# Optimal State Estimation Using Balanced Truncation

**Brian F. Farrell[1] and Petros J. Ioannou[2]**

*[1]Harvard University, Cambridge, Massachusetts*
*[2]Athens University, Athens, Greece*

## 1. Introduction

An important component of forecast error is error in the analysis of the initial state from which the forecast is made. Analysis error can be reduced by taking more observations, by taking more accurate observations, by taking observations at locations chosen to better constrain the forecast, and by extracting more information from the observations that are available. The last of these, obtaining the maximum amount of information from observations, is attractive because it makes existing observations more valuable and because, at least for linear systems,there is a solution to the problem of extracting the maximum information from a given set of observations: under appropriate assumptions the problem of extracting the maximum amount of information from a set of observations of a linear system in order to minimize the uncertainty in the state estimate is solved by the Kalman filter (KF) (Kalman, 1960,Ghil and Malanotte-Rizzoli, 1991).

Unfortunately, the Kalman filter requires statistical description of the forecast error in the form of the error covariance and obtaining the required error covariance involves integrating a system with dimension equal to the square of the dimension of the forecast system. Direct integration of a system of such high dimension is not feasible. Attempts to circumvent this difficulty have involved various approximations to the error covariance (Bishop et al, 1999; Tippett et al, 2000) and approximate integration methods (Evensen, 1994; Dee, 1995; Fukumori and Malanotte-Rizzoli, 1995; Cohn and Todling, 1996; Verlaan and Heemink, 1997; Houtekamer and Mitchell, 1998).

While the formal dimension of the forecast error system obtained by linearizing the forecast model about a base trajectory is the same as that of the forecast system itself, there are reasons to believe that the effective dimension is far lower.In the case of the tangent linear forecast error system the spectrum of optimal perturbations of the error propagator over the forecast interval typically comprise a few hundred growing structures and Lyapunov spectra for error growth have shown similar numbers of positive exponents suggesting that the effective dimension of the error system is O(103).

The problem of reducing the order of a linear dynamical system can be cast mathematically as that of finding a finite dimensional representation of the dynamical system so that the Eckart-Schmidt-Mirsky (ESM) theorem can be applied to obtain an approximate truncated system with quantifiable error. The ESM theorem states that the optimal k order truncation of an n dimensional matrix in the euclidean or Frobenius norm is the matrix formed by truncating the singular value decomposition of the matrix into its first k singular vectors and singular values. A method for exploiting the ESM theorem to obtain a reduced order approximation to a dynamical system was developed in the context of controlling lumped parameter engineering systems and is called balanced truncation (Moore, 1981; Glover, 1984). Balanced truncation was applied to the set of ordinary differential equations approximating the partial differential equations governing perturbation growth in time independent atmospheric flows by Farrell and Ioannou (2001)(FI01).

In this work a reduced order Kalman filter is derived based on balanced truncation and applied to a time dependent Lyapunov unstable quasi-geostrophic model of a forecast tangent linear error system. We first review the method of balanced truncation and then use it to construct the reduced order Kalman filter

## 2. Reduction by balanced truncation

The error dynamics are assumed to be governed by the linear system:

$$\frac{d\psi}{dt} = A\psi$$

where $\psi$ is the error state vector and $A$ is the matrix dynamical operator which may be time dependent, but will for the time being be assumed to be time independent. Because of the high dimension of the error system in forecast applications we are interested in exploring the accuracy of reduced order approximations to this system.

Before proceeding with the method of balanced order reduction we must first choose the norm that will be used to measure the accuracy of the approximation. The accuracy is measured by the norm of the euclidean length of the errors incurred in a chosen variable. This norm is the square root of the Euclidean inner product in this variable. If another norm is selected to measure the accuracy of the approximation then the most direct method of accounting for this choice is to transform the variable used to represent the state of the system so that the Euclidean inner product in the transformed variable corresponds to the new norm. The reduced order approximate system resulting from balanced truncation will in general depend on the norm. As discussed in FI01, optimal order reduction for stable normal systems is immediate: it is a Galerkin model based on projection of the dynamics onto the least damped modes. Difficulties in the reduction process arise in cases for which the system is non-normal in the variable corresponding to the chosen norm. Then a model based on projection on the least damped modes is sub optimal and the reduction must proceed by including in the retained subspace the distinct subspaces of the preferred excitations and preferred responses of the system.

The preferred structures of response of a non-normal system are revealed by stochastically forcing the system with spatially and temporally uncorrelated unitary forcing and calculating the eigenfunctions of the resulting mean covariance matrix $P = \langle \psi\psi^{+} \rangle$ (the brackets denote an ensemble average, and $+$ the hermitian transpose of a vector or a matrix). The covariance matrix under such forcing is given by:

$$P = \int_{0}^{\infty} e^{At} e^{A^{+}t} dt$$

and this integral is readily calculated by solving the Lyapunov equation (FI96):

$$AP + PA^{+} = -I$$

which $P$ satisfies, as can be easily verified. The hermitian and positive definite matrix $P$ characterizes the response of the system and its orthogonal eigenvectors, ordered in decreasing magnitude of their eigenvalue, are the empirical orthogonal functions (EOF's) of the system under spatially and temporally uncorrelated forcing.

The preferred structures of excitation of the system are determined from the stochastic optimal matrix:

$$Q = \int_{0}^{\infty} e^{A^{+}t} e^{At} dt$$

the orthogonal eigenvectors of which, in decreasing magnitude of their eigenvalue, order the forcing structures according to their effectiveness in producing the statistically maintained variance. The eigenvectors of $Q$ are called the stochastic optimals (SO's) and because of the non-normality of the system are distinct from the EOF's. The stochastic optimal matrix $Q$ satisfies the back Lyapunov equation:

$$A^{+}Q + QA = -I$$

The Lyapunov equations for $P$ and $Q$ have unique positive definite solutions if $A$ is stable. The covariance matrix $P$ and stochastic optimal matrix $Q$ need to be determined in order to proceed with order reduction by balanced truncation.

A successful order reduction must accurately approximate the dynamics of the system which can be expressed as the mapping of all past (square integrable) forcings to all future responses of the system. This linear mapping of inputs to outputs is called the Hankel operator. Application of the ESM theorem to the Hankel operator provides the optimal low order truncation of the dynamics. Remarkably, because of the separation between past forcings and future responses in the Hankel operator representation of the dynamics this operator has finite rank equal to the order of the system; its singular values, denoted by h, are the square root of the eigenvalues of the product of the covariance and stochastic optimal matrix $P$ and $Q$.

The balanced truncation transforms the internal coordinates of the system so that the transformed covariance matrix $P$ and stochastic optimal matrix $Q$ become identical and diagonal (while preserving the inner product of the physical variables). The dynamical system is then truncated in these transformed balanced coordinates. The balanced truncation retains a leading subset of empirical orthogonal functions (EOF's) and stochastic optimals (SO) of the dynamical system and preserves the norm.

Balanced truncation preserves the stability of the full system and provides an approximation with known error bounds which is found in practice to be nearly optimal (Moore, 1981; Glover, 1984; FI01)

# 3. The reduced order Kalman filter

In the previous section we showed how to reduce the order of an autonomous linear system by obtaining an accurate balanced truncation of the time independent operator $A$. Consider now a time dependent operator of the form $A(t) = A_0 + A_1(t)$, and assume that the mean operator $A_0$ is dominant. In previous work we showed that perturbation growth in time dependent non-normal systems occurs primarily in the non-normal subspace of the time mean operator (Farrell and Ioannou, 1999). In the following we take advantage of this result to obtain an approximate reduced order model of a time dependent system by reducing the order of the time dependent operator using the balancing transformation derived for the mean operator. This procedure is found to produce an accurate model order reduction for example non-autonomous stationary systems at far less computational cost than is incurred by balancing at each time. In the following the error covariance matrix and the Kalman gain are obtained using this reduced model and transformed back to the full space for use in updating the state estimate.

Let the dimension of the full system perturbation vector, $\psi$, be N, and of the reduced system, $\psi_k$, be k with k < N. The variables in the reduced system, $\psi_k$, are related to the variables in the full system, $\psi$, by the transformation

$$\psi_k = Y^{+}\psi .$$

The evolution equation in the $\psi_k$ coordinates is

$$\frac{d\psi_k}{dt} = A_k(t)\psi$$

where

$$A_k(t) = Y^+ A(t) X .$$

In this approximation the biorthogonal bases $Y$ and $X$ remain the balancing transformation of the mean operator $A_0$ instead of being recalculated at each step.

In the reduced order system variables the observation matrix is $H^k = HX$, so that the error covariance matrix in the reduced order system $P^k(t_i) = Y^+ P^k(t_i) Y$, is evolved according to the reduced order dynamics:

$$P^{kf}(t_{i+1}) = M^k(t_i) P^{ka}(t_i) M^{k^T}(t_i) + Q^k$$

where $Q^k$ is the model error covariance projected on the reduced order space, i.e. $Q^k = Y^+ Q Y$, and $M^k$ is the reduced order propagator. The reduced order covariance matrix is corrected by the observation statistics:

$$P^{ka}(t_{i+1}) = P^{kf}(t_{i+1}) - K^k(t_{i+1}) H P^{kf}(t_{i+1})$$

in which the reduced order Kalman gain is

$$K^k(t_{i+1}) = P^{kf}(t_{i+1}) H^{k^T} \left( H^k P^{kf}(t_{i+1}) H^{k^T} + R^k \right)^{-1}$$

where $R^k = Y^+ R Y$ .

The Kalman gain that will be used in the full system in order to entrain the observations is obtained from the reduced order system gain:

$$K^r = X K^k$$

with the superscript indicating that the Kalman gain is obtained from the reduced order system.

A model of a time dependent tangent linear system with dimension 400 is formed by varying the zonal flow in an Eady channel model according to Fig. 1. This model is Lyapunov unstable and the performance of state estimation methods is compared in the four panels of Fig. 2 where we show the true state (top panel), and the estimates of the state based on the full Kalman, the order 40 reduced order Kalman and the estimate of the state by direct substitution, all at time $t$ =150. Both the Kalman filter based on integrating the full error covariance matrix and the order 40 reduced order Kalman filter give a good estimate of the true state, while direct substitution fails.

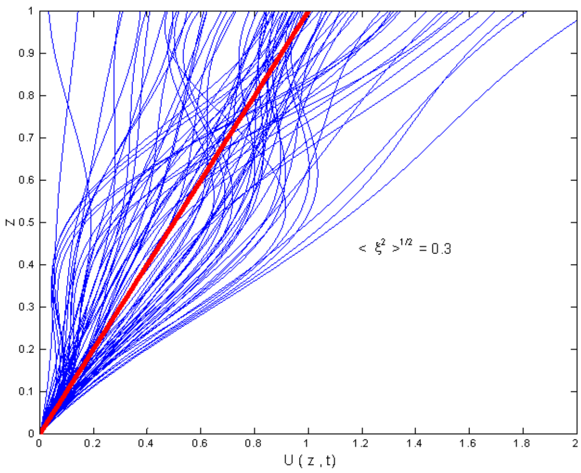


*Figure 1: Realization of the time dependent velocity. The mean is a constant shear wind, and the fluctuations have r.m.s amplitude* 0.3.

## 4. Conclusions

Optimal utilization of observing resources requires that the structure of the time dependent error is taken into account in identifying the state. The error covariance matrix contains the required information but the high dimension of the forecast system precludes directly obtaining it. In this work we described a method for obtaining an approximate error covariance and an approximate state identification using a Kalman filter based on balanced truncation of the tangent linear forecast error system.

Comparison of the performance of a full Kalman filter and approximate filters obtained by balanced truncation on the order 400 storm track model reveals that truncation at order 40 is sufficient to provide accurate flow dependent covariances for the purpose of approximating the Kalman gain.

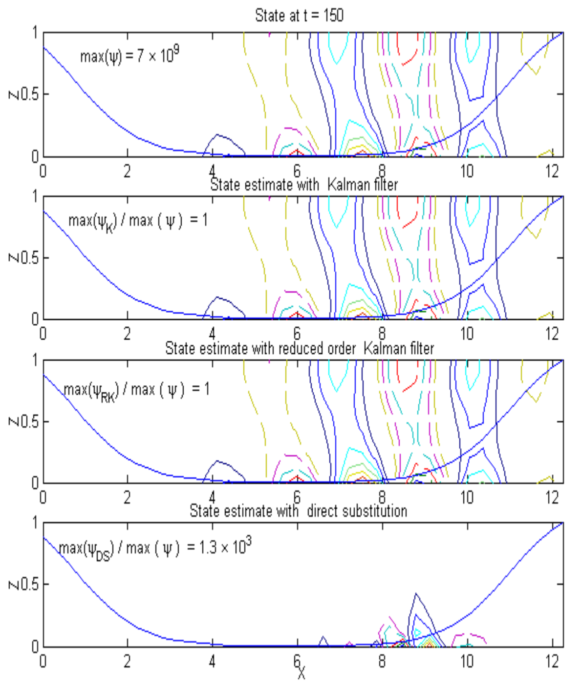


*Figure 2: First panel: streamfunction of the state at $t$=150. The state is growing with Lyapunov exponent $\lambda$=0.075. Second panel: streamfunction of the analyzed state obtained using a full Kalman filter. Third panel: streamfunction of the analyzed state obtained using a Kalman filter calculated from a balanced truncation of the full system to 40 degrees of freedom. Fourth panel: streamfunction obtained by direct substitution. Observations are made at one location $x$=4, $z$=0.8.*

This work was supported by NSF ATM-0123389 and by ONR N00014-99-1-0018.

## 5. References

Bishop, C. H., B. J. Etherton, and S. J. ajumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects., *Mon. Wea. Rev*., **129**, 420-436.

Cohn, S. E. and R. Todling, 1966: Approximate Kalman filters for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **75**, 257-288.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. Mon. Wea. Rev., **123**, 1128-1145.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res*., **99 (C5)**, 10143-10162.

Farrell, B. F., and P. J. Ioannou, 1999: Perturbation growth and structure in time dependent flows., *J. Atmos. Sci*., **56**, 3622-3639.

Farrell, B. F., and P. J. Ioannou, 2001: Accurate Low Dimensional Approximation of the Linear Dynamics of Fluid Flow. *J. Atmos. Sci***., 58**, 2771-2789.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation; an example with an idealized Gulf Stream model. *, J. Geophys. Res*. Oceans, **100**, 6777-6793.

Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. in Geophys*., **33**, 141-266.

Glover, K., 1984: An optimal Hankel-norm approximation of linear multivariable systems and their $L^{\infty}$ error bounds*. Int. J. Control*, **39**, 1115-1193.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique, *Mon. Wea. Rev*., **126**, 796-811.

Kalman, R.E., 1960: A new approach to linear filtering and prediction problems. J. Basic Eng., **82D**, 35-45.

Moore, B. C., 1981: Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. on Automatic Control,* **AC-26**, 17-31.

Tippett, M. K., S. E. Cohn, R. Todling, and D. Marchesin, 2000: Low-dimensional representation of error covariance. *Tellus*, **52** 533-553.

Verlaan, M. and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics***, 11**, 349-368.