

# Comparison of Modifier Adaptation Schemes in Real-Time Optimization

Weihua Gao, Simon Wenzel, Sebastian Engell

*TU Dortmund, Emil-Figge-Str. 70, 44221 Dortmund  
Germany (Tel: +49-231-7555131; E-mail: weihua.gao@bci.tu-dortmund.de).*

**Abstract:** In model-based real-time optimization, plant-model mismatch can be handled by applying gradient- and bias-corrections to the cost and constraint functions in an iterative optimization procedure. One of the major challenges in practice is the estimation of the plant gradients from noisy measurement data, in particular for the case of several optimization variables. In this paper we compare four modifier adaptation schemes which were proposed to handle noisy data, iterative gradient-modification optimization, dual modifier adaptation, nested modifier adaptation and modifier adaptation with quadratic approximation. Simulation studies for the optimization of the Otto-Williams reactor with plant-model mismatch are used to illustrate the performance of the different schemes.

**Keywords:** Real-time optimization, Model mismatch, Gradient estimation, Modifier adaptation.

## 1. INTRODUCTION

With increasing global competition, companies in the process industries face intense pressure to improve production efficiency, product quality and process safety. As a result, real-time optimization (RTO) is attracting considerable industrial interest. RTO is a model based upper-level optimization system that is operated iteratively in closed loop and provides set-points to the lower-level regulatory control system in order to maintain the process operation as close as possible to the economic optimum. As for any model and optimization based scheme, the success of RTO depends on the quality of the model which is used in the optimization. The effort required for building and maintaining the model is the bottleneck in the deployment of RTO solutions, and even when sophisticated models are used, they will never exactly represent the real process. It is therefore highly desirable to combine the use of models and of the data which is collected during the operation of the plant in order to obtain a real-time optimization scheme that drives the plant to its optimal operation without having to represent each and every phenomenon in the plant accurately in the model.

Modifier adaptation can handle considerable plant-model mismatch by applying empirical gradient- and bias-corrections to the objective and constraint functions in an iterative optimization procedure (Chachuat et al., 2009). One of the major challenges in practice is the estimation of the process gradients from noisy measurement data, in particular for the case of several optimization variables. Finite-difference based approaches suffer from the problem of choosing the right step-size, using a large step-size may decrease the effect of noisy data on the one hand, on the other hand it leads to considerable approximation errors. Different schemes have been developed for handling noisy data in modifier adaptation, iterative gradient-modification optimization (Gao and Engell, 2005), dual modifier adaptation (Marchetti et al., 2010), nested modifier adaptation (Navia, 2012) and modifier adaptation with quadratic approximation (Gao et al., 2015). All schemes are

targeted to acquire reliable and accurate gradient modifiers for modifier adaptation in the presence of noisy. This motivated us to analyse the characteristics of each scheme and make comparisons of their performance.

The rest of this paper is organized as following. First modifier adaptation is reviewed together with the calculation of empirical gradients from data. The four modifier schemes are then described by flow diagrams. Simulation results for the optimization of the Otto-Williams reactor are presented to compare the performance of each scheme.

## 2. MODIFIER ADAPTATION OPTIMIZATION AND GRADIENT ESTIMATION

The general model-based set-point optimization problem can be stated as

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_m(\mathbf{u}) \\ \text{s. t.} \quad & \mathbf{C}_m(\mathbf{u}) \leq \mathbf{0}, \end{aligned} \quad (1)$$

where  $\mathbf{u}$  is a vector of manipulated variables to be optimized,  $J_m(\mathbf{u})$  and  $\mathbf{C}_m(\mathbf{u})$  are the objective and the vector of constraint functions, assumed to be twice differentiable with respect to  $\mathbf{u}$ . The plant objective and constraints are represented by  $J_p(\mathbf{u})$  and  $\mathbf{C}_p(\mathbf{u})$  and their values are only available via plant evaluations. To handle plant-model mismatch, the optimization problem (1) is iteratively adapted as

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_m(\mathbf{u}) + \varepsilon_J^{(k)} + \lambda_J^{(k)}(\mathbf{u} - \mathbf{u}^{(k)}) \\ \text{s. t.} \quad & \mathbf{C}_m(\mathbf{u}) + \varepsilon_C^{(k)} + \lambda_C^{(k)}(\mathbf{u} - \mathbf{u}^{(k)}) \leq \mathbf{0}, \end{aligned} \quad (2)$$

where the superscript ( $k$ ) represents the iteration index. The adaption modifiers are defined by

$$\varepsilon_J^{(k)} = J_p^{(k)} - J_m^{(k)} \quad (3)$$

$$\lambda_J^{(k)} = \left( \nabla J_p^{(k)} - \nabla J_m^{(k)} \right)^T \quad (4)$$

$$\varepsilon_C^{(k)} = \mathbf{C}_p^{(k)} - \mathbf{C}_m^{(k)} \quad (5)$$

$$\lambda_C^{(k)} = \left( \nabla \mathbf{C}_p^{(k)} - \nabla \mathbf{C}_m^{(k)} \right)^T, \quad (6)$$

where  $(\mathbf{u}^{(k)})$  is replaced by the superscript  $(k)$  for simplicity. Note that the bias modifier  $\varepsilon_j^{(k)}$  does not influence the optimum and it is used to correct the objective values which are displayed to users during an optimization run. Let  $\hat{\mathbf{u}}^{(k)}$  denote the solution of (2), the next set-point is updated as

$$\mathbf{u}^{(k+1)} = \hat{\mathbf{u}}^{(k)} + \mathbf{K}(\mathbf{u}^{(k)} - \hat{\mathbf{u}}^{(k)}), \quad (7)$$

where  $\mathbf{K}$  is a diagonal matrix of the damping factors that are within interval  $[0,1]$ .

Modifier adaptation was originally derived from iterative system optimization and parameter estimation (ISOPE) (Roberts, 1979; Brdys and Tatjewski, 2005). Gao and Engell (2005) extended ISOPE to handle process-dependent constraints and proposed a strategy for gradient estimation. This was called iterative gradient-modification optimization (IGMO). Marchetti et al. (2009) studied IGMO both theoretically and experimentally and coined the name ‘‘Modifier Adaptation’’.

The use of the 1<sup>st</sup>-order modifiers requires the computation of the process gradients at the current set-point. In order to decrease the effort for perturbations of the process, the collected data at the previous set-points can be used to estimate the gradients by a finite difference approximation

$$\nabla J_p^{(k)} = [\mathbf{S}^{(k)}]^{-1} \cdot \left[ \left( J_p^{(k)} - J_p^{(k-1)} \right) \quad \dots \quad \left( J_p^{(k)} - J_p^{(k-n_u)} \right) \right]^T$$

$$\nabla \mathbf{C}_p^{(k)} = [\mathbf{S}^{(k)}]^{-1} \cdot \left[ \left( \mathbf{c}_p^{(k)} - \mathbf{c}_p^{(k-1)} \right) \quad \dots \quad \left( \mathbf{c}_p^{(k)} - \mathbf{c}_p^{(k-n_u)} \right) \right]^T, \quad (8)$$

where  $n_u$  is the number of dimensions of  $\mathbf{u}$ , and the matrix  $\mathbf{S}^{(k)}$  is defined as

$$\mathbf{S}^{(k)} = \left[ \left( \mathbf{u}^{(k)} - \mathbf{u}^{(k-1)} \right) \quad \dots \quad \left( \mathbf{u}^{(k)} - \mathbf{u}^{(k-n_u)} \right) \right]^T. \quad (9)$$

### 3. DIFFERENT MODIFIER ADAPTATION SCHEMES

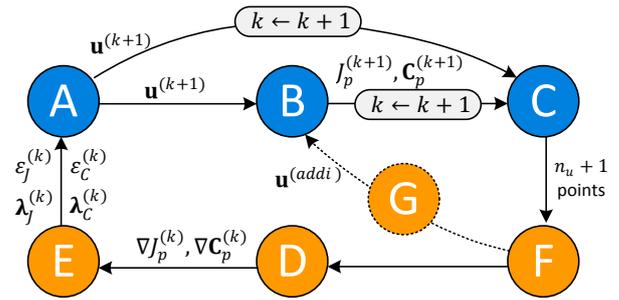
The flow diagrams of four modifier adaptation schemes are compared in Fig. 1, where the circles marked by different letters are used to represent different functional blocks and the lines represent the information flow. The common blocks are listed as follows:

- ‘‘A’’: Adapted optimization (2)
- ‘‘B’’: Plant evaluation
- ‘‘C’’: Collected data,  $(\mathbf{u}^{(i)}, J_p^{(i)}, \mathbf{C}_p^{(i)})$ ,  $i=0 \dots k$
- ‘‘D’’: Finite difference approximation (8)
- ‘‘E’’: Modifier evaluation (3 - 6)
- ‘‘ $k \leftarrow k + 1$ ’’: Run delay.

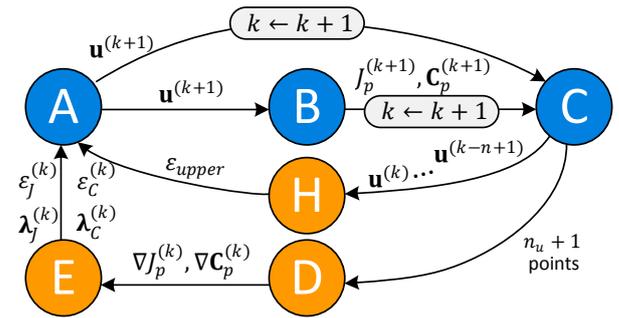
All the other blocks will be defined when they appear for the first time.

#### 3.1 Iterative Gradient-Modification Optimization

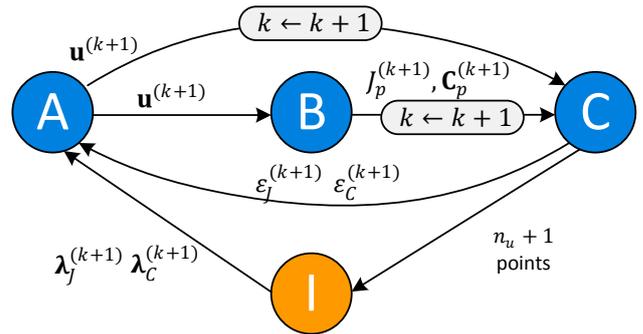
Iterative gradient-modification optimization (Fig. 1a) was proposed by Gao and Engell (2005) for set-point optimization of batch chromatographic separation. In this scheme, the plant gradients are calculated by the finite difference approximation (block ‘‘D’’) from the collected information at



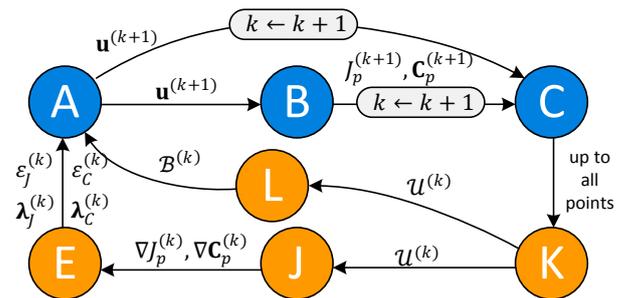
(a) Iterative gradient-modification optimization.



(b) Dual modifier adaptation.



(c) Nested modifier adaptation.



(d) Modifier adaptation with quadratic approximation.

Fig. 1. Illustration of modifier adaptation schemes.

the previous  $(n_u + 1)$  set-points. In order to decrease the influence of measurement noise on the gradient calculation, the conditioning of the set-point change matrix  $\mathbf{S}^{(k)}$  is monitored by the block ‘‘F’’ as

$$\kappa^{-1}(\mathbf{S}^{(k)}) \geq \delta_{cond}. \quad (10)$$

An additional set-point is required when (10) cannot be satisfied. The additional perturbation is optimized (block ‘‘G’’)

by maximizing the inverse of the condition number of  $\mathbf{S}^{a(k)}$  subject to the adapted process constraints:

$$\max_{\mathbf{u}^{(addi)}} \kappa^{-1}(\mathbf{S}^{a(k)}) \quad (11)$$

$$\text{s. t. } \mathbf{C}_m(\mathbf{u}^{(addi)}) + \boldsymbol{\varepsilon}_C^{(k-1)} + \boldsymbol{\lambda}_C^{(k-1)}(\mathbf{u}^{(addi)} - \mathbf{u}^{(k-1)}) \leq 0,$$

where

$$\mathbf{S}^{a(k)} = \left[ \left( \mathbf{u}^{(k)} - \mathbf{u}^{(addi)} \right) \quad \dots \quad \left( \mathbf{u}^{(k)} - \mathbf{u}^{(k-n_u+1)} \right) \right]^T. \quad (12)$$

### 3.2 Dual Modifier Adaptation

Dual modifier adaptation (Fig. 1b) was proposed by Marchetti et al. (2010) to pay attention to the accuracy of the gradients which are calculated with the next set-point move. The adapted optimization problem (2) is augmented by explicitly upper bounding the norm of the gradient estimation error (block ‘‘H’’) that consists of the truncation error and the measurement noise errors

$$\frac{\sigma_{max}}{2} \left\| \mathbf{S}^{-1} \cdot \text{diag}(\mathbf{S} \cdot \mathbf{S}^T) \right\| + \frac{\delta_{noise}}{l_{min}} \leq \varepsilon_{upper}, \quad (13)$$

where the first term on the left side of (13) is the truncation error ( $\sigma_{max}$  is the upper bound on the spectral radius of the Hessian matrix of the process mapping), and the second term is the error due to the measurement noise ( $l_{min}$  is the shortest distance between all possible pairs of complement affine subspaces that can be generated from the set-point set for the gradient estimation, and  $\delta_{noise}$  is the interval of the measurement noise). For simplicity the superscript of the iteration index was dropped here. (13) leads to a dual control choice of the next set-point that balances the convergence to the optimum with the gradient calculation. Since (13) is non-convex, an additional constraint is required to ensure convex feasibility regions:

$$\left( \mathbf{n}^{(k)} \right)^T \mathbf{u} \notin \left( b^{(k)} - \rho^{(k)} \left\| \mathbf{n}^{(k)} \right\|, b^{(k)} + \rho^{(k)} \left\| \mathbf{n}^{(k)} \right\| \right), \quad (14)$$

where  $\mathbf{n}^{(k)}$  is the vector normal to the hyperplane generated by the  $n_u$  most recent set-points, and the hyperplane is defined by  $\left( \mathbf{n}^{(k)} \right)^T \mathbf{u} = b^{(k)}$ , with  $b^{(k)} = \left( \mathbf{n}^{(k)} \right)^T \mathbf{u}^{(k)}$ .  $\rho^{(k)}$  is the minimal point-to-hyperplane distance used to remove the non-convex part of the regions generated by (13). All variables are taken from (Marchetti et al., 2010).

### 3.3 Nested Modifier Adaptation

Nested modifier adaptation (Fig. 1c) was proposed by Navia (2012) to avoid the explicit calculation of plant gradients with noisy data. It directly optimizes the gradient modifiers (block ‘‘I’’) using the Nelder-Mead simplex algorithm with the information collected at the previous  $(n_u + 1)$  set-points. Note that the previous information should be provided with the gradient modifiers as the input variables. Nested modifier adaptation transforms the optimization problem from the set-point space to the gradient-modifier space and applies the simplex algorithm to solve it. Its optimality is based on the assumption that the optimum of the gradient modifier adaptation corresponds to the optimum of the set-point.

### 3.4 Modifier Adaptation with Quadratic Approximation

Modifier adaptation with quadratic approximation (Fig. 1d) was recently proposed by Gao et al. (2015) to combine the

robustness of derivative-free optimization to noisy data with the convergence to the true optimum of modifier adaptation using empirical gradients. Different from the other schemes, which only rely on the latest  $(n_u + 1)$  set-points, all the previous data is involved in this scheme. The data is first selected by a screening algorithm (block ‘‘K’’), which takes distribution, age and distance to the current set-point into consideration, to formulate a suitable regression set  $\mathcal{U}^{(k)}$ . Based on the regression set, the plant mapping is approximated by a quadratic function and the gradients at the current set-point are then evaluated analytically from the quadratic approximation (block ‘‘J’’). A trust region  $\mathcal{B}^{(k)}$  for the next move is also determined by a covariance analysis of the regression set (‘‘L’’) and will be applied in the adapted optimization ‘‘A’’.

## 4. SIMULATION STUDIES

The Otto-Williams reactor (Williams and Otto, 1960) has been used as a benchmark problem to evaluate RTO schemes in Marchetti et al. (2010), Navia (2012), Gao et al. (2015), and Roberts (1979). The real plant is described by three irreversible reactions:



where  $k_1$ ,  $k_2$  and  $k_3$  are the reaction constants. The mismatched model is based on only two reactions:



The ignorance of the reactions with C leads to a structural plant-model mismatch which cannot be handled by model parameter adaptation. The optimization objective is the steady-state profit

$$J = (F_A + F_B)(P_P x_P + P_E x_E) - C_A F_A - C_B F_B. \quad (20)$$

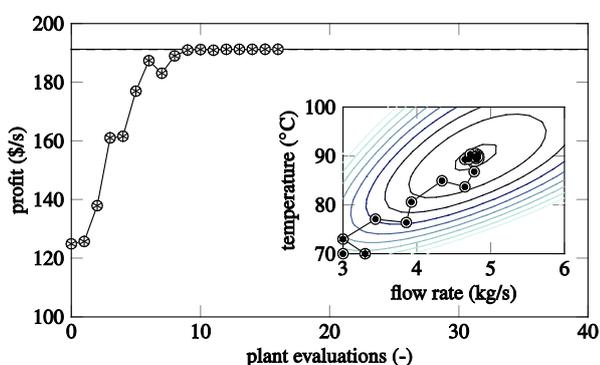
The optimization variables are the flow rate  $F_B$  and the reaction temperature  $\vartheta$ . The variable definitions and parameter values are taken from (Marchetti et al., 2010). In order to test how the new scheme behaves in the presence of noisy data, random noise (Gaussian normal distribution, standard deviation: 0.5) is added to all plant evaluations. The optimization variables are normalized by their operating intervals. The optimization results of four modifier adaptation schemes are presented in Figs. 2 - 6. No damping is applied. All optimizations start from the left-bottom corner (3.0, 70) of the operating region and are run for up to 40 plant evaluations after the starting point. Two initial perturbations are used to estimate the gradients for the starting point by the finite difference approximation (8).

In IGMO, the conditioning threshold  $\delta_{cond} = 0.1$  was used. At each iteration, a bound  $\mathbf{u} \in [\mathbf{u}^{(k)} - 0.2, \mathbf{u}^{(k)} + 0.2]$  is applied to limit the search range for the next set-point. The termination criterion is  $\left\| \mathbf{u}^{(k+1)} - \mathbf{u}^{(k)} \right\| \leq 0.01$ . The noise-free optimization (Fig. 2a) terminates after 16 plant evaluations, including 3 additional perturbations (4<sup>th</sup>, 7<sup>th</sup> and 11<sup>th</sup>). The inset shows the set-point trajectory on the objective

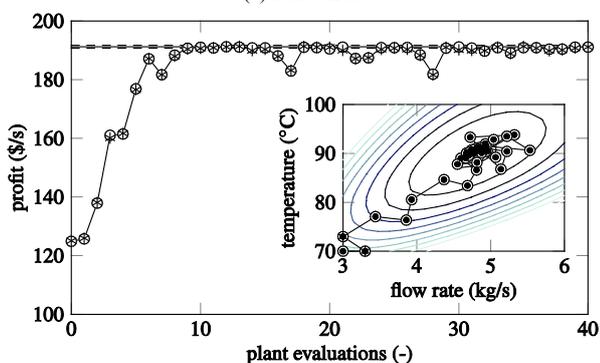
contours. It stays in the vicinity of the optimum after the 9<sup>th</sup> evaluation. The optimization for one realization of the noise (Fig. 2b) does not terminate after 40 plant evaluations. The “star” symbols represent the noisy data, and the “circle” symbols represent the “clean” data. The dashed lines mark the  $\pm 3\sigma$  noise interval centered at the maximal profit. Compared to the noise-free optimization, there is only slight difference during the first 9 evaluations. The noise does not lead to considerable gradient estimation error with large set-point moves. When entering the vicinity of the optimum, the moves are small and lead to large gradient errors. As a result, irregular zig-zag moves appear. Since the gradient error is dominated by the effect of the small step size, the increase of the conditioning threshold  $\delta_{cond}$  will not decrease the influence of the noise. The results of 100 realizations of the noise (Fig. 2c) show that this scheme is sensitive to the noise, where the “crossed-circle” symbols mark the termination set-points. More than 60% of the runs do not satisfy the termination criterion within 20 plant evaluations, and around

20% runs do not terminate after 40 plant evaluations. The set-point distribution of all the runs after 40 plant evaluations (Inset) shows that most set-points are within the 190 (\$/s) contour.

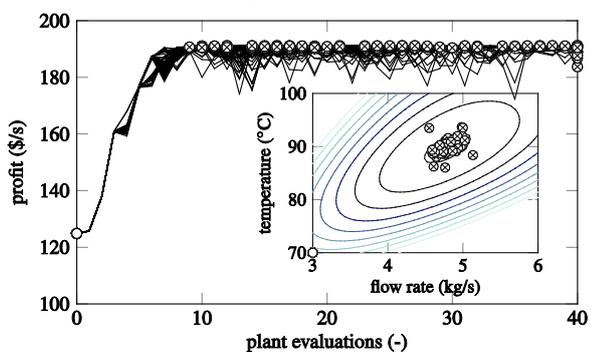
In dual modifier adaptation the upper error norm  $\varepsilon_{upper} = 111.2$ , which was used in (Marchetti et al., 2010), is considered first. For the noise interval  $\delta_{noise} = 3.0$ , the step size should be greater than 0.054 in order to satisfy (13). The termination criterion  $\|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\| \leq 0.01$  cannot be applied and the optimizations are run for 40 plant evaluations. The noise-free optimization (Fig. 3a) takes 14 plant evaluations to reach the vicinity of the optimum (8 taken by IGMO). The zig-zag trajectory around the optimum is due to the lower limit on the step size. The optimization of one realization of the noise (Fig. 3b) shows an enlarged zig-zagging (compared to the noise-free situation) when approaching the optimum. This can be clearly observed from the change of the objective values. The oscillation interval of



(a) Noise free.

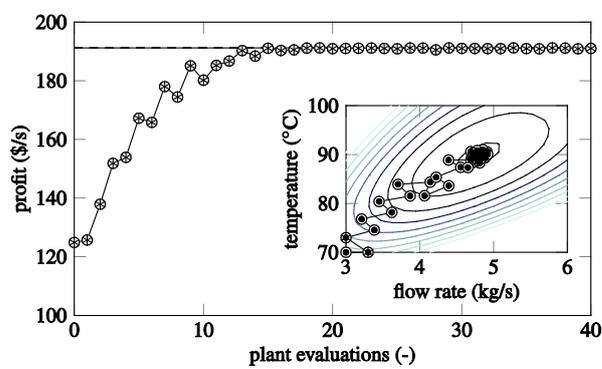


(b) Single run with noise.

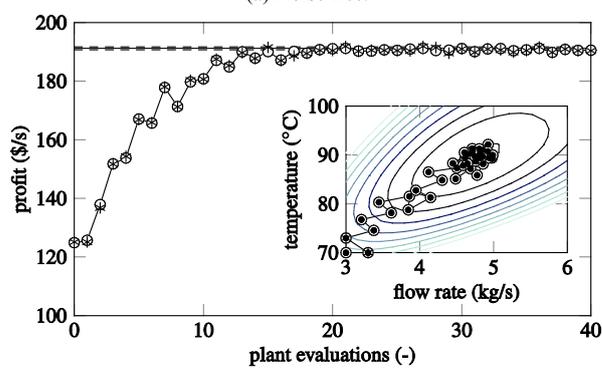


(c) Multiple runs with 100 realizations of the noise.

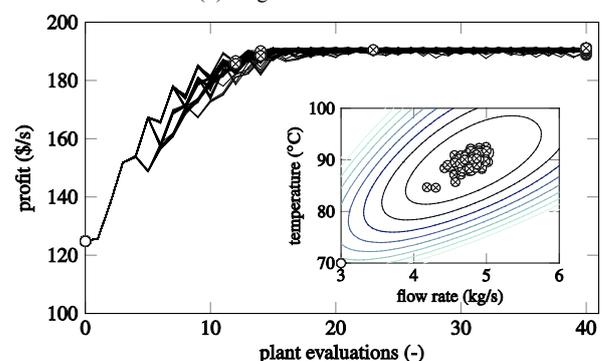
Fig. 2. Iterative gradient-modification optimization.



(a) Noise free.



(b) Single run with noise.



(c) Multiple runs with 100 realizations of the noise.

Fig. 3. Dual modifier adaptation,  $\varepsilon_{upper} = 111.2$ .

multiple runs (Fig. 3c) is much narrower than for IGMO (Fig. 2c). This is attributed to the upper bounding of the gradient error norm, which limits the minimal step size allowed. The maximal step size, on the other hand, is also limited by (13) and it leads to a slow rate of convergence to the optimum. Note that five runs terminated unsuccessfully because MATLAB optimizer FMINCON with SQP algorithm cannot find a feasible point. The unsuccessful runs are assumed to be caused by the implementation of numerical optimization under the tight upper error norm.

Next the upper error norm was relaxed to 200, hence the interval of the admitted step size is enlarged. The minimal step size is 0.03 in order to satisfy (13). Since this is still greater than the termination criterion, the optimizations are again run for 40 plant evaluations. Because large steps are allowed, the noise-free optimization (Fig. 4a) takes only 6 plant evaluations to reach the vicinity of the optimum, less than half of the case with 111.2. The optimization of one realization of the noise (Fig. 4b) shows large zig-zags, as the result of smaller step-sizes used there. The multiple runs

(Fig. 4c) and the final distribution of set-points show that dual modifier adaptation is more sensitive to the noise now.

Nested modifier adaptation transforms the optimization problem from the set-point space to the gradient-modifier space. Four plant evaluations are required to initialize the scheme, of which three are used to calculate the initial values of the gradient modifiers, and the 4<sup>th</sup> corresponds to the plant evaluation at the initial gradient modifiers. The noise-free optimization (Fig. 5a) shows a slow and irregular rate of convergence to the optimum. It takes 25 plant evaluations to reach the vicinity of the optimum. The optimization for one realization of the noise (Fig. 5b) shows that the nested modifier adaptation does not converge to the optimum. Similar results can be observed from the multiple runs (Fig. 5c) and the final distribution of set-points. Although the optimum of the gradient modifiers corresponds to the optimal set-point, there is no guarantee that the convexity of the optimization problem in the set-point space is inherited in the gradient-modifier space.

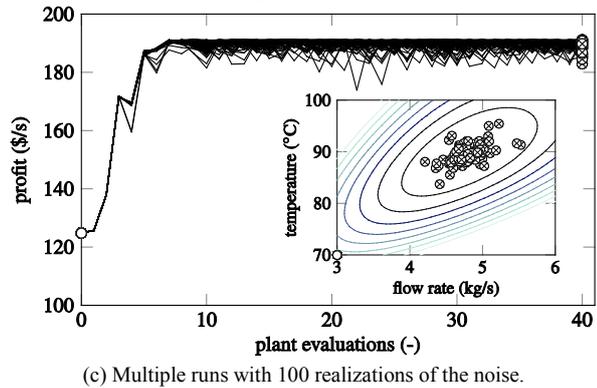
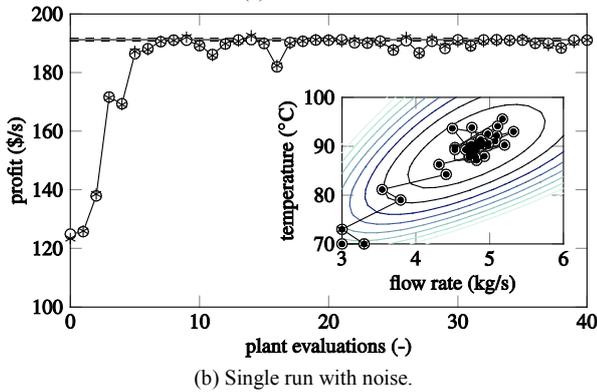
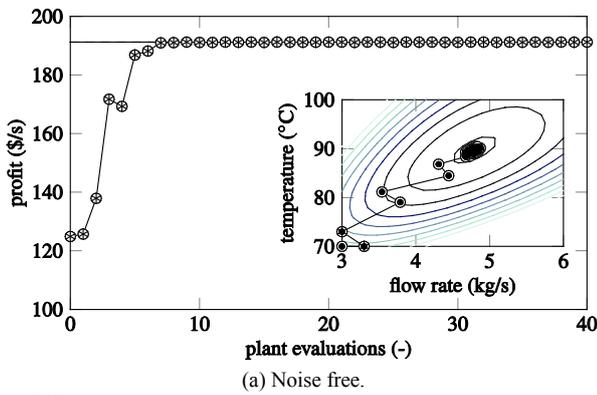


Fig. 4. Dual modifier adaptation,  $\epsilon_{upper} = 200$ .

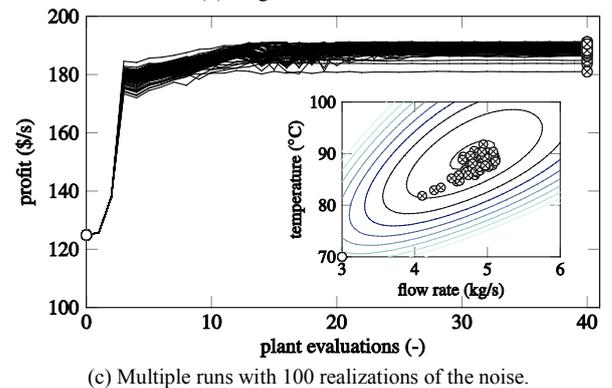
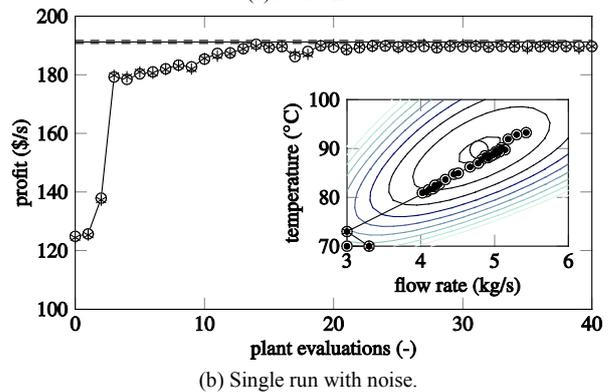
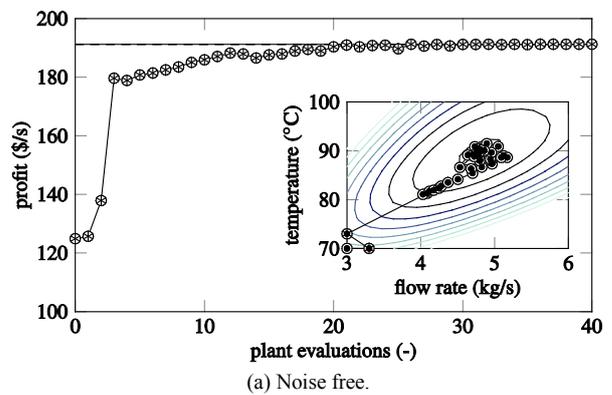


Fig. 5. Nested modifier adaptation optimization.

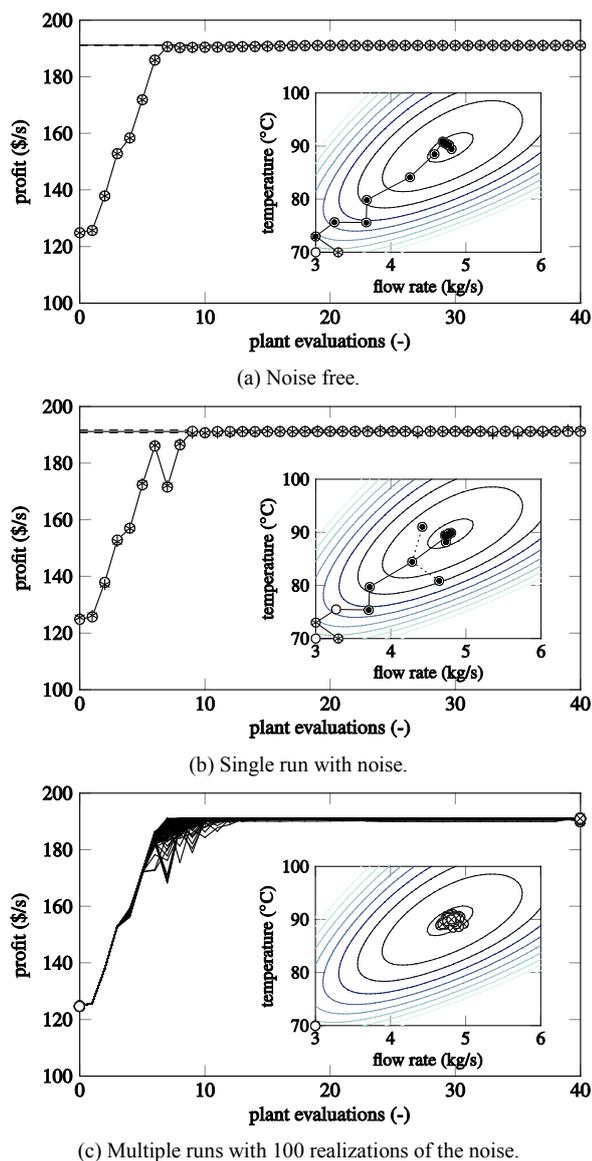


Fig. 6. Modifier adaptation with quadratic approximation.

Fig. 6 shows the optimization results of modifier adaptation with quadratic approximation. The parameters are taken from (Gao et al., 2015) and no termination criterion is applied. Before enough points (in this case 6) are acquired for the quadratic approximation, IGMO is run. After the 6<sup>th</sup> plant evaluation, the quadratic approximation starts to provide gradient estimates. The noise-free optimization (Fig. 6a) reaches the vicinity of the optimum after 7 plant evaluations, and 9 evaluations are performed by the optimization of one realization of the noise (Fig. 6b). Note that the 8<sup>th</sup> and 9<sup>th</sup> plant evaluations are on each sides of the centre contour (inset of Fig. 6b). They are generated by the optimization scheme to provide curvature information along the perpendicular direction to the previous set-point trajectory. Similar plant evaluations, but with smaller step sizes, can be observed in the noise-free run. These phenomena illustrate the characteristics of the modifier adaptation with quadratic approximation: it makes full use of the previous information and make only additional probes when necessary. All optimization runs of the 100 realizations of the noise (Fig. 6c) converge to the optimum accurately after 14 plant evaluations.

The inset shows the final distribution of the set-points and a considerable improvement of the spread of the solutions can be observed.

## 6. CONCLUSIONS

This paper provided a comparison of four different modifier adaptation schemes. All schemes can converge to the plant optimum when noise-free data is used to calculate the gradient modifiers. In the presence of noise, iterative gradient-modification optimization can efficiently improve the set-point when far away from the optimum. In the vicinity of the optimum, the decrease of step-size leads to considerable gradient errors and therefore irregular zig-zag moves. Dual modifier adaptation suffers from the problem of choosing the upper error norm, in which using a small value decreases the rate of convergence considerably, and using a large value leads to oscillations near the optimum. Nested modifier adaptation cannot guarantee optimality and is also characterized by a slow rate of convergence. To nest modifier adaptation with a derivative-free optimization approach may lead to non-convexity of the optimization problem. The best performance, in terms of efficiency and reliability, is attained by modifier adaptation with quadratic approximation. This scheme combines the quadratic approximation of derivative-free optimization with the iterative gradient-modification approach and integrates recent advances in both areas. The simulation results for the optimization of the Otto-Williams reactor demonstrate an impressive performance compared to the other schemes. Further studies with more degrees of freedom, process-dependent constraints, and varying disturbances will be performed to investigate the limitations of this scheme.

## REFERENCES

- Brdys, M. and Tatjewski, P. (2005). *Iterative algorithms for multilayer optimizing control*. London, UK: Imperial College Press.
- Chachuat, B., Srinivasan, B. and Bonvin, D. (2009). Adaptation strategies for real-time optimization. *Computers & Chemical Engineering*, 1557-1567.
- Gao, W. and Engell, S. (2005). Iterative set-point optimization of batch chromatography. *Computers & Chemical Engineering*, 1401-1409.
- Gao, W., Wenzel, S. and Engell, S. (2015). Modifier adaptation with quadratic approximation in iterative optimizing control. Accepted for *2015 European Control Conference*, Linz, Austria.
- Marchetti, A., Chachuat, B. and Bonvin, D. (2009). Modifier-adaptation methodology for real-time optimization. *I&EC Research*, 6022-6033.
- Marchetti, A., Chachuat, B. and Bonvin, D. (2010). A dual modifier-adaptation approach for real-time optimization. *Journal of Process Control*, 1027-1037.
- Navia, D. (2012). *Handling Uncertainties in Process Optimization*, PhD Thesis, Universidad de Valladolid.
- Roberts, P.D. (1979). An algorithm for steady-state system optimization and parameter estimation. *International Journal on System Science*, 719-734.
- Williams, T.J. and Otto, R.E. (1960). A Generalized Chemical Processing Model for the Investigation of Computer Control. *AIEE Trans*, 458 - 473.