Acceleration of the Finite Element Gaussian Belief Propagation Solver Using Minimum Residual Techniques

Yousef El-Kurdi[1], David Fernández[1,2], Warren J. Gross[1], and Dennis Giannacopoulos[1]

[1]Department of Electrical and Computer Engineering, McGill University, Montreal (QC), H3A 0E9, Canada
{yousef.el-kurdi@mail., david.fernandezbecerra@mail., warren.gross@, dennis.giannacopoulos@}mcgill.ca
[2]Escuela de Ingenieriá Eléctrica, Universidad del Zulia, Maracaibo, 4002, Venezuela

The Finite Element Gaussian Belief Propagation method (FGaBP) introduced recently provides a powerful parallel alternative to conventional Finite Element Method (FEM) solvers. In this work we accelerate the FGaBP convergence by combining it with two methods based on residual minimization techniques. Numerical results show considerable reductions in the total number of operations while maintaining the scalability features of FGaBP.

***Index Terms*—Finite Element Method, iterative methods, Krylov subspace methods, Gaussian Belief Propagation.**

I. Introduction

PARALLEL methods, such as the recently introduced Finite Element Gaussian Belief Propagation (FGaBP) method [1], address the challenging problem of attaining high computational scalability on manycore computing architectures used in High Performance Computing (HPC) platforms. The FGaBP algorithm, when adapted in a multigrid setting [2], demonstrated considerable scalability over conventional Finite Element Method (FEM) software. This was a direct consequence of the FGaBP probabilistic reformulation of the FEM computation using distributed message updates on a matrix-free data-structure. Most importantly, the multigrid adapted FGaBP (FMGaBP) solver [2] eliminates the need to perform global algebraic operations such as Sparse Matrix Vector Multiplies (SMVMs). Nonetheless, to help extend the applicability of the FGaBP solver to a wider range of applications, we here explore accelerating it using residual minimization with variants of Krylov subspace methods. While this may reduce the distributed behavior of FGaBP by introducing global algebraic operations after a number of FGaBP message update sweeps, our results show that a considerable reduction in number of operations can be realized making this solution very beneficial.

A key challenge emerges when accelerating FGaBP in the context of Krylov methods. The FGaBP solver uses distributed message computation supporting arbitrary update schedules, thus generating an iteration matrix is not tractable. As a result, a conventional approach Krylov preconditioner can not be used. In this work, a residual minimization technique and a flexible Krylov subspace method are used with the FGaBP solver to accelerate its convergence. Our results demonstrate considerable reductions in the overall computational load of FGaBP using these techniques.

II. Background

The FGaBP formulates the FEM as a variational inference problem by modifying its functional form as follows:

$$\mathcal{P}(U) = \frac{1}{Z}\prod_{s\in\mathcal{S}}\Psi_s(U_s) \tag{1}$$

where Z is a normalizing constant, $\Psi_s(U_s)$ are local factor functions corresponding to each finite element indexed by s in the S set of finite elements, U is the variables vector, and U_s is the subset variables connected to factor Ψ_s. For Symmetric Positive Definite problems, Ψ_s takes a multivariate Gaussian form albeit unnormalized. Correspondingly, the nodal variables are each assumed to model a random Gaussian variable. It can be shown that the solution to the underlying FEM problem can alternatively be obtained by inferring the marginal mean and variance parameters of each of the Gaussian variables in U. This, in turn, motivates the use of the Belief Propagation (BP) [3] as a computational inference algorithm. The BP is a recursive message passing algorithm that exhibits highly distributed computations by using intermediate results, generally referred to as local beliefs. The resulting FGaBP algorithm communicates messages between variable nodes representing U and factor nodes representing each finite element in a localized matrix-free form. Since the FGaBP messages take Gaussian forms, each message is composed of two parameters, a first order parameter (β) and a second order parameter (α). In [2], it was shown that the FGaBP can be adapted into a completely distributed and stationary multigrid process resulting in high computational scalability.

III. FGaBP Acceleration Using Iterative Recombination and Krylov Subspaces

The FGaBP can be accelerated using message relaxation, as shown in [4], resulting in considerable iteration reductions. However, such an approach can be limited since it uses information from only the previous iteration solution. In this work, we aim to obtain better solution approximations using a longer history of previous approximations. The successive solution approximations are obtained using the criterion to minimize the residual. The framework of this method is highlighted in [5, pp. 280-282] and [6] and is referred to as acceleration by Iterant Recombination (IR). The successive solution estimates $\bar{u}^{(m)}$ at iteration m is obtained as a linear combination of $\tilde{m}$

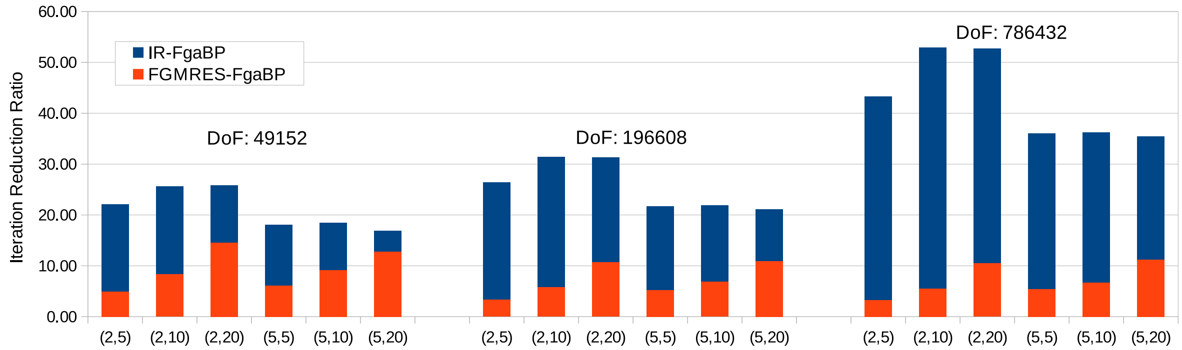

Fig. 1. Iteration reduction ratios of the FGaBP accelerated using the IR and the GMRES methods. The top numbers represent the DoF for each set, the lower numbers in parentheses $(\cdot\ ,\ \cdot)$ represent the inner FGaBP iterations and the size of the subspace, respectively.

previous solutions as follows:

$$\bar{u}^{(m)} = u^{(m)} + \sum_{i=1}^{\tilde{m}} a_i (u^{(m-i)} - u^{(m)}). \tag{2}$$

Here the factors a_i are chosen such that the residual L_2-norm is minimized as follows:

$$a_o = \arg\min_a \| d^{(m)} + \sum_{i=1}^{\tilde{m}} a_i (d^{(m-i)} - d^{(m)}) \|_2. \tag{3}$$

The IR method, while presented for multigrid in [5], exhibits great flexibility for the FGaBP algorithm. The FGaBP algorithm, as shown in [2], can be restarted from an arbitrary solution. At the IR step, the method needs to perform global operations such as SMVM and dot products; however in between the IR iterations, the FGaBP can perform a number of arbitrary update sweeps maintaining its distributed nature. Here, the SMVM operation utilizes the FGaBP matrix-free data-structure without a major impact on memory other than storing the truncated Krylov subspace since typically $\tilde{m} <= 10$.

The next approach we consider is to use the FGaBP to the effect of a preconditioner to a Krylov subspace method. A modified Krylov method of the classical Generalized Minimum Residual (GMRES) referred to as the Flexible GMRES (FGMRES) method [7, pp. 287-290] can accommodate a dynamically-varying preconditioner such as the FGaBP. The FGMRES computes the solution vector using a linear combination of the preconditioned orthonormalized subspace $z_j = M^{-1} v_j$. The FGaBP here is used as a right preconditioner by resetting its right-hand-side using the v_j vectors and restarting only its β messages from zero since the α messages depend only on the operator rather than the right-hand-side.

IV. Results

The behavior of the new algorithms is tested using the well-known 2D square conductor Laplace potential problem. The problem uses Dirichlet and Neumann boundary conditions and has a dimension of 1 cm. A quadrilateral mesh is used to mesh the domain containing one of the corners of the square conductor along the two lines of symmetry. All the experiments were terminated when the normalized residual's L_2-norm dropped below 10^{-6}.

The plots in Fig. 1 show the iteration reduction ratios of the FGaBP accelerated by IR (IR-FGaBP) and FGMRES (FGMRES-FGaBP). Experiments are performed for three sets of Degrees of Freedom (DoF), each with six variations on FGaBP inner iterations and three subspace sizes. The reduction ratios are obtained by dividing the total number of FGaBP iterations by itself with relaxation [4] over the total iterations of the accelerated method, which indicate the reductions on total Floating Point Operations (FLOPs). The IR-FGaBP obtained the highest ratios on all experiments showing a growing trend of reduction ratios with increasing DoFs. As DoFs increase the FGMRES-FGaBP benefits from more inner FGaBP iterations, but stagnates rapidly compared to the IR-FGaBP.

V. Conclusion

The highly parallel FGaBP algorithm was demonstrated to also be amicable for acceleration using variants of Krylov methods resulting in considerable iteration reductions. The IR method showed considerably higher iteration reductions than FGMRES preconditioning, without considerable impact on the scalability of the combined algorithms. The long version paper will include more detailed analysis and results of this approach along with an extension to FMGaBP in order to address large scale problems. Also, other residual minimization schemes can be explored for the IR method such as Gram-Schmidt and Householder orthogonalizations.

References

[1] Y. El-Kurdi, W. Gross, and D. Giannacopoulos, "Parallel solution of the finite element method using Gaussian belief propagation," in *The 15th Biennial IEEE Conference on Electromagnetic Field Computation*, 2012, p. 141.

[2] ——, "Parallel multigrid acceleration for the finite-element Gaussian belief propagation algorithm," *Magnetics, IEEE Transactions on*, vol. 50, no. 2, pp. 581–584, Feb 2014.

[3] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[4] Y. El-Kurdi, D. Giannacopoulos, and W. Gross, "Relaxed Gaussian belief propagation," in *IEEE Int. Symp. on Inform. Theory (ISIT)*, July 2012, pp. 2002 –2006.

[5] U. Trottenberg, C. Oosterlee, and A. Schüller, *Multigrid*. Academic Press, 2001.

[6] T. Washio and C. W. Oosterlee, "Krylov subspace acceleration method for nonlinear multigrid schemes," *ETNA, Electron. Trans. Numer. Anal*, vol. 6, pp. 271–290, 1997.

[7] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia, PA: SIAM, 2003.