An Empirical Local Convergence Study of Alternative Coordination Schemes in Analytical Target Cascading

Nikos Tzevelekos[1], Michael Kokkolaras[1], Panos Y. Papalambros[1], Martijn F. Hulshof [2], L.F. Pascal Etman[2], and J.E. (Koos) Rooda[2]

[1] Department of Mechanical Engineering, The University of Michigan
2350 Hayward, Ann Arbor, Michigan 48109-2125
[2] Department of Mechanical Engineering, Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven, The Netherlands
{babis,mk,pyp}@umich.edu, m.f.hulshof@student.tue.nl, {l.f.p.etman,j.e.rooda}@tue.nl

1. Abstract
The objective of this work is to investigate empirically the numerical behavior of the analytical target cascading process and local convergence properties of different coordination strategies. We adopt the χ language to specify and implement various coordination schemes, and employ a simple analytical example. We examine the effect of linking variables, subproblem solution accuracy, and amount of significant digits to represent quantities passed among the subproblems on numerical stability. Four different coordination schemes for a three-level hierarchy are evaluated in terms of accuracy and computational cost. We believe that the findings of this study aid us improve our understanding and initiate an effort to study multilevel algorithms and coordination methods for decomposition-based optimization.
2. Keywords: Multilevel hierarchical optimization, analytical target cascading, coordination strategies, local convergence properties, χ language.

3. Introduction

Analytical target cascading (ATC) is a methodology for optimal system design by decomposition [1]. A hierarchical, multilevel, model-based decomposition of the system is defined. Elements at adjacent levels in this hierarchy are coupled by means of their responses (responses of lower-level elements are inputs to higher-level elements). Elements at the same level can share so-called linking design variables. Optimization problems associated with the elements are formulated to minimize deviations from propagated targets subject to consistency constraints.

Global convergence properties of the ATC process have been proven theoretically under standard convexity and smoothness assumptions for a specific class of coordination strategies [2]. However, local convergence properties have not been studied either theoretically or empirically. Moreover, there exist alternative coordination schemes that are not supported yet by convergence theory but work quite well in practice.

In previous work [3], we demonstrated that the computer science language χ, originally developed to model manufacturing systems by means of discrete-event simulation, is an appropriate and useful tool to specify and implement precisely alternative coordination schemes of ATC in a rapid and efficient manner. An analytical example was used to illustrate how χ can be used to enable empirical local convergence studies.

Preliminary computations exhibited inconsistent results and trends that could not be interpreted systematically. Therefore, in this paper we use the same analytical example to study the numerical behavior of the ATC process. We investigate two possible causes for this inconsistent numerical behavior: a) the presence of linking variables and b) the accuracy of the subproblems in relation to the amount of significant digits to represent the quantities communicated among the subproblems. In addition, we evaluate the performance of the four different coordination schemes that can be employed in the case of a three-level hierarchy.

4. Brief overview of analytical target cascading

The analytical target cascading process is outlined in detail in [1,2,3]. Here, we present the formulation briefly. Assuming a multilevel, hierarchical decomposition of the original design problem, a deviation minimization problem is formulated and solved for each element j at a level i in the hierarchy. The vector of optimization variables $\bar{\mathbf{x}}_{ij}$ consists of local design variables $\mathbf{x}_{ij}$, linking design variables $\mathbf{y}_{ij}$, and response variables $\hat{\mathbf{R}}_{ij}$, where $\hat{\mathbf{R}}_{ij} = \left[\tilde{\mathbf{R}}_{ij}^t, \mathbf{R}_{ij}^t\right]^t = \mathbf{r}_{ij}(\hat{\mathbf{R}}_{(i+1)k_1}, \ldots, \hat{\mathbf{R}}_{(i+1)k_{c_{ij}}}, \mathbf{x}_{ij}, \mathbf{y}_{ij})$ with c_{ij} being the number of "children" of this element, and t denoting transpose. Responses $\tilde{\mathbf{R}}_{ij}$ are associated with local targets $\mathbf{T}_{ij}$, and responses $\mathbf{R}_{ij}$ are associated with targets $\mathbf{R}_{ij}^U$ that are cascaded down to the element from its parent. Furthermore, $\mathbf{R}_{(i+1)k}^L$ are response values passed up to the element from its k-th child, and $\mathbf{y}_{ij}^U$ and $\mathbf{y}_{(i+1)k}^L$ are linking design variable values cascaded down and passed up from the parent and children elements, respectively. Finally, we define tolerance optimization variables ϵ_{ij}^R and ϵ_{ij}^y for coordinating responses and linking variables, respectively. The mathematical problem formulation is then

$$\begin{aligned}
\min_{\bar{\mathbf{x}}_{ij}} \|\tilde{\mathbf{R}}_{ij} - \mathbf{T}_{ij}\|_2^2 + \|\mathbf{R}_{ij} - \mathbf{R}_{ij}^U\|_2^2 + \|\mathbf{y}_{ij} - \mathbf{y}_{ij}^U\|_2^2 + \epsilon_{ij}^R + \epsilon_{ij}^y \\
\text{subject to} \qquad \sum_{k \in \mathcal{C}_{ij}} \|\mathbf{R}_{(i+1)k} - \mathbf{R}_{(i+1)k}^L\|_2^2 \leq \epsilon_{ij}^R \\
\sum_{k \in \mathcal{C}_{ij}} \|\mathbf{y}_{(i+1)k} - \mathbf{y}_{(i+1)k}^L\|_2^2 \leq \epsilon_{ij}^y \\
\mathbf{g}_{ij}(\hat{\mathbf{R}}_{ij}, \mathbf{x}_{ij}, \mathbf{y}_{ij}) \leq 0, \\
\mathbf{h}_{ij}(\hat{\mathbf{R}}_{ij}, \mathbf{x}_{ij}, \mathbf{y}_{ij}) = 0,
\end{aligned} \tag{1}$$

where $\mathbf{g}_{ij}$ and $\mathbf{h}_{ij}$ are vector functions representing inequality and equality design constraints, respectively.

5. Example

We use the geometric programming problem we presented in [3]:

$$\begin{aligned}
\min_{\mathbf{z} \geq \mathbf{0}} \qquad z_1^2 + z_2^2 \\
\text{s.t.} \quad z_3^{-2} + z_4^2 - z_5^2 \leq 0 \;\; ; \;\; z_5^2 + z_6^{-2} - z_7^2 \leq 0 \;\; ; \;\; z_8^2 + z_9^2 - z_{11}^2 \leq 0 \\
z_8^{-2} + z_{10}^2 - z_{11}^2 \leq 0 \;\; ; \;\; z_{11}^2 + z_{12}^{-2} - z_{13}^2 \leq 0 \;\; ; \;\; z_{11}^2 + z_{12}^2 - z_{14}^2 \leq 0 \\
z_1^2 - z_3^2 - z_4^{-2} - z_5^2 = 0 \;\; ; \;\; z_2^2 - z_5^2 - z_6^2 - z_7^2 = 0 \\
z_3^2 - z_8^2 - z_9^{-2} - z_{10}^{-2} - z_{11}^2 = 0 \;\; ; \;\; z_6^2 - z_{11}^2 - z_{12}^2 - z_{13}^2 - z_{14}^2 = 0
\end{aligned} \tag{2}$$

The solution of problem (2) is $\mathbf{z}^* = [2.84\ 3.09\ 2.36\ 0.76\ 0.87\ 2.81\ 0.94\ 0.97\ 0.87\ 0.80\ 1.30\ 0.84\ 1.76\ 1.55]$. We decompose the original problem into three levels as shown in Fig. 1.

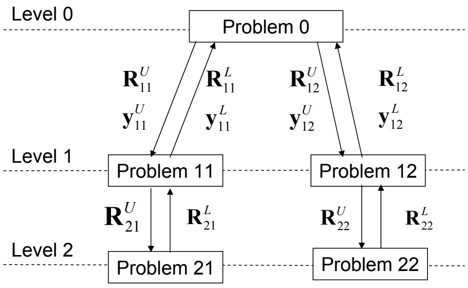

Figure 1: Hierarchical structure of decomposed problem

The subproblems are formulated following the notation presented in the previous section; the index j is dropped at the top-level problem since there is only one element.

The top-level problem P_0 is formulated as

$$\begin{aligned}
\min_{\mathbf{R}_{11}, \mathbf{R}_{12}, \mathbf{y}_0, \epsilon_0^y, \epsilon_0^R} \qquad \|\mathbf{R}_0 - \mathbf{T}_0\|_2^2 + s\epsilon_0^y + s\epsilon_0^R \\
\text{s.t.} \qquad \|\mathbf{y}_0 - \mathbf{y}_{11}^L\|_2^2 + \|\mathbf{y}_0 - \mathbf{y}_{12}^L\|_2^2 \leq \epsilon_0^y \\
\|\mathbf{R}_{11} - \mathbf{R}_{11}^L\|_2^2 + \|\mathbf{R}_{12} - \mathbf{R}_{12}^L\|_2^2 \leq \epsilon_0^R,
\end{aligned} \tag{3}$$

where $\mathbf{R}_{11} := z_1$, $\mathbf{R}_{12} := z_2$, $\mathbf{y}_0 := z_5$, $\mathbf{R}_0 = r_0(\mathbf{R}_{11}, \mathbf{R}_{12}) = z_1^2 + z_2^2$, and $\mathbf{T}_0 = 0$. Note that z_1, z_2, and z_5, correspond to the formulation of the original problem, and that z_5 is a linking variable computed at the problems of the intermediate level and coordinated at the top level. Also note that we have introduced a scaling parameter s to scale the tolerance optimization variables.

There are two problems at the intermediate level. Problem P_{11} is formulated as

$$\begin{aligned}
\min_{\mathbf{R}_{21}, \mathbf{y}_{11}, \mathbf{x}_{11}, \epsilon_{11}^R} \quad \|\mathbf{R}_{11} - \mathbf{R}_{11}^U\|_2^2 + \|\mathbf{y}_{11} - \mathbf{y}_0^U\|_2^2 + s\epsilon_{11}^R \\
\text{s.t.} \qquad \|\mathbf{R}_{21} - \mathbf{R}_{21}^L\|_2^2 \leq \epsilon_{11}^R \\
\mathbf{g}_{11}(\mathbf{R}_{21}, \mathbf{x}_{11}, \mathbf{y}_{11}) \leq \mathbf{0},
\end{aligned} \tag{4}$$

where $\mathbf{R}_{21} := z_3$, $\mathbf{x}_{11} := z_4$, $\mathbf{y}_{11} := z_5$, $\mathbf{R}_{11} = r_{11}(\mathbf{R}_{21}, \mathbf{x}_{11}, \mathbf{y}_{11}) = \sqrt{z_3^2 + z_4^{-2} + z_5^2}$, and $\mathbf{g}_{11}(\mathbf{R}_{21}, \mathbf{x}_{11}, \mathbf{y}_{11}) = (z_3^{-2} + z_4^2)z_5^{-2} - 1$.

Problem P_{12} is stated as

$$\begin{array}{cr} \min\limits_{\mathbf{R}_{22}, \mathbf{y}_{12}, \mathbf{x}_{12}, \epsilon_{12}^R} & \|\mathbf{R}_{12} - \mathbf{R}_{12}^U\|_2^2 + \|\mathbf{y}_{12} - \mathbf{y}_0^U\|_2^2 + s\epsilon_{12}^R \\ \text{s.t.} & \|\mathbf{R}_{22} - \mathbf{R}_{22}^L\|_2^2 \leq \epsilon_{12}^R \\ & \mathbf{g}_{12}(\mathbf{R}_{22}, \mathbf{x}_{12}, \mathbf{y}_{12}) \leq \mathbf{0}, \end{array} \tag{5}$$

where $\mathbf{R}_{22} := z_6$, $\mathbf{x}_{12} := z_7$, $\mathbf{y}_{12} := z_5$, $\mathbf{R}_{12} = r_{12}(\mathbf{R}_{22}, \mathbf{x}_{12}, \mathbf{y}_{12}) = \sqrt{z_5^2 + z_6^2 + z_7^2}$, and $\mathbf{g}_{12}(\mathbf{R}_{22}, \mathbf{x}_{12}, \mathbf{y}_{12}) = (z_5^2 + z_6^{-2})z_7^{-2} - 1$.

Finally, there are two problems at the bottom level. Note that z_{11} should be treated as a linking variable coupling the two problems of the bottom level. However, the ATC formulation presented in [1,2,3] does not allow subproblems to share variables unless they are children of the same parent. Although we have implemented "non-traditional" schemes in which the variable coupling the two problems at the bottom level can be treated as a linking variable that is coordinated either at the grandparent level or at the parent level by the introduction of an additional process, we do not present results here since this is out of the scope of this paper. Instead, we treat z_{11} as a parameter p using its known optimal value. Problem P_{21} is given then by

$$\begin{array}{ll} \min\limits_{\mathbf{x}_{21}} & \|\mathbf{R}_{21} - \mathbf{R}_{21}^U\|_2^2 \\ \text{s.t.} & \mathbf{g}_{21}(\mathbf{x}_{21}) \leq \mathbf{0}, \end{array} \tag{6}$$

where $\mathbf{x}_{21} := [z_8, z_9, z_{10}]$, $p = 1.3 (= z_{11})$, $\mathbf{R}_{21} = r_{21}(\mathbf{x}_{21}) = \sqrt{z_8^2 + z_9^{-2} + z_{10}^{-2} + p^2}$, and

$$\mathbf{g}_{21}(\mathbf{x}_{21}) = \begin{bmatrix} (z_8^2 + z_9^2)p^{-2} - 1 \\ (z_8^{-2} + z_{10}^2)p^{-2} - 1 \end{bmatrix}$$

and problem P_{22} is

$$\begin{array}{ll} \min\limits_{\mathbf{x}_{22}} & \|\mathbf{R}_{22} - \mathbf{R}_{22}^U\|_2^2 \\ \text{s.t.} & \mathbf{g}_{22}(\mathbf{x}_{22}) \leq \mathbf{0}, \end{array} \tag{7}$$

where $\mathbf{x}_{22} := [z_{12}, z_{13}, z_{14}]$, $p = 1.3 (= z_{11})$, $\mathbf{R}_{22} = r_{22}(\mathbf{x}_{22}) = \sqrt{z_{12}^2 + z_{13}^2 + z_{14}^2 + p^2}$, and

$$\mathbf{g}_{22}(\mathbf{x}_{22}) = \begin{bmatrix} (p^2 + z_{12}^{-2})z_{13}^{-2} - 1 \\ (p^2 + z_{12}^2)z_{14}^{-2} - 1 \end{bmatrix}.$$

6. Coordination strategies

The possible coordination schemes for a three-level hierarchy are depicted on Figure 2. Schemes I and II are

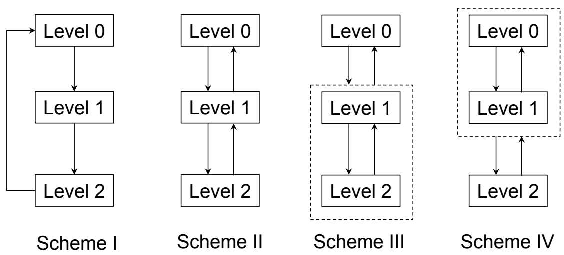

Figure 2: Possible coordination strategies for three-level hierarchy

not supported by the convergence theory presented in [2], which requires convergence between adjacent levels (as in schemes III and IV) before proceeding to the next level. The χ language has been used to specify and implement the four coordination schemes as discussed in [3]. This language has been developed originally to model manufacturing systems by means of discrete-event simulation, and is based on communicating parallel processes. In [4] we demonstrate that such a language is highly suitable to specify the coordination in decomposed multidisciplinary optimization problems. In [3] we show how the ATC decomposition framework can be modelled using χ. Each ATC optimization subproblem is viewed as a process. The various processes may be executed in parallel; however, they require data from other processes before they can proceed. Using χ, we are able to specify and implement different coordination schemes for solving the subproblems. Note that scheme IV was not implemented in [3].

We have employed the Matlab 6.0 implementation (`fmincon`) of the sequential quadratic programming (SQP) algorithm for the solution of all the subproblems. The optimization solver accuracy (represented by the three `fmincon` termination criteria `TolX`, `TolFun`, and `TolCon`) was set to 10^{-7}. We performed numerical experiments by varying both the tolerance associated with the termination criterion of the ATC process (defined by the change in the solution between two successive ATC iterations, i.e. $\|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k\|_2^2 \leq tol$ for all subproblems) and the scaling parameter values of the tolerance optimization variables. We used both finite differences and analytical expressions to compute gradients. We measured accuracy by comparing the ATC solution to the known solution of the original problem. Computational cost was determined as a function of ATC iterations n. For scheme I, total computational cost is $3n + 1$ (the top-level problem is solved $n + 1$ times, and each intermediate-level and bottom-level problem is solved n times). Similarly, for scheme II the total computational cost is given by $4n + 1$. For scheme III the total computational cost is given by $n + 2\max(n_l, n_r) + 1$, where n_l and n_r are the numbers of total iterations required for the left and right branches of the hierarchy tree. Finally, for scheme IV the computational cost is given by $2n_t + n_b + 1$, where n_t and n_b are the numbers of total problems solved for the top and bottom problems, respectively. Note that the ATC process starts at the bottom level for scheme IV, as opposed to the the other three schemes.

7. Preliminary observations

The following observations were made based on obtained preliminary results: a) the termination tolerance did not affect accuracy (although it obviously affected computational cost), b) no scaling parameter value could be identified for which accuracy was consistently high, c) using finite differences to compute gradients sometimes yielded more accurate solutions (although no systematic trend could be identified), and d) different Matlab versions (5.3, 6.0, and 6.1) yielded slightly different results. Figure 3 illustrates observations a) and b), while Figure 4 illustrates observation c).

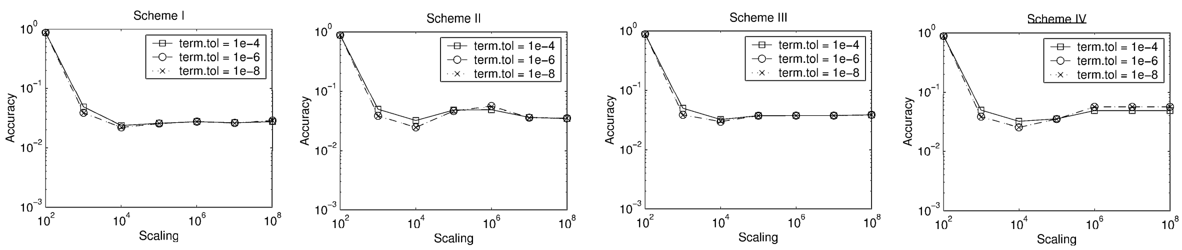

Figure 3: Preliminary accuracy results using analytical gradients

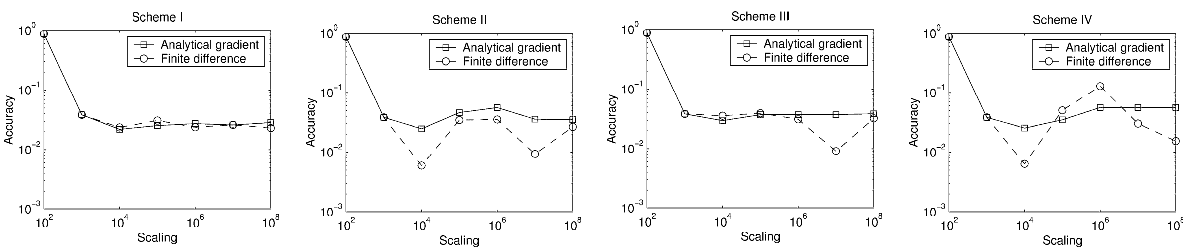

Figure 4: Preliminary accuracy results sometimes differ significantly when using finite differences instead of analytical gradients (with ATC termination tolerance $tol = 10^{-6}$)

8. Investigations, experimentations, and final results

Based on the above observations, we decided to investigate two possible reasons for these numerical discrepancies: the presence of linking variables and the numerical inaccuracies of the quantities communicated among the subproblems. We found that scheme I fails to converge when the linking variables are removed and large values of the scaling parameters are used, while schemes II, III, and IV yielded accurate results after few ATC iterations. Although we cannot explain why scheme I failed in this case while it did converge at the presence of linking variables, this may indicate that it is not appropriate for general use on other problems. Note that scheme I is not supported by the convergence theory of [2]. The fast ATC convergence without linking variables implies that the presence of the latter is indeed the basic cause for the significant increase of ATC iterations. The more tightly coupled the subproblems are the more difficulties the ATC process will probably have to quickly converge to an accurate solution.

A less expected finding is associated with the accuracy of the subproblem solutions and the inaccuracies of the communicated data. We should not have obtained different ATC results when using finite difference gradients instead of analytical expressions, or different Matlab versions. Therefore, we have experimented with the amount of significant digits used to represent the quantities passed among the subproblems. The plot on the left of Figure 5 shows the accuracy obtained for different significant digits using Scheme I. We can see how the accuracy deteriorates as more digits are used while keeping the optimizer tolerance fixed at 10^{-7}. Furthermore, we can see from the plot in the middle of Figure 5 that if the optimization tolerance is tightened (in this case from 10^{-7} to 10^{-9}), i.e., the subproblem solution accuracy is higher, we can achieve same ATC solution accuracy using more significant digits. This relation between subproblem solution accuracy and significant digits is illustrated in the right plot of Figure 5. Similar results were obtained for all schemes. We conclude that we

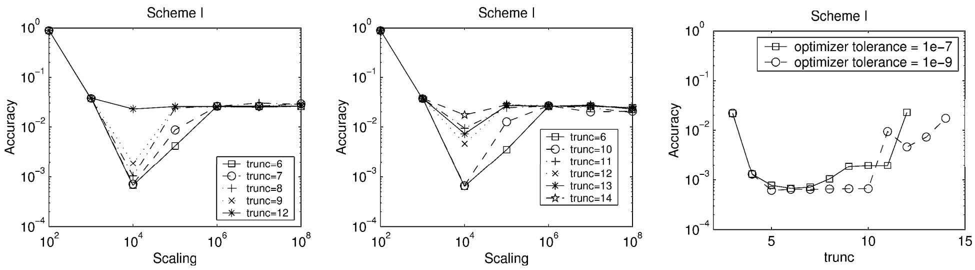

Figure 5: Accuracy results obtained using different number of significant digits to represent quantities passed among subproblems (with ATC termination tolerance $tol = 10^{-8}$ and scaling parameter value for the plot at the right $s = 10^4$)

must use an appropriate amount of significant digits that is related to the accuracy of the subproblem solutions. Doing this, the ATC process indeed yields results that follow systematic trends. Specifically, a) tightening the ATC termination tolerance does improve accuracy, b) a scaling parameter value ($s = 10^4$) is identified for which accuracy is highest, and c) analytical and finite difference gradients yield consistent results. Using more significant digits than the available accuracy of the subproblem solutions to represent the communicated quantities causes numerical instabilities within the ATC process. This is the case both when using finite differences and analytical expressions to compute gradients.

Figures 6 and 7 depict the accuracy and computational cost, respectively, for all four schemes using analytical gradients, optimization tolerance $tol = 10^{-7}$, and 6 significant digits to represent quantities passed among subproblems. It can be seen that all four schemes exhibit almost identical qualitative and quantitative behavior. Scheme IV seems to require the lowest number of ATC iterations (note the logarithmic scale). On the other hand, Scheme IV is more difficult to implement and one may prefer a more simple scheme. As expected, the results obtained using finite differences differ only slightly. Based on the observed trends, one can not predict that a certain scheme will outperform the others when using finite differences. Scheme I failed for certain scaling values when the linking variables were removed from this example. We must also note that measuring the computational cost in terms of ATC iterations may not be the best way. It may be more appropriate to record total number of iterations and/or function evaluations of all subproblems. This is an issue that we will address in future work.

9. Conclusions

The ATC process is highly sensitive to numerical inaccuracies in the data communicated among the subproblems. It is imperative to use only as many significant digits as necessary to represent the accuracy of the subproblem solutions. Only after this is ensured can one compare accuracy and computational cost of different coordination schemes. For the particular problem, all four schemes have similar local convergence properties.

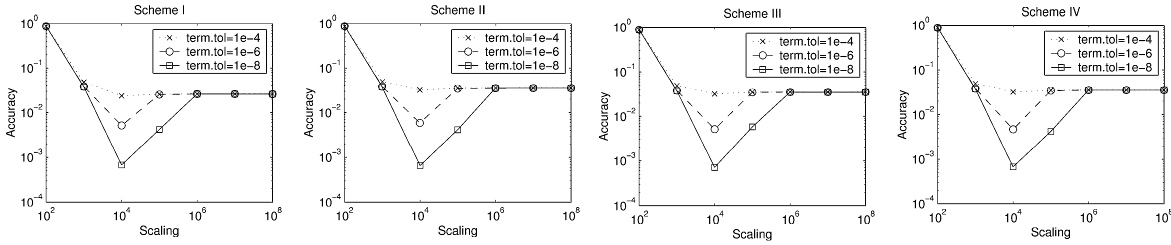

Figure 6: Accuracy results for all four schemes

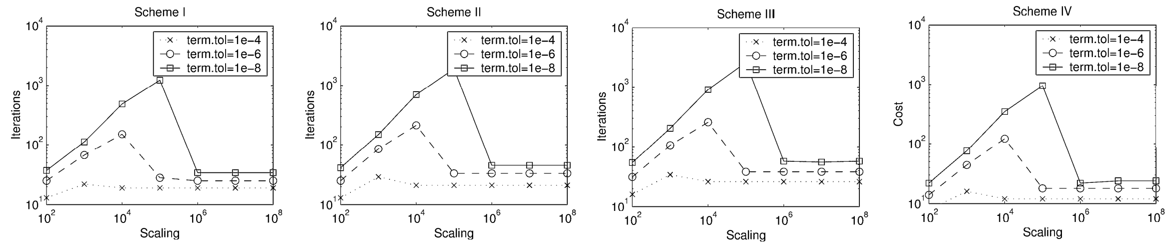

Figure 7: Computational cost for all four schemes

In such a case, it is suggested to use the simplest or most inexpensive alternative. The latter would indicate a slight preference for Scheme IV. Here we “measured” computational cost in terms of ATC iterations or total number of subproblems solved. However, there are alternative ways (e.g., record the total number of SQP iterations and/or function evaluations) to assess the computational cost. As expected, the presence of linking variables affects the performance of the coordination schemes. The required number of ATC iterations may increase significantly when a higher solution accuracy is required. An interesting finding was that coordination scheme I converges in the presence of linking variables, while it fails when these are removed. However, the fact that this coordination scheme is not supported by convergence theory indicates that such schemes will only work for some problems.

References

[1] H.M. Kim, N.F. Michelena, P.Y. Papalambros, and T. Jiang. Target Cascading in Optimal System Design. To appear in the *Journal of Mechanical Design*, 2003. An earlier conference version of this paper was presented at the *26th ASME Design Automation Conference*, Baltimore, MD, 2000, paper no DAC-14265.

[2] N. Michelena, H. Park, and P.Y. Papalambros. Convergence Properties of Analytical Target Cascading. To appear in the *AIAA Journal*. An earlier conference version of this paper was presented at the *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, 2002, paper no. AIAA-2002-5506.

[3] L.F.P. Etman, M. Kokkolaras, P.Y. Papalambros, A.T. Hofkamp, and J.E. Rooda. Coordination Specification of the Analytical Target Cascading Process Using the χ Language. Submitted to *Structural and Multidisciplinary Optimization*. An earlier conference version of this paper was presented at the *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, 2002, paper no. AIAA-2002-5637.

[4] L.F.P. Etman, A.T. Hofkamp, J.E. Rooda, M. Kokkolaras, and P.Y. Papalambros, . Coordination Specification for Distributed Optimal Design Using the χ Language. Submitted to *Structural and Multidisciplinary Optimization*. An earlier conference version of this paper was presented at the *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, 2002, paper no. AIAA-2002-5410.