A few reminiscences about the genesis of 4D-Var, and what followed

Olivier Talagrand
Laboratoire de Météorologie Dynamique, École Normale Supérieure, Paris, France

Symposium: *20 years of 4D-Var at ECMWF*

European Centre for Medium-Range Weather Forecasts
26 January 2018

Around 1979-80, staff member in Assimilation Section at ECMWF

At that time, assimilation was based on *Optimal Interpolation* (poor man's Kalman Filter, A. Lorenc)

Question. *How to propagate information in assimilation, not only from the past into the future, but also from the future into the past (did not know about Kalman Smoother) ? Variational approach ?*

Two visitors from the USSR at ECMWF, G. R. Kontarev and V. N. Lykossov (coming from Academician G. I. Marchuk's group)

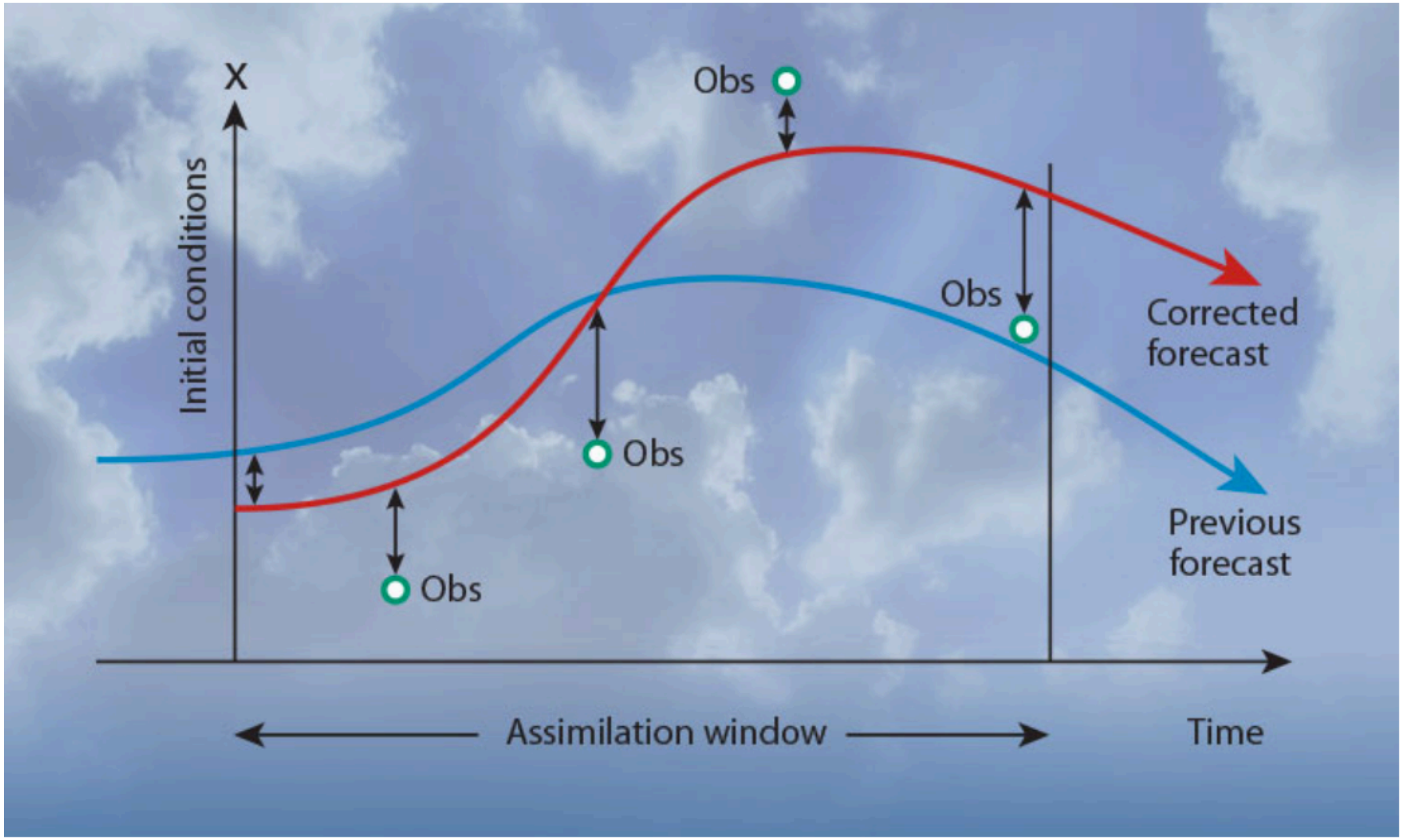
G. R. Kontarev gave a seminar on the Technique of Adjoint Equation (*The adjoint equation technique applied to meteorological problems*, ECMWF Technical Report 21)

Could it be used for assimilation ?

1981. Back in France

Met F.-X. Le Dimet. Had studied optimal control with J.-L. Lions, and was interested in applying optimization methods to meteorology and atmospheric sciences.

Gave me a number of documents, describing various algorithms for optimization. Among these, found one, based on adjoint equations, that seemed to give the solution to ‘my’ problem.


X
Initial conditions
Obs
Obs
Obs
Obs
Corrected
forecast
Previous
forecast
Assimilation window
Time

State vector $x(t)$

Evolves in time according to

$$\frac{dx}{dt} = F(x,t) \tag{1}$$

Small perturbation $\delta x(t_0)$ at time t_0. To first order, perturbation evolves in time according to *tangent linear equation*

$$\frac{d\delta x}{dt} = F'(x(t))\,\delta x$$

where $F'(x(t))$ is matrix of partial derivatives (*Jacobian*) of F wrt to x, taken at point $x(t)$

Scalar objective function $J[x(t)], \quad t_0 \leq t \leq t_1$

Through evolution equation (1), $J[x(t)]$ *is* a compound function of initial state $u \equiv x(t_0)$. $J[x(t)] = K(u)$

Direct dependence of $K(u)$ with respect to u ?

Integrate *adjoint equation* backwards in time from $\lambda(t_1) = 0$

$$\frac{d\lambda}{dt} = -F'^{*}\lambda - \nabla_{x(t)}J \qquad (2)$$

where F'^* is the adjoint of F', and $\nabla_{x(t)} J$ is the gradient of J wrt $x(t)$

Then $\lambda(t_0)$ is the gradient of the function K wrt u. $\lambda(t_0) = \nabla_u K$

Computation of gradient requires one forward integration of direct equation (1) followed by one backward integration of adjoint equation (2)

Can easily be extended to dependence of K with repect to, *e.g.*, parameters in F or lateral boundary conditions.

Particular case of much more general approach

$$u \rightarrow v \equiv G(u)$$

Scalar function $\quad J(v) = J[G(u)] \equiv K(u)$

Then

$$\nabla_u K = G'^* \nabla_v J$$

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

Idea was there, and looked all right. But real work had to be done !

I submitted the idea to (what was not yet called) Météo-France as the subject of a student project. Ph. Courtier, who had studied optimization, expressed his interest.

Collaboration with Institut national de recherche en informatique et en automatique (INRIA, C. Lemaréchal, J.-C. Gilbert)

In the mean time, I became aware of a number of facts

- Variational assimilation had already been studied and numerically implemented by Penenko, V. and Obraztsov, N. N., *Meteorologiya i Gidrologiya*, 1976.

  Did use adjoint approach, in an exactly linear situation (but with real, gridded, observations)

- The use of adjoint equations had already been suggested, for sensitivity studies in the context of atmospheric and oceanic dynamcis, by several authors

  Cacuci, D. G., *J. Math. Phys.*, 1981

  Other works with M. C. G. Hall, M. E. Schlesinger

Link with estimation theory

Minimization of an objective function of the form

$$\mathcal{J}(\boldsymbol{\xi}) \equiv (1/2)\ [\Gamma\boldsymbol{\xi} - z]^{\mathrm{T}}\ S^{-1}\ [\Gamma\boldsymbol{\xi} - z] \qquad (3)$$

for data of the form $z = \Gamma x + \zeta$

or, in a form that may be more familiar to a meteorological audience,

$$\mathcal{J}(\xi_0) = (1/2)\ (x_0^{\,b} - \xi_0)^{\mathrm{T}}\ [P_0^{\,b}]^{-1}\ (x_0^{\,b} - \xi_0) + (1/2)\ \Sigma_k[y_k - H_k\xi_k]^{\mathrm{T}}\ R_k^{-1}\ [y_k - H_k\xi_k]$$

defines, in the case of a linear observation operator Γ, the linear variance-minimizer (*Best Linear Unbiased Estimate*) of the unknown x from the data z. Variational assimilation is in this case exactly equivalent to *Kalman smoothing*, which is just another numerical algorithm for solving the same problem.

If in addition the data error ζ is Gaussian, mimation of (3) solves the problem of Bayesian estimation, in the sense that that minimizer is the conditional expectation $\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{y})$ of $\boldsymbol{x}$ given $\boldsymbol{y}$.

At about the same time, Hoffman, R. N., *Mon. Wea. Rev.*, 1986, performed variational assimilation, computing the gradient through finite differences. Everything was all right, but numerical cost was prohibitive.

All that led to papers

Lewis, J. M. and Derber. J . C., *Tellus A*, 1985

F.-X. Le Dimet and O. Talagrand, *Tellus A*, 1986

And then, Ph. Courtier, who used real observations, showed that variational assimilation based on the use of adjoint equations could be successfully implemented on a realistic meteorological model.

Philippe and I presented our results at a workshop at ECMWF in 1987 (?)

Successful transfer of technology

Other actors : J. Pailleux, F. Rabier, J.-N. Thépaut, F. Bouttier, …

The rest of the story of 4DVar left to the other speakers …

Adjoint approach has been used in many applications, some of which do not require minimizations (sensitivity studies), or are not even meant to compute the gradient of an explicitly defined function (computation of singular modes)

Development of adjoint codes

Meteorological applications were the first ones (I think) which confronted people with the problem of developing the adjoint of an already existing major code.

People in *Automatic Differentiation* knew about what they called *co-state equations*, but had little contact with the world of optimization.

Definition of strict rules for writing the adjoint of an existing code (beware of 'intuition')

Adjoint compilers

- FastOpt (R. Giering and colleagues)
- Tapenade (L. Hascoët and colleagues, INRIA)

Difficulties lie in the management of memory (what to keep in memory from the direct integration ?)

Physical significance of adjoint equations and adjoint computations ?

My answer. Adjoint equations compute purely numerical sensitivities. Possible physical significance in the results in the results of the adjoint computations, not in the process. Recipe : integrations by parts.

Until ....

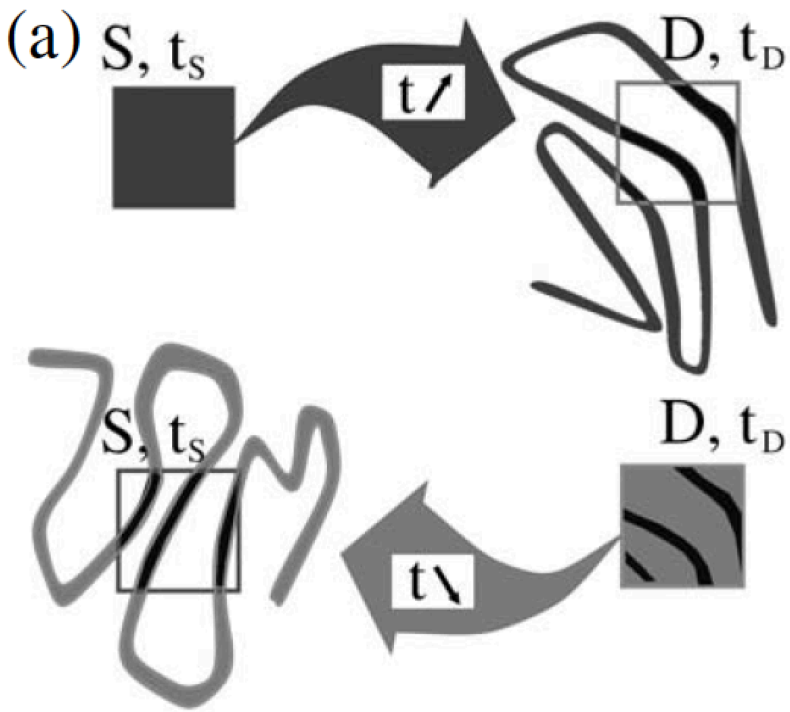
F. Hourdin (interested in verification of the Comprehensive Test Ban Treaty)

Radioactive tracer is observed. What is the guilty part ?

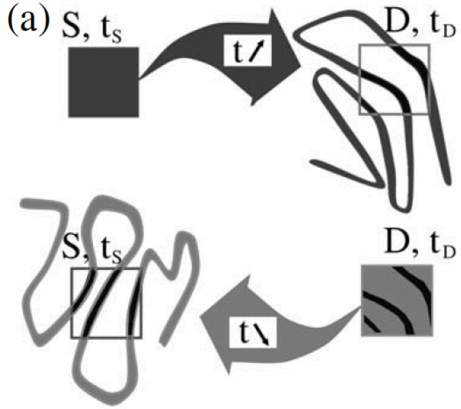
Transport equation

$$\frac{\partial c}{\partial t} + \mathbf{v.grad}\, c + \lambda c = \sigma$$

where σ is source of tracer.



Reciprocity of motion



Direct transport

$$\frac{\partial c}{\partial t} + \mathbf{v}.\mathbf{grad}\, c + \lambda c = \sigma$$

Retro-transport

$$-\frac{\partial c^*}{\partial t} - \mathbf{v}.\mathbf{grad}\, c^* + \lambda c^* = \mu$$

Extends to any form of (linear) dissipation of diffusion, or (linear) sources or sinks, as long as motion is reversible.

Parametrisation of subgrid scale motions

- if parametrisation is invariant in a change of sign of velocities, keep as it is
- if not, change signs of velocity (*e.g.*, in parametrisation of convection, downdraughts become updraughts, and vice-versa)

Retro-transport equation is adjoint of direct transport equation for the air-mass-weighted scalar product (mass of transporting air must be conserved):

$$\langle \phi, \psi \rangle = \int_{\Omega \times \tau} \rho \phi \psi \, \mathbf{dx} \, \mathrm{d}t$$

Adjoint equation has been obtained in this case on the basis of purely physical (actually kinematic) considerations, with no need for an explicit scalar product, nor for the machinery of integration by parts *et al*..

Thanks !