School of Mathematics, Meteorology
and Physics

University of Reading

Department of Mathematics

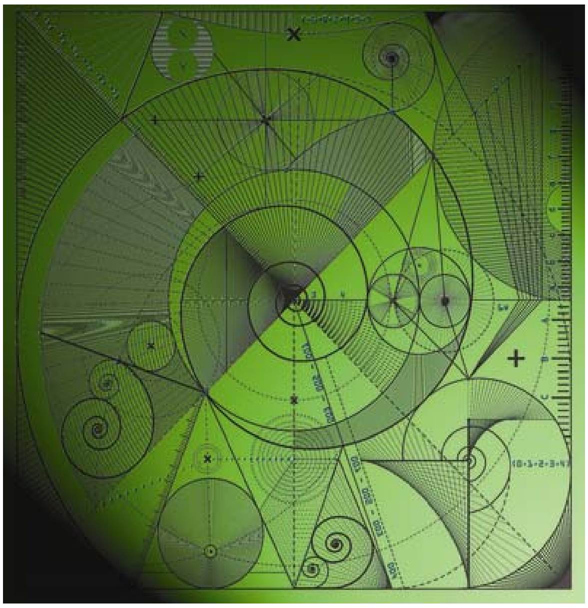
Preprint MPS_2010_15

21 April 2010

State estimation using model order reduction for unstable systems

by

C. Boess, A.S. Lawless, N.K. Nichols,
A. Bunse-Gerstner



State estimation using model order reduction for unstable systems

C.Boess[a,*], A.S. Lawless[a], N.K. Nichols[a], A. Bunse-Gerstner[b]

[a] *Department of Mathematics, University of Reading, UK*
[b] *ZeTeM, Universitaet Bremen, Germany*

Abstract

The problem of state estimation occurs in many applications of fluid flow. For example, to produce a reliable weather forecast it is essential to find the best possible estimate of the true state of the atmosphere. To find this best estimate a nonlinear least squares problem has to be solved subject to dynamical system constraints. Usually this is solved iteratively by an approximate Gauss-Newton method where the underlying discrete linear system is in general unstable. In this paper we propose a new method for deriving low order approximations to the problem based on a recently developed model reduction method for unstable systems. To illustrate the theoretical results, numerical experiments are performed using a two dimensional Eady model – a simple model of baroclinic instability, which is the dominant mechanism for the growth of storms at mid-latitudes. It is a suitable test model to show the benefit that may be obtained by using model reduction techniques to approximate unstable systems within the state estimation problem.

Keywords: State estimation, Gauss-Newton methods, variational data assimilation, unstable models, balanced truncation

1. Introduction

The problem of state estimation occurs in many applications of fluid flow. For example, to make reliable predictions of weather and ocean systems, of reservoir dynamics, or even of traffic flows, it is essential to find the best possible estimate of the true state of the system with which to initialize

*Corresponding author: Caroline Boess (C.Boess@reading.ac.uk)

a forecast. *Data assimilation* techniques aim to find this best estimate by combining observational data with a numerical model of the system. In the popular technique of four-dimensional variational data assimilation (4D-Var) the assimilation problem is posed as a large nonlinear least squares problem of the form

$$\min_x \phi(x) = f(x)^T f(x), \tag{1}$$

where the function $f(x)$ includes the nonlinear forecast model. This is often solved by applying a few iterations of an approximate Gauss-Newton method, in an algorithm known as incremental 4D-Var [1, 2]. We note that in order to apply such a method we need the Jacobian of the function $f(x)$ and hence the Jacobian of the nonlinear forecast model, which is referred to as the tangent linear model (TLM). Significant properties of the numerical prediction models are their large dimensions and their unstable behaviour. Instability in the TLM arises where the underlying nonlinear forecast model is linearly unstable over a finite time interval. This motivates the necessity to be able to approximate large unstable models by low order systems.

A commonly used approach to reduce the complexity of the problem (1) is to approximate the full TLM by a linearized model at low spatial resolution. Whilst this leads to an algorithm that is practical to compute in real-time, the approximations made do not take into account whether the most important parts of the dynamical system are being retained. Recently it has been shown that model reduction techniques from the field of control theory can be used to approximate the solution of the full order problem if the underlying dynamical system is asymptotically stable [3, 4, 5].

However, the systems occurring in the state estimation problem (1) are in general unstable whilst most of the known model reduction methods are for asymptotically stable systems only. We consider the recently derived model reduction method of *α-bounded balanced truncation* [6]. In contrast to existing approaches for model reduction of unstable systems, this new method approximates the input-output behaviour of the asymptotically stable as well as of the unstable part of the full order system, no matter how many unstable poles there are.

Employing this new method of α-bounded balanced truncation we show how model reduction may be used in the inner step of the Gauss-Newton algorithm to give an approximate iteration procedure that retains the most important properties of the dynamical system response if the underlying system is unstable. We demonstrate the application of this technique to the problem of variational data assimilation, which corresponds to an optimal state estimation problem. The results are illustrated by numerical experiments with a two dimensional Eady model, a simple model of baroclinic instability, which is the dominant mechanism for the growth of storms at mid-latitudes. It is a suitable test model (with a considerable number of unstable poles) to show the benefit that may be obtained by using the new balanced truncation method for unstable systems within the Gauss-Newton algorithm. In the experiments we compare the new model reduction technique with the standard low resolution approach and with the standard extension of balanced truncation for unstable systems within the state estimation problem.

In the next section we describe the state estimation problem as it occurs in the field of data assimilation. Section 3 then presents how the model reduction technique of α-bounded balanced truncation can be used to solve the state estimation problem approximately when the underlying TLM is unstable. In Section 4 we illustrate the method by numerical experiments with a 2-dimensional Eady model. Finally we draw conclusions in section 5.

2. State estimation within data assimilation

The aim of variational data assimilation is to match the output response of a dynamical system model to observed measurements of the outputs over a specified time window $[t_0, t_N]$. For a discrete dynamical system, we let $x_j \in \mathbb{R}^n$ be the model state vector, $y_j \in \mathbb{R}^p$ be a vector of p observations and $h_j : \mathbb{R}^n \to \mathbb{R}^p$ be a nonlinear observation operator that relates the system states to the observations at time t_j. The model state vectors satisfy the nonlinear model equations $x_j = m_{j,0}(x_0)$.

The 4D-Var method can be stated as a nonlinear least squares problem of the form (1) with the nonlinear function $f : \mathbb{R}^n \to \mathbb{R}^{n+(N+1)p}$ defined as

$$f(x_0) := \begin{bmatrix} B_0^{-\frac{1}{2}}(x_0 - x_0^{(b)}) \\ R_0^{-\frac{1}{2}}(h_0[x_0] - y_0) \\ \cdots \\ R_N^{-\frac{1}{2}}(h_N[x_N] - y_N) \end{bmatrix}. \tag{2}$$

The background estimate $x_0^{(b)}$ of the initial state x_0 is known and the initial errors $(x_0 - x_0^{(b)})$ and the observational errors $(h_j[x_j] - y_j)$ are assumed to be unbiased, Gaussian random vectors with covariance matrices B_0 and R_j, respectively.

In practice, the nonlinear least squares problem (1) with f as defined in (2) can be solved by applying the Gauss-Newton method. This is an iterative algorithm that minimizes in each iteration step (k) the linear least squares function

$$\tilde{\phi}(\delta x_0^{(k)}) = \|J_f \delta x_0^{(k)} + f(x_0^{(k)})\|_2^2, \tag{3}$$

where J_f denotes the Jacobian of f. The new iterate is then defined as $x_0^{(k+1)} = x_0^{(k)} + \delta x_0^{(k)}$. It follows from (2) that the Jacobian J_f is given by

$$J_f = \left[\begin{array}{ccccc} \left(B_0^{-\frac{1}{2}}\right)^T & \left(R_0^{-\frac{1}{2}} H_0\right)^T & \left(R_1^{-\frac{1}{2}} H_1 M_{1,0}\right)^T & \ldots & \left(R_N^{-\frac{1}{2}} H_N M_{N,0}\right)^T \end{array} \right]^T, \tag{4}$$

with linearized observation and model matrices $H_i := \frac{\partial h_i}{\partial x_i}(x_i^{(k)})$, $i = 0, \ldots, N$, and $M_{i,0} := \frac{\partial m_{i,0}}{\partial x_0}(x_0^{(k)})$ for $i = 1, \ldots, N$, respectively.

In cases where the dimension of the state vector is very large it is computationally expensive to solve (3). A common approach for simplifying the problem is to consider the linear model at a lower spatial resolution. This reduces the complexity substantially and thus makes a solution of (3) feasible for very large dimension, but it is not assured that the most important information in the full order model is retained. In [4, 5] it has been shown that the model reduction method of balanced truncation can be used instead to solve (3) approximately, but only if the system is asymptotically stable. This method projects the tangent linear system

$$\begin{array}{rcl} \delta x_{i+1}^{(k)} & = & M_{i+1,i} \delta x_i^{(k)}, \\ d_i^{(k)} & = & H_i \delta x_i^{(k)}, \end{array} \tag{5}$$

where $M_{i+1,i} = \frac{\partial m_{i+1,i}}{\partial x_i}(x_i^{(k)})$, to a lower dimensional space, where the cost function is then minimized approximately. Compared to the low resolution approach, the use of model reduction techniques supplies better approximations to the solution of the full order problem. The next section shows how this idea can be extended to unstable systems.

3. Model reduction for unstable systems

The aim is to solve (3) by finding low order approximations to the linear system (5) within the $(k) - th$ iteration step of the Gauss-Newton method.

For ease of notation the iteration index (k) is omitted in the following. We consider a time-invariant approximation

$$\mathcal{S} : \left\{ \begin{array}{lll} \delta x_{i+1} & = & M \delta x_i, \\ d_i & = & H \delta x_i, \end{array} \right. \tag{6}$$

to the linear system (5), where the constant matrices M and H are approximations to the time-varying operators $M_{i+1,i}$ and H_i over the time window $[t_0, t_N]$. The initial state $\delta x_0 = B_0^{\frac{1}{2}} \omega$ is a normally distributed random variable with mean zero and covariance matrix $B_0 \in \mathbb{R}^{n \times n}$, where $\omega \sim \mathcal{N}(0, I)$.

The transfer function T of the system (6) is a complex matrix valued function describing the behaviour of the system in frequency domain. It is defined as

$$T : \mathbb{C} \rightarrow \mathbb{R}^{p \times m}, \tag{7}$$

$$z \mapsto T(z) := H(zI - M)^{-1} B_0^{\frac{1}{2}}. \tag{8}$$

The system (6) is in general unstable, i.e. the eigenvalues of the system matrix M may lie outside as well as inside the unit circle. To be able to find low order approximations to (6) we need a reliable model reduction technique for unstable systems. The method of α-bounded balanced truncation has recently been developed [6] for *α-bounded* systems, i.e. for systems where all eigenvalues of the system matrix M lie in a circle around the origin with real positive radius α. For any regular unstable system $\mathcal{S}$ it is possible to find such an α. The α-bounded balanced truncation then computes restriction and prolongation matrices U and V, respectively, such that the projected system,

$$\hat{\mathcal{S}} : \left\{ \begin{array}{lll} \delta \hat{x}_{i+1} & = & U^T M V \delta \hat{x}_i, \\ \hat{d}_i & = & H V \delta \hat{x}_i, \end{array} \right. \tag{9}$$

with low order state vector $\delta \hat{x}_i = U^T \delta x_i \in \mathbb{R}^r$, $r \ll n$, approximates the full order system $\mathcal{S}$ very well. More precisely, there exists a global error bound with respect to the $h_{\infty,\alpha}$-norm of the error system [6]:

$$\|T - \hat{T}\|_{h_{\infty,\alpha}} \leq 2(\sigma_{r+1}^{(\alpha)} + \ldots + \sigma_n^{(\alpha)}), \tag{10}$$

where $\|T\|_{h_{\infty,\alpha}} := \sup_{\theta \in [0,2\pi]} \sigma_{max} \left(T(\alpha e^{i\theta}) \right)$ and σ_{max} denotes the largest singular value of the operator T. The $h_{\infty,\alpha}$-norm is well-defined for any function

$F : \mathbb{C} \to \mathbb{R}^{m\times p}$ that is holomorphic in the complement of the circle around the origin with radius α and, therefore, for α-bounded systems. The scalars $\sigma_{r+1}^{(\alpha)}, \ldots, \sigma_n^{(\alpha)}$ are the Hankel singular values of the α-scaled system $\mathcal{S}_\alpha$, given by the transfer function

$$T_\alpha = \frac{1}{\alpha} H(zI - \frac{1}{\alpha} M)^{-1} B_0^{\frac{1}{2}},$$

that are not matched by the corresponding reduced order system.

We show next that the low order system (9) can be used to solve the least squares problem within the Gauss-Newton iteration approximately. Instead of minimizing (3) we minimize the cost function

$$\hat{\phi}(\delta\hat{x}_0) = \|\hat{J}\delta\hat{x}_0 + \hat{f}\|_2^2 \tag{11}$$

where

$$\hat{J} = \begin{bmatrix} (U^T B_0 U)^{-\frac{1}{2}} \\ R_0^{-\frac{1}{2}} HV \\ R_1^{-\frac{1}{2}} HV(U^T MV) \\ \vdots \\ R_N^{-\frac{1}{2}} HV(U^T MV)^N \end{bmatrix}, \hat{f} = - \begin{bmatrix} U^T \delta x_0^{(b)} \\ R_0^{-\frac{1}{2}} d_0 \\ R_1^{-\frac{1}{2}} d_1 \\ \vdots \\ R_N^{-\frac{1}{2}} d_N \end{bmatrix}, \tag{12}$$

with $\delta x_0^{(b)} := x_0^{(b)} - x_0$ and $d_i := y_i - h_i[x_i]$. The minimization of (11) is computationally much more efficient than solving (3); the solution $\delta\hat{x}_0$ has dimension $r \ll n$. The prolongation operator V is then applied to lift the solution $\delta\hat{x}_0$ back into the space $\mathbb{R}^n$. The lifted state vector $\delta\hat{x}_0^{(lift)} := V\delta\hat{x}_0$ is in general a good approximation to the solution δx_0 of the full problem (3).

In [6] the superior performance of the α-bounded balanced truncation method of model reduction has been shown over a standard extension of balanced truncation for unstable systems based on an additive decomposition of the system into its asymptotically stable and its unstable parts. We investigate in the next section whether this result continues to hold when these two model reduction methods are used to solve the state estimation problem (3) approximately.

4. Numerical experiments

We now perform numerical experiments to illustrate the benefit of the new α-bounded model reduction method for approximating the unstable linear system inside the Gauss-Newton procedure. As a test model we consider a two dimensional Eady model in the $x-z$ plane – a simple model of baroclinic instability, which is the dominant mechanism for the growth of storms at mid-latitudes.

4.1. Experimental design

The nondimensional equations for the 2D Eady model [7] are now described. The basic state is given by a linear zonal wind shear with height, z, in a domain between two rigid horizontal boundaries, $z = \pm\frac{1}{2}$.

Following [8] it is assumed that the interior quasi-geostrophic potential vorticity (QGPV) is zero. The initial state is given by the perturbation buoyancy, $b = b(x, z, t)$, on the boundaries, $z = \pm\frac{1}{2}$, at time $t = 0$. This is used to calculate the corresponding perturbation geostrophic stream function, $\psi = \psi(x, z, t)$, which satisfies:

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial z^2} = 0, \quad \text{in } z \in [-\frac{1}{2}, \frac{1}{2}],\ x \in [0, X], \tag{13}$$

with boundary conditions

$$\frac{\partial\psi}{\partial z} = b, \quad \text{on } z = \pm\frac{1}{2},\ x \in [0, X]. \tag{14}$$

Perturbations to the basic state are advected zonally by the basic shear flow as described by the non-dimensional QG thermodynamic equation:

$$\left(\frac{\partial}{\partial t} + z\frac{\partial}{\partial x}\right) b = \frac{\partial\psi}{\partial x}, \quad \text{on } z = \pm\frac{1}{2},\ x \in [0, X]. \tag{15}$$

The spatial boundary conditions in the x-direction are taken to be periodic such that at any time, t, and height, z, $b(0, z, t) = b(X, z, t)$ and $\psi(0, z, t) = \psi(X, z, t)$. As in [8] we use dimensionless values for x, z, and t.

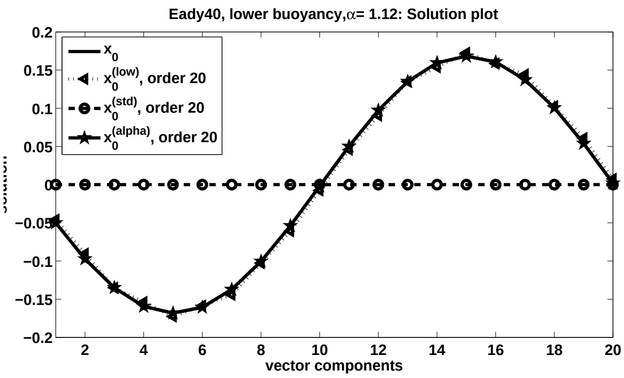
In the experimental studies here, the Eady model is discretized using 11 vertical levels for stream function. There are 20 grid points in the horizontal, giving 40 degrees of freedom in the state vector b. The advection equations are discretized using a leap-frog advection scheme. We refer to [8] for further details. The observation matrix H is chosen such that the observations are taken from the lower-level buoyancy only.

4.2. Comparison of low order and low resolution approximations

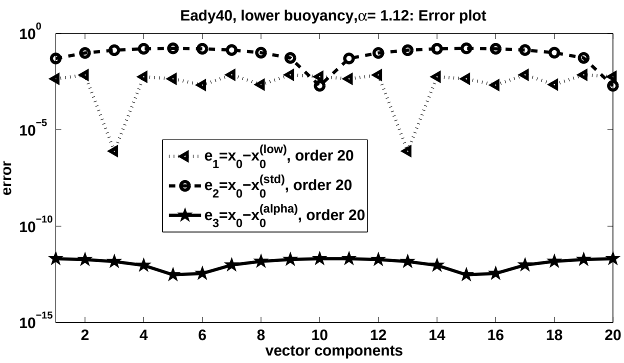
To investigate the quality of the approximations of the tangent linear model we compare the lifted solutions $\delta\hat{x}_0^{(lift)}$ of the low order least squares problem (11) with the solution δx_0 of the full order least squares problem (3). We define a true solution $\delta x_0^{(true)}$ as the growing mode of the Eady model, i.e. the eigenvector associated with the largest eigenvalue in absolute value. Then perfect observations are generated from the true solution: $d_i = HM^i\delta x_0^{(true)}$, where M represents the linear model operator $M_{i+1,i}$. For the computation of the low order solutions $\delta\hat{x}_0$ of (11) we consider three different approximation techniques: the low resolution approach and two model reduction methods that use the standard extension of balanced truncation for unstable systems and also the new α-bounded balanced truncation method.

Figure 1(a) shows the buoyancy b on the lower boundary and its low order approximations where the reduction order is half the size of the dimension of the full order problem. We see that the true solution (solid line) is approximated very well by the solution computed by α-bounded balanced truncation with $\alpha = 1.12$ (solid line with stars). We note that the solid line and the solid line with stars lie on top of each other and, thus, they are indistinguishable in the figure. The error is of order 10^{-12} (see Figure 1(b), solid line with stars). In contrast, the standard balanced truncation method for unstable systems (dashed line with circles) and the low resolution approach (dotted line with triangles) supply solutions that approximate the true solution (solid line) rather poorly with average errors of order of magnitude 10^{-1} (Figure 1(b), dashed line with circles) and 10^{-3} (Figure 1(b), dotted line with triangles), respectively.

The standard balanced truncation method supplies a very poor approximation. This is caused by the fact that the Eady model has a large number of unstable poles (i.e. eigenvalues that are larger than one in absolute value). Only two of the 40 eigenvalues lie inside the unit circle (see Figure 3(a)). The standard balanced truncation approach additively decomposes the system into an asymptotically stable subsystem (of order 2) and and unstable subsystem (of order 38). The attempt to reduce the order of the system to an order smaller than the number of unstable poles leads to a low order system that only captures a part of the unstable modes. It is not even assured that at least the most dominant unstable part is kept. Additionally, the asymptotically stable components are ignored completely. This explains why this technique is not capable of computing a good approximation when we reduce the order of the system to 20, which is much smaller than the number of unstable poles.



(a) Solution on lower boundary

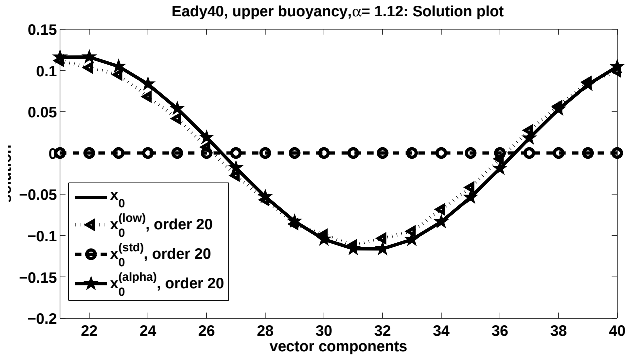


(b) Error on lower boundary

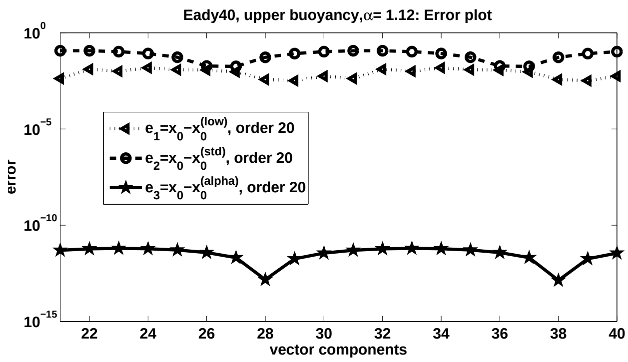
Figure 1: Comparison of low resolution (dotted line with triangles), standard balanced truncation (dashed line with circles) and α-bounded balanced truncation (solid line with stars) approximations to the buoyancy on the lower boundary.

Very similar results continue to hold for the buoyancy on the upper boundary. In Figure 2(a) the approximation by the α-bounded method (solid line with stars) is indistinguishable from the true solution (solid line). The error is still of order of magnitude 10^{-11} to 10^{-12}. This is a good result taking into account that in our experimental setting we have no observations at the upper boundary. The errors in the low resolution approximation (dotted line with triangles) and in the standard balanced truncation approximation (dashed line with circles) are of order of magnitude 10^{-1} to 10^{-2}. This means that the approximation using the new α-bounded approach is approximately

10 orders of magnitude more accurate.



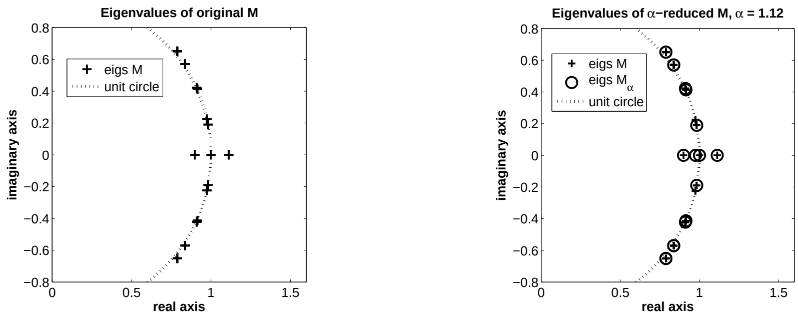
(a) Solution on upper boundary



(b) Error on upper boundary

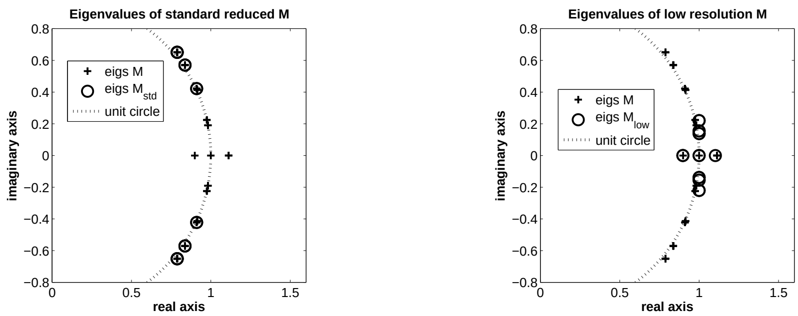
Figure 2: Comparison of low resolution (dotted line with triangles), standard balanced truncation (dashed line with circles) and α-bounded balanced truncation (solid line with stars) approximations to the buoyancy on the upper boundary

The experiments have shown the clear superiority of the α-bounded approximation technique. This benefit can be explained in part by examining the eigenstructure of the reduced dimensional systems. With the α-bounded method it is possible to match more of the significant eigenvalues of the full system than is the case for the low resolution model and for the standard balanced truncation model. Figure 3(a) shows the eigenvalues of the system matrix of the original full order system (crosses) while Figures 3(b), 4(a) and 4(b) show the eigenvalues of the different low order approximations. The α-bounded approximation method matches almost all of the eigenvalues of the full order system, inside as well as outside the unit circle (see Figure 3(b), circles). In contrast, the standard balanced truncation method (Figure 4(a), circles) is capable of matching only some of the eigenvalues outside the unit circle, but none inside. Thus, we cannot expect this method to capture the dominant behaviour of the full order system successfully. The low resolution method, on the other hand, matches eigenvalues inside and outside the unit circle (Figure 4(b), circles), but there is still a considerable number of eigenvalues that are not captured at all. We see in the figures that the α-bounded method matches the widest range of eigenvalues of the full order system.



(a) (b)

Figure 3: Eigenvalues of (a) full order M and (b) α-reduced M.



(a) (b)

Figure 4: Eigenvalues of (a) standard reduced M and (b) low resolution M.

5. Conclusions

State estimation problems occur in many different applications. Within numerical weather prediction, data assimilation techniques seek to find the best estimate of the true state of the atmosphere at the initial step of a given time window. In the well-known 4D-Var method this is achieved by solving a nonlinear least squares problem constrained by nonlinear model equations that describe the evolution of the state of the atmosphere with time. In operational weather forecasting, this complex problem is solved using an approximate Gauss-Newton procedure. Each step of this iterative method contains a linear least squares problem subject to linear model equations, the tangent linear model (TLM). The TLM is deduced from a nonlinear system and may be unstable over a finite time window. The state vectors have very large dimension and further approximations are therefore indispensable. Usually the TLM is approximated by using a model with a lower spatial resolution.

In this paper we have proposed that the model reduction method of α-bounded balanced truncation may be employed to obtain better approximations to the unstable TLM within the Gauss-Newton procedure. This model reduction technique computes a low order approximation to the TLM while still capturing its most important properties. It can be applied independently of the number of unstable poles of the full order system. The existence of a global error bound can be proved.

The proposed method is computationally expensive, however, and more work is needed in order to make it feasible for operational systems. However, it is possible to make the method more practical for very large systems by using Krylov subspace techniques for finding the projections.

We have compared the α-bounded balanced truncation method with the standard balanced truncation approach for unstable systems and with the low resolution approximation using numerical experiments with a 2-dimensional Eady model. The Eady model is a simple model of baroclinic instability, which is the dominant mechanism for the growth of storms at mid-latitudes. In the numerical experiments we demonstrate the clear superiority of the α-bounded approximation method. It captures the dominant behaviour of the full order system. The low order approximation of the buoyancy on the lower and upper boundary is hardly distinguishable from the full order solution. In the experiments performed, the error was found to be on average ten orders of magnitude smaller than the errors of the other two approximation techniques.

Acknowledgements
This work was supported in part by the NERC National Centre for Earth Observation.

References

[1] Courtier P, Thépaut JN, Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. Q J R Meteorol Soc 1994;120:1367–87.

[2] Lawless AS, Gratton S, Nichols NK. Approximate iterative methods for variational data assimilation. Int J for Numerical Methods in Fluids 2005;47:1129–35.

[3] Boess C. Using model reduction techniques within the incremental 4D-Var method. Ph.D. thesis; Universitaet Bremen; 2008.

[4] Lawless AS, Nichols NK, Boess C, Bunse-Gerstner A. Using model reduction methods within incremental 4D-Var. Monthly Weather Review 2008;136:1511–22.

[5] Lawless AS, Nichols NK, Boess C, Bunse-Gerstner A. Approximate Gauss-Newton methods for optimal state estimation using reduced order models. Int J Numer Methods in Fluids 2008;56:1367–73.

[6] Boess C, Nichols NK, Bunse-Gerstner A. Model reduction for discrete unstable control systems using a balanced truncation approach. Preprint MPS_2010_06 , Department of Mathematics, University of Reading 2010;.

[7] Eady ET. Long waves and cyclone waves. Tellus 1949;1:33–52.

[8] Johnson C, Hoskins BJ, Nichols NK. A singular vector perspective of 4D-Var: Filtering and interpolation. Q J R Meteorol Soc 2005;131:1–19.