Data assimilation.
Advanced Methods (1)

Olivier Talagrand
ESA summer school
Earth System Monitoring & Modelling
ESRIN, Frascati, Italy
1 August 2006

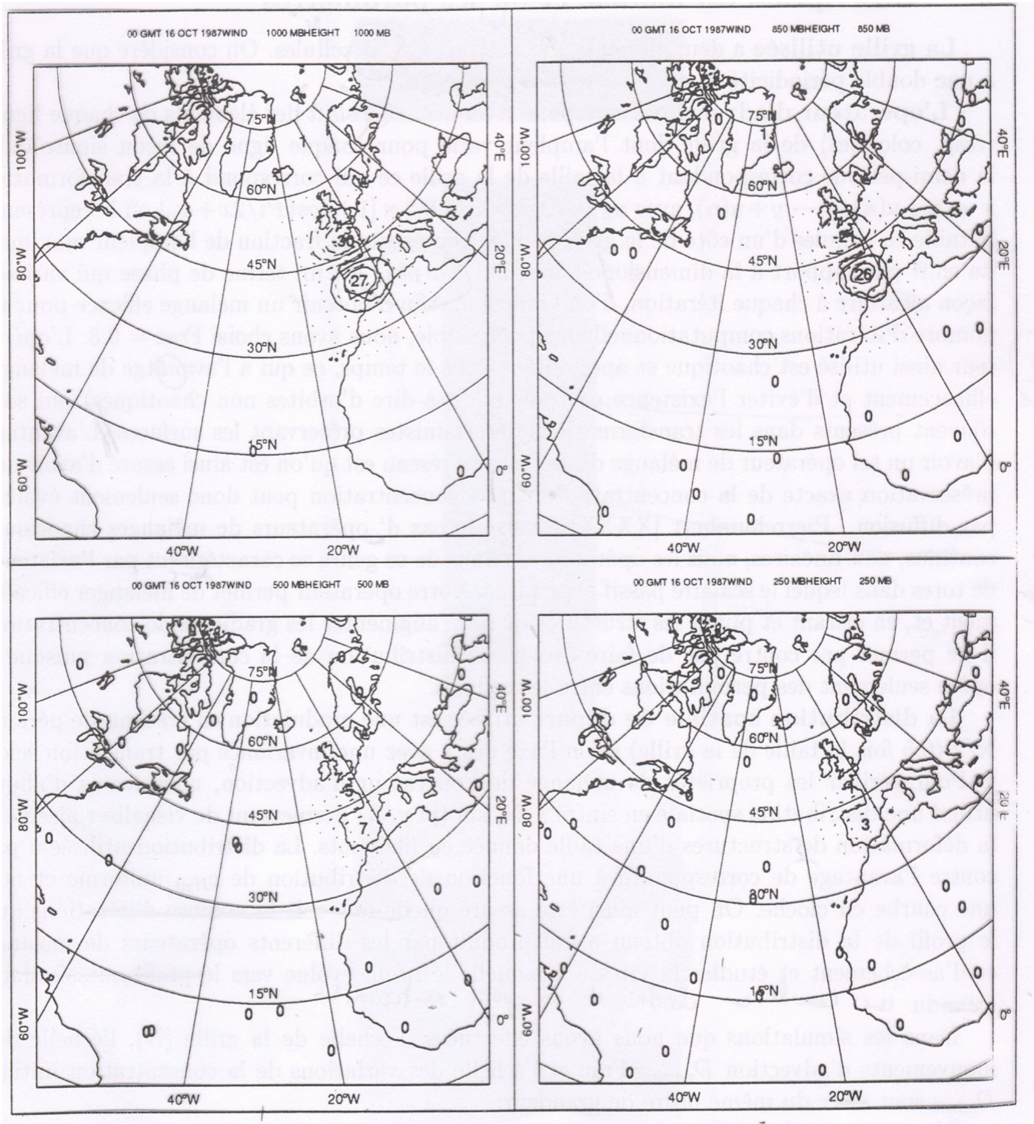
$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b \boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$$

$\boldsymbol{m} \equiv [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$ is a p-vector, with components (m_i)

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \Sigma_i \, m_i \boldsymbol{w}_i$$

where $\boldsymbol{w}_i$ is the i-th column vector of matrix $\boldsymbol{P}^b \boldsymbol{H}^T$.

$\boldsymbol{w}_i$, which belongs to state space is the *representer* associated with i-th observation y_i. It is entirely determined by the i-th row of $\boldsymbol{H}$ (*i*. *e*. the link between y_i and state vector $\boldsymbol{x}$) and background error covariance matrix $\boldsymbol{P}^b$. It represents the impact of y_i on analysis.



Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa presssure level (no temporal evolution of background error covariance matrix)

Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414

Best Linear Unbiased Estimate

Available data consist of

Background $\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b$

'Observations' $\boldsymbol{y} = \boldsymbol{H}x + \boldsymbol{\varepsilon}$

Errors assumed to be unbiased, $E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{b\mathrm{T}}) = \boldsymbol{P}^b$, $E(\boldsymbol{\varepsilon\varepsilon}^{\mathrm{T}}) = \boldsymbol{R}$, $E(\boldsymbol{\zeta}^b\boldsymbol{\varepsilon}^{\mathrm{T}}) = 0$

BLUE

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{Hx}^b)$$

$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} \boldsymbol{HP}^b$$

Equivalent set of formulæ

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^a \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{Hx}^b)$$

$$[\boldsymbol{P}^a]^{-1} = [\boldsymbol{P}^b]^{-1} + \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1}\boldsymbol{H}$$

Best Linear Unbiased Estimate (continuation 1)

$\boldsymbol{H}$ can be any linear operator

Example : (scalar) satellite observation

$\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ temperature profile

Observation $y = \Sigma_i h_i x_i + \varepsilon = \boldsymbol{Hx} + \varepsilon$, $\boldsymbol{H} = (h_1, \ldots, h_n)$, $E(\varepsilon^2) = r$

Background $\boldsymbol{x}^b = (x_1{}^b, \ldots, x_n{}^b)^{\mathrm{T}}$, error covariance matrix $\boldsymbol{P}^b = (p_{ij}{}^b)$

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{Hx}^b)$$

$[\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{Hx}^b) = (y - \Sigma_\iota h_\iota x_\iota{}^b) / (\Sigma_{ij} h_i h_j p_{ij}{}^b + r)^{-1} \equiv \mu$ scalar !

- $\boldsymbol{P}^b = p^b \boldsymbol{I}_n$ $x_i^a = x_i^b + p^b h_i \mu$
- $\boldsymbol{P}^b = \mathrm{diag}(p_i{}^b)$ $x_i^a = x_i^b + p_i^b h_i \mu$
- General case $x_i^a = x_i^b + \Sigma_j p_{ij}{}^b h_j \mu$

  Each level i is corrected, not only because of its own contribution to the observation, but because of the contribution of the other levels to which its background error is correlated.

Best Linear Unbiased Estimate (continuation 2)

Variational form of the *BLUE*

BLUE x^a minimizes following scalar *objective function*, defined on state space

$$\xi \in \mathcal{S} \rightarrow$$

$$\begin{aligned} \mathcal{J}(\xi) &= (1/2)\,(\boldsymbol{x}^b - \xi)^{\mathrm{T}}\,[\boldsymbol{P}^b]^{-1}\,(\boldsymbol{x}^b - \xi) + (1/2)\,(\boldsymbol{y} - \boldsymbol{H}\xi)^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,(\boldsymbol{y} - \boldsymbol{H}\xi) \\ &= \mathcal{J}_b + \mathcal{J}_o \end{aligned}$$

'*3D-Var*'

Can easily, and heuristically, be extended to the case of a nonlinear observation operator H.

Used operationally in USA, Australia, China, …

Question. How to introduce temporal dimension in estimation process ?

- Logic of Optimal Interpolation can be extended to time dimension.

- But we know much more than just temporal correlations. We know explicit dynamics.

Real (unknown) state vector at time k (in format of assimilating model) x_k. Belongs to state space $\mathcal{S}$ ($\dim \mathcal{S} = n$)

Evolution equation

$$\boldsymbol{x}_{k+1} = \boldsymbol{M}_k(\boldsymbol{x}_k) + \boldsymbol{\eta}_k$$

$\boldsymbol{M}_k$ is (known) model, $\boldsymbol{\eta}_k$ is (unknown) model error

Sequential Assimilation

- Assimilating model is integrated over period of time over which observations are available. Whenever model time reaches an instant at which observations are available, state predicted by the model is updated with new observations.

Variational Assimilation

- Assimilating model is globally adjusted to observations distributed over observation period. Achieved by minimization of an appropriate scalar *objective function* measuring misfit between data and sequence of model states to be estimated.

- Observation vector at time k

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\varepsilon}_k \qquad k = 0, \ldots, K$$

$$E(\boldsymbol{\varepsilon}_k) = 0 \quad ; \quad E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = \mathbf{R}_k \, \delta_{kj}$$

$\mathbf{H}_k$ linear

- Evolution equation

$$\mathbf{x}_{k+1} = \mathbf{M}_k \mathbf{x}_k + \boldsymbol{\eta}_k \qquad k = 0, \ldots, K\text{-}1$$

$$E(\boldsymbol{\eta}_k) = 0 \quad ; \quad E(\boldsymbol{\eta}_k \boldsymbol{\eta}_j^{\mathrm{T}}) = \mathbf{Q}_k \, \delta_{kj}$$

$\mathbf{M}_k$ linear

- $E(\boldsymbol{\eta}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = 0$ (errors uncorrelated in time)

At time k, background $\boldsymbol{x}^b{}_k$ and associated error covariance matrix $\boldsymbol{P}^b{}_k$ known

- Analysis step

$$\boldsymbol{x}^a{}_k = \boldsymbol{x}^b{}_k + \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}^{\mathrm{T}} [\boldsymbol{H}_k \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} (\boldsymbol{y}_k - \boldsymbol{H}_k \boldsymbol{x}^b{}_k)$$
$$\boldsymbol{P}^a{}_k = \boldsymbol{P}^b{}_k - \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}^{\mathrm{T}} [\boldsymbol{H}_k \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \boldsymbol{H}_k \boldsymbol{P}^b{}_k$$

- Forecast step

$$\boldsymbol{x}^b{}_{k+1} = \boldsymbol{M}_k \boldsymbol{x}^a{}_k$$
$$\begin{aligned}\boldsymbol{P}^b{}_{k+1} &= E[(\boldsymbol{x}^b{}_{k+1} - \boldsymbol{x}_{k+1})(\boldsymbol{x}^b{}_{k+1} - \boldsymbol{x}_{k+1})^{\mathrm{T}}] = E[(\boldsymbol{M}_k \boldsymbol{x}^a{}_k - \boldsymbol{M}_k \boldsymbol{x}_k - \boldsymbol{\eta}_k)(\boldsymbol{M}_k \boldsymbol{x}^a{}_k - \boldsymbol{M}_k \boldsymbol{x}_k - \boldsymbol{\eta}_k)^{\mathrm{T}}] \\ &= \boldsymbol{M}_k E[(\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)(\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)^{\mathrm{T}}]\boldsymbol{M}_k{}^{\mathrm{T}} - E[\boldsymbol{\eta}_k (\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)^{\mathrm{T}}] - E[(\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)\boldsymbol{\eta}_k{}^{\mathrm{T}}] + E[\boldsymbol{\eta}_k \boldsymbol{\eta}_k{}^{\mathrm{T}}] \\ &= \boldsymbol{M}_k \boldsymbol{P}^a{}_k \boldsymbol{M}_k{}^{\mathrm{T}} + \boldsymbol{Q}_k\end{aligned}$$

To sum up

- Analysis step

$$\boldsymbol{x}^a_k = \boldsymbol{x}^b_k + \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} [\boldsymbol{H}_k \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} (\boldsymbol{y}_k - \boldsymbol{H}_k \boldsymbol{x}^b_k)$$
$$\boldsymbol{P}^a_k = \boldsymbol{P}^b_k - \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} [\boldsymbol{H}_k \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \boldsymbol{H}_k \boldsymbol{P}^b_k$$

- Forecast step

$$\boldsymbol{x}^b_{k+1} = \boldsymbol{M}_k \boldsymbol{x}^a_k$$
$$\boldsymbol{P}^b_{k+1} = \boldsymbol{M}_k \boldsymbol{P}^a_k \boldsymbol{M}_k^{\mathrm{T}} + \boldsymbol{Q}_k$$

Kalman filter (KF, Kalman, 1960)

Must be started from some initial estimate $(\boldsymbol{x}^b_0, \boldsymbol{P}^b_0)$

If all operators are linear, and if errors are uncorrelated in time, Kalman filter produces at time k the *BLUE* $\boldsymbol{x}^b{}_k$ (resp. $\boldsymbol{x}^a{}_k$) of the real state $\boldsymbol{x}_k$ from all data prior to (resp. up to) time k, plus the associated estimation error covariance matrix $\boldsymbol{P}^b{}_k$ (resp. $\boldsymbol{P}^a{}_k$).

If in addition errors are gaussian, the corresponding conditional probability distributions are the respective gaussian distributions

$\mathcal{N}[\boldsymbol{x}^b{}_k, \boldsymbol{P}^b{}_k]$ and $\mathcal{N}[\boldsymbol{x}^a{}_k, \boldsymbol{P}^a{}_k]$.

Nonlinearities ?

Model is usually nonlinear, and observation operators (satellite observations) tend more and more to be nonlinear.

- Analysis step

$$\boldsymbol{x}^a_k = \boldsymbol{x}^b_k + \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} [\boldsymbol{H}_k' \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} [\boldsymbol{y}_k - \boldsymbol{H}_k(\boldsymbol{x}^b_k)]$$
$$\boldsymbol{P}^a_k = \boldsymbol{P}^b_k - \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} [\boldsymbol{H}_k' \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \boldsymbol{H}_k' \boldsymbol{P}^b_k$$

- Forecast step

$$\boldsymbol{x}^b_{k+1} = \boldsymbol{M}_k(\boldsymbol{x}^a_k)$$
$$\boldsymbol{P}^b_{k+1} = \boldsymbol{M}_k' \boldsymbol{P}^a_k \boldsymbol{M}_k'^{\mathrm{T}} + \boldsymbol{Q}_k$$

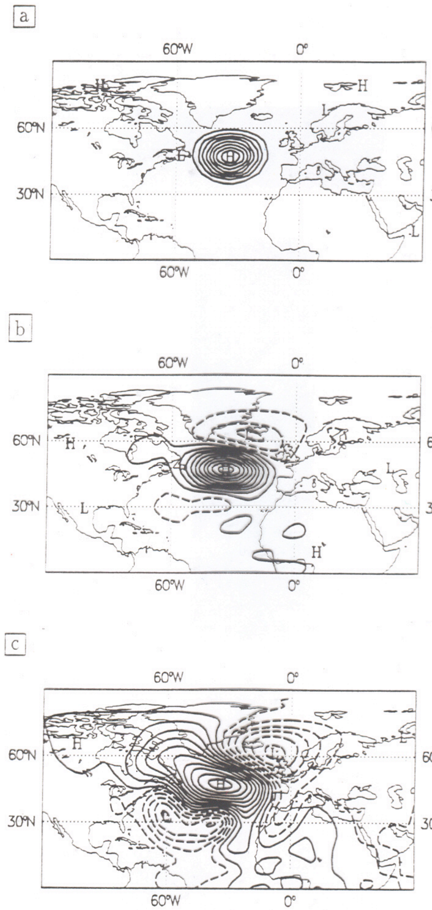
Extended Kalman Filter (EKF, heuristic !)

Costliest part of computation

$$\boldsymbol{P}^{b}{}_{k+1} = \boldsymbol{M}_{k}\boldsymbol{P}^{a}{}_{k}\boldsymbol{M}_{k}{}^{\mathrm{T}} + \boldsymbol{Q}_{k}$$

Multiplication by $\boldsymbol{M}_k$ = one integration of the model between times k and $k+1$.
Computation of $\boldsymbol{M}_{k}\boldsymbol{P}^{a}{}_{k}\boldsymbol{M}_{k}{}^{\mathrm{T}} \approx 2n$ integrations of the model

Need for determining the temporal evolution of the uncertainty on the state of the system is the major difficulty in assimilation of meteorological and oceanographical observations



Temporal evolution of the 500-hPa geopotential autocorrelation with respect to point located at 45N, 35W. From top to bottom: initial time, 6- and 24-hour range. Contour interval 0.1. After F. Bouttier.

Variational approach

Available data consist of

- Background estimate at time 0

  $\boldsymbol{x_0}^b = \boldsymbol{x}_0 + \zeta_0^b \qquad E(\zeta_0^b \zeta_0^{bT}) = \boldsymbol{P}_0^b$

- Observations at times $k = 0, \ldots, K$

  $\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{\varepsilon}_k \qquad E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = \boldsymbol{R}_k$

- Model (supposed now, and for the time being, to be exact)

  $\boldsymbol{x}_{k+1} = \boldsymbol{M}_k \boldsymbol{x}_k \qquad k = 0, \ldots, K\text{-}1$

Errors assumed to be unbiased and uncorrelated in time, $\boldsymbol{H}_k$ and $\boldsymbol{M}_k$ linear

Then objective function

$\xi_0 \in \mathcal{S} \rightarrow$

$$\mathcal{J}(\xi_0) = (1/2)\,(\boldsymbol{x}_0^b - \xi_0)^{\mathrm{T}}\,[\boldsymbol{P}_0^b]^{-1}\,(\boldsymbol{x}_0^b - \xi_0) + (1/2)\,\Sigma_k[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]^{\mathrm{T}}\,\boldsymbol{R}_k^{-1}\,[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]$$

subject to $\xi_{k+1} = \boldsymbol{M}_k \xi_k, \qquad k = 0, \ldots, K\text{-}1$

$$\mathcal{J}(\xi_0) = (1/2)\,(\boldsymbol{x}_0^{\,b} - \xi_0)^{\mathrm{T}}\,[\boldsymbol{P}_0^{\,b}]^{-1}\,(\boldsymbol{x}_0^{\,b} - \xi_0) + (1/2)\,\Sigma_k[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]^{\mathrm{T}}\,\boldsymbol{R}_k^{-1}\,[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]$$

Background is not necessary, if observations are in sufficient number to overdetermine the problem. Nor is strict linearity.

How to minimize objective function with respect to initial state $\boldsymbol{u} = \xi_0$ ($\boldsymbol{u}$ is called the *control variable* of the problem) ?

Use iterative minimization algorithm, each step of which requires the explicit knowledge of the local gradient $\nabla_{\boldsymbol{u}}\mathcal{J} \equiv (\partial\mathcal{J}/\partial u_i)$ of $\mathcal{J}$ with respect to u.

Gradient computed by *adjoint method*.

How to numerically compute the gradient $\nabla_u \mathcal{J}$?

Direct perturbation, in order to obtain partial derivatives $\partial\mathcal{J}/\partial u_i$ by finite differences ? That would require as many explicit computations of the objective function $\mathcal{J}$ as there are components in u. Practically impossible.

Adjoint Method

Input vector $\boldsymbol{u} = (u_i)$, $\dim \boldsymbol{u} = n$

Numerical process, implemented on computer (*e. g.* integration of numerical model)

$$\boldsymbol{u} \rightarrow \boldsymbol{v} = \boldsymbol{G}(\boldsymbol{u})$$

$\boldsymbol{v} = (v_j)$ is *output vector* , $\dim \boldsymbol{v} = m$

Perturbation $\delta \boldsymbol{u} = (\delta u_i)$ of input. Resulting first-order perturbation on $\boldsymbol{v}$

$$\delta v_j = \Sigma_i \, (\partial v_j / \partial u_i) \; \delta u_i$$

or, in matrix form

$$\delta \boldsymbol{v} \;=\; \boldsymbol{G}' \, \delta \boldsymbol{u}$$

where $\boldsymbol{G}' \equiv (\partial v_j / \partial u_i)$ is local matrix of partial derivatives, or jacobian matrix, of $\boldsymbol{G}$.

Adjoint Method (continued 1)

$$\delta \boldsymbol{v} \;=\; \boldsymbol{G}' \,\delta \boldsymbol{u} \qquad \text{(D)}$$

Scalar function of output

$$\mathcal{J}(\boldsymbol{v}) \;=\; \mathcal{J}[\boldsymbol{G}(\boldsymbol{u})]$$

Gradient $\nabla_{\boldsymbol{u}} \mathcal{J}$ of $\mathcal{J}$ with respect to input $\boldsymbol{u}$?

'Chain rule'

$$\partial \mathcal{J}/\partial u_i = \Sigma_j \, \partial \mathcal{J}/\partial v_j \,(\partial v_j/\partial u_i)$$

or

$$\nabla_{\boldsymbol{u}} \mathcal{J} \;=\; \boldsymbol{G}'^{\mathrm{T}} \, \nabla_{\boldsymbol{v}} \mathcal{J} \qquad \text{(A)}$$

Adjoint Method (continued 2)

$\boldsymbol{G}$ is the composition of a number of successive steps

$$\boldsymbol{G} = \boldsymbol{G}_N \circ \ldots \circ \boldsymbol{G}_2 \circ \boldsymbol{G}_1$$

'Chain rule'

$$\boldsymbol{G}' = \boldsymbol{G}_N' \ldots \boldsymbol{G}_2' \boldsymbol{G}_1'$$

Transpose

$$\boldsymbol{G}'^{\mathrm{T}} = \boldsymbol{G}_1'^{\mathrm{T}} \boldsymbol{G}_2'^{\mathrm{T}} \ldots \boldsymbol{G}_N'^{\mathrm{T}}$$

Transpose, or *adjoint*, computations are performed in reversed order of direct computations.

If $\boldsymbol{G}$ is nonlinear, local jacobian $\boldsymbol{G}'$ depends on local value of input $\boldsymbol{u}$. Any quantity which is an argument of a nonlinear operation in the direct computation will be used gain in the adjoint computation. It must be kept in memory from the direct computation (or else be recomputed again in the course of the adjoint computation).

If everything is kept in memory, total operation count of adjoint computation is at most 4 times operation count of direct computation (in practice about 2).