Brief paper

Application of multi-level signals to the identification of direction-dependent processes\footnote{This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Johan Schoukens under the direction of Editor T. Söderström. * Corresponding author. Fax: +603-8318-3029. E-mail addresses: hta@mmu.edu.my (A.H. Tan), k.godfrey@warwick.ac.uk (K.R. Godfrey), h.a.barker@swansea.ac.uk (H.A. Barker).}

A.H. Tan\textsuperscript{a,\footnote{This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Johan Schoukens under the direction of Editor T. Söderström. * Corresponding author. Fax: +603-8318-3029. E-mail addresses: hta@mmu.edu.my (A.H. Tan), k.godfrey@warwick.ac.uk (K.R. Godfrey), h.a.barker@swansea.ac.uk (H.A. Barker).}, K.R. Godfrey\textsuperscript{b}, H.A. Barker\textsuperscript{c}

\textsuperscript{a}Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia
\textsuperscript{b}School of Engineering, University of Warwick, Coventry, CV4 7AL, UK
\textsuperscript{c}School of Engineering, University of Wales Swansea, Swansea, SA2 8PP, UK

Received 16 April 2003; received in revised form 30 November 2003; accepted 10 December 2003

Abstract

Theoretical results for the output and the input–output crosscorrelation function are derived for first order direction-dependent processes perturbed using pseudo-random maximum length ternary (MLT) signals with unsymmetrical signal levels. The analytical results are valid if the process output is either always increasing or always decreasing when the input is at its median level of zero. It is shown that, although this is not normally the case for the whole of the signal period, good approximations to the analytical results are obtained when it is the case for most of the signal period, as can be achieved with a suitable choice of signal levels. MLT signals with unequal spacing between signal levels can also be used to minimise the nonlinear distortion. This is equivalent to compensating for the direction-dependent behaviour of the system by preceding it with a static nonlinearity. Based on the theoretical and simulation results obtained, a novel technique is proposed to allow the best linear approximation of the process to be estimated from the ratio between the amount of time when the output increases to that when it decreases. The proposed method is shown to be applicable even when the number of signal levels is greater than three, and is less susceptible to the effects of noise than the method of correlation analysis and least squares.

\textsuperscript{\copyright} 2003 Elsevier Ltd. All rights reserved.

Keywords: Correlation analysis; Direction-dependent dynamics; Maximum length signals; Multisine signals; Pseudo-random signals; System identification

1. Introduction

Direction-dependent processes have dynamics which depend on whether the process output is increasing or decreasing. Examples in industry include steam-raising plants (Godfrey & Briggs, 1972), gas turbines, chemical processes and nuclear reactors (Godfrey & Moore, 1974), polymerisation reactors (Xu, Jiang, & Zhu, 1994), distillation columns (Turner, Montague, & Morris, 1996), furnaces (Hågglund & Tengvall, 1995), thermomechanical pulp refiners (Rosenqvist, Eriksson, & Karlström, 2001) and automotive suspensions (Tan & Godfrey, 2001).

Hågglund and Tengvall (1995) describe these processes as “unsymmetrical”.

Theoretical results for direction-dependent processes are very limited compared with their common occurrence in industry. The complexity involved in obtaining analytical results is such that theoretical results have only ever been obtained for first order processes perturbed by binary and ternary signals. Godfrey and Briggs (1972), Tan and Godfrey (2001) and Barker, Tan, and Godfrey (2003) have described theoretical results for the input–output cross-correlation function obtained with pseudo-random maximum length binary (MLB) signals and their corresponding inverse-repeat signals. Tan (2003) has derived similar results with pseudo-random maximum length ternary (MLT) signals.

The theoretical crosscorrelation function is very important as it can be used to detect the direction-dependent behaviour by observing the presence of coherent discontinuities in the function. In addition, the rate of decay from the linear dynamics peak can be used to obtain the best linear approximation of the process, in response to a specified
perturbation signal. In the present paper, the best linear approximation is referred to as the related linear dynamics, following the terminology proposed by Schoukens, Dobrowiecki, and Pintelon (1998).

In Section 2, the results of Tan (2003) are extended to the case in which the levels of the MLT perturbation signal are unequally spaced about zero. As before, it is necessary to assume that the process output is unidirectional when the input is zero. It is, however, shown in Section 3 that by appropriate choice of the nonzero signal levels, this assumption can be made valid for more than 96% of the signal period. Although theoretical results are still obtainable for first order processes only, they serve to validate the experimental results, which may be extended to processes of higher order. This is important for applications of these unсимmetrical signals. It is shown in Section 3.2, for example, that the direction-dependent behaviour of the process may be minimised by appropriate choice of the nonzero signal levels, thus facilitating estimation of the related linear dynamics that represent the overall dynamics for a particular excitation signal, when a linear model is assumed.

In Section 4, it is shown that the related linear dynamics of the process can be reasonably well estimated using the ratio of the amount of time the output increases to that when it decreases; these dynamics are dependent on the perturbation signal applied.

2. Derivation of process output and input–output crosscorrelation function

Consider a first order direction-dependent process with unity gain and time constant $T_U$ when the output is increasing; and unity gain and time constant $T_D$ when the output is decreasing. If the process is perturbed using a ternary signal with levels $+x_U$, $0$ and $-x_L$, where the subscripts U and L denote ‘upper’ and ‘lower’ respectively, then the relationship between the sampled input $x$ and output $y$ is given by

$$y_t = x_t - (x_t - y_{t-1}) \exp(-T/T_U),$$

$$y_t = x_t - (x_t - y_{t-1}) \exp(-T/T_D),$$

where $t$ denotes time and $T$ is the clock-pulse interval which is also the sampling interval. Note that (1) excludes the trivial case where $x_t = y_{t-1} = 0$ as this represents the situation of zero input and zero output, and $y_t = 0$. Throughout this paper, it is assumed that $T_U < T_D$, that is the dynamics are faster when the output is increasing, but a similar analysis can also be carried out when $T_U > T_D$.

There is no known method to simplify (1) in general. In order to proceed, it is necessary to assume that $y_{t-1} > 0$ when $x_t = 0$. This assumption implies that the output is decreasing when the input is zero. Introducing two parameters

$$a = 0.5(\exp(-T/T_U) + \exp(-T/T_D)),$$

$$b = 0.5(\exp(-T/T_U) - \exp(-T/T_D)),$$

gives

$$y_t = (a + b)y_{t-1} + x_U(1 - a - b), \quad x_t = +x_U,$$

$$y_t = (a - b)y_{t-1}, \quad x_t = 0,$$

$$y_t = (a - b)y_{t-1} - x_L(1 - a + b), \quad x_t = -x_L.$$  

Eq. (3) can now be combined into a single expression for $y_t$. In order to simplify the analysis, it is advantageous to map the input levels $+x_U$, $0$ and $-x_L$ into symmetrical levels $+1$, $0$ and $-1$. This can be achieved by defining two additional parameters

$$A = (x_U + x_L)/2 \quad \text{and} \quad F = (x_U - x_L)/2.$$  

Combining (3) and (4),

$$y_t = [(a - b) + bx_t + bx_t^2]y_{t-1} + \left[\{A(1 - a) - Fb\} - \{Ab - F(1 - a)\}x_t x_t\right],$$

where $x_t$ now takes the values $+1$, $0$ and $-1$ following the transformation of the signal levels. This recursive form can be rewritten in a non-recursive form (Tan, 2003), in which the output is defined only by the past values of the input;

$$y_t = \sum_{p=0}^v \left[\{A(1 - a) - Fb\} - \{Ab - F(1 - a)\}x_{t-p}\right] x_{t-p} \prod_{l=0}^{p-1} ((a - b) + bx_{t-l} + bx_{t-l}^2),$$

where $v$ is an integer beyond which the terms in the series are negligible.

Eq. (6) can be simplified if the input signal possesses the shift-and-add property. This property enables the input product terms at the output to be replaced by a single delayed input term and is unique to maximum length signals based on Galois fields (GFs) (Barker, 1993, Chapter 11). An example is an MLT signal generated from an MLT sequence on Galois Nfields (GFs) (Barker, 1993, Chapter 11). An example is an MLT signal generated from an MLT sequence. Such a signal has length $N = 3^a - 1$, where $n$ is the number of stages in the shift register used to generate the sequence. Applying the shift-and-add property, contributions of terms of different orders can be summed to give the terms in the input–output crosscorrelation function (Tan, 2003). Since the MLT signal is inverse-repeat, having the second half period the negative of the first half, even order terms are zero. The linear term is given by

$$\Phi_1(iT) = \left(A(1 - a) - Fb - (Ab - F(1 - a)) \times \left(\frac{2b}{3(1 - d)} - \frac{2b}{3}\right)^i\right) R \left(\frac{2b}{3}\right)^i,$$

$$i = 0, 1, 2, \ldots,$$
where \( d = a - b \) and \( R = 2(3^n - 1)/N \). The derivation of (7) is very similar to the case for symmetrical inputs as shown in Tan (2003), and is omitted here due to space constraints. The linear term in the crosscorrelation function approximates the impulse response of the system, as the input MLT signal has an impulse-like autocorrelation function. This yields information about the related time constant, which is the time constant assuming the process is linear. From (7),

\[
T_{C,MLT} = -\frac{T}{\ln(d + 2b/3)}
\]

\[
= -\frac{T}{\ln\left(\frac{1}{3} \exp(-T/T_U) + \frac{2}{3} \exp(-T/T_D)\right)}.
\]

The subscript ‘MLT’ is used to show that the related time constant varies depending on the input signal applied. The coefficients \( \frac{1}{3} \) and \( \frac{2}{3} \) in the denominator of (8) reflect the assumption that the process output is decreasing when the input is zero.

Higher odd order terms contribute to discontinuities in the input–output crosscorrelation function (Barker & Obidegwu, 1973). The positions and magnitudes of these discontinuities are predictable and this can be used as a method to identify direction-dependent processes. The magnitudes of the discontinuities decrease with increasing order.

To illustrate the theory, a first order process was chosen in which \( T_U = 3T \) and \( T_D = 15T \). For this process, \( a = 0.8260 \) and \( b = -0.1095 \) (and therefore \( d = a - b = 0.9355 \)), from (2). The process was perturbed by an MLT signal generated from an MLT sequence in GF(3) with characteristic polynomial \( f(D) = 1 \oplus_3 D \oplus_3 D^3 \oplus_3 D^4 \oplus_3 D^5 \), where \( \oplus_3 \) denotes modulo-3 addition. The sequence elements 0, 1 and 2 were mapped into signal levels 0, +4 and -1 respectively. Hence, \( x_U = 4 \), \( x_L = 1 \), \( A = 2.5 \) and \( F = 1.5 \) from (4). The signal has length \( N = 242 \), with 81 clock-pulse intervals at each of the signal levels +4 and -1, and 80 clock-pulse intervals at signal level 0. For this signal, \( R = 2(3^{2.5} - 1)/242 = 0.6694 \). From (7), the linear dynamics term is given by

\[
\Phi_1(iT) = 0.2111(0.8625)^i, \quad i = 0, 1, 2, \ldots
\]

Hence, the related time constant \( T_{C,MLT} \) is 6.7611\( T \), calculated from (8).

The input–output crosscorrelation function, the process output and the magnitude of the output discrete Fourier transform (DFT) are plotted in Figs. 1–3 respectively. From Fig. 1, the effects of third order terms with a larger magnitude can be observed as discontinuities starting at lags 92, 68, 34, 32, 26 and 83. From Fig. 3, it can be seen from the output DFT (+) that the process is considerably nonlinear, as the magnitudes of the even harmonics are nonnegligible, and those at odd harmonics are significantly distorted. The output DFT of the best linear approximation (o) is included for comparison, and will be explained in Section 4.

It is worth noting that the input–output crosscorrelation function will remain unchanged if the dynamics in
the upward and downward directions of the output are interchanged. This is due to the inverse-repeat nature of the perturbation signal. Such a signal contains only odd harmonics, and has the second half of the signal the negative of the first half.

3. Application of maximum length ternary signals with unsymmetrical signal levels

By using an MLT signal with unequal spacing between signal levels and correctly choosing these levels, two important objectives, which are mutually exclusive, can be achieved. First, it is possible to increase the fraction of time in a period when the assumption given in Section 2, that \( y_{t-1} > 0 \) when \( x_t = 0 \), is valid; this facilitates the detection of the direction-dependent behaviour. Second, it is possible to minimise the nonlinear effects; this is useful in the estimation of related linear dynamics. Both may be viewed as preceding the direction-dependent process with a static nonlinearity, as shown in Fig. 4, which can be set to either maximise or minimise the overall nonlinearity. In the former case, the effect of the static nonlinearity enhances the nonlinearity in the direction-dependent process. In the latter case, the effects of the two nonlinearities compensate for each other, thus resulting in an approximately linear overall process. It should be noted that this technique is applicable only if the process can accept unsymmetrical signal levels; this is likely to be the case for most, if not all, of the direction-dependent processes listed in Section 1.

3.1. Signal design to ensure the validity of the assumption required for theoretical analysis

The assumption that \( y_{t-1} > 0 \) when \( x_t = 0 \) was investigated using the system shown in Fig. 4, perturbed by the MLT signal defined in Section 2. In this experiment, \( x_L = 1 \), \( T_U = 1 \) s and \( T = 1 \) s while \( x_U \) and \( T_D \) were varied. A measure of the validity of the assumption is the ratio between the number of clock-pulse intervals in a period when the assumption breaks down. \( N_{BK} \), and the signal period, \( N \). This can also be given in percentage as \( N_{PC} = 100(N_{BK}/N)\% \). The values for \( N_{BK} \) are plotted in Fig. 5.

From Fig. 5, it can be seen that the value of \( x_U \) for a given \( N_{BK} \) decreases as \( T_D \) increases. In general, the minimum difference between \( x_U \) and \( x_L \) decreases as the difference between \( T_U \) and \( T_D \) increases. The overall result is such that \( N_{BK} \) is closer to zero when \( x_U \gg x_D \) and \( T_U \ll T_D \), that is a combination of a greater input and a shorter time constant when the output is increasing. This is due to the fact that the output now stays positive for a large part of a period. Hence, when the input is at zero, the output will decrease towards this value. The graph of Fig. 5 is not particularly smooth due to the fact that \( N_{BK} \) is confined to integer values. It should also be noted that if \( N_{BK} \leq 12 \) for this input signal, \( N_{PC} \) is less than 5%. In general, for such a situation, the corresponding error introduced in the theoretical cross-correlation function expression would be reasonably small (Tan, 2003).

3.2. Signal design to minimise nonlinear distortion

The optimum values of \( x_U \) for the processes considered in Section 3.1, such that the nonlinear distortion is minimised in the frequency domain, were obtained by using the function \textit{fmin} in MATLAB. Since an MLT signal with symmetrical levels is inverse-repeat, its frequency spectrum contains only odd harmonics. The program attempts to minimise the ratio between the sum of the absolute magnitudes of the even harmonic components in the system output and the corresponding sum of the odd harmonic components. In doing this, the constant (DC) and Nyquist frequencies were excluded from the optimisation. The static nonlinearity as shown in Fig. 4 therefore acts as a compensator for the direction-dependent process by cancelling part of the nonlinear distortion caused by the latter. The optimum values of the upper signal level \( x_U \) are plotted as circles in Fig. 5, from which it can be seen that the optimum value of \( x_U \) decreases as \( T_D \) increases.

3.3. Extension of nonlinear compensation to other forms of nonlinearity

The technique of minimising the nonlinear distortion by cascading the process with a static nonlinearity has several
potential applications. Such a compensation scheme can be applied in feedback control strategies. One form of this, using a proportional-integral-derivative (PID) controller with a static nonlinearity, has been proposed by Burnham, Dunoyer, and Marcroft (1999). They showed that this control strategy provides an improved consistency of performance over a wider operating range. Application to higher order direction-dependent and bilinear processes, and other nonlinear processes, is a topic of further research.

4. Estimation of related linear dynamics

When the nonlinear distortion is minimised by the static nonlinearity, the assumption incorporated in (3B) breaks down for a significant part of the period. Thus, Eq. (8), which was derived for the case when the output increases approximately one-third of the time, is no longer valid, as the output now increases approximately half of the time. The experimental values of the related time constant $T_C$, obtained using correlation analysis and least squares (Isermann & Baur, 1974), are therefore closer to the theoretical result for a first order direction-dependent process perturbed using an MLB signal, or its corresponding inverse-repeat signal:

$$T_{C,MLB} = -T/\ln(a).$$

The argument above suggests that it may be possible to estimate the related time constant using the ratio between the number of clock-pulse intervals when the output is increasing to that when it is decreasing, leading to the expression

$$T_{CR} = -T/\ln(f r_U \exp(-T/T_U) + f r_D \exp(-T/T_D)),$$

where $f r_U$ and $f r_D$ denote the fraction of time in a period when the output increases and decreases, respectively.

The proposed formula was tested on the first order processes of Section 3. The related time constants calculated according to (8)–(10) are tabulated in Table 1. In all cases, the values of $T_C$ are closest to those obtained using (10), the percentage deviations being shown in the right-hand column. These deviations are all less than 2%, thus demonstrating the validity of the formula proposed.

<table>
<thead>
<tr>
<th>$T_D$</th>
<th>$x_U$</th>
<th>$x_L$</th>
<th>$T_C$</th>
<th>$T_{C,MLT}$</th>
<th>$T_{C,MLB}$</th>
<th>$T_{CR}$</th>
<th>Dev. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.7320</td>
<td>1.409</td>
<td>1.561</td>
<td>1.391</td>
<td>1.403</td>
<td>-0.426</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.5624</td>
<td>1.684</td>
<td>1.960</td>
<td>1.634</td>
<td>1.677</td>
<td>-0.416</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.4474</td>
<td>1.876</td>
<td>2.255</td>
<td>1.798</td>
<td>1.887</td>
<td>+0.586</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.3707</td>
<td>2.018</td>
<td>2.483</td>
<td>1.916</td>
<td>2.024</td>
<td>+0.297</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.3164</td>
<td>2.129</td>
<td>2.663</td>
<td>2.004</td>
<td>2.171</td>
<td>+1.973</td>
<td></td>
</tr>
<tr>
<td>4.264</td>
<td>2.215</td>
<td>2.890</td>
<td>2.074</td>
<td>2.258</td>
<td>+1.941</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$x_L = 1, T_U = T = 1 s$ and $x_U$ is the optimum upper signal level for a particular value of $T_U$ in order to minimise nonlinear distortion. An MLB signal with $N = 242$ was used.

As a further study, the first order process with $T_U = 3T$ and $T_D = 15T$, considered previously in Section 2, was again simulated. MLB signals with $N = 242$ and two different sets of signal levels ($x_U = 4$, $x_L = 1$ (Case 1) and $x_U = 0.2755$, $x_L = 1$ (Case 2)) were applied. Case 2 corresponds to that for a minimum amount of nonlinear distortion. The input–output crosscorrelation function and the process output are illustrated in Figs. 6 and 7 respectively. It can be seen that the crosscorrelation function in Fig. 6 is much smoother than that in Fig. 1, while the output in Fig. 7 is more symmetrical about the zero value compared with that in Fig. 2.

Applying (10), it was found that $T_{CR}$ equals $6.745T$ and $5.650T$ for Cases 1 and 2 respectively while $T_C$ was estimated at $6.778T$ and $5.565T$ respectively. The DFTs of the linear systems with the above values of $T_{CR}$ were calculated.
and appropriately scaled to give the same power at the output as the direction-dependent systems. In Case 1, the scaling was such that $x_U = x_L = 2.281$ while in Case 2, $x_U = x_L = 0.406$. These linear approximations were then compared with those of the original direction-dependent process with added static nonlinearity. These are plotted in Figs. 3 and 8 for Cases 1 and 2, respectively. It can be seen that the fit is reasonable in Fig. 3 for the odd harmonics, but is considerably better in Fig. 8 for both odd and even harmonics, due to the nonlinear suppression. The average error for the magnitude of the odd harmonic components $E_{\text{odd}}$ and that for the even harmonic components $E_{\text{even}}$ are 0.4827 and 3.5153 respectively for Case 1; and 0.0459 and 0.2126 respectively for Case 2. The errors for Case 1 are larger than those for Case 2, which is evident from Figs. 3 and 8. In particular, $E_{\text{odd}}$ has improved by 90.5% and $E_{\text{even}}$ has improved by 94.0% from Case 1 to Case 2.

Finally, two signals with more than three levels and symmetrical around zero were considered; for both, the maximum and minimum amplitudes were set to +1 and −1 respectively. The first was a five-level signal with $N = 342$ based on GF(7) (Barker, 2001). The signal was mapped from sequence elements to signal levels such that $u(0) = 0$, $u(1) = 0.5$, $u(2) = -1$, $u(3) = 1$, $u(4) = -1$, $u(5) = 1$ and $u(6) = -0.5$. The second signal was a computerto-optimised multisine specified with $N = 256$ and uniform odd harmonics (Kollár, 1994). The even harmonics were suppressed in order to generate a signal with inverse-repeat property, to allow a direct comparison with the other signals above. Results obtained are given in Table 2.

From Table 2, it can be observed that, for both signals, a higher estimation accuracy is achieved when the difference between $T_U$ and $T_D$ is smaller. This is due to the nonlinear distortion being less apparent. The results obtained show that Eq. (10) gives a good approximation of the value of $T_C$, even in the presence of noise. It is interesting to note that the formula for $T_{CR}$ is in fact more robust than the estimation using correlation analysis and least squares, as the estimates obtained using the former method do not change very much in the presence of noise.

5. Conclusions

In the paper, theoretical results for first order direction-dependent processes have been extended to the case where the input is an MLT signal with unsymmetrical signal levels. These results are only valid if the process output meets a certain constraint regarding its direction. With a correct choice of signal levels, the constraint above can be more easily fulfilled. It was also shown that by correctly choosing the signal levels, the nonlinear distortion can be suppressed by minimising the even harmonics in the process output. This has potential applications in the modelling and control of other forms of nonlinearity as well, such as the bilinear nonlinearity. It also allows a more accurate estimation of the linear dynamics to be carried out. In the final part of the paper, a formula was proposed to enable the related linear dynamics of the process to be estimated, which is very useful particularly when exact analytical solutions are not available. This method makes use of the ratio between the amount of time the output increases to that when it decreases. The formula proposed was shown to be applicable even when the number of signal levels is greater than three, and is robust in the presence of noise.

References


**Ai Hui Tan** graduated with first class honors in Electronic Engineering from the University of Warwick in 1999. She was awarded a Ph.D. in 2002 for research done in the Systems Modelling and Simulation Research Group at the University of Warwick. Her research interests are in the fields of system identification, in particular the effects of nonlinearities on process models, and of control. She is currently a Lecturer at Multimedia University, Cyberjaya, Malaysia.

**Professor Keith Godfrey** is Head of the Systems Modelling and Simulation Research Group in the School of Engineering at the University of Warwick. He received the Doctor of Science degree from the University of Warwick in 1990 for publications with the collective title “Applications of Modelling, Identification and Parameter Estimation in Engineering and Biomedicine”. He is author, or co-author, of more than 170 papers and is a member of the IFAC Technical Committees on Biomedical Engineering and Control, and on Modelling, Identification and Signal Processing. He was awarded the Honeywell International Medal (2000/2001) of the Institute of Measurement and Control for distinguished contributions to control engineering.

**H. Anthony Barker** is Emeritus Professor of Electrical Engineering at the University of Wales Swansea. He received his B.Sc. degree from the University of Nottingham in 1952, and his Ph.D. from the University of Cambridge in 1963, having in the meanwhile worked for Rolls-Royce on engine and aircraft control systems. Since 1963, he has worked in the Universities of Glasgow, Aston and Wales, with research interests in System Identification and Computer-Aided Control System Design, and is the author of over 200 papers. He was awarded the Sir Harold Hartley Medal of the Institute of Measurement and Control in 1992, and was the Institute’s President in 1995. He has been a Member of Council of the International Federation of Automatic Control, and is currently a Member of their Policy Committee and the Technical Committee on Modelling, Identification and Signal Processing.