Computers and Chemical Engineering 98 (2017) 61–69

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Modifier adaptation with guaranteed feasibility in the presence of gradient uncertainty

A.G. Marchetti*[,1], M. Singhal, T. Faulwasser[2], D. Bonvin

Laboratoire d'Automatique, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 15 June 2016
Received in revised form 17 October 2016
Accepted 24 November 2016
Available online 30 November 2016

Keywords:
Real-time optimization
Modifier adaptation
Feasible operation
Gradient uncertainty

ABSTRACT

In the context of real-time optimization, modifier-adaptation schemes use estimates of the plant gradients to achieve plant optimality despite plant-model mismatch. Plant feasibility is guaranteed upon convergence, but not at the successive operating points computed by the algorithm prior to convergence. This paper presents a strategy for guaranteeing rigorous constraint satisfaction of all iterates in the presence of plant-model mismatch and uncertainty in the gradient estimates. The proposed strategy relies on constructing constraint upper-bounding functions that are robust to the gradient uncertainty that results when the gradients are estimated by finite differences from noisy measurements. The performance of the approach is illustrated for the optimization of a continuous stirred-tank reactor.

1. Introduction

The optimization of process operations is key to the economic success of continuous and batch industrial processes, for which the goal is to maximize profit or minimize cost subject to a number of operating constraints. An optimal operating point is typically found by solving a model-based optimization problem. Unfortunately, due to plant-model mismatch, the optimal solution obtained using the model does not, in general, correspond to the plant optimum, and may not even be feasible for the plant.

In order to deal with plant-model mismatch, real-time optimization (RTO) schemes use measurements to iteratively improve the quality of the model used for optimization. The most common RTO scheme consists in the two-step approach of parameter estimation followed by optimization using the updated model (Chen and Joseph, 1987; Darby et al., 2011). The performance of the two-step approach is highly dependent on the accuracy of the model, the choice of the adjustable parameters, and the choice of the measurements (Yip and Marlin, 2004; Quelhas et al., 2013). In the presence of structural plant-model mismatch, the model optimum will typically differ from the plant optimum. It is well known that a model that is adequate for optimization needs to be able to accurately approximate the plant necessary conditions of optimality (NCO), which involves the characterization of the set of active constraints and the gradients of the cost and constraint functions (Biegler et al., 1985; Marchetti et al., 2009). Hence, it is well understood in the RTO community that gradient estimates for the uncertain plant are key in achieving optimal plant operation in the presence of plant-model mismatch (Roberts, 1995; Marchetti et al., 2009; Gao and Engell, 2005).

For instance, modifier-adaptation (MA) schemes enforce plant optimality and feasibility upon convergence by making first-order corrections to the cost and constraint functions based on estimates of the plant cost and constraint gradients (Gao and Engell, 2005; Marchetti et al., 2009). Several gradient estimation methods can be considered for RTO applications (Mansour and Ellis, 2003; Srinivasan et al., 2011). Dual RTO schemes estimate the gradients by finite-difference approximation based on the past operating points (Brdyś and Tatjewski, 1994, 2005; Marchetti et al., 2010; Marchetti, 2013). This requires accommodating two conflicting objectives: the "primal objective" aims at improving the cost, while the "dual objective" aims at estimating the gradients accurately. The dual objective is dealt with by introducing a *duality constraint* that considers the accuracy of the gradient estimates.

In addition to enforcing plant optimality upon convergence, it is desirable for an RTO algorithm to generate only feasible iterates, that is, the uncertain plant constraints are satisfied at all RTO iterates (Bunin et al., 2013). For continuous processes, it is possible to generate only feasible steady-state operating points by

* Corresponding author.
E-mail addresses: marchetti@cifasis-conicet.gov.ar (A.G. Marchetti), martand.singhal@epfl.ch (M. Singhal), timm.faulwasser@kit.edu (T. Faulwasser), dominique.bonvin@epfl.ch (D. Bonvin).
[1] Also with the French-Argentine International Center for Information and Systems Sciences (CIFASIS), CONICET-Universidad Nacional de Rosario (UNR), S2000EZP Rosario, Argentina.
[2] Also with the Institute for Applied Computer Science, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany.

http://dx.doi.org/10.1016/j.compchemeng.2016.11.027
0098-1354/

implementing the RTO results via a feedback control layer that controls the constrained quantities (Tatjewski et al., 2001; Marchetti et al., 2011; Navia et al., 2012). In many cases, however, the use of feedback controllers is not possible or inappropriate, for instance, when the constrained quantities cannot be measured with the required frequency. Another possibility to enforce feasibility of the points generated by the RTO algorithm is by using constraint functions that upper bound the constraints of the real system. The use of constraint upper bounds generates feasibility regions that are inner approximations of the true plant feasibility region. Bunin et al. (2013) proposed constraint upper bounds based on Lipschitz constants. These bounds match the value of the plant constraints at the current operating point and do not require the knowledge of the corresponding gradients. Nevertheless, since the approach may converge to a suboptimal point if the RTO iterates get too close to constraint activation, a projection is typically implemented to stay sufficiently deep inside the feasible region. This projection does require the knowledge of the gradients of the plant constraints. More recently, Singhal et al. (2016) considered the use of constraint upper bounds based on quadratic data-driven surrogate models. These constraint upper bounds match the values and the gradients of the plant constraints at the current operating point, and are valid under the assumption that the gradients of the plant constraints are perfectly known. However, in any real-world application, the plant gradients can only be approximated, and the feasibility guarantees provided by the constraint upper bounds are lost.

In this paper, we derive constraint upper bounds based on second-order corrections of a first-principles model for use in modifier-adaptation schemes, and we tackle the problem of rigorous feasibility guarantees for all RTO iterates in the presence of gradient uncertainty. In general, the level of gradient uncertainty depends on the approach used to estimate the gradients. Here, we consider finite-difference gradient estimation from the measurements obtained at the current and past operating points. For this case, an upper-bound on the gradient-error norm was obtained in Marchetti et al. (2010). In the present paper, we rely on this approach to construct bounds on the constraint gradient errors. We show how these bounds can be used to robustly guarantee plant feasibility of modifier-adaptation iterates in the presence of plant-model mismatch and additive gradient uncertainty.

The paper is organized as follows. Section 2 presents preliminary material that is used in this work, including the formulation of the static optimization problem, the introduction to modifier adaptation, the approach used to estimate the gradients from past operating points, the computation of upper bounds on the gradient error, and the introduction to dual modifier adaptation. The proposed approach for guaranteeing feasibility in RTO schemes by constructing robust constraint upper bounds is presented in Section 3. A dual modifier-adaptation algorithm is formulated that combines all the elements previously introduced. The performance of the RTO algorithm is illustrated via the case study of the Williams-Otto reactor in Section 4. Finally, conclusions are drawn in Section 5.

2. Preliminaries

2.1. Optimization problem

The steady-state optimization problem for the plant can be formulated as the nonlinear program (NLP)

$$\begin{aligned} \min_{\mathbf{u}} \quad & \Phi_p(\mathbf{u}) := \phi(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \\ \text{s.t.} \quad & G_{p,i}(\mathbf{u}) := g_i(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \le 0, \quad i = 1, \ldots, n_g, \\ & \mathbf{u} \in \mathcal{U}, \end{aligned} \tag{1}$$

where $\mathbf{u} \in \mathbb{R}^{n_u}$ are the decision (or input) variables; $\mathbf{y}_p \in \mathbb{R}^{n_y}$ are the measured output variables; $\phi : \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}$ is the cost function to be minimized; $g_i : \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}$, $i = 1, \ldots, n_g$, is the set of process-dependent inequality constraint functions; and $\mathcal{U} \subset \mathbb{R}^{n_u}$ is a compact set. $\mathcal{U}$ is typically determined by lower and upper bounds on the input variables, $\mathcal{U} = \{\mathbf{u} \in \mathbb{R}^{n_u} : \mathbf{u}^{\mathrm{L}} \le \mathbf{u} \le \mathbf{u}^{\mathrm{U}}\}$.

The feasible sets associated with the constraints $G_{p,i}$ are defined as

$$\mathcal{G}_{p,i} = \{\mathbf{u} \in \mathbb{R}^{n_u} : G_{p,i}(\mathbf{u}) \le 0\}, \quad i = 1, \ldots, n_g, \tag{2}$$

with which the feasible set of the plant can be written as

$$\mathcal{F}_p = \left(\cap_{i=1}^{n_g} \mathcal{G}_{p,i}\right) \cap \mathcal{U}. \tag{3}$$

Note that ϕ and g_i are known functions of $\mathbf{u}$ and $\mathbf{y}_p$. On the other hand, the input-output steady-state mapping of the plant $\mathbf{y}_p : \mathbb{R}^{n_u} \to \mathbb{R}^{n_y}$ is typically unknown, and only an approximate model is available, which gives the steady-state mapping $\mathbf{y} : \mathbb{R}^{n_u} \to \mathbb{R}^{n_y}$. The model may consist of mass and energy balances and other physical relationships. Using this model, the solution to Problem (1) can be approached by solving the following model-based optimization problem:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \Phi(\mathbf{u}) := \phi(\mathbf{u}, \mathbf{y}(\mathbf{u})) \\ \text{s.t.} \quad & G_i(\mathbf{u}) := g_i(\mathbf{u}, \mathbf{y}(\mathbf{u})) \le 0, \quad i = 1, \ldots, n_g, \\ & \mathbf{u} \in \mathcal{U}. \end{aligned} \tag{4}$$

The solution to Problem (4) does not in general agree with the solution to Problem (1) in the presence of plant-model mismatch and disturbances. The purpose of RTO schemes is to iteratively approach the plant optimum by updating the model based on measurements. The standard strategy for updating the model is to estimate some of the model parameters by solving a parameter estimation problem (Chen and Joseph, 1987; Darby et al., 2011). However, it is well known that this two-step approach does not in general converge to the plant optimum (Roberts, 1995; Marchetti et al., 2009). The modifier-adaptation approach described next introduces a parameterization that is specially tailored to match the NCO of the plant upon convergence.

2.2. Modifier adaptation

Modifier adaptation introduces first-order correction terms that are added to the cost and constraint functions predicted by the nominal model. At the kth RTO iteration, the next RTO inputs are computed by solving the following *modified* optimization problem (Marchetti et al., 2009):

$$\mathbf{u}_{k+1} = \operatorname{argmin}_{\mathbf{u}} \quad \Phi_{m,k}(\mathbf{u}) := \Phi(\mathbf{u}) + (\boldsymbol{\lambda}_k^{\Phi})^{\mathsf{T}}\mathbf{u} \tag{5a}$$

$$\text{s.t.} \quad G_{m,i,k}(\mathbf{u}) := G_i(\mathbf{u}) + \varepsilon_{i,k} + (\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k) \le 0, \quad i = 1, \ldots, n_g, \tag{5b}$$

$$\mathbf{u} \in \mathcal{U}, \tag{5c}$$

with

$$\varepsilon_{i,k} = G_{p,i}(\mathbf{u}_k) - G_i(\mathbf{u}_k), \tag{6a}$$

$$(\boldsymbol{\lambda}_k^{\Phi})^{\mathsf{T}} = \frac{\partial \Phi_p}{\partial \mathbf{u}}(\mathbf{u}_k) - \frac{\partial \Phi}{\partial \mathbf{u}}(\mathbf{u}_k), \tag{6b}$$

$$(\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}} = \frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k) - \frac{\partial G_i}{\partial \mathbf{u}}(\mathbf{u}_k), \tag{6c}$$

where $\Phi_{m,k}$ and $G_{m,i,k}$ are the *modified* cost and constraint functions at the kth RTO iteration, the scalars $\varepsilon_{i,k}$, $i = 1, \ldots, n_g$, are the constraint value modifiers, and the vectors $\boldsymbol{\lambda}_k^{\Phi}$, and $\boldsymbol{\lambda}_k^{G_i}$, $i = 1, \ldots, n_g$, are the cost and constraint gradient modifiers.

These modifiers may be filtered, as proposed by Marchetti et al. (2009).

Modifier adaptation has the attractive property that it reaches a KKT point of the plant Problem (1) upon convergence. The difficulty lies in the need to estimate the plant gradients $\frac{\partial \Phi_p}{\partial \mathbf{u}}(\mathbf{u}_k)$ and $\frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k)$ at each RTO iteration.

2.3. Gradient estimation by finite-difference approximation using past operating points

We recall the gradient estimation used in dual ISOPE and dual modifier-adaptation methods (Brdyś and Tatjewski, 1994, 2005; Marchetti et al., 2010; Marchetti, 2013). This estimation uses finite-difference approximation based on the measurements obtained at the current and past operating points $\mathbf{u}_k, \mathbf{u}_{k-1}, \ldots, \mathbf{u}_{k-n_u}$. At the kth RTO iteration, the following matrix can be constructed:

$$U_k := [\ \mathbf{u}_k - \mathbf{u}_{k-1}, \quad \mathbf{u}_k - \mathbf{u}_{k-2}, \quad \ldots, \quad \mathbf{u}_k - \mathbf{u}_{k-n_u}\ ] \in \mathbb{R}^{n_u \times n_u}.$$

Assuming that measurements of the cost Φ_p and constrained quantities $G_{p,i}$ are available at each iteration, we construct the following vectors:

$$\boldsymbol{\delta}\Phi_{p,k} := [\ \tilde{\Phi}_k - \tilde{\Phi}_{k-1}, \quad \tilde{\Phi}_k - \tilde{\Phi}_{k-2}, \quad \ldots, \quad \tilde{\Phi}_k - \tilde{\Phi}_{k-n_u}\ ] \in \mathbb{R}^{1\times n_u},$$

$$\boldsymbol{\delta}G_{p,i,k} := [\ \tilde{G}_{i,k} - \tilde{G}_{i,k-1}, \quad \tilde{G}_{i,k} - \tilde{G}_{i,k-2}, \quad \ldots, \quad \tilde{G}_{i,k} - \tilde{G}_{i,k-n_u}\ ] \in \mathbb{R}^{1\times n_u}, \quad i = 1, \ldots, n_g,$$

where the superscript $\tilde{(\cdot)}$ denotes noisy measurements with measurement noise v:

$$\tilde{G}_{i,k} = G_{p,i}(\mathbf{u}_k) + v_k. \tag{7}$$

The cost and constraint gradients at $\mathbf{u}_k$ can then be estimated by finite-difference approximation as follows:

$$\hat{\nabla}\Phi_{p,k} = \boldsymbol{\delta}\Phi_{p,k}(U_k)^{-1}, \tag{8}$$

$$\hat{\nabla}G_{p,i,k} = \boldsymbol{\delta}G_{p,i,k}(U_k)^{-1}, \quad i = 1, \ldots, n_g, \tag{9}$$

which requires the columns of U_k to be linearly independent.

2.4. Upper bound on the gradient-error norm

The gradient uncertainty associated with the aforementioned finite-difference approximation was analyzed by Marchetti et al. (2010). The gradient estimation error is defined as the difference between the estimated gradient and the true gradient:

$$\mathbf{e}^{\nabla}_{i,k} = \hat{\nabla}G_{i,k} - \frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k). \tag{10}$$

From (7) and (9), this error can be split as

$$\mathbf{e}_{i,k} = \mathbf{e}^t_{i,k} + \mathbf{e}^n_{i,k}, \tag{11}$$

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

$$(\mathbf{e}^t_{i,k})^{\mathsf{T}} = [G_{p,i}(\mathbf{u}_k) - G_{p,i}(\mathbf{u}_{k-1}), \quad \ldots, \quad G_{p,i}(\mathbf{u}_k) - G_{p,i}(\mathbf{u}_{k-n_u})](U_k)^{-1} - \frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k), \tag{12a}$$

$$(\mathbf{e}^n_{i,k})^{\mathsf{T}} = [v_k - v_{k-1}, \quad \ldots, \quad v_k - v_{k-n_u}](U_k)^{-1}. \tag{12b}$$

Assuming that $G_{p,i}$ is twice continuously differentiable with respect to $\mathbf{u}$, the norm of the gradient error due to truncation can be upper bounded as

$$\|\mathbf{e}^t_{i,k}\| \le d_{g_i} r_k, \tag{13}$$

where d_{g_i} is an upper bound on the spectral radius of the Hessian of $G_{p,i}$ for $\mathbf{u} \in \mathcal{U}$, and r_k is the radius of the unique n_u-sphere that can be generated from the points $\mathbf{u}_k, \quad \mathbf{u}_{k-1}, \quad \ldots, \quad \mathbf{u}_{k-n_u}$:

$$r_k = r(\mathbf{u}_k, \quad \mathbf{u}_{k-1}, \quad \ldots, \quad \mathbf{u}_{k-n_u}) = \frac{1}{2}\left\| \left[(\mathbf{u}_k - \mathbf{u}_{k-1})^{\mathsf{T}}(\mathbf{u}_k - \mathbf{u}_{k-1}), \ldots, (\mathbf{u}_k - \mathbf{u}_{k-n_u})^{\mathsf{T}}(\mathbf{u}_k - \mathbf{u}_{k-n_u})\right](U_k)^{-1} \right\|. \tag{14}$$

In turn, assuming that the noisy measurements $\tilde{G}_i$ remain within an interval δ_{g_i} at steady-state operation, the norm of the gradient error due to measurement noise can be bounded from above:

$$\|\mathbf{e}^n_{i,k}\| \le \frac{\delta_{g_i}}{l_{\min,k}}, \qquad l_{\min,k} = l_{\min}(\mathbf{u}_k, \mathbf{u}_{k-1}, \quad \ldots, \quad \mathbf{u}_{k-n_u}), \tag{15}$$

where $l_{\min,k}$ is the minimum distance between all possible pairs of complement affine subspaces that can be generated from the set of points $\mathcal{S}_k = \{\mathbf{u}_k, \quad \mathbf{u}_{k-1}, \quad \ldots, \quad \mathbf{u}_{k-n_u}\}$. Using (13) and (15), the gradient-error norm can be bounded from above by

$$\|\mathbf{e}_{i,k}\| \le \|\mathbf{e}^t_{i,k}\| + \|\mathbf{e}^n_{i,k}\| \le E_{i,k} := d_{g_i} r_k + \frac{\delta_{g_i}}{l_{\min,k}}. \tag{16}$$

We refer to Marchetti et al. (2010) for detailed derivations of the aforementioned relations. In addition, we remark that the gradient-error bound $E_{i,k}$ will, in general, be different for each constraint $g_{p,i}$ since the values of the parameters d_{g_i} and δ_{g_i} are different for each constraint. Furthermore, the bound $E_{i,k}$ will change from one RTO iteration to the next because the set of points $\mathcal{S}_k$ changes.

2.5. Dual modifier adaptation

Dual modifier-adaptation schemes estimate the gradients based on the measurements obtained at the current and past operating points. In order to obtain accurate gradient estimates, a duality constraint enforcing the accuracy of gradient estimates of the Lagrangian has been proposed (Marchetti et al., 2010; Marchetti, 2013). To this end, consider the set of points $\mathcal{S}_k(\mathbf{u}) = \{\mathbf{u}, \quad \mathbf{u}_k, \quad \ldots, \quad \mathbf{u}_{k-n_u+1}\}$, for which we can compute the radius $r_k(\mathbf{u}) = r(\mathbf{u}, \quad \mathbf{u}_k, \quad \ldots, \quad \mathbf{u}_{k-n_u+1})$, and the minimum distance $l_{\min,k}(\mathbf{u}) = l_{\min}(\mathbf{u}, \quad \mathbf{u}_k, \quad \ldots, \quad \mathbf{u}_{k-n_u+1})$. This way, it is possible to ensure a given accuracy of the estimated gradient, for example $\|\mathbf{e}^L(\mathbf{u})\| \le E^{\mathrm{U}}$, by selecting $\mathbf{u}$ such that,

$$\|\mathbf{e}^L(\mathbf{u})\| \le E^L_k(\mathbf{u}) := d^L r_k(\mathbf{u}) + \frac{\delta^L}{l_{\min,k}(\mathbf{u})} \le E^{\mathrm{U}}, \tag{17}$$

where $\mathbf{e}^L$ is the Lagrangian gradient error, d^L is an upper bound on the spectral radius of the Hessian of the Lagrangian function, and δ^L is the range of error in the Lagrangian function that results from noisy measurements of the cost and constraint functions (Marchetti, 2013). In dual modifier adaptation, this can be achieved by including the duality constraint

$$\mathcal{D}_k(\mathbf{u}) = E^L_k(\mathbf{u}) - E^{\mathrm{U}} \le 0 \tag{18}$$

in the modified optimization problem. The following problem is solved at the kth RTO iteration:

$$\begin{aligned} \mathbf{u}_{k+1} = \operatorname{argmin}_{\mathbf{u}} \quad & \Phi_{m,k}(\mathbf{u}) := \Phi(\mathbf{u}) + (\hat{\boldsymbol{\lambda}}^{\Phi}_k)^{\mathsf{T}}\mathbf{u} \\ \text{s.t.} \quad & G_{m,i,k}(\mathbf{u}) := G_i(\mathbf{u}) + \varepsilon_{i,k} + (\hat{\boldsymbol{\lambda}}^{G_i}_k)^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k) \le 0, \\ & i = 1, \ldots, n_g, \\ & \mathcal{D}_k(\mathbf{u}) \le 0, \quad \mathbf{u} \in \mathcal{U}, \end{aligned} \tag{19}$$

where $\hat{\boldsymbol{\lambda}}_k^{\Phi}$ and $\hat{\boldsymbol{\lambda}}_k^{G_i}$ are the estimated cost and constraint gradient modifiers

$$(\hat{\boldsymbol{\lambda}}_k^{\Phi})^{\mathsf{T}} = \hat{\nabla}\Phi_{p,k} - \frac{\partial \Phi}{\partial \mathbf{u}}(\mathbf{u}_k), \tag{20a}$$

$$(\hat{\boldsymbol{\lambda}}_k^{G_i})^{\mathsf{T}} = \hat{\nabla}G_{p,i,k} - \frac{\partial G_i}{\partial \mathbf{u}}(\mathbf{u}_k), \quad i = 1, \ldots, n_g. \tag{20b}$$

Remark 1 *(Feasible Region of Duality Constraint).* The duality constraint $\mathcal{D}_k(\mathbf{u}) \leq 0$ generates two disjoint feasible regions (Marchetti et al., 2010). Because of this, the solution to Problem (19) is tackled by solving two separate optimization problems, one for each of the disjoint regions. For simplicity, we have presented the RTO problem (19) as a single optimization problem.

Remark 2 *(Initialization).* The initialization of the dual modifier-adaptation scheme requires relying initially on $n_u + 1$ points that satisfy the duality constraint $E_k^L(\mathbf{u}) \leq E^U$ in order to obtain a first estimate of the gradients. In this paper, the initial $n_u + 1$ points are obtained by introducing deviations from the starting point along the Cartesian axes, that is, using a forward-finite difference approach to initialize the algorithm. An optimized initial phase has been proposed for the dual ISOPE algorithm (Brdyś and Tatjewski, 2005), which uses a different duality constraint. A similar initialization phase has not been developed for the dual modifier-adaptation scheme considered in this work.

3. Modifier adaptation with guaranteed feasibility

The modifier-adaptation scheme (5) and (6) guarantees feasibility upon convergence. However, even in the absence of measurement noise and with perfect gradient estimates, the RTO points computed prior to convergence may violate the constraints since the realistic dual modifier-adaptation scheme (19) and (20), which accounts for measurement error and estimated gradients, does not in general guarantee feasibility of the computed RTO inputs. In this section, we propose an approach for guaranteeing feasibility at every iteration in (dual) modifier adaptation.

3.1. Feasibility guarantee based on upper-bounding functions

Let us assume that a non-empty inner approximation $\mathcal{F}_k$ to the plant feasible set $\mathcal{F}_p$ in (3) can be computed at the RTO point $\mathbf{u}_k$. Then, an RTO algorithm will guarantee feasibility if it enforces the constraint

$$\mathbf{u}_{k+1} \in \mathcal{F}_k \subseteq \mathcal{F}_p$$

at each RTO iteration. Next, we propose a procedure for computing the feasible set $\mathcal{F}_k$ by constructing the inner approximations $\mathcal{G}_{i,k}$ of the feasibility sets for each individual constraint $g_{p,i}$

$$\mathcal{G}_{i,k} \subseteq \mathcal{G}_{p,i}, \quad i = 1, \ldots, n_g.$$

Intersecting these sets yields

$$\mathcal{F}_k = \left(\cap_{i=1}^{n_g} \mathcal{G}_{i,k}\right) \cap \mathcal{U} \tag{21}$$

and the following trivial result.

Lemma 1 (Feasible Set). *Let $\mathcal{G}_{i,k} \subseteq \mathcal{G}_{p,i}, \quad i = 1, \ldots, n_g$, then the set $\mathcal{F}_k$ from (21) is a subset of the feasible set of the plant (3), that is, $\mathcal{F}_k \subseteq \mathcal{F}_p$.*

To compute the sets $\mathcal{G}_{i,k}$, one needs to generate the upper-bounding functions $G_{i,k}^U$ on the unknown constraint functions $G_{p,i}$,

$$G_{p,i}(\mathbf{u}) \leq G_{i,k}^U(\mathbf{u}), \forall \mathbf{u} \in \mathcal{U}, \quad i = 1, \ldots, n_g.$$

Noting that enforcing the right-hand side of the above inequality to be non-positive also enforces $G_{p,i}(\mathbf{u}) \leq 0$, the inner approximations $\mathcal{G}_{i,k}$ can be written as:

$$\mathcal{G}_{i,k} = \left\{\mathbf{u} \in \mathcal{U} : G_{i,k}^U(\mathbf{u}) \leq 0\right\}. \tag{22}$$

We start by defining a quadratic upper-bound matrix and a quadratic upper-bounding function to be constructed at the kth RTO iteration.

Definition 1 *(Quadratic Upper Bound).* Consider the function $f: \mathbb{R}^{n_u} \to \mathbb{R}$. Let $\overline{Q} \in \mathbb{R}^{n_u \times n_u}$ be a symmetric matrix and construct $q_k^U : \mathbb{R}^{n_u} \to \mathbb{R}$ as follows:

$$q_k^U(\mathbf{u}) = f(\mathbf{u}_k) + \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u}_k)(\mathbf{u} - \mathbf{u}_k) + \frac{1}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}\overline{Q}(\mathbf{u} - \mathbf{u}_k). \tag{23a}$$

If q_k^U satisfies

$$f(\mathbf{u}) \leq q_k^U(\mathbf{u}), \quad \forall \ \mathbf{u}, \mathbf{u}_k \in \mathcal{U}, \tag{23b}$$

then q_k^U is said to be a quadratic upper bound of f on $\mathcal{U}$ constructed at $\mathbf{u}_k$, and $\overline{Q}$ is called a quadratic upper-bound matrix.

Different quadratic bounds based on the Lipschitz constants of the partial derivatives of the function f have been presented in Bunin et al. (2014). Note that some of the bounds presented therein are piecewise quadratic in the sense that the matrix $\overline{Q}$ varies for different regions of the input space. The following lemma presents an alternative quadratic bound that uses a single design parameter.

Let f be twice continuously differentiable over $\mathcal{U}$ so that,

$$\underline{\lambda} \leq \min_{\mathbf{u} \in \mathcal{U}} \min_{\lambda} \left\{\lambda \in \sigma(H(\mathbf{u}))\right\}, \tag{24a}$$

$$\overline{\lambda} \geq \max_{\mathbf{u} \in \mathcal{U}} \max_{\lambda} \left\{\lambda \in \sigma(H(\mathbf{u}))\right\}, \tag{24b}$$

where $H(\mathbf{u}) = \frac{\partial^2 f}{\partial \mathbf{u}^2}(\mathbf{u})$ is the Hessian matrix of f evaluated at $\mathbf{u}$, and $\sigma(\cdot)$ stands for the spectrum of a matrix. The scalar $\underline{\lambda}$ is a lower bound on the smallest eigenvalue of the Hessian of f, while $\overline{\lambda}$ is an upper bound on the largest eigenvalue of the Hessian of f.

Lemma 2 (Construction of Quadratic Bounds). *Assume that f is twice continuously differentiable on $\mathcal{U}$. Then, for any $\mathbf{u}, \mathbf{u}_k \in \mathcal{U}$,*

$$q_k^L(\mathbf{u}) = f(\mathbf{u}_k) + \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u}_k)(\mathbf{u} - \mathbf{u}_k) + \frac{\underline{\lambda}}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k), \tag{25a}$$

$$q_k^U(\mathbf{u}) = f(\mathbf{u}_k) + \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u}_k)(\mathbf{u} - \mathbf{u}_k) + \frac{\overline{\lambda}}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k), \tag{25b}$$

satisfy

$$q_k^L(\mathbf{u}_k) = f(\mathbf{u}_k) = q_k^U(\mathbf{u}_k), \tag{26a}$$

$$\frac{\partial q_k^L}{\partial \mathbf{u}}(\mathbf{u}_k) = \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u}_k) = \frac{\partial q_k^U}{\partial \mathbf{u}}(\mathbf{u}_k), \tag{26b}$$

$$q_k^L(\mathbf{u}) \leq f(\mathbf{u}) \leq q_k^U(\mathbf{u}), \quad \forall \ \mathbf{u}, \mathbf{u}_k \in \mathcal{U}. \tag{26c}$$

Proof. Eqs. (26a) and (26b) follow directly from the definition of $q_k^L(\mathbf{u})$ and $q_k^U(\mathbf{u})$ in (26a) and (26b). To prove (26c), we apply Taylor's theorem at $\mathbf{u}_k$, together with the mean-value theorem,

$$f(\mathbf{u}) = f(\mathbf{u}_k) + \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u}_k)(\mathbf{u} - \mathbf{u}_k) + \frac{1}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}H(\overline{\mathbf{u}})(\mathbf{u} - \mathbf{u}_k), \tag{27}$$

with $\overline{\mathbf{u}} = \mathbf{u}_k + \gamma(\mathbf{u} - \mathbf{u}_k)$, for some $\gamma \in [0, 1]$. Since the Hessian matrix $H(\overline{\mathbf{u}})$ is symmetric, the proof proceeds by constructing upper and lower bounds of the Hessian using its smallest and largest eigenvalue. The inequalities in (26c) follow directly from the combination of these bounds with (24) and (27). □

The quadratic bounding functions q_k^L and q_k^U can be compared with the different quadratic bounds presented in Bunin et al. (2014). An advantage of the quadratic functions q_k^L and q_k^U is that they require

only a single design parameter, namely, either $\underline{\lambda}$ or $\overline{\lambda}$. The proposed bounding functions use the worst-case curvature (given by $\underline{\lambda}$ and $\overline{\lambda}$) in all input directions. This has the disadvantage that it can be very conservative for many functions. Conservatism can be reduced by using quadratic bounding functions with a diagonal or a full Hessian matrix, or by considering piecewise-quadratic functions. Note that the quadratic upper bound proposed in Bunin et al. (2013), which uses a diagonal matrix, is convex regardless of whether the function f is convex, concave, or neither of both. If f is strictly concave, having a convex upper bound may be overly conservative. In contrast, the upper bound given by (25b) may be strictly concave for a strictly concave function f, since we may choose $\overline{\lambda} < 0$.

3.2. Upper bounds for modifier adaptation

In this section, we propose to construct constraint upper bounds for use in modifier adaptation. Let us start by introducing the following assumption.

Assumption 1 *($\mathcal{C}^2$ plant and model constraints).* The plant constraints $G_{p,i}$, $i = 1, \ldots, n_g$, and the model constraints G_i, $i = 1, \ldots, n_g$, are twice continuously differentiable on $\mathcal{U}$.

The next proposition provides a theoretical derivation of the constraint upper-bounding functions.

Proposition 1 (Upper Bounds for Plant Constraints). *Let Assumption 1 hold, so that there exists a finite β_i satisfying*

$$\beta_i \geq \max_{\mathbf{u}\in\mathcal{U}} \max_{\lambda} \{\lambda \in \sigma(H_{p,i}(\mathbf{u}) - H_i(\mathbf{u}))\}, \quad i = 1, \ldots, n_g, \tag{28}$$

where $H_{p,i}(\mathbf{u}) = \frac{\partial^2 G_{p,i}}{\partial \mathbf{u}^2}(\mathbf{u})$ is the Hessian matrix of the plant constraint $G_{p,i}$ evaluated at $\mathbf{u}$, $H_i(\mathbf{u}) = \frac{\partial^2 G_i}{\partial \mathbf{u}^2}(\mathbf{u})$ is the Hessian matrix of the model constraint G_i evaluated at $\mathbf{u}$, and $\sigma(\cdot)$ stands for the spectrum of a matrix. The scalar β_i is an upper bound on the largest eigenvalue of the difference between the Hessian matrices of the plant and the model.

Then, at the kth RTO iteration, the following upper-bounding function

$$G^{\mathrm{U}}_{i,k}(\mathbf{u}) = G_i(\mathbf{u}) + \varepsilon_{i,k} + (\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k) + \frac{\beta_i}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k) \tag{29}$$

can be used to upper bound the evolution of $G_{p,i}$, for any $\mathbf{u}$ and $\mathbf{u}_k \in \mathcal{U}$, such that

$$G^{\mathrm{U}}_{i,k}(\mathbf{u}_k) = G_{p,i}(\mathbf{u}_k), \tag{30}$$

$$\frac{\partial G^{\mathrm{U}}_{i,k}}{\partial \mathbf{u}}(\mathbf{u}_k) = \frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k), \tag{31}$$

$$G^{\mathrm{U}}_{i,k}(\mathbf{u}) \geq G_{p,i}(\mathbf{u}), \quad \forall \mathbf{u} \in \mathcal{U}. \tag{32}$$

Proof. Evaluating $G^{\mathrm{U}}_{i,k}(\mathbf{u})$ at $\mathbf{u}_k$ gives

$$G^{\mathrm{U}}_{i,k}(\mathbf{u}_k) = G_i(\mathbf{u}_k) + \varepsilon_{i,k} = G_{p,i}(\mathbf{u}_k), \tag{33}$$

which proves Condition (30). Computing the gradient of $G^{\mathrm{U}}_{i,k}(\mathbf{u})$ gives

$$\frac{\partial G^{\mathrm{U}}_{i,k}}{\partial \mathbf{u}}(\mathbf{u}) = \frac{\partial G_i}{\partial \mathbf{u}}(\mathbf{u}) + (\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}} + \beta_i(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}, \tag{34}$$

which, when evaluated at $\mathbf{u}_k$, results in Condition (31). In order to prove Condition (32), let us consider the auxiliary function

$$f_{i,k}(\mathbf{u}) = G_{p,i}(\mathbf{u}) - G_i(\mathbf{u}) - \varepsilon_{i,k} - (\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k). \tag{35}$$

In view of Assumption 1, $f_{i,k}$ is twice continuously differentiable over $\mathcal{U}$. Hence, we can apply Taylor's theorem at $\mathbf{u}_k$, together with the mean-value theorem, in order to obtain the expression given in Eq. (27) for $f = f_{i,k}$. From

$$f_{i,k}(\mathbf{u}_k) = G_{p,i}(\mathbf{u}_k) - G_i(\mathbf{u}_k) - \varepsilon_{i,k} = 0, \tag{36}$$

and

$$\frac{\partial f_{i,k}}{\partial \mathbf{u}}(\mathbf{u}_k) = \frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k) - \frac{\partial G_i}{\partial \mathbf{u}}(\mathbf{u}_k) - (\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}} = \mathbf{0}, \tag{37}$$

Eq. (27) reduces to

$$f_{i,k}(\mathbf{u}) = \frac{1}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}} \left[H_{p,i}(\overline{\mathbf{u}}) - H_i(\overline{\mathbf{u}})\right](\mathbf{u} - \mathbf{u}_k), \tag{38}$$

with $\overline{\mathbf{u}} = \mathbf{u}_k + \gamma(\mathbf{u} - \mathbf{u}_k)$, for some $\gamma \in [0, 1]$. Considering (36)–(38) and using Lemma 2, it follows that $f_{i,k}$ can be upper bounded as

$$f_{i,k}(\mathbf{u}) \leq \frac{\beta_i}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k), \quad \forall \mathbf{u} \in \mathcal{U}. \tag{39}$$

Using (35) and (39) gives

$$G_{p,i}(\mathbf{u}) \leq G_i(\mathbf{u}) + \varepsilon_{i,k} + (\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k) + \frac{\beta_i}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k), \quad \forall \mathbf{u} \in \mathcal{U}, \tag{40}$$

which proves the Condition (32). □

Proposition 1 shows that the addition of the quadratic term $\frac{\beta_i}{2}(\mathbf{u} - \mathbf{u}_k)^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_k)$ to the modified constraints $G_{m,i,k}$ can be used to upper bound the plant constraints $G_{p,i}$. Note that a similar quadratic term has been added to an input-affine modified cost function in what is known as the *augmented* ISOPE method (Brdyś et al., 1987; Brdyś and Tatjewski, 2005). However, the purpose therein is to convexify the cost function and not to obtain an upper-bounding function. Although this paper is only concerned with feasibility, we note that convexification may also be important for convergence purpose. For instance, François and Bonvin (2013) showed that the use of convex cost and constraint functions guarantees that a KKT point of the plant is a fixed point of the modifier-adaptation algorithm.

Assuming that the constraint value $G_{p,i}(\mathbf{u}_k)$ and the constraint plant gradient $\frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k)$ are perfectly known, it is possible to guarantee feasibility of the next operating point $\mathbf{u}_{k+1}$ by selecting the feasibility sets $\mathcal{G}_{i,k}$ in (22) with $G^{\mathrm{U}}_{i,k}$ given by (29), and the feasibility set $\mathcal{F}_k$ as given by (21).

Remark 3 *(Perfect Model).* If the model is perfect in the sense that it matches the plant exactly, then the modifiers are always equal to zero, that is, $\varepsilon_{i,k} = 0$ and $\boldsymbol{\lambda}_k^{G_i} = \mathbf{0}$, and setting $\beta_i = 0$ satisfies Proposition 1 with equality in (32). Hence, at least theoretically, the constraint upper bounds (29) can match the plant constraints, and thus introduce no conservatism, in the case of a perfect model.

In general, it can be expected that the larger the value of β_i required to upper bound the constraint $G_{p,i}$, the smaller the resulting feasibility set $\mathcal{G}_{i,k}$. Note that the value of β_i depends on the accuracy of the model constraint G_i, in particular on the accuracy with which the model predicts the second-order derivatives of the plant. If the second-order derivatives of the plant are much larger than those of the model, then the required value of β_i will be large and result in a conservative upper bound.

Remark 4 *(Quadratic Data-Driven Model).* If no model is available, we have

$$G_i(\mathbf{u}) = 0, \quad \varepsilon_{i,k} = G_{p,i}(\mathbf{u}_k), \quad (\boldsymbol{\lambda}_k^{G_i})^{\mathsf{T}} = \frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k),$$

$$\beta_i \geq \max_{\mathbf{u}\in\mathcal{U}} \max_{\lambda} \{\lambda \in \sigma(H_{p,i}(\mathbf{u}))\},$$

and the upper-bounding function (29) reduces to the quadratic upper bound

$$G^{U}_{i,k}(\mathbf{u}) = G_{p,i}(\mathbf{u}_k) + \frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k)(\mathbf{u}-\mathbf{u}_k) + \frac{\beta_i}{2}(\mathbf{u}-\mathbf{u}_k)^{\mathsf{T}}(\mathbf{u}-\mathbf{u}_k),$$

which is the same as the upper bound given in (25b). An RTO approach that uses similar quadratic data-driven surrogate models has recently been proposed (Singhal et al., 2016).

3.3. Robust upper bounds

In practice, only the estimate $\hat{\nabla} G_{p,i,k}$ of the constraint gradient $\frac{\partial G_{p,i}}{\partial \mathbf{u}}(\mathbf{u}_k)$ is available, and thus the evaluation of (29) would require the knowledge of the gradient estimation error $\mathbf{e}_{i,k}$ from (10). Although the exact value of $\mathbf{e}_{i,k}$ is unknown, we shall assume that the gradient error is bounded, that is, $\mathbf{e}_{i,k} \in \mathcal{E}_{i,k}$, and that the uncertainty set $\mathcal{E}_{i,k}$ can be computed. In general, $\mathcal{E}_{i,k}$ will depend on the approach used to estimate the gradient. Hence, given the gradient estimate $\hat{\nabla} G_{p,i,k}$, the upper bound (29) is recast as a function of $\mathbf{u}$ and the gradient error variables $\mathbf{e}$:

$$G^{U}_{i,k}(\mathbf{u},\mathbf{e}) = G_i(\mathbf{u}) + \varepsilon_{i,k} + \left[\hat{\nabla} G_{p,i,k} - \mathbf{e}^{\mathsf{T}} - \frac{\partial G_i}{\partial \mathbf{u}}(\mathbf{u}_k)\right](\mathbf{u}-\mathbf{u}_k) + \frac{\beta_i}{2}(\mathbf{u}-\mathbf{u}_k)^{\mathsf{T}}(\mathbf{u}-\mathbf{u}_k), \tag{41}$$

and $G_{p,i}(\mathbf{u}_{k+1}) \le 0$ can be guaranteed if $\mathbf{u}_{k+1}$ is selected such that $\mathbf{u}_{k+1} \in \mathcal{G}_{i,k}$, where the set $\mathcal{G}_{i,k}$ is constructed as

$$\mathcal{G}_{i,k} = \left\{\mathbf{u} \in \mathcal{U} : \max_{\mathbf{e}\in\mathcal{E}_{i,k}} \; G^{U}_{i,k}(\mathbf{u},\mathbf{e}) \le 0\right\}. \tag{42}$$

By considering the sets $\mathcal{G}_{i,k}$ for each constraint $G_{p,i}$, the feasibility set $\mathcal{F}_k$ is given by (21). Since $G^{U}_{i,k}(\mathbf{u},\mathbf{e}) \le 0$ holds for the worst-case gradient error, it follows that $\mathcal{G}_{i,k} \subseteq \mathcal{G}_{p,i}$, and Lemma 1 implies $\mathcal{F}_k \subseteq \mathcal{F}_p$.

Next, we shall consider the special case where the bounded gradient uncertainty is given by an upper bound on the gradient-error norm

$$\mathcal{E}_{i,k} = \left\{\mathbf{e} \in \mathbb{R}^{n_u} : \|\mathbf{e}\| \le E_{i,k}\right\}, \tag{43}$$

whose computation in case of finite-difference gradient estimation is detailed in Section 2.4.[3] Furthermore, the use of norm bounds simplifies the construction of the feasibility sets $\mathcal{G}_{i,k}$ in (42), as given in the following proposition.

Proposition 2 (Robust Upper Bounds for Plant Constraints). *Let Assumption 1 hold and assume that, at the RTO point $\mathbf{u}_k$, an upper bound on the gradient-error norm $E_{i,k}$ is available for the constraint $G_{p,i}$. Then, the worst-case feasibility set $\mathcal{G}_{i,k}$ defined in (42) reduces to*

$$\mathcal{G}_{i,k} = \left\{\mathbf{u} \in \mathcal{U} : \overline{G}^{U}_{i,k}(\mathbf{u}) \le 0\right\} \tag{44}$$

with

$$\overline{G}^{U}_{i,k}(\mathbf{u}) = G_i(\mathbf{u}) + \varepsilon_{i,k} + (\hat{\boldsymbol{\lambda}}^{G_i}_k)^{\mathsf{T}}(\mathbf{u}-\mathbf{u}_k) + E_{i,k}\|\mathbf{u}-\mathbf{u}_k\| + \frac{\beta_i}{2}(\mathbf{u}-\mathbf{u}_k)^{\mathsf{T}}(\mathbf{u}-\mathbf{u}_k), \tag{45}$$

where $\hat{\boldsymbol{\lambda}}^{G_i}_k$ is the estimated constraint gradient modifier given in (20b).

[3] Other descriptions of gradient uncertainty are also possible. For instance, Bunin et al. (2013) considered box constraints on the true gradient values, but did not address the question of how these gradient box constraints can be obtained for some particular gradient estimation method.

Proof. The proof is based on the solution to the following optimization problem

$$\begin{aligned} &\max_{\mathbf{e}} \quad \left(\hat{\nabla} G_{p,i,k} - \mathbf{e}^{\mathsf{T}}\right)(\mathbf{u}-\mathbf{u}_k) \\ &\text{s.t.} \quad \mathbf{e}^{\mathsf{T}}\mathbf{e} \le (E_{i,k})^2. \end{aligned} \tag{46}$$

The KKT conditions of Problem (46) read:

$$(\mathbf{u}-\mathbf{u}_k)^{\mathsf{T}} + 2\lambda\mathbf{e}^{\mathsf{T}} = \mathbf{0}, \tag{47a}$$

$$\lambda \ge 0, \tag{47b}$$

$$\lambda(\mathbf{e}^{\mathsf{T}}\mathbf{e} - (E_{i,k})^2) = 0, \tag{47c}$$

$$\mathbf{e}^{\mathsf{T}}\mathbf{e} - (E_{i,k})^2 \le 0. \tag{47d}$$

If the inequality constraint is not active, then $\lambda = 0$ and (47a) leads to $\mathbf{u} = \mathbf{u}_k$. Therefore, if $\mathbf{u} \ne \mathbf{u}_k$, we must have $\lambda > 0$ and we conclude that the inequality constraint is (strongly) active, that is, $\mathbf{e}^{\mathsf{T}}\mathbf{e} = (E_{i,k})^2$. Hence, the gradient error $\mathbf{e}$ that maximizes (46) has a direction opposite to the vector $\mathbf{u}-\mathbf{u}_k$ and the magnitude $E_{i,k}$:

$$\mathbf{e}^{\mathsf{T}} = -E_{i,k}\frac{(\mathbf{u}-\mathbf{u}_k)^{\mathsf{T}}}{\|\mathbf{u}-\mathbf{u}_k\|}. \tag{48}$$

Substitution of this worst-case gradient error into Eq. (41) results in Eq. (45). □

3.4. Dual modifier adaptation with guaranteed feasibility

The robust constraint upper bounds (45) can now be used in the dual modifier-adaptation scheme in order to guarantee feasibility of the RTO solutions. The following problem is solved at the kth RTO iteration:

$$\mathbf{u}_{k+1} = \operatorname{argmin}_{\mathbf{u}} \quad \Phi_{m,k}(\mathbf{u}) = \Phi(\mathbf{u}) + (\hat{\boldsymbol{\lambda}}^{\Phi}_k)^{\mathsf{T}}\mathbf{u} \tag{49a}$$

$$\text{s.t.} \quad \overline{G}^{U}_{i,k}(\mathbf{u}) \le 0, \quad i = 1,\ldots,n_g \tag{49b}$$

$$\mathcal{D}_k(\mathbf{u}) \le 0, \quad \mathbf{u} \in \mathcal{U}, \tag{49c}$$

where a first-order filter with gain $K \in (0, 1]$ is applied to the cost gradient modifier:

$$(\hat{\boldsymbol{\lambda}}^{\Phi}_k)^{\mathsf{T}} = (1-K)(\hat{\boldsymbol{\lambda}}^{\Phi}_{k-1})^{\mathsf{T}} + K\left(\hat{\nabla}\Phi_{p,k} - \frac{\partial\Phi}{\partial\mathbf{u}}(\mathbf{u}_k)\right). \tag{50}$$

The overall dual modifier-adaptation algorithm comprises the three steps given next.

Algorithm 1. (Dual MA with Guaranteed Feasibility)

1) *Initialization.* Set the values of the constants

δ_{g_i}, $i = 1,\ldots,n_g$.	
d_{g_i}, $i = 1,\ldots,n_g$.	Used to compute $E_{i,k}$.
d^L, δ^L, E^U	Used to define the duality constraint.
β_i, $i = 1,\ldots,n_g$.	Used to obtain constraint upper bounds.
K	Filter gain.

2) *Initial points.* Start from the initial feasible (conservative) point $\mathbf{u}_0$. Perturb each input individually around $\mathbf{u}_0$:

$$\mathbf{u}_{-j} = \mathbf{u}_0 + h\,\mathbf{a}_j, \quad j = 1,\ldots,n_u. \tag{51}$$

Apply the perturbed points $\mathbf{u}_{-1},\ldots,\mathbf{u}_{-n_u}$ to the plant. Here, h is the step size and $\mathbf{a}_j$ is the jth unit vector. Select h such that $E^L_{-1}(\mathbf{u}_0) \le E^U$. Set $\boldsymbol{\lambda}^{\Phi}_{-1} = \mathbf{0}$.

3) *Iterations.* For $k = 0, 1, 2, \ldots$
   a) Obtain the gradient estimates $\hat{\nabla}\Phi_{p,k}$, $\hat{\nabla} G_{p,i,k}$ from (8) and (9).
   b) Evaluate $E_{i,k}$, $i = 1,\ldots,n_g$ from (16).
   c) Compute the constraint biases $\varepsilon_{i,k} = \tilde{G}_{i,k} - G_i(\mathbf{u}_k)$, $i = 1,\ldots,n_g$.
   d) Compute the constraint gradient modifiers from (20b) and the filtered cost gradient modifier from (50).
   e) Solve Problem (49) and apply the solution $\mathbf{u}_{k+1}$ to the plant.

Table 1
Prices of the different species.

Prices	P_P	P_E	P_A	P_B
Scenario 1	1073.25	25.92	94.18	95
Scenario 2	1043.38	20.92	79.23	118.34

Remark 5. The approach presented in this work does not guarantee feasibility of the perturbed points (51) used for initialization. In general, the perturbed points can be guaranteed to be feasible if $\mathbf{u}_0$ is sufficiently conservative in terms of the sizes of its constraint back-offs. Recently, an approach was proposed for deriving constraint back-off sizes that guarantee that one can always perturb in a ball of a given radius without violating the constraints (Bunin, 2016).

4. Case study – Williams-Otto reactor

In this section, we apply the dual modifier-adaptation scheme with guaranteed feasibility to the Williams-Otto reactor (Williams and Otto, 1960). The reactor is a continuous stirred tank reactor with the following three reactions occurring in the plant (simulated reality):

$$A + B \rightarrow C,$$
$$B + C \rightarrow P + E,$$
$$C + P \rightarrow G.$$

The reactants A and B are fed with the mass flowrates F_A and F_B, respectively. The desired products P and E are formed along with the intermediate product C and the undesired byproduct G. The mass flowrate of the product is $F = F_A + F_B$. Since it is assumed that the reaction scheme is not well known, the following two reactions have been proposed for the model (Forbes et al., 1994):

$$A + 2B \rightarrow P + E,$$
$$A + B + P \rightarrow G.$$

The material balance equations for the plant and the model can be found in Zhang and Forbes (2000). The optimization problem consists in maximizing the steady-state profit subject to an upper limit on the concentration of the undesired by-product G, and is taken from the work of Marchetti (2013). To further maintain the purity of the product, an additional constraint on the concentration of A is considered. The decision variables are the flowrate F_B and the reactor temperature T_R. The optimization problem then reads:

$$\begin{aligned} &\max_{F_B, T_R} \quad \phi_p = P_P X_P F + P_E X_E F - P_A F_A - P_B F_B, \\ &\text{s.t.} \qquad\quad F_B \in [4, 7], \quad T_R \in [70, 100], \\ &X_A \leq 0.12, \\ &X_G \leq 0.08, \end{aligned} \tag{52}$$

where X_i and P_i are the concentration and price of Species i. The concentrations are functions of the decision variables (inputs) F_B and T_R. The flowrate of reactant A is constant at $F_A = 1.8275$ kg/s. The above optimization problem is solved for the two different price scenarios given in Table 1. In Scenario 1, the constraint on X_G is active at the plant optimum, while, in Scenario 2, both constraints on X_A and X_G are active at the plant optimum.

We shall compare the performances of the following modifier-adaptation schemes:

- *Scheme 1:* Standard dual modifier-adaptation scheme (19), which does not include feasibility guarantees.

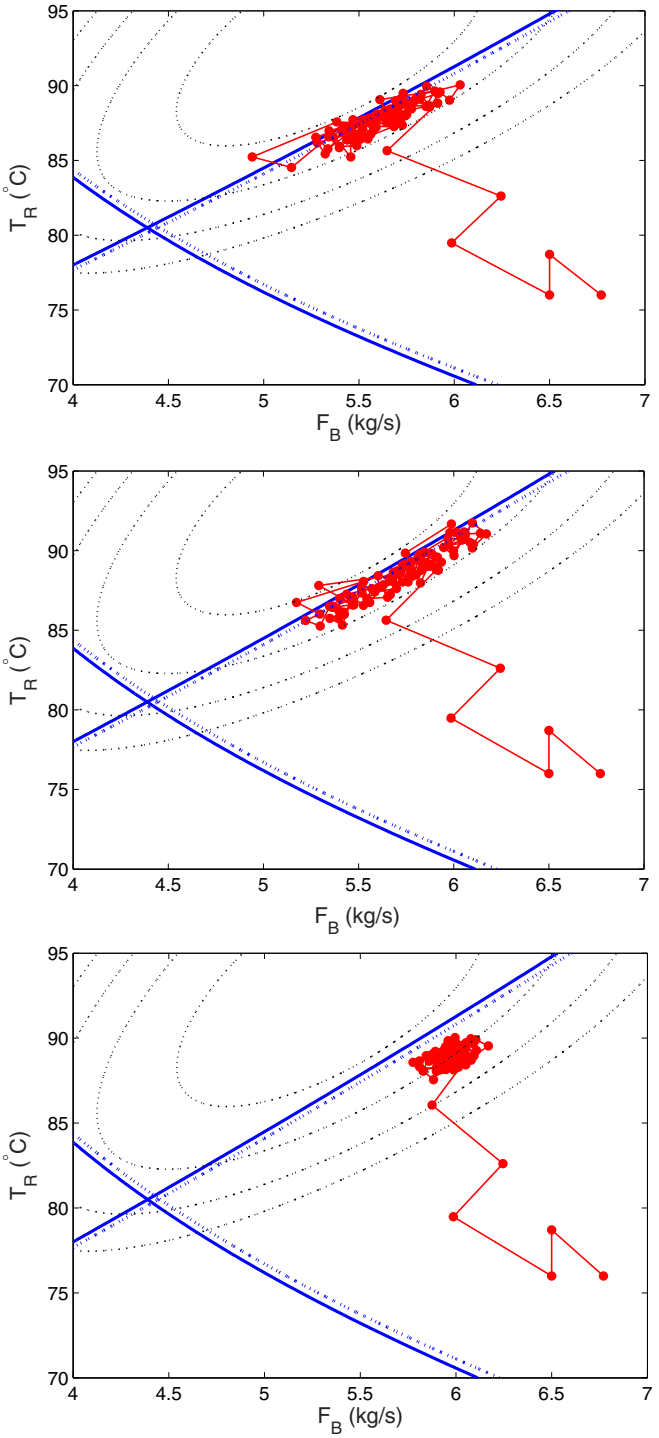

Fig. 1. First 100 RTO iterations for the price Scenario 1. Top plot: Scheme 1. Middle plot: Scheme 2. Bottom plot: Scheme 3. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

- *Scheme 2:* Dual modifier-adaptation scheme (49) that uses the constraint upper bound $G^{\mathrm{U}}_{i,k}(\mathbf{u}) \leq 0$ instead of (49b). Feasibility is guaranteed assuming perfect knowledge of the plant gradients.
- *Scheme 3:* Robust dual modifier-adaptation scheme (49). Feasibility is guaranteed in the presence of gradient uncertainty.

Following the work of Marchetti (2013), the measurements of the plant profit ϕ_p and the concentrations X_A, X_G are subject to zero-mean Gaussian noise with standard deviations $\sigma_\phi = 0.5$ and

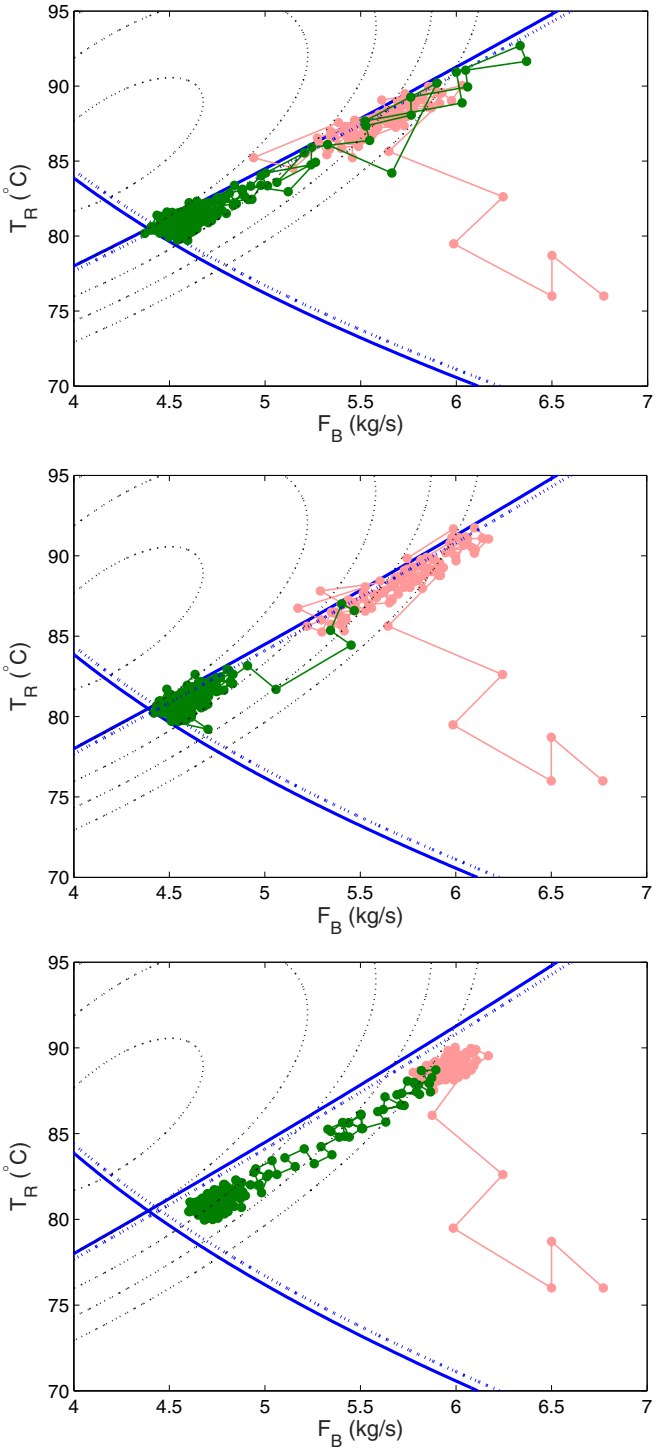

Fig. 2. Next 200 RTO iterations for the price Scenario 2. Top plot: Scheme 1. Middle plot: Scheme 2. Bottom plot: Scheme 3. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

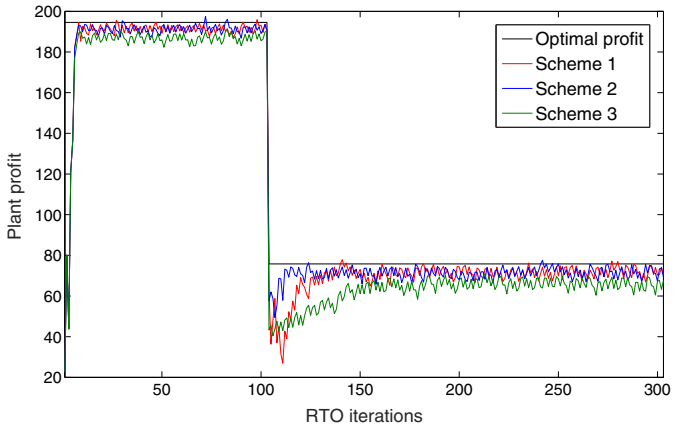

Fig. 3. Evolution of the plant profit for the different dual MA schemes with $E^{U} = 155$.

$\sigma_g = \sigma_{X_A} = \sigma_{X_G} = 0.0005$, respectively. The inputs are centered and scaled to belong to the interval [0, 1] for $F_B \in [4, 7]$ and $T_R \in [70, 100]$.

To compute the duality constraint from (17), we use the same parameter values as Marchetti (2013), that is, $\delta^L = 3.1$ and $d^L = 1085$. We use the upper bound on the Lagrangian gradient error corresponding to $E^{U} = 155$. At each RTO iteration k and for each constraint i, our approach requires computing from (16) the upper bound $E_{i,k}$ on the constraint gradient error, which is then used in the robust upper-bound constraint (45). For this, the noise intervals for the measurements of the concentrations X_A and X_G are selected as $\delta_g = 6\sigma_g = 0.003$. The upper bound on the spectral radius of the Hessian of X_A on the input space is estimated as $d_{X_A} = 0.2326$ for the scaled inputs, and the corresponding value for X_G is $d_{X_G} = 0.3301$. The values of the parameters β_i are set as $\beta_{X_A} = 0.0132$ and $\beta_{X_G} = 0.2349$. These Hessian upper-bound parameters are estimated numerically based on the model equations representing the simulated reality and the model, with the purpose of illustrating the applicability of the theoretical results developed in this work.

A constraint backoff $b_g = 3\sigma_g$ is added to the constraints on the concentrations X_A and X_G to avoid constraint violations due to measurement noise.

Starting from the initial conservative feasible point $\mathbf{u}_0 = [6.5, 76]^T$, simulations are performed for 300 RTO iterations. The first 100 RTO iterations corresponding to the price Scenario 1 are shown in red in Fig. 1, while the next 200 iterations for the new prices of Scenario 2 are shown in green in Fig. 2. Both, Figs. 1 and 2 present the resulting input sequences for Schemes 1, 2 and 3. The contour lines of the plant profit corresponding to Scenarios 1 and 2 are shown in black dotted curves in Figs. 1 and 2, respectively. The blue solid lines in both figures correspond to the limit values of $X_A = 0.12$ and $X_G = 0.08$ for the plant constraints. The blue dotted lines represent the backoff for the constraints on X_A and X_G.

The top plots of Figs. 1 and 2 illustrate that the RTO iterates obtained using Scheme 1, which does not include feasibility guarantees, violate the constraints at many RTO iterations. The middle plots of Figs. 1 and 2 show that Scheme 2, which uses constraint upper bounds that are not robust to gradient uncertainty, still results in constraint violations at many RTO iterations. On the other hand, it can be verified in the bottom plots of Figs. 1 and 2 that the proposed RTO Scheme 3, with robust constraint upper bounds, strictly satisfies the constraints on X_A and X_G at all RTO iterations.

Comparing the first 100 iterations for Scenario 1 (red iterates) in Fig. 1, no noticeable differences are observed in the rate at which the three schemes approach the constraint boundary for X_G when starting at the initial point $\mathbf{u}_0$. We observe that the three schemes approach the constraint boundary in a very similar way. This behaviour can be explained by the fact that when the operating point is far from the constraint boundaries, the iterations are dominated by the duality constraint, and not by the conservatism of the constraint upper bounds.

A change in the prices of the objective function does not affect the feasibility guarantees as the constraints for the plant do not

change. At the RTO iteration 101 the prices are changed to those corresponding to Scenario 2, and the plant optimum moves to the intersection of the constraint boundaries for X_A and X_G, as can be seen if Fig. 2. Hence, the three dual MA schemes start moving towards the new plant optimum by moving alongside the constraint boundary for X_G, which can be observed in the green iterates in Fig. 2. In Fig. 3 we observe that, in this case, the three schemes approach the new plant optimum at different rates. In particular, the proposed Scheme 3 converges much slower than Scheme 1. The difficulty here comes mainly from the fact that we are moving alongside the constraint boundary for X_G. Moving alongside the constraint while guaranteeing not to violate it results in the observed conservatism.

5. Conclusions

This paper has introduced feasibility guarantees in modifier adaptation by using constraint upper-bounding functions. We have shown that, if one combines the usual first-order modification of the model constraints with quadratic upper-bounding terms, it is possible to obtain constraint upper bounds that match the true process constraints up to first order. However, in practice, the gradients are typically estimated with a certain level of uncertainty, and the feasibility guarantees are lost. This paper presented an approach for robustifying the constraint upper bounds by using estimates of the gradient uncertainty region. Note that, in the presented approach, the following two principles hold: (i) the gradient uncertainty region is centered at the estimated gradient, and not at the true gradient, which is unknown; and (ii) the gradient uncertainty region depends on the gradient estimation method. In our case, we estimate the gradients via finite differences using past operating points, and we compute upper bounds on the gradient-error norm to define the gradient uncertainty regions. The effectiveness of the proposed robust constraint bounds was demonstrated in simulation for the optimization of the Williams-Otto reactor using a dual modifier-adaptation algorithm. We have shown that, even in the case of unanticipated changes in the objective function, due for example to changes in the prices, plant feasibility is not lost.

Note that Propositions 1 and 2 constitute theoretical results. In the case study of Section 4, the parameters β_i and d_{g_i} were estimated numerically to illustrate the applicability of the theoretical results developed in this work. Future work should investigate how the upper bounds of the Hessian can be reliably estimated from measured plant data. Furthermore, it would also be interesting to investigate whether online adaptation of the Hessian upper bounds can be used to speed up RTO convergence.

References

Biegler, L.T., Grossmann, I.E., Westerberg, A.W., 1985. A note on approximation techniques used for process optimization. Comp. Chem. Eng. 9 (2), 201–206.

Brdyś, M., Ellis, J.E., Robert, P.D., 1987. Augmented integrated system optimization and parameter estimation technique: derivation, optimality and convergence. IEE Proc.-D 134 (3), 201–209.

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. Imperial College Press, London, UK.

Bunin, G.A., 2016. Constraint back-offs for safe, sufficient excitation: a general theory with application to experimental optimization. Comp. Chem. Eng. 93, 353–360.

Bunin, G.A., François, G., Bonvin, D., 2013. Sufficient conditions for feasibility and optimality of real-time optimization schemes – I. Theoretical foundations, ArXiv:1308.2620.

Bunin, G.A., François, G., Bonvin, D., 2013. Sufficient conditions for feasibility and optimality of real-time optimization schemes – II. Implementation issues, ArXiv:1308.2625.

Bunin, G.A., François, G., Bonvin, D., 2014. On linear and quadratic Lipschitz bounds for twice continuously differentiable functions, ArXiv:1406.3991.

Chen, C.Y., Joseph, B., 1987. On-line optimization using a two-phase approach: an application study. Ind. Eng. Chem. Res. 26, 1924–1930.

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

Forbes, J.F., Marlin, T.E., MacGregor, J.F., 1994. Model adequacy requirements for optimizing plant operations. Comp. Chem. Eng. 18 (6), 497–510.

François, G., Bonvin, D., 2013. Use of convex model approximations for real-time optimization via modifier adaptation. Ind. Eng. Chem. Res. 52, 11614–11625.

Gao, W., Engell, S., 2005. Iterative set-point optimization of batch chromatography. Comp. Chem. Eng. 29, 1401–1409.

Mansour, M., Ellis, J.E., 2003. Comparison of methods for estimating real process derivatives in on-line optimization. Appl. Math. Model. 27, 275–291.

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

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

Marchetti, A.G., 2013. A new dual modifier-adaptation approach for iterative process optimization with inaccurate models. Comp. Chem. Eng. 59, 89–100.

Marchetti, A.G., Luppi, P.A., Basualdo, M.S., 2011. Real-time optimization via modifier adaptation integrated with model predictive control. In: Preprints of the 18th IFAC World Congress, Milano, Italy, pp. 9856–9861.

Navia, D., Martí, R., Sarabia, D., Gutiérred, G., de Prada, C.,2012. Handling infeasibilities in dual modifier-adaptation methodology for real-time optimization. In: Preprints of the 8th IFAC Symposium on Advanced Control of Chemical Processes. Furama Riverfront, Singapore, pp. 537–542.

Quelhas, A.D., de Jesus, N.J.C., Pinto, J.C., 2013. Common vulnerabilities of RTO implementations in real chemical processes. Can. J. Chem. Eng. 91 (April), 652–668.

Roberts, P.D., 1995. Coping with model-reality differences in industrial process optimisation – a review of integrated system optimisation and parameter estimation (ISOPE). Comput. Ind. 26, 281–290.

Singhal, M., Marchetti, A.G., Faulwasser, T., Bonvin, D., 2016. Real-time optimization based on adaptation of surrogate models. In: 11th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems, Trondheim, Norway.

Srinivasan, B., François, G., Bonvin, D., 2011. Comparison of gradient estimation methods for real-time optimization. In: 21st European Symposium on Computer Aided Process Engineering, ESCAPE 21, Chalkidiki, Greece, pp. 607–611.

Tatjewski, P., Brdyś, M.A., Duda, J., 2001. Optimizing control of uncertain plants with constrained feedback controlled outputs. Int. J. Control 74 (15), 1510–1526.

Williams, T.J., Otto, R.E., 1960. A generalized chemical processing model for the investigation of computer control. AIEE Trans. 79, 458.

Yip, W.S., Marlin, T.E., 2004. The effect of model fidelity on real-time optimization performance. Comp. Chem. Eng. 28, 267–280.

Zhang, Y., Forbes, J.F., 2000. Extended design cost, a performance criterion for real-time optimization systems. Comp. Chem. Eng. 24, 1829–1841.