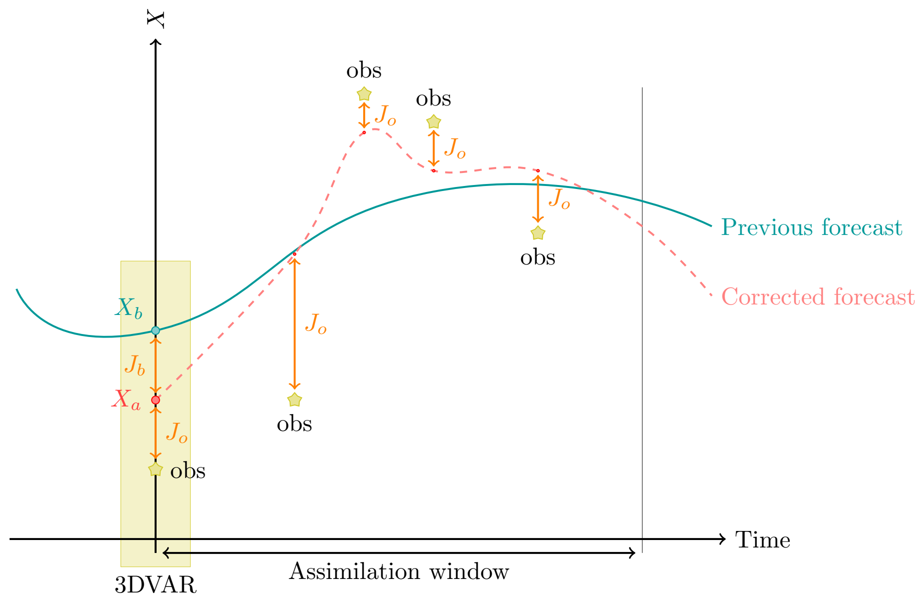
Variational Data Assimilation

Mark Asch - CSU/IMU/2023



Outline of the course (I)

Adjoint methods and variational data assimilation (4h)

1. Introduction to data assimilation: setting, history, overview, definitions.

2. Adjoint method.

3. Variational data assimilation methods:

   (a) 3D-Var,
   (b) 4D-Var.

Outline of the course (II)

Statistical estimation, Kalman filters and sequential data assimilation (4h)

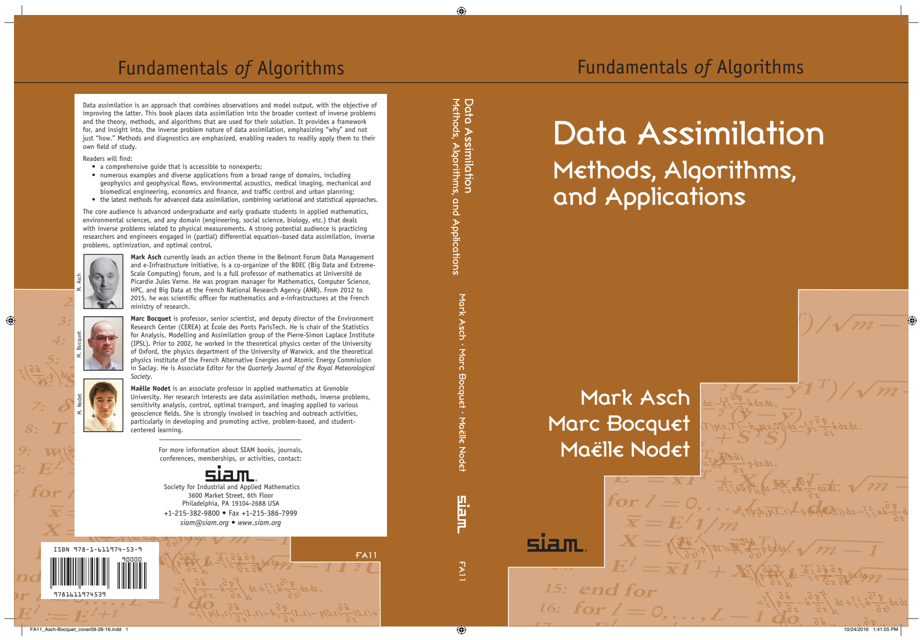
1. Introduction to statistical DA.

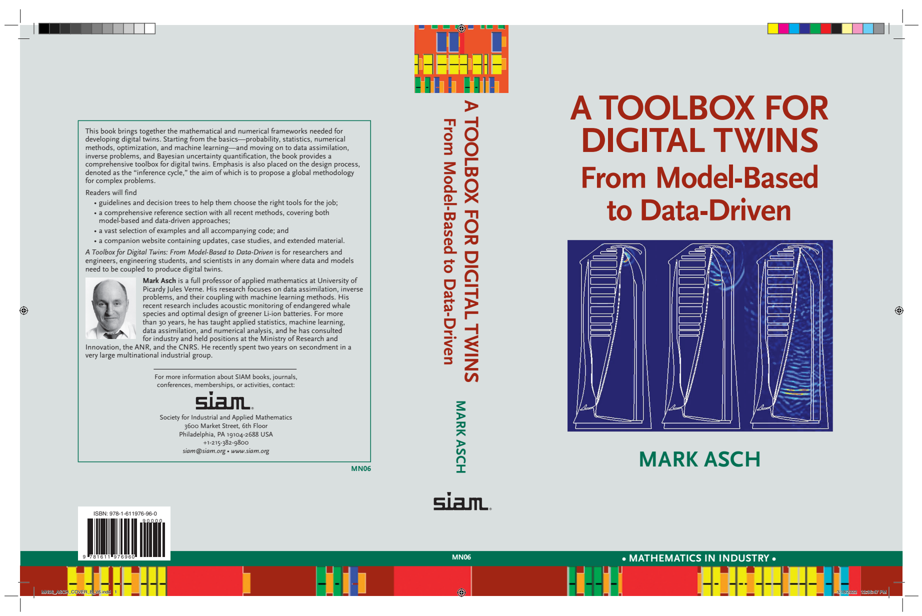
2. Statistical estimation.

3. The Kalman filter.

4. Nonlinear extensions and ensemble filters.

Reference Textbooks





Variational DA - formulation

- In variational data assimilation we describe the state of the system by

  $\Rightarrow$ a state variable, $\mathbf{x}(t) \in \mathcal{X}$, a function of space and time that

  $\Rightarrow$ represents the physical variables of interest, such as current velocity (in oceanography), temperature, sea-surface height, salinity, biological species concentration, chemical concentration, etc.

- Evolution of the state is described by a system of (in general nonlinear) differential equations in a region Ω,

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathcal{M}(\mathbf{x}) \quad \text{in } \Omega \times [0, T]\,, \\ \mathbf{x}(t = 0) = \mathbf{x}_0, \end{cases} \tag{1}$$

where the initial condition is unknown (or poorly known).

- Suppose that we are in possession of observations $\mathbf{y}(t) \in \mathcal{O}$ and an observation operator $\mathcal{H}$ that describes the available observations.

- Then, to characterize the difference between the observations and the state, we define the objective (or cost) function,

$$J(\mathbf{x}_0) = \frac{1}{2}\int_0^T \|\mathbf{y}(t) - \mathcal{H}\left(\mathbf{x}(\mathbf{x}_0, t)\right)\|_{\mathcal{O}}^2 \,\mathrm{d}t + \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}\right\|_{\mathcal{X}}^2 \tag{2}$$

where

⇒ $\mathbf{x}^{\mathrm{b}}$ is the background (or first guess)

⇒ and the second term plays the role of a regularization (in the sense of Tikhonov—see previous Lecture.

⇒ The two norms under the integral, in the finite-dimensional case, will be represented by the error covariance matrices $\mathbf{R}$ and $\mathbf{B}$ respectively, and will be described below.

$\Rightarrow$ Note that for mathematical rigor we have indicated, as subscripts, the relevant functional spaces on which the norms are defined.

- In the continuous context, the data assimilation problem is formulated as follows:

Find the analyzed state $\mathbf{x}_0^{\mathrm{a}}$ that minimizes J and satisfies

$$\mathbf{x}_0^{\mathrm{a}} = \operatorname{argmin} J(\mathbf{x}_0).$$

- The necessary condition for the existence of a (local) minimum is (as usual...)

$$\nabla J(\mathbf{x}_0^{\mathrm{a}}) = 0.$$

Variational DA - adjoint method

- Variational DA is based on an adjoint approach that is explained inthe previous Lecture.

- The particularity here is that the adjoint is used to solve an inverse problem for the unknown *initial condition*.

Variational DA - 3D Var

- Usually, 3D-Var and 4D-Var are introduced in a finite dimensional or discrete context—this approach will be used in this section.

- For the infinite dimensional or continuous case, we must use the calculus of variations and (partial) differential equations, as was done in the previous Lectures.

- We start out with the finite-dimensional version of the cost function (2),

$$
\begin{aligned}
J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right) \qquad (3) \\
+ \frac{1}{2}(\mathbf{H}\mathbf{x}-\mathbf{y})^{\mathrm{T}} \mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y}), \qquad (4)
\end{aligned}
$$

  where

- $\Rightarrow$ $\mathbf{x}$, $\mathbf{x}^{\mathrm{b}}$, and $\mathbf{y}$ are the state, the background state, and the measured state respectively;
- $\Rightarrow$ $\mathbf{H}$ is the observation matrix (a linearization of the observation operator $\mathcal{H}$);
- $\Rightarrow$ $\mathbf{R}$ and $\mathbf{B}$ are the observation and background error covariance matrices respectively.

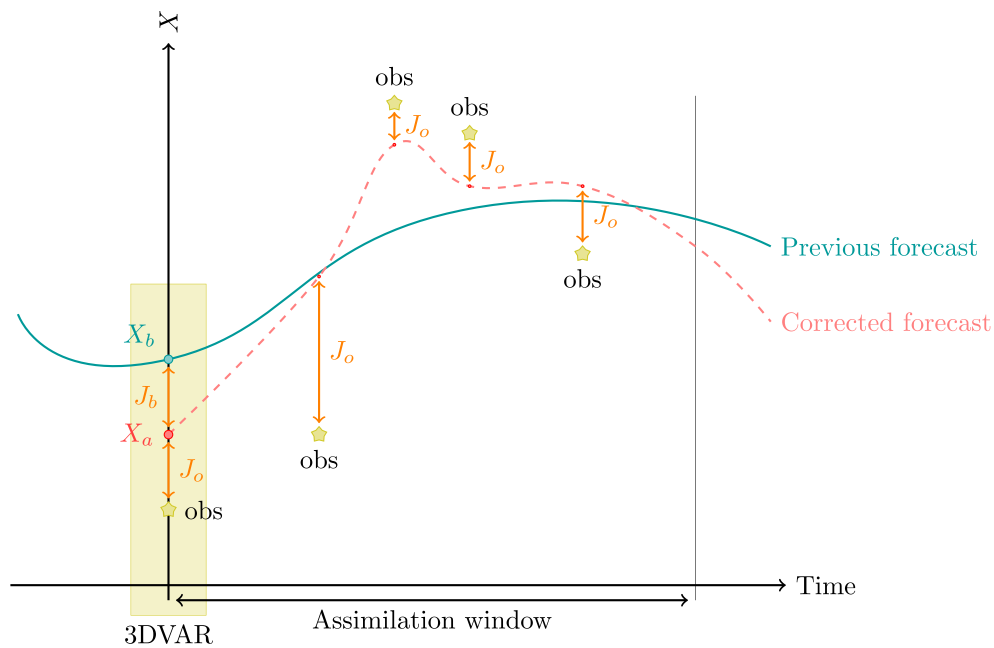
- This quadratic function attempts to strike a balance between some *a priori* knowledge about a background (or historical) state and the actual measured, or observed, state.

- It also assumes that we know and that we can invert the matrices $\mathbf{R}$ and $\mathbf{B}$. This, as we will be pointed out below, is not always obvious.

- Furthermore, it represents the sum of the (weighted) background deviations and the (weighted) observation deviations. The basic methodology is presented in the Algorithm below, which is nothing more than a classical gradient descent algorithm.

Variational DA - 3D Var Algorithm

$j = 0, \;\; x = x_0$
`while` $\|\nabla J\| > \epsilon$ `or` $j \le j_{\max}$
    `compute` J
    `compute` ∇J
    `gradient descent and update of` x_{j+1}
    $j = j + 1$
`end`



Variational DA - 3D Var (II)

- We note that when
  - $\Rightarrow$ the background $\mathbf{x}^{\mathrm{b}} = \mathbf{x}^{\mathrm{b}} + \epsilon^{\mathrm{b}}$ is available at some time t_k, together with
  - $\Rightarrow$ observations of the form $\mathbf{y} = \mathbf{H}\mathbf{x}^{\mathrm{t}} + \epsilon^{\mathrm{o}}$ that have been acquired at the same time (or over a short enough interval of time when the dynamics can be considered stationary),
  - $\Rightarrow$ then the minimization of (3) will produce an estimate of the system state at time t_k.
- In this case, the analysis is called "three-dimensional variational analysis" and is abbreviated by *3D-Var*.
- Borrowing from control theory—see [Asch2022]—the optimal gain can be shown to take the form

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1},$$

where $\mathbf{B}$ and $\mathbf{R}$ are the covariance matrices.

- We obtain the analyzed state,

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}(\mathbf{y} - \mathbf{H}(\mathbf{x}^{\mathrm{b}})).$$

- This is the state that minimizes the 3D-Var cost function.

- We can verify this by taking the gradient, term by term, of the cost function (3) and equating to zero,

$$\nabla J(\mathbf{x}^{\mathrm{a}}) = \mathbf{B}^{-1}\left(\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}}\right) - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{a}}\right) = 0, \tag{5}$$

where

$$\mathbf{x}^{\mathrm{a}} = \operatorname{argmin}\ J(\mathbf{x}).$$

- Solving the equation, we find

$$
\begin{aligned}
\mathbf{B}^{-1}\left(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{b}}\right) &= \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{H}\mathbf{x}^{\mathrm{a}}\right)\\
\left(\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{x}^{\mathrm{a}} &= \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}+\mathbf{B}^{-1}\mathbf{x}^{\mathrm{b}}\\
\mathbf{x}^{\mathrm{a}} &= \left(\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}+\mathbf{B}^{-1}\mathbf{x}^{\mathrm{b}}\right)\\
&= \left(\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\left(\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{x}^{\mathrm{b}}\right.\\
&\quad \left.-\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{x}^{\mathrm{b}}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}\right)\\
&= \mathbf{x}^{\mathrm{b}}+\left(\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{H}\mathbf{x}^{\mathrm{b}}\right)\\
&= \mathbf{x}^{\mathrm{b}}+\mathbf{K}\left(\mathbf{y}-\mathbf{H}\mathbf{x}^{\mathrm{b}}\right), \qquad (6)
\end{aligned}
$$

where we have simply added and subtracted the term $\left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{x}^{\mathrm{b}}$ in the third-last line and in the last line we have brought out what are known as the *innovation* term,

$$\mathbf{d}=\mathbf{y}-\mathbf{H}\mathbf{x}^{\mathrm{b}},$$

and the *gain matrix*,

$$\mathbf{K}=\left(\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}.$$

- This matrix can be rewritten as

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} \right)^{-1} \tag{7}$$

  using a well-known Sherman-Morrison-Woodbury formula of linear algebra that completely avoids the direct computation of the inverse of the matrix $\mathbf{B}$.

- The linear combination in (6) of a background term plus a multiple of the innovation is a classical result of linear-quadratic (LQ) control theory and shows how nicely DA fits in with and corresponds to (optimal) control theory

- The form of the gain matrix (7) can be explained quite simply.

  $\Rightarrow$ The term $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ is the background covariance transformed to the observation space.

  $\Rightarrow$ The "denominator" term $\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ expresses the sum of observation and background covariances.

$\Rightarrow$ The "numerator" term $\mathbf{BH}^{\mathrm{T}}$ takes the ratio of $\mathbf{B}$ and $\mathbf{R}+\mathbf{HBH}^{\mathrm{T}}$ back to the model space.

- This recalls (and is completely analogous to) the variance ratio

$$\frac{\sigma_{\mathrm{b}}^2}{\sigma_b^2+\sigma_{\mathrm{o}}^2}$$

that appears in the optimal BLUE (Best Linear Unbiased Estimate) that will be derived below in the statistical DA Lecture.

- This corresponds to the case of a single observation $y^{\mathrm{o}} = x^{\mathrm{o}}$ of a quantity x,

$$\begin{aligned} x^{\mathrm{a}} &= x^{\mathrm{b}} + \frac{\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{b}}^2+\sigma_{\mathrm{o}}^2}(x^{\mathrm{o}}-x^{\mathrm{b}}) \\ &= x^{\mathrm{b}} + \frac{1}{1+\alpha}(x^{\mathrm{o}}-x^{\mathrm{b}}), \end{aligned}$$

where

$$\alpha = \frac{\sigma_{\mathrm{o}}^2}{\sigma_{\mathrm{b}}^2}.$$

- In other words, the best way to estimate the state is to take a weighted average of the background (or prior) and the observations of the state. And the best weight is the ratio of the mean squared errors (variances).

- The statistical viewpoint is thus perfectly reproduced in the 3D-Var framework.

Variational DA - 4D Var

- A more realistic, but complicated situation arises when one wants to assimilate observations that are acquired over a time interval, during which the system dynamics (flow, for example) cannot be neglected.

- Suppose that the measurements are available at a succession of instants, $t_k,\ k = 0, 1, \ldots, K$ and are of the form

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\epsilon}_k^{\mathrm{o}}, \tag{8}$$

  where

  ⇒ $\mathbf{H}_k$ is a linear observation operator and
  ⇒ $\boldsymbol{\epsilon}_k^{\mathrm{o}}$ is the observation error with covariance matrix $\mathbf{R}_k$,
  ⇒ and suppose that these observation errors are uncorrelated in time.

- Now we add the dynamics described by the state equation,

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}_k, \tag{9}$$

where we have neglected any model error.[1]

- We suppose also that at time index $k = 0$ we know

$\Rightarrow$ the background state $\mathbf{x}_0^{\mathrm{b}}$ and
$\Rightarrow$ its error covariance matrix $\mathbf{P}_0^{\mathrm{b}}$
$\Rightarrow$ and we suppose that the errors are uncorrelated with the observations in (8).

- Then a given initial condition, $\mathbf{x}_0$, defines a unique model solution $\mathbf{x}_{k+1}$ according to (9).

- We can now generalize the objective function (3),

[1] This will be taken into account below.

which becomes

$$J(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}}\left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \tag{10}$$

$$+\frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right). \tag{11}$$

- The minimization of $J(\mathbf{x}_0)$ will provide the initial condition of the model that fits the data most closely.

- This analysis is called "strong constraint four-dimensional variational assimilation," abbreviated as *strong constraint 4D-Var*. The term "strong constraint" implies that the model found by the state equation (9) must be exactly satisfied by the sequence of estimated state vectors.

- In the presence of model uncertainty, the state equation becomes

$$\mathbf{x}_{k+1}^{\mathrm{t}} = \mathbf{M}_{k+1}\mathbf{x}_k^{\mathrm{t}} + \boldsymbol{\eta}_{k+1}, \qquad (12)$$

where

⇒ the model noise $\boldsymbol{\eta}_k$ has covariance matrix $\mathbf{Q}_k$,

⇒ which we suppose to be uncorrelated in time and uncorrelated with the background and observation errors.

- The objective function for the best, linear unbiased estimator (BLUE) for the sequence of states

$$\{\mathbf{x}_k,\ k = 0, 1, \ldots, K\}$$

is of the form

$$J(\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_K) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}}\left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)$$
$$+\frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right)$$
$$+\frac{1}{2}\sum_{k=0}^{K-1}\left(\mathbf{x}_{k+1} - \mathbf{M}_{k+1}\mathbf{x}_k\right)^{\mathrm{T}}\mathbf{Q}_{k+1}^{-1}\left(\mathbf{x}_{k+1} - \mathbf{M}_{k+1}\mathbf{x}_k\right). \quad (13)$$

- This objective function has become a function of the complete sequence of states

$$\{\mathbf{x}_k,\ k = 0, 1, \ldots, K\},$$

and its minimization is known as "weak constraint four-dimensional variational assimilation," abbreviated as *weak constraint 4D-Var*.

- Equations (10) and (13), with an appropriate reformulation of the state and observation spaces, are special cases of the BLUE objective function.

Variational DA - 4D Var Applications

- All the above forms of variational assimilation, as defined by (3), (10) and (13), have been used for real-world data assimilation, in particular in meteorology and oceanography.

- However, these methods are directly applicable to a vast array of other domains, among which we can cite

  ⇒ geophysics and environmental sciences,
  ⇒ seismology,
  ⇒ atmospheric chemistry, and terrestrial magnetism.
  ⇒ Many other examples exist.

- We remark that in real-world practice, variational assimilation is performed on nonlinear models. If the extent of the nonlinearity is sufficiently small (in some sense), then variational assimilation, even if it does not solve the correct estimation problem, will still produce useful results.

Variational DA - 4D Var Implementation

- Now, our problem reduces to
  - ⇒ quantifying the covariance matrices and then, of course,
  - ⇒ computing the analyzed state.

- The quantification of the covariance matrices must result from extensive data studies, or the use of a Kalman filter approach—see below.

- The computation of the analyzed state will be described next—this will not be done directly, but rather by an adjoint approach for minimizing the cost functions.

- There is of course the inverse of $\mathbf{B}$ or $\mathbf{P}^{\mathrm{b}}$ to compute, but we remark that there appear only matrix-vector products of $\mathbf{B}^{-1}$ and $\left(\mathbf{P}^{\mathrm{b}}\right)^{-1}$ and we can thus define operators (or routines) that compute these efficiently without the need for large storage capacities.

Variational DA - 4D Var Adjoint

- We explain the adjoint approach in the case of strong constraint 4D-Var, taking into account a completely general nonlinear setting for the model and for the observation operators.

- Let $\mathbf{M}_k$ and $\mathbf{H}_k$ be the nonlinear model and observation operators respectively.

- We reformulate (9) and (10) in terms of the nonlinear operators as

$$\begin{aligned} J(\mathbf{x}_0) &= \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \\ &+ \frac{1}{2}\sum_{k=0}^{K} \left(\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right)^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right), \end{aligned} \tag{14}$$

with the dynamics

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\left(\mathbf{x}_k\right), \quad k = 0, 1, \ldots, K-1. \tag{15}$$

- The minimization problem requires that we now compute the gradient of J with respect to $\mathbf{x}_0$.

- The gradient is determined from the property that, for a given perturbation $\delta\mathbf{x}_0$ of $\mathbf{x}_0$, the corresponding first-order variation of J is

$$\delta J = \left(\nabla_{\mathbf{x}_0} J\right)^{\mathrm{T}} \delta\mathbf{x}_0. \tag{16}$$

- The perturbation is propagated by the tangent linear equation,

$$\delta\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\delta\mathbf{x}_k, \quad k = 0, 1, \ldots, K-1, \tag{17}$$

  obtained by differentiation of the state equation (15), where $\mathbf{M}_{k+1}$ is the Jacobian matrix (of first-order partial derivatives) of $\mathbf{x}_{k+1}$ with respect to $\mathbf{x}_k$.

- The first-order variation of the cost function is obtained similarly by differentiation of (14),

$$\delta J = \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1} \delta \mathbf{x}_0 \tag{18}$$

$$+ \sum_{k=0}^{K} \left(\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right)^{\mathrm{T}} \mathbf{R}_k^{-1} \mathrm{H}_k \delta \mathbf{x}_k, \tag{19}$$

where H_k is the Jacobian of $\mathbf{H}_k$ and $\delta \mathbf{x}_k$ is defined by (17).

⇒ This variation is a compound function of $\delta \mathbf{x}_0$ that depends on all the $\delta \mathbf{x}_k$'s.

⇒ But if we can obtain a direct dependence on $\delta \mathbf{x}_0$ in the form of (16), eliminating the explicit dependence on $\delta \mathbf{x}_k$, then we will (as in the previously seen examples) arrive at an explicit expression for the gradient $\nabla_{\mathbf{x}_0} J$ of our cost function J.

⇒ This will be done, as we have done before, by introducing an adjoint state and requiring that it satisfy certain conditions—namely, the adjoint equation. Let us now proceed with this program.

- We begin by defining, for $k = 0, 1, \dots, K$, the adjoint state vectors $\mathbf{p}_k$ that belong to the dual of the state space.

- Now we take the null products (according to the tangent state equation (17)),

$$\mathbf{p}_k^{\mathrm{T}} \left(\delta\, \mathbf{x}_k - \mathbf{M}_k \delta \mathbf{x}_{k-1}\right),$$

  and subtract them from the right-hand side of the cost function variation (18),

$$\delta J = \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1} \delta$$

$$\mathbf{x}_0 + \sum_{k=0}^{K} \left(\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right)^{\mathrm{T}} \mathbf{R}_k^{-1} \mathrm{H}_k \delta \mathbf{x}_k$$

$$- \sum_{k=0}^{K} \mathbf{p}_k^{\mathrm{T}} \left(\delta \mathbf{x}_k - \mathbf{M}_k \delta \mathbf{x}_{k-1}\right).$$

- Rearranging the matrix products, using the symmetry of $\mathbf{R}_k$ and regrouping terms in $\delta \mathbf{x}.$, we obtain,

$$\begin{aligned}
\delta J \quad = \quad & \left[\left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) + \mathrm{H}_0^{\mathrm{T}} \mathbf{R}_0^{-1} \left(\mathbf{H}_0(\mathbf{x}_0) - \mathbf{y}_0\right) + \mathbf{M}_0^{\mathrm{T}} \mathbf{p}_1 \right] \delta \mathbf{x}_0 \\
& + \left[\sum_{k=1}^{K-1} \mathrm{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right) - \mathbf{p}_k + \mathbf{M}_k^{\mathrm{T}} \mathbf{p}_{k+1} \right] \delta \mathbf{x}_k \\
& + \left[\mathrm{H}_K^{\mathrm{T}} \mathbf{R}_K^{-1} \left(\mathbf{H}_K(\mathbf{x}_K) - \mathbf{y}_K\right) - \mathbf{p}_K \right] \delta \mathbf{x}_k .
\end{aligned}$$

- Notice that this expression is valid for any choice of the adjoint states $\mathbf{p}_k$ and, in order to "kill" all $\delta \mathbf{x}_k$ terms, except $\delta \mathbf{x}_0$, we must simply impose that,

$$\mathbf{p}_K = \mathrm{H}_K^{\mathrm{T}} \mathbf{R}_K^{-1} \left(\mathbf{H}_K(\mathbf{x}_K) - \mathbf{y}_K\right), \tag{20}$$

$$\mathbf{p}_k = \mathrm{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right) + \mathbf{M}_k^{\mathrm{T}} \mathbf{p}_{k+1}, \quad k = K-1, \ldots, 1, \tag{21}$$

$$\mathbf{p}_0 = \left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) + \mathrm{H}_0^{\mathrm{T}} \mathbf{R}_0^{-1} \left(\mathbf{H}_0(\mathbf{x}_0) - \mathbf{y}_0\right) + \mathbf{M}_0^{\mathrm{T}} \mathbf{p}_1 \tag{22}$$

- We recognize the backward, adjoint equation for $\mathbf{p}_k$ and the only term remaining in the variation of J is then

$$\delta J = \mathbf{p}_0^{\mathrm{T}} \delta \mathbf{x}_0,$$

so that $\mathbf{p}_0$ is the sought for gradient, $\nabla_{\mathbf{x}_0} J$, of the objective function with respect to the initial condition $\mathbf{x}_0$ according to (16).

- The system of equations (20)-(22) is the adjoint of the tangent linear equation (17).

- The term *adjoint* here corresponds to the transposes of the matrices $\mathrm{H}_k^{\mathrm{T}}$ and $\mathbf{M}_k^{\mathrm{T}}$ that, as we have seen before, are the finite-dimensional analogues of an adjoint operator.

Variational DA - 4D Var Algorithm

- We can now propose the "usual" algorithm for solving the optimization problem by the adjoint approach:

1. For a given initial condition $\mathbf{x}_0$, integrate forwards the (nonlinear) state equation (15) and store the solutions $\mathbf{x}_k$ (or use some sort of checkpointing).

2. From the final condition, (20), integrate backwards in time the adjoint equations (21).

3. Compute directly the required gradient (22).

4. Use this gradient in an iterative optimization algorithm to find a (local) minimum.

- The above description for the solution of the 4D-Var data assimilation problem clearly covers the case of 3D-Var, where we seek to minimize (3). In this case, we only need the transpose Jacobian $\mathbf{H}^{\mathrm{T}}$ of the observation operator.

Variational DA - roles of $\mathbf{R}$ and $\mathbf{B}$

- The relative magnitudes of the errors due to measurement and background provide us with important information as to how much "weight" to give to the different information sources when solving the assimilation problem.

- For example, if background errors are larger than observation errors, then the analyzed state, solution to the DA problem, should be closer to the observations than to the background and vice-versa.

- The background error covariance matrix, $\mathbf{B}$, plays an important role in DA. This is illustrated in the following examples.

3D- and 4D-VAR EXAMPLES

3D-VAR - effect of a single observation

- Suppose that we have a single observation, at a point corresponding to the j-th element of the state vector.

- The observation operator is then

$$\mathbf{H} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}^{\mathrm{T}}.$$

- The gradient of J is

$$\nabla J = \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^{\mathrm{b}}\right) + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{x} - \mathbf{y}^{\mathrm{o}}\right).$$

- Since it must be equal to zero at the minimum $\mathbf{x}^{\mathrm{a}}$, we must have

$$\left(\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}}\right) = \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{a}}\right).$$

- But $\mathbf{R} \doteq \sigma^2$, $\mathbf{H}\mathbf{x}^{\mathrm{a}} = x_j^{\mathrm{a}}$ and $\mathbf{B}\mathbf{H}^{\mathrm{T}}$ is the j-th column of $\mathbf{B}$ whose elements are denoted by $B_{i,j}$ with $i = 1, \ldots, n$.

- So we see that

$$\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}} = \frac{y^{\mathrm{o}} - x_k^{\mathrm{a}}}{\sigma^2} \begin{pmatrix} B_{1,j} \\ B_{2,j} \\ \vdots \\ B_{n,j} \end{pmatrix}.$$

- The increment is proportional to a column of $\mathbf{B}$.

- The choice of $\mathbf{B}$ is thus crucial and will determine how this observation provides information about what happens around the j-th variable.

- In the 4D-Var case, the increment at time t will be proportional to a single column of $\mathbf{M}\mathbf{B}\mathbf{M}^{\mathrm{T}}$ which describes the error covariances of the background at the time of the observation t.

4D-VAR - for a single observation

- We consider an example with
  - ⇒ a single observation at time step 3
  - ⇒ and a known background at time step 0.
- In this case, the 4D-Var cost function (10) for determining the initial state becomes scalar,

$$J(x_0) = \frac{1}{2}\frac{\left(x_0 - x_0^{\mathrm{b}}\right)^2}{\sigma_B^2} + \frac{1}{2}\sum_{k=1}^{K}\frac{\left(x_k - x_k^{\mathrm{o}}\right)^2}{\sigma_R^2},$$

  where σ_B^2 and σ_R^2 are the (known) background and observation error variances respectively.

- With a single observation, say at time step 3, the cost function is

$$J(x_0) = \frac{1}{2}\frac{\left(x_0 - x_0^{\mathrm{b}}\right)^2}{\sigma_B^2} + \frac{1}{2}\frac{\left(x_3 - x_3^{\mathrm{o}}\right)^2}{\sigma_R^2}.$$

- The minimum is reached when the gradient of J disappears,

$$J'(x_0) = 0,$$

which can be computed as

$$\frac{(x_0 - x_0^{\mathrm{b}})}{\sigma_B^2} + \frac{(x_3 - x_3^{\mathrm{o}})}{\sigma_R^2}\frac{\mathrm{d}x_3}{\mathrm{d}x_2}\frac{\mathrm{d}x_2}{\mathrm{d}x_1}\frac{\mathrm{d}x_1}{\mathrm{d}x_0} = 0. \quad (23)$$

- We now require a dynamic relation between the x_k's in order to compute the derivatives.

- To this end, let us take the most simple linear forecast model,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\alpha x,$$

with α a known positive constant.

- This is a typical model for describing decay, for example of a chemical compound whose behavior over time is then given by

$$x(t) = x(0)\mathrm{e}^{-\alpha t}.$$

- To obtain a discrete representation of the dynamics, we can use an upstream finite difference scheme,

$$x(t_{k+1}) - x(t_k) = (t_{k+1} - t_k)\left[-\alpha x(t_{k+1})\right], \tag{24}$$

which can be rewritten in the explicit form

$$x(t + \Delta t) = \left(\frac{1}{1 + \alpha \Delta t}\right) x(t)$$

and we have assumed a fixed time-step $\Delta t = t_{k+1} - t_k$ for all k.

- We thus have the scalar relation

$$x_{k+1} = M(x_k) = \gamma x_k \tag{25}$$

where the constant

$$\gamma = \frac{1}{1+\alpha\Delta t}.$$

- The necessary condition (23) then becomes

$$\frac{\left(x_0 - x_0^{\mathrm{b}}\right)}{\sigma_B^2} + \frac{\left(x_3 - x_3^{\mathrm{o}}\right)}{\sigma_R^2}\gamma^3 = 0.$$

- This can be solved for x_0 and then for x_3 to obtain the analyzed state,

$$\begin{aligned}
x_0 &= x_0^{\mathrm{b}} + \frac{\gamma^3\sigma_B^2}{\sigma_R^2}\left(x_3^{\mathrm{o}} - x_3\right)\\
&= x_0^{\mathrm{b}} + \frac{\gamma^3\sigma_B^2}{\sigma_R^2}\left(x_3^{\mathrm{o}} - \gamma^3 x_0\right)\\
&= \frac{\sigma_R^2}{\sigma_R^2+\gamma^6\sigma_B^2}x_0^{\mathrm{b}} + \frac{\gamma^3\sigma_B^2}{\sigma_R^2+\gamma^6\sigma_B^2}x_3^{\mathrm{o}}\\
&= x_0^{\mathrm{b}} + \frac{\gamma^3\sigma_B^2}{\sigma_R^2+\gamma^6\sigma_B^2}\left[x_3^{\mathrm{o}} - \gamma^3 x_0^{\mathrm{b}}\right],
\end{aligned}$$

where we have added and subtracted x_0^{b} to obtain the last line and used the system dynamics (25).

- Finally, by again using the dynamics, we find the 4D-Var solution,

$$x_3 = \gamma^3 x_0^{\mathrm{b}} + \frac{\gamma^6 \sigma_B^2}{\sigma_R^2 + \gamma^6 \sigma_B^2} \left[x_3^{\mathrm{o}} - \gamma^3 x_0^{\mathrm{b}} \right]. \quad (26)$$

- Let us examine some asymptotic cases.

  ⇒ If the parameter α tends to zero then the dynamic gain γ tends to one and the model becomes stationary, with

$$x_{k+1} = x_k.$$

  The solution then tends to the 3D-Var case, with (see above)

$$x_3 = x_0 = x_0^{\mathrm{b}} + \frac{\sigma_B^2}{\sigma_R^2 + \sigma_B^2} \left[x_3^{\mathrm{o}} - x_0^{\mathrm{b}} \right]. \quad (27)$$

- ⇒ If the model is stationary, we can thus use all observations whenever they become available, exactly as in the 3D-Var case.
- ⇒ The other asymptotic occurs when the step-size tends to infinity and the dynamic gain goes to zero. The dynamic model becomes

$$x_{k+1} = 0$$

  with the initial condition $x_0 = x_0^{\mathrm{b}}$ and there is thus no connection between states at different time steps.
- ⇒ Finally, if the observation is perfect, then $\sigma_R^2 = 0$ and

$$x_3 = x_3^{\mathrm{o}}.$$

  But there is no link to x_0 and there is once again no dynamical connection between states at two different instants.

- This example will be repeated below using a Kalman filter and will bring out the equivalence between the two approaches—4D-Var and statistical—under the hypothesis of a noise-free model.

Codes

Various open-source repositories and codes are available for both academic and operational data assimilation.

1. DARC: `https://research.reading.ac.uk/met-darc/` from Reading, UK.
2. DAPPER: `https://github.com/nansencenter/DAPPER` from Nansen, Norway.
3. DART: `https://dart.ucar.edu/` from NCAR, US, specialized in ensemble DA.
4. OpenDA: `https://www.openda.org/`.
5. Verdandi: `http://verdandi.sourceforge.net/` from INRIA, France.

6. PyDA: `https://github.com/Shady-Ahmed/PyDA`, a Python implementation for academic use.

7. Filterpy: `https://github.com/rlabbe/filterpy`, dedicated to KF variants.

8. EnKF; `https://enkf.nersc.no/`, the original Ensemble KF from Geir Evensen.

References

1. K. Law, A. Stuart, K. Zygalakis. *Data Assimilation. A Mathematical Introduction*. Springer, 2015.

2. G. Evensen. *Data assimilation, The Ensemble Kalman Filter*, 2nd ed., Springer, 2009.

3. A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* SIAM. 2005.

4. O. Talagrand. Assimilation of observations, an introduction. *J. Meteorological Soc. Japan*, **75**, 191–209, 1997.

5. F.X. Le Dimet, O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus,* **38**(2), 97–110, 1986.

6. J.-L. Lions. Exact controllability, stabilization and perturbations for distributed systems. *SIAM Rev.*, **30**(1):1–68, 1988.

7. J. Nocedal, S.J. Wright. *Numerical Optimization*. Springer, 2006.

8. F. Tröltzsch. *Optimal Control of Partial Differential Equations*. AMS, 2010.