Theory and Practice of Data Assimilation for Oceanography

Robert N. Miller

College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, OR 97330

Introduction and Overview

"Data assimilation refers to three problems in time series analysis. Given a time series ω_k, or possible a continuous function of space and time $\omega(x,t)$ which may be noisy or incomplete, beginning with time $t = -T$ and ending at $t = 0$, the "present," define three problems:

- The prediction problem What will ω be in the future?
- The filtering problem What is the best estimate of ω *now*, i.e., at $t = 0$?
- The smoothing problem: What is the best estimate of ω for the entire time series?

Origins of Data Assimilation

Data assimilation probably started with Gauss (1826)

Carl Friedrich Gauss, 1777-1855

Origins of Data Assimilation

...at least he gets the credit. But Legendre published first:

Adrien-Marie Legendre, 1752-1833

Origins of Data Assimilation

Gauss and Legendre were interested in *planetary orbits*.

- These are specified by 6 parameters, the *orbital elements*.
- Three observations are necessary to determine the orbital elements.
- If more than three observations are available choose elements to minimize:

$$\sum (predicted\ position - observed\ position)^2$$

This is the *least squares method*, the most basic concept in data assimilation.

The Least Squares Method

Gauss and Legendre solved the *smoothing problem* for planetary orbits

- They assumed the motion of the planets was described exactly by a solution to the classical two-body problem.
- The six parameters are equivalent to three initial velocity components and three initial position coordinates.
- In the context of data assimilation today, we would call that a strong constraint method.

Variational Methods

Given

- A model: $\mathbf{u}_t - L\mathbf{u} = \mathbf{f}$
- Chosen to mimic the "true" state $\mathbf{u}^{(t)}$ assumed to evolve according to $\mathbf{u}_t^{(t)} - L\mathbf{u}^{(t)} = \mathbf{f} + \mathbf{b}$ for some random function $\mathbf{b}$
- Estimated initial condition $\mathbf{u}(0)$ with random error $\mathbf{e}_0$
- Observations $\mathbf{z} = H\mathbf{u}^{(t)} + \mathbf{e}_{obs}$

Variational Methods

Minimize the cost function:

$$\begin{aligned} J(\mathbf{u}) &= \int (\mathbf{u}_t - L\mathbf{u} - \mathbf{f})^T W^{-1} (\mathbf{u}_t - L\mathbf{u} - \mathbf{f}) dt + \\ & (\mathbf{u}(0) - \mathbf{u}_0)^T V^{-1} (\mathbf{u}(0) - \mathbf{u}_0) + \\ & (\mathbf{z} - H\mathbf{u})^T R^{-1} (\mathbf{z} - H\mathbf{u}) \end{aligned}$$

The minimizer of J is the BLUE of $\mathbf{u}^{(t)}$ if:

$$\begin{aligned} E(\mathbf{b}\mathbf{b}^T) &= W \\ E(\mathbf{e}_0\mathbf{e}_0^T) &= V \\ E(\mathbf{e}_{obs}\mathbf{e}_{obs}^T) &= R \end{aligned}$$

Variational Methods

We begin with $\mathbf{u}$ a (possibly) vector-valued function of time.
This formulation generalizes naturally to functions of time and space, in which case:

- L would be a partial differential operator
- The constraint on the initial condition would be an integral
- There might be a constraint on the boundary conditions.

We will derive all of the linearized methods from here.

The Representer Method

Without loss of generality, we can set $f = u_0 = 0$. so:

$$\begin{aligned} J(\mathbf{u}) &= \int (\mathbf{u}_t - L\mathbf{u})^T W^{-1} (\mathbf{u}_t - L\mathbf{u}) dt \\ &\quad + \mathbf{u}(0)^T V^{-1} \mathbf{u}(0) + \sum_{j=1}^{N} R_j^{-1} (z_j - H_j \mathbf{u}(t_j))^2 \\ &\equiv \; < \mathbf{u}, \mathbf{u} > + \sum_{j=1}^{N} R_j^{-1} (z_j - H_j \mathbf{u}(t_j))^2 \end{aligned}$$

So the cost function defines a positive definite bilinear form $< \cdot, \cdot >$
(Think dot product)

The Representer Method

Define the j^{th} *representer* r_j:

$$< r_j, u >= H_j u(t_j)$$

for any admissible function u

- The representer *represents* the measurement functional in terms of the new inner product.
- This allows us to form an orthogonal decomposition of the space of admissible functions.

Orthogonal Decomposition of State Space

Write the minimizer $\hat{u}$ of the functional J, as:

$$\hat{u} = \sum_{j=1}^{N} b_j r_j + G$$

where the b_j are constants and

$$< r_j, G >= 0, \; j = 1, \ldots, N$$

Solution in Representer Space

The cost function then becomes:

$$J(u) \;=\; \sum_{i,j=1}^{N} b_i b_j < r_i, r_j > + < G, G > + \\ \sum_{j=1}^{N} R_j^{-1} (z_j - \sum_i b_i < r_i, r_j >)^2$$

- We might as well pick $G = 0$
- Picking nonzero G doesn't change the data misfit and can only increase the cost.

The Representer Method

The original infinite dimensional problem is reduced to finding a finite number of coefficients b_j:

$$\frac{\partial J}{\partial b_k} = 2\sum_j b_j < r_j, r_k > - \\ 2\sum_j R_j^{-1}(z_j - < r_j, \sum_i b_i r_i >) < r_j, r_k >$$

Setting $\partial J/\partial b_k = 0$ leads to:

$$\sum_j < r_j, r_k > \left(R_j b_j + \sum_i < r_i, r_j > b_i - z_j \right) = 0$$

The Representer Method

$$\sum_j < r_j, r_k > \left(R_j b_j + \sum_i < r_i, r_j > b_i - z_j \right) = 0$$

In matrix form. Define $R = diag(R_j)$ and $M_{i,j} =< r_i, r_j >$ the *representer matrix*. The solution is then defined by:

$$(M + R)\, b = z$$

where b is the vector of representer coefficients and z is the vector of observations.

What Value Should the Cost Function Be at Minimum?

At the minimum,

$$
\begin{aligned}
J &= z^T(M+R)^{-1}M(M+R)^{-1}z + \\
&\quad (z - M(M+R)^{-1}z)^T R^{-1}(z - M(M+R)^{-1}z \\
&\quad (lots\ of\ algebra\ldots) \\
&= z^T(M+R)^{-1}z
\end{aligned}
$$

So z should be a random variable with covariance $M+R$ and J is a random variable with χ^2 distribution on M degrees of freedom.

Computing Representers

Begin with the simplest case: a linear, scalar ODE:

$$\dot{u} - au = F$$

F, $u(0)$ unknown. First guess: $F = 0$; $u(0) = 0$
Given measurements y_j of the system at times t_j

$$\begin{aligned} J &= \int_0^T (\dot{u} - au) W^{-1} (\dot{u} - au) dt + u(0) V^{-1} u(0) + \\ & \sum (y_j - u(t_j))^2 / R_j \\ &\equiv < u, u > + \sum (y_j - u(t_j))^2 / R_j \end{aligned}$$

Computing Representers

The j^{th} representer is defined by
$< r_j, v >= v(t_j) = \int_0^T \delta(t - t_j)v(t)dt$
Step 1:
Define the *representer adjoint* $\alpha_j = (r_j - ar)W^{-1}$, so:

$$\begin{aligned} < r_j, v > &= \int_0^T \alpha(\dot{v} - av)dt + r(0)V^{-1}v(0) \\ &= \int_0^T \delta(t - t_j)v(t)dt \end{aligned}$$

Computing Representers

Step 2:
Integrate by parts:

$$\int_0^T (-\dot{\alpha} - a\alpha)v dt + \alpha v|_0^T + r_j(0)V^{-1}v(0) = v(t_j)$$

Step 3: Solve

$$\begin{aligned}
-\dot{\alpha} - a\alpha &= \delta(t - t_j) \\
\alpha(T) &= 0 \\
r_j(0) &= \alpha(0)V \\
\dot{r}_j - ar_j &= W\alpha
\end{aligned}$$

Remarks

- α is the Green's function for the initial value problem
- As such, in general, α is the solution to an adjoint problem
- Generalization to vector ODEs and PDEs is straightforward
- Generalization to different measurement functionals is also straightforward.

Summary of the Representer Method

- The linear inverse problem is potentially a minimization problem over ∞ dimensions
- In practice the observations determine only a finite number of degrees of freedom
- A quadratic cost function can define a useful orthogonal decomposition of state space into two components:
  - The space spanned by the representers
  - Its orthogonal complement, all members of which are *unobservable*, i.e., they give measurements with value zero, by construction.

Summary, continued

- The minimization can thus be carried out over the space of representers
- The representers can (but need not be) calculated explicitly
- The representers do not depend on the data weights

The Variational Approach

Calculate the first variation δJ of the cost function J and set $\delta J = 0$ A slightly more general cost function:

$$
\begin{aligned}
J(u) \;=\; & \frac{1}{2}\int_0^T \int_\Omega \int_\Omega (u_t(x_1,t) - Lu)W^{-1} \\
& (u_t(x_2,t) - Lu)dx_1 dx_2 dt + \\
& \frac{1}{2}\int_\Omega \int_\Omega u(x_1,0)V^{-1}u(x_2,0)dx_1 dx_2 + \\
& \frac{1}{2} z^T R^{-1} z
\end{aligned}
$$

where z is the innovation vector, with components $z_j = y_j - H_j u$.

The Variational Approach

As before, write:

$$\lambda = (u_t - Lu)W^{-1}$$

For $u \longrightarrow u + \delta u$ set $\delta J = J(u + \delta u) - J(u) = O(\delta u^2)$

The Euler-Lagrange Equations

$$
\begin{aligned}
-\lambda_t - L^*\lambda &= z^T R^{-1} H \\
\lambda(T) &= 0 \\
u(x,0) &= \lambda(x,0)v(0) \\
u_t - Lu &= W\lambda
\end{aligned}
$$

Write $\lambda = \sum_j a_j \alpha_j$ where the α_j are the *representer adjoints*:

$$
\begin{aligned}
-\alpha_{jt} - L^*\alpha_j &= H_j \delta(t - t_j) \\
\alpha(T) &= 0
\end{aligned}
$$

$\longrightarrow$ the representer solution: Bennett (1992, 2002) or the tutorial at http://iom.asu.edu.

Filtering

Recall the *filtering problem*

Given a time series ω_k, or possible a continuous function of space and time $\omega(x,t)$ which may be noisy or incomplete, beginning with time $t = -T$ and ending at $t = 0$, the "present," What is the best estimate of ω?

Given current observations, we will *not* revise our estimate of past states.

Filtering

Consider a model with state vector $\mathbf{v}$.
Consider a single step of a prediction-analysis cycle:

1. Given an initial condition $\mathbf{u}_0$ at $t = t_0$, predict the new state $\mathbf{u}_1$ at the next time t_1: $\mathbf{u}_1^f = L\mathbf{u}_0$.
2. Given observations $\mathbf{y}$ at time t_1, form an improved estimate $\mathbf{u}_1^a = \mathbf{u}_1^f + \mathbf{v}_1$ of the state $\mathbf{u}_1$
3. As before, if full system is linear, the corrections $\mathbf{v}_{0,1}$ go by the same dynamics as $\mathbf{u}$.

Filtering: Variational Formulation

Cost function:

$$
\begin{aligned}
J &= \mathbf{v}_0^T P_0^{-1} \mathbf{v}_0 + (\mathbf{v}_1 - L\mathbf{v}_0)^T Q^{-1} (\mathbf{v}_1 - L\mathbf{v}_0) \\
&\quad + (\mathbf{z} - H\mathbf{v}_1)^T R^{-1} (\mathbf{z} - H\mathbf{v}_1)
\end{aligned}
$$

$$
\mathbf{z} = \mathbf{y} - H\mathbf{u}_1^f
$$

Filtering: Variational Formulation

Minimization of J by the representer method leads to:

$$\mathbf{v}_1 = (LP_0L^* + Q)H^T \left[H(LP_0L^* + Q)H^T + R\right]^{-1} \mathbf{z}$$

Recall $\mathbf{v}_1$ is the correction to the first guess $\mathbf{u}_1^f$.

Putting it all together

$$\mathbf{u}_1^a = \mathbf{u}_1^f + (LP_0L^* + Q)H^T \left[H(LPL^* + Q)H^T + R\right]$$

This is usually broken down into steps:

1. $\mathbf{u}_1^f = L\mathbf{u}_0$
2. $P_1^f = LP_0L^* + Q$
3. $K = P_1^f H^T \left[HP_1^f H^T + R\right]^{-1}$
4. $\mathbf{u}_1^a = \mathbf{u}_1^f + K(\mathbf{y} - H\mathbf{u}_1^f)$

Statistics

We assume our model, given by:

$$\mathbf{u}_{j+1} = L\mathbf{u}_j$$

differs from the "truth" by some random error ϵ

$$\mathbf{u}^t_{j+1} = L\mathbf{u}^t_j + \epsilon$$

ϵ is white in time with covariance $E(\epsilon\epsilon^T) = Q$
The error in the state is given by $\mathrm{e}_0 = \mathbf{u}^t_0 - \mathbf{u_0}$
with covariance $P_0 = E(e_0 e_0^T)$ at time $t = 0$.
The observation error is white with mean zero and covariance R.

Filtering: Statistics

Then:
The state error covariance evolves according to:

$$P_1^f = E(\mathbf{e}_1\mathbf{e}_1^T) = LE(\mathbf{e}_0\mathbf{e}_0^T)L^* + Q$$

The error in the corrected state should be smaller than the error in the original state. The covariance of the error in the updated state is:

$$P_1^a = (I - KH)P_1^f$$

The Filter Solution

Putting it all together:

1. $\mathbf{u}_1^f = L\mathbf{u}_0$
2. $P_1^f = LP_0L^* + Q$
3. $K = P_1^f H^T \left[H P_1^f H^T + R \right]^{-1}$
4. $\mathbf{u}_1^a = \mathbf{u}_1^f + K(\mathbf{y} - H\mathbf{u}_1^f)$
5. $P_1^a = (I - KH)P_1^f$

This is the *Kalman Filter.*

Remarks

- This is one of many ways to derive the Kalman filter
- Implementation is straightforward, but potentially very expensive
- Not necessary to write complex adjoint code

Remarks

- There are many natural generalizations and simplifications of the KF:
  - Use a nonlinear model for the state evolution and linearized dynamics to calculate the evolution of the error covariance; this is the *extended Kalman filter*
  - Use a static form of the error covariance P and eliminate the repeated calculations.
  - Use a collection of model runs with randomly chosen initial conditions and forcing to calculate an approximate covariance. This is the *ensemble Kalman filter*
  - Neglect errors outside of a low-dimensional subspace of the full state space. This is the *reduced state space Kalman filter*.

Summary

- We have explored solving the linear inverse problem by the least squares method
- In variational form, the cost function gives a natural orthogonal decomposition of space and allows us to reduce the problem to manageable size.
- The representer is one way to derive the Kalman filter.

Final Thought

- Data assimilation is a highly technical subject
- When you understand the technical aspects, you are at the *beginning*, *not the end* of the subject.