Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physics Letters A 336 (2005) 448–458
PHYSICS LETTERS A
www.elsevier.com/locate/pla

Estimation of parameters in one-dimensional maps from noisy chaotic time series

Dmitry A. Smirnov [a,*], Vladislav S. Vlaskin [b], Vladimir I. Ponomarenko [a,b]

[a] *Saratov Department, Institute of Radioengineering and Electronics of the Russian Academy of Sciences, 38, Zelyonaya Street, 410019 Saratov, Russia*

[b] *Department of Nonlinear Processes, Saratov State University, 83, Astrakhanskaya Street, 410012 Saratov, Russia*

Received 24 August 2004; received in revised form 16 December 2004; accepted 17 December 2004

Available online 21 January 2005

Communicated by C.R. Doering

Abstract

The problem of parameter estimation in model maps from noisy time series is addressed. We suggest a new technique for a special case of one-dimensional maps and chaotic signals. It is based on the maximum likelihood (ML) principle and evaluation of the cost function via *backward* iterations of a model map. We demonstrate in numerical experiments and, in part, justify theoretically that this "backward ML technique" gives more accurate estimates than previously known techniques for low and moderate noise levels. In particular, global optimisation of the cost function becomes much easier; biases in the estimates vanish as the time series length N increases; variances of the estimates decrease as fast as $N^{-\alpha}$ where α depends on the original system, typical values being about $\alpha = 2.0$ under mild conditions on the original systems.

PACS: 05.45.Tp

Keywords: Time series analysis; Modelling from time series; Parameter estimation; One-dimensional maps; Deterministic chaos

1. Introduction

Apart from their basic importance, the methods of mathematical modelling from time series find numerous applications ranging from estimation of nonlinear characteristics of electric circuit elements [1] to the description of biochemical processes in cells [2]. The problem of modelling from time series is known as "system identification" [3] and "reconstruction of dynamical systems" [4,5]. Nowadays, approaches to its solution are developed within the framework of nonlinear dynamics [1,2,6–18], the use of maps being quite widespread [12–18] since even low-dimensional maps are capable of describing complex dynamics

* Corresponding author.
E-mail addresses: sbire@sgu.ru, smirnovda@info.sgu.ru (D.A. Smirnov).

0375-9601/$ – see front matter
doi:10.1016/j.physleta.2004.12.092

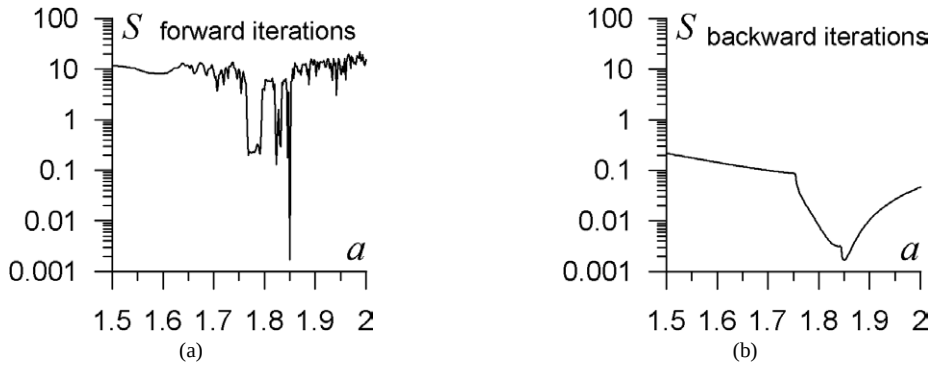

Fig. 1. The graphs of the cost functions (2) and (4) for the time series of the quadratic map (6) at $N = 20$, $a^{(0)} = 1.85$, and $x_1^{(0)} = 0.3$. We specified $x_1 = x_1^{(0)}$ in (2) and $x_N = x_N^{(0)} \equiv f^{(N-1)}(x_1^{(0)})$ in (4).

of many real-world systems. Thus, the Poincaré map of a strongly dissipative nonlinear flow is often one-dimensional to a high accuracy, e.g., the Lorenz and Rössler systems [19].

Our work is devoted to parameter estimation in a one-dimensional map from a noisy chaotic time series. This problem was considered in a series of works [20–24] where different estimation techniques were suggested. Each of them involves optimisation of some cost function and is related to a certain extent to the ML principle. Theoretically, a longer time series must provide opportunity to get more accurate estimates. However, even the ML method encounters serious practical difficulties when it is applied to chaotic signals. The cause is that exponential sensitivity of a chaotic map orbit to initial conditions and parameters results in a very complicated, jagged "relief" of the cost function graph, see Fig. 1(a) for an illustration and [24] for details. Hence, the optimisation problem becomes practically unsolvable, especially for multiple estimated parameters. Therefore, modification of the method was suggested in [24] and called "piecewise" ML technique. It consists in segmentation of the time series and averaging the estimates obtained from different segments. Despite such a modified approach seems to be the most efficient among all others for estimation from chaotic signals, it is not "perfect" also since the length of the segments must be sufficiently small that inevitably leads to bias in the estimates. This bias may not vanish even when the time series length N tends to infinity. Besides, the variance of the estimates decreases with N only as $1/N$ like for a usual regression problem.

In the present Letter for the particular case of one-dimensional maps we suggest a new technique which is also a special version of the ML method. It is based on the use of *backward iterations* of a map to evaluate the cost function. Such an idea was exploited earlier for nonlinear noise reduction [25] and separation of the sum of chaotic signals into its components [26]. The new technique allows the use of the entire chaotic time series without dividing it into segments due to typically simple relief of the cost function graph, see Fig. 1(b) for an illustration. As we show below, apart from the ease of global optimisation, the new estimates are more accurate for moderate noise. They are almost unbiased and their variances often decrease with N faster than $1/N$.

The Letter is organised as follows. In Section 2 we describe the "piecewise" and "backward" ML techniques. In Section 3 we present the results of their comparison in numerical experiments and formulate conditions for the superiority of the "backward" technique, while more rigorous mathematical justification of the results is presented in Appendix A. We summarise in Section 4.

2. Problem and methods

2.1. Problem formulation

Let us consider a one-dimensional map

$$x_{n+1} = f(x_n, \mathbf{a}), \tag{1}$$

where x is a dynamical variable, $\mathbf{a}$ is a parameter vector of dimensionality P. Let a time series be generated by the map (1) in a chaotic regime at some $\mathbf{a} = \mathbf{a}^{(0)}$ and $x_1 = x_1^{(0)}$ in the presence of additive observational noise. We denote the observed time series $\{\eta_1, \ldots, \eta_N\}$, where $\eta_n = x_n + \xi_n$, ξ_n is a sequence of i.i.d. Gaussian random values with zero mean and variance σ_ξ^2. The form of f is known, the values of $\mathbf{a}^{(0)}$, $x_1^{(0)}$, and σ_ξ^2 are unknown. It is necessary to get an estimate $\hat{\mathbf{a}}$ of the parameter vector, which is as close to the true value $\mathbf{a}^{(0)}$ as possible.

Throughout the Letter, we use a usual mean squared error (MSE) to quantify "goodness" of the estimates. Namely, the MSE in the estimate $\hat{a}$ of a single parameter a is $\varepsilon^2 = \langle(\hat{a} - a^{(0)})^2\rangle$ where angle brackets stand for mathematical expectation, i.e., for the average over different noise realisations at fixed $\mathbf{a}^{(0)}$ and $x_1^{(0)}$. As is well known from the theory of probabilities, the MSE can be expressed as $\varepsilon^2 = (\langle\hat{a}\rangle - a^{(0)})^2 + \sigma_{\hat{a}}^2$, where $(\langle\hat{a}\rangle - a^{(0)})$ is the bias in the estimate and $\sigma_{\hat{a}}^2$ is its variance. For multiple estimated parameters, $P > 1$, we define the total MSE as $\varepsilon^2 = \sum_{i=1}^{P} \varepsilon_i^2$ where ε_i^2 is the MSE in the individual estimate $\hat{a}_i$.

2.2. ML method and its "piecewise" version

Theoretically, one expects the ML method to be the most efficient estimation technique under certain "general" conditions, see [24] and references therein. It is tantamount to the least-squares method for Gaussian observational noise considered here. Namely, one minimises a cost function

$$S(x_1, \mathbf{a}) = \sum_{n=0}^{N-1} \bigl(\eta_{n+1} - f^{(n)}(x_1, \mathbf{a})\bigr)^2, \tag{2}$$

where $f^{(n)}$ is the nth iteration of the map f, $f^{(0)}(x, \mathbf{a}) \equiv x$.

If the observed regime is stable periodic, then the ML-method gives consistent, asymptotically unbiased, and efficient estimates. So, both biases and variances of the estimates vanish as $N \to \infty$. The variances $\sigma_{\hat{x}_1}^2$ and $\sigma_{\hat{a}_i}^2$ for moderate noise level can be estimated accurately as diagonal elements of the matrix $\sigma_\xi^2 \cdot J_N^{-1}$, where J_N is a cumulative Fisher information matrix [24]. E.g., for the simplest case of $P = 1$ and known $x_1^{(0)}$, J_N is just a real number and reads

$$J_N = \sum_{n=1}^{N} I_n^2, \tag{3}$$

where $I_n = \frac{\partial f^{(n-1)}(x_1,a)}{\partial a}\big|_{x_1=x_1^{(0)}, a=a^{(0)}}$. It can be shown that for a periodic regime $J_N \propto N$ and, hence, the variance $\sigma_{\hat{a}}^2$ scales as $\sigma_{\hat{a}}^2 \propto N^{-1}$, see Appendix A. Besides, the cost function graph is typically not complicated, so that a global minimum can be found easily. Everything is the same for $P > 1$ and unknown $x_1^{(0)}$. Seemingly, nothing better can be achieved in this case.

The situation is quite different for a chaotic regime. First of all, some of the "general" conditions mentioned above are not fulfilled, so that the ML estimates obtained via minimisation of (2) are no more guaranteed to be consistent [24]. The cause is that the cost function becomes fast oscillating and, perhaps, is not smooth asymptotically. Fig. 1(b) illustrates this for a very moderate N. So, the theoretical property of consistency and others are under the question. It would not be crucial in practice since for sufficiently large *finite* N the cost function (2) is smooth and biases and variances of the ML estimates are negligibly small. E.g., for a single unknown parameter a in the quadratic map and known $x_1^{(0)}$, the variance $\sigma_{\hat{a}}^2$ is shown in [23,24] to decrease exponentially with N. For unknown $x_1^{(0)}$, the variance $\sigma_{\hat{a}}^2$ does not decrease so fast that is explained in [24] by ill-conditioning of the matrix J_N. Nevertheless, one observes empirically that $\sigma_{\hat{a}}^2 \propto N^{-\alpha}$ with $\alpha = 3.2$ [24]. This relatively fast decrease in $\sigma_{\hat{a}}^2$ along with a very small bias would suffice to claim that the ML estimate is "practically consistent and efficient".

However, such efficiency takes place only if one succeeds in searching for the global minimum of (2). In practice, the minimisation is carried out with the aid of an iterative technique [27], where certain starting guesses for x_1 and $\mathbf{a}$ are generated and a "descent" to a minimum is realised stepwise. To find the global minimum of (2) for a long chaotic time series is practically impossible [24] since its relief exhibits variety of local minima, see illustration in Fig. 1(a).

The starting guesses determine which of the minima will be found. Usually, one tries big number M of starting guesses and chooses the deepest minimum among all minima found. Those guesses are taken from a certain domain where the true values of

the estimated quantities are expected to be located. One often uses a $(P+1)$-dimensional parallelepiped $\Delta x \times \Delta a_1 \times \Delta a_2 \times \cdots \times \Delta a_P$, where Δx is its projection onto the initial condition axis and Δa_i—onto the respective parameter axes. Let us call Δx and Δa_i "the intervals of starting uncertainty". To find the global minimum of (2) for a long chaotic time series is possible only for astronomically large M or extremely small Δx and Δa_i [22], both conditions being unrealistic.

According to "piecewise" ML technique, which seems the best approach known so far [24], one fixes certain reasonable M, Δx, and Δa_i, and divides the original time series into the segments of a moderate length L. Thus, it turns out practically possible to find the global minimum of the cost function (2) for each individual segment. The ultimate estimate $\hat{\mathbf{a}}_f$ is an average value over the estimates obtained from N/L individual segments. Since the estimates obtained from each segment may be significantly biased, $\hat{\mathbf{a}}_f$ is also probable to be biased. Technical details of the method implementation, such as the choice of L, are concerned in Section 3.

2.3. "Backward" ML technique

We suggest to use the ML principle in combination with backward iterations of a map, i.e., to minimise the cost function

$$S(x_N, \mathbf{a}) = \sum_{n=1}^{N} \bigl(\eta_{N-(n-1)} - f^{(-(n-1))}(x_N, \mathbf{a})\bigr)^2, \quad (4)$$

where $f^{(-n)}$ is the nth backward iteration of the map (1). The *only* Lyapunov exponent of the *one-dimensional* map (1) becomes negative under the time reversal. Thus, sensitivity of a backward orbit of the map to its "initial" condition x_N disappears. Therefore, one can expect much smaller number of the local minima for the cost function (4), see Fig. 1(b) for a numerical illustration.

Since a chaotic regime can be demonstrated only by a *nonmonotone* one-dimensional map, the problem of the nonuniqueness of the inverse map $f^{(-1)}$ arises inevitably. To implement backward iterations, one must decide which root of the equation

$$x_{n+1} - f(x, \mathbf{a}) = 0, \quad (5)$$

where the value x_{n+1} is the result of the previous backward iteration, should be taken as x_n. Ideally, one would choose the root closest to the true value $x_n^{(0)}$. Since $x_n^{(0)}$ is not observed, quite a natural and simple way is to take the root which is the closest to the corresponding observed value η_n. For the values of $\mathbf{a}$ and x_N close to the true values, such a choice can be sometimes erroneous due to observational noise. To avoid the erroneous choice, one can try to perform m backward iterations for every root of (5) thereby generating several alternative m-point "branches" $x_n, x_{n-1}, \ldots, x_{n-m+1}$ of the backward orbit. Finally, the branch closest to the observed segment $\eta_n, \eta_{n-1}, \ldots, \eta_{n-m+1}$ can be selected. The depth m cannot be very big in practice since the number of branches rises exponentially with m. We show in Section 3 that $m = 1$ is a good choice, while increase in m gives almost no gain, and explain this observation.

The minimisation of (4) is also performed numerically at some M, Δx, and Δa_i, but the technique appears insensitive to these settings as shown below. Therefore, we use $M = 1$, while Δx and Δa_i are taken to be the same as for $\hat{\mathbf{a}}_f$. Let us denote the new estimator as $\hat{\mathbf{a}}_b$.

3. Numerical experiment

Throughout this section we calculate the MSE in $\hat{\mathbf{a}}_f$ and $\hat{\mathbf{a}}_b$ via averaging over an ensemble of 1000 time series generated at fixed values of $\mathbf{a}^{(0)}$, $x_1^{(0)}$, N, and σ_ξ^2. $N = 100$, if not stated otherwise. The starting guesses are obtained as random values uniformly distributed in the respective intervals of starting uncertainty. All the calculations are performed in double precision.

3.1. Estimation of a single parameter

Time series is generated by the quadratic map

$$x_{n+1} = 1 - a x_n^2, \quad (6)$$

with $a^{(0)} = 1.85$ and $x_1^{(0)} = 0.3$, exactly as in [20–22, 24] to facilitate the comparison.

Piecewise ML technique We fixed somewhat arbitrarily $M = 10$, $\Delta x = [x_n^{(0)} - 0.1, x_n^{(0)} + 0.1]$, and

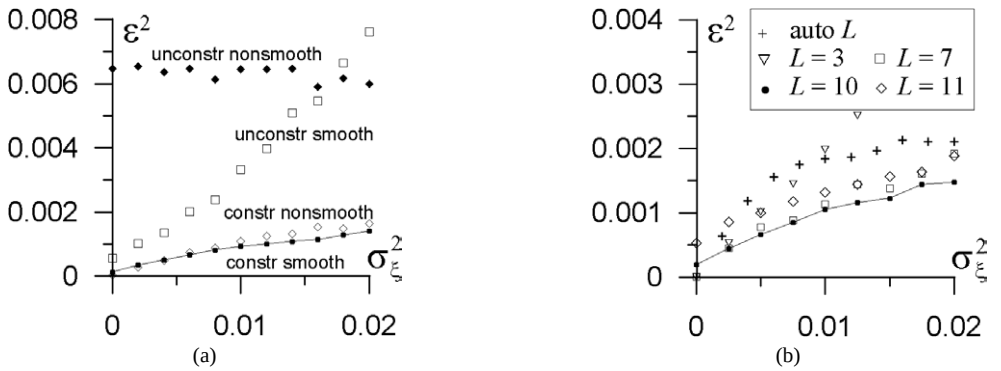

Fig. 2. MSE in the estimates $\hat{a}_f$ of the parameter a of the quadratic map (6) at $N = 100$: (a) for different minimisation techniques at $L = 10$; (b) for constrained smooth minimisation at different L.

$\Delta a = [1.5, 2.0]$. First, let us consider what minimisation technique to use. We performed unconstrained minimisation of (2) using both a smooth technique[1] and a nonsmooth one.[2] Besides, in practice one may be sometimes confident that the values of estimated quantities cannot lie outside of certain intervals, e.g., intervals of starting uncertainty Δx and Δa. Then, it is natural to perform constrained minimisation by enforcing $\hat{x}_1$ and $\hat{a}_f$ to belong to Δx and Δa, respectively. Thus, we used also constrained versions of smooth[3] and nonsmooth[4] techniques. We compare the obtained estimates $\hat{a}_f$ for the noise level $\sigma_\xi^2 = 0.01$ and the segment length $L = 10$ in Fig. 2(a). All the techniques give comparable results except for the unconstrained nonsmooth one which is significantly worse. Constrained techniques give lower MSE since the constraints decrease the variance of the estimates. The constrained nonsmooth technique seems to be the best one for the case considered and, therefore, it is used further.

Selection of the segment length L is performed also via trials and errors. In Fig. 2(b) the MSE for different L are presented. The length $L = 10$ provides the smallest MSE for a wide range of noise levels. It can be explained as follows. For smaller L, e.g., $L = 3$, the global minimum of (2) can be found easily, but since the segment is short, the exact location of the minimum strongly fluctuates around the true value $a^{(0)}$ due to noise. The latter results in a relatively large variance of the estimates and, hence, a large MSE. For very long segments, e.g., $L = 20$, local minima are found very often since the cost function graph becomes jagged, that leads again to a large MSE. Therefore, there is an optimal L in between, when local minima are found only rarely. We note also that it is very difficult to choose the best segment length L in practice when one has only a single time series. One opportunity is to choose L intuitively. The second opportunity is to choose such L that gives the minimal empirical variance of the individual estimates over N/L segments as mentioned in [24]. The results for the latter (automatic) choice of L are moderately good and are shown in Fig. 2(b) with crosses: MSE is higher than for $L = 10$ but is comparable to "worse" L. Further we always use the best $L = 10$ to compare good version of the piecewise ML technique of [24] to our backward ML technique.[5]

[1] Quasi-Newton method with finite-difference gradient and BFGS formula as implemented in the subroutine UMINF/DUMINF of the IMSL library, default settings.

[2] Direct search polytope algorithm as implemented in the subroutine UMPOL/DUMPOL of the IMSL library, default settings, tolerance for the relative error is specified to be 10^{-7}.

[3] Quasi-Newton method with finite-difference gradient, BFGS formula, and active set strategy as implemented in the subroutine BCONF/DBCONF of the IMSL library, default settings.

[4] Direct search complex algorithm as implemented in the subroutine BCPOL/DBCPOL of the IMSL library, default settings, tolerance for the relative error is specified to be 10^{-7}.

[5] Pisarenko and Sornette identified $L = 4$ as the optimal segment length [24]. However, we believe that this holds true only for very low noise levels.

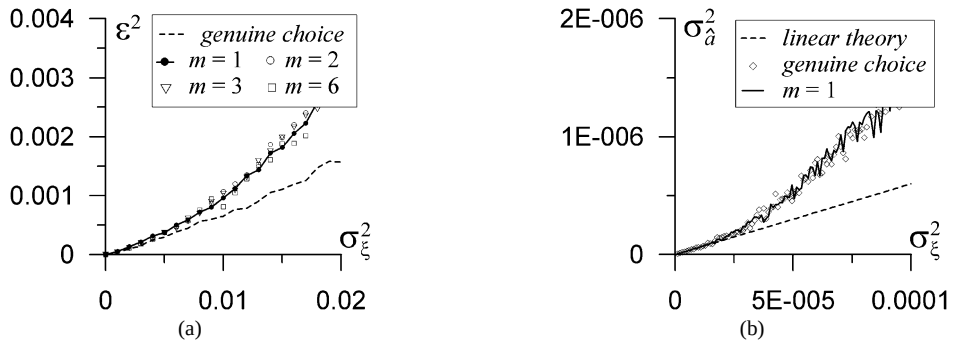

Fig. 3. (a) MSE in the estimates $\hat{a}_b$ of the parameter a of the quadratic map (6) at $N = 100$ for different depths m of backward branches and genuine choice of the root of Eq. (5). (b) Variances of $\hat{a}_b$ and approximate "linear" theory (the dashed line) at low noise levels.

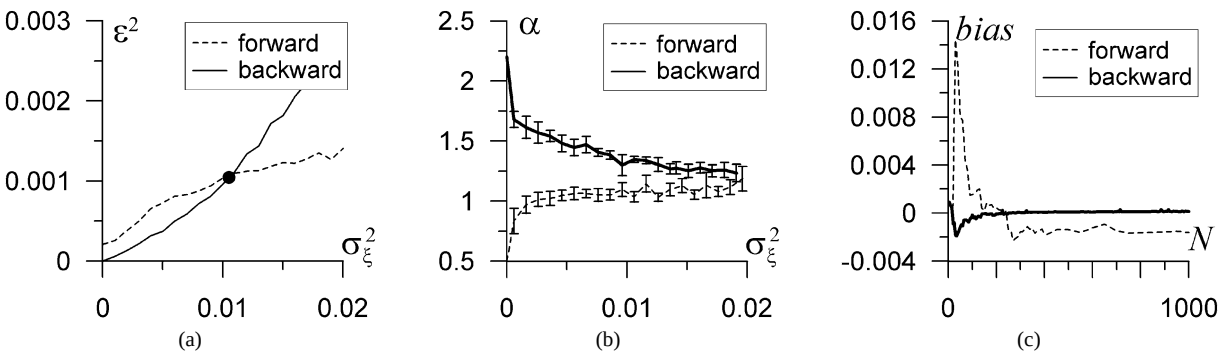

Fig. 4. Comparison of the estimates $\hat{a}_f$ (at $L = 10$) and $\hat{a}_b$ (at $m = 1$) of the parameter a of the quadratic map (6): (a) MSE at $N = 100$, (b) scaling exponent α in the "asymptotic" law $\sigma^2_{\hat{a}} \propto N^{-\alpha}$, (c) biases at $\sigma^2_\xi = 0.0001$.

Backward ML technique We use the smallest number of starting guesses $M = 1$. The MSE in the estimates $\hat{a}_b$ are shown in Fig. 3(a) for different choices of the root of the Eq. (5): genuine choice based on knowing the true values $x_n^{(0)}$ and different depths m of backward branches. The MSE for genuine choice is the least, but not drastically different from others. The results for different m are almost indistinguishable. It can be explained as follows. Let us suppose that for $m = 1$ a deviation of a model orbit from the true orbit is induced at time instant n by the erroneous choice of the root. This deviation decreases under further backward iterations due to negativity of the Lyapunov exponent in reverse time. Therefore, deeper backward branches add relatively small amount of information to recognise the wrong choice, i.e., $m > 1$ does not result in better estimates. Further we always use $m = 1$ that is much faster and simpler.

Additional peculiarities of $\hat{a}_b$ are illustrated in Fig. 3(b) where its variance is presented for very weak noise. Theoretically, the linear dependence $\sigma^2_{\hat{a}} = k_N \sigma^2_\xi$ holds for $\sigma^2_\xi \to 0$, k_N is determined from the cumulative Fisher information matrix J_N, see Appendix A. However, the variance of $\hat{a}_b$ observed empirically follows the theoretical straight line only up to $\sigma^2_\xi \approx 2.0 \times 10^{-5}$. After that, it rises with noise level much faster. This is the case even for the genuine choice of the root of Eq. (5), so this fast increase is not due to wrong choice of the root. It is not also due to local minima problem since our experiments showed that nothing is changed if more starting guesses ($M > 1$) are used. It means that a global minimum of (4) is always found even for $M = 1$ (that is a very important advantage of the backward ML technique). Thus, the fast increase in $\sigma^2_{\hat{a}}$ with noise reflects seemingly the nonlinear properties of the cost function (4) around its global minimum.

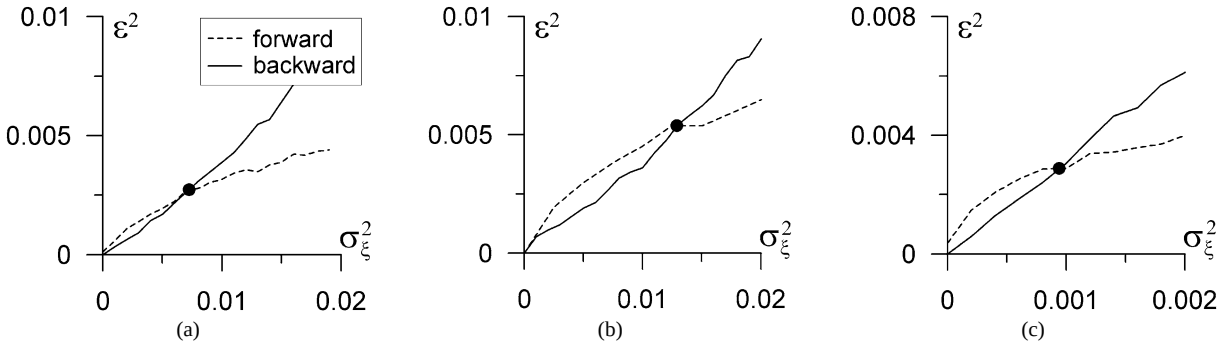

Fig. 5. MSE in $\hat{\mathbf{a}}_f$ (dashed lines) and $\hat{\mathbf{a}}_b$ (solid lines) for estimation of multiple parameters at $N = 100$: (a) 3 parameters of the map (7), constrained nonsmooth minimisation; (b) 3 parameters of the map (7), constrained smooth; (c) 4 parameters of the map (8), constrained nonsmooth.

Comparison of the techniques To facilitate the comparison, the MSE in $\hat{a}_f$ ($L = 10$, $M = 10$) and $\hat{a}_b$ ($m = 1$, $M = 1$) are reproduced together in Fig. 4(a). The backward ML technique gives more accurate estimates for noise levels up to a certain threshold value of $\sigma^2 \approx 0.011$ shown with a filled circle. In terms of noise-to-signal ratio this threshold level is equal to 17% (ratio of the standard deviation of noise $\sigma_\xi = 0.105$ to the standard deviation of the "pure" signal $\sigma_x = 0.616$).

An important property of the estimates is the behaviour of their variances and biases when N increases. We calculated variances of $\hat{a}_f$ and $\hat{a}_b$ for different noise levels and the time series lengths N in the range $[10^2, 10^3]$. They follow a power law $\sigma^2_{\hat{a}} \propto N^{-\alpha}$ quite well. We determined the values of the exponent α together with their 95% confident intervals via regressing $\log \sigma^2_{\hat{a}}$ onto $\log N$, Fig. 4(b). For high nose levels $\sigma^2_\xi > 0.015$, one observes approximately the same values $\alpha \approx 1.0$. Significantly faster decrease in the variance of $\hat{a}_b$ is observed for lower noise. Vice versa, the exponent α for $\hat{a}_f$ becomes even less than 1.0 due to the local minima problem. For $\sigma^2_\xi \to 0$ one observes quite fast decrease of the variance of $\hat{a}_b$ ($\alpha \approx 2.0$) which is due to close returns of the map orbit into the vicinity of the maximum of f as explained in Appendix A. Those returns provide high values of partial Fisher information I^2_n under backward iterations.

The behaviour of biases with increase in N is also different for the two techniques. For a low noise level $\sigma^2_\xi = 0.0001$, the bias in $\hat{a}_f$ does not decrease with N while the bias in $\hat{a}_b$ is significantly smaller and vanishes as $N \to \infty$ as illustrated in Fig. 4(c). This is an expected result. For the piecewise ML technique, systematic errors from each segment may not compensate each other, so the bias in $\hat{a}_f$ may remain asymptotically. For the backward ML technique the entire time series is used "at once", so the usual property of asymptotic unbiasedness of ML estimates looks natural, even though it is not proven rigorously for the considered problem.

3.2. Estimation of multiple parameters

To study what changes when the number of the estimated parameters increases, we take as an object again the quadratic map but with $P = 3$:

$$x_{n+1} = a_1 + a_2 x_n - a_3 x_n^2. \tag{7}$$

Time series were generated at the same values of $a_1^{(0)} = 1.0$, $a_2^{(0)} = 0.0$, $a_3^{(0)} = 1.85$, and $x_1^{(0)} = 0.3$. We use constrained nonsmooth minimisation and unchanged M, L, m to calculate $\hat{\mathbf{a}}_f$ and $\hat{\mathbf{a}}_b$. The uncertainty intervals are $\Delta a_1 = [0.8, 1.2]$ and $\Delta a_2 = [-0.2, 0.2]$, the others are unchanged.

Backward ML estimate $\hat{\mathbf{a}}_b$ is again more accurate for low noise levels, the threshold noise variance being approximately $\sigma_c^2 \approx 0.008$, Fig. 5(a). The threshold is smaller as compared to the case of $P = 1$ where $\sigma_c^2 \approx 0.011$, but the difference is not large. Moreover, the use of constrained smooth minimisation leads to unchanged $\sigma_c^2 \approx 0.011$ for single parameter estimation (not shown) and to greater $\sigma_c^2 \approx 0.013$ for estimation of three parameters (Fig. 5(b)). Thus, one can conclude that the observed estimation results for $P = 1$

and $P = 3$ are approximately the same even quantitatively.

Let us finally report the results of estimation of four parameters for the same object and orbit

$$x_{n+1} = a_1 + a_2 x_n - a_3 x_n^2 + a_4 x_n^3. \quad (8)$$

As the only additional setting we used $\Delta a_4 = [-0.2, 0.2]$. The results shown in Fig. 5(c) are qualitatively the same but the threshold noise level is definitely smaller now. It is $\sigma_c^2 \approx 0.001$ or about 5%, i.e., approximately three times as small as for $P = 1$ and $P = 3$. This is *again quite a strong noise* since, say, typical levels of measurement noise in electronics are about 0.05–0.1%.

It is difficult to reveal a universal law in the dependence of the threshold noise level on P. One could expect that superiority of the backward ML technique should become more prominent for larger P since the global minimum of (4) is easily found for any P and so the accuracy of $\hat{\mathbf{a}}_b$ could remain the same. On the other hand, minimisation of (2) is far more difficult for big P at fixed M and the accuracy of $\hat{\mathbf{a}}_f$ should decrease. However, our experience shows that the accuracy of $\hat{\mathbf{a}}_b$ also decreases with P due to the properties of the global minimum of (4) similar to that illustrated in Fig. 3(b) (not due to local minima problem). We would conjecture that the threshold noise level dependence on P is specific to the form of the original map (1), rather than being universal.

Finally, we would like to note that only *enormous* increase in M can *more or less* refine the piecewise ML estimates, since it would allow an increase in the segment length L. But it would introduce only relatively weak *quantitative* change while *qualitatively* all the results, such as an outlook of the comparison plot in Fig. 4(a), would remain the same. Similarly, increase in the number of the estimated parameters may lead only to quantitative changes. In fact, Fig. 4(a) illustrating *robust* superiority of the backward ML method for moderate noise levels is the main result of the Letter.

4. Conclusions

Modelling from time series is always finished with parameter estimation for a selected model structure. Intuitively, accurate estimates could be achieved in the case of a chaotic signal since a model orbit is very sensitive to parameters. However, if the usual ML method is applied and forward iterations of a model map are used, inevitable practical difficulties arise for a long chaotic time series. The problem is that one cannot practically find the global minimum of the cost function. A "piecewise" ML method is a good attempt to circumvent the problem, but the estimates may be asymptotically biased and their variance decreases with the time series length N (only as $1/N$).

In this Letter we suggested new "backward" ML technique for the case of one-dimensional maps and chaotic time series. It consists in the use of backward iterations of a map for evaluation of the cost function. It is shown to have several essential advantages over the "piecewise" ML technique for weak and moderate observational noise levels. First, the cost function graph is much simpler and to find the global minimum is very easy. Second, the MSE in the new estimates are (much) less than in the "piecewise" ML estimates for noise levels up to a certain threshold level. This threshold is typically quite significant, of the order of 5–15% of the signal in rms values. Third, the bias in the new estimates vanishes as $N \to \infty$ while for the piecewise ML estimates it remains approximately constant. Fourth, the variance of the new estimates typically decrease as $N^{-\alpha}$ with quite a typical value $\alpha \approx 2$. The third and fourth properties hold true at least as an intermediate asymptotic, but only for sufficiently low noise level. The results are qualitatively the same when multiple parameters are estimated.

We stress that the backward ML technique is applicable for *one-dimensional* maps. Generalisation to higher dimensions is problematic since an orbit of a multidimensional dissipative system becomes even more sensitive to the parameters under the time reversal. Nevertheless, there also exist multidimensional systems to which the method is applicable. These are hyperchaotic maps without negative Lyapunov exponents, e.g., several weakly coupled quadratic maps (6).

Acknowledgements

We are grateful to anonymous referees whose comments allowed to improve the quality of the Letter significantly. The work is supported by the President of Russia (grant MK-1067.2004.2), the RFBR (grant 05-

02-16305), CRDF (REC-006), and Russian Science Support Foundation.

Appendix A. Asymptotic scaling laws for variances of estimates

Here, we focus on the dependence of the cumulative Fisher information J_N [24] on N for a single estimated quantity a ($x_1^{(0)}$ is assumed to be known for the ease of illustration). Thereby, we find out the dependence of the variances of the estimates on N since $\sigma_{\hat{a}}^2 = \sigma_\xi^2 / J_N$ for weak noise.

The cumulative Fisher information is given by Eq. (3). Using the chain differentiation rule, it is straightforward to get the following recurrent relation for I_n

$$I_{n+1} = \mu_n \cdot I_n + \nu_n, \tag{A.1}$$

where

$$\nu_n = \left.\frac{\partial f(x_n, a)}{\partial a}\right|_{x_n = x_n^{(0)}, a = a^{(0)}},$$

$$\mu_n = \left.\frac{\partial f(x_n, a)}{\partial x_n}\right|_{x_n = x_n^{(0)}, a = a^{(0)}}.$$

The question is whether I_n is bounded or what is the law of its increase with n. If I_n were independent of μ_n and ν_n then the Eq. (A.1) would describe a linear system. Boundedness or rise of I_n would be determined by the Lyapunov exponent of (A.1). Even though I_n *depends* in general on μ_n and ν_n, we further *assume* that I_n, μ_n and ν_n are *mutually independent*. This is a strong approximation but it is justified in part by the following considerations:

(1) the dependence of I_n on μ_n and ν_n can be nonlinear and very complicated so that they can be at least almost uncorrelated;

(2) the results obtained under the assumption are confirmed by numerical experiments.

The values of μ_n and ν_n are *bounded* for the considered case of continuously differentiable map f. Hence, under our assumption and according to the Lyapunov theorem, there exist the Lyapunov exponent of the system (A.1). It is $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \ln|\mu_n|$ which coincides with the Lyapunov exponent λ of the original map orbit. Thus, if the observed regime is periodic then λ is negative and I_n is bounded. Hence, mean squared value $\langle I_n^2 \rangle$ is finite. Then one derives from (3) that asymptotically $J_N = \langle I_n^2 \rangle \cdot N$, i.e., $\sigma_{\hat{a}}^2 \propto 1/N$. For chaotic regime the Lyapunov exponent λ is positive that leads to unbounded rise of I_n. In practice, a piecewise ML technique is applied. The variance of the estimate obtained from a single segment of the length L is given by σ_ξ^2 / J_L. Since the estimates obtained from different segments are independent, then empirical averaging of them gives final estimate $\hat{a}_f$ with the variance $\sigma_{\hat{a}}^2 = \sigma_\xi^2 / (J_L(N/L))$, i.e., again $\sigma_{\hat{a}}^2 \propto 1/N$.

For the backward ML technique the similar considerations hold with some changes. The function f in (A.1) should be replaced with its inverse $f^{-1}(x_n, a)$. $f^{-1}(x_n, a)$ depends on n, since the root of Eq. (5) is chosen based on the observed value η_{n-1}. For weak noise such a choice is almost equivalent to genuine choice of the root as closest to $x_{n-1}^{(0)}$. One can rewrite (A.1) in reverse time as

$$I_{n-1} = \mu_n \cdot I_n + \nu_n \tag{A.2}$$

with

$$\nu_n = \left.\frac{\partial f^{-1}(x_n, a)}{\partial a}\right|_{x_n = x_n^{(0)}, a = a^{(0)}},$$

$$\mu_n = \left.\frac{\partial f^{-1}(x_n, a)}{\partial x_n}\right|_{x_n = x_n^{(0)}, a = a^{(0)}}.$$

Let us consider only chaotic regime. If μ_n in (A.2) is bounded, then it is easy again to derive that the Lyapunov exponent of (A.2) in reverse time is equal to $-\lambda$, where λ is the positive Lyapunov exponent of the original "forward-time" orbit. Since the Lyapunov exponent is negative, one gets again the scaling law $\sigma_{\hat{a}}^2 \propto 1/N$. However, μ_n in (A.2) is bounded only if extrema of the smooth function f are *not* embedded into the observed orbit and are not visited arbitrarily close. Otherwise, μ_n is unbounded so that one cannot guarantee boundedness of I_n under the iterations of (A.2). E.g., for the quadratic map (6) its maximum is visited arbitrarily close by a chaotic orbit.

Let us consider the most typical situation with unbounded μ_n:

(1) The extremum of f at the point x^* is quadratic so that $f(x) \approx f(x^*) + c \cdot (x - x^*)^2/2$ and $f'(x) \approx c \cdot (x - x^*)$ in its vicinity;

(2) Probability density function (invariant measure derivative) does not vanish at x^*: $p(x^*) = b > 0$.

The values of ν_n are bounded for a smooth map f and oscillate irregularly since the original orbit $x_n^{(0)}$ is chaotic. At the first glance, the product $|\mu_1\mu_2 \dots \mu_N|$ is small for large N so even the initially large values of I_n should decay. If μ_n were bounded, then I_n would fluctuate around the constant value C proportional to $\langle \nu_n \rangle$. But sometimes μ_n are enormously large and the corresponding values of I_{n-1} contribute the most essentially to the cumulative information J_N (3). To estimate this contribution, let us choose an interval $[x^* - d, x^* + d]$ with d being so small that $p(x)$ is approximately constant within the interval. The probability for $x_n^{(0)}$ to visit this interval is equal to $2p(x^*)d$. Among the N observed values $x_n^{(0)}$, $n^* = 2p(x^*)dN$ values fall into this interval. If $N \to \infty$, then $n^* \to \infty$ and the points $x_{n_j}^{(0)}$ falling into the interval $[x^* - d, x^* + d]$ tend to fill it uniformly. Similarly, the respective n^* values of the distance $|x_{n_j}^{(0)} - x^*|$ tend to be distributed uniformly in the interval $[0, d]$. Hence, those n^* distances $|x_{n_j}^{(0)} - x^*|$ are well represented by the sample $\{\frac{d}{n^*}, \frac{2d}{n^*}, \dots, \frac{(n^*-1)d}{n^*}, d\}$. These n^* time instants give the largest values of $|I_{n-1}|$. Namely, one can assume that the previous values $|I_n|$ are equal to C by the order of magnitude, since the close returns to x^* are rare and between them I may return to the assumed mean level C. The latter assumption is approximate and may lead finally to an underestimated J_N. Thus, the values of $|I_{n-1}|$ corresponding to the n^* close returns to x^* are equal approximately to $|I_{n-1}| = |\mu_n||I_n| \equiv |I_n|/(c|x_n - x^*|)$, they are represented well by the sample $C/(cd) \cdot \{n^*, n^*/2, \dots, n^*/(n^* - 1), 1\}$. Their contribution to J_N (3) is equal to $(\frac{Cn^*}{cd})^2(1 + \frac{1}{4} + \dots + \frac{1}{(n^*)^2}) \approx (\frac{2Cp(x^*)\cdot 1.5}{c})^2 \cdot N^2 = \text{const} \cdot N^2$. Even if one neglects the contribution of all remaining terms in (3), then the scaling law $J_N \propto N^2$ is derived. That is, the variance $\sigma_{\hat{a}}^2$ of the estimate $\hat{a}_b$ decreases *at least* as fast as N^{-2}.

As numerical experiments show, the scaling law for $\sigma_{\hat{a}}^2$ is unchanged if $x_1^{(0)}$ is also unknown since the ill-conditioning of the matrix J_N^{-1} is not observed even for very large N. However, one cannot guarantee that this scaling holds true in the limit $N \to \infty$ since J_N^{-1} must become finally ill-conditioned. The law $\sigma_{\hat{a}}^2 \propto N^{-2}$ holds true at least as intermediate asymptotic.

Table 1
Coefficients k_N and scaling exponents α calculated for the example of the quadratic map (6). Backward ML estimates are obtained for the reverse time orbit starting from $x_{10000}^{(0)} \approx 0.306$

	Piecewise ML ($L = 10$)	Backward ML
k_N, $N = 10$	2.25	4.608
k_N, $N = 100$	0.117	0.0211
k_N, $N = 1000$	0.0127	0.00109
k_N, $N = 10000$	0.00149	0.00000869
α	0.93	2.10

Using quite similar arguments, we formulate a *conjecture* about the scaling laws for J_N in case when the extremum is not quadratic and it is visited not so often, i.e., $p(x) \to 0$ as $x \to x^*$. Let x^* be a critical point of the map f and $|f(x) - f(x^*)| = \text{const} \cdot |x - x^*|^\beta$ with $\beta \geqslant 2$. Let the probability density function be equal to $p(x) = \text{const} \cdot |x - x^*|^\gamma$ with $\gamma \geqslant 0$ in the vicinity of x^*. Then via similar manipulations one can derive that $J_N \propto N^\alpha$ with $\alpha \geqslant 2(\beta - 1)/(\gamma + 1)$. Thus, for the above example $\beta = 2$, $\gamma = 0$, and, hence, $\alpha \geqslant 2$. So, the cause for fast decrease of $\sigma_{\hat{a}}^2$ with N is close returns to critical point of the map f which are rich of information about parameter value under backward iterations.

We stress that all above considerations are heuristic arguments and not a rigorous proof of asymptotic properties of the estimates. Finally, let us illustrate their plausibility with the main numerical example of Section 3 which is the quadratic map (6) with $a^{(0)} = 1.85$ and $x_1^{(0)} = 0.3$. We calculated the coefficients k_N in the law $\sigma_{\hat{a}}^2 = k_N \sigma_\xi^2$ as respective diagonal elements of the inverse Fisher matrix J_N^{-1} for piecewise and backward ML techniques and different N in the case of unknown $x_1^{(0)}$. The results are shown in Table 1, where the exponent α is also estimated as $\alpha \approx \log_{10}(k_{1000}/k_{10000})$. $\sigma_{\hat{a}}^2$ decreases approximately as N^{-2} for the backward ML estimate, while the scaling is N^{-1} for the piecewise ML technique. Besides, k_N for the backward ML estimate and $N = 100$ is 20 times as small as for the piecewise ML estimate. Everything is in a very good agreement with the above

heuristic considerations and the results of numerical experiments (Section 3, Fig. 4(b)).

References

[1] R. Hegger, H. Kantz, F. Schmuser, et al., Chaos 8 (1998) 727.
[2] I. Swameye, T.G. Muller, J. Timmer, et al., Proc. Natl. Acad. Sci. USA 100 (2003) 1028.
[3] L. Ljung, System Identification. Theory for the User, Moscow, 1991.
[4] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge Univ. Press, Cambridge, 1997.
[5] G. Gouesbet, S. Meunier-Guttin-Cluzel, O. Menard (Eds.), Chaos and its Reconstructions, Nova Science Publishers, New York, 2003.
[6] E. Baake, M. Baake, H.J. Bock, K.M. Briggs, Phys. Rev. A 45 (1992) 5524.
[7] R. Brown, N.F. Rulkov, E.R. Tracy, Phys. Lett. A 194 (1994) 71.
[8] V.S. Anishchenko, A.N. Pavlov, N.B. Janson, Chaos Solitons Fractals 8 (1998) 1267.
[9] J. Timmer, H. Rust, W. Horbelt, H.U. Voss, Phys. Lett. A 274 (2000) 123.
[10] B.P. Bezruchko, D.A. Smirnov, Phys. Rev. E 63 (2001) 016207.
[11] A. Sitz, U. Schwartz, J. Kurths, H.U. Voss, Phys. Rev. E 66 (2002) 016210.
[12] M. Casdagli, Physica D 35 (1989) 335.
[13] L.A. Smith, Physica D 58 (1992) 50.
[14] K. Judd, A.I. Mees, Physica D 82 (1995) 426.
[15] L.A. Aguirre, E.M.A.M. Mendes, Int. J. Bifur. Chaos 6 (1996) 279.
[16] U. Parlitz, C. Merkwirth, Phys. Rev. Lett. 84 (2000) 1890.
[17] A.M. Feigin, Ya.I. Mol'kov, D.N. Mukhin, Ye.M. Loskutov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. 44 (5–6) (2001) 376.
[18] D. Smirnov, B. Bezruchko, Phys. Rev. E 68 (2003) 046209.
[19] G. Gouesbet, S. Meunier-Guttin-Cluzel, O. Ménard (Eds.), Chaos and its Reconstruction, Cambridge Univ. Press, Cambridge, 2003, pp. 1–160.
[20] L. Jaeger, H. Kantz, Chaos 6 (1996) 440.
[21] P.E. McSharry, L.A. Smith, Phys. Rev. Lett. 83 (1999) 4285.
[22] K. Judd, Phys. Rev. E 67 (2003) 026212.
[23] W. Horbelt, J. Timmer, Phys. Lett. A 310 (2003) 269.
[24] V.F. Pisarenko, D. Sornette, Phys. Rev. E 69 (2004) 036122.
[25] J.D. Farmer, J.J. Sidorowich, Physica D 47 (1991) 373.
[26] Yu.V. Andreyev, A.S. Dmitriev, E.V. Efremova, Phys. Rev. E 65 (2002) 046220.
[27] J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.