Journal of Process Control 24 (2014) 129–145

Contents lists available at ScienceDirect

Journal of Process Control

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

Steady-state target optimization designs for integrating real-time optimization and model predictive control

A.G. Marchetti [a,*], A. Ferramosca [b], A.H. González [b]

[a] French-Argentine International Center for Information and Systems Sciences (CIFASIS), CONICET-Universidad Nacional de Rosario (UNR), 27 de Febrero 210bis, S2000EZP Rosario, Argentina
[b] Institute of Technological Development for the Chemical Industry (INTEC), CONICET-Universidad Nacional del Litoral (UNL), Güemes 3450, 3000 Santa Fe, Argentina

ARTICLE INFO

Article history:
Received 16 April 2013
Received in revised form 1 November 2013
Accepted 4 November 2013
Available online 14 December 2013

Keywords:
Real-time optimization
Steady-state optimization
Target optimization
Constraint control
Model predictive control

ABSTRACT

In industrial practice, the optimal steady-state operation of continuous-time processes is typically addressed by a control hierarchy involving various layers. Therein, the real-time optimization (RTO) layer computes the optimal operating point based on a nonlinear steady-state model of the plant. The optimal point is implemented by means of the model predictive control (MPC) layer, which typically uses a linear dynamical model of the plant. The MPC layer usually includes two stages: a steady-state target optimization (SSTO) followed by the MPC dynamic regulator. In this work, we consider the integration of RTO with MPC in the presence of plant-model mismatch and constraints, by focusing on the design of the SSTO problem. Three different quadratic program (QP) designs are considered: (i) the standard design that finds steady-state targets that are as close as possible to the RTO setpoints; (ii) a novel optimizing control design that tracks the active constraints and the optimal inputs for the remaining degrees of freedom; and (iii) an improved QP approximation design were the SSTO problem approximates the RTO problem. The main advantage of the strategies (ii) and (iii) is in the improved optimality of the stationary operating points reached by the SSTO-MPC control system. The performance of the different SSTO designs is illustrated in simulation for several case studies.

1. Introduction

Optimization of process operations is continuing to receive attention in the process industries. The goal is to achieve the economic optimal operation of an industrial process in the presence of operating constraints, disturbances, and plant-model mismatch. In real-time optimization (RTO) schemes, optimal process operation is approached by solving the economic optimization problem on-line (or in real time), based on a model of the process. In highly automated plants, RTO is typically implemented within an automation decision hierarchy involving several *layers* (or *levels*), as shown in Fig. 1 [1,2]. There is a time scale separation between the different layers in terms of the frequencies of the disturbances that are rejected and of the decisions that are made at each layer [2]. At the upper layer the *planning and scheduling* addresses long term issues such as production rate targets and raw material allocation. At the lowest layer, basic flow, pressure, and temperature control is implemented, possibly via advanced regulatory controllers. Linear model predictive control (MPC) is widely employed because of its ability to handle large multivariable systems with operating constraints [3,4]. For this reason, the advanced control layer is often called the *MPC layer* [5]. In between the other two layers, the *RTO layer* computes the optimal steady-state operating policy based on a rigorous (first-principle) steady-state model of the process. The operating policy is characterized by setpoints that are passed to the controllers at the lower layer, and/or by values of manipulated variables that are applied directly to the plant.

In response to plant-model mismatch and process disturbances several adaptation strategies have been proposed at the RTO layer. The two-step approach consists in an iteration between parameter estimation and optimization [2,4]. The objective of the parameter estimation step is to find values of selected adjustable model parameters for which the steady-state model gives a good prediction of the measured plant outputs. Next, in the optimization step, the updated model is used to determine a new operating point by solving a model-based optimization problem that typically consists in a nonlinear programming (NLP) problem. As an alternative to the classical two-step approach, the *constraint adaptation* approach does not require updating the model parameters [6,7]. Constraint adaptation uses measurements of the constrained quantities to bias the constraints in the optimization problem. This guarantees that a feasible, yet possibly suboptimal, operating point is reached

* Corresponding author. Tel.: +54 341 4237 248.
E-mail address: marchetti@cifasis-conicet.gov.ar (A.G. Marchetti).

0959-1524/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jprocont.2013.11.004

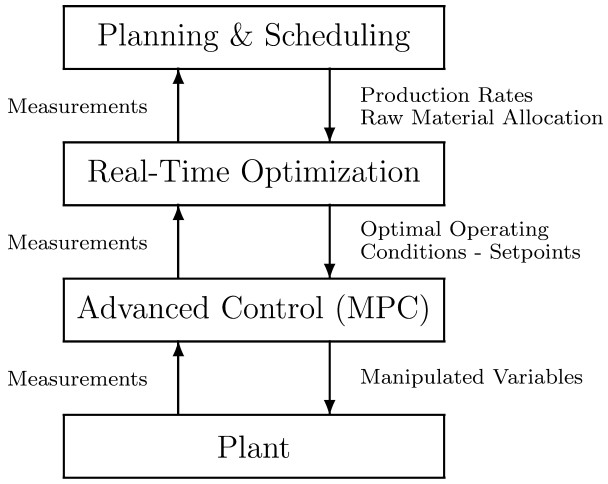

Fig. 1. Multilayer automation decision hierarchy.

upon convergence of the RTO iterations [7]. These steady-state RTO schemes use steady-state data for model adaptation, which requires the online identification of steady state [8,9].

In order to combine RTO with MPC, the optimal input and output setpoints computed at the RTO layer could be passed to the MPC layer as targets in the input and output error terms included in the objective function of the MPC regulator. However, the inconsistency between the steady-state RTO model and the dynamic MPC model may cause the MPC regulator to distribute the final steady-state errors for both the inputs and outputs, resulting in steady-state offsets in the system [10]. Note also that, in the presence of plant-model mismatch and disturbances, the input and output RTO setpoints might not correspond to a feasible operating point for the plant. In order to address these issues, a two-stage approach is usually adopted at the MPC layer. Namely, a *steady-state target optimization* (SSTO) stage, followed by the standard MPC regulator [3,10–13]. The SSTO problem is executed at the same frequency as the MPC problem. It consists of a linear program (LP) or a quadratic program (QP), which results in an LP-MPC or a QP-MPC cascade control system, respectively [10]. The SSTO uses a linear steady-state model, which is typically the steady-state version of the dynamic linear model used in the MPC regulator. Its purpose is to correct the RTO setpoints by computing steady-state targets that are attainable by the MPC regulator. Feedback to the SSTO problem is incorporated by including the estimated output bias in the SSTO model. This way, SSTO can react faster than RTO when disturbances occur. The RTO layer repeatedly modifies the operating point setpoints in response to disturbances and changing operating conditions. Hence, the MPC should be designed to work well in a wide range of operating points and disturbance values, which represents a difficult challenge. For highly nonlinear processes or moderately nonlinear processes with large operating regimes, a nonlinear model-predictive controller (NMPC), or an adaptive linear MPC controller, is required [14]. In other applications that do not exhibit strong nonlinearities in the normal range of operation, a robustly tuned linear MPC may provide satisfactory control action. This has proven to work well in many industrial applications [14].

Kadam and Marquardt [15], pointed out that many of the technologies that have been developed in the last decades in order to solve the operational problems at the different layers of the automation decision hierarchy have been segregated techniques. For instance, at the RTO layer we notice that many RTO algorithms have been studied by analyzing their open-loop properties. The RTO setpoints are assumed to be effectively implemented at each RTO iteration, i.e., without correcting their values by means of the SSTO in the MPC layer. In the RTO literature, the RTO setpoints are typically filtered in order to enforce stability of the open-loop RTO system. Yet, the same kind of filtering might not be required in a closed-loop RTO implementation. On the other hand, at the MPC layer, SSTO-MPC cascade control systems have been studied assuming that the optimal input and output setpoints are provided by the RTO optimizer. The SSTO problem is often formulated so as to drive the steady-state inputs and outputs as close as possible to the RTO setpoints, without selecting which inputs and outputs to track based on the structure of the RTO solution. Additional information from the RTO solution, such as the knowledge of the active constraints, and gradient and Hessian information, could allow the SSTO to improve decisions in between RTO executions. Only a few design approaches where the SSTO problem approximates the NLP problem solved at the RTO layer have been mentioned in the literature [10,5]. Ying and Joseph [10], describe a QP design that is based on the QP approximation used in successive quadratic programming (SQP) approaches for solving NLP problems [16,17]. Surprisingly, the analysis of the properties of this SQP approximation design has received little attention in the MPC literature. Even in [10], the economic SSTO problem in the illustrative case study considered (the control of the heavy oil fractionator in the Shell control problem) consists of an LP problem, which does not cover many of the features of SQP approximations.

The goal of this paper is to study the integration of RTO and MPC in the presence of plant-model mismatch and process disturbances by focusing on the design of the SSTO problem. The RTO and SSTO-MPC problems are formulated and analyzed in a unified framework, and special emphasis is put on how the operating constraints are handled at both automation layers. Three different QP designs of the SSTO problem are investigated: (i) Design A, which considers the standard formulation in which the QP problem finds feasible steady-state targets for the input variables that are as close as possible to the optimal RTO inputs; (ii) Design B, which considers a novel optimizing control design in which the QP problem tracks the active constraints in the RTO solution as well as the optimal inputs along selected input directions; and (iii) Design C, which considers an economic optimization design where the SSTO problem approximates the RTO problem. This work includes several contributions to the RTO/MPC literature: the SSTO Design B is novel (to the authors best knowledge, an optimizing control design of the SSTO problem has not been previously proposed in the literature). An SQP approximation design of the SSTO problem was described in [10], but not further investigated. The SSTO Design C, presented in this paper, considers the SQP approximation design described in [10], including additional novel features that are regarded here as important, such as the approximation of the Lagrangian Hessian (described in Section 3.3.1), and the adaptation of the static gain matrix (described in Section 3.3.2). In addition, the link between Designs B and C is established in Section 3.3.4, and Design C is viewed as a first-order approximation to the solution given by constraint adaptation at the RTO layer. This paper also presents a novel analysis on a necessary condition for matching the RTO optimal point upon convergence of the RTO iterations. This analysis is carried out for each of the three investigated SSTO designs.

The paper is organized as follows. Preliminary results are given in Section 2: the economic optimization problem is formulated as an NLP problem in Section 2.1; RTO via constraint adaptation is reviewed in Section 2.2; Section 2.3 presents the MPC regulator and a general formulation of the SSTO problem; the main steady-state properties of the SSTO-MPC cascade control system are discussed in Section 2.4. Section 3 presents the three different QP designs of the SSTO problem; the performance of the different SSTO designs is tested in simulation in Section 4 by considering three different case studies which comprise: a linear system with first-order dynamics,

a continuous stirred-tank reactor, and a numerical example. Finally, Section 5 concludes the paper.

2. Preliminaries

2.1. Formulation of the optimization problem

The input–output mapping corresponding to the operation of the plant at steady state is represented as $\mathbf{y}_p(\mathbf{u}, \mathbf{d}_p)$, where $\mathbf{u} \in \mathbb{R}^{n_u}$ is the vector of decision (or input) variables, $\mathbf{y}_p \in \mathbb{R}^{n_y}$ is the vector of measured (or output) variables, and $\mathbf{d}_p \in \mathbb{R}^{n_{d_p}}$ is the vector of process disturbances. The notation $(\cdot)_p$ is used throughout for the variables associated with the plant.

In practice, only an (approximate) steady-state model is available:

$$\mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}) = \mathbf{0} \tag{1a}$$

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \boldsymbol{\theta}) \tag{1b}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ are the model state variables, and $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ is a set of model parameters. We denote by $\mathbf{y}(\mathbf{u}, \boldsymbol{\theta})$ the input–output mapping representing the steady-state behavior predicted by the model. In order to obtain $\mathbf{y}(\mathbf{u}, \boldsymbol{\theta})$ one has to first solve the model equations (1a) in order to compute the states $\mathbf{x}$ and then obtain $\mathbf{y}$ by evaluating (1b).

The steady-state economic optimum of the plant is given by the solution of the following optimization problem:

$$\begin{aligned} \mathbf{u}_p^\star = \underset{\mathbf{u}}{\operatorname{argmin}} \quad & \Phi_p(\mathbf{u}, \mathbf{d}_p) := \phi(\mathbf{u}, \mathbf{y}_p(\mathbf{u}, \mathbf{d}_p)) \\ \text{s.t.} \quad & \mathbf{y}_p^{\text{eq}}(\mathbf{u}, \mathbf{d}_p) = \mathbf{y}^{\text{S}}, \\ & \mathbf{y}^{\text{L}} \le \mathbf{y}_p^{\text{in}}(\mathbf{u}, \mathbf{d}_p) \le \mathbf{y}^{\text{U}}, \\ & \mathbf{u}^{\text{L}} \le \mathbf{u} \le \mathbf{u}^{\text{U}}, \end{aligned} \tag{2}$$

where $\phi : \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}$ is the scalar cost function to be minimized; $\mathbf{y}_p^{\text{eq}} \in \mathbb{R}^{n_y^{\text{eq}}}$ is the set of equality constrained outputs for which $\mathbf{y}^{\text{S}}$ are the setpoint values; $\mathbf{y}_p^{\text{in}} \in \mathbb{R}^{n_y^{\text{in}}}$ is the set of inequality constrained outputs for which $\mathbf{y}^{\text{L}}$ and $\mathbf{y}^{\text{U}}$ are the lower and upper bounds, respectively; and $\mathbf{u}^{\text{L}}$ and $\mathbf{u}^{\text{U}}$ are the lower and upper bounds on the decision variables. The constrained outputs $\mathbf{y}_p^{\text{eq}}$ and $\mathbf{y}_p^{\text{in}}$ are subsets of the measured outputs $\mathbf{y}_p$.

In this paper, we limit our study to the case where the following assumptions hold:

Assumption 1. It is assumed that $n_y^{\text{eq}} < n_u$, i.e., the number of equality constrained outputs is lower than the number of manipulated inputs.

Assumption 2. There exists a feasible solution to Problem (2) for $\mathbf{d}_p \in \mathcal{D}$, where the disturbance set $\mathcal{D}$ comprises the disturbances that may be encountered during the operation of the plant.

Assumption 1 is included so that there are degrees of freedom in Problem (2). If $n_y^{\text{eq}} > n_u$ then there will be in general no feasible solution to Problem (2). On the other hand, the case $n_y^{\text{eq}} = n_u$ is not excluded for technical reasons, but because it is not interesting in the framework of this paper. Notice that, these two assumptions exclude the possibility of not having a feasible solution to Problem (2).

The plant mapping $\mathbf{y}_p(\mathbf{u}, \mathbf{d}_p)$ is not known accurately, and only the approximate model $\mathbf{y}(\mathbf{u}, \boldsymbol{\theta})$, is available. Using the model, the solution of the original problem (2) can be approached by solving the following NLP problem:

$$\begin{aligned} \mathbf{u}^\star = \underset{\mathbf{u}}{\operatorname{argmin}} \quad & \Phi(\mathbf{u}, \boldsymbol{\theta}) := \phi(\mathbf{u}, \mathbf{y}(\mathbf{u}, \boldsymbol{\theta})) \\ \text{s.t.} \quad & \mathbf{y}^{\text{eq}}(\mathbf{u}, \boldsymbol{\theta}) = \mathbf{y}^{\text{S}}, \\ & \mathbf{y}^{\text{L}} \le \mathbf{y}^{\text{in}}(\mathbf{u}, \boldsymbol{\theta}) \le \mathbf{y}^{\text{U}}, \\ & \mathbf{u}^{\text{L}} \le \mathbf{u} \le \mathbf{u}^{\text{U}}. \end{aligned} \tag{3}$$

Due to plant-model mismatch and unmeasured disturbances, the solution to Problem (3) might be neither optimal nor feasible for the plant. Hence, the need for some kind of adaptation.

2.2. Real-time optimization layer

At the RTO layer, Problem (3) is updated based on measurements. The classical adaptation strategy consists in a two-step approach involving an iteration between parameter estimation and optimization [2,4]. The idea is to estimate repeatedly the model parameters $\boldsymbol{\theta}$ of the nonlinear steady-state model, and to use the updated model in the model optimization to generate new inputs. This way, the model is expected to represent the plant at its current operating point more accurately.

Let us denote by $\mathbf{u}_k$ the steady-state operating point applied to the plant at the kth RTO iteration, and by $\boldsymbol{\theta}_k$ the parameter estimates obtained at $\mathbf{u}_k$. A desirable property of the model-update scheme is that the predicted constrained outputs match the corresponding plant measurements at each RTO execution:

$$\mathbf{y}^{\text{eq}}(\mathbf{u}_k, \boldsymbol{\theta}_k) = \mathbf{y}_p^{\text{eq}}(\mathbf{u}_k, \mathbf{d}_p), \tag{4}$$

$$\mathbf{y}^{\text{in}}(\mathbf{u}_k, \boldsymbol{\theta}_k) = \mathbf{y}_p^{\text{in}}(\mathbf{u}_k, \mathbf{d}_p). \tag{5}$$

When the RTO results are implemented in open loop, matching the constrained outputs guarantees that a feasible operating point is reached upon convergence. However, the satisfaction of (4) and (5) will depend on the flexibility of the selected model parameterization [18].

It is also possible to satisfy (4) and (5) without updating the model parameters, by simply biasing the constraint values in the optimization problem. Assuming that measurements are available for every constrained quantity, the following plant-model bias terms can be computed at $\mathbf{u}_k$:

$$\boldsymbol{\varepsilon}_k^{\text{eq}} := \mathbf{y}_p^{\text{eq}}(\mathbf{u}_k, \mathbf{d}_p) - \mathbf{y}^{\text{eq}}(\mathbf{u}_k, \boldsymbol{\theta}), \tag{6}$$

$$\boldsymbol{\varepsilon}_k^{\text{in}} := \mathbf{y}_p^{\text{in}}(\mathbf{u}_k, \mathbf{d}_p) - \mathbf{y}^{\text{in}}(\mathbf{u}_k, \boldsymbol{\theta}). \tag{7}$$

The constraint-adaptation approach is to obtain the RTO optimal points by solving an NLP problem similar to (3), which includes the bias constraint corrections [6,7]:

$$\begin{aligned} \mathbf{u}_{k+1}^\star = \underset{\mathbf{u}}{\operatorname{argmin}} \quad & \Phi(\mathbf{u}, \boldsymbol{\theta}) := \phi(\mathbf{u}, \mathbf{y}(\mathbf{u}, \boldsymbol{\theta})) \\ \text{s.t.} \quad & \mathbf{y}^{\text{eq}}(\mathbf{u}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k^{\text{eq}} = \mathbf{y}^{\text{S}}, \\ & \mathbf{y}_k^{\text{L}} \le \mathbf{y}^{\text{in}}(\mathbf{u}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k^{\text{in}} \le \mathbf{y}_k^{\text{U}}, \\ & \mathbf{u}_k^{\text{L}} \le \mathbf{u} \le \mathbf{u}_k^{\text{U}}. \end{aligned} \tag{8}$$

The bounds on the decision variables, $\mathbf{u}_k^{\text{L}}$ and $\mathbf{u}_k^{\text{U}}$, can be varied in order to limit the input range for the next operating point:

$$u_{i,k}^{\text{L}} = \max\{u_{i,k} - \Delta u_i,\ u_i^{\text{L}}\}, \quad i = 1, \ldots, n_u, \tag{9}$$

$$u_{i,k}^{\text{U}} = \min\{u_{i,k} + \Delta u_i,\ u_i^{\text{U}}\}, \quad i = 1, \ldots, n_u, \tag{10}$$

where $\Delta u_i > 0, i = 1, \ldots, n_u$ are the upper bounds on the input moves. The bounds on the inequality constrained outputs are varied in a similar way, based on their current measured values:

$$y_{j,k}^{\text{L}} = \max\{y_{p,j}^{\text{in}}(\mathbf{u}_k, \mathbf{d}_p) - \Delta y_j^{\text{in}},\ y_j^{\text{L}}\}, \quad j = 1, \ldots, n_y^{\text{in}}, \tag{11}$$

$$y^{U}_{j,k} = \min\{y^{in}_{p,j}(\mathbf{u}_k, \mathbf{d}_p) + \Delta y^{in}_j,\ y^{U}_j\}, \quad j = 1, \ldots, n^{in}_y, \tag{12}$$

where $\Delta y^{in}_j > 0$, $j = 1, \ldots, n^{in}_y$, are fixed increment values. Upon convergence of the RTO algorithm, the active constraints of the RTO problem reach their actual constraint bounds.

Implementation of the full biases in (6), (7) may lead to excessive correction, thereby compromising the convergence of the algorithm [7]. In some cases, a better strategy consists in filtering the constraint biases, using for example a first-order exponential filter [7]:

$$\boldsymbol{\varepsilon}^{eq}_k = (\mathbf{I} - \mathsf{K}^{eq})\boldsymbol{\varepsilon}^{eq}_{k-1} + \mathsf{K}^{eq}\left(\mathbf{y}^{eq}_p(\mathbf{u}_k, \mathbf{d}_p) - \mathbf{y}^{eq}(\mathbf{u}_k, \boldsymbol{\theta})\right), \tag{13}$$

$$\boldsymbol{\varepsilon}^{in}_k = (\mathbf{I} - \mathsf{K}^{in})\boldsymbol{\varepsilon}^{in}_{k-1} + \mathsf{K}^{in}\left(\mathbf{y}^{in}_p(\mathbf{u}_k, \mathbf{d}_p) - \mathbf{y}^{in}(\mathbf{u}_k, \boldsymbol{\theta})\right), \tag{14}$$

where K^{eq} and K^{in} are diagonal gain matrices with gain entries in $(0, 1]$.

Remark 1. In this work, the integration of RTO with MPC is developed using constraint adaptation at the RTO layer because of the following reasons: (i) constraint adaptation uses a simple adaptation strategy; (ii) constraint adaptation satisfies (4) and (5) by biasing the constraint values in the optimization problem. This will prove to be important in order for the closed-loop RTO system to reach the RTO optimum point upon convergence of the RTO iterations; (iii) A link will be established between constraint adaptation and one of the SSTO designs that will be presented in this work.

If the RTO results are implemented in open loop, the new operating point is obtained by applying the optimal inputs $\mathbf{u}^{\star}_{k+1}$ directly to the plant:

$$\mathbf{u}_{k+1} = \mathbf{u}^{\star}_{k+1}.$$

In this case, the constraint-adaptation algorithm guarantees feasibility upon convergence and the filter on the constraint biases can be used to enforce stability [7]. Yet, the constraints might be violated at the RTO points prior to convergence.

In practice, RTO results are seldom applied to the plant in open-loop fashion. Instead, they are usually implemented by means of an MPC regulator that takes care not to violate the constraints. In this case, the values $\mathbf{u}_{k+1}$ to be used in the next RTO iteration correspond to the input values reached by the controlled plant at steady state.

2.3. Model predictive control layer

We now go down in the general control structure shown in Fig. 1 and concentrate on the MPC layer. The MPC regulator should be designed with zero offset. A formulation of the MPC regulator is presented, as well as a general formulation of the SSTO problem.

2.3.1. MPC regulator

The prediction model used by the model predictive controller is a linear, time invariant, discrete-time system:

$$\boldsymbol{\delta}\mathbf{x}(t+1) = \mathsf{A}\,\boldsymbol{\delta}\mathbf{x}(t) + \mathsf{B}\,\boldsymbol{\delta}\mathbf{u}(t) \tag{15}$$

$$\boldsymbol{\delta}\mathbf{y}(t) = \mathsf{C}\,\boldsymbol{\delta}\mathbf{x}(t) + \hat{\mathbf{d}}(t) \tag{16}$$

The linearization is carried out at a steady state point ($\mathbf{u}_0$, $\mathbf{x}_0$, $\mathbf{y}_0$), and deviation variables are used in (15) and (16). The disturbance estimate $\hat{\mathbf{d}}(t)$ is used to add integral control, which is required for offset-free MPC. Implementing offset-free MPC typically involves augmenting the prediction model with a constant output disturbance model, and estimating the augmented state using a state observer. Further details on this topic can be found in [19,20].

The integration of RTO with MPC considered in this paper relies on the following assumption:

Assumption 3 *(Validity of Linear MPC).* It is assumed that a linear MPC regulator based on the system (15), (16) is valid for all the different operating regimes that are targeted by the upper layer RTO problem.

This assumption limits the applicability of the results presented in this paper. However, notice that the use of linear MPC in conjunction with nonlinear steady-state RTO is current industrial practice, and has shown to be adequate for many industrial control problems [14]. In particular, the two-layer RTO approach with linear MPC may be adequate for the case of a linear (or approximately linear) process with a nonlinear economic cost function.

At the current time t, MPC predicts the behavior of the process over N future time steps and determines the optimal sequence of N input moves $\mathbf{U} = \{\mathbf{u}(t|t), \mathbf{u}(t+1|t), \ldots, \mathbf{u}(t+N-1|t)\}$, that minimizes a given objective function. The parameter N denotes the prediction horizon. Following the receding horizon policy, only the first element of the optimal control sequence, $\mathbf{u}(t|t)$, is implemented, and at time $t+1$ the computation is repeated, moving the prediction window one step ahead. The MPC objective function to be minimized is usually given by the tracking error of the predicted trajectories to given setpoints. These setpoints are the steady-state targets $\mathbf{u}_s(t)$ and $\mathbf{y}_s(t)$, which are determined at each sampling time t by solving the SSTO problem (see Section 2.3.2). A general formulation of the MPC problem reads[1]:

$$\min_{\mathbf{U}} \quad J_t = \sum_{l=0}^{N-1}\left(\left\|\mathbf{y}(t+l|t) - \mathbf{y}_s(t)\right\|^2_{\mathsf{Q}} + \left\|\mathbf{u}(t+l|t) - \mathbf{u}_s(t)\right\|^2_{\mathsf{R}}\right) \tag{17a}$$

s.t.

$$\boldsymbol{\delta}\mathbf{x}(t+l+1|t) = \mathsf{A}\,\boldsymbol{\delta}\mathbf{x}(t+l|t) + \mathsf{B}\,\boldsymbol{\delta}\mathbf{u}(t+l|t), \quad l = 0, \ldots, N-1, \tag{17b}$$

$$\mathbf{y}(t+l|t) = \mathsf{C}\,\boldsymbol{\delta}\mathbf{x}(t+l|t) + \mathbf{y}_0 + \hat{\mathbf{d}}(t|t), \quad l = 1, \ldots, N, \tag{17c}$$

$$\mathbf{u}^{L}_k \le \mathbf{u}(t+l|t) \le \mathbf{u}^{U}_k, \quad l = 0, \ldots, N-1, \tag{17d}$$

$$\boldsymbol{\delta}\mathbf{x}(t+N|t) = \boldsymbol{\delta}\mathbf{x}_s(t), \tag{17e}$$

where $\mathbf{y}(t+l|t) = \boldsymbol{\delta}\,\mathbf{y}(t+l|t) + \mathbf{y}_0$ are the model predictions of the output variables at time $t+l$ based on information available at time t; $\mathbf{u}(t+l|t) = \boldsymbol{\delta}\mathbf{u}(t+l|t) + \mathbf{u}_0$ is the control sequence to be computed; Q is a positive semidefinite weighting matrix on the outputs; R is a positive definite weighting matrix on the inputs; the input bounds are included in (17d); $\boldsymbol{\delta}\mathbf{x}(t+N|t)$ are the state deviations predicted at time $t+N$ based on information available at time t, and $\boldsymbol{\delta}\mathbf{x}_s(t) = (\mathbf{I} - \mathsf{A})^{-1}\mathsf{B}\boldsymbol{\delta}\mathbf{u}_s(t)$. Notice also that, the terminal equality constraint (17e) is included for stability reasons. However, several alternative MPC designs can be proposed to ensure closed-loop stability and recursive feasibility. They mainly depend on the choice of the terminal constraint, and eventually on the choice of a terminal penalty term in the cost function [21].

2.3.2. Steady state target optimization

The SSTO problem uses a steady-state model that is consistent with the dynamical model used by the MPC regulator,

$$\boldsymbol{\delta}\mathbf{y}_s = \mathsf{A}_s\boldsymbol{\delta}\mathbf{u}_s + \hat{\mathbf{d}}(t|t), \tag{18}$$

$$\mathsf{A}_s = \mathsf{C}(\mathbf{I} - \mathsf{A})^{-1}\mathsf{B}, \tag{19}$$

[1] The notation $\|\mathbf{v}\|^2_{\mathsf{M}} := \mathbf{v}^{\mathsf{T}}\mathsf{M}\mathbf{v}$ is used thereafter.

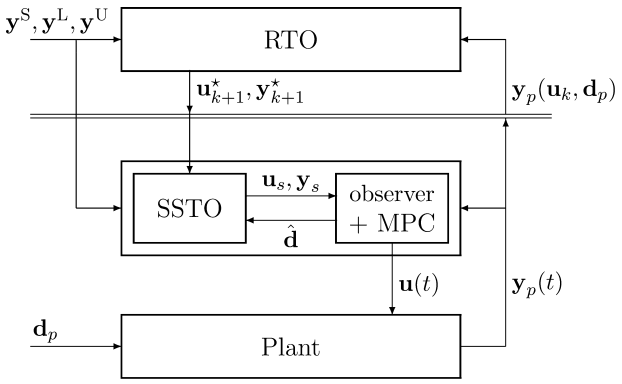

Fig. 2. Integration of RTO with MPC using SSTO.

where $\delta\mathbf{u}_s$ and $\delta\mathbf{y}_s$ are the steady-state input and output deviation variables, respectively, and $\hat{\mathbf{d}}(t|t)$ is the current estimate of the output disturbance.

Let $\mathbf{u}^\star_{k+1}$ be the optimal inputs determined at the kth RTO execution, which takes place when the plant is at the steady-state operating point $\mathbf{u}_k$, and let $\mathbf{y}^\star_{k+1}$ be the optimal outputs predicted by the updated model, that is, $\mathbf{y}^\star_{k+1} := \mathbf{y}(\mathbf{u}^\star_{k+1}, \boldsymbol{\theta}_k)$, if the classical two-step approach is applied, while $\mathbf{y}^\star_{k+1} := \mathbf{y}(\mathbf{u}^\star_{k+1}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k$, if the constraint-adaptation problem (8) is applied. The SSTO problem is updated based on this optimal solution, and the SSTO-MPC control system is implemented until the next RTO execution takes place at the new steady-state point $\mathbf{u}_{k+1}$. A general formulation of the SSTO problem is given by the following QP problem [10]:

$$\min_{\mathbf{u}_s, \mathbf{y}_s} \quad \left\|\mathbf{y}_s - \mathbf{y}^\star_{k+1}\right\|^2_{\mathsf{Q}_s} + \left\|\mathbf{u}_s - \mathbf{u}^\star_{k+1}\right\|^2_{\mathsf{R}_s} + \mathsf{q}_s^\mathsf{T}(\mathbf{y}_s - \mathbf{y}^\star_{k+1}) + \mathsf{r}_s^\mathsf{T}(\mathbf{u}_s - \mathbf{u}^\star_{k+1}) \tag{20a}$$

s.t.

$$\mathbf{y}_s = \mathsf{A}_s(\mathbf{u}_s - \mathbf{u}_0) + \mathbf{y}_0 + \hat{\mathbf{d}}(t|t), \tag{20b}$$

$$\mathbf{y}_s^{\text{eq}} = \mathbf{y}^S, \tag{20c}$$

$$\mathbf{y}^L_k \le \mathbf{y}_s^{\text{in}} \le \mathbf{y}^U_k, \tag{20d}$$

$$\mathbf{u}^L_k \le \mathbf{u}_s \le \mathbf{u}^U_k, \tag{20e}$$

where Q_s is a positive semidefinite weighting matrix on the outputs and R_s is a positive semidefinite (or positive definite) weighting matrix on the inputs. The overall automation system hierarchy is illustrated in Fig. 2.

Notice that the constraints on the output variables (20c) and (20d) have been included in the SSTO problem but not in the MPC controller. This way, constraint violations are allowed during the transients but not at the steady-state points reached by the controlled plant. In case it becomes critical to minimize constraint violations during the transients, it is possible to include the output constraints in the MPC regulator as well. However, this would require the use of a soft-constraint approach in order to avoid possible infeasible solutions to the MPC problem [22]. This approach consists in adding a penalty term to the objective function of the MPC problem that penalizes a measure of constraint violation. The inclusion of the output constraints in the MPC regulator may also degrade the performance of the controller. Several commercially available MPC algorithms also provide soft constraints in the SSTO problem [3]. This is because if a large disturbance enters the process, it may not be possible, given the available input space, to completely remove the disturbance at steady state. However, since relaxing the output constraints in the SSTO will result in steady-state targets that continuously violate the output constraints [13], we choose not to relax them, and in agreement with [13], we suggest using an infeasible solution as an indicator of a process exception.

2.4. Steady-state properties of the SSTO-MPC control system

The SSTO stage confers important steady-state properties to the controlled plant. These include offset-free behavior of the MPC regulator, and guaranteed feasibility at steady-state operation.

2.4.1. Zero steady-state offset

Lemma 1. *In the presence of stable step-wise disturbances, the controlled plant will reach a steady-state operating point $\bar{\mathbf{u}}$ with no offsets with respect to the SSTO targets [10], i.e., with $\bar{\mathbf{u}} = \mathbf{u}_s$, and $\mathbf{y}_p(\bar{\mathbf{u}}, \mathbf{d}_p) = \mathbf{y}_s$.*

Proof. At a steady-state operating point $\bar{\mathbf{u}}$ the output disturbance $\hat{\mathbf{d}}(t|t)$ takes the value $\bar{\mathbf{d}}$ given below:

$$\bar{\mathbf{d}} = \mathbf{y}_p(\bar{\mathbf{u}}, \mathbf{d}_p) - \mathbf{y}_0 - \mathsf{C}\,\delta\bar{\mathbf{x}}, \tag{21}$$

Also, the constraints (17b) and (17c) in the MPC problem reduce to:

$$\delta\bar{\mathbf{x}} = \mathsf{A}\,\delta\bar{\mathbf{x}} + \mathsf{B}\,\delta\bar{\mathbf{u}} \tag{22}$$

$$\bar{\mathbf{y}} = \mathsf{C}\,\delta\bar{\mathbf{x}} + \bar{\mathbf{d}} + \mathbf{y}_0. \tag{23}$$

Using (21) in (23) we have

$$\bar{\mathbf{y}} = \mathbf{y}_p(\bar{\mathbf{u}}, \mathbf{d}_p). \tag{24}$$

Eqs. (22) and (22) can be written jointly as $\bar{\mathbf{y}} = \mathsf{A}_s(\bar{\mathbf{u}} - \mathbf{u}_0) + \bar{\mathbf{d}} + \mathbf{y}_0$, which is identical to the constraint (20b) in the SSTO problem with $\hat{\mathbf{d}}(t|t) = \bar{\mathbf{d}}$.

Since the solution $(\mathbf{u}_s, \mathbf{y}_s)$ of the SSTO problem satisfies the constraints (23) and (17d) of the MPC problem, the targets $\mathbf{u}_s$ and $\mathbf{y}_s$ are feasible for the MPC problem. Therefore, the solution to the MPC problem is $\bar{\mathbf{u}} = \mathbf{u}_s$, and $\bar{\mathbf{y}} = \mathbf{y}_s$ since this solution is feasible and it minimizes the MPC quadratic objective function in (17a), which becomes equal to zero. □

In order to fully specify the operating point, the number of linearly independent input and output targets that are included in the objective function of the MPC regulator should be equal to the total number of input variables, n_u. However, since the zero-offset condition is independent of the number of targets included in (17a), it is possible to exceed n_u targets without incurring in offset. In fact, one can include in (17a) targets for all the input and all the output variables. Zero steady-state offset results from the fact that there is no model mismatch between the SSTO and MPC stages.

2.4.2. Feasibility

An important property of the SSTO-MPC control system is that the controlled plant will reach a feasible steady-state operating point, i.e., a point that satisfies the constraints in Problem (2).

Lemma 2. *In the presence of stable step-wise disturbances, the SSTO-MPC control system will reach a steady-state operating point $\bar{\mathbf{u}}$ that is a feasible point for the plant.*

Proof. The proof follows directly from Lemma 1 and the fact that $\mathbf{u}_s$, $\mathbf{y}_s$ satisfy the output constraints (20c) and (20d), and the input bounds. □

3. Alternative designs of the SSTO problem

The SSTO-MPC control system presented in Section 2.3 permits to implement RTO results while paying attention to the feasibility of the steady-state operating points reached by the controlled plant. However, which feasible steady-state is actually reached will depend on the design of the SSTO problem. In this section, we describe three different QP design approaches that vary according to the formulation of the QP problem and the information that is passed from the RTO solution.

3.1. Design approach A

We present first the standard SSTO design where the QP problem finds feasible steady-state targets for the input variables that are as close as possible (in a least-square sense) to the optimal RTO inputs. In this approach, which is adapted from [11,12], the QP problem reads:

$$\begin{aligned} \mathbf{u}_s(t), \mathbf{y}_s(t) = \underset{\mathbf{u}_s, \mathbf{y}_s}{\operatorname{argmin}} \quad & \left\| \mathbf{u}_s - \mathbf{u}^{\star}_{k+1} \right\|^2_{\mathsf{R}_s} \\ \text{s.t.} \quad & \mathbf{y}_s = \mathsf{A}_s(\mathbf{u}_s - \mathbf{u}_0) + \mathbf{y}_0 + \hat{\mathbf{d}}(t|t), \\ & \mathbf{y}^{\text{eq}}_s = \mathbf{y}^{\text{S}}, \\ & \mathbf{y}^{\text{L}}_k \le \mathbf{y}^{\text{in}}_s \le \mathbf{y}^{\text{U}}_k, \\ & \mathbf{u}^{\text{L}}_k \le \mathbf{u}_s \le \mathbf{u}^{\text{U}}_k, \end{aligned} \tag{25}$$

where R_s is a positive definite weighting matrix.

3.1.1. Matching the RTO optimal point upon convergence

Notice that, the steady-state inputs $\mathbf{u}_{k+1}$ reached by the SSTO-MPC system will not in general match the RTO optimal inputs $\mathbf{u}^{\star}_{k+1}$. In particular, using any SSTO design, $\mathbf{u}_{k+1}$ will not match $\mathbf{u}^{\star}_{k+1}$ whenever $\mathbf{u}^{\star}_{k+1}$ is infeasible for the plant (e.g., due to plant-model mismatch). Yet, a desirable property of the overall RTO/SSTO-MPC two-layer system is that, upon convergence of the RTO algorithm, the steady-state operating point $\mathbf{u}_{\infty} = \lim_{k\to\infty} \mathbf{u}_k$ reached by the controlled plant will match the RTO solution $\mathbf{u}^{\star}_{\infty} = \lim_{k\to\infty} \mathbf{u}^{\star}_k$, i.e., $\mathbf{u}_{\infty} = \mathbf{u}^{\star}_{\infty}$. In this context, let us consider the following uniqueness assumption:

Assumption 4. The RTO problem has a unique solution point $\mathbf{u}^{\star}_{k+1}$ at each RTO execution k.

This assumption implies that $\mathbf{u}^{\star}_{k+1}$ is a strict local minimum. Using this assumption, a convergent RTO implementation will converge to the unique solution point $\mathbf{u}^{\star}_{\infty}$, i.e., the possibility of converging to an optimal solution set (that is not a singleton) is excluded. The following proposition gives a necessary condition that the RTO model adaptation scheme must satisfy in order to satisfy $\mathbf{u}_{\infty} = \mathbf{u}^{\star}_{\infty}$.

Proposition 1. *In order to match the RTO optimal point upon convergence of the overall RTO/SSTO-MPC two-layer system (i.e., to have $\mathbf{u}_{\infty} = \mathbf{u}^{\star}_{\infty}$) it is necessary that the RTO model updated at $\mathbf{u}^{\star}_{\infty}$ predicts with zero offset the equality constrained outputs and the inequality constrained outputs that are active at $\mathbf{u}^{\star}_{\infty}$.*

Proof. From Lemma 1 we know that the SSTO-MPC control system will reach a steady state with $\mathbf{y}_p(\mathbf{u}_{\infty}, \mathbf{d}_p) = \mathbf{y}_s$. For the case of the equality constrained outputs, Lemma 2 implies that $\mathbf{y}^{\text{eq}}_s = \mathbf{y}^{\text{S}}$. Hence, if the RTO model updated at $\mathbf{u}_{\infty}$ does not predict with zero offset the equality constrained output variables, then $\mathbf{u}^{\star}_{\infty}$ cannot be equal to $\mathbf{u}_{\infty}$, because $\mathbf{u}_{\infty}$ will not be a feasible point for the RTO problem. Similarly, if the active equality constrained output variables are not predicted with zero offset, then $\mathbf{u}_{\infty}$ will be either infeasible of suboptimal for the RTO problem, and therefore $\mathbf{u}^{\star}_{\infty}$ will not be equal to $\mathbf{u}_{\infty}$. □

Notice that, any RTO model adaptation scheme that matches the constrained outputs of the plant upon convergence of the RTO iterations will satisfy the necessary condition in Proposition 1. In particular, the constraint-adaptation approach satisfies this condition. Next, let us consider the following definition:

Definition 1. A fixed point of the overall RTO/SSTO-MPC two-layer system is a steady-state operating point that is mapped to itself by the RTO/SSTO-MPC two-layer system.

Let $\overline{\mathbf{u}}^{\star}$ and $\overline{\mathbf{y}}^{\star}$ be the optimal RTO inputs and outputs evaluated when the plant is at steady-state at the operating point $\overline{\mathbf{u}}$. Then, $\overline{\mathbf{u}}$ is a fixed point of the overall RTO/SSTO-MPC system if $\mathbf{u}_s = \overline{\mathbf{u}}$ is the unique solution to the SSTO problem with the RTO setpoints $\overline{\mathbf{u}}^{\star}$, $\overline{\mathbf{y}}^{\star}$, and with the disturbance value $\overline{\mathbf{d}}$ in (21) evaluated at $\overline{\mathbf{u}}$.

A necessary condition for matching the RTO solution upon convergence is that the point $\mathbf{u}^{\star}_{\infty}$ be a fixed point of the overall RTO/SSTO-MPC two-layer system.

Proposition 2. *If constraint adaptation is implemented at the RTO layer, Assumption 4 holds, and Design A is implemented at the SSTO-MPC layer, then $\mathbf{u}_{\infty} = \mathbf{u}^{\star}_{\infty}$ is a fixed point of the overall RTO/SSTO-MPC two-layer system.*

Proof. The condition $\mathbf{u}_{\infty} = \mathbf{u}^{\star}_{\infty}$ implies that $\mathbf{u}^{\star}_{\infty}$ is the solution to the constraint-adaptation problem (8) with constraint biases $\boldsymbol{\epsilon}_{\infty}$ evaluated at $\mathbf{u}^{\star}_{\infty}$. Constraint adaptation guarantees that the necessary condition in Proposition 1 is satisfied. In order for $\mathbf{u}^{\star}_{\infty}$ to be a fixed point of the overall RTO/SSTO-MPC system it is required that $\mathbf{u}_s = \mathbf{u}^{\star}_{\infty}$ be also the unique solution to the QP problem (25) with the disturbance value $\overline{\mathbf{d}}$ in (21) evaluated at $\mathbf{u}^{\star}_{\infty}$. This is the case since $\mathbf{u}^{\star}_{\infty}$ is feasible and R_s is positive definite. □

The SSTO design approach A is simple and easy to understand. However, it does not economically update the targets in response to process disturbances, and it has no optimizing control properties. An optimizing control design of the SSTO problem is presented next.

3.2. Design approach B

In terms of optimality there is often much to win by controlling the active inequality constrained variables to their optimal boundary values [7]. Controlling the active constrained quantities has been widely adopted in implicit process optimization approaches such as *constraint control* [23], *optimizing control* [24–26] and *NCO-tracking* [27]. The idea of the SSTO design described next is to take advantage of the optimizing control properties that stem from controlling the active constraints in the RTO solutions.

Let the vector $\mathbf{z}_s \in \mathbb{R}^{n^a_{k+1}}$ include the targets for all the inequality constrained output and input variables that are active at the optimal RTO point $\mathbf{u}^{\star}_{k+1}$, and let $\mathbf{z}^{\text{S}}_{k+1} \in \mathbb{R}^{n^a_{k+1}}$ be the vector of active boundary values in Problem (8), corresponding to the variables in $\mathbf{z}_s$. We also define $\mathbf{z}(\mathbf{u}, \boldsymbol{\theta}) \in \mathbb{R}^{n^a_{k+1}}$ as the vector of active inequality constrained output and input variables predicted by the RTO model. For example, if the third inequality constrained output in $\mathbf{y}^{\text{in}}$ is active at its lower bound, and the second input variable in $\mathbf{u}$ is active at its upper bound, then we select $\mathbf{z}_s = [y^{\text{in}}_{s,3},\ u_{s,2}]^{\text{T}}$, $\mathbf{z}^{\text{S}}_{k+1} = [y^{\text{L}}_{3,k},\ u^{\text{U}}_{2,k}]^{\text{T}}$, and $\mathbf{z}(\mathbf{u}, \boldsymbol{\theta}) = [y^{\text{in}}_3(\mathbf{u}, \boldsymbol{\theta}),\ u_2]^{\text{T}}$. This way, the active constraints can be controlled by controlling the steady-state targets $\mathbf{z}_s$ to their setpoint values $\mathbf{z}^{\text{S}}_{k+1}$ in the SSTO problem. The active constraints might change from one RTO execution to the other, thus

requiring a change in $\mathbf{z}_s$ and $\mathbf{z}^S_{k+1}$. The SSTO problem is given by the following QP:

$$\min_{\mathbf{u}_s,\mathbf{y}_s} \quad \left\|\mathbf{z}_s - \mathbf{z}^S_{k+1}\right\|^2_{\mathsf{Q}_s} + \left\|\mathbf{u}_s - \mathbf{u}^\star_{k+1}\right\|^2_{\mathsf{R}_s} \tag{26a}$$

s.t.

$$\mathbf{y}_s = \mathsf{A}_s(\mathbf{u}_s - \mathbf{u}_0) + \mathbf{y}_0 + \hat{\mathbf{d}}(t|t), \tag{26b}$$

$$\mathbf{y}^{\text{eq}}_s = \mathbf{y}^S, \tag{26c}$$

$$\mathbf{y}^L_k \le \mathbf{y}^{\text{in}}_s \le \mathbf{y}^U_k, \tag{26d}$$

$$\mathbf{u}^L_k \le \mathbf{u}_s \le \mathbf{u}^U_k, \tag{26e}$$

with

$$\mathsf{R}_s = \mathsf{V}_{k+1}\mathsf{C}_s(\mathsf{V}_{k+1})^{\mathsf{T}}. \tag{27}$$

The quadratic term $\|\mathbf{u}_s - \mathbf{u}^\star_{k+1}\|^2_{\mathsf{R}_s}$ is included in (26a) only if the number of equality constraints plus the number of active inequality constraints is lower than the number of inputs, i.e., if $n^b_{k+1} := n_u - n^{\text{eq}}_y - n^a_{k+1} > 0$. The columns in the matrix $\mathsf{V}_{k+1} \in \mathbb{R}^{n_u \times n^b_{k+1}}$ correspond to directions in the input space, and the weighting matrices $\mathsf{Q}_s \in \mathbb{R}^{n^a_{k+1} \times n^a_{k+1}}$ and $\mathsf{C}_s \in \mathbb{R}^{n^b_{k+1} \times n^b_{k+1}}$ are positive definite. V_{k+1} can be selected from information given by the steady-state model used at the RTO layer. Let us define the Jacobian matrix of the controlled constrained quantities, evaluated at $\mathbf{u}^\star_{k+1}$, as:

$$\mathcal{G}_{k+1} := \begin{bmatrix} \dfrac{\partial \mathbf{y}^{\text{eq}}}{\partial \mathbf{u}}(\mathbf{u}^\star_{k+1}, \boldsymbol{\theta}) \\ \dfrac{\partial \mathbf{z}}{\partial \mathbf{u}}(\mathbf{u}^\star_{k+1}, \boldsymbol{\theta}) \end{bmatrix} \tag{28}$$

Notice that the rows of $\mathcal{G}_{k+1}$ are the gradients of the controlled constrained quantities $\mathbf{y}^{\text{eq}}$ and $\mathbf{z}$ evaluated at $\mathbf{u}^\star_{k+1}$. Assuming that $\mathcal{G}_{k+1}$ is full row rank, the columns in V_{k+1} can be selected as an orthonormal basis of the null space of the constraint gradients, i.e., $\mathsf{V}_{k+1} = \mathsf{N}_{k+1}$, with $\mathcal{G}_{k+1}\mathsf{N}_{k+1} = \mathbf{0}$. This choice of V_{k+1} is based on the variational analysis carried out in [7], where it is shown that upon enforcing the active constraints and the optimal inputs along the columns in N_{k+1}, the loss of optimality is only $\mathcal{O}(\boldsymbol{\delta}\mathbf{d}^2)$. This selection of V_{k+1} is convenient, but it is not optimal. An optimal selection of V_{k+1} will be presented later, in Section 3.3.4.

Remark 2. Notice that $\|\mathbf{u}_s - \mathbf{u}^\star_{k+1}\|^2_{\mathsf{R}_s} = \|\boldsymbol{\beta}_s - \boldsymbol{\beta}^S_{k+1}\|^2_{\mathsf{C}_s}$, where $\boldsymbol{\beta}_s := (\mathsf{V}_{k+1})^{\mathsf{T}}\mathbf{u}_s \in \mathbb{R}^{n^b_{k+1}}$ are the inputs along the directions given by the columns in V_{k+1}, and $\boldsymbol{\beta}^S_{k+1} := (\mathsf{V}_{k+1})^{\mathsf{T}}\mathbf{u}^\star_{k+1}$ are their corresponding optimal values. Hence, the QP Problem (26) represents a square control problem where the number of setpoints in $\mathbf{y}^S$, $\mathbf{z}^S_{k+1}$, and $\boldsymbol{\beta}^S_{k+1}$, equals the number of decision variables, n_u. The constraint (26c) is used to control the equality constrained variables, the quadratic term $\|\mathbf{z}_s - \mathbf{z}^S_{k+1}\|^2_{\mathsf{Q}_s}$ is used to control the active inequality constrained variables, and $\|\mathbf{u}_s - \mathbf{u}^\star_{k+1}\|^2_{\mathsf{R}_s}$ is used to exploit the additional degrees of freedom towards optimality.

Example 1. Consider the following constraint-adaptation problem:

$$\begin{aligned} \mathbf{u}^\star_{k+1} = \arg\min_{\mathbf{u}} \quad & \Phi(\mathbf{u}, \boldsymbol{\theta}) \\ \text{s.t.} \quad & y^{\text{in}}(\mathbf{u}, \boldsymbol{\theta}) + \varepsilon^{\text{in}}_k \le y^U \end{aligned} \tag{29}$$

where the input $\mathbf{u}$ has two components u_1 and u_2, and there is a single inequality constraint y^{in}. Fig. 3 illustrates the case where, at the current operating point $\mathbf{u}_k$, Problem (29) is solved for $\mathbf{u}^\star_{k+1}$. In the presence of plant-model mismatch, the predicted constraint boundary $y^{\text{in}}(\mathbf{u}, \boldsymbol{\theta}) + \varepsilon^{\text{in}}_k = y^U$ does not match the constraint boundary for the plant, $y^{\text{in}}_p(\mathbf{u}) = y^U$. Assuming that at $\mathbf{u}^\star_{k+1}$ the predicted inequality constraint is active, the objective function of the SSTO problem is chosen as follows:

$$J_{SSTO} = q_s\left(y^{\text{in}}_s - y^U\right)^2 + c_s\left((\mathsf{V}_{k+1})^{\mathsf{T}}\mathbf{u}_s - (\mathsf{V}_{k+1})^{\mathsf{T}}\mathbf{u}^\star_{k+1}\right)^2$$

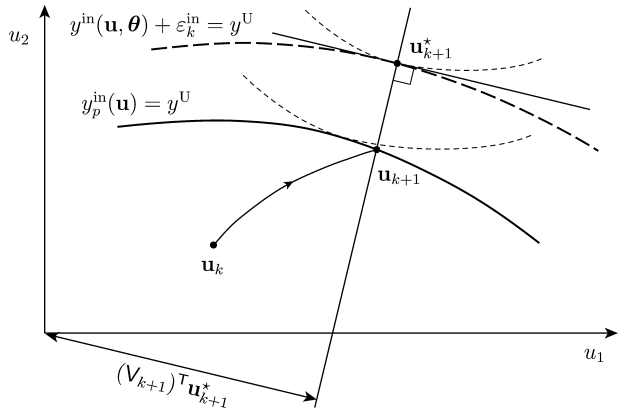

Fig. 3. Sketch of the integration between RTO and SSTO-MPC using Design B with $\mathsf{V}_{k+1} = \mathsf{N}_{k+1}$. *Dotted lines*: contours of the cost function $\Phi(\mathbf{u}, \boldsymbol{\theta})$.

Hence, the SSTO-MPC controller will take the operation from $\mathbf{u}_k$ to the new steady state $\mathbf{u}_{k+1}$, for which the inequality constraint of the plant is active, and the input along the direction V_{k+1} is at its optimal value given by the RTO point $\mathbf{u}^\star_{k+1}$. Notice that, Fig. 3 depicts where the SSTO-MPC control system will take the operation with respect to the RTO optimum point using Design B (for a single RTO execution and prior to convergence of the RTO iterations).

3.2.1. Matching the RTO optimal point upon convergence

As discussed in Section 3.1.1, a necessary condition for matching the RTO solution upon convergence is that the point $\mathbf{u}_\infty = \mathbf{u}^\star_\infty$ be a fixed point of the overall RTO/SSTO-MPC two-layer system.

Proposition 3. *If constraint adaptation is implemented at the RTO layer, Assumption 4 in Section 3.1.1 holds, and Design B is implemented at the SSTO-MPC layer, then $\mathbf{u}_\infty = \mathbf{u}^\star_\infty$ is a fixed point of the overall RTO/SSTO-MPC two-layer system.*

Proof. Constraint adaptation guarantees that the necessary condition in Proposition 1 is satisfied. We need to show that if $\mathbf{u}_\infty = \mathbf{u}^\star_\infty$ then $\mathbf{u}_s = \mathbf{u}^\star_\infty$ is the unique minimizing solution of Problem (26). Notice that $\mathbf{u}^\star_\infty$ is feasible and it is a zero of the objective function of Problem (26), i.e., it satisfies the setpoints $\mathbf{y}^S$, $\mathbf{z}^S_\infty$, and $\boldsymbol{\beta}^S_\infty$. Since Q_s and C_s are positive definite, it follows that $\mathbf{u}^\star_\infty$ is a minimum of Problem (26). Assuming that $\mathcal{G}_{k+1}$ is full row rank, there is by construction a unique target that satisfies the setpoints $\mathbf{y}^S$, $\mathbf{z}^S_\infty$, and $\boldsymbol{\beta}^S_\infty$, and therefore $\mathbf{u}^\star_\infty$ is a unique minimum. □

The SSTO design approach B is an optimizing control design, and as most optimizing control schemes it relies on the knowledge of the set of active constraints. If the set of active constraints changes as a result of process disturbances, then Design B may lose its optimizing properties until the next RTO execution recomputes the active set. An economic optimization design of the SSTO problem is presented next, where the SSTO problem approximates the nonlinear RTO problem. This may allow the SSTO problem to detect changes in the optimal set of active constraints during the transient.

3.3. Design approach C

Design approaches where the SSTO problem approximates the NLP problem solved at the RTO layer have been described in the literature [10,5]. Here, we consider the QP design described in [10], which is based on the QP approximation used in successive

quadratic programming (SQP) approaches for solving NLP problems [16,17]. The SQP approximation of the RTO NLP problem at the optimal point $\mathbf{u}^\star_{k+1}$ is given by:

$$\min_{\mathbf{u}} \quad \frac{\partial \Phi}{\partial \mathbf{u}}(\mathbf{u}-\mathbf{u}^\star_{k+1}) + \frac{1}{2}(\mathbf{u}-\mathbf{u}^\star_{k+1})^{\mathrm{T}} \frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}(\mathbf{u}-\mathbf{u}^\star_{k+1}) \tag{30a}$$

s.t.

$$\mathbf{y}^{\mathrm{eq}}(\mathbf{u}^\star_{k+1},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}^{\mathrm{eq}}_k + \frac{\partial \mathbf{y}^{\mathrm{eq}}}{\partial \mathbf{u}}(\mathbf{u}-\mathbf{u}^\star_{k+1}) = \mathbf{y}^S, \tag{30b}$$

$$\mathbf{y}^{\mathrm{L}}_k \le \mathbf{y}^{\mathrm{in}}(\mathbf{u}^\star_{k+1},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}^{\mathrm{in}}_k + \frac{\partial \mathbf{y}^{\mathrm{in}}}{\partial \mathbf{u}}(\mathbf{u}-\mathbf{u}^\star_{k+1}) \le \mathbf{y}^{\mathrm{U}}_k, \tag{30c}$$

$$\mathbf{u}^{\mathrm{L}}_k \le \mathbf{u} \le \mathbf{u}^{\mathrm{U}}_k, \tag{30d}$$

where $\frac{\partial \Phi}{\partial \mathbf{u}}$, $\frac{\partial \mathbf{y}^{\mathrm{eq}}}{\partial \mathbf{u}}$, $\frac{\partial \mathbf{y}^{\mathrm{in}}}{\partial \mathbf{u}}$ are evaluated at $(\mathbf{u}^\star_{k+1},\boldsymbol{\theta})$, $\mathcal{L}$ is the Lagrangian function, and $\frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}$ is evaluated at $(\mathbf{u}^\star_{k+1},\boldsymbol{\mu}^\star_{k+1},\boldsymbol{\theta})$, where $\boldsymbol{\mu}^\star_{k+1}$ are the optimal values of the Lagrange multipliers. In this formulation, the output constraints are linearized at $\mathbf{u}^\star_{k+1}$, and the curvature of the constraints is taken into account by using the Hessian of Lagrangian function in the quadratic term of the objective function.

In the framework of SQP, it is well known that if $\mathbf{u}^\star_{k+1}$ is a KKT (Karush–Kuhn–Tucker) solution to the RTO NLP problem (8), then $\mathbf{u}^\star_{k+1}$ is also a KKT solution to the SQP approximation (30) [16,17]. Indeed, the first and second order necessary conditions of optimality of Problem (30) match those of the RTO problem (8) at $\mathbf{u}^\star_{k+1}$.

Based on the SQP approximation (30), the SSTO problem can be formulated as follows:

$$\min_{\mathbf{u}_s,\mathbf{y}_s} \quad \frac{\partial \Phi}{\partial \mathbf{u}}(\mathbf{u}_s-\mathbf{u}^\star_{k+1}) + \frac{1}{2}\left\|\mathbf{u}_s-\mathbf{u}^\star_{k+1}\right\|^2_{\mathsf{H}_k} \tag{31a}$$

s.t.

$$\mathbf{y}_s = \mathsf{A}_{s,k}(\mathbf{u}_s-\mathbf{u}_0) + \mathbf{y}_0 + \hat{\mathbf{d}}(t|t), \tag{31b}$$

$$\mathbf{y}^{\mathrm{eq}}_s = \mathbf{y}^S, \tag{31c}$$

$$\mathbf{y}^{\mathrm{L}}_k \le \mathbf{y}^{\mathrm{in}}_s \le \mathbf{y}^{\mathrm{U}}_k, \tag{31d}$$

$$\mathbf{u}^{\mathrm{L}}_k \le \mathbf{u}_s \le \mathbf{u}^{\mathrm{U}}_k, \tag{31e}$$

where

$$\mathsf{A}_{s,k} = \mathsf{C}_k(\mathbf{I}-\mathsf{A}_k)^{-1}\mathsf{B}_k = \frac{\partial \mathbf{y}}{\partial \mathbf{u}}(\mathbf{u}^\star_{k+1},\boldsymbol{\theta}), \tag{32}$$

and H_k is a positive definite approximation of $\frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}(\mathbf{u}^\star_{k+1},\boldsymbol{\mu}^\star_{k+1},\boldsymbol{\theta})$. Condition (32) is included so as to match the gradients of $\mathbf{y}^{\mathrm{eq}}_s$ and $\mathbf{y}^{\mathrm{in}}_s$ (recall that $\mathbf{y}^{\mathrm{eq}}_s$ and $\mathbf{y}^{\mathrm{in}}_s$ are subsets of $\mathbf{y}_s$) with the gradients of the corresponding linear approximations (30b) and (30c). For simplicity, the gradients of all the output variables are matched in (32), although there is no need to match the gradients of output variables that do not belong to $\mathbf{y}^{\mathrm{eq}}_s$ or $\mathbf{y}^{\mathrm{in}}_s$.

3.3.1. Approximation of the Lagrangian Hessian

We would like Problem (31) to have a unique solution point. If it has infinite solutions, the optimizer would randomly pick either of these solutions at any sampling time, preventing the controlled plant from reaching steady-state operation. Problem (30) has a unique solution point $\mathbf{u}_s$ if the rows in matrix $\mathcal{G}_{k+1}$, defined in (28), are linearly independent, and the reduced Hessian matrix $\mathsf{N}^{\mathrm{T}}_{k+1}\left(\frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}\right)\mathsf{N}_{k+1}$ is positive definite (see Lemma 16.1 in [16]). Hence, H_k could be selected as an approximation of $\frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}$ such that $\mathsf{N}^{\mathrm{T}}_{k+1}\mathsf{H}_k\mathsf{N}_{k+1}$ is positive definite. However, since the set of active constraints in Problem (31) might change with the disturbance values $\hat{\mathbf{d}}(t|t)$, we use the stronger condition that H_k be positive definite.

One approach is to specify $\mathsf{H}_k = \left(\epsilon\mathbf{I} + \frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}\right)$, where ϵ is determined as follows [17]: fix $\delta > 0$, and let $\epsilon \ge 0$ be the smallest scalar that would make all the eigenvalues of the matrix $\left(\epsilon\mathbf{I} + \frac{\partial^2 \mathcal{L}}{\partial \mathbf{u}^2}\right)$ greater than or equal to δ.

3.3.2. Static gain matrix adaptation

In order to meet the condition (32) it is necessary to adapt the linear dynamic state-space model (15), (16) used by the MPC regulator at each RTO execution, such that its static gain matrix matches the output derivatives of the RTO model evaluated at $\mathbf{u}^\star_{k+1}$. The problem to be solved is to find new model matrices A_k, B_k, and C_k, such that:

- They represent as close as possible the old dynamic system used by the MPC, which is given by matrices A, B, and C.
- They satisfy condition (32).

Several options arise at this point. Assuming that $n_x \ge n_u$, one option that seems to be simple is to keep matrices A and B unmodified (i.e., $\mathsf{A}_k = \mathsf{A}$ and $\mathsf{B}_k = \mathsf{B}$), and to adapt only matrix C_k. Notice that condition (32) is a set of $n_y \times n_u$ equations, and matrix C_k has $n_y \times n_x$ entries that can be updated. Hence, one can search for the matrix C_k that is closest to C (in some metric) by solving an optimization problem where condition (32) enters as the constraints, and where the entries of C_k are the decision variables. For example, one can obtain C_k by solving the following problem:

$$\begin{aligned} \min_{c_{ij,k}} \quad & \sum_{i=1}^{n_y} (\mathbf{c}_{i,k}-\mathbf{c}_i)^{\mathrm{T}}(\mathbf{c}_{i,k}-\mathbf{c}_i) \\ \text{s.t.} \quad & \mathsf{C}_k(\mathbf{I}-\mathsf{A})^{-1}\mathsf{B} = \frac{\partial \mathbf{y}}{\partial \mathbf{u}}(\mathbf{u}^\star_{k+1},\boldsymbol{\theta}), \end{aligned} \tag{33}$$

where $c_{ij,k}$ is the ijth entry of matrix C_k, $\mathbf{c}^{\mathrm{T}}_{i,k}$ is the ith row of C_k, and $\mathbf{c}^{\mathrm{T}}_i$ is the ith row of matrix C. In this way, the model parameters affecting the transient regime of the system remain unmodified with respect to the original model (i.e., the modes and the input matrix are not changed).

3.3.3. Matching the RTO optimal point upon convergence

As before, a necessary condition for matching the RTO solution upon convergence is that the point $\mathbf{u}_\infty = \mathbf{u}^\star_\infty$ be a fixed point of the overall RTO/SSTO-MPC two-layer system.

Lemma 3. *If in Problem (31) the disturbance is such that* $\hat{\mathbf{d}}(t|t) = \mathbf{d}_{k+1}$, *with*

$$\mathbf{d}_{k+1} := \mathbf{y}(\mathbf{u}^\star_{k+1},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k - \mathsf{A}_{s,k}(\mathbf{u}^\star_{k+1}-\mathbf{u}_0) - \mathbf{y}_0, \tag{34}$$

then $\mathbf{u}_s = \mathbf{u}^\star_{k+1}$ *is the (unique) solution to Problem (31).*

Proof. The proof follows by verifying that the first-order KKT conditions of Problem (31) match those of the RTO NLP problem (8), and that the second-order sufficient conditions for a strict minimum of Problem (31) are also satisfied at $\mathbf{u}^\star_{k+1}$. Matching the output values and the output gradients is sufficient for matching the first-order KKT conditions [28,29]. Using (34) in (31b) we obtain:

$$\mathbf{y}_s = \mathbf{y}(\mathbf{u}^\star_{k+1},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k + \mathsf{A}_{s,k}(\mathbf{u}_s-\mathbf{u}^\star_{k+1}). \tag{35}$$

This is the same linearization of the output variables at $\mathbf{u}^\star_{k+1}$ that is used in (30b) and (30c) for the equality and inequality constrained outputs. Evaluating (35) at $\mathbf{u}_s = \mathbf{u}^\star_{k+1}$ we obtain $\mathbf{y}_s = \mathbf{y}(\mathbf{u}^\star_{k+1},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k$, which matches the values of the constrained outputs of Problem (8), evaluated at $\mathbf{u}^\star_{k+1}$. The output gradients are matched via condition

(32). On the other hand, the second-order sufficient conditions for a strict minimum are satisfied since H_k is positive definite. □

In general, $\hat{\mathbf{d}}(t|t)$ will not take the value $\mathbf{d}_{k+1}$, except upon convergence of the overall RTO/SSTO-MPC two-layer system as will be shown next. However, the theoretical disturbance $\mathbf{d}_{k+1}$ will prove itself useful in the forthcoming analysis of Section 3.3.4.

Proposition 4. *If constraint adaptation is implemented at the RTO layer, Assumption 4 in Section 3.1.1 holds, and Design C is implemented at the SSTO-MPC layer, then $\mathbf{u}_\infty = \mathbf{u}^\star_\infty$ is a fixed point of the overall RTO/SSTO-MPC two-layer system.*

Proof. We need to show that if $\mathbf{u}_\infty = \mathbf{u}^\star_\infty$ then $\mathbf{u}_s = \mathbf{u}^\star_\infty$ is the unique minimizing solution of Problem (31). From (21) it follows that the steady state value of the disturbance vector evaluated at $\mathbf{u}^\star_\infty$ is

$$\bar{\mathbf{d}} = \mathbf{y}_p(\mathbf{u}^\star_\infty, \mathbf{d}_p) - \mathbf{y}_0 - \mathsf{C}_\infty\, \delta\mathbf{x}_\infty = \mathbf{y}_p(\mathbf{u}^\star_\infty, \mathbf{d}_p) - \mathbf{y}_0 - \mathsf{A}_{s,\infty}(\mathbf{u}^\star_\infty - \mathbf{u}_0).$$

Constraint adaptation guarantees that $\mathbf{y}(\mathbf{u}^\star_\infty, \boldsymbol{\theta}) + \boldsymbol{\epsilon}_\infty = \mathbf{y}_p(\mathbf{u}^\star_\infty, \mathbf{d}_p)$. Hence, we have

$$\bar{\mathbf{d}} = \mathbf{y}(\mathbf{u}^\star_\infty, \boldsymbol{\theta}) + \boldsymbol{\epsilon}_\infty - \mathbf{y}_0 - \mathsf{A}_{s,\infty}(\mathbf{u}^\star_\infty - \mathbf{u}_0) = \mathbf{d}_\infty = \lim_{k\to\infty} \mathbf{d}_{k+1}. \tag{36}$$

In view of Lemma 3, it follows that $\mathbf{u}_s = \mathbf{u}^\star_\infty$ is the unique minimizing solution of Problem (31). □

KKT matching in Lemma 3 in general requires the satisfaction of condition (32), i.e., matching of the output derivatives between the RTO and MPC models. Notice that, if the dynamical model used in MPC is not adapted so as to satisfy condition (32), then $\mathbf{u}_\infty = \mathbf{u}^\star_\infty$ will not in general be a fixed point of the overall RTO/SSTO-MPC system using Design C. A particular case for which precise matching of the output gradients is not required was analyzed by Forbes et al. [6]. The analysis in [6] is restricted to problems where there are as many independent active constraints at the optimum as decision variables. In this case, the output gradients in the MPC model (rows in A_s) can vary with respect to the RTO model as long as the Lagrange multipliers corresponding to the active constraints remain positive.

3.3.4. Link between Designs B and C

The link between Designs B and C will be established for the case where the set of active inequality constraints in Problem (31) does not change with the disturbance values $\hat{\mathbf{d}}(t|t)$. By taking the active inequality constraints as equality constraints and removing the inactive constraints, the KKT conditions of Problem (31) reduce to:

$$\frac{\partial \Phi}{\partial \mathbf{u}} + (\mathbf{u}_s - \mathbf{u}^\star_{k+1})^\mathsf{T} \mathsf{H}_k + \boldsymbol{\lambda}^\mathsf{T} \mathcal{G}_{k+1} = \mathbf{0}, \tag{37}$$

$$\tilde{\mathbf{z}}_s = \mathcal{G}_{k+1}(\mathbf{u}_s - \mathbf{u}_0) + \tilde{\mathbf{z}}_0 + \tilde{\mathbf{d}}(t|t) = \tilde{\mathbf{z}}^S_{k+1}, \tag{38}$$

with $\tilde{\mathbf{z}}_s = [(\mathbf{y}^{\mathrm{eq}}_s)^\mathsf{T},\ (\mathbf{z}_s)^\mathsf{T}]^\mathsf{T}$, and $\tilde{\mathbf{z}}^S_{k+1} = [(\mathbf{y}^S)^\mathsf{T},\ (\mathbf{z}^S_{k+1})^\mathsf{T}]^\mathsf{T}$. Here, $\mathbf{z}_s$ and $\mathbf{z}^S_{k+1}$ were previously defined in Section 3.2; $\boldsymbol{\lambda}$ are the Lagrange multipliers associated with $\tilde{\mathbf{z}}_s$; $\tilde{\mathbf{d}}(t|t)$ contains the disturbance values in $\hat{\mathbf{d}}(t|t)$ corresponding to $\tilde{\mathbf{z}}_s$; and $\mathcal{G}_{k+1}$ is the Jacobian matrix previously defined in (28).

If $\hat{\mathbf{d}}(t|t) = \mathbf{d}_{k+1}$, then from Lemma 3 the solution of Problem (31) is $\mathbf{u}^\star_{k+1}$, and the KKT conditions (37), (38), evaluated at $\mathbf{u}^\star_{k+1}$, reduce to:

$$\frac{\partial \Phi}{\partial \mathbf{u}} + \boldsymbol{\lambda}^\mathsf{T}_{k+1} \mathcal{G}_{k+1} = \mathbf{0},$$

$$\tilde{\mathbf{z}}_s = \mathcal{G}_{k+1}(\mathbf{u}^\star_{k+1} - \mathbf{u}_0) + \tilde{\mathbf{z}}_0 + \tilde{\mathbf{d}}_{k+1} = \tilde{\mathbf{z}}^S_{k+1},$$

where $\tilde{\mathbf{d}}_{k+1}$ contains the disturbance values in $\mathbf{d}_{k+1}$ corresponding to $\tilde{\mathbf{z}}_s$.

Introducing the deviation variables $\delta\mathbf{d} = \tilde{\mathbf{d}}(t|t) - \tilde{\mathbf{d}}_{k+1}$, $\delta\mathbf{u}_s = \mathbf{u}_s - \mathbf{u}^\star_{k+1}$, and $\delta\boldsymbol{\lambda} = \boldsymbol{\lambda} - \boldsymbol{\lambda}_{k+1}$, conditions (37) and (38) can be written as:

$$\mathsf{M}_{k+1} \begin{bmatrix} \delta\mathbf{u}_s \\ \delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -\delta\mathbf{d} \end{bmatrix}, \quad \text{with} \quad \mathsf{M}_{k+1} = \begin{bmatrix} \mathsf{H}_k & \mathcal{G}^\mathsf{T}_{k+1} \\ \mathcal{G}_{k+1} & \mathbf{0} \end{bmatrix}. \tag{39}$$

If $\mathcal{G}_{k+1}$ has full row rank and the reduced Hessian matrix $\mathsf{N}^\mathsf{T}_{k+1}\mathsf{H}_k\mathsf{N}_{k+1}$ is positive definite, then the KKT matrix M_{k+1} is nonsingular [16], and there is a unique vector pair $(\delta\mathbf{u}_s, \delta\boldsymbol{\lambda})$ satisfying:

$$\begin{bmatrix} \delta\mathbf{u}_s \\ \delta\boldsymbol{\lambda} \end{bmatrix} = \mathsf{M}^{-1}_{k+1} \begin{bmatrix} \mathbf{0} \\ -\delta\mathbf{d} \end{bmatrix}, \quad \text{with} \quad \mathsf{M}^{-1}_{k+1} = \begin{bmatrix} \mathsf{M}_{1,k+1} & \mathsf{M}_{2,k+1} \\ \mathsf{M}_{3,k+1} & \mathsf{M}_{4,k+1} \end{bmatrix}. \tag{40}$$

Hence, if the active set is invariant, the optimal input targets are a linear function of the bias disturbances:

$$\delta\mathbf{u}_s = -\mathsf{M}_{2,k+1}\, \delta\mathbf{d}. \tag{41}$$

If H_k is positive definite, then H_k is invertible, and $\mathsf{M}_{2,k+1} = (\mathsf{H}_k)^{-1}\mathcal{G}^\mathsf{T}_{k+1}\left[\mathcal{G}_{k+1}(\mathsf{H}_k)^{-1}\mathcal{G}^\mathsf{T}_{k+1}\right]^{-1}$ (see exercise 3.7.11 in [30]).

This result motivates an alternative selection of matrix V_{k+1} in Design B, and the link between Designs B and C.

Proposition 5. *Let the rows in matrix $\mathcal{G}_{k+1}$ be linearly independent, and the reduced Hessian matrix $\mathsf{N}^\mathsf{T}_{k+1}\mathsf{H}_k\mathsf{N}_{k+1}$ be positive definite. Assume that the active set in Problem (31) does not change with the disturbance values $\mathbf{d}(t|t)$. If the columns in V_{k+1} are selected as an orthonormal basis of the null space of the rows in matrix $\mathsf{M}_{2,k+1}$, defined in (40), then Designs B and C reach the same steady-state targets in the presence of stable step-wise disturbances.*

Proof. The assumptions in the Proposition imply that matrix M_{k+1} in (39) is nonsingular, and hence, (41) holds. Let the columns in $\mathcal{N}_{k+1} \in \mathbb{R}^{n_u \times n^b_{k+1}}$ be an orthonormal basis of the null space of the rows in $\mathsf{M}_{2,k+1}$, then from (41) we have

$$(\mathcal{N}_{k+1})^\mathsf{T}\mathbf{u}_s - (\mathcal{N}_{k+1})^\mathsf{T}\mathbf{u}^\star_{k+1} = (\mathcal{N}_{k+1})^\mathsf{T}\delta\mathbf{u}_s = -(\mathcal{N}_{k+1})^\mathsf{T}\mathsf{M}_{2,k+1}\,\delta\mathbf{d} = \mathbf{0}. \tag{42}$$

Notice that, by means of (38), Design C is controlling the equality and active inequality constrained variables to their optimal set point values; and by means of (42), Design C is controlling the input targets along the directions given by the columns in $\mathcal{N}_{k+1}$ to their optimal RTO setpoints. These input targets do not depend on the disturbances $\delta\mathbf{d}$. Hence, if in Design B we select $\mathsf{V}_{k+1} = \mathcal{N}_{k+1}$, then (under the assumptions of the Proposition) we are enforcing the same steady-state targets with both SSTO designs. □

If in Design B we select $\mathsf{V}_{k+1} = \mathcal{N}_{k+1}$, and the set of active constraints does not change with the disturbance values, then the SSTO problem generates the same steady-state targets as with Design C. The advantage of Design C is that the approximated economic optimization problem has more chances to correctly detect changes in the set of active constraints. On the other hand, Design B has the advantage of not requiring to adapt the MPC model (so as to satisfy condition (32)) in order to reach the RTO solution upon convergence of the RTO iterations.

4. Illustrative case studies

4.1. Case Study 1

The behavior of the plant is described by the following continuous-time linear state-space system:

$$\begin{aligned}\dot{\mathbf{x}}_p &= \mathsf{A}_p\mathbf{x}_p + \mathsf{B}_p\mathbf{u} + \mathsf{G}_p\mathbf{d}_p,\\ \mathbf{y}_p &= \mathsf{C}_p\mathbf{x}_p + \mathbf{y}_c,\end{aligned} \tag{43}$$

with

$$\mathsf{A}_p = \begin{bmatrix} -\frac{7}{60} & -\frac{1}{300} & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & -\frac{19}{180} & -\frac{5}{1800}\\ 0 & 0 & 1 & 0 \end{bmatrix},\quad \mathsf{B}_p = \begin{bmatrix} 1 & 0\\ 0 & 0\\ 0 & 1\\ 0 & 0 \end{bmatrix},$$

$$\mathsf{G}_p = \begin{bmatrix} -0.186 & 1.556\\ 2.766 & -25.52\\ -0.126 & -1.055\\ -3.454 & 24.02 \end{bmatrix},$$

$$\mathsf{C}_p = \begin{bmatrix} \frac{2}{30} & \frac{1}{300} & -0.05 & -\frac{5}{1800}\\ 0.05 & \frac{1}{300} & \frac{5}{90} & \frac{5}{1800}\\ \frac{8}{75} & \frac{2}{375} & -0.012 & -\frac{2}{3000} \end{bmatrix},\quad \mathbf{y}_c = \begin{bmatrix} -0.35\\ -1.00\\ -1.10 \end{bmatrix},$$

which consists of four state variables $\mathbf{x}_p = [\,x_{p,1}\ \ x_{p,2}\ \ x_{p,3}\ \ x_{p,4}\,]^\mathsf{T}$, two input variables $\mathbf{u} = [u_1\ u_2]^\mathsf{T}$, two unmeasured disturbances $\mathbf{d}_p = [d_1\ d_2]^\mathsf{T}$, and three measured output variables $\mathbf{y}_p = [y_{p,1}\ y_{p,2}\ y_{p,3}]^\mathsf{T}$.

The steady-state optimization problem for the plant reads:

$$\begin{aligned}\max_{\mathbf{u}}\ & \mathbf{c}^\mathsf{T}\mathbf{u}\\ \text{s.t.}\ & \mathbf{y}_p(\mathbf{u},\mathbf{d}_p) = \mathbf{y}_c - \mathsf{C}_p\left(\mathsf{A}_p\right)^{-1}(\mathsf{B}_p\mathbf{u} + \mathsf{G}_p\mathbf{d}_p) \le \mathbf{0},\end{aligned} \tag{44}$$

with $\mathbf{c} = [6.85\ 2.95]^\mathsf{T}$.

The following discrete-time linear state-space model is available, which uses a time step, $t_{step} = 4$:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathsf{A}\mathbf{x}_k + \mathsf{B}\mathbf{u}_k,\\ \mathbf{y}_k &= \mathsf{C}\mathbf{x}_k + \mathbf{y}_o,\end{aligned} \tag{45}$$

with

$$\mathsf{A} = \begin{bmatrix} 1.4777 & -0.70124 & 2.8879\\ 0.0087227 & 0.81012 & 0.038218\\ -0.16573 & 0.16357 & 0.092586 \end{bmatrix},$$

$$\mathsf{B} = \begin{bmatrix} 7.1307 & -1.1426\\ -0.10167 & 3.6016\\ 1.9318 & 0.45105 \end{bmatrix},$$

$$\mathsf{C} = \begin{bmatrix} 0.024812 & -0.065606 & 0.10527\\ 0.0066459 & 0.032906 & 0.028499\\ 0.027378 & -0.0097313 & 0.11945 \end{bmatrix},\quad \mathbf{y}_o = \begin{bmatrix} -0.12\\ -0.847\\ -1.031 \end{bmatrix},$$

which consists of three state variables $\mathbf{x} = [x_1\ x_2\ x_3]^\mathsf{T}$, and includes the two input variables $\mathbf{u} = [u_1\ u_2]^\mathsf{T}$, and the three predicted output variables $\mathbf{y} = [y_1\ y_2\ y_3]^\mathsf{T}$.

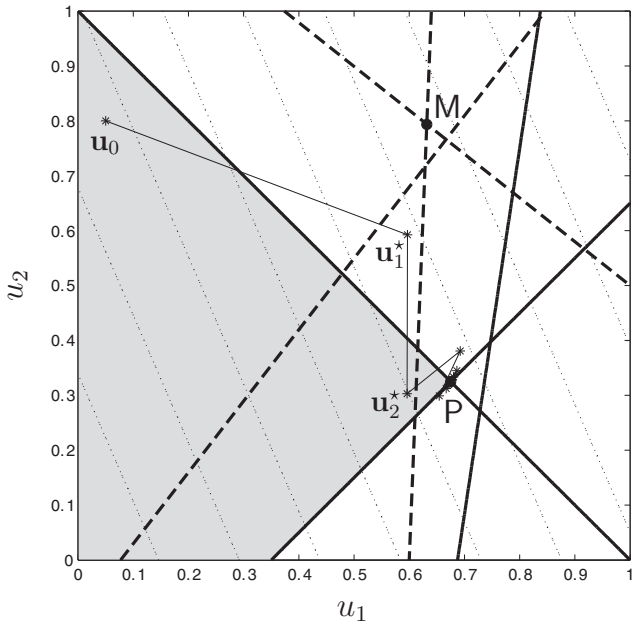

Fig. 4. Contour maps for Problem (44) with $\mathbf{d}_p = \mathbf{0}$. *Colored area*: feasible region; *Thick solid lines*: boundaries of the output constraints for the plant; *Thick dashed lines*: boundaries of the output constraints predicted by the model without adaptation; *Dotted lines*: contours of the cost function; *Point P*: optimum for the plant; *Point M*: optimum for the model without adaptation. *Iterations -*-*: constraint adaptation applied in open loop with $\mathsf{K}^{in} = \mathbf{I}_3$.

The nominal model-based steady-state optimization problem reads:

$$\begin{aligned}\max_{\mathbf{u}}\ & \mathbf{c}^\mathsf{T}\mathbf{u}\\ \text{s.t.}\ & \mathbf{y}(\mathbf{u}) = \mathbf{y}_o + \mathsf{C}(\mathbf{I}_3 - \mathsf{A})^{-1}\mathsf{B}\mathbf{u} \le \mathbf{0}.\end{aligned} \tag{46}$$

The contour maps of the plant (Problem (44)) with $\mathbf{d}_p = \mathbf{0}$, and the nominal model (Problem (46)) are presented in Fig. 4. The constraints on $y_{p,1}$ and $y_{p,2}$ are active at the plant optimum, while the constraints on y_2 and y_3 are active at the model optimum. Note that the optimal point determined from the nominal model is infeasible for the plant. The observed plant-model mismatch is purposely introduced in order to demonstrate the applicability and the behavior of the RTO/SSTO-MPC system using the different SSTO designs in a situation where such mismatch exists.

First, the RTO results using constraint-adaptation are applied in open-loop fashion. Constraint adaptation is executed every 100 time units, which leaves sufficient time for the controlled system to reach steady state after an input change. For simplicity, the RTO optimizer is executed using a fixed RTO period. In practice, however, a steady-state identification scheme should determine when the plant is operating at (near) steady-state condition and trigger the execution of the RTO optimizer [2,9]. Starting from $\mathbf{u}_0 = [0.05\ 0.8]^\mathsf{T}$, the iterations obtained without filtering the constraint biases (i.e., with $\mathsf{K}^{in} = \mathbf{I}_3$) are depicted in Fig. 4. The corresponding time responses of the input and output variables, and of the cost function, are shown in Fig. 5 for the first five RTO iterations. Notice that, due to plant-model mismatch, the optimal steady-state inputs determined at the RTO layer might be infeasible, besides from suboptimal, prior to convergence of the constraint-adaptation scheme.

Next, the RTO results are applied in closed loop by means of SSTO-MPC. In the three simulation scenarios studied next, constraint adaptation is implemented at the RTO layer, without filtering the constraint biases, and the MPC regulator described in Section 2.3.1 is implemented with the prediction horizon $N = 5$. The prediction model (45) is augmented with constant output

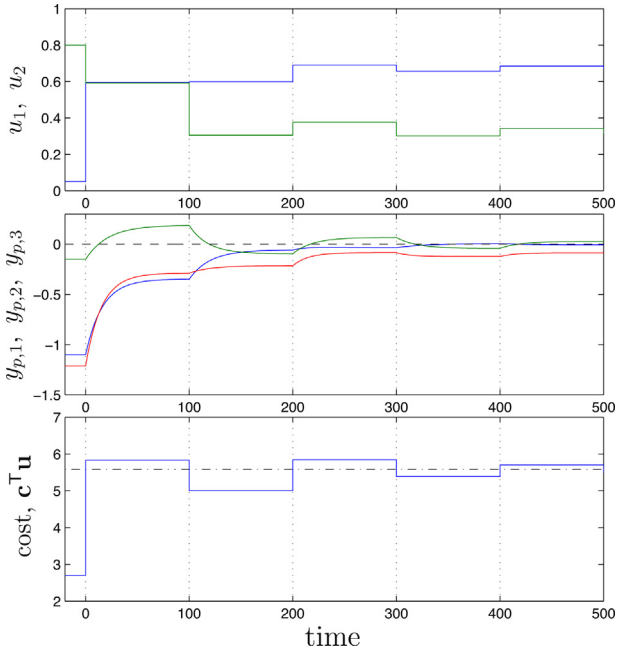

Fig. 5. Time response of the main variables in Case Study 1: open loop implementation of RTO (constraint adaptation).

disturbances, and a Luenberger observer is used. The terminal constraint (17e) is not included, as the MPC without the constraint is stable.

Since the cost function and the constraints are linear, Design C results in an LP problem rather than a QP problem. In fact, in this example the steady-state optimization problem solved at the SSTO stage is identical to the problem solved at the RTO layer. This issue will be discussed later.

Scenario 1. We analyze how the integrated RTO/SSTO-MPC schemes perform when starting from the conservative initial point $\mathbf{u}_0 = [0.05\ 0.8]^T$ without perturbations (i.e., with $\mathbf{d}_p = \mathbf{0}$). The results using SSTO Design A are shown in Fig. 6. The SSTO stage effectively

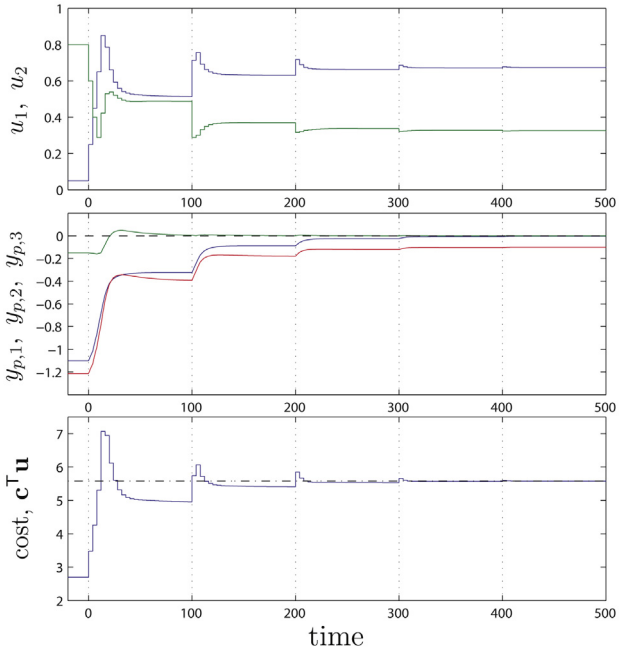

Fig. 6. Time response of the main variables in Case Study 1: Scenario 1 with Design A.

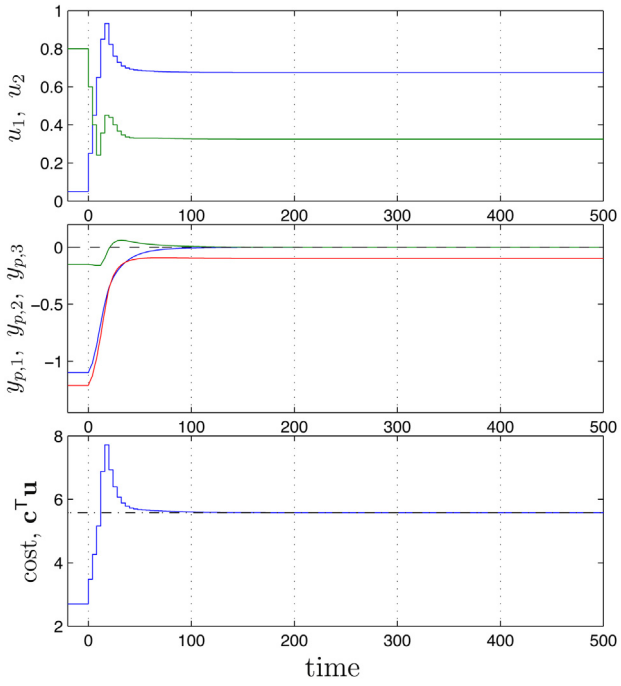

Fig. 7. Time response of the main variables in Case Study 1: Scenario 1 with Design B.

avoids constraint violations at the steady state operating points (compare with the open-loop implementation in Fig. 5). However, it takes about three RTO periods for the iterations to approach the plant optimum. The results using SSTO Design B are shown in Fig. 7. In this case, the correct set of active constraints is identified upon adaptation of the constraints at the initial point $\mathbf{u}_0$. Hence, by implementing Design B, the plant optimum is reached within the first RTO period. In this scenario, the response obtained with Design C is identical to that obtained with Design B.

Scenario 2. A unit step on disturbance d_1 is applied at time 20, when the integrated RTO/SSTO-MPC system is operating the plant at the plant optimum previous to the disturbance. The effect of this disturbance on the steady-state mapping of the plant is to move the constraint boundaries as illustrated in Fig. 8. Note that the plant optimum changes but the set of active constraints remains the same. Constraint adaptation is executed every 200 time units. The results using SSTO Designs A and B are shown in Figs. 9 and 10, respectively. Since the disturbance does not change the set of active constraints, the controller with Design B takes the operation directly to the new plant optimum. As in the previous scenario, Design C produces the same response as Design B.

Scenario 3. This time, a unit step on disturbance d_2 is applied at time 20. The effect of this disturbance on the steady-state mapping of the plant is to move the boundary of $y_{p,3}$ as illustrated in Fig. 11. The constraint on $y_{p,3}$ becomes active at the plant optimum, and the constraint on $y_{p,1}$ becomes inactive. Constraint adaptation is executed every 200 time units. The results using SSTO Design B are shown in Fig. 12. Previous to the disturbance, Design B is controlling $y_{p,1}$ and $y_{p,2}$ to their boundary values. These setpoints become infeasible after the disturbance, and therefore the SSTO-MPC controller brings the operation to a feasible steady-state point, which is at a compromise distance from meeting the boundaries on $y_{p,1}$ and $y_{p,2}$ depending on the corresponding weighting matrices used for these two outputs in the objective function of the SSTO problem. At time 200, when the controlled plant has reached a near steady-state condition, the RTO optimizer is executed and the correct set of active constraints is passed to the SSTO stage. The results using Design C are shown in Fig. 13. Since Design C is evaluating the

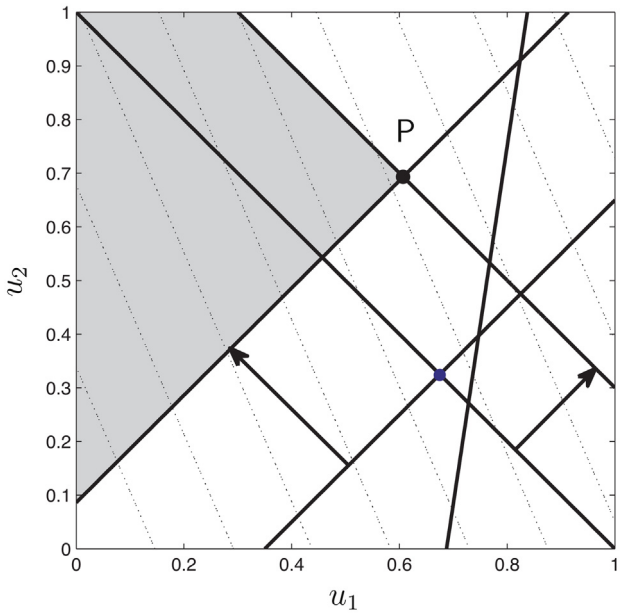

Fig. 8. Contour map for Problem (44) with $\mathbf{d}_p = [1\,0]^T$.

economic steady state during the transient, it is able to detect the change in the active constraints and bring the operation at steady state directly to the intersection of the new active constraints.

As we mentioned earlier, when Design C is used in this case study, the economic optimization carried out at the SSTO stage is identical to the economic optimization at the RTO layer (in both cases an LP program). Hence, the RTO layer plays no role and can be eliminated. The inclusion of the RTO layer makes sense when the RTO layer uses a more detailed and accurate model than the SSTO stage.

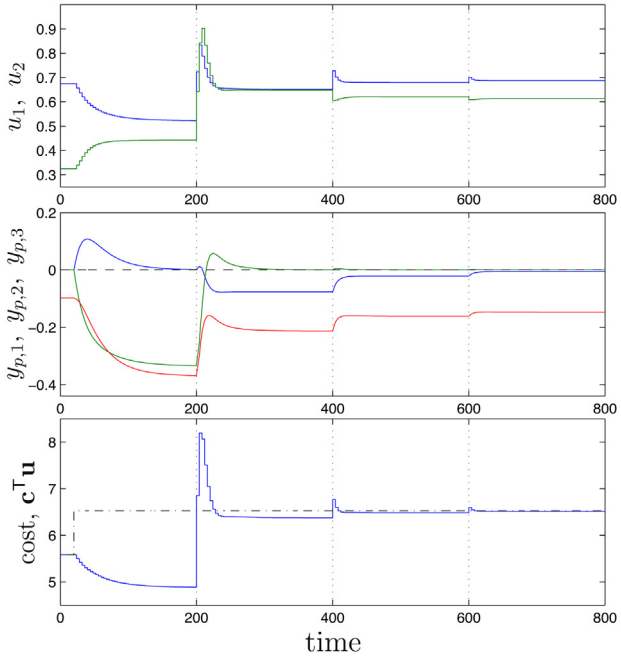

Fig. 9. Time response of the main variables in Case Study 1: Scenario 2 with Design A.

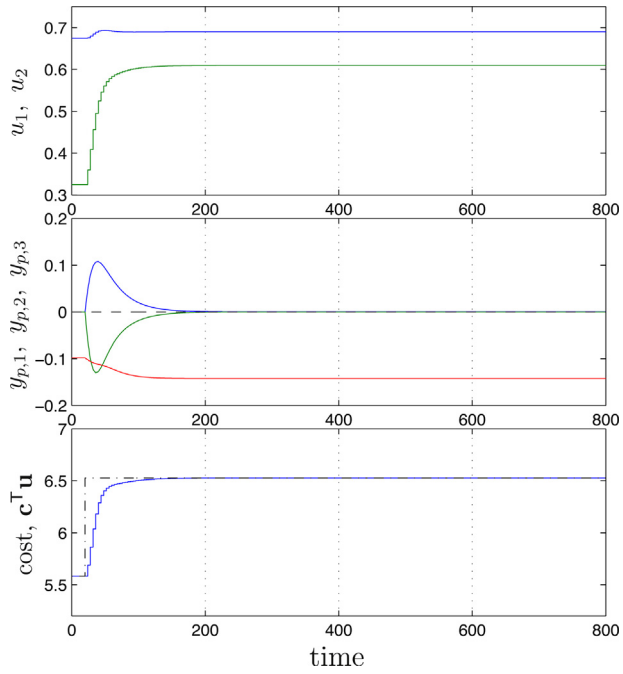

Fig. 10. Time response of the main variables in Case Study 1: Scenario 2 with Design B.

4.2. Case Study 2

The reactor in the Williams-Otto plant is considered [31]. It consists of an ideal CSTR in which the following reactions occur:

$$A + B \longrightarrow C, \qquad k_1 = 1.6599 \times 10^6 e^{-6666.7/(T_R+273)}$$

$$B + C \longrightarrow P + E, \qquad k_2 = 7.2117 \times 10^8 e^{-8333.3/(T_R+273)}$$

$$C + P \longrightarrow G, \qquad k_3 = 2.6745 \times 10^{12} e^{-11111/(T_R+273)}$$

where the reactants A and B are fed with the mass flow rates 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 reactor mass holdup is 2105 kg, and the product stream has the mass flowrate $F = F_A + F_B$. The material balance equations for this CSTR can be found in [32].

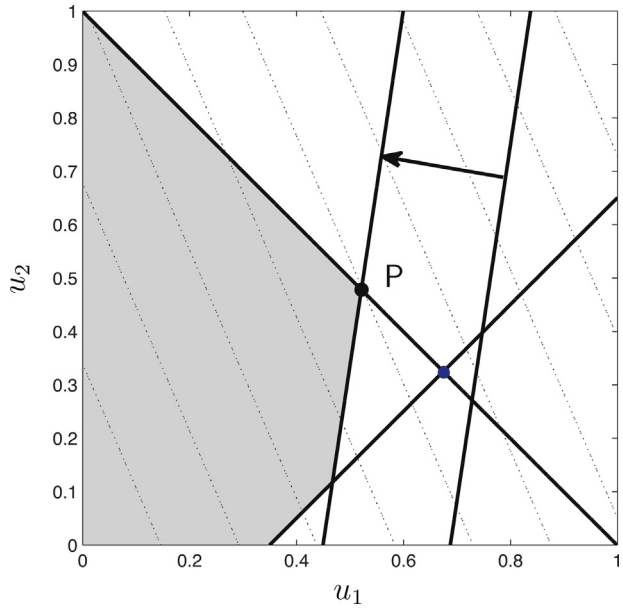

Fig. 11. Contour map for Problem (44) with $\mathbf{d}_p = [0\,1]^T$.

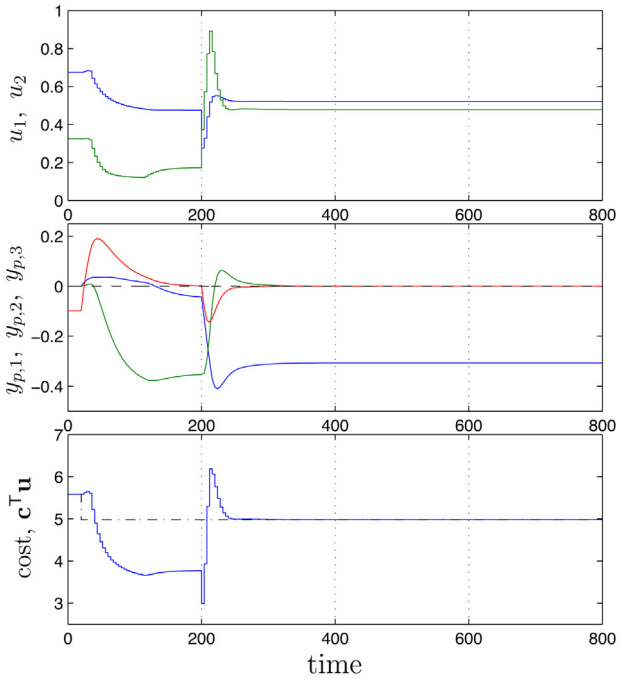

Fig. 12. Time response of the main variables in Case Study 1: Scenario 3 with Design B.

The decision variables are F_B and the reactor temperature, T_R, i.e., $\mathbf{u} = [F_B\ T_R]^T$. The states are $\mathbf{x} = [X_A\ X_B\ X_C\ X_P\ X_G\ X_E]^T$, where X_j is the mass fraction of species j. The objective is to maximize profit at steady state operation, which is expressed as the price difference between the products and the reactants:

$$\phi = 1200X_P F + 80X_E F - 76F_A - 114F_B, \tag{47}$$

The disturbance variable considered is F_A, i.e., $d_p = F_A$. The steady-state mapping for the plant mass fraction j is represented as $X_{j,p}(\mathbf{u}, d_p)$. The nominal model uses $F_A = 1.4$ kg/s. Hence, the steady-state mapping for the model is given by $X_j(\mathbf{u}) = X_{j,p}(\mathbf{u}, 1.4)$. Similarly, the steady-state mappings for the plant and nominal model profit are represented by $\Phi_p(\mathbf{u}, d_p)$, and $\Phi(\mathbf{u}) = \Phi_p(\mathbf{u}, 1.4)$, respectively. The steady-state optimization problem for the plant reads:

$$\begin{aligned} \max_{F_B\ T_R}\ & \Phi_p(\mathbf{u}, d_p) \\ \text{s.t.}\ & X_{B,p}(\mathbf{u}, d_p) \le 0.25 \\ & 2 \le F_B \le 4, \quad 70 \le T_R \le 100 \end{aligned} \tag{48}$$

where $X_{B,p}$ is constrained to be lower than or equal to 0.25. The constraint-adaptation approach is implemented at the RTO layer, using the nominal model with $F_A = 1.4$ kg/s. The steady-state map for the nominal model is illustrated in Fig. 14. The RTO solution map as a function of the constraint bias ε is given by:

$$\begin{aligned} U^\star(\varepsilon) = \operatorname*{argmax}_{F_B,\ T_R}\ & \Phi(\mathbf{u}) \\ \text{s.t.}\ & X_B(\mathbf{u}) + \varepsilon \le 0.25 \\ & 2 \le F_B \le 4, \quad 70 \le T_R \le 100 \end{aligned} \tag{49}$$

Let us assume that the initial value of F_A is 1.4 kg/s. Hence, there is initially no plant-model mismatch, and the reactor is initially operating at the nominal optimum $\bar{\mathbf{u}}^\star = [85.6, 2.549]^T = U^\star(0)$, which is indicated in Fig. 14. The curve $U^\star(\varepsilon)$ indicates the possible locations of the RTO setpoints if a stable disturbance enters the reactor.

Using Design C, the SQP approximation of the RTO problem is taken at $\bar{\mathbf{u}}^\star$. Let $\hat{d}$ denote the estimated disturbance for the output X_B. As a function of $\hat{d}$, the solution of the SSTO problem using Design C is given by the linear relation $U_s^C(\hat{d})$, which is computed from (41). $U_s^C(\hat{d})$ is shown in Fig. 14. Notice that $U_s^C(\hat{d})$ is the tangent of $U^\star(\varepsilon)$ at $\bar{\mathbf{u}}^\star$. This tangency indicates that the solution of the SSTO problem using Design C represents a first-order approximation to the solution of the RTO problem using constraint adaptation. This is the case only because both problems use the same adaptation strategy, which is based on biasing the output predictions. In general, Design C will not give a first-order approximation of the RTO solution if a different adaptation strategy is used at the RTO layer, e.g., such as the classical two-step approach.

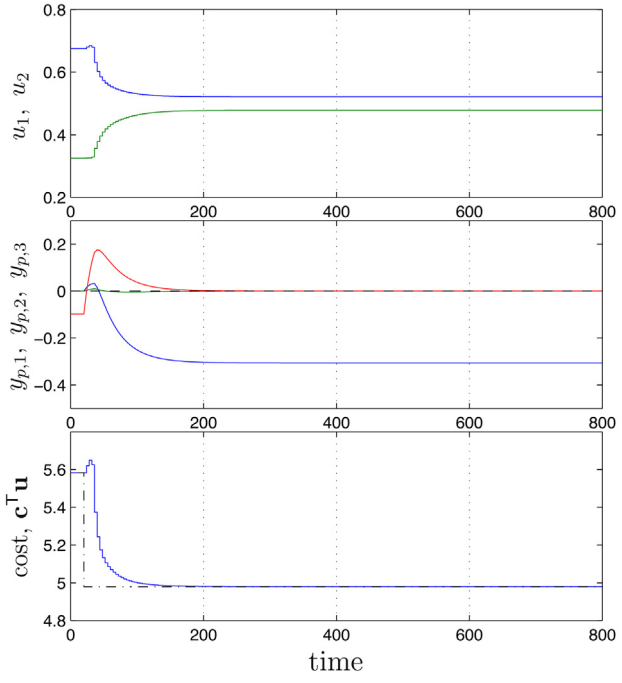

Fig. 13. Time response of the main variables in Case Study 1: Scenario 3 with Design C.

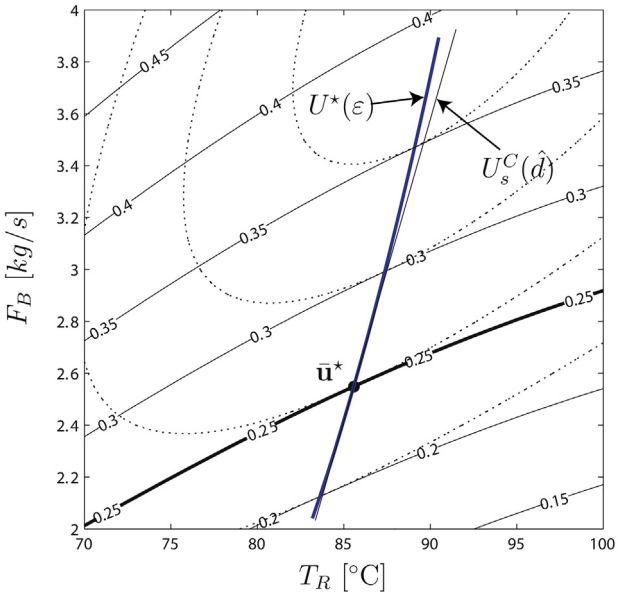

Fig. 14. Case Study 2. Steady-state map with $F_A = 1.4$ kg/s. *Dotted curves:* contours of the profit function; *Solid curves:* contours of X_B.

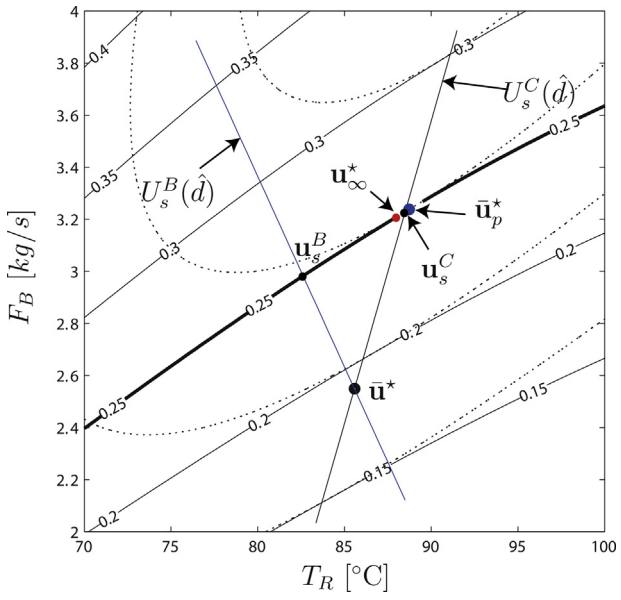

Fig. 15. Case Study 2. Steady-state map with $F_A = 1.83$ kg/s. *Dotted curves:* contours of the profit function; *Solid curves:* contours of X_B.

Now, let us assume that the value of F_A increases to 1.83 kg/s. This disturbance introduces plant-model mismatch in the system, which modifies the steady-state map for the true reactor, as shown in Fig. 15. The constraint on $X_{B,p}$ shifts, and the plant optimum moves to point $\bar{\mathbf{u}}_p^\star$. The constraint-adaptation scheme will converge to the point $\mathbf{u}_\infty^\star$, i.e., the point where $U^\star(\varepsilon)$ intersects the constraint contour $X_{B,p} = 0.25$. However, right after the disturbance takes place, the SSTO problem will take the operation to different points, depending on the design used. Design A will keep the operation at $\bar{\mathbf{u}}^\star$, since this point is feasible. Design C will take the operation to the point $\mathbf{u}_s^C$, which is where $U_s^C(\hat{d})$ intersects $X_{B,p} = 0.25$. The point reached with Design B will depend on the choice of V_{k+1}. One possibility is to select V_{k+1} as proposed in Section 3.2. Since the direction in V_{k+1} is dependent on the scaling of the inputs, these are scaled as $T_R^s = (T_R - 70)/30$, and $F_B^s = (F_B - 2)/2$. Next, the derivative $\frac{dX_B}{d\mathbf{u}^{\text{scaled}}}$ is computed from the nominal model at $\bar{\mathbf{u}}^\star$ using the scaled inputs, and V_{k+1} is obtained such that $\frac{dX_B}{d\mathbf{u}^{\text{scaled}}}\mathsf{V}_{k+1} = 0$. Using this choice of V_{k+1}, Design B will take the operation to point $\mathbf{u}_s^B$, which is at the intersection of $U_s^B(\hat{d})$ and $X_{B,p} = 0.25$, where $U_s^B(\hat{d})$ is orthogonal to V_{k+1}. Another possibility is to select the optimal V_{k+1}, as proposed in Proposition 5, in which case Design B will take the operation to $\mathbf{u}_s^C$, the same point as Design C.

The close proximity of $\mathbf{u}_s^C$ with $\bar{\mathbf{u}}_p^\star$ is just a lucky coincidence in this example. The point $\mathbf{u}_s^C$ should be viewed as an approximation of $\mathbf{u}_\infty^\star$, not of $\bar{\mathbf{u}}_p^\star$. Also, notice that the close proximity of the line $U_s^C(\hat{d})$ with the curve $U^\star(\varepsilon)$, which can be visualized in Fig. 14, indicates that in this simple case study the constraint-adaptation RTO algorithm brings no meaningful improvement over the steady-state points that can be reached by the SSTO problem alone using Design C.

4.3. Case Study 3

Let us assume that the steady-state optimization problem of a given process reads:

$$\begin{aligned} \max_{\mathbf{u}} \quad & -\Phi_p(\mathbf{u}) = 2,213u_1 + u_2 \\ \text{s.t.} \quad & y_p(\mathbf{u}, d_p) = \theta_{1,p} + \theta_{2,p}u_1 + \theta_{3,p}u_2 + \theta_{4,p}u_1^2 + \theta_{5,p}u_1u_2 + \theta_{6,p}u_2^2 + d_p \le 10 \\ & 0 \le u_1, u_2 \le 10 \end{aligned} \tag{50}$$

Table 1
Values of the parameters $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}$ in Case Study 3.

i	1	2	3	4	5	6
$\theta_{i,p}$ (plant)	4.154	−0.059	−0.067	0.0486	−0.0065	0.025
θ_i (model)	6.154	−0.049	−0.052	0.0436	−0.0165	0.03

with two decision variables $\mathbf{u} = [u_1\ u_2]^\mathsf{T}$, six unknown plant parameters $\boldsymbol{\theta}_p = [\theta_{1,p} \ldots \theta_{6,p}]^\mathsf{T}$, one uncertain constrained output variable y_p, and one disturbance variable d_p. Using constraint adaptation, the model-based optimization problem solved at the RTO layer reads:

$$\begin{aligned} \max_{\mathbf{u}} \quad & -\Phi(\mathbf{u}) = 2,213u_1 + u_2 \\ \text{s.t.} \quad & y(\mathbf{u}, \boldsymbol{\theta}) + \varepsilon = \theta_1 + \theta_2u_1 + \theta_3u_2 + \theta_4u_1^2 + \theta_5u_1u_2 + \theta_6u_2^2 + \varepsilon \le 10 \\ & 0 \le u_1, u_2 \le 10 \end{aligned} \tag{51}$$

where ε is the constraint bias, and $\boldsymbol{\theta} = [\theta_1 \ldots \theta_6]^\mathsf{T}$ are the nominal model parameters. The parameter values for the plant constraint y_p (simulated reality) and for the model constraint y are reported in Table 1. Notice that the cost function is a known linear function, whereas there is model uncertainty in the quadratic constraint.

Let us denote by $U_p^\star(d_p)$ the optimal solution map of Problem (50) as a function of the disturbance value, and by $U^\star(\varepsilon)$ the RTO solution (Problem (51)) as a function of the constraint bias. The initial disturbance value is $d_p = 0$, for which the corresponding steady-state map of the plant is shown in Fig. 16. In this case, we have $U_p^\star(0) = [10,\ 10]^\mathsf{T}$, i.e., the upper input bounds on u_1 and u_2 are both active at the optimum. On the other hand, the constraint $y_p(\mathbf{u}, 0) \le 10$ is inactive. Let us assume that initially the plant is operating at the optimum, i.e., $\mathbf{u}_0 = U_p^\star(0)$. At this point, the constraint bias is computed as $\varepsilon_0 = y_p(\mathbf{u}_0, 0) - y(\mathbf{u}_0, \boldsymbol{\theta})$. The steady-state map for the RTO model corrected by ε_0 is shown in Fig. 17. Notice that $\mathbf{u}_0 = U^\star(\varepsilon_0)$.

In this example, we shall consider that $y_p(\mathbf{u}, d_p)$ is the steady-state mapping corresponding to a dynamic process, and that a stable SSTO-MPC control system is implemented, where the MPC controller is designed with zero offset. When the steady-state map of the plant changes due to a change in the disturbance value d_p, it is possible to determine the steady-state point that will be reached by the SSTO-MPC control system for the different SSTO designs.

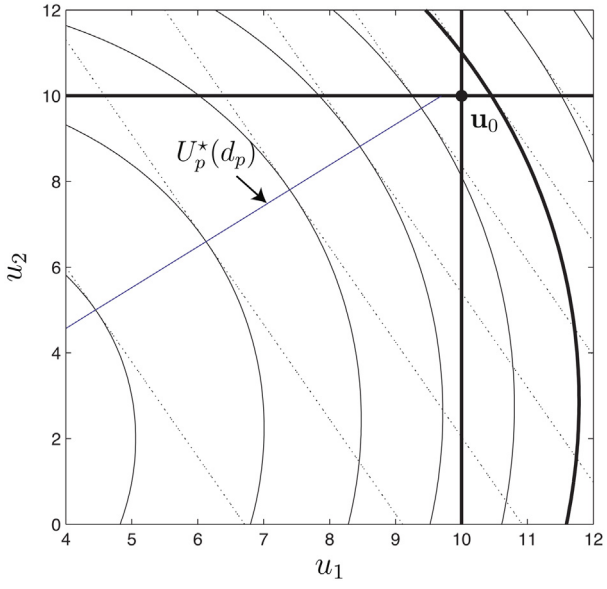

Fig. 16. Case Study 3. Steady state map for the plant with $d_p = 0$. *Dotted lines:* contours of the cost function; *Thin solid lines:* contours of $y_p(\mathbf{u}, 0)$; *Thick lines:* constraint boundaries.

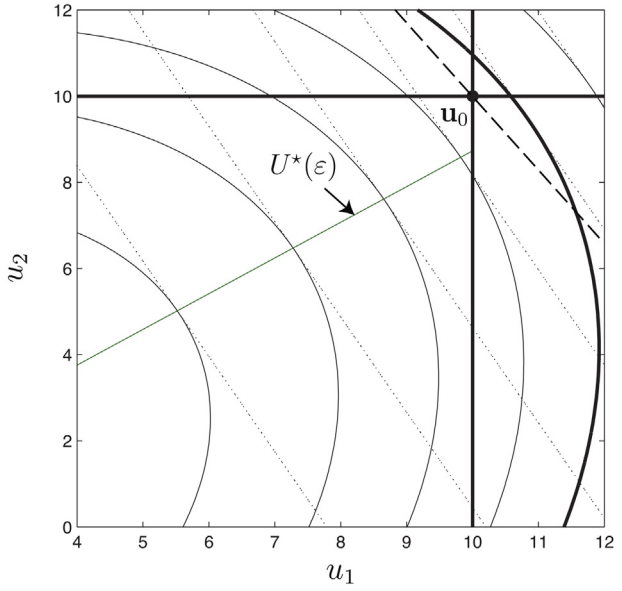

Fig. 17. Case Study 3. Steady state map for the RTO model with ε_0 computed at $\mathbf{u}_0$ (with $d_p = 0$ for the plant). *Dotted lines:* contours of the cost function; *Thin solid lines:* contours of $y(\mathbf{u}, \boldsymbol{\theta}) + \varepsilon_0$; *Thick lines:* constraint boundaries predicted by the RTO model at $\mathbf{u}_0$; *Dashed line:* contour of the linearized output constriant at $\mathbf{u}_0$.

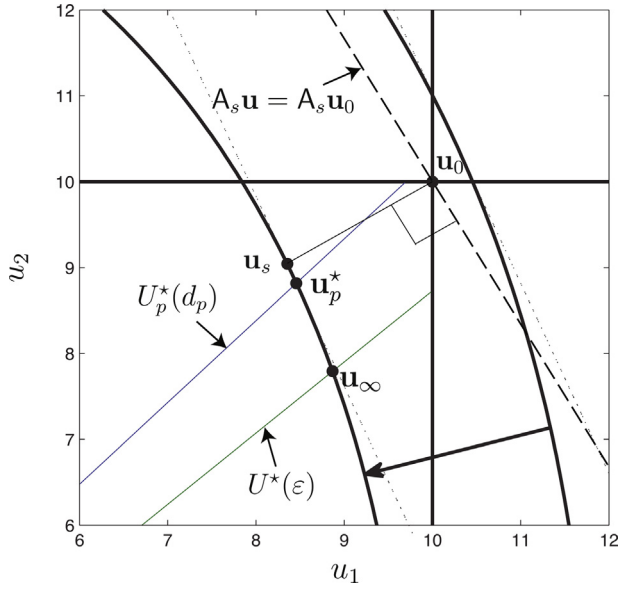

Fig. 19. Case Study 3. Shift in the boundary of the output constraint for the plant with $d_p = 2$. Stationary point reached using Designs A and B. *Dotted lines:* contours of the cost function; *Thick lines:* constraint boundaries.

In effect, the point $\mathbf{u}_s$ must satisfy the NCO of the SSTO problem, and from Section 2.4 we know that at steady state $\mathbf{u}_s$ must satisfy $\mathbf{y}_s = \mathbf{y}_p(\mathbf{u}_s, \mathbf{d}_p)$.

Now, coming back to the example, let us consider that the value of d_p changes to $d_p = 2$. This disturbance modifies the steady-state map for the plant by shifting the boundary of the output constraint as indicated in Fig. 18. The new plant optimum $\mathbf{u}_p^\star = U_p^\star(2)$ is now located on the boundary of the output constraint. After the disturbance takes place, Design C will take the operation to point $\mathbf{u}_s'$, which is at the intersection of the lower input bound on u_2 and the upper bound on y_p. Since this point is very distant from the RTO solutions $U^\star(\varepsilon)$, one can infer that the SQP approximation at $\mathbf{u}_0$ fails to approximate the RTO solution when this change in the set of active constraints takes place. As a matter of fact, before the disturbance takes place the plant is operating at $\mathbf{u}_0$. In the SQP approximation the constraint $y(\mathbf{u}, \boldsymbol{\theta}) + \varepsilon_0 \le 10$ is linearized at $\mathbf{u}_0$. Since this constraint is initially inactive, its Lagrange multiplier is equal to zero and its curvature is not captured in the objective function of the SQP approximation. In fact, the SQP approximation is an LP problem, for which the solution is always at the intersection of the constraints, which is why the SSTO problem takes the operation to point $\mathbf{u}_s'$. The RTO solution obtained at $\mathbf{u}_s'$ is $\mathbf{u}_1^\star$. Using the SQP approximation at $\mathbf{u}_1^\star$ the SSTO stage will take the operation from $\mathbf{u}_s'$ to $\mathbf{u}_1 = \mathbf{u}_\infty$. Notice that, in this example, the RTO/SSTO-MPC system using Design C converges to the RTO solution in a single RTO iteration. However, after the disturbance takes place the stationary point $\mathbf{u}_s'$ is reached. This is very undesirable because the SSTO problem abruptly changes the operating point from the upper to the lower input bound on u_2, taking the operation to a distant point with important loss in cost. One way to avoid this long excursion is to limit the input range for the next operating point by moving the input bounds as in (9), (10). If the bounds on the input moves are chosen as $\Delta u_1 = \Delta u_2 = 4$, then using Design C the SSTO stage will take the operation to the point $\mathbf{u}_s''$, instead of $\mathbf{u}_s'$.

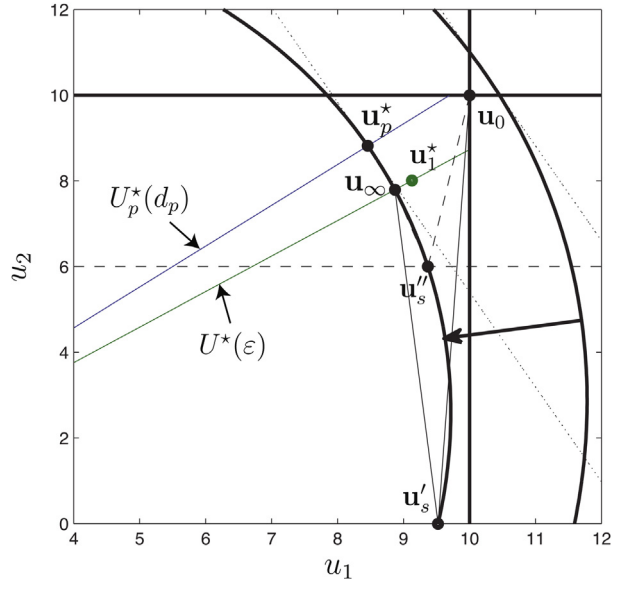

Fig. 18. Case Study 3. Shift in the boundary of the output constraint for the plant with $d_p = 2$. Stationary points reached using Design C. *Dotted lines:* contours of the cost function; *Thick lines:* constraint boundaries.

Next, let us consider the stationary points reached after the disturbance takes place using Designs A and B. At the initial point $\mathbf{u}_0$, the SSTO problem with Design A minimizes $\|\mathbf{u}_s - \mathbf{u}_0\|^2_{\mathsf{R}_s}$, since $\mathbf{u}_0 = U^\star(\varepsilon_0)$. On the other hand, Design B minimizes $\|\mathbf{u}_s - \mathbf{u}^{\mathrm{U}}\|^2_{\mathsf{Q}_s}$, since the upper input bounds are both active at $U^\star(\varepsilon_0)$. Noticing that $\mathbf{u}_0 = \mathbf{u}^{\mathrm{U}}$, it turns out that both SSTO problems are equivalent if $\mathsf{R}_s = \mathsf{Q}_s$. Assuming that $\mathsf{A}_s = \frac{\partial y}{\partial \mathbf{u}}(\mathbf{u}_0, \boldsymbol{\theta})$ and that $\mathsf{R}_s = \mathsf{Q}_s = \mathbf{I}$, the stationary point reached by Designs A and B after the disturbance takes place is $\mathbf{u}_s$ in Fig. 19, which is very close to the plant optimum. The RTO solution obtained at $\mathbf{u}_s$ is very close to $\mathbf{u}_\infty$, so for any practical purpose the RTO/SSTO-MPC system using Designs A and B will also reach $\mathbf{u}_\infty$ in a single RTO iteration. Notice that, unlike case studies 1 and 2, where in the presence of disturbances the SSTO problem using Design C approximated very well the constraint adaptation (RTO) solution, in this case, Design C fails to approximate the constraint adaptation solution when an inactive quadratic constraint becomes active. This case study illustrates a situation where Designs A and B outperform Design C.

5. Conclusions

This paper considered the integration between RTO and two-stage SSTO-MPC systems. In the presence of plant-model mismatch and constraints, the SSTO stage permits to correct the RTO setpoints so as to reach feasible operating points at steady state. However, the optimality of the steady state points reached will depend on the design of the SSTO problem. In the present study, two economic SSTO designs were presented (designs B and C) and contrasted with Design A, which does not have economic optimizing properties. The performance of the different SSTO designs was evaluated using three illustrative case studies involving different optimization and disturbance scenarios, with and without changes in the set of active constraints. This allows to better understand situations in which a given design may perform well, and situations in which a design may not perform as desired. Several remarks are in order.

- Designs B and C represent economic designs of the SSTO problem because, in the presence of plant-model mismatch in the RTO model, they may bring the operation directly to the intersection of the active constraints in a single RTO execution, while Design A may require several RTO executions for doing the same (this situation was illustrated in Case Study 1, Scenario 1). Furthermore, in the presence of stable disturbances Designs B and C may take the operation directly back to the active constraints (and therefore to near optimality), without waiting for the RTO execution to take place. In contrast, Design A has to wait for the controlled system to settle down to a possibly suboptimal steady-state point and for the RTO execution to update the RTO setpoints.
- Design B is an optimizing control design because it tracks setpoints for selected input and output variables, such that near optimality is achieved in the presence of disturbances. If the set of active constraints does not change, Design B can lead to (near) optimal operation without the intervention of the RTO layer (this situation was encountered in Case Study 1, Scenario 2). In the invariant active set case, Design B can even be made equivalent to Design C by selecting V_{k+1} as described in Proposition 5 (this situation was discussed in Case Study 2). When compared to Design C, Design B has the advantage of not requiring to adapt the MPC model in order to converge to the RTO solution.
- When the set of active constraints does not change with the disturbance values, the stationary solution of the SSTO problem using Design C can be viewed as a first-order approximation to the solution of the constraint-adaptation approach at the RTO layer (this situation was analyzed in Case Study 2). If the active set changes, and the constraints that become active or inactive are linear, the approximation given by Design C continues to be valid, since linear constraints appear unchanged in the SQP approximation, and they do not affect the Hessian of the Lagrangian in the objective function of the SQP problem (this situation was encountered in Case Study 1, Scenario 3). On the other hand, difficulties in approximating the constraint-adaptation solution might arise if the constraints that become active or inactive are nonlinear. If an active nonlinear constraint becomes inactive, its curvature will continue affecting the objective function of the SQP approximation until the next RTO execution takes place. Conversely, if an inactive nonlinear constraint becomes active, its curvature will not be captured in the objective function of the SQP approximation. This last situation was encountered in Case Study 3, where an inactive quadratic constraint becomes active. One might think that these difficulties may be overcomed if a more comprehensive nonlinear approximation is used at the SSTO stage. In [5] it has been suggested that, rather than using a quadratic approximation, the best solution would be to solve the NLP RTO problem at every sampling time of the MPC regulator (if this was computationally possible). However, we would like to point out that this would imply loosing the consistency between the SSTO and MPC models, and with it, the zero steady-state offset and steady-state feasibility properties, discussed in Section 2.4, would no longer hold. We view these properties as the main reasons for including the SSTO stage in the first place.
- The steady-state points reached by the SSTO-MPC system do not in general match the RTO setpoints. However, the three SSTO designs studied have the ability to match the RTO setpoints upon convergence of the RTO iterations. In the case of Design C, this requires to adapt the MPC model at each RTO iteration, so as to match the output gradients of the RTO model.
- Also, Proposition 1 shows that, in order to match the RTO setpoints upon convergence of the RTO iterations there are conditions that the RTO model adaptation scheme must satisfy. These conditions can be easily met using bias corrections of the constraints at the RTO layer (constraint adaptation).

Future developments may consider a more comprehensive theoretical analysis of the overall RTO/SSTO-MPC system, including the analysis of sufficient conditions under which a given SSTO design is guaranteed to reach the RTO optimum upon convergence, or how to deal with optimization problems for which the set of active constraints is ill conditioned.

References

[1] W. Findeisen, F. Bailey, M. Bryds, K. Malinowski, P. Tatjewski, A. Wozniak, Control and Coordination in Hierarchical Systems, John Wiley & Sons, New York, 1980.
[2] T.E. Marlin, A.N. Hrymak, Real-time operations optimization of continuous processes, in: AIChE Symposium Series – CPC-V, vol. 93, 1997, pp. 156–164.
[3] S.J. Qin, T.A. Badgwell, A survey of industrial model predictive control technology, Control Eng. Pract. 11 (2003) 733–764.
[4] M.L. Darby, M. Nikolaou, J. Jones, D. Nicholson, RTO: an overview and assessment of current practice, J. Process Contr. 21 (2011) 874–884.
[5] P. Tatjewski, Advanced Control of Industrial Processes. Structures and Algorithms, Springer-Verlag, London, 2007.
[6] J.F. Forbes, T.E. Marlin, Model accuracy for economic optimizing controllers: the bias update case, Ind. Eng. Chem. Res. 33 (1994) 1919–1929.
[7] B. Chachuat, A. Marchetti, D. Bonvin, Process optimization via constraints adaptation, J. Process Contr. 18 (2008) 244–257.
[8] S. Cao, R.R. Rhinehart, An efficient method for on-line identification of steady state, J. Process Contr. 5 (6) (1995) 363–374.
[9] S.A. Bhat, D.N. Saraf, Steady-state identification, gross error detection, and data reconciliation for industrial process units, Ind. Eng. Chem. Res. 43 (2004) 4323–4336.
[10] C.-M. Ying, B. Joseph, Performance and stability analysis of LP-MPC and QP-MPC cascade control systems, AIChE J. 45 (7) (1999) 1521–1534.
[11] K.R. Muske, J.B. Rawlings, Model predictive control with linear models, AIChE J. 39 (2) (1993) 262–287.
[12] K.R. Muske, Steady-state target optimization in linear model predictive control, in: Proceedings of the American Control Conference, Albuquerque, New Mexico, 1997, pp. 3597–3601.
[13] C.V. Rao, J.B. Rawlings, Steady states and constraints in model predictive control, AIChE J. 45 (6) (1999) 1266–1278.
[14] G. Gattu, S. Palavajjhala, D.B. Robertson, Are oil refineries ready for non-linear control and optimization? in: International Symposium on Process Systems Engineering and Control, Mumbai, India, 2003.
[15] J.V. Kadam, W. Marquardt, Integration of economical optimization and control for intentionally transient process operation, in: Assessment and Future Directions of Nonlinear Model Predictive Control, Vol. 358 of Lecture Notes in Control and Information Sciences, Springer, Berlin/Heidelberg, 2007, pp. 419–434.
[16] J. Nocedal, S.J. Wright, Numerical Optimization, Springer-Verlag, New York, 1999.
[17] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear Programming: Theory and Algorithms, 3rd ed., John Wiley and Sons, New Jersey, 2006.
[18] B. Chachuat, B. Srinivasan, D. Bonvin, Adaptation strategies for real-time optimization Comp. Chem. Eng. 33 (2009) 1557–1567.
[19] G. Pannocchia, J.B. Rawlings, Disturbance models for offset-free model predictive control, AIChE J. 49 (2) (2003) 426–437.
[20] U. Maeder, F. Borrelli, M. Morari, Linear offset-free model predictive control, Automatica 45 (2009) 2214–2222.
[21] D.Q. Mayne, J.B. Rawlings, C.V. Rao, P.O.M. Scokaert, Constrained model predictive control: stability and optimality, Automatica 36 (2000) 789–814.
[22] P.O.M. Scokaert, J.B. Rawlings, Feasibility issues in linear model predictive control, AIChE J. 45 (8) (1999) 1649–1659.
[23] A. Maarleveld, J.E. Rijnsdorp, Constraint control on distillation columns, Automatica 6 (1970) 51–58.

[24] C.E. Garcia, M. Morari, Optimal operation of integrated processing systems. Part II: Closed-loop on-line optimizing control, AIChE J. 30 (2) (1984) 226–234.
[25] Y. Arkun, G. Stephanopoulos, Studies in the synthesis of control structures for chemical processes: Part IV. Design of steady-state optimizing control structures for chemical process units, AIChE J. 26 (6) (1980) 975–991.
[26] P. Tatjewski, M.A. Brdyś, J. Duda, Optimizing control of uncertain plants with constrained feedback controlled outputs, Int. J. Control 74 (15) (2001) 1510–1526.
[27] G. François, B. Srinivasan, D. Bonvin, Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty, J. Process Contr. 15 (6) (2005) 701–712.
[28] L.T. Biegler, I.E. Grossmann, A.W. Westerberg, A note on approximation techniques used for process optimization, Comp. Chem. Eng. 9 (2) (1985) 201–206.
[29] A. Marchetti, B. Chachuat, D. Bonvin, Modifier-adaptation methodology for real-time optimization, Ind. Eng. Chem. Res. 48 (13) (2009) 6022–6033.
[30] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.
[31] T.J. Williams, R.E. Otto, A generalized chemical processing model for the investigation of computer control, AIEE Trans. 79 (1960) 458.
[32] Y. Zhang, J.F. Forbes, Extended design cost: a performance criterion for real-time optimization systems, Comp. Chem. Eng. 24 (2000) 1829–1841.