On Non-parametric Nonlinear System Identification Using Dispersion Functions

Ho-En Liao and William A. Sethares

Department of Electrical and Computer Engineering
University of Wisconsin-Madison

Abstract — This paper adopts dispersion functions as a tool for non-parametric nonlinear system identification. Dispersion functions have the meaning of "projection" and they make it possible to determine the order of a model, and the significance (or the existence) of delay elements within a nonlinear model.

1. Introduction

This paper proposes a method for nonlinear system identification that determines the order of the model and the possible delay elements that exist in a nonlinear ARMA (NARMA) model [6]. The tool used in this paper is the dispersion function defined by Rajbman [1]. The dispersion function can be used to determine the degree of nonlinearity of a nonlinear plant, as in [1] or [2]. Unlike correlation functions, dispersion functions measure the relationship or the strength among random processes which are nonlinearly related. Moreover, dispersion functions can be viewed as a "projection" that allows nonparametric identification even when the data is highly correlated. In this paper, nonlinear AR models [3] are considered for non-parametric system identification using dispersion functions, though it is easy to extend the idea to NARMA models. The proposed method evaluates dispersion functions of different orders to determine the order and the significant delay elements of a nonlinear model. The method is independent of the degree-of-nonlinearity of the model to be identified.

2. Dispersion Functions

For two random processes $Y(t)$, $t \in T_y$ and $X(s)$, $s \in T_x$, the cross-dispersion function, denoted by $\theta_{Y|X}(t, s)$, is

$$\theta_{Y|X}(t, s) \equiv \mathrm{cov}\{E[Y(t)|X(s)], E[Y(t)|X(s)]\}. \tag{1}$$

The auto-dispersion function $\theta_{X|X}(t, s)$ is the special case where $Y(t)=X(t)$ and $t \in T_x$,

$$\theta_{X|X}(t, s) \equiv \mathrm{cov}\{E[X(t)|X(s)], E[X(t)|X(s)]\}. \tag{2}$$

In the above definitions, $\mathrm{cov}\{\cdot, \cdot\}$ is the covariance operator of two random processes and $E[\cdot \mid \cdot]$ is the conditional expectation. The cross-dispersion function $\theta_{Y|X}(t, s)$ represents, for the given values of the arguments t and s, the variance of the conditional expectation of $Y(t)$ relative to $X(s)$, and it characterizes the overall variance of $Y(t)$ on the closed measure subspace generated by the random process $X(s)$. Similarly, the auto-dispersion function $\theta_{X|X}(t, s)$ represents, for the given values of the augments t and s, the conditional expectation of $X(t)$ relative to $X(s)$. In the case when all random processes are jointly stationary, the dispersion functions in (1) and (2) become functions of (t - s), i.e.,

$$\theta_{Y|X}(t, s) = \theta_{Y|X}(\tau)$$

and

$$\theta_{X|X}(t, s) = \theta_{X|X}(\tau) ,$$

where $\tau = t - s$.

Cross-dispersion (and auto-dispersion) functions have the following properties (proofs are omitted).

Property 1: $0 \le \theta_{Y|X}(t, s) \le \sigma_Y^2(t)$, where $\sigma_Y^2(t)$ is the variance of the random process $Y(t)$.

Property 2: If $Y(t)$ is independent of $X(s)$ then $\theta_{Y|X}(t, s)=0$.

Property 3: $\theta_{Y|X}(t, s) = \sigma_Y^2(t)$ iff $Y(t)$ and $X(s)$ are related by some (possiblely nonlinear) functional relation $f(\cdot)$, i.e., if there is a function $f(\cdot)$ such that $Y(t) = f(X(s), t, s)$.

Property 4: Let $\sigma_{Y|X}^2(t,s) = E\{ (Y(t) - E[Y(t)|X(s)])^2 \mid X(s) \}$, then

$$\theta_{Y|X}(t, s) = \sigma_Y^2(t) - E[\sigma_{Y|X}^2(t,s)] . \tag{3}$$

Property 5: $\theta_{Y|X_1}(t, s) \le \theta_{Y|X_1,X_2}(t, s, u)$, where

$$\theta_{Y|X_1,X_2}(t,s,u) = \mathrm{cov}\{E[Y(t)|X_1(s), X_2(u)],E[Y(t)|X_1(s), X_2(u)] \} .$$

In **Property 5**, dispersion functions have been extended to two dimensions. More generally, the multi-dimensional cross-dispersion function of order N can be defined as

$$\theta_{Y|X_1,X_2,\dots,X_N}(t, s_1, s_2, \dots, s_N) = \mathrm{cov}\{E[Y(t)|X_1(s_1), X_2(s_2), \dots, X_N(s_N)], E[Y(t)|X_1(s_1), X_2(s_2), \dots, X_N(s_N)] \}.$$

The auto-dispersion function of order N is defined similarly as

$\theta_{X|X,X,...,X}(t, s_1, s_2, ..., s_N)$
$= \text{cov}\{ E[X(t)|X(s_1), X(s_2), ..., X(s_N)],$
$E[X(t)|X(s_1), X(s_2), ..., X(s_N)] \}$

Among the properties presented above, **Properties 1** and **2** were shown in [1], [4]. Some remarks about the properties of dispersion functions are noted in the following.

Remark 1: **Property 1** provides upper and lower bounds for dispersion functions. Note from **Properties 2, and 3**, $\theta_{Y|X}(t, s)=0$ is a necessary but not a sufficient condition for Y(t) to be independent of X(s). But $\theta_{Y|X}(t, s) = \sigma_Y^2(t)$ is both necessary and sufficient for Y(t) to be expressible as some function of X(s). This plays a major role in determining the order of a given nonlinear AR model and determining the significance of a delay element in the model.

Remark 2: The quantity $E[\sigma^2_{Y|X}(t,s)]$ ($\sigma^2_{Y|X}(t,s)$ is the so-called "conditional variance of Y(t) relative to X(s)") defined in **Property 4** is closely related to $\theta_{Y|X}(t, s)$. $\theta_{Y|X}(t, s)$ is the variance of Y(t) after "projecting" [5] on the closed measure subspace generated by the random process X(s), denoted by $\mathbf{G}\{X(s)\}$. On the other hand, $E[\sigma^2_{Y|X}(t,s)]$ is a measure of the variance of Y(t) left after the "projection" on $\mathbf{G}\{X(s)\}$ by considering the definition of $E[\sigma^2_{Y|X}(t,s)]$. Hence $E[\sigma^2_{Y|X}(t,s)]$ inherits many of the properties of $\theta_{Y|X}(t, s)$, e.g. $0 \le E[\sigma^2_{Y|X}(t,s)] \le \sigma^2_Y(t)$ (from **Property 1** and (3)).

Remark 3: The meaning of **Property 5** is that the estimated variance of Y(t) on the measure subspace generated by $X_1(s_1)$ and $X_2(s_2)$ is greater than or equal to the estimated variance of Y(t) on the one generated by only $X_1(s_1)$ or $X_2(s_2)$ and thus, by comparing the value of $\theta_{Y|X_1}(t, s)$ or $\theta_{Y|X_2}(t, s)$ with $\theta_{Y|X_1,X_2}(t, s, u)$ explores the dependency of Y(t) on $X_1(s_1)$ and/or $X_2(s_2)$.

Remark 4: The properties of the auto-dispersion functions and the multi-dimensional dispersion functions are the same as those of single dimensional cross-dispersion functions. **Property 5**, for the multi-dimensional case, can be more general written in the following way, for $M \le N$

$$\theta_{Y|X_1,X_2,...,X_M}(t, s_1, s_2, ..., s_M) \le \theta_{Y|X_1,X_2,...,X_N}(t, s_1, s_2, ..., s_N) .$$

3. The Determination of the Existence of Delay Elements in Nonlinear AR Models

Consider a discrete time nonlinear AR model of order N, i.e., a model in which y(n-N) is the longest delay element,

$$y(n) = f\,(y(n-1), y(n-2), ..., y(n-N)) + v(n)\,, \qquad (4)$$

where $f: \mathbf{R}^N \to \mathbf{R}$ is nonlinear and v(n) is an added white noise which is independent of the output prior to time n. The delay elements y(n), y(n-1), y(n-2), ..., y(n-N), and the input v(n) can be considered to be random processes and are denoted by $Y_0, Y_1, Y_2, ..., Y_N$ and v_0 respectively , and these random processes are assumed to be jointly stationary. Equation (4) can be rewritten as a probabilistic model

$$Y_0 = f(Y_1, Y_2 ..., Y_N) + v_0 \ . \qquad (5)$$

It is known that given (4) or (5), there are a variety of ways to express the system equation when the order of the model is not restricted to N. Because the random processes are correlated with each other, y(n-i) can be expressed by delay elements of higher delay than i , i.e., y(n-j)'s with j>i. In particular, this means that the model to be identified is not unique. Thus the original equation can be expressed in a variety of ways. In order to narrow down the description of a nonlinear model, two constraints are imposed.

Constraint 1: The nonlinear function f(·) should be expressed as simply as possible;

Constraint 2: The number of delay elements should be minimized.

Both constraints tend to reduce the effort required in the subsequent parametric system identification. The first constraint tends to simplify the nonlinear function f(·) of the model in (4) or (5), which in general implies the use of shorter delay elements. The second constraint tends to reduce the number of delay elements in the model so that the complexity of the nonlinear functions f(·) is minimized. Hence delay elements with shorter delay should be considered first. For example, if two models describe a given set of input/output data with equal accuracy, the simpler model (in turns of complexity of nonlinear functionals and in turns of the number of delay elements) should be prefered.

The order of the model in (5) can be determined from **Properties 3** and **5** described in the following proposition. Denote the dispersion function of the random processes of the sequence Y_i, i= 0, 1, 2, ..., as

$\theta_{Y_0|Y_1,Y_2,\ldots,Y_j}(t, s_1, s_2, \ldots, s_j) \equiv \theta_{Y_0|Y_1,Y_2,\ldots,Y_j}$,

where j= 1, 2, ...

Proposition 1: Consider the model in (5). Then $\theta_{Y_0|Y_1,Y_2,\ldots,Y_M} < \sigma^2_{Y_0} - \sigma^2_{v_0}$ iff M<N.

The above proposition suggests a heuristic way to determine the order of the model described in (5). Evaluate $\theta_{Y_0|Y_1}$, then $\theta_{Y_0|Y_1,Y_2}$ ... and so on until $\theta_{Y_0|Y_1,Y_2,\ldots,Y_N} = \sigma^2_{Y_0} - \sigma^2_{v_0} \pm \varepsilon$ for some N and some small ε>0. Of course, this heuristic way of determining the order of the model may contain delay elements which are redundant. The following proposition provides a way to exclude such redundant delay elements. Let f_j be the nonlinear functional that specifies the effect of y(n-j) in (4) or Y_j in (5) to the output process.

Proposition 2: Consider the model of (5). $f_j(\cdot)$, for some $j \in \{1, 2, \ldots, N\}$, is a zero mapping iff

$$\theta_{Y_0|Y_1,Y_2,\ldots,Y_{j-1},Y_{j+1},\ldots,Y_N} = \sigma^2_{Y_0} - \sigma^2_{v_0}.$$

Note that if the order of the model in (5) is known to be N, then the delay element y(n-N) should always be included.

Although the two propositions theoretically suggest ways to determine the order and delay elements in a model, it is generally not possible to determine the order precisely, nor the existence or nonexistence of delay elements exactly. This is due to the high correlation of the output processes. Moreover $\theta_{Y|X} = 0$ is only a necessary condition for Y to be independent of X as mentioned in Remark 1. However, the dispersion functions still provide a valuable information that shows how significant a delay element is. The dispersion function $\theta_{Y|X}$ provides information about the "power" (or variance) of Y when "projected" on $\mathbf{G}\{X\}$, which is equivalent to saying E[Y|X] is the best estimation of Y using X. To see this more clearly, consider the following deterministic nonlinear AR model

$$\begin{aligned} y(n) &= f_1(y(n-1)) + f_2(y(n-2)) + \ldots + f_N(y(n-N)) \\ &\equiv f_j(y(n-j)) + F(y(n-1), \ldots, y(n-j+1), y(n-j-1), \ldots, y(n-N)) \\ &\equiv f_j + F. \end{aligned} \quad (7)$$

The model in (7) is different from the model in (4) since there are no cross-terms among the delay elements. In this case, it is straightforward to show that

$$E[\,\sigma^2_{Y_0|Y_j}\,] = \sigma^2_F - \theta_{F|Y_j}. \quad (8)$$

It is then clear from (8) that $E[\,\sigma^2_{Y_0|Y_j}\,]$ is equal to the variance of the functional $F(\cdot)$ left after it is estimated via $\mathbf{G}\{Y_j\}$. Since $E[\,\sigma^2_{Y_0|Y_j}\,] = \sigma^2_{Y_0} - \theta_{Y_0|Y_j}$, from (8)

$$\theta_{Y_0|Y_j} = (\sigma^2_{Y_0} - \sigma^2_F) + \theta_{F|Y_j},$$

which clearly shows , as stated in Remark 2, $\theta_{Y_0|Y_j}$ is the variance of the random process Y_0 after projecting on $\mathbf{G}\{Y_j\}$. Note that $\theta_{Y_0|Y_j}$ is not only equal to $(\sigma^2_{Y_0} - \sigma^2_F)$ but an extra term $\theta_{F|Y_j}$ is added since the Y_i's are correlated to each other. The term $\theta_{F|Y_j}$ represents the estimated variance of the random process F (note that F is independent of Y_j) using $\mathbf{G}\{Y_j\}$. It can also be shown that $\theta_{Y_0|Y_j}$ is the best estimation of Y_0 on $\mathbf{G}\{Y_j\}$ because

$E\{\,(Y_0 - E[Y_0|Y_j])^2\,\} \le E\{\,(Y_0 - Y_j')^2\}$ for all $Y_j' \in \mathbf{G}\{Y_j\}$ .Therefore, instead of determining the existence of delay elements, the significance in the sense of "power" of delay elements are checked to determine their inclusion or exclusion in a given model. To explain the strategy, a simple second order example is discussed. Consider the nonlinear model

$$y(n) = f(y(n-1), y(n-2)) + v(n). \quad (9)$$

Proposition 1 shows that $\theta_{Y_0|Y_1,Y_2} = \sigma^2_{Y_0} - \sigma^2_{v_0}$. Three cases can occur:

1. $\theta_{Y_0|Y_1} \approx \theta_{Y_0|Y_1,Y_2}$;
2. $\theta_{Y_0|Y_1} \ll \theta_{Y_0|Y_1,Y_2}$ and $\theta_{Y_0|Y_2} \approx \theta_{Y_0|Y_1,Y_2}$;
3. Neither $\theta_{Y_0|Y_1} \approx \theta_{Y_0|Y_1,Y_2}$ nor $\theta_{Y_0|Y_2} \approx \theta_{Y_0|Y_1,Y_2}$, and $\theta_{Y_0|Y_1} + \theta_{Y_0|Y_2} \approx \theta_{Y_0|Y_1,Y_2}$.

These three criteria corresponding to situations when
1. The delay element y(n-1) is much more significant than y(n-2), and y(n-2) may be ignored.
2. The delay element y(n-2) is much more significant than y(n-1), and y(n-1) may be ignored.
3. Both y(n-1) and y(n-2) should be included in the model (9).

Case 1 may also have $\theta_{Y_0|Y_2} \approx \theta_{Y_0|Y_1,Y_2}$, which means y(n-2) can be included instead of y(n-1), but the constraints made previously suggest y(n-1) should be included but not y(n-2).

By extending these ideas, rules can be made to determine the significance of delay-element candidates in a general Nth order model as (4). Under the constraints discussed previously, rules to include or exclude a delay element are

Rule 1: If y(n-1), y(n-2), ..., y(n-j+1) are not included and $\theta_{Y_0|Y_j} \approx 0$ then y(n-j) may be excluded from the model.

Rule 2: If $\theta_{Y_0|Y_1, \ldots, Y_{j-1}} \approx \theta_{Y_0|Y_1, \ldots, Y_j}$ then y(n-j) may be excluded from the model.

Rule 3: If $\theta_{Y_0|Y_1, \ldots, Y_{j-1}} \ll \theta_{Y_0|Y_1, \ldots, Y_j}$ then y(n-j) should be included in the model.

Rule 1 is used to determine the leading delay element (the shortest delay element) in a model, and the following two rules determine if a delay element should be excluded from or included in the model. Hence the rules suggest a forward method to determine the order and significance of the various delay elements to the given model. In order to make the identification more precise, another rule is added.

Rule 4: Suppose $y(n-i_1), y(n-i_2), \ldots, y(n-i_p)$ are selected. If $\theta_{Y_0|Y_{i1} \ldots, Y_{ip}, Y_j} > \theta_{Y_0|Y_{i1} \ldots, Y_{ip}, Y_m}$, where j and $m \neq i_1, \ldots, i_p$, then y(n-j) should be included.

4. Simulation Results

In the previous sections, dispersion functions were used because they are more convenient theoretically. It is simpler to evaluate the expectation of the conditional variances of Y_0, i.e., to calculate $E[\sigma^2_{Y_0|Y_1;Y_2;\ldots;Y_M}]$, instead of $\theta_{Y_0|Y_1;Y_2;\ldots;Y_M}$.

As an example of the identification procedure, consider the "unknown" system,

$$y(n)=0.25\cos(2\pi y(n-3))+\exp(-|y(n-1)|)+ v(n),$$

where v(n) is N(0, 0.01). The initial conditions are y(0)=v(0) and y(i)=0 for all i<0. Five independent experiments are conducted. In each experiment, 10,000 data points of the output process are collected. The first 1,000 points data are discarded so that the data represents the behaviour of the output process in steady state. For different initial conditions, the numerical results of the expectations of the conditional variances are listed in **Table 1** ($\sigma^2_{Y_0}$ in column 0, $E[\sigma^2_{Y_0|Y_1}]$ in column 1, $E[\sigma^2_{Y_0|Y_2}]$ in column 2, $E[\sigma^2_{Y_0|Y_1;Y_2}]$ in column 3 and $E[\sigma^2_{Y_0|Y_1;Y_3}]$ in column 4).

0	1	2	3	4
0.0418	0.0338	0.0356	0.0226	0.0099
0.0400	0.0331	0.0361	0.0243	0.0105
0.0418	0.0327	0.0358	0.0213	0.0102
0.0395	0.0322	0.0346	0.0251	0.0110
0.0400	0.0330	0.0358	0.0206	0.0106

Table 1

For the five independent simulations, **Table 1** shows that the results are consistent. In particular,

1. $E[\sigma^2_{Y_0|Y_1}]$ is decreased from $\sigma^2_{Y_0}$;
2. $E[\sigma^2_{Y_0|Y_2}] > E[\sigma^2_{Y_0|Y_1}]$;
3. although $E[\sigma^2_{Y_0|Y_1;Y_2}] > E[\sigma^2_{Y_0|Y_1}]$, $E[\sigma^2_{Y_0|Y_1;Y_3}] \approx \sigma^2_{V_0}$.

Thus, by the rules discussed in the previous section, the order of the model is 3 and the significant delay elements in the model are y(n-1) and y(n-3), as desired.

5. Conclusions

This paper proposes the use of dispersion functions to determine the order and the significant delay elements of a nonlinear AR model. Note that the output processes of nonlinear models are highly correlated, hence this tool can be further extended to nonparametric identification of nonlinear MA models (with correlated input) and nonlinear ARMA models. Moreover, the method proposed is independent of the degree of nonlinearity of the models, hence it provides valuable information for parametric identification problems.

References

[1] N. S. Rajbman and A. T. Terekhin, "Variance ratio methods for random processes and their application to the analysis of nonlinear control plants," *Automation and Remote Control*, vol. 35, no. 3, pp. 496-506, 1965.

[2] N. S. Rajbman and V. M. Chadeev, *Identification of industrial processes*, Noth-Holland Publishing Co., 1980.

[3] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems, Part II: stochastic non-linear systems," *Int. J. Control*, vol. 41, no. 2, pp. 329-344, 1985.

[4] P. Eykhoff (editor), *Trends and progress in system identification, IFAC* series for graduates, research workers & Practising Engineers, vol. 1, pp. 189-195, 1981.

[5] K. Chung, *A course in probability theory*, 2nd edition, Academic Press, 1974.

[6] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems, Part II: stochastic non-linear systems," *Int. J. Control*, vol. 41, no. 2, pp. 329-344, 1985.