The University of Reading

Variational Data Assimilation for Hamiltonian Problems

L.R. Stanton[1], A. Lawless[1], N.K. Nichols[1] and I. Roulstone[2]

NUMERICAL ANALYSIS REPORT 4/04

[1] *Department of Mathematics*
The University of Reading
Whiteknights, PO Box 220
Reading
Berkshire RG6 6AX

[2] *Department of Mathematics and Statistics*
School of Electronics and Physical Sciences
University of Surrey
Guilford
Surrey GU2 7XH

Paper presented at ICFD conference on Numerical Methods for Fluid Dynamics, Oxford, UK,
29 March - 1 April 2004.

Department of Mathematics

Abstract

We investigate the conservation properties of Hamiltonian systems in variational data assimilation. We set up a four dimensional data assimilation scheme for the two-body (Kepler) system using a symplectic scheme to model the non-linear problem. We use our completed scheme to investigate the observability of the system and the effect of different background constraints. We find that the addition of these constraints gives an improved solution for the cases we have investigated.

1 Introduction

Some features of atmospheric dynamics can be modelled using Hamiltonian methods. We investigate whether the conservation properties of such systems can be exploited when using data assimilation schemes. To do this we set up a full 4d variational (4D-Var) data assimilation scheme for the simpler problem of planetary orbits, which also has a Hamiltonian structure. The first section introduces variational data assimilation. We then discuss the two-body problem in its continuous and discrete form. Section 3 shows the results of our assimilation experiments.

1.1 4D Variational Data Assimilation

Data assimilation involves the integration of observations into a model to give a state that most accurately describes reality. These methods are used in numerical weather prediction where there is considerable amounts of data, and a very large state vector. A direct solution of the data assimilation problem would involve the inversion of a matrix that is too big to be achieved computationally. Thus many data assimilation schemes attempt to find ways to approximate the problem [1]. *Variational* data assimilation methods are optimisation problems where we find the optimal state that minimises an objective function, J, at the initial time. J has an observation term that measures the departures between the observations and the model state for all observations over a given assimilation time window. In addition it may often have a background term that accounts for the departure between a known background state and the model state at the initial time.

4D-Var data assimilation includes data that is distributed in time *and* space. If we initially assume that the only contribution to J is given by the observation term, we have

$$J(\mathbf{x}) = \sum_{n=0}^{N} (\mathbf{y}_n - H_n[\mathbf{x}_n])^T \mathbf{R}_n^{-1} (\mathbf{y}_n - H_n[\mathbf{x}_n]). \qquad (1)$$

Here n denotes quantities at time n; $\mathbf{y}_n$ are the observations and $\mathbf{x}_n$ the model state. H_n is the observation operator which transforms the model state to that of the observation. The matrix $\mathbf{R}_n$ is the observation error covariance matrix describing statistical information about the errors in the observations.

The minimisation of equation (1) is subject to the strong constraint that the model states, $\mathbf{x}_n$, are a solution to the numerical model. In addition the minimisation requires that the tangent linear hypothesis holds. This hypothesis states that the forward model can be linearised, and that the resulting linear model exhibits the same local behaviour as the original.

The minimisation of J requires an optimisation algorithm which requires the calculation of both J and its gradient, ∇J at each iteration. The gradient of the observation term can be found by running the adjoint model backwards, where the adjoint can be found as the transpose of the tangent linear model, but is more generally derived directly from the code of the linear model [2]. Thus to minimise J we require the non-linear forward model and the adjoint. However to find the adjoint we also need to find the tangent linear model.

In this paper we investigate whether we can retrieve the true state by using a good conservation method for the forward model of a 4D-Var scheme. We then compare two methods of including a background constraint. For both we use the same background information at the initial time. In the first case we constrain the background model state vector directly. In the second case this information is transformed to an energy thus using the Hamiltonian property as a constraint. For the first of these we add to the cost function a term of the form $\alpha_1(\mathbf{x}_{b_0} - \mathbf{x}_0)(\mathbf{x}_{b_0} - \mathbf{x}_0)^T$, for the second the constraint term is $\alpha_2(E(\mathbf{x}_{b_0}) - E(\mathbf{x}_0))^2$. Here α_1 and α_2 are parameters that allow the effect of each of the terms to be controlled.

2 Modelling the Two-Body Problem

2.1 The Continuous Problem

The two-body problem is one of the simplest Hamiltonian problems. Instead of writing the system as two particles of mass m_1 and m_2 in mutual orbit, we can set the origin at the centre of mass. This reduces the problem to one particle of reduced mass μ, where $\mu = \frac{m_1 m_2}{m_1+m_2}$, in orbit around one particle of total mass, m_1+m_2. Our continuous equations of motion can therefore be written as two first order, non-dimensional equations describing the evolution of position, $\mathbf{q} = (q_1, q_2)$, and momentum, $\mathbf{p} = (p_1, p_2)$,

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \mathbf{p} \qquad\qquad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\mathbf{q}}{(q_1^2 + q_2^2)^{\frac{3}{2}}}. \tag{2}$$

2.1.1 Conservation Properties

The two-body problem has two conserved quantities, the Hamiltonian, E, which for this problem is the total energy, and the angular momentum, L. These are given by,

$$E = \frac{1}{2}\left(p_1^2 + p_2^2\right) - \frac{1}{(q_1^2 + q_2^2)^{\frac{1}{2}}} = \text{constant} \qquad\qquad L = q_1 p_2 - p_1 q_2 = \text{constant}. \tag{3}$$

These characteristics are intrinsic to the physical problem, and will provide a useful test of the discretised equations.

2.2 The Discrete Problem

To test the effect of these conservation properties in the 4D-Var scheme it is essential that they are captured by the discrete model. In recent years *geometric integration* has

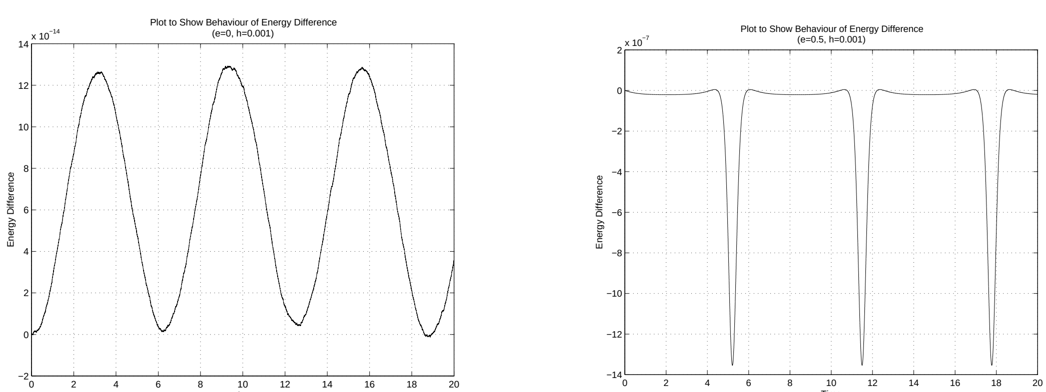

Figure 1: Difference between the energy given by the model and the true energy for (a) eccentricity, $e = 0$, and (b) $e = 0.5$

attempted to address the issue of preserving global features, and in particular *symplectic methods* are particularly good at conserving energy, and can also conserve angular momentum [3]. Following previous work on the two-body problem we use a second order, symplectic, Runge-Kutta scheme known as the Störmer-Verlet method [4].

Figures 1(a) and 1(b) illustrate how well the model captures the energy conservation for a circle (where eccentricity, $e = 0$) and an ellipse with $e = 0.5$. These figures show the difference between the energy given by the model and the truth. For a circle this difference is at a scale of 10^{-14}, and so here the scheme does well. However we see that if we increase the eccentricity the deviation increases around the point of closest approach. These deviations can be explained by considering the second of Kepler's three laws - a line joining the orbitting body and the central body will sweep out equal areas in equal times. Hence at closest approach the body will have a greater velocity. As we are using a fixed step method to model the problem, this means the trajectory will be modelled by fewer steps at this point giving a less accurate solution [3].

After finding a suitable non-linear forward model, we now use this to find our tangent linear model. From this we will be able to derive our adjoint and thus calculate the gradient of the cost function. To produce our linear model we linearise the discrete equations. We then test both the code and the validity of the tangent linear hypothesis using standard methods [5]. The *validity time* is a measure of how long the linear model is a good approximation to the non-linear model. To test this, we track the evolution of a perturbation in both models. We find that for a circle the validity time is reasonably long, whereas if we increase the eccentricity the validity time is reduced, suggesting that the more eccentric ellipses exhibit more non-linear behaviour. Once we have the linear model we can construct the adjoint directly from the code. This is then tested using standard techniques [5]. We are now able to set up the 4D-Var scheme using a quasi-Newton iterative scheme to find the optimal state.

3 Assimilation Experiments

For our investigation we carry out identical twin experiments. For an identical twin experiment the observations are generated by the forward non-linear model - this allows us to know the true solution. We assimilate these observations using the scheme starting from an incorrect initial guess. If we have perfect observations, we should thus be able to retrieve the true solution. These 'observations' can then be made more realistic by adding noise to them. For our experiments we add noise with a Gaussian distribution with variance 10^{-4} and no bias.

3.1 Observability

In general we do not have observations of all of the variables at every timestep. We investigate whether we still obtain a good solution if we use fewer observations by looking at the *observability* of the system. For observations at two timesteps the observability matrix is given by,

$$\tilde{\mathbf{H}} = \begin{pmatrix} \mathbf{H} \\ \mathbf{HM} \end{pmatrix} \tag{4}$$

where $\mathbf{H}$ is the linearised observation operator and $\mathbf{M}$ is the linear model. If the matrix $\tilde{\mathbf{H}}$ has full column rank then the system is said to be observable, that is we can construct

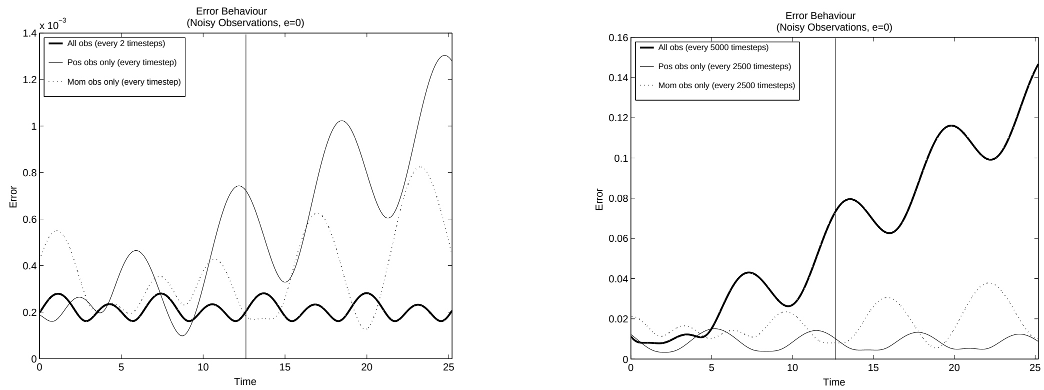

Figure 2: Error between the optimal solution and the truth for $e = 0$ and assimilation window $t = 12.6$, for (a) dense observations, and (b) sparse observations

the solution from the given observations. This limited analysis suggests that the system is fully observable if we have any two of the four variables as observations.

To look at this further we run our 4D-Var scheme using observations in all four variables, position observations only, and finally observations of momentum only. In the first instance we use observations at every other timestep where all four variables are used, and every timestep for the other two cases, so that we are using the same number of observations in each case. Our data assimilation window has length $t = 12.6$, with timestep, $\Delta t = 0.001$. Figure 2(a) compares the error in the trajectories of the three cases over the data assimilation window, and a subsequent forecast.

We see that over the data assimilation window the error for each of the three cases is of a similar magnitude. However where we have used only observations of position in the 4D-Var scheme, the error in the forecast is diverging. This contradicts the result found previously using the observability analysis for two timesteps only - although we have used observations in two of the four variables we have not reconstructed the true solution. This suggests that although we are using a good energy conserving model, the data assimilation scheme does not always produce a good solution. Thus we may wish to impose an explicit constraint on the Hamiltonian, E, to improve the forecast.

3.2 Constraints

To improve our solution we add a constraint to the cost function. Typically this is done using the background information directly, as discussed in section 1.1. Alternatively we transform this information to the background energy, and use the Hamiltonian as an explicit constraint. We now compare these constraints using sparse observations so that their effect is more easily seen. We repeat the experiment as in section 3.1 this time using observations every 5000 timesteps where all variables are observed, and every 2500 timesteps where only two of the four are observed, thus assimilating 12 observations. Figure 2(b) illustrates these results *without* using any constraint. We see that when we use observations of all variables but at fewer observation times, the solution is diverging. This suggests it is better to have more frequent observations, even if they are of position or momentum only. We use this diverging case to investigate the effect of constraints. In both cases we use perfect background information.

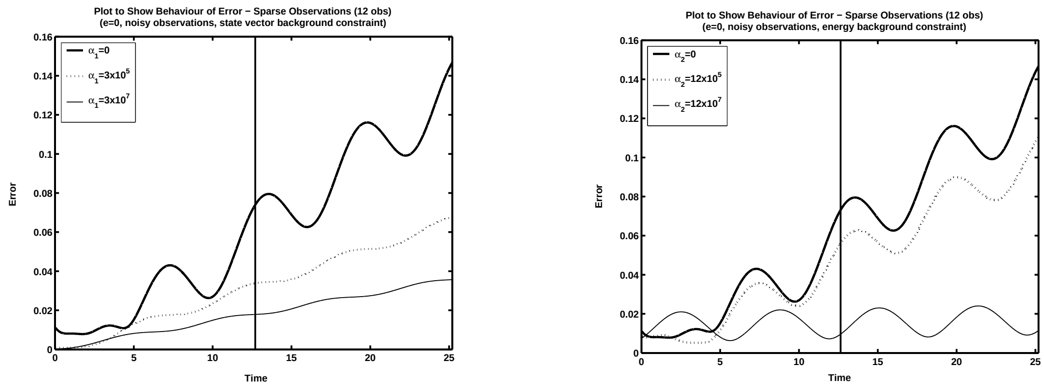

Figure 3: Error between the optimal solution and the truth using sparse observations of momentum with (a) a background constraint, and (b) an energy constraint ($e = 0$, assimilation window $t = 12.6$)

3.2.1 Constraint using $\mathbf{x}_b$

Here we alter the cost function and its gradient to include a background term. This means that the optimal solution must fit the observations and remain close to a background term produced by a previous forecast ($\mathbf{x}_{b_0}$). Figure 3(a) shows the effect of including a background constraint for the case where we have observations every 5000 timesteps. We note that there is an improvement everywhere over both the data assimilation window and the forecast, although the error is still diverging even with a large value of α_1.

3.2.2 Constraint using $E(\mathbf{x}_b)$

Here we constrain the energy of the assimilation solution to be close to the energy of the background used in section 3.2.1, $E(\mathbf{x}_{b_0})$. Figure 3(b) shows the effect of the energy constraint using the same observations as in section 3.2.1. Here we see that for large α_2 although the solution at the beginning of the assimilation window is worse than without any constraint, the forecast is considerably improved. In addition the error diverges less for this constraint than for that illustrated by figure 3a.

4 Conclusions

We have seen that when producing a model for this simple Hamiltonian system it is possible to find methods that are very good at preserving the physical characteristics of the problem - energy and angular momentum conservation. However in spite of this the data assimilation scheme produced using these methods does not produce a good solution in all cases. We have investigated whether these results can be improved using constraints, and have found that if we use our background state directly as a constraint we notice an improvement in the solution. If we transform this background information to the Hamiltonian property, the forecast is further improved.

Acknowledgements

I would like to thank Prof M. Baines and Prof M. Cullen for their help with this research.

References

[1] Nichols NK. Data assimilation: aims and basic concepts. *Data Assimilation for the Earth System* Swinbank R, Shutyaev V, Lahoz WA (eds). Kluwer Academic, 2003; 9-20

[2] Chao WC, Chang L. Development of a four-dimensional variational analysis system using the adjoint method at GLA. Part1: Dynamics. *Monthly Weather Review* August 1992; 1661-1673

[3] Budd CJ, Piggott MD. Geometric integration and its applications. *Proceedings of the ECMWF Workshop on Developments in Numerical Methods for Very High Resolution Global Models* 5-7 June 2000; 93-117

[4] Skeel RD, Biesiadecki JJ. Symplectic integration with variable stepsize. *Annals of Numerical Mathematics* 1994; **1**: 1-9

[5] Li Y, Navon MI, Yang W, Zou X, Bates JR, Moorthi S, Higgins RW. Four-dimensional variational data assimilation experiments with a multilevel semi-Lagrangian semi-implicit general circulation model. *Monthly Weather Review* May 1994; 966-983