Review report of "Point spread function approximation of high rank Hessians with locally supported non-negative integral kernels"

Many Hessian-based computational algorithms have been developed to solve large-scale inverse problems constrained by partial differential equations (PDEs) in the last decade, especially in the context of inferring infinite-dimensional parameters. However, it remains a critical computational challenge to solve such inverse problems when the Hessian of the data-misfit term does not have fast spectrum decay, or the Hessian is high-rank. This challenge often leads to the need to solve a considerable number of PDEs in many (preconditioned) CG iterations with an inexact Newton-CG algorithm, one of the most advanced optimization algorithms in solving infinite-dimensional inverse problems.

To address this challenge, this paper proposes an efficient preconditioned CG method with the preconditioner computed as an approximate Hessian at some suitable parameter point, which is demonstrated to significantly reduce the number of required CG iterations to achieve given accuracy. The key novelty of this paper is the development of an efficient hierarchical matrix approximation of the Hessian using only a small number of Hessian matrix-vector products, thus a small number of PDE solves. This novelty is made possible by exploiting the property of the Hessian operator with locally supported non-negative integral kernels. Specifically, the authors propose a point spread function (PSF) approximation of the Hessian by (1) computing the zeroth, first, and second-order impulse response moments of the Hessian by applying its product with constant, linear, and quadratic functions, (2) building local ellipsoid support estimate of the impulse response functions of the Hessian based on these moments, (3) selecting sample points of the impulse response from a candidate set by a greedy ellipsoid packing algorithm, (4) computing the impulse responses with disjoint ellipsoid in batches by applying the Hessian to Dirac combs, which plays a key role in reducing the total number of Hessian matrix-vector product, (5) approximating any integral kernel entries by a radial basis interpolation based on the computed impulse responses, and (6) building the hierarchical matrix approximation of the Hessian with the radial basis interpolation of the integral kernel entries before applying proper symmetrizing and flipping of the negative eigenvalues to make the approximate matrix symmetric positive semi-definite.

The PSF approximation in this paper is built for more general operators than Hessian. It demonstrates the effectiveness of the PSF-based method by using it to build preconditioners for the Hessian operator in two challenging inverse problems of basal friction coefficient inversion of ice sheet flow and initial condition inversion of advection-dominated transport. It shows significant reductions (5-10X) in the required number of PDE solves compared to classical regularization-based preconditioning and no preconditioning. The paper also presents a comprehensive numerical study on the influence of various parameters (data noises, # batchs, diffusion coefficients, terminal times) on the effectiveness and data-scalability of the proposed method, showing that the PSF-based preconditioners can form good approximations of high-rank Hessians using only a small number of operator applications.

The proposed method is very interesting and makes a great contribution to solving large-scale inverse problems with high-rank Hessians. The presentation of the proposed method is concise and illustrative. The numerical results are very convincing. Overall, the paper is very well written.

I recommend to accept the paper for publication with minor revisions. I suggest the authors properly address the following questions to help for its better understanding and broader applications.

- The author mentioned one limitation of the proposed method: the Hessian should have a local non-negative integral kernel, which is not satisfied for wave inverse problems that lead to a substantial amount of negative entries of the Hessian. It would be interesting if the author could elaborate more details with intuition or numerical evidence for this property and the limitation, especially if you also use Gauss–Newton approximate Hessian for the wave inverse problems as in this paper.

- It is interesting to see the effectiveness of the approximate Hessian used as a preconditioner. It would also be very interesting to see how accurate the approximate Hessian is compared to a full Hessian in terms of the number of batches, ellipsoid sizes, and in particular the total number of impulse responses. Can you say anything about the convergence of the approximation, either numerical or theoretical?

- The paper presented a nice complexity analysis of the steps in constructing the approximate Hessian, with the dominating cost arising from the Hessian matrix-vector product. It would be interesting to see the computational time of each step for the numerical examples to provide a good sense of the complexity and to support the 5-10X computational reduction by accounting also the overhead beyond the PDE solves.