A Global Optimization Approach to Nonlinear System Identification

A. Tiano[†*], F. Pizzocchero[†] and P. Venini[§]

† Department of Information and Systems, University of Pavia
Via Ferrata 1, I-27100 Pavia, Italy
and Institute of Ship Automation C.N.R.
Via de Marini 6, I-16149, Genova, Italy

§ Department of Structural Mechanics, University of Pavia
Via Ferrata 1, I-27100 Pavia, Italy

Abstract

This paper addresses the identification problem for nonlinear multivariable dynamical systems. A novel identification method is presented, which is based on a suitable modification of Simulated Annealing Algorithm. This method allows to formulate and solve numerically the related minimization problem by using an efficient random search minimization approach. The main features of the proposed identification method are illustrated through its application to a case study, which consists of the simulated hysteretic model of a vibrating civil structure under seismic excitation. The results show that the proposed identification method is quite efficient in comparison with other conventional identification methods.

1 Introduction

Many identification problems can be formulated as the estimation of an unknown n-dimensional parameter vector θ, which characterizes the system dynamics as expressed by a set of non linear continuous-time differential equations plus a set of discrete-time noisy observation equations of the form :

$$\dot{x} = f(x, u, \theta) \tag{1}$$

$$y(t_k) = g(x(t_k), u(t_k), \theta) + e(t_k) \tag{2}$$

on the basis of a finite number of discrete time measurements of the output vector $\{y(t_k)\}_{k=1}^{N}$ and input vector $\{u(t_k)\}_{k=1}^{N}$, where $e(t_k)$ is a discrete-time measurement noise vector.

The estimation problem is usually formulated in the form of the Prediction Error method (Ljung, 1987), which attempts to minimize directly a cost function associated to the unknown parameter vector θ, given by :

$$J(\theta) = \frac{1}{N} \sum_{k=1}^{N} \epsilon^T(t_k) W^{-1}(t_k) \epsilon(t_k) \tag{3}$$

which is constituted by a sum of squares of prediction errors , weighted through the positive definite matrix $W(t_k)$, where $\epsilon(t_k)$ denotes the prediction error vector at time t_k, i.e. the

*E-mail: *antonio@control1.unipv.it*

difference between the observed output vector $y(t_k)$ and the one-step prediction of the output vector $\hat{y}(t_k)$, which is supposed to be supplied by a suitable predictor, given by :

$$\epsilon(t_k) = y(t_k) - \hat{y}(t_k) \tag{4}$$

Under the hypothesis of linearity with respect to the parameter vector θ and of wide sense stationarity of measurement noise sequence $e(t_k)$, it can be shown (Zhou and Blanke, 1989) that prediction error vectors can be obtained by a set of Kalman filter equations.

According to Prediction Error identification method (Ljung, 1987) no particular assumptions are required on the probability distribution of measurement noise vector.

A number of identification methods based on Prediction Error have been widely used as parameter estimators for both linear systems (Ljung, 1987) and certain classes of non linear multivariable systems (Zhou and Blanke, 1989). Such methods are essentially based on the recursive minimization of the quadratic cost function expressed by Eq.(3), which takes prediction errors into account. It is worth noting that the minimization of such a cost function under constraints given by Eq. (1) and Eq. (2) is a rather difficult task, particularly in the case of nonlinear multivariable systems with a high number of parameters to be estimated. Such a minimization should be carried out with the aim of locating the global minimum of the likelihood function, not just simply a local minimum. Only the global minimum provides an estimate with good statistical properties: unbiasedness, consistency and efficiency, (Ljung,1987).

The minimization algorithms are generally based on a gradient-type iteration of the form

$$\theta_{k+1} = \theta_k - \gamma_k H_k^{-1} g_k \tag{5}$$

where θ_k is the parameter vector estimate at k-th iteration, g_k is the corresponding gradient vector of cost function $J(\theta)$, i.e. $g_k = \left[\frac{\partial J}{\partial \theta}\right]_{\theta=\theta_k}$ and H_k is a suitable approximation of the Hessian matrix $\left[\frac{\partial^2 J}{\partial \theta^2}\right]_{\theta=\theta_k}$. The positive scalar step size coefficient γ_k is chosen by a linear search procedure in such a way to ensure that cost function J is monotonically decreasing at each iteration, i.e. $J(\theta_{k+1}) < J(\theta_k)$.

Different numerical algorithms have been proposed, see, e.g. (Gupta, 1974), for the efficient implementation of Eq.(5), all of which can be regarded as variants of the basic Newton-Raphson method, in the sense that they try to reduce the numerical efforts required by computation of the Hessian, using an approximation of this matrix. At the same time it is required that such approximations result into a positive definite matrix, since only in this case a satisfactory convergence of the iterative algorithm to the global minimum of cost function $J(\theta)$ may be guaranteed.

It should be noted, however, that a number of drawbacks may seriously affect such algorithms. First of all, the approximations of the Hessian matrix, in the presence of a large number of variables or of an highly non linear cost function, is very critical, since it may become quite difficult to balance the influence of rounding errors and truncation errors, when using finite differences to estimate derivatives. As a consequence of that, the approximated Hessian matrix may become not positive definite, thus preventing convergence of the algorithm to the global minimum. Another issue of concern is that gradient-type algorithms used for minimization suffer from the fundamental weakness that the algorithm may get stuck in local minima that are not globally optimum.

It seems, therefore, that minimization methods which do not require derivatives should generally be preferable for multivariable identification problems characterized by a high number of parameters to be estimated or by an highly non linear system dynamical model. A global

minimization method, based on direct cost function evaluation and its application to nonlinear multivariable identification, will be described in the next section.

2 Simulated Annealing Algorithm

One of the most efficient random search methods for global minimization is Simulated Annealing, (Collins et al.,1988), (Aarts and Kurst,1989). In a previous paper (Tiano and Blanke,1997), Simulated Annealing was successfully applied to the identification of a multivariable linearized ship model. In the case of a nonlinear system such a method might require a fairly high number of iterations before achieving convergence. It is therefore necessary to find out suitable modifications capable to reduce the number of iterations before achieving convergence. Such modifications consist of an adaptive scheme used for carrying out the random search procedure associated with Simulated Annealing algorithm.

The determination of the solution of the global minimization problem is carried out by a direct random search procedure, which attempts to reduce the value of the cost function by means of proper tests near a preset initial parameter vector.

This method is based on an analogy between the global minimization problem and the problem of determining the lowest energy state of a physical system. A physical system has a large number of interacting atoms in thermal equilibrium at a specified temperature. If the system states are characterized by a parameter vector θ and $E(\theta)$ is the energy associated with state θ, τ is the temperature and k_b is the Boltzmann's constant, then according to statistical mechanics, the probability $P(\theta)$ that the system is in the state θ is given by :

$$P(\theta) = exp\left(-\frac{E(\theta)}{k_b \tau}\right) \tag{6}$$

Under equilibrium, the most probable states at any given temperature are those associated with the lowest energy. As it can be demonstrated by theoretical arguments as well as by experimental evidence, the most effective strategy for obtaining the state with globally minimum energy consists of slowly cooling a thermodynamic system. This enables it to achieve equilibrium during the transition from a given initial state to the lowest energy state. In fact, if the cooling process is carried out sufficiently slowly, the system is allowed to skip over locally stable states and reach the global minimum energy one. Usually, the Boltzmann's constant k_b is combined with the temperature and the term *temperature* is used with reference to their product. In this way, temperature can also be viewed as nothing but a control parameter for the minimization procedure.

Simulated Annealing consists of three distinct steps:

- a random search step;
- a minimization step;
- a stopping rule.

The random search step is basically given by the iterative generation of random vectors in a domain $S(\theta_k)$, constituted by neighboring vectors associated to the current vector θ_k, at k-th iteration. The minimization step consists of applying a local minimization routine to some of the sampled vectors, while the stopping rule is used to stop the algorithm if there is sufficient evidence that the global minimum has been detected within the limits of a specified accuracy or when some specified iterations number and/or computing resources have been exhausted.

The Simulated Annealing algorithm is sketched in Table 1.

```
begin
    θ := initial solution θ_0;
    τ := initial temperature τ_0;
    while (stopping criterion is not satisfied) do
      begin
        while (not yet in equilibrium) do
          begin
            ξ : = random vector selected in S(θ);
            ΔJ := J(ξ) − J(θ);
            P := min{1, e^{−ΔJ/τ}};
            ε := random generation uniform in [0, 1];
              if ε ≤ P then θ_k := ξ;
          end
        τ_k := updated temperature f(J(θ));
      end
    output of optimal solution
end
```

Table 1. Outline of Simulated Annealing algorithm

After the assignment of an initial value θ_0 to the unknown parameter vector, an initial temperature τ_0 is chosen high enough to ensure that virtually all transitions in the parameter space may be possible. The random search is then carried out iteratively. A new vector ξ is chosen belonging to the neighboring set $S(\theta_k)$ associated to the current parameter vector $\theta_k \in R^n$:

$$\theta_k = \left[\theta_k^{(1)} \quad \theta_k^{(2)} \quad \cdots \theta_k^{(n)}\right]^T \tag{7}$$

Before taking a new vector ξ into account, however, a statistical test is performed to decide if equilibrium has been reached. Such test essentially consists in the verification that a given finite sequence of vectors generated by the algorithm inside the inner loop can be regarded as a realization of a time-homogeneous Markov chain, (Feller, 1950).

In the implementation of Simulated Annealing described in this paper, the neighboring set $S(\theta_k)$ is constituted by the interior of a hyper-ellipsoid centered in θ_k and having semi-axes $\rho_k^{(i)}$, $i = 1, ...n$. These semi-axes are chosen according to an adaptive scheme, in such a way that when the cost function decreases, a step is made in the same direction, and when it increases, in the opposite direction:

$$\begin{aligned} \rho_{k+1}^{(i)} &= \lambda_1^{(i)} \rho_k^{(i)} \qquad if \qquad J(\xi) < J(\theta_k) \\ \rho_{k+1}^{(i)} &= \lambda_2^{(i)} \rho_k^{(i)} \qquad if \qquad J(\xi) \geq J(\theta_k) \end{aligned} \tag{8}$$

where $\lambda_1^{(i)} \geq 1$, $0 < \lambda_2^{(i)} < 1$ and $\lambda_1^{(i)} \lambda_2^{(i)} < 1$ are the parameters characterizing the adaptive scheme, (Rastrigin,1983).

The region can be set to depend on properties such as output sensitivity for each particular parameter. The direction of search is chosen as the realization of n independent, equally distributed random variables, sampled inside the hyper-ellipsoid with adaptively varying semi-axes.

The determination of the solution of the global minimization problem is carried out by a direct random search procedure, which attempts to reduce the value of the cost function by means of proper tests near a preset initial parameter vector.

At each iteration a decision has to be taken whether or not to accept the new vector ξ in place of the current one θ_k. Acceptance of the new vector is made with probability

$$\min\left\{1, exp\left(-\frac{J(\xi) - J(\theta_k)}{\tau_k}\right)\right\} \tag{9}$$

where τ_k is the current value of temperature. Every descent is thus accepted, but it is also possible, even if at a limited extent, to perform also *up-hill* transitions, which may allow the algorithm to escape out of local minima.

The temperature is gradually, at each iteration, reduced according to a proper *cooling schedule*. Such cooling should be carried out quite slowly, in order to enable the algorithm to achieve equilibrium. For this purpose, the cooling schedule is implemented automatically as a function of the current cost function, i.e.

$$\tau_k = f(J(\theta_k)) \tag{10}$$

After a number of simulation experiments, following also a suggestion (Bohachevsky, 1986), it has been used a piecewise linear schedule of the type:

$$f(J(\theta_k)) = \gamma_k(J(\theta_k) - J_{\text{inf}}) \tag{11}$$

where J_{inf} is the value of the cost function corresponding to the global minimum and the coefficients γ_k are positive constants. In this way the temperature parameter τ_k is automatically driven to zero, with a speed regulated by γ_k. It is immediately clear that the implementation of the method requires the knowledge of the solution J_{inf} itself, which unfortunately is unknown. A procedure for estimating recursively J_{inf} can, however, be developed, which is illustrated in (Tiano and Blanke,1997).

Two salient features of Simulated Annealing compared to conventional iterative minimization algorithms were shown to be

- The algorithm very rarely gets stuck into local minima, since transition out of a local minimum is always possible when the search operation is carried out at a nonzero temperature.
- Simulated Annealing exhibits an interesting temperature-dependent adaptive behavior, according to which the gross features of the global minimum become evident at higher temperatures, while fine details appear at lower temperatures.

As concerns convergence analysis of the algorithm, some theoretical results, (Aarts and Korst, 1989), show that convergence with probability one to a globally optimal solution can be achieved, under some conditions concerning the temperature cooling schedule. Unfortunately, however, such conditions are of limited practical use, since they are only sufficient but not necessary and the results are asymptotic by nature. Nevertheless, these convergence results give some confidence on the Simulated Annealing method and does offer an explanation of its success in optimization practice.

3 On the need of system identification in seismic design and retrofit

In the presence of strong earthquakes, civil structural systems undergo permanent inelastic deformations. A physically sound model able to describe the nonlinear relationship between the applied dynamical loads and the response of the structure in terms of stresses and strains is of utmost importance at design stage as well as for retrofitting existing damaged structures. Endochronic and standard plasticity models are among the most widely used by the civil engineering community to simulate the dynamics of structures under the action of strong earthquakes. Such idealizations are of parametric type. The parameters involved in the model are material–dependent and govern the yielding stress, i.e. the stress above which permanent deformations occur, the smoothness and the shape of the hysteresis loop that is of crucial importance in that it controls the amount of energy that the structure may dissipate in the inelastic range. As to this contribution, we have adopted the Bouc-Wen model that is described in the section to come.

3.1 The Bouc-Wen model

The Bouc–Wen model is often used to include hysteretic phenomena into the analysis of civil structures within a rather simple framework. Referring to shear–beam idealizations, the i–th interstory restoring force F_{s_i} reads

$$F_{s_i} = \alpha_i k_i x_i + (1-\alpha_i) k_i D_{y_i} v_i, \tag{12}$$

where α_i is the post–to–pre yielding stiffness ratio, k_i is the initial elastic stiffness, x_i is the relative displacement between adjacent storeys and D_{y_i} is the yielding displacement. The new hysteretic variable v_i is a state variable so that each story of the building is now described by a triplet of state variables, i.e. displacement, velocity and hysteretic variable. The nonlinear constitutive law is then introduced by means of the nonlinear equation

$$\dot{v}_i = D_{y_i}^{-1}\{A_i \dot{x}_i - \beta_i |\dot{x}_i||v_i|^{n_i-1} v_i - \gamma_i \dot{x}_i |v_i|^{n_i}\}, \tag{13}$$

where A_i, β_i and γ_i control the shape of the hysteresis loop and n_i determines the smoothness of the loop and the degree of nonlinearity. In particular, as n_i approaches infinity, an elastic perfectly plastic behavior is attained. Equations (12) and (13) define completely the Bouc–Wen endochronic model that need to be coupled to the classical second–order equation of motion. One should notice that, unlike standard plasticity models that call for an incremental formulation, the adoption of the Bouc–Wen model allows to use finite quantities likewise the linear case. This allows the usage of classical methods of integration for nonlinear differential equations avoiding expensive techniques based on Newton–Raphson iterative schemes.

3.2 Global equations of motion

Let us focus our attention on a n_d degree–of–freedom system which is excited by a ground base acceleration $\ddot{x}_g(t)$, see Figure 1 for a schematic of the system and Figure 2 for the El Centro seismic record that has been used for the numerical simulations to be presented next. The vector equation of motion of the system may be written as

$$M\ddot{x}(t) + C\dot{x}(t) + K_e x(t) + K_h v(t) = \Gamma \ddot{x}_g(t), \tag{14}$$

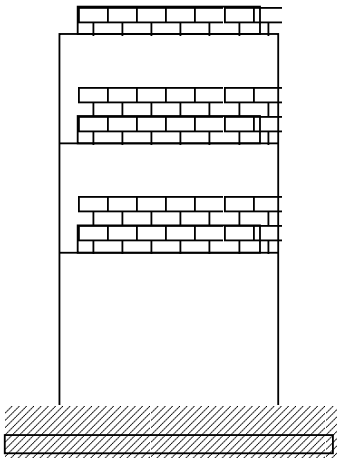

Figure 1: Typical 3DOF shear-type structure

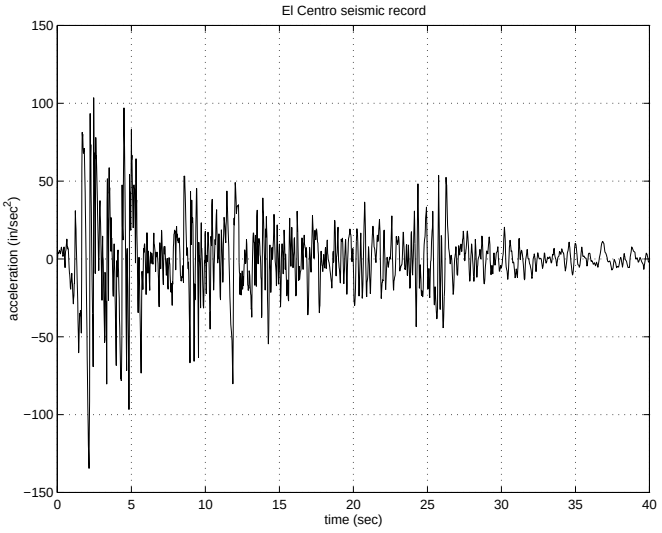

Figure 2: El Centro seismic record

where $M = n_d \times n_d$ mass matrix, $C = n_d \times n_d$ damping matrix, $K_e = n_d \times n_d$ initial stiffness matrix, $K_h = n_d \times n_d$ hysteretic stiffness matrix and $\Gamma = n_d \times 1$ vector distributing the inertial forces due to the base motion to each story. An n_d-dimensional vector f is then introduced that groups all the equations of type 13, one for each story of the structure,

$$\dot{v} = h(\dot{x}, v), \quad h_i(\dot{x}_i, v_i) = D_{y_i}^{-1}\{A_i\dot{x}_i - \beta_i|\dot{x}_i||v_i|^{n_i-1}v_i - \gamma_i\dot{x}_i|v_i|^{n_i}\}, i = 1, \ldots, n_d. \tag{15}$$

Toward the adoption of classical time integration schemes for first–order vector differential equations, we introduce a $3n_d$ state vector $z(t)$ and a $3n_d$ vector W as

$$z(t) = \left\{ \begin{array}{c} x \\ v \\ \dot{x} \end{array} \right\}, \quad W = \left\{ \begin{array}{c} 0 \\ 0 \\ M^{-1}\Gamma \end{array} \right\}. \tag{16}$$

One may then write the nonlinear state–space equation of motion as

$$\dot{z} = g(t, z) + W\ddot{x}_g(t) \equiv f(t, z), \tag{17}$$

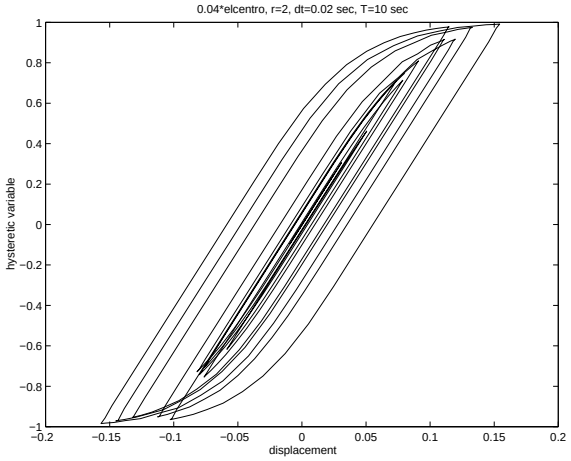

Figure 3: Hysteretic restoring force

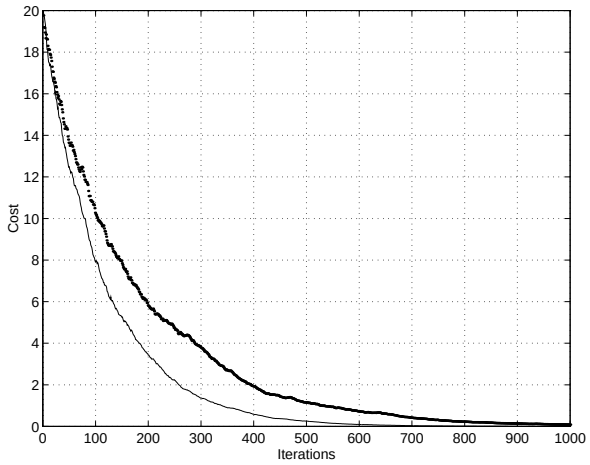

Figure 4: Comparison of SA (dotted line) and MSA (solid line) convergence properties

where $g(t,z)$ is a $3n$ vector, depending nonlinearly on the state $z(t)$, that reads

$$g(t,z) = \left\{ \begin{array}{c} \dot{x} \\ h(\dot{x},v) \\ -M^{-1}[C\dot{x} + K_e x + K_h v] \end{array} \right\}. \tag{18}$$

There exist several numerical schemes for the integration of (17). We have considered the uniform $r+1$-steps Adams-Bashforth family that approximates the state vector z with the vector η that is computed by

$$\eta_{i+1} = \eta_i + \Delta t \sum_{j=0}^{r} b_{j,r} f_{i-j}, \quad i \geq r, \quad f_i = f(t_i, \eta_i), \tag{19}$$

where $b_{j,r}$ are properly computed weight factors.

4 Identification Results

Figure 3 presents a typical hysteresis cycle that describes the restoring force of a SDOF system subject to the El Centro record. The latter was scaled by a factor 0.4 to match the features of

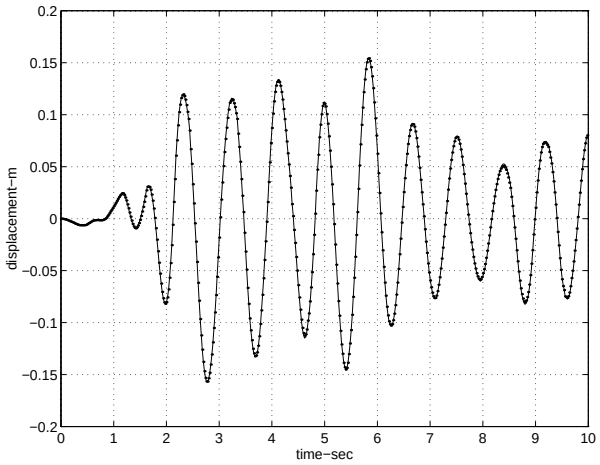

Figure 5: Fitting of MSA identified model (dotted line) and simulated (solid line) displacement

the system that were chosen as shown in Table 2.

M	C	K_e	K_h	D_y	A	β	γ	n
10000	692	600000	300000	0.1	1	0.5	0.5	3

Table 2: Physical properties of the system

In order to test the proposed identification method, a number of simulations were first carried out on a structure described by parameters shown in Table 2. The identification problem consisted in estimating from input/output simulated time histories the unknown parameter vector

$$\theta = [D_y \quad A \quad \beta \quad \gamma \quad n]'. \tag{20}$$

A standard Simulated Annealing algorithm and the one modified with the adaptive scheme above described were compared with respect to a varying range of initial parameter vector estimates. It has been verified that the modified SA method is capable to achieve an appreciably faster convergence to the true parameter vector. A typical behavior is shown in Fig. 4, where the cost function time histories of the standard SA (.) and the modified SA one (-) are plotted.

Such results are based on a 10 second time length seismic excitation record and corresponding civil structure response.

The rather good agreement between simulated displacement and identified one is shown in Fig. 5.

More complete and detailed result, not reported here, which will be published in the near future allow to draw the following conclusions:

- The modified SA identification algorithm can achieve an appreciable reduction in convergence time with respect to the standard SA method.
- The quite high reliability of the estimated parameters indicates that it can be used as a tool for nonlinear system identification and particularly in the area of seismic design.

References

Aarts, E. and J. Korst (1989). *Simulated Annealing and Boltzmann Machines*, J. Wiley & Sons, New York.

Baber, T. T. and Y.-K. Wen (1981). "Random vibrations of hysteretic degrading systems," *ASCE J. Engrg. Mech. Div.*, **107**, no. 6, pp. 1069–1087.

Bohachevsky, I. O., M. E. Johnson, and M. L. Stein (1986), "Generalized simulated annealing for function optimization," *Technometrics*, **28**, pp. 209–217.

Brent, R. P. (1972). *Minimizing a function without calculating derivatives*, Prentice Hall, London.

Collins, N. E., R. W. Eglese, and B. L. Golden (1988). "Simulated annealing – An annotated bibliography," *American Journal of Mathematical and Management Sciences*, **8**, pp. 209–308.

Eykhoff, P. (1974). *System Identification*, John Wiley and Sons, London, U.K.

Feller, W. (1950). *An introduction to probability theory and its applications*, Vol. 1, John Wiley & Sons, New York.

Gupta, N. K. and R. K. Mehra (1974). "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations," *IEEE Transactions on Automatic Control*, **19**, no. 6, pp. 774–783.

Hajek, B. (1985). "A tutorial survey of theory and applications of simulated annealing," *Proceedings of 2th Conference on Decision and Control*, Ft. Lauderdale, USA, A4, pp. 755–760.

van Laarhoven, P. J. M. and E. H. L. Aarts (1987). *Simulated Annealing : Theory and Practice*, Reidel, Dordrecht.

Ljung, L. (1987). *System Identification: Theory for the User*, Prentice-Hall, London.

Rastrigin, L. A. (1983). *Contemporary principles to control complex objects*, Mir Publisher, Moscow.

Tiano A. and M. Blanke (1997). "Multivariable identification of ship steering and roll motions," *Transactions of Institute of Measurement and Control*, **19**, no. 2, pp. 63–77.

Zhou, W. W. and M. Blanke (1989). "Identification of a class of nonlinear state space model using RPE techniques," *IEEE Transactions on Automatic Control*, **34**, no. 3, pp. 312–316.