Forward Sensitivity Approach to Dynamic Data Assimilation

S. Lakshmivarahan
School of Computer Science
University of Oklahoma

Joint work with John M. Lewis
NSSL, Norman and DRI, Reno

Source – LL(2010)

- S. Lakshmivarahan and J. M. Lewis, (2010) “Forward Sensitivity Approach to Dynamic Data Assimilation”, Advances in Meteorology, Volume xx, 2010, doi:10.1155/2010/375615

Introduction – Forecast error

- Models are **abstractions of reality**
- Observations are samplings from (**unknown**) **true state** of the system modulo (**unobservable**) **random** (measurement) **errors**
- **Goodness** of a model is judged by its ability to **explain** the observations
- Difference between the observation and model forecast is called **forecast error, e**

Genesis of forecast errors

- Use of **inadequate or defective** mathematical models, e.g. sub-grid processes are not captured by the model
- Use of **wrong values** for the control variables - initial and boundary conditions and parameters in the model
- The physical or the empirical mapping, h (also called the forward or observation operator) that relates the model state, x(t) to the observable, z(t) may **not fully capture** the inherent underlying relation
- Observations are corrupted by (unobservable) measurement error

Note: We can have a combination of one or more of the above factors.

Some of these are more **tractable** than others

Tractability requires basic assumptions

- **Thanks to modelers -** great strides have been made in the development of good, acceptable models – we assume that our **model is perfect**
- It is also assumed that the relation, h between model state and the observable has been well tested and is good/acceptable.
- Consequently, the forecast errors are only due to

(1) errors in I.C., B.C. and/or parameters and

(2) observational errors

Structure of forecast errors

- We posit that the forecast error

  e = e (det) + e (ran) where

- e (det) is the deterministic part called the **forecast error** and

- e (ran) is induced by the (unobservable) random measurement errors

- **Goal:** To isolate the deterministic component of the forecast error

Choices for the framework

- Models of interest in geophysical domain arise as a system of **nonlinear coupled PDEs**
- These can be converted into a system of nonlinear ODEs by Galerkin type projection methods, also known as **spectral methods**
- Discretized using standard finite-difference methods into a system of **nonlinear difference (algebraic) equations**
- Our frame work has been adapted to cover models described by a system of ODEs or difference equations – Refer to the Technical Report (2009)
- For concreteness, we use the ODE formulation

Model forecast

- $t \geq 0$ time
- $X(t) \in R^n$ denote the state of the system at time $t \geq 0$, R^n is called model space
- $\alpha \in R^p$ vector of parameters, R^p is called parameter space
- $f: R^n \times R^p \times R \to R^n$ $f(X, \alpha, t) = (f_1, f_2, ..., f_n)^T$, $f_i = f_i(X, \alpha, t)$
- $\dot{X} = f(X, \alpha, t)$ dynamic systems, $X(0)$ is the initial condition
- f satisfies the usual conditions for the existence and uniqueness of the solution
- $X(t) = X(X(0), \alpha, t)$ is smooth in $X(0), \alpha$ *and* t, $(X(0)^T, \alpha^T)^T \in R^n \times R^p$ is called control space

Observation operator

- $h: R^n \rightarrow R^m$ is a map that relates the model state to observation
- The **actual observation** at time t:

$$z(t) = h(x^*) + v(t)$$

- x^* is the (**unknown**) **true** state
- v is the (**unobservable**) measurement error
- $v \sim N(0, R)$ – R is the **known** error covariance matrix.

Forecast error

- **Model counterpart** of the observation:

$$z^M(t) = h(x(t))$$

  where $x(t)$ is the **predicted** model state at time starting from a given IC/BC/parameter values

- **Forecast error**: $e(t) = z(t) – z^M(t)$

A decomposition of the forecast error: definition of forecast bias

- Recall $z(t) = h(x^*(t)) + v(t)$
- **Forecast error**:

  $e(t) = [h(x^*(t)) + v(t)] – h(x(t)) = b(t) + v(t)$

- **Forecast bias** $b(t) = h(x^*(t)) – h(x(t))$
- $b(t)$ is the **deterministic part** of the forecast error called the **forecast bias**
- $v(t)$ is the (unobservable) **random part** of the forecast error

- **Note**: If the observational noise does not have a zero mean, this non-zero mean will add another component to the forecast bias

A classification of forecast errors

- **Case I**: If the model is perfect and IC and /BC and parameters are known exactly, then $x(t) = x^*(t)$ and $b(t) = 0$. In this case, $e(t) = v(t)$, random errors and $e(t) \sim N(0, R)$ and $E(e(t)) = 0$

- **Case II**: If the model is perfect but there are errors in IC and/or BC and/or parameters, then $x(t) \neq x^*(t)$ and $e(t) = b(t) + v(t)$ and $e(t) \sim N(b(t), R)$

- In this case $b(t) = b(c, t)$ is only a function of the control vector c (initial/boundary conditions and parameters and) time t and $E(e(t)) = b(t)$

Classification of forecast errors

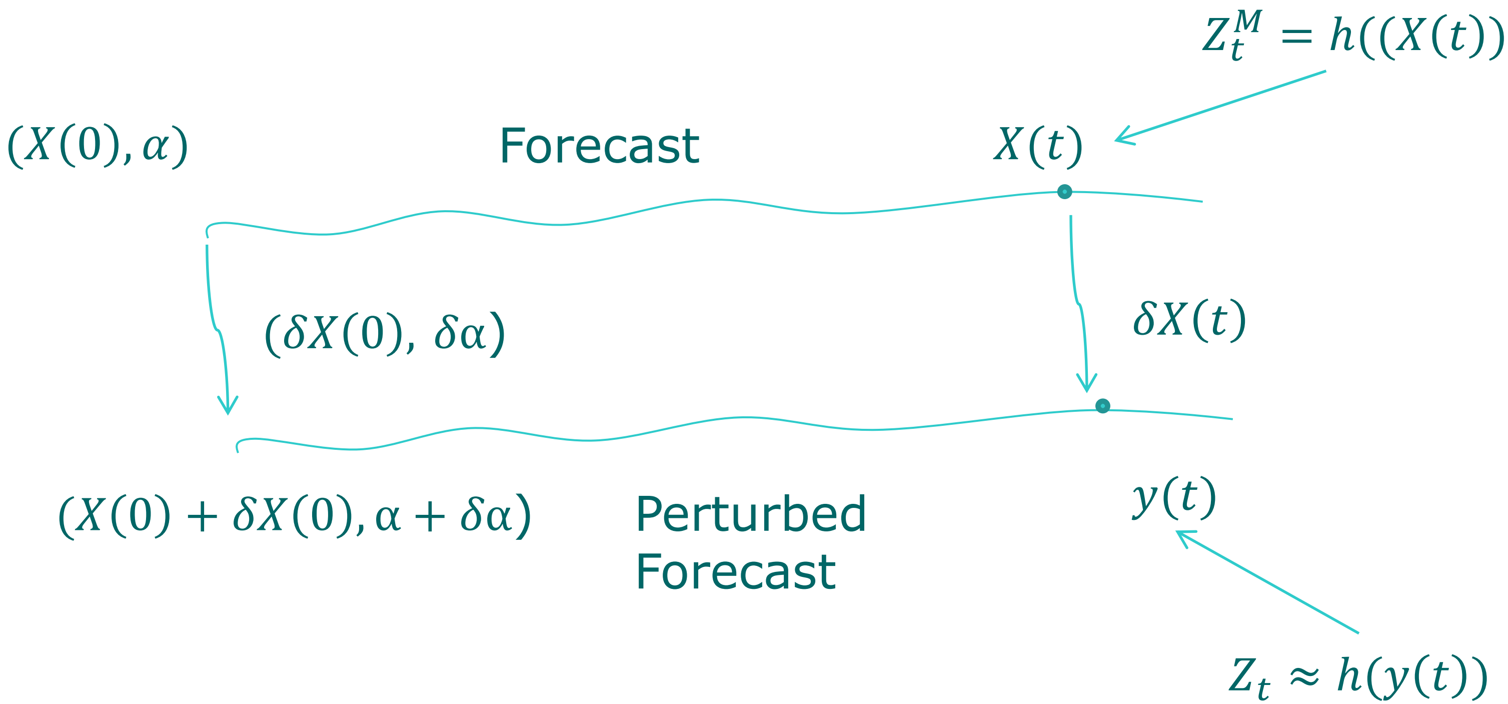
- Case III: If the model f(.) and/or the operator h(.) are in error, then $e(t) = b(t) + v(t) \sim N(b(t), R)$ with $b(t) = b(f, h, c, t)$ – a complex structure

- **Note**: This is the most challenging case and we avoid it by suitable assumptions: model is perfect, h(.) is good, etc.

Statement of problem – forecast error correction

- Given the perfect model f(.) and a set values of the IC, BC, and the parameters
- Given the noisy observation, z(t), the observation operator h(.), and observational error covariance, R
- Find **corrections or perturbations** to the initial conditions and parameters that **annihilates** the forecast bias

A path way to the solution- a pictorial view



Pick $(\delta X(0),\ \delta\alpha)$:

$$Z_t \approx h\big(y(t)\big) = h(x(t) + \delta\ x(t)) = h\big(x(t)\big) + \delta h = Z_t^M + \delta h$$

$$e(t) = Z_t - Z_t^M = \delta h$$

A pathway to the solution: a first-order method

- Let x(t) be the forecast from the control

  c = (x(0), α)

- Let $\partial X(0)\, and\, \partial\alpha$

  be the perturbations in the initial condition x(0) and parameter α

- Let y(t) be the forecast obtained from the perturbed state

  $X(0)+\partial X(0)\, and\, \alpha+\partial\alpha$

- If $\partial X(t)$ is **first variation** of the solution x(t) induced by

  the perturbations in the control, then y(t) = x(t) + $\partial X(t)$

A pathway to solution

- The first variation $\partial X(t)$ in turn induces a

  variation ∂h in h(x(t))

- Then
- h(y(t)) = h(x(t) + $\partial X(t)$)

  = h(x(t)) + ∂h

The defining equation for the first - order method

- To achieve the goal of annihilating the forecast error, set

$0 = z(t) – z^M(t)$

$= [h(x^*(t)) + v(t)] – h(y(t))$

$= [h(x^*(t)) +v(t)]- h(x(t)) – \partial h$

$= e(t) – \partial h$

or

$e(t) = \partial h$

Note: e(t) is the observed forecast error with respect to the forecast obtained from x(0) and alpha

It remains to compute del(h) to which we now turn

Compute δh

- From first principles

$$\delta h = D_X(h) \cdot \delta X \qquad (2)$$

- $D_X(h) \in R^{m \times n}$ is Jacobian matrix of h with respect to X

$$D_X(h) = \begin{bmatrix} \frac{\partial h_1}{\partial X_1} & \frac{\partial h_1}{\partial X_2} & \cdots & \cdots & \frac{\partial h_1}{\partial X_n} \\ \frac{\partial h_2}{\partial X_1} & \frac{\partial h_2}{\partial X_2} & \cdots & \cdots & \frac{\partial h_2}{\partial X_n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \frac{\partial h_m}{\partial X_1} & \frac{\partial h_m}{\partial X_2} & \cdots & \cdots & \frac{\partial h_m}{\partial X_n} \end{bmatrix} \in R^{m \times n}$$

Compute δh

- Also $\delta X(t) = D_{X(0)}(X(t)) \cdot \delta X_0 + D_\alpha(X(t)) \cdot \delta\alpha$ (3)
- $D_{X(0)}(X) \in R^{n \times n}$ is the Jacobian matrix of $X(t)$ with respect to $X(0)$
- $D_\alpha(X) \in R^{n \times p}$ is the Jacobian matrix of $X(t)$ with respect to α

$$D_{X(0)}(X(t)) = \begin{bmatrix} \frac{\partial X_1(t)}{\partial X_1(0)} & \frac{\partial X_1(t)}{\partial X_2(0)} & \cdots & \cdots & \frac{\partial X_1(t)}{\partial X_n(0)} \\ \frac{\partial X_2(t)}{\partial X_1(0)} & \frac{\partial X_2(t)}{\partial X_2(0)} & \cdots & \cdots & \frac{\partial X_2(t)}{\partial X_n(0)} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \frac{\partial X_n(t)}{\partial X_1(0)} & \frac{\partial X_n(t)}{\partial X_2(0)} & \cdots & \cdots & \frac{\partial X_n(t)}{\partial X_n(0)} \end{bmatrix} \in R^{n \times n} \quad D_\alpha(X(t)) = \begin{bmatrix} \frac{\partial X_1}{\partial \alpha_1} & \frac{\partial X_1}{\partial \alpha_2} & \cdots & \cdots & \frac{\partial X_1}{\partial \alpha_p} \\ \frac{\partial X_2}{\partial \alpha_1} & \frac{\partial X_2}{\partial \alpha_2} & \cdots & \cdots & \frac{\partial X_2}{\partial \alpha_p} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \frac{\partial X_n}{\partial \alpha_1} & \frac{\partial X_n}{\partial \alpha_2} & \cdots & \cdots & \frac{\partial X_n}{\partial \alpha_n} \end{bmatrix} \in R^{n \times p}$$

- $D_{X(0)}(X(t))$, $D_\alpha(X(t))$ are called first-order sensitivity functions

Compute δh

- Substitute δX in (2) with (3)

  We have $$\delta h = \left[D_X(h) D_{X(0)}(X) \quad D_X(h) D_\alpha(X) \right] \begin{bmatrix} \delta X(0) \\ \delta \alpha \end{bmatrix} \qquad (4)$$

- Let

$$H_1 = D_X(h) \cdot D_{X(0)}(X) \in R^{m \times n}$$

$$H_2 = D_X(h) \cdot D_\alpha(X) \in R^{m \times p}$$

$$H = \left[H_1, H_2 \right] \in R^{m \times (n+p)}$$

$$\beta_1 = \delta X_0 \in R^n$$

$$\beta_2 = \delta \alpha \in R^p$$

$$\beta = (\beta_1^T, \beta_2^T)^T = (\delta X(0)^T, \delta \alpha^T)^T \in R^{n+p}$$

(4) becomes $H\beta = e_t$, where $e_t \in R^m$

Least Square Method

- Using least square method (J. M. Lewis, S. Lakshmivarahan and S. K. Dhall *(2006)* **Dynamic Data Assimilation**, Cambridge University Press, 654 pages, Chapter 5) to solve

$$H\beta = e_t$$

$$\beta = \begin{cases} (H^T H)^{-1} H^T e_t, & if\ m > (n+p), overdetermined \\ H^{-1} e_t & if\ m = (n+p) \\ H^T (HH^T)^{-1} e_t, & if\ m < (n+p), underdetermined \end{cases}$$

Least square solution

- But recall that from case II

$$e(t) = b(t) + v(t)$$

- Substituting this we get

$$\beta = (H^T H)^{-1} H^T [b(t) + v(t)]$$

Taking expectations on both sides we get

$$E(\beta) = (H^T H)^{-1} H^T b(t)$$

Sensitivity Functions

- In Gulf problem we first analyzed, the solution x(t) is known explicitly using which we can compute the first-order sensitivity functions in closed form

- If the model equations are not solvable explicitly, then $D_{X(0)}(X(t))$ and $D_{\alpha}(X(t))$ have to be numerically computed

Solve Parameter Sensitivity Function

- Following H. Rabitz, M. Kramer, and D. Dacol (1983) "Sensitivity Analysis in Chemical Kinetics", Annual Review of Physical Chemistry, Vol 334, pp 419-461
- Consider

$$\dot{X} = f(X, \alpha, t)$$

$$\frac{\partial \dot{X}}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left(\frac{dX}{dt}\right) = \frac{d}{dt}\left(\frac{\partial X(t)}{\partial \alpha}\right) = \frac{d}{dt}(D_\alpha(X(t))$$

$$= D_X(f)D_\alpha(X) + D_\alpha(f)$$

$$\Rightarrow \quad \frac{d}{dt}(D_\alpha(X(t)) = D_X(f)D_\alpha(X) + D_\alpha(f)$$

- This is a linear ordinary differential equation that describes the time evolution of the first-order sensitivity function $D_\alpha(X(t)$, I.C. $D_\alpha(X(0) \equiv 0$

Solve I.C. Sensitivity Function

- Similarly,

$$\frac{\partial \dot{X}}{\partial X(0)} = \frac{\partial}{\partial X(0)}(\frac{dX}{dt}) = \frac{d}{dt}(\frac{\partial X}{\partial X_0}) = \frac{d}{dt}[D_{X(0)}(X(t))]$$
$$= D_X(f)D_{X(0)}(X(t))$$

$$\Rightarrow \quad \frac{d}{dt}[D_{X(0)}(X(t))] = D_X(f)D_{X(0)}(X(t))$$

- This is a linear ordinary differential equation that describes the evolution of the first-order sensitivity function $D_{X(0)}(X(t)$, I.C. $D_{X(0)}(X(0) \equiv I$, Identity Matrix

Summary of the framework: single observation

- **Step 1**

  Solve

$$\dot{X} = f(X, \alpha, t) \qquad X(0)$$

$$\frac{d}{dt}(D_\alpha(X(t)) = D_X(f)D_\alpha(X(t)) + D_\alpha(f) \qquad D_\alpha(X(0)) \equiv 0$$

$$\frac{d}{dt}[D_{X(0)}(X(t))] = D_X(f)D_{X(0)}(X(t)) \qquad D_{X(0)}(X(0) \equiv I$$

  Using 4th order Runge-Kutta methods and compute $D_{X(0)}(X(t))$ and $D_\alpha(X(t))$

- **Step 2**

  Solve for β , $H_{m\times(n+p)}\beta_{(n+p)\times 1} = e_t\,|_{m\times 1}$

Summary of the framework: multiple observations

- **Step 1**

  solve

$$\dot{X} = f(X, \alpha, t)$$

$$\frac{d}{dt}(D_\alpha(X(t)) = D_X(f)D_\alpha(X(t)) + D_\alpha(f)$$

$$\frac{d}{dt}[D_{X(0)}(X(t))] = D_X(f)D_{X(0)}(X(t))$$

  similarly as in single observation case

- **Step 2**

  solve β for

$$\begin{bmatrix} H_1(t_1) & H_2(t_1) \\ H_1(t_2) & H_2(t_2) \\ \vdots & \vdots \\ H_1(t_N) & H_2(t_N) \end{bmatrix}_{Nm\times(n+p)} \beta_{(n+p)\times 1} = \begin{bmatrix} e_{t_1} \\ e_{t_2} \\ \vdots \\ e_{t_N} \end{bmatrix}_{Nm}$$

Relation to 4-D Var

$$J(c) = \frac{1}{2}\sum_{k=1}^{N}[Z_k - h(X_k)]^T R_k^{-1}[Z_k - h(X_k)] \longrightarrow (1)$$

Then

$$J(c+\delta c) = \frac{1}{2}\sum_{k=1}^{N}[Z_k - h(X_k+\delta X_k)]^T R_k^{-1}[Z_k - h(X_k+\delta X_k)] \longrightarrow (2)$$

Relation to 4-D Var

But

$$h(X_k + \delta X_k) = h(X_k) + D_{X_k}[h]\delta X_k$$

and

$$\delta X_k = D_{X(0)}[X_k]\delta X(0) + D_\alpha[X_k]\delta\alpha$$

$$h(X_k + \delta X_k) = h(\mathrm{X_k}) + \mathrm{H_{1,k}}\,\delta X(0) + \mathrm{H_{1,k}}\delta\alpha \longrightarrow (3)$$

$$= h(\mathrm{X_k}) + \mathrm{H_k}\delta\zeta$$

Where

$$\mathrm{H_k} = [\mathrm{H_{1k}}, \mathrm{H_{2k}}]$$

Relation to 4-D Var

Substitute (3) in (2)

$\Longrightarrow$

$$Q(\zeta) = \frac{1}{2}\sum_{k=1}^{N}(e_k - \mathrm{H_k}\zeta)^T R_k^{-1}\,(e_k - H_k\zeta) \longrightarrow (4)$$

Minimization of (4) w.r.t ζ is exactly solution to FSM.

Relation to other known methods

- When there is only one observation, the forward sensitivity method described is equivalent to the well known 3D-VAR method
- When there are multiple observations, this method is equivalent to the 4D-VAR method
- In the case of 4D-VAR one has to solve the adjoint equation backward but in this method we need to compute the solution to the forward sensitivity equations
- Both the adjoint equation and the forward sensitivity equations are both linear however.
- It can be shown that the forward sensitivity method is better when there are multiple observations and when we operate sequentially

Extensions

- If a PDE based model (after discretization) is given by a nonlinear difference equations, there is a **discrete version** of the framework
- If f(.) and h(.) highly nonlinear, we may need to consider the **second-order variation** leading to the second-order method
- These variations are given in LL(2009)
- This framework has been applied:

  1) A three equation model for the gulf problem

  2) Burgers 1932 model

  3)Logistic equation for population model

  4) Burgers PDE describing the 1D turbulence

An example of return flow – a part of the model

- Cold continental air moves over a warm ocean surface at a constant temperature
- In the Lagrangian framework, the air moves with the prevailing wind at a low speed
- Turbulent transfer of heat from the ocean to the air warms the air and the warm returns back to Gulf states which in turn decides the nature and type of precipitation

$$d\theta / d\tau = (CV / H)(\theta_s - \theta)$$

A scalar model

- Nondimemsionalized version: $\theta = \pi x, \tau = Tt$
- $\tau = 1.0 \,\&\, T = 1.0$
- The model becomes: $dx/dt = f(x, \alpha) = k(x_s - x)$
- The solution is: $x(t) = (x_0 - x_s)\exp(-kt) + x_s$
- Observation z(t) = x(t) + v(t)
- Sensitivity of x(t):

$$\partial x(t)/\partial x_0 = \exp(-kt)$$

$$\partial x(t)/\partial x_s = 1 - \exp(-kt)$$

$$\partial x(t)/\partial k = (x_s - x_0)t\exp(-kt)$$

Behavior of the solution and three of its sensitivities

- A plot of the solution and its sensitivities

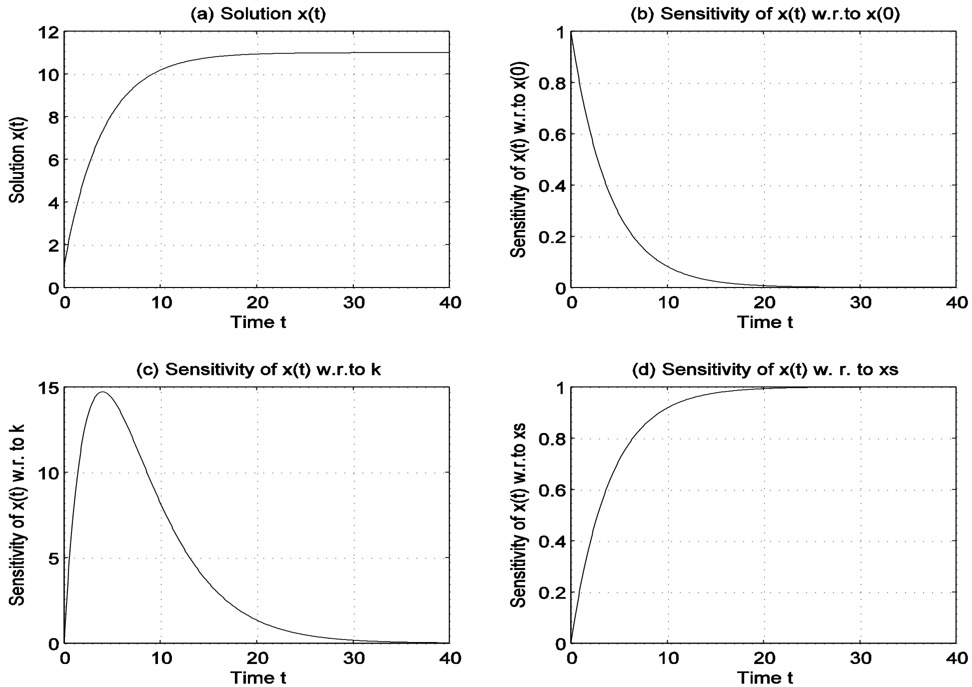


Figure 1. Evolution of the solution and its sensitivities

A cluster of corrections

Using 6 observations: t = 2, 7, 12, 17, 22 and 27

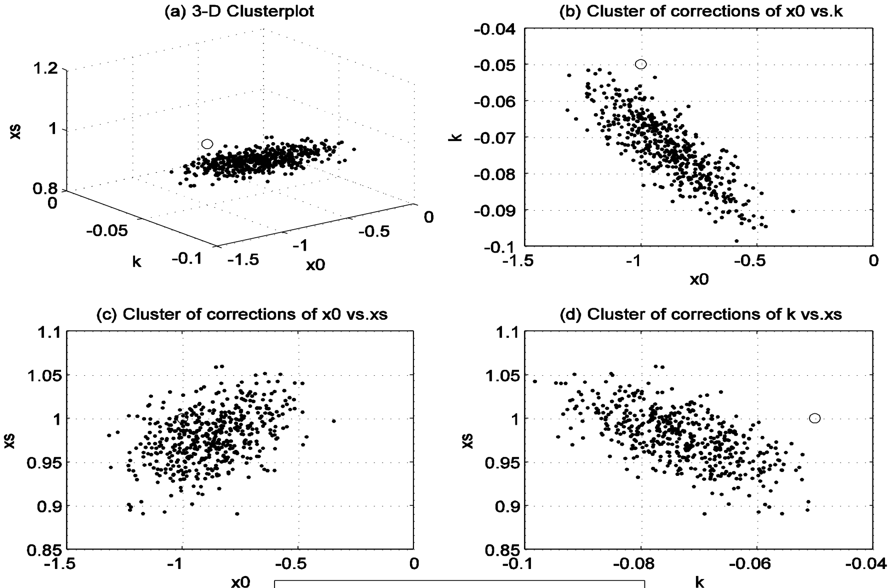


Figure 2. 3-D cluster and its projections

Iterative corrections

- Using 6 observations: t = 2,7,12,17,22 and 27

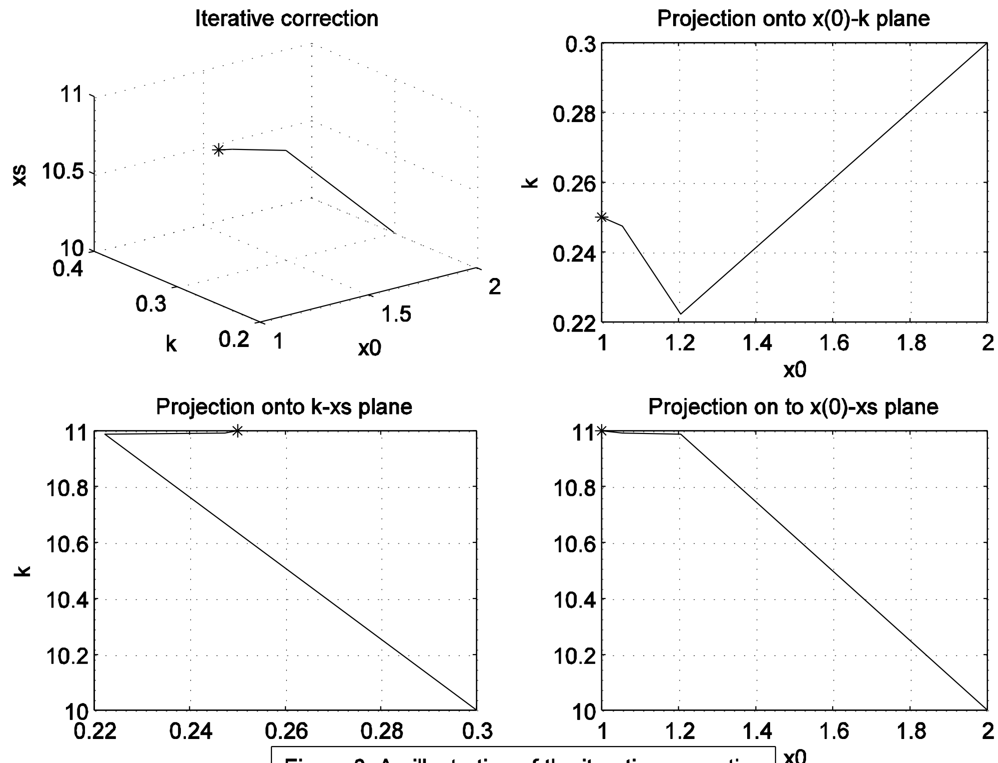


Figure 3. An illustration of the iterative correction

Comments

- Recall that in this example the solution exhibits saturation
- If we pick all the observations in the saturated region, the sensitivities exhibit linear dependence and hence leads to ill-conditioning
- In this case we cover the sea surface temperature well but we cannot recover the initial condition
- Thus, sensitivity based analysis would provide clue as to where to place the observations so that the estimation problem is well posed
- Currently working on return flow + chemistry problem