University of Reading

Optimal linearization trajectories for tangent linear models

Article

Accepted Version

Stappers, R. J. J. and Barkmeijer, J. (2012) Optimal linearization trajectories for tangent linear models. Quarterly Journal of the Royal Meteorological Society, 138 (662). pp. 170-184. ISSN 1477-870X doi: 10.1002/qj.908 Available at https://centaur.reading.ac.uk/28437/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/qj.908

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

University of Reading

Reading's research outputs online

Optimal linearization trajectories for tangent linear models

R.J.J. Stappers and J. Barkmeijer

Abstract

We examine differential equations where nonlinearity is a result of the advection part of the total derivative or the use of quadratic algebraic constraints between state variables (such as the ideal gas law). We show that these types of nonlinearity can be accounted for in the tangent linear model by a suitable choice of the linearization trajectory. Using this optimal linearization trajectory we show that the tangent linear model can be used to reproduce the exact nonlinear error growth of perturbations for more than 200 days in a quasi geostrophic model and more than (the equivalent of) 150 days in the Lorenz 96 model. We introduce an iterative method, purely based on tangent linear integrations, that converges to this optimal linearization trajectory.

The main conclusion from this paper is that this iterative method can be used to account for nonlinearity in estimation problems without using the nonlinear model. We demonstrate this by performing forecast sensitivity experiments in the Lorenz 96 model and show that we are able to estimate analysis increments that improve the two day forecast using only four backward integrations with the tangent linear model.

1 Introduction

The use of tangent linear and, in particular, adjoint models has been very useful in several applications in numerical weather prediction (NWP) (see Errico, 1997, 2003; Errico and Ehrendorfer, 2007, for an overview). For example at the European Centre for Medium-Range Weather Forecasts (ECMWF) these linear models play a crucial role in the computation of initial condition perturbations used in the ensemble prediction system (Leutbecher and Palmer, 2008) and in their 4D-VAR data assimilation system (Courtier *et al.*, 1994). One of the major limitations to the application of linear models is that the results are useful only when the linear approximation is valid (Errico, 1997). By this we mean that the difference between two runs of the nonlinear model can be described by the associated linearized version of the nonlinear model. To achieve this, great effort is taken to develop linearized models which capture as many as possible features of the full nonlinear model (Janisková *et al.*, 1999). Despite these efforts the use of tangent linear and adjoint models is restricted to "short" time spans. The time span for which the TL-model can be considered accurate will be referred to as the tangent linear regime.

The duration of the tangent linear regime depends on many factors. Typically the difference between two nonlinear forecasts is compared with the linear forecast by a scalar index, and it is said that the TL-assumption is violated when the index has reached a threshold value. So the measure which is employed to compare forecast fields is already important in the definition of the TL-regime. But also the size of the initial condition perturbation, the orientation of the perturbation, the background trajectory around which the TL-model is linearized and the physical processes taken into account in the TL-model play a role. Another issue which influences the usefulness of linear models is whether we are considering forecast problems, where error growth is determined by the singular value spectrum of the propagator, or estimation problems that are typically characterized by the reciprocal of the singular value spectrum. In general the spectrum of reciprocal of the singular values attains higher values (Reynolds and Palmer, 1998) and therefore the usefulness of the TL-model in esti-

mation problems is shorter. This effect becomes even more pronounced by the fact that the typical size of perturbations used in backward integrations is larger than in forward mode.

In this forward mode the tangent linear assumption is generally believed to be valid for 2-3 days at the synoptic scale. However, see Gilmour *et al.* (2001) who argue that 1 day is perhaps a better estimate. On the cloud resolving scale the tangent linear assumption probably holds for much shorter time periods on the order of 1.5 hours (Hohenegger and Schär, 2007). As models reach higher resolutions the validity of the TL assumption is therefore a major concern. We will show that for bilinear systems the usefulness of the TL-model can be greatly extended by modifying the linearization trajectory and therefore the one of the major limitations for using tangent linear models can be eliminated.

In section 2 the definition of bilinear differential equations is given. In section 3 we show that for bilinear systems there is an optimal linearization trajectory such that if the tangent linear model is linearized around this trajectory the perturbation growth in the tangent linear model is equal to the nonlinear perturbation growth. Knowing that such a trajectory exist we show in section 4 that there is an iterated map based purely on TL-integrations that converges to this linearization trajectory. In section 5 we show how the iterative method can be used in forecast sensitivity experiments using the inverse of the TL-model. In section 6 the experimental results using a quasi geostophic model (Marshall and Molteni, 1993, described in appendix A) and the Lorenz 96 model (Lorenz, 1996, described in appendix B) are given. In the discussion in section 7 the prospects for using the method in realistic NWP models and a method to regularize the error growth in the tangent linear model are discussed. The conclusions are given in section 8.

To keep the notation simple we use the convention that lower case variables are perturbations (also referred to as increments) to upper case variable, e.g. $\mathbf{x}$ is a perturbation to the state vector $\mathbf{X}$.

2 Bilinear differential equations

In this section some terms are defined that will be used throughout the rest of the paper. Let $\mathbf{U}, \mathbf{V}, \mathbf{W}$ and $\mathbf{X}$ be elements of a vector space $\mathbb{H}$.

Definition 1 (Bilinear map)
A map q

$$\begin{aligned} q : \mathbb{H} \times \mathbb{H} &\to \mathbb{H} \\ (\mathbf{U}, \mathbf{V}) &\mapsto \mathbf{W} = q(\mathbf{U}, \mathbf{V}), \end{aligned}$$

is called bilinear if q is linear in both arguments.

Definition 2 ((Anti)symmetric bilinear map)
A bilinear map s will be called symmetric if for any $\mathbf{U}, \mathbf{V} \in \mathbb{H}$

$$s(\mathbf{U}, \mathbf{V}) = s(\mathbf{V}, \mathbf{U}). \tag{1}$$

A bilinear map a will be called antisymmetric if

$$a(\mathbf{U}, \mathbf{V}) = -a(\mathbf{V}, \mathbf{U}). \tag{2}$$

Note that for any bilinear map q there is a unique decomposition

$$q(\mathbf{U}, \mathbf{V}) = s(\mathbf{U}, \mathbf{V}) + a(\mathbf{U}, \mathbf{V}), \tag{3}$$

where $s(\mathbf{U}, \mathbf{V}) = \frac{1}{2}\left(q(\mathbf{U}, \mathbf{V}) + q(\mathbf{V}, \mathbf{U})\right)$ is symmetric and $a(\mathbf{U}, \mathbf{V}) = \frac{1}{2}\left(q(\mathbf{U}, \mathbf{V}) - q(\mathbf{V}, \mathbf{U})\right)$ is antisymmetric.

Definition 3 (Bilinear differential equation)
A differential equation will be called bilinear if it is of the form

$$\dot{\mathbf{X}} = q(\mathbf{X}, \mathbf{X}) + b(\mathbf{X}) + c, \tag{4}$$

where q is a bilinear map, b is a linear map and c is a forcing. If $\mathbf{X}$ is finite dimensional this is an ordinary differential equation (ODE) while if $\mathbf{X}$ represents a (collection of) space and time dependent field(s) this is a partial differential equation (PDE). For PDE's the mappings q, b and forcing c are allowed to depend on space and time explicitly.

Definition 4 (Bilinear differential algebraic equation) A differential algebraic equation (Brenan *et al.*, 1996) will be called bilinear if it is of the form

$$e(\mathbf{X}, \dot{\mathbf{X}}) + d(\dot{\mathbf{X}}) = q(\mathbf{X}, \mathbf{X}) + b(\mathbf{X}) + c, \qquad (5)$$

where q and e are bilinear maps, d and b are linear maps and c is a forcing.

Example 1 (Barotropic vorticity equation) The barotropic vorticity equation (bve) is

$$\begin{aligned} \frac{\partial \eta}{\partial t} &= -J(\psi, \eta) \\ 0\frac{\partial \psi}{\partial t} &= \nabla^2 \psi + f - \eta \\ 0\frac{\partial \mathbf{u}}{\partial t} &= \mathbf{u} - \mathbf{k} \times \nabla \psi \end{aligned}$$

where the first equation is a prognostic equation for the absolute vorticity η, the second and third equation are algebraic constraints (diagnostic equations) for the stream function ψ and the two dimensional velocity $\mathbf{u}$ respectively (hence the zeros in front of the time derivatives), f is the Coriolis parameter, $\mathbf{k}$ is the vertical unit vector and J is an antisymmetric bilinear map defined as

$$J(\psi, \eta) = \frac{\partial \psi}{\partial x}\frac{\partial \eta}{\partial y} - \frac{\partial \psi}{\partial y}\frac{\partial \eta}{\partial x}. \qquad (6)$$

If we define $\mathbf{X} = (\eta, \psi)^T$ we see that the bve is a bilinear partial differential algebraic equation in the state vector $\mathbf{X}$ with $e = 0$ and $d = \mathrm{diag}(I, 0)$ and the velocity is a "post-processed" variable. Alternatively the equation for $\frac{\partial \eta}{\partial t}$ can be written as $\frac{\partial \eta}{\partial t} = -\mathbf{u} \cdot \nabla \eta$ in which case the state vector should be defined as $\mathbf{X} = (\eta, \psi, \mathbf{u})^T$.

Example 2 (Momentum equation) The momentum equation in a uniform rotating coordinate frame is (Pedlosky, 1987)[1]

$$\rho\frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u} \cdot \nabla)\mathbf{u} + 2\boldsymbol{\Omega} \times \rho\mathbf{u} = -\nabla p - \rho\nabla\phi + \mathcal{F}(\mathbf{u}),$$

where $\mathbf{u}$ is the three dimensional velocity vector, p is pressure, ρ is density, $\boldsymbol{\Omega}$ is the angular rotation vector, ϕ is the potential that represents conservative body forces, including gravity, $\mathcal{F}$ represents nonconservative (frictional) forces. The prognostic equation for $\mathbf{u}$ is a trilinear differential equation due to the term $\rho(\mathbf{u} \cdot \nabla)\mathbf{u}$ caused by the advection part of the total derivative. It is however easy to transform the trilinear equation to a bilinear differential algebraic equation by augmenting the state vector with the momentum density $\mathbf{p} = \rho\mathbf{u}$

$$\begin{aligned} \rho\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{p} \cdot \nabla)\mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{p} &= -\nabla p - \rho\nabla\phi + \mathcal{F}(\mathbf{u}) \\ 0\frac{\partial \mathbf{p}}{\partial t} &= \rho\mathbf{u} - \mathbf{p}. \end{aligned}$$

Alternatively the momentum density vector field $\mathbf{p}$ can be considered as the prognostic variable

$$\begin{aligned} \frac{d\mathbf{p}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{p} &= -(\nabla \cdot \mathbf{u})\mathbf{p} - \nabla p - \rho\nabla\phi + \mathcal{F}(\mathbf{u}) \\ 0\frac{\partial \mathbf{u}}{\partial t} &= \mathbf{p} - \rho\mathbf{u}, \end{aligned}$$

where we used the mass continuity equation.

Example 3 (Equation of state) The equation of state for an ideal gas can be formulated as an algebraic constraint as

$$0\frac{\partial p}{\partial t} = p - \rho RT. \qquad (7)$$

Examples 1 and 2 illustrate that in fluid dynamics bilinearity is typically a result of the advection part of the total derivative. Example 3 shows that another source for bilinearity is the use of algebraic constraints between state variables such as the ideal gas law. Example 2 further illustrates that it is easy to reduce multilinear systems (see appendix D) to bilinear systems by augmenting the state vector.

Notation 1 (Nonlinear integrations) Integrations with a nonlinear model starting from an initial condition $\mathbf{X}_0$ are denoted by

$$\mathbf{X}(t) = \mathcal{M}(t, 0, \mathbf{X}_0). \qquad (8)$$

[1] We introduced a minus sign in the $\rho\nabla\phi$ term such that potential energy is increasing with increasing height

Then, by definition, the exact increment trajectory for a given perturbation $\mathbf{x}_0$ of an initial condition $\mathbf{X}_0$ is given by

$$\mathbf{x}(t) \equiv \mathcal{M}(t, 0, \mathbf{X}_0 + \mathbf{x}_0) - \mathcal{M}(t, 0, \mathbf{X}_0). \quad (9)$$

Notation 2 (Tangent linear integrations) Integrations with the tangent linear model starting with an initial condition perturbation $\mathbf{x}_0$ are denoted by

$$\hat{\mathbf{x}}(t) = \mathbf{M}_{\mathbf{X}}(t, 0)\mathbf{x}_0, \quad (10)$$

where $\mathbf{M}_{\mathbf{X}}(t, 0)$ is known as the propagator and $\mathbf{X}$ is the trajectory around which the tangent linear model is linearized.

3 Optimal linearization trajectories

In this section we derive the tangent linear model for the general form of bilinear system and show how to modify the linearization trajectory to obtain an exact correspondence between the nonlinear time evolution and the corresponding tangent linear evolution of perturbations.

Consider the general form of a bilinear differential equation

$$\dot{\mathbf{X}} = q(\mathbf{X}, \mathbf{X}) + b(\mathbf{X}) + c, \quad (11)$$

where $\mathbf{X} \in \mathbb{H}$, q is a bilinear mapping, b is a linear mapping and c is a forcing and the mappings q and b and forcing c are allowed to explicitly depend on space and time. Solutions (trajectories in $\mathbb{H}$) of (11) are denoted as $\mathbf{X}(t)$. The time evolution of a perturbed run $\mathbf{X}(t) + \mathbf{x}(t)$ is given by

$$\dot{\mathbf{X}} + \dot{\mathbf{x}} = q(\mathbf{X} + \mathbf{x}, \mathbf{X} + \mathbf{x}) + b(\mathbf{X} + \mathbf{x}) + c. \quad (12)$$

Now using the bilinearity of q, the linearity of b and (11) to eliminate $\dot{\mathbf{X}}$ we obtain

$$\dot{\mathbf{x}} = \underbrace{q(\mathbf{X}, \mathbf{x}) + q(\mathbf{x}, \mathbf{X}) + b(\mathbf{x})}_{\mathbf{J}(\mathbf{X})\mathbf{x}} + q(\mathbf{x}, \mathbf{x}). \quad (13)$$

Here we used the linearity of q in both arguments and the linearity of b to define an operator

$$\mathbf{J}(\mathbf{X})\mathbf{x} \equiv q(\mathbf{X}, \mathbf{x}) + q(\mathbf{x}, \mathbf{X}) + b(\mathbf{x}). \quad (14)$$

For finite dimensional systems $\mathbf{J}(\mathbf{X})$ is the Jacobian of (11) evaluated along the trajectory $\mathbf{X}(t)$. In the tangent linear approximation the bilinear term $q(\mathbf{x}, \mathbf{x})$ is neglected and the system

$$\dot{\hat{\mathbf{x}}}^1 = \mathbf{J}(\mathbf{X})\hat{\mathbf{x}}^1 \quad (15)$$

is known as the tangent linear model. We use a hat to indicate that this is only an approximation to the true evolution $\mathbf{x}$ and the reason to add the superscript 1 will become apparent later.

The key observation in this section is that the exact time evolution of perturbations in (13) can also be written as

$$\dot{\mathbf{x}} = \mathbf{J}(\mathbf{X} + \mathbf{x}/2)\mathbf{x}, \quad (16)$$

i.e. we obtain the exact time evolution of perturbations if the tangent linear model is linearized around the trajectory $\mathbf{X} + \mathbf{x}/2$ instead of $\mathbf{X}$. The trajectory $\mathbf{X}+\mathbf{x}/2$ will be referred to as the optimal linearization trajectory. The previous results can be generalized to bilinear partial differential algebraic equations (BPDAE). Let

$$e(\mathbf{X}, \dot{\mathbf{X}}) + d(\dot{\mathbf{X}}) = q(\mathbf{X}, \mathbf{X}) + b(\mathbf{X}) + c, \quad (17)$$

substitution of $\mathbf{X} \to \mathbf{X} + \mathbf{x}$, using (17) and retaining only terms linear in $\mathbf{x}$ gives the TL-model

$$e(\hat{\mathbf{x}}^1, \dot{\mathbf{X}}) + e(\mathbf{X}, \dot{\hat{\mathbf{x}}}^1) + d(\dot{\hat{\mathbf{x}}}^1) = 2s(\mathbf{X}, \hat{\mathbf{x}}^1) + b(\hat{\mathbf{x}}^1).$$

where we used definition 2 to write $q(\mathbf{X}, \hat{\mathbf{x}}^1) + q(\hat{\mathbf{x}}^1, \mathbf{X}) = 2s(\mathbf{X}, \hat{\mathbf{x}}^1)$. It is easy to see that the neglected bilinear terms $e(\mathbf{x}, \dot{\mathbf{x}})$ and $q(\mathbf{x}, \mathbf{x})$ are recovered if the TL-model is linearized around the trajectory $\mathbf{X} + \mathbf{x}/2$. An important difference with the previous result is that to integrate the TL-model both the trajectory $\mathbf{X}$ and the tendencies $\dot{\mathbf{X}}$ are required if $e \neq 0$.

Integrations with the tangent linear model linearized around a trajectory $\mathbf{X} + \mathbf{x}/2$ starting from an initial condition $\mathbf{x}_0$ will be denoted by

$$\mathbf{x}(t) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t, 0)\mathbf{x}_0. \quad (18)$$

Note that, although the tangent linear model is used to propagate the increment $\mathbf{x}_0$, this is not a linear mapping from $\mathbf{x}_0$ to $\mathbf{x}(t)$ due to the dependence of the linearization trajectory on $\mathbf{x}_0$. In appendix C we discuss how to preserve bilinearity when higher than first order integration schemes are used to integrate (11) and show that bilinearity is preserved if a finite dimensional representation of the state vector $\mathbf{X}$ is obtained by truncating the coordinate vector w.r.t. a time independent orthonormal basis.

4 Iterative relinearization

In section 3 we observed that, for a given initial condition perturbation, there is an optimal linearization trajectory for the TL-model such that the tangent linear predictions become exactly equal to the nonlinear predictions. In this section we introduce an iterative method purely based on integrations with the tangent linear model that converges to this optimal trajectory. Section 5 shows how this iterative method can be used to update the linearization trajectory in forecast sensitivity experiments without using the nonlinear model.

We have seen that for bilinear systems

$$\mathbf{x}(t) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0, \tag{19}$$

describes the exact time evolution of perturbations. For a given initial condition perturbation $\mathbf{x}_0$ and a trajectory $\mathbf{X}(t)$ this equation can be written as a map $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}$ that maps increment trajectories to increment trajectories

$$\mathbf{x} = \mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{x}), \tag{20}$$

with $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{x}) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0$, i.e. the trajectory $\mathbf{x}(t)$ is a fixed point of $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}$. If for a fixed time interval $[0,T]$ there is a constant $0 < q < 1$ and a suitable metric d on the space of increments defined on the interval $[0,T]$ such that $d(\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{x}), \mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{y})) \leq q d(\mathbf{x},\mathbf{y})$ then $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}$ is known as a contraction mapping. The Banach fixed point theorem then guarantees that the fixed point $\mathbf{x}$ is unique and more over the iterated map

$$\hat{\mathbf{x}}^k = \mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\hat{\mathbf{x}}^{k-1}), \tag{21}$$

converges to this fixed point. This suggests that given an estimate of the trajectory $\hat{\mathbf{x}}^{k-1}(t)$ the tangent linear model can be integrated in the form

$$\hat{\mathbf{x}}^k(t) = \mathbf{M}_{\mathbf{X}+\hat{\mathbf{x}}^{k-1}/2}\mathbf{x}_0, \tag{22}$$

where the superscripts indicate the iteration number. With $\hat{\mathbf{x}}^0(t) = 0$, the first iteration $k = 1$ is equal to a standard TL-integration (as given by (15)) and gives a trajectory $\hat{\mathbf{x}}^1$. During the second iteration we integrate the tangent linear model with a modified trajectory $\mathbf{X} + \hat{\mathbf{x}}^1/2$ etc. Alternatively the iteration can be started with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ which has the advantage that the time derivatives in the TL-model become exact at $t = 0$. In the experiments both methods are compared.

In appendix D an analysis of (22) for multilinear models is given and we show that, independent of the order of the nonlinearities in the nonlinear model, at convergence (22) always gives better predictions of the time evolution of perturbations than the standard TL-model (15). In particular the bilinear terms are exactly taken into account. In section 6.2 we examine the rate of convergence for the iterated map (21) for the QG and the L96 model.

Remark 1 (Radius of convergence) *Iterated maps can exhibit a finite radius of convergence even though there is a fixed point valid for all t. Therefore, even though the fixed point trajectory $\mathbf{x}(t)$ is valid for all t, this does not imply that the iterated map* (22) *converges to this fixed point. As an example consider the system $\dot{X} = -2tX^2$ with $X(0) = 1$. The solution is given by the Witch of Agnesi $X(t) = \frac{1}{t^2+1}$. The Picard iteration $X^k(t) = X(0) + \int_0^t -2s(X^{k-1}(s))^2 ds$, with $X^0(t) = 1$ converges to the Taylor series of $X(t)$ but because $X(t)$ has poles at $t = \pm i$ the Picard iteration only converges to the fixed point for $|t| < 1$.*

5 Estimation using the inverse TL-model

The estimation problem considered in this paper is: given a forecast starting from an analysis $\mathbf{X}_0$

$$\mathbf{X}_T = \mathcal{M}(T, 0, \mathbf{X}_0), \tag{23}$$

and an analysis $\mathbf{X}_T + \mathbf{x}_T$ valid at time T. Can we determine an analysis increment $\mathbf{x}_0$ such that

$$\mathbf{X}_T + \mathbf{x}_T = \mathcal{M}(t, 0, \mathbf{X}_0 + \mathbf{x}_0). \tag{24}$$

These types of experiments are known as forecast sensitivity experiments and have been studied by Rabier *et al.* (1996); Klinker *et al.* (1998); Pu *et al.* (1997a,b). If the tangent linear assumption is valid we expect

$$\mathbf{x}_T = \mathbf{M}_{\mathbf{X}}(T, 0)\mathbf{x}_0, \tag{25}$$

and therefore we can obtain estimates of $\mathbf{x}_0$ from

$$\hat{\mathbf{x}}_0 = \mathbf{M}_{\mathbf{X}}^{-1}(T, 0)\mathbf{x}_T. \tag{26}$$

Besides giving estimates for $\mathbf{x}_0$ the integration with the inverse tangent linear model can be used to produce estimates of the complete trajectory $\mathbf{x}(t)$. The result from section 4 therefore suggests to use this method iteratively

$$\hat{\mathbf{x}}^k(t) = \mathbf{M}^{-1}_{\mathbf{X}+\hat{\mathbf{x}}^{k-1}/2}(T, t)\mathbf{x}_T, \tag{27}$$

with $\hat{\mathbf{x}}^0(t) = 0$ or $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$. This defines an iterated map on the space of increments trajectories

$$\hat{\mathbf{x}}^k = \mathbf{S}^{\mathbf{X}}_{\mathbf{x}_T}(\hat{\mathbf{x}}^{k-1}), \tag{28}$$

where $\mathbf{S}^{\mathbf{X}}_{\mathbf{x}_T}(\hat{\mathbf{x}}^{k-1}) = \mathbf{M}^{-1}_{\mathbf{X}+\hat{\mathbf{x}}^{k-1}/2}\mathbf{x}_T$. In section 6.3 the convergence rate of the iterated map (28) is investigated for the L96 model.

6 Applications

6.1 Indices

To highlight different aspects of the optimal linearization trajectories and the iterative relinearization method the exact time evolution of the perturbations $\mathbf{x}(t)$ and the corresponding tangent linear evolution $\hat{\mathbf{x}}^k(t)$ are compared using five indices l^k, α^k, R^k and d^k and R^k_d. The similarity index $l^k(t)$ is defined as

$$l^k(t) = \frac{\big(\mathbf{x}(t), \hat{\mathbf{x}}^k(t)\big)}{\|\mathbf{x}(t)\|\,\|\hat{\mathbf{x}}^k(t)\|}, \tag{29}$$

the angle $\alpha^k(t)$ is given by

$$\alpha^k(t) = \operatorname{acos}\left(l^k(t)\right), \tag{30}$$

the relative norm $R^k(t)$ is given by

$$R^k(t) = \frac{\|\hat{\mathbf{x}}^k(t)\|}{\|\mathbf{x}(t)\|}, \tag{31}$$

the error norm $d^k(t)$ by

$$d^k(t) = \|\mathbf{x}(t) - \hat{\mathbf{x}}^k(t)\|, \tag{32}$$

and the relative error norm $R_d(t)$ by

$$R^k_d(t) = \frac{\|\mathbf{x}(t) - \hat{\mathbf{x}}^k(t)\|}{\|\mathbf{x}(t)\|}. \tag{33}$$

For the QG-model the values of d^k, l^k, R^k and R^k_d are determined using the kinetic energy inner product. For the L96 model the Euclidean inner product is used. In the context of twin experiments values of $l = 0.7$, corresponding to an angle $\alpha = 45°$, are commonly used to indicate that the TL assumption is violated (e.g. Gilmour *et al.* (2001)).

We will say that $\hat{\mathbf{x}}^k(t)$ is more similar to $\mathbf{x}$ than $\hat{\mathbf{x}}^{k-1}(t)$ at time t if $\alpha^k(t) < \alpha^{k-1}(t)$ or, equivalently, if $l^k(t) > l^{k-1}(t)$. We say that $\hat{\mathbf{x}}^k(t)$ is closer to $\mathbf{x}$ than $\hat{\mathbf{x}}^{k-1}(t)$ at time t if $d^k(t) < d^{k-1}(t)$ or, equivalently, if $R^k_d(t) < R^{k-1}_d(t)$.

6.2 Iterative relinearization

In this section the rate of convergence of the iterated map $\mathbf{T}$ is examined in a quasi geostrophic model (described in appendix A) and the Lorenz 96 model (described in appendix B).

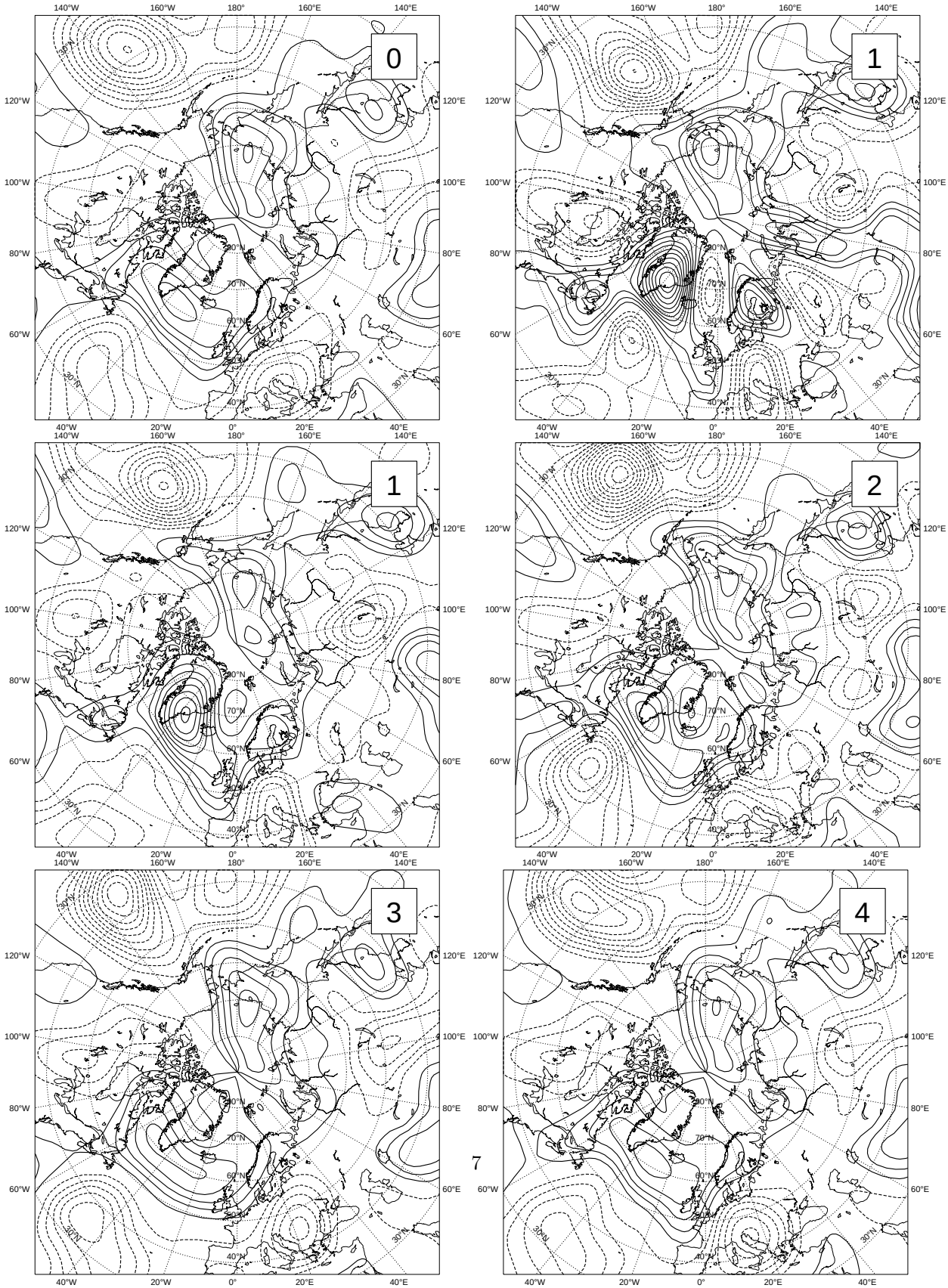

Figure 1: Stream function perturbation at 500 hPa after 2 days using the nonlinear model $\mathbf{x}$ (top left), the standard tangent linear model with $\hat{\mathbf{x}}^0(t) = 0$ (top right) and the iterative relinearization method with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ for iteration 1 (middle left) to iteration 4 (bottom right). The initial condition for the perturbed run and the control run are 100 days apart. The contour interval is $1 \cdot 10^{-3}\ \Omega a^2$ in all panels (with a and Ω the average radius and the angular velocity of the Earth, respectively). Positive values (solid) negative values (dashed).

6.2.1 QG model

In figure 1 we show the 2 day forecast difference of the stream function at 500 hPa (upper left). The initial condition for the control run and the perturbed run are 100 days apart and therefore we may assume that they are uncorrelated (see also figure 9). The size of the perturbations used in these experiments is therefore much larger than typical analysis increments. In the upper right we show the forecast of the standard TL-model $\hat{\mathbf{x}}^0(t) = 0$.

The other panels show the iterative method for 4 iterations with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$. Both the standard TL-integration ($l^1 = 0.55$) and the first iteration with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ ($l^1 = 0.73$) differ substantially from the truth, with large differences north of 60°N. In the first iteration with $\mathbf{x}^0 = 0$ there is a wave pattern over the North Atlantic ocean which is absent in the first iteration with $\mathbf{x}^0 = \mathbf{x}_0$. At subsequent iterations all positive and negative cells are gradually moved to their correct location and with the correct amplitude. At iteration 2 to 4 we have $l^2 = 0.90$, $l^3 = 0.95$ and $l^4 = 0.99$ respectively, indicating that the iterative method converges quickly with the largest improvement when going from iteration 1 to 2.

Figure 2 shows the similarity index l^k and the relative error norm R_d^k as a function of time and iteration number. The solid black line refers to the standard TL-model with $\hat{\mathbf{x}}^0(t) = 0$. The colored lines show the iterative relinearized results for 4 iterations with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$. The control run and the perturbed run are 2 days apart. From the standard TL-integration we see that the duration of the TL-regime is slightly larger than 1 day. Especially in the short range it is beneficial to use $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ because the derivatives in the TL-model become exact at $t = 0$. In figure 2 (right) this can be seen for example from the relative error norm where $\dot{R}_d(0) \neq 0$ when the standard TL-model is used. Observe that the iterative method adds approximately 0.5 days to the usefulness of the TL-model at each iteration.

6.2.2 L96 model

Figure 3 shows the similarity index and relative error norm (average over 50 experiments) as a function of time and iteration index for the L96 model. All experiments start with a random initial condition perturbation with $\|\mathbf{x}_0\| = 10$. Such an initial condition amplitude is approximately equal to the size of 12 hour forecast differences (see figure 10). From the first iteration using $\hat{\mathbf{x}}^0(t) = 0$ (black) we see that the duration of the tangent linear regime is slightly larger than 1.5 days (0.3 time units). Using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ this can be extended to 2 days. The iterative linearization method converges to the true increment at subsequent iterations. For a 2 day forecast (0.4 time units) of the order of 4 iterations are required to converge to the true time evolution of the increment, with the largest improvements when going from iteration 1 to 2. For longer lead times more iterations are needed. This is related to the fact that the TL-model produces large increments beyond the duration of tangent linear regime (see also figure 7). Therefore the corrections $\hat{\mathbf{x}}^1/2$ used in the second iteration are actually deteriorating the linearization trajectory at the end. As a result of this the second iteration is further away from the truth at the end of the optimization window, even though it is more similar to the truth. In section 7.2 we discuss a method to regularize this behaviour without affecting the fixed point of the iterated map.

6.3 Estimation using the inverse TL-model

Here we examine the iterated map (28) from section 5 The action of $\mathbf{M}_{\mathbf{X}}^{-1}$ on a vector $\mathbf{x}_T$ is obtained by integrating the tangent linear model backwards in time. In the L96 model the fourth order Runge Kutta (RK4) scheme is used to propagate the state. Theoretically the backwards integration requires the use of the inverse integration scheme (which will be an implicit scheme) to ensure $\mathbf{M}_{\mathbf{X}}\mathbf{M}_{\mathbf{X}}^{-1} = \mathbf{I}$. Here the adjoint of the RK4 scheme is used to integrate the TL-model backwards in time. In the L96 model we find experimentally that the angle between $\mathbf{M}_{\mathbf{X}}^{-1}\mathbf{M}_{\mathbf{X}}\mathbf{x}_0$ and $\mathbf{x}_0$ is of the order $O(10^{-3})$ degrees and the relative norm $\|\mathbf{M}_{\mathbf{X}}^{-1}\mathbf{M}_{\mathbf{X}}\mathbf{x}_0\|/\|\mathbf{x}_0\| - 1 \approx O(10^{-6})$ for an optimization time of 0.6 time units (3 days). So it appears that $\mathbf{M}_{\mathbf{X}}^{-1}\mathbf{M}_{\mathbf{X}}$ is close to the identity operator. We conclude that the adjoint RK4 scheme can

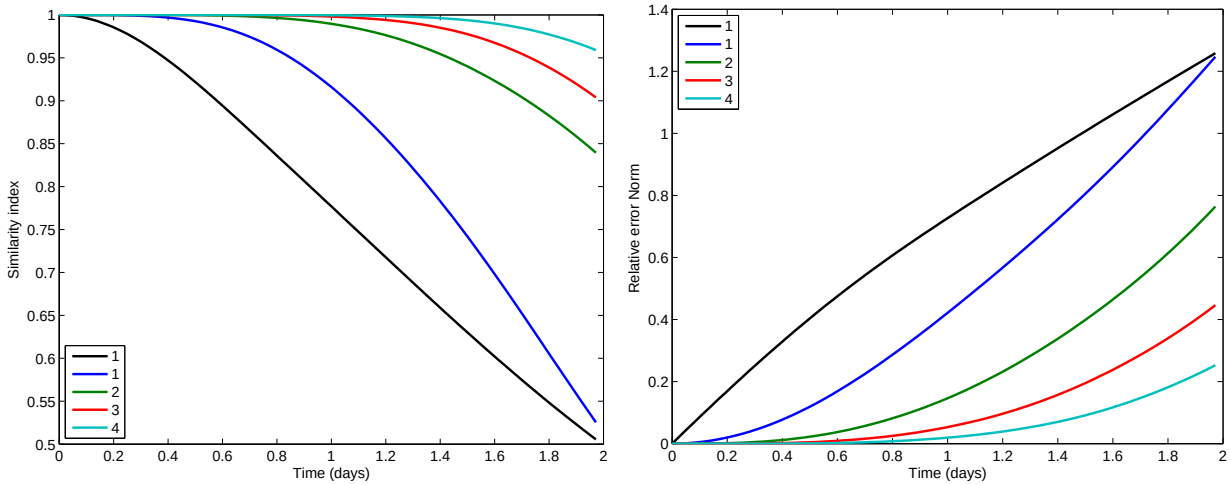

Figure 2: Similarity index $l^k(t)$ (left) and relative error norm $R_d^k(t)$ (right) as a function of time and iteration number for the QG model. Shown are average values for 20 experiments. The black line is the standard TL-model with $\hat{\mathbf{x}}^0(t) = 0$. The colored lines are iteration 1 to 4 with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$. The control run and perturbed run are two days apart.

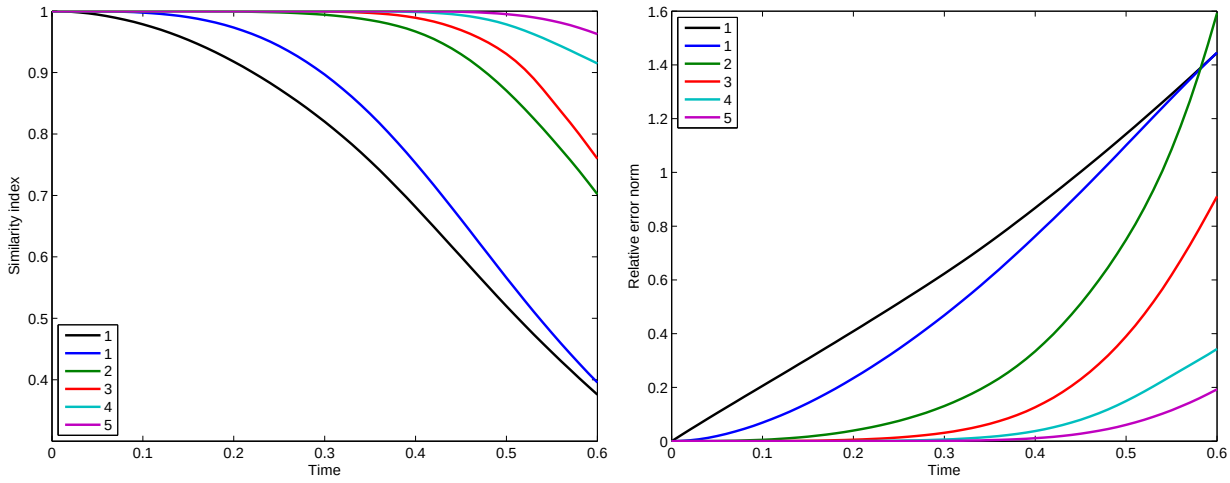

Figure 3: Similarity index $l^k(t)$ (left) and relative error norm $R_d^k(t)$ (right) as a function of time for the L96 model. Shown are average results for 50 experiments using random initial condition perturbations with $\|\mathbf{x}_0\| = 10$. The black line is the result for iteration 1 with $\hat{\mathbf{x}}^0(t) = 0$ the colored lines for iteration 1 to 5 with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$

be used for the inverse integrations.

Figure 4 shows the result when we iteratively solve (27) using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$. Even though the estimate from the first iteration differs substantialy from the truth with $l_1 = 0.4$, the method quickly converges and the subsequent iterations are more similar and closer to the truth. Approximately 4 iterations are required to obtain an almost perfect estimate. Note that during the inverse integration we also obtain the corrections needed for the next iterations. Therefore the computational cost is equal to four TL-integrations (backwards). This cost should be compared to the alternative of solving this estimation problem in terms of a cost function minimization (e.g. 4D-VAR) where a single inner loop iteration already involves two linear integrations (1 adjoint and 1 tangent linear integration). For comparison figure 4 also includes the result when the standard TL-model, i.e. $\hat{\mathbf{x}}^0(t) = 0$, is used to propagate the increment backwards in time (the black line). If $l = 0.7$ is used as threshold value then the gain of using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$ in the first iteration is 0.13 time units (0.65 days). From the time evolution of the error norm (figure 4 right) we see that this gain is mainly a result of the fact that, by using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$, the time derivatives at $t = 0.4$ become exact in the TL-model and thus $\dot{d}_1 = 0$ at $t = 0.4$. In particular for large perturbations we therefore expect to benefit from using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$.

From figure 4 it is also clear that for long optimization windows the estimated increment at $t = 0$ from the first iteration becomes uncorrelated with the true increment. As a result, the nonlinear forecast starting from $\mathbf{X}_0 + \hat{\mathbf{x}}_0^1$ bears low similarity to the truth (dashed lines in figure 4). For long windows therefore, the nonlinear model starting from $\mathbf{X}_0 + \hat{\mathbf{x}}_0^1$ can not be used to update the linearization trajectory. In a forthcoming paper applying optimal linearization trajectories in the context of 4D-VAR we will show that also in 4D-VAR it is better to update the linearization trajectory using the TL-model.

If we use $l < 0.7$ to indicate the breakdown of the TL-assumption, figure 4 indicates that the tangent linear assumption linearized around the control run is valid for 0.15 time units (i.e. from $t = 0.4$ to $t = 0.25$). This should be compared with the forward integration in figure 3 where the value 0.7 is reached after 0.3 time units. The duration of the TL-regime is shorter for inverse integrations. Partly this is a result of the fact that error growth in the backward integration is characterized by the reciprocal singular value spectrum and these values are larger than the singular values (see figure 6). Another reason is that typically $\|\mathbf{x}_T\| > \|\mathbf{x}_0\|$ and therefore the backward integration is started with larger initial conditions. integration is determined by the reciprocal of the singular values and these are in general larger than the singular values (see figure 6) and partly because we created $\mathbf{x}_T$ using the nonlinear model with a perturbations $\|\mathbf{x}_0\| = 10$ and in general we have $\|\mathbf{x}_T\| > \|\mathbf{x}_0\|$. Therefore we started the inverse integration with a large perturbations and for large perturbations the TL-assumption is violated on shorter time spans. The idea of using the inverse of the TL-model has been studied in Pu *et al.* (1997a) using a method called the quasi inverse. They reversed the sign of the dissipation terms in the TL-model as a form of regularization. As will be discussed in section 7.2 on the regularized prediction experiments, there is no need for bilinear systems to add regularization when the optimal linearization trajectory is used. Therefore the amount of regularization should depend on how close we are to the optimal linearization trajectory. If the linear term b in the nonlinear model (11)) is a purely dissipative term, i.e. $(\mathbf{X}, b(\mathbf{X})) < 0$, then the TL-model can be integrated in the form

$$\dot{\hat{\mathbf{x}}}^k = \mathbf{J}(\mathbf{X} + \hat{\mathbf{x}}^{k-1}/2)\hat{\mathbf{x}}^k + \alpha b(\hat{\mathbf{x}}^{k-1} - \hat{\mathbf{x}}^k) \qquad (34)$$

The choice $\alpha = 2$ amounts to reversing the sign of the dissipation terms (see also equation (13)) during the first iteration. However at subsequent iterations, at locations in space and time where the solution has converged, the unmodified TL is used.

6.4 Identification of bilinear systems

For bilinear systems the time evolution of the increment $\hat{\mathbf{x}}(t)$ in the TL-model linearized around the tra-

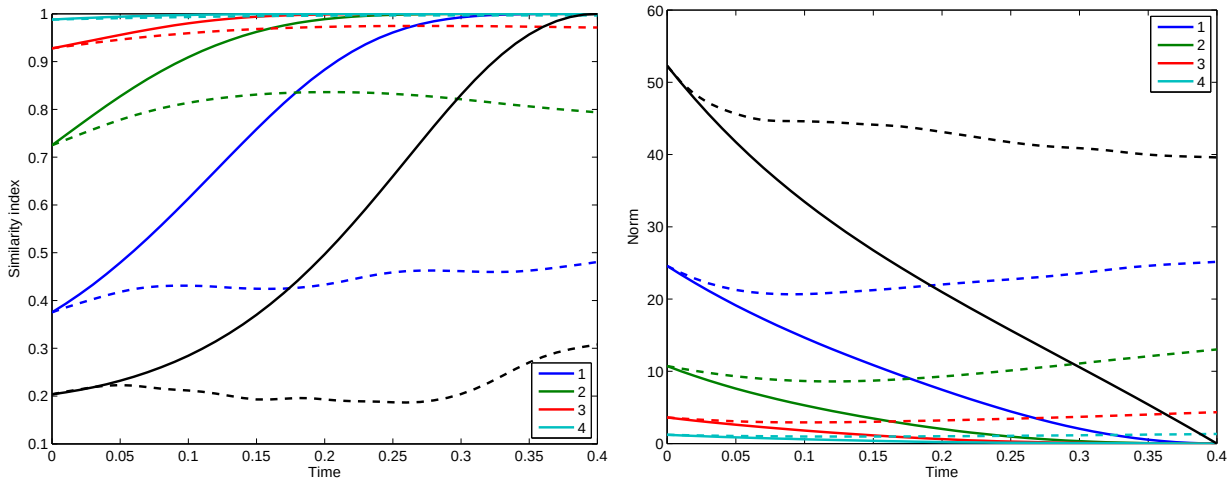

Figure 4: Similarity index (left) and error norm $d_k = \|\mathbf{x} - \hat{\mathbf{x}}^k\|$ (right) as a function of time for the inverse TL-model (solid) and the NL-model starting from $\mathbf{X}_0 + \hat{\mathbf{x}}_0^k$ (dashed). Shown are average results over 50 experiments with an OT=0.4 and random initial condition perturbations with norm $\|\mathbf{x}_0\| = 10$. The black line is the first iteration with $\mathbf{x}^0(t) = 0$, i.e the standard TL-model

jectory $\mathbf{X} + \mathbf{x}/2$ given by

$$\hat{\mathbf{x}}(t) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t, 0)\mathbf{x}_0, \tag{35}$$

is equal to the time evolution according to the nonlinear model: $\hat{\mathbf{x}}(t) - \mathbf{x}(t) = 0$. Therefore a necessary condition for the model $\mathcal{M}$ to be a bilinear system is that the error norm (or equivalent the relative error norm) is zero:

$$d(t) = R_d(t) = 0. \tag{36}$$

However, numerical integrations will be subject to round off error leading to nonzero values for d and R_d. To highlight different aspects, the time evolution of perturbations is examined in terms of the angle α (30) and the relative norm R (31). Note that $\alpha = 0$ and $R = 1$ if and only if $R_d = d = 0$. In the following sections we study the behaviour of R_d, α and R in the QG and L96 model.

6.4.1 QG model

Figure 5 shows the relative error norm R_d, the relative norm R and angle α as a function of time for the QG model for 10 experiments. The control run is obtained by integrating the nonlinear model for 300 days. Continuing the integration for another 300 days yields the perturbed run. The trajectory for the second experiment starts using the final condition of the previous perturbed run and so forth. Due to the long integration times the initial condition for the tangent linear model is given by the difference between two uncorrelated state vectors on the model attractor and is therefore larger than typical analysis increments. For these large amplitude perturbations the TL approximation is valid for 1 day.

The 10 experiments show exponential growth of the relative norm R after day 210. Before day 210 both $\alpha \approx 0$ and $R \approx 1$ and we conclude that the TL-model can be used for lead times shorter than 210 days. The time evolution of the relative error norm R_d (figure 5, left) shows no signal at day 210. Instead it merely indicates exponential growth beyond day 10 with an exponent of 0.148 day^{-1} (standard deviation 0.005 in 10 experiments) corresponding to an error doubling time of $\tau_d = 4.7$ days. Note that this is longer than the error doubling time based on linearization

of the TL-model around a control run, which gives a Lyapunov exponent of $\sigma = 0.254$ (with standard deviation 0.014 in 10 experiments) and a corresponding error doubling 2.7 days. This is in agreement with other studies, e.g. Swanson *et al.* (1998), where an approximate value of 3 days is given. The increase of the error doubling time when we linearize around the average trajectory of the control and perturbed run is consistent with Hoskins *et al.* (2000) who determined singular vector growth using different linearization trajectories in the TL-model. They found that the dominant factor for singular vector growth is the dynamic structure of the linearization trajectory and, in particular, its smoothness.

From figure 5 we see that the time evolution of the relative error norm R_d is approximately exponential beyond day 10. This suggests that we can model the time evolution of R_d for $t > 10$ days by

$$\dot{R}_d = \sigma R_d. \tag{37}$$

The values of σ and $R_d(0)$ are estimated using linear least squares on the experimental values of $\ln(R_d(t))$. The solid line in figure 5 (left) show the predictions of this model with the estimated values $R_d(0) = 2.9 \cdot 10^{-14}$ (standard deviation $2.2 \cdot 10^{-14}$ in 10 experiments) and $\sigma = 0.148$ (standard deviation 0.005 in 10 experiments). With the additional assumption that the error vector $\hat{\mathbf{x}} - \mathbf{x}$ is perpendicular to $\mathbf{x}$ the modelled time evolution of R_d can be used to predict values of the angle α and the relative norm R (solid lines in the middle and right panel of figure 5). We emphasize that the solid lines in the middle and right panel of figure 5 are not fitted to the experimental data but purely a result of the geometric assumption that the error vector $\hat{\mathbf{x}} - \mathbf{x}$ is perpendicular to $\mathbf{x}$. Experimentally we find that the angle between $\mathbf{x}$ and $\mathbf{x} - \hat{\mathbf{x}}$ is 89.6° with a standard deviation of 8.1°.

With the assumption that the error vector $\hat{\mathbf{x}} - \mathbf{x}$ is perpendicular to $\mathbf{x}$ the condition $\alpha = 45°$ is equivalent to the condition $R_d = 1$. Setting $R_d = 1$ in the error growth model gives the estimate

$$t_\mathrm{p} = \frac{-1}{\sigma} \ln R_d(0) = 212 \pm 3.8 \text{ days}. \tag{38}$$

This estimate is plotted in the left panel of figure 5. The same estimate is obtained from $\alpha = 45$ and $R = \sqrt{2}$. Note that in the absence of round-off error $\mathbf{x} = \hat{\mathbf{x}}$ and as such there is no reason to prefer the NL over the TL integration. Therefore these results also put a predictability limit on the nonlinear model due to round-off error of 212 days.

6.4.2 L96 model

In the L96 model we obtain the estimates $\hat{R}_d(0) = 4.0 \cdot 10^{-16}$ and $\sigma = 0.233$ day^{-1} equivalent to an error doubling time of 2.97 days. The error doubling times are higher than estimates based on the Lyapunov exponent (2.1 days) consistent with the reduced growth of singular vectors for smooth trajectories in Hoskins *et al.* (2000). The figures for R_d, α and R are similar to the results for the QG-model (not shown) For the L96 model the TL-model can be used for $t_\mathrm{p} = \frac{-1}{\sigma} \ln R_d(0) = 152$ days.

7 Discussion

7.1 Prospects for using the method in NWP

We demonstrated the advantage of using the optimal linearization trajectories in the context of two simple bilinear models. Although the analysis in appendix D shows that, independent of the order of the nonlinearities in the nonlinear model, the iteratively relinearized TL-model always gives better results at convergence, to get an exact correspondence between the TL and the NL model, the NL-model has to be bilinear. In example 2 it was shown that it is possible to transform multilinear systems to bilinear systems by augmenting the state vector.

There are other situations where apparent "infinite" order nonlinearities can be transformed to bilinear terms. Let $\dot{X} = e^{\alpha X}$ and define $Y = e^{\alpha X}$ then $\dot{X} = Y$ and $\dot{Y} = \alpha Y^2$ which is a bilinear system. One difference between the reduction of multilinear systems (appendix D) to bilinear system in example 2 is that in this case the newly introduced variable Y has to be a prognostic variable because the algebraic constraint $0\dot{Y} = Y - e^X$ is not bilinear and can therefore not be used. Similary it can be shown that

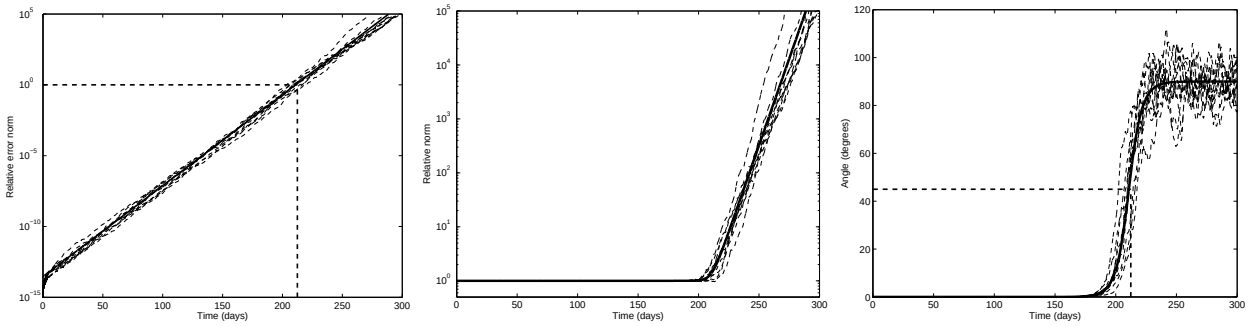

Figure 5: Relative error norm R_d (left), relative norm R (middle) and angle α (right) for the QG model with individual experiments dashed. The solid line in the left panel is the estimate R_d using (37) with $R_d(0) = 2.9 \cdot 10^{-14}$ and $\sigma = 0.148$. In the middle panel and right panel the solid lines are estimates using the assumption that the error vector is perpendicular to $\mathbf{x}$.

$\dot{X} = \sin(X)$ (Define $Y = \sin(X)$ and $Z = \cos(X)$), $\dot{X} = \ln X$ (define $Y = \ln X$) and $\dot{X} = X^\alpha$ (define $Y = X^{\alpha-1}$) can be written as bilinear systems. Although this does not show that realistic NWP models can be formulated as bilinear systems, it illustrates that both multilinear models and models that contain "infinite" order nonlinearites can be written as a bilinear system and demonstrates that the class of bilinear systems is very general. In a forthcoming paper we show that the restriction to bilinear systems can be lifted if the tangent linear model is linearized around an ensemble of trajectories simultaneously.

7.2 Regularized relinearization in the L96 model

The TL-model produces large increments for long lead times (see figure 7). This will deteriorate the linearization trajectory for the next iterations. In principle, this can be solved by increasing the dissipation in the TL-model, however, in that case the solution would no longer converge to the true solution during the iterative process. Here we propose to add a term $\alpha(\hat{\mathbf{x}}^{k-1} - \hat{\mathbf{x}}^k)$ to the TL-model leading to

$$\dot{\hat{\mathbf{x}}}^k = \mathbf{J}(\mathbf{X} + \hat{\mathbf{x}}^{k-1}/2)\hat{\mathbf{x}}^k + \alpha(\hat{\mathbf{x}}^{k-1} - \hat{\mathbf{x}}^k) \tag{39}$$

So dissipation is added to the model but at the same time the previous iteration is used as a forcing in the TL-model. At convergence of the algorithm $\hat{\mathbf{x}}^{k-1} = \hat{\mathbf{x}}^k$ and the added term becomes zero, i.e. the added term does not modify the fixed point of the iterated map $\mathbf{T}$ (21). In general α could be an operator (see also section 6.3), here we only discuss the situation where α is a scalar.

Using $\hat{\mathbf{x}}^0 = 0$ the first iteration is given by

$$\hat{\mathbf{x}}^1(t) = \mathbf{M}_\mathbf{X}(t,0)e^{-\int_0^t \alpha dt'}\mathbf{x}_0 \tag{40}$$

Where $\mathbf{M}_\mathbf{X}(t,0)$ is the propagator for the TL-model with $\alpha = 0$. If $\mathbf{M}_\mathbf{X} = \mathbf{UDV}^T$ is the singular value decomposition of $\mathbf{M}_\mathbf{X}$ we obtain

$$\hat{\mathbf{x}}^1(t) = \mathbf{U}(\mathbf{D}e^{-\int_0^t \alpha dt'})\mathbf{V}^T\mathbf{x}_0 \tag{41}$$

So the added term has no impact on the singular vectors but it changes the singular value spectrum. Let $\sigma_{\max}(t)$ denote the leading singular value of $\mathbf{M}_\mathbf{X}(t,0)$. By choosing α such that

$$\int_0^t \alpha dt' > \ln \sigma_{\max}(t) \tag{42}$$

we conclude that $||\hat{\mathbf{x}}^1(t)|| < ||\mathbf{x}_0||$ for all $\mathbf{x}_0$. In figure 6 we show the leading singular value as a function of the optimization time and the value for α when α is kept constant during the optimization window $\alpha = \frac{1}{t}\log\sigma_{\max}(t)$.

Figure 7 shows the impact of the added term by examining the norm $\|\hat{\mathbf{x}}^k\|$ as a function of time for

$\alpha = 0$ and $\alpha = 8$. The iterative method still convergences to the true solution but in a more controlled manner. At the first iteration the norm decreases monotonically as expected. At subsequent iterations the forcing ensures that we still converge to the true solution.

In NWP models we know that at each grid point in the integration domain the density ρ, absolute temperature T, pressure p and the specific humidity q are all positive quantities. Tangent linear integrations do not respect these types of constraints and therefore it is possible that in the linearization trajectory $\mathbf{X}+\mathbf{x}^{k-1}/2$ some of these variables are negative. We therefore suggest to use a projection operator $\mathbf{P}$ that sets negative values of ρ, T, p and q to zero and integrate the tangent linear model in the form

$$\hat{\mathbf{x}}^k(t) = \mathbf{M}_{\mathbf{P}(\mathbf{X}+\mathbf{x}^{k-1}/2)}\mathbf{x}_0 \tag{43}$$

Being solutions of the nonlinear model the trajectories $\mathbf{X}$ and $\mathbf{X}+\mathbf{x}$ do not contain negative values for ρ, T and q. At convergence of the iterated map the linearization trajectory $\mathbf{X}(t)+\mathbf{x}(t)/2$ is the average of $\mathbf{X}(t)+\mathbf{x}(t)$ and $\mathbf{X}(t)$ and therefore the linearization trajectory does not contain negative values for ρ, T, p and q, i.e. the projection operater does not modify the fixed point of the iterated map but ensures that during the iterations only "physically consistent" trajectories are used.

7.3 Identification of multilinear system

In section 6.4 we introduced a necessary condition (36) for a nonlinear model to have at most bilinear terms. Here we illustrate that this condition can be used to detect higher order multilinearities.

Consider the L96 model with modified dissipation

$$\dot{\mathbf{X}} = q(\mathbf{X},\mathbf{X}) - \mathbf{X} + \mathbf{F} + \alpha(1 - \frac{\|\mathbf{X}\|^2}{\|\mathbf{F}\|^2})\mathbf{X} \tag{44}$$

Where $q(\mathbf{X},\mathbf{X})$ is given in appendix B, $\mathbf{F} = 8$ and $\alpha \geq 0$. For $\alpha = 0$ we recover the L96 model and dissipation is linear. For $\alpha = 1$ the dissipation is a purely trilinear term and dependent on the total energy in the system. The factor $\|\mathbf{F}\|^2$ is introduced

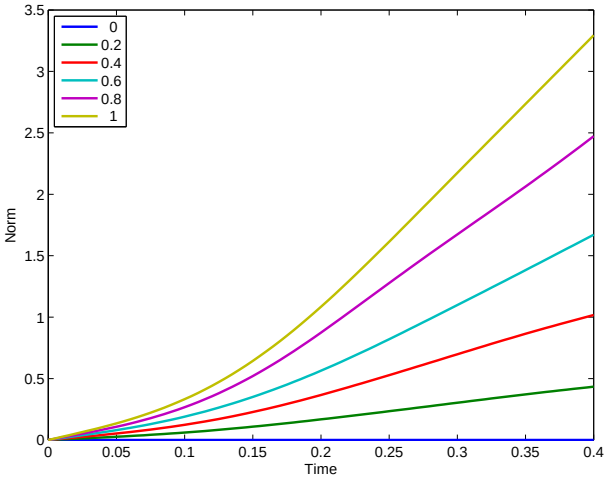

Figure 8: Error norm $\|\mathbf{x}(t) - \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0\|$ as a function of time for the trilinear Lorenz model for $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Shown are average results for 50 experiments

to ensure that the (unstable) steady state solution $\mathbf{X}^* = \mathbf{F}$ for the case $\alpha = 0$ is also a (unstable) steady state for $\alpha \neq 0$. For $\alpha \neq 0$ the additional steady state solutions are $\mathbf{X}^* = -\frac{\mathbf{F}}{2}\left(1 \pm \sqrt{1 - 4/\alpha}\right)$. For $0 < \alpha < 4$ the last expression gives two complex conjugate steady state solutions which can not be reached if we start with a real valued initial condition. The time derivative of the total energy is

$$\begin{aligned}\dot{E} &= -\|\mathbf{X}\|^2 + (\mathbf{X},\mathbf{F}) + \alpha(1 - \frac{\|\mathbf{X}\|^2}{\|\mathbf{F}\|^2})\|\mathbf{X}\|^2 \\ &\leq -\|\mathbf{X}\|^2 + \|\mathbf{X}\|\|\mathbf{F}\| + \alpha(1 - \frac{\|\mathbf{X}\|^2}{\|\mathbf{F}\|^2})\|\mathbf{X}\|^2\end{aligned}$$

For points outside the sphere with radius $\|\mathbf{F}\|$ we therefore have $\dot{E} < 0$ and we conclude that all trajectories eventually enter this ball and cannot escape afterwards.

We expect that for nonzero values of α we have $\|\mathbf{x}(t) - \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0\| > 0$ and this is indeed what we observe (see figure 8). This shows that nonbilinearity can be identified purely based on the model output and might be useful in realistic NWP models

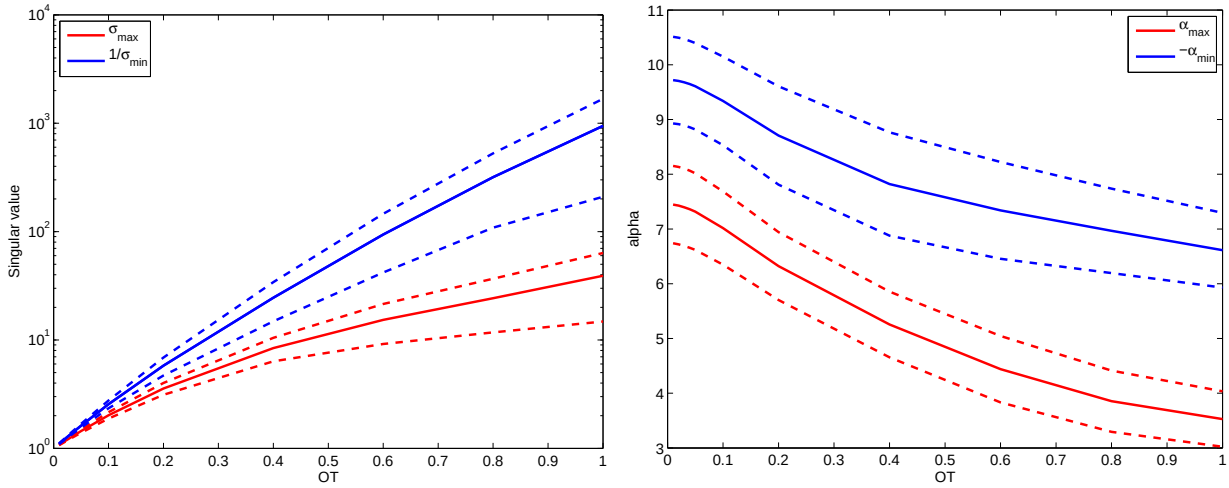

Figure 6: Average leading singular value $\sigma_{\max}$ and reciprocal of the trailing singular value $1/\sigma_{\min}$ (left) and corresponding mean values of α for the regularized prediction $\alpha_{\max}$ and regularized estimation $\alpha_{\min}$ values as a function of optimization time (right) for the Lorenz model. In both plots the range indicates the standard deviation in 50 experiments.

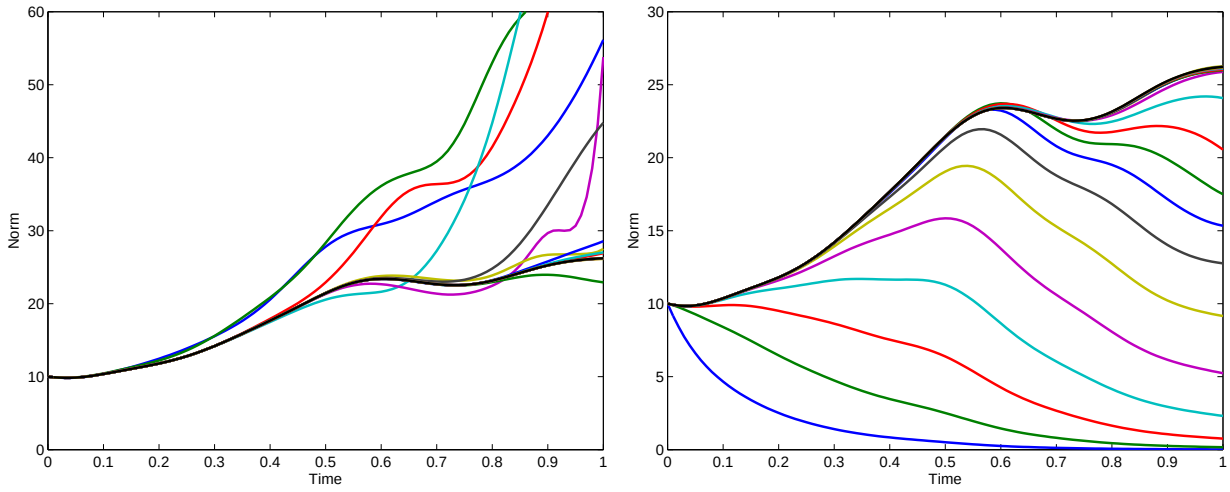

Figure 7: The norm $\|\hat{\mathbf{x}}^k\|$ as a function of time without regularization (left) and with regularization (right) using $\alpha = 8$. The black line is $\|\mathbf{x}\|$. The colored lines are the values for $\|\hat{x}^k\|$. Initial condition perturbation is random with norm $\|\mathbf{x}_0\| = 10$

where analysing the code to determine nonbilinearity might be prohibitive.

8 Conclusions

The nonlinearities in fluid dynamics as a result of the advection part of the total derivative and the use of algebraic constraints such as the ideal gas law give rise to bilinear differential equations. We have shown that for bilinear systems there exists an optimal linearization trajectory for the tangent linear model, such that the TL-model predicts the exact time evolution of the perturbations. Using a quasi geostrophic model and the Lorenz 96 model we showed that, when the optimal linearization trajectory is used, the tangent linear model can be used for more than 200 days in a quasi geostropic model and more than 150 days in the Lorenz 96 model.For bilinear systems therefore one of the major limitations to the application of linear models mentioned in the introduction can be eliminated by linearizing around the optimal linearization trajectory.

We introduced an iterative method that, based purely on TL-integrations, converges to this optimal linearization trajectory. We showed that the optimal linearization trajectory is a fixed point of this iterative method and using prediction experiments in the QG and L96 model we showed that the iterative method converges to the fixed point. In the discussion we introduced a method to regularize the error growth in the TL-model without affecting the fixed point of the iteration.

The main conclusion from this paper is that this iterative method can be used in estimation problems to account for nonlinearity without using the nonlinear model. In particular when long windows are used in forecast sensitivity experiments the estimated increment at $t = 0$ will be uncorrelated to the true increment and the nonlinear model can not be used to update the linearization trajectory. Using forecast sensitivity experiments in the Lorenz 96 model where we iteratively use the inverse of the tangent linear model we showed that the iterative method can be used for long windows and converges quickly. Typically 4 iterations (computation cost equal to 4 integrations with the linear model) are needed to find the optimal corrections for a two day forecast. In a forthcoming paper we will show that the same ideas can be used in multi incremental 4D-VAR.

9 Acknowledgements

We would like to thank Wim Verkley, Theo Opsteegh and two anonymous reviewers for carefully reading earlier versions of the manuscript.

A Quasi-geostrophic model

Marshall and Molteni (1993) introduced a spectral three-level quasi-geostrophic (QG) model with global domain and pressure as the vertical coordinate. The model is truncated at wave number 21 and the model levels are at 200 (level 1), 500 (level 2) and 800 hPa (level 3). The model integrates the system

$$\begin{aligned}
\frac{\partial q_1}{\partial t} &= -J(\psi_1, q_1) - D_1(\psi_1, \psi_2) + S_1 \\
\frac{\partial q_2}{\partial t} &= -J(\psi_2, q_2) - D_2(\psi_1, \psi_2, \psi_3) + S_2 \\
\frac{\partial q_3}{\partial t} &= -J(\psi_3, q_3) - D_3(\psi_2, \psi_3) + S_3,
\end{aligned}$$

where q_i is potential vorticity (PV), ψ_i streamfunction, D_i are linear operators that represent dissipative terms, S_i are constant PV sources and J the Jacobian of a two dimensional field. We refer to Marshall and Molteni (1993) for a complete description of the model.

Figure 9 shows the norm $\|X(t) - X(t - \delta t)\|$ as a function of δt averaged over 1 year for the quasi geostrophic model. In Bengtsson *et al.* (2008, their figure 3) a similar picture is shown for the RMSE of the geopotential height at 500 hPa for the ECMWF model but based on analyses instead of forecasts. If the trend due to seasonal variability is removed in the ECMWF model the RMSE reaches a maximum of 110.8m and the RMSE of analyses one day apart is 61m, i.e. at 1 day the error is already half of the value reached for large δt. The QG model the L96

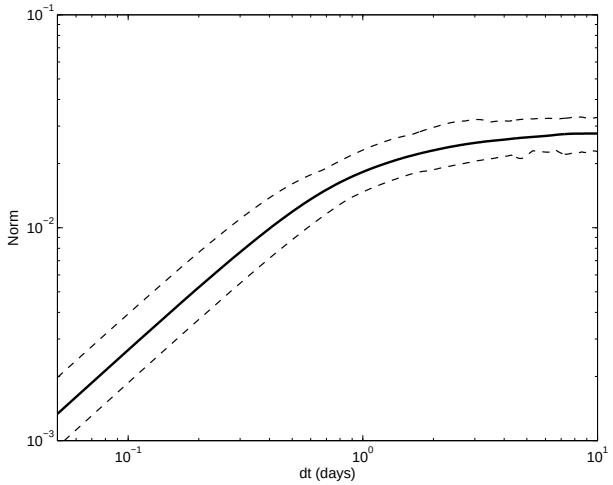

Figure 9: The norm $\|\mathbf{X}(t) - \mathbf{X}(t - \delta t)\|$ as a function of δt averaged over 1 year for the quasi geostrophic model. The solid line is the average the dashed lines are the maximum and minimum value that occurred during the 1 year period.

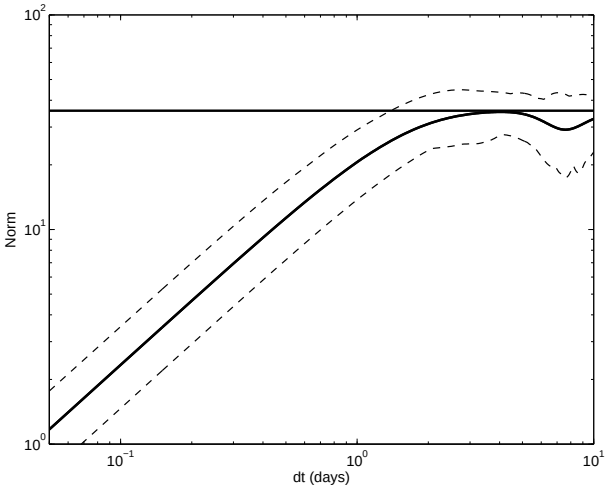

Figure 10: The norm $\|\mathbf{X}(t) - \mathbf{X}(t - \delta t)\|$ as a function of δt averaged over 1 year on a loglog scale for the L96 model. The solid line is the average and the dashed lines are the maximum and minimum value of the norm that occurred during the 1 year period. The straight line is the estimated value $\sqrt{2}\|\frac{\mathbf{F}}{2}\|$ based on the sphere $\mathcal{C}$ given by (48). For ease of comparison with figure 9 the time axis is scaled such that 1 time unit is 5 days.

model (figure 10) and the ECMWF model therefore show similar behaviour in this respect. Both in the QG model and the L96 model the growth of the error norm saturates at $\delta t = 10$ days.

B Lorenz 96 model

Lorenz (1996) introduced a simple system of the form (50) with $q^i_{i-1,i+1} = 1$, $q^i_{i-2,i-1} = -1$, all other $q^i_{jk} = 0$, $b^i_j = -\delta^i_j$ the Kronecker delta and $c^i = F$. Giving the system

$$\dot{X}^i = X^{i-1}(X^{i+1} - X^{i-2}) - X^i + F, \qquad (45)$$

where the dimension of the state vector is N and the cyclic convention $X^{i+N} = X^i$ is used. We will use the vector notation

$$\dot{\mathbf{X}} = q(\mathbf{X}, \mathbf{X}) - \mathbf{X} + \mathbf{F}. \qquad (46)$$

The nonlinear term conserves the total energy $\frac{1}{2}\|\mathbf{X}\|^2$, i.e. $(\mathbf{X}, q(\mathbf{X}, \mathbf{X})) = 0$. The linear term $-\mathbf{X}$, representing mechanical or thermal dissipation, decreases the total energy $-(\mathbf{X}, \mathbf{X}) < 0$ while the constant term $\mathbf{F}$, representing external forcing prevents the total energy from decaying to zero. We imagine that $\mathbf{X}$ represents some atmospheric variable around a latitude circle and X^i is the value at longitude $360i/N$. In all simulations we use $N = 40$ and $\mathbf{F} = 8$. If 1 time unit in the model is identified with 5 days the error doubling time of the model is 2.1 days (Lorenz and Emanuel, 1998).

Figure 10 shows the norm $\|\mathbf{X}(t) - \mathbf{X}(t - \delta t)\|$ as a function of δt average over 1 year. This should be compared with figure 9 for the QG-model. The forecast error norm saturates after day 10 in both models. The straight line in figure 10 is the estimated bound $\sqrt{2}\|\frac{\mathbf{F}}{2}\|$ which can be derived as follows.

The time evolution of the total energy $\frac{1}{2}\|\mathbf{X}\|^2$ is given by

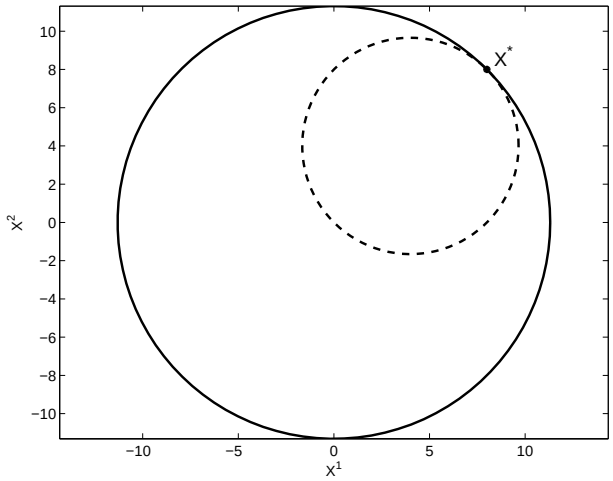

Figure 11: The boundary of ball $\mathcal{B}$ (solid) and the sphere $\mathcal{C}$ (dashed) and the steady state solution $\mathbf{X}^*$ for $N = 2$ and $F = 8$. Trajectories that start in $\mathcal{B}$ can not cross the boundary of $\mathcal{B}$. This figure is not equivalent to a cross section through the X^1 X^2 plane of the system with $N = 40$ because the center of $\mathcal{C}$ will not be contained in this cross section. In particular the sphere $\mathcal{C}$ will appear much smaller in such cross sections

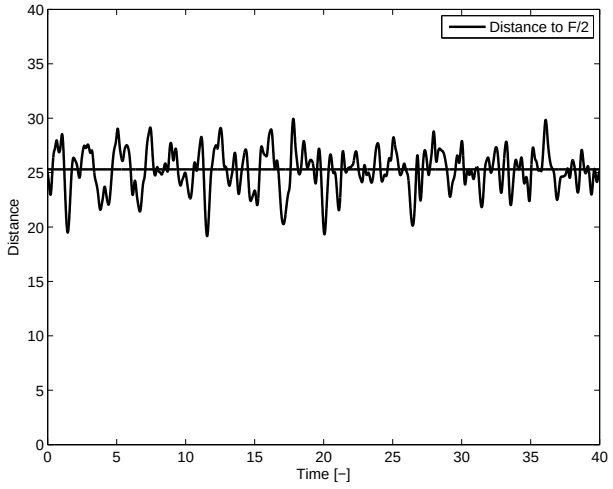

Figure 12: Distance to the center of the sphere $\mathcal{C}$ as a function of time. The straight line is the radius of $\mathcal{C}$.

$$\begin{aligned} \frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} &= -\|\mathbf{X}\|^2 + (\mathbf{X}, \mathbf{F}) \\ &\leq (-\|\mathbf{X}\| + \|\mathbf{F}\|)\,\|\mathbf{X}\|, \end{aligned}$$

where we used the Cauchy-Schwarz inequality. If we define the closed ball

$$\mathcal{B} = \{\mathbf{X} \mid \|\mathbf{X}\| \leq \|\mathbf{F}\|\}, \tag{47}$$

then for all $\mathbf{X} \notin \mathcal{B}$ we have $\frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} < 0$. For all $\mathbf{X}$ on the boundary of $\mathcal{B}$ we have $\frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} \leq 0$. So all trajectories that start in the interior of $\mathcal{B}$ at $t = 0$ remain in this interior for $t > 0$. Note that the steady state solution $\mathbf{X}^* = \mathbf{F}$ is on the sphere.

The time derivative of the energy can also be written as

$$\frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} = -\|\mathbf{X} - \frac{\mathbf{F}}{2}\|^2 + \|\frac{\mathbf{F}}{2}\|^2. \tag{48}$$

Therefore there is a sphere $\mathcal{C}$ with radius $R = \|\frac{\mathbf{F}}{2}\|$ and center $\mathbf{C} = \frac{\mathbf{F}}{2}$ on which the time derivative of the total energy is zero. Again note that the steady solution $\mathbf{X}^*$ is on this sphere (see figure 11). Trajectories that start in the interior of $\mathcal{B}$ stay in the interior for $t > 0$ and therefore the energy of the state is bounded as $T \to \infty$. This is only possible if either the state asymptotically approaches $\mathcal{C}$ or by crossing the surface of the sphere indefinitely. In either case this implies that the dynamics of the system takes place "near" the surface of the sphere $\mathcal{C}$. This is indeed what we observe, see figure 12.

Assume now that the state vectors for large δt are uncorrelated and on the sphere $\mathcal{C}$. By symmetry considerations the expected value for the angle between two vectors associated with two random points on a $N-1$ dimensional sphere is $\pi/2$ (see Borel E. (1914) where it is shown that for large N the probability density function tends to a normal distribution with mean $\pi/2$ and standard deviation $1/\sqrt{N}$) and therefore the expected distance between two random points on the sphere $\mathcal{C}$ is $\sqrt{2}R$. This estimate is shown in figure 10. Given the simplicity of the arguments that were used in the derivation this is remarkably good estimate of the asymptotic behaviour of the forecast error norm.

Before each experiment we started from a random point on the sphere $\mathcal{C}$ and integrate for 100 days (20 time units) to allow the system to reach the attractor. All integrations were performed using a Runge Kutta order 4 scheme with a time step of 0.01.

C Bilinearity preserving finite dimensional representations and time discretizations

If $\mathbf{e}_i$ is a complete time independent orthonormal basis of the phase space $\mathbb{H}$ w.r.t. an inner product $\langle \cdot, \cdot \rangle$ we can write $\mathbf{X}(t) = \sum_i X^i(t)\mathbf{e}_i$. Using the bilinearity of q and linearity of b, (11) can be written as

$$\dot{X}^j \mathbf{e}_j = X^j X^k q(\mathbf{e}_j, \mathbf{e}_k) + X^j b(\mathbf{e}_j) + c, \tag{49}$$

where we use the convention that there is an implied summation over a repeated upper and lower index in a single term. Taking the inner product of this equation with $\mathbf{e}_i$ gives the time evolution of the coordinates $X^i(t)$

$$\dot{X}^i = q^i_{jk} X^j X^k + b^i_j X^j + c^i, \tag{50}$$

where $q^i_{jk} = \langle \mathbf{e}_i, q(\mathbf{e}_j, \mathbf{e}_k) \rangle$, $b^i_j = \langle \mathbf{e}_i, b(\mathbf{e}_j) \rangle$ and $c^i = \langle \mathbf{e}_i, c \rangle$. We see that if the coordinate vector is truncated at a certain index N the truncated system is bilinear (e.g. if $\mathbf{e}_i$ is a spherical harmonic basis). Therefore the time evolution of the coordinates X^i w.r.t. a time independent truncated orthonormal basis is given by a bilinear differential equation and the optimal linearization trajectory can be obtained by adding the coordinates.

C.1 Integration schemes

The Euler forward scheme propagates the state vector as

$$\mathbf{X}_{k+1} = \mathbf{X}_k + h f(t_k, \mathbf{X}_k), \tag{51}$$

where h is the time step. If $f(t_k, \mathbf{X}_k) = q(\mathbf{X}_k, \mathbf{X}_k) + b(\mathbf{X}_k) + c$ then the highest order nonlinear term in the map from $\mathbf{X}_k$ to X_{k+1} is bilinear and therefore the time discretization by the integration scheme preserves the bilinearity of the underlying differential equation. This is no longer true if higher order integration schemes are used. For these schemes the value that is used to evaluate the right hand side of the differential equation at intermediate time steps needs to be stored in the linearization trajectory. In the tangent linear integration these values from the linearization trajectory should then be used in the evaluation of the right hand side of the tangent linear model.

D Multilinear systems

Definition 5 (Multilinear map)
A map $q_n(\mathbf{X}_1, \dots \mathbf{X}_n)$ is called multilinear if it is linear in each argument.

Definition 6 (Symmetric Multilinear map)
For a given multilinear map q_n we define a symmetric map s_n by

$$s_n(\mathbf{X}_1, \dots \mathbf{X}_n) = \frac{1}{n!} \sum_{\text{permutations}} q_n(\mathbf{X}_1, \dots, \mathbf{X}_n), \tag{52}$$

where the summation is over all possible permutations of the arguments $\mathbf{X}_1, \dots \mathbf{X}_n$.

Consider the general form of a multilinear system with at most N-th order multilinearities.

$$\dot{\mathbf{X}} = \sum_{n=0}^{N} q_n(\underbrace{\mathbf{X}, \dots, \mathbf{X}}_{n \text{ times}}), \tag{53}$$

where q_0 is the forcing term in the model. Substitution of $\mathbf{X} \to \mathbf{X} + \mathbf{x}$ and using (53) gives

$$\dot{\mathbf{x}} = \sum_{n=0}^{N} q_n(\mathbf{X} + \mathbf{x}, \dots, \mathbf{X} + \mathbf{x}) - q_n(\mathbf{X}, \dots, \mathbf{X}).$$

Using definition 6 and Newton's binomium this can be written as

$$\dot{\mathbf{x}} = \sum_{n=1}^{N}\sum_{k=1}^{n}\binom{n}{k} s_n(\underbrace{\mathbf{X},\ldots,\mathbf{X}}_{n-k \text{ times}},\underbrace{\mathbf{x},\ldots,\mathbf{x}}_{k \text{ times}}). \qquad (54)$$

The sum over k starts from $k = 1$ because the terms with only capital $\mathbf{X}$'s are cancelled. The summation over n starts from $n = 1$ because the constant term is cancelled. Retaining only the terms linear in $\mathbf{x}$ (terms with $k = 1$) gives the tangent linear model

$$\dot{\hat{\mathbf{x}}}^1 = \sum_{n=1}^{N} n s_n(\underbrace{\mathbf{X},\ldots,\mathbf{X}}_{n-1 \text{ times}},\hat{\mathbf{x}}^1). \qquad (55)$$

If we iteratively relinearized the TL model around the trajectory $\mathbf{X} + \hat{\mathbf{x}}^{k-1}/2$ we get at convergence of the algorithm a unique increment $\hat{\mathbf{x}}^*$ that satisfies

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N} n s_n(\underbrace{\mathbf{X}+\hat{\mathbf{x}}^*/2,\ldots,\mathbf{X}+\hat{\mathbf{x}}^*/2}_{n-1},\hat{\mathbf{x}}^*). \qquad (56)$$

Using Newton's binomium this can be written as

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N} n \sum_{k=0}^{n-1}\binom{n-1}{k} s_n(\underbrace{\mathbf{X},\ldots,\mathbf{X}}_{n-1-k},\underbrace{\frac{\hat{\mathbf{x}}^*}{2},\ldots,\frac{\hat{\mathbf{x}}^*}{2}}_{k},\hat{\mathbf{x}}^*).$$

Shifting the summation over k with 1 gives

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N}\sum_{k=1}^{n} 2^{1-k} n\binom{n-1}{k-1} s_n(\underbrace{\mathbf{X},\ldots,\mathbf{X}}_{n-k \text{ times}},\underbrace{\hat{\mathbf{x}}^*,\ldots,\hat{\mathbf{x}}^*}_{k \text{ times}}),$$

which can also be written as

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N}\sum_{k=1}^{n} k 2^{1-k}\binom{n}{k} s_n(\underbrace{\mathbf{X},\ldots,\mathbf{X}}_{n-k \text{ times}},\underbrace{\hat{\mathbf{x}}^*,\ldots,\hat{\mathbf{x}}^*}_{k \text{ times}}).$$

Therefore we have for the exact time evolution the coefficients

		k				
		1	2	3	4	5
	1	1	-	-	-	-
	2	2	1	-	-	-
n	3	3	3	1	-	-
	4	4	6	4	1	-
	5	5	10	10	5	1

while at convergence we get for the relinearized TL model

		k				
		1	2	3	4	5
	1	1	-	-	-	-
	2	2	1	-	-	-
n	3	3	3	$1\cdot\frac{3}{4}$	-	-
	4	4	6	$4\cdot\frac{3}{4}$	$1\cdot\frac{4}{8}$	-
	5	5	10	$10\cdot\frac{3}{4}$	$5\cdot\frac{4}{8}$	$1\cdot\frac{5}{16}$

The normal TL-model has only nonzero values in the first column. Therefore we see that the relinearized model takes into account all linear terms but also all terms quadratic in the perturbation $\mathbf{x}$. For terms higher than quadratic in $\mathbf{x}$ the relinearized model multiplies the exact coefficient with a factor $k2^{1-k}$. This is a number between 0 and 1 and therefore is always closer to the exact coefficient then setting the coefficient to zero as is done in the standard TL-model. We therefore conclude that the relinearization iteration will always give better approximations than the standard TL-model at convergence of the algorithm.

References

Bengtsson L, Magnusson L, Källén E. 2008. Independent estimations of the asymptotic variability in an ensemble forecast system. *Mon. Weather Rev.* **136**: 4105–4112.

Borel E. 1914. *Introduction géométrique à quelques théories physiques*. Gauthier-Villars.

Brenan K, Campbell S, Petzold L. 1996. *Numerical solution of initial-value problems in differential-algebraic equations*. Society for Industrial Mathematics.

Courtier P, Thépaut J, Hollingsworth A. 1994. A strategy for operational implementation of 4D-var

using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1387.

Errico R. 1997. What is an adjoint model? *Bull. Am. Meteorol. Soc.* **78**(11): 2577–2591.

Errico R. 2003. The Workshop on Applications of Adjoint Models in Dynamic Meteorology. *Bull. Am. Meteorol. Soc.* **84**: 795–798.

Errico R, Ehrendorfer M. 2007. Adjoint Model Applications in Dynamic Meteorology. *Bull. Am. Meteorol. Soc.* **88**(12): 1979–1981.

Gilmour I, Smith L, Buizza R. 2001. Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *J. Atmos. Sci.* **58**: 3525–3539.

Hohenegger C, Schär C. 2007. Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Am. Meteorol. Soc.* **88**(11): 1783–1793.

Hoskins B, Buizza R, Badger J. 2000. The nature of singular vector growth and structure. *Q. J. R. Meteorol. Soc.* **126**(566): 1565–1580.

Janisková M, Thépaut JN, Geleyn JF. 1999. Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Mon. Weather Rev.* **127**(1): 26–45.

Klinker E, Rabier F, Gelaro R. 1998. Estimation of key analysis errors using the adjoint technique. *Q. J. R. Meteorol. Soc.* **124**(550): 1909–1933.

Leutbecher M, Palmer T. 2008. Ensemble forecasting. *J. Comput. Phys.* **227**(7): 3515–3539.

Lorenz E. 1996. Predictability: A problem partly solved. In: *Proc. Seminar on Predictability*, vol. 1. ECMWF, pp. 1–18.

Lorenz E, Emanuel K. 1998. Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.* **55**(3): 399–414.

Marshall J, Molteni F. 1993. Toward a dynamical understanding of planetary-scale flow regimes. *J. Atmos. Sci.* **50**(12): 1792–1818.

Pedlosky J. 1987. *Geophysical fluid dynamics.* Springer, ISBN 0387963871.

Pu Z, Kalnay E, Derber J, Sela J. 1997a. Using forecast sensitivity patterns to improve future forecast skill. *Q. J. R. Meteorol. Soc.* **123**(540): 1035–1053.

Pu Z, Kalnay E, Sela J, Szunyogh I. 1997b. Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Mon. Weather Rev.* **125**(10): 2479–2503.

Rabier F, Klinker E, Courtier P, Hollingsworth A. 1996. Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.* **122**(529): 121–150.

Reynolds C, Palmer T. 1998. Decaying singular vectors and their impact on analysis and forecast correction. *J. Atmos. Sci.* **55**(19): 3005–3023.

Swanson K, Vautard R, Pires C. 1998. Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus* **50A**(4): 369–390.