\downarrow}$) means that the top row (respectively, the bottom row) has been deleted and where $\boldsymbol{D} = \operatorname{diag}\{z_1, \ldots, z_K\}$.

In practice, one cannot directly find the K Vandermonde vectors. Instead, one estimates the subspace spanned by the Vandermonde vectors and use the shift-invariance property of $\boldsymbol{S}$ to retrieve the set of signal poles $\{z_k\}_{k=1}^K$. A common way to obtain the subspace is to compute the *singular value decomposition* (SVD) of the matrix $\boldsymbol{H}$

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H \tag{7}$$

Manuscript received August 17, 2006; revised October 10, 20006. This work was supported by the Research Council of the Katholieke Universiteit Leuven (GOA-MEFISTO 666, GOA-AMBioRICS), the Flemish Government (FWO: projects G.0407.02/G.0269.02/G.0360.05, research communities ICCoS and ANMMM), the Belgian Federal Government [DWTC: IUAP IV-02 (1996–2001) and IUAP V-22 (2002–2006)], and the EU (PDT-COIL: contract NNE5/2001/887; BIOPATTERN: contract FP6–2002-IST 508803; eTUMOUR: contract FP6–2002-LIFESCIHEALTH 503094). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Olivier Besson.

J.-M. Papy is with the Flander's Mechatronics Technology Center, Celestijnenlaan 300D, 3001 Leuven-Heverlee, Belgium (e-mail: jean-michel.papy@fmtc.be).

L. De Lathauwer is with the Laboratory ETIS, UMR 8051 (ENSEA, UCP, CNRS), 95014 Cergy-Pontoise Cedex, France (e-mail: delathau@ensea.fr).

S. Van Huffel is with the Katholieke Universiteit Leuven, ESAT, Division SCD-SISTA, 3001 Leuven-Heverlee, Belgium (e-mail: Sabine.VanHuel@esat.kuleuven.be).

Color versions of Figs. 1 and 2 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2006.891324

1070-9908/$25.00 © 2007 IEEE

where $\boldsymbol{U} \in \mathbb{C}^{L\times\min(L,M)}$ and $\boldsymbol{V} \in \mathbb{C}^{M\times\min(L,M)}$ are orthogonal matrices, called singular vector matrices, and $\boldsymbol{\Sigma} \in \mathbb{R}^{\min(L,M)\times\min(L,M)}$ is the diagonal singular value matrix. The superscript ^{H} denotes the Hermitian transpose. The relevant vectors are those related to the nonzero singular values. In the ordered SVD they correspond to the K leftmost column vectors of the singular vector matrix $\boldsymbol{U}$. When the data are subject to some additive white noise, the K-dimensional dominant subspace of $\boldsymbol{H}$ may still yield a good approximation of $\mathrm{span}\{\boldsymbol{S}\}$. However in this case, the $\min(L, M) - K$ trailing singular values are related to the so-called noise subspace, and are not zero anymore. Hence, defining K, the size of the dominant subspace, becomes a difficult task. Consequently, without prior knowledge of K, no reliable estimation of the signal poles is possible. While underestimation of the model order yields biased estimates [7], overestimation will result in fitting the noise to EDS's. Therefore, the determination of the model order is a key-step in all high-resolution methods. A simple method consists of determining the numerical rank of $\boldsymbol{H}$ by looking at the gap between singular values. However this is a very rough method. More refined model order-selection methods are based on Information Theoretic Criteria (ITC), such as the Akaike Information Criterion (AIC) [8], the Minimum Description Length (MDL) [8], or the Efficient Detection Criterion (EDC) [9]. These so-called penalized likelihood methods assume that the noise is Gaussian and use this property to determine how many singular values are related to the noise subspace. Moreover, for short data records the ITC are no longer optimal. More recent methods have been proven to perform better. Among the most popular ones, there is the Bayesian method (BAY) [12], the maximum *a posteriori* method (MAP) [11], an approach based on constant false alarm rate (CFAR) hypothesis tests [14], and the method using a least square (LS) error function introduced by Cheng and Hua [13]. All these methods perform well on small datasets, but use a statistical approach assuming a Gaussian distribution of the noise. Moreover in [13] a confidence interval has to be chosen by the user and the noise variance has to be known. The Bayesian method [12], as well as the CFAR method [14], also make use of a user-chosen parameters. This makes these methods very difficult to automate. Finally, a common drawback of the methods previously cited is that they assume an undamped model (pure sinusoids or cisoids). Recently, a model-based method [7], called ESTimation ERror (ESTER), was developed to automatically find the order of a EDS model. Compared to the statistical approaches, ESTER has the great advantage of not assuming that the noise is Gaussian. It does not require any penalization term or any threshold since it is directly based on the underlying model. The signal may contain damped and undamped complex (or real) sinusoids. The ESTER method merely assumes that the noise and the signal subspace are sufficiently separated, which implies that the noise has to be white and zero-mean. It has been proven to be robust against noise and outperforms the ITC cited above. In this paper, we present a simple SVD-based method, easy to implement, which is shown to be more general, flexible and robust than ESTER. It mainly relies on the shift-invariance property of $\boldsymbol{S}$. The paper is organized as follows. In Section II, we summarize the ESTER method and we explain briefly some of its important properties. In Section III, we derive our new method. In Section IV, the two methods are compared by means of Monte-Carlo simulations, and in Section V, we conclude the paper.

II. Subspace Shift-Invariance Property and Least-Squares Approach

For exponential data modeling the shift-invariance of the column subspace of $\boldsymbol{H}$ is crucial. This property is induced by the shift-invariance of the Vandermonde matrix $\boldsymbol{S}$. In the noise-free case, the K first column vectors of $\boldsymbol{U}$ exactly span the same subspace as the vectors of $\boldsymbol{S}$ and hence there exists a square nonsingular matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{S} = \boldsymbol{U}_K\boldsymbol{Q} \tag{8}$$

where $\boldsymbol{U}_K = [U_1 U_2 \cdots U_K]$ is the semi-unitary matrix containing the K first columns of $\boldsymbol{U}$. From (6), it follows that

$$\boldsymbol{U}_K{}^{\uparrow} = \boldsymbol{U}_{K\downarrow}\boldsymbol{\Phi}(K) \tag{9}$$

where $\boldsymbol{\Phi}(K) = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{-1}$ is an unknown square nonsingular matrix whose eigenvalues are the signal poles. Let us call a property generic when it holds with probability one when the parameters of the problem are drawn from a continuous probability density function. The equality in (9) generically only holds for the true order K. Considering a subspace of size $k \neq K$ leads to an inconsistent set of equations

$$\boldsymbol{U}_k{}^{\uparrow} \approx \boldsymbol{U}_{k\downarrow}\boldsymbol{\Phi}(k)\,. \tag{10}$$

The ESTER method developed by Badeau *et al.* is based on this observation. ESTER consists of computing the residual error of the set $\boldsymbol{U}_k{}^{\uparrow} = \boldsymbol{U}_{k\downarrow}\boldsymbol{\Phi}(k)$ in the LS sense

$$E(k)^2 = \left\|\Delta\boldsymbol{U}_k{}^{\uparrow}\right\|_2^2 = \left\|\boldsymbol{U}_{k\downarrow}\widehat{\boldsymbol{\Phi}}(k) - \boldsymbol{U}_k{}^{\uparrow}\right\|_2^2 \tag{11}$$

with

$$\widehat{\boldsymbol{\Phi}}(k) = \left(\boldsymbol{U}_{k\downarrow}\right)^{\dagger}\boldsymbol{U}_k^{\uparrow} \tag{12}$$

where † denotes the pseudo-inverse. Indeed, in the noise-free case, the system has generically only an exact solution for $k = K$. Moreover, it was shown by means of simulations that, in the presence of noise, the number of columns k for which $E(k)^2$ is minimal, is still K, the order of the EDS model. The estimation of K can then be formulated as follows:

$$K = \arg\max_k \frac{1}{E(k)^2},\ 1 \leq k < \min(L, M). \tag{13}$$

III. SVD-Based Approach

In the presence of noise, always both $\boldsymbol{U}_{k\downarrow}$ and $\boldsymbol{U}_k{}^{\uparrow}$ are perturbed. Consequently, the LS approach is suboptimal in the sense that it supposes that only $\boldsymbol{U}_k{}^{\uparrow}$ is perturbed. Therefore, we will jointly consider both the subspaces $\mathrm{span}(\boldsymbol{U}_{k\downarrow})$ and

$\mathrm{span}(\boldsymbol{U}_k{}^{\uparrow})$ by starting from the matrix $\boldsymbol{U}_k^{\mathrm{tb}} = \left[\boldsymbol{U}_k{}^{\uparrow}|\boldsymbol{U}_{k\downarrow}\right]$. Indeed, this $(L-1)\times 2K$ matrix has some particular properties that are related to the shift-invariance property of $\boldsymbol{S}$. We first investigate the noise-free case and we will then discuss the presence of additive white noise. The subspace shift-invariance property of $\boldsymbol{H}$ described by (9), simply implies that the column vectors of $\boldsymbol{U}_K{}^{\uparrow}$ span exactly the same subspace as the column vectors of $\boldsymbol{U}_K{}^{\downarrow}$. Therefore, in the special case where $k = K$, we have that $\boldsymbol{U}_k^{\mathrm{tb}}$ is a rank-K matrix. Consequently, its last K singular values $\gamma_{K+1},\ldots,\gamma_{2K}$ are equal to zero. Assume $L-1 \geq 2K$ and let

$$\boldsymbol{U}_K^{\mathrm{tb}} = \boldsymbol{Y}\Gamma\boldsymbol{W}^H \tag{14}$$

be the SVD of $\boldsymbol{U}_K^{\mathrm{tb}}$, where $\boldsymbol{Y} \in \mathbb{C}^{(L-1)\times 2K}$, $\boldsymbol{W} \in \mathbb{C}^{2K\times 2K}$, and $\boldsymbol{\Gamma} = \mathrm{diag}\{\gamma_1,\ldots,\gamma_{2K}\}$, then we have $\{\gamma_i\}_{i=K+1}^{2K} = 0$. We will now investigate the case $k \neq K$ to see what happens to the rank in case of over- and underestimation. Our observations will lead to a new algorithm for the estimation of K.

A. Overestimation of the Model Order

This is the case where $k > K$. Consider

$$\boldsymbol{U}_{K+p}^{\mathrm{tb}} = \left[\boldsymbol{U}_K{}^{\uparrow}|U_{K+1}{}^{\uparrow}\ldots U_{K+p}{}^{\uparrow}|\boldsymbol{U}_{K\downarrow}|U_{K+1\downarrow}\ldots U_{K+p\downarrow}\right].$$

In general, the subspace spanned by $U_{K+1},\ldots,U_{K+p}$ has no shift invariance properties. Neither does the subspace spanned by $U_{K+1}{}^{\uparrow},\ldots,U_{K+p}{}^{\uparrow},U_{K+1\downarrow},\ldots,U_{K+p\downarrow}$ have a component in the column space of $\boldsymbol{U}_K{}^{\uparrow}$ (or $\boldsymbol{U}_{K\downarrow}$). This means that, in general, adding a column to $\boldsymbol{U}$ increases the rank of $\boldsymbol{U}^{\mathrm{tb}}$ by two. Then we have the following theorem, the proof of which is given in an extended version of the paper [10].

Proposition 1: Generically, $\boldsymbol{U}_{K+p}^{\mathrm{tb}}$, with $p \in [1\ldots\min(L,M)-K]$, is a rank-$(K+2p)$ matrix for $K+2p \leqslant L-1$ and a rank-$(L-1)$ matrix for $K+2p \geqslant L-1$.

B. Underestimation of the Model Order

This is the case where $k < K$. In general, the k columns of $\boldsymbol{U}_k$ contain contributions of all K Vandermonde vectors. This makes the columns of $\boldsymbol{U}_k{}^{\uparrow}$ and $\boldsymbol{U}_{k\downarrow}$ span a $(2k)$-dimensional subspace, if $2k \leq K$, or a K-dimensional subspace, if $2k \geq K$. We have the following theorem, the proof of which is given in an extended version of the paper [10].

Proposition 2: If $k < K$, the matrix $\boldsymbol{U}_k^{\mathrm{tb}}$ is generically of rank $\min(2k, K)$.

C. Order Determination

We have seen that, in the noise-free case, the $(L-1)\times 2k$ matrix $\left[\boldsymbol{U}_k{}^{\uparrow}\,\boldsymbol{U}_{k\downarrow}\right]$ is rank-k only for $k = K$. This implies that, for $k = K$, the k last singular values of $\boldsymbol{U}_k^{\mathrm{tb}}$ are zero. However, in the presence of noise, the column vectors of $\boldsymbol{U}$ are perturbed and the matrix $\left[\boldsymbol{U}_k{}^{\uparrow}\,\boldsymbol{U}_{k\downarrow}\right]$ is never rank deficient, for any k. If the SVD of $\boldsymbol{H}$ allows a good separation of the noise and the signal subspace then the k last singular values of $\left[\boldsymbol{U}_k{}^{\uparrow}\,\boldsymbol{U}_{k\downarrow}\right]$ will still be small while the reasoning in the previous sections still holds. This means that K can still be estimated by inspection of the amplitude of the k smallest singular values. We propose the following model order estimator:

$$\hat{K} = \arg\max_k \frac{1}{E(k)},\ 1 \leq k < \min\left(\mathrm{floor}\left(\frac{L-1}{2}\right), M\right) \tag{15}$$

with

$$E(k) = \frac{1}{k}\sum_{i=k+1}^{2k}\gamma_i \tag{16}$$

where γ_i are the singular values of $\left[\boldsymbol{U}_k{}^{\uparrow}\,\boldsymbol{U}_{k\downarrow}\right]$. This method is called subspace-based automatic model order selection (SAMOS).

TABLE I
SET OF PARAMETERS FOR THE FIVE-PEAK EXAMPLE

	ν [Hz]	α [s^{-1}]	a [a.u.]
Peak 1	0.2	0.01	1
Peak 2	0.3	0.02	1
Peak 3	-0.2	0.1	2
Peak 4	0.4	0.05	1
Peak 5	0.35	0.03	1

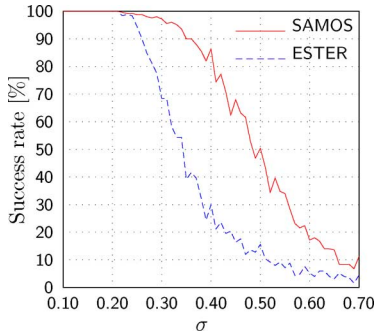

Fig. 1. Comparison of the performance in terms of the probability of correct determination of the true model order ($K = 5$).

IV. Results

In the first simulation, we generated a signal of 129 samples based on a five poles model whose parameters are given in Table I. In order to have a good separation between the noise and the signal subspace, it is preferable to use a Hankel matrix that is as square as possible [1]. Therefore, we arranged the signal in a 65×65 Hankel matrix. For the same reason, the dimension of the subspace spanned by the Vandermonde vectors should not be too high. Although we do not want to use any prior knowledge about the model order, it is reasonable not to consider $k > L/4$. Therefore we set the maximum order to $k = 16$. The comparison is performed by means of Monte Carlo simulations consisting of 250 independent realizations for each noise level. The result is expressed in terms of the percentage of runs in which the model order was successfully determinated as a function of the noise standard deviation σ. Fig. 1 compares the performance of the two subspace-based algorithms. It is clear that the new SVD-based rank detection

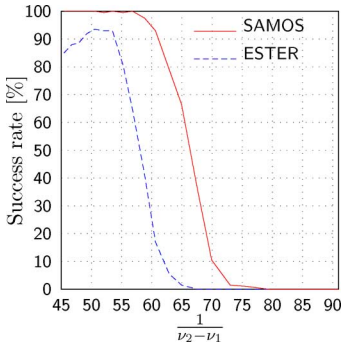

Fig. 2. Comparison of the performance in terms of the probability of correct determination of the true model order ($K = 2$).

algorithm yields a better performance; the probability of correct detection is higher over the full noise level range. To take a concrete example, ESTER achieves a 90% success rate for $\sigma = 0.25$, while the new method SAMOS has the same performance for $\sigma = 0.35$, which represents a gain of about 3 dB. We recall that ESTER already outperforms ITCs. The next simulation compares the robustness of the two methods when two signal poles are getting very close. We use a two-pole model $x_n = \exp\{(-0.02 + 2j\pi 0.2)n\} + \exp\{(-0.02 + 2j\pi\nu_2)n\}$, $n = 0, \ldots, 64$ ($\Delta t = 1$), where the frequency ν_2 of the second pole varies between 0.222 and 0.205. The noise standard deviation is set to 0.4. The result is displayed in Fig. 2. We observe again a higher robustness of the new method. Also, in this example, ESTER never achieves a 100% success rate.

V. Conclusion

We have presented a subspace-based method for determining the number of exponentially damped sinusoids present in a signal. This method, called SAMOS, which merely assumes a good separation of the signal and noise subspace, is based on the shift-invariance property of the dominant subspace of the Hankel data matrix. It has been successfully compared to the ESTER method, which was also based on the shift invariance property. A following step in this research is to investigate whether exploiting knowledge of the noise distribution allows to further improve the results.

References

[1] S. Van Huffel, "Enhanced resolution based on minimum variance estimation and exponential data modeling," *Signal Process.*, vol. 33, no. 3, pp. 333–355, 1993.

[2] S. Van Huffel, H. Chen, C. Decanniere, and P. Van Hecke, "Algorithm for time-domain NMR data fitting based on total least squares," *J. Mag. Res.*, vol. A110, pp. 228–237, 1994.

[3] L. Rippert, "Optical Fiber for Damage Monitoring in Carbon Fiber Reinforced Plastic Composite Materials," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, 2003.

[4] W. Verhelst, K. Hermus, P. Lemmerling, P. Wambacq, and S. Van Huffel, "Modeling audio of damped sinusoids using total least squares algorithm," in *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, S. Van Huffel and P. Lemmerling, Eds. Dordrecht, The Netherlands: Kluwer, 2002, pp. 331–340.

[5] R. Roy and T. Kailath, "ESPRIT – estimation of signal parameters via rotational invariance technique," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[6] S. Y. Kung, K. S. Arun, and D. V. B. Rao, "State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc. Amer.*, vol. 33, no. 12, pp. 1799–1811, 1983.

[7] R. Badeau, B. David, and G. Richard, "Selecting the modeling order for the ESPRIT high resolution method: An alternative approach," in *Proc. Int. Conf. Acoust. Speech Signal Processing*, 2004, vol. II, pp. 1025–1028.

[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech Signal Process.*, vol. SP-33, no. 2, pp. 387–392, Apr. 1985.

[9] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multi. Anal.*, vol. 20, no. 1, pp. 1–25, 1986.

[10] J.-M. Papy, L. De Lathauwer, and S. Van Huffel, A Shift Invariance-Based Order Selection Technique for Exponential Data Modelling, Tech. Rep. [Online]. Available: ftp://www.esat.kuleuven.ac.be/pub/sista/papy/reports/jmp-05-107.ps.gz

[11] P. M. Djurić, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1744–1751, Jul. 1996.

[12] P. M. Djurić, "Simultaneous detection and frequency estimation of sinusoidal signals," in *Proc. Int. Conf. Acoust. Speech Signal Processing*, 1993, vol. IV, pp. 53–56.

[13] Q. Cheng and Y. Hua, "Detection of cisoids using least square error function," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1584–1590, Jul. 1997.

[14] A. A. Shah and D. W. Tufts, "Determination of the dimension of a signal subspace from short datarecords," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2531–2535, Sep. 1994.

[15] H. Barkhuysen, R. de Beer, and D. van Ormondt, "Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals," *J. Magn. Reson.*, vol. 73, pp. 553–557, 1987.

[16] D. W. Tufts and R. Kumaresan, "Singular value decomposition and improved frequnecy estimation using linear prediction," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-30, no. 4, pp. 671–675, 1982.