Chemical Engineering Science, Vol. 47, No. 4, pp. 809–820, 1992.
Printed in Great Britain.

0009–2509/92 $5.00 + 0.00

NONMINIMAL PREDICTIVE CONTROL

WEIPING LU† and D. GRANT FISHER
Department of Chemical Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2G6

(*Received for publication* 6 *August* 1991)

Abstract—A new, nonminimal predictive controller (NPC) is formulated by combining the nonminimal finite-impulse response (FIR) output predictor developed by Lu *et al.* (1990) with a long-range predictive control strategy. The predictor structure used in the proposed NPC does not have the truncation error associated with the truncated, impulse- or step-response predictor structure used in DMC or MAC. Furthermore, the adaptive version of the proposed NPC does not require the on-line solution of a Diophantine equation as required in GPC and uses a single adaptive predictor to produce a trajectory of predicted output values with equal error variances at each predicted point. The conditions under which this can be achieved are formally generalized and proven as the *equal-variance principle*. Obeying this principle facilitates minimization of a parameter identification cost function which is the sum of long-range (multi-step) prediction error squares instead of the sum of single-step prediction error squares. Therefore, the identification algorithm is the dual of the control calculation which aims at minimizing the long-range tracking error squares. This results in improved, more robust system performance in the presence of noise and unmodeled dynamics. Reduced versions of adaptive NPC are also possible, which significantly reduce the number of parameters to be identified on-line. Simulation results confirm the advantage of using the nonminimal FIR predictor and show that adaptive NPC can outperform existing long-range adaptive predictive controllers such as GPC.

1. INTRODUCTION

Long-range predictive control (LRPC) schemes [e.g. Richalet *et al.* (1976), Cutler and Ramaker (1980), Garcia and Morari (1982), Spripada and Fisher (1985) and Clarke *et al.* (1987a)] are widely used in process control applications and represent an important class of model-based predictive control schemes. An LRPC uses a specific model structure for the prediction of the multi-step output trajectory. Given an *assumed* process input–output relationship there are different methods to formulate the prediction model structure. The most commonly used prediction model structures, when reduced to their basic forms, are essentially of two types: minimal finite-impulse response (FIR) predictors and truncated infinite-impulse response (IIR) predictors. For example, if the assumed process input–output relationship is given by

$$q^{-1}B_1(q^{-1})/A_1(q^{-1})$$

then the minimal FIR predictor for one-step ahead prediction is

$$\hat{y}^1(t+1) = q[1 - A_1(q^{-1})]y(t) + B_1(q^{-1})u(t)$$

and the truncated IIR predictor is

$$\hat{y}^{ir}(t+1) = H_N(q^{-1})u(t) = (h_0 + h_1 q^{-1} + \cdots + h_{N-1}q^{-N+1})u(t)$$

where H_N is the N-term truncation of B_1/A_1. The prediction model structure used in GPC (Clarke *et al.*, 1987a) belongs to the first type while those used in DMC (Cutler and Ramaker, 1980), MAC (Richalet *et al.*, 1976) and MOCCA (Spripada and Fisher, 1985) belong to the second type.

Unmeasurable process disturbances are usually handled by additional modeling or *ad hoc* methods. For example, in GPC (Clarke *et al.*, 1987b) a disturbance filter model (T-filter) is used while in MAC (Richalet *et al.*, 1976) and DMC (Cutler and Ramaker, 1980) values of the past (measured) output prediction error are added to the future output prediction values to heuristically account for the effect of unmeasurable disturbances. Output prediction errors can be lumped with the unmeasurable disturbances and handled by these *ad hoc* methods. Therefore, there is a tendency to underestimate the importance of selecting a good prediction model structure, which may be justified for nonadaptive control systems but not for adaptive control systems.

1.1. *Adaptive multi-step prediction and identification*

In adaptive LRPC systems the choice of a prediction model structure has a significant effect on overall system performance. Shook *et al.* (1989, 1990) correctly point out that the cost function of the parameter identification algorithm should match that of the LRPC law, i.e. the parameter identification scheme should minimize the multi-step prediction error squares:

$$J_{\mathrm{LRPI}}(N) = \sum_{t=0}^{N-N_2}\sum_{j=1}^{N_2}[y(t+j) - \hat{y}(t+j|t)]^2 \quad (1)$$

instead of the conventional one-step prediction error squares:

$$J_{\mathrm{LS}}(N) = \sum_{t=0}^{N-1}[y(t+1) - \hat{y}(t+1|t)]^2 \quad (2)$$

†Present address: Control Systems Services, Cominco Ltd., Trail, British Columbia, Canada V1R 4L8.

where $\hat{y}(t+j|t)$ is the jth step ahead predicted value of $y(t+j)$ at time instant t, given a model parameter vector θ. The minimization is taken with respect to θ. Shook *et al*'s recommendation of eq. (1) can be supported by simulation examples and by theoretical analysis. *Ideally*, the output cost of an LRPC scheme (assuming a zero output set point for simplicity) is

$$\sum_{j=1}^{N2} [y(t+j)]^2$$

at each time instant t. Heuristically, the time average of the above cost function can be closely approximated by the time average of

$$\sum_{j=1}^{N2} [\hat{y}(t+j|t)]^2 + \sum_{j=1}^{N2} [y(t+j) - \hat{y}(t+j|t)]^2$$

considering the fact that a prediction function is usually orthogonal to its prediction error function. The control law of an LRPC can minimize only the first sum at each time instant. The second sum must be minimized (in a time average sense) by the identification law. From this point of view, the cost functions of the identification and control law should be *matched* in the sense that a multi-step control or prediction error trajectory should be used in both cost functions. This provides the justification for using eq. (1) instead of eq. (2) as the cost for the identification. With the minimal FIR type of prediction model, using eq. (1) instead of eq. (2) leads to a nonlinear identification scheme or equivalently to the use of a data prefilter (which is not known *a priori*) for ordinary least squares (LS). Unfortunately, nonlinear optimization and the design of a data prefilter are both complex and inconvenient procedures. If a truncated IIR type of prediction model is used, then minimizing the conventional one-step ahead prediction error squares in eq. (2) can be shown to be equivalent to minimizing the multi-step prediction error squares in eq. (1). However, truncated IIR prediction models require a large number of parameters to reduce the truncation error and this leads to difficulties in adaptive control applications. Thus neither the FIR nor the truncated IIR prediction model is ideal for use in adaptive LRPC schemes.

The NPC developed in this paper can be conveniently formulated using the multi-step identification cost function (1) and offers significantly greater design flexibility. The specific nonminimal model (Lu *et al.*, 1990) used in NPC is only one example of the nonminimal models which could be effectively used with an adaptive LRPC scheme and still guarantee the cost function match. The principles and the development used in the NPC should be applicable to the derivation of an adaptive LRPC scheme based on other nonminimal models.

2. MULTI-STEP PREDICTED TRAJECTORY WITH EQUAL VARIANCES

The problem of minimizing the identification cost in eq. (1) can be considered as a simultaneous parameter fit for N_2 pairs of curves: $[y(t+1)$ vs $\hat{y}(t+1|t)$, $t = 0, 1, \ldots, N]$, $[y(t+2)$ vs $\hat{y}(t+2|t)$, $t = 0, 1, \ldots, N], \ldots, [y(t+N_2)$ vs $\hat{y}(t+N_2|t)$, $t = 0, 1, \ldots, N]$. It is quite natural that parameters of the N_2 prediction models are correlated, and hence, the minimization is nonlinear in most cases. However, by using a nonminimial FIR predictor model structure the nonlinear identification problem can be avoided. The conditions under which this is possible are stated in the following theorem.

Theorem 1 (equal-variance principle): *Let* $\hat{y}(t+j|t)$, $j = 1, 2, \ldots, N_2$, *be the predicted values of* $y(t+j)$ *produced at time instant* t *by a multi-step prediction model given a prediction model parameter vector* θ. *If the multi-step prediction model structure satisfies*

$$\hat{y}(t+j|t) = \hat{y}(t+j|t+i-1) \text{ for } 1 \leqslant i \leqslant j \leqslant N_2 \tag{3}$$

and is subject to the boundedness assumption of prediction errors, i.e. $[y(t+j) - \hat{y}(t+j|t)]^2 \leqslant c < \infty$ *for* $t \geqslant 0$ *and* $1 \leqslant j \leqslant N_2$, *then for any* $1 \leqslant i \leqslant N_2$ *the time averages of* i*-step prediction error squares are the same as those of multi-step prediction error squares, i.e.*

$$\lim_{N\to\infty} \left\{ (1/N) \sum_{t=0}^{N-1} (1/N_2) \sum_{j=1}^{N2} [y(t+j) - \hat{y}(t+j|t)]^2 - (1/N) \sum_{t=0}^{N-1} [y(t+i) - \hat{y}(t+i|t)]^2 \right\} = 0$$

$$i = 1, 2, \ldots, N_2. \tag{4}$$

This implies, in particular for $i = 1$, that the minimization of one-step prediction error squares in eq. (1) and the minimization of multi-step prediction error squares in eq. (2) are equivalent.

Proof: For $1 \leqslant i \leqslant j \leqslant N_2$ the range of $j - i$ and that of $i - 1$ are the same. Therefore from eq. (3) and the boundedness assumption of prediction errors

$$\sum_{t=0}^{N-1} [y(t+j) - \hat{y}(t+j|t)]^2 = \sum_{t=0}^{N-1} [y(t+j) - \hat{y}(t+j|t+j-i)]^2 = \sum_{t=j-i}^{N+j-1-i} [y(t+i) - \hat{y}(t+i|t)]^2 = \sum_{t=0}^{N-1} [y(t+i) - \hat{y}(t+i|t)]^2 + k_1(i,j,t) \quad 1 \leqslant i \leqslant j \leqslant N_2$$

where $k_1(i,j,t)/N \to 0$ uniformly as $N \to \infty$. For $1 \leqslant j \leqslant i \leqslant N_2$ interchanging i and j in the above equation gives

$$\sum_{t=0}^{N-1} [y(t+j) - \hat{y}(t+j|t)]^2 = \sum_{t=0}^{N-1} [y(t+i) - \hat{y}(t+i|t)]^2 - k_1(j,i,t) \quad 1 \leqslant j \leqslant i \leqslant N_2.$$

The expression inside the curly brackets of eq. (4) can

then be reorganized as

$$(1/N)(1/N_2)\sum_{j=1}^{N-2}\left\{\sum_{i=0}^{N-1}[y(t+j)-\hat{y}(t+j|t)]^2 - \sum_{i=0}^{N-1}[y(t+i)-\hat{y}(t+i|t)]^2\right\}$$

$$i = 1, 2, \ldots, N_2.$$

Equation (4) is then obtained by comparing this expression with the above two equations. ■

Note that the validity of eq. (4) also implies that the time average of the N_2 prediction error variances are the same (if these variances exist), i.e.

$$\lim_{N\to\infty}\left\{(1/N)\sum_{i=0}^{N-1}(1/N_2)\sum_{j=1}^{N-1}E\{[y(t+j) - \hat{y}(t+j|t)]^2\} - (1/N)\sum_{i=0}^{N-1}E\{[y(t+i)-\hat{y}(t+i|t)]^2\}\right\} = 0$$

$$i = 1, 2, \ldots, N_2. \quad (5)$$

The implications of theorem 1 are important, i.e. the minimization of the multi-step cost in eq. (1) can be achieved by minimizing the single-step cost in eq. (2) if the theorem conditions are satisfied. It should also be noted that this theorem should be used mainly in dealing with *model mismatch*. When external noise is involved in the process, a noise filter should be used in the prediction model structure so that the effect of the noise is appropriately separated from the effect of model mismatch. The theorem is also applicable in the sense that the multi-step prediction errors due to model mismatch have equal variances (cf. remark 4 in Section 4).

As shown below, the use of a nonminimal FIR predictor model structure satisfies condition (3) of theorem 1. Furthermore a linear estimation procedure such as recursive least squares (RLS) can be used since the prediction model is linear in its parameters.

3. NONMINIMAL FIR PREDICTION

In this section, three basic types of prediction models are briefly reviewed. To facilitate comparison, no additional disturbance handling method is included in the formulation such as filtering or feedback of the prediction errors.

The assumed process input–output relationship is given by the following transfer function

$$q^{-1}B_1(q^{-1})/A_1(q^{-1}) \quad (6)$$

where

$$A_1(q^{-1}) = 1 + a_{11}q^{-1} + \cdots + a_{1n}q^{-n} = \prod_{i=1}^{n}[1-(\lambda_i q)^{-1}] \quad \lambda_i \neq 0 \quad 1 \leqslant i \leqslant n \quad (7)$$

$$B_1(q^{-1}) = b_{10} + b_{11}q^{-1} + \cdots + b_{1s}q^{-s} \quad b_{10} \neq 0. \quad (8)$$

Truncated IIR predictor: The multi-step predictor is given by (Richalet *et al.*, 1976)

$$\hat{y}^{tr}(t+j|t) = H_N(q^{-1})u(t-1+j) \quad j = 1, 2, \ldots, N_2 \quad (9)$$

where $N \geqslant N_2$ and

$$H_N(q^{-1}) = h_0 + h_1q^{-1} + \cdots + h_{N-1}q^{-N+1} \quad (10)$$

is the N-term truncation of B_1/A_1.

Minimal FIR predictor: The multi-step predictor is given by

$$\hat{y}^1(t+j|t) = F_j(q^{-1})y(t) + E_j(q^{-1})B_1(q^{-1})u(t-1+j) \quad j = 1, 2, \ldots, N_2 \quad (11)$$

where

$$E_j(q^{-1}) = 1 + e_1q^{-1} + \cdots + e_{j-1}q^{-j+1} \quad j = 1, 2, \ldots, N_2 \quad (12)$$

and

$$F_j(q^{-1}) = f_{j0} + f_{j1}q^{-1} + \cdots + f_{j(n-1)}q^{-n+1} \quad j = 1, 2, \ldots, N_2 \quad (13)$$

satisfy the Diophantine equations

$$1 = E_j(q^{-1})A_1(q^{-1}) + q^{-j}F_j(q^{-1}) \quad j = 1, 2, \ldots, N_2. \quad (14)$$

Note that the predictor (11) is the positional version of the multi-step predictor given by Clarke *et al.* (1987a).

Nonminimal FIR predictor: The multi-step predictor is given by (Lu *et al.*, 1990)

$$\hat{y}^J(t+j|t) = [1 - A_J(q^{-J})]y(t+j) + B_J(q^{-1})u(t-1+j) \quad j = 1, 2, \ldots, N_2 \quad (15)$$

where $J \geqslant N_2$ and

$$A_J(q^{-J}) \equiv A_1(q^{-1})P_J(q^{-1}) = 1 + a_{J1}q^{-J} + \cdots + a_{Jn}q^{-Jn} \quad (16)$$

$$B_J(q^{-1}) = B_1(q^{-1})P_J(q^{-1}) = b_{J0} + b_{J1}q^{-1} + \cdots + b_{Jr}q^{-r} \quad r = (J-1)n + s \quad (17)$$

$$P_J(q^{-1}) = \prod_{i=1}^{n}[1 + (\lambda_i q)^{-1} + \cdots + (\lambda_i q)^{2-J} + (\lambda_i q)^{1-J}]. \quad (18)$$

3.1. *Main features of the nonminimal FIR predictor* (15)

(1) Only one prediction model is required to predict the entire multi-step output trajectory. More precisely, condition (3), of theorem 1 is satisfied by the prediction structure. Therefore there is no need to change the ordinary LS formulation in order to make the cost functions of parameter identification and LRPC match

each other. This is in contrast to the minimal FIR predictor (11), where for $i \neq j$ $(1 \leqslant i, j \leqslant N_2)$ the equation for calculating the i-step ahead prediction is different from that required to calculate the j-step ahead prediction, and the cost function match can be done only through a nonlinear optimization procedure.

(2) The realization of transfer function (6) by eq. (15) is accurate, i.e. no approximation or truncation is made. This is in contrast to the truncated IIR predictor (9), where a very large number of parameters is usually required to reduce the truncation error.

(3) For stable processes the predictor is robust to measurement noise without the addition of a noise filter (Lu *et al.*, 1990). This is in contrast to the minimal FIR predictor type, where the prediction is extremely sensitive to measurement noise when no filter is used.

(4) The prediction model structure leads naturally to the use of a disturbance filter (cf. the next section) which is designed based on *a priori* knowledge of the spectral density of the external disturbances. This is in contrast to the truncated IIR predictor (9), where the typical way to compensate for external disturbances is by feedback of the prediction error residual. This approach of residual feedback is not restricted to truncated IIR predictors and can be used as well with the minimal and nonminimal FIR predictors (11) and (15).

(5) The number of parameters used in the nonminimal prediction model is greater than in the minimal case. However, on-line parameter identification is still feasible and the prediction model structure guarantees the unique convergence of the estimated parameters if there is sufficient excitation and no unmodeled dynamics (Lu, 1990). This is in contrast to the truncated IIR models, where the large number of parameters is inconvenient to identify on-line and the limiting values of the parameter estimates are input-signal dependent.

(6) In the adaptive control context, the "extra" parameters in the nonminimal model provide extra degrees of freedom which reduce the effect of noise and unmodeled dynamics and improve the quality of the output prediction. This is in contrast to the minimal FIR type, where the use of a prefilter (Clarke *et al.*, 1987b; McIntosh, 1988; Shook *et al.*, 1989, 1990) is necessary to overcome the adverse effect of noise and unmodeled dynamics. Even though more parameters must be identified, the net computational load is not increased since there is no need to solve a Diophantine equation on-line.

(7) The nonminimal model structure provides an excellent basis for using "reduced-parameter" models in adaptive LRPC algorithms (cf. Section 5).

4. NONMINIMAL FIR PREDICTOR BASED LRPC (NPC)

This section develops a nonminimal FIR predictor based LRPC (NPC). Although there are many varieties of LRPCs in the literature, the derivations are all quite similar once the prediction model is specified. Therefore, only the derivation of the GPC is presented. Other LRPCs, DMC or MAC, can be easily derived using the same nonminimal predictor models but are not discussed in this paper.

It is assumed that the actual process can be described by the following CARIMA model

$$A_1(q^{-1})y(t) = B_1(q^{-1})u(t-1) + C_1(q^{-1})\omega(t)/\Delta_J \tag{19}$$

where A_1 and B_1 are given in eqs (7) and (8),

$$C_1(q^{-1}) = 1 + c_{11}q^{-1} + \cdots + c_{1p}q^{-p} \tag{20}$$

$$\Delta_J \equiv 1 - q^{-J} \quad J > 0 \quad \text{and} \quad \Delta_0 \equiv 1 \tag{21}$$

and $\omega(t)$ is white noise with zero mean. For $J = 0$, eq. (19) is a CARMA model, which cannot describe nonstationary disturbances. However, with $J \geqslant 1$ eq. (19) is capable of describing two types of nonstationary disturbances frequently encountered in industrial applications: random steps occurring at random times and Brownian motion. In GPC (Clarke *et al.*, 1987a) $J = 1$ is selected and C_1 is treated as a user-specified design polynomial (T-filter) since successful on-line identification of C_1 is difficult. The same approach is followed below. However, J is set equal to the prediction horizon N_2 of the algorithm and C_1 is always designed such that all its roots are strictly inside the unit circle on the complex plane.

4.1. *Multi-step prediction* ($J = N_2$)

It is not difficult to verify (Lu *et al.*, 1990) that a nonminimal model of system (19) is given by

$$A_J(q^{-J})\Delta_J y(t) = \bar{B}(q^{-1})\Delta u(t-1) + \bar{C}(q^{-1})\omega(t) \tag{22}$$

where

$$\bar{B}(q^{-1}) = B_J(q^{-1})Q_J(q^{-1}) = \bar{b}_0 + \bar{b}_1 q^{-1} + \cdots + \bar{b}_m q^{-m} \quad m = r + J - 1 \tag{23}$$

$$\bar{C}(q^{-1}) = C_1(q^{-1})P_J(q^{-1}) = \bar{c}_0 + \bar{c}_1 q^{-1} + \cdots + \bar{c}_l q^{-l} \quad \bar{c}_0 = 1 \quad l = p + n(J+1) \tag{24}$$

$$\Delta \equiv 1 - q^{-1} = \Delta_J / Q_J(q^{-1})$$

$$\text{and} \quad Q_J(q^{-1}) = 1 + q^{-1} + \cdots + q^{-J+1} \tag{25}$$

where $A_J(q^{-J})$, $B_J(q^{-1})$ and $P_J(q^{-1})$ are given in eqs (16), (17) and (18). Based on model (22) the proposed multi-step predictor is

$$\hat{y}^J(t+j|t) = [1 - A_J(q^{-J})]\Delta_J y(t+j) + \bar{B}(q^{-1}) \times \Delta u(t-1+j) + y(t+j-J) + \delta \bar{C}_j(q^{-1})[y(t) - \hat{y}^J(t|t-1)]$$

$$j = 1, 2, \ldots, N_2 \tag{26}$$

where $\delta = 0$ or 1 and

$$\begin{aligned}\bar{C}_j(q^{-1}) &= q^j[\bar{C}(q^{-1}) \\ &\quad - (\bar{c}_0 + \bar{c}_1 q^{-1} + \cdots + \bar{c}_{j-1} q^{-j+1})] \\ &= \bar{c}_j + \bar{c}_{j+1} q^{-1} + \cdots + \bar{c}_{i-j} q^{-(i-j)} \\ &\qquad j = 1, 2, \ldots, N_2. \end{aligned} \quad (27)$$

Note that the first term of $1 - A_J$ is $a_{J1} q^{-J}$ and $j \leqslant N_2 = J$, thus the right-hand side of eq. (26) does not contain future y-values. In order to derive the control law the RHS of eq. (26) has to be grouped into "future" and "past" terms, i.e.

$$\begin{aligned}\hat{y}^J(t+1|t) &= \bar{b}_0 \Delta u(t) + r_1(t) \\ \hat{y}^J(t+2|t) &= \bar{b}_0 \Delta u(t+1) + \bar{b}_1 \Delta u(t) + r_2(t) \\ &\vdots \end{aligned} \quad (28)$$

where

$$\begin{aligned} r_1(t) &= [\bar{B} - \bar{b}_0]\Delta u(t) + [1 - A_J]\Delta_J y(t+1) \\ &\quad + y(t+1+J) + \delta \bar{C}_1[y(t) - \hat{y}^J(t|t-1)] \\ r_2(t) &= [\bar{B} - (\bar{b}_0 + \bar{b}_1 q^{-1})]\Delta u(t+1) \\ &\quad + [1 - A_J]\Delta_J y(t+2) + y(t+2-J) \\ &\quad + \delta \bar{C}_2[y(t) - \hat{y}^J(t|t-1)] \\ &\vdots \end{aligned} \quad (29)$$

4.2. *Long-range predictive control law*

The following LRPC cost function is considered (Clarke *et al.*, 1987a):

$$\begin{aligned} J_c(t) &= \sum_{j=N_1}^{N2} [\hat{y}^J(t+j|t) - y_{sp}(t+j)]^2 \\ &\quad + \sum_{j=1}^{Nu} \lambda[\Delta u(t+j-1)]^2 \quad \lambda \geqslant 0. \end{aligned} \quad (30)$$

At time t, $J_c(t)$ is minimized with respect to $\Delta u(t+j-1)$, $j = 1, 2, \ldots, Nu$, and subject to $\Delta u(t+j-1) = 0$, $j > Nu$. Only the first control action $\Delta u(t)$ is implemented. At time $t+1$ the procedure repeats and so on. For convenience $N_1 = 1$ will be assumed as in Clarke *et al.* (1987a). The vector form of eq. (30) can be written as

$$\begin{aligned} J_c(t) &= [\hat{Y}^J(t) - Y_{sp}(t)]^T[\hat{Y}^J(t) - Y_{sp}(t)] \\ &\quad + \lambda \tilde{U}^T(t)\tilde{U}(t) \end{aligned} \quad (31)$$

where

$$Y_{sp}(t) = [y_{sp}(t+1), \ldots, y_{sp}(t+N_2)]^T$$

$$\begin{aligned} \hat{Y}^J(t) &= [\hat{y}^J(t+1|t), \ldots, \hat{y}^J(t+N_2|t)]^T \\ &= \mathbf{G}\tilde{U}(t) + R(t), \end{aligned} \quad (32)$$

$$\tilde{U}(t) = [\Delta u(t), \ldots, \Delta u(t+Nu-1)]^T \quad (33)$$

$$R(t) = [r_1(t), \ldots, r_{N2}(t+N_2-1)]^T \quad (34)$$

and

$$\mathbf{G} = \begin{bmatrix} \bar{b}_0 & 0 & 0 & \cdots & 0 \\ \bar{b}_1 & \bar{b}_0 & 0 & \cdots & 0 \\ \vdots & & \bar{b}_0 & \ddots & \vdots \\ & & & & \bar{b}_0 \\ \vdots & & & & \vdots \\ \bar{b}_{N2-1} & \bar{b}_{N2-2} & \cdots & & \bar{b}_{N2-Nu} \end{bmatrix}. \quad (35)$$

Note that, in eq. (35), $\bar{b}_j$, $j = 0, 1, \ldots, N_2 - 1$, are the first N_2 terms of the unit-step response of the transfer function $q^{-1}B_1(q^{-1})/A_1(q^{-1})$, the proof of which is not difficult and is omitted here. Minimizing $J_c(t)$ leads to the following control action:

$$\begin{aligned} \Delta u(t) &= [1 \quad 0 \quad \ldots \quad 0][\mathbf{G}^T\mathbf{G} + \lambda I]^{-1}\mathbf{G}^T \\ &\quad \times [Y_{sp}(t) - R(t)]. \end{aligned} \quad (36)$$

Remarks:

(1) When the actual process is accurately represented by the assumed process (19) and C_1 is stable, it can be shown that the multi-step predictors [eq. (26) with $\delta = 1$] are minimal-variance predictors with prediction error variances given by

$$\sum_{i=0}^{j-1} \bar{c}_i^2 \quad j = 1, \ldots, N_2. \quad (37)$$

(2) For applications, where the process and measurement noise is mild, $\delta = 0$ is recommended for easier implementation. This recommendation is based on the fact that the nonminimality of the predictor structure is usually capable of absorbing the effect of noise into the values of the predictor parameters. In this case the parameters of A_J and B_J should be estimated directly from the input–output data of the process instead of being calculated from A_1 and B_1 via eqs (16) and (17).

(3) For applications, where noise and external disturbances are significant, use of a satisfactory C_1 is necessary and $\delta = 1$ is recommended. Another way to cope with such a situation is to use $\delta = 0$ and separately model the noise and external disturbances in terms of prediction residuals similar to the approach taken in MAC (Richalet *et al.*, 1976) and DMC (Cutler and Ramaker, 1980).

(4) When $\delta = 1$ and NPC is used adaptively, it is recommended that $\bar{C}(q^{-1})$ be selected off-line similar to the T-filter in GPC (Clarke *et al.*, 1987b). If the one-step prediction error,

$$\tilde{y}(t) \equiv y(t) - \hat{y}^J(t|t-1)$$

in the last term of eq. (26) is invariant to the on-line identified parameters, then minimizing the multi-step prediction error squares (1) recursively will result in a linear recursive identifica-

tion scheme very similar to RLS. In this case the multi-step prediction errors due to model mismatch, i.e.

$$[y(t+j) - \bar{C}(q^{-1})\omega(t)] - [\hat{y}(t+j|t) - \bar{C}_j(q^{-1})\tilde{y}(t)] \quad j = 1, 2, \ldots, N_2$$

will have equal variances.

(5) Since $\tilde{y}(t)$, the measurable current prediction error, is essentially an approximation of the current value of the white signal $w(t)$, it need not be strictly represented by $y(t) - \hat{y}_J(t|t-1)$ and can be generated by an independent one-step ahead adaptive prediction algorithm, e.g. RLS with the one-step ahead cost function (2), so that it does not rely on (i.e. is invariant to) the on-line identified parameters produced by recursively minimizing the multi-step prediction error cost function (1).

(6) When there is no external noise ($\delta = 0$), predictor (26) uses the most recent u-data but not the most recent y-data. (Note that it is still possible to obtain accurate output prediction even though the latest measured value, $y(t)$, is not used. This is obvious from the prediction models used in DMC or MAC.) When there is external noise ($\delta = 1$), the most recent y-data are used. In either case, $\delta = 0$ or 1, if the predictor is adaptive, the most recent y-data are always used to update the parameters of A_j and B_j, which in turn will affect the prediction.

5. ADAPTIVE NPC: STANDARD AND REDUCED ALGORITHMS

In this section, the NPC developed in Section 4 is extended to its adaptive counterpart (standard algorithm). Then a "reduced parameter" adaptive NPC (reduced algorithm) is proposed, which uses a significantly smaller number of identified parameters without loss of the main characteristics of the standard adaptive NPC. For simplicity, only $\delta = 0$ is considered in the formulation.

5.1. *Standard algorithm*

The predictor (26) with $\delta = 0$ can be written as

$$\hat{y}^J(t+j|t) = [1 - \hat{A}_J(q^{-1}, t)]\Delta_J y(t+j) + \hat{B}_J(q^{-1}, t)\Delta_J u(t-1+j) + y(t+j-J) \quad (38)$$

where at each time t, A_J and B_J in eq. (26) are replaced by their on-line estimates $\hat{A}_J(q^{-J}, t)$ and $\hat{B}_J(q^{-J}, t)$. In vector form [eq. (38)] can be written as

$$\hat{y}^J(t+j|t) = \phi(t+j)^T\hat{\theta}(t) + y(t+j-J) \quad (39)$$

where

$$\hat{\theta}(t) = [-\hat{a}_{J1}(t) \quad \ldots \quad -\hat{a}_{Jn}(t) \quad \hat{b}_{J0}(t) \quad \ldots \quad \hat{b}_{Jr}(t)]^T \quad (40)$$

and

$$\phi(t) = [\Delta_J y(t-J) \quad \Delta_J y(t-2J) \quad \ldots \quad \Delta_J y(t-nJ) \quad \Delta_J u(t-1) \quad \Delta_J u(t-2) \quad \ldots \quad \Delta_J u(t-r)]^T. \quad (41)$$

The parameter adaptation can be done using any applicable identification method, e.g. using ordinary RLS:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P(t-2)\phi(t-1)}{1 + \phi(t-1)^T P(t-2)\phi(t-1)} \times [y(t) - \hat{y}(t+1|t) - \phi(t-1)^T\hat{\theta}(t-1)] \quad t \geq 1 \quad (42)$$

and

$$P(t-1) = P(t-2) - \frac{P(t-2)\phi(t-1)\phi(t-1)^T P(t-2)}{1 + \phi(t-1)^T P(t-2)\phi(t-1)} \quad (43)$$

with $\hat{\theta}(0)$ and $P(-1) > 0$ given. The control law is given by

$$\Delta u(t) = [1 \quad 0 \quad \ldots \quad 0][\hat{\mathbf{G}}(t)^T\hat{\mathbf{G}}(t) + \lambda\mathbf{I}]^{-1}\mathbf{G}^T(t)[Y_{sp}(t) - \hat{R}(t)] \quad (44)$$

where $\hat{\mathbf{G}}(t)$ and $\hat{R}(t)$ are calculated at each time t using the parameter estimates.

5.2. *Reduced algorithm*

The development of the reduced algorithm is described in the following four steps. {For simplicity $s = n - 1$ [cf. eq. (8)] is assumed.}

Step 1: (*selection of a set of nominal parameters*). Obtain an estimate of $B_J(q^{-1})$ by some means, e.g. by using open-loop step response data of the process; by an off-line identification procedure; or by running eqs (42) and (43) for an initial period of time. Let this estimate be a nominal representation of $B_J(q^{-1})$ and denote it by

$$B_J^*(q^{-1}) = B_{J0}^*(q^{-1}) + B_{J1}^*(q^{-1})q^{-J} + \cdots + B_{J(n-1)}^*(q^{-1})q^{-(n-1)J} \quad (45)$$

where

$$B_{Jj}^*(q^{-1}) = b_{Jj}^1 + b_{Jj}^2 q^{-1} + \cdots + b_{Jj}^J q^{-(J-1)} \quad j = 0, 1, \ldots, n-1. \quad (46)$$

Step 2 (*reduced adaptive multi-step predictors*). The multi-step predictor has the same form as eq. (38), but

$$\hat{B}_J(q^{-1}, t) = \hat{b}_0(t)B_{J0}^*(q^{-1}) + \cdots + \hat{b}_{n-1}(t)B_{J(n-1)}^*(q^{-1})q^{-(n-1)J} \quad (47)$$

where $\hat{b}_j(t)$, $j = 0, 1, \ldots, n-1$, are estimated while $\hat{B}_{Jj}^*(q^{-1})$, $j = 0, 1, \ldots, n-1$, are kept constant. This reduces the number of on-line estimated parameters by $n(J-1)$, but the approach will obviously be influenced by the accuracy of the constant B^* parameters.

Step 3 (*parameter identification*). In this case the vector form of the multi-step predictor is given by

$$\hat{y}^J(t+j|t) = \phi_r(t+j)^T\hat{\theta}_r(t) + y(t+j-J) \quad (48)$$

where

$$\hat{\theta}_r(t) = [\, -\hat{a}_{J1}(t) \quad \ldots \quad -\hat{a}_{Jn}(t) \quad \hat{b}_0(t) \quad \ldots \quad \hat{b}_{n-1}(t)]^T \quad (49)$$

and

$$\phi_r(t) = [\Delta_J y(t-J)\ \Delta_J y(t-2J) \ \ldots\ \Delta_J y(t-nJ)\ [B^*_{J0}\Delta_J u(t-1) \quad B^*_{J1}\Delta_J u(t-1-J) \ \ldots \quad B^*_{J(n-1)}\Delta_J u(t-1-nJ+J)]^T. \quad (50)$$

The parameter adaptation is similar to eqs (42) and (43) except that $\theta_r(t)$ and $\phi_r(t)$ are used instead.

Step 4 (*control law*). The control law has the same form as eq. (44) except that the coefficients of $\hat{B}_J(q^{-1}, t)$ are calculated from eq. (47).

Note that both predictors (39) and (48) obey the equal-variance principle, i.e. they both satisfy eq. (3) of theorem 1. Therefore, in both cases the estimated parameters generated using ordinary RLS [which in general only minimizes the single-step prediction cost of eq. (2)] guarantee equal-variance prediction errors and the cost functions for identification and control are matched. Under conditions of unmodeled dynamics and severe measurement noise, obeying the equal-variance principle significantly improves the performance and robustness of the adaptive control system as is shown in the simulation results presented in the next section.

6. SIMULATION RESULTS

In this section the implementation and importance of the equal-variance principle (theorem 1) are illustrated by simulation examples using four different predictors: a minimal FIR predictor (11), a truncated IIR predictor (9), a nonminimal FIR predictor (15) and a reduced nonminimal FIR predictor. Closed-loop LRPC performance using an adaptive NPC and its reduced version are then simulated and compared with a standard adaptive GPC scheme (Clarke *et al.*, 1987a).

The actual process is produced by

$$y(s) = \frac{2 \times 229}{(s+1)(s^2+30s+229)}u(s) + w(s) \quad (51)$$

where w is the measurement noise. Note that this process was used in the work of Rohr *et al.* (1985) on robustness of adaptive control systems. The sampling period of the discrete models used in all the simulations is 0.1 s. With this sampling rate the discretized system is a nonminimum phase system, i.e. there is a zero outside the unit circle.

6.1. *Open-loop evaluation of four predictors*

The actual third-order process [eq. (51)] is assumed to be unknown and a first-order transfer function is used as the identified relationship between u and y in the prediction models, i.e. the discrete system of eq. (6) is assumed to be first-order. This leads to the following structures for the four predictor models used to predict $y(t+j)$, $j = 1, \ldots, 10$.

(1) Minimal FIR predictor (11):

$$\hat{y}^1(t+1|t) = -a_1 y(t) + b_0 u(t) \quad (52)$$

which is for one-step prediction, and the expressions for $\hat{y}^1(t+j|t)$, $j = 2, \ldots, 10$, are given in eq. (11) with $N_2 = 10$. The parameters to be identified are a_1 and b_0. $F_j(q^{-1})$ and $E_j(q^{-1})$ in eq. (11) are calculated using the estimates of a_1 and b_0.

(2) Truncated IIR predictor (9):

$$\hat{y}^{tr}(t+j|t) = h_0 u(t-1+j) + h_1 u(t-2+j) + \cdots + h_{22}u(t-23+j) \quad j = 1, \ldots, 10 \quad (53)$$

i.e. $N = 23$ in eq. (9). The 23 parameters, $h_i i = 0, \ldots, 22$, are all to be identified.

(3) Nonminimal FIR predictor (15):

$$\hat{y}^J(t+j|t) = -a_{J1}y(t-J+j) + b_{J0}u(t-1+j) + \cdots + b_{J9}u(t-10+j) \quad j = 1, \ldots, 10 \quad (54)$$

where $J = 10$. The 10 parameters, a_{J1} and b_{Ji1} $i = 1, \ldots 9$, are to be identified.

(4) Reduced nonminimal FIR predictor:

$$\hat{y}^{re}(t+j|k) = -a_{J1}y(t-J+j) + b_0[b^*_{J0} \times u(t-1+j) + \cdots + b^*_{J9}u(t-10+j)] \quad j = 1, \ldots, 10 \quad (55)$$

where only a_{J1} and b_0 are to be identified. b^*_{Ji}, $i = 1, \ldots, 9$, are prespecified impulse response values of the actual process calculated from the first 11 points in the step-response curve shown in Fig. 1. Measurement noise was added so that the measured curve (solid line) differs from the noise-free output response curve (dotted line) of the process. [The reduced algorithm for adaptive NPC referred to in Section 5 groups the B_J polynomial in the same way as eq. (55). The parameters with superscript "*" are determined off-line by any means and only the parameters without the superscript are identified on-line.]

Note that the minimal FIR predictor (11) is the only one among the above four predictors which does not obey the equal-variance principle [condition (3) of theorem 1].

In order to compare the open-loop performance of the four predictors let $w(t) \equiv 0$ and the input signal be

$$u(t) = \begin{cases} 1 + e(t) & 0 \leqslant t < 100 \\ -1 + e(t) & 100 \leqslant t \leqslant 200 \end{cases} \quad (56)$$

where $e(t)$ is $N(0, 1)$ for 200 samples. A batch least-squares algorithm is used to identify the parameters for each of the above four predictors. The cost function for the off-line identification is the conventional one-step prediction error type, eq. (2).

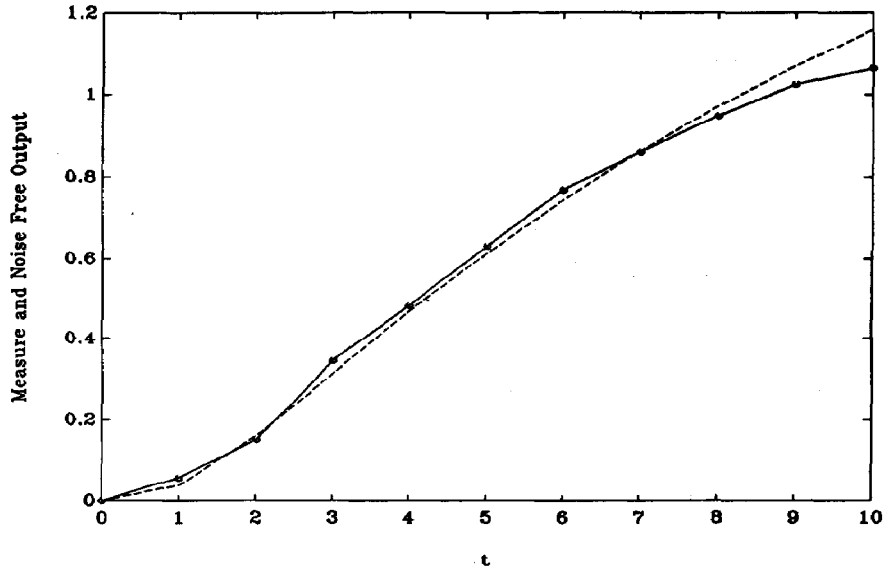

Fig. 1. Step response, (———) measured values, (– – – –) noise-free values.

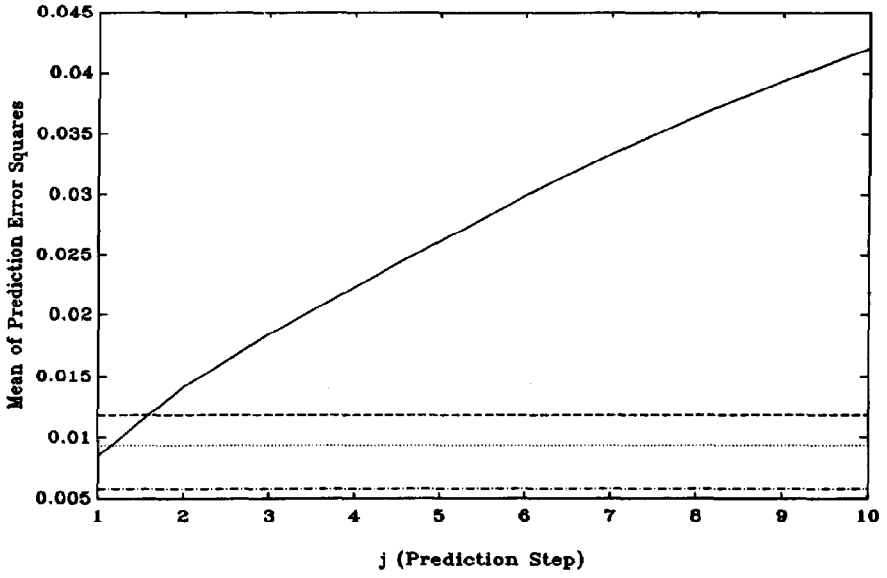

Fig. 2. Test of equal-variance principle, (———) minimal FIR, (– – – –) truncated IIR, (– · – · – · – ·) nonminimal FIR, (. . . . . .) reduced nonminimal FIR.

The four predictors with the parameters obtained using the batch least-squares algorithm are then compared using the same input sequence, eq. (50). For each point in the predictions, the mean of the prediction error squares of the four predictors, i.e.

$$(1/200)\sum_{t=1}^{200}[y(t+j)-\hat{y}(t+j|t)]^2 \qquad j=1,\ldots,N_2$$

is plotted versus the prediction step in Fig. 2, which shows:

(1) The minimal FIR predictor (11) has a very low mean of prediction error squares for one-step ahead prediction. However, the error increases significantly as the prediction range increases. Since the minimal FIR predictor (11) does not obey the equal-variance principle, this result is not a surprise.

(2) The other three predictors, (9), (15) and (49), all have an almost constant mean of prediction error squares for all prediction steps. These results verify the equal-variance principle of theorem 1.

(3) The truncated IIR predictor (9) does not yield the lowest mean of prediction error squares even though it uses the largest number of identified parameters, i.e. 23. In contrast the reduced nonminimal FIR predictor (49) gives a lower prediction error than the truncated IIR predictor (9) even though it uses only two identified parameters. This comparison implies that for prediction accuracy the structure of the predictor is more important than the number of its identified parameters.

(4) The nonminimal FIR predictor (15) with 10 identified parameters has the lowest prediction error squares.

6.2. *Closed-loop performance of adaptive NPC and GPC*

Figures 3–8 show the closed-loop responses of adaptive GPC (Clarke *et al.*, 1987a), adaptive standard NPC and adaptive reduced NPC. For all of these simulation runs the initial parameters were set equal to parameter estimates obtained from the off-line,

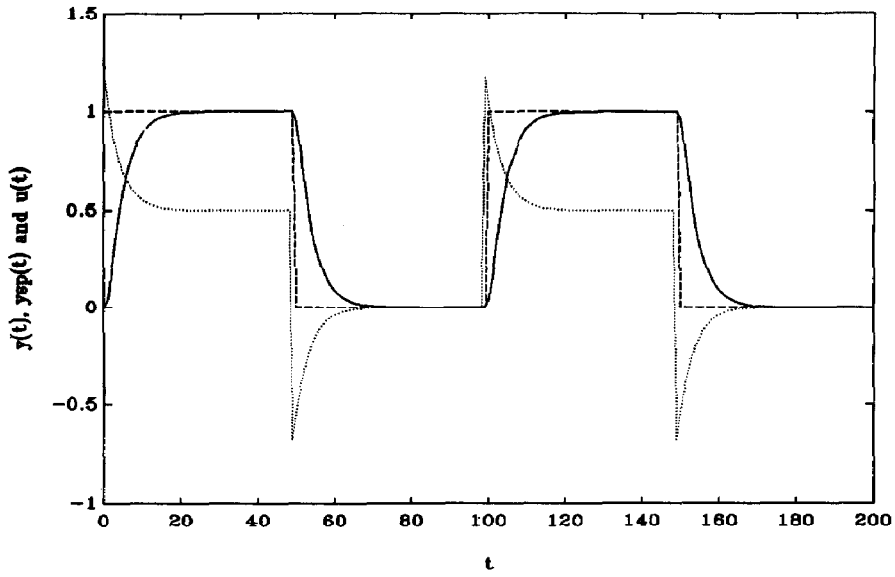

Fig. 3. Response of GPC, $\omega(t) = 0$, (———) $y(t)$, (– – – –) $y_{sp}(t)$, (. . . . .) $u(t)$.

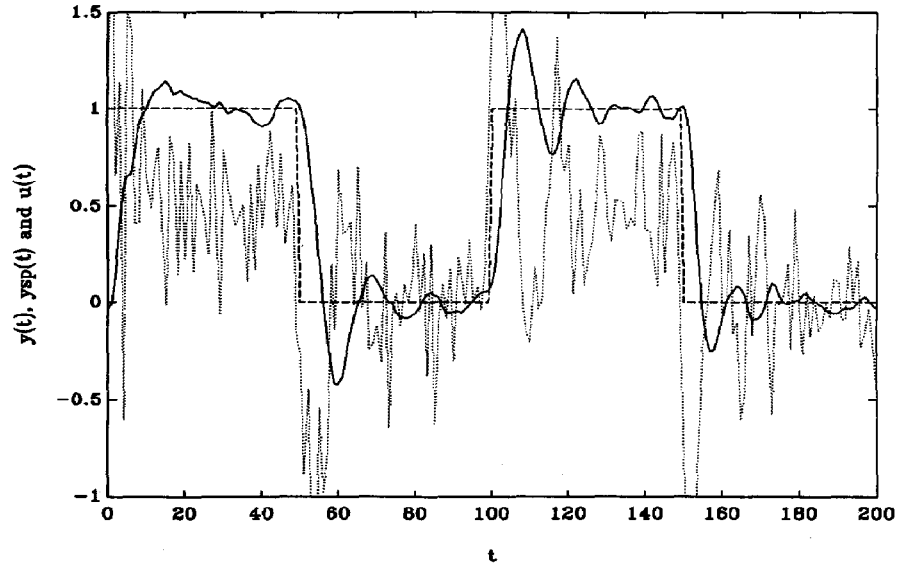

Fig. 4. Response of GPC, $\omega(t) \neq 0$, (———) $y(t)$, (– – – –) $y_{sp}(t)$, (. . . . .) $u(t)$.

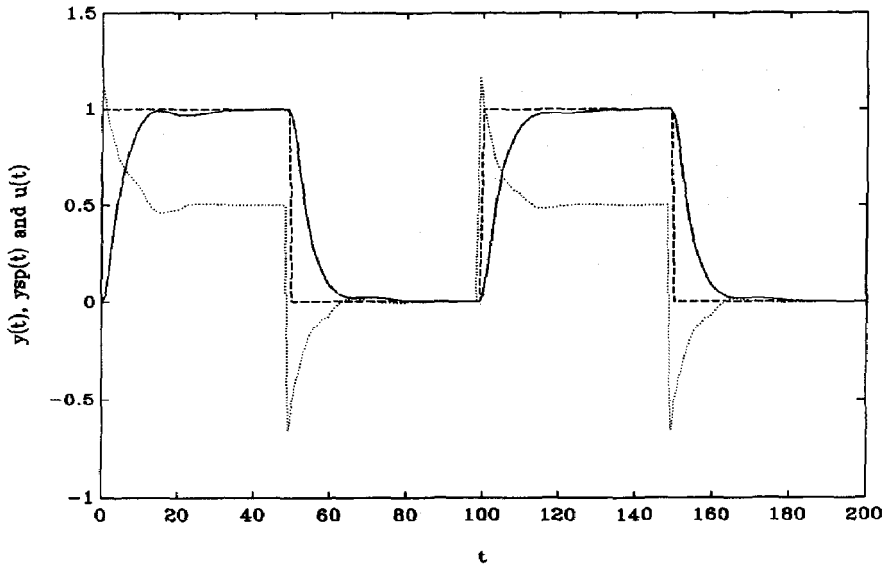

Fig. 5. Response of NPC, $\omega(t) = 0$, (———) $y(t)$, (– – – –) $y_{sp}(t)$, (. . . . .) $u(t)$.

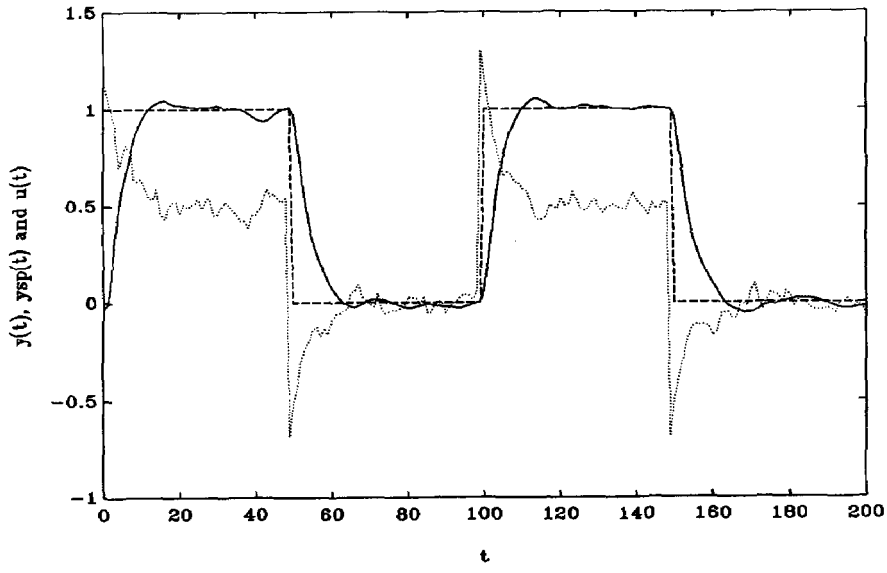

Fig. 6. Response of NPC, $\omega(t) \neq 0$, (———) $y(t)$, (– – – –) $y_{sp}(t)$, (. . . . .) $u(t)$.

open-loop test described above. Ordinary recursive least squares was used for the on-line identification and the controller settings for all these simulation runs were the same: $N_2 = 10$ and $\lambda = 0$. The controlled responses without measurement noise are shown by the solid lines in Figs 3, 5 and 7. Figures 4, 6 and 8 show the corresponding responses with measurement noise $N(0, 0.1)$. The manipulated variable, $u(t)$, is shown by the dotted line.

With GPC the use of a first-order prediction model resulted in *unstable responses* (not plotted). Therefore the adaptive GPC responses shown in Figs 3 and 4 are based on the use of a third-order prediction model. Since measurement noise is not present and there is no model structure error, the response shown in Fig. 3 represents the best performance achievable by GPC or NPC with the given controller settings. For the same measurement noise-free case, the response of the standard NPC (shown in Fig. 5) is almost identical even though it is based on a first-order model of the process. The response of the reduced NPC shown in Fig. 7 is not as good but is still impressive considering

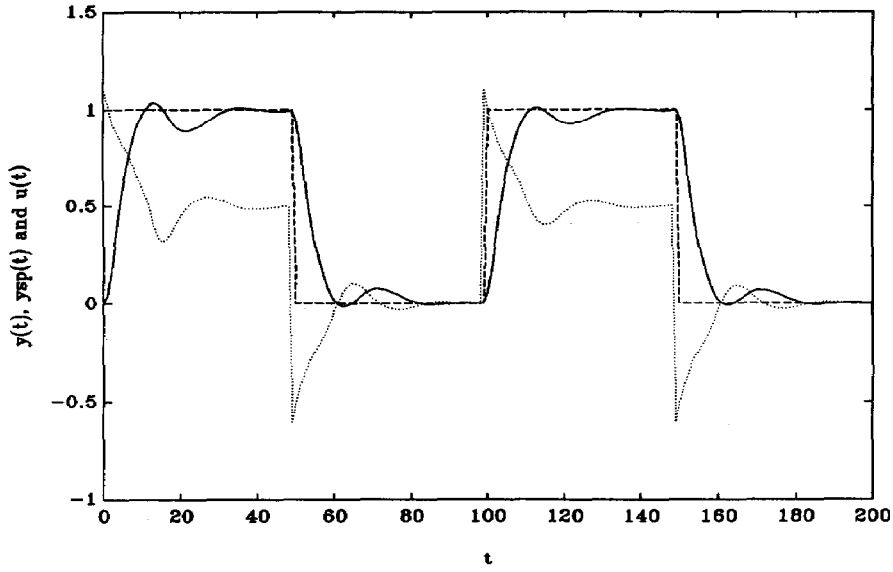

Fig. 7. Response of reduced NPC, $\omega(t) = 0$, (———) $y(t)$, (– – – –) $y_{sp}(t)$, (. . . . .) $u(t)$.

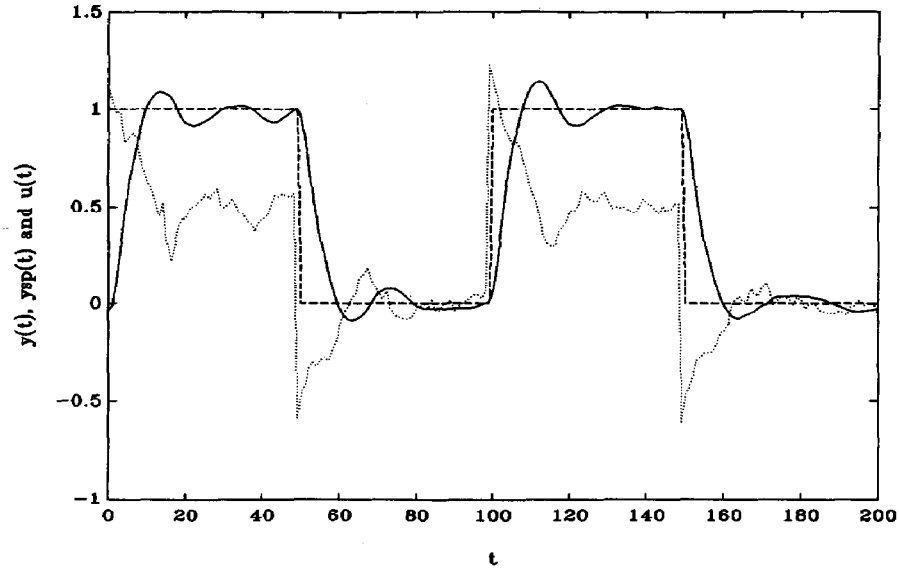

Fig. 8. Response of reduced NPC, $\omega(t) \neq 0$, (———) $y(t)$, (– – – –) $y_{sp}(t)$, (. . . . .) $u(t)$.

the fact that adaptive GPC with the *same number of* identified parameters (first-order prediction model with two identified parameters) resulted in an *unstable adaptive control system.*

When measurement noise is present, adaptive GPC (Fig. 4) even with a third-order prediction model produces a strong, noisy control action and an oscillatory process output with large overshoot magnitudes. In contrast, the two NPCs (cf. Figs 6 and 8) yield much smoother control action and more satisfactory process output performance.

Note that the response of all the adaptive controllers in the presence of noise and unmodeled dynamics could be improved by the use of appropriate filters, e.g. using a T-filter for GPC (Clarke *et al.*, 1987b) or similarly by using an appropriate C_1-polynomial with $\delta = 1$ for NPC. The topic of filter design is important but is beyond the scope of this paper.

Comparisons of adaptive controllers are of course application dependent and a large number of cases must be evaluated before generalizations can be made. However, results obtained to date, such as those pre-

sented above, plus the theoretical basis of the formulation indicate that NPC can outperform existing methods such as DMC (Cutler and Ramaker, 1980) and GPC (Clarke *et al.*, 1987a).

7. CONCLUSIONS

Better LRPC can be achieved by using a long-range (multi-step) cost function for both the parameter estimation and control calculations (Shook *et al.*, 1989, 1990). The equal-variance principle stated in theorem 1 indicates the conditions under which this can be done without resorting to a nonlinear optimization solution.

The NPC proposed herein uses a nonminimal FIR predictor which obeys the equal-variance principle. NPC therefore has the advantage of an identification cost function that is the dual of that used for control, i.e. does not require on-line solution of a Diophantine identity, requires only one equation to predict the entire output trajectory, and is less sensitive to noise. However, the main significance of NPC is it shows that the use of nonminimal models can be advantageous for LRPC and produces a different predictor structure which, for example, separates the model-based prediction from the noise modeling functions.

Acknowledgement—The authors would like to thank Professor Lixin Peng for a careful review of the original version of this paper. His constructive comments have helped to improve the quality of this paper.

REFERENCES

Anderson, B. D. O. and Moore, J. B., 1979, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ.

Astrom, K. J. and Wittenmark, B., 1990, *Computer-controlled Systems: Theory and Design*. Prentice-Hall, Englewood Cliffs, NJ.

Clarke, D. W., Mohtadi, C. and Tuffs P. S., 1987a, Generalized predictive control—Part 1. The basic algorithm. *Automatica* **21**, 137.

Clarke, D. W., Mohtadi, C. and Tuffs, P. S., 1987b, Generalized predictive control—Part II. Extensions and interpretations. *Automatica* **21**, 149.

Cutler, C. R. and Ramaker, B. L., 1980, *Dynamic Matrix Control—a Computer Control Algorithm*. JACC, San Francisco, CA.

Garcia, C. E. and Morari, M., 1982, Internal model control. 1. A unifying review and some new results. *Ind. Engng Process Des. Dev.* **21**, 308.

Elliot, D. F., 1987, *Handbook of Digital Signal Processing Engineering Application*. Academic Press, San Diego, CA.

Goodwin, G. C. and Sin, K. S., 1984, *Adaptive Filtering Prediction and Control*. Prentice-Hall, Englewood Cliffs, NJ.

Kailath, T., 1980, *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Lu, W., 1989, Parameter estimation and multirate adaptive control. Ph.D. thesis, University of Alberta, Edmonton.

Lu, W., Fisher, D. G., Shah, S. L. and Mohtadi, C., 1990, Nonminimal model based output predictors, in *Proceedings of the American Control Conferences* pp. 1607–1613. San Diego, CA.

McIntosh, A. R., 1988, Performance and tuning of adaptive generalized predictive control. M.Sc. thesis, University of Alberta, Edmonton.

Ogata, K., 1987, *Discrete-time Control Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Richalet, J., Rault, A., Testud, J. L. and Papon, J., 1976, Model predictive heuristic control: application to industrial processes. *Automatica* **18**, 413.

Rohr, C. E., Valavani, L., Athans, M. and Stein, G., 1985, Robustness of continuous-time adaptive control algorithms in the presence of unmodeled dynamic. *IEEE Trans. Autom. Control* **AC-30**, 881.

Shook, D. S., Mohtadi, C. and Shah, S. L., 1989, Identification for long range predictive control. *IEE Proc. D Control Theory Appl.* (in press).

Shook, D. S., Mohtadi, C. and Shah, S. L., 1990, Adaptive filtering and GPC, in *Proceedings of the American Control Conference*, pp. 556–561. San Diego, CA.

Spripada, N. R. and Fisher, D. G., 1985, Multivariable optimal constrained control algorithm (MOCCA): Part 1. formulation and application, in *Proceedings, International Conference on Industrial Process Modeling and Control*, Vol. 1, Hangzhou, China.