Identifying the Optimal Measurement Subspace for the Ensemble Kalman Filter

N. Zhou, Z. Huang, G. Welch, and J. Zhang

To reduce the computational load of the ensemble Kalman filter while maintaining its efficacy, an optimization algorithm based on the generalized eigenvalue decomposition method is proposed for identifying the most informative measurement subspace. When the number of measurements is large, the proposed algorithm can be used to make an effective trade-off between computational complexity and estimation accuracy.

Introduction: The ensemble Kalman filter (EnKF) [1] is an important tool for estimating dynamic states of a non-linear system. The computational load for a Kalman filter (KF) is significant when the number of measurements is large [2]. This computational problem is exacerbated for an EnKF because it uses a collection of samples (referred as ensembles) to represent and propagate uncertainty. To reduce the computational load of an EnKF while maintaining its efficacy, this paper proposes an optimization algorithm to identify the most informative measurement subspace. The paper is structured as follows. First, the EnKF algorithm is briefly reviewed. Next, the major computational load is identified and the problem of identifying the most informative measurement subspace is formulated to reduce the computational complexity. Finally, a solution based on the generalized eigenvalue decomposition method is proposed, and its properties in balancing computational complexity and estimation efficacy are discussed.

Review of EnKF: A non-linear system can be described by a discrete state space model as in (1).

$$\begin{cases} \tilde{x}_k = \tilde{f}(\tilde{x}_{k-1}) + \tilde{w}_{k-1} \\ \tilde{z}_k = \tilde{h}(\tilde{x}_k) + \tilde{v}_k \end{cases} \tag{1}$$

where $\tilde{x}_k$ is a state vector at time step k; $\tilde{z}_k$ is a measurement vector; $\tilde{f}$ is a system transition function; $\tilde{h}$ is a measurement function; and the vectors $\tilde{w}_{k-1}$ and $\tilde{v}_k$ represent the process and measurement noise respectively. When $\tilde{h}$ is non-linear, (1) is usually transformed into (2) which has a linear measurement model to facilitate computation [1]. The transformation can be done by augmenting the model states as $x_k = \left[\tilde{x}_k^T \quad \tilde{h}^T(\tilde{x}_k)\right]^T$ [1].

$$\begin{cases} x_k = f(x_{k-1}) + w_{k-1} \\ z_k = Hx_k + v_k \end{cases} \tag{2}$$

Here, $x_k \in \mathrm{R}^{n\times 1}$; $f : \mathrm{R}^{n\times 1} \to \mathrm{R}^{n\times 1}$; $z_k \in \mathrm{R}^{m\times 1}$; $H \in \mathrm{R}^{m\times n}$ is a measurement matrix; and $w_{k-1} \in \mathrm{R}^{n\times 1}$ and $v_k \in \mathrm{R}^{m\times 1}$ follow Gaussian white noise assumptions (i.e., $w_k \sim N(0,Q)$ and $v_k \sim N(0,R)$).

After initialization, an EnKF recursively estimates the states through prediction and correction steps. In the prediction step, states are propagated to the next time step using (3),

$$\hat{x}_k^-[i] = f\left(\hat{x}_{k-1}[i]\right) + \hat{w}_{k-1}[i] \tag{3}$$

where the index '[i]' is for the ith member of the total N ensembles, the circumflex (^) indicates estimation, and the superscript '-' indicates the *a priori* states. Unlike the traditional KF, the *a priori* error covariance $\hat{P}_k^- \in \mathrm{R}^{n\times n}$ for an EnKF does not have to be propagated explicitly because it can be calculated from the ensembles of states using $\hat{P}_k^- = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{x}_k^-[i] - \bar{x}_k^-\right)\left(\hat{x}_k^-[i] - \bar{x}_k^-\right)^T$. Here, $\bar{x}_k^- = \frac{1}{N}\sum_{i=1}^{N}\hat{x}_k^-[i]$.

In the correction step, each state member is updated by assimilating information from the measurement data using (4).

$$\hat{x}_k[i] = \hat{x}_k^-[i] + \hat{P}_k^- H^T (H\hat{P}_k^- H^T + \hat{R})^{-1}(z_k[i] - H\hat{x}_k^-[i]) \tag{4}$$

where $z_k[i]$ is the perturbed measurement defined as $z_k[i] \overset{\Delta}{=} z_k + v_k[i]$; and $\hat{R}$ is the estimate of R. The objective of the KF is to minimize the trace of the *a posteriori* covariance [3]. Equation (5) shows how the *a posteriori* covariance $\hat{P}_k$ is related to the *a priori* covariance $\hat{P}_k^-$.

$$\hat{P}_k = \hat{P}_k^- - \hat{P}_k^- H^T \left(H\hat{P}_k^- H^T + \hat{R}\right)^{-1} H\hat{P}_k^- \tag{5}$$

Problem Definition: Equation (4) can be computationally intensive when a model has a large number of measurements. As pointed out by [1], "If the number of measurements is larger than the number of ensemble members, the matrices $H\hat{P}_k^- H^T$ and $\hat{R}$ will be singular and a pseudo inversion must be used." Computing the pseudo inverse of $H\hat{P}_k^- H^T + \hat{R}$ in (4) is the dominant computational load in the EnKF algorithm, because the computational complexity of the pseudo inverse is $O(\tfrac{27}{4}m^4)$ [4].

To reduce the computational load, this paper proposes to formulate a measurement selection problem as described below. Given measurements $z_k \in \mathrm{R}^{m\times 1}$, we construct a new set of measurements $y_k \in \mathrm{R}^{d\times 1}$. These measurements are linear combinations of z_k and have a lower dimension (i.e., d<m), so that when y_k is used in EnKF (instead of z_k), the trace of the *a posteriori* covariance matrix (as defined in (5)) is minimized. Because y_k has a lower dimension than z_k, the computational complexity of the pseudo inverse can be reduced to $O(\tfrac{27}{4}d^4)$ from $O(\tfrac{27}{4}m^4)$.

Specifically, the problem can be defined as follows. The newly constructed measurements y_k can be written as (6).

$$y_k \overset{\Delta}{=} C^T z_k = C^T H x_k + C^T v_k \tag{6}$$

where $C \in \mathrm{R}^{m\times d}$ is a matrix to be determined. Because $v_k \sim N(0,R)$, it follows that $C^T v_k \sim N(0, C^T R C)$. Therefore, if y_k is used as the measurements instead of z_k, the covariance matrix of the *a posteriori* states can be calculated using (7).

$$\hat{P}_k(C) = \hat{P}_k^- - \hat{P}_k^- (C^T H)^T \left\{C^T H\hat{P}_k^- H^T C + C^T \hat{R} C\right\}^{-1} C^T H\hat{P}_k^- \tag{7}$$

Consistent with the KF objective, C can be determined by solving the optimization problem defined by (8):

$$\min_{C}\ tr\left\{\hat{P}_k(C)\right\} \tag{8}$$

where the notation 'tr' stands for the matrix trace operator.

Solution: Because $tr(A-B) = tr(A) - tr(B)$ and $tr(AB) = tr(BA)$, (8) and (9) are equivalents.

$$\max tr\left\{\hat{P}_k^- - \hat{P}_k(C)\right\} = \max tr\left\{\frac{C^T H\hat{P}_k^- \hat{P}_k^- H^T C}{C^T (H\hat{P}_k^- H^T + \hat{R}) C}\right\} \tag{9}$$

Note that (9) is a generalized Rayleigh-Ritz quotient problem. Therefore, its solution is the subspace spanned by the first 'd' number of the dominant eigenvectors of the matrix pair defined by (10) [5].

$$\left\{H\hat{P}_k^- \hat{P}_k^- H^T, \quad H\hat{P}_k^- H^T + \hat{R}\right\} \tag{10}$$

More specifically, assume that the generalized eigenvalues of the matrices in (10) are $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_\mu > \lambda_{\mu+1} = \ldots = \lambda_m = 0$, whose corresponding eigenvectors are $u_1, u_2, \cdots, u_m$ (i.e., $\left(H\hat{P}_k^- \hat{P}_k^- H^T\right)u_i = \lambda_i\left(H\hat{P}_k^- H^T + \hat{R}\right)u_i$). Here, the symbol 'μ' is the total number of the non-zero eigenvalues and $\mu \leq \min(m,n)$. Then, the solution to (9) is $C^* = \left[u_1 \quad u_2 \quad \cdots \quad u_d\right]K$. Here, $K \in \mathrm{R}^{d\times d}$ can be any full-rank constant matrix. In addition, for the optimal solution C*, the objective function of (9) takes the value of (11).

$$tr\left\{\hat{P}_k^- - \hat{P}_k(C^*)\right\} = tr\left\{\frac{C^{*T} HP^- P^- H^T C^*}{C^{*T}(HP^- H^T + \hat{R})C^*}\right\} = \sum_{j=1}^{d}\lambda_j \tag{11}$$

Discussion: To implement the proposed method, it is required to calculate 'd' dominant eigenvalues and eigenvectors for the matrices in (10), whose one-step computational complexity is $O(d^2 m)$ using the Arnoldi method [6]. In addition, using the measurement subspace only requires inverting a matrix of size dXd, whose computational complexity is $O(27/4\, d^4)$. Therefore, when d<<m, the computational complexity using the proposed measurement subspace is much less than that of the original matrix pseudo inverse, i.e. $O(27/4\, m^4)$. Note that the reduction of computational complexity may come at the cost of reduced estimation efficacy. The relative estimation efficacy of y_k^* can be defined as $eff\left(y_k^*\right) \overset{\Delta}{=} \frac{tr\left\{\hat{P}_k^- - \hat{P}_k\left\langle y_k^* \right\rangle\right\}}{tr\left\{\hat{P}_k^- - \hat{P}_k\left\langle z_k \right\rangle\right\}} \times 100\%$. As indicated by (11), when $y_k^* = C^{*T} z_k \in \mathrm{R}^{d\times 1}$ is used for the measurements, $tr\left\{\hat{P}_k^- - \hat{P}_k\left\langle y_k^* \right\rangle\right\} = \sum_{j=1}^{d} \lambda_j$. In comparison, when $z_k \in \mathrm{R}^{m\times 1}$ is used, $tr\left\{\hat{P}_k^- - \hat{P}_k\left\langle z_k \right\rangle\right\} = \sum_{j=1}^{m} \lambda_j = \sum_{j=1}^{\mu} \lambda_j$. Therefore, to retain 100% estimation efficacy, 'd' needs to be greater than or equal to 'µ'. The estimation efficacy will be less than 100% when d<µ.

Simulations were performed using MATLAB 2010a™. The function 'pinv' was used to calculate a pseudo-inversion, and the function 'eigs' was used to perform the generalized eigenvalue decomposition. To perform the calculations, we used a laptop computer with a 2.5-GHz CPU, and 8 GB of memory. For m=3000, n=300, µ=30, the computation time and estimation efficacy for the original pseudo inversion method, Goris method [2], and the proposed method (including the generalized eigenvalue decomposition and the reduced-size pseudo inverse) are compared in Table 1 for different number of measurements. Note that compared with the original method, both Goris method and the proposed method can significantly reduce the computation load by constructing a new set of measurements of lower dimension. To maintain 100% estimation efficacy, Goris method reduces the number of measurements to 'n'. In comparison, the proposed method can reduce the dimension to 'µ' (note that µ≤min(m,n)). In addition, the proposed method can further reduce the number of measurements to a value less than µ if the estimation efficacy is allowed to be less than 100%. Also note that when d=300, Goris method uses less computation time than the proposed method, which indicates that Goris method is computationally more efficient when the number of measurements is same.

Table 1: The method comparison

	# of Meas (d)	Comp Time (s)	$eff\left(y_k^*\right)$
Org Method	3000	259.95	100%
Goris Method	300	5.30	100%
Proposed Method	300	18.80	100%
	100	6.33	100%
	30	2.68	100%
	20	2.29	99%
	15	2.06	83%

Conclusion: A new algorithm is developed to identify the most informative measurement subspace for an EnKF. When the number of measurements is large and number of non-zero eigenvalues is small, the proposed algorithm can help users make a well-informed trade-off between computational complexity and estimation efficacy.

N. Zhou, Z. Huang (*Pacific Northwest National Laboratory, P.O. Box 999, MSIN K1-85, Richland, WA 99352, USA*)
G. Welch, J. Zhang (*The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA*)
G. Welch is also with The University of Central Florida, Orlando, FL 32816, USA

E-mail: ning.zhou@pnnl.gov

References

1. Evensen, G.,: 'The ensemble Kalman filter: Theoretical formulation and practical implementation', Ocean Dyn., 2003, 53, pp. 343-367

2. Goris, M.J., Gray, D.A., and Mareels, I.M.Y.: 'Reducing the computational load of a Kalman filter', Electron. Lett., 1997, 33, (9), pp. 1539-541

3. Welch, G., and Bishop, G.: 'An introduction to the Kalman filter.' Technical Report 95-041, University of North Carolina, Department of Computer Science, Chapel Hil, North Carolina, 1995

4. Hassibi, B.: 'An efficient square-root algorithm for BLAST', ICASSP, 2000, 2, pp. 737-740

5. Overton, M., and Womersley, R.: 'On the sum of the largest eigenvalues of a symmetric matrix', SIAM J. Matrix Anal. Appl., 1992, pp. 41-45

6. Lee, J., Balakrishnan, V., Koh, C., and Jiao, D.: 'From O(k2N) to O(N): A fast complex-valued eigenvalue solver for large-scale on-chip interconnect analysis', IEEE Trans. Microwave Theory Tech., 2009, 57, (6), pp. 3219–3228