Computers and Chemical Engineering 115 (2018) 246–261

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Process dynamics, control and monitoring

Active directional modifier adaptation for real-time optimization

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

[a] Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
[b] Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

ARTICLE INFO

Article history:
Received 12 September 2017
Revised 12 February 2018
Accepted 18 February 2018

Keywords:
Real-time optimization
Plant-model mismatch
Modifier adaptation
Input dimension reduction

ABSTRACT

Modifier adaptation is a real-time optimization (RTO) methodology that uses plant gradient estimates to correct model gradients, thereby driving the *plant* to optimality. However, obtaining accurate gradient estimates requires costly plant experiments at each RTO iteration. In directional modifier adaptation (DMA), the model gradients are corrected only in a small subspace of the input space, thus requiring fewer plant experiments. DMA selects the input subspace *offline* based on the *local* sensitivity of the Lagrangian gradient with respect to the uncertain model parameters. Here, we propose an extension, whereby the input subspace is selected *at each RTO iteration* via *global* sensitivity analysis, thus making the approach more reactive to changes and robust to large parametric uncertainties. Simulation results performed on the run-to-run optimization of two different semi-batch reactors show that the proposed approach finds a nice balance between experimental cost and optimality.

1. Introduction

Industrial plants target at optimizing process economics, while respecting operational constraints such as those on product quality, safety, and environmental regulations. In the presence of plant-model mismatch and process disturbances, real-time optimization (RTO) plays a pivotal role toward operating the plant optimally. RTO typically relies on the accuracy of the process model and/or the availability of plant measurements. RTO strategies differ in the way they exploit the available data and the model to update the operating point. For instance, the most common RTO strategy proceeds by first adapting the model parameters using experimental data and then optimizing the plant economics over the adapted model. This iterative approach is known as the two-step approach (Chen and Joseph, 1987). The two-step approach is intuitive and has become industrial practice in many process industries (Naysmith and Douglas, 1995). However, this approach typically converges to a sub-optimal solution in the presence of structural plant-model mismatch (Forbes and Marlin, 1996; Marchetti, 2009).

An alternative RTO strategy consists in adapting the optimization problem directly, while keeping the model parameters at their nominal values. This involves the adaptation of bias terms added to the constraints of the optimization problem (zeroth-order corrections). This approach, which is known as constraint adaptation (Chachuat et al., 2008), has shown promising results on an experimental solid-oxide fuel cell setup developed for industrial use (Bunin et al., 2012). In addition to these bias corrections, modifier-adaptation (MA) schemes include (first-order) gradient correction terms in the cost and constraint functions of the optimization problem (Marchetti et al., 2009). MA represents an appealing solution in the presence of plant-model mismatch as it guarantees the satisfaction of the plant first-order Karush–Kuhn–Tucker (KKT) conditions upon convergence. For the implementation of MA, plant measurements are expected to be sufficiently rich to allow good estimates of the plant cost and constraint values and of their gradients. The most straightforward way of estimating gradients is via finite differences, which requires evaluating the plant outputs at several (perturbed) operating points. The required number of perturbed points depends on the number of inputs and, as a consequence, the experimental cost of gradient estimation increases with increasing input dimension.

In the past years, several methods have been proposed to obtain gradient information. In dual MA (Marchetti et al., 2010), one considers an additional constraint in the RTO problem, which restricts the location of the next RTO inputs such that reliable gradient information can be extracted using the current and previously visited operating points. Dual ISOPE (Brdyś and Tatjewski, 2005) and

* Corresponding author.
E-mail addresses: martand.singhal@epfl.ch (M. Singhal), marchetti@cifasis-conicet.gov.ar (A.G. Marchetti), timm.faulwasser@kit.edu (T. Faulwasser), dominique.bonvin@epfl.ch (D. Bonvin).
[1] Current address: French-Argentine International Center for Information and Systems Sciences (CIFASIS), CONICET-Universidad Nacional de Rosario (UNR), S2000EZP Rosario, Argentina.

https://doi.org/10.1016/j.compchemeng.2018.02.016
0098-1354/

the approach proposed by Rodger and Chachuat (2011) also make use of 'duality constraints' so as to simultaneously estimate gradients and optimize the plant. Recently, Gao et al. (2016) proposed to combine a quadratic approximation used in derivative-free optimization (Conn et al., 2009) and MA to improve the quality of gradient estimates in the presence of noise. Alternatively, instead of estimating gradients, one can attempt to directly compute the first-order correction terms using an additional optimization layer as proposed by Navia et al. (2015). We refer to Marchetti et al. (2016) and the references therein for a detailed literature overview on MA.

Recently, Costello et al. (2016) proposed a MA approach that reduces the burden of gradient estimation by questioning the necessity of correcting in all input directions. The approach, labeled directional modifier adaptation (DMA), proposes to correct the model gradients only in 'privileged' directions that span a reduced subspace of the input space. This subspace is computed once offline by means of a local sensitivity analysis conducted on the gradient of the Lagrangian function predicted by the model. The sensitivities are evaluated with respect to variations around the nominal values of the model parameters.

In this paper, we extend the concept of DMA to cover the case where the parametric uncertainty is not local, but belongs to a fairly large uncertainty set. In this case, we argue that correcting the gradients only in the privileged directions identified offline via local sensitivity analysis may result in significant sub-optimality. Instead, we propose here to perform a global sensitivity analysis using ideas derived from *active subspaces* (Constantine, 2015; Russi, 2010). The concept of active subspaces has emerged as a set of techniques for reducing the dimension of the input space. Similar ideas are used in this paper to develop an *active* directional modifier-adaptation (ADMA) algorithm.

The contribution of this paper is in establishing the theoretical foundations of ADMA via the concepts derived from active subspaces. We extend our preliminary work described in Singhal et al. (2017) by providing a formal analysis of optimality upon convergence in ADMA. We discuss the practical aspects of ADMA and we demonstrate the effectiveness of the algorithm for the run-to-run optimization of two different semi-batch reactors.

The paper is structured as follows. Preliminary material including the formulation of the optimization problem, the description of the MA and DMA schemes, and background elements from active subspace theory, are presented in Section 2. The novel RTO approach that deals with large parametric uncertainty is then proposed in Section 3. In Section 4, two case studies dealing with semi-batch reactors are presented. The first case study considers only parametric uncertainty, while the second study deals with structural plant-model mismatch. We conclude the paper in Section 5.

2. Preliminaries

2.1. Problem formulation

The plant optimization problem can be written mathematically as:

$$\min_{\mathbf{u}} \quad \Phi_p(\mathbf{u}) := \phi(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \tag{1a}$$

$$\text{s.t.} \quad \mathrm{G}_{p,i}(\mathbf{u}) := \mathrm{g}_i(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \leq 0, \quad i = 1, \ldots, n_\mathrm{g}, \tag{1b}$$

where $\mathbf{u} \in \mathbb{R}^{n_\mathrm{u}}$ is the vector of input variables, $\mathbf{y}_p \in \mathbb{R}^{n_\mathrm{y}}$ are the measured output variables, ϕ: $\mathbb{R}^{n_\mathrm{u}} \times \mathbb{R}^{n_\mathrm{y}} \rightarrow \mathbb{R}$ is the cost to be minimized, g_i: $\mathbb{R}^{n_\mathrm{u}} \times \mathbb{R}^{n_\mathrm{y}} \rightarrow \mathbb{R}$, $i = 1, \ldots, n_\mathrm{g}$, are the inequality constraints. The solution to Problem (1) is denoted $\mathbf{u}_p^\star$.

The main challenge in solving this optimization problem stems from the fact that the input-output mapping $\mathbf{y}_p(\mathbf{u})$ is unknown. However, an approximate process model is assumed to be available, which gives the input-output mapping $\mathbf{y}(\mathbf{u}, \theta)$, where $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ are the model parameters. Then, using the model, Problem (1) can be approximated as:

$$\min_{\mathbf{u}} \quad \Phi\big(\mathbf{u}, \boldsymbol{\theta}\big) := \phi\big(\mathbf{u}, \mathbf{y}\big(\mathbf{u}, \boldsymbol{\theta}\big)\big) \tag{2a}$$

$$\text{s.t.} \quad \mathrm{G}_i(\mathbf{u}, \boldsymbol{\theta}) := \mathrm{g}_i(\mathbf{u}, \mathbf{y}(\mathbf{u}, \boldsymbol{\theta})) \leq 0, \quad i = 1, \ldots, n_\mathrm{g}. \tag{2b}$$

The nominal solution $\mathbf{u}^\star$ is found by solving Problem (2) for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is the vector of nominal model parameters. In the presence of plant-model mismatch, the model optimum $\mathbf{u}^\star$ may not be equal to the plant optimum $\mathbf{u}_p^\star$. The goal of RTO is to find $\mathbf{u}_p^\star$ by iteratively modifying and solving Problem (2).

2.2. Modifier adaptation

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

$$\min_{\mathbf{u}} \quad \Phi_{m,k}(\mathbf{u}) := \Phi(\mathbf{u}, \boldsymbol{\theta}) + (\boldsymbol{\lambda}_k^\Phi)^\mathrm{T}\mathbf{u} \tag{3a}$$

$$\text{s.t.} \quad \mathbf{G}_{m,k}(\mathbf{u}) := \mathbf{G}(\mathbf{u}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_k^\mathrm{G} + (\boldsymbol{\lambda}_k^\mathrm{G})^\mathrm{T}(\mathbf{u} - \mathbf{u}_k) \preceq \mathbf{0}, \tag{3b}$$

where $\mathbf{G} \in \mathbb{R}^{n_\mathrm{g}}$ is the vector of constraints G_i, $i = 1, \ldots, n_\mathrm{g}$; $\boldsymbol{\varepsilon}_k^\mathrm{G} \in \mathbb{R}^{n_\mathrm{g}}$ is the vector of zeroth-order modifiers for the constraints; and $\boldsymbol{\lambda}_k^\Phi \in \mathbb{R}^{n_\mathrm{u}}$ and $\boldsymbol{\lambda}_k^\mathrm{G} \in \mathbb{R}^{n_\mathrm{u} \times n_\mathrm{g}}$ are the first-order modifiers for the cost and constraint functions, respectively. At the kth RTO iteration, the modifiers are computed as follows:

$$\boldsymbol{\varepsilon}_k^\mathrm{G} = \mathbf{G}_p(\mathbf{u}_k) - \mathbf{G}(\mathbf{u}_k, \boldsymbol{\theta}), \tag{4a}$$

$$(\boldsymbol{\lambda}_k^\Phi)^\mathrm{T} = \nabla_\mathbf{u}\Phi_p(\mathbf{u}_k) - \nabla_\mathbf{u}\Phi(\mathbf{u}_k, \boldsymbol{\theta}), \tag{4b}$$

$$(\boldsymbol{\lambda}_k^\mathrm{G})^\mathrm{T} = \nabla_\mathbf{u}\mathbf{G}_p(\mathbf{u}_k) - \nabla_\mathbf{u}\mathbf{G}(\mathbf{u}_k, \boldsymbol{\theta}), \tag{4c}$$

where $\nabla_\mathbf{u}(\cdot)$ is the gradient of a scalar-valued function or the Jacobian of a vector-valued function with respect to $\mathbf{u}$. MA guarantees meeting the plant KKT conditions of Problem (1) upon convergence (Marchetti et al., 2009). Gradient adaptation via first-order modifiers plays a key role in meeting the plant KKT conditions. However, finding reliable plant gradients is a costly task as it requires additional plant evaluations. If, for instance, the forward finite-difference approach is used, then the number of plant evaluations at each RTO iteration increases linearly with the dimension of the input space.

2.3. Directional modifier adaptation

The dependency of MA on the knowledge of full plant gradients can be reduced with the help of a process model. As the model gradients are sensitive to model parameters, the input subspace in which the parametric uncertainty has the most influence on the solution to Problem (2) can be found via *local* sensitivity analysis. In Costello et al. (2016), this subspace is spanned by the so-called 'privileged directions' for the purpose of gradient estimation. DMA evaluates offline the sensitivity of the model Lagrangian gradient with respect to local parametric variations that are evaluated at the model optimum. To this end, the model Lagrangian function is defined as

$$\mathcal{L}(\mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\theta}) := \Phi(\mathbf{u}, \boldsymbol{\theta}) + \boldsymbol{\mu}^\mathrm{T}\mathbf{G}(\mathbf{u}, \boldsymbol{\theta}), \tag{5}$$

with $\boldsymbol{\mu} \in \mathbb{R}^{n_\mathrm{g}}$ the vector of Lagrange multipliers. Then, the sensitivity matrix $\boldsymbol{A}^\star \in \mathbb{R}^{n_\mathrm{u} \times n_\theta}$ is computed as follows:

$$\boldsymbol{A}^\star := \nabla_{\mathbf{u}\boldsymbol{\theta}}\mathcal{L}(\mathbf{u}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\theta}_0) = \left.\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta}\, \partial \mathbf{u}}\right|_{\mathbf{u}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\theta}_0}, \tag{6}$$