INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2005; **47**:1361–1367
Published online 12 January 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.844

Variational data assimilation for Hamiltonian problems

L. R. Watkinson[1,*,†], A. S. Lawless[1], N. K. Nichols[1] and I. Roulstone[2]

[1]*Department of Mathematics, The University of Reading, U.K.*
[2]*Department of Mathematics and Statistics, The University of Surrey, U.K.*

SUMMARY

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. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: data assimilation; Hamiltonian; conservation; two-body

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 four-dimensional variational (4D-Var) data assimilation scheme for the simpler problem of planetary orbits, which also has a Hamiltonian structure.

Data assimilation involves the integration of observations into a model to give a state that most accurately describes reality. This approach is used in numerical weather prediction where there is a very large state vector and many observations—a direct solution of the data assimilation problem cannot be achieved computationally. Thus, many data assimilation schemes attempt to find ways to approximate the problem [1]. For *variational* data assimilation methods our solution is obtained by finding the optimal state of an objective function, J, at the initial time. J includes an observation term measuring the departures between the observations

*Correspondence to: Laura R. Watkinson, Department of Mathematics, The University of Reading, P.O. Box 220, Whiteknights, RG6 6AX Reading, U.K.
†E-mail: l.r.stanton@reading.ac.uk

Contract/grant sponsor: EPSRC
Contract/grant sponsor: Met Office U.K.

Received 27 April 2004
Revised 10 November 2004
Accepted 18 November 2004

and the model state for all observations over a given assimilation time window. In addition, it usually has a background term that accounts for the departure between a known background state and the model state at the initial time. This term is required where the problem is very large, and not enough information can be gained from the observations alone. For our problem a background term is not *necessary* as the problem is sufficiently small.

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

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

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

The optimization 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, it is necessary that the forward model can be linearized, and that the resulting linear model exhibits the same local behaviour as the original. This is known as the tangent linear hypothesis.

Here we find the optimal state using a quasi-Newton optimization algorithm that requires the calculation of J and its gradient, ∇J at each iteration. The gradient of the observation term can be found by running the adjoint model backwards in time. The adjoint is determined from the transpose of the tangent linear model, but is more generally derived directly from the code of the linear model [2]. Thus to optimize J we require the non-linear forward model and the adjoint. The latter, however, also requires the derivation of 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. Where this in itself does not produce a good solution, we investigate the effect of adding two different constraints. For the first we add a term to the objective function that constrains the optimal state vector to be close to the background state vector. This term has the form $\alpha_1(\mathbf{x}_b - \mathbf{x}_0)(\mathbf{x}_b - \mathbf{x}_0)^{\mathrm{T}}$. Here $\mathbf{x}_b$ and $\mathbf{x}_0$ are the background state and the model state, respectively, both at the initial time. For our second case, instead of constraining the state vector directly, we add a term to the objective function such that the *energy* of the optimal state is close to the *energy* of the background. Here the constraint term is $\alpha_2(E(\mathbf{x}_b) - E(\mathbf{x}_0))^2$. The parameters α_1 and α_2 allow the weight of each of the terms to be controlled.

2. MODELLING THE TWO-BODY PROBLEM

2.1. The continuous problem

The Kepler problem describing the motion of two bodies, m_1 and m_2, in mutual orbit is one of the simplest examples of a Hamiltonian system. By setting the origin at the centre of mass we can reduce the problem to that of one particle of mass $\mu = m_1 m_2/(m_1 + m_2)$, 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, $\mathbf{p}=(p_1,p_2)$,

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

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

$$E=\frac{1}{2}(p_1^2+p_2^2)-\frac{1}{(q_1^2+q_2^2)^{1/2}}=\text{constant}, \quad L=q_1p_2-p_1q_2=\text{constant} \tag{3}$$

These characteristics are intrinsic to the physical problem and provide a useful test of the numerical model.

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 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]. Our initial conditions are given by

$$\mathbf{q}_0=(1-e,0), \quad \mathbf{p}_0\left(0,\sqrt{\frac{1+e}{1-e}}\right) \tag{4}$$

corresponding to an elliptical trajectory with eccentricity e, semi-major axis $a=1$ and energy $E=-0.5$.

We test the non-linear model by looking at the difference between the energy given by the model and the true energy for the given initial conditions. For a circle this difference is of the order 10^{-14}—here the scheme does well. If we increase the eccentricity, the difference between the truth and the model trajectory 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 orbiting 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 linearize the discrete scheme to find our tangent linear model. From this we are able to derive our adjoint and thus calculate the gradient of the cost function. 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 long, whereas if we increase the eccentricity the validity time is reduced, suggesting that the more eccentric ellipses exhibit more non-linear behaviour. From the linear model we 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 method to find the optimal state. For this optimization method we use the following stopping criteria [6]:

$$\frac{J_{k-1} - J_k}{1 + |J_k|} < \varepsilon, \quad \frac{\|\mathbf{x}_{k-1} - \mathbf{x}_k\|}{1 + \|\mathbf{x}_k\|} < \sqrt{\varepsilon}, \quad \frac{\|\nabla J_k\|}{1 + |J_k|} \leqslant \sqrt[3]{\varepsilon} \tag{5}$$

For our experiments all three criteria must be met, and we use a value of $\varepsilon = 10^{-6}$.

3. ASSIMILATION EXPERIMENTS

For our investigation we carry out identical twin experiments. Here, synthetic observations are generated by the forward, non-linear model, allowing us to know the true solution. These 'observations' can be made more realistic by adding noise to them. Here 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

$$\tilde{\mathbf{H}} = \begin{pmatrix} \mathbf{H} \\ \mathbf{HM} \end{pmatrix} \tag{6}$$

where $\mathbf{H}$ is the linearized observation operator and $\mathbf{M}$ is the linear model. If matrix $\tilde{\mathbf{H}}$ has full column rank then the system is said to be observable, that is, we can construct 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. Initially 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. For all our experiments, the data assimilation window has length $t = 12.6$, with timestep, $\Delta t = 0.001$. Figure 1(a) compares the error between the truth and the model trajectories of the three cases over the data assimilation window and a subsequent forecast. In this figure, and in Figure 1(b), the thick solid line denotes observations in all variables, the thin solid line observations in position only and the dotted line observations of momentum only.

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. We therefore consider the addition of an explicit constraint.

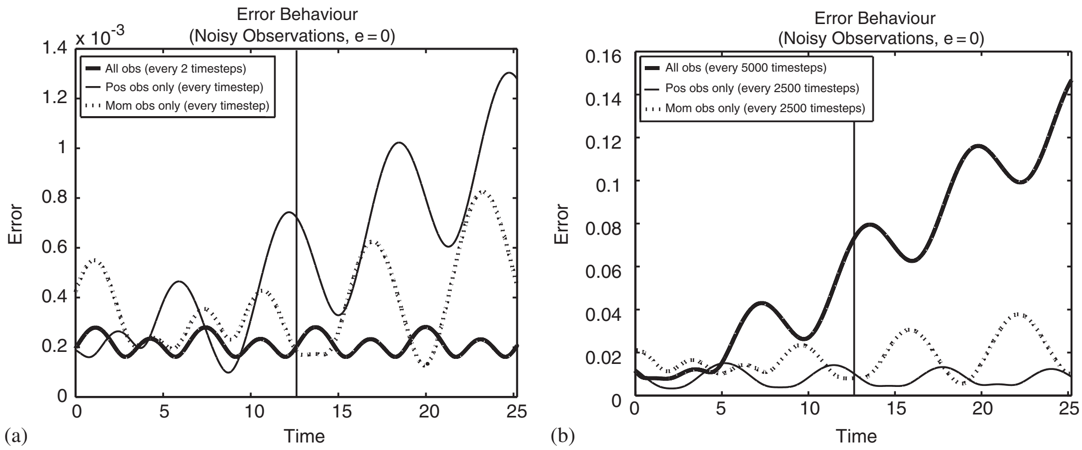

Figure 1. Error in trajectory between the optimal solution and the truth for $e = 0$ and assimilation window $t = 12.6$, for: (a) dense observations; and (b) sparse observations.

3.2. Constraints

Generally 4D-Var problems include a background term, as discussed in Section 1. Typically this uses the background state vector directly and so our first constraint is of this form. We also investigate the use of the Hamiltonian as an explicit constraint by adding a term that constrains the energy of the solution to be close to the energy of the background. We 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 1(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 the truth as the background. Our choice of the parameters α_1 and α_2, controlling the weights of the constraints, are chosen such that these are approximately in balance with the observation term in the cost function.

Constraint using $\mathbf{x}_b$: Here the optimal solution must fit the observations and remain close to the background state vector. Figure 2(a) shows the effect of including a background constraint where we have observations every 5000 timesteps. Here, and in Figure 2(b), the solid line and dotted line show the unconstrained and constrained results, respectively. There is improvement everywhere over the data assimilation window and the forecast but the error is still diverging.

Constraint using $E(\mathbf{x}_b)$: Using the same observations, we constrain the energy of the assimilation solution to be close to the energy of the background. In Figure 2(b), we can see that, while the solution at the beginning of the window is worse than without the constraint, the forecast is considerably improved. In addition the error diverges less than in Figure 2(a).

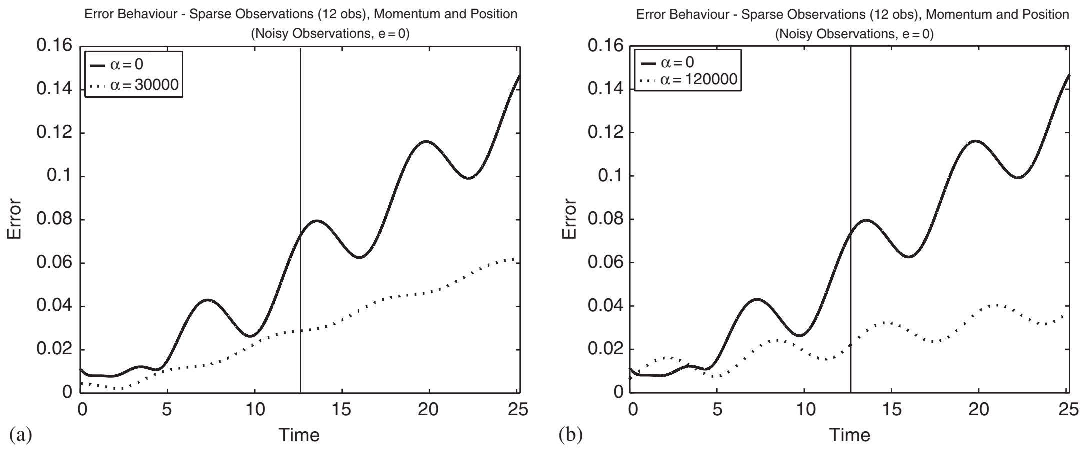

Figure 2. Error in trajectory between the optimal solution and the truth using sparse observations of momentum with: (a) a background constraint, $\alpha = 30\,000$; and (b) an energy constraint, $\alpha = 120\,000$ ($e = 0$, assimilation window $t = 12.6$).

4. CONCLUSIONS

We have seen that when producing a numerical 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 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 instead use a constraint on the energy, the forecast is further improved.

However, further investigation into the nature of the diverging forecast suggests that this behaviour is a result of the geometry of the problem. If the assimilation trajectory has a different semi-major axis to the truth, then the period of the orbit will be different (Kepler's 3rd law). Thus, the two trajectories will move out of and then into phase periodically and the error will increase and decrease, respectively. This effect is confirmed by producing a long forecast for the unconstrained example in Figure 2—after 500 orbits the error reaches a maximum of approximately 2.8 then begins to decrease. We have seen that adding a constraint on the energy reduces this divergence. Because any ellipse with the same energy will have the same semi-major axis and period [7], a constraint on the energy constrains these parameters as well. We can therefore see that these constraints affect the geometry of the assimilation solution.

Because of this phase problem, comparison between the two constraint types is difficult. It may prove useful to investigate alternative methods of measuring the accuracy of the assimilation solutions. In general, however, the results for the two-body problem need to be tested in a more complex system and work is in progress in this direction.

ACKNOWLEDGEMENTS

We would like to thank Prof. M. Baines and Prof. M. Cullen for their help with this research. This research is sponsored by the EPSRC and Met Office U.K.

REFERENCES

1. Nichols NK. Data assimilation: aims and basic concepts. In *Data Assimilation for the Earth System*, Swinbank R, Shutyaev V, Lahoz WA (eds). Kluwer: Dordrecht, 2003; 9–20.
2. Chao WC, Chang L. Development of a four-dimensional variational analysis system using the adjoint method at GLA. Part 1: Dynamics. *Monthly Weather Review* 1992; **120**: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, ECMWF, Reading, U.K., 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* 1994; **122**(5):966–983.
6. Gill PE, Murray W, Wright MH. *Practical Optimization*. Academic Press: New York, 1981.
7. scienceworld.wolfram.com, http://scienceworld.wolfram.com/physics/Eccentricity.html.

