



Reduced model-based global optimisation of largescale steady state nonlinear systems

DOI:

[10.1016/B978-0-12-818634-3.50174-0](https://doi.org/10.1016/B978-0-12-818634-3.50174-0)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Tao, M., Li, J., & Theodoropoulos, C. (2019). Reduced model-based global optimisation of largescale steady state nonlinear systems. In *Proceedings of the 29th European Symposium on Computer Aided Process Engineering* <https://doi.org/10.1016/B978-0-12-818634-3.50174-0>

Published in:

Proceedings of the 29th European Symposium on Computer Aided Process Engineering

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact openresearch@manchester.ac.uk providing relevant details, so we can investigate your claim.



Reduced model-based global optimisation of large-scale steady state nonlinear systems

Min Tao^a, Jie Li^a and Constantinos Theodoropoulos^{a*}

^a*School of Chemical Engineering and Analytical Science, The University of Manchester, M13 9PL, UK*

**k.theodoropoulos@manchester.ac.uk*

Abstract

Many engineering systems can be accurately simulated using partial differential equations (PDEs), resulting in large-scale distributed parameter systems. Deterministic global optimisation algorithms (GOP) can compute global optimal solutions offering theoretical guarantees on the global optimality. However, distributed parameter systems pose computational challenges for these optimisation methods. Model reduction techniques can produce low-order systems that are computationally amenable. In this work, a combined principal component analysis (PCA) and artificial neural networks (ANNs)-based model reduction methodology is employed for the global optimisation of large-scale distributed steady state systems. Still, the optimisation problem is hard to solve due to the high nonlinearity of activation functions in the reduced ANN structure. A novel piece-wise linear approximation reformulation is introduced to reduce the complexity of the original problem and to provide a good globally approximate solution. The performance of the proposed PCA-ANN-GOP framework is demonstrated through an illustrative example: a tubular reactor where an exothermic reaction takes place.

Keywords: Model reduction, Distributed parameter systems, Artificial neural networks, Global optimisation

1. Introduction

PDE-constrained black/grey box modelling and optimisation methodologies have wide applicability in industrial engineering areas where many finite approximation-based generic commercial simulators or input-output data from real complex systems are available (Boukouvala et al., 2017). However, many simulators solely provide input-output results of simulation process and can not perform optimisation tasks. Hence, the optimisation of real complex systems is problematic. Even if the high dimensional complex model codes are available, the cost of computing derivatives is often unacceptable. One of the most effective techniques to reduce computational costs for input/output systems is the use of *equation-free* methods (Theodoropoulos, 2011). Recently, equation-free based reduced SQP algorithms have been exploited for large-scale local optimisation with black-box steady-state simulators (Bonis and Theodoropoulos, 2012). With a small set of input-output data from the simulator, matrix-free reduced Jacobian and Hessian matrix techniques can be employed. The aim of this work is to construct deterministic global optimisation methods for large-scale input/output simulators. Such deterministic global optimisation methods are usually computationally intensive due to the repeated utilisation of branch-and-bound algorithms. Hence, in terms of computational cost, detailed models of large-scale problems are very hard or even impossible to deal with.

In this work, a data-driven methodology is employed first to create a good approximate surrogate model for large-scale steady-state systems. Principal component analysis (PCA) in combination with artificial neural networks (ANNs) is used to produce a reduced order model. The reduced model (ANN-constrained model) is then utilized by the global general-purpose optimisation solver BARON (Tawarmalani and Sahinidis, 2005). Furthermore, we present a novel piece-wise linear approximation reformulation for the nonlinear constraints in order to accelerate the computation of the global optimum for the reduced system.

2. Model Reduction

The focus of this work is on optimisation of spatially distributed processes, described by a set of steady-state dissipative PDEs with accompanying boundary and initial conditions:

$$\frac{\partial X}{\partial t} = D\left\{\frac{\partial X}{\partial y}, \frac{\partial^2 X}{\partial y^2}, \dots, \frac{\partial^n X}{\partial y^n}\right\} + R(X, P_m) \quad (1)$$

where D is a dissipative spatial differential operator, $R(X, P_m)$ is a nonlinear function, P_m are the parameter variables and X the set of state variables. If we consider the steady state analysis of these equations, we assume that $X(t, y) \rightarrow X(y)$, and we can set $\frac{\partial X}{\partial t} = 0$, giving the equations:

$$0 = D\left\{\frac{\partial X}{\partial y}, \frac{\partial^2 X}{\partial y^2}, \dots, \frac{\partial^n X}{\partial y^n}\right\} + R(X, P_m) \quad (2)$$

Due to the large number of variables in the corresponding discretised problems, the full model is usually too big for global optimisation (Houska and Chachuat, 2017). This barrier can be overcome by employing model reduction methodologies such as artificial neural networks and principal component analysis. Principal component analysis (PCA) is a powerful methodology for data compression. Common linear PCA is a transformation of the original data space spanned by the principal components (PCs). PCs are the basis vectors for the new space and are ranked by the magnitude of variance of the original data projected onto the corresponding PCs. PCs can be computed by matrix decomposition as in (Jackson, 2005):

$$Y = TP^T + E \quad (3)$$

where $Y \in R^{N \times m}$ is the high dimensional sample data from the large-scale steady state simulator given by Eq.(2), N and m , the number of sample data and data dimension, respectively. $T \in R^{N \times a}$, is the score matrix for the PCs and a ($a \leq m$), the number of PCs. $P \in R^{m \times a}$ is the loading matrix and $E \in R^{N \times m}$ the residual matrix.

In order to compute a representation for the outputs produced by the the black-box simulator, the Latin hyper-cube (LHC) sampling method is applied to collect snapshots. LHC sample method can maximize the sample information content for a limited number of samples. After snapshots collecting and PCA reduction steps applied, the high dimensional input-output data is transformed into lower dimensional PC scores. Subsequently a reduced order model is built through the use of the reduced data obtained. Here, we adopt artificial neural networks (ANNs) because standard

feed-forward artificial neural network (ANN) with one hidden layer can be an universal approximator for any smooth function (Hornik et al., 1989). The structure of the feed-forward ANN is relatively simple and easy to be constructed. In this work, the training process for ANN is performed in MATLAB R2017a with Neural Networks toolbox. The activation function is chosen to be the hyperbolic tangent function. The neural network is trained with the Levenberg-Marquardt back-propagation algorithm, which can reduce the mean square error quickly. The input is the low-dimensional input parameter matrix from the black-box simulators. The output is the reduced principal component scores of the output data from black-box simulators. At this point, data-driven PCA-ANN based surrogate model has been built that can be employed for global optimisation.

3. Model reduced global optimisation

The general nonlinear optimisation problem with embedded neural networks (reduced model) is considered. The general problem can be formulated as follows:

$$\begin{aligned} \min_{x \in R^c, y \in R^d} \quad & f(x, y) \\ \text{s.t.} \quad & h(x, y) = 0, \\ & g(x, y) \leq 0 \end{aligned} \tag{4}$$

where the objective function $f(x, y) : R^c \times R^d \rightarrow R$, $x \in R^c$ is an input vector of the black-box simulator, $y \in R^d$ is an output vector of the black-box simulator, $h(x, y) : R^c \times R^d \rightarrow R$ is ANN constraint derived from the reduced model, and $g(x, y) : R^c \times R^d \rightarrow R$ represents other constraints.

In the current work, $g(x, y)$ are the bound constraints for the state variables. The intrinsic non-linearity of the optimisation problem lies on $h(x, y)$ due to the activation function of the ANN structure. The feed-forward ANN uses the hyperbolic tangent function $\tanh(\cdot)$. The general global optimisation software BARON is employed to solve the above problem. General global solvers in GAMS such ANTIGONE and BARON can not identify $\tanh(\cdot)$, so its explicit algebraic form ($\tanh(x) = \frac{e^x - 1}{e^x + 1}$) is utilized. The basic formulation is transformed into $\tanh(x) = \frac{-2}{e^x + 1} + 1$ in order to produce a tighter under-estimator for the global solver.

However, the computational cost of global solvers increases exponentially when reducing the high non-convexity of the activation function. We propose a novel piece-wise linear approximation (PWA) reformulation for the activation function. With an accurate PWA model, a good approximation to the true global solution can be obtained. It should be, however, noted that a PWA reformulation transforms the high non-convex NLP problem into a set of MILP problems. The global solvers for both NLP and MILP problems are based on the branch and bound framework. The biggest difference is that they turn the continuous variables into auxiliary binary variables on the branch step. The advantage of the PWA reformulation lies on the use proven MILP software such as CPLEX. Here, we present an iterative way to collect piece-wise points for PWA models. $\tanh(x)$ is an odd function of central symmetry, which is concave on $[0, +\infty]$ and convex on $[-\infty, 0)$. So it is easy to compute the grid points for piece-wise approximation of $\tanh(x)$ on $[-\infty, 0)$ when the grid points on $[0, +\infty]$ are available. On the range of $[0, +\infty]$, $\tanh(x)$ experiences a 'short-time' increase and then tends to be "static" (very slow growth to its limit value). In the current work, we start from only two intervals and three grid points (original point and two static points) in this range. The linear model in a static interval can approximate the activation function accurately. For the 'initial short-time' increasing interval, we reduce the error between the multiple linear models and the nonlinear $\tanh(x)$ by iteratively narrowing the interval sizes.

The iteration procedure is performed by adding more partitioning points where the largest error is calculated. In addition, a sub-interval is acceptable if the biggest error or the number of intervals satisfies a set requirement. This iteration process efficiently produces a set of collecting points, which contribute to a tight piece-wise linear representation of the highly nonlinear activation function. Assume that there are N linear models, $N + 1$ generated grid points $x_1, x_2, \dots, x_{N+1} \in R$ and approximate function $f = \tanh(x_i)$ in the i th interval. By introducing auxiliary variables h_i and λ_i , we can generalize the formulation of a piece-wise linear approximation as follows (Floudas, 1995):

$$\begin{aligned}
f(x) &= \sum_{i=1}^{N+1} \lambda_i f(x_i) \\
x &= \sum_{i=1}^{N+1} \lambda_i x_i \\
\sum_{i=1}^{N+1} \lambda_i &= 1 \\
\lambda_1 &\leq h_1 \\
\lambda_i &\leq h_i + h_{i-1}, \forall i \in \{2, 3, \dots, N\} \\
\lambda_{N+1} &\leq h_N \\
\lambda_i &\geq 0, \forall i \in \{1, 2, \dots, N+1\} \\
\sum_{i=1}^N h_i &= 1 \\
h_i &\in \{0, 1\}^N
\end{aligned} \tag{5}$$

In the next section, this algorithm is employed to a system of steady-state PDEs after collecting a "good" set of snapshots in order to illustrate its computational capabilities.

4. Application

We first apply our methodology to an example with known global optima for illustration purposes. Then we apply our PCA-ANN-GOP framework to a large-scale steady state nonlinear PDE-based system.

4.1. Illustrative example

In this illustrative example, a two-dimensional multimodal function is chosen, obtained by translating and scaling Gaussian distributions:

$$\begin{aligned}
g_{peaks} &= 3(1-x)^2 \exp(-x^2 - (y+1)^2) - 10\left(\frac{x}{5} - x^3 - y^5\right) \exp(-x^2 - y^2) \\
&\quad - \frac{1}{3} \exp(-(x+1)^2 - y^2)
\end{aligned} \tag{6}$$

where variables $x, y \in [-3, 3]$. Multiple local solutions exist in this defined domain. The known unique global optimum lies on the point $(0.228, -1.626)$ where the function value is -6.551. We just use it to show the performance of our reduced global optimisation methodology.

Table 1: Comparative results of ANN-constrained model and PWA models

Model	Solver	Optimal value	CPU time (s)	Rel.gap
ANN-constrained model	BARON	-6.555	30294.26	0.002
PWA model1 (31 grid points)	CPLEX	-6.542	1004.71	0.002
PWA model2 (59 grid points)	CPLEX	-6.540	4190.16	0.002

After sampling and ANN learning, we obtain an ANN-constrained reduced model with 52 neurons. We do the global optimisation with reduced model and PWA models. ANN-constrained model is the surrogate model after ANN step. PWA model1 is the piece-wise approximation model for hyperbolic activation function with 31 grid points while PWA model2 is the piece-wise approximation model with 59 grid points. The number of grid points depends on the chosen error during the above iteration process. The NLP with ANN-constrained model is solved by the general global solver BARON 17.4.1/GAMS 24.8.5, and MILP with PWA model is dealt with CPLEX 12.7.1/GAMS 24.8.5 on a Desktop (Intel Core(TM) CPU 3.3 GHz, 8 GB memory, 64-bit operating system) running Windows 7. The results are given in Table 1. From Table 1, we can see that we obtain almost the same optimal solution, very close to the true global optimizer of the original problem. In addition, our proposed PWA formulation can reduce the computation time greatly with high accuracy.

4.2. Case study

The effectiveness of our PCA-ANN-GOP is illustrated using a chemical engineering application; a tubular reactor, where an exothermic reaction takes place. The mathematical formulation of the tubular reactor optimisation is given below:

$$\begin{aligned}
 & \max_{T_{wi}} C_{exit} \\
 s.t. & 0 = \frac{1}{Pe_1} \frac{\partial^2 C}{\partial y^2} - \frac{\partial C}{\partial y} + Da(1-C) \exp(T/(1+T/\gamma)) \\
 & 0 = \frac{1}{LePe_2} \frac{\partial^2 T}{\partial y^2} - \frac{1}{Le} \frac{\partial T}{\partial y} - \frac{\beta}{Le} T + BDa(1-C) \exp(T/(1+T/\gamma)) + \frac{\beta}{Le} T_w \\
 & 0 \leq T_{wi} \leq 5
 \end{aligned} \tag{7}$$

Here C and T are the dimensionless concentration and temperature respectively, while C_{exit} is dimensionless output concentration. The system parameters are $Pe_1 = 5, Pe_2 = 5, Le = 1, Da = 0.1, \beta = 1.5, \gamma = 10, B = 12$; T_w is the adiabatic wall temperature of the cooling problem with three cooling zones, expressed as following:

$$T_w(y) = \sum_{i=1}^N (H(y - y_{i-1}) - H(y - y_i)) T_{wi} \tag{8}$$

The resulting discretised 500 algebraic equations comprise our house-made simulator. After model reduction step (PCA and ANN steps), a small size ANN-constrained optimisation problem (3 input, 12 output, 20 neurons ANN) is generated, which is easy to be solved. In order to compare the computation efficiency of ANN-constrained model and PWA models, we compute the results with different numbers of neurons. All these computation cases converge to almost the same

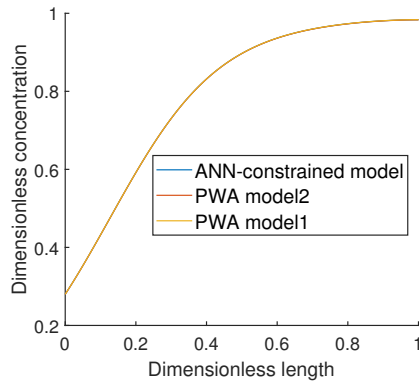


Figure 1: Solution profiles at optimum for concentration

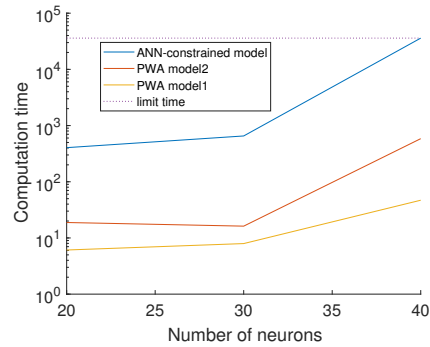


Figure 2: Computation time (sec) under different number of neurons

solution, whose objective values are extremely close to 0.999. Figure 1 gives the optimal solution profiles for concentration distribution under 30-neuron ANN structures. The three concentration distribution curves for different approximate models almost overlap. Hence, our PWA models show high computation accuracy. Figure 2 illustrates the comparison of the computation cost for the different models. The limit time (max time for computations to stop) is set to be 36000 seconds. As it can be seen, computation time increases rapidly as with increasing number of neurons. The computation cost reaches the limit time for the ANN-constrained model with only 40 neurons while the computation times are much smaller for both PWA models. This verifies the high computation efficiency of our PWA models.

5. Conclusion

We propose a reduced model-based global optimisation framework for large-scale steady state nonlinear systems. Data-driven PCA and ANN techniques are employed to produce a small or middle size surrogate model for the original system. To further reduce computation complexity, novel PWA models are utilized to replace the reduced ANN-constrained model. Illustrative examples are presented to verify the efficiency of our framework. Computation results reveal that our PCA-ANN-GOP framework can capture the global optimum with acceptable computational cost, especially when PWA approximations are used.

References

- I. Bonis, C. Theodoropoulos, 2012. Model reduction-based optimization using large-scale steady-state simulators. *Chemical engineering science* 69 (1), 69–80.
- F. Boukouvala, M. F. Hasan, C. A. Floudas, 2017. Global optimization of general constrained grey-box models: new method and its application to constrained pdes for pressure swing adsorption. *Journal of Global Optimization* 67 (1-2), 3–42.
- C. A. Floudas, 1995. *Nonlinear and mixed-integer optimization: fundamentals and applications*. Oxford University Press.
- K. Hornik, M. Stinchcombe, H. White, 1989. Multilayer feedforward networks are universal approximators. *Neural networks* 2 (5), 359–366.
- B. Houska, B. Chachuat, 2017. Global optimization in hilbert space. *Mathematical Programming*, 1–29.
- J. E. Jackson, 2005. *A user's guide to principal components*. Vol. 587. John Wiley & Sons.
- M. Tawarmalani, N. V. Sahinidis, 2005. A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming* 103 (2), 225–249.
- C. Theodoropoulos, 2011. Optimisation and linear control of large scale nonlinear systems: a review and a suite of model reduction-based techniques. In: *Coping with Complexity: Model Reduction and Data Analysis*. Springer, pp. 37–61.