MULTIGRID METHODS FOR LARGE-SCALE PROBLEMS: APPROXIMATE COARSEST-LEVEL SOLVES AND MIXED PRECISION COMPUTATION

by Petr Vacek

Let me start this report with a preliminary remark. It is not easy to evaluate the skills of a Ph.D. candidate when the manuscript is made of some articles, particularly when the articles have several co-authors. Was the candidate the main contributor and the main writer? It is even worse if, like in this manuscript, one of the articles has already been published in a journal. If I would make negative comments about that article, would it mean that the referees had not made a good job?

Let us now consider the manuscript which has three chapters. The topic of the first chapter is the study of a V-cycle multigrid method for a symmetric positive definite linear system when the coarsest level has a large number of nodes or unknowns. It may not be feasible to use a direct method to solve the linear system on the coarsest level and the solution has to be approximated with an iterative method.

The authors start with some motivating examples which arise from the discretization of simple second-order PDEs on the unit square. The conclusion is that that choice of coarsest-level solver accuracy can have a significantly effect on the convergence behavior. Then, using some assumptions on the accuracy of the coarsest-level solver, they study how the error is propagated in one V-cycle of the multigrid method. They do that by comparison with the V-cycle with an exact coarsest-level solver. Then, they study some stopping criteria for the coarsest-level solver and describe some numerical experiments illustrating the theory for the two simple problems used in the motivating examples.

If I would have been a referee for this article as it is in the manuscript, I would have made the following remarks. Excerpts from the manuscript are in italic.

p. 15 *respectively its transpose.* It seems that some words are missing in this sentence.

The authors never tell us which prolongation (interpolation) and restriction operators are used in the numerical experiments.

p. 16 $v_j = v_j + M_j(f_j A_j v_j)$ is not a correct mathematical expression. It must be $v_j + M_j(f_j A_j v_j) \to v_j$.

(1.2) seems a strong assumption that may not be satisfied by all smoothers.

p. 17 Why not showing a picture of the fine mesh used for the experiments?

We generate the sequence of stiffness matrices $A_{0:J}$, by discretizing the problems on each level of the hierarchy., Is (1.1) not used? We would have like to know which finite elements were used.

We use the standard prolongation matrices What are they?

We modify the stiffness matrices What does this mean?

We would have like to know how many levels were used. Was CG used without preconditioning?

p. 19 The conclusion could depend on the ingredients of the multigrid method. What about the loss of orthogonality in CG? What about preconditioning?

p. 20 What is the A-norm of a matrix? It is used before being defined.

p. 21 Last line is not really proved.

p. 22 Assumptions (1.11) and (1.12) seem very unrealistic for solving practical problems.

Theorem 1.1 is not formulated as it should have been. The inexact algorithm does not start from the same x_{prev}. Things must be presented in a different way like: exV-cycle starts from the same iterate...

p. 23 What is $v_0^{(k)}$? It is not obtained from exV-cycle.

p. 25 I do not fully understand (1.21).

p. 27 After (1.30), *this is not a significant overestimation* How do you know that?

p. 28 *We simulate the exV-cycle method by using MATLAB backslash operator as the solver on the coarsest level.* How is v_0 computed?

What is x_{ex}^{new} in (1.32)? This is not explained very clearly.

p. 30 *The rates are significantly lower than its bounds* → The rates are significantly lower than their bounds.

How do you get an approximation of the smallest eigenvalue of the coarsest matrix?

In conclusion, this paper contains interesting mathematical results. But, if I would have been a referee, I would have asked for a minor revision. However, I did not check if this chapter is completely similar to the published paper.

The goal of the second chapter is to develop a posteriori error estimates when the coarsest-level problem is solved approximately with the conjugate gradient method using an appropriate stopping criterion. This chapter is quite technical. The authors consider several a posteriori estimates on the total and algebraic errors based on decomposing the error on a sequence of finite element subspaces and using approximation properties of quasi-interpolation operators, stable splittings, or frames.

The main contribution of the paper is a procedure for approximating the term associated with the coarsest-level problem. This paper is not easy to read mainly because of all the constants that are involved in bounds. The authors do not tell us if they are computable or not, and if the bounds are sharp or not.

Some remarks are following.

p. 43 Note that approximations of $r_0^* A_0^{-1} r_0$ can be obtained by running a few iterations of the Lanczos algorithm.

p. 45 What is the equation satisfied by v_J?

The paper starts with a long review of works of other researchers. It is not easy to see if there are any new results in that part of the paper.

p. 46 Why using the same notation for different inner products?

We first recall the standard residual-based error estimator for the discretization error in a single-level setting assuming exact algebraic computations or to steer an adaptive mesh refinement. This is a strange sentence.

p. 47 What are the significations of the names of the constants?

p. 56 I do not fully understand what is P_j^J in (2.39).

p. 59 It seems that the main contribution of the paper is Section 2.5.4. Note that, as I wrote above, the authors could have used Lanczos to approximate $r_0^* A_0^{-1} r_0$.

p. 62 Is C_{numexp} an ad-hoc constant?

p. 63 *Recall, however, that using a direct solver is for large problems in practice unfeasible.* This is clearly not the case for this example.

What is the stopping criteria for CG?

p. 65 Why do we have this difference between other levels and the coarsest one?

The estimates for total and algebraic errors involve some constants that must be approximately determined, which involves heuristics. So, what do we have to do in practice?

it provides a justification for extrapolating the estimated values of the constants from smaller to larger problems. This does not seem very realistic.

This paper is an interesting contribution to the computation of error estimates in finite element problems solved with the multigrid method. Note that I did not check all the proofs in the appendix.

Chapter 3 studies some aspects of using a lower precision in parts of a multigrid method. It is proposed to use an incomplete Cholesky (IC) factorization in lower precision as smoother. Numerical experiments using a GPU are described.

Some remarks are following.

p. 88 *we assumed that the computation was done in infinite precision arithmetic.* There is nothing as "infinite precision arithmetic", this is just mathematics!

factorizarion → factorization.

p. 91 *the computation does not break down due to overflow or underflow,* This is a strong assumption when using low precision, particularly in fp16 which has only five bits for the exponent.

p. 94 *is transpose of the prolongation matrix* → is the transpose of the prolongation matrix.

p. 95 *There are other versions where smoothing is done both before and after* Smoothing before or after or both is more or less the same.

As I wrote above, there is nothing like *infinite precision*, this is just mathematics. Since everything is computed with the same precision in Algorithm 3.2, why mentioning it?

For (3.12), after how many iterations is this assumed since this cannot be used as a stopping criterion?

p. 96 In Theorem 3.1, δ_{TG} may not be small. The proof is written in a strange way. Why assuming (3.15) and (3.16)?

p. 97 The author should recall that y is the exact solution.

p. 100 *or a product of at least two of them.* Is this the product of ε with something else? What happens if Λ_M or Λ_C are large? What is K_α?

It would have been interesting to know on a given level which operations can be done in lower precision. (this was done for the smoothing)

Why choosing this strategy on the precisions?

p. 102 Explain why using (3.37) for the coarsest level.

p. 103 Is (3.38) a realistic assumption?

p. 104 Where is the hypothesis on the precision needed in the proof?

A smoothing routine computes an approximate solution The role of a smoothing routine is to smooth, not to compute an approximate solution.

p. 105 *there is no finite precision error analysis of incomplete Cholesky factorization for general SPD* IC does not always work for general SPD matrices. One needs a stronger hypothesis, like being an H-matrix.

It is likely that solving trianguler systems in mixed precision must be already known. Why not rounding b? Note that the author could have used the results of S.M. Rump and C.-P. Jeannerod, "Improved backward error bounds for LU and Cholesky factorizations", SIMAX, v. 35 n. 2 (2014), pp. 684-698.

p. 108 Scaling can be expensive. In practice, what is the overhead of scaling?

p. 109 It does not seem very necessary to solve 1D problems with multigrid. Moreover, it is not necessary to solve this smooth 1D problem with 409,599 nodes.

p. 111 Once again, the goal is not to have LL^T like A, but to smooth the residual vector.

p. 113 The explanations about the experiment are confusing. Using 64 bits for the exponent when simulating low precision is not very realistic if some checks on the range of numbers are not done. In low precision there can be problems because of the small number of bits used for the exponent.

p. 117 It would have been nice to explain what is done on the GPU and what is done on the CPU.

In conclusion, this paper shows that, at least on some problems, the incomplete Cholesky smoothers can be used in lower precision. It would have been nice to study if some other parts of the multigrid algorithm could be done in lower precision while maintaining the same accuracy as in double precision.

This manuscript contains interesting mathematical and numerical results about the multigrid method. The writing and the English could have been improved. Assuming that the candidate is the main contributor of the papers, he deserves to obtain a doctoral degree.

Gérard Meurant, November 19, 2024