MET6308: Zou

- Answer the following questions:

(a) What is atmospheric data assimilation?
(b) List 5-6 main classes of algorithms used in data analysis and assimilation.
(c) What is the role of adjoint model in data assimilation? Why is it important?
(d) Explain homogeneous, isotropic, posteriori weights, and observation operator.

Sol

(a) The procedure which takes observed data and creates homogeneous fields is usually called analysis, or assimilation when the data are distributed in time and the procedures uses an explicit dynamical model for the time evolution of the atmospheric flow.

(b)
① Subjective analysis
② function fitting
③ successive correction
④ optimal (statistical) Interpolation
⑤ variational method (3DVAR, 4DVAR)

(c) The adjoint model can produce the gradient of any forecast aspects with respect to intial condition. This is the only way to get ∇J in atmospheric data assimilation. So, we can obtain the minimization of J using one of several methods. Therefore, adjoint model is prerequisite to obtain ∇J.

(d) homogeneous assumption can be expressed as follows

$$B_{k,l}(\vec{r}_k, \vec{r}_l) = B_{kl}(r, \overset{\text{angle}}{\phi})$$

isotropic assumption is

$$B_{k,l}(\vec{r}_k, \vec{r}_l) = E_b^2 \, \overset{\text{correlation}}{\rho_B}(r)$$

posteriori weights

★ For example, after applying the function fitting or successive corrections, we will have the posteriori weights function like $U^{ana} = W_1 U_1^o + W_2 U_2^o$, so using observed data in some analysis (or assimilation), this weight function can be determined by some posteriori information, while the priori weights are coming from the analysis field (background) itself.

observation operator

$$\text{Analysis of model state} \underset{\text{retrieval or adjoint}}{\overset{\text{observation operator}}{\rightleftarrows}} \text{(Indirect) Observation}$$

(a) Atmospheric data assimilation deals with problems related to merging the measurements with a numerical forecasting model

The objective of 3DVAR/4DVAR is to determine the optimal initial conditions of a numerical weather prediction model by fitting the background information (the model dynamics and physics) to data over an interval of time, where the optimality is measured by a cost function.

* obs. operator ⟶ see (pp.29) (big page #)

$$\underset{\text{Obs.}}{Y} = H \underset{\text{parameter to estimate}}{X} + \overset{\text{observational noise}}{E}$$

H → observation operator

Operator which interpolates X into the obs location

and so on

MET6308: Zou

- For a linear equation

$$\frac{\partial \delta \mathbf{x}}{\partial t} = A(t,\mathbf{x})\delta \mathbf{x}$$
$$\delta \mathbf{x}\big|_{t=t_0} = \delta \mathbf{x}_0 \qquad (1)$$

answer the following questions:

(1) What is the initial condition?
(2) What's the Jacobian operator?
(3) What is the resolvent of eq. (1) (give the definition)?
(4) What is the adjoint equation of eq. (1)?
✓(5) What is the resolvent of the adjoint equation of eq. (1)?
✓(6) Is it true that the resolvent of the adjoint equation is the adjoint of the resolvent of the tangent linear equation (1)? Can you prove it (this is optional)?

Sol)

(1) The initial condition is $\delta\vec{X}|_{t=0} = \delta\vec{X}_0$, i.e. pertulation at time t_0

(2) $A(t,\vec{X})$: Jacobian operator

(3) $\delta\vec{X}(t) = \boxed{R(t,t_0)}\delta\vec{X}_0$ ✓

$\hookrightarrow$ resolvent

(4)
$$\begin{cases} -\dfrac{\partial\hat{X}}{\partial t} = (A(t,\vec{X}))^T\hat{X} \\ \hat{X}|_{t_R} = \text{forcing} \end{cases}$$

(5)
$$\left(\delta J = \int_t <2(\vec{X}-\vec{X}^{obs}), R(t,t_0)\delta\vec{X}_0>dt = \int_t <\boxed{R^*(t,t_0)}2(\vec{X}-\vec{X}^{obs}), \delta\vec{X}_0>dt\right)$$

$\hookrightarrow$ resolvent of adjoint model

(6) Tangent linear eq.

$$\frac{\partial\delta X}{\partial t} = A\delta X \quad (1)$$

resolvent of TGL model

$$\boxed{\delta X(t) = \underline{R(t,t_0)}\delta X(t_0)} \quad (3)$$

adjoint eq

$$-\frac{\partial\delta\hat{X}}{\partial t} = A^T\delta\hat{X} \quad (2)$$

resolvent of adjoint model

$$\boxed{\delta\hat{X}(t_0) = \underline{S(t_0,t)}\delta\hat{X}(t)} \quad (4)$$

$$<\delta\vec{X}(t), \delta^*\vec{X}(t)>$$

$$\frac{d}{dt}<\delta\vec{X}(t), \delta^*\vec{X}(t)> = <\frac{d\delta\vec{X}(t)}{dt}, \delta^*\vec{X}(t)> + <\delta\vec{X}(t), \frac{d}{dt}\delta^*X(t)>$$
$$\overset{(1)}{=} <A(t,\vec{X})\delta\vec{X}(t), \delta^*\vec{X}(t)> + <\delta\vec{X}(t), -\overset{(2)}{A^*}(t,\vec{X})\delta^*\vec{X}(t)>$$
$$= <A\delta\vec{X}(t), \delta^*X(t)> - <A\delta\vec{X}(t), \delta^*X(t)> \equiv 0$$

t_0: $\delta\vec{X}(t_0)$, $\delta^*\vec{X}(t_0) \overset{(4)}{=} S(t_0,t_i)\delta^*X(t_i)$

t_i: $\delta\vec{X}(t_i) \overset{(3)}{=} R(t_i,t_0)\delta X(t_0)$ and $\delta^*X(t_i)$

$$<\delta\vec{X}(t_0), S(t_0,t_i)\delta^*\vec{X}(t_i)>$$
$$= <R(t_i,t_0)\delta\vec{X}(t_0), \delta^*\vec{X}(t_i)>$$
$$= <\delta X(t_0), R^*(t_i,t_0)\delta^*\vec{X}(t_i)>$$

$\therefore \underline{S = R^*}$

MET6308: Zou (98', 99')

- To use a standard unconstrained minimization software, what are required from users to provide to the minimization routine in order to find the minimum of a given cost function:

$$J(\mathbf{x}_0) \qquad (2)$$

where $\mathbf{x}_0$ is a vector of dimension N serving as the control variable.

Just we need J and ∇J

☆ • Suppose J in eq. (2) is a cost function defined on the space R to which the control variable $\mathbf{x}_0$ belongs, and $<,>$ represents an inner product defined on R. The gradient ∇J of J with respect to the control variable $\mathbf{x}_0$ is defined as a vector in the same space R as $\mathbf{x}_0$, and is such that the first-order variation δJ resulting from a perturbation ~~of~~ $\delta \mathbf{x}_0$ ~~is~~ of $\mathbf{x}_0$ is equal to the inner product

$$\delta J \equiv \langle \nabla J, \delta \mathbf{x}_0 \rangle \qquad (3)$$

Circle the right statements in the following:

[✓ (a)] ∇J reflects how sensitive is J with respect to $\mathbf{x}_0$; ⟶ True

(b) The value of ∇J depends on perturbation $\delta \mathbf{x}_0$; ⟶ False

[✓ (c)] The value of ∇J depends on how the inner product is defined on R; ⟶ True

MET6308: Zou

- Let the inner produce [,] represent the L_2 norm, i.e.,

$$[\mathbf{x},\mathbf{y}] = \sum_i^N x_i y_i \qquad (4)$$

Assume that the inner product < , > is related to the inner product [,] through the following relation:

$$\langle \mathbf{x},\mathbf{y}\rangle = [\mathbf{x}, \mathbf{W}\mathbf{y}] \qquad (5)$$

$\delta J = \langle \nabla J, \delta x_0 \rangle$

where **W** is a weighting matrix. What is the relationship between the gradient ∇J associated with the inner produce < , > and the gradient $\nabla^{L_2} J$ associated with the inner product [,]?

Sol)

one method

$$[x, y] = \sum_i x_i y_i$$

$$\langle x, y\rangle = [x, Wy] = \sum_i x_i W_i y_i = \sum_i W_i^T x_i y_i$$

$$\delta J_1 = [x, y] \Rightarrow \nabla J_1 = x$$

$$\delta J_2 = \langle x, y\rangle \Rightarrow \nabla J_2 = W^T x = W^T \nabla J_1$$

$$\therefore \nabla J = W^T \nabla^{L_2} J$$

second method

based on [,]

$$\delta J = [\nabla^{L_2} J, \delta \vec{x}_0] = (\delta x_0)^T \nabla^{L_2} J$$

based on < , >

$$\delta J = \langle \nabla J, \delta \vec{x}_0 \rangle = [\nabla J, W \delta \vec{x}_0] = (W \delta x_0)^T \nabla J$$

$$= (\delta x_0)^T W^T \nabla J = [W^T \nabla J, \delta \vec{x}_0]$$

$$\therefore \nabla J = W^T \nabla^{L_2} J$$

$$\langle x, y\rangle = [x, Wy]$$

① Let $J_1 = \langle x - x^o, x - x^o \rangle$

$$\delta J_1 = \langle 2(x-x^o), \delta x \rangle$$

$$\delta x(t) = R(t, t_0)\delta x_0$$

$$\delta J_1 = \langle 2(x-x^o), R(t. t_0)\delta x_0 \rangle$$

$$= \langle \underbrace{R^*(t_0, t) 2(x-x^o)}_{\nabla^{L_2} J}, \delta x_0 \rangle$$

$$\therefore \nabla^{L_2} J = R^*(t_0, t) 2(x-x^o)$$

② Let $J_2 = \langle (x-x^o), W(x-x^o) \rangle$

$$= \langle W^T(x-x^o), (x-x^o) \rangle$$

$$\delta J_2 = \langle W^T 2(x-x^o), \delta x \rangle$$

$$= \langle W^T 2(x-x^o), R(t, t_0)\delta x_0 \rangle$$

$$= \langle \underbrace{R^*(t_0, t) W^T 2(x-x^o)}_{\nabla J}, \delta x_0 \rangle$$

Therefore

$\nabla J = W^T \nabla^{L_2} J$ (W^T is not func of time)

MET6308: Zou ('98, '99)

✓ * • Prove that using the optimal interpolation algorithm, there is no correlation between the analysis error and the observation increment $x^o\left(\vec{r}_k\right) - x^b\left(\vec{r}_k\right)$.

Hint: Given the background value $x^b\left(\vec{r}_i\right)$ of x at the analysis gridpoint $\vec{r}_i$, and the observed and background values $x^o\left(\vec{r}_k\right)$ and $x^b\left(\vec{r}_k\right)$ at the observation location $\vec{r}_k$, $k =$ 1,…,K, the optimal interpolation algorithm calculates the analysis $x^a\left(\vec{r}_i\right)$ using the following formula:

$$x^a\left(\vec{r}_i\right) = x^b\left(\vec{r}_i\right) + \sum_{k=1}^{K} W_{ik}\left(x^o\left(\vec{r}_k\right) - x^b\left(\vec{r}_k\right)\right) \qquad (6)$$

Where K is the total number of observation points affecting the analysis of x at the analysis point $\vec{r}_i$ and W_{ik} is the weighting determined by minimizing the expected analysis error variance:

$$\overline{\left(x^a\left(\vec{r}_i\right) - x'\left(\vec{r}_i\right)\right)^2} \qquad (7)$$

Where the overbar means the expectation and $x'\left(\vec{r}_i\right)$ is the true value of x at $\vec{r}_i$.

mine

analysis error $= X^a(\vec{r}_i) - X^t(\vec{r}_i) \equiv A$

observation increment $= X^o(\vec{r}_k) - X^b(\vec{r}_k) \equiv B$

$E\{(A-B)(A-B)\}$ has to be zero

$= E\{(X^a(\vec{r}_i) - X^t(\vec{r}_i) - X^o(\vec{r}_k) + X^b(\vec{r}_k))^2\} = E\{(X^a(\vec{r}_i) - X^t(\vec{r}_i) - (\sum_{k=1}^{K} W_{ik})^{-1}(X^a(\vec{r}_i) - X^b(\vec{r}_i)))^2\} = 0$

where we assume background errors and observation errors are not correlated and observation errors are uncorrelated

$$\Gamma = \overline{(X^a(\vec{r}_i) - X^t(\vec{r}_i))(X^o(\vec{r}_k) - X^b(\vec{r}_k))}$$

$$X^a(\vec{r}_i) = X^b(\vec{r}_i) + \sum_{k=1}^{K} W_{ik}(X^o(\vec{r}_k) - X^b(\vec{r}_k))$$

?

$$\sigma_a^2 = \overline{(X^o(\vec{r}_i) - X^t(r_i))^2}$$

$$\frac{\partial \sigma_a^2}{\partial W_{ik}} = 0 = 2\,\overline{(X^a(\vec{r}_i) - X^t(\vec{r}_i))(X^o(\vec{r}_k) - X^b(\vec{r}_k))}$$

↳ see back ?

MET6308: Zou

- Following is a linearized shallow-water model written in term of vorticity ζ, divergence D and geopotential height ϕ:

$$\left.\begin{aligned} &\frac{\partial\zeta}{\partial t}+f_0D=0\\ &\frac{\partial D}{\partial t}-f_0\zeta+\nabla^2\phi=0\\ &\frac{\partial\phi}{\partial t}+\tilde{\phi}\,D=0 \end{aligned}\right\} \quad (8)$$

where f_0 is a constant Coriolis parameter, $\tilde{\phi}=g\tilde{h}$, $\phi=gh$ is the deviation from $\tilde{\phi}$, h is the height of the free surface of the fluid, g the gravitational constant, $\tilde{h}$ a free surface height independent of space and time representing a basic state at rest, $\nabla^2=\partial^2/\partial x^2+\partial^2/\partial y^2$ is the horizontal Laplacian operator in Cartesian coordinates.

Derive the adjoint model of (8).

Hint: Use the *Green's second theorem*

$$\int_{(R)}\left(\psi\nabla^2\phi-\phi\nabla^2\psi\right)dR=\int_{(S)}\left(\psi\nabla\phi-\phi\nabla\psi\right)\cdot\mathbf{n}\,dS \qquad (9) \quad =0$$

where (R) is the two-dimensional space domain, (S) is the bounding surface of (R), $\mathbf{n}$ is the outward normal direction of S, ψ and ϕ are two scalar functions defined on (R), and ∇ is the gradient operator.

Sol)

$$\left(\begin{aligned} &\frac{\partial\zeta}{\partial t}+f_0D=0\\ &\frac{\partial D}{\partial t}-f_0\zeta+\nabla^2\phi=0\\ &\frac{\partial\phi}{\partial t}+\tilde{\phi}D=0 \end{aligned}\right.$$

Derive adjoint

$$L=\iint\left\{\hat{\zeta}\left(\frac{\partial\zeta}{\partial t}+f_0D\right)+\hat{D}\left(\frac{\partial D}{\partial t}-f_0\zeta+\nabla^2\phi\right)+\hat{\phi}\left(\frac{\partial\phi}{\partial t}+\tilde{\phi}D\right)\right\}dtdx$$

$$L=\iint\left\{-\zeta\frac{\partial\hat{\zeta}}{\partial t}+\hat{\zeta}f_0D-D\frac{\partial\hat{D}}{\partial t}-\hat{D}f_0\zeta\right\}+\phi\nabla^2\hat{D}-\phi\frac{\partial\hat{\phi}}{\partial t}+\hat{\phi}\tilde{\phi}D\}dtdx$$

$$\frac{\partial L}{\partial\zeta}=\iint\left\{-\frac{\partial\hat{\zeta}}{\partial t}-\hat{D}f_0\right\}dtdx=0$$

$$\frac{\partial L}{\partial D}=\iint\left\{\hat{\zeta}f_0-\frac{\partial\hat{D}}{\partial t}+\hat{\phi}\tilde{\phi}\right\}dtdx=0$$

$$\frac{\partial L}{\partial\phi}=\iint\left\{\nabla^2\hat{D}-\frac{\partial\hat{\phi}}{\partial t}\right\}dtdx=0$$

$$\therefore\begin{cases} -\dfrac{\partial\hat{\zeta}}{\partial t}-\hat{D}f_0=0\\ -\dfrac{\partial\hat{D}}{\partial t}+\hat{\zeta}f_0+\hat{\phi}\tilde{\phi}=0\\ -\dfrac{\partial\hat{\phi}}{\partial t}+\nabla^2\hat{D}=0 \end{cases}$$

$$\lambda=\begin{pmatrix}\hat{\zeta}\\ \hat{D}\\ \hat{\phi}\end{pmatrix}$$

$$L=\int_t\int_x\Big[\hat{\zeta}\left(\frac{\partial\zeta}{\partial t}+f_0D\right)+\hat{D}\left(\frac{\partial D}{\partial t}-f_0\zeta+\nabla^2\phi\right)+\hat{\phi}\left(\frac{\partial\phi}{\partial t}+\tilde{\phi}D\right)\Big]dxdt$$

$$=\int_t\int_x\Big[-\zeta\frac{\partial\hat{\zeta}}{\partial t}-D\frac{\partial\hat{D}}{\partial t}-\phi\frac{\partial\hat{\phi}}{\partial t}+\hat{\zeta}f_0D-\hat{D}f_0\zeta+\phi\nabla^2\hat{D}+\hat{\phi}\tilde{\phi}D\Big]dxdt$$

$$+\int_t\int_x\Big[\frac{\partial}{\partial t}(\hat{\zeta}\zeta)+\frac{\partial}{\partial t}(\hat{D}D)+\frac{\partial}{\partial t}(\hat{\phi}\phi)+\nabla^2(\phi\hat{D}$$

$$+\underbrace{\int_R\int_t\left(\hat{D}\nabla^2\phi-\phi\nabla^2\hat{D}\right)dRdt}_{\substack{\shortparallel\\ 0}}\ \text{by Green 2nd theorem}$$

$$\frac{\partial L}{\partial\zeta}=-\frac{\partial\hat{\zeta}}{\partial t}-f_0\hat{D}=0$$

$$\frac{\partial L}{\partial D}=-\frac{\partial\hat{D}}{\partial t}+f_0\hat{\zeta}+\hat{\phi}\tilde{\phi}D=0$$

$$\frac{\partial L}{\partial\phi}=-\frac{\partial\hat{\phi}}{\partial t}+\nabla^2\hat{D}=0$$

MET6308: Zou

- Write the adjoint of the following code which is part of Fortran codes of a forward linear model:

```
c input variables: X(N), QDOT(N)
c input constants: a(N), b(N)
        DO 10 I=1,N
            IF (I.NE.1) THEN
                W(I)=b(I)*QDOT(I)
            ENDIF
            IF (I.EQ.2) THEN
                W(I-1)=a(I)*QDOT(I)
            ENDIF
            Y(I)=W(I)
            W(I)=X(I)+QDOT(I)
10      CONTINUE
c output variables: Y(N), W(N)
```

adjoint code →

```
c input variable  Y(N), W(N)
c input constant  a(N), b(N)
    DO 10 I= N, 1, -1
      X(I) = W(I)
      QDOT(I) = W(I)          W(I) = 0
      W(I) = Y(I) (+ W(I))
      IF(I.EQ.2) THEN
        QDOT(I) = QDOT(I) + a(I) * W(I-1)
      ENDIF
      IF (I. NE. 1) THEN
        QDOT(I) = QDOT(I) + b(I) * W(I)
      ENDIF
  10 Continue
c output variable  X(N), QDOT(N)
```

If the second ("IF (I.NE.1) THEN") and the fourth ("ENDIF") lines are removed from the forward model, which has no impact on the output values of *Y*(*N*) and *W*(*N*), the forward code becomes

```
c input variables: X(N), QDOT(N)
c input constants: a(N), b(N)
        DO 10 I=1,N
            W(I)=b(I)*QDOT(I)
            IF (I.EQ.2) THEN
                W(I-1)=a(I)*QDOT(I)
            ENDIF
            Y(I)=W(I)
            W(I)=X(I)+QDOT(I)
10      CONTINUE
c output variables: Y(N), W(N)
```

✓

```
  DO 10  I = N, 1, -1
    X(I) = W(I)
    QDOT(I) = W(I)
    W(I) = Y(I) + W(I)
                                  IF(I. NE.1) THEN
                                    QDOT(I) = b(I) * W(I) + QD
                                  ENDIF
                                    └→ must be included !
    IF(I. EQ.2) Then
      QDOT(I) = a(I) * W(I-1) + QDOT(I)
    ENDIF
10  CONTINUE
```

What modification shall we make to the adjoint code you've written? Can we remove those two lines (the line "IF (I.NE.1) THEN" and the line "ENDIF") from the adjoint model in loop 10? We can not remove !

MET6308 Zou (1999)

- Answer the following questions:

(a) List 5-6 main classes of algorithms used in data analysis and assimilation.
(b) What is the input and output of these data analysis algorithms?
(c) What is the role of adjoint model in data assimilation? Why is it important?
✓(d) What are the definitions of homogeneous and isotropic background error covariances?

Sol)

(a) ① Subjective analysis
② function fitting
③ successive correction
④ optimal (statistical) Interpolation
⑤ variational method (3DVAR, 4DVAR)

(b) Input : observational data
output : uniformly gridded analysis data

(c) The adjoint model can produce the gradient of any forecast aspects with respect to initial condition. This is the only way to get ∇J in the atmospheric data assimilation. So we can obtain the minimization of J using one of several methods. Therefore, adjoint model is a prerequisite to obtain ∇J

(d) homogeneous background error

✓ $B_{kl}(\vec{r_k}, \vec{r_l}) = B_{kl}(r, \overset{\text{angle.}}{\phi})$

isotropic background error

✓ $B_{kl}(\vec{r_b}, \vec{r_l}) = \underset{\text{error}}{\bar{E}_b^2}\, \overset{\text{correlation}}{\rho_B}(r)$

$B = \overline{\{(x^b(\vec{r_k}) - x^t(\vec{r_k}))(x^b(\vec{r_l}) - x^t(\vec{r_l}))\}} = E_B^2 \rho_B(r_{lk})$ $\qquad E_B^2 = \overline{(x^b - x^t)^2}$

MET6308: Zou (1999)

- For a linear equation

$$\frac{\partial \delta \mathbf{x}}{\partial t} = A(t, \mathbf{x})\delta \mathbf{x}$$

$$\delta \mathbf{x}\big|_{t=t_0} = \delta \mathbf{x}_0 \qquad (1)$$

or in another form

$$\delta \mathbf{x}(t) = \mathbf{R}(t, t_0)\delta \mathbf{x}_0 \qquad (2)$$

give answers to the following terms:

1. The initial condition:
2. The Jacobian operator:
3. The resolvent:
4. The adjoint equation of eq. (1):
5. The resolvent of the adjoint equation:

Sol)

1. $\delta \vec{X}\big|_{t=t_0} = \delta \vec{X}_0$
2. $A(t, \vec{X})$
3. $R(t, t_0)$
4. $\begin{cases} -\dfrac{\partial \hat{X}}{\partial t} = (A(t, \vec{X}))^T \hat{X} \\ \hat{X}_{t_R} = \text{forcing} \end{cases}$
5. $\delta \hat{X}(t_0) = S(t_0, t)\,\delta \hat{X}(t)$

MET6308: Zou (1999)

- Suppose J is a cost function defined on the space R to which the control variable $\mathbf{x}_0$ belongs, and < , > represents an inner product defined on R. The gradient ∇J of J with respect to the control variable $\mathbf{x}_0$ is defined as a vector in the same space R as $\mathbf{x}_0$, and is such that the first-order variation δJ resulting from a perturbation $\delta\mathbf{x}_0$ of $\mathbf{x}_0$ is equal to the inner product

$$\delta J \equiv \langle \nabla J, \delta\mathbf{x}_0 \rangle \qquad (4)$$

Circle the right statements in the following:

① ∇J reflects ~~the~~ How sensitive is J with respect to $\mathbf{x}_0$. → True

② The value of ∇J depends on the value of $\mathbf{x}_0$. → True

3. The value of ∇J depends on the perturbation $\delta\mathbf{x}_0$. → False

④ The value of ∇J depends on how the inner product is defined on R. → True

5. Define a cost function based on the Euclidean L_2 norm:

$$J(\mathbf{x}_0) = \left[\mathbf{Hx}(T) - \mathbf{y}^{obs}(T), \mathbf{Hx}(T) - \mathbf{y}^{obs}(T)\right] = \sum_{i=1}^{N} \left(\mathbf{Hx} - y^{obs}\right)_i^2 \qquad (5)$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$ is an N-dimensional vector and represents the model forecast at the time T starting from $\mathbf{x}_0$ at time $t = 0$.

(i) Is it true that the gradient of J under the L_2 norm, ∇J, can be obtained by integrating the adjoint model $\mathbf{R}^T$ background in time once from T to t_0? True

(ii) What will be the "initial" condition at $t = T$ for the adjoint model integration to obtain ∇J? $\hat{\bar{X}}|_{t=T} =$ forcing

✓ MET6308: Zou (1999)

- Following is a simple nonlinear model:

$$\frac{\partial\phi}{\partial t}+\frac{\partial}{\partial x}(\phi u)=K\frac{\partial^2\phi}{\partial x^2}$$

$$\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(\phi+\frac{1}{2}u^2\right)=K\frac{\partial^2 u}{\partial x^2} \qquad (6)$$

where K is a constant, and the "geopotential" ϕ and the "zonal" velocity u are functions of time t and a one-dimensional periodic spatial coordinate x.

The model state vector can be expressed as:

$$\mathbf{x}=\begin{pmatrix}\phi(x,t)\\ u(x,t)\end{pmatrix}$$

and the inner product of the two state vectors

$$\mathbf{x}_1=\begin{pmatrix}\phi_1(x,t)\\ u_1(x,t)\end{pmatrix} \qquad \mathbf{x}_2=\begin{pmatrix}\phi_2(x,t)\\ u_2(x,t)\end{pmatrix}$$

is defined as

$$\checkmark \quad \langle\mathbf{x}_1,\mathbf{x}_2\rangle=\int_{x_a}^{x_b}(\phi_1\phi_2+\Phi_0 u_1 u_2)dx$$

where Φ_0 is an appropriate constant geopotential.

Derive the adjoint model of (6) in analytic form.

Hint: Use the result:

$$\int_{x_a}^{x_b}\left(\psi\frac{\partial^2\phi}{\partial x^2}-\phi\frac{\partial^2\psi}{\partial x^2}\right)dx=\left(\psi\frac{\partial\phi}{\partial x}-\phi\frac{\partial\psi}{\partial x}\right)\Bigg|_{x_a}^{x_b} = 0 \qquad (7)$$

Sol)

where ψ and ϕ are two scalar functions defined on $[x_a,x_b]$.

Inner product

$$\langle Y_1,Y_2\rangle=\int(\phi_1\phi_2+\Phi_0 u_1u_2)dx$$

$$Y_1=\begin{pmatrix}\phi_1\\ u_1\end{pmatrix},\quad Y_2=\begin{pmatrix}\phi_2\\ u_2\end{pmatrix}$$

adjoint varialle : $\hat\phi,\hat u$

$$L(\phi,u,\hat\phi,\hat u)=\iint\left\{\hat\phi\left[\frac{\partial\phi}{\partial t}+\frac{\partial(\phi u)}{\partial x}-K\frac{\partial^2\phi}{\partial x^2}\right]+\Phi_0\hat u\left[\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(\phi+\frac{1}{2}u^2\right)-K\frac{\partial^2 u}{\partial x^2}\right]\right\}dxdt$$

$$=\iint\left\{\left(\cancel{\frac{\partial\hat\phi\phi}{\partial t}}-\phi\frac{\partial\hat\phi}{\partial t}\right)+\left(\cancel{\frac{\partial\phi u\hat\phi}{\partial x}}-\phi u\frac{\partial\hat\phi}{\partial x}\right)+\Phi_0\left(\cancel{\frac{\partial u\hat u}{\partial t}}-u\frac{\partial\hat u}{\partial t}\right)+\Phi_0\left(\cancel{\frac{\partial\hat u(\phi+\frac{1}{2}u^2)}{\partial x}}-\left(\phi+\frac{1}{2}u^2\right)\frac{\partial\hat u}{\partial x}\right)\right\}dxdt$$

$$-\iint\left\{\hat\phi K\frac{\partial^2\phi}{\partial x^2}+\Phi_0\hat u K\frac{\partial^2 u}{\partial x^2}\right\}dxdt \longleftarrow \text{Green's theorem} \longrightarrow$$

$$\iint\hat\phi K\frac{\partial^2\phi}{\partial x^2}dxdt$$

$$=\iint\cancel{\frac{\partial^2}{\partial x^2}(\hat\phi K\phi)}-\phi K\frac{\partial^2\hat\phi}{\partial x^2}dxdt$$

$$=\iint\phi K\frac{\partial^2\hat\phi}{\partial x^2}+\cancel{K\left(\hat\phi\frac{\partial^2\phi}{\partial x^2}-\phi\frac{\partial^2\hat\phi}{\partial x^2}\right)}dt$$

Applying b.c.

$$L=\iint\left\{-\phi\frac{\partial\hat\phi}{\partial t}-\phi u\frac{\partial\hat\phi}{\partial x}-\Phi_0 u\frac{\partial\hat u}{\partial t}-\Phi_0\left(\phi+\frac{1}{2}u^2\right)\frac{\partial\hat u}{\partial x}-K\phi\frac{\partial^2\hat\phi}{\partial x^2}-\Phi_0 u K\frac{\partial^2 u}{\partial x^2}\right\}dxdt$$

$$\frac{\partial L}{\partial\phi}=\iint\left(-\frac{\partial\hat\phi}{\partial t}-u\frac{\partial\hat\phi}{\partial x}-\Phi_0\frac{\partial\hat u}{\partial x}-K\frac{\partial^2\hat\phi}{\partial x^2}\right)dtdx=0$$

$$\frac{\partial L}{\partial u}=\iint\left(-\phi\frac{\partial\hat\phi}{\partial x}-\Phi_0\frac{\partial\hat u}{\partial t}-\Phi_0 u\frac{\partial\hat u}{\partial x}-\Phi_0 K\frac{\partial^2 u}{\partial x^2}\right)dtdx=0$$

$$\therefore\begin{cases}-\dfrac{\partial\hat\phi}{\partial t}-u\dfrac{\partial\hat\phi}{\partial x}-\Phi_0\dfrac{\partial\hat u}{\partial x}=K\dfrac{\partial^2\hat\phi}{\partial x^2}\\[2ex] -\Phi_0\dfrac{\partial\hat u}{\partial t}-\phi\dfrac{\partial\hat\phi}{\partial x}-\Phi_0 u\dfrac{\partial\hat u}{\partial x}=\Phi_0 K\dfrac{\partial^2 u}{\partial x^2}\end{cases}$$

MET6308: Zou (1999)

- Write the adjoint of the following code which is part of Fortran codes of a linearized shallow-water model:

```
          SUBROUTINE LUVDECPL (M,PHA,U,V,UA,VA)
c input variables: PHA, UA, VA
c input basic state variables: PHA9, UA9, VA9
          REAL PHA(M), U(M), V(M), UA(M), VA(M)
          DIMENSION PHA9(M), UA9(M), VA9(M)
          DO 10 I=2,M-1
               U(I) = UA(I)/PHA9(I)-PHA(I)*UA9(I)/PHA9(I)**2
               V(I) = VA(I)/PHA9(I)-PHA(I)*VA9(I)/PHA9(I)**2
10        CONTINUE
c output variables: U,V
          RETURN
          END
```

If the arrays U and V are also input variables, i.e., they had values before calling this subroutine in its top subroutine, what modifications shall be made to the adjoint subroutine you've written?

Sol)

```
          SUBROUTINE AUVDECPL (M, PHA, U, V, UA, VA, PHA9, UA9, VA9)
 c input variables : U, V
 c input basic state variables : PHA9, UA9, VA9
          REAL PHA(M), U(M), V(M), UA(M), VA(M)
          DIMENSION PHA9(M), UA9(M), VA9(M)
          DO 10 I = 2, M-1
            VA(I) = VA(I) + V(I)/PHA9(I)
            PHA(I) = PHA(I) - V(I)*VA9(I)/PHA9(I)**2
            V(I) = 0
            UA(I) = UA(I) + U(I)/PHA9(I)
            PHA(I) = PHA(I) - U(I)*UA9(I)/PHA9(I)**2
            U(I) = 0
   10     Continue
   c output variables : PHA, UA, VA
            RETURN
            END
```

MET6308: Zou (1999, Final)

- Answer the following questions:

1. What is the background in data assimilation? Why is it needed?

background — forecast
Climatology.

data sparse region.

obs analys

Observations are often insufficient to completely determine all the components, i.e. $M<N$
In this case. the estimation problem cannot be solved in the absence of other sources of information. In numerical prediction, additional information is usually available in the form of a forecast that results from the previous analysis cycle.
This means that an estimate x_B, which does not take into account the new observations, is also available.

✓ 2. What is the incremental 4D-Var? Why is it needed?

* An approximation of the 4D-VAR, namely the incremental approach, similar to the linearization underlying the extended Kalman filter eqs, seems to achieve the amount of the computational reduction in a way consistent with the non-linear estimation theory.
The incremental approach considers a perturbation or an "increment" instead of the full model state.
variable → $\delta x(t_0) = x(t_0) - x_B$

✓ 3. Is it correct to say that the Kalman filter is equivalent to OI done at every time step? If the answer is no, then why?

Yes, each time step, we need to evaluate analysis error covariances

4. Is it correct say that for parameter estimation, one solves a problem similar to 4D-Var and the parameters to be estimated can be treated equally as the initial condition in 4D-Var?

Yes

5. What will you do to suppress undesirable small scale features in 4D-Var?

✓ Penalty term

✓ 6. Which one, the analysis or analysis increment, satisfies the imposed physical constraint in OI? both

Q.G. penalty

✓ 7. What and how is the physical constraint incorporated into the analysis procedure of 4D-Var?

From eq.

8. List the advantages and disadvantages of the FDA (finite-difference of adjoint) and AFD (adjoint of finite difference).

FDA → adv : smoothing gradient
dis : smoothing is approx.
hard to derive adjoint eq.

AFD → adv : at initial point → ∇ is exact
dis : oscillating ∇

✓ 9. What is the final condition (at time T) for the backward adjoint integration of the 1D shallow-water equation model to obtain the gradient

$$\nabla_{\mathbf{x}(0)} J \qquad (1)$$

where

$$J = \frac{1}{2}\left(u^2(T, II) + v^2(T, II)\right), \qquad T = 1h,\ II = 15 \qquad (2)$$

and $\mathbf{x}(t) = \left(u(t,1), \ldots, u(t,61), v(t,1), \ldots, v(t,61), \phi(t,1), \ldots, \phi(t,61)\right)^T$ represents the model state at time t.

$\frac{\partial J}{\partial X(t)}$

✓ 10. List three components that need to be developed to incorporate a new type of indirect observations into an existing 3D-Var and/or 4D-Var system.

① develop adj ✓
② define cost function ✓
③ connect the new obs. with the model ✓

For a specific type of observation, users of adjoint model may need to develop their own adjoint of the observation operator (H^T) for assimilating that observation.

→ Prelim question ?

MET6308: Zou (1998, 30 minutes)

- What is atmospheric data analysis and/or assimilation? List as much attributes as possible that need to be considered for atmospheric data analysis and/or assimilation.

- Following is a linearized shallow-water model written in term of vorticity ζ, divergence D and geopotential height ϕ:

$$\frac{\partial \zeta}{\partial t} + f_0 D = 0$$
$$\frac{\partial D}{\partial t} - f_0 \zeta + \nabla^2 \phi = 0$$
$$\frac{\partial \phi}{\partial t} + \tilde{\phi} D = 0 \qquad (1)$$

where f_0 is a constant Coriolis parameter, $\tilde{\phi} = g\tilde{h}$, $\phi = gh$ is the deviation from $\tilde{\phi}$, h is the height of the free surface of the fluid, g the gravitational constant, $\tilde{h}$ a free surface height independent of space and time representing a basic state at rest, $\nabla^2 = \partial^2 / \partial x^2 + \partial^2 / \partial y^2$ is the horizontal Laplacian operator in Cartesian coordinates.

Derive the adjoint model one (ADJ-1) of (1) with respect to the L_2 norm:

$$[\mathbf{x}, \mathbf{y}] = \sum_i^N x_i y_i \qquad (2)$$

And the adjoint model two (ADJ-2) of (1) with respect to an arbitrary norm:

$$\langle \mathbf{x}, \mathbf{y} \rangle = [\mathbf{x}, \mathbf{W}\mathbf{y}] \qquad (3)$$

where $\mathbf{W}$ is a weighting matrix, $\mathbf{x}$ and $\mathbf{y}$ represent vectors consisting of ζ, D and ϕ:

Compare the model ADJ-1 with ADJ-2, how can you still use ADJ-1 to calculate the gradient ∇J of J (a forecast aspect of eq. (1)) with respect to the norm defined in (3)? → $\nabla J = W^T \nabla^{L} J$ The vector of ∇J is defined as a vector such that the first-order variation δJ resulting from a perturbation $\delta \mathbf{x}_0$ of the initial condition $\mathbf{x}_0$ of model (1) is equal to the inner product

$$\delta J \equiv \langle \nabla J, \delta \mathbf{x}_0 \rangle \qquad (4)$$

Hint: Use the *Green's second theorem* to derive the adjoint model equations:

$$\int_{(R)} \left(\psi \nabla^2 \phi - \phi \nabla^2 \psi \right) dR = \int_{(S)} \left(\psi \nabla \phi - \phi \nabla \psi \right) \cdot \mathbf{n} dS \qquad (9)$$

where (R) is the two-dimensional space domain, (S) is the bounding surface of (R), $\mathbf{n}$ is the outward normal direction of S, ψ and ϕ are two scalar functions defined on (R), and ∇ is the gradient operator.

- The procedure which takes observed data and creates homogeneous fields is usually called analysis, or assimilation when the data are distributed in time and the procedures uses an explicit dynamical model for the time evolution of the atmospheric flow.

- Attributes to be sought for the estimates

memorize!

* ① observations are weighted according to their accuracy.

② Any error inherent in the original observations is not carried over to the grid-point values.

③ Dynamic and kinematic constraint are incorporated.

④ The field estimated are consistent with the other atmospheric and/or oceanic variables.

⑤ Observations other than the estimated field contributes to the analysis.

⑥ Include all available observations and past observations

⑦ Small scale features and obs. noises should be suppressed.

⑧ Continuity of fields across the boundaries of data dense regions to data sparse regions is maintained.

⑨ Quantities subject to unrepresentative fluctuations (like wind) are to require more input data than say pressure.

⑩ Algorithm are computationally efficient and the machine storage is affordable.

- adjoint variables : $\hat{\zeta}, \hat{\delta}, \hat{\Phi}$

→ **ADJ-1** $[x, y] = \sum_i^N x_i y_i$

$$L = \iint \left\{ \hat{\zeta}\left(\frac{\partial \zeta}{\partial t} + f_0 D\right) + \hat{\delta}\left(\frac{\partial D}{\partial t} - f_0\zeta + \nabla^2\Phi\right) + \hat{\Phi}\left(\frac{\partial \Phi}{\partial t} + \bar{\Phi} D\right)\right\} dt dx$$

$$L = \iint \left\{ -\zeta\frac{\partial \hat{\zeta}}{\partial t} + \hat{\zeta} f_0 D - D\frac{\partial \hat{\delta}}{\partial t} - \hat{\delta} f_0 \zeta + \Phi\nabla^2\hat{\delta} - \Phi\frac{\partial \hat{\Phi}}{\partial t} + \hat{\Phi}\bar{\Phi} D\right\} dt dx$$

$$\because \hat{\delta}\nabla^2\Phi = \nabla^2(\hat{\delta}\Phi) - \Phi\nabla^2\hat{\delta}$$

$$= \boxed{\hat{\delta}\nabla^2\Phi + \Phi\nabla^2\hat{\delta} - \Phi\nabla^2\hat{\delta}} \overset{\text{green's}}{\to 0}$$

$$= \Phi\nabla^2\hat{\delta}$$

$$\frac{\partial L}{\partial \zeta} = \iint \left\{ -\frac{\partial \hat{\zeta}}{\partial t} - \hat{\delta} f_0 \right\} dt dx = 0$$

$$\frac{\partial L}{\partial D} = \iint \left\{ \hat{\zeta} f_0 - \frac{\partial \hat{\delta}}{\partial t} + \hat{\Phi}\bar{\Phi} \right\} dt dx = 0$$

$$\frac{\partial L}{\partial \Phi} = \iint \left\{ \nabla^2\hat{\delta} - \frac{\partial \hat{\Phi}}{\partial t} \right\} dt dx = 0$$

$$\therefore \begin{cases} -\frac{\partial \hat{\zeta}}{\partial t} - \hat{\delta} f_0 = 0 \\ -\frac{\partial \hat{\delta}}{\partial t} + \hat{\zeta} f_0 + \hat{\Phi}\hat{\Phi} = 0 \\ -\frac{\partial \hat{\Phi}}{\partial t} + \nabla^2\hat{\delta} = 0 \end{cases} \leftarrow \text{ADJ-1}$$

→ **ADJ-2** $\langle x, y\rangle = [x, Wy] = \sum_i^N x_i W_i y_i = \sum_i W_i^T x_i y_i$

$$L = \iint \left\{ W^T\hat{\zeta}\left(\frac{\partial \zeta}{\partial t} + f_0 D\right) + W^T\hat{\delta}\left(\frac{\partial D}{\partial t} - f_0\zeta + \nabla^2\Phi\right) + W^T\hat{\Phi}\left(\frac{\partial \Phi}{\partial t} + \bar{\Phi} D\right)\right\} dt dx$$

?

$$[x, y] = \sum_i x_i y_i$$

$$\langle x, y\rangle = [x, Wy] = \sum_i x_i W_i y_i = \sum_i W_i^T x_i y_i$$

$$\delta J_1 = [x, y] \rightarrow \nabla J_1 = X$$

$$\delta J_2 = \langle x, y\rangle \Rightarrow \nabla J_2 = W^T X = W^T \nabla J_1$$

$$\therefore \nabla J = W^T \nabla^{L_2} J$$

or

based on [,]

$$\delta J = [\nabla^{L_2} J, \delta\vec{x}_0] = (\delta x_0)^T \nabla^{L_2} J$$

based on

$$\delta J = \langle \nabla J, \delta\vec{x}_0 \rangle = [\nabla J, W\delta\vec{x}_0] = (W\delta x_0)^T \nabla J$$

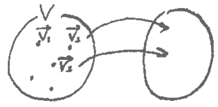
$$= (\delta x_0)^T W^T \nabla J = [W^T \nabla J, \delta\vec{x}_0]$$

$$\therefore \nabla J = W^T \nabla^{L_2} J$$

//

MET6308 Atmospheric data assimilation (Zou)

- Adjoint : transpose of a matrix
- linear transform ($\mathcal{T}$)



$\vec{V}_i = (x_1, x_2, \cdots, x_n)$

$\mathcal{T}(\alpha\vec{V}_1 + \beta\vec{V}_2) = \alpha\mathcal{T}(\vec{V}_1) + \beta\mathcal{T}(\vec{V}_2)$

$(\mathcal{T}(x), y) = (x, \overset{\text{adjoint}}{\mathcal{T}^*}(y))$ (,) : inner product.

$L_2 : (x, y) = \sum_i x_i y_i$

diff. eq. form

$$\frac{\partial X}{\partial t} = AX \Rightarrow -\frac{\partial \hat{X}}{\partial t} = A^T \hat{X}$$

- $\begin{cases} \dfrac{\partial \vec{X}}{\partial t} = F(\vec{X}) \\ \vec{X}|_{t=0} = \vec{X}_0 \end{cases}$ } non-linear

$\vec{X}_1, \vec{X}_2 \longrightarrow \vec{X}_2 - \vec{X}_1 = \vec{X}^{ptb}(t) = \vec{X}'$ perturbation

↓ tangent linear model

$\begin{cases} \dfrac{\partial \vec{X}'}{\partial t} = \dfrac{\partial F}{\partial x}\vec{X}' \\ \vec{X}'|_{t=0} = \vec{X}_0' \end{cases}$

※ $\dfrac{\partial \vec{X}_1}{\partial t} = F(\vec{X}_1)$

$\dfrac{\partial \vec{X}_2}{\partial t} = F(\vec{X}_2)$

$\dfrac{\partial(\vec{X}_2 - \vec{X}_1)}{\partial t} = F(\vec{X}_2) - F(\vec{X}_1)$

$\approx \dfrac{\partial F}{\partial x}(\vec{X}_2 - \vec{X}_1)$ Taylor expansion

↓

$$\boxed{-\frac{\partial \hat{\vec{X}}}{\partial t} = \left(\frac{\partial F}{\partial x}\right)^T \hat{\vec{X}}}$$

- $\vec{X}_{tr} = Q(\vec{X})\vec{X}_0$ } non-linear

$\vec{X}_{tr}' = P(\vec{X})\vec{X}_0'$ ← linear

$$\boxed{\hat{\vec{X}}_0^{(t_0)} = P^T(\vec{X})\hat{\vec{X}}_{tr}}$$

- Example 1

```
DO 10 I=1,3
  Y(I)=a(I)*X(I)
10 CONTINUE
```

output, input

what is the adjoint coding?

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \underline{P_{3\times3}} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \Longrightarrow \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = P^T \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

$P_{transpose}$ → $P = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix} = P^T$

So,

```
DO 20 I=1,3
  X(I) = a(I) *Y(I)
20 CONTINUE
```

✓ Example 2 The coding in ex1 can be different, if input is reused

* input = X

* output = Y, also X

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ X_1 \\ X_2 \\ X_3 \end{pmatrix} = P_{6\times3} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \Longrightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = P^T \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

↓

$$P = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \longrightarrow P^T = \begin{pmatrix} a_1 & 0 & 0 & 1 & 0 & 0 \\ 0 & a_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & a_3 & 0 & 0 & 1 \end{pmatrix}$$

Therefore,

```
DO 20 I=1,3
  X(I) = a(I)*Y(I) + X(I)
20 CONTINUE
```

✓ Example 3

$\begin{cases} A = B*C \\ D = A*C + A \\ B, C \text{ are variables} \end{cases}$

$\begin{cases} A' = B*C' + C*B' \\ D' = A'*C + C'*A + A' \end{cases}$ (output D'; input C', A')

→ $\begin{cases} \hat{C}' = \hat{A}' * B \\ \hat{B}' = \hat{A}' * C \end{cases}$

↓ adjoint

$\begin{cases} \hat{A}' = \hat{D}' * C \\ \hat{C}' = \hat{D}' * A \\ \hat{A}' = \hat{D}' + \hat{A}' \end{cases}$

※ $D' = A' * (1+C) + C' * A$

↓

$\hat{A}' = \hat{D}' * (1+C)$

$= \hat{D}' + \hat{D}'C$

- $P = P_1 P_2 P_3$

1) $P^T = P_3^T P_2^T P_1^T$

✱ 2) $(P\vec{X})^T \boxed{(P\vec{X})} = \vec{X}^T \boxed{P^T P\vec{X}} = \text{scalar}\left(= (\cdots\cdots)\begin{pmatrix} \cdot \\ \vdots \end{pmatrix}\right)$

output of TGL (tangent linear model)

output of adjoint.

TGL Input where $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \end{pmatrix}$

- NLM $\dfrac{\partial X}{\partial t} = N(X) \longrightarrow X_0 \longrightarrow X_{tr}$

↓ ∵ $P(X) = \dfrac{\partial N}{\partial X}$

TGL $\dfrac{\partial X'}{\partial t} = P(X)X' \longrightarrow X_0' \longrightarrow X_{tr}'$ perturbation

↓

ADJ $-\dfrac{\partial \hat{X}}{\partial t} = P^T(X)\hat{X}$

- Ex

$\underline{\dfrac{\partial u}{\partial t} = -cu}$: linear form (1) $\longrightarrow u = u_0 e^{-ct}$

$\left(\begin{array}{l} u|_{t=0} = u_0 \\ c : \text{constant} \end{array} \right.$

$\underline{\dfrac{\partial u'}{\partial t} = -cu'}$ (tangent linear) $\longrightarrow u' = u_0' e^{-ct} \rightarrow u_f' = u_0' e^{-ct_f}$

$u'|_{t=0} = u_0'$

$\underline{-\dfrac{\partial \hat{u}}{\partial t} = -c\hat{u}}$ (adjoint model) $\longrightarrow \hat{u}_0 = u_f' e^{ct}$ ← ⊜ → $u_0' = u_f' e^{ct_f}$

$\hat{u}|_{t=t_f} = u_f'$

(1) can be written in general form

$$\frac{\partial \vec{u}}{\partial t} = -A\vec{u} \quad \because A = \text{const.}$$

- Ex

NL

$\underline{\dfrac{du}{dt} = -cu^2}$

$u|_{t=0} = u_0$

$\because \int_{u_0}^{u} \frac{1}{u^2} du = -\int_0^t c\,dt \rightarrow -\left[\frac{1}{u}\right]_{u_0}^{u} = -ct$

$\longrightarrow \dfrac{1}{u} - \dfrac{1}{u_0} = ct \rightarrow u = \dfrac{u_0}{1 + ctu_0} = U$

TGL

$\underline{\dfrac{du'}{dt} = -2cUu'}$

$u'|_{t=0} = u_0'$

ADJ

$\underline{-\dfrac{d\hat{u}}{dt} = -2cU\hat{u}}$

$\hat{u}_f = u_f'$

} later! ⟶ p2 right side!

- $$\begin{tikzcd} X_0 \arrow[r, "2\,day"] & X^{(2day)} \\ & X_{ana} \leftarrow \text{analysis} \arrow[d] \\ X_0' \arrow[uu] & (X^{2day} - X_{ana}) \arrow[l, "\substack{\text{adjoint model}\\ \text{Inverse TGL}}"'] \end{tikzcd}$$

- Response function

$f(\vec{X}_{t_0}) \longrightarrow$ Scalar

$\left.\begin{matrix} X_0 \\ \alpha \end{matrix}\right\} \nabla_{\vec{X}_0} f(\vec{X})$, $f(X_0+\Delta X_0) - f(X_0) = \nabla_{X_0} f(X)\,\Delta X_0 + O((\Delta x)^2)$

$f = (X_{t_i} - X^{obs})^2$: forecast error

↑ model forecast

$X_{t_i} = Q(x) X_0$, $X_{t_i}^{(2)} = Q(x)(x_0 + \Delta X_0)$ ← not square

$X_{t_i}^{(2)} - X_{t_i} = P(x)\Delta X_0$

$f(X_0 + \Delta X_0) - f(X_0)$ → fcst error

fcst error

$$= (X_{t_i}^{(2)} - X^{obs})^2 - (X_{t_i} - X^{obs})^2$$
$$= (X_{t_i} + P(x)\Delta X_0 - X^{obs})^2 - (X_{t_i} - X^{obs})^2$$
$$= ((X_{t_i} - X^{obs}) + P(x)\Delta X_0)^2 - (X_{t_i} - X^{obs})^2$$
$$= (p(x)\Delta X_0)^2 + 2p(x)\Delta X_0 (X_{t_i} - X^{obs})$$
$$= (p(x)\Delta X_0)(p(x)\Delta X_0 + 2(X_{t_i} - X^{obs}))$$
$$\left.\begin{aligned} &= \langle p(x)\Delta X_0 ,\ 2(X_{t_i} - X^{obs}) \rangle \\ &= [(X - X^{obs}) \ ;\ P(x)\Delta X_0] \\ &= [P^T(x)(X - X^{obs}) \ ;\ \Delta X_t] \\ &= [\nabla f(x) \ ,\ \Delta X_i] \end{aligned}\right\}$$? see later!

$$\nabla_{X_0} f(x) = P^T[2(X - X^{obs})]$$

↓ general form

$$\boxed{\nabla_{X_0} f(x) = P^T \underbrace{\frac{\partial f(x)}{\partial X_t}}_{\text{forcing term.}}}$$

- $$\begin{tikzcd} \underset{t_0}{\vec{X}_0} \arrow[r, "\text{foreward}"] & \underset{t_R}{\vec{X}_{(t_R)}} \end{tikzcd}$$

$\frac{\partial \vec{X}'}{\partial t} = A\vec{X}'$ ※ $\begin{cases} \vec{X}_{(t_R)} = Q(\vec{X})\vec{X}_0 & : \text{NLM} \\ \vec{X}'_{(t_R)} = P(\vec{X})\vec{X}_0' & : \text{TGL} \\ \hat{X}_0 = P^T \hat{X}_{(t_R)} & \end{cases}$

- $f(\vec{X}_{(t_R)}) = J(\vec{X}_0)$ ← initial

↑ Scalar

$\nabla_{\vec{X}_0} J$

$$J(\vec{X}_0 + \vec{X}_0') - J(\vec{X}_0) = (\nabla_{\vec{X}_0} J)^T \vec{X}_0' + O(\|\vec{X}_0'\|^2)$$
$$= f(\vec{X}_{(t_R)} + \vec{X}'_{(t_R)}) - f(\vec{X}_{(t_R)})$$
$$= (\nabla_{\vec{X}_{(t_R)}} f)^T \vec{X}'_{(t_R)} \quad \text{TGL}$$
$$= (\nabla_{\vec{X}_{(t_R)}} f)^T P(x)\vec{X}_0'$$
$$= (P^T \nabla_{\vec{X}_{(t_R)}} f)^T \vec{X}_0'$$

※ $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{pmatrix}$

∴ $\boxed{\nabla_{\vec{X}_0} J = P^T \nabla_{\vec{X}_{t_R}} f}$

Adjoint model

$$\boxed{\begin{aligned} &\begin{cases} \hat{X}_0 = P^T \hat{X}_{(t_R)} \\ \hat{X}_{(t_R)} = \nabla_{\vec{X}_{(t_R)}} f \leftarrow \text{initial} \end{cases} \\ &\hat{X}_0 = \nabla_{\vec{X}_0} J \end{aligned}}$$

- ex)

$J(u_0) = f = u_{t_R}$ ∵ $\frac{\partial f}{\partial u_{t_R}} = 1$

$\begin{cases} -\frac{\partial \hat{u}}{\partial t} = -2cu\hat{u} \\ \hat{u}|_{t=t_R} = 1 \end{cases}$

$$u'_{t_R} = \nabla_{u_0} J \cdot u_0'$$
$$u'_{t_R} = (u_{t_R} + u'_{t_R}) - u_{t_R}$$
$$= J(u_0 + u_0') - J(u_0)$$
$$= \nabla_{u_0} J \cdot u_0'$$

∴ $u_0' = \frac{u'_{t_R}}{\nabla_{u_0} J}$ ⟵ 0 dimensional problem

- $\begin{cases} J_1(X_0) = (X_{t_R} - X_{t_R}^{obs})^2 \\ J_2(X_0) = \frac{1}{X_{t_R}} \\ J_3(X_0) = X_{t_R} \end{cases}$

$$\frac{\partial J_1}{\partial X_{t_R}} = 2(X_{t_R} - X_{t_R}^{obs})$$
$$\frac{\partial J_2}{\partial X_{t_R}} = -\frac{1}{X_{t_R}^2}$$

- NLM

(∵ $t_0 = 0$

$\begin{cases} \frac{du}{dt} = -cu^2 \\ u|_{t=t_0} = u_0 \end{cases}$ $\frac{du}{u^2} = -c\,dt \rightarrow \int_{t_0}^{t} d\frac{1}{u} = \int_{t_0}^{t} c\,dt \rightarrow \frac{1}{u(t)} - \frac{1}{u_0} = ct$

$$\therefore u(t) = \frac{u_0}{1 + cu_0 t}$$

TGL

$$\frac{du'}{dt} = -2cuu'$$
$$\frac{du'}{dt} = -2c\frac{u_0}{1+cu_0t}u' \rightarrow \frac{du'}{u'} = \frac{-2cu_0}{1+cu_0t}dt$$
$$\rightarrow \int_{t_0}^{t} d\ln u' = \int_{t_0}^{t} d\ln(1+cu_0t)^{-2}$$
$$\rightarrow \ln u'(t) - \ln u_0' = \ln\frac{1}{(1+cu_0t)^2} \rightarrow \frac{u'}{u_0'} = \frac{1}{(1+cu_0t)^2}$$
$$\rightarrow u'(t) = \frac{u_0'}{(1+cu_0t)^2} \longrightarrow u_0' = (1+cu_0t)^2 u'(t)$$

ADJ

$$-\frac{d\hat{u}}{dt} = -\frac{2cu_0}{1+cu_0t}\hat{u}$$
$$\frac{d\hat{u}}{\hat{u}} = \frac{2cu_0}{1+cu_0t}dt$$
$$\int_{t_R}^{t_0} d(\ln\hat{u}) = \int_{t_R}^{t_0} d\ln(1+cu_0t)^2 \quad ?$$
$$\ln\hat{u}_{t_0} \overset{1}{\cancel{-\ln\hat{u}|_{t=t_R}}} = -\ln(1+cu_0t_R)^2 = \ln(1+cu_0t_R)^{-2}$$
$$\therefore \hat{u}(t_0) = \frac{1}{(1+cu_0t_R)^2}$$

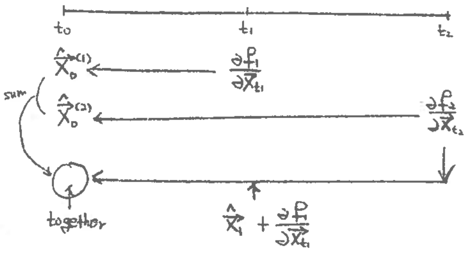
- $$\begin{cases} \dfrac{\partial \vec{X}}{\partial t} = F(\vec{X}) \ : \text{NLM} \\ \vec{X}|_{t=0} = \vec{X}_0 \end{cases}$$

$$\begin{cases} -\dfrac{\partial \hat{\vec{X}}}{\partial t} = \left(\dfrac{\partial F(\vec{X})}{\partial \vec{X}}\right)^T \hat{\vec{X}} \\ \hat{\vec{X}}_{tr} = \dfrac{\partial J}{\partial \vec{X}_{tr}} = \dfrac{\partial f(\vec{X}_{tr})}{\partial \vec{X}_{tr}} \end{cases}$$

$$\left\langle \begin{array}{l} J(\vec{X}_0) = f(\vec{X}_{tr}) \\ \nabla_{\vec{X}_0} J = \hat{\vec{X}}_0 \end{array} \right.$$

- $J(\vec{X}_0) = f(\vec{X}_{t_1}) + f(\vec{X}_{t_2}) = P_{t_2} - P_{t_1}$

$\nabla_{\vec{X}_0} J(\vec{X}_0) = \nabla_{\vec{X}_0} J_1 + \nabla_{\vec{X}_0} J_2 \qquad \because t_2 > t_1$



- <u>TGL Test</u> : Taylor expension

$$\vec{X}(t : \underbrace{\vec{X}_0 + \overset{\vec{X}_0'}{\alpha\Delta\vec{X}_0}}_{\text{I.C}}) - \vec{X}(t : \underbrace{\vec{X}_0}_{\text{I.C.}}) = \vec{X}'(t : \vec{X}_0)\,\alpha\Delta\vec{X}_0 + O((\alpha\Delta\vec{X}_0)^2)$$

$$= \alpha(\Delta\vec{X}_0)^T \vec{X}'(t) + O((\alpha\Delta\vec{X}_0)^2)$$

$$\varphi(\alpha) \equiv \frac{\vec{X}(t ; \vec{X}_0 + \alpha\Delta\vec{X}_0) - \vec{X}(t : \vec{X}_0)}{\alpha(\Delta\vec{X}_0)^T X'(t)} = 1 + O(\alpha)$$

The value of $\varphi(\alpha)$ approaches 1 linearly with respect to α

make a table

α	$\varphi(\alpha)$
1	1.3456
10^{-1}	1.0345
10^{-2}	1.0034
10^{-3}	1.0003
⋮	⋮

- <u>Gradient test</u> : same method as TGL test.

$$\vec{X} \longrightarrow J$$

$$J(\vec{X}_0 + \alpha\Delta\vec{X}_0) - J(\vec{X}_0) = (\alpha\Delta\vec{X}_0)^T \nabla_{\vec{X}_0} J + O((\alpha\Delta\vec{X}_0)^2)$$

$$\varphi(\alpha) = \frac{J(\vec{X}_0 + \alpha\Delta\vec{X}_0) - J(\vec{X}_0)}{\alpha(\Delta\vec{X}_0)^T \nabla_{\vec{X}_0} J} = 1 + O(\alpha)$$

$$\because \Delta\vec{X}_0 = \nabla_{\vec{X}_0} J$$

- <u>ADJ Test</u>

$$(P\vec{X}_0')^T \underbrace{P\overset{\text{input of TGL}}{\vec{X}_0'}}_{\text{TGL}} = \vec{X}_0'^T \underbrace{P^T P \vec{X}_0'}_{\text{ADJ}}$$

(output of TGL)^2 = (input of TGL)^T(output of adjoint with the TGL input as the input of ADJ)

↓

$(\vec{y}^2 = \vec{y}^T\vec{y})$

✓ - <u>EX</u>

$$\begin{cases} \dfrac{du}{dt} = -Cu^2 \\ u|_{t=0} = u_0 \end{cases}$$

0-dimensional model

$$\frac{U_{n+1} - U_n}{\Delta t} = -CU_n^2$$

```
                    ↓  ↓  ↓   ↑
Subroutine NLM (U0, C, DT, U, N)
DIMENSION U(N)
U(1) = U0
DO 10 J = 1, N-1
   U(J+1) = U(J) - C*DT*U(J)*U(J)
10 Continue
   return
   end
```

$u(N) = Q(u)u_0$

```
                    ↓  ↓  ↓   ↑
Subroutine TGL (U0, C, DT, U, N, U9)
DIMENSION U(N), U9(N)
U(1) = U0
DO 10 J=1, N-1
   U(J+1) = -2*C*DT*U9(J)*U(J)+U(J)
10 continue
   return
   end
```

→ nonlinear model Solution (basic state)

※ $\dfrac{du'}{dt} = -2cu'u$

$u'_{j+1} = -(2cu_j'u_j)\Delta t + u_j'$

$u(N) = Pu_0$

※ ADJ : there are two method

- one → using above discretized form
- two → using analytic eq

- From analytic eq. $-\dfrac{d\hat{u}}{dt} = -2c\hat{u}u \rightarrow \dfrac{\hat{u}_{n+1} - \hat{u}_n}{\Delta t} = 2cu_n\hat{u}_{n+1}$

```
Subroutine ADJ (U0, C, DT. U, N, U9)
DO 10 J=N-1, 1
   U(J) = (1-2*C*U9(J)*DT)*U(J+1)
10 continue
   return
   end
```

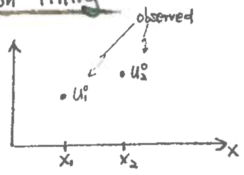
※ another → think about it!

✳ <u>classes</u> of data analysis/assimilation

① Subjective analysis
② function fitting
③ Successive corrections
④ Statistic interpolation
⑤ 3DVAR/4DVAR

(②–④) → these days, use <u>background field</u>

② <u>function fitting</u>



① <u>$f(x) = ax + b$</u> determine a and b!

<u>cost function</u> (I)

$$I = (f(x_1) - u_1^o)^2 + (f(x_2) - u_2^o)^2$$

min. I with repect to a and b

✓ $\dfrac{\partial I}{\partial a} = 0$, ✓ $\dfrac{\partial I}{\partial b} = 0$

$$I = (ax_1 + b - u_1^o)^2 + (ax_2 + b - u_2^o)^2$$

$$\begin{cases} \dfrac{\partial I}{\partial a} = 2(ax_1 + b - u_1^o)x_1 + 2(ax_2 + b - u_2^o)x_2 = 0 \\ \dfrac{\partial I}{\partial b} = 2(ax_1 + b - u_1^o) + 2(ax_2 + b - u_2^o) = 0 \end{cases}$$

$$\begin{cases}(x_1^2+x_2^2)a+(x_1+x_2)b = u_1^o x_1 + u_2^o x_2 \\ (x_1+x_2)a + 2b = u_1^o + u_2^o\end{cases}$$

$$G\begin{pmatrix}a\\b\end{pmatrix} = \begin{pmatrix}u_1^o x_1 + u_2^o x_2 \\ u_1^o + u_2^o\end{pmatrix}$$

Gram $\rightarrow G^{-1}$

derive ?

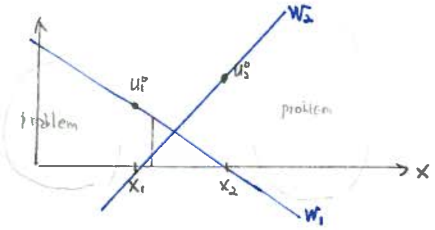
$$\begin{cases}a = \dfrac{u_1^o}{x_1-x_2} + \dfrac{u_2^o}{x_2-x_1} \ ✓\\ b = \dfrac{x_2 u_1^o}{x_2-x_1} + \dfrac{x_1 u_2^o}{x_1-x_2} \ ✓\end{cases}$$

Therefore

$$f(x) = \frac{x_2-x}{x_2-x_1}u_1^o + \frac{x_1-x}{x_1-x_2}u_2^o \equiv u^{ana}(x)$$

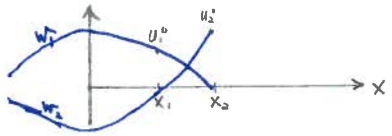
$$u^{ana}(x) = W_1 u_1^o + W_2 u_2^o$$

where W_1, W_2 : posteriori weighting functions



② $f(x) = ax^2+b$

$$\hookrightarrow u^{ana}(x) = \underbrace{\frac{x_2^2-x^2}{x_2^2-x_1^2}}_{W_1}u_1^o + \underbrace{\frac{x_1^2-x^2}{x_1^2-x_2^2}}_{W_2}u_2^o \rightarrow \text{proof!}$$



cost function

$$I = (ax_1^2+b-u_1^o)^2 + (ax_2^2+b-u_2^o)^2$$

$$\frac{\partial I}{\partial a} = 2(ax_1^2+b-u_1^o)x_1^2 + 2(ax_2^2+b-u_2^o)x_2^2 = 0$$

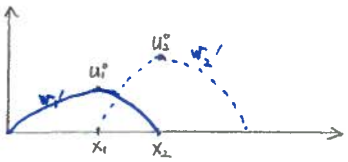
$$\frac{\partial I}{\partial b} = 2(ax_1^2+b-u_1^o) + 2(ax_2^2+b-u_2^o) = 0$$

$$\begin{cases}(x_1^4+x_2^4)a+(x_1^2+x_2^2)b = u_1^o x_1^2 + u_2^o x_2^2 \\ (x_1^2+x_2^2)a + 2b = u_1^o + u_2^o\end{cases}$$

→ same as before

③ $f(x) = ax^2+bx+c$

$$\hookrightarrow u^{ana}(x) = \underbrace{\frac{(x_2-x_1)^2-(x-x_1)^2}{(x_2-x_1)^2}}_{W_1'}u_1^o + \underbrace{\frac{(x_2-x_1)^2-(x-x_2)^2}{(x_2-x_1)^2}}_{W_2'}u_2^o$$



$$u^{ana}(x) = \begin{cases}u_1^o, & x=x_1 \\ u_2^o, & x=x_2\end{cases}$$ → This rule is applied above ①, ② + ③

- Using "background field" → { climatology / previous forecast }

$$\begin{cases}u_i'^{\,ana} \equiv (u_i^{ana} - \overset{\text{background}}{u_i^b}) & : \text{analysis increment} \\ u_k'^{\,obs} \equiv (u_k^o - u_k^b) & : \text{observation increment}\end{cases}$$

where i : analysis grid k : observation grid

- observations at one point → for successive correction

$$\overset{\text{variable}}{x_1} \longrightarrow \overset{\text{error}}{\varepsilon_1}$$

$$x_2 \longrightarrow \varepsilon_2$$

$$x^{ana} = f(x_1, x_2, \varepsilon_1, \varepsilon_2)$$

linear unbias estimate
minimum variance ㊂

Cost function

$$I = \frac{1}{\varepsilon_1^2}(f(x_1)-u_1^o)^2 + \frac{1}{\varepsilon_2^2}(f(x_2)-u_2^o)^2$$

→ depends on observation error

- review

$$\vec{X}(t_r ; \vec{X_0}+\alpha\Delta\vec{X_0}) - \vec{X}(t_r : \vec{X_0}) = \overbrace{\underbrace{P(x)\,\alpha\Delta\vec{X_0}}_{\text{tangent linear}}}^{\vec{X}'_{tr}} + O(\text{high})$$

$$\left(\nabla_{\vec{X_0}}\vec{X}\right)_{N\times M}\alpha\Delta\vec{X_0}$$

$$\vec{X_0} = \begin{pmatrix}x_{01}\\x_{02}\\ \vdots \\ x_{0N}\end{pmatrix} \qquad \begin{pmatrix}(\nabla_{\vec{X_0}} x_{01})^T \\ (\nabla_{\vec{X_0}} x_{02})^T \\ \vdots \\ (\nabla_{\vec{X_0}} x_{0N})^T\end{pmatrix}\alpha\Delta\vec{X_0}$$

$$\begin{cases}\dfrac{\partial\vec{X}}{\partial t} = F(\vec{X}) \\ \vec{X}|_{t=0} = \vec{X_0} = \begin{pmatrix}x_{01}\\ \vdots \\ x_{0N}\end{pmatrix}\end{cases} \qquad \vec{X}(t_r : \vec{X_0}) = Q(\vec{X})\vec{X_0}$$

$$\begin{cases}\dfrac{\partial\vec{X}'}{\partial t} = \dfrac{\partial F(\vec{X})}{\partial\vec{X}}\vec{X}' \\ \vec{X}'|_{t=0} = \vec{X}_0'\end{cases} \qquad \vec{X}'(t_r : \vec{X_0}, \vec{X}_0') = \mathbb{P}(\vec{X})\vec{X}_0'$$

◎ Successive Correction

- x_t : real value

obs.	error	expected error variance	errors are unbiased
x_1	ε_1	σ_1^2	$\overset{\text{mean}}{\overline{\varepsilon_1}} = 0$
x_2	ε_2	σ_2^2	$\overline{\varepsilon_2} = 0$

- find an estimate x_e

x_e is linear with respect to x_1, x_2
x_e estimate is unbiased
The error variance is minimum

$$\begin{cases}\varepsilon_1 = x_1 - x_t & \sigma_1^2 = \overline{(x_1-x_t)^2} \\ \varepsilon_2 = x_2 - x_t & \sigma_2^2 = \overline{(x_2-x_t)^2}\end{cases}$$

① $x_e = c_1x_1 + c_2x_2$

② $\overline{\varepsilon_e} = \overline{x_e - x_t} = 0 \Rightarrow c_1+c_2 = 1 \ (1-c_1-c_2=0)$

③ $\sigma_e^2 = \overline{(x_e-x_t)^2}$ minimum

- $J(c_1,c_2) = \sigma_e^2 + \lambda(1-c_1-c_2)$

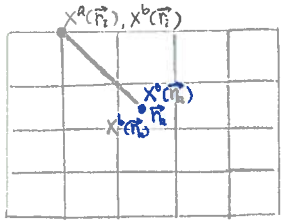
$$\frac{\partial J}{\partial c_1} = 0, \quad \frac{\partial J}{\partial c_2} = 0, \quad \frac{\partial J}{\partial \lambda} = 0$$

- $$x_e = \frac{\sigma_1^{-2}x_1 + \sigma_2^{-2}x_2}{\sigma_1^{-2}+\sigma_2^{-2}} \quad \text{———} \quad (1)$$

$$c_1 = \frac{\sigma_1^{-2}}{\sigma_1^{-2}+\sigma_2^{-2}} \qquad c_2 = \frac{\sigma_2^{-2}}{\sigma_1^{-2}+\sigma_2^{-2}}$$

↳ posteriori weight

- error
  1. background field X^b, E_b^2
  2. assign the posteriori weights
  3. obs. $X^o(\vec{r}_k)$, E_o^2



k : observation location
i : analysis grid point

$X^o(\vec{r}_k) - X^b(\vec{r}_k)$: observation increment

$$\left\langle \begin{array}{l} X^b(\vec{r}_i) \\ X^b(\vec{r}_i) + [X^o(\vec{r}_k) - X^b(\vec{r}_k)] \end{array} \right. \qquad \begin{array}{c} E_b^2 \\ E_c^2 \\ \Uparrow \end{array}$$

assume these are known

Similar to (1)

$$* \quad X^{ana}(\vec{r}_i) = \frac{E_b^{-2} X^b(\vec{r}_i) + E_c^{-2}\{X^b(\vec{r}_i) + [X^o(\vec{r}_k) - X^b(\vec{r}_k)]\}}{E_b^{-2} + E_c^{-2}} \quad (2)$$

$$E_c^2 = E_o^2 \underbrace{W(\vec{r}_k - \vec{r}_i)}_{\text{weight}}$$ → different in all lab.

ex)

$$W(\vec{r}_k - \vec{r}_i) = \begin{cases} 1 & \text{if } \vec{r}_k = \vec{r}_i \\ W(\hat{r}) & \text{if } \hat{r} = |\vec{r}_k - \vec{r}_i| \le R : \text{influence radius} \\ 0 & \text{if } \hat{r} > R \end{cases}$$

$$W(\hat{r}) = \frac{R^2 - \hat{r}^2}{R^2 + \hat{r}^2}$$: "Cressman-type" ✓

Eq.(2) can be written as follows

$$* \quad \boxed{X^{ana}(\vec{r}_i) = X^b(\vec{r}_i) + \frac{E_o^{-2} W(\vec{r}_k - \vec{r}_i)}{E_b^{-2} + E_o^{-2} W(\vec{r}_k - \vec{r}_i)} (X^o(\vec{r}_k) - X^b(\vec{r}_k))}$$

If there are K observations. How the above eq look like?

$$\rightarrow X^a(\vec{r}_i) = X^b(\vec{r}_i) + \underbrace{\frac{\sum_{k=1}^{K_i} E_o^{-2} W(\vec{r}_k - \vec{r}_i)}{E_b^{-2} + \sum_{k=1}^{K_i} E_o^{-2} W(\vec{r}_k - \vec{r}_i)}}_{\text{weighting}} \underbrace{(X^o(\vec{r}_k) - X^b(\vec{r}_k))}_{\text{obs. increment}}$$

✓ ● Optimal interpolation (Statistical interpolation)

$$* \quad X^a(\vec{r}_i) = X^b(\vec{r}_i) + W_{ik}[X^o(\vec{r}_k) - X^b(\vec{r}_k)]$$

→ How to define?

Determine W_{ik} by minimizing $\overline{(X^a - \underbrace{X^t}_{\text{real value}})^2} = f(W_{ik})$

✓ assumption

* no correlation between obs. and background

✓ $\overline{(X^b(\vec{r}_i) - X^t(\vec{r}_i))(X^o(\vec{r}_i) - X^t(\vec{r}_i))} = 0$

- $\dfrac{\partial f}{\partial W_{ik}} = 0$

diagonal matrix(E_o^2)

How can we get? → $\underbrace{[\underbrace{B}_{\text{background covariance matrix}} + \underbrace{O}_{\text{obs. covariance matrix}}]}\vec{W}_i = \overset{\text{known}}{\vec{B}_i}$ ← some function of obs.

$\vec{B}_i$ → $X^o(\vec{r}_k)$

?

- dense radiosonde network

$$R_{lk} = \frac{\overline{(X^o(\vec{r}_k) - X^b(\vec{r}_k))(X^o(\vec{r}_l) - X^b(\vec{r}_l))}}{\sqrt{\overline{(X^o(\vec{r}_k) - X^b(\vec{r}_k))^2}\ \overline{(X^o(\vec{r}_l) - X^b(\vec{r}_l))^2}}}$$

→ long sequence of obs. on $\vec{r}_k$, $\vec{r}_l$

→ assumption
  1. obs has no correlation
  2. obs and background have no correlation
  3. homogeneous, isotropic

- variational approach

$\vec{X}(t_1), \vec{X}(t_2), \cdots, \vec{X}(t_r)$

$\dfrac{\partial \vec{X}}{\partial t} = F(X) \sim f(X) = 0$

model solution.

$$\text{min. } J = \sum_{i=1}^{N} W_{(t_i)} (X_{(t_i)} - X^o_{(t_i)})^2$$

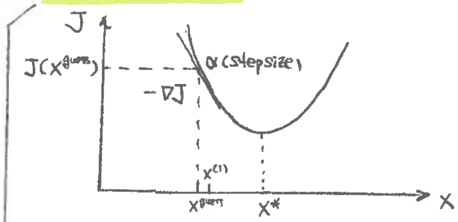
- min $J(\vec{X})$?

J, ∇J

all

$\vec{X}^*$, $J(\vec{X}^*) \le J(\vec{X})$ $\forall \vec{X}$

* ⟨descent direction⟩



$$X^{(1)} = X^{guess} + \alpha(-\nabla J)$$

↓ general form

iteration

$$X^{(k+1)} = X^{(k)} - \alpha_k \nabla_{X^{(k)}} J$$

$$d_k \equiv -\nabla_{X^{(k)}} J$$

$$X^{(k+1)} = X^{(k)} + \alpha_k d_k$$

how can we determine
  - i) d_k = ?
  - ii) α_k
    - → min $J(\alpha) = J(X^{(k)} + \alpha d_k)$
      ↓
      $\alpha^* = \alpha_k$

ex) $\begin{cases} \dfrac{du}{dt} = -cu^2 \\ u|_{t=0} = u_0 \end{cases}$

truth u_0 ⟶ "$u^{obs}(t_1)$, $u^{obs}(t_2)$"

$u_0 + u_0' = u_0^{guess}$

∇J ; $J = (u_{(t_1)} - u^{obs}_{(t_1)})^2 + (u_{(t_2)} - u^{obs}_{(t_2)})^2$

how to get α (stepsize)

⟨Newton method⟩

$J(X^{(k)})$: k = 0 guess
k = 1, ⋯ iterations

$$* \quad J(X) = J(X_k) + \nabla J(X_k)(X - X_k) + \frac{1}{2}(X - X_k)\nabla^2 J(X - X_k)$$

↙ $g(X)$

ignore

$$\nabla q(x) = \nabla J(x_k) + \nabla^2 J(x_k)(\overset{k+1}{x} - x_k) = 0$$

$$\nabla^2 J \cdot (x_{k+1} - x_k) = -\nabla J(x_k)$$

$$\underline{x_{k+1} = x_k - (\nabla^2 J(x_k))^{-1} \nabla J(x_k)}$$

$$\begin{cases} d_k = -(\nabla^2 J)^{-1} \nabla J \\ \alpha = 1 \end{cases}$$

Again?

- Problem

min. $J(x)$

Find x^*, $J(x^*) \le J(x)$ for any x

↳ minimum

→ We can approximate

$$J(x) = \underbrace{J(x^k) + \nabla J(x^k)(x - x^k)}_{\equiv q(x)} + \cancel{\tfrac{1}{2}\nabla^2 J(x^k)(x-x^k)^2 + \cdots}$$

→ ignore

$$\begin{cases} \text{given } x^k,\ x^0 \leftarrow \text{guess} \\ \text{find } x^{k+1},\ J(x^{k+1}) < J(x^k) \end{cases}$$

→ iteration number

$$\nabla q(x) = 0 = \nabla J(x^k) + \nabla^2 J(x^k)(x - x^k)$$

✓ $\underline{g_k \equiv \nabla J(x^k)}$

$$g_k + \nabla^2 J(x^k)(x - x^k) = 0$$

$$-(\nabla^2 J)^{-1} g_k = x - x^k$$

$$* \quad \underline{x^{k+1} = x^k - (\nabla^2 J)^{-1} g_k}$$

└ difficult (not easy)

→ How to get

$$(\nabla^2 J)^{-1} \equiv F^{-1}$$

$* \nabla^2 J$: Hessian $(=F)$ ✓

approximate

$$\underline{x^{k+1} = x^k - \boxed{\alpha_k H_k} g_k}$$

stepsize matrix

※ $\underline{(\nabla^2 J)^{-1} \equiv F^{-1} \equiv \alpha_k H_k}$

Can we find an approximation to the inverse of the Hessian F? → H_k and

a stepsize α_k which minimize $\phi(\alpha) = J(x^k - \alpha H_k g_k)$?

↳ α_k

$$\begin{cases} d_k = -H_k g_k \\ \langle \text{descent direction} \rangle \end{cases}$$

→ Construct H_k?

notation

vector
$$\begin{cases} g_k = \nabla J(x^k), \quad g_{k+1} = \nabla J(x^{k+1}) \\ p_k = x^{k+1} - x^k \\ q_k = g_{k+1} - g_k = \nabla J(x^{k+1}) - \nabla J(x^k) \end{cases}$$

matrix
$$\begin{cases} F = \nabla^2 J \\ H : \text{the approximation to the inverse Hessian}, F^{-1} \\ B : \quad \text{〃} \quad \text{to the Hessian } F \end{cases}$$

$$※ \quad p_k = \begin{pmatrix} p_{1k} \\ p_{2k} \\ \vdots \\ p_{nk} \end{pmatrix}$$

$p_k^T q_k \longrightarrow$ number

$p_k q_k^T \longrightarrow$ matrix

6

$$g_{k+1} - g_k = F(x^k)(x^{k+1} - x^k)$$ ← Taylor's expansion

$$\underline{q_k = F(x^k) p_k}$$

$F(x)$ is a constant matrix F

$$\boxed{q_i = F p_i \qquad 0 \le i \le k} \text{ ——— ⓐ}$$

if given $H_k \xrightarrow{\text{derive}} H_{k+1}$ ※ Let's put $H_0 = I$ ← unit mat

→ rank one correction

$$\underline{H_{k+1} = H_k + \underbrace{a_k z_k z_k^T}_{\text{unknown}}} \text{ ——— ⓑ}$$

constant

$q_k = F p_k$ ← from ⓐ

$H_k = F^{-1}$

$p_k = F^{-1} q_k$

$\underline{p_k = H_{k+1} q_k}$

substituting ⓑ into

$$p_k = (H_k + a_k z_k z_k^T) q_k$$

$$\underline{(p_k - H_k q_k) = a_k z_k z_k^T q_k} \text{ ——— ⓒ}$$

take inner product

$$(p_k - H_k q_k)(p_k - H_k q_k)^T$$

$$= (a_k z_k z_k^T q_k)\boxed{(a_k z_k z_k^T q_k)^T} \rightarrow (a_k q_k^T z_k z_k^T)$$

$$= a_k^2 z_k (z_k^T q_k)(q_k^T z_k) z_k^T = a_k^2 z_k (z_k^T q_k)(z_k^T q_k)^T z_k^T$$

$$= a_k^2 z_k z_k^T (z_k^T q_k)^2$$

$$\left(\begin{array}{l} q_k^T \times ⓒ \rightarrow \\ \underline{q_k^T p_k - q_k^T H_k q_k = a_k (z_k^T q_k)^2} \end{array} \right)$$

$$(p_k - H_k q_k)(p_k - H_k q_k)^T = a_k^2 z_k z_k^T (z_k^T q_k)^2$$

$$\therefore \quad a_k z_k z_k^T = \frac{(p_k - H_k q_k)(p_k - H_k q_k)^T}{q_k^T p_k - q_k^T H_k q_k}$$

↓ substitute ⓑ

$$\therefore \quad \boxed{H_{k+1} = H_k + \frac{(p_k - H_k q_k)(p_k - H_k q_k)^T}{q_k^T p_k - q_k^T H_k q_k}}$$

<quasi-Newton>

↑ input : H_k, x^k, x^{k+1}, g^k, g^{k+1}

→ in practice, more-complicated method is used (limited-memory)

$$※ \begin{cases} H_{k+1} \\ d_{k+1} = -H_{k+1} g_{k+1} \\ \alpha_{k+1} \text{ minimize } J(x^{k+1} + \alpha_{k+1} d_{k+1}) \\ \quad \downarrow \\ H_{k+2} \\ \vdots \end{cases}$$

The problem

Calculating the least value of a given function :

$$J(\vec{X}), \quad \vec{X} = (X_1, X_2, \cdots, X_N)^T$$

in the case that just

$$J(\vec{X}), \quad \frac{\partial J}{\partial X_i} = g_i \quad , i=1,2,\cdots,N$$

are available.

- A rank one update :

$$H_{k+1} = H_k + \frac{(P_k - H_k g_k)(P_k - H_k g_k)^T}{g_k^T P_k - g_k^T H_k g_k}$$

$$H_{k+1} = f(H_k, P_k, g_k)$$

$$P_k = X_{k+1} - X_k$$

$$g_k = g_{k+1} - g_k$$

$$X_{k+1} = X_k + \alpha_k \underbrace{(-H_k g_k)}_{d_k}$$

- A rank two update

$$H_{k+1} = H_k + \frac{P_k P_k^T}{\underbrace{(P_k^T g_k)}_{\to 0 \text{ (sometimes prollem)}}} - \frac{H_k g_k g_k^T H_k}{g_k^T H_k g_k}$$

DFP method (1970)

→ (David – Fletcher – Powell method)

1) If H_k is positive definite, then H_{k+1} is positive definite.
2) If J is quadratic function with constant Hessian F, then P_k^T are F-orthogonal
3) $P_0, P_1, \cdots, P_k$ are linearly independent eigenvectors of the matrix $H_{k+1}F$.

↓ new method

※ $g_k = F P_k$

$F^{-1} g_k = P_k$

$H_k \longrightarrow (\nabla^2 J)^{-1}$

$B_k \longrightarrow \nabla^2 J$

$$B_{k+1} = B_k + \frac{g_k g_k^T}{g_k^T P_k} - \frac{B_k P_k P_k^T B_k}{P_k^T B_k P_k}$$

∵ Sherman – Morrison formula

apply twice

$$\left((\underset{\text{matrix}}{A} + \underset{\text{vector}}{ab^T})^{-1} = A^{-1} - \frac{A^{-1} a b^T A^{-1}}{1 + b^T A^{-1} a} \right)$$

$$H_{k+1}^{BFGS} = H_k + \underbrace{\left(\frac{1 + g_k^T H_k g_k}{g_k^T P_k} \right) \frac{P_k P_k^T}{P_k^T g_k} - \frac{P_k g_k^T H_k + H_k g_k P_k^T}{g_k^T P_k}}_{f(P_k, g_k, H_k)} \quad ㊟$$

<quasi-Newton method>

※ BFGS = Broyden – Fletcher – Goldfard – Shanno

- "memoryless"

$H_0 = I$ ← unit matrix

$$H_1 = I + f(g_0, P_0)$$

$$H_2 = H_1 + f(g_1, P_1, H_1) = I + f_1(P_0, P_1, g_0, g_1)$$

$$\vdots$$

$$H_{m+1} = H_m + f(p_m, g_m, H_m) = I + f_m(P_0, g_0, \cdots P_m, g_m)$$

$$\underbrace{H_{(m+1)+1}}_{k>m} = \overset{k-m}{m} \text{ pairs of vectors } P_l, g_l$$

L-BFGS (Limited-meory quasi-Newton method) procedure.

1. Start at an initial guess X_0
   Compute the gradient $g_0 = \nabla J(X_0)$
2. Set $H_0 = I$ and $k=0$
3. set $d_k = -H_k g_k$
4. Do a line search to find the stepsize α_k
   $$J(X_k + \alpha_k d_k) = \min_\alpha J(X_k + \alpha d_k)$$
5. Update the variable
   $$X_{k+1} = X_k + \alpha_k d_k$$
6. Compute the new gradient value
   $$g_{k+1} = \nabla J(X_{k+1})$$
7. Update the H_k (using eq ㊟)
   $$H_{k+1} = \begin{cases} I + f(P_0, g_0, I) + \sum_{l=1}^{k} f(P_l, g_l, H_l), & k \le m \\ I + f(P_{k-m}, g_{k-m}, I) + \sum_{l=k-m+1}^{k} f(P_l, g_l, H_l), & k > m \end{cases}$$
   ※ m → depend on machine accuracy
   $5 < m \le 11$
   we can use $m=5$
8. Check for convergence : $\|g_{k+1}\| \le \varepsilon \max\{1, \|X_k\|\}$
   $\varepsilon = 10^{-5}$ → Stop, otherwise, return to step3.

- min $J(\vec{X}_k)$ ← find out ?

→ iterative (k) procedure

$$\vec{X}_{k+1} = \vec{X}_k + \alpha_k \vec{d}_k$$

look for α_k → Let's find this

$\vec{d}_k$ (search direction) → $H_{k+1} = H_k + f(P_k, g_k)$

$\xrightarrow{\vec{d}_k}$ minimize $f(\alpha) = J(\vec{X}_k + \alpha \vec{d}_k)$

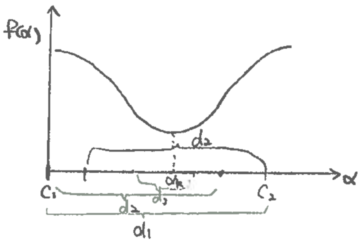
$\alpha > 0$

one-dimensional minimization problem.

- There are two method to get α_k

一. line search by reduce the region of uncertainty
二. line search by curve fitting

一.



① Select initial interval $[C_1, C_2]$

② find a sequence of the length of intervals d_k → there are several.

ex) Fibonacci search

$$F_N = F_{N-1} + F_{N-2} \quad , \quad F_0 = F_1 = 1$$

$$\begin{Bmatrix} F_0, & F_1, & F_2, & F_3, & \cdots \\ 1, & 1, & 2, & 3, & \cdots \end{Bmatrix}$$

define

$$d_k = \frac{F_{N-k+1}}{F_N} \quad , \quad d_1 = C_2 - C_1 \quad (\because d_2 < d_1)$$

三. <Newton method>

Suppose : $f'(\alpha)$, $f''(\alpha)$ known

$\alpha_k \longrightarrow \alpha_{k+1}$

$$g(\alpha) = f(\alpha_k) + f'(\alpha_k)(\alpha-\alpha_k) + \frac{1}{2}f''(\alpha_k)(\alpha-\alpha_k)^2 \sim f(\alpha)$$

$$g'(\alpha) = f'(\alpha_k) + f''(\alpha_k)(\alpha-\alpha_k)$$

$$g'(\alpha_{k+1}) = 0 \Rightarrow \underline{\alpha_{k+1} = \alpha_k - f'(\alpha_k)/f''(\alpha_k)}$$

<False position method>

Suppose we have : $f'(\alpha_k)$, $f'(\alpha_{k-1})$ at α_k, α_{k-1}

$$f''(\alpha_k) = \frac{f'(\alpha_{k-1}) - f'(\alpha_k)}{\alpha_{k-1} - \alpha_k}$$

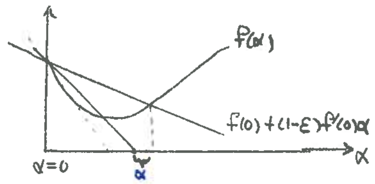
$$\rightarrow \alpha_{k+1} = \alpha_k - f'(\alpha_k)\left[\frac{\alpha_{k-1}-\alpha_k}{f'(\alpha_{k-1}) - f'(\alpha_k)}\right]$$

✓ • **convergence criteria**

1. Armijo's test :

$$\underline{f(\alpha) < f(0) + (1-\varepsilon)f'(0)\alpha}$$

$\because 0 < \varepsilon < 1$ $\Rightarrow$ ⓐ $J(X_k + \alpha_k d_k) < J(X_k) + (1-\varepsilon)\alpha_k g_k^T d_k$



$\because 0 < \varepsilon < 1$
$0 < 1-\varepsilon < 1$

$f(0) = J(\vec{X_k})$

2. $f(\alpha) > f(0) + (1-\varepsilon)f'(0)$ (— line)

3. Wolf test :

$$|f'(\alpha_{k+1})| \le \eta |f'(0)| \Rightarrow ⓑ \frac{|\nabla J(X_k + \alpha_k d_k)^T d_k|}{|g_k^T d_k|} \le \eta$$

$0 \le \eta < 1$

if $\eta = 0 \longrightarrow$ exact solution

※ Scaling → convergence speed ↑

• EX) Input : B
output : C

$A = 0.1 * B$
$C = 0.3 * A^2 + B$

→ TGL

input : B , B9 (basic state) , output : C

$A = 0.1 * B$
$A9 = 0.1 * B9$
$C = 0.6 * A9 * A + B$

→ ADJ

Input : C, B9 , output : B

$A9 = 0.1 * B9$
$A = 0.6 * A9 * C$
$B = C$
$B = \bigcirc\!\!\!\!B + 0.1 * A$

※ be careful !

① basic state in TGL and adjoint
② variable reused or not reused

✓ • EX)

$A = 0.1 * B$
$B = B$

input : B
output : A, B

$\downarrow$

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 0.1 \\ 1 \end{pmatrix} B$$

$$B^{adj} = (0.1 \quad 1)\begin{pmatrix} A^{adj} \\ B^{adj} \end{pmatrix}$$

$$B^{adj} = 0.1 A^{adj} + B^{adj}$$

• be careful !

$A = C$
$A = C_1 A + C_2 B$

constant

→ $\boxed{B = C_2 A \\ A = C_1 A}$ correct

→ ~~$A = C_1 A$~~
~~$B = C_2 A$~~ incorrect.

三• < Variational Analysis >

→ optimizing observed values under certain subsidiary conditions.

• $$\frac{\partial \varphi}{\partial t} + c\frac{\partial \varphi}{\partial x} = 0 \qquad (1)$$

suppose we have

① observation : $\varphi^{obs}(t,x)$
② background field : $\varphi(t,x)$
③ eq.(1)

$[t_0, t_R]$ ⟵ assimilation window
$[x_a, x_b]$

Define the variance of the difference between the obs. and analysis

$$\boxed{J = \int_t\int_x \alpha(\varphi - \varphi^{obs})^2 dxdt + J_c} \qquad (2)$$

constraint.

Find φ^* which minimizes J.

⇝ There are three diffent ways.

<1> diagnostic equetion

$J_1 = f(\varphi(t,x))$

if small

$$J_1 = \int_t\int_x \left\{\alpha(\varphi - \varphi^{obs})^2 + \alpha_p\left(\frac{\partial \varphi}{\partial t}\right)^2\right\} dxdt$$

<2> "week" constraint

$J_2 = f(\varphi(t,x))$

$$J_2 = \int_t\int_x \left\{\alpha(\varphi - \varphi^{obs})^2 + \alpha_p\left[\frac{\partial \varphi}{\partial t} + c\frac{\partial \varphi}{\partial x}\right]^2\right\} dxdt$$

∴ α_p called weighting or penalty coef.

<3> "Strong" constraint

$J_3 = f(\varphi(t,x), \lambda$

$$J_3 = \int_t\int_x \left\{\alpha(\varphi - \varphi^{obs})^2 + \lambda\left(\frac{\partial \varphi}{\partial t} + c\frac{\partial \varphi}{\partial x}\right)\right\} dxdt$$

∴ λ : Lagrangian multiplier

• diagnostic eq.

$\delta J \longleftarrow \delta\varphi$

eq.(1)

$$J = \int_t\int_x \left\{\alpha(\varphi - \varphi^{obs})^2 + \alpha_p \boxed{c^2\left(\frac{\partial \varphi}{\partial x}\right)^2}\right\} dtdx$$

$$\delta J = \int_t\int_x \left\{2\alpha(\varphi - \varphi^{obs})\delta\varphi + 2\alpha_p c^2 \frac{\partial \varphi}{\partial x}\frac{\partial \delta\varphi}{\partial x}\right\} dtdx$$

$\hookrightarrow (\quad\quad)\delta\varphi = 0$
↑ any $\delta\varphi$

$$\because \int_x\int_t 2\alpha_p c^2 \frac{\partial\varphi}{\partial x}\frac{\partial\delta\varphi}{\partial x}\,dx\,dt$$

part integration?

$$= \int_x\int_t 2\alpha_p c^2 \left[\frac{\partial(\varphi\delta\varphi)}{\partial x} - \frac{\partial^2\varphi}{\partial x^2}\delta\varphi\right] dx\,dt$$

$$= \int_t 2\alpha_p c^2 \varphi\delta\varphi\Big|_{x_a}^{x_b} - \int_x\int_t 2\alpha_p c^2 \frac{\partial^2\varphi}{\partial x^2}\delta\varphi\,dx\,dt$$

$$\int_t\int_x \left[2\alpha(\varphi-\varphi^{obs}) - 2\alpha_p c^2\frac{\partial^2\varphi}{\partial x^2}\right]\delta\varphi\,dx\,dt + \int_t 2\alpha_p c^2\varphi\delta\varphi\Big|_{x_a}^{x_b} dt = 0$$

$$\therefore \begin{cases} 2\alpha(\varphi-\varphi^{obs}) - 2\alpha_p c^2 \frac{\partial^2\varphi}{\partial x^2} = 0 \\ \varphi|_{x_b} = 0 \quad , \quad \varphi|_{x_a} = 0 \end{cases}$$

⇐ 2nd-order PDE

↑ can be solved by iteration method(k)

If we assume

$\varphi^{obs} = A\cos kx$ $\quad \because k$: wavenumber

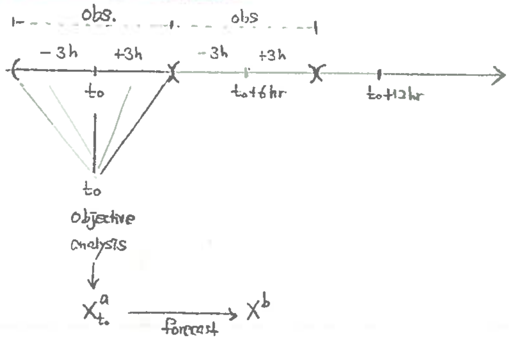
$$\varphi = \frac{\alpha A}{\alpha+\alpha_p c^2 k^2}\cos kx = \underbrace{\frac{\alpha}{\alpha+\alpha_p c^2 k^2}}_{r}\varphi^{obs}$$

$\because ck = \nu$

$r \leq 1$

if $r\downarrow \longrightarrow k\uparrow$

✓ • < Data assimilation cycle >



① quality control : reject and modify bad data

② objective analysis

③ Initialization

Control gravity-wave oscillation

ex) nonlinear normal mode initialization

(digital filter

$X^a \longrightarrow$ balance field

④ Short-term forecast

using an assimilation model (6h ~ 12h)

• $$\begin{cases}\frac{\partial\vec{X}}{\partial t} = F(\vec{X}) \\ \vec{X}|_{t=t_0} = \vec{X}_0\end{cases}$$

NLM

$$J(\vec{X}_0) = \int_t < \vec{X}(t)-\vec{X}^{obs}, \vec{X}(t)-\vec{X}^{obs} > dt$$

where < , > : inner product.

* gradient definition

If an inner product has been defined on the space of $\mathcal{R}$ to which the (new) control variable $\vec{X}_0$ belongs, the gradient ∇J of J is defined as a vector in the same space $\mathcal{R}$ as $\vec{X}_0$, and is such that the first-order variation δJ resulting from a perturbation $\delta\vec{X}_0$ of $\vec{X}_0$ is equal to the inner product: $\delta J = <\nabla J, \delta\vec{X}_0>$

∇J ↳ (independent of $\delta\vec{X}_0$
(dependent on $\vec{X}_0$

※ $\nabla J \longleftarrow \vec{X}_0$ or $\delta\vec{X}$?

<question> $<\vec{x},\vec{y}>_1 = \sum_{i=1}^{N} x_i y_i = \vec{y}^T\vec{x}$

$\because \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N\end{pmatrix} \in R_n$

$<\vec{x},\vec{y}>_2 = \vec{y}^T W\vec{x}$ $\quad \because W$ is a matrix

find the relation between the gradient with $< >_1$ and $< >_2$ definition.

$\rightarrow \delta J = ? = < ? , \delta\vec{X}_0 >$

✓ $$\delta J = \int_t < 2(\vec{X}-\vec{X}^{obs}), \delta\vec{X}(t) > dt$$

substitute

$$\text{TGL}\begin{cases}\frac{\partial\delta\vec{X}}{\partial t} = \frac{\partial F}{\partial x}\delta\vec{X} \\ \delta\vec{X}|_{t=t_0} = \delta\vec{X}_0\end{cases} \longrightarrow \quad \frac{\partial F}{\partial x} : \text{Jacobian operator}, \quad \delta\vec{X}(t) = R(t,t_0)\,\delta\vec{X}_0$$

$\because R(t,t_0)$: resolvant between time t_0 and t

$$\delta J = \int_t < 2(\vec{X}-\vec{X}^{obs}), R(t,t_0)\delta\vec{X}_0 > dt$$

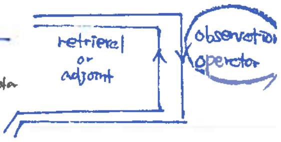
$$= \int_t < R^*(t,t_0)2(\vec{X}-\vec{X}^{obs}), \delta\vec{X}_0 > dt$$

$$\therefore \nabla J = \int_t R^*(t,t_0)2(\vec{X}-\vec{X}^{obs})dt$$

※ ($<\vec{x},\vec{y}> = <\vec{z},\vec{y}>$
($\forall\vec{y} \rightarrow \vec{x}=\vec{z}$

$$\begin{cases} -\frac{\partial\hat{X}}{\partial t} = \left(\frac{\partial F}{\partial x}\right)^T\hat{X} \\ \hat{X}|_{t_R} = \text{forcing}\end{cases} \longrightarrow Z(t_R,t_0) = R^*$$

• obs. ~~related~~ analysis of model state



obs. (direct obs. : wind, T, q, q_L, q_R (cloud, rainwater)
(indirect obs. (surface fluxes, rainfall, GPS-PW
(ozone, GPS occulation α (bending angle)

• (Adjoint of finite-difference (AFD) : from coding
(finite-difference of adjoint (FDA) : from analytic solution.

$$\frac{\partial X}{\partial t} = F(X) \quad (1)$$

$<x, y>$: inner product over space R $\quad x \in R$

$$<\lambda, \frac{\partial X}{\partial t} - F(X)> = 0$$

✓ $$L \equiv \int_0^T <\lambda, \frac{\partial X}{\partial t} - F(X)> dt$$

$\frac{\partial L}{\partial x} = 0 \longleftarrow$ adjoint eq.

$$L = \int_0^T <\lambda, \frac{\partial X}{\partial t} - F(X)> dt$$

$$= \int_0^T \left(\frac{\partial\lambda X}{\partial t} - X\frac{\partial\lambda}{\partial t}\right)dt - \int_0^T \lambda F(X)\,dt$$

$$= \lambda X\Big|_0^T - \int_0^T \left(X\frac{\partial\lambda}{\partial t} + \lambda F(X)\right)dt$$

$\because \lambda(T) = 0$ (define)

$$L = -\lambda X_0 - \int_0^T \left(x\frac{\partial\lambda}{\partial t} + \lambda F(x)\right)dt$$

$$\frac{\partial L}{\partial x} = -\int_0^T \left(\frac{\partial\lambda}{\partial t} + \frac{\partial F(x)}{\partial x}\lambda\right)dt = 0$$

$$\therefore \frac{\partial\lambda}{\partial t} + \frac{\partial F(x)}{\partial x}\lambda = 0$$

$$-\frac{\partial\lambda}{\partial t} = \frac{\partial F(x)}{\partial x}\lambda$$

Ex 1

$$\begin{cases} \dfrac{\partial\phi}{\partial t} + \dfrac{\partial(\phi u)}{\partial x} = 0 \\ \dfrac{\partial u}{\partial t} + \dfrac{\partial}{\partial x}(\phi + \frac{1}{2}u^2) = 0 \end{cases}$$

Inner product

$$\langle Y_1, Y_2 \rangle = \int (\phi_1\phi_2 + \phi_0 u_1 u_2)dx$$

$$Y_1 = \begin{pmatrix}\phi_1\\ u_1\end{pmatrix}, \quad Y_2 = \begin{pmatrix}\phi_2\\ u_2\end{pmatrix}$$

adjoint variable : $\hat\phi$, $\hat u$

$$L(\phi, u, \hat\phi, \hat u) = \iint \left\{ \hat\phi\left(\frac{\partial\phi}{\partial t} + \frac{\partial(\phi u)}{\partial x}\right) + \phi_0 \hat u\left(\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(\phi + \tfrac{1}{2}u^2)\right)\right\} dtdx$$

$$= \iint \left\{ \left(\cancel{\frac{\partial\hat\phi\phi}{\partial t}} - \phi\frac{\partial\hat\phi}{\partial t}\right) + \left(\cancel{\frac{\partial\phi u\hat\phi}{\partial x}} - \phi u\frac{\partial\hat\phi}{\partial x}\right) + \phi_0\left(\cancel{\frac{\partial u\hat u}{\partial t}} - u\frac{\partial\hat u}{\partial t}\right) + \phi_0\left(\cancel{\frac{\partial\hat u(\phi + \frac{1}{2}u^2)}{\partial x}} - (\phi + \tfrac{1}{2}u^2)\frac{\partial\hat u}{\partial x}\right)\right\} dtdx$$

∵ boundary

$$\hat\phi\phi\Big|_0^T = 0, \quad \phi u\hat\phi\Big|_{x_2}^{x_1} = 0, \quad \phi_0 u\hat u\Big|_0^T = 0, \quad \hat u(\phi + \tfrac{1}{2}u^2)\Big|_{x_1}^{x_2} = 0$$

$$\rightarrow \frac{\partial L}{\partial\phi} = 0$$

$$\rightarrow \frac{\partial L}{\partial u} = 0$$

$$\frac{\partial L}{\partial\phi} = -\iint_{x\,t} \left(\frac{\partial\hat\phi}{\partial t} + u\frac{\partial\hat\phi}{\partial x} + \phi_0\frac{\partial\hat u}{\partial x}\right)dtdx = 0$$

$$\frac{\partial L}{\partial u} = -\int_x\int_t \left(\phi_0\frac{\partial\hat u}{\partial t} + \phi\frac{\partial\hat\phi}{\partial x} + \phi_0 u\frac{\partial\hat u}{\partial x}\right)dtdx = 0$$

Ex2 $\dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} = 0$ ——→ adjoint eq. ?

$$\langle u_1, u_2 \rangle \equiv \int u_1 u_2 dx$$

adjoint variable : $\hat u$

$$L = \iint \hat u\left(\frac{\partial u}{\partial t} + \overset{\frac{\partial}{\partial x}\left(\frac{u^2}{2}\right)}{\boxed{u\frac{\partial u}{\partial x}}}\right)dtdx$$

$$= \iint \left\{ \cancel{\frac{\partial\hat u u}{\partial t}} - u\frac{\partial\hat u}{\partial t} + \cancel{\frac{\partial\hat u(\frac{u^2}{2})}{\partial x}} - \frac{u^2}{2}\frac{\partial\hat u}{\partial x}\right\} dtdx$$

$$\frac{\partial L}{\partial u} = 0$$

$$\frac{\partial L}{\partial u} = -\iint \left(\frac{\partial\hat u}{\partial t} + u\frac{\partial\hat u}{\partial x}\right)dtdx = 0$$

$$\therefore \frac{\partial\hat u}{\partial t} + u\frac{\partial\hat u}{\partial x} = 0$$

- Attributes to be sought for the estimates

① observations are weighted according to their accuracy.

② Any error inherent in the original observations is not carried over to the grid-point values.

③ Dynamic and kinematic constraint are incorporated.

④ The field estimated are consistent with the other atmospheric and/or oceanic variables.

⑤ observations other than the estimated field contributes to the analysis.

⑥ Include all available observations and past observations

⑦ Small scale features and obs. noises should be suppressed.

⑧ Continuity of the fields across the boundaries of data dense regions to data sparse regions is maintained.

⑨ Quantities subject to unrepresentative fluctuations (like wind) are to require more input data than say pressure.

⑩ Algorithm are computationally efficient and the machine storage is affordable.

Given (1) M height obs. : h_m^o, $m = 1, \cdots, M$

(2) N wind obs. : (u_n^o, v_n^o), $n = 1, \cdots, N$

Questions

i) If the height h on a regular mesh of points are to be estimated can wind obs. be used for the h estimation?

ii) Can we obtain a simultaneous analysis of h, u, v which are balanced?

→ Function fitting approach

$$J = \sum_{m=1}^{M}\alpha_m\left(h_{(x,y)} - h^o\right)_m^2 + \sum_{n=1}^{N}\beta_n\left[(u-u^o)_n^2 + (v-v^o)_n^2\right] \quad \text{—— (}$$

$$u_g = -\frac{g}{f}\frac{\partial h}{\partial y}, \quad v_g = \frac{g}{f}\frac{\partial h}{\partial x} \quad \text{———— (}$$

assuming,

$$h = \sum_{j=0}^{JJ}\sum_{i=0}^{II}\overset{\text{to be determined.}}{\boxed{a_{ij}}}\underbrace{h_{ij}(x,y)}_{\text{known basis function}} \quad \text{———— (}$$

(3) ⟶ (2)

$$\left.\begin{aligned} u_g(x,y) &= \sum_{ij}\left(-\frac{g}{f}a_{ij}\frac{\partial h_{ij}}{\partial y}\right) \\ v_g(x,y) &= \sum_{ij}\left(\frac{g}{f}a_{ij}\frac{\partial h_{ij}}{\partial x}\right) \end{aligned}\right) \quad \text{———— (4}$$

(4) ⟶ (1)

o o o

Solve $\dfrac{\partial J}{\partial a_{ij}} = 0$, $0 \le i \le II$, $0 \le j \le JJ$

advantage :

① wind obs. can contribute to height analysis

② number of obs. ≪ number of a_{ij} (degree of freedom)

M+2N $\quad (II+1)*(JJ+1)$

originally. $3*(II+1)*(JJ+1)$

so, reduce the # of degrees of freedom

→ Variational approach

⟶ thick about this!

◉ Penalty term

consider the equality-constrained problem :

$$\begin{cases} \text{minimize } J(\vec X) \\ \text{subject to } \vec g(\vec X) = 0 \end{cases}$$

$\vec X$ ⟶ control variable

⟶ dimension of $\vec X$ = degree of freedom

- Penalty method :

minimize $J(\vec X) + \boxed{\frac{1}{2}\rho_k\sum_{i=1}^{m}\vec g_i^{\,2}(\vec X)}$ ⟶ penalty term

for an increasing sequence of $\{\rho_k\}$ tending to infinity

$\rho_k > 0$

∵ ρ_k called penalty coefficient

→ How can this penalty be applied to ADA

I. Suppress small scale features and observational noises ✓

$\alpha_t \|\nabla_t \vec{X}\|^2$

$\alpha_x \|\nabla_x \vec{X}\|^2$

ex) $\min J = \int_0^T \int_x \alpha (\phi - \phi^e)^2 dx dt$

$\frac{\partial \phi}{\partial t} + C_x \frac{\partial \phi}{\partial x} = 0$

change problem

$\because \phi^e$: given estimation or obs.

strong constraint

$= J_2$

$\min J_3 = \min \left\{ J + \int_0^T \int_x \lambda \left(\frac{\partial \phi}{\partial t} + C \frac{\partial \phi}{\partial x} \right) dt dx + \alpha_t \left(\frac{\partial \phi}{\partial t} \right)^2 + \alpha_x \left(\frac{\partial \phi}{\partial x} \right)^2 \right\}$

given $\phi^e = \phi^o \cos k(x - C_x^o t)$

$$\begin{cases} \frac{\partial J_3}{\partial \phi} = 0 \\ \frac{\partial J_3}{\partial \lambda} = 0 \end{cases} \rightarrow \phi(t,x) = \boxed{\frac{\alpha}{\alpha + \alpha_t k^2 C_x^2 + \alpha_x k^2}} \frac{\phi^o}{k(C_x^o - C_x)}$$

$\{ \sin k(C_x^o - C_x)T \cos k(x - C_x^o t) + [1 - \cos k(C_x^o - C_x)T] \sin k(x - C_x^o t) \}$

$$\begin{cases} \frac{\partial J_2}{\partial \phi} = 0 \\ \frac{\partial J_2}{\partial \lambda} = 0 \end{cases} \rightarrow \phi(t,x) = \frac{\phi^o}{k(C_x^o - C_x)} \{ \quad " \quad \}$$

$r \equiv \frac{\alpha}{\alpha + \alpha_t k^2 C_x^2 + \alpha_x k^2}$ $\quad \because$ frequency $\nu = kC_x$

$= \frac{\alpha}{\alpha + \alpha_t \nu^2 + \alpha_x k^2}$ (damping effect)

high freq ⟶ more damping

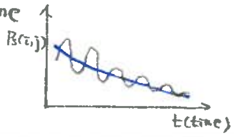
II. Control the gravity wave oscillations ✓

III. Biasing a fit toward smooth solution using bogus data ✓

- $\min J$

$\min J_p = J + \alpha_t \|\nabla_t \vec{X}\|^2 + \alpha_x \|\nabla_x \vec{X}\|^2$

$J_{P_1} = r_{P_s} \sum_{r=0}^{R-1} \left\| \frac{\partial P_s}{\partial t} \right\|^2$ $\quad \because P_s$: surface pressure



↳ suppress external gravity wave oscillation

$J_{P_2} = r_D \sum_{r=0}^{R-1} \left\| \frac{\partial D}{\partial t} \right\|^2$ $\quad \because$ D : divergence

$\min J$

$\min J + J_{P_1} + J_{P_2}$

We need $J, \nabla J, \nabla J_{P_1}, \nabla J_{P_2}$

$\frac{\partial P_s}{\partial t} = \frac{P_s(t_r+1) - P_s(t_r)}{\Delta t}$

$J_{P_s}(\vec{X}_0) = \frac{\partial J}{\partial P_{t_r}, P_{t_{r+1}}}$

$J = \frac{r_{ps}}{\Delta t^2} (P_s(t_r+1) - P_s(t_r))^T (P_s(t_r+1) - P_s(t_r))$

forcing term ⟶ $\frac{2 r_{ps}}{(\Delta t)^2} (P_s(t_r+1) - P_s(t_r)) \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ for t_{r+1} / for t_r

$\because J(\vec{X}_0) = f(\vec{X}_t)$

$t_0 \xleftarrow[\text{linear}]{\text{adjoint}} \frac{\partial f}{\partial x(t)}$

- Nudging

$\frac{\partial \alpha}{\partial t} = F(x,t) + G_\alpha (\hat{\alpha}_o - \alpha)$ $\quad \because \alpha \rightarrow \begin{pmatrix} T \\ u \\ v \\ q \\ \vdots \end{pmatrix}$ any variable

$\alpha \in x$

ex) $\frac{\partial T}{\partial t} = F(x,t) + \underbrace{G_T (T^{obs} - T)}_{\text{nudging term}}$

where G_T : nudging coefficient

⟶ determine the magnitude of the nudging term relative to all the other model processes in F

based on
- not too large
  - a. is similar in magnitude to the Coriolis parameter
    - Slowest physical adjustment process
  - b. must satisfy numerical stability criteria $G_\alpha < \frac{1}{\Delta t}$
- not too small

typical value : $G_\alpha = 3 \times 10^{-4} s^{-1} \sim 6 \times 10^{-4} s^{-1}$

✓ • Singular vectors ~ called "optimal perturbation"

→ definition : perturbations maximizing the growth of a chosen norm over a finite time interval.

✓ norm
- total energy norm
- kinetic energy norm
- enstrophy norm
- L_2 norm

$\|X'(t)\|_E = \langle X'(t), X'(t) \rangle_E = E(t)$

perturb. — any norm — function of time

time interval : $[t_0, \tau]$

$E(t) = \langle X'(t), X'(t) \rangle_E = \langle P_t X'(0), P_t X'(0) \rangle$

$\uparrow$ $X'(t) = P_t X'(0)$ (TGL model)

$= \langle P_t^{*E} P_t X'(0), X'(0) \rangle$

$\because P_t^{*E}$ is the adjoint operation of P_t with respect to E-norm

$\lambda \equiv \frac{E(\tau)}{E(0)} = \frac{\langle P_\tau^{*E} P_\tau X_0', X_0' \rangle}{\langle X_0', X_0' \rangle}$ ———— ⊛

☆ $K = P_\tau^{*E} P_\tau$ ⟶ a self-adjoint operator

$K \nu_i = \sigma_i^2 \nu_i$ $\quad$ where ν_i — eigenvectors, σ_i^2 — eigenvalues

The eigenvector corresponds to the largest eigenvalue of K is the perturbation which maximizes the E-norm

→ Let's prove

$X_0' = \sum a_i \nu_i$

⊛ →

$\lambda = \frac{E(\tau)}{E(0)} = \frac{\langle \sum_i a_i K \nu_i, \sum_i a_i \nu_i \rangle}{\langle \sum_i a_i \nu_i, \sum_i a_i \nu_i \rangle}$

$\because \nu_i$ orthogonal $\langle \nu_i, \nu_j \rangle = 0$

$\lambda = \frac{\sum_i a_i^2 \sigma_i^2 \|\nu_i\|}{\sum_i a_i^2 \|\nu_i\|} \leq \sigma_1^2$

→ applications
(1) predictability study
(2) instability study
(3) Adaptive observations.

© Ensemble forecast

I. Why ensemble?

1. Atmosphere is a chaotic system

2 analysis is never perfect due to

- (a) measurement error
- (b) incomplete data coverage
- (c) approximation one has to use in analysis technique

II. How to generate ensemble perturbation?

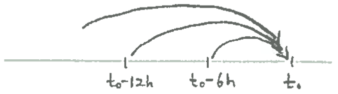
background
- Climatology
- model forecast
  - transporter of information
    - data-rich → data poor
    - variable measured → others
  - analysis error growth
    - fast-grow errors
    - slow-growing errors

(1) random (Monte Carlo) perturbation

random number
- grid points
- spectral coeff
- empirical orthogonal function

Simple, cheaper

(2) Lagged average forecasting (LAF)



$$X_{6h} - X^a_{t_0}$$

$$X_{12h} - X^a_{t_0}$$

(3) Singular vector perturbations

ECMWF

48-h window, energy norm

16 SVs ⟶ 4 leading SV

32 ensemble perturbations + control

(4) Breeding of growing modes

Step1. Generate a small arbitrary perturbation $\vec{X}_0$ to start the breeding cycle at time t_0

Step2: Add the small perturbation $\vec{X}_0'$ to the analysis $\vec{X}_0^a$ at t_0

Step3: Integrate the model for 6h from both the control analysis $\vec{X}_0$ and the perturbed analysis

$$\vec{X}^c_{6h} = M(t_{6h}, t_0)\vec{X}^a_0$$

$$\vec{X}^p_{6h} = M(t_{6h}, t_0)(\vec{X}^a_0 + \vec{X}_0')$$

Step4 Subtract the 6-h forecast:

$$\vec{X}_0' = \alpha(\vec{X}^p_{6h} - \vec{X}^c_{6h})$$

Step5 set $t_0 = t_0 + 6h$ and goto step 2.

III Applications

(i) Improve the forecast skill

by reducing the nonlinear error growth and averaging out unpredicable components.

(ii) Predict the forecast skill, by relating it to the agreement among ensemble forecast members.

© Four-dimensional variational data assimilation (4DVAR)

- Problem

min. J

background info about $\vec{X}_0$

N

N×N

$$J(\vec{X}_0) = \underbrace{(\vec{X}_0 - \vec{X}_b)B^{-1}(\vec{X}_0 - \vec{X}_b)}_{J_b} + \underbrace{\sum_{i=0}^{n}(H(\vec{X}_i) - \vec{d}_i^{obs})O^{-1}(H(\vec{X}_i) - \vec{d}_i^{obs})}_{J_o}$$

control variable
(I.C.

time

→ if
- wind, … → just interpolation
- indirect obs. → physics …

where B : background error covariance matrix

$\vec{d}_i^{obs}$: obs vector at time t_i

O : obs. error coveriance matrix

$H(\vec{X}_i)$: observation operator

$\vec{X}_i$: model forecast at time t_i starting from $\vec{X}_0$, a model state at t_0

→
- How do you define J?
- How do you solve the problem of minimizing J?

-

$$B = E\{(\vec{X}^b(\vec{r}_k) - \vec{X}^t(\vec{r}_k))(\vec{X}^b(\vec{r}_l) - \vec{X}^t(\vec{r}_l))\}$$

$$O = E\{(\vec{X}^o(\vec{r}_k) - \vec{X}^t(\vec{r}_k))(\vec{X}^o(\vec{r}_l) - \vec{X}^t(\vec{r}_l)\}$$

∵ $\vec{X}^t$: truth (what will be a truth?)

→ To do this, lets look at OI.

Optimal interpolation

$$\vec{X}^a(\vec{r}_i) = \vec{X}^b(\vec{r}_i) + \underbrace{W_{ik}}_{\vec{X}^a - \vec{X}^b \text{ will be minimum}}[\vec{X}^o(\vec{r}_k) - \vec{X}^b(\vec{r}_k)]$$

$$\Rightarrow \min E\{(\vec{X}^a - \vec{X}^t)^T(\vec{X}^a - \vec{X}^t)\} \longrightarrow W_{ik}$$

↓ result eg.

grid point

$$\boxed{[B+O]\vec{W}_i = \vec{B}_i}$$

$$\because \vec{W}_i = \begin{pmatrix} W_{i1} \\ W_{i2} \\ \vdots \\ W_{iK} \end{pmatrix}$$

total number of obs

$$\vec{B}_i = \begin{pmatrix} E\{(\vec{X}^b(\vec{r}_1) - \vec{X}^t(\vec{r}_1))^T(\vec{X}^b(\vec{r}_i) - \vec{X}^t(\vec{r}_i) \\ \vdots \\ E\{(\vec{X}^b(\vec{r}_k) - \vec{X}^t(r_k))^T(\vec{X}^b(\vec{r}_i) - \vec{X}^t(\vec{r}_i) \end{pmatrix}$$

dimension K×K, data selection

- notations

$$B = \{B_{kl}\}$$

$$\vec{r} = (x, y) \text{ — horizontal coord.}$$

$$r_{lk} = \sqrt{(x_l - x_k)^2 + (y_l - y_k)^2}$$

assumptions

① The background errors are homogeneous

$$B_{k.l}(\vec{r}_k, \vec{r}_l) = B_{kl}(r, \phi)$$

angle

② The background errors are isotropic

$$B_{k.l}(\vec{r}_k, \vec{r}_l) = E_b^2 \rho_B(r)$$

③ The background errors and observation errors are not correlat

$$E\{(X^b(\vec{r}_k) - X^t(\vec{r}_k))(X^o(\vec{r}_k) - X^t(\vec{r}_k)\} = 0$$

④ The observation errors are uncorrelated.

——— END my note ———

Zous NCAR technical note

① Simple statistical estimation problem

x (parameter) : measured at the same location + time with two different methods.

- y_1, σ_1 : measured value of x, uncertainty
- y_2, σ_2 : 〃 〃

→ use both observations (y_1, y_2) to derive an estimate of x in order to reduce the measurement uncertainty : estimation theory.

Y_1 (random variable) ⟶ y_1 (realizations)

Y_2 (〃 〃) ⟶ y_2 (〃)

→ assume : measurements are unbiased

$$Y_1 = x + \underbrace{E_1}, \quad Y_2 = x + \underbrace{E_2}_{\text{observational noise}} \qquad \text{——— ①}$$

$$y_1 = x + \epsilon_1, \quad y_2 = x + \epsilon_2 \qquad \text{——— ②}$$

realization of random variable E

→ assume : the instruments are unbiased

$$E\{E_1\} = 0, \quad E\{E_2\} = 0 \qquad \text{——— ③}$$

$$E\{E_1^2\} = \sigma_1^2, \quad E\{E_2^2\} = \sigma_2^2 \qquad \text{——— ④}$$

$$E\{E_1 E_2\} = 0 \qquad \text{——— ⑤}$$

→ no correlation between the observational noises

→ estimator(X^*)

$$X^* = a_1 Y_1 + a_2 Y_2 \qquad \text{——— ⑥}$$

→ $E\{X^*\} = x$: the estimator is unbiased.

→ ①, ⑥ ⟶

$$E\{X^*\} = x = E\{a_1 Y_1 + a_2 Y_2\} = a_1 E\{Y_1\} + a_2 E\{Y_2\}$$

$$= a_1 x + a_2 x \qquad \text{——— ⑦}$$

$$\Rightarrow \boxed{a_1 + a_2 = 1} \qquad \text{——— ⑧}$$

$$\to \sigma^2 = E\{(X^* - E\{X^*\})(X^* - E\{X^*\})\} \qquad \text{——— ⑨}$$

⑤, ⑥, $E\{X^*\} = x$ ⟶

minimize

$$\sigma^2 = E\{(X^* - E\{X^*\})(X^* - E\{X^*\})\}$$

$$= E\{(a_1 Y_1 + a_2 Y_2 - x)^2\}$$

$$= E\{(a_1(Y_1 - x) + a_2(Y_2 - x) + (a_1 + a_2 - 1)x)^2\}$$

$$= E\{(a_1(Y_1 - x) + a_2(Y_2 - x))^2\}$$

$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 = a_1^2\sigma_1^2 + (1-a_1)^2\sigma_2^2 \qquad \text{——— ⑩}$$

$$= (1-a_2)^2\sigma_1^2 + a_2^2\sigma_2^2$$

$$\frac{\partial\sigma^2}{\partial a_1} = 0, \quad \frac{\partial\sigma^2}{\partial a_2} = 0$$

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \qquad \text{——— ⑪}$$

⑥ →

$$X^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} Y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} Y_2 \qquad \text{——— ⑫}$$

estimate (x^*)

$$x^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y_2 \qquad \text{——— ⑬}$$

If $\sigma_1 \to \infty$, $x^* = y_2$

If $\sigma_1 \to 0$, $x^* = y_1$

⑩, ⑪ ⟶

$$\sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad : \text{estimation error} \qquad \text{——— ⑭}$$

↳ $\sigma^2 \le \sigma_1^2$, $\sigma^2 \le \sigma_2^2$

⑭ →

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \qquad \text{——— ⑮}$$

↳ precision

⇒ the best linear unbiased estimator (BLUE)

- a different approach to estimate x^*

$$\min. \ J(x) = \frac{(x - y_1)^2}{\sigma_1^2} + \frac{(x - y_2)^2}{\sigma_2} \qquad \text{——— ⑯}$$

↳ cost function

$$\frac{dJ}{dx}(x^*) = 0 = 2\frac{(x^* - y_1)}{\sigma_1^2} + 2\frac{(x^* - y_2)}{\sigma_2^2} \qquad \text{——— ⑰}$$

$$\therefore \ x^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y_2 \qquad \text{——— ⑱}$$

there is no counterpart of eq⑭

② optimal statistical estimation (based on the above note)

- X (random vector) $X^1, \ldots X^N$
- Y (observation) $Y^1, \ldots, Y^M$

Similar to ①

$$Y = HX + E \qquad \text{——— ⑲}$$

observation operator (M×N matrix)

the obs. could be different from the field to be estimated. In such a case H would also represent some physical transformation for instance the calculation of satellite radiance from pressure, T and q profil

where H : operator which interpolate X into the observation locations.

X, Y, E : random vectors

the observational noise is assumed to have zero mean

$$E\{E\} = [E\{E^1\}, \ldots E\{E^M\}]^T = [0, \cdots, 0]^T \qquad \text{——— ⑳}$$

variance-covariance matrix

$$O = E\{(E - E\{E\})(E - E\{E\})^T\} = E\{EE^T\} \qquad \text{——— ㉑}$$

↳ M×M matrix

↳ assumed to be known

the previous analysis cycle ⟶ x_B

X_B (N dimensional random vector estimator)

↳ B (background information)

$$E\{X_B\} = E\{X\} = \bar{x}$$

$$B = \{(X_B - X)(X_B - X)^T\}$$

↳ a priori probability distribution

→ the quantities assumed known in this estimation problem are (the observation operator H, a realization y with its uncertainty matrix O of the observations random vector Y and a prior estimate x_B (a realization of the random variable X_B) with its uncertainty matrix B.)

$$X^* = \underset{N \times N}{A_1} X_B + \overset{N \times M}{A_2} Y$$

$$E\{X^*\} = \bar{x}$$

$$\bar{x} = E\{X^*\} = E\{A_1 X_B + A_2 Y\} = A_1\bar{x} + A_2 H\bar{x}$$

$$A_1 = \underset{\text{unit matrix}}{I_N} - A_2 H \qquad \text{——— ⓐ}$$

$$\to X^* = \underset{\text{background}}{X_B} + A_2\underbrace{\boxed{(Y - \overset{M \times N}{H} X_B)}}_{\text{correction term}}$$

innovation vector

This vector represents the difference between the new information brought in by the observations Y and what is already known from the priori information X_B

✓ estimation error ($\tilde{X}$)

$$\tilde{X} = X^* - X = [X_1^* - X_1, \cdots, X_N^* - X_N]$$

$$\sigma = |\tilde{X}| = \left(\sum_{i=1}^{N} E\{\tilde{X}^i \tilde{X}^i\}\right)^{1/2}$$: norm of the error vector

✓ error covariance matrix

$\hookrightarrow P = E\{\tilde{X}\tilde{X}^T\}$

→ the norm of the estimation error vector is equal to the square root of the sum of the diagonal terms or trace of P

⇒

$$\tilde{X} = X^* - X = A_1 X_B + A_2 Y - X$$

$$= \underset{\leftarrow ⓐ}{A_1}(X_B - X) + A_2(Y - HX) + (\overset{\to 0}{A_1 + A_2 H - I_N})X$$

✓ $$\tilde{X} = (I_N - A_2 H)(X_B - X) + A_2(Y - HX)$$

$$P = E\{\tilde{X}\tilde{X}^T\}$$

(the new measurements are not correlated with the past estimation)

✓ $$P = (I_N - A_2 H)B(I_N - A_2 H)^T + A_2 O A_2^T$$

※ theorems on trace

$$\begin{cases} \mathrm{trace}(A+B) = \mathrm{trace}(A) + \mathrm{trace}(B) \\ \frac{\partial}{\partial A}\mathrm{trace}(ABA^T) = 2AB \end{cases}$$ for any symmetric matrix B

$$\frac{\partial}{\partial A_2}\mathrm{trace}(P) = \frac{\partial}{\partial A_2}\mathrm{trace}\left((I_N - A_2 H)B(I_N - A_2 H)^T\right) + \frac{\partial}{\partial A_2}\mathrm{trace}(A_2 O A_2^T)$$

$$= -2(I_N - A_2 H)BH^T + 2A_2 O$$

$$= 0$$

$$0 = -BH^T + A_2[HBH^T + O]$$

$$A_2 = BH^T[HBH^T + O]^{-1}$$

↳ the Kalman Gain

∴ the unbiased minimum variance estimator

★ $$X^* = X_B + BH^T[HBH^T + O]^{-1}(Y - HX_B)$$

↳ Kalman filter.

↓ realization

$$x^* = x_B + BH^T[\underbrace{HBH^T + O}_{\text{computational burden}}]^{-1}(y - Hx_B) \text{ ——— } ⊛$$

$$P = B - BH^T[HBH^T + O]^{-1}HB$$

⇒ Optimal Interpolation (OI)

A numerical meteorological model is used to integrate the results of the previous analysis to the time of the new analysis. This provides the background term x_B ; while all the observations that are made available during the corresponding analysis time period are collected to build the vector y. At the end of the model integration, the statistical estimate x^* is computed using the expression ⊛. This implies that the matrices O and B are known.

★ ⊙ Variational approach – 3DVAR

Cost function

✓ $$J_{VAR} = \frac{1}{2}(x - x_B)^T B^{-1}(x - x_B) + \frac{1}{2}(y - Hx)^T O^{-1}(y - Hx)$$

※ Precision corresponds to the inverse of the variance-covariance matrix

Find the minimum of J_{VAR}

$$\frac{\partial J}{\partial x} = B^{-1}(x - x_B) - H^T O^{-1}(y - Hx)$$ → can be $<^{\max}_{\min}$.

At the minimum $x = x^*$

$$0 = B^{-1}(x^* - x_B) - H^T O^{-1}(y - Hx^*)$$

$$[B^{-1} + H^T O^{-1} H]x^* = B^{-1}x_B + H^T O^{-1} y$$

$$\therefore x^* = [B^{-1} + H^T O^{-1} H]^{-1}[B^{-1}x_B + H^T O^{-1} y]$$ → Same as ⊛

↳ minimum.

✓ Hessian matrix (second derivative matrix)

★ $$\frac{\partial^2 J}{\partial x^2} = \frac{\partial}{\partial x}\left(B^{-1}(x - x_B) - H^T O^{-1}(y - Hx)\right)$$

$$= B^{-1} + H^T O^{-1} H \quad > 0$$

→ since B + O are symmetric positive definite, the matrix $[B^{-1} + H^T O^{-1} H]$ is also a positive definite.

Inverse of Hessian matrix → estimation of error

$$\hookrightarrow \left(\frac{\partial^2 J}{\partial x^2}\right)^{-1} = B - BH^T[O + HBH^T]^{-1}HB = P$$

$$\therefore \boxed{P = \left(\frac{\partial^2 J}{\partial x^2}\right)^{-1}}$$

$$\rightarrow P^{-1} = \left(\frac{\partial^2 J}{\partial x^2}\right) = B^{-1} + H^T O^{-1} H$$

→ Iterative technique

★ ⊙ Physical space analysis system – 3D-PSAS

$$x^* = x_B + BH^T\underbrace{[O + HBH^T]^{-1}}_{\text{solved by Iterative algorithm}}(y - Hx_B)$$

$$J_{PSAS} = \frac{1}{2}(W - W_B)^T[O + HBH^T](W - W_B) - (W - W_B)^T(y - Hx_B)$$

where $W = (BH^T)^{-1}x$, $W_B = (BH^T)^{-1}x_B$

→ the variational solution is expressed in the phase space while the 3DPSAS uses the physical space for its operation.

★ ⊙ 4-Dimensional variational data assimilation (4D-VAR)

- numerical model

$$\frac{\partial X(t)}{\partial t} = F(X(t)) + W(t) \text{ ——— } Ⓐ$$

where F : all the mathematical functions involved in the meteorological model,

W(t) : random variable figuring the model error

↳ zero mean

covariance matrix Q(t)

white process $E\{W(t)W^T(t')\} = 0$ if $t \neq t'$

- The information available in the 4DVAR

i) a background term X_B with its covariance matrix B

ii) the meteorological model (eq Ⓐ)

iii) observations distributed in time

$$Y(t) = H(X(t)) + E(t)$$

→ zero mean

covariance matrix O(t)

white

$E\{EW^T\} = 0$

✓two different approaches

- filtering solution → kalman filter ✓
- Smoothing solution → better estimate ✓

· assimilation time window $[t_0, t_R]$

$$J = \frac{1}{2}(x_{t_0}-x_B)^T B^{-1}(x_{t_0}-x_B) + \frac{1}{2}\int_{t_0}^{t_R}(y_t - H(x_t))^T O_t^{-1}(y_t - H(x_t))\,dt$$

↳ meteorological information is not used

So,

$$J = J + \frac{1}{2}\int_{t_0}^{t_R}(\overset{\frac{dx}{dt}}{\dot{x}} - F(x_t))^T Q_t^{-1}(\dot{x} - F(x_t))\,dt \quad Ⓐ$$

$$= \boxed{J} + \frac{1}{2}\int_{t_0}^{t_R} W_t^T Q_t^{-1} W_t\,dt \quad \text{———} \quad Ⓑ$$

· The 4D constraint minimization can be reduced to an unconstrained optimization problem by considering the minimization of the Augmented Lagrangian function of J, instead of J itself
↳ (L)

$$L = J + \int_{t_0}^{t_R}\lambda_t^T(\dot{x} - F(x_t) - W_t)\,dt \quad \text{———} \quad Ⓒ$$

↳ input variables $(x_{t_0}, \dot{x}_{t_0}, x_t, \dot{x}_t, x_{t_R}, \dot{x}_{t_R}, W_t, \dot{W}_t, \lambda_t, \dot{\lambda}_t$

where λ_t : a N dimensional vector to be defined

→ The solution to the problem of minimizing J in Ⓑ or L in Ⓒ can be found by introducing an adjoint model and a standard unconstrained minimization algorithm

⊙ A continuous form of adjoint model

· to get an expression of the Euler-Lagrange eqs.

$$L = \frac{1}{2}(x_{t_0}-x_B)^T B^{-1}(x_{t_0}-x_B) + \frac{1}{2}\int_{0}^{t_R}(y_t - H(x_t))^T O_t^{-1}(y_t - H(x_t))\,dt + \frac{1}{2}\int_{t_0}^{t_R} W_t^T Q_t^{-1} W_t\,dt + \boxed{\int_{t_0}^{t_R}\lambda_t^T(\dot{x} - F(x_t) - W_t)\,dt}$$

$$\int_{t_0}^{t_R}\lambda_t^T(\dot{x} - F(x_t) - W_t)\,dt = \int_{t_0}^{t_R}\lambda_t^T\dot{x}\,dt - \int_{t_0}^{t_R}\lambda_t^T(F(x_t)+W_t)\,dt$$

$$= [\lambda_t^T x_t]_{t_0}^{t_R} - \int_{t_0}^{t_R}\dot{\lambda}_t^T x_t\,dt - \int_{t_0}^{t_R}\lambda_t^T(F(x_t)+W_t)\,dt$$

$$L = \frac{1}{2}(x_{t_0}-x_B)^T B^{-1}(x_{t_0}-x_B) + \frac{1}{2}\int_{t_0}^{t_R}(y_t - H(x_t))^T O_t^{-1}(y_t - H(x_t))\,dt + \frac{1}{2}\int_{t_0}^{t_R} W_t^T Q_t^{-1} W_t\,dt$$
$$+ \lambda_{t_R}x_{t_R} - \lambda_{t_0}x_{t_0} - \int_{t_0}^{t_R}\dot{\lambda}_t^T x_t\,dt - \int_{t_0}^{t_R}\lambda_t^T(F(x_t)+W_t)\,dt$$
$$= J + \lambda_{t_R}x_{t_R} - \lambda_{t_0}x_{t_0} - \int_{t_0}^{t_R}\dot{\lambda}_t^T x_t\,dt - \int_{t_0}^{t_R}\lambda_t^T(F(x_t)+W_t)\,dt$$

→ derivatives of L with respect to all the input variables and arrange

$$\begin{cases} x_{t_0}^* = x_B + B\lambda_{t_0} \\ \frac{\partial x^*}{\partial t} = F(x_t^*) + Q(t)\lambda_t \\ -\frac{\partial \lambda}{\partial t} = \left(\frac{\partial F}{\partial x}\right)^T\lambda_t + \frac{\partial H^T}{\partial x}O_t^{-1}(y_t - H(x_t^*)) \\ \lambda_{t_R} = 0 \end{cases}$$

$$\frac{\partial L}{\partial x_0} = 0 \rightarrow$$

$$\frac{\partial J}{\partial x_0} - \lambda_{t_0} = 0$$

↳ the value of the adjoint state λ_{t_0} at the initial time is equal to the value of the gradient of cost function J with respect to the initial conditions x_{t_0}.

⊙ A discretized form of adjoint model

The discretized form of the numerical model eq.

$$x(t_r) = Q_r(x)x_0 \quad \text{———} \quad ⓘ$$

Cost function

$$J(x_0) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{r=0}^{n}(\overset{\text{[illegible]}}{H_r(x_r) - y_r})^T O_r^{-1}(H_r(x_r) - y_r + \boxed{J^P}$$

↳ a penalty term controlling gravity wave oscillations.

$$J = J^b + J^o + J^P = J^b + \sum_{r=0}^{n}(J^o)_r + J^P$$

→ In order to obtain the optimal IC(x_0^*) then minimize J.

$$\nabla_{x_0}J = \nabla J^b + \boxed{\nabla J^o} + \nabla J^P$$

→ requires the adjoint model integration

↳ $\nabla J^b = B^{-1}(x - x_b)$

Let's consider the change in J resulting from a small perturbation x_0'.

$$J^{o\prime}(x_0) = J^o(x_0 + x_0') - J^o(x_0)$$

$$J^{o\prime}(x_0) = (\nabla J^o(x_0))^T x_0' + O(\|x_0\|^2)$$

forecast difference

$$J^{o\prime}(x_0) = \sum_{r=0}^{n} H_r^T(O_r^{-1}(H_r(x_r) - y_r))^T \boxed{x_r'} + O(\|x_r\|^2)$$

$$(\nabla J^o(x_0))^T x_0' = \sum_{r=0}^{n} H_r^T(O_r^{-1}(H_r(x_r) - y_r))^T x_r' \quad \text{———} \quad ⓘⓘ$$

ⓘ → tangent linear model (TLM)

$$x'(t_r) = P_r(x)x_0$$

ⓘⓘ →

$$(\nabla J^o(x_0))^T x_0' = \sum_{r=0}^{n} H_r^T(O_r^{-1}(H_r(x_r) - y_r))^T P_r x_0'$$

$$\|x_0'\| \rightarrow 0$$

$$\nabla J^o(x_0) = \sum_{r=0}^{R} P_r^T H_r^T O_r^{-1}(H_r(x_r) - y_r)$$

$$\therefore \nabla_{x_0}J^o = \sum_{r=0}^{R}\hat{x}_0^r$$

where
$$\hat{x}_0^r = \boxed{P_r^T(x)}\hat{x}(t_r)$$
$$\hat{x}(t_r) = H_r^T O_r^{-1}(H_r(x_r) - y_r), \quad r = 0, 1, \cdots, R$$

⊙ Parameter estimation

Lagrangian multiplier method.

$$J(\alpha^*) \le J(\alpha), \ \forall\alpha$$

↳ model parameter

$$\frac{\partial x}{\partial t} = F(x, \alpha)$$

vector of Lagrangian multipliers.

$$L(x, \alpha, \hat{x}) = J + \left\langle \hat{x}, \frac{\partial x}{\partial t} - F(x, \alpha)\right\rangle$$

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial \hat{x}} = 0, \quad \frac{\partial L}{\partial \alpha} = 0$$

→ stationary point of L

$$\begin{cases} \frac{\partial x}{\partial t} = F(x, \alpha) \text{ (nonlinear model)} \\ -\frac{\partial \hat{x}}{\partial t} = -\left(\frac{\partial F(x)}{\partial x}\right)^T \hat{x} + \frac{\partial J}{\partial x} \text{ (adjoint model)} \\ \hat{x}\big|_{t=t_R} = 0 \\ \frac{\partial J}{\partial \alpha} + \int_{t_0}^{t_R} < \hat{x}, \frac{\partial F(x,\alpha)}{\partial \alpha} > dt = 0 \end{cases}$$

$$\nabla_\alpha J(\alpha) = \frac{\partial L}{\partial \alpha}$$

END