Lecture notes on assimilation algorithms

Elías Valur Hólm
European Centre for Medium-Range Weather Forecasts
Reading, UK

April 18, 2008

1 Basic concepts

1.1 The analysis

In meteorology and other branches of geophysics the process of approximating the true state of a physical system at a given time is called analysis. The information on which the analysis is based includes observational data and a model of the physical system, together with some background information on initial and boundary conditions and possibly additional constraints on the analysis. The analysis is useful in itself as a description of the physical system, but it can also be used for example as an initial state for studying the further time evolution of the system.

- *Example: Classical analysis methods.* Meteorologists were producing meteorological analyses by hand long before the advent of the computer (and some do it still today). The observations were all made roughly simultaneously (12UTC surface pressure reports from ships for example), and bore on the physical model variables like temperature, wind and pressure. These hand analyses actually contained far more than just a spatial interpolation of observations, and [6] and others formalized this approach for use in the early numerical weather prediction models. The approach is the following. First-guess fields defined at the gridpoints of the forecasting model are interpolated to the observation location. The difference between the observation and the interpolated value are then interpolated back onto the gridpoints to define a correction (which decreases with the distance from the observation). The important step here is to use a first-guess to guide the eye (or the computer) when interpreting observational information. These early works lead to *optimal interpolation* which in different guises still are widely used analysis methods (see [8] for a comprehensive discussion of classical analysis methods).

- *Example: Ozone analysis.* Total ozone is measured by remote sensing from space, and one days worth of data covers the Earth with some gaps. An analysis will fill in all gaps and produce a global map of total ozone. We can make a rather poor analysis by interpolating the observations in space, not least because the observations are up to 24 hours apart and the total ozone

field may have changed significantly at some locations in that time. A better analysis would take the dynamic evolution of ozone and the distribution of observations in time into account, and in that way produce a picture of total ozone which is consistent with the dynamic state of the whole atmosphere.

- *Example: Weather forecasts.* The analysis prepares the initial conditions for weather forecasts. In addition to the newly arrived observations, a background estimate of the state of the atmosphere is available as a forecast from the previous analysis. In the preparation of the analysis, a particular effort is usually made to damp gravity waves in the analysis by additional constraints.

A particular challenge in the forecasts of the time-evolution of geophysical systems is the nonlinearity of the system and the corresponding sensitivity to initial conditions. It is well known that even if we had a perfect forecast model, all forecasts start to diverge from the truth after a finite time. This is illustrated in Fig. 1 which shows how integrations of a nonlinear equation starting from slightly different initial states follow each other for some time and then start to diverge from each other. If we take one of the integrations to be the truth and interpret the difference in initial conditions as the analysis error, we can see that increasing the analysis error reduces the time period for which the forecast remains close to the truth.

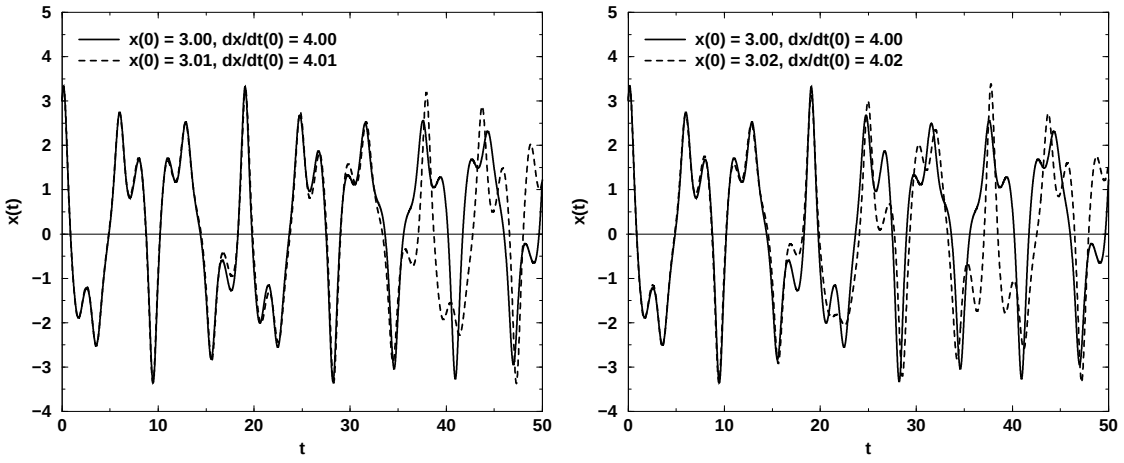

Figure 1: *Integrations of Duffing's equation,* $\ddot{x} + 0.05\dot{x} + x^3 = 7.5\cos t$*, from slightly different initial conditions. We take* $x(0) = 3$ *and* $\dot{x}(0) = 4$ *to be the true initial state and interpret differences in initial conditions as 'analysis errors'. In the left figure, with 'analysis errors' of ca.* 0.3%*, the 'forecast' is good until* $t \approx 35$*, whereas in the right figure with 'analysis errors' of ca.* 0.6% *the forecast starts to diverge already at* $t \approx 20$*. See [20] for this and other interesting nonlinear systems.*

1.2 The assimilation

As discussed above, an analysis can be very simple, for example a spatial interpolation of observations. However, much better results can be obtained by including the dynamic evolution of the physical system in the analysis. An analysis which combines time distributed observations and a dynamic model is called assimilation or data assimilation.

The assimilation problem can be discussed from many angles, depending on background and preferences (control theory, estimation theory, probability theory, variational analysis, etc.). A few excellent introductions to data assimilation from different points of view are given by [1, 2, 5, 13, 15, 16, 17, 18, 19]. A discrete formulation is the most common basis for developing the algorithms in numerical models. However, in this introduction I want to show how different assimilation algorithms can be derived from one common source, and what the approximations are which lead to each algorithm (OI, 3DVAR, 4DVAR and Kalman filters in particular). One attractive way to achieve this is to consider the continuous assimilation problem from the point of view of variational analysis. Following [1] and [2] a *generalized inverse* solution to the assimilation problem will be derived, which has many of the common assimilation algorithms as special cases. See [14] for an introduction to variational analysis.

It is illustrative to see how the assimilation problem differs from the familiar initial value problem. The assimilation problem can be seen as an initial value problem with some added features: the model equations for the *model state* $\vec{x}$ are approximate with model errors $\vec{\varepsilon}_m$; the initial condition is approximate with error $\vec{\varepsilon}_b$ (background error); and there are observations y_n present with observation errors $\varepsilon_{o,n}$. Expressed in mathematical form,

$$\frac{\partial \vec{x}}{\partial t} + \vec{M}(\vec{x}) = \vec{\varepsilon}_m \tag{1}$$

$$\vec{x}(0) = \vec{x}_b(0) + \vec{\varepsilon}_b \tag{2}$$

$$y_n = H_n(\vec{x}) + \varepsilon_{o,n} \quad n = 1, \ldots, N \tag{3}$$

In Eq. 3 we have introduced the *observation operator* H_n, which calculates the model equivalent of the observation y_n. Even if y_n measures an explicit model state variable (say temperature), the observation operator will still be needed to interpolate the model state, which always has finite resolution, to the position of the observation. When the measurement is indirect, like radiances measured by satellites from space, the observation operator can be a radiative transfer model which takes a whole model column and calculates the radiance emitted to space from Earth in the particular waveband measured. Observation operators are central to data assimilation, and much of the work in developing data assimilation algorithms is related to the observation operators.

Frequently $\vec{M}(\vec{x})$ is used to denote the total model operator, but here we keep the time derivative of the model state separate to better distinguish the dynamic evolution within the assimilation.

Most of the discussion could also have been made in terms of the discrete state vector $\mathbf{x}$, and it is good to keep this in mind and try to figure out discrete equivalents of some of the material we will go through.

1.3 Formulation in terms of probability

1.3.1 Probability density functions

The assimilation problem outlined above includes unknown errors, so it is obvious that the solution we aim at is statistical. Before we can go any further, we need to

know something about the statistics of the errors. Characterizing the errors of the particular model and observations of the system under study is another one of the cornerstones of data assimilation. In the best of worlds, we have a full description of the probability density function $P(\vec{\varepsilon}_m, \vec{\varepsilon}_b, \vec{\varepsilon}_o, t)$ and include the equation for the time-evolution of P in the formulation of our assimilation system,

$$\frac{\partial P(\vec{\varepsilon}, t)}{\partial t} + \frac{\partial}{\partial \vec{\varepsilon}} \left(\frac{\partial \vec{\varepsilon}}{\partial t} P(\vec{\varepsilon}, t) \right) = 0 \tag{4}$$

This is the Liouville equation, which in this case says that the total probability of all errors integrated over the phase space is conserved (= 1 by definition). Note how similar this is to a continuity equation, with the 'error' transporting 'probability'. In practical applications, we have a very limited knowledge of P, for example only the mean and covariances.

We can in most cases assume model, background, and observation errors to be independent. In that case the total probability density function (pdf) is a product of the component pdf's for the observations, $P_o(\vec{\varepsilon}_o, t)$, the background, $P_b(\vec{\varepsilon}_b, t)$, and the model, $P_m(\vec{\varepsilon}_m, t)$,

$$P = P_m P_b P_o = \exp(\log P_m + \log P_b + \log P_o) \tag{5}$$

It may seem obvious that the optimal analysis corresponds to the maximum of the pdf, that is, the maximum value of the exponent in Eq. 5. However, this is not always the case as the following example shows.

- *Example: Non-Gaussian P.* Let us consider a 'one-dimensional' pdf, say the conditional probability $P(x|y)$, that is the probability of finding the system at x given the observation y. A Gaussian and a non-Gaussian pdf are shown in Fig. 2. In the Gaussian case, the maximum likelihood point and the mean of x coincide, so choosing the maximum of the pdf as our analysis seems the best solution. In the particular non-Gaussian case shown here, the mean and the maximum likelihood estimate are different, and it is not obvious what the best analysis is. The lesson is that we need to study the actual pdf enough to be sure that a maximum likelihood solution really gives us the analysis we want.

1.3.2 The costfunction

The maximizing of the probability is often expressed in terms of minimizing the negative of the exponent of the pdf, which is called the *costfunction* J,

$$\min J = \min(-\log P_m - \log P_b - \log P_o) = \min(J_m + J_b + J_o) \tag{6}$$

where we have introduced the model, background and observation costfunctions. In passing we note that when additional constraints are added to the assimilation, these also appear as additional terms in the costfunction, for example J_c is frequently used to denote the costfunction for gravity wave constraints ('reduction of small scale noise in the meteorological fields').

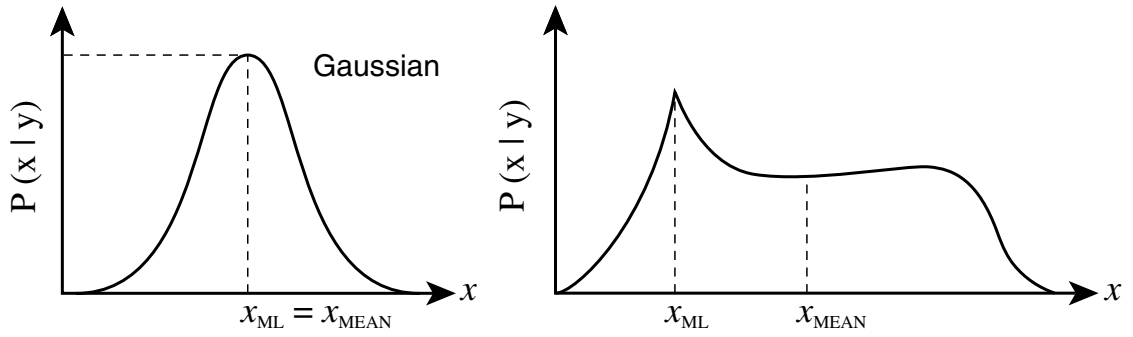

Figure 2: *Representativeness of maximum likelihood solutions for a Gaussian and a non-Gaussian pdf $P(x|y)$, that is the conditional probability of finding the model state at x, given an observation y. In the non-Gaussian case, the maximum likelihood point may be a poor representative of the best solution.*

- *Example: Quadratic costfunction.* A costfunction which appears frequently in meteorological and oceanographic applications is the following. Assume that the model errors can be neglected, and that the observation and background errors can be modelled by a Gaussian (also called normal) unbiased distribution. Using the definitions $\vec{\varepsilon}_b = \vec{x} - \vec{x}_b$ and $\vec{\varepsilon}_o = \vec{y} - \vec{H}(\vec{x})$ from the formulation of the assimilation problem, Eqs. 1– 3, we get

$$P_b = \frac{1}{\sqrt{2\pi|\vec{B}|}} e^{-\frac{1}{2}(\vec{x}(0)-\vec{x}_b(0))^T \vec{B}^{-1}(\vec{x}(0)-\vec{x}_b(0))} \tag{7}$$

$$P_o = \frac{1}{\sqrt{2\pi|\vec{R}|}} e^{-\frac{1}{2}(\vec{y}-\vec{H}(\vec{x}))^T \vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}))} \tag{8}$$

  where $\vec{B}$ and $\vec{R}$ are matrices of background and observation covariance functions, which reduce to the covariance matrices $\mathbf{B}$ and $\mathbf{R}$ in the discrete formulation, and $(\cdot)^T$ is the transpose of $(\cdot)$. The costfunction is then

$$J = \frac{1}{2}(\vec{x}(0)-\vec{x}_b(0))^T \vec{B}^{-1}(\vec{x}(0)-\vec{x}_b(0)) + \frac{1}{2}(\vec{y}-\vec{H}(\vec{x}))^T \vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x})) + C \tag{9}$$

  We do not need to consider the constant C further, since it has no effect on finding the minimum of the costfunction. There are some very nice consequences of having a Gaussian pdf. First, we only need to know the covariances of the errors instead of the whole pdf to completely characterize the errors (if the errors are biased, we also need the mean of the errors, which we can subtract before doing the analysis). Second, if the observation operators and the model are linear, the costfunction is quadratic in $\vec{x}$, and there is only one minimum, which simplifies the analysis. We will show how we can deal with nonlinearities in the next example. Third, there are very effective methods for minimizing quadratic problems. This is very important, since the main feature of meteorological and oceanographic assimilation problems is their large dimension ($\approx 10^7$).

- *Example: Nonlinear costfunctions.* In real applications we always have nonlinear model and very often the observation operators H_n are nonlinear. In the example above, Eq. 9, a nonlinear $\vec{H}(\vec{x})$ will make the costfunction nonlinear in $\vec{x}$. The nonlinearity of the model can also contribute to the nonlinearity of J in the following way. The observations used in the analysis are collected over a short time window (the *assimilation window*), usually 6–12 hours long. Instead of minimizing J with respect to $\vec{x}$ in the whole assimilation window, we only minimize it with respect to $\vec{x}(0)$, which greatly reduces the dimension of the problem. Since the model is perfect in this example, a forward model integration from $\vec{x}(0)$ gives $\vec{x}(t)$, and in this way the nonlinearity of the model enters $\vec{H}(\vec{x})$ when the observations are compared with the model at appropriate time (as in 4DVAR). Figure 3 shows the approach which we can follow in this case. First, linearize the costfunction around the background state $\vec{x}_b$ and minimize the quadratic costfunction J_i to get an estimate $\vec{x}_i$. Repeat the procedure by linearizing around $\vec{x}_i$, until the procedure has converged to required accuracy. This approach is widely used in assimilation applications. We see that in general there is no guarantee that we find the minimum of the costfunction. First of all, we need an accurate background estimate in order to end up in the right valley of the costfunction. Otherwise we may end up in a local minimum which is far away from the optimum solution. Second, the accuracy required of the background estimate really depends on the nonlinearity of the costfunction. In meteorological and oceanographic applications the costfunction is not more nonlinear than that forecasts from previous analysis are in most cases a good enough background to reach a good analysis (not always though, there is the odd bad forecast).

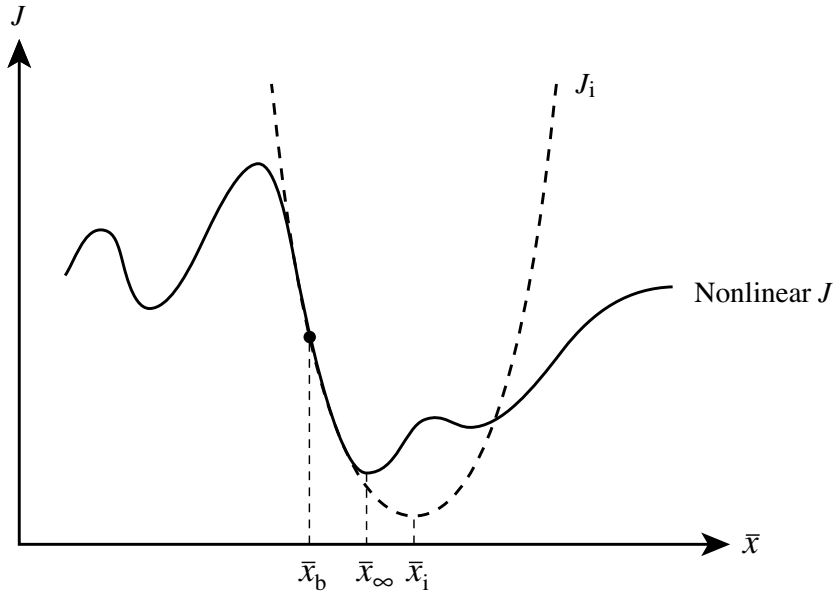

Figure 3: *Iterative solution for a nonlinear costfunction. In each iteration the costfunction is linearized around a previous estimate $\vec{x}_i$, which gives a quadratic costfunction J_i. The convergence depends on the accuracy of the first estimate of $\vec{x}$ (the background $\vec{x}_b$) and the nonlinearity of J.*

1.3.3 Illustrations with some special cases

We can now use the knowledge about the statistical distribution of the errors to solve the assimilation problem. The relative magnitude of different errors tells us which aspects of the information provided to the assimilation algorithm to give most weight. For example, if the background errors are larger than the observation errors, then the analysis will most probably be closer to the observations than the background. In Figs. 4– 8 we will illustrate the effect of a single observation y_n, with observation operator H_n, on the analysis $\vec{x}_a$ in different cases which can arise in applications. The figures are adapted from [17].

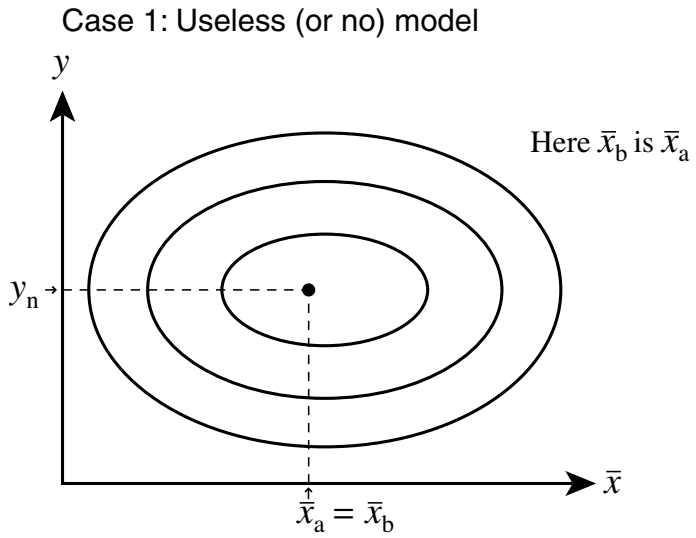

Figure 4: *Useless (or no) model. There is no background information available, and the analysis is at the maximum of* $P(x|y)$.

2 Variational data assimilation

In this section we will show how the assimilation problem, Eqs. 1– 3, can be solved for the special case of Gaussian probability density functions. Gaussian error distributions take a special place in applications because of their abundance in nature. From studying the error characteristics of background and observation errors arising in meteorological applications for example, we know that the Gaussian distribution is very often a good approximation. We have also mentioned that nonlinear assimilation problems with Gaussian error distributions can be formulated as an iteration of linear problems minimizing a quadratic costfunctions, for which very efficient numerical algorithms exist. Even non-Gaussian error distributions can in most cases be transformed to Gaussian error distributions by a suitable change of variable (see [18] for interesting examples). For these reasons most assimilation algorithms developed for atmospheric and oceanographic applications take the assumption of Gaussian errors as their starting point, and we will devote the rest of these notes to this case.

Below we will apply variational analysis to the nonlinear assimilation problem, and arrive at a full nonlinear solution which includes the Euler-Lagrange equation

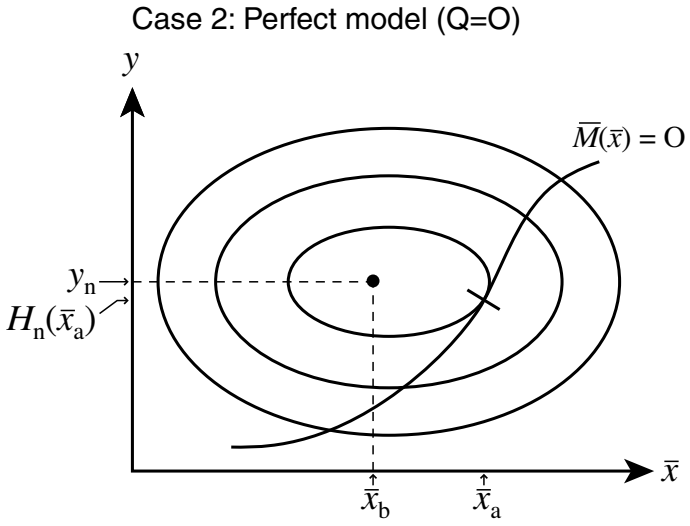

Figure 5: *Perfect model. In this case the model error is zero, and the analysis is constrained to lie along the model trajectory. We see that because of this the maximum probability has moved to a state consistent with the model.*

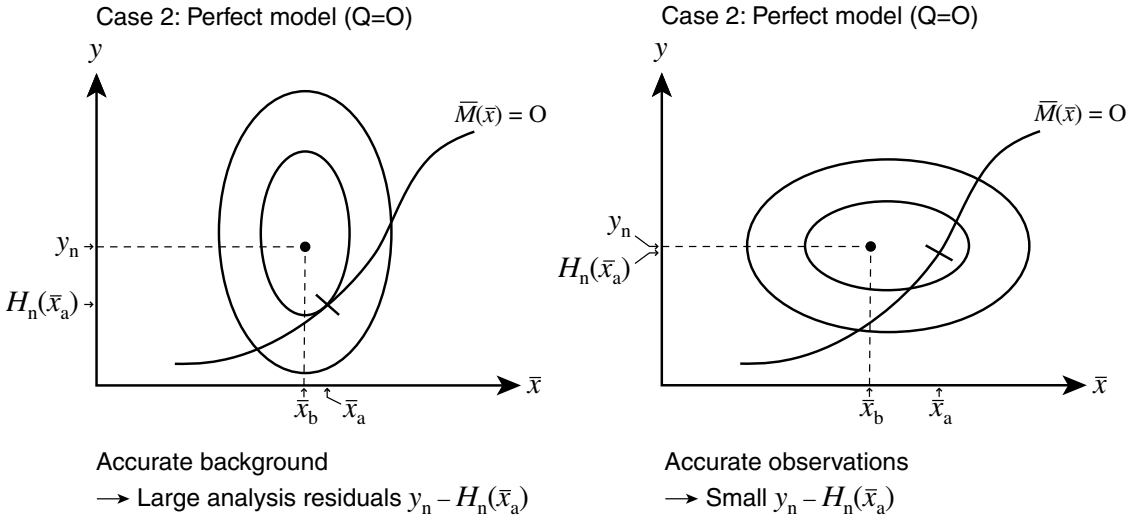

Figure 6: *The relative accuracy of background and observations. If the background is more accurate than the observation (left), then the analysis will be closer to the background, giving large* analysis residuals $y_n - H(\vec{x}_a)$. *Conversely, if the observations are more accurate (right), the analysis will be closer to the observation, giving small analysis residuals.*

for the assimilation problem (see [1], which uses the term generalized inverse for the solution). From this theoretical solution, we will see how different assumptions and simplifications can be made to arrive at some common data assimilation algorithms. It is worth pointing out that different theoretical approaches to the assimilation problem often end up with similar assimilation algorithms once all the assumptions necessary to achieve fast and accurate algorithms have been made (see for example [16, 7]).

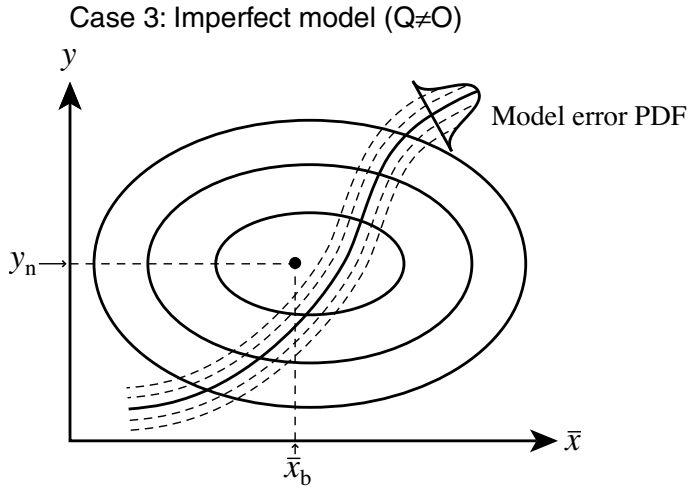

Figure 7: *Imperfect model. Now the analysis does no longer have to lie exactly on the model trajectory, but can deviate from the model. How much depends on the relative size of the model error. The model error pdf shown has a maximum at the model trajectory (the same as in previous examples), and isolines of smaller probability run parallel to the trajectory.*

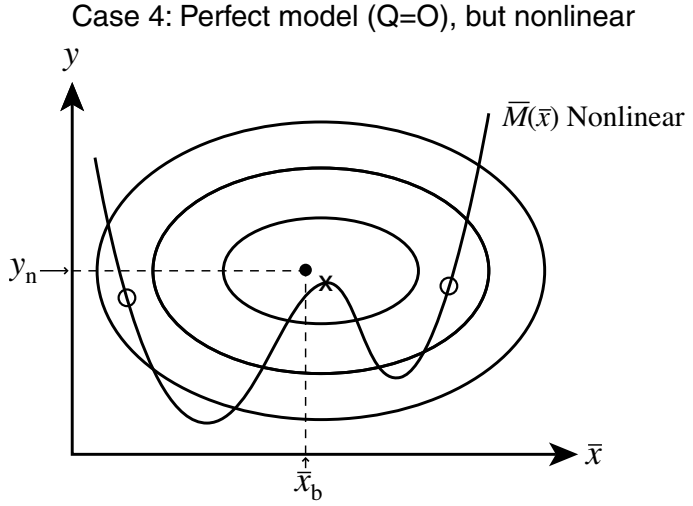

Figure 8: *Nonlinear model. With nonlinear models, the extremes of the pdf are not unique, and convergence of iterative methods is not guaranteed. We need to linearize and solve nonlinear assimilation problems as a series of approximate linear assimilation problems. Alternatively, apply nonlinear methods like the Monte Carlo method (see for example [12, 11]).*

We will start by discussing covariances and biases, then formulate the assimilation problem variationally and derive the solution, followed by discussions and examples.

2.1 Error covariances and biases

2.1.1 Covariances

As we have seen, the pdf of an unbiased Gaussian distribution is completely characterized by its covariance function, or its covariance matrix in the discrete case. The common notations for the covariances we will encounter are:

- $\vec{Q}(\xi, t, \xi', t')$ or $\mathbf{Q}$: *Model error covariance.* In the continuous case $\vec{Q}$ describes the covariances between errors in the model at any two locations ξ and ξ', and any two time instances t and t'. The model errors can be caused by processes not described by the model equations, or by an inaccurate estimate of some model parameters. If there are M model variables, then $\vec{Q}$ is a matrix with dimension $M \times M$, where each element is a covariance function. $\vec{B}$ and $\vec{A}$ have the same dimension as $\vec{Q}$.

- $\vec{B}(\vec{\xi}, \vec{\xi'})$ or $\mathbf{B}$: *Background error covariance.* Most geophysical systems remain close to some approximate dynamic (or other) balances. This translates into a correlation between the errors of different model variables. For example, in large scale atmospheric flow, the geostrophic balance translates into a strong correlation between wind and temperature errors. These cross correlations between variables always need to be included in $\vec{B}$. Otherwise the assimilation will produce an unbalanced analysis, which is not useful as an initial condition for forecasts, since the correction added by the analysis will be lost in transient noise caused by the model adjusting itself to a more balanced state. In most assimilation algorithms the formulation is much simpler if these cross correlations can be avoided. One widely used method is to use an analytical or statistical description of the balance between errors to transform the model variables into another set of variables which are uncorrelated. This is an example of a so called *control variable*, which is a term used to describe the variables which are actually used in the minimization step of an analysis. The control variable can even be a completely different set of variables (vorticity and divergence instead of wind) or a different representation can be used (spectral instead of gridpoint). See [10] for an introduction.

  - *Example: Uncorrelated control variable.* Let $\vec{x} = (u, \varphi)^T$ with background error covariance functions B_{uu}, $B_{\varphi\varphi}$ and $B_{u\varphi}$. To eliminate the cross-correlation between errors of different control variables, a transform is chosen so that

$$\begin{pmatrix} u - u_b \\ \varphi - \varphi_b \end{pmatrix} = \begin{pmatrix} \delta u \\ \delta\varphi \end{pmatrix} \longrightarrow \begin{pmatrix} \tilde{\delta u} \\ \tilde{\delta\varphi} \end{pmatrix} \tag{10}$$

$$\begin{pmatrix} B_{uu} & B_{u\varphi} \\ B_{\varphi u} & B_{\varphi\varphi} \end{pmatrix}^{-1} \begin{pmatrix} \delta u \\ \delta\varphi \end{pmatrix} \longrightarrow \begin{pmatrix} B_{\tilde{u}\tilde{u}} & 0 \\ 0 & B_{\tilde{\varphi}\tilde{\varphi}} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\delta u} \\ \tilde{\delta\varphi} \end{pmatrix} \tag{11}$$

- $\mathbf{R}$: *Observation error covariance.* The dimension of the observation error covariance matrix is $N \times N$. In $\mathbf{R}$ are included the effects of measurement errors, errors in the design of the observation operators and representativeness errors (this is just saying that a model can not represent all the small scale

variations of say temperature measured by a thermometer somewhere within a model gridpoint of say 50 km square).

- $\vec{A}(\vec{\xi},\vec{\xi'})$ or $\mathbf{A}$: *Analysis error covariance.* Some measure of $\vec{A}(\vec{\xi},\vec{\xi'})$ is always useful to diagnose how much of an improvement the analysis is upon the background.

  - *Example: The Hessian J'' as a measure of $\vec{A}$.* The probability density function for the analysis error $\vec{\varepsilon}_a$ is for the case of uncorrelated observation, background, and model errors discussed above $P_a(\vec{\varepsilon}_a) = P_m P_b P_o$. The corresponding costfunction $J_a = -\log P_a$ has a minimum at the analysis, so the gradient is zero, $J'_a = -P'_a/P_a = \vec{0}$. The second derivative (the Hessian) is $J''_a = -P''_a/P_a + P'_a/P_a^2 = -P''_a/P_a$. For the Gaussian case we are considering

$$P_a = \frac{1}{\sqrt{2\pi|\vec{A}|}} e^{-\frac{1}{2}\vec{\varepsilon}_a^T \vec{A}^{-1}\vec{\varepsilon}_a} \tag{12}$$

    and we get $P''_a = -\vec{A}^{-1}P_a - \vec{A}^{-1}\vec{\varepsilon}_a P'_a = -\vec{A}^{-1}P_a$ where we have used that $P'_a = \vec{0}$ for the optimal analysis (the maximum of the pdf). Collecting together these results, we see that for the Gaussian case, the analysis error covariance matrix is the inverse of the Hessian of the costfunction,

$$\vec{A} = (J''_a)^{-1} \tag{13}$$

2.1.2 Biases

Biases are the black sheep of the data assimilation family. Including biases in addition to covariances in the formulation of an assimilation algorithm is possible in principle, but in practice it is extremely difficult to determine the bias. We may find that a given set of observations and a model forecast (the background) are biased with respect to each other, but is it the observation or the model forecast that is biased? How to include biases in assimilation algorithms and how to estimate the bias in different components of an assimilation system is a challenging area of research.

Biases must be dealt with in one way or another for the error characterization to be acceptable for an assimilation system. The following methodology is often used. If the bias of an observation is known, subtract it from the observation, and feed the unbiased estimate of the observation to the analysis. This process is called bias removal. Much research is usually needed to characterize the bias for different observation types, giving rise to different bias correction parameters for each observation type, where the parameters $\vec{\beta}$ are a function of the measuring geometry and the model state.

Once the bias correction models with parameters $\vec{\beta}$ are known, these parameters can alternatively be estimated by including them in the assimilation ([9]). The bias parameters are assumed slowly varying (which is different from the biases being slowly varying!) and are added to the costfunction by modifying the observation operator from $\vec{H}(\vec{x})$ to $\vec{H}(\vec{x}) + \vec{b}(\vec{x},\vec{\beta})$, where $\vec{b}(\vec{x},\vec{\beta})$ is the bias. If the pdf of the

bias parameters is Gaussian with covariance matrix $\vec{C}$ and previous estimate $\vec{\beta}_0$ is given, then the costfunction (without the model error term) is

$$\begin{aligned} J &= \frac{1}{2}(\vec{x}(0)-\vec{x}_b(0))^T\vec{B}^{-1}(\vec{x}(0)-\vec{x}_b(0)) && (14)\\ &+ \frac{1}{2}(\vec{y}-\vec{b}(\vec{x},\vec{\beta})-\vec{H}(\vec{x}))^T\vec{R}^{-1}(\vec{y}-\vec{b}(\vec{x},\vec{\beta})-\vec{H}(\vec{x})) && (15)\\ &+ \frac{1}{2}(\vec{\beta}-\vec{\beta}_0)^T\vec{C}^{-1}(\vec{\beta}-\vec{\beta}_0) && (16)\end{aligned}$$

But there still remains the fact that the model may be biased. This bias is often ignored or simply absorbed into the observation bias. Alternatively, the model bias can also be estimated as part of the assimilation by including its statistical description in the model error term. This is an area of active research.

In the remainder of these notes we will emphasize assimilation algorithms without bias, assuming that appropriate bias removal techniques are used, and just note that bias removal remains one of the hardest aspects of the assimilation.

2.2 Formulation of the variational assimilation problem

From the discussion on probability density functions we know that the most likely solution is also the best solution for a Gaussian pdf. The most likely solution also gives the minimum of the costfunction defined by a pdf. This costfunction actually defines what norm best measures the accuracy of the solution (see [17, 16, 18]). For a Gaussian pdf the costfunction is quadratic, which implies the L_2, or least square, norm. So the optimum solution minimizes the sum of the errors in the model, background and observations as measured by the L_2 norm integrated over the model volume V and the assimilation time window $[0,\tau]$ (which is just a fancy way of saying that the optimum solution minimizes the following costfunction),

$$\begin{aligned} J(\vec{x}) &= \frac{1}{2}\int_V d\xi\int_0^\tau dt\int_V d\xi'\int_0^\tau dt'(\frac{\partial\vec{x}}{\partial t}+\vec{M}(\vec{x}))^T\vec{Q}^{-1}(\xi,t,\xi',t')(\frac{\partial\vec{x}}{\partial t}+\vec{M}(\vec{x}))\\ &+ \frac{1}{2}\int_V d\xi\int_V d\xi'(\vec{x}(\xi,0)-\vec{x}_b(\xi,0))^T\vec{B}^{-1}(\xi,\xi')(\vec{x}(\xi',0)-\vec{x}_b(\xi',0))\\ &+ \frac{1}{2}\int_V d\xi\int_0^\tau dt\int_V d\xi'\int_0^\tau dt'(\vec{y}-\vec{H}(\vec{x}))^T\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x})) \qquad (17)\end{aligned}$$

First a couple of clarifications. The double integration over time and space just expresses all possible correlations between errors at (ξ,t) (terms to the left of the correlation functions) and errors at (ξ',t') (terms to the right). The observation error in particular may look odd, but we have to express precisely where and when to evaluate the departures $y_n - H_n(\vec{x})$. We do this by multiplying $y_n - H_n(\vec{x})$ by a 'localization function' $\phi_n(\xi,t)$ so that element n of $(\vec{y}-\vec{H}(\vec{x}))$ is

$$(\vec{y}-\vec{H}(\vec{x}))_n = \phi_n(\xi,t)(y_n - H_n(\vec{x})) \qquad (18)$$

The localization function, which has the following property,

$$\int_V d\xi\int_0^\tau dt\phi_n(\xi,t) = 1 \qquad (19)$$

can be anything from a delta function at the observation time and location to more advanced averaging in space or time, say over a satellite footprint. A particular form of the localization function, which is often used to simplify assimilation algorithms, is to evaluate all the departures at the same time (the analysis time) irrespective of when the observations were made within the assimilation window.

Now the standard tools of variational analysis can be used to obtain the optimum solution (see e. g. [14]). Let $\vec{x}_a$ (our analysis) be the model state which gives the minimum of J with respect to a variation of $\vec{x}$. Define

$$\vec{x}(\xi, t) = \vec{x}_a(\xi, t) + \gamma\vec{\eta}(\xi, t) \tag{20}$$

where γ is a constant and $\vec{\eta}(\xi, t)$ is an arbitrary vector (with dependent components however). The minimum of Eq. 17 is found from

$$\lim_{\gamma \to 0} \frac{dJ}{d\gamma} = 0 \tag{21}$$

In Fig. 9 an interpretation of the variational procedure is shown.

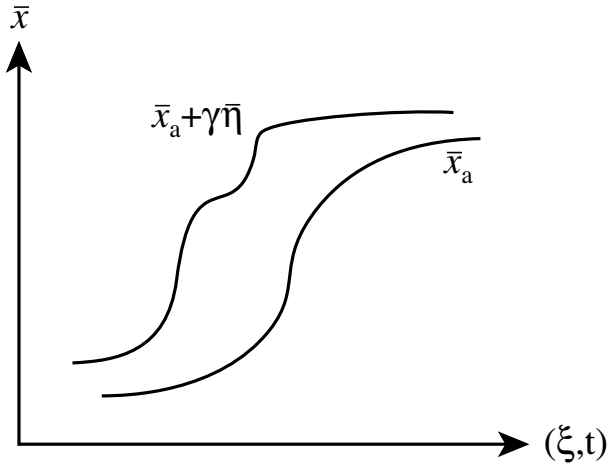

Figure 9: *Variation of $\vec{x}(\xi, t)$ to get the minimum of $J(\vec{x})$.*

2.3 Solution of the variational problem

Using Eq. 20 in Eq. 17 we get the following condition at the minimum of J,

$$\begin{aligned}
\lim_{\gamma\to 0}\frac{d}{d\gamma} \quad (\quad & \frac{1}{2}\int_V d\xi \int_0^\tau dt(\frac{\partial \vec{x}_a}{\partial t}+\gamma\frac{\partial\vec{\eta}}{\partial t}+\vec{M}(\vec{x}_a+\gamma\vec{\eta}))^T \\
& \times\int_V d\xi' \int_0^\tau dt'\vec{Q}^{-1}(\frac{\partial \vec{x}_a}{\partial t}+\gamma\frac{\partial\vec{\eta}}{\partial t}+\vec{M}(\vec{x}_a+\gamma\vec{\eta})) \\
& +\frac{1}{2}\int_V d\xi\int_V d\xi'(\vec{x}_a+\gamma\vec{\eta}-\vec{x}_b)^T\vec{B}^{-1}(\vec{x}_a+\gamma\vec{\eta}-\vec{x}_b) \\
& +\frac{1}{2}\int_V d\xi\int_0^\tau dt\int_V d\xi'\int_0^\tau dt'(\vec{y}-\vec{H}(\vec{x}_a+\gamma\vec{\eta}))^T\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a+\gamma\vec{\eta}))) \\
= \quad & \int_V d\xi\int_0^\tau dt(\frac{\partial\vec{\eta}}{\partial t}+\frac{\partial\vec{M}}{\partial\vec{x}_a}\vec{\eta})^T\vec{\lambda}(\xi,t) \\
+ \quad & \int_V d\xi\int_V d\xi'\vec{\eta}^T(\xi,0)\vec{B}^{-1}(\vec{x}_a(\xi',0)-\vec{x}_b(\xi',0)) \\
- \quad & \int_V d\xi\int_0^\tau dt\int_V d\xi'\int_0^\tau dt'(\frac{\partial\vec{H}}{\partial\vec{x}_a}\vec{\eta})^T\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a)) \\
= \quad & 0
\end{aligned} \tag{22}$$

where we have introduced the *'adjoint variable'* $\vec{\lambda}$ as

$$\vec{\lambda}(\xi,t)=\int_V d\xi'\int_0^\tau dt'\vec{Q}^{-1}(\xi,t,\xi',t')(\frac{\partial\vec{x}_a(\xi',t')}{\partial t}+\vec{M}(\vec{x}_a(\xi',t'))) \tag{23}$$

and the *jacobians* $\frac{\partial\vec{M}}{\partial\vec{x}_a}$ and $\frac{\partial\vec{H}}{\partial\vec{x}_a}$ of the model and the observation operators, which are defined by

$$\lim_{\gamma\to 0}\frac{d\vec{M}(\vec{x}_a+\gamma\vec{\eta})}{d\gamma}=\lim_{\vec{x}\to\vec{x}_a}\frac{\partial\vec{M}(\vec{x})}{\partial\vec{x}}\vec{\eta}=\frac{\partial\vec{M}}{\partial\vec{x}_a}\vec{\eta} \tag{24}$$

and the same for $\vec{H}$.

- *Example: Model errors.* From Eq. 1 we recognize the model error $\vec{\varepsilon}_m$ as the term to the right of $\vec{Q}$ in the definition of the adjoint variable. By inverting the relationship we get an explicit expression for the model errors at the minimum of the costfunction

$$\vec{\varepsilon}_m(\xi,t)=\int_V d\xi'\int_0^\tau dt'\vec{Q}(\xi,t,\xi',t')\vec{\lambda}(\xi',t') \tag{25}$$

- *Example: The jacobian of an advection operator.* In the one-dimensional advection equation $\frac{\partial\varphi}{\partial t}+u\frac{\partial\varphi}{\partial\xi}=0$, the jacobian of the operator $M_\varphi=u\frac{\partial\varphi}{\partial\xi}$ is, with the state vector $\vec{x}=(u,\varphi)$,

$$\frac{\partial M_\varphi}{\partial\vec{x}_a}\vec{\eta}=\begin{pmatrix}0 & 0\\ \frac{\partial\varphi_a}{\partial\xi} & u_a\frac{\partial}{\partial\xi}\end{pmatrix}\begin{pmatrix}\eta_u\\ \eta_\varphi\end{pmatrix}=\frac{\partial\varphi_a}{\partial\xi}\eta_u+u_a\frac{\partial\eta_\varphi}{\partial\xi} \tag{26}$$

  We note that the jacobian of a nonlinear operator acts linearly on $\vec{\eta}$. The jacobian also appears when operators are linearized, and is then referred to as the *tangent linear* operator.

If we can rearrange the expression for the minimum in Eq. 22 so that $\vec{\eta}$ only appears as a multiplicative factor, then the remaining expressions must be zero at the minimum of J, since $\vec{\eta}$ is arbitrary. This can be achieved by integration by parts. Term by term we get

$$\begin{aligned}\int_V d\xi \int_0^\tau dt(\frac{\partial\vec{\eta}}{\partial t})^T\vec{\lambda} &= \int_V d\xi\vec{\eta}^T(\xi,\tau)\vec{\lambda}(\xi,\tau) - \int_V d\xi\vec{\eta}^T(\xi,0)\vec{\lambda}(\xi,0) \\ &\quad - \int_V d\xi \int_0^\tau dt\vec{\eta}^T(\xi,t)\frac{\partial\vec{\lambda}}{\partial t} \qquad (27)\end{aligned}$$

$$\int_V d\xi \int_0^\tau dt(\frac{\partial\vec{M}}{\partial\vec{x}_a}\vec{\eta})^T\vec{\lambda} = \int_V d\xi \int_0^\tau dt\vec{\eta}^T(\xi,t)(\frac{\partial\vec{M}}{\partial\vec{x}_a})^*\vec{\lambda} \qquad (28)$$

$$\begin{aligned}&\int_V d\xi \int_0^\tau dt \int_V d\xi' \int_0^\tau dt'(\frac{\partial\vec{H}}{\partial\vec{x}_a}\vec{\eta})^T\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a)) = \\ &\int_V d\xi \int_0^\tau dt \int_V d\xi' \int_0^\tau dt'\vec{\eta}^T(\xi,t)(\frac{\partial\vec{H}}{\partial\vec{x}_a})^*\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a)) \qquad (29)\end{aligned}$$

Here we have introduced a compact notation for the integration by parts of the jacobians, namely the *adjoint operators* $(\frac{\partial\vec{M}}{\partial\vec{x}_a})^*$ and $(\frac{\partial\vec{H}}{\partial\vec{x}_a})^*$.

- *Example: The adjoint of* $\frac{\partial}{\partial t}$. We have already seen an example of an adjoint operator in Eq. 27. If we use the simplified notation $<\cdot,\cdot>$ for the integration of the product of the two vectors over time and space (that is, we define a scalar product in this way) we see that

$$<\frac{\partial}{\partial t}\vec{\eta},\vec{\lambda}>=<\vec{\eta},(\frac{\partial}{\partial t})^*\vec{\lambda}> \qquad (30)$$

  where the adjoint operator $(\frac{\partial}{\partial t})^*$ is

$$(\frac{\partial}{\partial t})^* = \delta(t-\tau) - \delta(t) - \frac{\partial}{\partial t} \qquad (31)$$

  and $\delta(t-\tau)$ and $\delta(t)$ are Dirac delta functions.

- *Example: The adjoint of the advection jacobian.* We assume that the one-dimensional advection takes place in a periodic domain $[0,L]$. Integration by parts in ξ gives

$$\begin{aligned}&\int_0^L d\xi \int_0^\tau dt \left(\begin{pmatrix} 0 & 0 \\ \frac{\partial\varphi_a}{\partial\xi} & u_a\frac{\partial}{\partial\xi}\end{pmatrix}\begin{pmatrix}\eta_u\\ \eta_\varphi\end{pmatrix}\right)^T\begin{pmatrix}\lambda_u\\ \lambda_\varphi\end{pmatrix} \\ =\ &\int_0^L d\xi \int_0^\tau dt(\frac{\partial\varphi_a}{\partial\xi}\eta_u\lambda_u + u_a\frac{\partial\eta_\varphi}{\partial\xi}\lambda_\varphi) \\ =\ &\int_0^\tau dt[u_a\eta_\varphi\lambda_\varphi]_0^L + \int_0^L d\xi \int_0^\tau dt(\eta_u\frac{\partial\varphi_a}{\partial\xi}\lambda_u - \eta_\varphi\frac{\partial}{\partial\xi}(u_a\lambda_\varphi)) \\ =\ &\int_0^L d\xi \int_0^\tau dt\begin{pmatrix}\eta_u\\ \eta_\varphi\end{pmatrix}^T\begin{pmatrix}\frac{\partial\varphi_a}{\partial\xi} & 0\\ 0 & -(\frac{\partial u_a}{\partial\xi}+u_a\frac{\partial}{\partial\xi})\end{pmatrix}\begin{pmatrix}\lambda_u\\ \lambda_\varphi\end{pmatrix} \qquad (32)\end{aligned}$$

  Because of the periodicity, all boundary terms disappear, but it is straightforward to allow for any boundary conditions.

The rearranged expression for the minimum of J is now

$$\begin{aligned} 0 &= \int_V d\xi \int_0^\tau dt \vec{\eta}^T(\xi,t)(-\frac{\partial\vec{\lambda}}{\partial t} + (\frac{\partial\vec{M}}{\partial\vec{x}_a})^*\vec{\lambda} - \int_V d\xi' \int_0^\tau dt'(\frac{\partial\vec{H}}{\partial\vec{x}_a})^*\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a))) \\ &+ \int_V d\xi \vec{\eta}^T(\xi,0)(\int_V d\xi' \vec{B}^{-1}(\vec{x}_a(\xi',0)-\vec{x}_b(\xi',0)) - \vec{\lambda}(\xi,0)) \\ &+ \int_V d\xi \vec{\eta}^T(\xi,\tau)\vec{\lambda}(\xi,\tau) \end{aligned} \tag{33}$$

2.4 Summary of the variational solution

The final equations now follow from that the arbitrary vector function $\vec{\eta}$ can be varied independently in each of the three integrals in Eq. 33, so the expressions following $\vec{\eta}$ in each of the integrals must be identically zero. Add to this the definition of the adjoint variable $\vec{\lambda}$ in Eq. 23 and we have the *generalized inverse* solution consisting of the forward model equation with initial conditions and the Euler-Lagrange (adjoint) equation with end conditions

$$\vec{\lambda}(\xi,\tau) = \vec{0} \tag{34}$$

$$-\frac{\partial\vec{\lambda}}{\partial t} + (\frac{\partial\vec{M}}{\partial\vec{x}_a})^*\vec{\lambda} = \int_V d\xi' \int_0^\tau dt'(\frac{\partial\vec{H}}{\partial\vec{x}_a})^*\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a)) \tag{35}$$

$$\int_V d\xi' \vec{B}^{-1}(\xi,\xi')(\vec{x}_a(\xi',0)-\vec{x}_b(\xi',0)) - \vec{\lambda}(\xi,0) = \vec{0} \tag{36}$$

$$\frac{\partial\vec{x}_a}{\partial t} + \vec{M}(\vec{x}_a) = \int_V d\xi' \int_0^\tau dt' \vec{Q}(\xi,t,\xi',t')\vec{\lambda}(\xi',t') \tag{37}$$

These are the equations fulfilled by the optimum solution $\vec{x}_a$ at the minimum of the Gaussian costfunction. The solution includes an estimate of the model errors in terms of the adjoint variable. An inverse problem of this form, where the model is not assumed to be exact, is called a *weak constraint* problem. The forward and adjoint equations are coupled, since $\vec{\lambda}$ appears in the forward equation and $\vec{x}_a$ appears in the adjoint equation. This makes it harder to construct assimilation algorithms for this general case, but it can be done (see [3, 4]).

A simplification is to assume the model to be exact ($\vec{Q} = \vec{0}$) which gives a *strong constraint* problem. Now the adjoint only influences the initial condition of the solution, and the forward model integration is simplified. This assumption is made in the formulation of most assimilation algorithms. This is fine when it can be demonstrated that the model error is small relative to observation and background errors. Even when model errors are not quite negligible, they are often not included in the assimilation either because they are too difficult to characterize or too expensive to calculate.

The solution for the generalized inverse problem is sketched in Fig. 10.

The generalized solution can be translated directly onto a discrete form, replacing continuous fields and operators by corresponding vectors and matrices. For the discrete case it is often an advantage to let the model operator include the time derivatives. One particular simplification in the discrete case is that the *adjoint of a linear operator, expressed as a matrix, is equal to the transpose of that matrix.* This property is very useful when coding the adjoints of discrete operators.

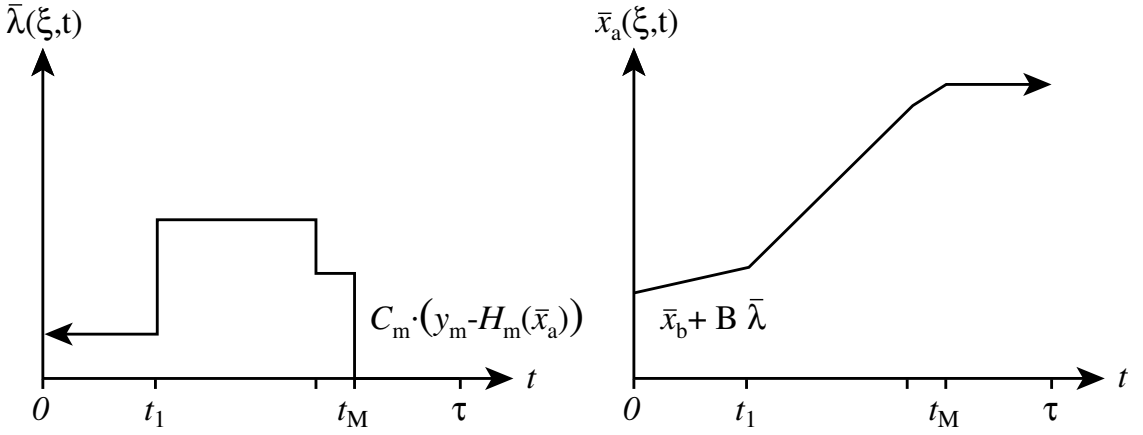

Figure 10: *The solution of the generalized inverse problem. Observations are a source term in the backward-in-time integration of the adjoint* $\vec{\lambda}$*. This gives a jump at each observation which translates into a change in the forcing in the forward integration of* $\vec{x}_a$*. The initial condition receives a contribution* $\vec{B}\vec{\lambda}$*. For the strong constraint problem the adjoint solution is unchanged, but there is no contribution to the forcing of the forward integration. When error correlations in time are included, the changes in* $\vec{\lambda}$ *and* $\vec{x}_a$ *go via smooth transitions.*

- *Example: Observations in the Euler-Lagrange equation.* The right hand side of the Euler-Lagrange equation Eq. 35 contains observations as source terms in the backward-in-time integration. Going back to the steps in the solution above, we recall that everything to the right of $\vec{R}^{-1}$ is a function of the primed coordinates (ξ', t'), and everything to the left is a function of (ξ, t), so we should have written

$$r.h.s. = (\frac{\partial \vec{H}}{\partial \vec{x}_a})^* \vec{R}^{-1} \int_V d\xi' \int_0^\tau dt' (\vec{y} - \vec{H}(\vec{x}_a))$$

  A departure

$$\int_V d\xi' \int_0^\tau dt' \phi_n(\xi', t')(y_n - H_n(\vec{x}_a))$$

  is spread by $\vec{R}^{-1}$ to give departures at all (correlated) observation points. Then $(\frac{\partial \vec{H}}{\partial \vec{x}_a})^*$ moves the resulting departures at observation points to a continuous increment in the model variables. In the discrete case the right hand side would become $H^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}_a))$.

- *Example: Observation and model space.* In observation space all quantities are comparable with the observations: y_n and $H_n(\vec{x})$ are both in observation space. In model space all quantities are comparable with the state vector: $\vec{x}$ and $(\frac{\partial H_n}{\partial \vec{x}_a})^*(y_n - H_n(\vec{x}_a))$ are examples. Assimilation algorithms can be formulated in either model or observation space: for each model space algorithm there is a corresponding *dual* form of the algorithm in observation space (see [7]). The link between the two spaces are the observation operators. In observation space, the assimilation problem dimension is determined by the number of observations. This can be an advantage when there are relatively

few observations, like in some oceanographic applications. In model space, the assimilation problem is determined by the model dimension. This can be an advantage when more and more observations are assimilated, like satellite data in meteorological assimilation systems.

- *Example: Implementation of assimilation algorithms.* The generalized inverse gives the exact relation for the analysis in the continuous case. When implementing assimilation algorithms we are always working with discrete models. The discrete equivalent of the generalized inverse must be coded so that it is internally consistent. For example, the adjoint of a linear operator must be exactly the transpose of the matrix for that operator. This we cannot guarantee by discretizing the continuous form of the generalized inverse directly. A practical approach is to derive the discrete tangent linear equivalents of all operators, and then derive the adjoint operators as the transpose of the tangent linear operators.

3 Common assimilation algorithms

We will now turn our attention to specific assimilation algorithms: optimal interpolation (OI), three-dimensional variational assimilation (3DVAR), four-dimensional variational data assimilation (4DVAR) and the Kalman filter. We will interpret each of the algorithms in terms of the generalized inverse solution, in particular what aspects of the generalized inverse they capture and what the approximations are. Starting from the generalized inverse, we will translate the terms remaining after approximations into commonly encountered expressions for the algorithms, moving to discrete notation where appropriate.

3.1 Optimal interpolation and 3DVAR

OI and 3DVAR are the simplest algorithms of the above – they do not include the dynamic evolution of the model in the assimilation. While OI and 3DVAR are similar in this respect, there are significant differences which are important in applications.

3.1.1 Relation to the generalized inverse

- *Approximation 1:* No model errors, $\vec{Q} = \vec{0}$.
- *Approximation 2:* The evolution of $\vec{x}_a$ is described perfectly by $\frac{\partial \vec{x}_a}{\partial t} = \vec{0}$, i. e. $\vec{M} = \vec{0}$.
- *Approximation 3:* Evaluate all observation operators at $t = \tau/2$. This means the localization function for an observation y_n is $\phi_n(\xi)\delta(t-\tau/2)$, with $\delta(t-\tau/2)$ a Dirac delta function at $\tau/2$.

All that now remains of the generalized inverse solution is the following (see Fig. 11)

$$\vec{\lambda}(\xi, \tau) = \vec{0}$$

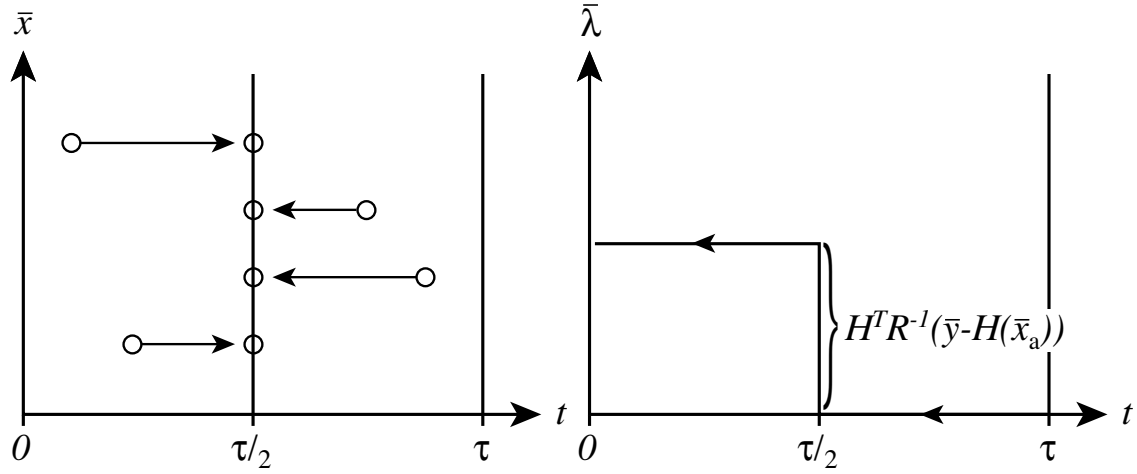

Figure 11: *The OI/3DVAR analysis as seen by the generalized inverse. All departures are evaluated at $\tau/2$ and the model operator is identity.*

$$-\frac{\partial\vec{\lambda}}{\partial t} = \delta(t-\tau/2)\int_V d\xi'(\frac{\partial\vec{H}}{\partial\vec{x}_a})^*\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a))$$
$$\int_V d\xi'\vec{B}^{-1}(\xi,\xi')(\vec{x}_a(\xi',0)-\vec{x}_b(\xi',0)) - \vec{\lambda}(\xi,0) = \vec{0}$$
$$\frac{\partial\vec{x}_a}{\partial t} = \vec{0}$$

Integrating the adjoint equation backward in time from the zero end condition we get a jump in $\vec{\lambda}$ at $t = \tau/2$ time where all the observation operators are evaluated,

$$\vec{\lambda}(\xi,\tau/2) = \int_V d\xi'(\frac{\partial\vec{H}}{\partial\vec{x}_a})^*\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a)) \tag{38}$$

Here the time dimension of the covariance function has disappeared, since all departures are evaluated at the same time. The solution then remains constant between $t = 0$ and $t = \tau/2$ ($\vec{\lambda}(\xi,0) = \vec{\lambda}(\xi,\tau/2)$ and $\vec{x}_a(\xi,0) = \vec{x}_a(\xi,\tau/2)$) so the initial condition is also valid at $t = \tau/2$. By defining the analysis time in the middle of the assimilation window we use background $\vec{x}_b$ which is never further away than $\tau/2$ from any observation,

$$\int_V d\xi'(\vec{B}^{-1}(\xi,\xi')(\vec{x}_a(\xi',\tau/2)-\vec{x}_b(\xi',\tau/2)) - (\frac{\partial\vec{H}}{\partial\vec{x}_a})^*\vec{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_a))) = \vec{0} \tag{39}$$

3.1.2 Discrete formulation

We now continue the description of OI and 3DVAR using the discrete formulation in which they are mostly expressed. Converting the above expression to the discrete case gives

$$\mathbf{B}^{-1}(\mathbf{x}_a-\mathbf{x}_b) - H^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}_a)) = \vec{0} \tag{40}$$

We need to solve for $\mathbf{x}_a$, but $H(\mathbf{x}_a)$ is nonlinear in $\mathbf{x}_a$, and the adjoint operator H^T contributes to the nonlinearity as well (remember adjoints and transposes are equivalent in the discrete case).

- *Approximation 4:* Let $\mathbf{H}$ be a linear approximation to the jacobian of H (a *tangent linear* approximation). Replace $H(\mathbf{x}_a)$ by the linear expression $H(\mathbf{x}_b) + \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b)$ and the (nonlinear) adjoint H^T of the jacobian by the (linear) adjoint $\mathbf{H}^T$ of the tangent linear approximation.

The linearized equation is

$$\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}_b) - \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b)) = \vec{0} \tag{41}$$

After some rearrangement we arrive at the following equivalent expressions (see [5]) for the optimal solution $\mathbf{x}_a$,

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H(\mathbf{x}_b)) \tag{42}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{43}$$

where $\mathbf{K}$ is the *gain* or *weight* matrix. A measure of the accuracy of the analysis is given by the analysis error covariance matrix $\mathbf{A}$ (see [5])

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \tag{44}$$

- *Example: Interpretation of the gain matrix.* $\mathbf{H}\mathbf{B}\mathbf{H}^T$ is just the background error covariance expressed in observation space. $\mathbf{H}^T$ takes the departures divided by the sum of observation and background covariances back to model space, where it is multiplied by the background error. This is analogous to the factor $\sigma_b^2/(\sigma_b^2 + \sigma_o^2)$ which arises in the optimal least square solution of a single direct observation y with error variance σ_o^2 of a quantity x, given a background estimate x_b with error variance σ_b^2,

$$x_a = x_b + \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)}(y - x_b) \tag{45}$$

  We see here again how the observation operators translate between observation and model space, making it possible to compare matrixes of different dimensions (e. g. $\mathbf{R}$ which is in observation space and $\mathbf{B}$, which is in model space).

We have thus shown that the solution of the variational assimilation problem under approximations 1–4 is equivalent to Eqs. 42– 43 for the solution at the minimum. This suggests two types of algorithms for solving the assimilation problem on this form. We can solve the expression for $\mathbf{x}_a$ directly by matrix inversions. But we can also solve the variational problem which leads to $\mathbf{x}_a$ iteratively, starting from $\mathbf{x}_b$. We now briefly discuss optimal interpolation and three-dimensional variational assimilation as examples of these two approaches.

3.1.3 Optimal interpolation – OI

In optimal interpolation the analysis equation is solved directly by inversion. In the inversion the gain matrix $\mathbf{K}$ is simplified by assuming that only the closest observations determine the analysis increment at ξ. The global analysis problem is thus divided into blocks, and a local optimal analysis is found for each block

(see Fig. 12). This reduces the size and the time it takes to solve the analysis problem numerically. However, this is not a global solution, and jumps can occur in the analysis when solutions from different blocks are joined together. Another disadvantage of optimal interpolation is that it is difficult to use observations with complex observation operators. This is because background errors need to be formed for the model equivalent of each observation. For this reason only observations with simple observation operators are used in OI.

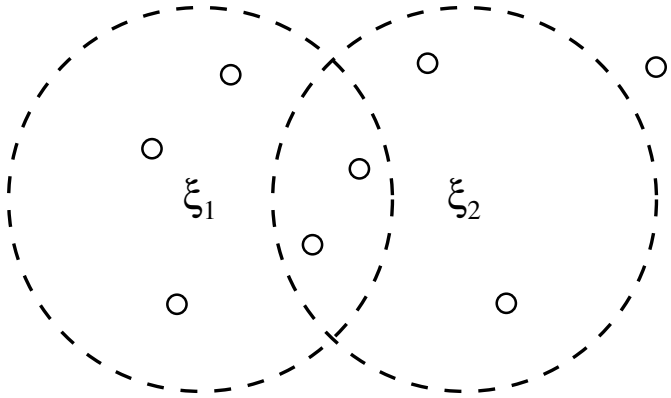

Figure 12: *An example of data selection in OI. There is some overlap between observations used to determine the analysis at* ξ_1 *and* ξ_2.

- *Example: Satellite radiances* The observation operator for satellite radiances includes a radiative transfer model. The model equivalent background error has to take into account the effect of the radiative transfer model on a column of temperature, humidity, ozone and many more parameters. Instead of radiances, retrievals of temperature, humidity and ozone are used in OI. This brings other problems, since retrievals use background estimates of the meteorological fields, leading to correlations between observation and background errors.

3.1.4 Three-dimensional variational assimilation – 3DVAR

In 3DVAR the analysis equation is solved iteratively, starting from $\vec{x} \neq \vec{x}_a$,

$$\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}_b)-\mathbf{H}(\mathbf{x}-\mathbf{x}_b)) \neq \vec{0} \tag{46}$$

We recognize this expression as the gradient $J'(\mathbf{x})$ of the (linearized) costfunction $J(\mathbf{x})$,

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}(\mathbf{y}-H(\mathbf{x}_b)-\mathbf{H}(\mathbf{x}-\mathbf{x}_b))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}_b)-\mathbf{H}(\mathbf{x}-\mathbf{x}_b)) \\ &+ \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) \end{aligned} \tag{47}$$

The linearized analysis problem has thus a quadratic costfunction, and we can use J and J' in a minimization algorithm to approach the unique minimum of J where $J' = 0$. This gives us the analysis $\mathbf{x}_a$ (see Fig. 13).

Compared with OI, 3DVAR gives a global analysis and it is easy to use any observation. We only need the observation operator and its tangent linear and adjoint in the analysis. However, it requires significant effort to develop observation operators for complex observations. For satellite radiances for example, a 'fast' radiative transfer model together with its tangent linear and adjoint need to be developed.

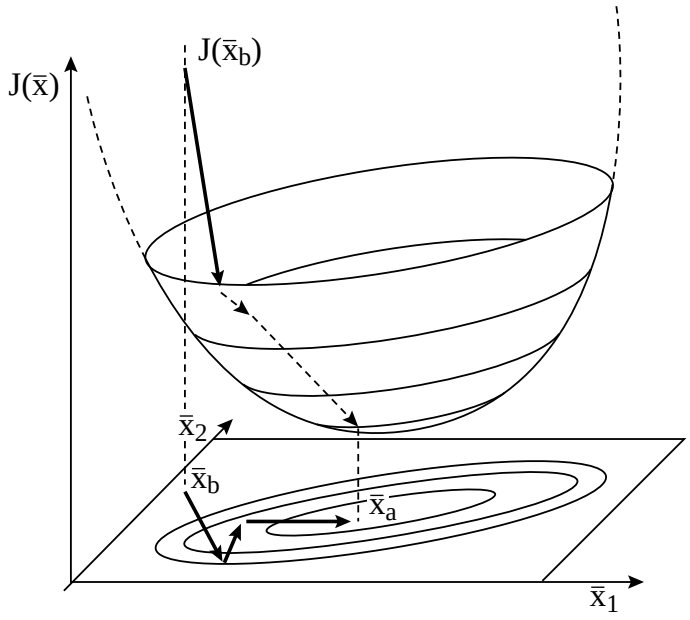

Figure 13: *Minimization of a quadratic costfunction $J(\mathbf{x})$. The gradient of the costfunction and the costfunction itself are supplied to a minimization algorithm which works out how to change* $\mathbf{x}$ *to obtain a smaller value of the costfunction.*

3.1.5 Dual formulation of OI/3DVAR

In the above analysis equation $\mathbf{K}(\mathbf{y} - H(\mathbf{x}_b))$ is an analysis increment in model space. An alternative way to solve the problem is to find analysis increments in observation space and then map them back to model space. This *dual* formulation (often referred to as 'PSAS') is obtained by rewriting the analysis equation as follows

$$\mathbf{w}_a = (\mathbf{HBH}^T + \mathbf{R})^{-1}(\mathbf{y} - H(\mathbf{x}_b)) \tag{48}$$

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{BH}^T\mathbf{w}_a \tag{49}$$

This system can be solved directly or iteratively, leading to the dual formulation of OI and 3DVAR. See [7] for a further discussion.

3.2 Four-dimensional variational assimilation – 4DVAR

In contrast to 3DVAR and OI, 4DVAR includes the dynamic evolution of the model in the assimilation. The 4DVAR algorithm is very close to the generalized inverse – only model errors are neglected. The essential difference between 3DVAR and 4DVAR is illustrated in Fig. 14.

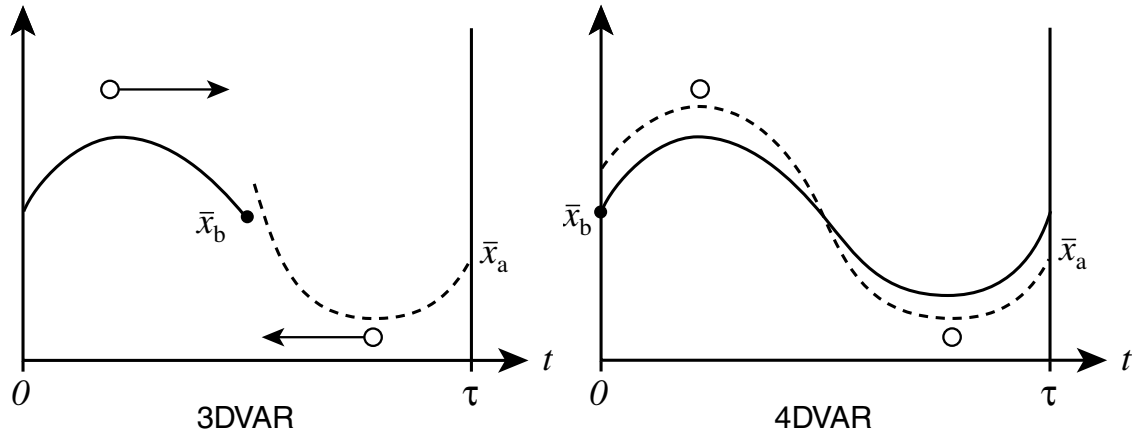

Figure 14: *The difference between 3DVAR and 4DVAR – 4DVAR is a time/model consistent interpolator of departures.*

3.2.1 Relation to the generalized inverse

- *Approximation 1:* No model errors, $\vec{Q} = \vec{0}$.

The only change to the generalized inverse solution is that the forward model equation no longer depends on the adjoint variable $\vec{\lambda}$,

$$\frac{\partial \vec{x}_a}{\partial t} + \vec{M}(\vec{x}_a) = \vec{0} \tag{50}$$

This makes the assimilation problem much easier to solve. Although $\vec{x}_a$ still appears in the adjoint equation, an iterative scheme can be designed to arrive at the analysis. To see this, let us rewrite the equations in terms of $\vec{x}_b$ and the *analysis increment* $\delta\vec{x}_a$,

$$\delta\vec{x}_a = \vec{x}_a - \vec{x}_b \tag{51}$$

Inserting this in the model and observation operators, we get

$$\vec{M}(\vec{x}_a) = \vec{M}(\vec{x}_b) + \frac{\partial \vec{M}}{\partial \vec{x}_a}\delta\vec{x}_a \tag{52}$$

and similar for $\vec{H}(\vec{x}_a)$, with the nonlinear jacobians $\frac{\partial \vec{M}}{\partial \vec{x}_a}$ and $\frac{\partial \vec{H}}{\partial \vec{x}_a}$ of the model and observation operators appearing. This gives

$$\vec{\lambda}(\xi, \tau) = \vec{0} \tag{53}$$

$$-\frac{\partial \vec{\lambda}}{\partial t} + (\frac{\partial \vec{M}}{\partial \vec{x}_a})^*\vec{\lambda} = \int_V d\xi' \int_0^\tau dt' (\frac{\partial \vec{H}}{\partial \vec{x}_a})^* \vec{R}^{-1}(\vec{y} - \vec{H}(\vec{x}_b) - \frac{\partial \vec{H}}{\partial \vec{x}_a}\delta\vec{x}_a) \tag{54}$$

$$\int_V d\xi' \vec{B}^{-1}(\xi, \xi')\delta\vec{x}_a(\xi', 0) - \vec{\lambda}(\xi, 0) = \vec{0} \tag{55}$$

$$\frac{\partial \delta\vec{x}_a}{\partial t} + \frac{\partial \vec{M}}{\partial \vec{x}_a}\delta\vec{x}_a = \vec{0} \tag{56}$$

$$\frac{\partial \vec{x}_b}{\partial t} + \vec{M}(\vec{x}_b) = \vec{0} \tag{57}$$

Here the evolution of the background $\vec{x}_b$ is determined by the model operator $\vec{M}$, whereas the (nonlinear) jacobian $\frac{\partial \vec{M}}{\partial \vec{x}_a}$ of the model operator determines the evolution of the analysis increment $\delta\vec{x}_a$. The 4DVAR integration is shown in Fig. 15.

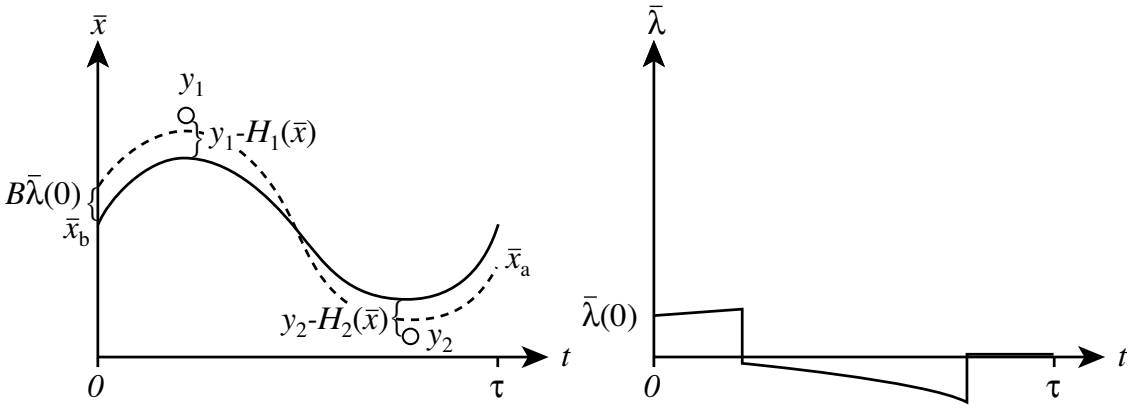

Figure 15: *4DVAR adjoint and forward integration.*

3.2.2 Iterative solution of 4DVAR

An iterative procedure for solving the generalized inverse problem without model errors could be formulated as follows:

1. Integrate $\vec{x}_b$ forwards $0 \rightarrow \tau$, and evaluate the departures $\vec{y} - \vec{H}(\vec{x}_b)$ along the way at the proper time.
2. Integrate $\vec{\lambda}$ backwards $\tau \rightarrow 0$, with $\delta\vec{x}_a = 0$ on the first iteration.
3. Update $\delta\vec{x}_a(\xi, 0)$.
4. Integrate $\delta\vec{x}_a$ forwards $0 \rightarrow \tau$, and evaluate $\frac{\partial\vec{H}}{\partial\vec{x}_a}\delta\vec{x}_a$ along the way at the proper time.
5. Repeat the iteration of $\delta\vec{x}_a$ until the solution converges.

The problem with this iterative procedure is that it will normally fail to converge to the optimal analysis for nonlinear systems. For linear problems however, this procedure will always converge. We discussed earlier, see Fig. 3, that *the nonlinear problem has a non-quadratic costfunction with multiple minima, which we can try to solve as a sequence of linear problems with quadratic costfunctions* by linearizing each problem in the sequence around the solution of the preceding problem. This is the approach which we do apply for solving the 4DVAR problem.

- *Approximation 2:* Solve the nonlinear analysis problem as a sequence of linear analysis problems.

Although there is still no formal guarantee that we will converge towards the optimal analysis, we know that with a good initial estimate of the background and for weakly nonlinear systems we stand a good chance to arrive at a solution which is accurate enough.

3.2.3 4DVAR solved as a sequence of linear problems

We will now show how 4DVAR is solved as a sequence of problems linear in $\delta\vec{x}$. Let us assume $\vec{x}_{i-1}$ is the solution to linear problem $i-1$, starting from $\vec{x}_0 = \vec{x}_b$. To

obtain $\vec{x}_i = \vec{x}_{i-1} + \delta\vec{x}_i$ we formulate a linearized problem for the solution of $\delta\vec{x}_i$. The model and observation operator and their jacobians are made linear in $\delta\vec{x}_i$ by using $\vec{x}_{i-1}$ instead of $\vec{x}_a$ in their expression, giving the following

$$\vec{\lambda}_i(\xi, \tau) = \vec{0} \tag{58}$$

$$-\frac{\partial\vec{\lambda}_i}{\partial t} + (\frac{\partial\vec{M}}{\partial\vec{x}_{i-1}})^*\vec{\lambda}_i = \int_V d\xi' \int_0^\tau dt' (\frac{\partial\vec{H}}{\partial\vec{x}_{i-1}})^* \vec{R}^{-1}(\vec{y} - \vec{H}(\vec{x}_{i-1}) - \frac{\partial\vec{H}}{\partial\vec{x}_{i-1}}\delta\vec{x}_i) \tag{59}$$

$$\int_V d\xi' \vec{B}^{-1}(\xi, \xi')\delta\vec{x}_i(\xi', 0) - \vec{\lambda}_i(\xi, 0) = \vec{0} \tag{60}$$

$$\frac{\partial\delta\vec{x}_i}{\partial t} + \frac{\partial\vec{M}}{\partial\vec{x}_{i-1}}\delta\vec{x}_i = \vec{0} \tag{61}$$

$$\frac{\partial\vec{x}_{i-1}}{\partial t} + \vec{M}(\vec{x}_{i-1}) = \vec{0} \tag{62}$$

Now we can apply the iterative procedure outlined above and converge to a minimum for this linear analysis problem as follows:

1. Integrate $\vec{x}_{i-1}$ forwards $0 \to \tau$, and evaluate the departures $\vec{y} - \vec{H}(\vec{x}_{i-1})$ along the way at the proper time.

2. Solve $\delta\vec{x}_i$ iteratively, using $\delta\vec{x} \neq \delta\vec{x}_i$ in the remaining equations. This is the same approach as used in 3DVAR, and again we can interpret finding $\delta\vec{x}_i$ as minimizing a costfunction J_i whose gradient is given by the initial condition of the equation system,

$$J_i'(\delta\vec{x}) = \int_V d\xi' \vec{B}^{-1}(\xi, \xi')\delta\vec{x}(\xi', 0) - \vec{\lambda}(\xi, 0) \neq \vec{0} \tag{63}$$

   Here $\vec{\lambda}(\xi, 0)$ is the result of transporting all departures $(\vec{y} - \vec{H}(\vec{x}_{i-1}) - \frac{\partial\vec{H}}{\partial\vec{x}_{i-1}}\delta\vec{x})$ back in time to $t = 0$ with linear operators. An important observation is that $\vec{\lambda}(\xi, 0)$ is a linear function of $\delta\vec{x}(\xi, t)$, and even stronger, it can be expressed as a linear function of $\delta\vec{x}(\xi, 0)$ only through the linear forward and backward integrations of the system. This simplifies the costfunction to only depend on $\delta\vec{x}(\xi, 0)$.

When we interpret the above iterative procedure as a minimization of the costfunction J_i with respect to $\delta\vec{x}_i(\xi, 0)$, the forward–backward integrations of $\delta\vec{x}$ and $\vec{\lambda}$ are seen only as means to calculate the gradient J_i'. One advantage of looking at the 4DVAR problem from the generalized inverse point of view is that we see the adjoint model integration as transporting observational information back to the analysis time. The adjoint integration, followed by the forward integration, spreads the influence of a single observation to all times within the assimilation window (see Fig. 16).

- *Example: Inner and outer loops.* In the iterative solution of the 4DVAR problem we often use the terms inner and outer loops. Outer loop refers to the sequence of linear problems which give $\vec{x}_i$. Inner loops refer to the iterations performed to find the solution to each linear problem. Usually only a few

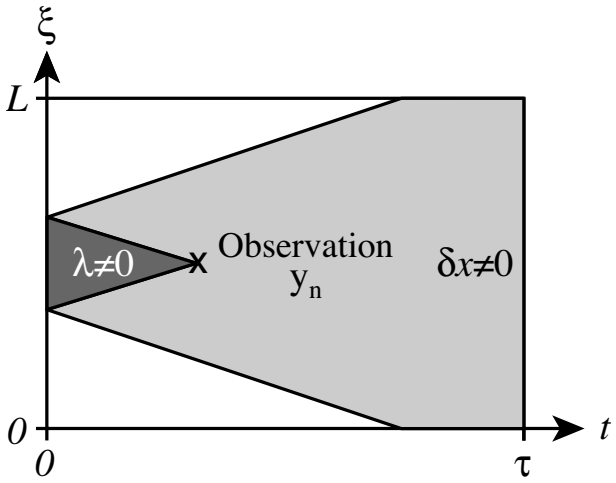

Figure 16: *Each observation influences the solution at all times within the assimilation window in 4DVAR.*

outer loops are enough (2–3), and for each outer loop a few tens of inner loops are performed. Compared to 3DVAR, the consistent use of observations in 4DVAR thus comes at a considerable extra computational cost. The computational cost can be reduced considerably by solving the inner loops (or equivalently, the iterations of the costfunction) at a lower resolution. This procedure is referred to as the *incremental method*. It can of course also be applied to the iterations of the costfunction in 3DVAR.

3.3 The Kalman filter

3.3.1 Cycling the analysis

In the generalized inverse solution of the analysis problem the background field and its error covariance matrix were given to the analysis without any reference to how they were obtained. A more general view of the analysis problem is to include the production of the background field. This process is referred to as *cycling* the analysis. Since background fields are usually obtained as model forecasts from a previous analysis, their error characteristics depend on the errors in the previous analysis as well as the model errors in the forecast. In other words, the background error covariance matrices are evolving from one analysis to the other. The time evolution of the background errors can be important in three ways. First, we can distinguish regions where the background errors were reduced by the use of observations in the previous analysis. Second, we can distinguish regions where the background errors grew because of model errors. Third, the physics and dynamics of the model integration can both amplify and attenuate the background errors. Below we will discuss the Kalman filter which includes the time evolution of the background errors.

As an aside, it is worth pointing out that there are hybrid assimilation algorithms which combine features of 4DVAR and the Kalman filter, for example by including the evolution of the background error in a 4DVAR framework.

3.3.2 Relation to the generalized inverse

Although the Kalman filter is formulated quite differently from the generalized inverse, we can use the present framework to interpret the Kalman filter. The Kalman filter allows for model errors and is in that way close to the generalized inverse. An important difference is that the Kalman filter performs an analysis at each timestep of the (discrete) model, using only the observations available during that timestep $[t_i, t_{i+1}]$. In the generalized inverse all observations in the assimilation window were used simultaneously. If we compare the Kalman filter integration and the 4DVAR integration inside the assimilation window $[0, \tau]$ of 4DVAR, the 4DVAR solution is a continuous curve, whereas the Kalman filter has jumps at each timestep (see Fig. 17). We can interpret the Kalman filter as solving the generalized inverse problem at each timestep of the model, so that the assimilation window shrinks to $[t_i, t_{i+1}]$. In practice, the 4DVAR and the Kalman filter solutions are close to each other at the end of the 4DVAR assimilation window. The solutions are identical in the case of a linear system if there are no model errors and the background errors are the same at the beginning of the assimilation window.

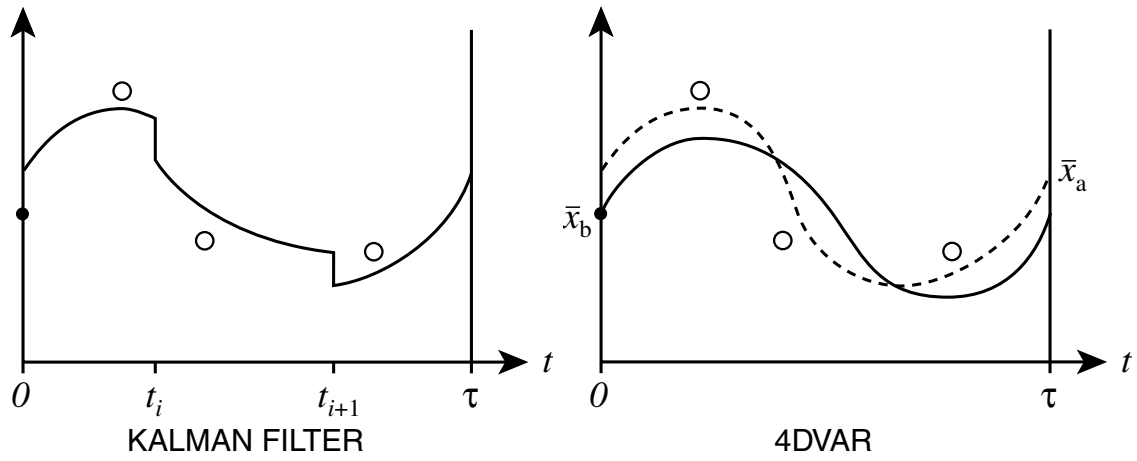

Figure 17: *The difference between the Kalman filter and 4DVAR. The Kalman filter performs an analysis at each model timestep. 4DVAR analyses all observations within a larger assimilation window simultaneously.*

Let us now consider a single timestep of the (discrete) model as the assimilation window, replacing $[t_i, t_{i+1}]$ by $[0, \tau]$. All departures are evaluated at the begin of the timestep $t = 0$. This means the localization function for an observation y_n is $\phi_n(\xi)\delta(t)$, with $\delta(t)$ a Dirac delta function at $t = 0$. This eliminates the adjoint equation from the generalized inverse and $\vec{\lambda}(\xi, 0)$ is obtained directly by evaluating all observation departures at $t = 0$. This gives an equation for the analysis at $t = 0$ which is identical to the OI/3DVAR case with $t = \tau/2$ replaced by $t = 0$,

$$\int_V d\xi' (\vec{B}^{-1}(\xi, \xi')(\vec{x}_a(\xi', 0) - \vec{x}_b(\xi', 0)) - (\frac{\partial \vec{H}}{\partial \vec{x}_a})^* \vec{R}^{-1}(\vec{y} - \vec{H}(\vec{x}_a))) = \vec{0} \qquad (64)$$

As background for the next analysis at $t = \tau$ we make a forecast $\vec{x}_f(\xi, t)$ $0 \rightarrow \tau$ starting from the analysis $\vec{x}_a(\xi, 0)$. In the forecast step of the Kalman filter, the

model error estimate given by the generalized inverse is *not* used,

$$\vec{x}_f(0) = \vec{x}_a(0) \tag{65}$$

$$\frac{\partial \vec{x}_f}{\partial t} + \vec{M}(\vec{x}_f) = \vec{0} \tag{66}$$

Here we have written $\vec{x}(0)$ instead of $\vec{x}(\xi, 0)$ to emphasize the time dimension.

3.3.3 Evolution of the forecast error

The forecast error evolution is obtained by writing $\vec{x}_f = \hat{x} + \vec{\varepsilon}_f$, where $\hat{x}$ is the true (but unknown) state of the system, whose evolution is given by

$$\frac{\partial \hat{x}}{\partial t} + \vec{M}(\hat{x}) = -\delta(t)\vec{\varepsilon}_m \tag{67}$$

Inserting this in the forecast equations gives

$$\vec{\varepsilon}_f(0) = \vec{\varepsilon}_a(0) \tag{68}$$

$$\frac{\partial \vec{\varepsilon}_f}{\partial t} + \frac{\partial \vec{M}}{\partial \hat{x}}\vec{\varepsilon}_f = \delta(t)\vec{\varepsilon}_m \tag{69}$$

We notice that the forecast errors are propagated as perturbations on top of the true model state by the jacobian of the model operator. In the Kalman filter *the forecast errors are considered to evolve linearly during a single timestep of the model.* Thus the jacobian is linearized by replacing $\hat{x}$ by $\vec{x}_a$,

$$\frac{\partial \varepsilon_f}{\partial t} + \frac{\partial \vec{M}}{\partial \vec{x}_a}\vec{\varepsilon}_f = \delta(t)\vec{\varepsilon}_m \tag{70}$$

Now introduce the following notation for the integration of the model equations over one timestep $0 \rightarrow \tau$,

$$\vec{x}_f(\tau) = \mathcal{M}_{0\rightarrow\tau}(\vec{x}_a(0)) \tag{71}$$

$$\vec{\varepsilon}_f(\tau) = \mathbf{M}_{0\rightarrow\tau}\vec{\varepsilon}_a(0) + \vec{\varepsilon}_m(0) \tag{72}$$

where $\mathcal{M}_{0\rightarrow\tau}$ stands for the nonlinear integration and $\mathbf{M}_{0\rightarrow\tau}$ stands for the linearized tangent linear integration. The forecast error covariance at τ, which is the new background error covariance, is obtained by multiplying Eq. 72 by its transpose and making the assumption that *analysis and forecast errors are uncorrelated.* This gives

$$\vec{B}(\tau) = \mathbf{M}_{0\rightarrow\tau}\vec{A}(0)\mathbf{M}^T_{0\rightarrow\tau} + \vec{Q}(0) \tag{73}$$

Here we have used that $\vec{\varepsilon}_f\vec{\varepsilon}_f^T = \vec{B}$ and analogous definitions of $\vec{A}$ and $\vec{Q}$.

3.3.4 The discrete formulation of the Kalman filter

We now come to the proper formulation of the Kalman filter as a discrete assimilation algorithm. The equation for the initial condition is similar to the OI/3DVAR case, with $t = \tau/2$ replaced by $t = 0$. As before, replace $H(\mathbf{x}_a)$ by $H(\mathbf{x}_b) + \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b)$, where $\mathbf{H}$ is a a linear approximation to the jacobian of H (the tangent linear approximation of H). The analysis equations for the Kalman filter are then identical to the 3DVAR/OI case. To emphasize the cycling nature of the Kalman filter, we replace $0 \rightarrow \tau$ by $i \rightarrow i + 1$ in the notation,

$$\mathbf{x}_a(i) = \mathbf{x}_b(i) + \mathbf{K}(i)(\mathbf{y}(i) - H(\mathbf{x}_b(i))) \tag{74}$$

$$\mathbf{K}(i) = \mathbf{B}(i)\mathbf{H}^T(i)(\mathbf{H}(i)\mathbf{B}(i)\mathbf{H}^T(i) + \mathbf{R}(i))^{-1} \tag{75}$$

$$\mathbf{A}(i) = (\mathbf{I} - \mathbf{K}(i)\mathbf{H}(i))\mathbf{B}(i) \tag{76}$$

where $\mathbf{K}$ is the *Kalman gain matrix*. The analysis error covariance matrix $\mathbf{A}$ gives an a posteriori estimate of the quality of the analysis. The forecast to the next step of the analysis gives

$$\mathbf{x}_b(i+1) = \mathcal{M}_{i \rightarrow i+1}(\mathbf{x}_a(i)) \tag{77}$$

$$\mathbf{B}(i+1) = \mathbf{M}_{i \rightarrow i+1}\mathbf{A}(i)\mathbf{M}^T_{i \rightarrow i+1} + \mathbf{Q}(i) \tag{78}$$

The Kalman filter is very expensive to evaluate directly, but several approximations and simplifications can be made to achieve practical assimilation algorithms.

4 Acknowledgements

Thanks to Anabel Bowen and Rob Hine for the artwork.

References

[1] A. F. Bennett. *Inverse Methods in Physical Oceanography*. Cambridge University Press, 1992.

[2] A. F. Bennett. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press, 2002.

[3] A. F. Bennett., B. S. Chua and L. M. Leslie. Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys.*, 60:165–178, 1996.

[4] A. F. Bennett., B. S. Chua and L. M. Leslie. Generalized inversion of a global numerical weather prediction model, II: Analysis and implementation. *Meteor. Atmos. Phys.*, 62:129–140, 1997.

[5] F. Bouttier and P. Courtier. *Data Assimilation Concepts and Methods*. Lecture Notes, ECMWF, 1999.

[6] P. Bergthórsson and B. Döös. Numerical weather map analysis. *Tellus*, 5:329–340, 1955.

[7] P. Courtier. Dual formulation of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, 123:2449–2461, 1997.

[8] R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, 1991.

[9] D. Dee. Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, 131:3323–3343, 2005.

[10] J. Derber and F. Bouttier. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, 51A:195–221, 1999.

[11] M. Ehrendorfer. Prediction of the uncertainty of numerical weather forecasts: problems and approaches. *Proceedings ECMWF Workshop on Predictability*, 27–99, 1997.

[12] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods for forecast error statistics. *J. Geophys. Res.*, 99(C5):10143–10162, 1994.

[13] J. R. Eyre *Variational Assimilation of Remotely-Sensed Observations of the Atmosphere*. ECMWF Tech. Mem. 221, 1995

[14] H. Goldstein. *Classical Mechanics. Second Edition*. Addison-Wesley, 1980.

[15] X. Y. Huang and X. Yang. *Variational Data Assimilation with the Lorenz Model*. Hirlam Technical Report Number 26, 1996.

[16] A. C. Lorenc. Analysis methods for numerical weather prediction. *Mon. Wea. Rev.*, 112:1177–1194, 1986.

[17] W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, Inc., 1984.

[18] A. Tarantola. *Inverse Problem Theory. Methods for data fitting and model parameter estimation*. Elsevier, 1987.

[19] A. Tarantola and B. Valette. Inverse problems = Quest for information. *J. Geophys.*, 50:159–170, 1982.

[20] J. M. T. Thompson and H. B. Stewart. *Nonlinear Dynamics and Chaos*. John Wiley and Sons, 1986.