Data assimilation with filtering methods that use Kullback-Leibler distance

Sam Pimentel[1] and Youssef Qranfal[2]

[1]Department of Mathematics, Trinity Western University, B.C., Canada (sam.pimentel@twu.ca)

[2]Department of Applied Mathematics, Wentworth Institute of Technology, Boston, MA, USA.

Abstract

In this paper we will present two filtering algorithms that minimize a cost function of weighted Kullback-Leibler (cross entropy) distance rather than the standard Euclidean distance. Because Kullback-Leibler distance is non-symmetric two filtering methods emerge, the EM (expectation maximization) filter and the SMART filter (simultaneous multiplicative algebraic reconstruction technique). These filters were originally developed to solve an ill-posed inverse problem that arises in reconstructing a time-varying medical image. The algorithms hold potential for data assimilation applications in geophysical fluid problems where we are also interested in time-varying variables of large-scale systems. These new methods have advantages over traditional methods, such as the Kalman filter, in that they do not involve matrix-matrix multiplication or matrix inversion and thus are computationally more efficient. We introduce the EM and SMART filter as a solution to the data assimilation problem and implement these methods on a few simple data assimilation applications. Results are compared with those from more standard approaches. We will highlight the advantages and disadvantages of the EM and SMART filters and demonstrate the potential benefits of the algorithms for geophysical data assimilation applications.

Data Assimilation

Best Linear Unbiased Estimator (BLUE)

Given partial observations $y = Hx^t + \mu$ of a true state, x^t, where $\mathbb{E}(\mu) = 0$ and $\mathbb{E}(\mu^T\mu) = R$, and a background estimate of the state x^b, where $\mathbb{E}(x - x^b) = 0$ and $\mathbb{E}((x - x^b)^T(x - x^b)) = B$ then the BLUE is found by minimizing the cost function

$$J(x) = ||x - x^f||^2_{B^{-1}} + ||Hx^f - y||^2_{R^{-1}}$$

which has solution:

$$x^a = x^b + BH^T(HBH^T + R)^{-1}(y - Hx^b)$$

Kalman Filter

The Kalman Filter evolves the BLUE analysis sequentially

$$x^f_{k+1} = M_{k,k+1}x^a_k$$
$$P^f_{k+1} = M_{k,k+1}P^a_kM^T_{k,k+1} + Q_k$$
$$x^a_k = x^f_k + K_k(y_k - H_kx^f_k)$$

where

$$K_k = P^f_kH^T_k(H_kP^f_kH^T_k + R_k)^{-1}$$
$$P^a_k = (I - K_kH_k)P^f_k$$

An approximation sometimes used is to fix the error covariance matrix, i.e. let $P_k = P_0, \forall k$.

SMART

The simultaneous multiplicative algebraic reconstruction technique (SMART) ([1], [2]) determines the solution to the following problem: for $x > 0$ and $0 \le \alpha \le 1$ minimize

$$J(x) = \alpha\mathrm{KL}(Tx, d) + (1 - \alpha)\mathrm{KL}(x, q)$$

subject to $Tx = d$ and where q is a priori estimate of x.

The solution, equivalent to maximizing Shannon entropy, is found iteratively as follows:

$$x^{\ell+1}_j = (q_j)^{1-\alpha}\left[x^\ell_j\prod_{i=1}^{M}\left(\frac{d_i}{(Tx^\ell)_i}\right)^{T_{ij}}\right]^\alpha$$

EM

The expectation maximization (EM) algorithm ([1], [3]) determines the solution to the problem: for $x > 0$ and $0 \le \alpha \le 1$ minimize

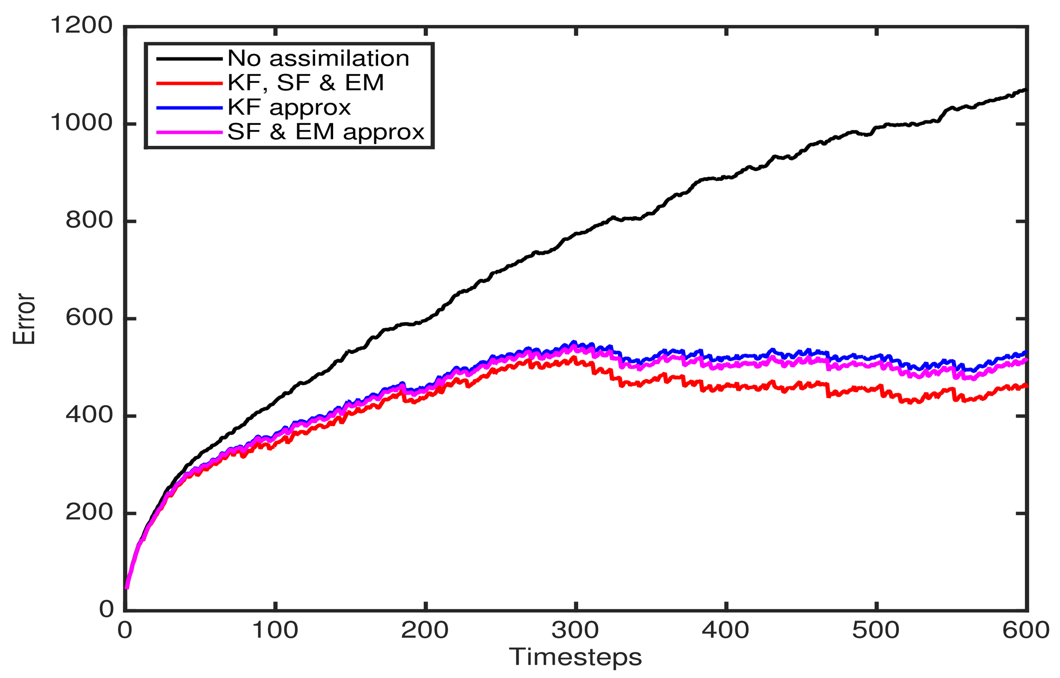
$$J(x) = \alpha\mathrm{KL}(d, Tx) + (1 - \alpha)\mathrm{KL}(q, x)$$

subject to $Tx = d$ and where q is a priori estimate of x.

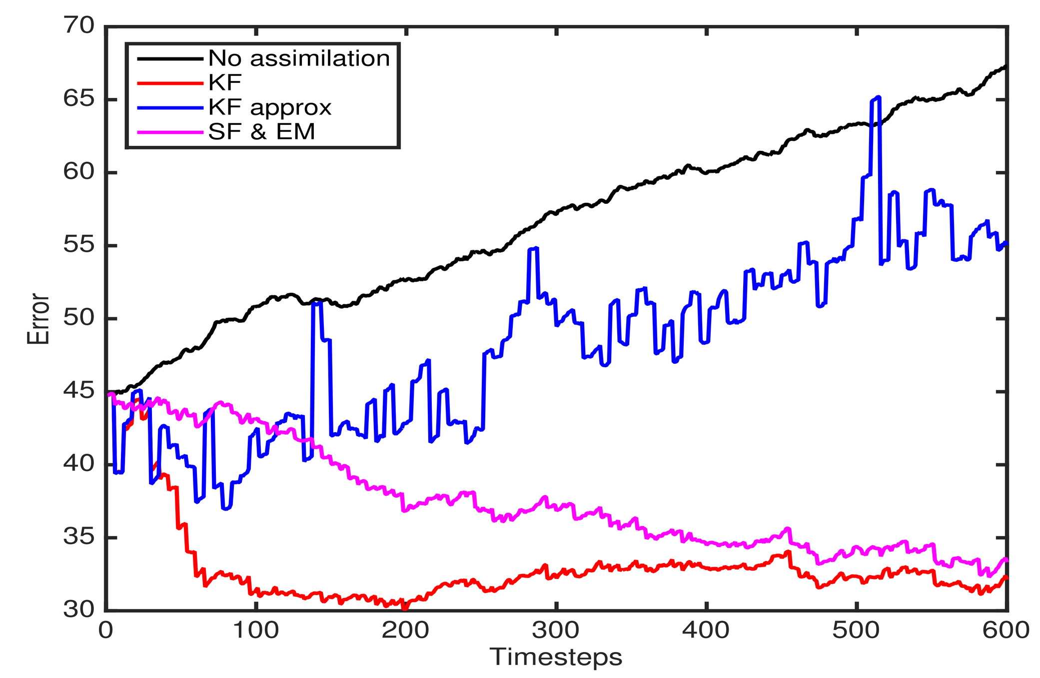
The solution, equivalent to maximizing Burg entropy, is found iteratively as follows:

$$x^{\ell+1}_j = \alpha x^\ell_j\sum_{i=1}^{M}\frac{T_{ij}d_i}{(Tx^\ell)_i} + (1 - \alpha)q_j$$

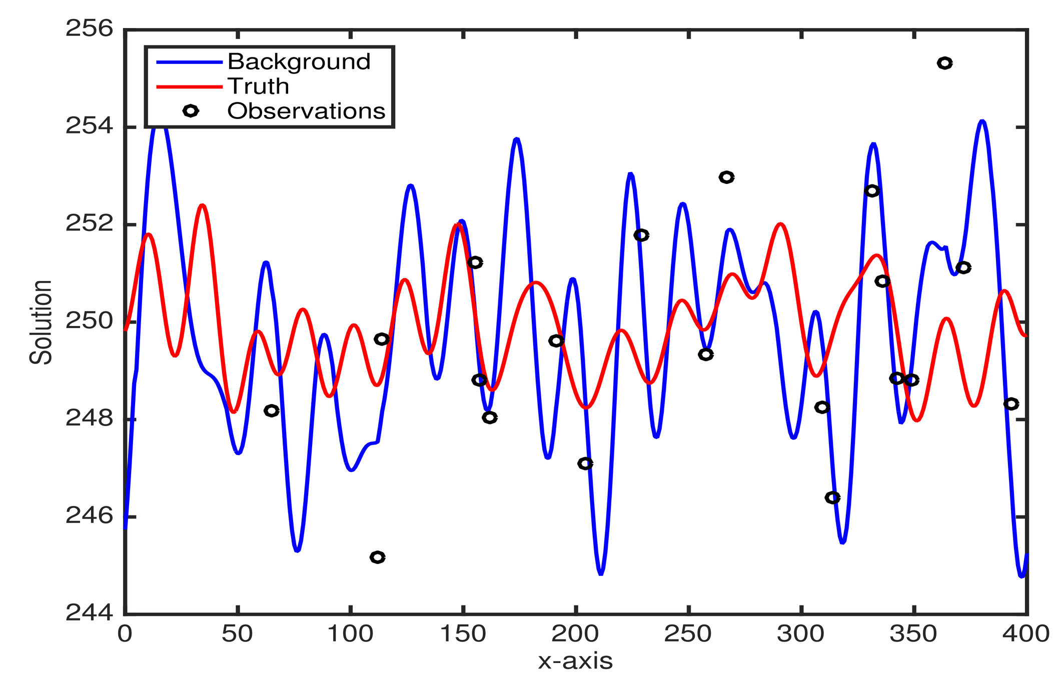
Results



Experiment 1: $\sigma^2_{\mathrm{m}} = 5$, $\sigma^2_{\mathrm{o}} = 5$, and $\sigma^2_{\mathrm{b}} = 0$.



Experiment 2: $\sigma^2_{\mathrm{m}} = 0.01$, $\sigma^2_{\mathrm{o}} = 5$, and $\sigma^2_{\mathrm{b}} = 5$



A pseudo-random wave (red) linearly advected $(1\,\mathrm{m\,s^{-1}})$ from left to right. 20 noisy observations (black circles) at randomly selected locations are assimilated every 6 timesteps. In experiment 2 the initial condition is offset from the truth by adding a normally distributed error with mean 0, variance 1, and decorrelation length of 20 (blue).

Exp.	Filter	Error Propagation	Time (s)	Error
1	KF	$P_k = M_{k-1,k}P_{k-1}M^T_{k-1,k} + Q_k$	8.54	498
1	SF	$(\sigma^2_{\mathrm{m}})_k = (\sigma^2_{\mathrm{m}})_{k-1}\|\|M_{k-1,k}\|\|^2_2 + (\sigma^2_{\mathrm{m}})_k$	0.35	489
1	EM	$(\sigma^2_{\mathrm{m}})_k = (\sigma^2_{\mathrm{m}})_{k-1}\|\|M_{k-1,k}\|\|^2_2 + (\sigma^2_{\mathrm{m}})_k$	0.30	494
1	KF	$P_k = \|\|M_{k-1,k}\|\|^2_2P_{k-1} + Q_k$	3.57	504
1	KF	$P_k = Q_k$	3.36	558
1	SF	$(\sigma_{\mathrm{m}})_k = (\sigma_{\mathrm{m}})_0 + \sigma_{\mathrm{b}}$	0.35	552
1	EM	$(\sigma_{\mathrm{m}})_k = (\sigma_{\mathrm{m}})_0 + \sigma_{\mathrm{b}}$	0.33	549
2	KF	$P_k = M_{k-1,k}P_{k-1}M^T_{k-1,k} + Q_k$	8.70	34
2	SF	$(\sigma^2_{\mathrm{m}})_k = (\sigma^2_{\mathrm{m}})_{k-1}\|\|M_{k-1,k}\|\|^2_2 + (\sigma^2_{\mathrm{m}})_k$	0.34	39
2	EM	$(\sigma^2_{\mathrm{m}})_k = (\sigma^2_{\mathrm{m}})_{k-1}\|\|M_{k-1,k}\|\|^2_2 + (\sigma^2_{\mathrm{m}})_k$	0.30	40
2	KF	$P_k = \|\|M_{k-1,k}\|\|^2_2P_{k-1} + Q_k$	3.49	36
2	KF	$P_k = Q_k$	3.35	66
2	KF	$P_k = \mathrm{diag}(B)$	3.21	37
2	SF	$(\sigma_{\mathrm{m}})_k = (\sigma_{\mathrm{m}})_0 + \sigma_{\mathrm{b}}$	0.32	38
2	EM	$(\sigma_{\mathrm{m}})_k = (\sigma_{\mathrm{m}})_0 + \sigma_{\mathrm{b}}$	0.32	38

These are mean values over 10 realizations.

Kullback-Leibler Distance

The Kullback-Leibler or cross-entropy distance [4] between two non-negative vectors $a = (a_1, \cdots, a_N)^T$ and $b = (b_1, \cdots, b_N)^T$ is given by

$$\mathrm{KL}(a, b) = \sum_{i=1}^{N} a_i\log\frac{a_i}{b_i} + b_i - a_i$$

Note, that the Kullback-Leibler distance is not symmetric, hence in general $\mathrm{KL}(a, b) \ne \mathrm{KL}(b, a)$.

Filtering

The cost functions can be reformulated to solve the data assimilation problem, for example:

$$\begin{aligned} J(x) &= \mathrm{KL}_{R^{-1}}(Hx, y) + \mathrm{KL}_{P^{-1}}(x, x^f) \\ &= \mathrm{KL}(U_1Hx, U_1y) + \mathrm{KL}(U_2x, U_2x^f) \\ &= \frac{\sigma_{\mathrm{m}} - 1}{\sigma_{\mathrm{m}}}\mathrm{KL}(Hx, y) + \frac{1}{\sigma_{\mathrm{m}}}\mathrm{KL}(x, x^f) \end{aligned}$$

This can be solved sequentially as a filtering algorithm by evolving the model state and updating the weights α.

Experiments

We perform 'twin experiments' whereby noisy pseudo-observations, $y_k = H_kx^t_k + \mu_k$ are taken from the truth $x^t_{k+1} = M_{k,k+1}x^t_k + \epsilon_k$. Such that

$$\mathbb{E}(\mu) = 0 \qquad \mathbb{E}(\mu^T\mu) = R = \sigma^2_{\mathrm{o}}$$
$$\mathbb{E}(\epsilon_k) = 0 \qquad \mathbb{E}(\epsilon^T_k\epsilon_k) = P_k = (\sigma_{\mathrm{m}})^2_k$$

$$R^{-1} = U^T_1U_1 \qquad P^{-1} = U^T_2U_2$$

- The physical model $M_{k,k+1}$ is the one-dimensional linear advection equation, solved on 400 gridpoints over 600 timesteps
- White noise for model error with different variance at each timestep
- White noise for observations, observation locations are randomly sampled, 20 observations every 6 timesteps

Experiment 1: Perfect initial condition
Experiment 2: Non-white initial condition error

Conclusions

1. EM and SMART Filters are computationally faster than the Kalman Filter, with EM being fastest
2. EM and SMART Filters have similar accuracy to the Kalman Filter
3. EM and SMART Filters can easily employ a positive constraint

References: [1] Bryne, 1993, [2] Qranfal & Bryne, 2011b, [3] Qranfal & Bryne, 2011a, [4] Kullback & Leibler, 1951