Computers and Chemical Engineering 91 (2016) 318–328

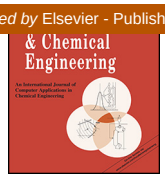
View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Computers and Chemical Engineering

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



A reliable modifier-adaptation strategy for real-time optimization

Weihua Gao*, Simon Wenzel, Sebastian Engell

Biochemical and Chemical Engineering Department, TU Dortmund, Emil-Figge-Str. 70, 44221 Dortmund, Germany

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form 19 May 2016
Accepted 24 May 2016
Available online 13 April 2016

Keywords:
Real-time optimization
Model mismatch
Quadratic approximation
Modifier adaptation

ABSTRACT

In model-based real-time optimization, plant-model mismatch can be handled by applying bias- and gradient-corrections to the cost and constraint functions in an iterative optimization procedure. One of the major challenges in practice is the estimation of the plant gradients from noisy measurement data, in particular for several optimization variables. In this paper we propose a new real-time optimization scheme that explores the inherent smoothness of the plant mapping to enable a reliable optimization. The idea here is to combine the quadratic approximation approach used in derivative-free optimization techniques with the iterative gradient-modification optimization scheme. The convergence of the scheme is analyzed. Simulation studies for the optimization of a ten-variable synthetic example and a reactor benchmark problem with considerable plant-model mismatch show its promising performance.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

With increasing global competition, companies in the process industries face intense pressure to improve production efficiency, product quality and process safety. As a result, real-time optimization (RTO) is attracting considerable industrial interest. RTO is a model based upper-level optimization system that is operated iteratively in closed loop and provides set-points to the lower-level regulatory control system in order to maintain the process operation as close as possible to the economic optimum. The introduction of the RTO system provides a clear separation between the tasks of optimization and control. The RTO system optimizes the plant economics on a medium timescale while the control system provides tracking and disturbance rejection on shorter timescales from seconds to hours. The economical optimization is normally based on a steady-state model of the plant while on the regulatory layer, simple controllers or linear MPC controllers are used.

As for any model and optimization based scheme, the success of RTO depends on the quality of the model which is used in the optimization. The effort required for building and maintaining the model is the bottleneck in the deployment of RTO solutions, and even when sophisticated models are used, they will never exactly represent the real process. It is highly desirable to combine the use of models and of the data which is collected during the operation of the plant in order to obtain a RTO scheme that drives the plant to its optimal operation without having to represent each and every phenomenon in the plant accurately in the model.

Several schemes have been proposed towards this goal, in particular the parameter adaptation or two-step scheme (Chen and Joseph, 1987), bias update with constraints (Forbes and Marlin, 1994), tracking of necessary conditions of optimality (NCO tracking) which are derived from the model but are used in a model-free scheme (François et al., 2005), extremum seeking control (Ariyur and Krstic, 2003), and the adaptation of the optimization objective (and constraints) using empirical gradients (Roberts, 1979; Gao and Engell, 2005; Marchetti et al., 2010). Derivative-free optimization (DFO) can also be considered as a RTO technique that determines the optimum using "probing" of the response of the process (Conn et al., 2009).

These schemes all have certain drawbacks and limitations. Parameter adaptation only works well if the process model is structurally correct and if the parameters can be estimated reliably from the available data. NCO tracking is well suited for not too large problems where the optimum is defined largely by the constraints. Extremum seeking control converges slowly and is not suitable for large problems with many degrees of freedom. DFO is known for less efficient iterations compared with derivative-based schemes.

Gradient adaptation (or modifier adaptation) schemes can handle considerable plant-model mismatch by applying empirical bias- and gradient-corrections to the objective and constraint functions in an iterative optimization procedure. One of the major challenges in practice is the estimation of the process gradients from noisy measurement data, in particular for the case of several optimization variables. Finite-difference based approaches suffer from the problem of choosing the right step-size, using a large step-size may

* Corresponding author.
E-mail address: weihua.gao@bci.tu-dortmund.de (W. Gao).

http://dx.doi.org/10.1016/j.compchemeng.2016.03.019
0098-1354/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

decrease the effect of noisy data on the one hand side, on the other hand side it leads to considerable approximation errors. In this paper we propose a new scheme that employs a quadratic approximation of the process mapping to enable an efficient and reliable optimization even if the data is subject to a high level of noise. The key idea is to combine the quadratic approximation approach used in DFO with the iterative gradient-modification optimization (IGMO) scheme (Gao and Engell, 2005).

The rest of this paper is organized as following. Firstly the IGMO scheme is reviewed together with a survey of approaches to estimate the empirical gradients from data, and the effect of the error of the gradients on the IGMO performance is studied. The DFO scheme is then introduced, followed by a detailed description of the new scheme. Its convergence is analyzed. Simulation results for the optimization of a ten-variable synthetic example and the Williams-Otto reactor benchmark problem with considerable plant-model mismatch are presented to illustrate the performance of the new scheme.

2. Iterative gradient-modification optimization and gradient estimation

2.1. The IGMO scheme

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

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

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

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_{ad}^{(k)}(\mathbf{u}) = J_m(\mathbf{u}) + J_p^{(k)} - J_m^{(k)} + \left(\nabla J_p^{(k)} - \nabla J_m^{(k)}\right)^T \left(\mathbf{u} - \mathbf{u}^{(k)}\right) \\ \text{s.t.} \quad & \mathbf{C}_{ad}^{(k)}(\mathbf{u}) = \mathbf{C}_m(\mathbf{u}) + \mathbf{C}_p^{(k)} - \mathbf{C}_m^{(k)} + \left(\nabla \mathbf{C}_p^{(k)} - \nabla \mathbf{C}_m^{(k)}\right)^T \left(\mathbf{u} - \mathbf{u}^{(k)}\right) \le \mathbf{0}, \end{aligned} \tag{2}$$

where the superscript (k) represents the iteration index. Note that the bias adaptation of the objective function does not influence the optimum and it just corrects the objective value at $\mathbf{u}^{(k)}$. Let $\hat{\mathrm{u}}^{(k)}$ denote the solution of (2), then the next set-point is updated as

$$\mathbf{u}^{(k+1)} = \hat{\mathrm{u}}^{(k)} + \mathbf{K}(\mathbf{u}^{(k)} - \hat{\mathrm{u}}^{(k)}), \tag{3}$$

where $\mathbf{K}$ is a diagonal matrix of damping factors

$$0 \le k_{ii} < 1. \tag{4}$$

The IGMO scheme generates a sequence of set-points that converges to a Karush–Kuhn–Tucker point of the plant (Kuhn and Tucker, 1951).

2.2. Estimation of empirical gradients

The use of gradient adaptation in the IGMO scheme requires the computation of the process gradients at the current set-point. In order to decrease the effort for perturbations of the process, the collected data at the previous set-points can be used to estimate the gradients by a finite difference approximation

$$\nabla J_p^{(k)} = \left[\mathbf{S}^{(k)}\right]^{-1} \cdot \left[\left(J_p^{(k)} - J_p^{(k-1)}\right) \cdots \left(J_p^{(k)} - J_p^{(k-n_u)}\right)\right]^T \tag{5}$$

$$\nabla \mathbf{C}_p^{(k)} = \left[\mathbf{S}^{(k)}\right]^{-1} \cdot \left[\left(\mathbf{C}_p^{(k)} - \mathbf{C}_p^{(k-1)}\right) \cdots \left(\mathbf{C}_p^{(k)} - \mathbf{C}_p^{(k-n_u)}\right)\right]^T, \tag{6}$$

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

$$\mathbf{S}^{(k)} = \left[\left(\mathbf{u}^{(k)} - \mathbf{u}^{(k-1)}\right) \cdots \left(\mathbf{u}^{(k)} - \mathbf{u}^{(k-n_u)}\right)\right]^T. \tag{7}$$

Since the matrix $\mathbf{S}^{(k)}$ can be singular, Roberts (2000) proposed to use Broyden's formula to evaluate the gradients. A dual control approach that implicitly considers the requirement of the gradient estimation in the set-point optimization was proposed by Brdyś and Tatjewski (1994). The kth set-point is required to additionally satisfy a lower bound on the inverse of the condition number of the matrix $\mathbf{S}^{(k)}$

$$\kappa^{-1}(\mathbf{S}^{(k)}) \ge \delta_{cond}. \tag{8}$$

Eq. (8) leads to a dual control optimization of the next set-point that balances the convergence to the optimum with the gradient estimation. Marchetti et al. (2010) extended this approach by explicitly upper bounding the norm of the gradient estimation error that consists of the approximation and the measurement noise errors

$$\frac{\sigma_{\max}}{2} \|\mathbf{S}^{-1} \cdot \mathrm{diag}\left(\mathbf{S} \cdot \mathbf{S}^T\right)\| + \frac{\delta_{noise}}{l_{\min}} \le \epsilon_{upper}, \tag{9}$$

where the first term on the left side of (9) is the approximation error (σ_{max} is the upper bound on the spectral radius of the Hessian matrix of the process mapping), and the second term is the error due to the measurement noise (l_{min} is the shortest distance between all possible pairs of complement affine subspaces that can be generated from the set-point set for the gradient estimation, and δ_{noise} is the range of the measurement noise). For simplicity, the superscript of the iteration index was dropped here.

In the IGMO scheme, Gao and Engell (2005) used (8) to decide whether to perturb the process additionally. The perturbation $\mathbf{u}^{(addi)}$ is optimized by maximizing the inverse of the condition number of $\mathbf{S}^{a(k)}$ subject to the adapted process constraints

$$\begin{aligned} \max_{\mathbf{u}^{(addi)}} \quad & \kappa^{-1}(\mathbf{S}^{a(k)}) \\ \text{s.t.} \quad & \mathbf{C}_m(\mathbf{u}^{(addi)}) + \mathbf{C}_p^{(k-1)} - \mathbf{C}_m^{(k-1)} + (\nabla \mathbf{C}_p^{(k-1)} - \nabla \mathbf{C}_m^{(k-1)})^T (\mathbf{u}^{(addi)} - \mathbf{u}^{(k-1)}) \le \mathbf{0}, \end{aligned} \tag{10}$$

where

$$\mathbf{S}^{a(k)} = \left[\left(\mathbf{u}^{(k)} - \mathbf{u}^{(addi)}\right) \cdots \left(\mathbf{u}^{(k)} - \mathbf{u}^{(k-n_u+1)}\right)\right]^T. \tag{11}$$

2.3. Effect of gradient error on IGMO

Since the use of empirical gradients plays a critical role in the IGMO scheme, it is important to analyze the effect of the gradient error on the IGMO performance. From (9), the set-points should neither be too distant nor too close to each other for a good gradient estimation. How the step-size influences the IGMO can be demonstrated by the optimization problem

$$\min_{u_{1,2}} \quad J_m(u_1, u_2) = (u_1 + 2)^2 + (u_2 + 2)^2. \tag{12}$$

The data for gradient estimation are generated by the "true" process mapping

$$J_p(u_1, u_2) = u_1^2 + u_2^2 + \nu, \tag{13}$$

where ν is the noise with a normal distribution $N(0.0, 0.1)$. The IGMO scheme starts from $(-2.0, -2.0)$ and for simplicity, we consider only one Newton step with the Hessian matrix computed at $J_m(u_1, u_2)$. Three different sets of set-points are compared:

(i) $\{(-2.0, 0.0), (0.0, -2.0)\}$, $\nabla J_p^{(i)} : (-2.0, -2.0)$
(ii) $\{(-2.0, -1.0), (-1.0, -2.0)\}$, $\nabla J_p^{(ii)} : (-3.0, -3.0)$
(iii) $\{(-2.0, -1.8), (-1.8, -2.0)\}$, $\nabla J_p^{(iii)} : (-3.8, -3.8)$

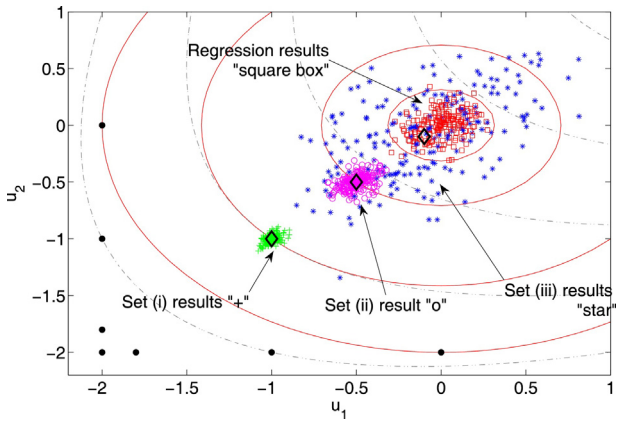


Fig. 1. Illustration of the IGMO results on the process contours (solid lines). Dash-dot lines: contours of the quadratic approximation. Adapted from (Gao et al., 2015).

where $\nabla J_p^{(i)}$, $\nabla J_p^{(ii)}$ and $\nabla J_p^{(iii)}$ are calculated from noise-free data. The exact gradient is $(-4.0, -4.0)$ and therefore the smallest step size provides the most accurate gradient estimation. The results of IGMO in case of noise-free data are represented by the diamond markers in Fig. 1. The largest movement to the "process" optimum at (0, 0) is obtained with the most accurate gradient estimation. When noisy data is used in the gradient estimation, IGMO generates set-point moves with different levels of precision. For each set, the gradients were repeatedly estimated for 200 noise realizations and the results of IGMO are represented by three clusters of points in Fig. 1. Now the smallest step size leads to less precise set-point moves (mean: -0.08, -0.12, variances: 0.12, 0.11). Some of the values are even worse than the IGMO results based on set (ii).

The idea of our new scheme is to regress all available data to obtain a quadratic approximation and to evaluate the gradient at $(-2, -2)$ analytically from it. The IGMO results for 200 noise realizations with this approach are shown by the square symbols in Fig. 1, leading to much better accuracy (means: 0.008, -0.010) and precision (variances: 0.015, 0.017) than the results using set (iii). Compared with the finite-difference approximation, the quadratic approximation can, on the one hand, capture the curvature information from more distant points to decrease the approximation error, and, on the other hand, provides a smooth mapping to decrease the influence of the noise. Note that there is a considerable mismatch between the contours of the quadratic approximation that were regressed from the noise-free data and those of the process mapping. The reason for this is explained in the next section. An optimization only based on the quadratic approximation will therefore also not lead to convergence to the true optimum. An improvement of the performance can be expected from the combination of the locally accurate gradients extracted from the data with the globally more accurate Hessian matrix that is predicted by the model in the IGMO scheme.

3. Derivative-free optimization and its combination with IGMO

3.1. Derivative-free optimization

The interpolation-based trust-region DFO methods provide reliable convergence in the presence of data errors and a relatively fast rate of convergence (Conn et al., 2009). These methods resemble the derivative-based trust-region methods, except that the values of the objective function are used to construct a quadratic function, on which the optimization step is based. Global convergence is ensured via repeatedly constructing local quadratic functions so that a sufficient decrease of the objective function can be obtained within each iteration. The rate of convergence is mainly determined by how well the objective function can be represented by the quadratic function and how many evaluations of the objective function are used to make the approximation sufficiently accurate. The geometry (or distribution of points) of the interpolation set is critical to the quality of the quadratic approximation. This was also evident from the example in the previous section. Although the true objective function is also quadratic, the approximating quadratic function is not accurate because the selected points did not cover the space of the variables sufficiently well. To ensure the so-called well-poisedness of the data set is very important in DFO.

Improving the IGMO by gradients that are extracted from quadratic approximations in our approach is complemented by three additional elements:

1. Selecting points from the collected data set for a well-poised regression set.
2. Introducing a constrained search space for the next move basing on a covariance analysis of the regression set.
3. Tracking the prediction accuracies of the adapted model-based mappings ($J_{ad}(\mathbf{u})$ and $\mathbf{C}_{ad}(\mathbf{u})$) and the approximating quadratic functions, and switching between model-based and data-based optimizations according to the observed accuracies.

Next the choice of the regression set and the evaluation of the search space are discussed in detail. The accuracy tracking and the optimization switching are presented in the algorithm section.

3.2. Selecting data for regression

The geometry of the regression set is critical for the accuracy of the quadratic approximation. The criteria for choosing the regression set can be summarized as follows:

1. Well-distributed and sufficiently distant points are indispensable for capturing the curvature reliably from noisy data.
2. The use of many points in a neighborhood can improve the accuracy of the gradient estimation.
3. Avoid to use too old points in the presence of varying disturbances.

Let $\mathbb{U}$ represent the set of all the collected data. According to the criteria 1 and 2, the regression set is defined as $\mathcal{U}^{(k)} = \mathcal{U}_{nb} \cup \mathcal{U}_{dist}$, where the neighboring set

$$\mathcal{U}_{nb} = \{\mathbf{u} : \|\mathbf{u} - \mathbf{u}^{(k)}\| \leq \Delta\mathbf{u}; \mathbf{u} \in \mathbb{U}\}$$

and the distant set $\mathcal{U}_{dist}$ are determined by

$$\min_{\mathcal{U}_{dist}} \quad \frac{\sum_{\mathbf{u}\in\mathcal{U}_{dist}} \|\mathbf{u} - \mathbf{u}^{(k)}\|}{\varphi(\mathcal{U}_{dist})} \qquad \text{s.t.} \quad \text{size}(\mathcal{U}_{dist}) = (n_u + 1)(n_u + 2)/2 - 1 \quad \mathcal{U}_{dist} \subset \mathbb{U} \setminus \mathcal{U}_{nb}. \tag{14}$$

$\Delta\mathbf{u}$ is a parameter of the screening algorithm and $\varphi(\mathcal{U}_{dist})$ is the minimal angle between all possible vectors that are defined by $\mathbf{u} - \mathbf{u}^{(k)}$, where $\mathbf{u} \in \mathcal{U}_{dist}$. Let $\mathcal{U}_{outer} = \mathbb{U} \setminus \mathcal{U}_{nb}$. (14) can be approximately solved by Algorithm 1:

Here $\varphi^{(i)}(\mathcal{U}_{outer})$ is the minimal angle between $\mathbf{u}^{(i)} - \mathbf{u}^{(k)}$ and all the other vectors $\mathbf{u}^{(j)} - \mathbf{u}^{(k)}$, where $\mathbf{u}^{(j)} \subset \mathcal{U}_{outer} \setminus \mathbf{u}^{(i)}$. The effect of using the screening algorithm to select points is illustrated using an example with 2 decision variables in Fig. 2. The collected data consists of 25 random points which are drawn from the standard uniform distribution on the open interval (0, 1). The current point $\mathbf{u}^{(k)}$ is firstly located at (0.5, 0.5), and then is moved to (1.0, 0.5). The screening algorithm is run with $\Delta\mathbf{u} = 0.1$ and $\Delta\mathbf{u} = 0.3$, which are illustrated by the circles centered at $\mathbf{u}^{(k)}$. All the points within the circles are treated as local points and are included in $\mathcal{U}_{nb}$.

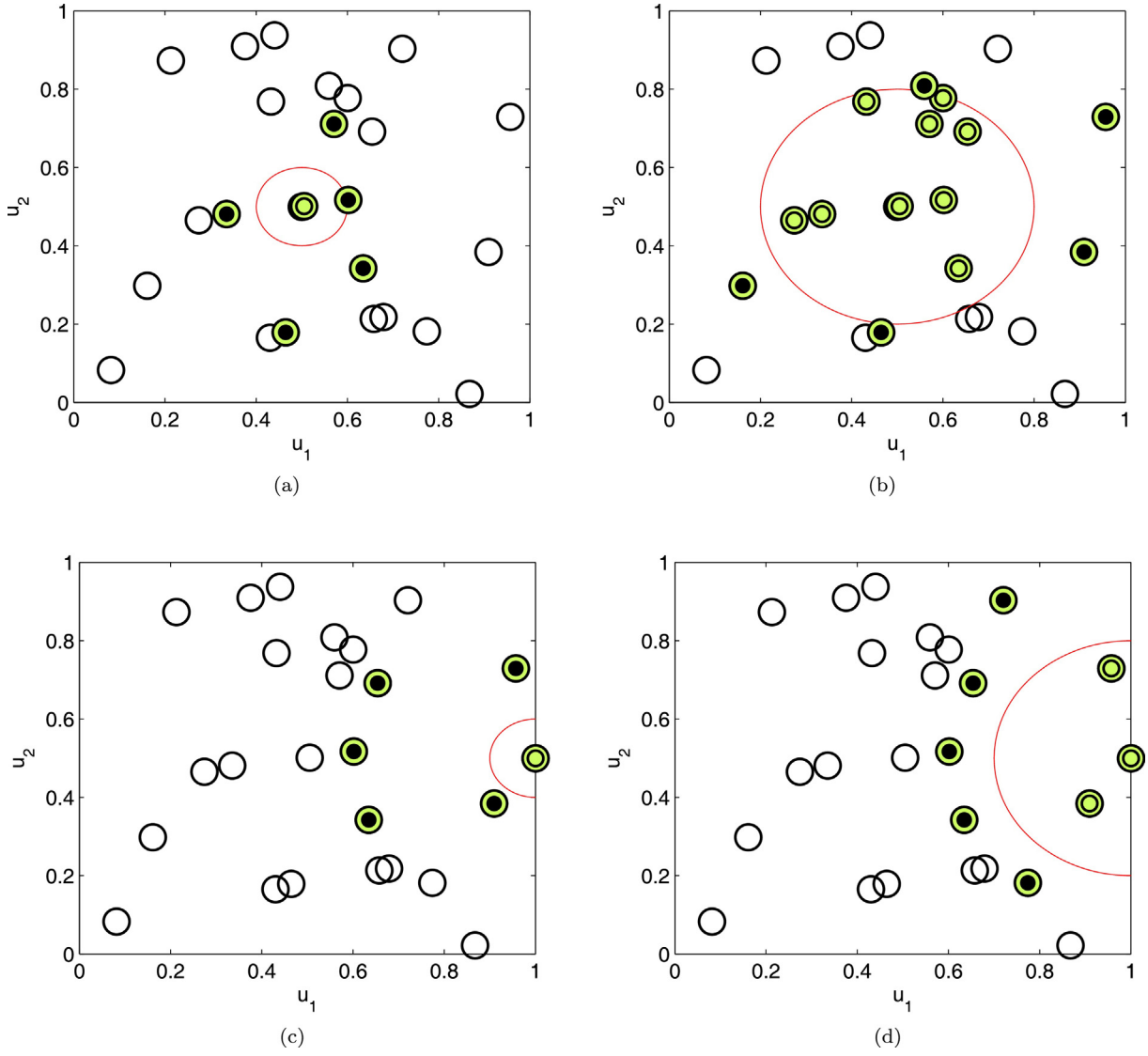


Fig. 2. Illustration of the screening of the regression set. ◉ local points, ◉ chosen outer points, ○ unchosen points.

Algorithm 1

$n_{reg} = (n_u+1)(n_u+2)/2$;
If $dim\mathcal{U}_{outer} > n_{reg} - 1$ **then**
  $w = \frac{n_{reg}-2}{dim\mathcal{U}_{outer}-1}$;
  for $i = 1 : dim\mathcal{U}_{outer}$ **do**
    Evaluate $\varphi^{(i)}(\mathcal{U}_{outer})$;
    $\zeta^{(i)} = \frac{\|\mathbf{u}^{(i)}-\mathbf{u}^{(k)}\|+w\sum_{\mathbf{u}\in\mathcal{U}_{outer}\setminus\mathbf{u}^{(i)}}\|\mathbf{u}-\mathbf{u}^{(k)}\|}{\varphi^{(i)}(\mathcal{U}_{outer})}$;
  end for
  Find $\mathbf{u}^{(worst)}$: point with the maximal ζ;
  $\mathcal{U}_{outer} \leftarrow \mathcal{U}_{outer} \setminus \mathbf{u}^{(worst)}$;
end if
$\mathcal{U}_{dist} \leftarrow \mathcal{U}_{outer}$;

The points outside are screened by the proposed algorithm until $(n_u+1)(n_u+2)/2-1=5$ points are selected for $\mathcal{U}_{dist}$. Different symbols are used in Fig. 2 to differentiate $\mathcal{U}_{nb}$, $\mathcal{U}_{dist}$ and the unchosen set.

As shown in Fig. 2, the value of $\Delta\mathbf{u}$ influences the screened regression set. In the presence of noise, a larger value of $\Delta\mathbf{u}$ is preferred for reliably capturing the curvature information. However, there is no guarantee that the plant mapping is quadratic and a wide distribution of the regression set may lead to a large approximation error. A practical method to choose $\Delta\mathbf{u}$ is to relate its value to the level of noise.

With the screened regression set $\mathcal{U}^{(k)} = \{\mathbf{u}^{(r_1)}, \ldots, \mathbf{u}^{(r_{n_r})}\}$, a quadratic approximation of the objective function is constructed by solving

$$\min_{\mathcal{P}} \sum_{i=1}^{n_r} \left(J_p\left(\mathbf{u}^{(r_i)}\right) - J_\phi\left(\mathbf{u}^{(r_i)}, \mathcal{P}\right)\right)^2, \tag{15}$$

where $J_\phi(\mathbf{u}, \mathcal{P})$ is the quadratic approximation. It is defined by

$$J_\phi(\mathbf{u}, \mathcal{P}) = \sum_{i=1}^{n_u}\sum_{j=1}^{i} a_{i,j}u_iu_j + \sum_{i=1}^{n_u} b_iu_i + c \tag{16}$$

with the coefficient set $\mathcal{P} = \left\{a_{1,1}, \ldots, a_{n_u,n_u}, b_1, \ldots, b_{n_u}, c\right\}$. The quadratic approximations of the constraint functions $\mathbf{C}_\phi(\mathbf{u})$ are

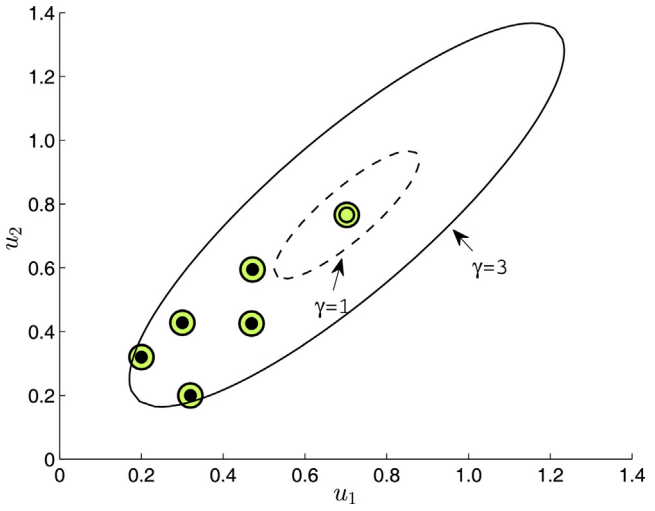


Fig. 3. Illustration of the covariance-based search space. ◎ current point $\mathbf{u}^{(k)}$, ◉ regression points.

constructed similarly. The process gradients are then determined by evaluating ∇J_ϕ and $\nabla \mathbf{C}_\phi$ at $\mathbf{u}^{(k)}$.

3.3. Covariance-based search space definition

Trust-region methods are widely used in numerical optimization algorithms and the size of the trust-region is critical to the effectiveness of the optimization steps (Nocedal and Wright, 2006). In practical algorithms, the size of the trust-region is adjusted by a feedback mechanism according to the performance of the algorithm during previous iterations. The performance of the IGMO scheme is not only related to the model used, but also sensitive to the gradient modifiers, which are here extracted via the quadratic approximation from the regression set. Consider a 2-D optimization problem where the previous set-points are distributed along a specific direction. The extracted gradient along that direction is more reliable than that along the perpendicular direction. In this case, it is reasonable to define an elliptical search space rather than a circular search space. We propose to use the covariance matrix of the regression set $M = cov(\mathcal{U}^{(k)})$ to define a constrained search space $\mathcal{B}^{(k)}$ for the next set-point move:

$$\mathcal{B}^{(k)} : (\mathbf{u} - \mathbf{u}^{(k)})^T M^{-1} (\mathbf{u} - \mathbf{u}^{(k)}) \le \gamma^2, \tag{17}$$

where M is the covariance matrix and γ is a scaling parameter. $\mathcal{B}^{(k)}$ is a n_u-axial ellipsoid centered at $\mathbf{u}^{(k)}$. The axes of the ellipsoid are aligned with the eigenvectors of the covariance matrix. The semi-axis lengths of the ellipsoid are related to the eigenvalues of the covariance matrix by the scaling parameter γ.

An illustration of this idea for a 2-D optimization problem is shown in Fig. 3, where the regression points are shown together with a conservative definition ($\gamma = 1$, dashed line) and a normal definition ($\gamma = 3$, solid line) of the search space. γ can also be adjusted online according to prediction accuracies, as in trust-region approaches.

3.4. Algorithm

Let $\mathcal{U}^{(k)}$ and $\mathcal{B}^{(k)}$ represent the regression set and the search space at the kth iteration. $\rho_m^{(k)}$ and $\rho_\phi^{(k)}$ are the prediction accuracies of the adapted model-based mappings and the quadratic functions that are computed according to

$$\rho_m^{(k)} = \max \left\{ \left| 1 - \frac{J_{ad}^{(k)} - J_{ad}^{(k-1)}}{J_p^{(k)} - J_p^{(k-1)}} \right|, \left| 1 - \frac{C_{ad,1}^{(k)} - C_{ad,1}^{(k-1)}}{C_{p,1}^{(k)} - C_{p,1}^{(k-1)}} \right|, \ldots, \left| 1 - \frac{C_{ad,n_c}^{(k)} - C_{ad,n_c}^{(k-1)}}{C_{p,n_c}^{(k)} - C_{p,n_c}^{(k-1)}} \right| \right\} \tag{18}$$

$$\rho_\phi^{(k)} = \max \left\{ \left| 1 - \frac{J_\phi^{(k)} - J_\phi^{(k-1)}}{J_p^{(k)} - J_p^{(k-1)}} \right|, \left| 1 - \frac{C_{\phi,1}^{(k)} - C_{\phi,1}^{(k-1)}}{C_{p,1}^{(k)} - C_{p,1}^{(k-1)}} \right|, \ldots, \left| 1 - \frac{C_{\phi,n_c}^{(k)} - C_{\phi,n_c}^{(k-1)}}{C_{p,n_c}^{(k)} - C_{p,n_c}^{(k-1)}} \right| \right\}, \tag{19}$$

where J_{ad} and $C_{ad,i}$ are the adapted objective and constraint functions from (2). n_c is the number of constraints. J_ϕ and $C_{\phi,i}$ are the regressed objective and constraint functions from the quadratic approximation.

Step 1. Choose an initial set-point $\mathbf{u}^{(0)}$ and probe the plant at $\mathbf{u}^{(0)}$ and $\mathbf{u}^{(0)} + h\mathbf{e}_i$, where h is a suitable step size and $\mathbf{e}_i \in \mathbb{R}^{n_u} (i = 1, \ldots, n_u)$ are mutually orthogonal unit vectors. Use the finite difference approach to calculate the gradients at $\mathbf{u}^{(0)}$ and run IGMO until $[(n_u + 1)(n_u + 2)/2$ set-points have been generated. Run the screening algorithm to acquire the regression set $\mathcal{U}^{(k)}$. Initialize $\rho_m^{(k)} = 0$ and $\rho_\phi^{(k)} = 0$.

Step 2. Calculate the quadratic functions $J_\phi^{(k)}$ and $\mathbf{C}_\phi^{(k)}$ by (15). Determine the search space $\mathcal{B}^{(k)}$ by (17).

Step 3. Extract the gradients from the quadratic functions. Adapt the model-based optimization problem and determine $\hat{\mathbf{u}}^{(k)}$ as follows:

a) If $\rho_m^{(k)} \le \rho_\phi^{(k)}$, run the adapted model-based optimization (2) under the constraint $\mathbf{u} \in \mathcal{B}^{(k)}$.

b) Else perform an optimization based on the quadratic approximation

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_\phi^{(k)}(\mathbf{u}) \\ \text{s.t.} \quad & \mathbf{C}_\phi^{(k)}(\mathbf{u}) \le \mathbf{0} \\ & \mathbf{u} \in \mathcal{B}^{(k)}. \end{aligned} \tag{20}$$

Step 4. If $\|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^{(k)}\| < \Delta\mathbf{u}$, where $\Delta\mathbf{u}$ is the parameter used by the screening algorithm to handle the influence of measurement noise, and there exists at least one point $\mathbf{u}^{(j)} \in \mathcal{U}^{(k)}$ such that $\|\mathbf{u}^{(j)} - \mathbf{u}^{(k)}\| > 2\Delta\mathbf{u}$, set $\hat{\mathbf{u}}^{(k)} = (\mathbf{u}^{(j)} + \mathbf{u}^{(k)})/2$. Note that this step is similar as the criticality step used in (Conn et al., 2009) to improve the quadratic approximation by shrinking the regression region.

Step 5. Evaluate the process at $\hat{\mathbf{u}}^{(k)}$ to acquire $J_p(\hat{\mathbf{u}}^{(k)})$ and $\mathbf{C}_p(\hat{\mathbf{u}}^{(k)})$. Define the next iterate by

a) Successful iteration. If $\hat{J}_p^{(k)} < J_p^{(k)}$, where $\hat{J}_p^{(k)} = J_p(\hat{\mathbf{u}}^{(k)})$, define $\mathbf{u}^{(k+1)} = \hat{\mathbf{u}}^{(k)}$ and run the screening algorithm to define the next regression set $\mathcal{U}^{(k+1)}$. Update the quality indices $\rho_m^{(k+1)}$ and $\rho_\phi^{(k+1)}$. Increase k by one and go to Step 2.

b) Unsuccessful iteration. If $\hat{J}_p^{(k)} \ge J_p^{(k)}$, run the screening algorithm to update the regression set for $\mathbf{u}^{(k)}$ with the new probe $\hat{\mathbf{u}}^{(k)}$ in the collected data. Due to the screening algorithm, $\hat{\mathbf{u}}^{(k)}$ will be included in the regression set to achieve an improved quadratic approximation around $\mathbf{u}^{(k)}$. Go to Step 2.

Note that, since the level of measurement noise influences the choice of the value of $\Delta\mathbf{u}$, it will indirectly influence the accuracy of the optimum via Step 4 and $\|\mathbf{u}^{(\infty)} - \mathbf{u}^*\| < \Delta\mathbf{u}$. In case of noise-free optimization, we have $\mathbf{u}^{(\infty)} \to \mathbf{u}^*$ as $\Delta\mathbf{u} \to 0$. This will be proven in the next section.

3.5. Convergence analysis of the unconstrained optimization

In order to keep the convergence analysis simple and intuitive, we consider an unconstrained optimization problem with the following assumptions:

Assumption 1. *J_p is twice continuously differentiable and its gradient and Hessian are bounded, which implies that there exist constants $\kappa_{pg} > 0$ and $\kappa_{ph} > 0$ such that $\|\nabla J_p\| \le \kappa_{pg}$ and $\|\nabla^2 J_p\| \le \kappa_{ph}$.*

Assumption 2. *J_p is convex and bounded from below.*

Assumption 3. *The Hessian of J_m is bounded, that is there exists a constant $\kappa_{mh} > 0$ such that $\|\nabla^2 J_m\| \le \kappa_{mh}$.*

Assumption 4. *The Hessian of $J_\phi^{(k)}$ is bounded, that is there exists a constant $\kappa_{\phi h} > 0$ such that $\|\nabla^2 J_\phi^{(k)}\| \le \kappa_{\phi h}$ for all $\mathbf{u} \in \mathcal{B}^{(k)}$.*

In what follows, we use the constant $\kappa_h = \max[\kappa_{pg}, \kappa_{ph}, \kappa_{mh}, \kappa_{\phi h}]$. The analysis of convergence comprises two theorems, which state that

1. The convergence to a KKT point is ensured by adequate regression sets and successful iterations.
2. Adequate regression sets and successful iterations are achievable via unsuccessful iterations.

We use the following definition of well-poised sets in Conn et al. (2009)

Let $\Lambda > 0$ and a set $\mathcal{B} \in \mathbb{R}^n$ be given. Let $\phi = \{\phi_0(x), \phi_1(x), \ldots, \phi_p(x)\}$ be a basis in $\mathcal{P}_n^d$. A poised set $Y = \{y^0, y^1, \ldots, y^p\}$ is said to be Λ-poised in $\mathcal{B}$ if and only if for the basis of Lagrange polynomials associated with Y

$$\Lambda \ge \max_{0\le i\le p} \max_{x\in\mathcal{B}} |\ell_i(x)|. \tag{21}$$

Theorem 1. *Assume that Assumptions 1–4 hold and consider only successful iterations and that the regression set $\mathcal{U}^{(k)}$ is Λ-poised (or adequate) in $\mathcal{B}^{(k)}$, where the definition of Λ-poised is given in (Conn et al., 2009), then one has*

$$\liminf_{k\to\infty} \|\nabla J_p^{(k)}\| = 0 \tag{22}$$

for a noise-free optimization if $\Delta\mathbf{u} \to 0$.

Proof. For contradiction, suppose that there is $\epsilon > 0$ and an index K such that $\|\nabla J_p(\mathbf{u}^{(k)})\| \ge \epsilon$ for all $k \ge K$. Since the regression set $\mathcal{U}^{(k)}$ is adequate in $\mathcal{B}^{(k)}$, from Theorem 4 in (Conn et al., 1997), we get that

$$|J_\phi(\hat{\mathbf{u}}^{(k)}) - J_p(\hat{\mathbf{u}}^{(k)})| \le \kappa_{e\phi} \max[(\Delta^{(k)})^2, (\Delta^{(k)})^3] \tag{23}$$

and

$$\|\nabla J_\phi(\mathbf{u}^{(k)}) - \nabla J_p(\mathbf{u}^{(k)})\| \le \kappa_{eg} \max[\Delta^{(k)}, (\Delta^{(k)})^2] \tag{24}$$

for some constants $\kappa_{e\phi}, \kappa_{eg} > 0$ independent of k. Here $\Delta^{(k)}$ is the radius of a hypersphere that includes the regression set $\mathcal{U}^{(k)}$ and the point $\hat{\mathbf{u}}^{(k)}$. Since $\hat{\mathbf{u}}^{(k)} \in \mathcal{B}^{(k)}$, which is defined by the covariance-based search space evaluation, we have

$$\Delta^{(k)} \approx (3+\gamma)\sqrt{D_{11}}. \tag{25}$$

$J_\cdot(\mathbf{u}^{(k)})$ and $J_\cdot(\hat{\mathbf{u}}^{(k)})$ are represented by $J_\cdot^{(k)}$ and $\hat{J}_\cdot^{(k)}$ in the remaining proof. Since $\hat{\mathbf{u}}^{(k)}$ can be determined either by the adapted model-based optimization or by the quadratic function-based optimization, we need to discuss both cases.

(i) If the adapted model-based mappings are more accurate than the quadratic function approximation, we have

$$\left|\hat{J}_{ad}^{(k)} - \hat{J}_p^{(k)}\right| \le \left|\hat{J}_\phi^{(k)} - \hat{J}_p^{(k)}\right|. \tag{26}$$

In this case $\hat{\mathbf{u}}^{(k)}$ is determined by the adapted model-based optimization and we get that

$$J_{ad}^{(k)} - \hat{J}_{ad}^{(k)} \ge \kappa_C \|\nabla J_\phi^{(k)}\| \min\left[\frac{\|\nabla J_\phi^{(k)}\|}{\kappa_h}, \Delta^{(k)}\right], \tag{27}$$

where κ_C is some constant in $(0, 1)$. (27) implies that the reduction of the adapted objective function should be at least a fraction of the maximum reduction that is achievable along the steepest descent direction while remaining in the search space. A proof of (27) can be found in Moré (1983). From (23) and (26), we obtain that

$$\left|\hat{J}_{ad}^{(k)} - \hat{J}_p^{(k)}\right| \le \kappa_{e\phi} \max[(\Delta^{(k)})^2, (\Delta^{(k)})^3], \tag{28}$$

and consequently

$$\hat{J}_{ad}^{(k)} \ge \hat{J}_p^{(k)} - \kappa_{e\phi} \max[(\Delta^{(k)})^2, (\Delta^{(k)})^3]. \tag{29}$$

Recall (27) and noting that $J_p^{(k)} = J_{ad}^{(k)}$, we obtain that

$$J_p^{(k)} - \hat{J}_p^{(k)} \ge \kappa_C \|\nabla J_\phi^{(k)}\| \min\left[\frac{\|\nabla J_\phi^{(k)}\|}{\kappa_h}, \Delta^{(k)}\right] - \kappa_{e\phi} \max[(\Delta^{(k)})^2, (\Delta^{(k)})^3]. \tag{30}$$

(ii) If $\hat{\mathbf{u}}^{(k)}$ is determined by the quadratic function-based optimization, we have

$$J_\phi^{(k)} - \hat{J}_\phi^{(k)} \ge \kappa_C \|\nabla J_\phi^{(k)}\| \min\left[\frac{\|\nabla J_\phi^{(k)}\|}{\kappa_h}, \Delta^{(k)}\right]. \tag{31}$$

From (23), we get that

$$\hat{J}_\phi^{(k)} \ge \hat{J}_p^{(k)} - \kappa_{e\phi} \max[(\Delta^{(k)})^2, (\Delta^{(k)})^3]. \tag{32}$$

From (31) and $J_p^{(k)} = J_\phi^{(k)}$, we can also obtain (30).

(30) implies that the reduction of the plant objective value is at least the difference between the optimization reduction and the regression error. From (24), we get

$$\kappa_{eg} \max[\Delta^{(k)}, (\Delta^{(k)})^2] \ge \|\nabla J_\phi^{(k)} - \nabla J_p^{(k)}\| \ge \left|\|\nabla J_\phi^{(k)}\| - \|\nabla J_p^{(k)}\|\right|, \tag{33}$$

and consequently

$$\|\nabla J_\phi^{(k)}\| \ge \|\nabla J_p^{(k)}\| - \kappa_{eg} \max[\Delta^{(k)}, (\Delta^{(k)})^2] \ge \epsilon - \kappa_{eg} \max[\Delta^{(k)}, (\Delta^{(k)})^2], \tag{34}$$

where the second "≥" is due to the contradiction assumption $\|\nabla J_p^{(k)}\| \ge \epsilon$. Recall (30), we obtain

$$J_p^{(k)} - \hat{J}_p^{(k)} \geq \kappa_C \left(\epsilon - \kappa_{eg}\max[\Delta^{(k)}, (\Delta^{(k)})^2]\right) \times \min\left[\frac{\left(\epsilon - \kappa_{eg}\max[\Delta^{(k)}, (\Delta^{(k)})^2]\right)}{\kappa_h}, \Delta^{(k)}\right] - \kappa_{e\phi}\max[(\Delta^{(k)})^2, (\Delta^{(k)})^3]. \quad (35)$$

Now we derive that $\Delta^{(k)} \to 0$ as $k \to \infty$. We first show that a sequence of set-points of successful iterations must converge. Since J_p is assumed to be convex and be bounded from below, we obtain that

$$\lim_{k\to\infty} J_p^{(k)} - \hat{J}_p^{(k)} = 0. \quad (36)$$

If the sequence of set-points does not converge, we have that

$$\|\mathbf{u}^{(k)} - \hat{\mathbf{u}}^{(k)}\| \geq \delta, \quad (37)$$

where $\delta > 0$. If $J_p(\mathbf{u})$ is monotonous between $\mathbf{u}^{(k)}$ and $\hat{\mathbf{u}}^{(k)}$, the convexity of $J_p(\mathbf{u})$ implies

$$\|\nabla \hat{J}_p^{(k)}\| < \left|\frac{J_p^{(k)} - \hat{J}_p^{(k)}}{\mathbf{u}^{(k)} - \hat{\mathbf{u}}^{(k)}}\right|. \quad (38)$$

From (36) and (37), we have that $\|\nabla \hat{J}_p^{(k)}\| \to 0$ as $k \to 0$. It contradicts the assumption that $\|\nabla J_p^{(k)}\| \geq \epsilon$ for $k > K$. Thus, $J_p(\mathbf{u})$ should not be monotonous between $\mathbf{u}^{(k)}$ and $\hat{\mathbf{u}}^{(k)}$. Since the regression set $\mathcal{U}^{(k)}$ is adequate, the quadratic function $J_\phi(\mathbf{u})$ is also not monotonous. A new probe in between will be deployed and therefore (37) cannot hold. We conclude that the sequence of set-points must converge.

Next, for contradiction suppose that there is $\epsilon_1 > 0$ and an index K_1 such that $\Delta^{(k)} \geq \epsilon_1$ for all $k \geq K_1$. Recall (17) and (25), we need that there exists at least one point $\mathbf{u}^{(j)} \in \mathcal{U}^{(k)}$ for $k \geq K_1$ such that $\|\mathbf{u}^{(j)} - \mathbf{u}^{(k)}\| \geq \delta(\epsilon_1)$, where $\delta(\epsilon_1) > 0$. When the sequence of set-points converges to $\mathbf{u}^{(\infty)}$, there exists an index $K_2 \geq K_1$ such that $\|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^{(k)}\| < \Delta\mathbf{u}$ for $k \geq K_2$, where $\Delta\mathbf{u} \ll \delta(\epsilon_1)/2$. Since $\|\mathbf{u}^{(j)} - \mathbf{u}^{(k)}\| \geq \delta(\epsilon_1) > 2\Delta\mathbf{u}$, the plant will be evaluated at the middle point between $\mathbf{u}^{(k)}$ and $\mathbf{u}^{(j)}$ (see Step 4). We therefore conclude that if the sequence of set-points converges, the regression set must concentrate at $\mathbf{u}^{(\infty)}$. As a result of this, the assertion of $\Delta^{(k)} \geq \epsilon_1$ for all $k \geq K_1$ does not hold.

As $\Delta^{(k)} \to 0$, recall (35) and (36) we get

$$\epsilon \leq \left(\kappa_{eg} + \frac{\kappa_{e\phi}}{\kappa_C}\right)\Delta^{(\infty)}. \quad (39)$$

Hence, our original assertion of the existence of $\epsilon > 0$ such that $\|\nabla J_p^{(k)}\| \geq \epsilon$ for all $k \geq K$ must be false. We therefore obtain

$$\liminf_{k\to\infty} \|\nabla J_p^{(k)}\| = 0. \quad (40)$$

The proof of Theorem 1 is completed. □

Theorem 2. *Suppose Assumptions 1, 3 and 4 hold, then a subsequence of unsuccessful iterations must eventually result in a well-poised geometry, i.e. $\mathcal{U}^{(k)}$ is adequate in $\mathcal{B}^{(k)}$.*

Proof. We first show that due to a combining effect of the covariance-based search space evaluation and the screening algorithm, an unsuccessful probe $\hat{\mathbf{u}}^{(k)}$ is always included in the updated regression set. This is illustrated in Fig. 4. Since the objective value at $\mathbf{u}^{(k)}$ is less than at all the other regression points, the optimum from Step 3 locates either in the region between $\mathbf{u}^{(k)}$ and the other regression points (represented by A) or in the region which is uncovered by the regression set (represented by B). When determining the new regression set centered at $\mathbf{u}^{(k)}$, there are three possibilities for $\hat{\mathbf{u}}^{(k)}$ to be included in the regression set:

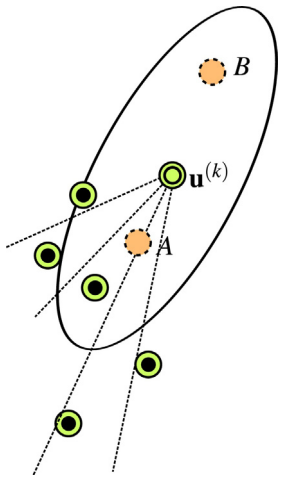


Fig. 4. Illustration of possible positions of unsuccessful probes. ◉ current point $\mathbf{u}^{(k)}$, ◉ regression points, ◌ unsuccessful probes $\hat{\mathbf{u}}^{(k)}$.

(i) As a local point, i.e. $\|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^{(k)}\| < \Delta\mathbf{u}$.
(ii) As the nearest point in an existing segment, i.e. replacing the previous point in that segment.
(iii) As an outer point in a new segment.

It is easy to see that case (iii) improves the poisedness of the regression set. We therefore need only to prove that cases (i) and (ii) yield an adequate regression set. For simplicity, we only consider the case that the adapted model-based optimization is run.

For an unsuccessful iteration $\hat{J}_p^{(k)} > J_p^{(k)}$, recall (27) and note that $J_p^{(k)} = J_{ad}^{(k)}$, we obtain

$$\hat{J}_p^{(k)} - \hat{J}_{ad}^{(k)} \geq \kappa_C \|\nabla J_\phi^{(k)}\| \min\left[\frac{\|\nabla J_\phi^{(k)}\|}{\kappa_h}, \Delta^{(k)}\right]. \quad (41)$$

From the Lemma 2.5 in Scheinberg and Toint (2010)

> "Given $\mathcal{B}(x, \Delta) \stackrel{\text{def}}{=} \{ y \in \mathbb{R}^n \mid \|y - x\| \leq \Delta \}$, a poised interpolation set $\mathcal{Y} \in \mathcal{B}(x, \Delta)$, and its associated basis of Lagrange polynomials $\{\ell_j(y)\}_{j=0}^p$, there exists a constant $\kappa_{ef} > 0$ such that, for any interpolating polynomial $m(y) = \sum_{j=0}^p f(y_j)\ell_j(y)$ and any given point $y \in \mathcal{B}(x, \Delta)$, $|f(y) - m(y)| \leq \kappa_{ef} \sum_{j=0}^p \|y_j - y\|^2 |\ell_j(y)|$",

we get that

$$|\hat{J}_p^{(k)} - \hat{J}_\phi^{(k)}| \leq \kappa_{ef} \sum_{i=1}^{n_r} \|\mathbf{u}^{(r_i)} - \hat{\mathbf{u}}^{(k)}\|^2 |\ell_{r_i}(\hat{\mathbf{u}}^{(k)})|, \quad (42)$$

where n_r is the number of points in the regression set and r_i is the iteration index of the ith regression point. From (26), we get that

$$\hat{J}_p^{(k)} - \hat{J}_{ad}^{(k)} \leq \kappa_{ef} \sum_{i=1}^{n_r} \|\mathbf{u}^{(r_i)} - \hat{\mathbf{u}}^{(k)}\|^2 |\ell_{r_i}(\hat{\mathbf{u}}^{(k)})|. \quad (43)$$

From (41) and (43), we obtain

$$\kappa_{ef} \sum_{i=1}^{n_r} \|\mathbf{u}^{(r_i)} - \hat{\mathbf{u}}^{(k)}\|^2 |\ell_{r_i}(\hat{\mathbf{u}}^{(k)})| \geq \kappa_C \|\nabla J_\phi^{(k)}\| \min\left[\frac{\|\nabla J_\phi^{(k)}\|}{\kappa_h}, \Delta^{(k)}\right]. \quad (44)$$

Assume $\|\mathbf{u}^{(r_i)} - \mathbf{u}^{(k)}\| \leq \beta\Delta^{(k)}$, where $\beta > 1$, we get that

$$\|\mathbf{u}^{(r_i)} - \hat{\mathbf{u}}^{(k)}\| \leq \|\mathbf{u}^{(r_i)} - \mathbf{u}^{(k)}\| + \|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^{(k)}\| \leq (\beta + 1)\Delta^{(k)}. \quad (45)$$

From (44), we obtain that

$$\sum_{i=1}^{n_r}|\ell_{r_i}(\hat{\mathbf{u}}^{(k)})| \geq \frac{\kappa_C\|\nabla J_\phi^{(k)}\|\min\left[\frac{\|\nabla J_\phi^{(k)}\|}{\kappa_h}, \Delta^{(k)}\right]}{\kappa_{ef}(\beta+1)^2(\Delta^{(k)})^2}. \tag{46}$$

Recall the three cases for $\hat{\mathbf{u}}^{(k)}$ to be chosen into the updated regression set. Case (i) and (ii) lead to a concentration of the regression points to $\mathbf{u}^{(k)}$, and consequently $\Delta^{(k)}$ is reduced by the covariance-based search space evaluation. When

$$\Delta^{(k)} \leq \min\left[\frac{\|\nabla J_\phi^{(k)}\|}{\kappa_h}, \frac{\kappa_C\|\nabla J_\phi^{(k)}\|}{\kappa_{ef}(\beta+1)^2 n_r}\right], \tag{47}$$

we get that

$$\sum_{i=1}^{n_r}|\ell_{r_i}(\hat{\mathbf{u}}^{(k)})| \geq n_r. \tag{48}$$

It is easy to see that

$$\max_{i=1,\ldots,n_r}|\ell_{r_i}(\hat{\mathbf{u}}^{(k)})| > 1. \tag{49}$$

Using Lemma 2.4 in Scheinberg and Toint (2010)

> "Given a closed bounded domain $\mathcal{B}$, any initial interpolation set $\mathcal{Y} \in \mathcal{B}$ and a constant $\Lambda > 1$, consider the following procedure: find $j \in \{0, \ldots, p\}$ and a point $x \in \mathcal{B}$ such that $|\ell_j(x)| > \Lambda$ (if such a point exists), and replace y_j by x to obtain a new set $\mathcal{Y}$. Then this procedure terminates after a finite number of iterations with a model which is Λ-poised (or adequate) in $\mathcal{B}$",

we have that a subsequence of unsuccessful iterations eventually results in a well-poised geometry. It also leads to a successful iteration because $\Delta^{(k)}$ cannot become too small as long as a KKT point is not approached.

The proof of Theorem 2 is completed. □

4. Simulation studies

4.1. Example 1: Application to the Williams-Otto Reactor

The Williams-Otto Reactor has been used as a benchmark problem to evaluate RTO schemes in Roberts (1979), Forbes and Marlin (1994), Marchetti et al. (2010), Navia (2012). The real plant is described by three irreversible reactions:

$$A + B \xrightarrow{k_1} C \tag{50a}$$

$$C + B \xrightarrow{k_2} P + E \tag{50b}$$

$$P + C \xrightarrow{k_3} G, \tag{50c}$$

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

$$A + 2B \xrightarrow{\tilde{k}_1} P + E \tag{51a}$$

$$A + B + P \xrightarrow{\tilde{k}_2} G + E. \tag{51b}$$

The ignorance of the reactions which involve C leads to a structural plant-model mismatch. The optimization objective is the steady-state profit

$$J = (F_A + F_B)\left(P_p x_P + P_E x_E\right) - C_A F_A - C_B F_B. \tag{52}$$

The variable definitions and parameter values are taken from (Marchetti et al., 2010). The optimization variables are the flow rate F_B and the reaction temperature ϑ. They are normalized by their operating intervals (F_B: 3–6 kg/s, ϑ: 70–100 °C).

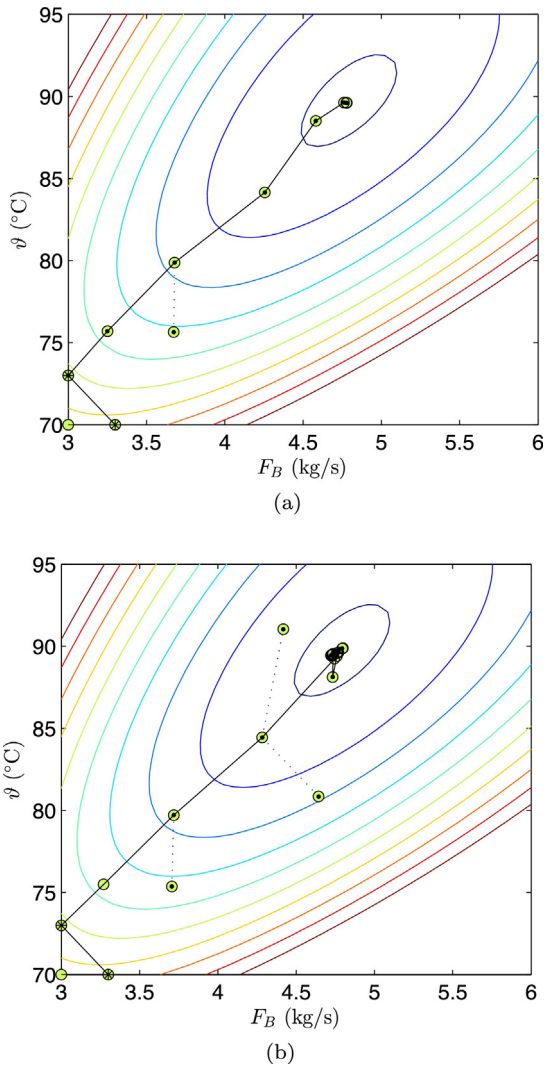


Fig. 5. Illustration of the set-point moves of the new scheme with noise-free (a) and noisy data (b). ⊛: initial perturbations, ◉: regression points.

Table 1
Parameters and variables for the new scheme.

Description	Symbol	Value
Screening parameter	$\Delta\mathbf{u}$	0.1
Search space parameter	γ	3
Perturbation step size	Δh	0.1
Noise standard deviation	σ	0.5

Fig. 5a and b shows the results of the new scheme with noise-free and noisy data. The parameters and variables used in the simulation are listed in Table 1. The noise is assumed to be normally distributed and is superimposed directly on the profit. Starting from set-point (3, 70), two perturbations with step size h are applied to generate data for an initial gradient estimation. At the 4th iteration, an additional perturbation is added by the IGMO approach. For the noise-free case, the optimum is reached after 9 plant evaluations. In the presence of noise, two data-collecting moves are taken after the 5th iteration.

We then study the performance of the new scheme for different values of noise, screening parameter and perturbation step size. All cases are run for 100 realizations of the noise. To quantify the performance, the average numbers of moves $\overline{N}_{outer}$ after which set-points stays within a circle (radius: 0.1) centered at the optimum are evaluated. The average numbers of unsuccessful moves $\overline{N}_{unsu}$,

Table 2
Performance under different levels of noise.

σ	0.25	0.5	0.75	1.0
$\overline{N}_{unsu}$	0.27	0.57	1.10	1.57
$\overline{N}_{outer}$	8.90	9.40	10.65	11.44

Table 3
Influence of the screening parameter $\Delta\mathbf{u}$.

$\Delta\mathbf{u}$	0.05	0.075	0.1	0.125	0.15
$\overline{N}_{unsu}$	1.13	0.72	0.57	0.56	0.51
$\overline{N}_{outer}$	11.18	9.60	9.40	9.20	9.25

Table 4
Influence of the perturbation step size Δh.

Δh	0.05	0.1	0.15
$\overline{N}_{unsu}$	1.08	0.57	1.06
$\overline{N}_{outer}$	12.18	9.40	9.26

which can be taken as a measure of the data-collecting moves, are also indicated.

Table 2 lists the performance under different levels of noise. Both $\overline{N}_{unsu}$ and $\overline{N}_{outer}$ increase with the increasing of the noise. This implies that the new scheme can automatically adapt to the quality of the data.

Table 3 lists the performance for different values of the screening parameter $\Delta\mathbf{u}$. A small value of $\Delta\mathbf{u}$ leads to more nearby points are chosen in the regression set. The quadratic approximation is more sensitive to the noise. Therefore, more data-collecting moves are taken than a large value of $\Delta\mathbf{u}$ is used. One disadvantage of increasing the value of $\Delta\mathbf{u}$ is the decrease of the accuracy of the final optimum.

Table 4 lists the performance for different step sizes of the initial perturbations. A small value of the step-size leads to a low quality of gradient estimation and later more data-collecting moves are required. However, a too large value of the step-size might also lead to more data-collecting moves, since the plant mapping is not quadratic over a large range.

4.2. Example 2: Application to a synthetic example with 10 variables

This example was used by Caballero and Grossmann (2008) to illustrate the performance of their algorithm for a function with 10 variables with relatively high noise

$$\begin{aligned}
\min_{\mathbf{x}} \quad & f(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}) \\
\text{s.t.} \quad & -105 + 4x_1 + 5x_2 - 3x_7 + 9x_8 \le 0 \\
& 10x_1 - 8x_2 - 17x_7 + 2x_8 \le 0 \\
& -8x_1 + 2x_2 + 5x_9 - 2x_{10} - 12 \le 0 \\
& z(x_1, x_2, x_3, x_4) \le 0 \\
& 5x_1^2 + 8x_2 + (x_3 - 6)^2 - 2x_4 - 40 \le 0 \\
& x_1^2 + 2(x_2 - 2)^2 - 2x_1x_2 + 14x_5 - 6x_6 \le 0 \\
& 0.5(x_1 - 8)^2 + 2(x_2 - 4)^2 + 3x_5^2 - x_6 - 30 \le 0 \\
& -3x_1 + 6x_2 + 12(x_9 - 8)^2 - 7x_{10} \le 0 \\
& 0 \le x_i \le 10, \quad i = 1, \ldots, 10.
\end{aligned} \tag{53}$$

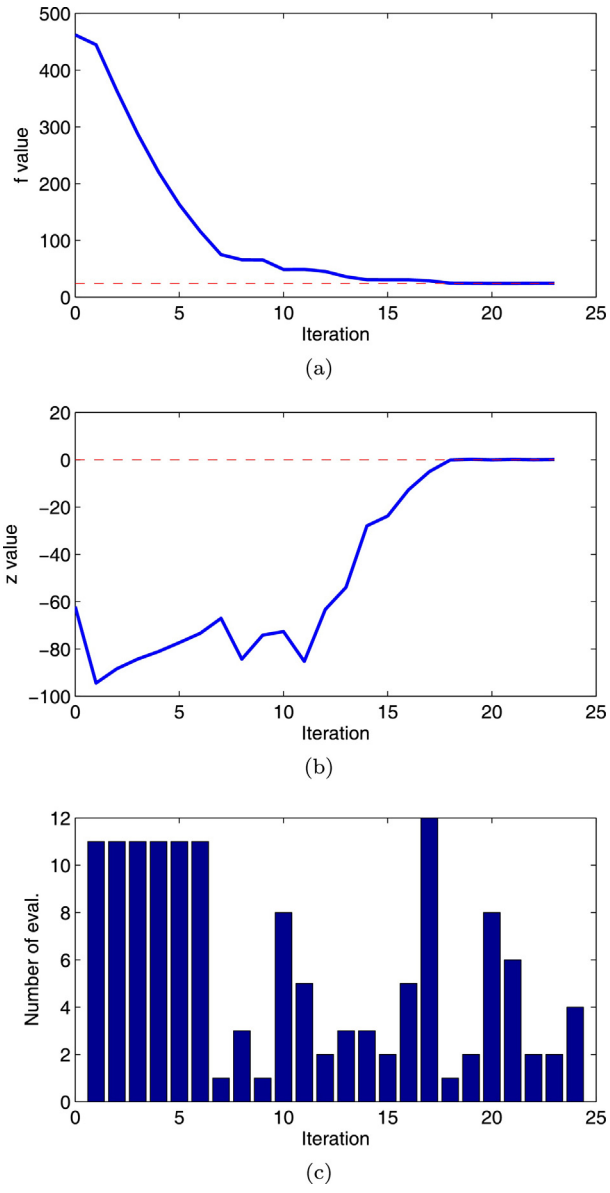


Fig. 6. Optimization results of the synthetic example using the new scheme, (a) and (b): evolutions of f and z (the dashed line marks the optimal solution), (c): number of point evaluations per iteration.

The cost function f and the constraint function z are represented by noisy black box functions

$$\begin{aligned}
f = {} & x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^4 \\
& + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 \\
& + (x_{10} - 7)^2 + 45 + \mathcal{N}(0, 0.01)
\end{aligned} \tag{54}$$

$$z = 3(x_1 - 2)^2 + 4(x_2 - 3)^2 + 2x_3^2 - 7x_4 - 120 + \mathcal{N}(0, 0.01),$$

where $\mathcal{N}(\mu, \sigma^2)$ is a normally distributed random variable with mean μ and standard deviation σ. The optimal solution if the noise is removed is $\mathbf{x} = (2.18, 2.37, 8.83, 5.36, 0.99, 1.43, 1.32, 9.82, 8.29, 8.37)$. In order to handle the noisy implicit black box functions, Caballero and Grossmann (2008) developed an algorithm based on fitting response surfaces. The algorithm needs more than 800 sampled points in f to reach a point which is close to the solution without noise.

In the test of the new scheme, the implicit functions $f(\mathbf{x})$ and $z(\mathbf{x})$ correspond to $J_p(\mathbf{u})$ and $C_p(\mathbf{u})$ respectively, where $\mathbf{u} = \mathbf{x}$. For simplicity, we set $J_m(\mathbf{u}) = 0$ and $C_m(\mathbf{u}) = 0$. So the optimizer does not

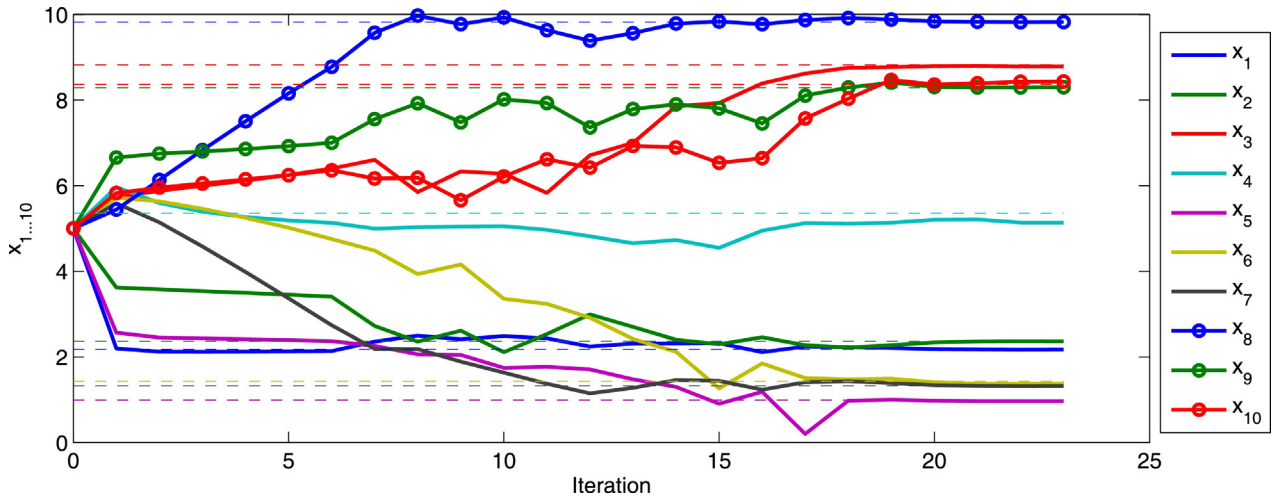


Fig. 7. Evolutions of the decision variables (the dashed lines mark the optimal solutions).

have explicit information about f and z and the model gradient is unknown. The screening parameter $\Delta\mathbf{u}$ and the initial step-size h are preliminarily set to 1, which is 10% of the operating region. The scaling parameter γ is set to 3. The initial set-point is chosen at the center of the operating region, i.e. $\mathbf{u}^{(0)} = (5, 5, 5, 5, 5, 5, 5, 5, 5, 5)$.

Figs. 6 and 7 illustrate the optimization results over the iterations. The new scheme converges to the optimal solution within 24 iterations. 136 points were evaluated in total, compared with 800 sampled points used in Caballero and Grossmann (2008). In the first 6 iterations, the IGMO scheme, which is based on the finite-difference approximation of the gradients, is run. For simplicity, 10 mutually orthogonal perturbations with step size h are applied around the current point in each iteration. Note that the approach proposed in (Gao and Engell, 2005) can be used to decrease the number of additional perturbations. After the 6th iteration, enough sampled points (here $(n_u+1)(n_u+2)/2 = 66$) are available for the quadratic approximation. Step 2 to Step 5 of the scheme are run repeatedly. Note that the iterations in Figs. 6 and 7, after the 6th iteration, correspond to the "successful" iterations, which are defined by $\hat{J}_p^{(k)} + 3\sigma < J_p^{(k)} - 3\sigma$ in the presence of noise. All the "unsuccessful" iterations are counted and are illustrated by the bar plot in Fig. 6c.

5. Conclusions

This paper proposed a new scheme for using the collected plant data together with an inaccurate model to converge to the optimum of the real plant in real-time optimization. The new scheme combines the quadratic approximation of derivative-free optimization with the iterative gradient-modification approach and integrates recent advances in both areas. Compared to the finite-difference calculation of the plant gradients, the quadratic approximation method can decrease the influence of the noise by capturing the curvature information from more distant points to provide a smooth mapping. The screening algorithm which selects well-distributed and sufficiently distant points for the current iterate ensures a better quadratic approximation of the plant mapping than the regression result based on all the collected data. Note that the parameter $\Delta\mathbf{u}$ of the screening algorithm is assumed to influence both the accuracy of the optimum and the number of plant evaluations to enter the vicinity of the optimum. The optimal tuning of $\Delta\mathbf{u}$ according to the level of the noise and the curvature of the plant can be an interesting research direction.

The use of the covariance-based restriction of the search space evaluation is twofold in the new scheme. On the one hand side, large set-point moves along a direction, in which more data has been collected previously, are allowed and this will lead to a fast rate of convergence. On the other hand side, moves along a direction, in which the plant needs to be probed, are bounded to avoid an unnecessary deterioration of the plant performance. Note that the size parameter γ can be adjusted according to the prediction accuracies of the adapted objective and constraint functions.

The switching between model-based and data-based optimizations in the new scheme is also twofold. On the one hand, the global convergence is ensured by running data-based optimizations in case of inadequate models which are defined for the modifier adaptation approach in Marchetti et al. (2009). Therefore, the use of convex model approximations, which are proposed by François et al. (2013), is not necessary. On the other hand, the accuracy of model-predicted second-order derivatives of the cost and constraint functions with respect to the decision variables can be used to achieve a faster rate of convergence than model-free schemes. A detailed description of the accuracy-induced optimization switching was presented in (Gao et al., 2015).

The new scheme relies on the collected data to construct quadratic approximations. One issue is that the previously collected data might be invalid in the presence of varying disturbances. Wenzel et al. (2015) studied the effect of considering only recently collected data in the regression set for quadratic approximations. By adjusting an age parameter in the screening algorithm, a trade-off between noise attenuation and rejecting time-varying disturbances can be made.

One of the difficulties in the application of this scheme in practice is that for each new set-point, one has to wait for the plant to reach a steady state after the set-point change, which makes the convergence slow. Future studies will be focused on the use of transient measurements to accelerate the steady-state optimization.

Acknowledgments

The research leading to these results has received funding from the European Commission under grant agreement number 291458 (ERC Advanced Investigator Grant MOBOCON).

References

Ariyur, K.B., Krstic, M., 2003. Real-Time Optimization by Extremum-Seeking Control, 1st ed. Wiley-Interscience.

Brdyś, M.A., Tatjewski, P., 1994. An algorithm for steady-state optimising dual control of uncertain plants. In: Proceedings of the First IFAC Workshop on New Trends in Design of Control Systems, pp. 249–254.

Caballero, J.A., Grossmann, I.E., 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. AIChE J. 54, 2633–2650.
Chen, C.Y., Joseph, B., 1987. On-line optimization using a two-phase approach: an application study. Industrial. Eng. Chem. Res. 26 (9), 1924–1930.
Conn, A.R., Scheinberg, K., Toint, P.L., 1997. On the Convergence of Derivative-Free Methods for Unconstrained Optimization. Cambridge University Press, Cambridge, England, pp. 83–108.
Conn, A.R., Scheinberg, K., Vicente, L.N., 2009. Introduction to Derivative-Free Optimization. SIAM, Philadelphia.
Forbes, J.F., Marlin, T.E., 1994. Model accuracy for economic optimizing controllers: the bias update case. Ind. Eng. Chem. Res. 33 (8), 1919–1929.
François, G., Srinivasan, B., Bonvin, D., 2005. Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. J. Process Control 15 (6), 701–712.
François, G., Srinivasan, B., Bonvin, D., 2013. Use of convex model approximations for real-time optimization via modifier adaptation. Ind. Eng. Chem. Res. 52, 11614–11625.
Gao, W., Engell, S., 2005. Iterative set-point optimization of batch chromatography. Comput. Chem. Eng. 29 (6), 1401–1409.
Gao, W., Wenzel, S., Engell, S., 2015. Modifier adaptation with quadratic approximation in iterative optimizing control. In: Proceedings of the 14th European Control Conference, Linz, Austria, pp. 2532–2537.
Gao, W., Wenzel, S., Engell, S., 2015. Integration of gradient adaptation and quadratic approximation in real-time optimization. In: Proceedings of the 34th Chinese Control Conference, Hangzhou, China, pp. 2780–2785.
Kuhn, H.W., Tucker, A.W.,1951. Nonlinear programming. In: Proceedings of 2nd Berkeley Symposium. University of California Press, Berkeley, pp. 481–492.
Marchetti, A., Chachuat, B., Bonvin, D., 2009. Modifier-adaptation methodology for real-time optimization. Ind. Eng. Chem. Res. 48, 6022–6033.
Marchetti, A., Chachuat, B., Bonvin, D., 2010. A dual modifier-adaptation approach for real-time optimization. J. Process Control 20 (9), 1027–1037.
Moré, J.J., 1983. Recent developments in algorithms and software for trust region methods. In: Bachem, A., Korte, B., Grötschel, M. (Eds.), Mathematical Programming The State of the Art. Springer, Berlin, Heidelberg, pp. 258–287.
Navia, D., 2012. Handling uncertainties in process optimization. Universidad de Valladolid (Ph.D. thesis).
Nocedal, J., Wright, S., 2006. Numerical Optimization. Springer-Verlag, Berlin.
Roberts, P., 2000. Broyden derivative approximation in ISOPE optimising and optimal control algorithms. Control Applications of Optimization, vol. 1., pp. 283–288.
Roberts, P.D., 1979. An algorithm for steady-state system optimization and parameter estimation. Int. J. Syst. Sci. 10 (7), 719–734.
Scheinberg, K., Toint, P., 2010. Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. SIAM J. Optim. 20 (6), 3512–3532.
Wenzel, S., Gao, W., Engell, S., 2015. Handling disturbances in modifier adaptation with quadratic approximation. In: Proceedings of the 16th IFAC Workshop on Control Applications of Optimization, Garmisch-Partenkirch, Germany.