Preprints, 10th IFAC International Symposium on
Advanced Control of Chemical Processes
Shenyang, Liaoning, China, July 25-27, 2018

On an Aspect of Implementing Real-Time Optimization: Establishing the Suspending and Activating Conditions Incorporating Process Monitoring ⋆

Lingjian Ye * **Feifan Shen** * **Zhiqiang Ge** ** **Zhihuan Song** **

* *Ningbo Institute of Technology, Zhejiang University, 315100, Ningbo, China (e-mail: lingjian.ye@gmail.com, zju.sff@gmail.com)*
** *Department of control science and engineering, Zhejiang University, 310027, Hangzhou, China (e-mail: gezhiqiang@zju.edu.cn, songzhihuan@zju.edu.cn)*

Abstract: For a large class of real-time optimization (RTO) schemes where online experimental gradients are evaluated for convergence to the plant optimum, the input signals are sufficiently excited in the noisy environment. Furthermore, the evaluations are typically persistent even if convergence is attained, for handling varying operating conditions caused by disturbances. The unsettled operation around the optimum leads to oscillations and extra economic loss. In this paper, we propose a strategy that establishes the suspending and activating conditions for RTO schemes. The conditions are developed based on process monitoring methods, which can in a passive way detect operating condition changes. Using the conditions, the RTO implementation is allowed to be suspended upon convergence and further restarted to approach the new optimum when the operating condition changes. The Williams-Otto reactor is studied to show the usefulness of the new idea.

Keywords: Chemical processes, real-time optimization, process monitoring, uncertainty

1. INTRODUCTION

Optimization has been long prevalent for operating chemical processes (Edgar et al., 2001). The rigourous process model is built and then numerically optimized regarding the economic index. In this case, an accurate mathematical model matching the real plant is fundamental for the optimization performance. Unfortunately, plant-model mismatches widely exist in real applications due to, *e.g.* unclear chemical mechanisms and unknown disturbances. Optimization based on an inaccurate process model leads to an inferior performance, or worse, infeasibility.

Real-time optimization (RTO) deals with optimization with uncertainties. In literatures, different RTO approaches have been proposed, including the two-step approach (Chen and Joseph, 1987), modifier adaptation (Marchetti et al., 2009, 2016), necessary conditions of optimality (NCO) tracking (Srinivasan et al., 2003; Francois et al., 2005; Srinivasan and Bonvin, 2007), extremum seeking (Ariyur and Krstic, 2003), self-optimizing control (Skogestad, 2000; Jäschke et al., 2017; Ye et al., 2015), and so forth. These RTO schemes generally pursue (1) constraints satisfaction, and (2) convergence to the (neighborhood of) true optimum. Comparisons among different RTO schemes can be found in recent literatures (Chachuat et al., 2009; Francois et al., 2012).

For a large class of RTO schemes, including the modifier adaptation and NCO tracking, the plant gradients for various quantities need to be evaluated for convergence to the true optimum. Typically, online experiments are carried out by perturbing the inputs in certain way and then the plant response is observed. The input signals should be sufficiently excited to extract information for calculating the gradients (Brdys and Tatjewski, 2005; Marchetti et al., 2010). Intuitively, when the operation converges into the optimum neighborhood within a tolerable level, the inputs should however be kept stationary to stay at the optimum. This is nonetheless opposed to the exciting constraint imposed on the optimization formulations of RTO. Another important fact is that the operating condition may change unexpectedly (caused by, for example, varying disturbances), hence the gradients need to be evaluated continuously, such that the new optimum is still approached in the new operating condition. The oscillations caused by input excitements are unfavored from a control point of view. Economically, this further leads to the loss of profit. The mentioned contradiction is similar to the well-known dual problem for control and identification, where the similarity has been notified (Marchetti et al., 2010). Solutions have been proposed on this issue (Brdys and Tatjewski, 2005; Marchetti et al., 2010), however, the focus is placed on finding reasonable magnitudes and directions of perturbed

⋆ The author Lingjian Ye gratefully acknowledge the National Natural Science Foundation of China (NSFC) (61673349, 61304081), Talent project of Zhejiang Association of Science and Technology (2017YCGC014).

inputs such that the two sides are consolidated. However, the contradiction still in essence remains.

Focusing on the mentioned RTO schemes with evaluating experimental gradients, the perspective in this paper is to take a more practical way, by means of establishing the suspending and activating conditions for performing RTO. In specific, we consider to suspend the RTO when the operation has converged into the optimum neighborhood, such that unnecessary oscillations are avoided upon convergence. On the other hand, when the operating condition changes, the RTO is reactivated to find the new optimum. While the suspending condition for judging convergence to optimum can be established along the RTO realizations, detection of operating condition change to keep suspending and activate RTO is not straightforward. To this end, we introduce the process monitoring theory in this paper, which is traditionally used for fault diagnosis. Previously, the process monitoring methodology has been novelly utilized to detect non-optimal status for processes with parametric uncertainties (Ye et al., 2014b,a). The advantage is that process monitoring methods run in a passive way, without the necessity of exciting the process. Furthermore, the monitoring models can be established based on the historical process data, which is purely data-driven. Combing the suspending and activating conditions into the RTO scheme can, from a decision level, handle the contradiction mentioned above. To the best knowledge of authors, such strategy has not been proposed in recent publications.

The rest of this paper is organized as follows. In section 2, the methodology of modifier adaptation is briefly reviewed, which is used as a particular RTO investigation throughout this paper. But the proposed strategy can be readily extended to RTO schemes with performing online gradient estimations, such as the NCO tracking. The main idea in this contribution is elaborated in Section 3, then the implementing logic and established conditions are presented in detail. In Section 4, the Williams-Otto reactor is investigated, where the merits of the new strategy are confirmed through simulations. The work is concluded with discussions in Section 5.

2. THE MODIFIER ADAPTATION

2.1 Method formulation

Consider the following static optimization problem

$$\begin{aligned} &\min_u \ \phi(u,d) \\ \text{s.t.} \quad &y = f(u,d) \\ &G(u,y(u,d)) \le 0 \end{aligned} \tag{1}$$

where ϕ is the economic cost to be minimized, u and d are the manipulated variables and uncertain disturbances, respectively. y is the measurements and f is the input-output mapping function, G is the process constraints. Due to plant-model mismatch, suppose that the actual optimization problem for the real plant is

$$\begin{aligned} &\min_u \ \phi(u,y_p(u,d_p)) \\ \text{s.t.} \quad &y_p = f_p(u,d_p) \\ &G_p(u,y_p(u,d_p)) \le 0 \end{aligned} \tag{2}$$

where the subscript $(\cdot)_p$ denotes the variables/functions associated with the real plant. Instead of disturbance estimation and model refinement, modifier adaptation (MA) solves a modified optimization problem on basis of the inaccurate process model, by making corrections of both bias and gradients on the constraints and cost function. Specifically, at the kth iteration, the following modified optimization is solved

$$\begin{aligned} u_{k+1} \in arg\min_u \ &\phi_m := \phi(u,d) + (\lambda_k^J)^{\mathrm{T}} u \\ \text{s.t.} \quad &G_m := G(u,d) + \varepsilon_k + (\lambda_k^G)^{\mathrm{T}}(u-u_k) \le 0 \end{aligned} \tag{3}$$

where the subscript "$(\cdot)_m$" represents for a modified function, ε_k and λ_k^G are the constraint-value and constraint-gradient modifiers, respectively; λ_k^J are the cost-gradient modifiers. The modifiers denote the differences between the true and predicted KKT elements at the kth operating point u_k, defined as follows

$$\varepsilon_k := G_p(u_k, y_{p,k}) - G(u_k, d) \tag{4}$$

$$\lambda_k^G := \frac{\partial G_p}{\partial u}(u_k) - \frac{\partial G}{\partial u}(u_k, d) \tag{5}$$

$$\lambda_k^J := \frac{\partial J_p}{\partial u}(u_k) - \frac{\partial J}{\partial u}(u_k, d) \tag{6}$$

To prevent from excessive corrections, the optimized inputs u_{k+1} is filtered as

$$u_{k+1}^* = K_u u_{k+1} + (I - K_u) u_k \tag{7}$$

where K_u is a diagonal matrix with elements bounded between 0 and 1, the filtered signals u_{k+1}^* are implemented instead as the actual inputs. Notice that, it is also the case to filter the modifiers such that smaller movements are implemented (Marchetti et al., 2009).

One of the most advantages of MA is that, upon convergence, the ultimate operation u_∞ approaches to the true plant optimum in the noise-free environment, even there exist structural plant-model mismatches, as stated in the following *KKT matching* theorem (Marchetti et al., 2009).

KKT matching theorem (Marchetti et al., 2009): Assume that the iterative procedure (3) *converges to u_∞ with a nonsingular filter K_u, the KKT of modified optimization problem matches with the plant at operating point u_∞, which implies that u_∞ is an optimum for the real plant.*

2.2 Gradient evaluation

A key element is the gradient evaluation. At all iteration points in MA, the gradients $\frac{\partial \phi_p}{\partial u}(u_k)$ and $\frac{\partial G_p}{\partial u}(u_k)$ needs to be evaluated. The simplest method for this purpose is the forward finite difference (FFD) method, which perturbs all inputs in turn along independent input directions and then evaluate the output differences. For example, the gradient of measurements with regarding to the jth input u_j, $\frac{\partial y_p}{\partial u_j}(u_k)$ is calculated as

$$\frac{\partial y_p}{\partial u_j}(u_k) = \frac{y_p(u_k + h e_j) - y_p(u_k)}{h} \tag{8}$$

where h is a small positive step size and e_j is the jth unit vector. The cost gradient is calculated similarly by replacing y_p with ϕ_p in above equation. However, the shortcoming of FFD is that the speed of gradient evaluation is slow, particularly for multi-input systems. This is because for a complete evaluation of the full gradients,

the FFD method requires n_u+1 operating points to collect sufficient information.

To speed up the evaluations, gradient estimation on basis of past operating points have been proposed in literatures. In the dual ISOPE algorithm (Brdys and Tatjewski, 2005), for instance, let the inputs and outputs for the past n_u operating points denoted as $\{u_k, \ldots, u_{k-n_u+1}\}$ and $\{y_{p_k}, \ldots, y_{p_{k-n_u+1}}\}$, respectively. Suppose that current inputs and outputs are u and y_p, two difference matrices are then defined as

$$\mathscr{U}(u) := [u - u_k \ \cdots \ u - u_{k-n_u+1}] \tag{9}$$

$$\mathscr{Y}(u) := \left[y_p - y_{p_k} \ \cdots \ y_p - y_{p_{k-n_u+1}}\right] \tag{10}$$

By linearization around the current point, it can be derived that a first order estimate of gradients is $\frac{\partial y_p}{\partial u}(u) = \mathscr{Y}(u)\mathscr{U}^{-1}(u)$, then both the iterative modified optimization and gradient estimation can be performed at successive operating points.

However, it was recognized that online gradient estimation is a *dual* problem. Basically, the differences of u in matrix $\mathscr{U}$ should be small enough to guarantee the accuracy of first order gradient approximation and on the other hand, big enough to ensure a significant signal-to-noise ratio and well conditioned $\mathscr{U}(u)$. To deal with this issue, the dual ISOPE algorithm introduces a lower bound for the condition number of the inverse of $\mathscr{U}(u)$. Equivalently, the condition number of $\mathscr{U}(u)$ should be lower than an upper bound κ_{max} as follows

$$\kappa(u) := \frac{\sigma_{max}(\mathscr{U}(u))}{\sigma_{min}(\mathscr{U}(u))} \leq \kappa_{max} \tag{11}$$

where σ_{max} and σ_{min} stand for the largest and smallest singular values of a matrix, respectively. The condition number constraint is added into the optimization problem (3). However, it was pointed out that the condition number constraint in the dual ISOPE algorithm does not convey information for the gradient error (Marchetti et al., 2010), κ_{max} is often set in a heuristic way.

Recently, an improved gradient evaluation method, termed as the dual MA, is proposed (Marchetti et al., 2010). Instead of using the condition number bound, the upper bounds of norm of gradient errors caused by finite-difference approximation and measurement noise were derived. Simulations show that the optimizing speed of MA is accelerated and further leads to enhanced economic performance. Due to the space limitation, the dual MA algorithm is not further presented, and the condition number bound is used in this paper.

3. SUSPENDING AND ACTIVATING CONDITIONS

Logically, the RTO implementation should terminate once the convergence is attained, such that unnecessary input perturbations are avoided. However, without persistent gradient evaluations, the operation loses optimality in case of operating condition change. Therefore, a key point lies in how to know the change of operating condition.

3.1 Detection of operating condition changes

We propose to detect the change of operating condition based on the process monitoring theory. Previously, we have used a similar strategy to judge the optimality status of chemical processes influenced by parametric uncertainties(Ye et al., 2014b,a). Among numerous process monitoring methods for chemical plants (Ge, 2017; Zhu et al., 2017; Ge et al., 2017), we introduce the classic linear PCA, which is the simplest one but sufficient to demonstrate our purpose.

Denote the data samples for PCA modelling in a matrix form $Y \in \mathbb{R}^{M\times n_y}$, where M is the number of samples. The row vector $(y_i)^{\mathrm{T}}$ in Y denotes a group of measurements. Y is firstly scaled to have zero mean values and unit variance for each variable. The PCA form is represented as

$$Y = TP^{\mathrm{T}} + E = TP^{\mathrm{T}} + \tilde{T}\tilde{P}^{\mathrm{T}} \tag{12}$$

where $T \in \mathbb{R}^{M\times k}$ and $P \in \mathbb{R}^{n_y\times k}$ is the score and loading matrix of principal components, respectively. $E \in \mathbb{R}^{M\times n_y}$ is the residual matrix, $\tilde{T} \in \mathbb{R}^{M\times(n_y-k)}$ and $\tilde{P} \in \mathbb{R}^{n_y\times(n_y-k)}$ is the score and loading matrix of residual components, respectively. k is the number of principal components, whose value can be determined from cross validation or cumulative percent variance (CPV).

The various matrices in (12) can be obtained through the symmetric eigenvalue decomposition for the covariance matrix of Y, $\Sigma = Y^{\mathrm{T}}Y/(M-1)$

$$\begin{aligned} \Sigma &= \begin{bmatrix} P & \tilde{P} \end{bmatrix} \Lambda \begin{bmatrix} P & \tilde{P} \end{bmatrix}^{\mathrm{T}} \\ T &= YP \\ \tilde{T} &= Y\tilde{P} \end{aligned} \tag{13}$$

where Λ is a diagonal matrix consists of all eigenvalues of Σ, i.e.

$$\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, ..., \lambda_{n_y}\} \tag{14}$$

Through above steps, each row vector $(y_i)^{\mathrm{T}}$ in Y can be projected onto the principal and residual spaces

$$\begin{aligned} \hat{y}_i &= PP^{\mathrm{T}}y_i \\ \tilde{y}_i &= (I - PP^{\mathrm{T}})y_i \end{aligned} \tag{15}$$

where $\hat{y}_i$ and $\tilde{y}_i$ is the projected vector onto the principal and residual spaces, respectively, while satisfying

$$\begin{aligned} (\hat{y}_i)^{\mathrm{T}}\tilde{y}_i &= 0 \\ y_i &= \hat{y}_i + \tilde{y}_i \end{aligned} \tag{16}$$

Using the PCA, the original data can be described in the reduced k dimensional uncorrected principal directions, with most of the information retained. For PCA, the T^2 and SPE statistics can be constructed in the principal and residual spaces, respectively. The T^2 statistic indicates the variation extent of data in principal space, which is defined as

$$T^2 = x^{\mathrm{T}}P\Lambda^{-1}P^{\mathrm{T}}x \tag{17}$$

Assuming the process data is Gaussian distributed, the T^2 statistic is demonstrated to obey an F distribution with k and $N-k$ as the degrees of freedom in normal conditions. Given a significance level α, the control limit of T^2 statistic can be calculated and the process is monitored as

$$T^2 \leq T_\alpha^2 = \frac{k(N-1)}{N-k}F_{k,N-k;\alpha} \tag{18}$$

Meanwhile, the SPE statistic indicates the distribution of data in the residual spaces, which is defined as the norm of projected residual vector

$$SPE = \|\tilde{x}\|^2 = \|(I - PP^{\mathrm{T}})x\|^2 \tag{19}$$

Similarly, the corresponding SPE control limit δ_α^2 can also be computed as (Jackson and Mudholkar, 1979)

$$SPE \leq \delta_\alpha^2 = \theta_1 \left(1 + \frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{1/h_0} \tag{20}$$

where

$$\theta_i = \sum_{j=k+1}^{n_y} \lambda_j^i, \quad i = 1, 2, 3 \tag{21}$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \tag{22}$$

and c_α is the normal deviate corresponding to the upper $1 - \alpha$ percentile.

3.2 Implementation logic

The flowchart of the implementation logic is showed in Figure 1. In general, we proceed with the common RTO schemes, but add the following logics:

(1) Judge whether the plant has converged to the neighborhood of optimum. If so, (i) determine a steady operation and then suspend the RTO; (ii) collect the associated measurements for PCA modelling. Otherwise, continue performing RTO;
(2) When the plant is operated with the RTO suspended, monitoring the operating condition with the built PCA model. When either T^2 or SPE limit is violated, which indicate an operating condition change, activate the RTO scheme so as to find the new optimum.

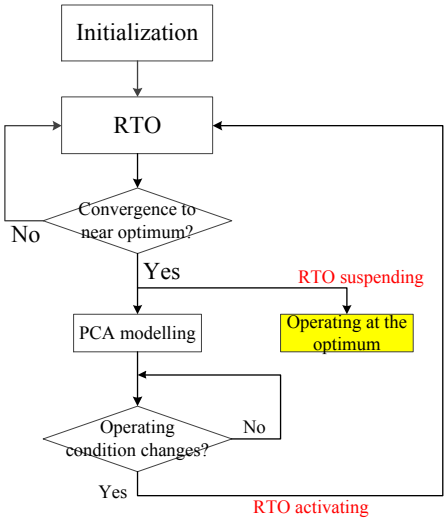

Fig. 1. Implementation logic

Confirmation of convergence to near optimum. Several criterions can be used to verify the convergence. For example, (1) the increments of inputs between successive iterations; (2) the plant gradients during the iterations. In this paper, the gradients of plant objective function are used for this purpose. Besides of the convergence, the cost gradients additionally reflect the plant optimality. The following condition is used

$$\|\xi\| = \|\partial \phi_p / \partial u\| \leq \xi_{tol} \tag{23}$$

where ξ_{tol} is a defined tolerance for the norm of derivatives. Hence, when the plant enters and remains within the region where (23) is satisfied, the dual algorithm for gradient evaluation stops and the whole RTO scheme is suspended. During the iterations, all associated quantities satisfying (23) (including u_k, y_k and $\frac{\partial \phi_p}{\partial u}$), are collected.

Locating the true optimum. The convergence is only into the neighborhood around the true optimum, owing to the dual algorithm. Here, we seek a final settlement for the true optimum as much as possible. A locally linear relationship between u and $\frac{\partial \phi_p}{\partial u}$ is assumed

$$\frac{\partial \phi_p}{\partial u} = Au + b \tag{24}$$

where the coefficient matrix A and bias vector b can be estimated via least-square regression. To achieve optimality, setting $\frac{\partial \phi_p}{\partial u} = 0$ leads to

$$u_{set} = -A^{-1} b \tag{25}$$

which is implemented as the settled operation during the suspense of RTO. Note that, since only a linear model for $\frac{\partial \phi_p}{\partial u}$ is assumed, it is desired that the collected data are in a small neighborhood around the optimum point. Hence, it is proposed to use data satisfying (23) for regression, which improves the model accuracy and then the quality of u_{set}.

Detection of operating condition change. The output measurements collected around the optimum are used to construct the PCA model for the detection purpose. Then, the RTO scheme will be activated if either T^2 or SPE indices exceed their control limits, thus the new plant optimum can be approached for the new operating condition. Otherwise, the operation keeps stationary and no RTO needs to be implemented.

4. CASE STUDY: THE WILLIAMS-OTTO REACTOR

4.1 Process descriptions

The Williams-Otto reactor is considered (Williams and Otto, 1960; Roberts, 1979), which has been frequently used for the RTO study. The reactor consists the following reactions

$$A + B \xrightarrow{k_1} C, \; k_1 = 1.660 \times 10^6 e^{-6666.4/(T_R + 273.15)} \tag{26}$$

$$C + B \xrightarrow{k_2} P + E, \; k_2 = 7.212 \times 10^8 e^{-8333.3/(T_R + 273.15)} \tag{27}$$

$$C + P \xrightarrow{k_3} G, \; k_3 = 2.675 \times 10^{12} e^{-11111/(T_R + 273.15)} \tag{28}$$

where A and B are reactants, P and E are the desired products, C is an intermediate product and G is the undesired by product. The mass feedrates of A and B are F_A and F_B, respectively. F_A keeps at a constant of 1.8275 kg/s. T_R is the maintained reactor temperature. The reactor mass holdup is 2105 kg and the outlet flow is $F = F_A + F_B$ to maintain the holdup balance.

The operational objective is to maximize the following profit

$$\phi = 1143.38 X_P F + 25.92 X_E F - 76.23 F_A - 114.34 F_B \tag{29}$$

where X_X stands for the concentration of material X. The manipulated variable vector for optimization is $u = [F_B\ T_R]^T$. However, due to insufficient process knowledge, the plant is understood as

$$A + 2B \xrightarrow{k_1^*} P + E,\ k_1^* = 2.189 \times 10^8 e^{-8077.6/(T_R+273.15)} \tag{30}$$

$$A + B + P \xrightarrow{k_2^*} G,\ k_2^* = 4.310 \times 10^{13} e^{-12438/(T_R+273.15)} \tag{31}$$

which is the model known by the designer. The optimum for the inaccurate model is $u_{model}^{opt} = [4.77\ 78.2]^T$, which is distant from the true plant optimum, $u_{plant}^{opt} = [4.79\ 89.7]^T$.

4.2 RTO with suspending condition

The modifier adaptation is implemented as the RTO realization. The convergence criterion for gradient norm is chosen as $\xi_{tol} = 60$, which is selected on basis of the data information of converged operating points.

The RTO realizations are shown in Figure, with the evolvements of inputs, gradient norms and profit ϕ. It can be seen that MA is effective, where the inputs are soon adjusted toward the optimum (red "o" in subfigure (a)). Within 10 iterations, convergence is nearly attained and the inputs are bounded around the optimum afterwards. However, the operation is unsettled due to the condition number constraint.

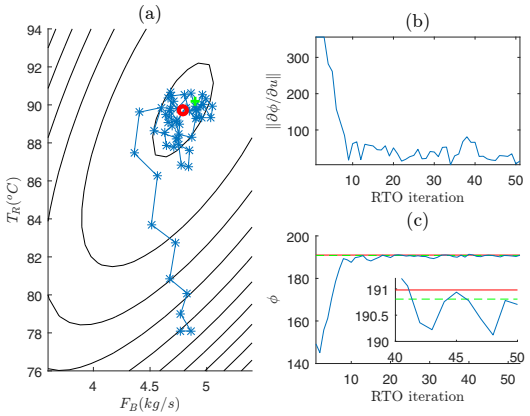

Fig. 2. Implementation of MA with suspending condition: (a) input evolvements (red "o": true optimum, green "+": settled operation); (b) norm of gradients; (c) profit (blue solid line: profit evolvements; red solid line: optimal profit; green dashed line: profit associated with the settled operation)

Using the proposed strategy, we obtain a local model of the gradient (24) using the least-square regression based on the data samples satisfying $\|\xi\| \leq \xi_{tol}$. The location of settled operation is $u_{set} = [4.90\ 90.19]^T$ (green "+" in subfigure (a)), calculated using (25). The settlement is very close to the true optimum, u_{plant}^{opt}. With this final settlement, the profit loss is 0.17. On the other hand, when the MA continues in the conventional way, the plant suffers from oscillation and the average loss is about 0.7, which is larger than the new method.

4.3 RTO with activating conditions

The measurements at operating points satisfying $\|\xi\| \leq \xi_{tol}$ are collected to build the PCA model. The measurements are assumed to be concentrations of all materials, which are corrupted by Gaussian noises with zero mean and 1% standard deviation. The number of principal component is chosen as 4, which explains more than 98% variance in the data set. The confident level is set as 0.98. The T^2 and SPE control limits are then calculated as 17.7 and 0.39, respectively.

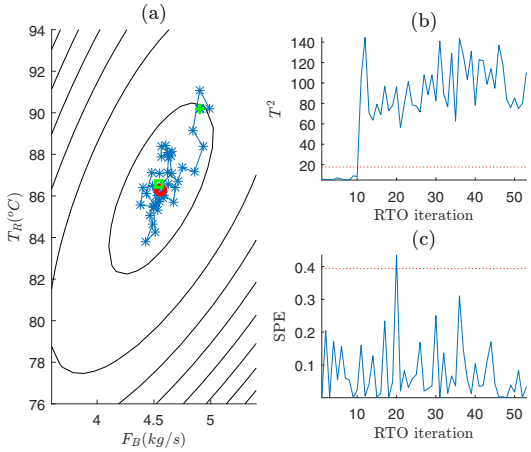

Fig. 3. Implementation of MA with activating condition: (a) input evolvements (red "o": new optimum, green "□": new settled operation); (b) monitoring with T^2 statistic; (c) monitoring with SPE statistic

The afore investigated scenario is continued by implementing the settled operation, $u_{set} = [4.90\ 90.19]^T$, meanwhile, the process is monitored by the obtained PCA model. To further simulate a new scenario, the reaction coefficient k_3 is introduced with a +20% step change after 10 iterations. As shown in Figure 3 (b) and (c), in the first 10 iterations, both the T^2 and SPE statistics stay within normal bounds, then the MA is kept suspended. However, from iteration 11, the changed operating condition is successfully detected by the T^2 statistic, while the SPE limit is not violated in this case. The alarm triggers the activating condition so that MA starts to work again, the inputs are then adjusted toward the new optimum, $u_{plant,new}^{opt} = [4.55\ 86.3]^T$, as shown in Figure 3 (a). Note that, once a change of operating condition has been detected, the monitoring model can be discarded, and a new one waits to be built for the new operating condition, upon convergence around the optimum. Furthermore, the settlement can be found in the same way as done before to locate the optimum. In this case, we obtain $u_{set,new} = [4.545\ 86.6]^T$ (green "□" in Figure 3 (a)), which is again very close to the optimum.

The profit curves with 10 random realizations are shown in Figure 4. Additionally, we compare the the curves implementing the normal MA. In general, both of the two method can effectively approach to the new optimum, besides, the following facts are observed: (1) In the first 10 iterations, the normal MA oscillates while the new method is steady. This has been previously confirmed in Figure 2 (c) where the profit is reasonably improved in the new method. (2) Between iterations 11-16, the

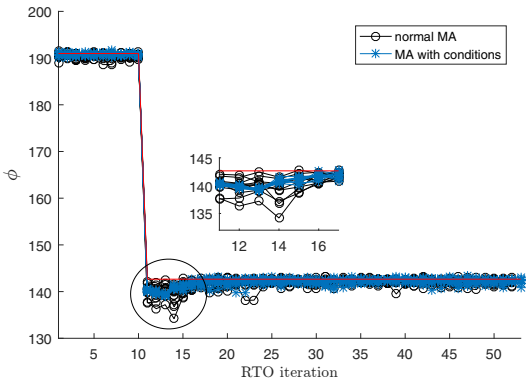

Fig. 4. Profit curves of the new method and normal MA

normal MA experiences more severe fluctuations than the new method. This is because the dual algorithm used in the normal MA uses past working points to calculate the gradients, which should be in the same operating condition. Since at iteration 11 the operating condition changes, the discrepancy leads to very wrongly calculated gradients. However, in the new method the MA is initiated with the FFD upon detection of a new operating condition, hence no such problem exists. (3) After iteration 17 both the two methods attain convergence, where their performances are similar. Notice that, it is not shown in the figure how the new settlement $u_{set,new}$ is implemented, which can further enhance the economic performance.

5. CONCLUSIONS

In this paper, we discussed a practical aspect of implementing RTO schemes. Both the suspending and activating conditions are established to enhance RTO schemes. Unnecessary oscillations around the plant optimum are avoided, whilst the capacity of finding the new optimum is still preserved when the operating condition changes. Furthermore, the local model of cost gradients around the optimum is proposed through regression, which can efficiently locate the true optimum. The final settled operation improves the operational economy. For the activating condition, the process monitoring theory was employed to detect the change of operating condition, which runs in a passive way without perturbing the process. The application to the Williams-Otto reactor was successful, various advantages of the proposed method were confirmed.

REFERENCES

Ariyur, K.B. and Krstic, M. (2003). *Real-time optimization by extremum-seeking control*. John Wiley & Sons.

Brdys, M.A. and Tatjewski, P. (2005). *Iterative algorithms for multilayer optimizing control*. World Scientific.

Chachuat, B., Srinivasan, B., and Bonvin, D. (2009). Adaptation strategies for real-time optimization. *Comput. Chem. Eng.*, 33(10), 1557–1567.

Chen, C.Y. and Joseph, B. (1987). On-line optimization using a two-phase approach: An application study. *Industrial & engineering chemistry research*, 26(9), 1924–1930.

Edgar, T.F., Himmelblau, D.M., and S.Lasdon, L. (2001). *Optimization of Chemical Processes*. McGraw-Hill, Second edition.

Francois, G., Srinivasan, B., and Bonvin, D. (2005). Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. *J. Process Control*, 15(6), 701–712.

Francois, G., Srinivasan, B., and Bonvin, D. (2012). Comparison of Six Implicit Real-Time Optimization Schemes. *J. Eur. Syst. Autom.*, 46(2-3), 291–305. doi: 10.3166/JESA.46.291-305.

Ge, Z. (2017). Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 171, 16 – 25.

Ge, Z., Song, Z., Ding, S.X., and Huang, B. (2017). Data mining and analytics in the process industry: the role of machine learning. *IEEE Access*, 5, 20590–20616.

Jackson, J.E. and Mudholkar, G.S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3), 341–349.

Jäschke, J., Cao, Y., and Kariwala, V. (2017). Self-optimizing control–a survey. *Annual Reviews in Control*, 43, 199–223.

Marchetti, A., Chachuat, B., and Bonvin, D. (2009). Modifier-adaptation methodology for real-time optimization. *Ind. Eng. Chem. Res.*, 48(13), 6022–6033.

Marchetti, A., Chachuat, B., and Bonvin, D. (2010). A dual modifier-adaptation approach for real-time optimization. *J. Process Control*, 20(9), 1027–1037.

Marchetti, A.G., François, G., Faulwasser, T., and Bonvin, D. (2016). Modifier adaptation for real-time optimizationmethods and applications. *Processes*, 4(4), 55.

Roberts, P. (1979). An algorithm for steady-state system optimization and parameter estimation. *International Journal of Systems Science*, 10(7), 719–734.

Skogestad, S. (2000). Plantwide control: The search for the self-optimizing control structure. *J. Process Control*, 10(5), 487–507.

Srinivasan, B. and Bonvin, D. (2007). Real-time optimization of batch processes by tracking the necessary conditions of optimality. *Industrial & engineering chemistry research*, 46(2), 492–504.

Srinivasan, B., Bonvin, D., Visser, E., and Palanki, S. (2003). Dynamic optimization of batch processes - II. role of measurements in handling uncertainty. *Comput. Chem. Eng.*, 27(1), 27–44.

Williams, T.J. and Otto, R.E. (1960). A generalized chemical processing model for the investigation of computer control. *Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics*, 79(5), 458–473.

Ye, L., Cao, Y., Ma, X., and Song, Z. (2014a). A novel hierarchical control structure with controlled variable adaptation. *Ind. Eng. Chem. Res.*, 53(38), 14695–14711.

Ye, L., Cao, Y., and Xiao, Y. (2015). Global approximation of self-optimizing controlled variables with average loss minimization. *Ind. Eng. Chem. Res.*, 54(48), 12040–12053.

Ye, L., Cao, Y., Ma, X., and Song, Z. (2014b). A non-optimality detection technique for continuous processes. *IFAC Proceedings Volumes*, 47(3), 7616–7621.

Zhu, J., Ge, Z., and Song, Z. (2017). Distributed parallel pca for modeling and monitoring of large-scale plant-wide processes with big data. *IEEE Transactions on Industrial Informatics*, 13(4), 1877–1885.