Multi-Output System Identification using Evolutionary Programming

B.L. Andersen, W.C. Page, and J.R. McDonnell
Naval Ocean Systems Center
San Diego, CA

Abstract

Evolutionary programming (EP) has been demonstrated to be an effective method of system identification of single-input-single-output (SISO) systems. This paper investigates the use of EP in system identification of single-input-multi-output (SIMO) systems. EP is used to identify parameters of a linear, time-invariant system. Specifically, this paper examines the identification of SIMO systems whose measurements can contain different amounts of noise. A cost function is proposed to take into account disparate noisy observations.

I. Introduction

System identification is the process of developing an accurate model of a dynamic system. This involves selecting the appropriate model structure and choosing the model parameters which optimize an objective function. The usual approach taken in system identification is to assume a model structure based on the process physics and then to determine the model parameters. Oftentimes, nonlinear models are linearized so that an appropriate state space formulation may be used.

Typically, methods of system identification involve minimizing the objective function using gradient descent techniques. Use of the gradient may result in the identified parameters being only locally optimal. Continuity constraints are another limitation associated with pure gradient search techniques. A Kalman filter approach is given in [1] along with cautions on its application to system identification. Many algorithms have been proposed which attempt to alleviate the inherent problems associated with gradient methods. Recent approaches have included simulated annealing, genetic algorithms, neural networks and evolutionary programming. The latter method has proven extremely successful for function optimization.

Evolutionary programming (EP) has been successfully applied to the problem of system identification. Fogel [2] has proposed the use of EP in system identification for evolving ARMAX process coefficients. His investigations extend to simultaneous parameter and model structure estimation. This paper extends the work on SISO systems to SIMO systems. The extension to multi-input, multi-output (MIMO) systems is obvious.

Identification of a SIMO system could be divided into identifying a set of SISO system transfer functions between the input and each of the outputs. However, each of these transfer functions would be identified using only the measurements of its output and not the full set of observations which contains more information. Also, it is not clear how one might combine these individual transfer functions with parameter estimates of varying degrees of uncertainty into one combined system model that would be required to control all the states of the system.

The remainder of this paper is organized as follows. Section II contains an overview of the EP paradigm. Section III formulates EP for SIMO system identification and describes the model structure. Section IV presents a linear, time-invariant system that is used for computer experiments. Section V discusses the results and Section VI provides a summary of the results and conclusions.

II. Evolutionary Programming

EP is a parallel stochastic search technique that maps natural evolution to an effective multi-agent search strategy. This search technique is not based on gradient methods and is therefore not susceptible to entrapment in local minima. Search by simulated evolution was first described by Fogel [3]. This investigation follows the search strategy outlined in Fogel [2] and described by the sequence below

1. Create the initial population P consisting of N parents.
2. Assess the fitness of each parent p_i in the population.
3. Mutate each element in the population by a $N(0,\sigma)$ random variable, where σ is

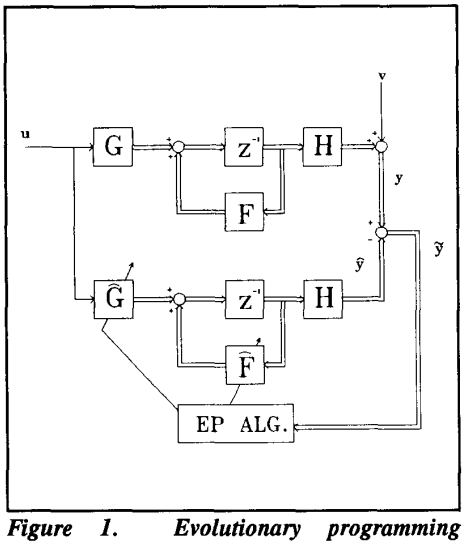

Figure 1. Evolutionary programming parameter estimation block diagram.

proportional to the fitness.

4. Assign a new fitness score to each element of the population.
5. Hold a stochastic competition between the members of the population.
6. Rank the members of the population according to the results (number of "wins") of step 5.
7. Repeat step 3 with the highest ranked N members from the population.

This algorithm is applied to system identification in the following section.

III. EP for System Identification

The linear time-invariant SIMO structure is assumed to be of the form

$$x(k+1) = Fx(k)+Gu(k)$$

$$y(k) = Hx(k)+v(k)$$

where $x(k)$ is the state vector of length n, $u(k)$ is the input, $y(k)$ is the vector of measurements of length l, $v(k)$ is the vector of $N(0,\sigma_v)$ measurement noise. F is the system matrix, G is the input weighting matrix and H is the measurement scaling matrix. The block diagram for estimating parameters using EP is shown in Figure 1.

The vectors to be evolved will consist of the elements of the F and G matrices. The H (measurement) matrix will not be evolved because finding all three of the matrices is an underdefined problem since an infinite number of combinations of F, G and H matrices apply to the same system dynamics. For this reason, the measurements will be assigned to the first l states of the system. This results in an H matrix of the form

$$H = \left[I_{l \times l} \mid 0_{l \times (n-l)} \right]$$

Given this formulation, the method of applying EP to the identification problem is straightforward. The remaining issues to be resolved are the mutation strategy and the objective function.

Two mutation strategies were investigated. The first incorporates a standard deviation of the mutation for each vector that is linearly proportional to the fitness score as suggested by Fogel [2]. This results in the size of the mutations decreasing as the population approaches the optimum value. In addition to ensuring convergence, this strategy also assures an effective search by allowing larger mutations for elements that have poor fitness scores. Atmar [4] suggests a variety of mechanisms to accelerate the search process.

The second mutation strategy investigated is based on the number of iterations. The mutation of any element in the vector can be described by the perturbation

$$\delta = \mu e^{-\alpha \tau k}$$

where μ is a $N(0,\sigma)$ random variable (r.v.), τ is a $U[0,1]$ r.v., α is a constant and k is the generation number. This mutation is a function of the number of iterations only. A random walk is maintained over a high number of generations while still preserving the capability of infinitesimal perturbations to achieve optimum values as k increases. Reasonable results were obtained using this mutation strategy for optimizing mobile manipulator configurations [5].

The objective function must be carefully chosen for this search. Not only will the measurements generally be of different units, but the amount of confidence in each measurement may differ due to sensor characteristics. A cost function

has been formulated to take into account disparate sensor noise and the difference between state estimates and observations. The error vector between observation and estimate is given by

$$\tilde{y} = y - \hat{y}$$

where y is the observation and $\hat{y}$ is the estimate. The cost function incorporates sensor uncertainty (noise) using the diagonal standard deviation matrix

$$\Omega = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & . & . & . \\ 0 & \frac{1}{\sigma_2} & 0 & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \frac{1}{\sigma_n} \end{bmatrix}$$

where σ_i corresponds to the noise characteristics of the i^{th} sensor. The cost associated with each parameter estimate can be described as

$$\Gamma = \frac{1}{n}\tilde{y}^T R \tilde{y}$$

where the variances are selected according to sensor specifications, n is the length of the error vector and $R = \Omega^T\Omega$ or the identity matrix I depending on the experiment. Using $R = \Omega^T\Omega$ accomplishes two things. First, it allows the cost to be a dimensionless quantity. Second, it scales each measurement by the amount of confidence in its accuracy. If the measurement is noisy, its contribution to the cost function is reduced since the prediction errors could be due primarily to measurement noise. If an observation is relatively noise-free, its contribution to the cost function will be large since the error is due primarily to estimation errors, assuming the model structure is correct. For these trials, it is assumed that the variances are constant but not necessarily equal. If the variances are equal, then the cost function is simply scaled by this magnitude. Using $R = I$ and scaling the cost function will yield the same results.

IV. Simulated System

A simple second-order system based on spring-mass-damper dynamics was used for computer experiments. Parameters were chosen so that the system was critically damped. The continuous state equations can be written as

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x + \nu$$

The equivalent discretized state-space formulation, assuming a zero-order-hold and a sample rate of T=0.05 sec, is given by

$$x(k+1) = \begin{bmatrix} 0.9988 & 0.0476 \\ -0.0476 & 0.9037 \end{bmatrix} x(k) + \begin{bmatrix} 0.0012 \\ 0.0476 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x(k) + \nu(k)$$

The forcing function is ensured to be persistently exciting by defining $u(k)$ as an r.v. with a uniform distribution. The system was propagated forward 100-200 time steps for each simulation from an initial state randomly chosen from a U[-1,1] distribution.

V. Results

Simulations were conducted using both mutation strategies. In general, the second mutation strategy outperformed the first. The reason for this is still under investigation. The state variables had to be constrained to prevent numeric overflow. This is not unreasonable given that the time integration of an unstable system is not bounded. However, it does have implications in system identification of unstable plants.

The first experiment for the second mutation strategy used numerically equal measurement noise terms with N(0,0.05) statistics and α=0.01. The resulting parameters are given in Table I under experiment 1. The average percentage error for all

parameters is approximately 8%.

The magnitude of the forcing function was increased tenfold in an attempt to obtain better system identification. The parameters resulting from this experiment are given in Table I in the experiment 2 column. As expected, these parameters are have a lower overall average error than those obtained in experiment 1. For this case, the average overall parameter error is roughly 3.14%.

Figure 2 shows the mean and best population scores using the second mutation strategy. The cost function is simply the mean sum squared error (MSE) of both observations. At roughly 300 iterations the identification process converged. This run had a signal-to-noise ratio of approximately 20 and equivalent noise statistics on each measurement. Figure 3 illustrates the convergence of the first parameter (f_{11}) in the highest ranked F matrix. The other parameters converge in a similar fashion.

An experiment was conducted to compare the effect of the two cost functions using the second mutation strategy. The standard deviation of the position measurement noise was 0.01 and the standard deviation of the velocity measurement noise was 0.05. In both cases α was set at 0.05. The best values of the parameters are shown in Table I. The parameter values under the experiment 3 heading are for the cost function that included the variance term ($R = \Omega^T\Omega$). The parameter values given under the experiment 4 heading are for R equal to the identity matrix. The experiment that incorporated a cost function without the variance term unexpectedly had better results than the experiment which incorporated a confidence weighting term. The reasons for this result are still being investigated.

VI. Conclusions

EP is a powerful search technique that can be applied to system identification. This research has demonstrated that EP can be extended to identify linear SIMO systems of known order. The extension to identification of MIMO systems is obvious. It is also clear that the scaling of the cost function must be chosen carefully to account for disparate sensor data and measurement noise. The results given in this work have yet to be statistically verified and therefore should not be generalized. Investigations on the algorithmic properties of EP are being continued for validation in further experiments.

The formulation described here is a general one for the identification of state-space models, but there are still additional considerations to be addressed. First, the model ignored process noise, the terms that account for differences from the actual dynamic equations. Second, this work assumed that the order of the system is known. To identify the order of the model as well as the parameters, a multi-dimensional equivalent of Akaike's information criterion (AIC) or minimum defining length (MDL) principle similar to that used by Fogel [6] for neural networks must be developed.

The application of EP for adaptive control is readily apparent. Issues to be addressed in using EP for adaptive control pertain to the bandwidth of the process being controlled. That is, of course, assuming a traditional adaptive control architecture. EP could be implemented in an arrangement such that system identification would not have to be explicit.

References

[1] B.D.O. Andersen and J.B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.

[2] D.B. Fogel, *System Identification through Simulated Evolution: A Machine Learning Approach to Modeling*, Ginn Press, Needham, MA, 1991.

[3] L.J. Fogel, A.J. Owens, and M.J. Walsh, *Artificial Intelligence through Simulated Evolution*, John Wiley and Sons, New York, 1966.

[4] W. Atmar,"Natural Processes which Accelerate the Evolutionary Search", IEEE Proc. of Twenty-Fourth Asilomar Conf. on Signals, Systems, and Computers, 1990.

[5] J.R. McDonnell, W.C. Page, and B.L. Andersen,"Configuration Optimization for Platform-Manipulator Systems using Evolutionary Programming", Proc. AMSE Conf. on Neural Networks, San Diego, 1991.

[6] D. B. Fogel,"An Information Criterion for Optimal Neural Net Selection", *IEEE Trans. on Neural Networks*, Vol. 2, No. 5, 1991.

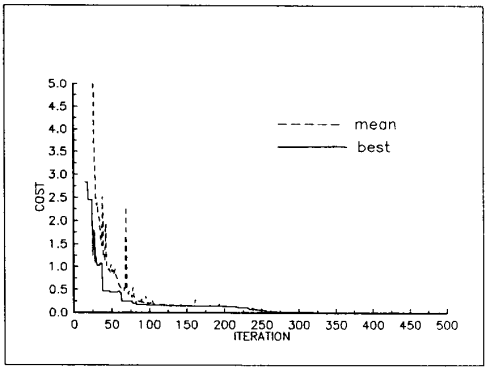

Figure 2. The mean and minimum (best) population costs.

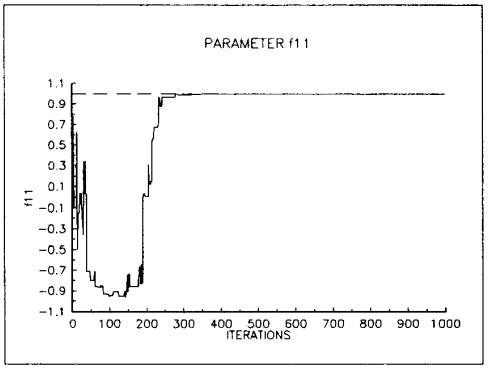

Figure 3. Evolution of parameter f_{11}.

Para-meters	Actual values	Final Best Value Statistics							
		1st Experiment		2nd Experiment		3rd Experiment		4th Experiment	
		value	%error	value	%error	value	%error	value	%error
f_{11}	0.9988	0.9965	0.23	0.9986	0.02	1.0193	2.05	0.9973	0.15
f_{12}	0.0476	0.0527	10.7	0.0480	0.84	0.1015	113.2	0.0422	11.3
f_{21}	-0.0476	-0.0449	5.67	-0.0481	1.05	-0.0745	56.51	-0.504	5.88
f_{22}	0.9037	0.90163	0.23	0.9033	0.04	0.8315	7.99	0.9062	0.28
g_1	0.0012	0.0009	25.0	0.0010	16.7	-0.0152	1367	0.0009	25.0
g_2	0.0476	0.0507	6.51	0.0477	0.21	0.0480	0.84	0.0440	7.56

Table I. Comparison of final value results of experiments and actual parameters.