Implementation issues in variational data assimilation

A.S. Lawless

National Centre for Earth Observation
University of Reading

Aim

The aim of this talk is to give you a tour of some of the issues associated with implementing data assimilation schemes. Details can be followed up in questions or in discussions tomorrow.

Aim

The aim of this talk is to give you a tour of some of the issues associated with implementing data assimilation schemes. Details can be followed up in questions or in discussions tomorrow.

You don't have to deal with all these issues before building a data assimilation scheme, but it is a good idea to have them in mind.

Table of contents

3D-Var and 4D-Var

In variational data assimilation we must minimize cost functions of the form

- Three-dimensional variational data assimilation (3D-Var)

$$\mathcal{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x}))$$

- Four-dimensional variational data assimilation (4D-Var)

$$\mathcal{J}(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \sum_{i=0}^{n} (\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i))^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i))$$

Minimization and convergence

In order to minimize the cost function we need to consider

1. Minimization algorithm
2. Stopping criterion
3. Gradient calculation
4. Preconditioning

We treat each of these in turn.

Descent algorithms

There are many different minimization algorithms. The most practical to use are those that require only the first derivative of the cost function, such as quasi-Newton or conjugate gradient.

Descent algorithms

There are many different minimization algorithms. The most practical to use are those that require only the first derivative of the cost function, such as quasi-Newton or conjugate gradient.

Numerical optimization is a huge field and so it is much better to take a package written by somebody else than to code your own minimization routine.
Do not try to write your own!

- For routines in Fortran or C++ see the Guide to Available Mathematical Software
  http://gams.nist.gov/
  (look under unconstrained minimization)
- For routines in Matlab a good place to start is the Stanford site
  http://www.stanford.edu/group/SOL/software.html

Stopping the minimization

The iterative process is stopped when some stopping criterion or termination criterion is satisfied. Common criteria are to test when one or more of the following is less than a given tolerance

- Change in function between two iterates

$$|\mathcal{J}(\mathbf{x}^{(k+1)}) - \mathcal{J}(\mathbf{x}^{(k)})| < tol$$

- Change in value of state

$$||\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}|| < tol$$

- The norm of the gradient

$$||\nabla\mathcal{J}(\mathbf{x}^{(k)})|| < tol$$

- The ratio between the norm of the gradient at the current iterate and its initial value

$$||\nabla\mathcal{J}(\mathbf{x}^{(k)})||/||\nabla\mathcal{J}(\mathbf{x}^{(0)})|| < tol$$

Note that we cannot satisfy these to the same tolerance. If we satisfy the change in function to accuracy ϵ then the change in state will be of the order $\sqrt{\epsilon}$ and the gradient will be of order $\epsilon^{1/3}$.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Note that we cannot satisfy these to the same tolerance. If we satisfy the change in function to accuracy ϵ then the change in state will be of the order $\sqrt{\epsilon}$ and the gradient will be of order $\epsilon^{1/3}$.

Note also that the tolerance we can achieve will depend on the noise on our inputs.

Gradient calculation

In order to calculate the gradient of the cost function we usually require the adjoint of the observation operators (and of the model in 4D-Var).

To form these we usually use the method of deriving the adjoint model from the nonlinear model source code (automatic differentiation).

It is very important to test the gradient calculation. If the gradient code is wrong the minimization will not converge!

Conditioning

When using an iterative method the speed of convergence and accuracy of the solution after a given time is dependent on the condition number of the problem.

Preconditioning is a technique to reduce the condition number of the problem first.

Preconditioning by variable transformation

Consider again the 3D-Var cost function

$$\mathcal{J}(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x}))$$

The matrix $\mathbf{B}$ is badly conditioned. If we define uncorrelated variables $\mathbf{v}$ such that

$$\mathbf{v} = \mathbf{B}^{-1/2}(\mathbf{x}-\mathbf{x}_b)$$

then we have

$$\hat{\mathcal{J}}(\mathbf{v}) = \mathbf{v}^T\mathbf{v} + (\mathbf{y}-\mathcal{H}(\mathbf{B}^{1/2}\mathbf{v}+\mathbf{x}_b))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{B}^{1/2}\mathbf{v}+\mathbf{x}_b))$$

which is better conditioned.

Other preconditioning

Other preconditioning is possible and relies on using knowledge of the Hessian of the cost function.
For example, ECMWF use a Lanczos algorithm to minimize the cost function and save the leading eigenvectors to precondition the next assimilation.

Factors affecting conditioning

The conditioning of the problem has been shown to be worse for

- certain correlation matrices (e.g. Gaussian);
- large correlation length scales in the **B** matrix;
- closely spaced observations;
- more accurate observations.

(Haben, Lawless and Nichols, 2011)

Choice of prior constraint

So far we have assumed a prior constraint of the form

$$(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b)$$

We have several choices for the implementation of this term:

- Variable transformation to model $\mathbf{B}$.
- Choice of spatial correlation function e.g. Gaussian, SOAR, Laplacian (for $\mathbf{B}^{-1}$).
- What is $\mathbf{x}_b$ for your problem?

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Spatial correlation functions

The spatial correlation functions in the prior constraint require the specification of variances and length scales for each variable. These must be estimated from knowledge of your prior.

If the prior $\mathbf{x}_b$ is a short term forecast, then two common methods are

- Generate statistics from a large sample of forecast differences.
- Generate statistics from an ensemble assimilation.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Other constraints

Other constraints may be added to the cost function. These may be

- strong constraints, which have to be exactly satisfied.
- weak constraints, which may be only approximately satsfied.

Examples

1. Spatial smoothing term - Penalise second derivative of field, e.g.

$$\alpha||\nabla^2 \mathbf{x}_0||$$

Examples

1. Spatial smoothing term - Penalise second derivative of field, e.g.

$$\alpha||\nabla^2 \mathbf{x}_0||$$

2. Digital filter constraint - Penalise distance to filtered solution, e.g.

$$\alpha \sum_{i=0}^{n} (\mathbf{x}_i - \mathbf{x}_i^f)^T (\mathbf{x}_i - \mathbf{x}_i^f)$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Examples

1. Spatial smoothing term - Penalise second derivative of field, e.g.

$$\alpha||\nabla^2 \mathbf{x}_0||$$

2. Digital filter constraint - Penalise distance to filtered solution, e.g.

$$\alpha \sum_{i=0}^{n} (\mathbf{x}_i - \mathbf{x}_i^f)^T (\mathbf{x}_i - \mathbf{x}_i^f)$$

3. In a standard 4D-Var the model is imposed as a strong constraint, i.e. the states must satisfy the model equations.

Examples

1. Spatial smoothing term - Penalise second derivative of field, e.g.
$$\alpha||\nabla^2 \mathbf{x}_0||$$
2. Digital filter constraint - Penalise distance to filtered solution, e.g.
$$\alpha \sum_{i=0}^{n} (\mathbf{x}_i - \mathbf{x}_i^f)^T (\mathbf{x}_i - \mathbf{x}_i^f)$$
3. In a standard 4D-Var the model is imposed as a strong constraint, i.e. the states must satisfy the model equations.
4. In weak constraint 4D-Var we add a term to the cost function of the form
$$(\mathbf{x}_{i+1} - \mathcal{M}(\mathbf{x}_i))^T \mathbf{Q}^{-1} (\mathbf{x}_{i+1} - \mathcal{M}(\mathbf{x}_i))$$

Observation errors

Recall the 3DVar cost function

$$\mathcal{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x}))$$

How do we represent the observation error covariance matrix **R**?

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Observation error covariance matrix

The observation error covariance matrix represents several different sources of error

1. Instrument errors
2. Observation preprocessing errors
3. Forward model errors
4. Representativity errors

In practice these are not always considered separately.

Instrument errors

Any measuring instrument will only measure to a certain accuracy. The known accuracy of the instrument (variance) must be included in the $\mathbf{R}$ matrix.

Observation preprocessing errors

Sometimes we assimilate observations that have been preprocessed in some way. For example, we may process a satellite radiance measurement to obtain a derived quantity, such as temperature, and then assimilate that. This introduces extra errors into the measurement.

Forward model errors

In order to compare an observation with the model forecast we use the observation operator $\mathcal{H}$. In theory the matrix $\mathbf{R}$ should account for errors in this operator.

Representativity errors

Representativity errors are errors that arise when observations can resolve scales that the model cannot.

Observation error correlations

The off-diagonal elements of the **R** matrix represent correlations between observation errors. These arise due to the errors discussed previously. How can we specify these?

Observation error correlations

Common methods for dealing with observation error correlations are

- Thin the data and assume **R** is diagonal (especially used with satellite data).

Observation error correlations

Common methods for dealing with observation error correlations are

- Thin the data and assume **R** is diagonal (especially used with satellite data).
- Assume **R** is diagonal but increase the variance to allow for the fact that it is not.

Observation error correlations

Common methods for dealing with observation error correlations are

- Thin the data and assume **R** is diagonal (especially used with satellite data).
- Assume **R** is diagonal but increase the variance to allow for the fact that it is not.
- Use of a Markov matrix. This has a tridiagonal inverse in 1D and so makes $\mathbf{R}^{-1}$ easy to represent.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Observation error correlations

Common methods for dealing with observation error correlations are

- Thin the data and assume $\mathbf{R}$ is diagonal (especially used with satellite data).
- Assume $\mathbf{R}$ is diagonal but increase the variance to allow for the fact that it is not.
- Use of a Markov matrix. This has a tridiagonal inverse in 1D and so makes $\mathbf{R}^{-1}$ easy to represent.
- By the leading eigenvectors of the matrix.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Observation error correlations

Common methods for dealing with observation error correlations are

- Thin the data and assume **R** is diagonal (especially used with satellite data).
- Assume **R** is diagonal but increase the variance to allow for the fact that it is not.
- Use of a Markov matrix. This has a tridiagonal inverse in 1D and so makes $\mathbf{R}^{-1}$ easy to represent.
- By the leading eigenvectors of the matrix.

Observation error correlations

Common methods for dealing with observation error correlations are

- Thin the data and assume **R** is diagonal (especially used with satellite data).
- Assume **R** is diagonal but increase the variance to allow for the fact that it is not.
- Use of a Markov matrix. This has a tridiagonal inverse in 1D and so makes $\mathbf{R}^{-1}$ easy to represent.
- By the leading eigenvectors of the matrix.

(Stewart, 2010, PhD thesis)

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

References for minimization and convergence

- Gill, P.E., Murrey, W. and Wright, M.H. (1982), *Practical Optimization*, Emerald Group Publishing Limited.
- Haben, S., Lawless, A.S. and Nichols, N.K. (2011), Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus*, in press.
- Katz, D., Lawless, A.S., Nichols, N.K., Cullen, M.J.P. and Bannister, R.N. (2011), Correlations of control variables in variational data assimilation. *Quart. J. Royal Met. Soc.*, 137, 620–630.
- Tshimanga, J., Gratton, S., Weaver, A. T. and Sartenaer, A. (2008), Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. *Quart. J. Royal Met. Soc.*, 134: 751–769.

References for constraints

- Nichols, N.K. (2009), Mathematical Concepts of Data Assimilation. In: Data Assimilation: Making Sense of Observations, eds. W Lahoz, R Swinbank and B Khattatov, Springer.
- Polavarapu, S., Tanguay, M. and Fillion, L. (2000), Four-Dimensional Variational Data Assimilation with Digital Filter Initialization. *Mon. Wea. Rev.*, 128, 2491–2510.
- Trémolet, Y. (2007), Model-error estimation in 4D-Var. *Quart. J. Royal Met. Soc.*, 133: 1267–1280
- Watkinson, L.R., Lawless, A.S., Nichols, N.K. and Roulstone, I. (2007), Weak constraints in four-dimensional variational data assimilation. *Met. Zeit.*, 16, 767–776.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

References for observation errors

- Fisher, M. (2005), Accounting for correlated observation error in the ECMWF analysis. ECMWF Technical Memoranda, MF/05106.
- Healy, S.B. and White, A.A. (2005), Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quart. J. Royal Met. Soc.*, 131:63–72.
- Liu, Z.-Q., and Rabier, F. (2002), The interaction between model resolution, observation resolution and observation density in data assimilation: A one- dimensional study. *Quart. J. Royal Met. Soc.*, 128, 1367–1386.
- Stewart, L.M. (2010), Correlated observation errors in data assimilation, PhD thesis, Dept of Mathematics, University of Reading.