An Efficient Matrix-free Ensemble Kalman Filter

Humberto C. Godinez, J. David Moulton, T-5

In this work we describe an efficient matrix-free implementation of the EnKF for the assimilation of large data sets. To reduce complexity and the cost of assimilation of large data sets, we propose a matrix-free implementation of the EnKF where a Sherman-Morrison solver, which exploits the outer-product form of the matrix, is used. The cost of the resulting algorithm scales linearly in the number of observations (data) being assimilated. Numerical experiments with a 2D SW model on the sphere confirm this scaling. In addition, these results show that this approach is significantly faster than the popular algorithm developed by Evensen [1], which is based on the SVD.

The Ensemble Kalman Filter (EnKF) is a sequential data assimilation method. It uses an ensemble of model integrations or forecasts to calculate the background mean and error covariance matrix needed to compute an improved model state. Hence it is a computationally feasible method for nonlinear models whose dependent variables, or state, are high-dimensional. Due to its ease of implementation and effectiveness it has gained wide acceptance.

The assimilation step in EnKF computes a correction term, which is a weighted difference between the model state and observations. The formulation of the weight matrix, referred to as the Kalman gain matrix, involves an inverse matrix whose dimension is equal to the number of observations being assimilated. Fortunately, the inverse is not needed explicitly, so common approaches, such as Cholesky decomposition and singular value decomposition (SVD), are used to factor the matrix, and then solve the corresponding linear system; however, these approaches scale superlinearly with the number of observations, making them too costly for large data sets.

A popular algorithm that partially addresses this issue of large data sets was proposed by Evensen [1]. In this algorithm an additional assumption is introduced that significantly reduces the cost of the SVD. In particular, this approach consists of two parts: an SVD that reportedly scales linearly in the number of observations, followed by four dense matrix multiplications. Although these dense matrix operations do scale linearly with the number of data points, their scaling with respect to the number of ensemble members and the number of degrees of freedom in the model (state) vector makes the overall cost unacceptably high.

A promising alternative to these traditional approaches is to solve the underlying system using a matrix-free approach that exploits the particular form of the matrix. Specifically, this matrix is expressed as the sum of a diagonal matrix plus a series of rank-one matrices. Recently, Egidi and Maponi [2] developed a powerful direct solver for matrices of this form. This algorithm recursively applies the well-known Sherman-Morrison and, hence, is completely matrix-free. We propose to use this solver to develop an efficient matrix-free EnKF assimilation algorithm. This approach will minimize the memory requirements because neither the overall matrix nor the matrices in the Sherman-Morrison formula need to be formed. Only the action of these matrices times a vector is required, and this action is readily expressed as a combination of inner products.

Furthermore, much like the factorization and decomposition approaches, the Sherman-Morrison solver can be thought of as having two distinct parts. In particular, the first solve is the most expensive part, and is analogous to the factorization or decomposition step. In the second step, subsequent solves with the same linear system for an efficient matrix-free EnKF, multiple right-hand-sides have a cost proportional to a vector dot product. This is analogous to solving a linear system once the matrix has been factored. The ability of the Sherman-Morrison solver to perform subsequent solves very efficiently is critical because the assimilation step in the EnKF solves the same linear system for each

For more information contact J. David Moulton at moulton@lanl.gov.

member of the ensemble of model states. In combination, these properties lead to an efficient matrix-free EnKF algorithm that scales linearly in the number of observations being assimilated.

To compare scaling and performance, the SVD-based EnKF algorithm developed by Evensen [1] and the proposed matrix-free EnKF algorithm were implemented within a single data assimilation framework. A global 2D shallow-water (SW) model on a sphere was used to provide the model predictions. The assimilation was performed under a twin experiment framework, where the initial condition for the control state was taken by shifting the reference initial condition one gridpoint to the east. An ensemble of 100 members was formed, where each ensemble member was initialized with a perturbed initial condition. Then EnKF assimilation for each method was performed with the number of data points varying from 100 to 30,000. Further details and references can be found in [3].

Although both methods produced similar assmilated states, the results confirm that the proposed matrix-free EnKF algorithm is significantly faster than the SVD-based approach. In Fig. 1 we plot the CPU times for the two stages of each algorithm as a function of the number of data points being assimilated. The times in the first plot (Fig. 1, top) show that the cost and scaling of the first step in both algorithms is comparable. In contrast, the times in the second plot (Fig. 1, bottom) show that the second stage of the proposed matrix-free EnKF approach is significantly more efficient than the SVD-based approach. This increased cost is directly attributed to the dense matrix multiplications used here, and gives a very steep slope to the linear dependence on the number of observations. In fact, the matrix-free EnKF is over 100 times faster when assimilating 30,000 observations.

The results show that the matrix-free implementation, using a Sherman-Morrison linear solver, is an efficient alternative to matrix decomposition methods by a wide margin. Future work will focus on the parallelization of this matrix-free approach, as well as the development of a proper localization scheme to prevent filter divergence.

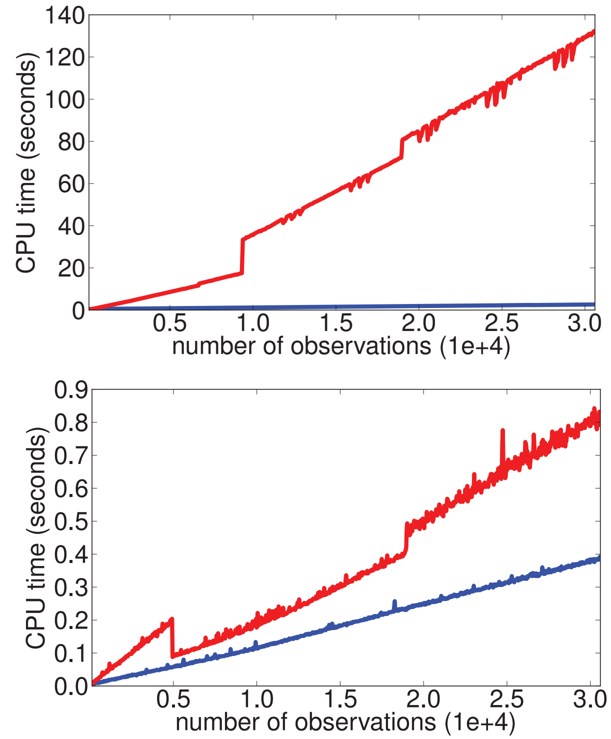

Fig. 1. Timing comparison between the matrix-free (blue line) and SVD (red line) implementations of EnKF as a function of the number of observations. The top figure shows the CPU time of the matrix-free first solve (blue line) and the SVD computation (red line). The bottom figure shows the CPU time for the subsequent matrix-free solves (blue line) and the matrix multiplications (red line) in Evensen's SVD method. The plots clearly show the efficiency and linear scaling of the matrix-free EnKF algortihm.

[1] Evensen, G., *Ocean Dynam* **53,** 343 (2003).
[2] Egidi, N. and P. Maponi, *J Comput Appl Math* **189,** 703 (2006).
[3] Godinez, H. and J. Moulton, "An Efficient Matrix-free Implementation of the Ensemble Kalman Filter," in preparation.

Funding Acknowledgments
DOE, Office of Science, Advanced Computing Research (ASCR) program in Applied Mathematical Sciences