Impact of Observations on Adjoint Sensitivity

S. Lakshmivarahan

School of Computer Science
University of Oklahoma
Norman, OK - 73019, USA
varahan@ou.edu

January 11, 2012

- Acknowledgement: Joint work with John Lewis of NSSL/DRI

Outline

- Motivation
- Computation of adjoint sensitivity
- Forward Sensitivity
- Relation between the two sensitivities
- Impact of observations
- Example

Motivation

- Quantifying the impact of observations on forecast aspect - rainfall, has a long and cherished history
- Observation System Simulation Expriments (OSSE) are still being used to assess the improvement in the analyses using various sets of model generated observations
- Targeted Observation approach on the otherhand, strives to identify the impact of a particular subset of observations on a specified model forecast aspect
- This latter approach relies on expresing the sensitivity of the forcast aspect w.r.to observation using the adjoint sensitivity
- By expressing the adjoing sensitivity as a linear combination of the the forward sensitivities, we uncover the fine structure of the adjoint sensitivity using which we can identify the temporal zones where the observations are likely to have little or no influence in the adjoint sensitivity

Literature on OSSE

- E. Andersson, A. Hollingsworth, G. Kelly, J. Pailleux and Z. Zhang (1991) Global observation system experiments on operational staistical retrivals of satellite data, MWR, Vol 119, 1851-1865
- R. Atlas (1997) Atmospheric observations and experiments to assess their usefulness in data assimilation, J. Meteor. Soc. Japan 75, 111-130
- C. Cardinali (2009) Monitoring the observations impact on the short range forecast, Q.J.R.M.S. Vol 135, 1469-1480

Literature on Targeted observations

- N.L. Barker and R. Daley (2000) Observations and background adjoint sensitivity in adaptive observation-targeting probelm, QJR.M.S., Vol 126, pp1431-1451
- R. Errico (2007) Interpretations of an adjoint-derived observational impact measure, Tellus, Vol 59-A, 272-276
- D. N. Daescu and R. Toddling (2009) Adjoint estimation of the variation in model functional output due to assimilation of data, MWR, Vol 137, 1705-1716
- R. Gelaro, Y. Zhu, and R. Errico (2007) Examination of various-order adjoint based approximation of observation impact, Meteor. Zeit. Vol 16, 685-692

Adjoint method - a resume

- $x \in R^n, a \in R^p, M : R^n \times R^p \to R^n$
- Model: $x(k+1) = M(x(k), a)$
- x(0) Initial condition, a is parameter
- $h : R^n \to R^m$ is the forward operator
- Observation: $z(k) = h(x^*(k)) + v(k)$
- x*(k) is the unknown but true state of the system
- v(k) is the white Gaussain noise with known covariance (i.e.) $N(0, R)$

Forecast error and cost functional

- Error $e(k) = z(k) - h(x(k)) = [h(x^*(k)) - h(x(k))] + v(k)$
- Cost functional: $J : R^n \times R^p \to R$
- $J(x(0), a) = \frac{1}{2} \sum_{k=1}^{N} \left\langle e(k), R^{-1} e(k) \right\rangle$
- $\langle x, Ay \rangle$ denotes the inner product of two vectors x and Ay
- Goal is to minimize J w.r.to (x(0), a)

Optimization problem

- This is an off-line (batch) constrained minimization where the model equations define the equality constraints
- There are two ways to formulate it - storng or weak constrained problem
- 4-D VAR first-order adjoint method is a result of the strong constrained formulation
- Solved by using the Lagrangian multiplier technique or can also be solved by using the method of first variation which is followed below
- Weak constrained formulation is useful to account for the model errors and uses a penalty functional

First variation of J

- $\delta J(x(0), a)) = \sum_{k=1}^{N}(f(k), \delta x(k))$
- $f(k) = -D_h^T(k)R^{-1}e(k)$
- $D_h(x(k)) \in R^{m \times n}$ is the Jacobian of h(x) evaluated at x(k)
- $f(k)$ is normalized forecast error viewed from model space
- $\delta x(k)$ is the first variation of x(k) induced by the variation $\delta x(0)$ in x(0) and δa in a propagated by the model dynamics

Tangent linear model - propagation of perturbations

- Model: $x(k+1) = M(x(k), a)$
- $\delta x(0)$ and δa are given initial perturbations
- $\delta x(k+1) = A(k)\delta x(k) + B(k)\delta a$
- $A(k) = D_M(x(k)) \in R^{n\times n}$ is the Jacobian of M w.r.to the state x
- $B(k) = D_M(a) \in R^{n\times p}$ is the the Jacobian of M w.r.to parameter a
- The above recurrence relation for $\delta x(k)$ can be solved
- $\delta x(k) = A(k-1:0)\delta x(0) + \sum_{j=0}^{k-1} A(k-1:j+1)B(j)\delta a$
- $A(i:j) = A(i)A(i+1)...A(j)$ is the product of the Jacobian along the trajectory from time i to j when $i \leq j$

Two properties of inner products

- Adjoint: $\langle x, Ay \rangle = \left\langle A^T x, y \right\rangle$
- Summation: : $\langle x, Az \rangle + \langle y, Az \rangle = \langle x + y, Az \rangle$
- Linear operators in finite dimensional vector spaces are represented by matrices. For these operators, transpose is the adjoint

Gradient of J w.r.to IC X(0)

- Substituting $\delta x(k)$ in the expression for δJ and simplifying:
- $\delta J = \left\langle \sum_{k=0}^{N} A^T(k-1:0)f(k), \delta x(0) \right\rangle$

  $+ \left\langle \sum_{k=0}^{N} \sum_{j=0}^{k-1} B^T(j)A^T(k-1:j+1)f(k), \delta a \right\rangle$
- From first principles:
- $\delta J = \left\langle \nabla_{x(0)} J, \delta x(0) \right\rangle + \langle \nabla_a J, \delta a \rangle$
- Comparing term by term we get the expressions for the adjoint sensitivities
- $\nabla_{x(0)} J = \sum_{k=1}^{N} A^T(k-1:0)f(k)$

Adjoint dynamics - initial conditon

- Computation of the adjoint sensitivity using the above summation is very inefficient
- An efficient method is to compute it recursively using the adjoint dynamics
- $\lambda(k) = A^T(k)\lambda(k+1) + f(k)$
- Final condition: $\lambda(N) = f(N)$
- Iterating backward from N to 0, we get
- Adjoint sensitivity: $\lambda(0) = \nabla_{x(0)} J$

- Complexity is $O[N(n^2 + n)]$

Gradient of J w.r.to a

- Recall $\nabla_a J = \sum_{k=0}^{N} \sum_{j=0}^{k-1} B^T(j) A^T(k-1:j+1) f(k)$
- This expression, since it involves double summation, takes longer time to compute
- we can likewise rewrite it as a recursive computation using a doubly nested loop
- Complexity is $O[N^2(n^2+n)+N(n^2)]$
- Computation of the adjoint sensitivity w.r.to parameters is considerably more expensive compared to that w.r.to the initial condition

Adjoint dynamics - parameters

- for $k = 1 : N$
- $\quad \lambda(N) = f(N)$
- $\quad$ for j = N-1:k
- $\quad\quad \lambda(j) = A^T(j)\lambda(j+1) + f(j)$
- $\quad$ end - j loop
- $\overline{\lambda}(k) = B(k-1)\lambda(k)$
- $sum = 0.0$
- $\quad$ for i = 1:N
- $\quad$ sum = sum + $\overline{\lambda}(i)$
- $\quad$ end - i loop
- end - k loop

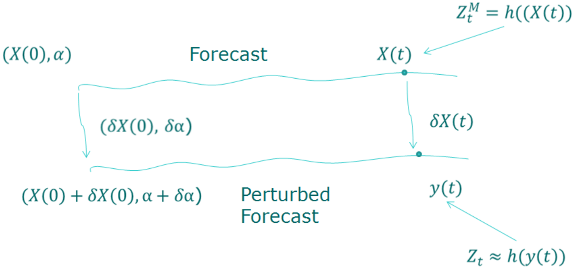
Reference on adjoint methods

- E. Kalnay (2003) Atmospheric Modeling, Data Assimilation, and Predictability, Cambridge University Press, 341 pages
- J. Lewis, S. Lakshmivarahan and S. K. Dhall (2006) Dynamic Data Assimilation: a least squares approach, Cambridge University Press, 654 pages

Forward Sensitivity Method - a summary

- Recall that error $e(k) = z(k) - h(x(k)) = [h(x^*(k)) - h(x(k))] + v(k)$
- This is contingent on the IC, $x(0)$ and parameter a
- Goal is to find corrections $\delta x(0)$ and δa :
- The solution from $(x(0) + \delta x(0))$ and $(a + \delta a)$ will annihilate the error $e(k)$

A pictorial view-single observation



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

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

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

First variation in h(x)

- $\delta h(x(k)) = D_h(x)\delta x(k)$
- $D_h(x) \in R^{m \times n}$ is the Jacobian of h(x)
- Next step is to compute $\delta x(k)$

Computation of $\delta x(k)$

- Let $\delta x(0)$ and δa be the perturbations in x(0) and a
- Let $\delta x(k)$ be the induced perturbation in x(k)
- From the Model: $x(k+1) = M(x(k), a)$
- we get $\delta x(k) = V(k)\delta x(0) + U(k)\delta a$
- $V(k)$ is the (forward) sensitivity of the solution x(k) w.r.to x(0)
- $V(k) = [V_{ij}(k)] = [\frac{\partial x_i(k)}{\partial x_j(0)}] \in R^{n \times n}$
- Likewise U(k) is the (forward) sensitivity of x(k) w.r.to the parameters a
- $U(k) = [U_{ij}(k)] = [\frac{\partial x_i(k)}{\partial a_j}] \in R^{n \times p}$

Basic Equation in FSM

- Substituting$\delta x(k)$ in δh, we get
- $e(k) = \delta h = D_h[V\delta x(0) + U\delta a]$
- $e(k) = [H_1, H_2][\delta x(0), \delta a]^T$
- where $H_1 = D_h V \in R^{m\times n}$, $H_2 = D_h U \in R^{m\times p}$,
- Define $H = [H_1, H_2] \in R^{m\times(n+p)}$
- We get the standard linear least squares problem:
- $e(k) = H\zeta$ where $\zeta = [\delta x(0), \delta a]^T$

Multiple Observations

- Let there be N observations z(k) for k = 1 to n
- Let $e = (e^T(1), e^T(2), ...e^T(N))^T \in R^{Nm}$ be the collection of all the errors
- Let $H = [H_1, H_2, ...H_N] \in R^{Nm \times (n+p)}$ matrix products of Jacobian of h (x) and the forward senstivities
- Solve the resulting linear least squares problem:
- $H\zeta = e$

Evolution of forward sensitivities w.r.to x(0)

- Model: $x(k+1) = M(x(k), a)$
- Differentiate this model equation w.r.to x(0):
- Forward Sensitivity dynamics w.r.to $x(0)$ in component form:
- $\frac{\partial x_i(k+1)}{\partial x_j(0)} = \sum \frac{\partial M_i(k)}{\partial x_q(k)} \frac{\partial x_q(k)}{\partial x_j(0)}$
- Thus we get $V(k+1) = A(k)V(k)$ with I.C. $V(0) = I$
- Iterating $V(k) = A(k-1)A(k-2)...A(1)A(0)$
- Denote $V(k) = A(k-1:0)$

Evolution of forward sensitivities w.r.to a

- Model: $x(k+1) = M(x(k), a)$
- Differenciate the above model equations w.r.to a in component form:
- Forward Sensitivity dynamics w.r.to a:
- $\frac{\partial x_i(k)}{\partial a_j} = \sum \frac{\partial M_i(k)}{\partial x_q} \frac{\partial x_q(k)}{\partial a_j} + \frac{\partial M_i}{\partial a_j}$
- We get $U(k+1) = A(k)U(k) + B(k)$ with $U(0) = 0$
- Iterating: $U(k) = \sum_{j=0}^{k-1} A(k-1+j)B(j)$

Reference on FSM

- S. Lakshmivarahan and J. Lewis (2010) Forward sensitivity approach to dynamic data assimilation"' Advances in Meteorology, doi: 10.1155/2010/375615
- S. Lakshmivarahan and J. Lewis (2011) Impact of observation on adjoint sensitivity, J. Meteor. Soc. Japan (under review)

Fine Structure of adjoint sensitivity w.r.to $x(0)$

- $\nabla_{x(0)} J = \sum_{k=1}^{N} A^T(k-1:0) f(k)$
- Using the defintion of V(k):
- $$\nabla_{x(0)} J = \sum_{k=1}^{N} V^T(k) f(k)$$
- In words:
- adjoint sensitivity w.r.to I.C. is the sum of the products of the transpose of the forward sensitivity w.r.to the IC, $V(k)$ and the normalized forecast error, $f(k) = -D_h^T(k) R^{-1} e(k)$ viewed from the model space where $e(k) = z(k) - h(x(k))$

Fine Structure of adjoint sensitivity w.r.to a

- $\nabla_a J = \sum_{k=0}^{N} [\sum_{j=0}^{k-1} B^T(j) A^T(k-1:j+1)] f(k)$
- Using the defition of U(k): $\nabla_a J = \sum_{k=0}^{N} U^T(k) f(k)$
- In words:
- adjoint sensitivity w.r.to parameters is the sum of the products of the transpose of the forward sensitivity w.r.to the IC, $U(k)$ and the normalized forecast error, $f(k) = -D_h^T(k) R^{-1} e(k)$ viewed from the model space where $e(k) = z(k) - h(x(k))$

Adjoint vs. forward sensitivity

- Recall adjoint sensitivity is computed using one forward run of the model, compute the normalized forecast errors $f(k)$ and using it as the forcing, run the backward adjoint model which is always linear
- In FSM on the otherhand, the forward sensitivity $V(k)$ is also also computed using the model and the sinsitivity equations both forward. There is no need to compute the adjoint model.
- The adjoint model related the evolution of the Lagrangian multiplier which is a vector, but the forward sensitivity is a matrix which has a system of n-vector equations to be solved simultaneously.
- When done serially, FSM will require more time, but $V(k)$ computation can be readily implemented in parallel - it is embarassingly parallel
- Despite its computational demands, it provides very useful information on the impact of observations on adjoint sensitivity

Impact of observations

- Despite its computational demands, FSM provides very useful information using which we can in addition also use it to evaluate the impact of the observation.
- For example if the qth column of $V(k)$ is very small, that is, when the qth compnent $x_q(k)$ is insenitive to the elements of $x_j(0)$ for all j= 1 to n, then the qth row of $V(k)$ is zero and the qth element of the product $V^T(k)f(k)$ will have its qth component also very close to zero.
- That is, the observation at time k does not materially affect the adjoint sensitivity
- A similar commnets apply to the adjoint sensitivity w.r.to the parameter a
- Here in lies the advantage FSM based approach. By running the model, FSM for $V(k)$ and $U(k)$ off-line in parallel, we should be able to assess the impact of the location of the observation on the adjoint sensitivity

Example

- Scalar dynamics: $\frac{\partial x}{\partial t} = C_T(\theta - x)$
- Scalar dynamics: $x(k+1) = (1-c)x(k) + c\theta$
- $x(0), c, \theta$ are three control variables
- Solution: $x(k) = (1-c)^k(x(0) - \theta) + \theta$
- Sensitivities of the solution:
- $\frac{\partial x(k)}{\partial x(0)} = (1-c)^k$ - exponentially decreases to zero with k
- $\frac{\partial x(k)}{\partial \theta} = [1 - (1-c)^k]$ - exponentially increases to 1 with k
- $\frac{\partial x(k)}{\partial c} = -k(1-c)^{k-1}[x(0) - \theta]$ - decreases to a minimum and then increases to zero

Solution and its sensitivities

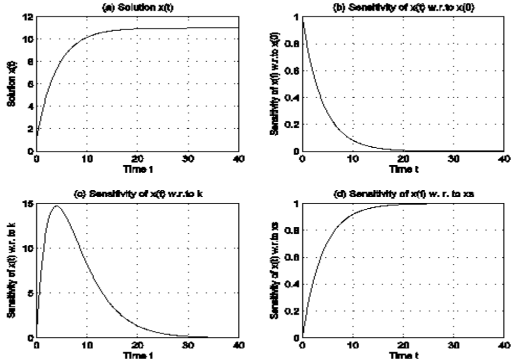


Figure 1. Evolution of the solution and its sensitivities

Solution saturates, two sensitivities go to zero and the third one tends to θ

Fine structure of adjoint sensitivity - Well placed observations

- Example 1 : Generate observation from x(0) = 1, $\theta = 11$, c=0.25
- Observations are chosen at times k =1, 2, 17, 18
- Forecast started from $x(0) = 2.0, \theta = 10, c = 0.30$
- $\nabla_{x(0)} J = 0.971, \nabla_{\theta} J = -1.267$ and $\nabla_c J = 1.436$
- Notice that the gradients are of the correct sign

Fine structure of adjoint sensitivity - Ill placed observations

- Example 2: Generate observation from x(0) = 1, θ = 11, c=0.25
- Observations are chosen at times k =15, 16, 17, 18
- Forecast started from $x(0) = 2.0, \theta = 10, c = 0.30$
- $\nabla_{x(0)} J = -0.011, \nabla_\theta J = -3.721$ and $\nabla_c J = -0.203$
- Notice that the gradient of w.r.to θ dominates and the other gradients are small
- The small values of the two gradients is a direct consequence of the relative insensitivity of the solution w.r.t. x(0) and c when the solution reaches a saturation stage
- Also the sign of the gradients $\nabla_{x(0)} J$ and $\nabla_c J$ are inconsistent
- Smallness of the gradients further indicate that inherent flatness of the J function the the θ and c directions

Conclusions

- Forward sensitivities of the solution can be used in two ways
- First, it can be used to isolate the effective from the ineffective temporal regions for placing the observations in order that they have a good impact on the adjoint gradient with respect to the chosen forecast aspect
- Second, once the sensitivities are made available they can be readily used to compute the adjoint sensitivity for use in an optimization procedure
- FSM requires more computations compared to the calssical adjoint method, but it is amenable to embarassingly parallel computation. This extra cost is well worth the effort since it helps to identify the "best" rigions to place the observations. It is a great off-line tool to do diagnostic studies

Other applocations of FSM

- E. Tromble, S. Lakshmivarahan, R. Kollar and K.M. Drescback (2011) "Application of Forward Sensitivity Mtehod to a GWCE-based Shallow Water Model", Computational Geosciences, (Under Review)
- D. Phan, S. Lakshmivarahan. E. Weng and Y. Luo (2011) "Forecast Error Correction in Ecosystem Carbon Pool Model using Forward Sensitivity Method", Under Preparation