REPORT ON A DIPLOMA THESIS

Title: Detection and correction of silent errors in pipelined Krylov subspace methods
Author: Bc. Jakub Hercík

Thesis summary

The thesis deals with the problem of detection of a class of errors that can appear in iterative methods to solve systems of linear algebraic equations (linear systems) and do not lead to a method breakdown. These errors called the silent errors may result in an unexpected behavior of the method, possibly resulting in slowing down (delaying) its convergence. The author discusses the problem for a specific subclass of Krylov space iterative methods called the pipelined ones. These methods derived as modifications of the seminal iterative approaches seem to offer more potential parallelism. And more parallelism, in particular, more massive parallelism on the emerging computational tools, motivates the discussion on the silent errors that can occasionally appear.

After mentioning basic theoretical background, the author introduces the pipelined methods to solve linear systems with the system matrix positive definite, that is, methods derived from the starting point of the conjugate gradient method.
One of these methods, Pipe-PR-CG method is than used as a starting point to develop new strategies to detect the silent errors. The main idea behind the approach is to compute some quantities of the method in more alternative ways that should provide the same results in exact arithmetic. In case these quantities differ more than predicted by floating-point-based estimates, a correcting procedure that recovers the iterations is used.
The thesis is a balanced mixture of theoretical derivations and implementations. My overall opinion is that it is a very good basis for the defense.

Overall evaluation

Subject of the thesis. Well chosen. Corresponds to the original plan.

Own contribution of the author. Yes. A couple of items like the detection of silent errors are new.

Mathematical contents. I appreciate this. I really like that the codes can be taken from the thesis and implemented in a simple interpreter like Matlab. I did this for a part of the schemes.

Citing the sources. Well done.

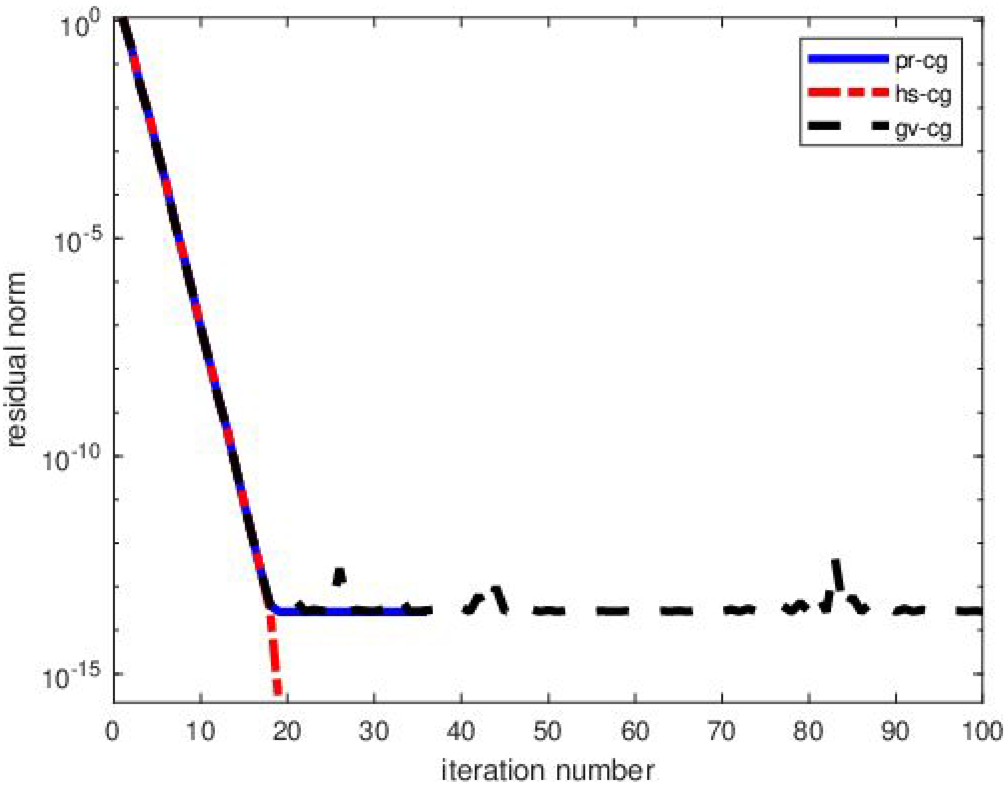
Formal shape of the thesis. Very good. If I would like to improve something, I would significantly decrease the number of figures. Instead, the results should be quantified/summarized in a much simpler way. For example, is it possible to summarize the results at figures 3.5-3.18? Similarly for the gaps?

At this point I would like to mention that I like the thesis, for me it is a nice example how the diploma thesis should look like and I do not have any other comment to what was done.

But still I am in the mood to comment on the subject further. And, during the defense author could just answer/comment on some of the points. Since the questions go mostly beyond the thesis, they do not need to be answered fully.

Questions and comments

1. *Do the figures 2.4 and 2.5 capture the average/exceptional behavior of the methods? Consider* *the following figure for the PR-CG method preconditioned by IC0 for the matrix BCSSTM07 (Jacobi scaled). You can see that the only method that still decreases is the classical CG by Hestenes and Stiefel. Here the plots show absolute residual norms but very similar expressions can be obtained for unpreconditioned method, showing relative residual norms etc. (I made more runs).*



2. While the PR-CG method here attains similar final level of accuracy as GV-CG, a problem that I had to fight with was that the runs were typically finished by getting NaNs since the value of \nu goes to zero very quickly. Also some of the other scalar quantities. Of course, this can be monitored (as there are checks in BiCGStab, for example – but note, here we should go more down than in standard codes) , but it should have some implications.
*For example, for **low** precision computations that may be also naturally connected (and may be in the future even more) to parallel iterative solvers of large problems. Did you noticed such effects? Note that I **did not** noticed a similar effect when running the same problems by the algorithm that uses two coupled recurrences.*
I know that at this point I am more commenting on PR-CG (and I have not studied the original paper). But still I think that an (unusual for me) sequence of scalar quantities may be worth to mention.

3. What about scaling? Did you try this? Did it influence the computation? Do we need such small residual (scaled residual) norms? Where?

*4. What about an analysis in case of preconditioning (considering effect of another solves or matvecs). Would this be useful? Note that the floating-point analysis (that **I appreciate** in the thesis) is based (modulo my weak understanding of it) on rough **(?)** estimates and preconditioning (as a must for solving real problems) may make the analysis even more inaccurate. How much inaccurate?*

5. Yes, computation in two or more possible ways may be a way if errors are more frequent. This can be done for preconditioner components as well and there are so many nice possibilities 😊. Would that be really useful in practice? Or the silent errors are nowadays so rare? (Here I would *appreciate if some effort should be done to mention whether we are really entering world of less reliable hardware/communication lines (because of the hardware/communication complexity).*

6. Do you have an idea how expensive are the checks? Note that my comments related to possible preconditioning are trying to emphasize (behind the lines) that the convergence curves from solving preconditioned problems will be shorter and the different methods may provide more similar curves. This was behind my original intention to play with algorithms.

7. What about trying to detect the behavior by checking the convergence curve and its behavior (delay)? Using error estimates? Note that this may be an alternative. Then, using state-of-the-art implementation based on a compiled programming language might provide a realistic quantitative comparison.

8. Ad the specific simulation. I understand that it is a difficult task to make a realistic choice for simulation of silent errors. My opinion is more on the side that such errors described may extinguish since development of hardware and software may lead to ubiquitous error correcting primitives available nearly everywhere. But it could happen that for efficiency one will have consider perturbation of iterative process by a "limited asynchronicity" in iterations in highly parallel environment. This may be something, combining this possibly in a combination with other changes, worth to follow. Could you comment on this?

Conclusion

Repeating my point, the thesis is very good. I recommend to accept the thesis as a very good basis for the defense. Nice work.

Report written by

Miroslav Tůma, Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University

prof. Ing. Miroslav Tůma, CSc.

Digitálně podepsal prof. Ing. Miroslav Tůma, CSc.
Datum: 2024.05.20 15:08:03 +02'00'