17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)

A Hybrid Optimization Approach to Parameter Estimation

Richard Faber, Harvey Arellano-Garcia, Günter Wozny

Department of Process Dynamics and Operation Sekr. KWT-9, Technical University of Berlin, Straße des 17. Juni 135, D-10623 Berlin, Germany, richard.faber@tu-berlin.de

Abstract

Many parameter estimation problems in chemical or biochemical engineering lead to ill-conditioned and nonconvex optimization problems. For bad starting values the use gradient based result in local optimal solutions. To overcome this drawback, a global optimization approach, Simulated Annealing, has been coupled with a gradient-based SQP approach. To improve the accuracy of the parameter estimates, sensitivity information has been included into the objective function by iteratively adjusting the weighting matrix with the variance-covariance matrix of the model prediction. The hybrid approach has been applied to a case study of biochemical nonlinear parameter estimation problem.

Keywords: hybrid optimization approach, simulated annealing, parameter estimation

1. Introduction

The basis for the design, control and optimization of process systems is a detailed mathematical model. To satisfy the increasing demand on model accuracy, a nonlinear model has to be used in most cases to describe the process as accurately as possible. These models frequently contain process parameters which have to be determined from available measurement data. To estimate the parameters, a model-based error-in-variables (EVM) approach can be used, where a weighted sum of squared errors is minimized subject to the model

equations. Due to the nonlinear nature of the problem the resulting optimization problem is, in many cases, ill-conditioned and nonconvex, resulting in local optimal solutions. Therefore, deterministic gradient-based optimization techniques fail to arrive at satisfactory solutions.

To solve the problem of finding the global optimal solution, a series of global optimization approaches has been developed. Simulated Annealing (SA) is a (physically inspired) stochastic optimization technique where the analogy of the arrangement of atoms during the cooling of metals, leading to the most stable configuration, is used as an optimization procedure. The drawback of this method is that the convergence toward the optimal solution is very slow. Even if the approach is near the global solution it still needs a large number of function evaluations due to the logarithmic cooling rate.

To overcome this drawback, the SA approach has been coupled with a gradient-based SQP approach. The developed optimization framework includes a coordinator which is used to determine the switching time between the two approaches. To improve the convergence properties of the SQP approach, an iterative adjustment of the weighting matrix of the objective function has been used. In the original weighted least-squares formulation of the objective function only the variances of the measurement errors are used. The proposed approach uses the variance-covariance matrix of the model prediction as an additional weighting matrix in order to include sensitivity information into the objective function. This leads to better convergence properties and earlier switching times can be realized.

2. Solution approach

To estimate parameters $\boldsymbol{\Theta}$ in a nonlinear implicit equation system, we usually have several measured data sets of some output (dependent) variables, **y**. A general parameter estimation problem with multiple sets of data can be formulated as follows:

$$\min_{\boldsymbol{\theta}} \Phi = \rho(y) = \sum_{i=1}^{I} \left(\hat{\mathbf{y}}_i - \mathbf{y}_i(\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\theta})\right)^T \mathbf{W}_i \left(\hat{\mathbf{y}}_i - \mathbf{y}_i(\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\theta})\right)$$

$$s.t.$$

$$\mathbf{g}_i(\mathbf{x}_i, \mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\Theta}) = \mathbf{0} \tag{1}$$

$$\mathbf{h}_i(\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\Theta}) \geq \mathbf{0}$$

$$\boldsymbol{\theta}^{LB} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}^{UB} \quad i = 1 \ldots I$$

Where $\mathbf{x} \in X \subseteq \Re^n$, $\mathbf{y} \in Y \subseteq \Re^m$, $\mathbf{u} \in U \subseteq \Re^l$, $\boldsymbol{\Theta} \in \Re^p$, $\mathbf{g} \subseteq \Re^{n+m}$, $\mathbf{h} \subseteq \Re^k$

W is the weighting matrix of the objective function. $\hat{\mathbf{y}}_i$ are the measured values of the dependent variables for data set i. **g** is the vector of the model equations

and **h** the vector of inequality constraints. The inequality constraints represent the process restrictions which should be considered in the parameter estimation.
Two major issues in parameter estimation have been studied in the past. One important issue is the high dimension of the estimation problem. As errors are allowed for all (dependent and independent) variables, the degrees of freedom as well as the number of constraints of the optimization problem become high and they increase with the number of data sets involved. Thus, even for medium-sized systems the optimization problem can become too large to handle with standard optimization software. Therefore, in a series of previous studies decomposition algorithms have been used [1,2]. Another important issue of finding the global optimum has been investigated by several authors (e.g. Esposito and Floudas [3], Gau and Stadtherr [4]), since the problem is generally non-convex.

2.1. Sequential SQP-based parameter estimation with iterative weight adjustment

In our contribution an sequential, SQP-based optimization approach for solving an EVM parameter estimation problem has been used [2]. In this approach the process model equations are decoupled from the optimization problem thus reducing the size of the optimization problem and making this approach applicable to large-scale systems with multiple data sets. In addition the weighting matrix in (1) is iteratively adjusted by introducing sensitivity information of the measured variables with respect to the optimization parameters into the weighting matrix.
In the general weighted least-squares formulation the variance-covariance matrix of the measurements is used as a weighting matrix in the objective function. Especially in nonlinear estimation problems however the sensitivity of the measured variables with respect to the optimization parameters is very important in deterministic approaches. If the sensitivity is very low, the optimization algorithm might result in a poor solution. To consider this sensitivity information in the objective function, the variance-covariance matrix of the model prediction **VC** was introduced into the weighting matrix:

$$\mathbf{W}_i = \left(\mathbf{MV}_i + \mathbf{VC}_i\right)^{-1} \tag{2}$$

With the variance-covariance matrix of the measurements **MV** and

$$\mathbf{VC}_i = \mathbf{V}_i \ \mathbf{PV}^{-1} \ \mathbf{V}_i^T \tag{3}$$

With the variance-covariance matrix of the parameters **PV** and the sensitivity matrix

$$\mathbf{V}_i = \frac{\partial \mathbf{y}_i}{\partial \mathbf{z}} \tag{4}$$

As **PV** in eq. (3) can only be calculated from historical experiments or analytically computed at convergence of the parameter estimation problem an iterative procedure has been introduced, where **PV** is calculated with first order approximations at each iteration of the SQP algorithm (see [6]).
The drawback of gradient-based approaches as SQP is that good starting values for the parameters are needed which are inside the attraction domain of the global optimum. In many nonlinear, nonconvex parameter estimation problems these starting values are not available and only local solutions can be achieved.

2.2. Simulated Annealing

Simulated annealing was developed in the 1980s by Kirkpatricket al. [5] and various other authors. SA is based on first principles of thermodynamics, it analogizes annealing processes of liquids and metals to the minimization of an objective function. If the system is cooled very slowly, it can reach thermal equilibrium on every temperature level.
Simulated annealing is considerably simpler than gradient-based approaches. Convergence can be guaranteed with suitable values for the initial annealing temperature, the cooling rate and the number of cycles performed.
However, SA has some serious drawbacks when it comes to computational efficiency. As a global solution can only be guaranteed for a small cooling rate the number of function evaluations is very large. Even if the algorithm has reached a near optimal solution, still a large number of iterations are necessary to achieve the optimal solution.

2.3. Hybrid approach

To overcome the drawback of both algorithms which is the local applicability of gradient-based methods and the low computational efficiency of Simulated Annealing a hybrid approach has been used, which couples both approaches. The SA procedure is used in the first stage to find a solution within the attraction domain of the global optimum. This solution is then used as starting values for the SQP approach. A coordinator is used to determine the switching time between the two algorithms. The switching time must be chosen as early as possible to decrease the number of function evaluations performed by the SA stage as much as possible but as late as necessary to compute starting values for the SQP stage that are in the attraction domain of the global solution. Several switching criteria have been studied in this contribution:

- The annealing temperature and the SA step size.
  If these values are small enough, the possibility of escaping a local minimum is very low and further improvement can be achieved using a gradient-based approach. Although it is very difficult to quantify the temperature and step size at which the switching should be performed, as these values are case

sensitive. If they are too small, then the improvement in the overall computational time is low, on the other hand if they are too large, the starting values provided by SA might not be in the attraction domain of the global optimum.

- The number of function evaluations performed.
  This criterion is connected to analyzing the annealing temperature but does not take the ratio of accepted function evaluations over the last Markov-chain into account. Therefore there is no information if SA is near a local (global) optimal solution. With this switching criterion the overall CPU time can be restricted, but the solution of the SA algorithm may still not be in the attraction domain of the optimal solution.
- The relative change of the objective function value over the last *n* accepted SA trials.
  This is a common termination criterion for The SA algorithm. By relaxing this criterion the algorithm is terminated at an earlier stage which reduces the computational time, but also might result in a poor solution. Again the value for this switching criterion is case sensitive and is therefore difficult to chose.
- Analysis of the confidence region at the best SA solution.
  With this criterion it is possible to analyze the quality of the SA solution. If it is near a local (global) optimum, the size of the confidence region decreases. The drawback is that it is not possible to analyze if it is a local or global solution. Therefore this criterion should be coupled with other switching criteria mentioned before.

2.4. Case study

To demonstrate the applicability of the hybrid optimization approach, it has been applied to a case study of biochemical nonlinear parameter estimation problem considered by Mendes and Kell [7]. The optimization problem consists of the estimation of 36 kinetic parameters of a nonlinear biochemical dynamic model formed by 8 ODE's that describe the variation of the metabolite concentrations with time. The problem is nonconvex and gradient-based methods could not find the global solution from any arbitrary starting vector. In a contribution by Moles et al. [8] several global optimization approaches have been tested on that problem, including evolutionary strategies that led to the best solution. To compare the results, the same conditions as in [8] have been used in this study for starting values, parameter boundaries and measurement data.

2.5. Results and discussion

Simulated Annealing was able to find a solution close enough to the global solution to enable the gradient-based SQP approach to converge. By choosing an appropriate switching criterion the number of function evaluations needed by

the SA stage could be reduced significantly. By using the iteratively adjusted weighting matrix in the objective function of the SQP approach the quality of the solution could be improved significantly. In Table 1 the average relative residuals for the Parameters $abs\left[\left(\boldsymbol{\theta}^{*}-\boldsymbol{\theta}^{true}\right)/\boldsymbol{\theta}^{true}\right]$ and the maximum relative residual $\max\left(abs\left[\left(\boldsymbol{\theta}^{*}-\boldsymbol{\theta}^{true}\right)/\boldsymbol{\theta}^{true}\right]\right)$ at the solution $\boldsymbol{\theta}^{*}$ of the parameter estimation problem for the biochemical pathway problem are presented for both cases. It can be seen that by using the sensitivity information in the objective function the quality of the estimated could be improved.

Table 1: Relative residuals at convergence

	Average residual	Maximum residual
Original weighting matrix	$8.63*10^{-3}$	0.109
Adjusted weighting matrix	$3.05*10^{-3}$	$4.49*10^{-2}$

3. Conclusions

A hybrid optimization approach has been used for estimating parameters in a nonlinear nonconvex implicit equation system. A stochastic global optimization approach, Simulated Annealing, has been used to find parameter values in the attraction domain of the global optimum. The global approach has been coupled with a gradient based local SQP optimization approach to decrease the number of function evaluations toward the end of the optimization procedure. To improve the accuracy of the parameter estimates, sensitivity information has been included into the objective function by iteratively adjusting the weighting matrix with the variance-covariance matrix of the model prediction. The algorithm was applied to a case study of biochemical nonlinear parameter estimation problem. For this case study the hybrid approach was able to achieve the global solution. By using the iteratively adjusted weighting matrix, the quality of the estimates could be improved significantly.

References

1. Tjoa, I.B.; Biegler, L.T. Comp. Chem. Eng. 1992, 16(6), 523-533.
2. Faber, R.; Li, P.; Wozny, G. Ind. Eng. Chem. Res. 2003, 42(23), 5850 - 5860.
3. Esposito, W. R.; Floudas, C. A. Ind. Eng. Chem. Res. 1998, 37, 1841-1858.
4. Gau, C.-Y.; Stadtherr, M. A. AIChE J. 2002, 48(6), 1192-1197.
5. Kirkpatrick, S., Gelatt Jr., C. D., Vecchi, M. P. (1982). IBM Research Report RC 9355.
6. Faber, R.; Arellano-Garcia, H.; Wozny, G. Chemie Ingenieur Technik (2006), 78(10), 1509-1522.
7. Mendes, P.; Kell, D.B. Bioinformatics (1998), 14, 869-883.
8. Moles, C.G.; Mendes, P.; Banga, J,R. Genome Research (2003), 13, 2467-2474.