COMPARISON OF ITERATIVE SET-POINT OPTIMISATION STRATEGIES UNDER STRUCTURAL PLANT-MODEL MISMATCH

Weihua Gao and Sebastian Engell

Process Control Laboratory, Department of Biochemical and Chemical Engineering, Universität Dortmund
Emil-Figge-Str. 70, D-44221 Dortmund, Germany

Abstract: This paper considers the set-point optimisation problem under plant-model mismatch. Three iterative optimisation strategies, the two-step method, the integrated system optimisation and parameter estimation (ISOPE) method and the gradient-modification optimisation method are compared for a nonlinear chromatographic separation process using different structurally mismatched models as nominal process models. The gradient-modification optimisation method is proven to be very attractive when a large structural plant-model mismatch exists. Its ability to deal with disturbances is also shown. Copyright © 2005 IFAC

Keywords: Set-point optimisation, iterative optimisation, batch-to-batch control, plant-model mismatch, chromatography.

1. INTRODUCTION

In the chemical, food, consumer care, and pharmaceutical industries, many production processes are performed in a repetitive fashion where a certain quantum of material (called a batch) is processed in a piece of equipment (for example a mixer, a reactor, or a crystalliser) for a certain period of time, then the products are transferred to another piece of equipment, processed again, etc., until the product has the desired properties and shape. Similar processes are found in the metal industry, e.g. a blast furnace or the forming of large parts of specially shaped metal. Also the treatment of a number of parts of one type on a machine for a certain period of time, followed by operations on another type of parts can be considered as a sort of a batch or, more generally, repetitive process.

Usually, continuous control of certain important process parameters over the course of one batch is performed. In order to operate the processes efficiently, the set-points of the control loops should also be optimised. Due to the existence of model errors and of unmeasurable disturbances, pre-computed operation policies will not lead to an optimal operation, and the constraints on the process may be violated. This paper is concerned with the iterative (or batch-to-batch or repetitive) adaptation of the operating parameters of repetitive processes using both a process model and the data collected from previous batches. The focus of attention is on the issue of plant-model mismatch – an optimisation solely based upon a fixed process model will not lead to an optimal performance of the real process.

In the classical two-step approach, the model parameters are updated by a parameter estimation procedure so that the model represents the plant at current operating conditions more accurately. The updated model is used in the optimisation procedure to generate a new set-point. This method works well for parametric mismatch. However, it does not guarantee an improvement of the set-point when structural errors in the model are present. A modified two-step method was proposed by Roberts (Roberts, 1979; Roberts and Williams, 1981), known as Integrated System Optimisation and Parameter Estimation (ISOPE). Beside the parameter estimation procedure, a gradient-modification term is added to the objective
function of the optimisation problem. ISOPE generates set-points which converge towards the true optimum despite parametric and structural plant-model mismatch. Tatjewski pointed out that the parameter estimation is not necessary when a model-shift term is added to the objective function and redesigned the ISOPE to a new form (Tatjewski, 2002). It was extended recently to handle also process-dependent constraints and improved with respect to the estimation of the gradients of the real system (Gao and Engell, 2004a, b).

In this paper, we compare these three different iterative optimisation strategies. The comparison is performed for the example of batch chromatography, a widespread and economically important separation technology which is frequently applied in the production of life-science products which are subject to strict purity requirements and tight regulations because it provides high product purities at moderate temperatures. The optimal operation of these processes with respect to throughput and solvent consumption is an important factor to reduce the overall production cost in the life science industry.

2. ITERATIVE OPTIMISATION STRATEGIES

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

$$\min_u J(u, y) \quad \text{s. t. } y = f(u, \beta) \leq 0,$$

$$u_{\text{min}} \leq u \leq u_{\text{max}}$$

(1)

where $J(u, y)$ is a scalar objective function, $u$ is a vector of optimisation variables (set-points), $y$ is a vector of measurable output variables. The relationship between $u$ and $y$ is represented by the model $y = f(u, \beta)$, where $\beta$ is a vector of model parameters. $g(u)$ is a vector of constraints. We assume that the real mapping of the plant is $y = f(u)$, which is different from the model because of modelling errors and changes in the plant behaviour due to disturbances and degradation. Therefore, offline optimisation of the operating conditions using the nominal plant model may yield a set-point far away from the true optimum.

3.1 Two Step Method

The two step method is also called Real-Time Optimisation (RTO). The measured output of the plant after the last set-point has been applied is used to update the model. Let $y^{(k)}$ denote the plant output at set-point $u^{(k)}$, where $k$ is the iteration index. The updating of the model can be written as an optimisation problem:

$$\min_\Delta \| y^* - f(u, \beta) \|_{u^{(k)}}.$$  

(2)

Then the updated model is used in the optimisation procedure. A bound $u^{(k)} - \Delta u \leq u \leq u^{(k)} + \Delta u$ is usually added to the optimisation problem to avoid overly aggressive changes of the set-points.

3.2 ISOPE

The Integrated System Optimisation and Parameter Estimation strategy was proposed first by Roberts (Roberts, 1979). Besides the model updating, the objective function in the optimisation problem (1) is adjusted by a gradient modification term

$$\min_u J(u, y) + \lambda^{(k)} u \quad \text{s. t. } 

(3)

where

$$\lambda^{(k)} = \left[ (y^* - y_u) J_y(u, y) \right]_{u^{(k)}}.$$  

Let $u^{(k)}$ denote the solution of (3), then the next set-point is chosen as

$$u^{(k+1)} = u^{(k)} + D \left[ u - u^{(k)} \right],$$

(4)

where $D$ is a damping term. ISOPE generates set-points which converge towards the true optimum despite parametric and structural model mismatch. Theoretical optimality and convergence of the method were proven by Brdys (Brdys et al., 1987).

3.3 Gradient Modification Method

The optimality of the result of ISOPE results from the gradient modification term. Tatjewski redesigned the ISOPE method resulting in a new algorithm that does not require parameter estimation (Tatjewski, 2002). The key idea is to introduce a model shift term $a^{(k)} = y^{(k)} - y^{(k)}$ in the modified objective function:

$$\min_u J(u, y + a^{(k)}) + \lambda^{(k)} u \quad \text{s. t. } 

(5)

As the optimality of the result is solely due to the gradient-modification in the optimisation problem, the redesigned ISOPE could be better termed "iterative gradient-modification optimisation method".

Handling of Constraints

If the constraints involve the behaviour of the real process (not only the inputs to the plant), they cannot be assumed to be met using a model for the computation of the constraint functions. In the original derivation of the ISOPE method, constraints were assumed to be process-independent. An extension of the ISOPE strategy which considers output-dependent constraints can be found in (Brdys et al., 1986). Tatjewski proposed to use a follow-up constraint controller which is responsible for satisfying the output constraints (Tatjewski et al., 2001).

A new method to handle the process-dependent constraints was proposed recently (Gao and Engell, 2004a, b). It is based on the idea to use information acquired at the last set-point to modify the model-based constraint functions at the current iteration.
The modified constraint functions approximate the true constraint functions of the process in the vicinity of the last set-point. Let \( g(u) \) denote the model-based constraint function and \( g'(u) \) denote the actual constraint function of the process. Then the modified constraint function is formulated as:

\[
\hat{g}^{(k)}(u) = g(u) + g'(u^{(k)}) - g(u^{(k)}) + \left((g')'(u^{(k)}) - g''(u^{(k)}))\right)(u - u^{(k)}).
\]  

(6)

The modified constraint function has the following properties at \( u^{(k)} \):

- The modified constraint has the same value as the real constraint function:

\[
\hat{g}^{(k)}(u^{(k)}) = g^{(k)}(u^{(k)}).
\]

- The modified constraint has the same first order derivative as the real constraint function:

\[
\left(\hat{g}^{(k)}\right)'(u^{(k)}) = (g')'(u^{(k)}).
\]

Therefore, the modified constraint function approximates the real constraint functions in the vicinity of the last set-point. The bound which is used to avoid aggressive changes also guarantees that the real constraint is not violated greatly. The optimisation problem in the gradient modification method can thus be stated as:

\[
\min_{u} J(u, y + a^{(k)}) + \lambda^{(k)}(u)
\]

s. t. \( \hat{g}^{(k)}(u) \leq 0 \)

\[
u^{(k)} - \Delta u \leq u \leq u^{(k)} + \Delta u
\]

\[
u^{\min} \leq u \leq u^{\max}.
\]

(7)

The value of \( \Delta u \) should be chosen according to the quality of the model. If the model is reliable, the bound can be loosened by increasing the value of \( \Delta u \).

**Estimation of the Gradients**

A key element of the iterative gradient-modification optimisation method is to estimate the gradient of the plant mapping. A novel method was proposed by (Gao and Engell, 2004a, b). It uses the previous set-points as a basis for the finite difference approximation of the gradient at the current set-point. Additional set-point perturbations at optimised locations are added to decrease the influence of measurement errors.

3. BATCH CHROMATOGRAPHY

The principle of batch chromatography in the elution mode is shown in figure 1. A mixture is periodically injected into a column filled with solid adsorption particles. Due to different adsorption affinities, the components in the mixture migrate at different velocities and therefore they are gradually separated. At the outlet of the column, the purified components are collected between the so-called cutting points \( t_1 - t_4 \). The locations of \( t_1 \) and \( t_2 \) are determined by a minimum concentration limit. The locations of \( t_3 \) and \( t_4 \) are determined by the purity requirements on the products as follows:

\[
\begin{align*}
\int_{t_1}^{t_2} c_1(t) dt &= Pur_1 \\
\int_{t_1}^{t_4} c_2(t) dt &= Pur_2 \\
t_{w,f,2} - \int_{t_1}^{t_3} c_1(t) dt &= Pur_2 \\
t_{w,f,2} - \int_{t_1}^{t_4} c_2(t) dt + t_{w,f,1} - \int_{t_1}^{t_3} c_1(t) dt \\
\end{align*}
\]

(8)

where \( c_1(t) \) and \( c_2(t) \) are the elution profiles. \( Pur_1 \) and \( Pur_2 \) are the purity requirements, \( c_{f1} \) and \( c_{f2} \) the feed concentrations. The flow rate \( Q \) and the injection period \( t_{inj} \) are considered as the manipulated variables here. The cycle period \( t_{cy} \) is fixed to the duration of the chromatogram \( (t_4 - t_1) \).

The set-point optimisation is based on the general rate model (Guiochon, et al., 1994), abbreviated as GRM\((Q, t_{inj})\):

\[
\begin{align*}
\frac{\partial c_i}{\partial t} + u \frac{\partial c_i}{\partial x} + 3(1-\varepsilon) k_{ij} r (c_i - c_{p,j}) (r_p) &= D_{ax,i} \frac{\partial^2 c_i}{\partial x^2} \\
(1-\varepsilon) \frac{\partial c_{p,i}}{\partial t} + \varepsilon \frac{\partial c_{p,j}}{\partial t} &= \varepsilon D_{p,i} \left[ \frac{1}{r_p} \frac{\partial c_{p,j}}{\partial r} \right] .
\end{align*}
\]

(9)

These two partial differential equations describe the concentrations in the mobile phase \( c_i \) and in the stationary phase \( (q_i \) and \( c_{p,i} \). The adsorption isotherms are used to relate the concentrations \( q_i \) (substance i adsorbed by the solid) and \( c_{p,i} \) (substance i in the stationary liquid phase). A commonly assumed isotherm functional form is the Bi-Langmuir isotherm:

\[
q_i = \frac{a_i c_{p,i}}{1 + \sum b_{p,j} c_{p,j}} + \frac{a_i c_{p,i}}{1 + \sum b_{p,j} c_{p,j}} .
\]

(10)

The set-point optimisation of batch chromatography can be formulated as

\[
\begin{align*}
\max_{Q, t_{inj}} \left( \int_{t_1}^{t_4} c_1(t) dt + \int_{t_1}^{t_4} c_2(t) dt \right) \frac{Q}{(t_4 - t_1)} \\
\text{s. t.} \quad [c_1(t) c_2(t)] &= \text{GRM}(Q, t_{inj}) \\
\text{Rec}(Q, t_{inj}) \geq \text{Rec}_{\text{min}} \\
0 \leq Q \leq Q_{\text{max}} \\
0 \leq t_{inj} \\
\end{align*}
\]

(11)

Fig. 1. Illustration of a chromatographic batch separation process.
The constraint on the recovery yield \( \text{Rec} \) is a process-dependent constraint while the other constraints are input constraints.

An efficient numerical solution of the general rate model (9) incorporating arbitrary nonlinear isotherms was proposed by Gu (1995). The mobile phase and the stagnant phase are discretised using a Galerkin method on finite elements for the mobile phase and orthogonal collocation for the stagnant phase. The resulting ode system is solved using a solver which is based on Gear’s method for stiff odes.

The numerical solution yields the concentrations of the components in the column at different locations and times. The concentration information at the outlet of the column is used to generate the chromatogram from which the production rate and the recovery yield can be computed.

A model-based online optimisation strategy was proposed for batch chromatography in (Dünnebier et al., 1999, 2001). In their approach, online parameter estimation is performed to improve the model accuracy and to track changes in the plant. It is thus a representative of the two-step approach. This technique has been tested experimentally at pilot scale plants and works well for separations with linear adsorption isotherms (Dünnebier et al., 1999). However, most chromatographic separation processes are characterised by nonlinear adsorption isotherms which are often not reproduced exactly by the standard isotherm models (e.g. Langmuir, Bi-Langmuir). One important reason for this is the presence of additional components in the mixture. If a mismatch between the model and the plant occurs, model-based optimisation may not give good results and the constraints must be established by an additional control layer (Hanisch, 2002), causing a loss of performance.

4. COMPUTATIONAL STUDY

The iterative optimisation strategies were tested in a simulation study of a batch chromatographic separation of enantiomers with highly nonlinear adsorption isotherms which has been used as a test case in laboratory experiments before (Hanisch, 2002). A GRM model with Bi-Langmuir isotherm (12) that was fitted to measurement data is considered as the plant in the simulation study:

\[
q_i = \frac{a_i}{1 + \sum_{j \neq i} b_{ij} c_j} + \frac{23.412 c_i}{1 + 1000 c_i + 1008 c_i} + \frac{23.412 c_i}{1 + 333.1 c_i + 94.5 c_i} \\
q_2 = \frac{8.979 c_2}{1 + 211.4 c_2 + 43.46 c_2} .
\]

(12)

Three different isotherm forms were used in the nominal optimisation model:

a) Langmuir isotherm with common saturation capacity (3 parameters)

\[
b_i = \frac{a_i}{1 + \sum_{j \neq i} b_{ij} c_j} i = 1, 2
\]

b) Langmuir isotherm with different saturation capacities (4 parameters)

\[
b_i = \frac{a_i c_i}{1 + \sum_{j \neq i} b_{ij} c_j} i = 1, 2
\]

c) Asymmetric Langmuir isotherm (6 parameters)

\[
b_i = \frac{a_i c_i}{1 + \sum_{j \neq i} b_{ij} c_j} i = 1, 2
\]

In the study, only the second component was considered to be a valuable product. The aim of the set-point optimisation was to maximise its production rate at 98% purity. The recovery yield should be greater than 80%, i.e. no more than 20% of the valuable component should be present in the waste fraction. The maximal permitted flow rate is 2.06 cm/s, corresponding to a maximal velocity of 0.42 cm/s. The length of the column is 10.8 cm.

We used the fluid velocity instead of the flow rate in the set-point optimisation and normalised it as well as the injection period to the interval \([0, 1]\), which represents \([0, 0.42]\) cm/s for the velocity and \([50, 150]\) s for the injection period. The bound \(\Delta u\) was set to 0.08. There was no damping, i.e. \(D = 1\). The starting set-point was \((0.20, 100)\). The iterations were stopped when the distance of sequential set-points was less than 0.01. For the two-step method and the ISOPE method, the parameters of the isotherms were estimated by a least squares method to minimise the difference of the elution profiles between the plant and the nominal model:

\[
\sum_{i} \sum_{j} (c_{i,j}^{\text{plant}} - c_{i,j}^{\text{model}})^2
\]

(16)
where $i$ is the index of the components, $j$ is the index of the sampling points.

Table 1 Simulation results: final set-point.

<table>
<thead>
<tr>
<th>Optim. strategy</th>
<th>Isotherm form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two Step</td>
<td>0.37, 116</td>
</tr>
<tr>
<td>ISOPE</td>
<td>0.42, 100</td>
</tr>
<tr>
<td>Gradient</td>
<td>0.42, 99</td>
</tr>
<tr>
<td>Modif.</td>
<td>0.42, 100</td>
</tr>
</tbody>
</table>

The simulation results are summarised in table 1. The optimium for the “real” plant is (0.42, 100). The two-step method stopped at different set-points in the first two cases because the structural mismatch between the plant and the model can not be eliminated by the parameter estimation procedure. In case c, it arrived at the real optimum because the estimation procedure improved the model efficiently. Actually, the mismatch in case c is nearly parametric. The final estimated isotherm is shown below (17), the difference to the true isotherm is very small.

$$q_i = \frac{23.421c_i}{1 + 333.1c_i + 94.739c_i^2}$$

$$q_2 = \frac{8.98c_2}{1 + 211.44c_2 + 43.46c_2^2}.$$ (17)

The ISOPE method and the gradient modification converged to the real optimum in all three cases. This is due to the influence of the gradient modification term. It corrects the mapping used by the optimisation problem in each iteration by using the gradient information of the plant. Therefore, although there exists a large plant-model mismatch, the true optimum was attained. A drawback is that additional batch runs with perturbed set-points are required in order to estimate the true plant gradient. Table 2 shows the number of iterations and the overall number of batch experiments for each of the methods.

Table 2 Simulation results: number of iterations /number of experiments.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Isotherm form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Two Step</td>
<td>16</td>
</tr>
<tr>
<td>ISOPE</td>
<td>11</td>
</tr>
<tr>
<td>Gradient Modif.</td>
<td>11</td>
</tr>
</tbody>
</table>

Fig. 3 and Fig 4 show the trajectories of the set-point for case a and b. The cross and star symbols denote the set-point perturbations for each method. The contours of production rate and recovery yield of the plant are also depicted. The recovery limit was kept well by using the new method (Gao and Engell, 2004b). Compared with the gradient modification method, the ISOPE method did not improve the optimisation by additionally performing parameter estimation. This demonstrates that the gradient modification term is the key ingredient in both methods. When the structural mismatch between plant and model is large, the parameter estimation does not help much any more. By avoiding the model update, the computation time is decreased considerably.

Fig. 3. Trajectories of the set-point for case a. ISOPE method - , perturbations × ; gradient modification method - , perturbations ±.

Fig. 4. Trajectories of the set-point for case b. ISOPE method - , perturbations × ; gradient modification method - , perturbations ±.

When unmodelled disturbances are present, the gradient modification method can also adjust the set-point to a new optimum. Fig. 5 and Fig. 6 show the set-point trajectory when ±10% disturbances of the feed concentrations $c_{i1}$ and $c_{i2}$ were added. At the beginning there was no disturbance. The iterative optimisation arrived at the optimum in 7 iterations. Then the feed concentrations were changed to 110% of the initial values. This increased the overlap of the elution profiles, and therefore decreased the recovery yield to 70% at the original optimal set-point. The constraint violation activated the iterative optimisation procedure. Two set-point perturbations were applied to estimate the new gradient of the plant at the former optimum. The process was moved to the new optimum in 3 iterations. The production rate decreased slightly and the recovery yield returned to 80%. When the feed concentrations were changed to 90% of the initial values, the overlap of the elution profiles decreased. The recovery yield (now 92%) was too high for a cost optimal operation. Again the iterative optimisation was activated and converged to the new optimum in a few iterations. The recovery decreased back to 80% while the production rate increased from $6.38 \times 10^4$g/s to $7.32 \times 10^4$g/s.
**Fig. 5. Trajectories of the set-point for feed concentration disturbances.** $c_{f0} - \circ$, $1.1\times c_{f0} - \times$, $0.9\times c_{f0} - \bullet$.  

**Fig. 6. Trajectories of the production rate and of the recovery yield for feed concentration disturbances.**

**CONCLUSIONS**

In many batch processes, e.g. in chromatographic separations, the identification of a precise model for process optimisation requires considerable efforts and costly experiments and often even is practically impossible. Inaccurate models can be used together with an iterative optimisation method to operate the real plant at the optimum. We compared different iterative optimisation strategies in this paper. The two-step method works well for a parametric plant-model mismatch. Its application in cases with a structural plant-model mismatch fails due to the impossibility of improving the model by the parameter estimation. The ISOPE method and the gradient modification methods give good results for both parametric and structural plant-model mismatch. The latter method does not require a parameter re-estimation, and therefore it is a better choice when large structural mismatches exist. Also time-varying disturbances can be dealt with efficiently. The price paid in both methods is that they require additional set-point perturbations and hence batch runs at suboptimal operating points to estimate the plant gradient, which may also be sensitive to measurement errors.

**ACKNOWLEDGEMENTS**

The work of the first author was supported by a scholarship of the NRW Graduate School of Production Engineering and Logistics at Universität Dortmund. This support is very gratefully acknowledged.

**REFERENCES**


