

Review on Mathematical Perspective for Data Assimilation Methods: Least Square Approach

Muhammed Eltahan

Aerospace Engineering, Cairo University, Cairo, Egypt

Email: muhammed_eltahan@hotmail.com

How to cite this paper: Eltahan, M. (2017) Review on Mathematical Perspective for Data Assimilation Methods: Least Square Approach. *Journal of Applied Mathematics and Physics*, 5, 1589-1606.
<https://doi.org/10.4236/jamp.2017.58131>

Received: July 27, 2017

Accepted: August 28, 2017

Published: August 31, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Environmental systems including our atmosphere oceans, biological... etc. can be modeled by mathematical equations to estimate their states. These equations can be solved with numerical methods. Initial and boundary conditions are needed for such of these numerical methods. Predication and simulations for different case studies are major sources for the great importance of these models. Satellite data from different wide ranges of sensors provide observations that indicate system state. So both numerical models and satellite data provide estimation of system states, and between the different estimations it is required the best estimate for system state. Assimilation of observations in numerical weather models with data assimilation techniques provide an improved estimate of system states. In this work, highlights on the mathematical perspective for data assimilation methods are introduced. Least square estimation techniques are introduced because it is considered the basic mathematical building block for data assimilation methods. Stochastic version of least square is included to handle the error in both model and observation. Then the three and four dimensional variational assimilation 3dvar and 4dvar respectively will be handled. Kalman filters and its derivatives Extended, (KF, EKF, ENKF) and hybrid filters are introduced.

Keywords

Least Square Method, Data Assimilation, Ensemble Filter, Hybrid Filters

1. Introduction

In general, models could be classified as [1] [2]:

- 1) Process specific models (Causality Process) are based on conservation of laws of nature e.g.: Shallow water modeling, Navier Stokes Equation.
- 2) Data specific models (correlation based) are based on developed experimental models e.g.: Time Series models, Machine Learning, Neural Networks.

etc.

Atmospheric and oceans models are process specific models which Navier Stokes equations are the core of the solver to predict, simulate and estimate system states. Process specific models have different classifications from different perspectives, time, space, and structure of the model, in addition to another classification which is deterministic and stochastic.

It is assumed that the state of dynamical system evolves according to first order nonlinear equation

$$X_{K+1} = M(X_K) \quad (1)$$

where X_K is the current state of the system; $M(\cdot)$ is the mapping function from the Current state X at time K to the next state X at $K + 1$

- If the mapping function $M(\cdot)$ doesn't depend on the time index K then it is called a time invariant or autonomous system
- If the $M(\cdot)$ varies with time index k that is $X_{K+1} = M_K(X_K)$, then it is called time-varying system (dynamic system)
- If $X_{K+1} = MX_K$ for M non-singular matrix, then it is called time invariant linear system.
- If matrix M varies with time, that is $X_{K+1} = M_K X_K$, then it is called time varying linear or non-autonomous system.
- In the special cases when $M(\cdot)$ is an identity map, that is $M(x) = x$, then it is called static system

The mentioned above is the deterministic case; the Randomness in the model can enter in three ways:

(A) Random initial conditions (B) Random forcing (C) Random Coefficients

So, a random or a stochastic model is given by:

$$X_{K+1} = M(X_K) + W_{K+1} \quad (2)$$

where the random sequences $\{W_K\}$, denotes to the external forcing; typically $\{W_K\}$ captures uncertainties in the model including model error.

So, the classes of models that are used in the data assimilation could be classified as:

- | | |
|---------------------------|------------------------|
| (1) Deterministic-Static | (3) Stochastic-Static |
| (2) Deterministic-Dynamic | (4) Stochastic-Dynamic |

The paper is organized as follow: Section 2 states the mathematical foundation for data assimilations which describes linear and nonlinear least square in addition to weighted least square and Recursive Least square. Section 3 introduces the deterministic static model linear and nonlinear cases. Section 4 shows stochastic static model: Both linear and nonlinear cases. Section 5 describes deterministic dynamic linear case and the recursive case. Section 6 explains stochastic dynamic linear, nonlinear, reduced and hybrid filters.

2. Mathematical Background

Mainly all the data assimilation techniques are based on least square estimation, so classification of the estimation problems based on different criteria. The Es-

estimation problems can be underdetermined problem when number of observations (m) are less than number of state (n) ($m < n$). And can be over determined when number of observations (m) is larger than number of states (n) ($m > n$) [1]. Also the Estimation problem can be classified according to the mapping function from the state space to observation space which can be linear or nonlinear. and nonlinearity should be handled

Other type of classification is offline or online problems. Where offline problems if the observations are known priori. In other words we have historical set of data and we treat with them. While Online/Sequential problems are computing a new estimate of the unknown state X as a function of the latest estimate and the current observation. So, online formulation is most useful in real time applications. The last type of classifications is Strong and Weak Constraints, The Strong Constraint is occurred when the estimation is performed under the perfect model assumption. While in case of allowing for errors in the model dynamics, it is considered Weak Constraint [1]

Since, Data Assimilation based on Least Square Approach [1] [2] [3] [4]. First we introduce linear version of least square then will move to nonlinear version. After that weighted version for both linear and nonlinear will be introduced.

2.1. Linear Least Square Method

Let $Z = HX$ (3)

Given the observation vector Z and the mapping function/Interpolation matrix H (Full Rank) find the unknown state vector X .

Then the error Vector which represents the difference between the Observations Z and the required estimated state X can be represented as follow

$$e = Z - HX \quad (4)$$

The Previous term is called also error term or Innovation term. So we need to find the best estimate that minimizes the error. So the problem is converted from Estimation Problem to Optimization Problem. One of the basic classifications for problem is over determined Problem ($m > n$) or under determined problem ($m < n$) and we consider first over determined case ($m > n$). To measure The vector e using The Euclidean norm and the fact of Minimization of norm is equal to minimization of square norm. The cost function will be as follow

$$J(x) = \|Z - HX\|_2^2 = (Z - HX)^T (Z - HX) \quad (5)$$

And minimization of X is Obtained under the following condition $\nabla J(x) = 0$. This lead to

$$X = (H^T H)^{-1} H^T Z \quad (6)$$

And this equation was known as normal equation method. In case of over determined problem $m > n$. While in case under determined case ($m < n$). The estimation for X will be as follow.

$$X = H^T (HH^T)^{-1} Z \quad (7)$$

And this undetermined problem where m number of observation is less than n , number of unknowns is known in geophysical (physical process physical properties of the earth and its surrounding space Environment) because of the cost of collection of the observation is high. And in case uniquely determined case ($m = n$) which mean that the error e is zero. The estimated state vector X will be

$$X = H^{-1}Z \quad (8)$$

As we can see from the above three formulation that minimization for the cost function ends to Solving system of linear equation, and solving the linear system ($A = Xb$) of equations can be done using:

- Direct methods (Cholesky decomposition, Q-R decomposition, Singular Value decomposition, ...) [5].
- Iterative methods (Jacobi Method, Gauss-Seidel Method, Successive Over-Relaxation method, ...) [5].

2.2. Non Linear Least Square Method

The problem here will be: Given set of observation Z and knowing the function form h which is nonlinear function. Find the state vector X as shown

$$Z = h(X) \quad (9)$$

And the Innovation term will be

$$e = Z - h(X) \quad (10)$$

And the cost function will be

$$J(X) = \|Z - h(X)\|_2^2 = (Z - h(X))^T (Z - h(X)) \quad (11)$$

And so that, the idea here is extension for the linear case by replacing the non-linear term $h(x)$ with its Linear approximation of Taylor series expansion around an Operating Point let it X_c and in this case it is called first order approximation for nonlinear least square as follow

$$h(X) = h(X_c) + D_h(X_c)(X - X_c) \quad (12)$$

where $D_h(X_c)$ is the Jacobin matrix of h which is an $m \times n$ matrix given by:

$$D_h(X_c) = \nabla h(X_c) = \left[\frac{\partial h_i}{\partial X_j} \right] \text{ where } 1 \leq i \leq m; 1 \leq j \leq n. \quad (13)$$

And by substitution of first order approximation of Taylor series expansion in the cost function and giving it name $Q_1(x)$ and the index one refer for first order

$$Q_1(x) = J(X) = \|Z - h(X)\|_2^2 = (z - h(X) - D_h(X_c)(X - X_c))^T (z - h(X) - D_h(X_c)(X - X_c)) \quad (14)$$

And simplifying the notation by define $g(X) = (Z - h(X))$

$$Q_1(x) = J(X) = \|Z - h(X)\|_2^2 \quad (15)$$

$$= (g(x) - D_h(X_c)(X - X_c))^T (g(x) - D_h(X_c)(X - X_c))$$

And by comparing the previous equation by the linear version.

$$J(x) = \|Z - HX\|_2^2 = (Z - HX)^T (Z - HX) \quad (16)$$

You will find that every Z was replaced by $g(x)$ and every H is replaced by $D_h(X_c)$ also every X is replaced by $(X - X_c)$

And if the gradient is $\nabla Q_1 = 0$; we will obtain

$$X - X_c = [D_h^T(X_c)D_h^T(X_c)]^{-1} [D_h^T(X_c)g(X_c)] \quad (17)$$

And this is iterative approach and given an initial value for X_c and solving the above equation using direct methods or iterative methods then iterate again until $X - X_c <$ prescribed threshold

And by similarity the second order algorithm for non Linear Least Square will be same steps except that Taylor series expansion will be full Quadratic approximation

$$h(X) = h(X_c) + D_h(X_c)(X - X_c) + \frac{1}{2}D_h^2(X_c)((X_c)(X - X_c))^2 \quad (18)$$

where the $D_h^2(X_c) = \nabla^2 h(X)$ is the hessian of $h(X)$ and by substituting this second order expansion in the nonlinear cost function and giving it name $Q_2(x)$ since the index 2 refer to second order approximation and Putting the gradient $\nabla Q_2 = 0$ equal zero you will get the following

$$X - X_c = [D_h^T(X_c)D_h^T(X_c) + g(X_c)\nabla^2 h(X)]^{-1} [D_h^T(X_c)g(X_c)] \quad (19)$$

And this is iterative approach and given an initial value for X_c and solving the above equation using direct methods or iterative methods then iterate again until $X - X_c <$ prescribed threshold. All the above methods were for deterministic least square where

$$Z = HX \quad \text{or} \quad Z = h(X) \quad (20)$$

2.3. The Weighted Least Square Method

If there additive random noise was V presented. Which mean the observation is noisy with as follow

$$Z = HX + V \quad \text{or} \quad Z = h(X) + V \quad (21)$$

- Mean $E(V) = \text{zero}$ which mean that the instrument is well calibrated Un-biased and If it is $\neq \text{zero}$ which mean that there is bias (for example under or over reading).
- Covariance $\text{Cov}(V) = R$, The covariance matrix for the instrument. which it is property for the instrumentation.

So, for the linear form of Least Square, the cost function will be for linear form will be

$$J(x) = \|Z - HX\|_{R^{-1}}^2 = (Z - HX)^T R^{-1} (Z - HX) \quad (22)$$

And the non Linear form

$$J(x) = \|Z - h(X)\|_{R^{-1}}^2 = (Z - h(X))^T R^{-1} (Z - h(X)) \quad (23)$$

Also by the same methodology that used above, the best estimate form for Stochastic linear least square after equality for the Gradient to zero will be

$$X = (H^T R^{-1} H)^{-1} H^T R^{-1} Z \quad (24)$$

And for the stochastic nonlinear least square first order approximation will be

$$X - X_c = [D_h^T(X_c) R^{-1} D_h^T(X_c)]^{-1} [D_h^T(X_c) R^{-1} g(X_c)] \quad (25)$$

And for the stochastic nonlinear least square Second order approximation will be

$$X - X_c = [D_h^T(X_c) R^{-1} D_h^T(X_c) + b(X_c) \nabla^2 h(X_c)]^{-1} [D_h^T(X_c) R^{-1} g(X_c)] \quad (26)$$

where $R^{-1} g(X_c) = b(X_c)$ for simplifying the notation.

2.4. The Recursive Least Square Estimation Approach (Offline/Online Approach)

All the above analysis was assumed that the number m of observations is fixed. which mean by another term it is offline version of least square. So if we don't know number observation m and they are arrive sequentially in time. There're two ways to add this new observations:

- The first approach is to solve the system of linear equation repeatedly after arrival of every new observation. But this approach is very expensive from Computational point of view
- The second approach is to formulate the following problem which is based on Knowing the Optimal estimate $X^*(m)$ based on the m observations. we need to compute $X^*(m+1)$ based on $m+1$ observation. In more clear words we need to calculate to reach to formulation that can compute the new estimate function of the old estimate plus sequantional term. And this approach is called situational or recursive framework

It was known as mentioned above that the optimal linear least square estimate for $Z = HX$ is $X^*(m) = (H^T H)^{-1} H^T Z$. Let Z_{m+1} be the new observation then the $Z = HX$ can be expanded in the form of matrix-vector relation as:

$$\begin{bmatrix} Z \\ Z_{m+1} \end{bmatrix} = \begin{bmatrix} H \\ h_{m+1}^T \end{bmatrix} X \quad (27)$$

So, the Innovation will be:

$$e_{m+1}(X) = \begin{bmatrix} Z \\ Z_{m+1} \end{bmatrix} - \begin{bmatrix} Z \\ Z_{m+1} \end{bmatrix} X \quad (28)$$

So, the Cost function will be:

$$J(X) = e_{m+1}^T e_{m+1} \quad (29)$$

$$J(X) = (Z - HX)^T (Z - HX) + (Z_{m+1} - h_{m+1}^T X)^T (Z_{m+1} - h_{m+1}^T X) \quad (30)$$

So, by taking the gradient and equal it to zero $\nabla J(x) = 0$. We can get the following

$$X^*(m+1) = X^*(m) + K_{m+1} h_{m+1} [Z_{m+1} - h_{m+1}^T X^*(m)] \quad (31)$$

where

$$K_m = H^T H, \quad K_{m+1} = K_m + h_{m+1} h_{m+1}^T \quad (32)$$

The cost of computation for the second term of this equation is less than solving the all system again.

The basic building block for understanding the Data assimilation based on Least Square approach was introduced.

3. Deterministic-Static Models

In Atmospheric science, when we want to assimilate observation data to the model at time step, Two source of information are available, one of them after mapping the observation Z to X and the other source is the Prior information.

The Linear/Non Linear Case

The formulation for the problem, we need the best estimate for X given the two source of information

- The first source is the given the Observation Z and mapping function H , where the innovation term was $e = Z - h(X)$
- The Second source is given the Prior or Background information X_B where the innovation term is $e = X - X_B$

So the cost function for this case will

$$J(X) = J_B(X) + J_o(X) \quad (33)$$

For linear Case $h(.) = H$

$$J(X) = \frac{1}{2}(X - X_B)^T (X - X_B) + \frac{1}{2}(Z - HX)^T (Z - HX) \quad (34)$$

For Nonlinear case $h(.) = h(X)$

$$J(X) = \frac{1}{2}(X - X_B)^T (X - X_B) + \frac{1}{2}(Z - h(X))^T (Z - h(X)) \quad (35)$$

4. Stochastic-Static Models

The formulation for the problem will be the same as in Deterministic/Static problem, we need the best estimate for X given the two source of information

- The first source is the given the Observation Z and mapping function H , where the innovation term was $e = Z - h(X)$ and the observation has noise with its covariance \mathbf{R}

- The Second source is given the Prior or Background information X_B where the innovation term is $e = X - X_B$, and the Background has its Covariance **B**

So the cost function for this case will

$$J(X) = J_B(X) + J_o(X) \tag{36}$$

For linear Case: $h(.) = H$

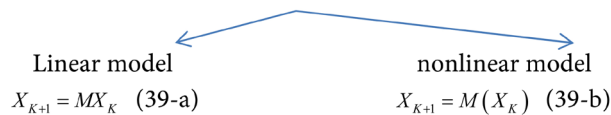
$$J(X) = \frac{1}{2}(X - X_B)^T B^{-1}(X - X_B) + \frac{1}{2}(Z - HX)^T R^{-1}(Z - HX) \tag{37}$$

For Nonlinear case: $h(.) = h(X)$

$$J(X) = \frac{1}{2}(X - X_B)^T B^{-1}(X - X_B) + \frac{1}{2}(Z - h(X))^T R^{-1}(Z - h(X)) \tag{38}$$

5. Deterministic-Dynamic

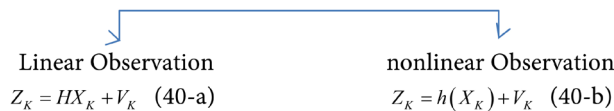
The Dynamical models can be classified as:



where X_k is the state of the dynamical system, so if X_o the initial condition is known, so computing the X_k is a forward problem

It is assumed that X_o is not known. And Estimating X_o based on noisy indirect information is the inverse problem that it is required to solve.

The observations also can be classified as:



And assume the V_k is White noise, has zero mean and Known covariance matrix R , which depend on the nature and the type of instruments used

So, formulation for statement of problem is: given set of noisy observations and the model equations, it is required to estimate the initial

Condition X_o that give best fit between the background states and noisy observations

To conclude there are four different types of problems:

- 1) Linear Model-Linear observation
- 2) Linear model-non Linear observation
- 3) Nonlinear model-linear observation
- 4) Non linear model-Non linear observation

We will consider only one case only the which is the simplest formulation with both model and observations are linear and the other cases could be checked in [1] [2].

The Definition of the cost function which is weighted sum of the squared errors is given as:

For linear case

$$J(X) = \frac{1}{2} \sum_{i=1}^N (Z_{Ki} - HX_{Ki})^T R^{-1} (Z_{Ki} - HX_{Ki}) \quad (41)$$

For nonlinear case

$$J(X) = \frac{1}{2} \sum_{i=1}^N (Z_{Ki} - h(X_{Ki}))^T R^{-1} (Z_{Ki} - h(X_{Ki})) \quad (42)$$

Depend on the whether the observations are linear or not linear. And the goal is to minimize the $J(X)$ w.r.t X_o

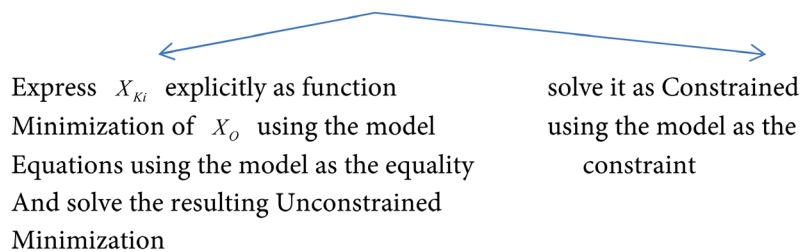
In case there background will be included for linear case

$$J(X) = \frac{1}{2} (X - X_B)^T B^{-1} (X - X_B) + \frac{1}{2} \sum_{i=1}^N (Z_{Ki} - HX_{Ki})^T R^{-1} (Z_{Ki} - HX_{Ki}) \quad (43)$$

And for nonlinear case

$$J(X) = \frac{1}{2} (X - X_B)^T B^{-1} (X - X_B) + \frac{1}{2} \sum_{i=1}^N (Z_{Ki} - h(X_{Ki}))^T R^{-1} (Z_{Ki} - h(X_{Ki})) \quad (44)$$

There are two approaches for minimization of this cost functions



5.1. Deterministic-Dynamic Linear Case

5.1.1. Linear Case-Method of Elimination

This method is mainly based on substitute the following equation $X_K = M_K X_o$ in the cost function $J(X)$ then, get the following equation

$$J(X_o) = \frac{1}{2} \sum_{i=1}^N (Z_{Ki} - HM_K X_o)^T R^{-1} (Z_{Ki} - HM_K X_o) \quad (45)$$

Then we get the gradient for the $\nabla J(X_o) = 0$

For simplicity

$$J(X_o) = \frac{1}{2} X_o^T A X_o - b^T X_o + C \quad \text{and it is quadratic in } X_o \quad (46)$$

$$A = \sum_{i=1}^N M_K^T H^T R^{-1} H M_K \quad (46-a)$$

$$b = \sum_{i=1}^N M_K^T H^T R^{-1} Z_K \quad (46-b)$$

$$C = \frac{1}{2} \sum_{i=1}^N M_K^T R^{-1} Z_K \quad (46-c)$$

So, the gradient is $\nabla J(X_o) = A X_o - b = 0$ which leads to

$$X_o = A^{-1} b \quad (47)$$

But, this approach is not practical, since it involves matrix-matrix products in the computation of A and b , so there is need to another way

5.1.2. Linear Case-Lagrangian Multipliers Formulation

Define Lagrangian L :

$$L(X_o, \lambda) = \underbrace{\frac{1}{2} \sum_{i=1}^N (Z_K - HX_K)^T R^{-1} (Z_K - HX_K)}_{\text{Objective function}} + \underbrace{\sum_{i=1}^N \lambda_K^T (X_K - MX_{K-1})}_{\text{Model}} \quad (48)$$

And the Necessary conditions for the minimum

$$\nabla_{X_o} L = 0 \quad (49-a)$$

$$\nabla_{X_K} L = 0 \quad (49-b)$$

$$\nabla_{\lambda_K} L = 0 \quad (49-c)$$

$$\text{So } \nabla_{\lambda_K} L = 0 \rightarrow X_K - MX_{K-1} = 0 \rightarrow X_K = MX_{K-1} \quad (50-a)$$

$$\nabla_{X_K} L = 0 \rightarrow H^T R^{-1} [HX_K - Z_K] + \lambda_K - M^T \lambda_{K+1} = 0 \quad (50-b)$$

$$\nabla_{X_N} L = 0 \rightarrow H^T R^{-1} [HX_N - Z_N] + \lambda_N = 0 \quad (50-c)$$

Defining:

$$f_K = H^T R^{-1} [Z_K - HX_K] = \text{normalized forecast error viewed from model space} \quad (51)$$

And substitute it in the last two equations

$$-f_K + \lambda_K - M^T \lambda_{K+1} = 0 \rightarrow \lambda_K = M^T \lambda_{K+1} + f_K \quad (52-a)$$

$$-f_N + \lambda_N = 0 \rightarrow \lambda_N = f_N \quad (52-b)$$

So, as shown that the formulation for calculate λ is backward relation. we can iterate backward starting from λ_N to Compute λ_1 . And this technique is known as backward adjoint dynamics. Then after getting λ_1 substitute in $\nabla_{X_o} J(X_o) = -M^T \lambda_1$ and get the gradient then use any minimization algorithm to get X_o , The 4-D Var Algorithm (First order adjoint method) can be summarized as follow:

- 1) Start with an arbitrary X_o , and compute the model solution using $X_{K+1} = MX_K$
- 2) Given the observation $\{Z_K\}$ where $1 \leq K \leq N$
- 3) Set $\lambda_N = f_N$ and Solve $\lambda_K = M^T \lambda_{K+1} + f_K$ to find λ_1
- 4) Compute the gradient $\nabla_{X_o} J(X_o) = -M^T \lambda_1$
- 5) Use this gradient in minimization algorithm to find the optimal X_o by repeating the steps 1 through 4 until convergence.

5.2. Recursive Least Squares Formulation of 4D Var (Online Approach)

In the previous part of 4Dvar section the solution for off-line 4Dvar problem of assimilating given set of observations in deterministic-dynamic model using classical least square method.

Now there is need to develop an online or recursive method for computing the estimate of the state as new observations arrive. which mean we need to compute the X_{N+1} in terms of X_N and the new observation Z_{N+1} .

Consider linear deterministic dynamical system without model noise

$$X_{K+1} = MX_K \quad (53)$$

where the initial condition X_o is random variable with $E(X_o) = m_o$ and $Cov(X_o) = P_o$. and the observations Z_K for $K = 1, 2, 3, \dots$ are given as

$$Z_K = H_K X_K + V_K \quad (54)$$

where H_K is full rank and V_K is the observation vector noise with the following known properties

$$E(V_k) = 0 \quad \text{and} \quad Cov(V_K) = R_K \quad (55)$$

So, the objective function that

$$J_N = J_N^p + J_N^o \quad (56)$$

where

$$J_N^p = \frac{1}{2} (m_o - X_o)^T P_o^{-1} (m_o - X_o) \quad (57)$$

$$J_N^o = \frac{1}{2} \sum_{K=1}^N (Z_K - H_K X_K)^T R_K^{-1} (Z_K - H_K X_K) \quad (58)$$

Since, our goal to find an optimal X_N that minimize the J_N , it is needed to express m_o and X_K

In tem of the corresponding N values X_N and m_N . So,

$$\text{Since, } X_N = M_{N-1} M_{N-2} \dots M_K X_K = M(N-1:K) X_K \quad (59)$$

$$\text{Then } X_K = M^{-1}(N-1:K) X_N = B(N-1:K) X_N \quad (60)$$

$$\text{Since } m_{K+1} = M_K m_K = M(K:0) m_o \quad (61)$$

Hence, the trajectory of the model starting from $m_o = B(N-1:0) m_N$

Substitute for X_K and m_o into J_N

$$\begin{aligned} J_N(X_N) &= \frac{1}{2} (m_N - X_N)^T \left[B^T(N-1:0) P_0^{-1} B(N-1:0) \right] (m_N - X_N) \\ &\quad + \frac{1}{2} \sum_{K=1}^N (Z_K - H_K B(N-1:K) X_N)^T R_K^{-1} (Z_K - H_K B(N-1:K) X_N) \end{aligned} \quad (62)$$

And differentiate $J_N(X_N)$ w.r.t X_N twice to get the gradient and Hessian. Then setting the gradient to zero and simplifying the notation

$$(F_N^p + F_N^o) X_N = (f_N^p + f_N^o) \quad (63)$$

where

$$F_N^p = B^T(N-1:0) P_0^{-1} B(N-1:0) \quad (64)$$

$$F_N^o = \sum_{K=1}^N B^T(N-1:K) H_K^T R_K^{-1} H_K B(N-1:K) \quad (65)$$

$$f_N^p = F_N^p m_N \quad (66)$$

$$f_N^o = \sum_{K=1}^N B^T(N-1:K) H_K^T R_K^{-1} Z_K \quad (67)$$

By induction the minimization for X_{N+1} of $J_{N+1}(X_{N+1})$ is given by

$$(F_{N+1}^p + F_{N+1}^o)X_{N+1} = (f_{N+1}^p + f_{N+1}^o) \tag{68}$$

The goal of the recursive framework is to express X_{N+1} as function of X_N and Z_{N+1} . this calls when expressing $F_{N+1}^p, F_{N+1}^o, f_{N+1}^p, f_{N+1}^o$ in terms of $F_N^p, F_N^o, f_N^p, f_N^o$.

So using equations from (61) to (64) and $B_K = M_K^{-1}$ and the following equation

$$M(j:i) = \begin{cases} M_j M_{j-1}, \dots, M_{I+1} M_I & \text{if } j \geq i \\ & \text{if } j < i \end{cases} \tag{69}$$

Then $N + 1$ formula in terms of N can be get

$$F_{N+1}^p + F_{N+1}^o = B_N^T [F_N^p + F_N^o] B_N + H_{N+1}^T R_{N+1}^{-1} H_{N+1} \tag{70}$$

$$f_{N+1}^p + f_{N+1}^o = B_N^T F_N^p B_N m_{N+1} + B_N^T f_N^o + H_{N+1}^T R_{N+1}^{-1} Z_{N+1} \tag{71}$$

And since

$$(P_{N+1})^{-1} = F_{N+1}^p + F_{N+1}^o \tag{72}$$

$$(P_N)^{-1} = F_N^p + F_N^o \tag{73}$$

$$B_N m_{N+1} = B_N M_N m_N = m_N \tag{74}$$

$$X_{N+1}^f = M_N X_N \text{ or } X_N = B_N X_{N+1}^f \tag{75}$$

$$f_{N+1}^p + f_{N+1}^o = B_N^T (P_N)^{-1} B_N X_{N+1}^f + H_{N+1}^T R_{N+1}^{-1} Z_{N+1} \tag{76}$$

By combining Equations (70) and (76) into Equation (68) and defining $(P_{N+1}^f)^{-1} = B_N^T (P_N)^{-1} B_N$ after combination

$$X_{N+1} = \left[(P_{N+1}^f)^{-1} + H_{N+1}^T R_{N+1}^{-1} H_{N+1} \right]^{-1} \left[(P_{N+1}^f)^{-1} X_{N+1}^f + H_{N+1}^T R_{N+1}^{-1} Z_{N+1} \right] \tag{77}$$

The right hand side of this equation is sum of two terms the first one is

$$\left[(P_{N+1}^f)^{-1} + H_{N+1}^T R_{N+1}^{-1} H_{N+1} \right]^{-1} \left[(P_{N+1}^f)^{-1} X_{N+1}^f \right] \tag{78}$$

So adding and subtracting $H_{N+1}^T R_{N+1}^{-1} H_{N+1} X_{N+1}^f$ this term is equal to

$$\begin{aligned} & \left[(P_{N+1}^f)^{-1} + H_{N+1}^T R_{N+1}^{-1} H_{N+1} \right]^{-1} \left[(P_{N+1}^f)^{-1} + H_{N+1}^T R_{N+1}^{-1} H_{N+1} - H_{N+1}^T R_{N+1}^{-1} H_{N+1} \right] X_{N+1}^f \\ & = X_{N+1}^f - K_{N+1} H_{N+1} X_{N+1}^f \end{aligned} \tag{79}$$

where

$$K_{N+1} = \left[(P_{N+1}^f)^{-1} + H_{N+1}^T R_{N+1}^{-1} H_{N+1} \right]^{-1} H_{N+1}^T R_{N+1}^{-1} \text{ is called kalman gain matrix and}$$

combine it with Equation (77). The desired recursive expression will be gained

$$X_{N+1} = X_{N+1}^f + K_{N+1} [Z_{N+1} - H_{N+1} X_{N+1}^f] \tag{80}$$

6. Stochastic-Dynamic Model

This type of data assimilation problems is same as deterministic/dynamic prob-

lems, except that, this type introduces an additional term in the forecast equation which is noise vector that associated with the model (*i.e.* model error).

For Stochastic-dynamic model we can divide the filters to Linear and Nonlinear filters Linear filter has evolution function $M(\cdot)$ in the model and the mapping function $h(\cdot)$ is linear while in Nonlinear filters those two functions are nonlinear

6.1. Linear Filters

Kalman Linear Filter

Kalman filter approach was first introduced at reference [6] [7]

Problem formulation:

This section will show how model and observation with error will be presented then will formulate the algorithm:

A1-Dynamic model: it will be assumed linear, non autonomous, dynamical system that evolves according to

$$X_{K+1} = M_K X_K + w_{K+1} \quad (81)$$

where M_K is nonsingular system matrix that varies with time K and w_K denote to model error. It is assumed that X_o and w_K satisfy the following conditions (A) X_o is random variable with known mean vector $E(X_o) = m_o$ and known covariance matrix $E[(X_o - m_o)(X_o - m_o)^T] = P_o$ (B) The model error is unbiased, mean $E(w_K) = 0$ for all k , and temporally uncorrelated (white noise) $E(w_K w_j^T) = Q_K$ when if $j = k$ and $E(w_K w_j^T) = 0$ otherwise (C) The model error w_K and the initial state is uncorrelated $E(w_K X_o^T) = 0$ for all k

B1-Observations: The observation Z_K is the observation at time k and related to X_K via

$$Z_K = H_K X_K + V_K \quad (82)$$

where H_K represent the time varying measurement system and V_K represent the measurement noise with the following properties (A) V_K has mean zero $E(V_K) = 0$ (B) V_K is temporary uncorrelated: $E(V_K V_j^T) = R_K$ if $j = k$ while otherwise $E(V_K V_j^T) = 0$. (C) V_K is uncorrelated with the initial state X_o and the model error w_K which mean $E(X_o V_K^T) = 0$ for all $K > 0$ and $E(V_K W_j^T) = 0$ for all K and j

C1-statement of the filtering problem: Given that X_K evolves according to equation 81 and set of observations. Our goal to find an estimate X_K that minimize the mean square error

And the following is the summary of the Kalman filter procedure (Covariance Form)

Model $x_{k+1} = M_k x_k + w_{k+1}$

$$E(w_k) = 0, \quad Cov(w_k) = Q_k$$

x_0 is random with mean m_0 and $Cov(x_0) = P_0$

Observation $z_k = H_k x_k + v_k$

$$E(v_k) = 0, \quad Cov(v_k) = R_k$$

Model Forecast $\hat{x}_0 = E(x_0)$, $\hat{P}_0 = P_0$

$$x_k^f = M_{k-1} \hat{x}_{k-1}$$

$$P_k^f = M_{k-1} \hat{P}_{k-1} M_{k-1}^T + Q_k$$

Data Assimilation

$$\hat{x}_k = x_k^f + K_k [z_k - H_k x_k^f]$$

$$K_k = P_k^f H_k^T [H_k P_k^f H_k^T + R_k]^{-1} = \hat{P}_k H_k^T D_k^{-1}$$

$$\hat{P}_k = P_k^f - P_k^f H_k^T [H_k P_k^f H_k^T + R_k]^{-1} H_k P_k^f = [I - K_k H_k] P_k^f$$

The computation of the covariance matrices P_{k+1}^f and P_{k+1} is the most time-consuming part since in many of the applications $n > m$. so that reduced order filters has been introduced [1].

6.2. Non Linear Filters

6.2.1. First Order Filter/Extended Kalman Filter (EKF)

The extended kalman filter is an extension for kalman filter idea in case the evolution function $M(\cdot)$ in the model and the mapping function $h(\cdot)$ is non linear

For nonlinear model

$$X_{K+1} = M(X_K) + w_{K+1}$$

For nonlinear observation

$$Z_K = h(X_K) + V_K$$

The main idea for First order filter/extended kalman filter is to expand the $M(X_K)$ around

X_K and $h(X_{K+1})$ around X_{K+1}^f in first order taylor series expansion. when $M(X)$ and $h(X)$ are linear it reduces to kalman filter. The following is the summary steps for Extended Kalman filter

Model $x_{k+1} = M(x_k) + w_{k+1}$

Observation $z_k = h(x_k) + v_k$

Forecast Step $x_{k+1}^f = M(\hat{x}_k)$

$$P_{k+1}^f = D_M \hat{P}_k D_M^T + Q_{k+1}$$

Data Assimilation Step

$$\hat{x}_{k+1} = x_{k+1}^f + K [z_{k+1} - h(x_{k+1}^f)]$$

$$K = P_{k+1}^f D_h^T [D_h P_{k+1}^f D_h^T + R_{k+1}]^{-1}$$

$$\hat{P}_{k+1} = (I - K D_h) P_{k+1}^f$$

6.2.2. Second Order Filter

The second order filter is the same idea of the first order filter except that the expansion of $M(X_K)$ around X_K and $h(X_{K+1})$ around X_{K+1}^f in Second order Taylor series expansion. And the following is summary of the second or-

der nonlinear filter

Model $x_{k+1} = M(x_k) + w_{k+1}$

Observation $z_k = h(x_k) + v_k$

Forecast Step

$$x_{k+1}^f = M(\hat{x}_k) + \frac{1}{2} \partial^2 (M, \hat{P}_k)$$

$$P_{k+1}^f = D_M \hat{P}_k D_M^T + Q_{k+1}$$

Data Assimilation Step

$$\hat{x}_{k+1} = x_{k+1}^f + K \left[z_{k+1} - h(x_{k+1}^f) - \frac{1}{2} \partial^2 (h, P_{k+1}^f) \right]$$

$$K = P_{k+1}^f D_h^T [D_h P_{k+1}^f D_h^T + R_{k+1}]^{-1}$$

$$\hat{P}_{k+1} = (I - KD_h) P_{k+1}^f$$

6.3. Reduced Rank Filters

Ensemble Kalman Filter

The ensemble Kalman filter [8] originated from the merger of Kalman filter theory and Monte Carlo estimation methods. It was introduced the basic principles of linear and nonlinear filtering but these are not used in day by day operations at the national centers for weather predictions. Because of the cost of the updating the covariance matrix was very high.

So, there are mainly two ways to avoid the high cost of computing the covariance matrix

- 1) The first method was the Parallel computation which mainly dependent on
 - a) The Algorithm
 - b) The number of Processors
 - c) The topology of the interconnection of the network
 - d) How the tasks of the algorithm are mapped on the processor
- 2) The second method which became more popular which to compute low/reduced rank approximation to the full rank covariance matrix. and most of low rank filters differs only in the way in which the approximation are derived. Excellent review on the Ensemble Kalman filter was introduced

Formulation of the problem

It is assumed that the model is nonlinear and observations are linear functions of the state

$$X_{K+1} = M(X_K) + w_{K+1}$$

$$Z_K = h(X_K) + V_K$$

And it is assumed that

- 1) The initial conditions $X_0 \sim N(m_0, P_0)$
- 2) The dynamic system noise w_k is white Gaussian noise with $w_k \sim N(0, Q_k)$
- 3) The observation noise V_K is White noise with $V_K \sim N(0, R_K)$
- 4) $X_0, \{w_k\}, \{V_k\}$ are mutually uncorrelated

Model $x_{k+1} = M(x_k)x_k + w_{k+1}$

Observation $z_k = H_k x_k + v_k$

Initial ensemble

- Create the initial ensemble

Forecast step

1) Create the ensemble of forecasts at time $(k + 1)$ using the following

The N members of the ensemble forecast at time $(k + 1)$ are generated

$$\xi_{k+1}^f(i) = M(\hat{\xi}_k(i)) + w_{k+1}(i) \text{ where } w_{k+1}(i) \sim N(0, Q_{k+1})$$

2) Compute $x_{k+1}^f(N)$ and $P_{k+1}^f(N)$ using

$$K = P_{k+1}^f(N) H_{k+1}^T [H_{k+1} P_{k+1}^f(N) H_{k+1}^T + R_{k+1}]^{-1}$$

$$P_{k+1}^f(N) = \frac{1}{N-1} \sum_{i=1}^N e_{k+1}^f(i) [e_{k+1}^f(i)]^T$$

Data assimilation step

1) Create the ensemble of estimates at time $(k + 1)$ using

$$\hat{\xi}_{k+1}(i) = \xi_{k+1}^f(i) + K [z_{k+1}(i) - H_{k+1} \xi_{k+1}^f(i)].$$

and

$$K = P_{k+1}^f(N) H_{k+1}^T [H_{k+1} P_{k+1}^f(N) H_{k+1}^T + R_{k+1}]^{-1}$$

2) Compute $\hat{x}_{k+1}(N)$ and $\hat{P}_{k+1}(N)$ using

The sample mean of the estimate at time $(k + 1)$ is then given by

$$\hat{x}_{k+1}(N) = \frac{1}{N} \sum_{i=1}^N \hat{\xi}_{k+1}(i) = x_{k+1}^f(N) + K [\bar{z}_{k+1}(N) - H_{k+1} x_{k+1}^f(N)]$$

where

$$\bar{z}_{k+1}(N) = \frac{1}{N} \sum_{i=1}^N z_{k+1}(i) = z_{k+1} + \frac{1}{N} \sum_{i=1}^N v_k(i) = \bar{z}_{k+1} + \bar{v}_{k+1}(N)$$

where

$$P_{k+1}(N) = (I - KH_{k+1}) P_{k+1}^f(N) (I - KH_{k+1})^T + KR_{k+1}(N) K^T$$

where for large N

$$R_{k+1}(N) = (1/N - 1) [v_{k+1}(i) - \bar{v}_{k+1}(N)] [v_{k+1}(i) - \bar{v}_{k+1}(N)]^T$$

All summaries, derivation and details could be checked in reference [1] [9]. and full review on ensemble kalman filter for atmospheric data assimilation is introduced by P. L. Houtekamer and Fuqing Zhang, 2016 [10].

6.4. Hybrid Filters

3DVar uses static climate Background error while 4DVar uses implicit flow dependent information but still start with static background error. And since. The B-matrix affects the performance of the assimilation heavily [11] it is important to use a B-matrix that is a realistic representation of the actual forecast error covariance [12]. So many proposed hybrid filters were introduced. they are to use flow dependent background error in variational data assimilation system by com-

binning the 3Dvar climate background error covariance and error of the day from ensemble.

In Equation (37) replace the B by weighted sum of 3Dvar B and the ensemble covariance as follow [13]:

$$B = a_1 B_1 + a_2 B_2$$

where

$$a_1 = 1 - a_2$$

The Ensemble covariance is included in the 3DVAR cost function through Augmentation of control variables [14] and the following formula is mathematically equivalent to [13].

This is well known hybrid 3DVar-EnKF method. While 4DVar-EnKF method has the same idea if we substitute by Equation (43) by the same methodology in 3DVar-EnKF part. More advanced hybrid filters are highlighted in ref [15] [16].

7. Conclusions

This paper shows the mathematical perspective for the basic foundation of data assimilation modules starting from least square to advanced filters that used in data assimilation as journey. This work is the first of its type to summarize the mathematical perspective for data assimilation in extensive way and highlights both classical and advanced data assimilation methods. This paper could be used as reference to understand the mathematics behind data assimilation. It started by least square method and their different versions then explains on the classical method 3Dvar. 4DVar also is introduced. Advanced filters such kalman filter and its families were highlighted. The idea of hybrid filter was introduced finally.

For future work, detailed hybrid filters should be highlighted, since there are different hybrid filters structure were introduced. Generic case studies the evaluate performance of the different assimilation techniques.

References

- [1] Lewis, J.M. (2009) Dynamic Data Assimilation: A Least Square Approach. Cambridge University Press.
- [2] Dynamic Data Assimilation: An Introduction Prof S. Lakshmivaran, School of Computer Science, University of Oklahoma. <http://nptel.ac.in/courses/111106082/#>
- [3] Gibbs, B.P. (2011) Advanced Kalman Filtering, Least-Squares and Modeling: A Practical Handbook. Wiley.
- [4] Bjorck, A. (1996) Numerical Methods for Least Squares Problems, Linköping University, Linköping, Sweden. <https://doi.org/10.1137/1.9781611971484>
- [5] Kreyszig, E. (2011) Advanced Engineering Mathematics. <https://www-elec.inaoep.mx/~jmram/Kreyszig-ECS-DIF1.pdf>
- [6] Zarchan, P. and Musoff, H. (2000) Fundamentals of Kalman Filtering: A Practical Approach. American Inst of Aeronautics & Astronautics, United States, 2.
- [7] Kalman, R.E. (1960) A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82**, 35-45. <https://doi.org/10.1115/1.3662552>

- [8] Evensen, G. (1994) Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research*, **99**, 10143-10162. <https://doi.org/10.1029/94JC00572>
- [9] Evensen, G. (2003) The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation.
- [10] Houtekamer, P.L. and Zhang, F. (2016) Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **144**. <https://doi.org/10.1175/MWR-D-15-0440.1>
- [11] Bannister, R.N. (2008) A Review of Forecast Error Covariance Statistics in Atmospheric Variational Data Assimilation. I: Characteristics and Measurements of Forecast Error Covariances. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1951-1970. <https://doi.org/10.1002/qj.339>
- [12] Bannister, R.N. (2008) A Review of Forecast Error Covariance Statistics in Atmospheric Variational Data Assimilation. II: Modelling the Forecast Error Covariance Statistics. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1971-1996. <https://doi.org/10.1002/qj.340>
- [13] Hamill, T.M. and Snyder, C. (2000) A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme. *Monthly Weather Review*, **128**, 2905-2919. [https://doi.org/10.1175/1520-0493\(2000\)128<2905:AHEKFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2)
- [14] Lorenc, A.C. (2003) The Potential of the Ensemble Kalman Filter for NWP: A Comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **129**, 3183-3203. <https://doi.org/10.1256/qj.02.132>
- [15] Nawinda, *et al.* (2016) A Hybrid Ensemble Transform Particle Filter for Nonlinear and Spatially Extended Dynamical Systems.
- [16] Laura, S., *et al.* (2015) A Hybrid Particle-Ensemble Kalman Filter for High Dimensional Lagrangian Data Assimilation.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jamp@scirp.org