Optimal approximation of Kalman filtering with temporally local 4DVAR in operational weather forecasting

Harri Auvinen, Heikki Haario, Tuomo Kauranne,
Lappeenranta University of Technology

ECMWF Workshop on High-Performance Computing

ECMWF, Reading, 25 October 2004

1. Overview

1. Model error and background error covariance
2. Extended Kalman Filtering and 4D Variational Assimilation
3. Approximating forecast error covariance
4. Hessian approximation and singular vectors
5. Key Analysis Errors, observations and RRKF skill
6. Causes of failure of HSV's in RRKF
7. Variational Kalman Filtering
8. A numerical example
9. Conclusions

Model error and background error covariance

- 4D Variational Assimilation (4DVAR) is a least squares optimal approximation to the analysis under the assumptions that:

- The forecast model $\mathcal{M}$ is perfect

- The background forecast error covariance is constant in time

- Observation and background error covariance matrices are known

Model error and background error covariance

- The optimal **BLUE** (Best Linear Unbiased Estimator) to the analysis is the Extended Kalman Filter (EKF)

- 4DVAR is identical to EKF under the assumptions above: it performs Kalman smoothing for a perfect model with a constant background error covariance and known observation and background error covariances

- Classical EKF calls for tangent linear forecasts of all columns of the forecast error covariance matrix

- In operational NWP, this would mean 10^7 24 hour integrations every night - the equivalent of ten million parallel one day forecasts

Model error and background error covariance

- Optimal Interpolation (OI) and 3DVAR provide for an unbiased snapshot analysis that is independent of model error

- Continuous Data Assimilation (CDA) or "nudging" corrects for model errors, but is based on empirical structure functions rather than true model dynamics

Extended Kalman Filtering and 4D Variational Assimilation

- EKF corrects for model error and dynamically changing forecast error covariance every time step

- As a result EKF, like CDA, OI or 3DVAR, produces dynamically inconsistent model trajectories

- These dynamical inconsistencies provide for the degrees of freedom, or "moving joints" in the "rigid skeleton" of 4DVAR that are needed to correct for model bias or changing background error

Approximating forecast error covariance

- Variational and Kalman data assimilation minimize the cost function for the analysis x_0 at initial time (as modified from *Beck* (2004))

$$J(x_0) = \sum_{t=0}^{L}((x_t - x_t^b)^T (P_t^f)^{-1}(x_t - x_t^b) +$$

$$r_t(y_t - \mathcal{H}x_t)^T R_t^{-1}(y_t - \mathcal{H}x_t))$$

where $x_t = \mathcal{M}_t(x_0)$ by the nonlinear model evolution, $\mathcal{H}$ is the nonlinear observation operator and R the observation error covariance matrix. r_t is an indicator function to identify the time steps with observations y_t.

- The analysis error covariance matrix P^a is the inverse of the Hessian of J:

$$P^a = (J'')^{-1} = (\nabla\nabla J)^{-1}$$

- The choice of forecast error covariance matrix P_t^f varies between different methods

Approximating forecast error covariance

- **4DVAR**:

- P^f is the static background error matrix B and there is no model error

$$P_t^f \equiv B$$

Approximating forecast error covariance

- Full Kalman Filter **EKF**:

- P^f is dynamic and may contain model error with covariance matrix Q

- P^f is propagated in time by the tangent linear model M. The model error covariance matrix Q is added at each time step

$$P_t^f = M_t P^a M_t^T + Q$$

Approximating forecast error covariance

- Reduced Rank Kalman Filtering **RRKF**:

- RRKF behaves like the Extended Kalman Filter in a low dimensional subspace determined at the beginning of the assimilation window

$$P_t^f = M_t L^T \begin{bmatrix} E & F^T \\ F & I \end{bmatrix} L M_t^T$$

- L is the orthogonal matrix that transforms model variables into control variables. E is the model error covariance matrix in the chosen k-dimensional subspace of the control space and F is its cross-correlation matrix with the complement subspace

Approximating forecast error covariance

- Ensemble Kalman Filter **EnKF**:

- A low dimensional approximation $\text{En} - P_t^f$ to P_t^f is propagated by the full nonlinear model and there is no model error

$$\text{En} - P_t^f = \mathcal{M}_t(S)\mathcal{M}_t(S)^T$$

- S is a sample of vectors, such as a multi-normal sample modulated by a low rank approximation to the analysis error covariance matrix P^a at initial time

Hessian approximation and singular vectors

- The optimal low-rank approximation to forecast error covariance, up to any given rank k, and targeted at **a particular final time** (and with respect to a particular final time norm), are the dominant k singular vectors of the Hessian, i. e. the inverse of the corresponding analysis error covariance matrix P^a. They are called Hessian singular vectors, or HSV's.

- As shown by *Ehrendorfer and Beck*(2003), HSV's evolve into the **eigenvectors** of the corresponding forecast error covariance matrix P^f when the model dynamics are linear

Hessian approximation and singular vectors

- The orthogonal Hessian singular vectors z_0 are defined by

$$M^T C^T C M z_0 = \lambda (P^a)^{-1} z_0$$

$$z_0^T (P^a)^{-1} z_0 = I$$

where C represents the chosen final time norm

- With the perfect model assumption $P^f = M P^a M^T$, and we get

$$C P^f C^T z_t = \lambda z_t$$

The z_t are still orthogonal, when transported by the tangent linear model

Key Analysis Errors, observations and RRKF skill

- Key Analysis Errors of *Klinker, Rabier and Gelaro* (1998) are the initial time directions that cause the largest deviations between a 48 h forecast and the subsequent analysis, transported back to t = 0h by the adjoint to the tangent linear model M^T. They have been suggested to be an indication of dominant analysis errors.

- *Isaksen, Fisher and Andersson* (2004) have carefully analyzed the influence that an analysis can have through the HSV's computed for t = 48 h on forecast quality and found this connection problematic.

- Different final time norms at t= 48h produce very different Key Analysis Error patterns, yet they should be due to the same defects in the analysis.

- Also, corrections in Key Analysis Error directions actually deteriorate the forecast for the first 12 hours. An improved analysis should do the opposite.

- *Fisher and Andersson* (2001) and also *Leutbecher* have noted that even after the first 12 hours into the forecast, HSV's computed at initial and final time, respectively, project only weakly onto one another. They mention this as a possible cause to the failure of RRKF experiments at ECMWF to have a positive impact on forecast skill.

Causes of failure of HSVs in RRKF: the Roots

- Let's first see a video!

- Fluid flow is a three dimensional system of Newtonian rotation with a very large number ($\mathcal{O}$(exp(Re)) of rotational degrees of freedom

- Two-dimensional rotation is predictable: its underlying Newtonian Lie group $U(1)$ is commutative. Two-dimensional vortices represent barotropic flow.

- Two-dimensional vortices foliate the plane and their structure can be represented by a separable normed space and the corresponding evolution operator is homeomorphic to a unitary operator.

Causes of failure of HSVs in RRKF: the Roots

- Three-dimensional Newtonian rotation is chaotic. It can represent also baroclinic flows.

- The Lie group $SO(3)$ of three dimensional rotation is non-commutative and the corresponding inviscid fluid flow cannot be represented by any real linear space.

Causes of failure of HSV's in RRKF: the Consequences

- As a consequence of its chaoticity, 3D fluid flow cannot have any linear system of invariant subspaces. All subspaces "roll away" quite literally!

- Therefore HSV's will be different for all final times and all initial conditions. This could be the reason why fixed subspaces in RRKF, or Key Analysis Errors, do nor persevere for even 12 hours.

- EnKF is likely to be troubled by this same phenomenon, if its space of initial perturbations is fixed for the whole analysis window

- Yet *Ehrendorfer and Beck* (2003) have shown, that for a quasigeostrophic model, the full EKF is quite skillful.

- But we cannot hope to compute the full P^f even with the Earth Simulator 2!

Plumbing covariance leaks

- EKF takes into account model error every time step: $P^f = M^T P^a M + Q$

- As a result, it produces model trajectories that are **not** dynamically consistent, but still statistically **BLUE**

- In RRKF, the failure of the chosen subspace to contain the covariance (”It leaks covariance”, according to *Fisher and Andersson* (2001)), is analogous to model error in the RRKF subspace. The Q term should be present!

- We can incorporate model error also with Hessian singular vectors!

Variational Kalman Filtering (VKF)

- The obvious answer is to recompute the HSV's (almost) every time step. This could be done with a local Lanczos algorithm.

- But there is a computationally faster way to achieve the same result:
  - Run 4DVAR over a few time steps - or even just one - at a time and
  - Build an approximation to the Hessian from the search directions with the LBFGS update formula.
  - Use this LBFGS approximation to the Hessian as a temporally local approximation to P_t^f in a full EKF algorithm

Variational Kalman Filtering (VKF)

- The resulting matrix spans the same HSV subspace as Lanczos (*Kauranne* (1992)), but for a **temporally local** analysis error covariance matrix P_t^a. This matrix has correctly accumulated all past analysis, background and model errors in the dominant HSV's at each instant in time

Computational Cost of Variational Kalman Filtering

- The forecast error covariance P^f is kept piecewise constant over a VKF assimilation "mini window" of a few time steps.

- There is no need for overlapping 4DVAR periods, when a piecewise constant P^f is used over a mini window.

- The computational cost of VKF is only slightly larger than that of 4DVAR.

A numerical example

- A linear advection diffusion equation over a latitude circle with an ozone like source term

  $$u_t = -vu_x + ku_{xx} + r(x)$$

  where $r(x)$ is a Haar wavelet like step function with average value zero

- A small finite difference model with 64 grid points and a time step of 10 minutes

- Polar orbiter like satellite observations that span a swath of ten grid points 15 times a day

- Gaussian observation noise

- Model error introduced in a 50 per cent systematic overestimation of $r(x)$

- A 15 day run, first 4 days shown

- A diagonal constant component diag(B) (i.e. variance only) in the forecast error covariance matrix P^f.

Conclusions

- Problems with Hessian Singular Vector based low rank Kalman filtering methods, such as RRKF and EnKF, may be a result of rapid and nonunitary covariance matrix eigenspace rotation in 3D fluid flow

- Low rank Kalman filtering methods must take model error into account explicitly when propagating forecast error covariance. Otherwise covariance leaks away from any fixed subspace

- Variational Kalman Filtering is a computationally efficient method to do this

- First numerical results with very simple models are encouraging, but

- Numerical tests with more elaborate atmospheric models are needed