# Convergence Analysis of Iterative Identification and Optimization Schemes

B. Srinivasan and D. Bonvin

Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland.

## Abstract

The use of measurements to compensate for model uncertainty has received increasing attention in the context of process optimization. The idea consists of iteratively using the measurements for identifying model parameters and the updated model for optimization. This paper investigates the convergence of various iterative identification and optimization schemes in the presence of model mismatch. The optimization can be model-based, data-based or of mixed nature. Based on the advantages and drawbacks of the various approaches, a novel scheme is proposed, by which the optimization starts model-based so as to ensure fast improvement and finishes as a data-based approach so as to converge towards the true optimum. The performance improvement obtained with the proposed methodology is illustrated via the simulation of a semi-batch reaction system.

**Keywords:** Optimization, Iterative schemes, Convergence, Data-based optimization, Parameter identification, Run-to-run optimization.

## 1 Introduction

Optimization has received increasing attention in industry since its provides a unified framework for improving productivity while meeting safety and quality constraints. Many optimization techniques are model-based, and they must be able to cope with the following two sources of uncertainty: (i) modeling errors are invariably present since reliable models are rarely available at the industrial level, and (ii) disturbances arise naturally due for example to variations in initial conditions and process operation. Optimal operation under uncertainty requires either a robust optimization approach or the use of measurements to adjust on-line the off-line calculated strategy [9].

In order to compensate the effect of model mismatch, the idea of iterating between identification and optimization has been used extensively in the context of model-based optimization. Numerous studies in the literature document the identification and optimization approach, for example in the context of on-line optimization and model-predictive control [4, 3] or with respect to run-to-run optimization [2, 6]. However, the convergence of these schemes towards the true optimum requires: (i) the excitation generated by the optimal inputs to be sufficient to identify all uncertain parameters, and (ii) no model mismatch. Since these two assumptions are nearly impossible to verify in practice, the iterative identification and optimization schemes typically suffer from convergence problems.

One improvement has consisted in matching the objective functions of the identification and optimization problems, thereby providing more synergy between the two steps [8]. With this approach, the goal of identification is not to come up with a model that best predicts the true outputs, but rather to generate a model that allows computing inputs that are nearly optimal for the reality.

In this paper, the convergence issue using classical (i.e. non-matched) cost functions is considered. The convergence properties of various iterative identification and optimization schemes are analyzed in a deterministic framework. Thus, only model mismatch will be considered, without measurement noise. It will be shown that fast convergence to the true optimum can be obtained by using initially a model-based scheme (ensures fast approach but no convergence) and then a data-based scheme (experimentally expensive but converges to the true optimum).

The paper is organized as follows. Section 2 introduces the various model- and data-based optimization schemes. The unconstrained static optimization problem will be considered first. The results will be extended to the case of constrained dynamic optimization problems in the example section. Section 3 discusses the modification of these schemes when the model is inaccurate and needs to be refined during optimization. A novel scheme that combines the advantages of model- and data-based optimization approaches is proposed in Section 4. Section 5 compares the performance of the various schemes via the simulation of a semi-batch reaction system, and conclusions are drawn in Section 6.

## 2 Model-based vs. Data-based Optimization

Consider the following unconstrained static optimization:

$$\min_u J^{opt} = \phi(x,u) \qquad (1)$$

$$s.t. \qquad F(x,u) = 0 \qquad (2)$$

where $J^{opt}$ is the the cost function to be minimized, $x$ the state vector, $u$ the input vector, $F$ the system equations, and $\phi$ a smooth scalar function expressing the cost. Upon elimination of the states using (2), the optimal inputs read:

$$u^* = \arg\min_u J^{opt}(u) \qquad (3)$$

where the superscript $^*$ indicates a quantity calculated via optimization.

A typical optimization procedure is depicted in Figure 1. The first iteration starts with $k = 1$ and some input values $u_0$. Then, a direction $d_k$ in input space is found along which the cost can be reduced. This is typically done by evaluating the gradient of the cost function with respect to the inputs, $d_k = H \frac{\partial J^{opt}}{\partial u}|_{u_{k-1}}$, where $H$ is a positive definite matrix. The minimum along the decent direction is found iteratively using a line search procedure that results in the optimal step size $\alpha_k$. The new input vector becomes:

$$u_k = u_{k-1} + \alpha_k d_k \qquad (4)$$

The iterations in $k$ are repeated until there is no variation in $u_k$, the convergence being guaranteed by the line search.
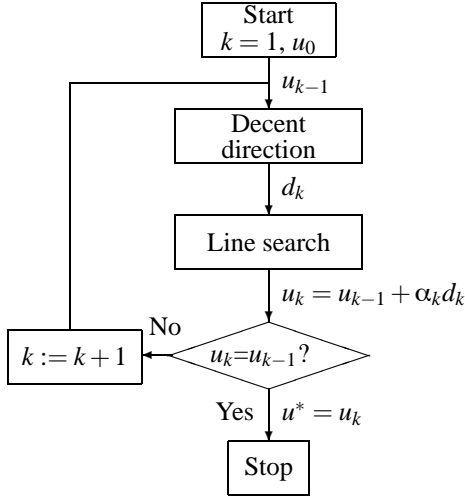


**Figure 1:** Iterative optimization scheme

The standard approach to optimization is model-based, whereby a model is used to evaluate the cost function. In contrast, in purely data-based optimization, the cost function is measured directly on the real plant [1]. There may also be hybrid approaches for which certain intermediary steps are model-based and the others are data-based. Since it will be important to distinguish between the reality and the model, the accent $(\bar{\cdot})$ will be used whenever necessary to denote a quantity related to the model.

### 2.1 Model-based numerical optimization
Numerical optimization of the model gives:

$$\bar{u}^* = \arg\min_{\bar{u}} \bar{J}^{opt}(\bar{u}, \theta) \qquad (5)$$

where $\theta$ are the model parameters. The iterations of the model-based numerical optimization follow exactly the

steps of Figure 1, with $u_k$, $d_k$ and $\alpha_k$ replaced by $\bar{u}_k$, $\bar{d}_k$, and $\bar{\alpha}_k$, respectively:

$$\bar{u}_k = \bar{u}_{k-1} + \bar{\alpha}_k \bar{d}_k \qquad (6)$$

Here, both the evaluation of the decent direction and the line search are done using the model and do not involve any experimentation. With such a procedure, the iterations do converge, but not necessarily to the true optimum due to model inaccuracy, i.e. $\lim_{k\to\infty} \bar{u}_k = \bar{u}^* \neq u^*$.

### 2.2 Data-based experimental optimization
In data-based optimization, $J^{opt}(u)$ is directly obtained from the real plant through an experiment [1]. The experimental optimization scheme also follows the algorithm in Figure 1, with the decent direction being evaluated using finite perturbations of the input elements. Also, the line search is performed through experimentation. Such a procedure is experimentally expensive but, in the absence of measurement noise, the algorithm converges to the true optimum $u^*$, i.e. $\lim_{k\to\infty} u_k = u^*$.

### 2.3 Model-based decent direction and data-based line search
An intermediate approach is possible, where the computation of the decent direction is model-based, while the line search is performed using experimental data. With this approach, the number of experiments can be considerably less than with the purely data-based optimization. The iterations do converge but, since the decent direction can be inaccurate, this procedure need not converge to the true optimum.

## 3 Optimization with Refined Models

The main drawback of model-based optimization is that the models are often inaccurate. Thus, possible improvement is via model identification using measurements. A typical identification problem consists of estimating the parameters $\theta$ so as to minimize a fit criterion:

$$\min_\theta J^{iden} = \sum_{i=1}^{N} \| y_i(u) - \bar{y}_i(u, \theta) \| \qquad (7)$$
$$s.t. \qquad F(x, u, \theta) = 0 \qquad (8)$$
$$\bar{y} = h(x, \theta) \qquad (9)$$

where $J^{iden}$ is the fit criterion, $y$ the measured outputs, $\bar{y}$ the outputs predicted by the model, and $N$ the number of data points available. Upon elimination of the states using (8), the optimal parameter vector can be written as:

$$\theta^* = \arg\min_\theta J^{iden}(u, \theta) \qquad (10)$$

Note that the computation of $J^{iden}$ requires an experimental run, while the computation of the gradient $\frac{\partial J^{iden}}{\partial \theta}$ does not require any further experimentation.

Since $J^{iden}$ is a function of $u$, so will the optimal parameters, i.e. $\theta^*(u)$. Two conditions are necessary for the identified parameters to be independent of $u$: (i) the model structure is correct so that $\theta_{true}$ exists, and (ii) the Hessian $\frac{\partial^2 J^{iden}}{\partial \theta^2}$ is positive definite for the given $u$ (such an input is called persistently exciting [10]). If these two conditions are verified, the identification generates an accurate model and the optimization calculates the optimum for this accurate model, i.e. the true optimum.

Unfortunately, this desirable situation is rarely found in practice since: (i) structural mismatch is always present, and (ii) persistency of excitation is difficult to achieve when the inputs are calculated for the sake of optimality (dual control problem [10]). Consequently, the identified parameters depend on the inputs, and the combination "identification/optimization" leads formally to the solution $(\bar{u}^*, \theta^*)$:

$$\theta^*(\bar{u}^*) = \arg\min_{\theta} J^{iden}(\bar{u}^*, \theta) \tag{11}$$

$$\bar{u}^*(\theta^*) = \arg\min_{\bar{u}} \bar{J}^{opt}(\bar{u}, \theta^*) \tag{12}$$

However, there is no guarantee that this solution exists.

## 3.1 Alternating regression

If a solution exists, one possibility to approach it is via an iterative scheme, where the result of one problem is used in the next. This structure is referred to as *alternating regression* [5]:

$$\theta_k^* = \arg\min_{\theta} J^{iden}(\bar{u}_{k-1}^*, \theta) \tag{13}$$

$$\bar{u}_k^* = \arg\min_{\bar{u}} \bar{J}^{opt}(\bar{u}, \theta_k^*) \tag{14}$$

This iterative procedure, which is depicted in Figure 2, starts with $k = 1$ and some input $\bar{u}_0^*$. Each iteration first solves the identification problem (13) to obtain the updated parameters $\theta_k^*$, which are then used in the optimization problem (14) to calculate the inputs $\bar{u}_k^*$. This procedure is repeated until there is no more variation in $\bar{u}_k^*$.

The main problem with this scheme is that convergence cannot be guaranteed. Due to the optimization step (14), $\bar{J}^{opt}(\bar{u}_k^*, \theta_k^*) \leq \bar{J}^{opt}(\bar{u}_{k-1}^*, \theta_k^*)$. However, after the identification step, there is no guarantee that $\bar{J}^{opt}(\bar{u}_{k-1}^*, \theta_k^*)$ is smaller than $\bar{J}^{opt}(\bar{u}_{k-1}^*, \theta_{k-1}^*)$. Thus, there is no defined relation between $\bar{J}^{opt}(\bar{u}_{k-1}^*, \theta_{k-1}^*)$ and $\bar{J}^{opt}(\bar{u}_k^*, \theta_k^*)$, i.e. between the costs in two successive iterations. This prevents analyzing the convergence in a general setting.

If alternating regression converges, it requires very little experimentation since only the identification problem requires experimentation, while the optimization problem (both decent direction evaluation and line search) is entirely model-based. However, there is no guarantee that $\bar{u}^*(\theta^*) = u^*$

## 3.2 Inner-outer optimization

Another possibility is to consider the identification problem (or its necessary conditions) acting as constraints in the optimization problem. In other words, the model used in the
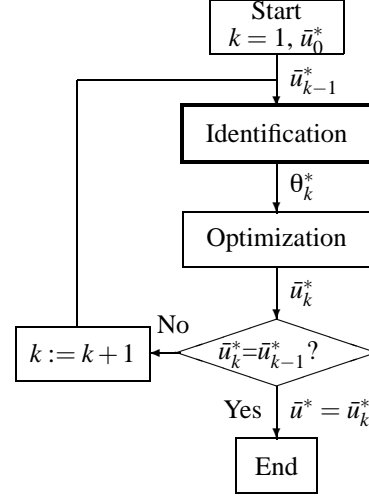


**Figure 2:** Alternating regression (experimentally-expensive blocks in bold)

optimization algorithm will always have optimal parameter values, $\theta^*(\bar{u})$. For this to happen, the identification problem needs to be repeated every time $\bar{u}$ changes. Such a structure is referred to as *inner-outer optimization* or lexicographic preference in the multi-criteria literature [11]:

$$\bar{u}^* = \arg\min_{\bar{u}} \bar{J}^{opt}(\bar{u}, \theta^*(\bar{u})) \tag{15}$$

$$\text{s.t.} \quad \theta^*(\bar{u}) = \arg\min_{\theta} J^{iden}(\bar{u}, \theta)$$

The procedure is depicted in Figure 3. The advantage is that convergence is guaranteed due to the fact that only the cost function $\bar{J}^{opt}$ is considered in the main iteration loop. However, the decent direction evaluation and line search require experimentation (indicated as bold blocks in Figure 3). This is because, for every new input vector $\bar{u}$, the corresponding optimal $\theta^*(\bar{u})$ is needed. Though the procedure converges, it is experimentally very expensive, and there is still no guarantee that $\bar{u}^* = u^*$ due to model mismatch.

## 3.3 Decent direction from a refined model and data-based line search

As in Section 2.3, a hybrid approach can be devised, where the decent direction is computed from a refined model and the line search is data-based [3]. The procedure follows Figure 3 with the following modifications: (i) no iteration between decent direction computation and identification for every new trial value $\bar{u}$, (ii) the second identification and model-based line search are replaced by a data-based line search that uses the measured cost $J^{opt}(\bar{u})$ instead of $\bar{J}^{opt}(\bar{u}, \theta^*(\bar{u}))$.

This scheme is experimentally less expensive than the inner-outer structure, while it still guarantees convergence (yet, not necessarily to the true optimum).
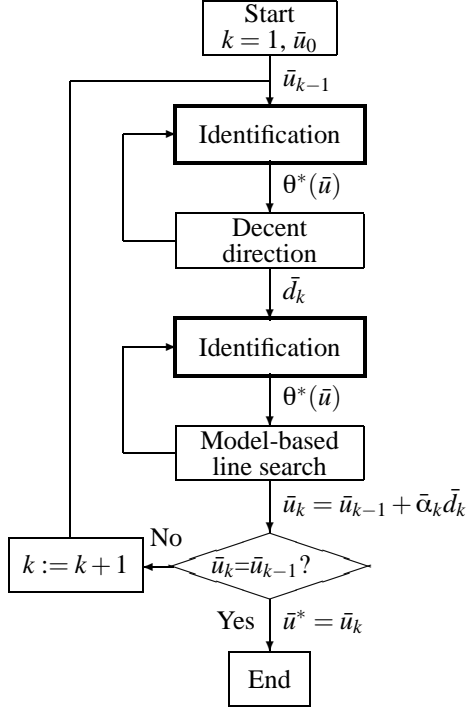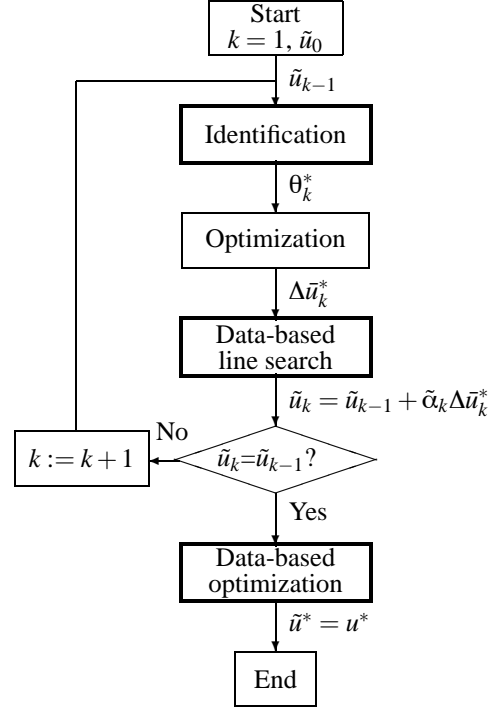
**Figure 3:** Inner-outer optimization



**Figure 4:** Optimization with smooth transition between model- and data-based schemes

## 4 Transition Between Model- and Data-based Schemes

A comparison of model- and data-based methods is presented in Table 1. The two main schemes of interest are: (i) model-based numerical optimization with refined model (Section 3.1) due to little experimentation and fast approach, and (ii) data-based experimental optimization with no model (Section 2.2) due to convergence to the true value.

| # | Decent direction | Line search | Model | Experimental effort | Convergence |
|---|---|---|---|---|---|
| 2.1 | Model | Model | Fixed | None | Not to optimum |
| 2.2 | Data | Data | None | Large | **True optimum** |
| 2.3 | Model | Data | Fixed | Some | Not to optimum |
| 3.1 | Model | Model | Refined | Little | *No convergence* **Fast approach** |
| 3.2 | Model | Model | Refined | Large | Not to optimum |
| 3.3 | Model | Data | Refined | Some | Not to optimum |
| 4 | Model/ Data | Model/ Data | Refined | Some | **True optimum** **Fast approach** |

**Table 1:** Comparison of model- and data-based optimization schemes

The approach proposed in this paper is depicted in Figure 4. It starts as model-based numerical optimization, and ends up as a data-based optimization scheme. To make a smooth transition between the two, a data-based line search as in Section 3.3 is incorporated.

The details of the procedure are as follows. Alternating re-gression is used to determine the optimal correction term $\Delta\tilde{u}_k^*$ (using model-based decent direction and model-based line search). As long as alternating regression progresses towards the optimum, the data-based line search simply implements the newly found optimum and verifies that the measured cost indeed decreases. Otherwise, the data-based line search adapts the step-size $\tilde{\alpha}_k$ to enforce a reduction in cost. At this stage, the computation of the decent direction is model-based and the line search is data-based. Continuing along, when the data-based line search finds that no further reduction in cost is possible in the direction pointed out by the model, i.e. $\tilde{\alpha}_k = 0$, then the scheme switches to purely data-based optimization, where both the decent direction computation and line search are done using the data. This way, a smooth transition is introduced from reliance on the model to reliance on the data.
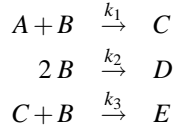
The features of this scheme are as follows:

- The main reason for the lack of convergence of the model-based iterative schemes is the absence of a global line search in the overall loop. In contrast, the new scheme proposes a global data-based line search.

- A parallel can be drawn with sequential quadratic programming, where a local optimization problem is solved to provide the decent direction for the global problem. The local problem in this case is model-based, while the global problem is data-based.

- The data-based optimization scheme may require a large number of additional runs to converge. However, since these runs start from an operating point relatively close to the optimum, the loss in optimality is low.

| $k_1$ | 0.1 | l/mol min | $t_f$ | 120 | min |
|---|---|---|---|---|---|
| $k_2$ | 0.2 | l/mol min | $c_{max}$ | 0.15 | mol/l |
| $k_3$ | 0.03 | l/mol min | $u_{min}$ | 0 | l/min |
| $c_{b_{in}}$ | 5 | mol/l | $u_{max}$ | 0.002 | l/min |
| $c_{a0}$ | 1 | mol/l | $\alpha$ | 0.01 | mol |
| $c_{b0}$ | 0 | mol/l | $\varepsilon$ | 0.01 | mol/l |
| $V_0$ | 1 | l | | | |

**Table 2:** Parameter values and initial conditions



**Figure 5:** Optimal feedrate (the singular arc is approximated by the constant profile $u_s$)

## 5 Run-to-run Optimization of a Semi-batch Reactor

The performance of different schemes will be compared in simulation of an isothermal semi-batch reaction system for the acetoacetylation of pyrrole with diketene [7]. Pyrrole is initially in the reactor and diketene is fed so as to maximize the production of acetoacetyl pyrrole. The simulated reality has the three reactions:

$$A + B \overset{k_1}{\to} C$$
$$2B \overset{k_2}{\to} D$$
$$C + B \overset{k_3}{\to} E$$

where A: pyrrole, B: diketene, C: 2-acetoacetyl pyrrole, D: dehydroacetic acid, E: undesired by-product.

The optimization problem is formulated mathematically as follows:

$$\max_{u(t)} J^{opt} = c_c(t_f)V(t_f) - p(c_d(t_f) + c_e(t_f) - c_{max}) \quad (16)$$

$$s.t. \quad \dot{c}_a = -k_1 c_a c_b - (u/V)\, c_a$$
$$\dot{c}_b = -k_1 c_a c_b - 2\, k_2 c_b^2 - k_3 c_b c_c$$
$$\qquad + (u/V)\, (c_{b_{in}} - c_b)$$
$$\dot{c}_c = k_1 c_a c_b - k_3 c_b c_c - (u/V)\, c_c$$
$$\dot{c}_d = k_2 c_b^2 - (u/V)\, c_d$$
$$\dot{c}_e = k_3 c_a c_b - (u/V)\, c_e$$
$$\dot{V} = u$$
$$u_{min} \leq u \leq u_{max}$$

where $c_a$, $c_b$, $c_c$, $c_d$, and $c_e$ are the concentrations of $A$, $B$, $C$, $D$, and $E$ in mol/l, respectively. The feed consists of Species $B$ with concentration $c_{b_{in}}$. The numerical values of the parameters used in this study are given in Table 2. The goal is to maximize the number of moles of $C$ at the given final time $t_f$ by manipulating the feedrate $u$ (l/min), whilst satisfying bounds on the input. An penalty barrier function $p(s)$ is added to the cost to penalize the production of the undesired species $D$ and $E$ above the prescribed limit $c_{max}$:

$$p(s) = \begin{cases} -\alpha \, \log(-s) & \text{if} \quad s \leq -\varepsilon \\ \alpha \left(1 - \log(\varepsilon) + \frac{s}{\varepsilon}\right) & \text{if} \quad s > -\varepsilon \end{cases} \quad (17)$$

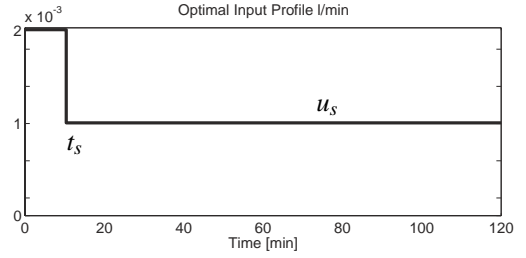The optimal solution computed numerically is shown in Figure 5. It is discontinuous with two intervals: (i) input at its upper bound $u_{max}$, and (ii) singular input approximated here by a constant profile of magnitude $u_s$. The singular arc represents the compromise between producing the desired product $C$ and the undesired side products $D$ and $E$. The maximum value of $J^{opt}$ is 0.371 mol, with the switching time $t_s = 10.84$ min, and $u_s = 9.98 \times 10^{-4}$ l/min.

To introduce model mismatch, it is assumed that the model used in the optimization does not know about the third reaction, i.e. $k_3^{model} = 0$. The concentration of $A$ is measured every 20 min. The parameters $k_1$ and $k_2$ are adapted from run-to-run using the measurements from the previous run only. Since convergence in the presence of model mismatch is studied, no measurement error is considered. The evolution of the cost function over 20 runs are depicted in Figures 6 and 7 for the various schemes. Furthermore, the value of the cost function after the 20th run, and its sum over the 20 runs, $J_{sum}^{opt} = \sum_{k=1}^{20} J_k^{opt}$, are presented in Table 3. It shows that the sum of the cost function over 20 runs is best for the proposed scheme.

| Section | $J^{opt}(20)$ | $J_{sum}^{opt}$ |
|---|---|---|
| 2.1 | 0.314 | 6.29 |
| 2.2 | *0.369* | 6.88 |
| 2.3 | 0.315 | 6.30 |
| 3.1 | 0.344 | 6.85 |
| 3.2 | 0.341 | 6.68 |
| 3.3 | 0.347 | 6.87 |
| 4 | *0.371* | 7.22 |

**Table 3:** Cost function [mol] after 20 runs and sum over the 20 runs

Figure 6 shows that none of the model-based schemes converges to the true optimum ($J^{opt} = 0.371$ mol). Even the data-based line search does not help when the model-based decent direction is incorrect (solid curve). However, the use of a refined model systematically leads to better results than with a fixed model. The estimated parameters are $k_1 = 0.11$ and $k_2 = 0.40$ l/mol min, where $k_2$ is increased to account for the loss of $B$ being consumed by the third (unmodeled) reaction. The inner-outer structure is not very performant and converges to 0.341 mol. Using the decent direction from a refined model and a data-based line search (Section 3.3) improves the cost only marginally to 0.347 mol.

Figure 7 shows that alternating regression converges *rapidly* to 0.344 mol (dashed line), and that both the purely data-based scheme and the proposed hybrid scheme converge (more or less slowly) to the true optimum. It is also seen that the proposed scheme starts out as alternating regression before leaving it to converge to the true optimum. The flat region for the proposed scheme corresponds to the alternating regression that improves very slowly. In this case, it would be appropriate to modify the algorithm and make it switch to data-based optimization earlier. The sluggish region in the data-based optimization corresponds to the runs necessary for decent direction evaluation and line searches.
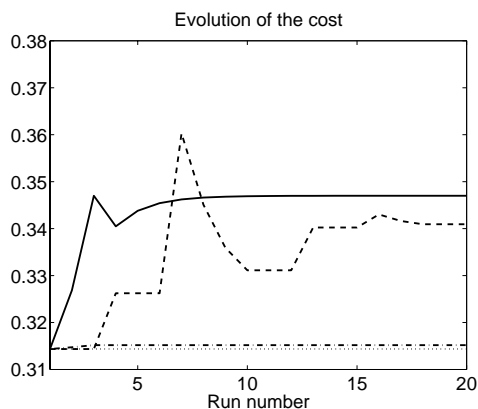


**Figure 6:** Comparison of various iterative model-based optimization schemes: Evolution of the cost over 20 runs. *Dotted* – model-based optimization (Section 2.1), *dash-dotted* – model-based decent direction and data-based line search (Section 2.3), *dashed* – inner-outer optimization (Section 3.2), *solid* – decent direction with refined model and data-based line search (Section 3.3)
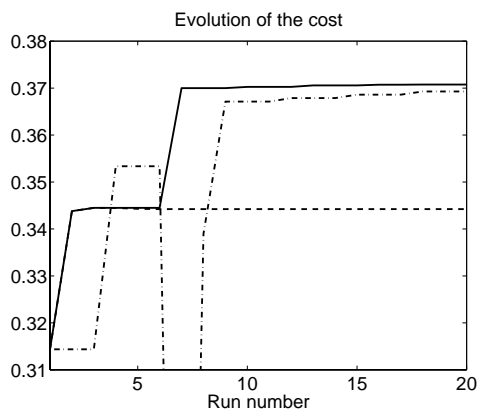


**Figure 7:** Performance of the proposed method compared to related existing schemes: Evolution of the cost over 20 runs. *Dashed* – alternating regression (Section 3.1), *dash-dotted* – data-based optimization (Section 2.2), *solid* - proposed method (Section 4)

## 6 Conclusions

The main advantage of model-based optimization methods is their fast approach towards the optimum while, in the vicinity of the optimum, they have difficulty converging to the true optimum due to the presence of model mismatch.

This paper has illustrated how data-based computation of the decent direction and line search (guaranteeing convergence to the true optimum) can be used in conjunction with model-based techniques so as to speed up converge.

In order to quantify convergence in a deterministic framework, only noise-free data have been considered in this work. Extension of these results to noisy data requires working in a stochastic framework and represents a challenging research direction.

### References

[1]   G. E. P. Box and N. R. Draper. *Empirical Model-building and Response Surfaces*. John Wiley, New York, 1987.

[2]   C. Filippi-Bossy, J. Bordet, J. Villermaux, S. Marchal-Brassely, and C. Georgakis. Batch reactor optimization by use of tendency models. *Comp. Chem. Eng.*, 13:35–47, 1989.

[3]   G. Gattu and E. Zafiriou. A methodology for on-line setpoint modification for batch reactor control in the presence of modeling error. *Chem. Eng. Journal*, 75(1):21–29, 1999.

[4]   A. Helbig, O. Abel, and W. Marquardt. Model predictive control for the on-line optimization of semi-batch reactors. In *American Control Conference*, pages 1695–1699, Philadelphia, PA, 1998.

[5]   E. J. Karjalainen. The spectrum reconstruction problem: Use of alternating regression for unexpected spectral components in two-dimensional spectroscopies. *Chemometerics and Intelligent Laboratory Systems*, 7:31–38, 1989.

[6]   M. V. Le Lann, M. Cabassud, and G. Casamatta. Modeling, optimization, and control of batch chemical reactors in fine chemical production. In *IFAC DYCOPS-5*, pages 751–760, Corfu, Greece, 1998.

[7]   D. Ruppen, D. Bonvin, and D. W. T. Rippin. Implementation of adaptive optimal operation for a semi-batch reaction system. *Comp. Chem. Eng.*, 22:185–189, 1998.

[8]   B. Srinivasan and D. Bonvin. Interplay between identification and optimization in run-to-run optimization schemes. In *American Control Conference*, pages 2174–2179, Anchorage, Alaska, 2002.

[9]   B. Srinivasan, D. Bonvin, E. Visser, and S. Palanki. Dynamic optimization of batch processes: II. Role of measurements in handling uncertainty. *Comp. Chem. Eng.*, 44:27–44, 2003.

[10]   B. Wittenmark. Adaptive dual control methods: An overview. In *IFAC Symposium on Adaptive Syst. in Control and Signal Proc.*, pages 67–72, Budapest, 1995.

[11]   P. L. Yu. Multiple criteria decision making : Five basic concepts. In G.L.Nemhauser, A. H. G. Rinnooy, and M. J. Todd, editors, *Handbooks in Operation Research and Management Science, Volume 1 - Optimization*, pages 663–699, Amsterdam, 1989. North-Holland.