Computers and Chemical Engineering 59 (2013) 89–100

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

# A new dual modifier-adaptation approach for iterative process optimization with inaccurate models

A.G. Marchetti *

*French-Argentine International Center for Information and Systems Sciences (CIFASIS), CONICET-Universidad Nacional de Rosario (UNR), 27 de Febrero 210bis, S2000EZP Rosario, Argentina*

ARTICLE INFO

*Article history:*
Received 31 July 2012
Received in revised form 16 March 2013
Accepted 19 March 2013
Available online 3 April 2013

*Keywords:*
Iterative process optimization
Real-time optimization
Modifier adaptation

ABSTRACT

In order to deal with plant-model mismatch, iterative process optimization schemes use some adaptation strategy based on measurements. The modifier-adaptation approach consists in performing first-order corrections of the cost and constraint functions in the model-based optimization problem. The approach has the ability to converge to the true process optimum but the first-order corrections require the experimental estimation of the process gradients. Dual modifier-adaptation algorithms estimate the gradients by finite difference approximation based on the measurements obtained at the current and past operating points. In order to guarantee the accuracy of the estimated gradients a constraint is added to the optimization problem in order to position the next operating points with respect to the previous ones. This paper presents an alternative first-order correction, which provides an improved approximation of the cost and constraint functions, together with a new gradient error constraint for use in dual modifier adaptation. By means of the Williams–Otto reactor case study, the new dual modifier-adaptation approach is compared in simulation with a previous approach found in the literature showing faster convergence to a neighborhood of the plant optimum.



## 1. Introduction

The optimal operation of many industrial chemical processes typically involves the solution of an optimization problem which relies on first-principle models and is subject to operating constraints. This paper considers the case where the optimization problem can be formulated as a nonlinear programming (NLP) problem (Bazaraa, Sherali, & Shetty, 2006). In the chemical process industries, NLP problems can be formulated for the real-time optimization (RTO) of the steady-state operation of continuous processes (Darby, Nikolaou, Jones, & Nicholson, 2011; Marlin & Hrymak, 1997), and for batch-to-batch optimization of the recipe of batch and semi-batch chemical processes (François, Srinivasan, & Bonvin, 2005).

In the presence of plant-model mismatch and unmeasured disturbances, the operating point obtained by optimization using the model is not in general an optimum point for the true process, or worse, might not be a feasible point for the true process (Gao & Engell, 2005; Marchetti, Chachuat, & Bonvin, 2009). In order to deal with model inaccuracies, iterative process optimization schemes use some adaptation strategy based on measurements. The modifier-adaptation approach is to make first-order corrections of the cost and constraint functions in the optimization problem by adapting the values of bias modifiers expressing the difference between the constraints and gradients of the plant and the model at the current operating point (Gao & Engell, 2005; Marchetti et al., 2009). The approach has the ability to converge to the true process optimum, but it requires the experimental estimation of the process gradients. In order to estimate the process gradients, one approach is to perturb the inputs around the current operating point, for example, one can use forward finite differencing (FFD) at each RTO iteration. An alternative approach proposed in Brdyś and Tatjewski (1994, 2005) in the context of the dual ISOPE (Integrated System Optimization and Parameter Estimation) algorithm is to estimate the gradients based on the output measurements at the current and past operating points. In order to obtain accurate gradient estimates, an additional constraint is added to the modified optimization problem. This constraint takes into account the location of the current and past operating points in order to position the next operating point. A constraint that ensures that the new operating point does not introduce ill-conditioning in the estimation of the gradients was used in Brdyś and Tatjewski (1994, 2005) and Gao and Engell (2005). However, this constraint is not directly related with the error in the gradient estimates (Marchetti, Chachuat, & Bonvin, 2010). Recently, a *gradient error constraint* that bounds the gradient error norm was introduced in Marchetti et al. (2010). This constraint takes into account the two main sources of gradient error, namely the truncation error introduced by

* Tel.: +54 341 4237 248.

0098-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2013.03.019

finite-difference approximation of the derivatives and the measurement noise. In general, the truncation error increases if the points are too separated from each other, the error due to measurement noise increases if the points are too close to each other, and both types of errors increase if all the points are close to the same hyperplane (i.e., the set of points is ill-conditioned). Since the first-order corrections are made at the current operating point, the gradient error constraint limits the input moves generated by the dual modifier-adaptation algorithm in order to limit the truncation error of the gradient estimates. This may result in a slow convergence to the plant optimum in some cases.

This paper proposes new gradient modifiers for correcting the cost and constraint gradients predicted by the model in the optimization problem, together with a new gradient error constraint for use in dual modifier-adaptation. When compared with the first-order corrections used in Gao and Engell (2005) and Marchetti et al. (2009), the first-order corrections of the cost and constraint functions obtained with the new gradient modifiers provide more accurate approximations of the cost and constraint functions for the real process. In turn, the new constraint produces larger feasible regions for the same upper bound on the gradient error norm. This allows for much larger input moves and therefore faster convergence to the optimum. This paper significantly extends the results presented in Marchetti and Basualdo (2012), where the proposed dual modifier-adaptation scheme is initially described for the particular case of an unconstrained optimization problem.

The paper is organized as follows. The optimization problem is formulated in Section 2 and the necessary conditions of optimality (NCO) are presented. The fundamentals of modifier-adaptation are discussed in Section 3, including two novel features: (i) the model adequacy criterion given in Marchetti et al. (2009) is restated including conditions for model inadequacy, and (ii) a sensitivity analysis is conducted which analyzes the optimality loss induced by small errors in the gradient of the Lagrangian function. Also in Section 3, the dual modifier-adaptation approach proposed in Marchetti et al. (2010) is presented, including an upper bound on the gradient error norm of the Lagrangian function. The sensitivity analysis previously conducted gives support to the choice of the Lagrangian function as the function for which the gradient error should be constrained. The new dual modifier-adaptation approach proposed in this paper is presented in Section 4, including the alternative gradient modifiers, and the new gradient error constraint previously mentioned. The proposed approach is illustrated for the case study of the Williams–Otto reactor in Section 5, and it is compared with the previous approach described in Marchetti et al. (2010). Finally, Section 6 concludes the paper.

## 2. Optimization problem

### 2.1. Problem formulation

The optimization problem for the plant (real process) can be formulated as follows:

$$\begin{aligned} \mathbf{u}_p^\star = \underset{\mathbf{u}}{\operatorname{argmin}} \quad & \Phi_p(\mathbf{u}) := \phi(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \\ \text{s.t.} \quad & \mathbf{G}_p(\mathbf{u}) := \mathbf{g}(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \le \mathbf{0}, \end{aligned} \tag{1}$$

where $\mathbf{u} \in \mathrm{IR}^{n_u}$ denotes the decision (or input) variables and $\mathbf{y}_p \in \mathrm{IR}^{n_y}$ the measured (or output) variables; $\phi : \mathrm{IR}^{n_u} \times \mathrm{IR}^{n_y} \to \mathrm{IR}$ is the cost function to be minimized; $g_i : \mathrm{IR}^{n_u} \times \mathrm{IR}^{n_y} \to \mathrm{IR}$, $i = 1, \ldots, n_g$, is the set of inequality constraint functions, which includes the input bounds. The notation $(\cdot)_p$ is used throughout for the variables associated with the plant.

This formulation assumes that $\phi(\mathbf{u}, \mathbf{y}_p)$ and $\mathbf{g}(\mathbf{u}, \mathbf{y}_p)$ are known functions of $\mathbf{u}$ and $\mathbf{y}_p$, i.e., they can be evaluated from the knowledge of $\mathbf{u}$ and measurement of $\mathbf{y}_p$. On the other hand, the steady-state input–output mapping of the plant, $\mathbf{y}_p(\mathbf{u})$, is typically unknown, and only an approximate model is available. The steady-state input–output mapping predicted by the model is denoted $\mathbf{y}(\mathbf{u}, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathrm{IR}^{n_\theta}$ is the set of model parameters. Using the model, the solution of the original problem (1) can be approached by solving the following NLP problem:

$$\begin{aligned} \mathbf{u}^\star = \underset{\mathbf{u}}{\operatorname{argmin}} \quad & \Phi(\mathbf{u}, \boldsymbol{\theta}) := \phi(\mathbf{u}, \mathbf{y}(\mathbf{u}, \boldsymbol{\theta})) \\ \text{s.t.} \quad & \mathbf{G}(\mathbf{u}, \boldsymbol{\theta}) := \mathbf{g}(\mathbf{u}, \mathbf{y}(\mathbf{u}, \boldsymbol{\theta})) \le \mathbf{0}. \end{aligned} \tag{2}$$

In the presence of plant-model mismatch, a model-based solution $\mathbf{u}^\star$ does not generally match the plant optimum $\mathbf{u}_p^\star$, so some adaptation is needed.

### 2.2. Necessary conditions of optimality

Provided that a constraint qualification holds at the solution point $\mathbf{u}^\star$ and the functions $\Phi$ and $\mathbf{G}$ are differentiable at $\mathbf{u}^\star$, there exist unique Lagrange multipliers $\boldsymbol{\mu}^\star \in \mathrm{IR}^{n_g}$ such that the so-called first-order Karush–Kuhn–Tucker (KKT) conditions hold at $\mathbf{u}^\star$, and $\mathbf{u}^\star$ is called a KKT point (Bazaraa et al., 2006):

$$\begin{aligned} & \mathbf{G} \le \mathbf{0}, \quad \boldsymbol{\mu}^{\mathrm{T}}\mathbf{G} = 0, \quad \boldsymbol{\mu} \ge \mathbf{0}, \\ & \frac{\partial \mathcal{L}}{\partial \mathbf{u}} = \frac{\partial \Phi}{\partial \mathbf{u}} + \boldsymbol{\mu}^{\mathrm{T}} \frac{\partial \mathbf{G}}{\partial \mathbf{u}} = \mathbf{0}, \end{aligned} \tag{3}$$

with $\mathcal{L}(\mathbf{u}, \boldsymbol{\mu}) := \Phi(\mathbf{u}) + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{G}(\mathbf{u})$ being the Lagrangian function.

The set of active constraints at $\mathbf{u}^\star$ is denoted by $\mathcal{A} := \{i : G_i(\mathbf{u}^\star) = 0, \quad i = 1, \ldots, n_g\}$. The Hessian of the Lagrangian function at $\mathbf{u}^\star$ is given by:

$$\frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}(\mathbf{u}^\star, \boldsymbol{\mu}^\star) = \frac{\partial^2 \Phi}{\partial \mathbf{u}^2}(\mathbf{u}^\star) + \sum_{i \in \mathcal{A}} \mu_i^\star \frac{\partial^2 G_i}{\partial \mathbf{u}^2}(\mathbf{u}^\star).$$

The active constraints at $\mathbf{u}^\star$ are denoted as $\mathbf{G}^a(\mathbf{u}^\star) \in \mathrm{IR}^{n_g^a}$, where $n_g^a$ is the number of elements of $\mathcal{A}$. Assuming that the Jacobian of the active constraints at $\mathbf{u}^\star$ has full row rank, we can write the relation:

$$\frac{\partial \mathbf{G}^a}{\partial \mathbf{u}}(\mathbf{u}^\star)\mathbf{Z} = \mathbf{0},$$

where the columns of $\mathbf{Z} \in \mathrm{IR}^{n_u \times (n_u - n_g^a)}$ are a set of basis vectors for the null space of the active constraint Jacobian. The reduced Hessian of the Lagrangian, $\nabla_r^2 \mathcal{L}(\mathbf{u}^\star) \in \mathrm{IR}^{(n_u - n_g^a) \times (n_u - n_g^a)}$, is given by (Gill et al., 2003):

$$\nabla_r^2 \mathcal{L}(\mathbf{u}^\star) := \mathbf{Z}^{\mathrm{T}} \left[ \frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}(\mathbf{u}^\star, \boldsymbol{\mu}^\star) \right] \mathbf{Z}.$$

A second-order necessary condition for a local minimum is the requirement that $\nabla_r^2 \mathcal{L}(\mathbf{u}^\star) \ge 0$ (positive semidefinite) at the solution point $\mathbf{u}^\star$. On the other hand, the condition $\nabla_r^2 \mathcal{L}(\mathbf{u}^\star) > 0$ (positive definite) is a second-order sufficient condition for a strict local minimum (Gill et al., 2003).

## 3. Dual modifier adaptation

### 3.1. The modifier-adaptation approach

The modifier-adaptation approach uses measurements to correct the predicted cost and constraints at each RTO iteration, in such a way that a KKT point for the plant is reached upon convergence (Marchetti et al., 2009). At the $k$th iteration, the next optimal input values are computed by solving a modified optimization problem

that includes first-order corrections of the plant and constraint functions:

$$\begin{aligned}\mathbf{u}_{k+1} = \underset{\mathbf{u}}{\mathrm{argmin}} \quad & \Phi_m(\mathbf{u},\boldsymbol{\theta}) := \Phi(\mathbf{u},\boldsymbol{\theta}) + \varepsilon_k^{\Phi} + (\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{G}_m(\mathbf{u},\boldsymbol{\theta}) := \mathbf{G}(\mathbf{u},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k^{\mathbf{G}} + (\boldsymbol{\lambda}_k^{\mathbf{G}})^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_k) \le \mathbf{G}^{\mathrm{U}},\end{aligned} \tag{4}$$

where $\varepsilon_k^{\Phi} \in \mathrm{IR}$ is the cost value modifier; $\boldsymbol{\varepsilon}_k^{\mathbf{G}} \in \mathrm{IR}^{n_g}$ are the constraint value modifiers; $\boldsymbol{\lambda}_k^{\Phi} \in \mathrm{IR}^{n_u}$ is the cost gradient modifier; and $\boldsymbol{\lambda}_k^{\mathbf{G}} \in \mathrm{IR}^{n_u \times n_g}$ are the constraint gradient modifiers. These modifiers represent the differences between the plant and predicted values of certain KKT-related elements at the current operating point $\mathbf{u}_k$:

$$\varepsilon_k^{\Phi} = \Phi_p(\mathbf{u}_k) - \Phi(\mathbf{u}_k,\boldsymbol{\theta}), \tag{5}$$

$$\boldsymbol{\varepsilon}_k^{\mathbf{G}} = \mathbf{G}_p(\mathbf{u}_k) - \mathbf{G}(\mathbf{u}_k,\boldsymbol{\theta}), \tag{6}$$

$$(\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}} = \frac{\partial \Phi_p}{\partial \mathbf{u}}(\mathbf{u}_k) - \frac{\partial \Phi}{\partial \mathbf{u}}(\mathbf{u}_k,\boldsymbol{\theta}), \tag{7}$$

$$(\boldsymbol{\lambda}_k^{\mathbf{G}})^{\mathrm{T}} = \frac{\partial \mathbf{G}_p}{\partial \mathbf{u}}(\mathbf{u}_k) - \frac{\partial \mathbf{G}}{\partial \mathbf{u}}(\mathbf{u}_k,\boldsymbol{\theta}). \tag{8}$$

The modifiers and KKT-related elements in (5)–(8) can be denoted collectively as $n_K$-dimensional vectors,

$$\boldsymbol{\Lambda}^{\mathrm{T}} := \left(\varepsilon^{\Phi}, \varepsilon^{G_1}, \ldots, \varepsilon^{G_{n_g}}, \boldsymbol{\lambda}^{\Phi\,\mathrm{T}}, \boldsymbol{\lambda}^{G_1\,\mathrm{T}}, \ldots, \boldsymbol{\lambda}^{G_{n_g}\,\mathrm{T}}\right)$$

$$\mathbf{C}^{\mathrm{T}} := \left(\Phi, G_1, \ldots, G_{n_g}, \frac{\partial \Phi}{\partial \mathbf{u}}, \frac{\partial G_1}{\partial \mathbf{u}}, \ldots, \frac{\partial G_{n_g}}{\partial \mathbf{u}}\right),$$

with $n_K = (n_u+1)(n_g+1)$. This way, (5)–(8) can be rewritten as:

$$\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}(\mathbf{u}_k) = \mathbf{C}_p(\mathbf{u}_k) - \mathbf{C}(\mathbf{u}_k,\boldsymbol{\theta}). \tag{9}$$

However, due to plant-model mismatch, this adaptation strategy may lead to excessive correction in some cases, thereby compromising the convergence of the algorithm (see the local convergence analysis and example 3 in (Marchetti et al., 2009)). For this reason, it is convenient to filter the modifiers using for example a first-order exponential filter as in Marchetti et al. (2009):

$$\boldsymbol{\Lambda}_{k+1} = (\mathbf{I} - \mathsf{K})\boldsymbol{\Lambda}_k + \mathsf{K}\left[\mathbf{C}_p(\mathbf{u}_{k+1}) - \mathbf{C}(\mathbf{u}_{k+1},\boldsymbol{\theta})\right] \tag{10}$$

where $\mathsf{K} \in \mathrm{IR}^{n_K \times n_K}$ is a gain matrix. A possible choice for K is the block-diagonal matrix

$$\mathsf{K} := \mathrm{diag}\left(1, d_1, \ldots, d_{n_g}, q\mathbf{I}_{n_u}, p_1\mathbf{I}_{n_u}, \ldots, p_{n_g}\mathbf{I}_{n_u}\right),$$

where the gain entries $d_1, \ldots, d_{n_g}, q, p_1, \ldots, p_{n_g}$ are taken in $(0, 1]$. Notice that the modifier $\varepsilon_k^{\Phi}$ does not require filtering. Indeed, since $\varepsilon_k^{\Phi}$ is a constant term added to the cost function, its value does not affect the solution to Problem (4).

Modifier adaptation has the appealing property that, upon convergence and in the absence of noise, the operating point $\mathbf{u}_\infty$ is a KKT point for Problem (1) (Marchetti et al., 2009). Its downside lies in the need to estimate the experimental gradients $\partial\mathbf{G}_p/\partial\mathbf{u}$ and $\partial\Phi_p/\partial\mathbf{u}$.

### 3.2. Model adequacy criterion

The problem of model adequacy for real-time model-based optimization is discussed in Forbes and Marlin (1996). A process model is said to be adequate for use in an RTO scheme if it is capable of producing a fixed point for that RTO scheme at the plant optimum $\mathbf{u}_p^\star$. Based on this definition, model adequacy for modifier-adaptation schemes was analyzed in Marchetti et al. (2009). The plant optimum $\mathbf{u}_p^\star$ satisfies the first- and second-order NCO for the original optimization problem (1). In order for $\mathbf{u}_p^\star$ to be a fixed point of the modifier-adaptation algorithm, $\mathbf{u}_p^\star$ must also satisfy the first- and second-order NCO for the modified optimization problem (4), with the modifiers (6)–(7) evaluated at $\mathbf{u}_p^\star$. Interestingly, by choosing $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\mathbf{u}_p^\star)$, the first-order NCO of problem (4) are automatically satisfied (since modifier adaptation matches the first-order KKT elements of the plant). Only the second-order NCO remain to be satisfied, that is, the reduced Hessian of the Lagrangian must be positive semidefinite at $\mathbf{u}_p^\star$: $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta}) \ge 0$. This is a requirement that the model must satisfy. The model adequacy criterion is restated below, including additional precisions concerning model inadequacy.

**Criterion 1** *(Model Adequacy for Modifier Adaptation).* Let $\mathbf{u}_p^\star$ be the unique plant optimum, which is assumed to be a regular point for the constraints. Let $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta})$ be the reduced Hessian of the Lagrangian at $\mathbf{u}_p^\star$,

- If $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta}) > 0$ (positive definite), then the process model is said to be **adequate** for use in the modifier adaptation RTO scheme.
- If $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta}) < 0$ (negative definite), then the process model is said to be **inadequate** for use in the modifier adaptation RTO scheme.
- If $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta})$ is singular, then the model adequacy criterion is not conclusive.

The condition $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta}) > 0$ represents a (local) sufficient condition for model adequacy. It can be viewed as a sufficient condition for $\mathbf{u}_p^\star$ to be a fixed point of the modifier-adaptation algorithm. However, this condition says nothing about the convergence of modifier adaptation to the plant optimum $\mathbf{u}_p^\star$ (a local convergence analysis for the modifier-adaptation algorithm is given in Marchetti (2009) and Marchetti et al. (2009)). On the other hand, the condition $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta}) < 0$ represents a sufficient condition for model inadequacy (since the second-order necessary conditions of optimality are not satisfied). Hence, if according to Criterion 1 the model is inadequate, $\mathbf{u}_p^\star$ is not a fixed point of the modifier-adaptation algorithm, and there is no chance modifier adaptation will converge to the plant optimum $\mathbf{u}_p^\star$. At last, if $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta})$ is singular, then the model can be adequate or inadequate depending on each particular case. For example, consider the problem $\min_u \ \Phi_p(u) = 2u^2$, for which $u_p^\star = 0$. The model $\Phi_1(u) = 2u$ has $d^2\Phi_1/du^2(u_p^\star) = 0$ and is inadequate, while the model $\Phi_2(u) = u^4$ also has $d^2\Phi_2/du^2(u_p^\star) = 0$ and is adequate (because $u_p^\star = 0$ minimizes $\Phi_2(u)$).

**Example 1.** Consider the following optimization problem for the plant:

$$\begin{aligned}\min_{\mathbf{u}} \quad & \Phi_p = -(u_1+4)^2 - (u_2+6)^2 \\ \text{s.t.} \quad & G_p = (u_1+1)^2 + (u_2+1)^2 \le 100, \\ & 0 \le u_1, \quad u_2 \le 10,\end{aligned}$$

which is comprised of two decision variables $\mathbf{u} = [u_1, u_2]^{\mathrm{T}}$. The inequality constraint on $G_p$ is active at the optimal solution point, $\mathbf{u}_p^\star = [4.1450, \quad 7.5749]^{\mathrm{T}}$, with the associated Lagrange multiplier $\mu^\star = 1.5831$. On the other hand, the optimization problem for the model reads:

$$\begin{aligned}\min_{\mathbf{u}} \quad & \Phi = -\theta_1(u_1+3)^2 - \theta_1(u_2+5)^2 \\ \text{s.t.} \quad & G = \theta_2(u_1+2)^2 + \theta_2 u_2^2 \le 100, \\ & 0 \le u_1, \quad u_2 \le 10.\end{aligned}$$

The parameter values $\theta_1$ and $\theta_2$ for the models A and B are reported in Table 1. The values of the reduced Hessian of the Lagrangian (which is a scalar in this case) are also given in Table 1, showing that Model A is adequate according to Criterion 1, while Model B is

**Table 1**
Adequate and inadequate models in Example 1.

| | $\theta_1$ | $\theta_2$ | $\nabla_r^2\mathcal{L}(\mathbf{u}_p^\star,\boldsymbol{\theta})$ | Model adequacy |
|---|---|---|---|---|
| Model A | 1.1 | 0.9 | 0.6496 | Adequate |
| Model B | 1.4 | 0.7 | -0.5837 | Inadequate |

inadequate. In this example the parameter values were placed so as to modify the second-order derivatives of the model with respect to the second-order derivatives of the plant (simulated reality) for the purpose of illustrating the adequacy conditions. The more the parameter values deviate from 1, the more the second-order derivatives of the model deviate from those of the plant. Notice that the deviations from 1 are larger for Model B, which is why this model turns out to be inadequate.

A difficulty in using this criterion as a tool for model selection in RTO technologies is that the criterion is point-wise, i.e., it is valid only at the plant optimum, which is typically unknown. In addition, the plant optimum might change in the presence of changing operating conditions and process disturbances. For instance, a model that is adequate might become inadequate if the optimization problem changes (e.g., if the boundary value for some inequality constraint is modified, or if the number of decision variables is increased), or if a perturbation takes place that modifies the steady-state input–output mapping of the plant. In both situations, the model adequacy conditions must be re-evaluated at the new plant optimum.

### 3.3. Principles of dual modifier adaptation

Dual modifier-adaptation algorithms estimate the gradients based on the measurements obtained at the current and past operating points $\mathbf{u}_k$, $\mathbf{u}_{k-1}$, ..., $\mathbf{u}_{k-n_u}$. At the $k$th RTO iteration, the following matrices can be constructed:

$$\mathcal{U}_k := [\,\mathbf{u}_k-\mathbf{u}_{k-1},\quad \mathbf{u}_k-\mathbf{u}_{k-2},\quad \ldots,\quad \mathbf{u}_k-\mathbf{u}_{k-n_u}\,]\in \mathrm{IR}^{n_u\times n_u}, \tag{11}$$

$$\mathcal{Y}_{p,k}^{\Phi} := [\,\tilde{\Phi}_{p,k}-\tilde{\Phi}_{p,k-1},\quad \tilde{\Phi}_{p,k}-\tilde{\Phi}_{p,k-2},\quad \ldots,\quad \tilde{\Phi}_{p,k}-\tilde{\Phi}_{p,k-n_u}\,]\in \mathrm{IR}^{1\times n_u}, \tag{12}$$

$$\mathcal{Y}_{p,k}^{\mathbf{G}} := [\,\tilde{\mathbf{G}}_{p,k}-\tilde{\mathbf{G}}_{p,k-1},\quad \tilde{\mathbf{G}}_{p,k}-\tilde{\mathbf{G}}_{p,k-2},\quad \ldots,\quad \tilde{\mathbf{G}}_{p,k}-\tilde{\mathbf{G}}_{p,k-n_u}\,]\in \mathrm{IR}^{n_g\times n_u}, \tag{13}$$

where the tilde over a variable is used to denote a noisy measurement. For example, the measured plant cost is:

$$\tilde{\Phi}_{p,k} = \phi(\mathbf{u}_k, \mathbf{y}_p(\mathbf{u}_k)+\boldsymbol{\nu}) = \Phi_p(\mathbf{u}_k)+\nu, \tag{14}$$

where $\boldsymbol{\nu}$ is the output measurement noise vector, and $\nu$ represents the resulting noise in the cost. Using these matrices, the plant gradients $\partial\Phi_p/\partial\mathbf{u}$ and $\partial\mathbf{G}_p/\partial\mathbf{u}$ required in (7) and (8) can be estimated by finite difference approximation as $\nabla_k\Phi_p := \mathcal{Y}_{p,k}^{\Phi}(\mathcal{U}_k)^{-1}$ and $\nabla_k\mathbf{G}_p := \mathcal{Y}_{p,k}^{\mathbf{G}}(\mathcal{U}_k)^{-1}$, respectively. Hence, the gradient modifiers at the current operating point $\mathbf{u}_k$ can be computed from the current and past operating points as follows:

$$(\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}} = \mathcal{Y}_{p,k}^{\Phi}(\mathcal{U}_k)^{-1} - \frac{\partial\Phi}{\partial\mathbf{u}}(\mathbf{u}_k,\boldsymbol{\theta}), \tag{15}$$

$$(\boldsymbol{\lambda}_k^{\mathbf{G}})^{\mathrm{T}} = \mathcal{Y}_{p,k}^{\mathbf{G}}(\mathcal{U}_k)^{-1} - \frac{\partial\mathbf{G}}{\partial\mathbf{u}}(\mathbf{u}_k,\boldsymbol{\theta}). \tag{16}$$

In order to obtain accurate gradient estimates, an additional constraint is added to the modified optimization problem:

$$\mathcal{D}_k(\mathbf{u}) := \mathcal{D}(\mathbf{u},\mathbf{u}_k,\mathbf{u}_{k-1},\ldots,\mathbf{u}_{k-n_u+1}) \le 0 \tag{17}$$

This constraint takes into account the location of the current and past operating points in order to position the next operating point. Let us introduce the matrix

$$\mathsf{U}_k := [\,\mathbf{u}_k-\mathbf{u}_{k-1},\quad \mathbf{u}_k-\mathbf{u}_{k-2},\quad \ldots,\quad \mathbf{u}_k-\mathbf{u}_{k-n_u+1}\,]\in \mathrm{IR}^{n_u\times(n_u-1)}.$$

Assuming that the columns in $\mathsf{U}_k$ are linearly independent, they span a unique hyperplane that contains the $n_u$ most recent points. Let $\mathbf{n}_k$ be a vector that is normal to this hyperplane, thus we have $(\mathsf{U}_k)^{\mathrm{T}}\mathbf{n}_k = \mathbf{0}$, and the hyperplane is $\mathcal{H}_k = \{\mathbf{u}\in\mathrm{IR}^{n_u} : \mathbf{n}_k^{\mathrm{T}}\mathbf{u} = b_k,\ \text{ with } b_k = \mathbf{n}_k^{\mathrm{T}}\mathbf{u}_k\}$. The nature of $\mathcal{D}_k(\mathbf{u})$ will be discussed later. We mention however that the proposed constraints used in dual ISOPE (Brdyś & Tatjewski, 1994, 2005) and in dual modifier adaptation (Gao & Engell, 2005; Marchetti et al., 2010) produce two disjoint feasible regions for locating the next operating point, one region at each side of the hyperplane $\mathcal{H}_k$ and strongly separated from $\mathcal{H}_k$. Notice that, if the next point $\mathbf{u}_{k+1}$ is located on $\mathcal{H}_k$, the matrix $\mathcal{U}_{k+1}$ will become singular.

In dual modifier adaptation a modified optimization problem including the constraint (17) is solved on each side of the hyperplane $\mathcal{H}_k$. The optimization problem corresponding to the half space $\mathbf{n}_k^{\mathrm{T}}\mathbf{u} \ge b_k$ reads:

$$\begin{aligned}\mathbf{u}_{k+1}^{+} = \underset{\mathbf{u}}{argmin}\quad & \Phi_m(\mathbf{u},\boldsymbol{\theta}) = \Phi(\mathbf{u},\boldsymbol{\theta}) + \varepsilon_k^{\Phi} + (\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_k)\\ \text{s.t.}\quad & \mathbf{G}_m(\mathbf{u},\boldsymbol{\theta}) = \mathbf{G}(\mathbf{u},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k^{\mathbf{G}} + \boldsymbol{\lambda}_k^{\mathbf{G}^{\mathrm{T}}}(\mathbf{u}-\mathbf{u}_k) \le \mathbf{0}\\ & \mathcal{D}_k(\mathbf{u}) \le 0,\qquad \mathbf{n}_k^{\mathrm{T}}\mathbf{u} \ge b_k\end{aligned} \tag{18}$$

and for the other half space $\mathbf{n}_k^{\mathrm{T}}\mathbf{u} \le b_k$:

$$\begin{aligned}\mathbf{u}_{k+1}^{-} = \underset{\mathbf{u}}{argmin}\quad & \Phi_m(\mathbf{u},\boldsymbol{\theta}) = \Phi(\mathbf{u},\boldsymbol{\theta}) + \varepsilon_k^{\Phi} + (\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_k)\\ \text{s.t.}\quad & \mathbf{G}_m(\mathbf{u},\boldsymbol{\theta}) = \mathbf{G}(\mathbf{u},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k^{\mathbf{G}} + \boldsymbol{\lambda}_k^{\mathbf{G}^{\mathrm{T}}}(\mathbf{u}-\mathbf{u}_k) \le \mathbf{0}\\ & \mathcal{D}_k(\mathbf{u}) \le 0,\qquad \mathbf{n}_k^{\mathrm{T}}\mathbf{u} \le b_k\end{aligned} \tag{19}$$

The modifiers $\varepsilon_k^{\Phi}$, $\boldsymbol{\varepsilon}_k^{\mathbf{G}}$, and $\boldsymbol{\lambda}_k^{\Phi}$, $\boldsymbol{\lambda}_k^{\mathbf{G}}$ are computed according to (5), (6), and (15), (16), respectively. These modifiers can be smoothed using an exponential filter as in (10). The next operating point $\mathbf{u}_{k+1}$ is chosen in the set $\{\mathbf{u}_{k+1}^{+}, \mathbf{u}_{k+1}^{-}\}$ as the value that minimizes $\Phi_m(\mathbf{u},\boldsymbol{\theta})$.

### 3.4. Optimality loss due to Lagrangian gradient error

Dual modifier adaptation requires estimates of the cost and constraint gradients, $\partial\Phi_p/\partial\mathbf{u}$, $\partial G_{p,1}/\partial\mathbf{u}$, ..., $\partial G_{p,n_g}/\partial\mathbf{u}$. The question arises as to what relative importance should be given to the error with which each of these individual gradients is estimated. For constrained optimization problems, Marchetti et al. (2010), proposed to pay attention to the gradient error of the Lagrangian function, wherein the Lagrange multipliers represent the weights given to the individual constraint gradients with respect to the gradient of the cost function. To limit the error in the estimated cost and constraint gradients is important only as a means to limit the loss in cost with respect to the infimum of the plant. Bearing this in mind, the choice of the Lagrangian function is given a theoretical support in this section by conducting a sensitivity analysis of the optimality loss (loss in optimal value or loss in cost) of the NLP problem (1) induced by small Lagrangian gradient errors.

**Theorem 1** (Optimality Loss). *Let the functions $\Phi_p$ and $G_{p,i}$, $i=1,\ldots,n_g$, be three times continuously differentiable with respect to $\mathbf{u}$. Let $\mathbf{u}^\star(\boldsymbol{\epsilon})$ denote the solution of Problem 1 with Lagrangian gradient error $\boldsymbol{\epsilon}$, and let $\mathbf{u}^\star(\mathbf{0}) = \mathbf{u}_p^\star$ be such that:*

1. *The second-order sufficient conditions for a local minimum of Problem (1) hold at $\mathbf{u}^\star(\mathbf{0})$ with the associated Lagrange multipliers $\boldsymbol{\mu}^\star(\mathbf{0})$;*
2. *The gradients $\partial G_{p,i}/\partial\mathbf{u}(\mathbf{u}_p^\star)$, for $i$ such that $G_{p,i}(\mathbf{u}_p^\star)=0$, are all linearly independent;*
3. *Strict complementary slackness holds, i.e., $\mu_i^\star(\mathbf{0})>0$, for each $i\in\mathcal{A}(\mathbf{u}_p^\star)$.*

*Then, for $\boldsymbol{\epsilon}$ in a neighborhood of $\mathbf{0}$, the loss in optimal value is given by:*

$$\Delta\Phi_p^\star(\boldsymbol{\epsilon}) := \Phi_p(\mathbf{u}^\star(\boldsymbol{\epsilon})) - \Phi_p(\mathbf{u}^\star(\mathbf{0})) = -\frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}}\left[\frac{\partial \mathbf{u}^\star}{\partial \boldsymbol{\epsilon}}(\mathbf{0})\right]\boldsymbol{\epsilon} + o(\|\boldsymbol{\epsilon}\|^2), \tag{20}$$

*with*

$$\begin{bmatrix} \dfrac{\partial \mathbf{u}^\star}{\partial \boldsymbol{\epsilon}}(\mathbf{0}) \\ \dfrac{\partial \boldsymbol{\mu}^\star}{\partial \boldsymbol{\epsilon}}(\mathbf{0}) \end{bmatrix} = -\mathsf{M}^{-1}\mathsf{N}, \tag{21}$$

*where the matrices $\mathsf{M} \in \mathrm{IR}^{(n_g+n_u)\times(n_g+n_u)}$, and $\mathsf{N} \in \mathrm{IR}^{(n_g+n_u)\times n_g}$ are defined as:*

$$\mathsf{M} := \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}_p}{\partial \mathbf{u}^2} & \dfrac{\partial G_{p,1}}{\partial \mathbf{u}}^{\mathrm{T}} & \cdots & \dfrac{\partial G_{p,n_g}}{\partial \mathbf{u}}^{\mathrm{T}} \\ \mu_1^\star \dfrac{\partial G_{p,1}}{\partial \mathbf{u}} & G_{p,1} & & \\ \vdots & & \ddots & \\ \mu_{n_g}^\star \dfrac{\partial G_{p,n_g}}{\partial \mathbf{u}} & & & G_{p,n_g} \end{pmatrix}_{(\mathbf{u}_p^\star)}, \qquad \mathsf{N} := \begin{pmatrix} \mathbf{I}_{n_u} \\ \mathbf{0}_{n_g\times n_u} \end{pmatrix}.$$

**Proof.** Assumption (1) implies the satisfaction of the first-order KKT conditions with the Lagrangian gradient error $\boldsymbol{\epsilon}$,

$$\frac{\partial \hat{\mathcal{L}}_p}{\partial \mathbf{u}}(\mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) := \frac{\partial \mathcal{L}_p}{\partial \mathbf{u}}(\mathbf{u}, \boldsymbol{\mu}) + \boldsymbol{\epsilon}^{\mathrm{T}} = \mathbf{0}, \tag{22}$$

$$\mu_i G_{p,i}(\mathbf{u}) = 0, \qquad i = 1, \ldots, n_g, \tag{23}$$

at $(\mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) = (\mathbf{u}^\star(\mathbf{0}), \boldsymbol{\mu}^\star(\mathbf{0}), \mathbf{0})$. The $\mathcal{C}^3$ assumption implies that the system of Eqs. (22), (23) is twice continuously differentiable in $\mathbf{u}$ and $\boldsymbol{\epsilon}$. From the Basic Sensitivity Theorem (see Theorem 3.2.2 and Corollary 3.2.5 in (Fiacco, 1983)), the assumptions imply that for $\boldsymbol{\epsilon}$ near $\mathbf{0}$ there exist unique twice continuously differentiable functions $\mathbf{u}^\star(\boldsymbol{\epsilon})$ and $\boldsymbol{\mu}^\star(\boldsymbol{\epsilon})$ satisfying (22), (23). Moreover, for $\boldsymbol{\epsilon}$ near $\mathbf{0}$ the set of active constraints does not change, and strict complementary slackness is preserved. Also, from Corollary 3.2.3 in Fiacco (1983), (21) holds with the matrices M and N as defined, where M is the Jacobian of (22), (23) with respect to $(\mathbf{u}, \boldsymbol{\mu})$ at $(\mathbf{u}^\star(\mathbf{0}), \boldsymbol{\mu}^\star(\mathbf{0}))$, and N is the Jacobian of (22), (23) with respect to $\boldsymbol{\epsilon}$.

From condition (23) it follows that in a neighborhood of $\boldsymbol{\epsilon} = \mathbf{0}$,

$$\mathcal{L}_p(\mathbf{u}^\star(\boldsymbol{\epsilon}), \boldsymbol{\mu}^\star(\boldsymbol{\epsilon})) = \Phi_p(\mathbf{u}^\star(\boldsymbol{\epsilon})). \tag{24}$$

Let us denote by $\mathbf{G}_p^a$ the set of active constraints at $\mathbf{u}_p^\star$, and by $\boldsymbol{\mu}^a$ their corresponding Lagrange multipliers. Since the set of active constraints does not change for $\boldsymbol{\epsilon}$ near $\mathbf{0}$, and since the Lagrange multipliers corresponding to the inactive constraints are equal to zero, the inactive constraints can be removed from the analysis, and thus from the Lagrangian function, i.e., we can write $\mathcal{L}_p(\mathbf{z}^\star(\boldsymbol{\epsilon})) = \mathcal{L}_p(\mathbf{u}^\star(\boldsymbol{\epsilon}), \boldsymbol{\mu}^\star(\boldsymbol{\epsilon}))$, where $\mathbf{z}^\star(\boldsymbol{\epsilon}) = \left[\mathbf{u}^\star(\boldsymbol{\epsilon})^{\mathrm{T}}, \boldsymbol{\mu}^{a\star}(\boldsymbol{\epsilon})^{\mathrm{T}}\right]^{\mathrm{T}}$.

A Taylor expansion of $\mathcal{L}_p(\mathbf{z}^\star(\boldsymbol{\epsilon}))$ in the neighborhood of $\boldsymbol{\epsilon} = \mathbf{0}$ gives,

$$\Phi_p(\mathbf{u}^\star(\boldsymbol{\epsilon})) = \Phi_p(\mathbf{u}^\star(\mathbf{0})) + \left[\frac{\partial \mathcal{L}_p}{\partial \boldsymbol{\epsilon}}(\mathbf{z}^\star(\boldsymbol{\epsilon}))\right]_{\boldsymbol{\epsilon}=\mathbf{0}} \boldsymbol{\epsilon} + \frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}}\left[\frac{\partial^2 \mathcal{L}_p}{\partial \boldsymbol{\epsilon}^2}(\mathbf{z}^\star(\boldsymbol{\epsilon}))\right]_{\boldsymbol{\epsilon}=\mathbf{0}} \boldsymbol{\epsilon} + o(\|\boldsymbol{\epsilon}\|^2). \tag{25}$$

Using the chain rule of differentiation,

$$\frac{\partial \mathcal{L}_p}{\partial \boldsymbol{\epsilon}}(\mathbf{z}^\star(\boldsymbol{\epsilon})) = \left[\frac{\partial \mathcal{L}_p}{\partial \mathbf{u}}(\mathbf{z}^\star(\boldsymbol{\epsilon})), \frac{\partial \mathcal{L}_p}{\partial \boldsymbol{\mu}^a}(\mathbf{z}^\star(\boldsymbol{\epsilon}))\right] \frac{\partial \mathbf{z}^\star}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) \tag{26}$$

From (22) we have $\partial \mathcal{L}_p / \partial \mathbf{u}(\mathbf{z}^\star(\boldsymbol{\epsilon})) = -\boldsymbol{\epsilon}^{\mathrm{T}}$, and noticing that $\partial \mathcal{L}_p / \partial \boldsymbol{\mu}^a(\mathbf{z}^\star(\boldsymbol{\epsilon})) = [\mathbf{G}_p^a(\mathbf{u}^\star(\boldsymbol{\epsilon}))]^{\mathrm{T}} = \mathbf{0}$, we get

$$\frac{\partial \mathcal{L}_p}{\partial \boldsymbol{\epsilon}}(\mathbf{z}^\star(\boldsymbol{\epsilon})) = -\left[\boldsymbol{\epsilon}^{\mathrm{T}}, \mathbf{0}\right] \frac{\partial \mathbf{z}^\star}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) = -\boldsymbol{\epsilon}^{\mathrm{T}} \frac{\partial \mathbf{u}^\star}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\epsilon}). \tag{27}$$

Differentiating (27) we obtain

$$\frac{\partial^2 \mathcal{L}_p}{\partial \boldsymbol{\epsilon}^2}(\mathbf{z}^\star(\boldsymbol{\epsilon})) = -\frac{\partial \mathbf{u}^\star}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) - \boldsymbol{\epsilon}^{\mathrm{T}} \frac{\partial^2 \mathbf{u}^\star}{\partial \boldsymbol{\epsilon}^2}(\boldsymbol{\epsilon}). \tag{28}$$

Evaluating (27) and (28) at $\boldsymbol{\epsilon} = \mathbf{0}$ we get

$$\left[\frac{\partial \mathcal{L}_p}{\partial \boldsymbol{\epsilon}}(\mathbf{z}^\star(\boldsymbol{\epsilon}))\right]_{\boldsymbol{\epsilon}=\mathbf{0}} = \mathbf{0}, \qquad \text{and} \qquad \left[\frac{\partial^2 \mathcal{L}_p}{\partial \boldsymbol{\epsilon}^2}(\mathbf{z}^\star(\boldsymbol{\epsilon}))\right]_{\boldsymbol{\epsilon}=\mathbf{0}} = -\frac{\partial \mathbf{u}^\star}{\partial \boldsymbol{\epsilon}}(\mathbf{0}),$$

and upon substitution in (25) we obtain

$$\Delta\Phi_p^\star(\boldsymbol{\epsilon}) = -\frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}}\left[\frac{\partial \mathbf{u}^\star}{\partial \boldsymbol{\epsilon}}(\mathbf{0})\right]\boldsymbol{\epsilon} + o(\|\boldsymbol{\epsilon}\|^2),$$

which completes the proof. □

**Remark 1** *(Unconstrained Problem).* In the case of an unconstrained optimization problem, the Jacobian matrices M and N in Theorem 1 reduce to

$$\mathsf{M} = \frac{\partial^2 \Phi_p}{\partial \mathbf{u}^2}(\mathbf{u}_p^\star), \qquad \mathsf{N} = \mathbf{I}_{n_u},$$

and the optimality loss induced by the cost function gradient error $\boldsymbol{\epsilon}$, for $\boldsymbol{\epsilon}$ near $\mathbf{0}$, reduces to

$$\Delta\Phi_p^\star(\boldsymbol{\epsilon}) = -\frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}}\left[\frac{\partial^2 \Phi_p}{\partial \mathbf{u}^2}(\mathbf{u}_p^\star)\right]^{-1}\boldsymbol{\epsilon} + o(\|\boldsymbol{\epsilon}\|^2).$$

The sensitivity analysis shows that when the available cost and constraint gradients are estimated quantities, the loss in cost induced will be determined by the resulting error in the gradient of the Lagrangian function. In practice, this suggests that in dual modifier-adaptation approaches, the constraint (17), which is used to place the next operating point, should take into account the accuracy with which the Lagrangian gradient is estimated.

**Example 2.** Consider the following optimization problem, which is adapted from Luenberger and Ye (2008):

$$\begin{aligned} \min_{\mathbf{u}} \quad & \Phi_p = -u_1u_2 - u_2u_3 - u_1u_3, \\ \text{s.t.} \quad & \mathbf{G}_p = u_1 + u_2 + u_3 - 3 \le 0, \end{aligned} \tag{29}$$

for which the KKT conditions (22), (23) with Lagrangian gradient error $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \epsilon_3]^{\mathrm{T}}$ read:

$$\begin{aligned} -u_2 - u_3 + \mu + \epsilon_1 &= 0, \\ -u_1 - u_3 + \mu + \epsilon_2 &= 0, \\ -u_1 - u_2 + \mu + \epsilon_3 &= 0, \\ \mu(u_1 + u_2 + u_3 - 3) &= 0, \end{aligned} \tag{30}$$

from where the solution $\mathbf{z}^\star(\mathbf{0}) = [1, 1, 1, 2]^{\mathrm{T}}$ is computed.

In this example, the loss in cost $\Delta\Phi_p^\star(\boldsymbol{\epsilon})$ can be obtained analytically. Knowing that the constraint is active at the optimum, the solution to (30) reads:

$$\mathbf{z}^\star(\boldsymbol{\epsilon}) = \mathsf{A}^{-1}\begin{bmatrix} -\boldsymbol{\epsilon} \\ 3 \end{bmatrix} = -\mathsf{B}\boldsymbol{\epsilon} + \mathbf{a}_4^{-1}3, \quad \text{with} \quad \mathsf{B} = [\mathbf{a}_1^{-1} \quad \mathbf{a}_2^{-1} \quad \mathbf{a}_3^{-1}], \tag{31}$$

where $\mathbf{a}_j^{-1}$ are the columns of $\mathsf{A}^{-1}$, with $\mathsf{A}$ defined below:

$$\mathsf{A} = \begin{pmatrix} 0 & -1 & -1 & 1 \\ -1 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}; \quad \mathsf{P} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$\mathsf{M} = \begin{pmatrix} 0 & -1 & -1 & 1 \\ -1 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 \\ 2 & 2 & 2 & 0 \end{pmatrix}. \tag{32}$$

The cost can be written as:

$$\Phi_p = -[u_2 \;\; u_3 \;\; u_1 \;\; 0][u_1 \;\; u_2 \;\; u_3 \;\; \mu]^{\mathrm{T}} = -\mathbf{z}^{\mathrm{T}}\mathsf{P}\mathbf{z}, \tag{33}$$

with matrix $\mathsf{P}$ defined in (32). Replacing (31) in (33), and noticing that $(\mathbf{a}_4^{-1}3)^{\mathrm{T}}\mathsf{PB} = \mathbf{0}$, the optimal cost becomes:

$$\Phi_p(\mathbf{u}^\star(\boldsymbol{\epsilon})) = -\mathbf{z}^\star(\boldsymbol{\epsilon})^{\mathrm{T}}\mathsf{P}\mathbf{z}^\star(\boldsymbol{\epsilon}) = -\boldsymbol{\epsilon}^{\mathrm{T}}\mathsf{B}^{\mathrm{T}}\mathsf{PB}\boldsymbol{\epsilon} - 3. \tag{34}$$

Hence, the optimality loss is given by:

$$\Delta\Phi_p^\star(\boldsymbol{\epsilon}) = -\boldsymbol{\epsilon}^{\mathrm{T}}\mathsf{B}^{\mathrm{T}}\mathsf{PB}\boldsymbol{\epsilon}. \tag{35}$$

It is easy to verify that in this example $\Delta\Phi_p^\star(\boldsymbol{\epsilon}) = -\boldsymbol{\epsilon}^{\mathrm{T}}\mathsf{B}^{\mathrm{T}}\mathsf{PB}\boldsymbol{\epsilon} = -1/2\boldsymbol{\epsilon}^{\mathrm{T}}[\partial\mathbf{u}^\star/\partial\boldsymbol{\epsilon}(\mathbf{0})]\boldsymbol{\epsilon}$. Compute $\partial\mathbf{u}^\star/\partial\boldsymbol{\epsilon}(\mathbf{0})$ from (21) with $\mathsf{M}$ given in (32). Next, define $\mathsf{Q}_1 = \mathsf{B}^{\mathrm{T}}\mathsf{PB}$ and $\mathsf{Q}_2 = 1/2\partial\mathbf{u}^\star/\partial\boldsymbol{\epsilon}(\mathbf{0})$, and verify that the corresponding symmetric matrices match, i.e., $1/2(\mathsf{Q}_1 + \mathsf{Q}_1^{\mathrm{T}}) = 1/2(\mathsf{Q}_2 + \mathsf{Q}_2^{\mathrm{T}})$.

### 3.5. Upper bound on gradient error norm

For the purpose of optimization, the constraint $\mathcal{D}_k(\mathbf{u})$ should be selected so as to obtain accurate Lagrangian gradient estimates. An upper bound on the gradient error norm was proposed in Marchetti et al. (2010). A measured value of the Lagrangian function can be expressed as:

$$\begin{aligned} \tilde{L}_p(\mathbf{u}) &= \phi(\mathbf{u}, \mathbf{y}_p(\mathbf{u}) + \boldsymbol{\nu}) + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{g}(\mathbf{u}, \mathbf{y}_p(\mathbf{u}) + \boldsymbol{\nu}) \\ &= \Phi_p(\mathbf{u}) + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{G}_p(\mathbf{u}) + \nu = L_p(\mathbf{u}) + \nu, \end{aligned} \tag{36}$$

where $\boldsymbol{\nu}$ is the output measurement noise vector, and $\nu$ represents the resulting noise in the Lagrangian function. The values of the Lagrange multipliers $\boldsymbol{\mu}$ at the plant optimum are unknown. However, approximate values can be obtained from the solution of the modified optimization problem. For simplicity, we shall assume that the selected values of $\boldsymbol{\mu}$ are kept constant, although it is also possible to update them during the iterative process. The Lagrangian gradient for the plant, $\partial L_p/\partial\mathbf{u}(\mathbf{u})$, can be estimated from the $n_u$ most recent operating points as:

$$\nabla L_p(\mathbf{u}) = \mathcal{Y}_p^L(\mathbf{u}) \;\; \mathcal{U}^{-1}(\mathbf{u}), \tag{37}$$

with:

$$\mathcal{U}(\mathbf{u}) = [\,\mathbf{u} - \mathbf{u}_k, \quad \ldots, \quad \mathbf{u} - \mathbf{u}_{k-n_u+1}\,] \in \mathrm{IR}^{n_u \times n_u}, \tag{38}$$

$$\mathcal{Y}_p^L(\mathbf{u}) = [\,\tilde{L}_p(\mathbf{u}) - \tilde{L}_p(\mathbf{u}_k), \quad \ldots, \quad \tilde{L}_p(\mathbf{u}) - \tilde{L}_p(\mathbf{u}_{k-n_u+1})\,] \in \mathrm{IR}^{1 \times n_u}, \tag{39}$$

The gradient estimation error is defined as

$$\boldsymbol{\epsilon}(\mathbf{u})^{\mathrm{T}} := \nabla L_p(\mathbf{u}) - \frac{\partial L_p}{\partial \mathbf{u}}(\mathbf{u}),$$

which, from (37) and using $\tilde{L}_p(\mathbf{u}_{k-j}) = L_p(\mathbf{u}_{k-j}) + \nu_{k-j}$ and $\tilde{L}_p(\mathbf{u}) = L_p(\mathbf{u}) + \nu$, can be split as

$$\boldsymbol{\epsilon}(\mathbf{u}) = \boldsymbol{\epsilon}^t(\mathbf{u}) + \boldsymbol{\epsilon}^n(\mathbf{u}), \tag{40}$$

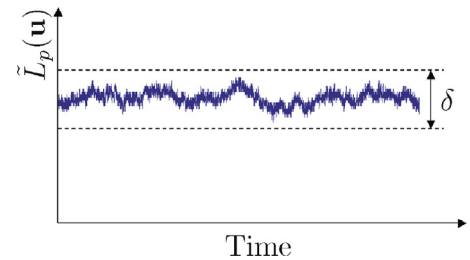


**Fig. 1.** Measured quantity at steady state with indication of the noise level $\delta$.

where $\boldsymbol{\epsilon}^t$ and $\boldsymbol{\epsilon}^n$ represent the errors due to finite-difference approximation (or truncation) and measurement noise, respectively,

$$\boldsymbol{\epsilon}^t(\mathbf{u})^{\mathrm{T}} = [L_p(\mathbf{u}) - L_p(\mathbf{u}_k), \quad \ldots \quad L_p(\mathbf{u}) - L_p(\mathbf{u}_{k-n_u+1})] \;\; \mathcal{U}^{-1}(\mathbf{u}) - \frac{\partial L_p}{\partial \mathbf{u}}(\mathbf{u}) \tag{41}$$

$$\boldsymbol{\epsilon}^n(\mathbf{u})^{\mathrm{T}} = [\nu - \nu_k, \quad \ldots \quad \nu - \nu_{k-n_u+1}]\mathcal{U}^{-1}(\mathbf{u}). \tag{42}$$

Assuming $L_p(\mathbf{u})$ is twice continuously differentiable with respect to $\mathbf{u}$, then the norm of the gradient error due to truncation can be upper bounded as follows (Marchetti et al., 2010):

$$\begin{aligned} \|\boldsymbol{\epsilon}^t(\mathbf{u})\| \le \mathcal{E}^t(\mathbf{u}) := \frac{\sigma_{max}}{2}\|[\;\; (\mathbf{u} - \mathbf{u}_k)^{\mathrm{T}}(\mathbf{u} - \mathbf{u}_k) \;\; \ldots \\ \ldots \;\; (\mathbf{u} - \mathbf{u}_{k-n_u+1})^{\mathrm{T}}(\mathbf{u} - \mathbf{u}_{k-n_u+1}) \;\; ] \;\; \mathcal{U}^{-1}(\mathbf{u})\|, \end{aligned} \tag{43}$$

where $\sigma_{max}$ is an upper bound on the spectral radius of the Hessian of $L_p(\cdot)$. Also, assuming that the noisy output $\tilde{L}_p(\mathbf{u})$ remains within an interval $\delta$ at steady-state operation, as illustrated in Fig. 1, then the norm of the gradient error due to measurement noise can be upper bounded as follows (Marchetti et al., 2010):

$$\|\boldsymbol{\epsilon}^n(\mathbf{u})\| \le \mathcal{E}^n(\mathbf{u}) := \frac{\delta}{l_{\min}(\mathbf{u})}, \tag{44}$$

where $l_{\min}(\mathbf{u})$ is the shortest distance between all possible pairs of complement affine subspaces that can be generated from the set of points $\mathcal{S} = \{\mathbf{u}, \mathbf{u}_k, \ldots, \mathbf{u}_{k-n_u+1}\}$ (see Marchetti et al. (2010) for the computation of $l_{\min}(\mathbf{u})$). In practice, it is possible to ensure a given accuracy of the estimated gradient, for example $\|\boldsymbol{\epsilon}(\mathbf{u})\| \le \mathcal{E}^U$, by selecting $\mathbf{u}$ such that $\mathcal{E}^t(\mathbf{u}) + \mathcal{E}^n(\mathbf{u}) \le \mathcal{E}^U$,

$$\|\boldsymbol{\epsilon}(\mathbf{u})\| \le \mathcal{E}(\mathbf{u}) := \mathcal{E}^t(\mathbf{u}) + \mathcal{E}^n(\mathbf{u}) \le \mathcal{E}^U, \tag{45}$$

In dual modifier adaptation, this can be achieved by choosing $\mathcal{D}_k(\mathbf{u}) = \mathcal{E}(\mathbf{u}) - \mathcal{E}^U$.

## 4. A new dual modifier-adaptation strategy

### 4.1. Alternative gradient modifiers

In addition to the matrices (11)–(13), the following matrices can be constructed at the $k$th RTO iteration, based on the model predictions:

$$\mathcal{Y}_k^\Phi := [\,\Phi(\mathbf{u}_k, \boldsymbol{\theta}) - \Phi(\mathbf{u}_{k-1}, \boldsymbol{\theta}), \quad \ldots, \quad \Phi(\mathbf{u}_k, \boldsymbol{\theta}) - \Phi(\mathbf{u}_{k-n_u}, \boldsymbol{\theta})\,] \in \mathrm{IR}^{1 \times n_u}, \tag{46}$$

$$\mathcal{Y}_k^{\mathbf{G}} := [\,\mathbf{G}(\mathbf{u}_k, \boldsymbol{\theta}) - \mathbf{G}(\mathbf{u}_{k-1}, \boldsymbol{\theta}), \quad \ldots, \quad \mathbf{G}(\mathbf{u}_k, \boldsymbol{\theta}) - \mathbf{G}(\mathbf{u}_{k-n_u}, \boldsymbol{\theta})\,] \in \mathrm{IR}^{ng \times n_u}, \tag{47}$$

Instead of the modifiers given in (15) and (16), we propose the following gradient modifiers for use in dual modifier-adaptation:

$$(\boldsymbol{\lambda}_k^\Phi)^{\mathrm{T}} = \left(\mathcal{Y}_{p,k}^\Phi - \mathcal{Y}_k^\Phi\right)(\mathcal{U}_k)^{-1}, \tag{48}$$

$$(\boldsymbol{\lambda}_k^{\mathbf{G}})^{\mathrm{T}} = \left(\mathcal{Y}_{p,k}^{\mathbf{G}} - \mathcal{Y}_k^{\mathbf{G}}\right)(\mathcal{U}_k)^{-1}. \tag{49}$$

**Proposition 1.** *Using the gradient modifiers (48) and (49) the modified cost and constraint functions $\Phi_m(\mathbf{u},\boldsymbol{\theta})$ and $G_{m,i}(\mathbf{u},\boldsymbol{\theta})$, $i=1\ldots n_g$, match the corresponding measured values for the plant at the current and past operating points $\mathbf{u}_k, \mathbf{u}_{k-1}, \ldots, \mathbf{u}_{k-n_u}$.*

**Proof.** Consider the modified cost function:

$$\Phi_m(\mathbf{u},\boldsymbol{\theta}) = \Phi(\mathbf{u},\boldsymbol{\theta}) + \varepsilon_k^{\Phi} + (\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_k). \tag{50}$$

From (5), (12), (46), and (48) we have

$$(\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}} = \left(\mathcal{Y}_{p,k}^{\Phi} - \mathcal{Y}_k^{\Phi}\right)(\mathcal{U}_k)^{-1} = [\,\varepsilon_k^{\Phi} - \varepsilon_{k-1}^{\Phi}, \quad \ldots, \quad \varepsilon_k^{\Phi} - \varepsilon_{k-n_u}^{\Phi}\,](\mathcal{U}_k)^{-1} \tag{51}$$

From (50) and (5) we have

$$\Phi_m(\mathbf{u}_k,\boldsymbol{\theta}) = \Phi(\mathbf{u}_k,\boldsymbol{\theta}) + \varepsilon_k^{\Phi} = \Phi_p(\mathbf{u}_k).$$

Noticing that $(\mathcal{U}_k)^{-1}(\mathbf{u}_k - \mathbf{u}_{k-j}) = \mathbf{e}_j$ for $j=1,\ldots,n_u$, where $\mathbf{e}_j$ is the $j$th unit vector, it follows from (51) that

$$(\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}(\mathbf{u}_{k-j} - \mathbf{u}_k) = \varepsilon_{k-j}^{\Phi} - \varepsilon_k^{\Phi}, \qquad j=1,\ldots,n_u. \tag{52}$$

Using (52) in (50), we have

$$\Phi_m(\mathbf{u}_{k-j},\boldsymbol{\theta}) = \Phi(\mathbf{u}_{k-j},\boldsymbol{\theta}) + \varepsilon_{k-j}^{\Phi} = \Phi_p(\mathbf{u}_{k-j}), \qquad j=1,\ldots,n_u,$$

which completes the proof for the case of the modified cost function. Since the constraints are modified in the same way as the cost function, an equivalent demonstration can be done for each individual modified constraint function. □

With the gradient modifiers (15) and (16) the modified cost and constraint functions are such that their values match the corresponding measured values for the plant at the current operating point $\mathbf{u}_k$, and their gradients match the corresponding estimated gradients $\mathcal{Y}_{p,k}^{\Phi}(\mathcal{U}_k)^{-1}$ and $\mathcal{Y}_{p,k}^{\mathbf{G}}(\mathcal{U}_k)^{-1}$ at $\mathbf{u}_k$. In contrast, with the gradient modifiers introduced in (48) and (49) the modified cost and constraint functions match the corresponding measured values for the plant at the current and past operating points $\mathbf{u}_k, \mathbf{u}_{k-1}, \ldots, \mathbf{u}_{k-n_u}$. We argue that this gives a better approximation of the cost and constraint functions for the plant, in particular for increased distances between the points. The following remark emphasizes the point that, with the new gradient modifiers, the modified cost and constraint functions should no longer be viewed as first-order corrections taken exclusively at the current operating point $\mathbf{u}_k$.

**Remark 2.** Notice that (52) can be rewritten as

$$\varepsilon_{k-j}^{\Phi} + (\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}\mathbf{u}_{k-j} = \varepsilon_k^{\Phi} + (\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}\mathbf{u}_k, \qquad j=1,\ldots,n_u,$$

which implies that

$$\Phi_m(\mathbf{u},\boldsymbol{\theta}) = \Phi(\mathbf{u},\boldsymbol{\theta}) + \varepsilon_{k-j}^{\Phi} + (\boldsymbol{\lambda}_k^{\Phi})^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_{k-j}), \qquad j=0,1,\ldots,n_u, \tag{53}$$

i.e., the modified cost function $\Phi_m(\mathbf{u},\boldsymbol{\theta})$ does not depend on which point in the set $\{\mathbf{u}_k, \mathbf{u}_{k-1}, \ldots, \mathbf{u}_{k-n_u}\}$ is used as the reference point in the first-order correction. The same applies to each modified constraint function.

The approximation given by both correction strategies is illustrated in Fig. 2 for the case of zero measurement noise. The top plot of Fig. 2 uses the old gradient modifier (15). It can be seen that the gradient (slope) of the modified cost function at $u_k$ matches the gradient (slope) of the secant line determined by the points $u_k$ and $u_{k-1}$. The bottom plot of Fig. 2 uses the new gradient modifier (48). In this case, the modified cost function matches the plant cost at the points $u_k$ and $u_{k-1}$. This latter strategy clearly gives a better approximation of the plant cost function.

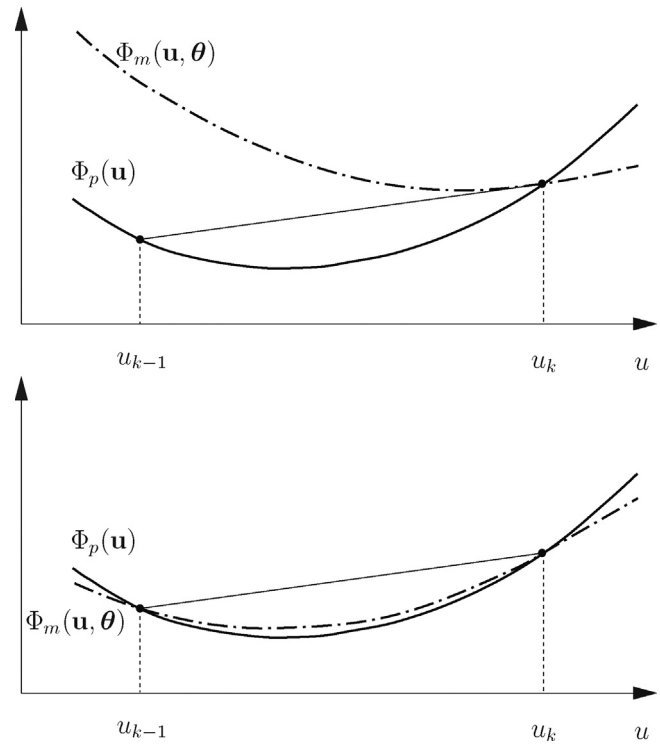


**Fig. 2.** Approximation of the plant cost function using modifier adaptation. **Top plot:** Previous gradient modifier. **Bottom plot:** New gradient modifier.

### 4.2. An auxiliary quadratic function

In this section, an auxiliary quadratic function is introduced that will be used to obtain conservative estimates of the gradient error due to truncation. The auxiliary quadratic function reads:

$$f(\mathbf{u},\mathbf{v}) := \frac{1}{2}(\mathbf{u}-\mathbf{v})^{\mathrm{T}}\mathsf{D}(\mathbf{u}-\mathbf{v}) + L_p(\mathbf{u}) + \boldsymbol{\lambda}(\mathbf{u})^{\mathrm{T}}(\mathbf{u}-\mathbf{v}), \tag{54}$$

with $\mathsf{D} = \sigma_{max}\mathbf{I}_{n_u}$, and

$$\boldsymbol{\lambda}(\mathbf{u})^{\mathrm{T}} = \left(\overline{\mathcal{Y}}_p^{L}(\mathbf{u}) - \mathcal{Y}^{f}(\mathbf{u})\right)\mathcal{U}^{-1}(\mathbf{u}), \tag{55}$$

where $\mathcal{U}(\mathbf{u})$ is given in (38), and $\overline{\mathcal{Y}}_p^{L}(\mathbf{u})$ and $\mathcal{Y}^{f}(\mathbf{u})$ are defined as follows:

$$\overline{\mathcal{Y}}_p^{L}(\mathbf{u}) := [\,L_p(\mathbf{u}) - L_p(\mathbf{u}_k), \quad \ldots, \quad L_p(\mathbf{u}) - L_p(\mathbf{u}_{k-n_u+1})\,] \in \mathbb{R}^{1\times n_u}, \tag{56}$$

$$\mathcal{Y}^{f}(\mathbf{u}) := -\frac{1}{2}[\,(\mathbf{u}-\mathbf{u}_k)^{\mathrm{T}}\mathsf{D}(\mathbf{u}-\mathbf{u}_k), \quad \ldots, \quad (\mathbf{u}-\mathbf{u}_{k-n_u+1})^{\mathrm{T}}\mathsf{D}(\mathbf{u}-\mathbf{u}_{k-n_u+1})\,]. \tag{57}$$

It is easy to show that $f(\mathbf{u},\mathbf{u}) = L_p(\mathbf{u})$, and $f(\mathbf{u},\mathbf{u}_{k-j}) = L_p(\mathbf{u}_{k-j})$, for $j=0,\ldots,n_u-1$. Therefore, $f(\mathbf{u},\cdot)$ is the quadratic function with Hessian matrix $\partial^2 f/\partial\mathbf{v}^2 = \mathsf{D}$ that matches the function $L_p(\cdot)$ at the $(n_u+1)$ points $\mathbf{u}, \mathbf{u}_k, \ldots, \mathbf{u}_{k-n_u+1}$. Notice that, since $\mathsf{D} = \sigma_{max}\mathbf{I}_{n_u}$, the auxiliary function $f(\mathbf{u},\mathbf{v})$ has a conservative curvature with respect to the function $L_p(\cdot)$.

For a given point $\mathbf{u}$, the gradient error due to truncation of the function $f(\mathbf{u},\mathbf{v})$ evaluated at point $\mathbf{v}$, is given by:

$$\begin{aligned}\boldsymbol{\epsilon}^{f}(\mathbf{u},\mathbf{v})^{\mathrm{T}} &= \overline{\mathcal{Y}}_p^{L}(\mathbf{u})\ \mathcal{U}^{-1}(\mathbf{u}) - \frac{\partial f}{\partial \mathbf{v}}(\mathbf{u},\mathbf{v}) \\ &= \overline{\mathcal{Y}}_p^{L}(\mathbf{u})\ \mathcal{U}^{-1}(\mathbf{u}) + (\mathbf{u}-\mathbf{v})^{\mathrm{T}}\mathsf{D} - \boldsymbol{\lambda}(\mathbf{u})^{\mathrm{T}} = \mathcal{Y}^{f}(\mathbf{u})\ \mathcal{U}^{-1}(\mathbf{u}) + (\mathbf{u}-\mathbf{v})^{\mathrm{T}}\mathsf{D},\end{aligned}$$

which can be rewritten as:

$$\begin{aligned}\boldsymbol{\epsilon}^{f}(\mathbf{u},\mathbf{v})^{\mathrm{T}} = &-\frac{\sigma_{max}}{2}[\ (\mathbf{u}-\mathbf{u}_k)^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_k)\ \ \ldots \\ &\ldots\ (\mathbf{u}-\mathbf{u}_{k-n_u+1})^{\mathrm{T}}(\mathbf{u}-\mathbf{u}_{k-n_u+1})\ ]\ \mathcal{U}^{-1}(\mathbf{u}) + \sigma_{max}(\mathbf{u}-\mathbf{v})^{\mathrm{T}}.\end{aligned} \tag{58}$$

Notice that, when evaluated at $\mathbf{v}=\mathbf{u}$, we obtain

$$\|\boldsymbol{\epsilon}^f(\mathbf{u},\mathbf{u})\| = \mathcal{E}^t(\mathbf{u}).$$

It was proved in Marchetti et al. (2010) that $\mathcal{E}^t(\mathbf{u})$ is an upper bound on the gradient error due to truncation for the true process, that is:

$$\|\boldsymbol{\epsilon}^t(\mathbf{u})\| \le \|\boldsymbol{\epsilon}^f(\mathbf{u},\mathbf{u})\| \tag{59}$$

It can also be shown that

$$\|\boldsymbol{\epsilon}^t(\mathbf{u}_{k-j})\| \le \|\boldsymbol{\epsilon}^f(\mathbf{u},\mathbf{u}_{k-j})\|, \quad j=0,\ldots,n_u-1. \tag{60}$$

As a justification, notice that for a given $\mathbf{u}$ the function $f(\mathbf{u},\cdot)$ is not modified by any rearrangement of the points $\mathbf{u},\mathbf{u}_k,\ldots,\mathbf{u}_{k-n_u+1}$ used in its definition. Hence, (60) follows by interchanging $\mathbf{u}$ and $\mathbf{u}_{k-j}$ in (58).

Eqs. (59) and (60) and indicate that at the points $\mathbf{u},\mathbf{u}_k,\ldots,\mathbf{u}_{k-n_u+1}$ the truncation gradient error evaluated with the auxiliary quadratic function is conservative with respect to the truncation error of the function $L_p(\cdot)$.

### 4.3. A new gradient error constraint

In the dual modifier-adaptation approach proposed in Marchetti et al. (2010) the modified cost and constraint functions provide a first-order correction of the corresponding cost and constraint functions for the plant at the current operating point $\mathbf{u}_k$. Hence, the constraint $\mathcal{E}(\mathbf{u}) \le \mathcal{E}^U$, used in Marchetti et al. (2010), pays attention to the accuracy of the gradient estimate obtained for the Lagrangian function $L_p(\mathbf{u})$ at $\mathbf{u}_k$. In contrast, the first-order corrections proposed in the present work are such that the cost and constraint functions match their corresponding measured values at the current and past operating points $\{\mathbf{u}_k,\mathbf{u}_{k-1},\ldots,\mathbf{u}_{k-n_u}\}$.Thus, the modified cost and constraint functions approximate the plant in a larger region of the input space. Hence, the interest in finding a constraint that pays attention to the accuracy of the Lagrangian gradient estimate in this larger region. In this paper, it is proposed to consider the approximation obtained in the polyhedral set that has $\{\mathbf{u}_k,\mathbf{u}_{k-1},\ldots,\mathbf{u}_{k-n_u}\}$ as its extreme points. With this purpose, let us introduce the gradient error constraint

$$\mathcal{E}^f(\mathbf{u}) := \mathcal{E}^{tf}(\mathbf{u}) + \mathcal{E}^n(\mathbf{u}) \le \mathcal{E}^U, \tag{61}$$

where the measurement noise component $\mathcal{E}^n(\mathbf{u})$ is the same as that previously defined in (44), whereas the truncation component is given by

$$\mathcal{E}^{tf}(\mathbf{u}) := \|\boldsymbol{\epsilon}^f(\mathbf{u},\bar{\mathbf{v}}_f^\star(\mathbf{u}))\|,$$

with $\bar{\mathbf{v}}_f^\star(\mathbf{u}) = [\mathbf{u},\mathbf{u}_k,\ldots,\mathbf{u}_{k-n_u+1}]\mathbf{d}_f^\star(\mathbf{u})$, where $\mathbf{d}_f^\star(\mathbf{u}) \in \mathrm{IR}^{n_u}$ is a minimizing solution to the following quadratic program:

$$\begin{aligned} \mathbf{d}_f^\star(\mathbf{u}) = \arg\min_{\mathbf{d}} \quad & [\boldsymbol{\epsilon}^f(\mathbf{u},\bar{\mathbf{v}})]^{\mathrm{T}}\boldsymbol{\epsilon}^f(\mathbf{u},\bar{\mathbf{v}}) \\ \text{s.t.} \quad & \bar{\mathbf{v}} = [\mathbf{u},\mathbf{u}_k,\ldots,\mathbf{u}_{k-n_u+1}]\mathbf{d}, \\ & \mathbf{d} \ge \mathbf{0}, \\ & \sum_{i=1}^{n_u+1} d_i = 1. \end{aligned} \tag{62}$$

The constraints in problem (62) ensure that $\bar{\mathbf{v}}$ belongs to the polyhedral set with extreme points $\{\mathbf{u},\mathbf{u}_k,\ldots,\mathbf{u}_{k-n_u+1}\}$. Therefore, $\bar{\mathbf{v}}_f^\star(\mathbf{u})$ is the point in the polyhedral set that minimizes the gradient error norm of the auxiliary function $f(\mathbf{u},\cdot)$. Problem (62) is given using a formulation that is aimed to facilitate its interpretation and can be transformed into a more standard form of a quadratic programming problem.

The constraint (61) can be used in the dual modifier-adaptation problems (18) and (19) by selecting $\mathcal{D}_k(\mathbf{u}) = \mathcal{E}^f(\mathbf{u}) - \mathcal{E}^U$. The following example shows that this constraint gives larger feasible regions than the constraint $\mathcal{E}(\mathbf{u}) \le \mathcal{E}^U$ for the same value of $\mathcal{E}^U$.

**Example 3.** In the two-input case ($n_u=2$), the contour lines produced by both constraints are depicted in Fig. 3 for two cases of most recent operating points. The top plots use $\mathbf{u}_k=[0\ 0.5]^{\mathrm{T}}$ and $\mathbf{u}_{k-1}=[0\ -0.5]^{\mathrm{T}}$, while the bottom plots use $\mathbf{u}_k=[0\ 0.1]^{\mathrm{T}}$ and $\mathbf{u}_{k-1}=[0\ -0.1]^{\mathrm{T}}$, which are closer to each other. In both cases, the hyperplane $\mathcal{H}_k$ is given by the line $u_1=0$. For a given value of $\mathcal{E}^U$, say $\mathcal{E}^U=2$, the constraints $\mathcal{E}(\mathbf{u}) \le \mathcal{E}^U$ and $\mathcal{E}^f(\mathbf{u}) \le \mathcal{E}^U$ both produce two feasible sets, one at each side of $\mathcal{H}_k$. However, the feasible regions obtained with the latter constraint are open and much larger than those obtained with the former constraint, which are closed.

## 5. Application to the Williams–Otto reactor

In this section, the dual modifier-adaptation approach is applied in simulation to the reactor of the Williams–Otto plant (Williams & Otto, 1960). Structural plant-model mismatch is introduced by considering an inaccurate two reaction approximation in the process model, as proposed by Forbes, Marlin, and MacGregor (1994). This reactor example has been used in the RTO literature to illustrate the concept of model adequacy and to study RTO performance (Forbes et al., 1994; Zhang & Forbes, 2000). In particular, the reactor was used to illustrate the previous dual modifier-adaptation approach in Marchetti et al. (2010), for the case of an unconstrained optimum. Here, the proposed dual modifier-adaptation approach will be compared with the previous approach using two different optimization problem formulations for the Williams–Otto reactor, the first one concerning an unconstrained optimum, and the second one concerning a constrained optimum.

### 5.1. Problem formulation

The reactor consists of an ideal CSTR in which the following reactions take place:

$$\begin{aligned} &A + B \xrightarrow{k_1} C && k_1 = 1.660\times10^6 e^{-6666.7/(T_R+273.15)} \\ &C + B \xrightarrow{k_2} P + E && k_2 = 7.212\times10^8 e^{-8333.3/(T_R+273.15)} \\ &C + P \xrightarrow{k_3} G && k_3 = 2.675\times10^{12} e^{-11111/(T_R+273.15)}. \end{aligned}$$

The two reactants, $A$ and $B$, are fed separately with the mass flowrates $F_A$ and $F_B$, respectively. The desired products are $P$ and $E$, $C$ is an intermediate product and $G$ is an undesired product. The outlet stream has the mass flowrate $F=F_A+F_B$. Operation is isothermal at the temperature $T_R$. The reactor mass holdup is 2105 kg, and the flowrate of reactant $A$ is fixed at 1.8275 kg/s.

The objective function is to maximize profit, which is expressed as the difference in price between the products and the reactants:

$$J(\mathbf{u},\mathbf{y}) = 1143.38X_PF + 25.92X_EF - 76.23F_A - 114.34F_B,$$

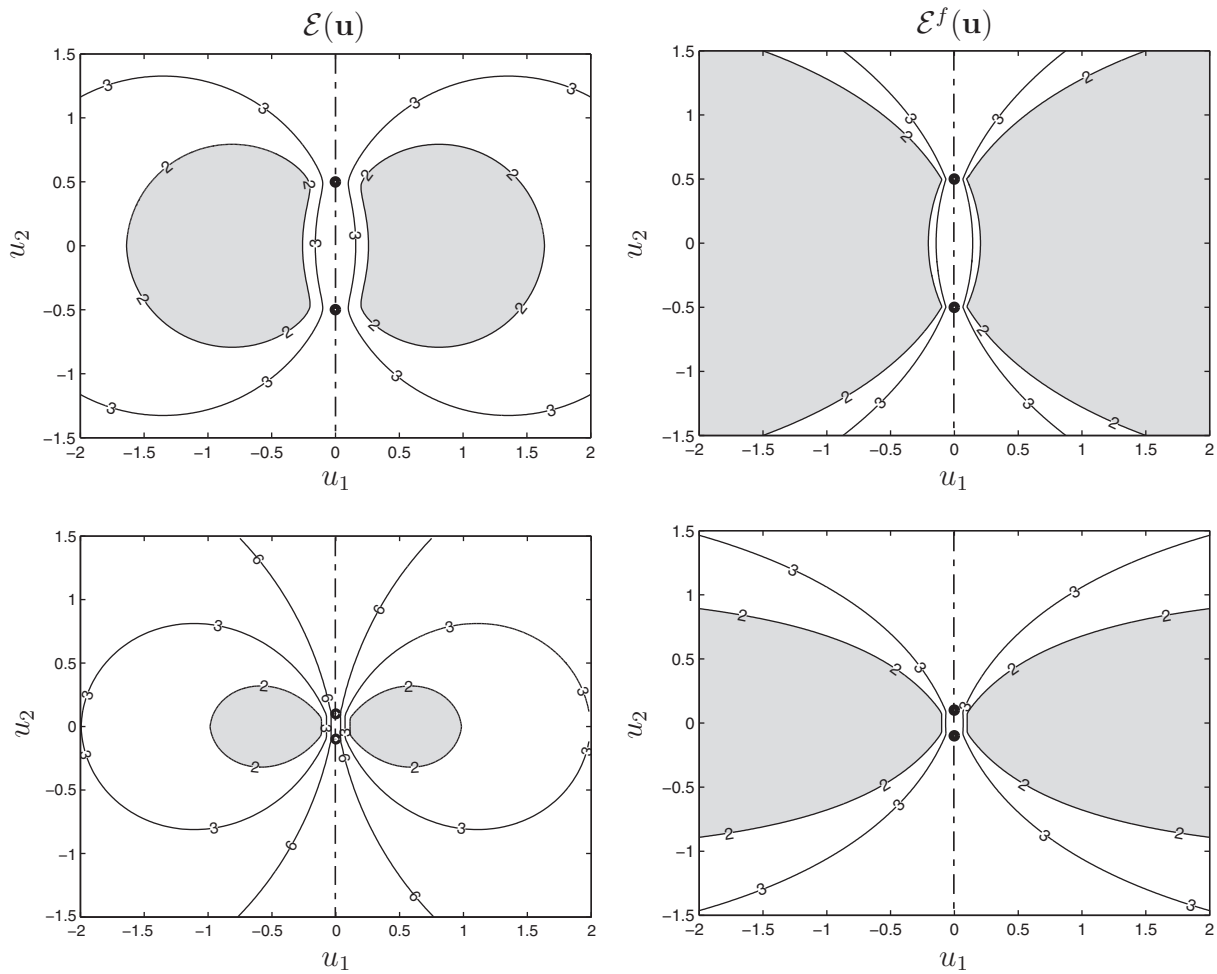


**Fig. 3.** Contours of $\mathcal{E}(\mathbf{u})$ and $\mathcal{E}^f(\mathbf{u})$ obtained with $\delta = 0.2$ and $\sigma_{max} = 2$ for two cases of most recent points (more distant on the top, and closer on the bottom). Shaded areas: Feasible regions of the constraint $\mathcal{E}(\mathbf{u}) \leq 2$ on the left, and $\mathcal{E}^f(\mathbf{u}) \leq 2$ on the right. Dash-dotted line: Hyperplane $\mathcal{H}_k$ determined by the two most recent points.

where $X_i$ represents the concentration of species $i$. As mentioned earlier, a two reaction approximation is considered in the reactor model (Forbes et al., 1994):

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

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

The material balance equations for the plant and the approximate model can be found in Zhang and Forbes (2000). The optimization problem reads:

$$\begin{aligned} \min_{F_B, T_R} \quad & \phi(\mathbf{u}, \mathbf{y}) = -J(\mathbf{u}, \mathbf{y}) \\ \text{s.t.} \quad & F_B \in [3, \quad 6]; \qquad T_R \in [70, \quad 100]; \\ & X_G \leq 0.08, \end{aligned} \tag{63}$$

where the decision variables are the flowrate of reactant $B$ and the reactor temperature, i.e., $\mathbf{u} = [F_B, T_R]^T$. Two optimization problem formulations will be considered depending on whether or not the upper bound on the concentration of the undesired product $G$ is included.

### 5.2. Scenario 1

In this scenario the inequality constraint $X_G \leq 0.08$ is removed from the optimization problem (63). Since no input bounds are active at the solution point, the solution is unconstrained, and the Lagrangian function reduces to the cost function. Hence, the noisy function to be considered in (36) reduces to:

$$\tilde{L}_p(\mathbf{u}) = \tilde{\Phi}_p(\mathbf{u}) = \phi(\mathbf{u}, \mathbf{y}_p(\mathbf{u}) + \boldsymbol{\nu}) = \Phi_p(\mathbf{u}) + \nu.$$

The inputs are scaled using the intervals [3, 6] for $F_B$, and [70, 100] for $T_R$. In this range, the largest value of the spectral radius of the Hessian of cost function predicted by the model $\Phi(\mathbf{u}, \boldsymbol{\theta})$, obtained with the scaled inputs, is 1030, while that of the (unknown) plant cost function $\Phi_p(\mathbf{u})$ is 1221. The simulations are carried out assuming that the noise $\nu$ has a Gaussian distribution with standard deviation $\sigma_\phi = 0.5$. In order to implement dual modifier adaptation, the noise interval $\delta = 6\sigma_\phi = 3$ is chosen, and we select $\sigma_{max} = 1030$. The exponential filter (10) is implemented for the cost gradient modifiers with $\mathrm{K} = \mathrm{diag}\,(1, \quad 0.6, \quad 0.6)$. Simulations are run with 200 RTO iterations, starting from $\mathbf{u}_0 = [3, 70]^T$. A realization of the input trajectories obtained by both, the previous and the new dual modifier-adaptation approaches, is illustrated in Fig. 4 for two different values of the upper bound $\mathcal{E}^U$. In each case, an initial gradient estimate is obtained by perturbing each input separately in the neighborhood of $\mathbf{u}_0$.

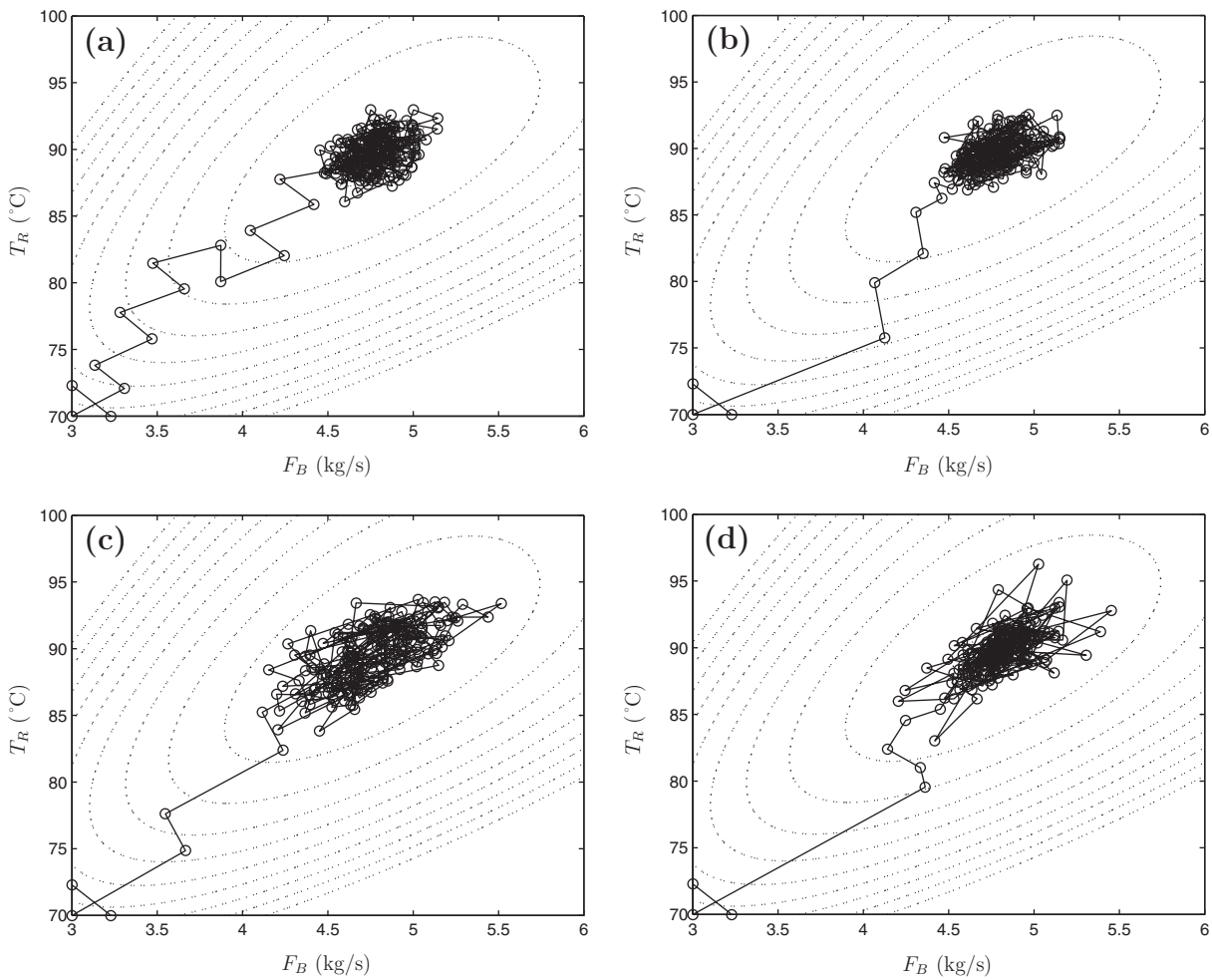


**Fig. 4.** Input trajectory for Scenario 1 with 200 operating points. Dotted lines: contours of the *plant* cost function. (a) Previous dual modifier adaptation with $\mathcal{E}^U = 111.2$. (b) New dual modifier adaptation with $\mathcal{E}^U = 111.2$. (c) Previous dual modifier adaptation with $\mathcal{E}^U = 200$. (d) New dual modifier adaptation with $\mathcal{E}^U = 200$.

The input trajectories obtained with $\mathcal{E}^U = 111.2$, which is the same value used in Marchetti et al. (2010), are shown in the top plots of Fig. 4: plot **a** shows the input trajectories obtained with the previous approach, while plot **b** shows the trajectories obtained with the new dual modifier-adaptation approach. The new approach shows a faster convergence to a neighborhood of the plant optimum in terms of the number of RTO iterations. The previous approach requires 13 iterations, while the new approach requires only 5 iterations. This improvement is due to the larger feasible regions produced by the gradient error constraint $\mathcal{E}^f(\mathbf{u}) \leq \mathcal{E}^U$ used in the new approach. The gradient error constraint $\mathcal{E}(\mathbf{u}) \leq \mathcal{E}^U$ used in the previous dual modifier-adaptation approach, represents a stronger restriction on the input moves, and thus slows down the convergence. The input trajectories obtained with $\mathcal{E}^U = 200$ are shown in the bottom plots of Fig. 4. Increasing the value of $\mathcal{E}^U$ allows for larger input moves in both approaches, since the feasible regions produced by the corresponding gradient error constraints are greater. However, this increases the allowed error in the estimated gradients, and thus the variability of the iterations with respect to the plant optimum. This increased variability results in a loss in economic profit.

### 5.3. Scenario 2

This scenario considers the optimization problem (63), including the inequality constraint $X_G \leq 0.08$. Since this constraint is active at the solution point, the noisy function to be considered in (36) reads:

$$\tilde{L}_p(\mathbf{u}) = \Phi_p(\mathbf{u}) + \mu G_p(\mathbf{u}) + \nu, \tag{64}$$

with $G_p = X_G - 0.08$. The Lagrange multiplier in (64) is selected as $\mu = 262.5$, which is the value obtained at the optimum of the nominal model. We assume that the measurements of the objective function and the concentration $X_G$ are Gaussian random variables with standard deviations $\sigma_\phi = 0.5$, and $\sigma_G = 0.0005$, respectively. Therefore, the noise $\nu$ of the Lagrangian function in (64) has a Gaussian distribution with standard deviation $\sigma_L = \sqrt{\sigma_\phi^2 + \mu^2\sigma_G^2}$. The noise interval is selected as $\delta = 6\sigma_L \approx 3.1$. Similar to Scenario 1, the inputs are scaled in the intervals [3, 6] for $F_B$, and [70, 100] for $T_R$. In this region, the largest eigenvalue of the Hessian of Lagrangian function predicted by the model, $L(\mathbf{u}, \boldsymbol{\theta}) = \Phi(\mathbf{u}, \boldsymbol{\theta}) + \mu G(\mathbf{u}, \boldsymbol{\theta})$, obtained with the scaled inputs, is 1085. Hence, we select $\sigma_{max} = 1085$.

The exponential filter (10) is implemented with $\mathsf{K} = \mathrm{diag}\,(1, 0.4\ \mathbf{I}_5)$. Simulations are run with 200 RTO iterations, starting from $\mathbf{u}_0 = [3.5,\ 72]^{\mathrm{T}}$. Three different noise realizations of the input trajectories obtained with $\mathcal{E}^U = 116$ are illustrated in Fig. 5 for the previous approach (left plot), and for the new approach (right plot). It can be seen that the new dual modifier-adaptation approach uses less steady-state operating points to reach a neighborhood of the constrained plant optimum. Similar to Scenario 1, by increasing the value of $\mathcal{E}^U$ it is possible to decrease

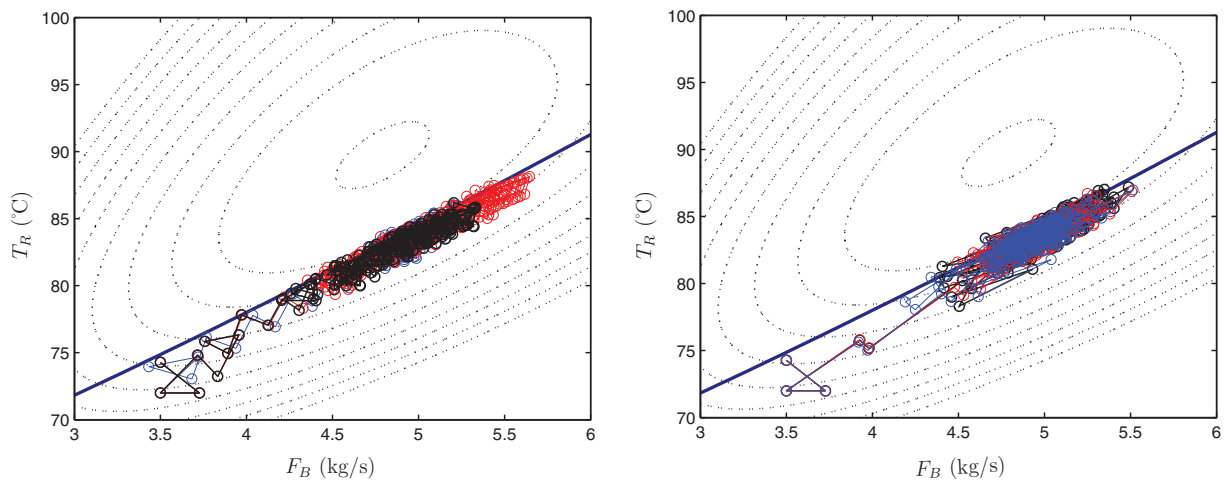


**Fig. 5.** Input trajectories for Scenario 2 with 200 operating points. Dotted lines: contours of the *plant* cost function. Thick solid line: Boundary of the inequality constraint ($X_G$ = 0.08). **Left plot:** Previous dual modifier adaptation with $\mathcal{E}^U = 116$. **Right plot:** New dual modifier adaptation with $\mathcal{E}^U = 116$.

the number of iterations required to approach the plant optimum. However, this will be at the expense of increased gradient errors, which will translate into a larger variability of the inputs with respect to the plant optimum. In the constrained case, this larger variability will result not only in a loss in economic profit, but also in larger constraint violations.

## 6. Discussion and conclusions

The contributions of this work to the modifier-adaptation literature can be summarized as follows:

- The model adequacy criterion for modifier-adaptation schemes has been extended in order to include conditions concerning an inadequate model. Previous model adequacy conditions in the RTO literature only include sufficient conditions for an adequate model (Forbes & Marlin, 1996; Marchetti et al., 2009).
- A sensitivity analysis of the optimality loss of a constrained NLP problem, induced by small Lagrangian gradient errors, was conducted. The analysis shows that, in the presence of inaccurate gradient estimates, the loss in optimality is determined by the resulting error in the gradient of the Lagrangian function. This result supports the idea that the gradient error constraint in dual modifier-adaptation schemes should pay attention to the accuracy with which the gradient of the Lagrangian function is estimated.
- New gradient modifiers were introduced for which the first-order corrections of the cost and constraint functions provide an improved approximation of the cost and constraint functions for the true process. The modified cost and constraint functions match the measured values of the cost and constraint functions at the current and $n_u$ past operating points.
- A new gradient error constraint was proposed, that pays attention to the accuracy with which the gradient of the Lagrangian function is estimated in the polyhedral set that has the current and $n_u$ past operating points as its extreme points. This constraint produces larger feasible regions than the constraint used in Marchetti et al. (2010), which allows larger input moves and thus, faster convergence to the optimum.

The main assumptions made in this work are listed and discussed next:

- The cost function and the constrained variables can be measured, or are known functions of the measured variables. If any of these variables is not measured, then the corresponding modifiers cannot be evaluated. In particular, if a constrained variable cannot be measured nor estimated from the measured variables, then some conservatism should be introduced. For example, one could impose fixed (conservative) constraint backoffs to the unmeasured constraints (Loeblein & Perkins, 1996).
- In evaluating the error in the gradient estimates, only the output variables are assumed to be subject to measurement noise. The input variables are assumed to be noise free. This assumption is valid if the input variables are setpoints of feedback controllers, or if they represent known variables applied to the plant, e.g., the voltage applied to a pump that is used to feed a tank, instead of the inlet flowrate to the tank.
- The frequency of the disturbances affecting the plant is sufficiently low with respect to the time required by the RTO scheme to approach the plant optimum. Notice that small disturbances can be accommodated within the noise level $\delta$ (see Fig. 1). However, larger disturbances that have a meaningful impact on the optimum point should occur at a very low frequency or slowly in time, in order for the gradient estimates to be meaningful and the dual modifier-adaptation scheme to effectively approach the plant optimum. This clearly limits the applicability of this approach, and in particular its scalability to large scale systems with many input variables.

For many complex chemical and biochemical processes it is very difficult, time consuming, and expensive to obtain accurate models for process optimization. The use of an inaccurate model can lead to a suboptimal or even infeasible operating point. If the frequency of disturbances affecting the plant is sufficiently low, it is possible to estimate the gradients from the current and past operating points and apply the dual modifier-adaptation approach in order to converge to the true process optimum. The dual modifier-adaptation approach proposed in this paper uses new gradient modifiers and a new gradient error constraint. The case study of the Williams–Otto reactor has served to compare the new approach with a previous dual modifier-adaptation approach proposed in the literature.

## References

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2006). *Nonlinear programming: Theory and algorithms* (3rd edition). New Jersey: John Wiley and Sons.

Brdyś, M., & Tatjewski, P. (1994). An algorithm for steady-state optimizing dual control of uncertain plants. In *Proc. 1st IFAC workshop on new trends in design of control systems* Smolenice, Slovakia, (pp. 249–254).

Brdyś, M., & Tatjewski, P. (2005). *Iterative algorithms for multilayer optimizing control*. London, UK: Imperial College Press.

Darby, M. L., Nikolaou, M., Jones, J., & Nicholson, D. (2011). RTO: an overview and assessment of current practice. *Journal of Process Control*, *21*, 874–884.

Fiacco, A. V. (1983). *Introduction to sensitivity and stability analysis in nonlinear programming, Vol. 165 of mathematics in science and engineering*. New York: Academic Press.

Forbes, J. F., & Marlin, T. E. (1996). Design cost: A systematic approach to technology selection for model-based real-time optimization systems. *Computers & Chemical Engineering*, *20*, 717–734.

Forbes, J. F., Marlin, T. E., & MacGregor, J. F. (1994). Model adequacy requirements for optimizing plant operations. *Computers & Chemical Engineering*, *18*(6), 497–510.

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

Gao, W., & Engell, S. (2005). Iterative set-point optimization of batch chromatography. *Computers & Chemical Engineering*, *29*, 1401–1409.

Gill, P. E., Murray, W., & Wright, M. H. (2003). *Practical optimization*. London: Academic Press.

Loeblein, C., & Perkins, J. D. (1996). Economical analysis of different structures of on-line process optimization systems. *Computers & Chemical Engineering*, *20*, S551–S556.

Luenberger, D. G., & Ye, Y. (2008). *Linear and nonlinear programming* (3rd edition). New York: Springer Science + Business Media, LLC.

Marchetti, A. (2009). *Modifier-adaptation methodology for real-time optimization (No. 4449)*. Ph.D. thesis, Ecole Polytechnique Fédérale de Lausanne.

Marchetti, A. G., & Basualdo, M. (2012). A new dual modifier-adaptation approach for real-time optimization with inaccurate models. In *22th European symposium on computer aided process engineering* ESCAPE 22, London, UK.

Marchetti, A., Chachuat, B., & Bonvin, D. (2009). Modifier-adaptation methodology for real-time optimization. *Industrial & Engineering Chemistry Research*, *48*(13), 6022–6033.

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

Marlin, T. E., & Hrymak, A. N. (1997). Real-time operations optimization of continuous processes. In *AIChE symposium series - CPC-V, Vol. 93* (pp. 156–164).

Williams, T. J., & Otto, R. E. (1960). A generalized chemical processing model for the investigation of computer control. *AIEE Transactions*, *79*, 458.

Zhang, Y., & Forbes, J. F. (2000). Extended design cost: A performance criterion for real-time optimization systems. *Computers & Chemical Engineering*, *24*, 1829–1841.