Global Optimization for Integrated Design and Control of Computationally Expensive Process Models

Jose A. Egea, Dirk Vries, Antonio A. Alonso and Julio R. Banga.

Abstract—The problem of integrated design and control optimization of process plants is discussed in this paper. We consider it as a nonlinear programming problem subject to differential-algebraic constraints. This class of problems is frequently (i) non-convex and (ii) “costly” (i.e. computationally expensive to evaluate). Thus, on the one hand, local optimization techniques usually fail to locate the global solution and, on the second hand, most global optimization methods require many simulations of the model, resulting in unaffordable computation times. As an alternative, one may consider global optimization methods which employ surrogate-based approaches to reduce computation times, and which require no knowledge of the problem structure. A challenging Wastewater Treatment Plant benchmark model (see [1] and references therein) is used to evaluate the performance of these techniques. Numerical experiments with different optimization solvers indicate that the proposed benchmark optimization problem is indeed non-convex, and that we can achieve an improvement of the controller performance compared to the best tuned controller settings available in the literature. Moreover, these results show that surrogate-based methods may indeed reduce computation times while, at the same time ensuring convergence to the best known solutions.

I. INTRODUCTION

Many optimization problems which arise during the design and/or operation of chemical and biochemical processes are frequently non-convex (also referred as multimodal). Thus, it is not surprising that global optimization (GO) has received much attention during the last decade from many researches in computer aided process engineering. In fact, many advances have been made regarding both deterministic and stochastic methods during recent years [2]. However, the current state of the art is far from satisfactory, especially when we consider the global optimization of complex process models. These models are typically complex due to their dynamic (non-linear) behaviour and high number of states. Therefore they are typically expensive to evaluate (i.e. each simulation can take minutes, or even hours, in CPU computation time of an ordinary PC) and, for a number of reasons, can only be treated as black-box models. Thus, several authors [3] have recently proposed the use of so-called surrogate models which are (computationally) cheaper to evaluate.

In the area of process systems engineering, Moles et al. [4] showed how stochastic global optimization methods can be applied for simultaneous design and control of process plants of low complexity. In this contribution, our aim is to apply a similar approach to more complex (and costly) models. In order to keep the computational burden acceptable, the capabilities of recent global surrogate model based optimization methods have been evaluated. Such evaluation is performed considering a challenging benchmark case study: the integrated design and control of a wastewater treatment plant (WWTP) for nitrogen removal, as developed by the COST 624 work group [5]. This paper is structured as follows: in the next section, the general statement of the integrated design problem is presented. Next, we briefly review the state of the art regarding global optimization for such problems, and we present the methods selected for our research. In the following section, the WWTP case is outlined. Finally, sections with results and discussion and conclusions are provided.

II. INTEGRATED DESIGN: PROBLEM STATEMENT

The general statement of the simultaneous (integrated) design and control problem takes into account the process and control superstructures indicating the different design alternatives [6]. This general approach results in mixed integer optimal control problems. In this work, we consider a simpler, yet non-trivial, subproblem, where it is assumed that the flowsheet (i.e. plant configuration) is given. It should be noted that this subproblem is challenging enough to serve as a case study for the comparison of surrogate-based global optimization methods, which is the main objective of this work. The aim is to simultaneously find the static variables of the process design, the operating conditions and the controllers’ parameters which optimize a combined measure of the plant economics and its controllability, subject to a set of constraints which ensure appropriate...
dynamic behaviour and process specifications. We state our problem as follows:

Find \( v \) to minimize:

\[
C = \sum w_i \cdot \phi_i \tag{1}
\]

subject to

\[
f(x, x, p, v) = 0 \tag{2}
\]

\[
x(t_0) = x_0 \tag{3}
\]

\[
h(x, p, v) = 0 \tag{4}
\]

\[
g(x, p, v) \leq 0 \tag{5}
\]

\[
v^L \leq v \leq v^U \tag{6}
\]

where \( v \) is the vector of decision variables (e.g. design variables, operating conditions, parameters of controllers, setpoints, etc.); \( C \) is the cost (objective function) to minimize (normally a weighted combination of capital, operation and controllability costs, \( \phi \)); \( f \) is the set of differential and algebraic equality constraints describing the system dynamics (i.e. the nonlinear process model), \( x \) is the vector of the states, \( t_0 \) the initial time for the integration of the ODE’s (and, consequently, \( x_0 \) is the vector of the states at that initial time); \( p \) is the vector of parameters of the dynamic model; \( h \) and \( g \) are possible equality and inequality path and/or point constraints which express additional requirements for the process performance; finally, \( v^L \) and \( v^U \) are the upper and lower bounds for the decision variables.

The formulation above is that of a non-linear programming problem (NLP) with differential-algebraic (DAEs) constraints. Due to the nonlinear and constrained nature of the system dynamics, these problems are very often multimodal. Further, it is known that using standard controllability measures such as the Integral Square Error (ISE) in the objective function often causes non-convexity [6]. Therefore, if this NLP-DAEs is solved via standard NLP methods, such as Sequential Quadratic Programming (SQP), it is very likely that the solution found will be of local nature. To avoid this, global optimization (GO) should be used.

III. OPTIMIZATION METHODS

A. Introduction

In principle, model based optimization can be successfully used to improve the design and/or operation of single units or full process plants. Typically, most of the problems in process engineering applications are highly constrained and exhibit nonlinear dynamics. These properties often result in non-convexity. Furthermore, in many complex process models some kind of noise and/or discontinuities (either due to numerical methods, or to intrinsic properties of the model) is often present. Therefore, there is great need of robust global optimization solvers which can (i) locate the vicinity of the global solution in a reasonable number of iterations, (ii) handle noise and/or discontinuities, and (iii) use some sort of approximation or reduction of the original process model, in order to keep the computational effort acceptable.

In general, (iterative) gradient-based local methods for constrained NLP problems are very efficient, but they can only handle smooth nonlinear functions subject to smooth constraints based on a set of continuous variables. Additionally, only convergence to local solutions is guaranteed. Therefore, one must use the so-called global optimization (GO) methods.

GO methods can be roughly classified as being deterministic [7] and stochastic [8]. Deterministic GO methods assure convergence to the global optimum for certain problems, although no algorithm can solve general GO problems with certainty in a finite time [8]. For these methods, the computational effort usually increases exponentially with the problem size. Further, they have requirements (e.g. smoothness, differentiability) which are rarely met in realistic applications.

Stochastic GO methods can find solutions in the vicinity of the global solution in relatively short computation times compared to deterministic GO. Note that with GO, the convergence to global optimality (in finite time) is not guaranteed either. Another advantage of these methods is that they are easy to implement, and they can treat the objective function as a black box (i.e. a simple connection between inputs and outputs, with no derivative information needed). This feature is specially appealing in the case of complex dynamic systems where the objective function is the result of e.g. a simulation provided by a third party software with restricted access for the user.

B. Surrogate-based Global Optimization

In general, all these GO approaches require a significant number of evaluations of the objective function and the constraints. In case of realistic problems, these models are costly to evaluate, posing a major challenge to the application of GO methods. In recent years, a number of approaches have been proposed to obtain surrogate models which are cheaper to evaluate than the original ones and which imitate the original model based on a reduced number of sampled points (simulations). Hence, these so called surrogate model based solvers try to approximate the original model over a region by a model that is cheaper to evaluate. Provided the surrogate model is accurate enough, the computation times can still be dramatically reduced. In addition, surrogate models go
C. Selected Optimization Methods

Regarding surrogate-based GO, we have considered two recent solvers which are Matlab® implementations of the two types of strategies mentioned above:

**rbfSolve:** this solver, included in the Tomlab® toolbox [9], solves costly global optimization problems using a Radial Basis Function (RBF) interpolation algorithm. It fits a response surface (based on splines) to data collected by evaluating the objective function at some points and then applies an optimization algorithm over that surrogate model. The initial points to create the response surface may be given by the user or selected by the algorithm based on different strategies.

**ego:** this solver, also included in the Tomlab® toolbox [9], solves costly global optimization problems using the Efficient Global Optimization (EGO) algorithm. The idea of the EGO algorithm is to first fit a response surface to data collected by evaluating the objective function at a few points. Then, EGO balances between finding the minimum of the surface and improving the approximation by sampling where the prediction error may be high.

In order to critically evaluate the performance of these two surrogate-based strategies, we have also considered selected local and global solvers which only rely on evaluations of the original (costly) model:

**fminsearch:** a local method implemented in the Matlab® Optimization Toolbox [10] that uses the simplex method instead of using gradient information. Although generally less efficient than gradient based methods, the simplex method may be more robust if the problem is highly discontinuous or presents noise.

**fmincon:** also part of the Matlab® Optimization Toolbox [10], this solver finds a local minimum of a constrained multivariable function by means of a SQP (Sequential Quadratic Programming) algorithm. The method uses numerical or, if available, analytical gradients.

**NOMADm:** Nonlinear Optimization for Mixed variables And Derivatives-Matlab, abbreviated as NOMADm, is a Matlab® code that runs various Generalized Pattern Search (GPS) algorithms to solve nonlinear and mixed variable optimization problems [11].

**Differential Evolution (DE):** This method, presented in [12], is a heuristic population based stochastic approach to global optimization for minimizing possibly nonlinear and non-differentiable continuous space functions.

IV. CASE STUDY: OPTIMIZATION OF CONTROLLERS SETTINGS IN A WWT PLANT

A number of control strategies have been proposed to meet the strict standards that Wastewater Treatment Plants (WWTP) must comply with, while also trying to reduce costs [13]. Relevant examples from the recent literature of attempts to optimize the controllers of these plants are:

1. *ad hoc* extensive simulation studies [14], (strictly speaking these may not be called optimizations, because there is no evidence that a locally or globally best solution is found).
2. dynamic optimizations of control or design strategies using local gradient-based optimization methods [15], often based on simplified or linearized models.
3. global optimization methods for simultaneously optimizing operation and design [4].
4. an integrated approach for the optimization of control strategies, where a small selection of global and local optimization methods was used [16].

Evaluation of these and similar strategies, either in practice or by simulation, is a real problem due to the lack of a standard with respect to evaluation criteria, process complexity and large variations in plant configuration. In order to enhance the development and acceptance of new control strategies, the International Water Association...
(IWA) Task Group on Respirometry, together with the European COST work group, proposed a standard simulation benchmarking methodology for evaluating the performance of activated sludge plants. The COST 624 work group defines the benchmark as “a protocol to obtain a measure of performance of control strategies for activated sludge plants based on numerical, realistic simulations of the controlled plant” [5]. According to this definition, the benchmark consists of a description of the plant layout, a simulation model and definitions of (controller) performance criteria.

The layout of this benchmark plant combines nitrification with predenitrification by a five compartment reactor with an anoxic zone (see Figure 1). A secondary settler separates the microbial culture from the liquid being treated. A basic control strategy consisting of 2 PI controllers is proposed to test the benchmark. Its aim is to control the dissolved oxygen level in the final compartment of the reactor (AS Unit 5) by manipulation of the oxygen transfer, and to control the nitrate level in the last anoxic compartment (AS Unit 2) by manipulating the internal recycle flow rate. A detailed description of the benchmark can be found in [5]. In this work, a Simulink® implementation of the benchmark model by Jeppsson was used for the simulations [17]. Each function evaluation consists of an initialization period of 100 days to achieve steady state, followed by a period of 14 days of dry weather and a third period of 14 days of rainy weather. Calculations of the controller performance criterion are based on data from the last 7 rain days.

As far as we know, global optimization methods have not been used for the integrated design and control of WWTPs models of this complexity. Further, since each simulation of this benchmark model takes a significant time on a standard PC (about 3 minutes in a PC-PIII 800 MHz), it is an illustrating example to evaluate the surrogate-based strategies mentioned previously.

The control performance is tested by using the ISE (Integral Square Error) as a controller performance criterion. Both the nitrate level and oxygen level controllers (further referred as N- and O-controller respectively) are optimized with respect to their controller parameters, that is, the gain $K$, integral time constant $\tau_i$ and anti-windup time constant $\tau_w$. The problem is formulated as follows:

$$\min J(v, t_f) = W \cdot ISE$$  \hspace{1cm} (7)

st.

$$\dot{x} = f(x, v, p, d)$$  \hspace{1cm} (8)

$$x(t_0) = x_0$$  \hspace{1cm} (9)

$$h(x, v, p) = 0$$  \hspace{1cm} (10)

$$g(x, v, p) \leq 0$$  \hspace{1cm} (11)

$$v^\ell \leq v \leq v^u$$  \hspace{1cm} (12)

where $W \in \mathbb{R}^{1 \times 2}$ contains the weighting coefficients; $ISE \in \mathbb{R}^{2 \times 1}$ contains the integral squared errors of the two PI controllers; $f \in \mathbb{R}^{150 \times 1}$ denotes the system dynamics; $h$ and $g$ are possible equality and inequality trajectory and/or endpoint constraints which express additional requirements for the process performance. The weighting vector $W$, the integral square error $ISE$ and the decision parameter vector are as follows:

$$W = \begin{bmatrix} w_1 & w_2 \end{bmatrix} = \begin{bmatrix} 1 & 0000 \\ 1000 & 1000 \end{bmatrix}$$  \hspace{1cm} (13)

$$ISE = \begin{bmatrix} ISE_N \\ ISE_O \end{bmatrix}$$  \hspace{1cm} (14)

$$ISE_{(i)} = \int_0^T \epsilon(t)^2 \, dt$$  \hspace{1cm} (15)

$$v = \begin{bmatrix} v_N \\ v_O \end{bmatrix} = \begin{bmatrix} K_{(i)} \\ \tau_{i(o)} \\ \tau_{i(o)} \\ \tau_{i(O)} \\ \tau_{i(O)} \end{bmatrix}$$  \hspace{1cm} (16)

The weighting vector is chosen such that the $ISE_{(i)}$ equals to the $ISE_{(N)}$ part when using the benchmark default settings (i.e. the tuned PI-parameters see(13)) provided by the COST project [5].

The system dynamics are described by algebraic mass balance equations, ordinary differential equations for the biological processes in the bioreactors as defined by the ASM1-model [18], and the double-exponential settling velocity function presented in [1] as a fair presentation of the settling process, with $x \in \mathbb{R}^{6 \times 1}$ the state vectors, $p \in \mathbb{R}^{150 \times 1}$ the system parameters and $d$ the influent disturbance. Due to the complexity of the system dynamics, the problem cannot be solved analytically, and the optimal values for the decision variables $v \in \mathbb{R}^{6 \times 1}$ (i.e. the PI controller parameters) have to be retrieved by
optimization techniques or open loop controller tuning. However, the problem remains that most tuning techniques are intended for linear systems so that model approximation by linearization around different operating points would be inevitable.

In the particular case considered here, no further dynamic restrictions are laid on the problem, i.e. \( h \) and \( g \) belong to empty sets. Boundaries on the decision variables are chosen such that the process behavior would not show (exceptional) unstable behavior:

\[
v^L = \left[ 100 \ 7.0 \times 10^{-1} \ 1.0 \times 10^{-4} \ 100 \ 1.0 \times 10^{-2} \ 1.0 \times 10^{-4} \right] (17)
\]

\[
v^U = \left[ 1000 \ 7.0 \times 10^{-1} \ 7.0 \times 10^{-4} \ 50000 \ 1.0 \ 7.0 \times 10^{-2} \right] (18)
\]

The objective function values are normalized with respect to the performance obtained with the tuned controller settings provided by the COST project [5], which are:

\[
v^{ref} = \left[ 500 \ 1.0 \times 10^{-3} \ 2.0 \times 10^{-4} \ 15000 \ 5.0 \times 10^{-1} \ 3.0 \times 10^{-3} \right] (19)
\]

V. RESULTS AND DISCUSSION

All runs were carried out on a PC Pentium-III 800 MHz computer. To illustrate the non-convexity of the problem, we investigated the performance and robustness of the local solvers \texttt{fminsearch} and \texttt{fmincon} by applying a multistart procedure. Hence, different initial values for the decision variables are randomly selected within their range. Both solvers converged to local solutions in all the runs, confirming the need of using global optimization methods. The histogram of the solutions obtained by \texttt{fmincon} in 14 runs is presented in Figure 2. It is interesting to note that, despite the large computational effort (about 16 h of CPU time per run), none of the runs were able to improve the controllers default settings (which we have normalized to \( J = 1.0 \)).

As expected, the other methods performed much better due to their capabilities to escape from local solutions. Typical convergence curves (i.e. objective function value versus CPU time) from the same initial point are presented in Figure 3 (note the log scale for \( J \)). The best final solutions were consistently found by \texttt{rbfSolve}, which also showed the best convergence rate. Both \texttt{DE} and \texttt{NOMADm} finally reached good values of the performance index, although their convergence rate was clearly worse than that of \texttt{rbfSolve}. It is somewhat surprisingly to the authors why the other surrogate-based method (\texttt{ego}) showed a rather good convergence rate but stopped prematurely in all runs. We are currently investigating if this behaviour can be avoided. In the meantime, as a preliminary conclusion, it seems that surrogate models based on \texttt{RBFs} perform better than \texttt{Kriging}.

The best solutions found by each solver are presented in Table 1, where it can be seen that several solvers were able to improve the controllers default settings. Simulations using the best solution obtained by \texttt{rbfSolve} and denoted by \( v^* \), were performed to compare other criteria like effluent quality and pollutant violations with those using \( v^{ref} \). Results are presented in Table 2, indicating that \( v^* \) is also better than \( v^{ref} \) for those criteria (except in the case of pumping energy). Hence, although we only minimized an ISE-based objective function in this work, the optimized plant also performed better in almost all relevant issues.
Optimization problems which arise during an integrated design and control approach are frequently non-convex. Thus, it is not surprising to that global optimization (GO) has received much attention during the last decade from many researchers in computer aided process engineering. However, the current state of the art is far from satisfactory, especially when we consider the global optimization of complex process models which are typically expensive to evaluate. In this contribution, we optimize of complex process models which are satisfactory, especially when we consider the global optimization (GO).

Results obtained indicate that surrogate-based methods, and in particular the ones based on RBFs, can indeed reduce computation times significantly while ensuring convergence to the best known solutions.

VI. CONCLUSIONS

REFERENCES


Table 2. Values of WWTP performance criteria for the reference and the optimized decision vector

<table>
<thead>
<tr>
<th>Criterion</th>
<th>v^ref</th>
<th>v*</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ Index</td>
<td>9031.7</td>
<td>8980.4</td>
<td>kg poll:units/day</td>
</tr>
<tr>
<td>Viol. N-level</td>
<td>11.16</td>
<td>9.97</td>
<td>% of time</td>
</tr>
<tr>
<td>Viol. NH4^+-level</td>
<td>25.92</td>
<td>25.15</td>
<td>% of time</td>
</tr>
<tr>
<td>Aeration Energy</td>
<td>7172.6</td>
<td>7164.4</td>
<td>kWh/day</td>
</tr>
<tr>
<td>Pumping Energy</td>
<td>1919.3</td>
<td>1984.6</td>
<td>kWh/day</td>
</tr>
</tbody>
</table>

Table 1. Best optimization results obtained by the different optimization algorithms

<table>
<thead>
<tr>
<th>Solver</th>
<th>K(O)</th>
<th>ϑ(O)</th>
<th>J</th>
<th>K(N)</th>
<th>ϑ(N)</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>fmincon</td>
<td>661.39</td>
<td>3.98·10^{-1}</td>
<td>6.59·10^{-1}</td>
<td>28478</td>
<td>3.74·10^{-1}</td>
<td>5.85·10^{-2}</td>
</tr>
<tr>
<td>fminsearch</td>
<td>819.08</td>
<td>4.87·10^{-1}</td>
<td>4.51·10^{-1}</td>
<td>26508</td>
<td>9.47·10^{-2}</td>
<td>2.55·10^{-2}</td>
</tr>
<tr>
<td>NOMADm</td>
<td>750.62</td>
<td>6.91·10^{-1}</td>
<td>6.30·10^{-1}</td>
<td>21175</td>
<td>3.07·10^{-2}</td>
<td>2.97·10^{-3}</td>
</tr>
<tr>
<td>DE</td>
<td>357.51</td>
<td>1.73·10^{-1}</td>
<td>1.00·10^{-4}</td>
<td>17876</td>
<td>2.00·10^{-2}</td>
<td>1.00·10^{-4}</td>
</tr>
<tr>
<td>rbfSolve</td>
<td>375.31</td>
<td>7.00·10^{-1}</td>
<td>1.00·10^{-4}</td>
<td>21018</td>
<td>2.71·10^{-2}</td>
<td>1.46·10^{-2}</td>
</tr>
<tr>
<td>ego</td>
<td>444.59</td>
<td>7.00·10^{-1}</td>
<td>1.00·10^{-4}</td>
<td>6129</td>
<td>1.00·10^{-2}</td>
<td>7.00·10^{-2}</td>
</tr>
</tbody>
</table>


