Rebuttal to Comments on "Global Optimization for the Parameter Estimation of Differential-Algebraic Systems"

William R. Esposito and Christodoulos A. Floudas

Department of Chemical Engineering, Princeton University, Princeton NJ 08544

Sir: Based on the comments of R. Luus, it is apparent that he has not fully understood what the paper by Esposito and Floudas[1] was trying to convey. In Esposito and Floudas[1], two novel deterministic global optimization methods are introduced for the parameter estimation of models that involve differential-algebraic systems. Under certain conditions these two new methods offer a theoretical guarantee of attaining an ϵ-global solution. This is the primary focus of Esposito and Floudas[1], and it represents the first rigorous contribution in the open literature that addresses such a large class of problems. We then explore the computational and theoretical aspects of our two proposed deterministic global optimization approaches by applying them to a series of example problems. R. Luus misinterpreted the context of the paper, treating the work as a comparison of parameter estimation methods. This is clearly not the case. Nevertheless, since R. Luus invites such a comparison, we will now take this as an opportunity to point out all the limitations of the approach by Luus and Jaakola[2]. The LJ[2] approach will be placed in the appropriate context based on (i) theoretical advances, (ii) algorithmic issues, and (iii) computational performance. Towards this goal we make the following points:

1. From the theory point of view, the LJ[2] procedure is *not* a deterministic global optimization method. As a result, it cannot guarantee convergence to the global minimum under *any* circumstances. In fact, convergence to even a *local* solution cannot be guaranteed since no check of first or second order optimality conditions is made. Our proposed methods[1] are *deterministic* in nature and can theoretically *guarantee* convergence to an ϵ-global minimum. The conditions required for proof of this claim are clearly presented in the paper. From the algorithmic and computational point of view, for the LJ[2] procedure to even come close to such convergence, the approach would have to be ran from an *infinite* number of starting points which would require *infinite time*. Therefore any comparisons made in terms of computational time are irrelevant.

2. In the application of the LJ[2] procedure, it was shown that the approach *failed* more than 15% of the time. It is stated in the last paragraph of the comment "LJ optimization procedure also avoids *most* of the numerous local optima that are present in these two problem." Our proposed approaches converge[1] to the global solution *every* time, and avoid *all* of the local minimum. It appears that R. Luus believes the local minima presented for these problems were determined using our global optimization

approach. In fact, it is clear that these minima were obtained through the application of a *local* method from numerous starting points. These minima are simply presented to illustrate the multiplicity of solutions to the examples.

3. The LJ^2 procedure is based on a collection of heuristics, which have a large number of adjustable parameters. These include the number of random points used at each iteration, the number of iterations in a pass, the number of passes, the size of the initial variable space, the rate of region reduction, and the starting point. The author himself states[3] "Three principle factors affecting the reliability of direct search optimization procedures are the starting point, the size of the initial search region, and the rate of search region reduction." The setting of these various parameters appears to be problem dependent, quite arbitrary, and done from prior experience with the approach. R. Luus touts the approach as easy to use, yet for a user unfamiliar with the method, the selection of these parameters can be quite difficult, and the selection of "wrong" values can easily lead to a local optimum.

4. It should be emphasized that, due to the nature of our approach (being a branch-and-bound based method), the global solution to the problem is actually identified as the upper bound very quickly In the Lotka-Volterra example, using the integration based method with sampling to determine the needed values of β with 25 points initially, at least 10 at each iteration, the global solution was identified as the upper bound in 4 iterations and 12.3 CPU sec on a Pentium III/600 (this run corresponds to the results presented in Table 20, columns 1 and 2 of Esposito and Floudas[1]). The remaining computational time is required to *prove* the global optimality of that solution. This compares very favorably with the solution time of 362.2 CPU sec (on a Pentium III/500) reported by R. Luus in his comment. Hence, our approach requires at least 20-fold less time to identify the global minimum than the LJ^2 procedure.

We would like to comment to other statements:

1. While discussing the Lotka-Volterra problem, R. Luus presents what he refers to as "more interesting results" as he increases the bounds on the parameters used in the problem. He shows that the global minimum objective function value occurs at regular intervals in the parameter space. It is quite obvious that these solutions exist as the result of frequency aliasing due to the cyclic nature of the dynamics. Since the global solution presented in the paper ($\theta_1 = 3.2434$, $\theta_2 = 0.9209$) has the least amount of "activity" between data points it would be chosen for the purposes of system identification. This is a common occurrence in sampled data systems and there really is nothing "interesting" about it.

2. Also, while discussing the Lotka-Volterra model, R. Luus refers to his work[4] in which the same model was studied with different data. In our work, the data used for the

example was generated by numerically integrating the model with a given set of parameters and adding random error to the resulting state values. In Luus[4] the data is generated in the same way, but used without any error being added. Our goal was to apply the approach to examples which represent *real* parameter estimation problems, and not the simple mathematical exercise of determining parameters with perfect data. For that reason, the data in Luus[4] was not used for our example and therefore the comment is not appropriate for our work. We would also like to point out that in the reference the additional "local" solutions have objective values of 2.5723×10^{-10} and 2.4449×10^{-10}. This gives a difference between a "local" and "global" objective value of less than 1×10^{-10}. It is difficult to believe the uniqueness of these solutions, unless strict numerical controls have been implemented in the LJ procedure. There is no mention of this in the reference.

3. While discussing example 5 in his comments, R. Luus chose to solve the problem by selecting 10 sets of random values for the parameters in the region $[0.1, 1.9]$. The bounds on these parameters are set to be $[0, 20]$ in the example in Esposito and Floudas [1]. Why was the approach applied using starting points from a region 10% of the size of the true variable space? In fact, both local minima fall outside of this range with the region being *much* closer to the global minimum then the next local minimum.

4. In the end, R. Luus states that "the global optimum of highly nonlinear optimization problems is very difficult to establish with absolute certainty". It is apparent that R. Luus is not familiar with the very significant advances and the large body of literature on *deterministic global optimization* methods. These methods can determine an ϵ-global solution to highly nonlinear problems with *theoretical guarantees*. The recent book by Floudas[5] provides an extensive exposition to rigorous deterministic global optimization approaches with applications ranging from nonconvex NLPs, to complex mixed integer problems with hundreds of local minima. Also the books by Sherali and Adams[6], Horst and Tuy[7], Horst and Pardalos[8], Horst et al.[9] and Tuy[10] describe theory, methods and applications of global optimization.

In sum, (a) the LJ[2] approach is not supported by theoretical advances that can result in guarantees of determining and ϵ-global solution, and hence it represents a heuristically based local search technique; (b) the computational performance in identifying, but not proving, the global solution is also inferior (at least a 20-fold increase in time if required for the Lotka-Volterra example) to the deterministic global optimization approaches of Esposito and Floudas[1].

References

(1) Esposito, W. R.; Floudas, C. A. Global Optimization for the Parameter Estimation of Differential–Algebraic Systems. *I&EC Res.* **2000**, *37* (5), 1841.

(2) Luus, R.; Jaakola, T. H. I. Optimization by direct search and systemic reduction of the size of search region. *AIChE J.* **1973**, *19* (4), 760.

(3) Wang, B.; Luus, R. Reliability of Optimization Procedures for Obtaining Global Optimium. *AIChE J.* **1978**, *24* (4), 619.

(4) Luus, R. Parameter Estimation of Lotka-Volterra Problem by Direct Search Optimization. *Hung. J. Ind. Chem.* **1998**, *26*, 287.

(5) Floudas, C. A. *Deterministic Global Optimization: Theory, Methods and Applications*; Nonconvex Optimization and its Applications. Kluwer Academic Publishers, 2000.

(6) Sherali, H. D.; Adams, W. P. *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*; Nonconvex Optimization and its Applications. Kluwer Academic Publishers: Dordrecht, 1999.

(7) Horst, R.; Tuy, H. *Global optimization: deterministic approaches*; Springer-Verlag: Berlin, 1993 2nd. rev. edition.

(8) Horst, R.; Pardalos, P. M. *Handbook of Global Optimization*; Nonconvex Optimization and its Applications. Kluwer Academic Publishers: Dordrecht, 1995.

(9) Horst, R.; Pardalos, P. M.; Thoai, N. V. *Introduction to Global Optimization*; Nonconvex Optimization and its Applications. Kluwer Academic Publishers: Dordrecht, 1995.

(10) Tuy, H. *Convex Analysis and Global Optimization*; Nonconvex Optimization and its Applications. Kluwer Academic Publishers: Dordrecht, 1998.