UTILIZATION OF THE ENSEMBLE KALMAN FILTER: AN OPTIMAL CONTROL LAW FOR THE ADAPTIVE OPTICS OF THE E-ELT

M. Gray[1] and B. Le Roux[1]

Abstract. Adaptive Optics (AO) systems require the implementation of techniques intended for real time identification of atmospheric turbulence. Nowadays there are several approaches. One of them is using the Kalman Filter (KF) and presents numerous advantages at the level of optimal control. However it will be impossible to install this process within the frame of an AO system for any ELT class telescope because of the quantitative leap in the number of parameters (high dimensional system) and consequently the quantitative leap in the cost in real time processes. First of all, we briefly give some backgrounds of AO on 8-10 m class telescopes and the utilization of the KF for an optimal control law. Then we present the Ensemble Kalman Filter (EnKF), a recent method tried and tested in Geophysics and which is particularly well suited for a transition to a very high number of parameters. After a description from a general point of view, we shortly present the numerical implementation and two main approaches for simplifying the matrix equations of the estimation in order to reduce the computational complexities of this technique. Finally, we propose some different perspectives and future works of this approach.

Keywords: adaptive optics, E-ELT, Kalman filter

1 Background

1.1 Adaptive Optics command on 8-10 m class telescopes

In a standard AO system, the wave front sensor which gives a measurement of the wave front shape is located after the corrector element, a Deformable Mirror (DM). The wave front sensor gives then access to the shape of the residual phase. From its measurements, one has to compute the new optimal voltages to apply on the DM. The whole system therefore works in a closed loop. The usual control law for classical AO on standard 8-10 m class telescopes is made of a simple integrator control law on which the integrator gain has eventually been optimized with respect of the signal to noise ratio, in order to minimize the propagation of the noise mode by mode. When the phase is decomposed on a basis of modes, it in fact appears that the signal to noise ratio can be different on each mode. It becomes then possible to adjust the integrator gain on each mode.
Unfortunately, this approach on special AO systems, such as wide field AO, does not allow to correct efficiently the turbulence because in those cases, some energetic modes have a very poor signal to noise ratio.
It becomes then necessary to estimate those modes and not only to filter them : the Kalman Filter (KF) based control law allows this estimation.

1.2 Utilization of the KF and notations used for the Multi Conjugate AO (MCAO) on 8-10 m class telescopes

On a 8-10 m class telescope, the KF based control law improves the performance of wide field of view AO systems (MCAO for example), thanks to its ability to estimate the badly seen modes (with a low signal to noise ratio). It is also very helpful in eXtreme AO systems to predict and compensate the telescope vibration's effects. We will use the notations and the structure of the control law presented for MCAO in Le Roux et al. (2004) and Petit et al. (2009). For the hidden state vector, we choose an expression with the turbulence phase at 3 successive instants and the DM's voltages at 2 successive instants :

[1] LAM/OAMP, Pôle de l'Étoile, site de Château-Gombert, 38 rue F. Joliot-Curie, 13388 Marseille Cedex 13, France

$$X_k = \left((\varphi_{k+1}^{tur})^T \quad (\varphi_{k}^{tur})^T \quad (\varphi_{k-1}^{tur})^T \quad (u_{k-1})^T \quad (u_{k-2})^T \right)^T$$

We have a linear state space model whose solution in the Gaussian case is given by the KF :

$$X_{k+1}^{(1)} = \begin{bmatrix} A_{tur} & 0 & 0 \\ Id & 0 & 0 \\ 0 & Id & 0 \end{bmatrix} X_k^{(1)} + \begin{bmatrix} Id \\ 0 \\ 0 \end{bmatrix} V_k \quad ; \quad X_{k+1}^{(2)} = \begin{bmatrix} 0 & 0 \\ Id & 0 \end{bmatrix} X_k^{(2)} + \begin{bmatrix} Id \\ 0 \end{bmatrix} u_k$$

and : $Y_k = \begin{bmatrix} 0 & 0 & DM_\alpha^L \end{bmatrix} \times X_k^{(1)} + \begin{bmatrix} 0 & -DM_\alpha^M N \end{bmatrix} \times X_k^{(2)} + W_k$

The estimation of the predicting state vector is : $\hat{X}_{k+1/k}^{(1)} = A^{(1)} \times \hat{X}_{k/k-1}^{(1)} + A^{(1)} \times H_k \times [Y_k - \hat{Y}_{k/k-1}]$

with : $\hat{Y}_{k/k-1} = C^{(1)} \times \hat{X}_{k/k-1}^{(1)} + C^{(2)} \times X_k^{(2)}$

The Kalman gain is : $H_k = \Sigma_{k/k-1}^{(1)} C^{(1)T} \times [C^{(1)} \Sigma_{k/k-1}^{(1)} C^{(1)T} + \sum_w]^{-1}$

with the Riccati equation : $\Sigma_{k+1/k}^{(1)} = A^{(1)} \Sigma_{k/k-1}^{(1)} A^{(1)T} - A^{(1)} H_k C^{(1)} \Sigma_{k/k-1}^{(1)} A^{(1)T} + \sum_v$

1.3 Limitations of the KF for the AO systems of the E-ELT

Such a control law would be very helpful on an ELT class telescope. But it is becoming to much computing demanding, as the number of parameters increases dramatically. If the dimension n of the state vector is large (about 10^5 for the E-ELT), then computing and storing large n×n covariance matrices is impossible and the products for the estimation error's covariance matrices are even more problematic to work out.

2 The Ensemble Kalman Filter (EnKF) for the AO systems of ELT class telescopes

2.1 Presentation and theoretical concepts of the EnKF

The idea will be ***to use Monte Carlo samples*** and *not to use* the exact estimation error's covariance matrices. EnKF represents a distribution of the system state using a random sample, called an ensemble, and replace the covariance matrices by the sample covariance matrices computed from the ensemble.
EnKF is a Monte Carlo approximation of the KF which avoids evolving the real covariance matrices.
First of all, an initial ensemble of N_s elements is simulated as Independent Identically Distributed (IID) Gaussian random vectors with the same statistics as the initial condition X_0 : $\hat{X}_{0/0}^1; \ ... \ ; \hat{X}_{0/0}^{N_s}$
*During the **prediction step***, given the previous analysis ensemble, each ensemble element i is propagated independently according to the state equation (i is an integer from 1 to N_s) :

$$\hat{X}_{k/k-1}^{i} = A \times \hat{X}_{k-1/k-1}^{i} + V_k^i$$

We have to notice that IID random vectors V_k^i are simulated with the same statistics as the additive Gaussian model noise V_k in the original state's equation. We can then calculate :
the empirical estimation mean vector : $m_{k/k-1}^{N_s} = \frac{1}{N_s} \times \sum_{i=1}^{N_s} \hat{X}_{k/k-1}^i$
and the empirical estimation error's covariance matrix :

$$P_{k/k-1}^{N_s} = \frac{1}{N_s - 1} \times \sum_{i=1}^{N_s} (\hat{X}_{k/k-1}^i - m_{k/k-1}^{N_s})(\hat{X}_{k/k-1}^i - m_{k/k-1}^{N_s})^T$$

*During the **correction step***, given the previous forecast ensemble, each ensemble element i is updated independently according to the estimation equation :

$$\hat{X}_{k/k}^{i} = \hat{X}_{k/k-1}^{i} + H_k(P_{k/k-1}^{N_s}) \times [Y_k + W_k^i - C \times \hat{X}_{k/k-1}^i]$$

We have to notice that IID random vectors W_k^i are simulated with the same statistics as the additive Gaussian measurement noise W_k in the original observation's equation. But there are 2 approaches (see 3.3 and 3.4) : the covariance matrix $\sum_w$ can be obtained from the randomized data or from the real measurement errors.
We can then calculate :
the empirical Kalman gain matrix : $H_k(P_{k/k-1}^{N_s}) = P_{k/k-1}^{N_s} \times C^T \times [C \times P_{k/k-1}^{N_s} \times C^T + \sum_w]^{-1}$
the empirical estimation mean vector : $m_{k/k}^{N_s} = \frac{1}{N_s} \times \sum_{i=1}^{N_s} \hat{X}_{k/k}^i$
The ensemble covariance matrix $P_{k/k-1}^{N_s}$ is computed from all ensemble members together which introduces dependence and destroys the normality of the ensemble distribution. But Mandel et al. (2009) gives a mathematical proof of the convergence of the EnKF in the limit for large ensembles to the KF (large ensembles are in fact nearly IID and nearly normal). In Geophysics, they usually use a value of N_s between 50 to 100.

2.2 Utilization of the EnKF for the AO systems of the ELT class telescopes

The EnKF allows to bring on an ELT all the advantages of the state space formalism of a KF based control law. The ability of estimating the unseen modes for wide field AO remains critical. On an ELT class telescope, the ability to filter out vibration's modes thanks to an adapted state space model can also become fundamental, as the vibrations of an ELT class telescope is a critical issue.

3 Numerical Implementation of the EnKF

3.1 Implementation of the EnKF

It can be shown (Evensen (2003); Mandel (2006)) that, during the correction step, the *vectorial equation* can be rewritten with this new *matrix equation* :

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + \frac{Z_k(CZ_k)^T}{N_s-1} \times [\frac{(CZ_k)(CZ_k)^T}{N_s-1} + \Sigma_w]^{-1} \times [D_k - C\hat{X}_{k/k-1}]$$

with : $Z_k = \hat{X}_{k/k-1} \times [I_{N_s} - \frac{1}{N_s} \times J_{N_s}]$ and J_{N_s} a matrix with each element is equal to 1.

3.2 Computational Complexities of the EnKF

For the estimation of the computational complexity of this formula, we will only consider the number of multiplications. We just have to know that the multiplication of a matrix of size $n_1 \times n_2$ by a matrix of size $n_2 \times n_3$ has a *numerical cost of* $n_1 \times n_2 \times n_3$ *multiplications*. This cost is noticed : $O(n_1 \times n_2 \times n_3)$.
We have also to remind that : n is the number of coordinates in the state's vector X_k, p is the number of coordinates in the observation's vector Y_k, and N_s is the number of elements in the ensemble.

3.3 Evensen's Approach

The observations D_k are treated as random variables having a distribution with mean equal to the first-guess observations and covariance matrix equal to Σ_w (the simulated random measurement errors W_k^i have a mean equal to zero). We define the ensemble covariance matrix of the measurements as : $\Sigma_w = \frac{1}{N_s-1} \times \sum_{i=1}^{N_s} W^i W^{iT}$
As Evensen wrote : "the actual observation error covariance matrix is poorly known and the errors introduced by the ensemble representation can be made less than the initial uncertainty in the exact form of Σ_w. Further, the errors introduced by using an ensemble representation for Σ_w have less impact than the use of an ensemble representation of matrix $P^{N_s}_{k/k-1}$."
Using a Single Value Decomposition (SVD), a pseudo inversion and with the assumption of uncorrelated forecast elements and measurement errors, for the correction step, the new estimation's formula is :

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + Z_k(CZ_k)^T \times O_1(\Sigma\Sigma^T)^{\dagger}O_1^T \times [D_k - C\hat{X}_{k/k-1}]$$

The total computational complexity obtained is : $O(N_s^2 \times (n + p))$
which is **linear** and **suitable** for large values of n and p.

3.4 Mandel's Approach

The matrix Σ_w is the covariance matrix of the real measurement errors rather than the sample covariance matrix of the randomized data. Because Σ_w is always positive definite, there will be no difficulty to compute $\Sigma_w = SS^T$ (very low cost) and calculate the inverse Σ_w^{-1}. Moreover, there will be no need to use a pseudo inversion or a SVD on a large matrix.
Using the Sherman-Morrison-Woodbury formula and a Cholesky decomposition on a small matrix (its size is only $N_s \times N_s$), for the correction step, the new estimation's formula is :

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + \frac{Z_k(S^{-1}CZ_k)^T}{N_s-1}\{I_p - \frac{S^{-1}CZ_k}{N_s-1}[I_{N_s} + \frac{(S^{-1}CZ_k)^T(S^{-1}CZ_k)}{N_s-1}]^{-1}(S^{-1}CZ_k)^T\}S^{-1}[D_k - C\hat{X}_{k/k-1}]$$

The total computational complexity obtained is : $O(N_s^3 + N_s^2 \times (n + p))$
which is **linear** and **suitable** for large values of n and p.

3.5 Comparisons

These two implementation techniques have a **linear computational complexity** *in the number of degrees of freedom* $\boldsymbol{n}$ and *in the number of degrees of observation* $\boldsymbol{p}$ (with a proportional factor N_s^2).
However the method used by Mandel involves symmetric products of matrices which is numerically more stable and allows to save memory.
The problems using a low rank measurement error covariance matrix pointed out by Kepert (2004) are resolved in Evensen (2004) where he introduced a new Square Root implementation of the EnKF.

4 Conclusions

We have made a brief and simple description of two different efficient implementations of the EnKF (with a linear complexity) in order to use it as an optimal control law for the AO systems of the ELT class telescopes. This version described here involves randomization of data. But some alternative methods (without randomization of data) based on Square Root analysis schemes and the Ensemble Transform Kalman Filter (ETKF) seem to be very promising : some theoretical studies on the links between ETKF and the AO/MCAO for the E-ELT will be therefore deepened. The next step will be to adapt the current routines in order to implement them on our AO simulator for different numerical simulations. Then some works on the hardware and software design will be made to obtain precious gain in the real time identification. And finally, we will compare our results with those obtained in the other existing methods for the AO systems of ELT class telescopes.

References

Burgers, G. et al. 1998, Analysis Scheme in the Ensemble Kalman Filter, Monthly Weather Review, 126, 1719-1724

Evensen, G. 2003, The Ensemble Kalman Filter : Theorical Formulation and Practical Implementation, Ocean Dynamics, 53, 343-367

Evensen, G. 2004, Sampling Strategies and Square Root Analysis Schemes for EnKF, Ocean Dynamics, 54, 539-560

Evensen, G. 2009, Data assimilation : The Ensemble Kalman Filter, Springer Verlag, Berlin

Kepert, J. 2004, On ensemble representation of the observation-error covariance in the Ensemble Kalman Filter, Ocean Dynamics, 54, 561-569

Le Roux, B. et al. 2004, Optimal Control Law for Classical and Multiconjugate Adaptive Optics, J. Opt. Soc. Am. A., Vol 21, n° 7, 1261-1276

Mandel, J. 2006, Efficient Implementation of the Ensemble Kalman Filter, Center for Computational Mathematics Reports, N° 231

Mandel, J. et al. 2009, On the Convergence of the Ensemble Kalman Filter, `http://arxiv.org/abs/0901.2951`

Petit, C. et al. 2009, LQG control for Adaptive Optics and Multiconjugate Adaptive Optics : experimental and numerical analysis, J. Opt. Soc. Am. A., Vol 26, n° 6, 1307-1325