Optimization Methods and Software

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/goms20

Towards global parameter estimation exploiting reduced data sets

Susanne Sass, Angelos Tsoukalas, Ian H. Bell, Dominik Bongartz, Jaromił Najman & Alexander Mitsos

To cite this article: Susanne Sass, Angelos Tsoukalas, Ian H. Bell, Dominik Bongartz, Jaromił Najman & Alexander Mitsos (2023): Towards global parameter estimation exploiting reduced data sets, Optimization Methods and Software, DOI: 10.1080/10556788.2023.2205645

To link to this article: https://doi.org/10.1080/10556788.2023.2205645

Published online: 09 May 2023.

OPTIMIZATION METHODS & SOFTWARE
https://doi.org/10.1080/10556788.2023.2205645

Taylor & Francis
Taylor & Francis Group

Towards global parameter estimation exploiting reduced data sets

Susanne Sass [a], Angelos Tsoukalas [b], Ian H. Bell [c], Dominik Bongartz [d], Jaromił Najman [a] and Alexander Mitsos [a,e,f]

[a]Process Systems Engineering (AVT.SVT), RWTH Aachen University, Aachen, Germany; [b]Department of Technology and Operations Management, RSM Erasmus University Rotterdam, Rotterdam, Netherlands; [c]Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, CO, USA; [d]Department of Chemical Engineering, KU Leuven, Leuven, Belgium; [e]JARA-CSD, Aachen, Germany; [f]Institute of Energy and Climate Research: Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, Jülich, Germany

ABSTRACT

We focus on deterministic global optimization (DGO) for nonconvex parameter estimation problems. Realistic and accurate solutions often require a fitting against very large measurement data sets, resulting in intractable size for DGO. Thus, we aim at accelerating the branch-and-bound algorithm by using reduced data sets for constructing valid bounds. We focus on fitting the equation of state for propane, which is of high interest for the chemical industry. The resulting estimation problem is a challenging nonconvex mixed-integer nonlinear optimization problem. We investigate the validity of using reduced data sets by comparing how the lower and upper bounds change when replacing the full data set with different reduced data sets. We account for regions with high and low quality fits by considering the results for the whole feasible region and 100 different subregions. Our results indicate that both regions containing solution candidates and regions containing only low quality fits can be identified based on reduced data sets. Moreover, we observe that the average CPU time per branch-and-bound iteration typically decreases if reduced data sets are used.

ARTICLE HISTORY
Received 4 April 2022
Accepted 5 April 2023

KEYWORDS
Deterministic global optimization; MINLP; parameter estimation; B&B algorithm; EOS

MATHEMATICS SUBJECT CLASSIFICATION (2010):
90C11; 90C26; 90C90; 80A23

1. Introduction

Parameter estimation aims at adjusting mathematical models to accurately describe real-world processes. However, inaccuracies in the measurement data, in the structure of the model, i.e. the equations used, or in the parameter values can result in model-data mismatch. Suboptimal parameter values as a reason for mismatch between a fixed data set and a specific model can only be excluded when using deterministic global optimization (DGO) methods [15,23]. The optimization of the resulting large-scale and potentially highly nonconvex models poses a significant challenge for existing DGO methods [4,7].

CONTACT Alexander Mitsos amitsos@alum.mit.edu Process Systems Engineering (AVT.SVT), RWTH Aachen University, 52074 Aachen, Germany

In particular, we focus on a common DGO method, namely the spatial branch-and-bound (B&B) algorithm [8,25].

The tighter the upper and lower bounds calculated in the B&B algorithm are, the more efficient the fathoming of the B&B tree becomes. Mere function evaluations or local solvers can be used for the upper bounding. The lower bounding typically applies convex relaxations of the model, e.g. obtained by the auxiliary variable method (AVM) [24–27] or McCormick relaxations [13,28]. The former relaxation technique is implemented, e.g. in the widely used commercial solver BARON [27], while we focus on the latter as implemented, e.g. in our open-source solver MAiNGO [2]. The effort for calculating good lower bounds rises with an increasing number of measurement data points considered. Similarly, the relaxations tend to get weaker as the number of functional operations in the objective rises, e.g. by taking a square or adding a term. Thus, we aim at using a reduced data set for the lower bounding procedure of the complete model, i.e. the model incorporating all available data points. When choosing the objective function such that lower bounds calculated based on reduced data sets are valid for the full data set, the proposed approach remains a DGO method. Note that reducing the data set affects the upper bounding (*i*) if the reduced data set is used to accelerate the local optimization and (*ii*) if the optimum point of the lower bounding problem is used as a start value for the upper bounding solver as done in MAiNGO. Note further that reducing the computational effort by using a sequence of simpler problems is an established approach in other scientific areas, e.g. in the context of tree search [18] or machine learning [e.g. [16]]. It is well known that data reduction allows for more sophisticated fitting techniques, for example, using nonlinear models rather than linear regression [10]. Similarly, Rulliére et al. [19] start with a small data set and include a sampling step in their B&B approach for a stochastic optimization problem.

We investigate the following hypotheses on the solution of a (specific) parameter estimation problem:

(H1): Both the final upper and lower bound do not change drastically when reducing the size of the data set.
(H2): The reduction of the data set allows for a faster global solution of the parameter estimation problem.

A B&B algorithm for minimization problems successively approaches the global solution by cutting off 'bad regions', which only contain parameter tuples yielding high objective values, and digging deeper into 'good regions', which contain parameter tuples giving low objective values. Thus, Hypothesis (H1) comprises (*i*) detecting good regions to avoid cutting off candidates for the global optimum of the complete problem, and (*ii*) detecting bad regions to early cut off unfavourable parts of the B&B tree based on reduced data sets. Still, our proposed approach can only save CPU time if the data reduction leads to a significant reduction of the computational effort, see Hypothesis (H2).

Our case study focuses on an equation of state (EOS) of a fluid, which is one of the fundamental models in chemical engineering. More precisely, we use the EOS of propane presented in Lemmon et al. [11]. For the verification of their model, they refer to about 200 data sources containing more than 10, 000 data points in total. As fitting this nonconvex model poses a great challenge, Lemmon et al. use a drastically reduced data set and a sophisticated fitting procedure based on nonlinear programming. Moreover, they manually adjust terms in the model formulation to get a more concise model. We, in turn, solve a predefined estimation problem with a B&B algorithm (*i*) to provide a measure on how

good the solution can get and (*ii*) to possibly find good solutions faster than by local or heuristic methods due to the iterative narrowing of the feasible region. As a first step, we consider a simplified problem statement where we reduced the number of free parameters from 162 to 9 including 1 binary and 2 integer variables. Aside from that, we simplify the problem further from the perspective of the DGO solver by using a reduced-space formulation wherein the optimizer only sees the unknown parameters as variables and does not see explicit equations [1,14]. As such, we have a solely box-constrained mixed-integer nonlinear optimization problem in 9 optimization variables. In contrast to MAiNGO, many common DGO solvers can not handle reduced-space formulations at all or at least not efficiently. In a full-space formulation of the EOS model discussed herein, at least one equality constraint and one auxiliary variable per data point would be required. Consequently, we expect an even more pronounced time gain when using a full-space formulation or an AVM-based solver like BARON. We investigate the DGO results for the whole feasible region and 100 subregions to check the hypotheses both in good and bad regions. Thus, we significantly extend our preliminary results presented at EUROPT 2021 where we only studied 6 subregions [20].

The remainder of this article is structured as follows. We give the numerical setup in Section 2 and the model for the case study in Section 3. In Section 4, we show and discuss DGO results of different regions in dependence of differently sized data sets for a parameter estimation problem based on the EOS of propane. Section 5 concludes the presented study. Additional numerical results are given in the appendix.

2. Setup

In the following, we call the parameter estimation problem using all data points *base scenario* and the problem formulations derived by reducing the number of data points to be fitted *sample scenarios*. The measurement data set is represented by $\mathcal{D}$. When dividing the feasible region into subregions, we use closed intervals for the continuous-valued unknown parameters and, in analogy, subsequent values for the discrete unknown parameters. Thus, we call subregions *parameter boxes*.

We run a complete DGO of the base and different sample scenarios over the whole feasible region and each of the subregions such that we obtain meaningful lower and upper bounds. For the optimization, we use our open-source DGO solver MAiNGO version 0.5.0.1 with an absolute and relative optimality tolerance of 10^{-2} as well as a CPU time limit of 2 h. In preprocessing, we run 100 local searches with LBFGS [12,17] implemented in the NLOPT toolbox v2.5.0 [9]. MAiNGO invokes CLP version 1.17.0 [5] for solving the lower bounding problem, and LFBGS with up to 10 iterations for solving the upper bounding problem. The relaxations are obtained from MC++ [3]. We perform all calculations on an Intel(R) Core(TM) i5-3570 processor with 3.4 GHz.

3. Equation of state of propane

We use DGO to estimate 9 out of 162 parameters of the EOS of propane based on the model given in Lemmon et al. [11]. In particular, we minimize the relative mean squared error of the predicted pressure. Our synthetic measurement set $\mathcal{D} = \{(p_j^{\mathrm{EOS}}, \delta_j^{\mathrm{EOS}}, \tau_j^{\mathrm{EOS}}), j = 1, \ldots, 262\}$ contains triplets of pressure values p, scaled density values δ, and scaled

temperature values τ which we generated with the model including the final parameter values of [11]. We obtain

$$\begin{aligned}
\min_{\substack{n_{10},t_{10},d_{10},c_{10},l_{10},\\ \eta_{10},\beta_{10},\gamma_{10},\varepsilon_{10}}} \quad & \frac{1}{|\mathcal{D}|}\sum_{j=1}^{|\mathcal{D}|}\left(\frac{p_j^{\mathrm{EOS}} - p(\delta=\delta_j^{\mathrm{EOS}},\tau=\tau_j^{\mathrm{EOS}})}{p_j^{\mathrm{EOS}}}\right)^2 \\
\text{s.t.} \quad & p(\delta,\tau) = \rho RT\cdot(1+\delta\cdot D_{\alpha^r}) \\
& D_{\alpha^r} = \sum_{i=1}^{18} n_i\tau^{t_i}\delta^{d_i-1}\cdot(d_i - c_i l_i\delta^{l_i} - 2\eta_i\delta^2 + 2\eta_i\varepsilon_i\delta) \\
& \qquad \cdot\exp\left(-c_i\delta^{l_i} - \eta_i(\delta-\varepsilon_i)^2 - \beta_i(\tau-\gamma_i)^2\right) \\
& \rho = \rho_{\mathrm{c}}\cdot\delta \\
& T = \frac{T_{\mathrm{c}}}{\tau}
\end{aligned} \tag{1}$$

with constants $\rho_{\mathrm{c}} = 5000\,\mathrm{mol\,m^{-3}}$, $T_{\mathrm{c}} = 369.89\,\mathrm{K}$, and $R = 8.314472\,\mathrm{J\,mol^{-1}\,K^{-1}}$, constant parameters n_i, t_i, d_i, c_i, l_i, η_i, β_i, γ_i, and ε_i fixed to the values given in Table 4 of [11] for any $i \in \{1,\ldots,9,11,\ldots,18\}$ as well as parameter bounds

$$\begin{aligned}
& n_{10} \in [-3,2], \quad t_{10} \in [10^{-5},22], \quad d_{10} \in \{1,\ldots,10\}, \\
& c_{10} \in \{0,1\}, \quad l_{10} \in \{1,2\}, \quad \eta_{10} \in [0,20], \\
& \beta_{10} \in [0,550], \quad \gamma_{10} \in [0,2], \quad \varepsilon_{10} \in [0,1.5]
\end{aligned}$$

based on parameter values for different fluids reported in [11,22,29]. Note that we use symbol D_{α^r} for the derivative of the residual Helmholtz energy with respect to scaled density δ and allow for a Gaussian term in term number 10 via $c_{10} \in \{0,1\}$. Further, we use the value reported in [11] for gas constant R rather than the standard value of the international system of units (SI) to stay consistent with the model including the parameter values given in [11]. The choice to optimize the parameters of addend $i = 10$ is a compromise between the excessively long runtimes for optimizing the shape of the first addends, which determine the general properties of the model predictions, and the excessively short runtimes for optimizing the shape of the last addends, which yield a fine tuning. Note again that we use a reduced-space formulation, meaning that all model equations are inserted into each other resulting in a solely box-constrained mixed-integer nonlinear optimization problem with a nonconvex objective function.

We pseudo randomly pick 15 data samples from the base measurement data set to get 15 sample scenarios. In particular, in Scenarios 100_1 to 100_5 we fit the parameters based on 100 data points, in Scenarios 50_1 to 50_5 based on 50 data points, and in Scenarios 10_1 to 10_5 based on 10 data points, see Figure 1.

The required files for solving Model (1) with MAiNGO and the specific measurement data points are provided online [21]. For interested readers, we provide the model statement based on the base scenario in the more common modelling language GAMS [6] as well.

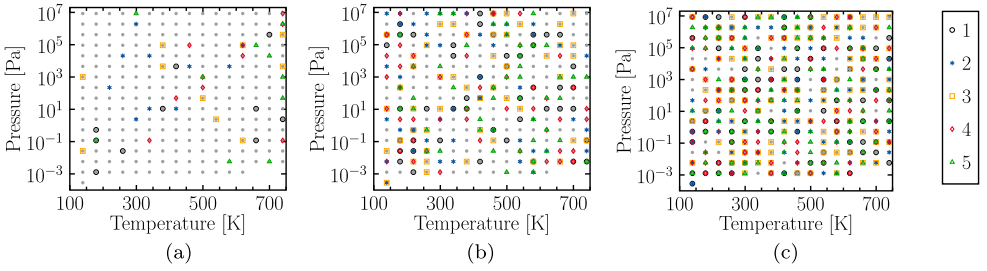

Figure 1. Pressure-temperature-pairs used as measurement data in the sample scenarios (see legend) and base scenario with a logarithmic y-axis. (a) Scenario 10_1 to 10_5. (b) Scenario 50_1 to 50_5. (c) Scenario 100_1 to 100_5.

4. Results and discussion

For investigating the impact of data reduction on the optimal solution, we study DGO results over the whole feasible set of the selected 9 parameters and 100 different parameter boxes. The different parameter boxes represent different nodes of a potential B&B tree. By (binary) branching, the feasible intervals of the parameters are subsequently split into halves until nodes containing solution candidates (good regions) and nodes which can not contain the optimal solution (bad region) are identified. Therefore, we identified Boxes 1, 2, and 3 which contain parameter values resulting in up to high quality fits (good region), up to medium quality fits, and only low quality fits (bad region), respectively, by function evaluations. To also cover larger regions, Boxes 4 to 50 are systematic combinations of subregions of the feasible set of different unknown parameters, e.g. we use bounds $c_{10} \in \{0\}$ in Boxes 4 and 5 as well as $c_{10} \in \{1\}$ in Boxes 6 and 7 in combination with bounds $l_{10} \in \{1\}$ in Boxes 4 and 6 as well as $l_{10} \in \{2\}$ in Boxes 5 and 7. For the exact bounds of all parameters, please refer to the supplementary material [21]. Since we can not foresee the actual B&B tree, we added Boxes 51 to 100 which contain pseudo randomly generated parameter bounds. For simplification of notation, the name 'Box 0' will refer to the whole feasible range in the following.

By computing lower and upper bounds for the boxes via DGO, which are as tight as possible given our CPU time limit, we aim at an informative representation of possible nodes of a B&B tree and the corresponding pruning behaviour of the B&B algorithm. Analogously, we do not investigate the optimal parameter values of the different boxes as these do not affect the pruning behaviour.

Before running DGO, we evaluated the objective function for all 15 sample scenarios and the base scenarios at 6 different points covering regions with high and low objective values. When using too few data points, namely Scenarios 10_1 to 10_5 in this case study, sample scenarios may give objective function values which are orders of magnitude smaller than the final lower bound of the base scenario [cf., 21]. We therefore run DGO only for Scenarios 50_1 to 50_5 and 100_1 to 100_5. Hereby, we use the last results reported rather than the final results for Scenario 50_5 in Boxes 32 and 33 as well as the base scenario in Box 34 as the DGO within MAiNGO aborts unexpectedly for these 3 out of $(100 + 1) \cdot 10 = 1100$ cases.

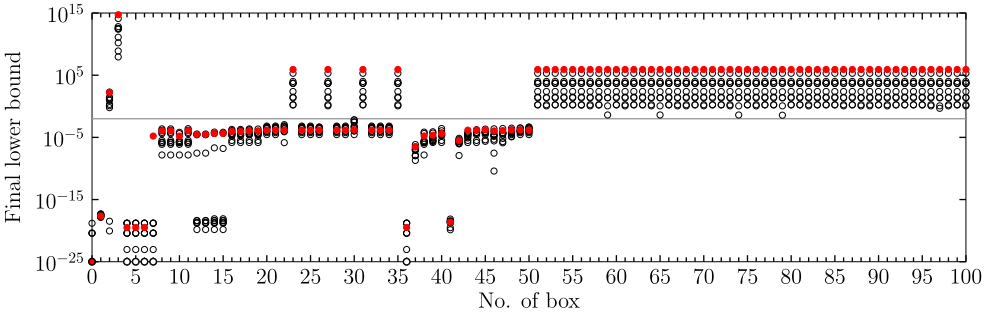

Figure 2. Final lower bound obtained by DGO of Model (1) for the base scenario (•) as well as sample scenarios 50_1 to 50_5 and 100_1 to 100_5 (o) when optimizing over different parameter boxes, where Box 0 is the whole feasible region, with an optimality tolerance of 0.01 (–); marks on the x-axis are equal to 0.

At first, we focus on Hypothesis (H1). Let us recall that we aim for an algorithm using a lower bound calculated based on a reduced data set for solving the complete model. Transferred to the study presented herein, this means to compare the lower bound of a sample scenario with the upper bound of the base scenario for deciding whether to prune the respective node. In this study, we are interested in the order of magnitude of the bounds on the optimum solution rather than the exact values, since we want to investigate whether the data reduction changes the pruning behaviour drastically.

Overall, the order of magnitude for the final upper and lower bounds is consistent, i.e. close to the optimal solution value 0 or significantly larger, for the base scenario and Scenarios 50_1 to 50_5, 100_1 to 100_5 within each of the parameter boxes, cf. Figures 2 and 3 as well as Table A2 in the appendix. Note that values smaller than the machine precision of 10^{-16} are numerically equal to 0 and correspond therefore to a perfect fit. In cases with a significant absolute deviation between the lower bounds of a sample scenario and the corresponding base scenario, the sample scenarios always yield smaller final lower bounds. In these cases, the current upper bound of the B&B tree, i.e. the best solution found so far, may lie between the lower bound of the base scenario and the lower bound of a sample scenario. Hence, data reduction may prevent pruning of such a node and, consequently, lead to a larger B&B tree. To investigate this in more detail, we assume a B&B tree, where we have the whole feasible range, i.e. Box 0, in the root node and one of the Boxes 1 to 100 in the active child node. Only if Box 2 were the active node, a child node would be pruned when using the base scenario but kept when using 2 out of the 10 sample scenarios, namely Scenarios 50_3 or 50_5. Note that Box 2 is right at the boundary between pruning and keeping, since it is deliberatily picked to contain only medium quality fits. For all other boxes, the decision whether or not to prune does not change when replacing the base scenario with any of the sample scenarios.

The reverse is seen in many of the cases where the final lower bound is smaller than the optimality tolerance, i.e. the solution is a(n) (almost) perfect fit: some of the sample scenarios give larger lower bounds than the base scenario. However, we can replace the mean squared error in the objective function by the sum of squared errors, which corresponds to scaling the objective presented in this article by the size of the data set. In this case, we add more error terms when augmenting the sampled data set until the full data set is reached.

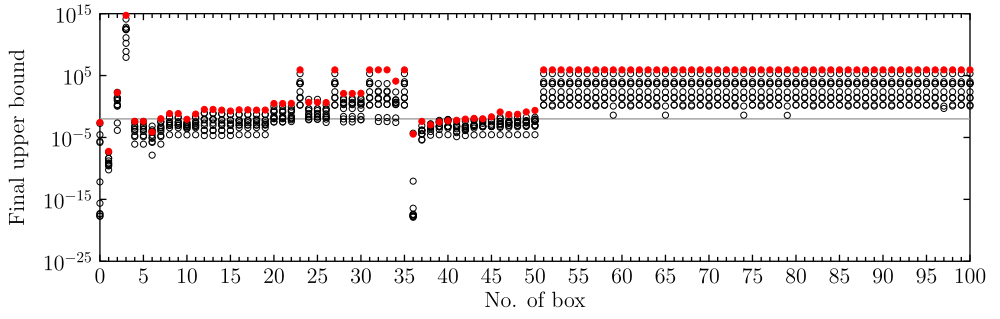

Figure 3. Final upper bound obtained by DGO of Model (1) for the base scenario (•) as well as sample scenarios 50_1 to 50_5 and 100_1 to 100_5 (o) when optimizing over different parameter boxes, where Box 0 is the whole feasible region, with an optimality tolerance of 0.01 (–).

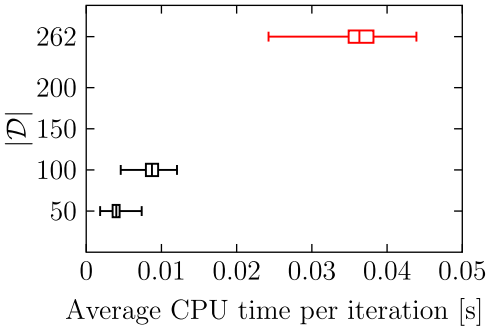

Figure 4. Box-Whisker-plot of the average CPU time per B&B iteration in dependence of the size of the data set when solving Model (1) for the base scenario (_) as well as sample scenarios 50_1 to 50_5 and 100_1 to 100_5 (–) over 100 different parameter boxes and the whole feasible region.

Due to the nonnegativity of a squared function, the lower bound of each sample scenario will therefore be a lower bound on the base scenario.

Finally, we check Hypothesis (H2) regarding the reduction of the CPU time by data reduction. We aim at developing an extension of the B&B algorithm which applies different reduced data sets in different B&B nodes, starting with very small data sets in the first nodes and, finally, approaching the full data set at the end of the B&B algorithm. Thus, we investigate the CPU time per B&B iteration rather than the CPU time of the complete DGO, whereby B&B iteration means processing one node including, beneath other steps, lower and upper bounding. Figure 4 shows that we can typically expect a CPU time reduction for the B&B iterations when using a reduced data set. In fact, the scenarios with 100 data points are on average 4 times faster and the scenarios with 50 data points are on average 9 times faster than the base scenario in the respective box. However, single scenarios may be exceptions of this rule. For example, the quickest case with $|\mathcal{D}| = 100$, namely Scenario 100_2 in Box 61, is faster than the slowest 18% cases with $|\mathcal{D}| = 50$.

Note that the results for the average CPU time per iteration cannot be easily transferred to the total CPU time of the DGO as the data reduction changes the optimization problem. Although the reduced data set may allow for a better fit, this solution may be harder to find for the upper bounding solver. At the same time, the convergence of the lower bound may

take more time for a reduced data set if the DGO solver needs the lower bound to approach the global solution rather than a potentially better, i.e. lower, upper bound obtained by another node with the full or another reduced data set. Nevertheless, all sample scenarios are faster than the base scenario when optimizing over the whole feasible range, and almost 98% of the sample scenarios are faster in Boxes 1–100, see Table A1 in the appendix. Note that the vast majority of both the base and the sample scenarios converges within 1 minute, while almost all of the remaining cases hit the CPU time limit, see Figure A1 in the appendix. Since the EOS model studied in this case study requires a very low optimality tolerance, the computational performance of the complete DGO of the scenarios depends strongly on the solutions found in the pre-processing, the choice of linearization points and the actual heuristic solutions found within the box.

5. Conclusion

We investigated whether a reduction of the measurement data set can make the deterministic global optimization (DGO) of parameter estimation problems tractable without distorting the solution. For this, we fitted the equation of state for propane described by [11] based on random data scenarios with 10, 50, and 100 data points as well as the full data set with 262 data points. In particular, we confirmed that

(H1): the order of magnitude of both objective function values and lower bounds is preserved
(H2): the average CPU time per iteration can be expected to decrease

when reducing the size of the data set.

Regarding Hypothesis (H1), good and bad regions of the base scenario can be distinguished well by sample scenarios with a sufficient number of data points, in particular, $|\mathcal{D}| \in \{50, 100\}$ in this case study. In general, data reduction may allow for lower objective values and, thus, blow up the B&B tree by keeping more nodes than required. However, the order of magnitude of the lower bound in different subregions of the feasible set is identified correctly by the sample scenarios, which already allows for pruning nodes containing only low quality fits. On the other hand, some of the sample scenarios give slightly larger lower bounds than the corresponding base scenario in good regions, which can result in pruning the global solution in the worst case. For an algorithm exploiting reduced data sets, we therefore propose data augmentation when approaching good regions and modifying the objective function such that valid lower bounds on the base scenario can be obtained even when using reduced data sets. Note again that the reduced data set should represent the complete data set adequately as solving a parameter estimation problem often means to balance model prediction errors from different regions. This means, for example, considering measurements in both liquid and gas phase for fitting an EOS. Otherwise, data reduction may distort the optimal solution when using realistic error-prone measurement data instead of noisefree data as we did in this preliminary study. Note that our positive results regarding Hypothesis (H1) indicate that the presented samples preserve the properties of the full data set sufficiently, even though they are chosen randomly.

Hypothesis (H2) is confirmed for the average CPU time per B&B iteration, while the conclusion for the total CPU time of the DGO is more diverse. The DGO of most of the sample scenarios converges faster than DGO of the corresponding base scenario, while the optimization of some sample scenarios takes as long as or even longer than the base

scenario in some of the subregions of the feasible set. In this case study, the overall speed of convergence highly depends on when and where the local solver succeeds in finding a good, i.e. a small, upper bound. Approaches for improving the upper bounding comprise the comparison of different local solvers, as easily possibly within MAiNGO, using multistart in pre-processing, as done in this case study, and initializing the local searches with good solutions found in other nodes. It remains for future studies whether narrowing the feasible region by the integration of constraints, e.g. for guaranteeing thermodynamic stability, is also beneficial or even counterproductive.

Since (*i*) the sample scenarios can identify good and bad regions and (*ii*) the average CPU time per B&B iteration tends to decrease when using reduced data sets, we expect an acceleration of the B&B algorithm when replacing the full data set with reduced data sets in some of the B&B nodes. In our case study, the B&B tree could process 9 or 4 times the number of iterations within the same CPU time limit when reducing the data set from 262 to 50 or 100 data points. Since the solution of complex models often takes multiple thousand iterations, fathoming the B&B tree may be sped up even if Hypotheses (H1) and (H2) fail in a small number of nodes. Moreover, the augmentation step will eventually close the gap between sample scenarios and base scenario. Consequently, replacing the base scenario by sample scenarios at the beginning of the B&B algorithm and successively augmenting data until we reach the full data set at the end appears promising for making the DGO of large-scale parameter estimation problems tractable.

In future work, we will therefore implement the proposed algorithm. A major challenge will be to find out when to augment the data set to achieve good convergence properties. Finally, it would be interesting to compare the performance of random and carefully selected reduced data sets for different case studies.

Acknowledgments

Susanne Sass is grateful for her association to the International Research Training Group (DFG) IRTG-2379 'Modern Inverse Problems'. We thank the two anonymous reviewers whose comments helped to improve and clarify this article. Commercial equipment, instruments, or materials are identified only in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose. Conceptualization: Alexander Mitsos, Susanne Sass, Angelos Tsoukalas; Methodology: Ian H. Bell, Dominik Bongartz, Alexander Mitsos, Jaromił Najman, Susanne Sass, Angelos Tsoukalas; Formal analysis and investigation: Susanne Sass; Writing—original draft preparation: Susanne Sass; Writing—review and editing: Susanne Sass; Funding acquisition: Alexander Mitsos, Susanne Sass, Angelos Tsoukalas; Resources: Ian H. Bell, Alexander Mitsos; Supervision: Alexander Mitsos.

Data availability statement

The data that support the findings of this study are openly available via RWTH publications at https://doi.org/10.18154/RWTH-2022-03252 [cf. [21]].The open-source version of MAiNGO [2] is available at https://git.rwth-aachen.de/avt-svt/public/maingo [accessed Apr 21, 2023].

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant MI 1851/10-1 'Parameter estimation with (almost) deterministic global optimization'.

ORCID

Susanne Sass http://orcid.org/0000-0001-9556-1721
Angelos Tsoukalas http://orcid.org/0000-0003-2679-3953
Ian H. Bell http://orcid.org/0000-0003-1091-9080
Dominik Bongartz http://orcid.org/0000-0003-1790-0235
Jaromił Najman http://orcid.org/0000-0001-7960-8542
Alexander Mitsos http://orcid.org/0000-0003-0335-6566

References

[1] D. Bongartz and A. Mitsos, *Deterministic global optimization of process flowsheets in a reduced space using McCormick relaxations*, J. Global Optim. 69 (2017), pp. 761–796. https://doi.org/10.1007/s10898-017-0547-4
[2] D. Bongartz, J. Najman, S. Sass, and A. Mitsos, *MAiNGO – McCormick-based algorithm for mixed-integer nonlinear global optimization*, Tech. Rep., Process Systems Engineering (AVT.SVT), RWTH Aachen University, 2018. http://permalink.avt.rwth-aachen.de/?id = 729717. Accessed Apr 21, 2023.
[3] B. Chachuat, B. Houska, R. Paulen, N. Perić, J. Rajyaguru, and M.E. Villanueva, *Set-theoretic approaches in analysis, estimation and control of nonlinear systems*, IFAC-PapersOnLine 48 (2015), pp. 981–995. https://doi.org/10.1016/j.ifacol.2015.09.097
[4] C.A. Floudas, *Deterministic Global Optimization: Theory, Methods and Applications*, Vol. 37, Springer US, New York, 2000. https://doi.org/10.1007/978-1-4757-4949-6
[5] J. Forrest, D. de La Nuez, and R. Lougee-Heimer, *CLP User Guide*, 2004.
[6] GAMS Development Corporation, *General Algebraic Modeling System (GAMS) Release 33.1.0*, Bussieck, Michael and Nelißen, Franz, Fairfax, VA, USA, 2021.
[7] R. Horst and P.M. Pardalos, *Handbook of Global Optimization*, Vol. 2, Springer US, New York, 1995. https://doi.org/10.1007/978-1-4615-2025-2
[8] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, 3rd revised and enlarged edition ed., Springer Berlin Heidelberg, Berlin, Heidelberg and s.l., 1996. https://doi.org/10.1007/978-3-662-03199-5
[9] S.G. Johnson, *The NLopt nonlinear-optimization package*, 2007. http://github.com/stevengj/nlopt Accessed Dec 22, 2020.
[10] E.W. Lemmon and R.T. Jacobsen, *A new functional form and new fitting techniques for equations of state with application to Pentafluoroethane (HFC-125)*, J. Phys. Chem. Ref. Data 34 (2005), pp. 69–108. https://doi.org/10.1063/1.1797813
[11] E.W. Lemmon, M.O. McLinden, and W. Wagner, *Thermodynamic properties of propane. III. A reference equation of state for temperatures from the melting line to 650 K and pressures up to 1000 MPa*, J. Chem. Eng. Data 54 (2009), pp. 3141–3180. https://doi.org/10.1021/je900217v
[12] D.C. Liu and J. Nocedal, *On the limited memory BFGS method for large scale optimization*, Math. Program. 45 (1989), pp. 503–528. https://doi.org/10.1007/BF01589116
[13] G.P. McCormick, *Computability of global solutions to factorable nonconvex programs: Part I* – Convex underestimating problems, Math. Program. 10 (1976), pp. 147–175. https://doi.org/10.1007/BF01580665
[14] A. Mitsos, B. Chachuat, and P.I. Barton, *McCormick-based relaxations of algorithms*, SIAM J. Optim.20 (2009), pp. 573–601. https://doi.org/10.1137/080717341
[15] A. Mitsos, B. Chachuat, and P.I. Barton, *Towards global bilevel dynamic optimization*, J. Global Optim. 45 (2009), pp. 63–93. https://doi.org/10.1007/s10898-008-9395-6

[16] K.P. Murphy, *Probabilistic Machine Learning: An Introduction*, Adaptive Computation and Machine Learning, The MIT Press, Cambridge, MA, 2022.
[17] J. Nocedal, *Updating quasi-Newton matrices with limited storage*, Math. Comput. 35 (1980), pp. 773–782. https://doi.org/10.2307/2006193
[18] A. Reinefeld and T.A. Marsland, *Enhanced iterative-deepening search*, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1994), pp. 701–710. https://doi.org/10.1109/34.297950
[19] D. Rulliere, A. Faleh, F. Planchet, and W. Youssef, *Exploring or reducing noise? A global optimization algorithm in the presence of noise*, Struct. Multidiscip. Optim. 47 (2013), pp. 921–936. https://doi.org/10.1007/s00158-012-0874-5
[20] S. Sass, A. Tsoukalas, I.H. Bell, D. Bongartz, J. Najman, and A. Mitsos, *Towards accelerating global parameter estimation with reduced data sets*, 2021. Presentation at EUROPT 2021, 18th Workshop on Advances in Continuous Optimization, Toulouse (virtual), 7–9 July 2021.
[21] S. Sass, A. Tsoukalas, I.H. Bell, D. Bongartz, J. Najman, and A. Mitsos, *Dataset From: Towards Global Parameter Estimation Exploiting Reduced Data Sets*, RWTH Publications, Aachen, Germany, 2022. https://doi.org/10.18154/RWTH-2022-03252.
[22] U. Setzmann and W. Wagner, *A new equation of state and tables of thermodynamic properties for Methane covering the range from the melting line to 625 K at pressures up to 100 MPa*, J. Phys. Chem. Ref. Data 20 (1991), pp. 1061–1155. https://doi.org/10.1063/1.555898
[23] A.B. Singer and P.I. Barton, *Global optimization with nonlinear ordinary differential equations*, J. Global Optim. 34 (2006), pp. 159–190. https://doi.org/10.1007/s10898-005-7074-4
[24] E.M. Smith and C.C. Pantelides, *Global optimisation of nonconvex MINLPs*, Comput. Chem. Eng.21 (1997), pp. S791–S796. https://doi.org/10.1016/S0098-1354(97)87599-0
[25] M. Tawarmalani and N.V. Sahinidis, *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*, Vol. 65, Springer US, New York, 2002. https://doi.org/10.1007/978-1-4757-3532-1
[26] M. Tawarmalani and N.V. Sahinidis, *Global optimization of mixed-integer nonlinear programs: A theoretical and computational study*, Math. Program. 99 (2004), pp. 563–591. https://doi.org/10.1007/s10107-003-0467-6
[27] M. Tawarmalani and N.V. Sahinidis, *A polyhedral branch-and-cut approach to global optimization*, Math. Program. 103 (2005), pp. 225–249. https://doi.org/10.1007/s10107-005-0581-8
[28] A. Tsoukalas and A. Mitsos, *Multivariate McCormick relaxations*, J. Global Optim. 59 (2014), pp. 633–662. https://doi.org/10.1007/s10898-014-0176-0
[29] W. Wagner and A. Pruß, *The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use*, J. Phys. Chem. Ref. Data 31 (2002), pp. 387–535. https://doi.org/10.1063/1.1461829

Appendix. Additional numerical results

This appendix contains more details on the computational performance of the different DGO runs. Note again that we use the last values reported for lower bound, upper bound, and CPU time as the final results for the 3 out of 1100 runs for which the DGO within MAiNGO aborts unexpectedly, cf. Section 4.

The total CPU times for each optimization run are depicted in Figure A1. All of the scenarios optimized over the whole feasible region and about 83% of the runs over boxes 1 to 100 converge within a minute, see also Table A1. The most of the remaining runs hit the CPU time limit, while only 19 runs in total have a runtime of multiple minutes up to 2 h. Even though, we can see the tendency that the base scenario takes longer than the sample scenarios. For further investigations, we define CPU time and data ratio

$$r^{\mathrm{CPU}} := \frac{\text{CPU time of sample scenario}}{\text{CPU time of base scenario}} \quad \text{and} \quad r^{\mathrm{data}} := \frac{|\,\mathcal{D}_{\mathrm{sample}}\,|}{|\,\mathcal{D}_{\mathrm{base}}\,|},$$

respectively, for evaluating the time gain for a sample scenario. In case of equality $r^{\mathrm{CPU}} = r^{\mathrm{data}}$, the CPU time increases linearly in the number of data points. If $r^{\mathrm{CPU}} \ll r^{\mathrm{data}}$, we can solve comparatively many samples while still reducing the overall CPU time. Table A1 shows that the total CPU

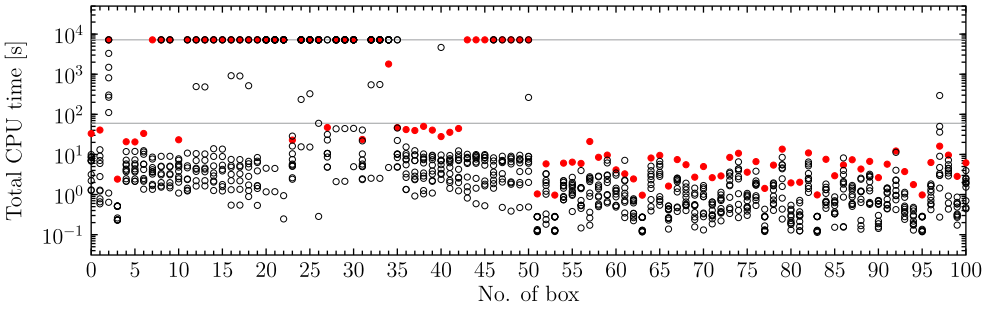

Figure A1. Total CPU time for DGO of Model (1) for the base scenario (●) as well as sample scenarios 50_1 to 50_5 and 100_1 to 100_5 (o) when optimizing over different parameter boxes, where Box 0 is the whole feasible region, with lines (–) indicating the CPU time limit of 2 h (at the top) and 60 s (at the bottom).

Table A1. Computational performance of sample scenarios 50_1 to 50_5 and 100_1 to 100_5 in comparison to the base scenario when running DGO over the whole feasible set and the parameter boxes.

Scenario	r^{data}	Median of r^{CPU}	Total no. scenarios	No. of scenarios where CPU time is/yields				
				< 60 s	$= 2$ h	$<$ CPU base[1]	$r^{\text{CPU}} < r^{\text{data}}$	$r^{\text{CPU}} < \frac{1}{2} r^{\text{data}}$
Whole feasible region								
Base	1	1	1	1	0	–[2]	–[2]	–[2]
100_1 to 100_5	0.38	0.25	5	5	0	5	5	0
50_1 to 50_5	0.19	0.07	5	5	0	5	4	4
Boxes 1 to 100								
Base	1	1	100	67	32	–[2]	–[2]	–[2]
100_1 to 100_5	0.38	0.27	500	404	91	491	354	138
50_1 to 50_5	0.19	0.09	500	447	40	488	413	263

[1] CPU time required for optimizing the base scenario over the same parameter box
[2] False by definition for the base scenario

time of 76.7% of the sample scenarios in the 100 boxes is reduced even more than the size of the data set, i.e. $r^{\text{CPU}} < r^{\text{data}}$. The same holds for 9 out of the 10 sample scenarios when optimizing over the whole feasible range. For the majority of optimization runs of Scenarios 50_1 to 50_5, the CPU time ratio is even twice as small as the data ratio.

Additionally, we provide an analysis of the order of magnitude of the final lower and upper bounds in Table A2. We cluster the boxes into two categories: the *promising boxes* where the base scenarios may contain a perfect fit according to the final lower bound, and the *inauspicious boxes* where the base scenario can only contain low quality fits as proven by a large final lower bound. The final lower bound in promising boxes is for all sample scenarios also smaller than the optimality tolerance. In contrast to that, there are sample scenarios indicating an inauspicious box as promising, see the minimum value of 8.93×10^{-21}. However, the median values over the sample scenarios are consistent with the categories, i.e. we have a small final lower bound for promising boxes and a large final lower bound for inauspicious boxes. Similarly, we can find large final upper bound values in promising boxes, indicating non-converged optimization runs, and—for some sample scenarios—small final upper bounds in inauspicious boxes. But again, the median values of the final upper bounds of the sample scenarios are consistent to the categories defined via the base scenario.

Table A2. Order of magnitude of final lower and upper bound for the base scenario and sample scenarios 50_1 to 50_5 and 100_1 to 100_5 indicated by the minimum value, maximum value, and median when dividing the parameter boxes into two categories.

	Base scenario			Sample scenarios		
	min	max	median	min	max	median
Final lower bound						
Promising boxes[1]	0.00	1.71×10^{-4}	1.12×10^{-4}	0.00	6.68×10^{-3}	2.77×10^{-5}
Inauspicious boxes[2]	1.66×10^{2}	5.38×10^{14}	8.30×10^{5}	8.93×10^{-21}	1.31×10^{14}	2.61×10^{3}
Final upper bound						
Promising boxes[1]	5.67×10^{-8}	8.30×10^{5}	1.31×10^{-1}	1.27×10^{-18}	5.01×10^{3}	2.08×10^{-3}
Inauspicious boxes[2]	1.72×10^{2}	5.38×10^{14}	8.30×10^{5}	1.39×10^{4}	1.31×10^{14}	2.61×10^{3}

[1] Set of all boxes with base LB $<$ 0.01 $=$ optimality tolerance, i.e. the boxes which potentially contain a perfect fit
[2] Set of all boxes with base LB $\geq$ 0.01 $=$ optimality tolerance, i.e. the boxes which can not contain a perfect fit