MPC closed-loop identification without excitation *

Yun Zhu, Wengang Yan, Yucai Zhu

State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027 China (e-mail: 21832028@zju.edu.cn).

Abstract: This paper presents a method of closed-loop identification for multivariable systems without external excitation. The method is specially designed for model predictive control (MPC) systems. Without using external excitation (test signals), the method ensures the informativity of the closed-loop data and, at the same time, improve the control performance during the test period. The purpose of the study is to reduce the cost of identification test. The basic idea is to switch the input weighting matrix in the MPC controller which leads to the informativity of the data-set. A preliminary test is carried out in order to find a new input weighting matrix which improve the control performance; then a switching scheme is developed based on the two weighting matrices. Traditional simulation based model validation no longer works in closed-loop identification without excitation, and model error bounds on the frequency responses can be used instead. The effectiveness of the proposed method is demonstrated by a simulation study.

Keywords: model predictive control; control performance; model mismatch; persistent excitation condition.

1. INTRODUCTION

In process industries, model predictive control (MPC) has been proven the most powerful control technology since it was proposed by Richalet et al. (1978) and Cutler and Ramaker (1980). Compared to some traditional methods like PID (proportional-integral-derivative) control, the advantages of MPC are its ability to solve multivariable control problems and to take operation constraints into consideration. For a model-based control algorithm like MPC, the performance strongly depends on the quality of the model. The model predicts the output of the process and the control behavior is determined by optimizing a performance index. An accurate model is important for MPC to work properly (Zhu et al. (2013)). In practice, the process would slowly change over time which could lead to control performance degradation due to the mismatch between the model and the process. Therefore, in order to maintain MPC performance, process model need to be maintained or re-identified from time to time.

In process industries, the MPC modeling is most often done by using system identification. In MPC control, it is natural to use closed-loop data for model identification in order to ensure the stability and the stable operation of the process. In industrial MPC closed-loop identification tests, test signals are added to the controlled process in order to ensure informativity of the data-set which in turn guarantees the consistency of the model. However, adding test signals will increase the variances of controlled variables (CVs) which degrades the control performance and is a cost of identification tests. The larger the CV variances, the higher the cost of identification.

Thus, it would be desirable to achieve closed-loop data informativity without using test signals. The study of identification without external excitation has attracted the attention of many researchers for several decades. For closed-loop identification in the MPC framework, in order to solve the issue of identifiability, Genceli and Nikolosou (1996) first introduced the notion called simultaneous constrained model predictive control and identification (MPCI). The basic idea is to add additional constraints of input that assure persistent excitation, trying to obtain sufficient model information, while minimally disturb the process. Aggelogiannaki and Sarimveis (2006) and Rathousky and Havlena (2013) also proposed similar methods. However, adding extra constraints to the original optimization proposition still more or less affects the control performance of the closed-loop system. Besides, it turns the optimization into a non-convex problem, which increases the difficulty in the optimization (Larsson et al. (2016)).

In the 1970s, some researchers studied on data informativity (then called model identifiability) of closed-loop identification without external excitation. Soderstrom et al. (1976) pointed out that switching the controller can ensure data informativity (or model identifiability). Yan and Zhu (2018) used switching strategy in a PID closed-loop system and obtained good models without using test signals. This paper extends the controller switching idea of Yan and Zhu (2018) to multivariable MPC systems. Through a proper design of control switching, one can not...
only meet the informativity conditions, but also increase control performance during the identification test, hence achieving a negative identification cost.

The rest of the paper is structured as follows: In section 2, an MPC algorithm is briefly introduced; Section 3 discusses the informativity conditions for multivariable closed-loop identification without excitation; Section 4 presents a switching scheme that ensures the informativity for MPC controlled closed-loop identification and proposes a method of model validation; Section 5 illustrates effectiveness of the proposed method in simulation. Section 6 is the conclusion.

2. INTRODUCTION TO AN MPC ALGORITHM

2.1 Optimization proposition of DMC

The identification method developed in this paper will work for most of the MPC algorithms. To make the presentation more illustrative the well known DMC algorithm is used and briefly discussed. DMC (dynamic matrix control) is a predictive control algorithm developed by Cutler and Ramaker (1980) where process model is given in the form of the step responses of the process which is called dynamic matrix. The DMC algorithm determines the future control moves (ΔU_M(k)) over the control horizon(M) to drive the model predicted outputs as closely as possible to a desired future trajectory over prediction horizon(P). The computation of DMC is to minimize the cost function:

$$\min J_k = \|W(k) - \hat{y}_{p,m}(k)\|^2_Q + \|\Delta U_M(k)\|^2_R$$  \hspace{1cm} (1)

where

$$W(k) = [w_1(k) \cdots w_q(k)]^T$$  \hspace{1cm} (2)

$$w_i(k) = [w_i(k+1) \cdots w_i(k+P)]^T$$  \hspace{1cm} (3)

$$\Delta U_M(k) = [\Delta U_{1, M}(k) \cdots \Delta U_{j, M}(k)]^T$$  \hspace{1cm} (4)

$$\Delta U_{j, M}(k) = [\Delta u_j(k) \cdots \Delta u_j(k+M-1)]^T$$  \hspace{1cm} (5)

where i = 1, ..., q, j = 1, ..., m, m and p represent the number of inputs and outputs respectively; Denote P and M as the prediction horizon and the control horizon; ΔU_j,M(k) as the decision variable and ω as the reference trajectory; \(\hat{y}_{p,m}(k)\) represents the predicted output; Q represents the output weighting matrix which is used to handle the delay part and the inverse part; R represents input weighting matrix which affects the cost of MVs.

$$Q = \text{block - diag}(Q_1, ..., Q_p)$$  \hspace{1cm} (6)

$$R = \text{block - diag}(R_1, ..., R_m)$$  \hspace{1cm} (7)

Increasing the sizes of the elements of R will lead the decreasing of the MV movements, which is equivalent to the decrease of the controller gain. Decreasing the sizes of the elements of R will do the opposite. For most of the industrial processes, one can assume that decreasing the input weighting R will increase the robust stability of the closed-loop system; increasing the input weighting R will decrease the robust stability of the closed-loop system. This assumption will be used in developing the switching method in the identification test.

2.2 Predictive model

The process model is used to provides predictive outputs which consists of two part,

$$\hat{y}_{po,k}(k) = \hat{y}_{po,k} + A\Delta U_M(k)$$  \hspace{1cm} (8)

where \(\hat{y}_{po,k}(k)\) is the natural response under \(u_j(k-1)\). A is a \(p \times m\) Matrix called dynamic matrix which consists of step response coefficient \(a_{ij}(s)\),

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{p1} & \cdots & A_{pm} \end{bmatrix}$$  \hspace{1cm} (9)

$$A_{ij} = \begin{bmatrix} a_{ij}(1) & 0 \\ \vdots & \ddots \\ a_{ij}(P) & \cdots & a_{ij}(P-M+1) \end{bmatrix}$$  \hspace{1cm} (10)

2.3 Moving horizon optimization

If the optimization has no constraints or does not trigger the constraint, the proposition has an analytical solution:

$$\Delta U_{j,k}(k) = (A^TQA + R)^{-1}A^TQ[W(k) - \hat{y}_{p,m}(k)]$$  \hspace{1cm} (11)

where \(\Delta U_{j,M}(k)\) contains M steps. However, the controller only takes the first step increment for actual control, namely:

$$u_j(k) = u_j(k-1) + \Delta u_j(k)$$  \hspace{1cm} (12)

at the next moment, a similar optimization proposition will be proposed. This is so-called ”Moving horizon optimization” strategy.

3. IDENTIFIABILITY CONDITIONS UNDER FEEDBACK

Ljung (1999) proposed the concept of identifiability and informativity which concerns the model structure and operation data respectively. And the second property is focused on this paper, which deals with the question whether the data generated by the system is rich enough to distinguish different models in the given model set. For single-variable closed-loop control system, Gevers et al. (2009) proposed that as long as the structure of the controller in the closed-loop system is complex enough, then the model can be identified without external excitation. Similarly for multivariable closed-loop control systems, Bazanella et al. (2010) proved that when the controller satisfies certain complexity requirement, then identification by noise excitation alone is possible. In order to solve closed-loop identification problem, Söderström and Stoica (1989) and Ljung (1999) summarized three methods as follows:

(1) To add external excitation signal;
(2) To add a time delay in the feedback controller;
(3) To use a nonlinear controller.

Adding external excitation would sacrifice the control performance. Similarly, adding delays in the feedback controller will slow down the control actions which also leads
to control performance degradation. In this work, the third way is adopted, that is, to generate nonlinearity through switching the controller settings among different ones. In this way, it is possible to ensure data informativity and to increase control performance. The method developed in this work is based on the following theory of informativity of the closed-loop system test.

**Theorem 3.1** (Soderstrom et al. (1976)) Consider a multivariable closed-loop system with \( m \) inputs and \( p \) outputs, and the system is switched between \( k \) different linear controllers, as shown in figure 1. When there is no external excitation, the input output data-set is informative if the following criteria is met,

\[
k \geq 1 + \frac{m}{p}
\]  

(13)

4. CONTROLLER SWITCH SCHEME

Here it is intended to find a switching scheme that can: 1) satisfy data informativity condition for closed-loop identification without excitation; 2) improve the control performance during the test period. Yan and Zhu (2018) designed a switching scheme for PID control system, which achieved a win-win effect between informativity and control performance. For DMC system, one can consider switching the weighting coefficient \( R \).

4.1 Relationship between \( R \) and output variance

In this research, the output variance of the system is chosen as an indicator to evaluate the control performance. This is because in the process industry, one of the MPC’s main purposes is to reduce the system’s output variance. A small output variance could ensure the system is operated in a safer and more cost-effective situation. Here, the system output variance \( \tau \) is defined as

\[
\tau = E \left[ \sum_{i=1}^{p} (y_i(k) - r_i(k))^2 \right]
\]  

(14)

Where \( p \) is the number of outputs. Regarding the system output variance and weighting coefficient \( R \), the following result can be used.

**Theorem 4.1** (Tran et al. (2014)) For a closed-loop system under DMC controller, in the case that the optimization proposition has no constraint or constraints will not be triggered, and there exists a model mismatch, then the system’s output variance will first decrease and then increase with an increase in the closed-loop bandwidth \( \omega \), as shown in figure 2.

When the closed-loop bandwidth is small, the system will be sensitive to the disturbance. Conversely, the system will be sensitive to the model mismatch which also causes a bad control performance, even instability. Therefore, there exists an optimal closed-loop bandwidth. Further, fixing the input weights and increasing the output weights, or fixing the output weights and reducing the input weights, both raise the closed-loop bandwidth. So when there exists a model mismatch, one could try to find a more suitable weight when the system is not working on the optimal point.

4.2 Preliminary test for finding a better input weighting

In general, a multi-input and multi-output (MIMO) process identification with \( p \) inputs can be divided into \( p \) times of multi-input single output (MISO) process identification. According to formula 13, to identify a MISO model with \( m \) inputs, at least \( m+1 \) different controllers are needed to ensure the identifiability of the model. The purpose of the preliminary is to determine the values of a new weighting matrix which could improve the performance. Assume that the current weighting matrices is

\[
R_0 = \text{block - diag}\{r_{1M}, ..., r_{mM}\}
\]  

(15)

where

\[
r_{1M} = \text{diag}\{r_1, ..., r_1\}
\]  

(16)

and

\[
r_{mM} = \text{diag}\{r_m, ..., r_m\}
\]  

(17)

**Step 1:** The closed-loop system is controlled by the existing linear DMC controller; Compute and record the current output variance \( \tau_0 \).

**Step 2:** Increase the weighting matrix. In general, increasing the input weighting slows down the control action and increase the closed-loop system robustly stability. So the user is on the safe side by starting with increasing the input weighting. Set the weighting matrix as

\[
R_1 = K_1R_0, K_1 > 1
\]  

(18)

where \( K_1 \) can be in the range from 2 to 5. Compute the output variance \( \tau_1 \) in this period. If \( \tau_0 > \tau_1 \), it means that the new input weighting matrix \( R_1 \) has a higher control performance and it will be used as the new input weighting. Set \( K_s = K_1 \) and stop the pretest. If \( \tau_0 < \tau_1 \), then go to Step 3 and increase the input weighting matrix.

**Step 3:** Decrease the input weighting matrix. Set the weighting matrix as

\[
R_2 = K_2R_0, 0 < K_2 < 1
\]  

(19)

where \( K_2 \) can be in the range from 0.3 to 0.8. If \( \tau_0 > \tau_2 \), it means that decreasing the input matrix increase the performance. Then set \( K_s = K_2 \) and stop the pretest.

Note that although one can achieve better control using a new input weighting in most of the cases, one cannot guarantee the increase of control performance in all situations. In a rare case, it may happen that \( \tau_0 < \tau_1, \tau_2 \).
4.3 Switch method

Based on the result of the pre-test, one obtains a new input weighting matrix $K_s$ that is in general a better controller. In order to create sufficient number of controllers, every input weighting is switched individually between its corresponding elements in the two weighting matrices $R_0$ and $R_1 (R_2)$ by following a switch sequence. Now for every input, there exists 2 different settings. When the test time is sufficient long, by combination, one obtains $2^m$ settings of controllers, which suffices to meet the informativity condition (11). For example, a 3-input system has 8 different settings. The switching sequence of each input can be arranged in different ways. Here, it is proposed that the switching of the inputs follows $m$ independent GBN (generalized binary noise) signals; see figure 3. The average switching time of GBN signals can be set as the time to steady state (response time) of the process.

4.4 Model validation

In identification for control, it is necessary to evaluate the quality of the model identified. When test signals are used in identification tests and when the signal-to-noise ratios are high, model output errors (simulation errors) can be used to measure model quality. For example, if the simulation error variance is smaller than 10% of the output variance, then the identified model is considered accurate. This means that model simulation errors cannot be used to evaluate the quality of the model identified. When test signals are used in identification tests and when the signal-to-noise ratio is larger than 10%, output error residual $\hat{v}(t)$, and prediction error residual $\hat{e}(t)$.

\[
\text{then we have}
\]

\[
RE = \frac{\text{var}[\hat{v}(t)]}{\text{var}[y(t)]} \times 100% = \frac{\text{var}[\hat{e}(t)]}{\text{var}[y(t)]} \times 100%
\]

A well performing controller ensures that the output variance in the closed-loop operation will be smaller than the open loop disturbance variance. In this case one has $RE > 100%$ when the model is identified perfectly. This means that model simulation errors cannot be used for model validation in closed-loop identification without excitation.

In the framework of prediction error method, the identified model is considered a random variable( Ljung (1999)). When the distribution of the model is known, model quality can be evaluated using the corresponding confidence intervals. Based on the prediction error method( Ljung (1999)), one can obtain

\[
\hat{\theta}_N \xrightarrow{w.p.1} \theta^o
\]

and as $N$ tends to infinity,

\[
\sqrt{N}(\hat{\theta} - \theta^o) \in As N(0,P_B)
\]

From (26), we get the covariance of $\hat{\theta}_N$

\[
\text{cov}(\hat{\theta}_N) \approx \frac{1}{N} P_B
\]

As $\theta$ becomes sufficiently close to $\theta^o$, by Taylor’s expansion,

\[
\hat{G}(e^{i\omega}, \hat{\theta}) - G^o(e^{i\omega}, \theta^o) = [G^o_{\theta}(e^{i\omega}, \theta^o)](\hat{\theta} - \theta^o) + O|\hat{\theta} - \theta^o|
\]

where $G^o_{\theta}(e^{i\omega}, \theta^o)$ is the derivative of $G(e^{i\omega}, \theta)$ with respect to $\theta$ at $\theta^o$. According to (28), the asymptotic distribution in the frequency domain could be obtained as $N$ tends to infinity.

\[
\sqrt{N}[\hat{G}(e^{i\omega}, \hat{\theta}) - G^o(e^{i\omega}, \theta^o)] \in As N(0, P_G)
\]

and (28) directly gives

\[
P_G \approx [G^o_{\theta}(e^{i\omega}, \theta^o)]^T P_B G^o_{\theta}(e^{i\omega}, \theta^o)
\]

and

\[
\text{cov}(\hat{G}(e^{i\omega}, \hat{\theta})) \approx [G^o_{\theta}(e^{i\omega}, \theta^o)]^T \text{cov}(\hat{\theta}_N) G^o_{\theta}(e^{i\omega}, \theta^o)
\]

Based on this property, to evaluate the quality of the identified model, a common choice is to plot the confidence region with 3 standard deviation. A relative small bound means a good model.

\[
\left|\hat{G}(e^{i\omega}, \hat{\theta}) - G^o(e^{i\omega}, \theta^o)\right| \leq 3\sqrt{\text{var}(\hat{G}(e^{i\omega}, \hat{\theta}))}
\]

Another error model bound is proposed using the asymptotic theory in Zhu (2001)

\[
\left|\hat{G}(e^{i\omega}, \hat{\theta}) - G^o(e^{i\omega}, \theta^o)\right| \leq \Delta(\omega)
\]

\[
\Delta(\omega) \leq 3 \sqrt{\frac{n_h}{N} \Phi_{ue}(\omega) \sigma_e^2}
\]

where $n_h$ is the order of a high-order model. $\Phi_{ue}(\omega)$ is the power spectrum of disturbance. $\Phi_{ue}(\omega)$ is the power spectrum of input. $\Phi_{ue}(\omega)$ is the cross power spectrum between input and white noise $e(t)$. $\sigma_e^2$ is the variance of white noise $e(t)$. The auto and cross-spectra used in (34) can be estimated using input signal $u(t)$, output error residual $\hat{v}(t)$, and prediction error residual $\hat{e}(t)$.
To use the model error bounds in model validation, an engineering approach of Zhu (1998) is used here: Compare the relative size between the model and the error bound over the low and middle frequencies; a transfer function is graded A (very good), if bound ≤ 30% of model, B (good), if bound ≤ 60% of model, C (marginal), if bound ≤ 90% of model, and D (poor), if bound > 90% of model.

5. SIMULATION

In order to confirm the effectiveness of the approach proposed, a simulation study is performed.

5.1 Description of simulation

The process is given as a 3-input and 2-output ARMAX system:

\[
A(q^{-1}) \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = B(q^{-1}) \begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{bmatrix} + C(q^{-1}) \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix}
\]

where,

\[
B(q^{-1}) = \begin{bmatrix} 0.15q^{-1} + 0.1q^{-2} \\ 0.2q^{-1} + 0.25q^{-2} \\ 0.6q^{-1} + 0.3q^{-2} \\ 0.45q^{-1} + 0.4q^{-2} \\ 0.35q^{-1} + 0.2q^{-2} \\ 0.1q^{-1} + 0.2q^{-2} \end{bmatrix}
\]

\[
A(q^{-1}) = \begin{bmatrix} 1 - 1.3q^{-1} + 0.38q^{-2} \\ 1 - 1.64q^{-1} + 0.69q^{-2} \end{bmatrix}
\]

\[
C(q^{-1}) = \begin{bmatrix} 1 + 0.2q^{-1} + 0.1q^{-2} \\ 1 + 0.1q^{-1} + 0.1q^{-2} \end{bmatrix}
\]

5.2 Description of simulation

(1) Two different cases would be compared. Case 1: System operates using a linear DMC controller. Case 2: System operates under switching DMC controllers.

(2) As shown in figure 4, a preliminary test is performed first to fix the \(K_s\), then GBN signals are added to direct the weight switching sequence of MVs.

(3) There is no external excitation. White noise signal \(e_1(t)\) and \(e_2(t)\) are mutually independent. The variances of the white noises are both 0.01.

(4) The simulation was run for 100 times; the sample length is 30000. The length of the GBN signals is 30000 and its average switching time is 200.

(5) The controller’s initial setting is as follows: Model horizon \(N = 40\), the prediction horizon \(P = 15\), control horizon \(M = 6\), the output weighting matrix \(Q = diag\{1,...,1\}\), the current input weighting matrix \(R = I\). The extent of a and the two settings are \(K_1 = 5\), \(K_2 = 0.3\).

(6) There exists a mismatch between the predictive model and the process, and the step responses of the process and the model are shown in figure 5.

5.3 Simulation results

Table 1 shows the parameters estimation results of 100 simulations in two different situations; The step responses of identified model and the process are shown in figure 6 (the figure only shows results of 10 models in order not to make it too messy). Figure 8, Tables 1 and 2 show that the identified models are poor with a linear DMC controller. On the contrary, by switching DMC controllers, one can indeed obtain accurate models. Further, figure 7 and figure shows the output variance of the system during each period. Obviously, with a good switching strategy, the control performance during the experimental period has been improved, at least not degraded, which verifies the effectiveness of the preliminary test. After the predictive model updated, the control performance has been improved again with the same parameter settings. Figure 8 shows that the relative error is very high when the model is accurate. As mentioned in section 4, it’s feasible to evaluate the quality from the estimated variance in frequency response. Figure 9 shows the frequency response of one simulation with confidence bound of 3 standard deviation. The confidence bounds (the shaded area) signifies the estimated variance is small and the identified model is desirable.

6. CONCLUSION

A method of multivariable MPC closed-loop identification without excitation is developed. A switching scheme of the MV weighting matrix is proposed to assure data informativity (model identifiability) during the identification test. Compared to the test method in process industries where test signals are used, this method can improve the control performance during the test period which decrease.
Table 1. Parameter estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Switch controller</th>
<th>Linear controller</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_1</td>
<td>1</td>
<td>-1.30 ±0.0315</td>
<td>-1.30 ±0.0441</td>
</tr>
<tr>
<td>b_1</td>
<td>0.38</td>
<td>0.38 ±0.0273</td>
<td>0.38 ±0.0473</td>
</tr>
<tr>
<td>b_2</td>
<td>0.15</td>
<td>0.15 ±0.0810</td>
<td>0.12 ±0.0176</td>
</tr>
<tr>
<td>b_3</td>
<td>0.1</td>
<td>0.0966 ±0.0798</td>
<td>0.119 ±0.3333</td>
</tr>
<tr>
<td>b_4</td>
<td>0.6</td>
<td>0.5991 ±0.1344</td>
<td>0.5976 ±0.4875</td>
</tr>
<tr>
<td>b_5</td>
<td>0.45</td>
<td>0.4598 ±0.1366</td>
<td>0.4347 ±0.3382</td>
</tr>
<tr>
<td>b_6</td>
<td>0.4</td>
<td>0.3877 ±0.1489</td>
<td>0.4108 ±0.3053</td>
</tr>
<tr>
<td>c_1</td>
<td>0.2</td>
<td>0.1967 ±0.0326</td>
<td>0.1988 ±0.0428</td>
</tr>
<tr>
<td>c_2</td>
<td>0.1</td>
<td>0.0983 ±0.0140</td>
<td>0.0996 ±0.0174</td>
</tr>
</tbody>
</table>

Fig. 7. Variance comparison totally

Fig. 8. Relative error of identified model

Fig. 9. Model validation in frequency domain

the cost of identification. It is pointed out that, in closed-loop identification without excitation, the simulation error cannot be used to evaluate the quality of the identified model and confidence intervals (error bounds) of model frequency responses can be used in model validation. The effectiveness of the proposed method is verified by simulating a 3-input and 2-output process.

REFERENCES