6.3 Tangent Linear Model, Adjoint Model, Singular Vectors, and Lyapunov Vectors

Lorenz (1965) published in 1965 another paper based on a low-order model that behaved like the atmosphere with just 28 variables. In this fundamental paper, Lorenz introduced for the first time (without using their current names) the concepts of tangent linear model, adjoint model, singular vectors, and Lyapunov vectors for the low order atmospheric model, and their consequences for ensemble forecasting. He also introduced "errors of the day": the predictability of the model is not constant with time: it depends on the stability of the evolving atmospheric flow (the basic trajectory or reference state).

a) Tangent linear model (TLM) and adjoint model

Consider a nonlinear model **discretized in space.**The model can be written as a set of n nonlinear coupled ordinary differential equations (ODEs):

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}),\ \mathbf{x} = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, \mathbf{F} = \begin{bmatrix} F_1 \\ \cdot \\ \cdot \\ F_n \end{bmatrix} \qquad (1.1)$$

This is the model in (time) differential form. Once we choose a time-difference scheme (e.g., Crank-Nicholson, see Table 3.1), it becomes a set of nonlinear-coupled **difference** equations (DEs). Typically, an atmospheric model consists of one such system of DEs which, for example, using a 2-time level Crank-Nicholson scheme would be of the form

$$\mathbf{x}^{n+1} = \mathbf{x}^{n} + \Delta t \mathbf{F}\left(\frac{\mathbf{x}^{n} + \mathbf{x}^{n+1}}{2}\right) \qquad (1.2)$$

"Running the model" gives us a **nonlinear model solution** that depends only on the initial conditions:

$$\mathbf{x}(t) = M(\mathbf{x}(t_0)) \qquad (1.3)$$

where *M* is the time integration of the numerical scheme from the initial condition to time t. A small perturbation **y**(t) can be added to the basic model integration **x**(t):

$$M(\mathbf{x}(t_0) + \mathbf{y}(t_0)) = M(\mathbf{x}(t_0)) + \frac{\partial M}{\partial \mathbf{x}} \mathbf{y}(t_0) + O(\mathbf{y}(t_0)^2) = \mathbf{x}(t) + \mathbf{y}(t) + O(\mathbf{y}(t_0)^2$$

(1.4)

At any given time, the linear evolution of the small perturbation $\mathbf{y}(t)$ will be given by

$$\frac{d\mathbf{y}}{dt} = \mathbf{J}\mathbf{y} \qquad (1.5)$$

where $\mathbf{J} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}}$ is the Jacobian of $\mathbf{F}$.

This system of linear ODEs is the tangent linear model (TLM) in differential form. Its solution between t_0 and t can be obtained by integrating (1.5) in time using the same time difference scheme used in the nonlinear model (1.3):

$$\mathbf{y}(t) = \mathbf{L}(t_0, t)\mathbf{y}(t_0) \tag{1.6}$$

Here $\mathbf{L}(t_0, t) = \dfrac{\partial M}{\partial \mathbf{x}}$ is an (nxn) matrix known as the **resolvent** or **propagator** of the TLM: it propagates an initial perturbation at time t_0 into the final perturbation at time t. Because it is linearized over the flow from t_0 to t, **L** depends on the **basic trajectory** ***x(t)*** (the solution of the nonlinear model), but it does not depend on the perturbation **y**. (The original nonlinear model is autonomous since **F(x)** depends on x(t) but not explicitly on time, but the linear tangent model is non-autonomous). Lorenz (1965) introduced the concept of the TLM of an atmospheric model, but he actually obtained it directly from (1.4), neglecting terms quadratic or higher order in the perturbation **y**:

$$M(\mathbf{x}(t_0)) + \mathbf{L}(t_0, t)\mathbf{y}(t_0) = \mathbf{x}(t) + \mathbf{y}(t) \approx M(\mathbf{x}(t_0) + \mathbf{y}(t_0)) \tag{1.7}$$

He did so by creating as initial perturbations a "sphere" of small perturbations of size ε along the n unit basis vectors $\mathbf{y}_i(t_0) = \varepsilon\,\mathbf{e}_i$ and applying (1.7) to each of these perturbations. With this choice of initial perturbations, subtracting (1.3) he obtained the matrix that defines the TLM:

$$\mathbf{L}(t_0,t)[\varepsilon\mathbf{e}_1,...,\varepsilon\mathbf{e}_n] = \varepsilon\mathbf{L}(t_0,t) = \left[\mathbf{y}_1(t),...,\mathbf{y}_n(t)\right] \tag{1.8}$$

The Euclidean **norm** of a vector is the inner product of the vector with itself:

$$\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = <\mathbf{y},\mathbf{y}> \tag{1.9}$$

The Euclidean norm of **y**(t) is therefore related to the initial perturbation by

$$\|\mathbf{y}(t)\|^2 = (\mathbf{L}\mathbf{y}(t_0))^T\mathbf{L}\mathbf{y}(t_0) = <\mathbf{L}\mathbf{y}(t_0),\mathbf{L}\mathbf{y}(t_0)> = <\mathbf{L}^T\mathbf{L}\mathbf{y}(t_0),\mathbf{y}(t_0)> \tag{1.10}$$

The **adjoint** of an operator $\mathbf{K}$ is defined by the property $<\mathbf{x},\mathbf{K}\mathbf{y}> \equiv <\mathbf{K}^T\mathbf{x},\mathbf{y}>$. In this case of a model with real variables, the **adjoint** of the TLM $\mathbf{L}(t_0,t)$ is simply the **transpose** of the TLM.

Now assume that we separate the interval (t_0,t) into two successive time intervals. For example, if $t_0 < t_1 < t$,

$$\mathbf{L}(t_0,t) = \mathbf{L}(t_1,t)\mathbf{L}(t_0,t_1) \tag{1.11}$$

Since the **adjoint** of the tangent linear model is the transpose of the TLM, the property of the transpose of a product is also valid:

$$\mathbf{L}^T(t_0,t) = \mathbf{L}^T(t_0,t_1)\mathbf{L}^T(t_1,t)$$
(1.12)

Equation (1.11) shows that the TLM can be cast as a product of the TLM matrices corresponding to short integrations, or even single time steps. Equation (1.12) shows that the adjoint of the model can also be separated into single time steps, but they are executed backwards in time, starting from the last time step at t, and ending with the first time step at t_0. For low order models the tangent linear model and its adjoint can be constructed by repeated integrations of the nonlinear model for small perturbations, as done by Lorenz (1965), equation (1.7) and by Molteni and Palmer (1993) with a global quasi-geostrophic model.

For large NWP models this approach is too time consuming, and instead it is customary to develop the linear tangent and adjoint codes from the nonlinear model code following some rules discussed in Appendix B. An example of a FORTRAN code for a nonlinear model, and the corresponding TLM and adjoint models are also given in Appendix B. See also the lecture on 4D-Var by Shu-Chih Yang.

b) Singular Vectors

Recall that for a given basic trajectory and an interval (t_0,t_1) the TLM is a matrix that when applied to a small initial perturbation $\mathbf{y}(t_0)$ produces the final perturbation $\mathbf{y}(t_1)$:

$$\mathbf{y}(t_1) = \mathbf{L}(t_0,t_1)\mathbf{y}(t_0)$$
(1.13)

Singular Value Decomposition (SVD) theory (e.g., Golub and Van Loan, 1989) indicates that for any matrix **L** there exist two orthogonal matrices **U, V** such that

$$\mathbf{U}^{\mathbf{T}}\mathbf{L}\mathbf{V} = \mathbf{S} \qquad (1.14)$$

where
$$\mathbf{S} = \begin{bmatrix} \sigma_1 & 0 & . & 0 \\ 0 & \sigma_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \sigma_n \end{bmatrix}$$

and

$$\mathbf{U}\mathbf{U}^{T} = \mathbf{I}; \mathbf{V}\mathbf{V}^{T} = \mathbf{I} \qquad (1.15)$$

S is a diagonal matrix whose elements are the **singular values** of **L.**

If we left multiply (1.14) by **U**, we obtain

$$\mathbf{L}\mathbf{V} = \mathbf{U}\mathbf{S}, \quad \text{i.e., } \mathbf{L}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = (\sigma_1\mathbf{u}_1, \ldots, \sigma_n\mathbf{u}_n) \qquad (1.16)$$

where $\mathbf{v}_i$ are the columns of V and $\mathbf{u}_i$ the columns of U. This implies that

$$\mathbf{L}\mathbf{v}_i = \sigma_i\mathbf{u}_i \qquad (1.17)$$

Eq. (1.17) defines the $\mathbf{v}_i$'s as the **right singular vectors of L**, hereafter referred to as **initial singular**

vectors, since they are indeed valid at the beginning of the optimization interval over which **L** is defined.

We now right multiply (1.14) by $\mathbf{V}^T$ and obtain:

$$\mathbf{U}^{\mathbf{T}}\mathbf{L} = \mathbf{S}\mathbf{V}^{\mathbf{T}} \tag{1.18}$$

Transposing (1.18), we obtain

$$\mathbf{L}^{\mathbf{T}}\mathbf{U} = \mathbf{VS}, \quad i.e., \quad \mathbf{L}^T(\mathbf{u}_1,...,\mathbf{u}_n) = (\sigma_1\mathbf{v}_1,...,\sigma_n\mathbf{v}_n) \tag{1.19}$$

so that

$$\mathbf{L}^T\mathbf{u}_i = \sigma_i\mathbf{v}_i \tag{1.20}$$

The $\mathbf{u}_i$'s are the **left singular vectors** of **L** and will be referred to as **final** (or **evolved**) **singular vectors,** since they correspond to the end of the interval of optimization.

From (1.17) and (1.20) we obtain

$$\mathbf{L}^T\mathbf{L}\mathbf{v}_i = \sigma_i\mathbf{L}^T\mathbf{u}_i = \sigma_i{}^2\mathbf{v}_i \tag{1.21}$$

Therefore the initial SVs can be obtained as the eigenvectors of $\mathbf{L}^T\mathbf{L}$, a normal matrix whose eigenvalues are the squares of the singular values. Since **U, V** are orthogonal matrices, the vectors $\mathbf{v}_i$ and $\mathbf{u}_i$ that form them constitute orthonormal bases, and any vector can be written in the following form:

$$\mathbf{y}(t_0) = \sum_{i=1}^{n} <\mathbf{y}_0, \mathbf{v}_i > \mathbf{v}_i,$$

$$\mathbf{y}(t_1) = \sum_{i=1}^{n} <\mathbf{y}_1, \mathbf{u}_i > \mathbf{u}_i \qquad (1.22)$$

where $< \mathbf{x}, \mathbf{y} >$ is the inner product of two vectors $\mathbf{x}, \mathbf{y}$.

Therefore, using (1.22)a and (1.17)

$$\mathbf{y}(t_1) = \mathbf{L}(t_0, t_1)\mathbf{y}(t_0) = \sum_{i=1}^{n} <\mathbf{y}_0, \mathbf{v}_i > \sigma_i \mathbf{u}_i \qquad (1.23)$$

If we now take the inner product of (1.23) with $\mathbf{u}_i$ we obtain

$$< \mathbf{y}(t_1), \mathbf{u}_i >= \sigma_i < \mathbf{y}(t_0), \mathbf{v}_i > \qquad (1.24)$$

This indicates that by applying the TLM **L** each initial vector $\mathbf{v}_i$ will be stretched by an amount equal to the singular value σ_i (or contracted if $\sigma_i < 1$), and the direction will be rotated to that of the evolved vector $\mathbf{u}_i$., i.e., that by applying the adjoint of the TLM, $\mathbf{L}^T$**,** each initial vector $\mathbf{v}_i$ will be stretched by an amount equal to the singular value σ_i

Exercise: Use (1.20) and (1.22)b to show that

$$< \mathbf{y}(t_0), \mathbf{v}_i >= \sigma_i < \mathbf{y}(t_1), \mathbf{u}_i >$$

If we consider all the perturbations $\mathbf{y}(t_0)$ of size 1, from (1.24) we obtain that for each of them

$$\sum_{i=1}^{n} \frac{< \mathbf{y}(t), \mathbf{u}_i >}{\sigma_i} = \sum_{i=1}^{n} < \mathbf{y}(t_0), \mathbf{v}_i > = \left\| \mathbf{y}(t_0) \right\|^2 = 1 \qquad (1.25)$$

so that an initial sphere of radius one becomes a hyperellipsoid of semi-axes σ_i . The first initial singular vector $\mathbf{v}_1$ is also called an "optimal vector" since it gives the direction in phase space (i.e., the shape in physical space) of the perturbation that will attain maximum growth σ_1 in the interval (t_0,t_1) (Fig. 6.3).

Fig. 6.3: Schematic of the application of the TLM to a sphere of perturbations of size 1 for a given interval (t_0,t_1).

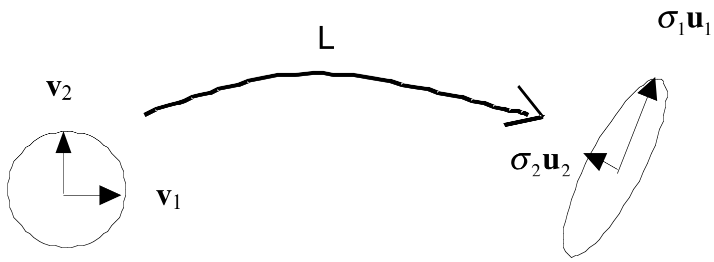
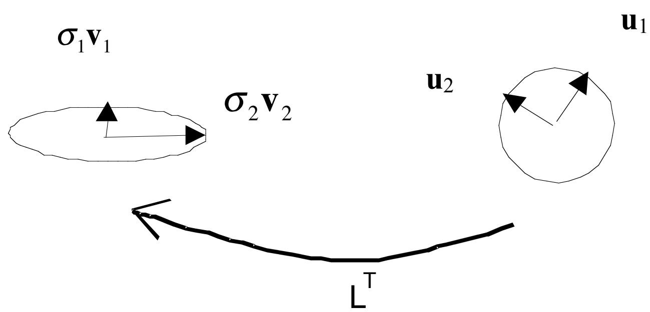


Fig. 6.4: Schematic of the application of the adjoint of the TLM to a sphere of perturbations of size 1 at the final time.



Note that applying **L** is the same as running the TLM forward in time, from t_0 to t_1. Applying $\mathbf{L}^T$ is like running the adjoint model backwards, from t_1 to t_0. From (1.21) we see that if we apply the adjoint model to a sphere of final perturbations of size one (expanded on the basis formed by the evolved or left SVs), they also become stretched and rotated into a hyperellipsoid of semiaxes in the directions of the $\mathbf{v}_i$ with length σ_i (Fig. 6.4)

Therefore, if we apply $\mathbf{L}^T\mathbf{L}$ (i.e., run the TLM forward in time, and then the adjoint backwards in time, the first initial SV will grow by a factor σ_1^2 (see schematic Fig. 6.5), and the other initial SVs will grow or decay by their corresponding singular value squared σ_i^2. In other words, the (initial) singular vectors $\mathbf{v}_i$ are the eigenvectors of $\mathbf{L}^T\mathbf{L}$ with singular values σ_i^2. Conversely, if we apply the adjoint model first (integrate the adjoint model backwards from the final to the initial time), followed by the TLM (integrate forward to the final time), the final singular vectors $\mathbf{u}_i$ will grow both backward and forward, by a total factor also equal to σ_i^2. In other words, the final SVs are the eigenvectors of $\mathbf{L}\mathbf{L}^T$, and again they have eigenvalues equal to the square of the singular values of **L.** Alternatively, once the initial singular vectors are obtained using, for example the Lanczos algorithm, the final singular vectors can be derived by integrating the TLM (equation (1.17)).

Fig. 6.5: Schematic of the application of the TLM forward in time followed by the adjoint of the TLM to a sphere of perturbations of size 1 at the initial time.

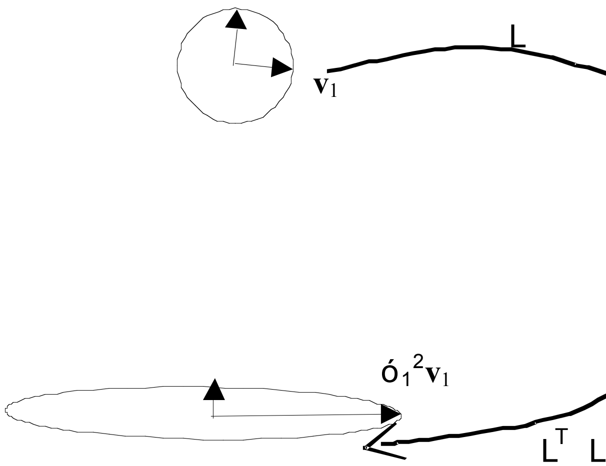
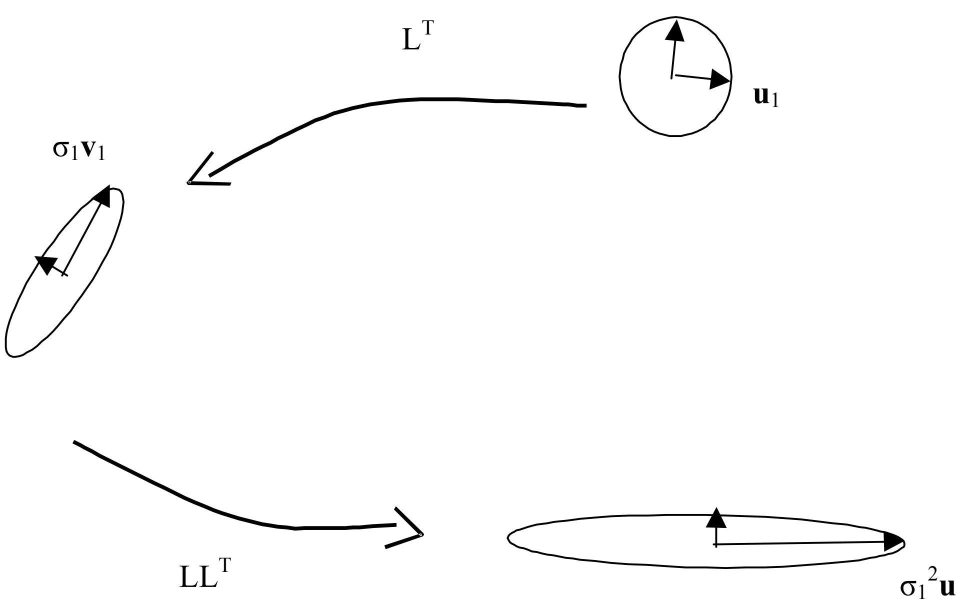


Fig. 6.6: Schematic of the application of the adjoint of the TLM backward in time followed by the TLM forward to a sphere of perturbations of size 1 at the final time.



If we apply $\mathbf{L}^T\mathbf{L}$ repeatedly over the same interval (t_0, t), we obtain the **leading initial SV**, or first optimal vector.

Additional leading SVs can be obtained by a generalization of the power method (Lanczos algorithm, Golub and Van Loan, 1989), which requires running the TLM and its adjoint about 3 times the number of SVs required. For example, to get the leading 30 SVs optimized for $t_1 - t_0$=36 hours, the ECMWF performed 100 iterations, equivalent to running the TLM for about 300 days (Molteni et al, 1996).

It is important to note that the adjoint model and the singular vectors are defined with respect to a given norm. So far we have used an Euclidean norm in which the weight matrix that defines the inner product is the identity matrix:

$$\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = <\mathbf{y}, \mathbf{y}> \qquad (1.26)$$

The leading (initial) singular vectors are the vectors of equal size, with initial norm equal to one

$$\|\mathbf{y}(t_0)\|^2 = \mathbf{y}(t_0)^T\mathbf{y}(t_0) = <\mathbf{y}(t_0), \mathbf{y}(t_0)> = 1,$$

that grow fastest during the optimization period (t_o, t_1), i.e., the initial vectors that maximize the norm at the final time:

$$J(\mathbf{y}(t_0)) \equiv \|\mathbf{y}(t_1)\|^2 = [\mathbf{L}\mathbf{y}(t_0)]^T\mathbf{L}\mathbf{y}(t_0) = <\mathbf{L}^T\mathbf{L}\mathbf{y}(t_0), \mathbf{y}(t_0)> \qquad (1.27)$$

If we define a norm using any other weight matrix **W** applied to **y**, then the requirement that the initial perturbations be of equal size implies:

$$\left\|\mathbf{y}(t_0)\right\|^2 = (\mathbf{W}\mathbf{y}(t_0))^T \mathbf{W}\mathbf{y}(t_0) = \mathbf{y}(t_0)^T \mathbf{W}^T \mathbf{W}\mathbf{y}(t_0) = 1$$

(1.28)

We can use a different norm to define the size of the perturbation to be maximized at the final time than the norm **W** used for the initial time (1.28). For example the final norm could be a projection operator **P** at the end of the interval. (At ECMWF the projection operator was 1 poleward of 30^o and zero equatorward).

Then the function that we want to maximize is, instead of (1.27):

$$J(\mathbf{y}(t_0)) = (\mathbf{PL}\mathbf{y}(t_0))^T \mathbf{PL}\mathbf{y}(t_0) = \mathbf{y}(t_0)^T \mathbf{L}^T \mathbf{P}^T \mathbf{PL}\mathbf{y}(t_0)$$

(1.29)

subject to the **strong constraint** (1.28).

From calculus of variations, the maximum of (1.29) subject to the strong constraint (1.28) can be obtained by the unconstrained maximum of another function:

$$K(\mathbf{y}(t_0)) = J(\mathbf{y}(t_0)) + \lambda[1 - \mathbf{y}(t_0)^T \mathbf{W}^T \mathbf{W}\mathbf{y}(t_0)] =$$
$$\mathbf{y}(t_0)^T \mathbf{L}^T \mathbf{P}^T \mathbf{PL}\mathbf{y}(t_0) + \lambda[1 - \mathbf{y}(t_0)^T \mathbf{W}^T \mathbf{W}\mathbf{y}(t_0)]$$

(1.30)

where the λ are the Lagrange multipliers multiplying the square brackets (equal to zero due to the constraint (1.28)).

The unconstrained minimization of *K* is obtained by computing its gradient with respect to the control variable **y**(t_o) and making it equal to zero. From the remark d) in section 5.4.1, we can compute this gradient as:

$$\nabla_{y(t_o)} K = \mathbf{L}^T \mathbf{P}^T \mathbf{P} \mathbf{L} \mathbf{y}(t_0) - \lambda \mathbf{W}^T \mathbf{W} \mathbf{y}(t_0) = 0$$

(1.31)

It is convenient, given the constraint (1.28), to change variables:

$$\mathbf{W}\mathbf{y}(t_o) = \hat{\mathbf{y}}(t_o), \text{ or } \mathbf{y}(t_o) = \mathbf{W}^{-1}\hat{\mathbf{y}}(t_o)$$

(1.32)

Then, (1.31) becomes

$$(\mathbf{W}^{-1})^T \mathbf{L}^T \mathbf{P}^T \mathbf{P} \mathbf{L} \mathbf{W}^{-1} \hat{\mathbf{y}}(t_0) = \lambda \hat{\mathbf{y}}(t_0)$$

(1.33)

subject to the constraint

$$\hat{\mathbf{y}}^T (t_o) \hat{\mathbf{y}}(t_o) = 1. \qquad (1.34)$$

Therefore, the transformed vectors $\hat{\mathbf{y}}(t_o)$ are the eigenvectors of the matrix $(\mathbf{W}^{-1})^T \mathbf{L}^T \mathbf{P}^T \mathbf{P} \mathbf{L} \mathbf{W}^{-1}$ in (1.33) with eigenvalues equal to the Lagrange multipliers λ_i .

After the leading eigenvectors $\hat{\mathbf{y}}(t_o)$ are obtained (using, for example, the Lanczos algorithm), the variables are transformed back to $\mathbf{y}(t_o)$ using (1.32). The eigenvalues of this problem are the square of the singular values of the TLM: $\lambda_i = \sigma_i^2$.

This allows great generality (as well as arbitrariness[1]) in the choice of initial norm and final projection operator. Errico and Vukicevic (1992), showed that the singular vectors are very sensitive to both the choice of norm and to the length of the optimization interval (the interval from t_0 to t_1). In another example, Palmer et al (1997) tested different weight matrices **W** defining the initial norm. They used "streamfunction", "enstrophy", "kinetic energy", and "total energy" norms, which measured, as the "initial size" the square of the perturbation streamfunction, vorticity, wind speed and weighted temperature, wind and surface pressure, respectively. They found that the use of different initial norms resulted in extremely different initial singular vectors, and concluded that the total energy was the norm of choice for ensemble forecasting. In 1995, ECMWF included in their

[1] Jon Ahlquist (2000, pers. communication) showed, given a linear operator L , a set of arbitrary vectors $\mathbf{x}_i$, and a set of non-negative numbers σ_i arranged in decreasing order, how to construct an inner product and a norm such the σ_i and the $\mathbf{x}_i$ are ith singular values and singular vectors of L. He pointed out that “Because anything not in the null space can be a singular vector, even the leading singular vector, one cannot assign a physical meaning to a singular vector simply because it is a singular vector. Any physical meaning must come from an additional aspect of the problem. Said in another way, nature evolves from initial conditions without knowing which inner products and norms the user wants to use.”

ensemble system a projection operator **P** that measures only the growth of perturbations north of 30^0N, i.e., a matrix that multiplies variables that correspond to latitudes greater or equal to 30N by the number one, and by zero otherwise) (Buizza and Palmer, 1995).

One could use any other pair of initial **W** and final **P** weights (norms) to answer the related question of **forecast sensitivity**. An example of a forecast sensitivity problem is: "What is the optimal (minimum size) initial perturbation (measured by the square of the change in surface pressure over the states of Oklahoma and Texas) that produces the maximum final change after a one day forecast (measured by the change in vorticity between surface and 500hPa over the eastern US)?" ECMWF has been routinely carrying out experiments to find out "What is the change in the initial conditions from 3 days ago that would lead to the best verification of today's analysis?" (see Errico, 1997, Rabier et al, 1996, Pu et al, 1997a,b for more details).

c) Lyapunov Vectors

As we saw in section 6.2, if we start a set of perturbations on a sphere of very small size, it will evolve into an ellipsoid. The growth of the axis of the hyperellipsoid after a finite interval s is given by the singular values $\sigma_i(t_0 + s)$. The (global) Lyapunov exponents (LEs) describe the linear *long-term* growth of the hyperellipsoid:

$$\lambda_i = \lim_{s\to\infty} \frac{1}{s} \ln(\sigma_i(t_0 + s)) \tag{1.35}$$

In other words, the LEs describe the long-term average exponential rate of stretching or contraction in the attractor. (We call the Lyapunov exponents "global" to distinguish them from the finite time or "local" LEs which are useful in predictability applications). There are as many LEs as the dimension of the model (number of independent variables or degrees of freedom). If the model has at least one λ_i greater than zero, then the system can be called chaotic, i.e., there is exponential separation of trajectories. In other words, there is at least one direction of the ellipsoid that continues to be stretched, and therefore two trajectories will diverge in time and eventually become completely different. Conversely, a system with all negative LEs is stable, and will remain predictable at all times. The first LE can be estimated by running the TLM for a long time starting from any randomly chosen initial perturbation $\mathbf{y}(t_0)$. During a long integration the growth rate of any random perturbation will converge to the first LE:

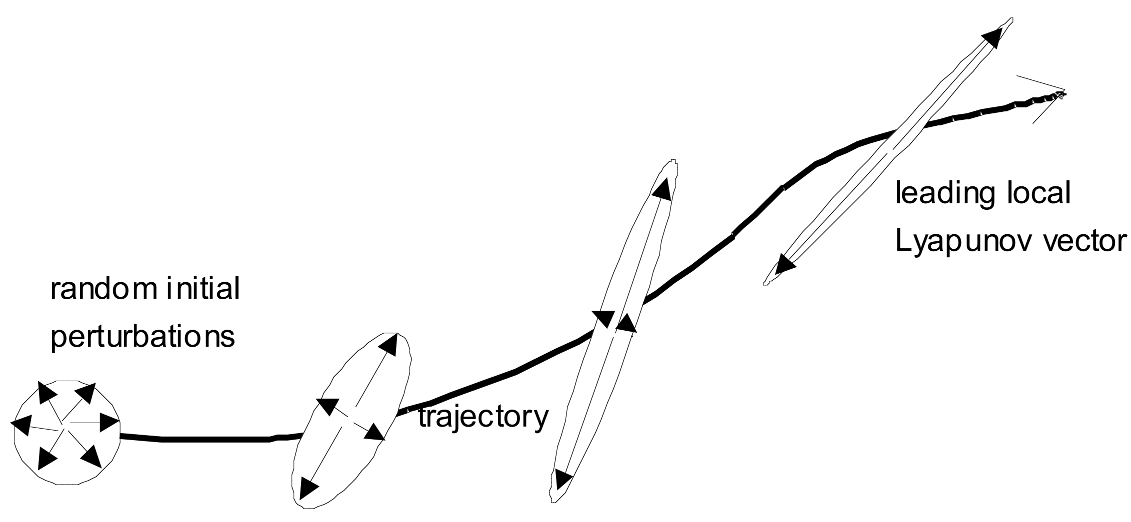
$$\lambda_1 = \lim_{s\to\infty} \frac{1}{s} \ln\left(\frac{\|\mathbf{y}(t_0+s)\|}{\|\mathbf{y}(t_0)\|} \right). \tag{1.36}$$

which is independent of the norm. In practice, the first LE is obtained by running the TLM for a long period from random initial conditions, and renormalizing the perturbation vector periodically in order to avoid computational overflow.

When we are dealing with atmospheric predictions, we are not really interested in the **global** growth properties, which correspond to the atmosphere's attractor (climatology), i.e., relevant average properties over many decades. Instead, in predictability problems we are interested in the growth rate of perturbations at a given time and space: we need to know the **local** stability properties in space and time, which are related to our ability to make skillful forecasts. We can define the leading Local Lyapunov Vector (LLV) at a certain time *t*, as the vector towards which *all* random perturbations $\mathbf{y}(t-s)$ started a long time *s* before *t* will converge (Fig. 6.7).

$$\mathbf{l}_1(t) = \lim_{s\to\infty} \mathbf{L}(t-s,t)\mathbf{y}(t-s) \qquad (1.37)$$

Fig. 6.7: Schematic of how all perturbations will converge towards the leading LLV



Once a perturbation has converged to the leading LLV $\mathbf{l}_1(t)$, the leading Local Lyapunov exponent can be computed from the rate of change of its norm. In practice, the local leading Lyapunov exponent, also known as finite time Lyapunov exponent, can be estimated over a finite period τ:

$$l_1 \approx \frac{1}{\tau}\ln\left(\frac{\left\|\mathbf{l}_1(t+\tau)\right\|}{\left\|\mathbf{l}_1(t)\right\|}\right) \qquad (1.38)$$

The argument of the logarithm is defined as the amplification rate A(t,τ). In practice, since we "live locally" in the atmosphere, it is the local (in time) instabilities (LLV) and growth rates that we are interested in.

The **first LLV is independent of the definition of norm**, and represents the direction in which maximum sustainable growth (or minimum decay) can occur in a system without external forcing. In fact, after a finite transition period T takes place, every initial perturbation will turn in the direction of the leading Lyapunov vector at every point of the trajectory. This even includes the final singular vectors $\mathbf{u}_i$ for a sufficiently long optimization interval.

Trevisan and Legnani (1995) introduced the notion of the leading LLV. Additional LLVs can be obtained by Gramm-Schmidt orthogonalization, and this would seem to indicate that they are norm-dependent. However, Trevisan and Pancotti (1998) showed that it is also possible, at least in theory, to define additional LLVs (denoted **characteristic** vectors by Legras and Vautard,1996) **without the use of norms**. The Local Lyapunov Vectors are therefore a fundamental characteristic of dynamical systems. It should be noted that unfortunately, at this time, there is not a universally accepted nomenclature for LLVs. Legras and Vautard (1996) denote the LLVs as "Backward Lyapunov Vectors", since they were started an infinitely long time in the past. Unfortunately, this name is extremely confusing, since they represent forward evolution rather than backward evolution as this name would imply. The LLVs are also the final singular vectors optimized for an infinitely long time, i.e., the eigenvectors (valid at time t) of $\mathbf{L}(t-T,t)\mathbf{L}^T(t-T,t)$ for $T \to \infty$. Similarly, Legras and Vautard define as "Forward Lyapunov Vectors" the **initial** singular vectors obtained from a very long **backward** integration with the adjoint of the model, i.e., they are the eigenvectors (valid at time t) of $\mathbf{L}^T(t,t+T)\mathbf{L}(t,t+T)$ for very large T.

Legras and Vautard (1996) showed (as did Trevisan and Pancotti, 1998) that a complete set of LLVs (which they denote **characteristic** Lyapunov vectors) can be defined from the intersection of the subspaces spanned by the “forward” and “backward” LVs. The (characteristic) Local Lyapunov vectors are therefore independent of the norm, and grow in time with a rate given by the local Lyapunov exponents. As such, they are a fundamental characteristic of dynamical systems.

Several authors have shown for that the leading (first few) LLVs of low-dimensional dynamical systems span the attractor, i.e., they are parallel to the hypersurface in phase space that the dynamical system visits again and again ("realistic solutions"). Leading Singular Vectors, on the other hand, have very different properties. They can grow much faster than the leading LLV, but are initially off the attractor:

they point to areas in the phase space where solutions do not naturally take place (e.g., Legras and Vautard, 1996, Trevisan and Legnani, 1995, Trevisan and Pancotti, 1998, Pires et al, 1996), see also next section.

For ensemble forecasting, Ehrendorfer and Tribbia (1997) showed that if **V** is the initial analysis error covariance (which unfortunately we don't know and can only estimate, except within Ensemble Kalman Filter), then the initial singular vectors defined with the norm $\mathbf{W} = \mathbf{V}^{-1/2}$ evolve into the eigenvectors of the evolved error covariance matrix. This implies that the leading singular vectors, defined using the initial error covariance, are optimal in describing the forecast errors at the end of the optimization period. The initial error covariance norm yields singular vectors quite different from those derived using the energy norm. Barkmeijer et al (1998) used the ECMWF estimated 3D-Var error covariance as initial norm (instead of the total energy norm) and obtained initial perturbations with structures closer to the bred vectors (i.e., leading local Lyapunov vectors) used at NCEP.

d) Simple examples of singular vectors and eigenvectors

In order to get a more intuitive feeling of the relationship between singular vectors and Lyapunov vectors, we consider a simple linear model in 2-dimensions:

$$\begin{bmatrix} x_1(t+T) \\ x_2(t+T) \end{bmatrix} = \mathbf{M}_T(x(t)) = \begin{bmatrix} 2x_1(t) + 3x_2(t) + 7 \\ 0.5x_2(t) - 4 \end{bmatrix} \tag{1.39}$$

We compute the 2-dimensional tangent linear model (TLM), constant in time:

$$\mathbf{L} = \begin{bmatrix} \dfrac{\partial M_1}{\partial x_1} & \dfrac{\partial M_1}{\partial x_2} \\ \dfrac{\partial M_2}{\partial x_1} & \dfrac{\partial M_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 0 & 0.5 \end{bmatrix} \tag{1.40}$$

The propagation or evolution of any perturbation (difference between two solutions) over a time interval (t, t+T) is given by

$$\delta\mathbf{x}(t+T) = \mathbf{L}\delta\mathbf{x}(t) \tag{1.41}$$

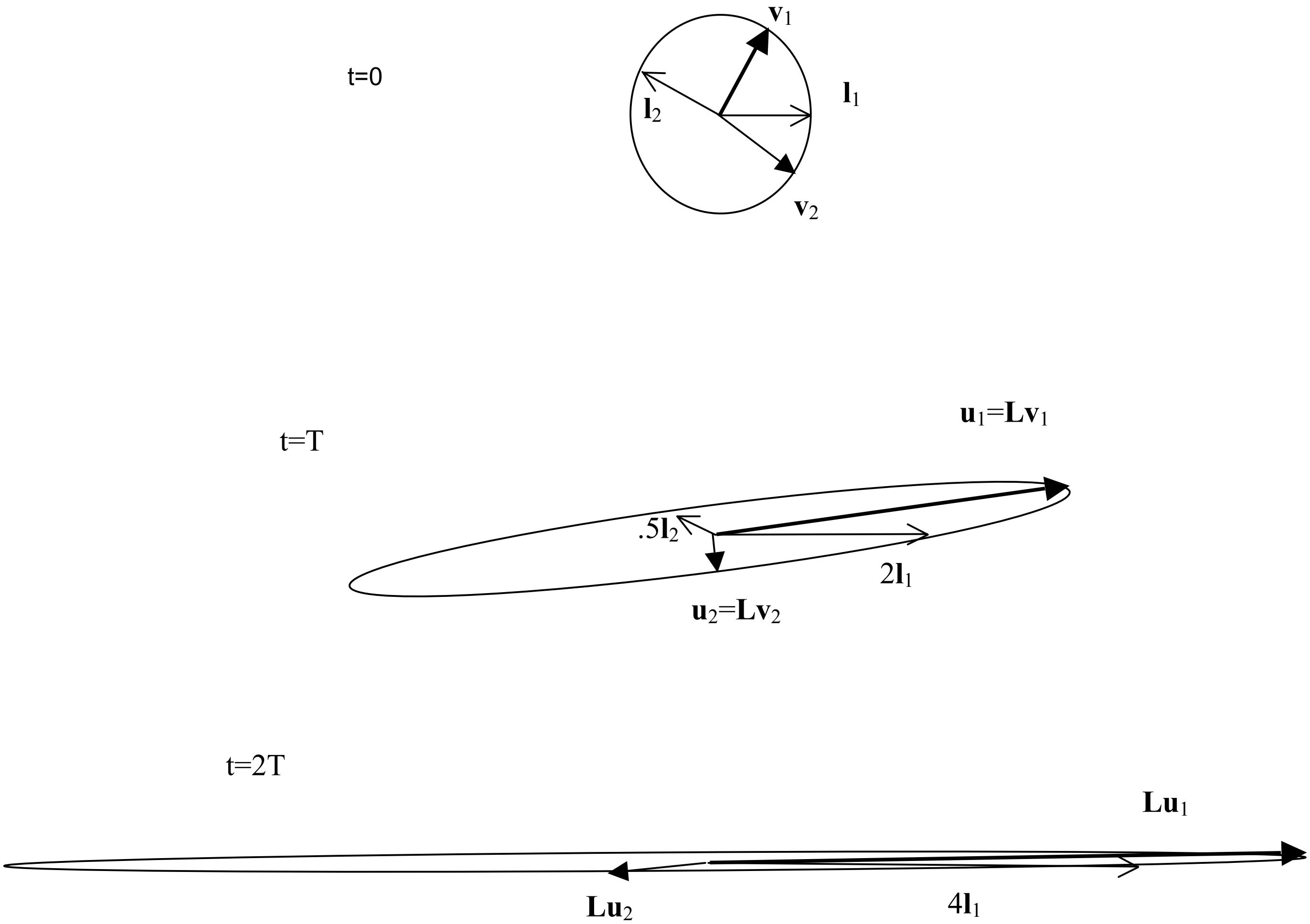
Note that the translation terms in (1.39) do not affect the perturbations. The eigenvectors of *L* (which for this simple constant TLM are also the Lyapunov vectors) are proportional to

$\mathbf{l}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{l}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$ corresponding to the eigenvalues $\lambda_1 = 2, \lambda_2 = 0.5$, respectively, which in this case are the two Lyapunov numbers (their logarithms are the Lyapunov exponents).

If we normalize them, so that they have unit length, the LVs are

$$\mathbf{l}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{l}_2 = \begin{pmatrix} -.89 \\ .45 \end{pmatrix} \tag{1.42}$$

The LVs are not orthogonal, they are separated by an angle of 153.4° (Fig. 6.8a). We will see that because they are not orthogonal it is possible to find linear combinations of the LVs that grow faster than the leading LV. We will also see that the leading LV is the attractor of the system, since repeated applications of **L** to any perturbation makes it evolve towards $\mathbf{l}_1$.



Applying first **L** and then its transpose $\mathbf{L}^\mathbf{T}$ we obtain the symmetric matrix

$$\mathbf{L}^T\mathbf{L} = \begin{bmatrix} 4 & 6 \\ 6 & 9.25 \end{bmatrix} \qquad (1.43)$$

whose eigenvectors are the **initial singular vectors**, and whose eigenvalues are the squares of the singular values. The initial SVs (eigenvectors of $\mathbf{L}^T\mathbf{L}$) are

$$\mathbf{v}_1 = \begin{pmatrix} .55 \\ .84 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} .84 \\ -.55 \end{pmatrix} \qquad (1.44)$$

with eigenvalues $\sigma_1^2 = 13.17, \sigma_2^2 = 0.076$. As indicated before, the **singular values** of $\mathbf{L}$ are the square roots of the eigenvalues of $\mathbf{L}^T\mathbf{L}$, i.e., $\sigma_1 = 3.63, \sigma_2 = 0.275$. Note that this implies that during the optimization period (0,T) the leading singular vector grows almost twice as fast as the leading Lyapunov Vector (3.63 vs. 2). The angle that the leading **initial SV** has with respect to the leading LV is 56.82°, whereas the second initial SV is perpendicular to the first one (Fig.6.8a).

The final or **evolved SVs** *at the end of the optimization period (0,T)* are the eigenvectors of $\mathbf{L}\mathbf{L}^T = \begin{bmatrix} 13 & 1.5 \\ 1.5 & 0.25 \end{bmatrix}$

(1.45)

and after normalization, they are

$$\mathbf{u}_1 = \begin{pmatrix} .99 \\ .12 \end{pmatrix}, u_2 = \begin{pmatrix} .12 \\ -.99 \end{pmatrix}. \qquad (1.46)$$

Note again that the operators $\mathbf{L}^T\mathbf{L}$ and $\mathbf{L}\mathbf{L}^T$ are quite different, and the final SVs are different from the initial SVs, but they have the same singular values $\sigma_1^2 = 13.17, \sigma_2^2 = 0.076$.

Alternatively, the evolved SVs at the end of the optimization period can also be obtained by applying **L** to the initial SVs, which is computationally inexpensive. In this case, $\mathbf{u}_1(T) = \mathbf{L}\mathbf{v}_1(0) = \begin{bmatrix} 3.6 \\ .42 \end{bmatrix}, \mathbf{u}_2(T) = \mathbf{L}\mathbf{v}_2(0) = \begin{bmatrix} .03 \\ -.27 \end{bmatrix}$, same as (1.46) but without normalization.

The final leading SV has strongly rotated towards the leading LV: at the end of the optimization period the angle between the leading SV and the leading LV of only 6.6° (Fig. 6.8b), and because the SVs have been optimized for this period, the final SVs are still orthogonal.

To obtain the evolution of the SVs *beyond the optimization period (0,T)* we apply **L** again to the evolved SV valid at t=T and
obtain

$$\mathbf{u}_1(t+2T) = \mathbf{L}\mathbf{u}_1(t+T) = \begin{bmatrix} 8.46 \\ 0.31 \end{bmatrix}, \mathbf{u}_2(2T) = \mathbf{L}\mathbf{u}_2(T) = \begin{bmatrix} -0.75 \\ -0.14 \end{bmatrix}.$$

During the interval (T,2T) the leading SV grows by a factor of just 2.33, not very different from the growth rate of the leading LV. At the end of this second period (Fig.6.8c) the angle with the leading LV is only 1.41°. The angle of the second evolved SV at time T, after applying the linear tangent model $\mathbf{L}$ and the leading LV is also quite small (10.24°), and because it was further away from the attractor, the second singular vector (whose original, transient, singular value was 0.5), grows by a factor of 2.79. This example shows how quickly **all perturbations, including all SVs, evolve towards the leading LV, which is the attractor of the system**. It is particularly noteworthy that during the optimization period (0,T), the first SV grows very fast as it rotates towards the attractor, but once it gets close to the leading LV, its growth returns to the normal leading LV's growth.

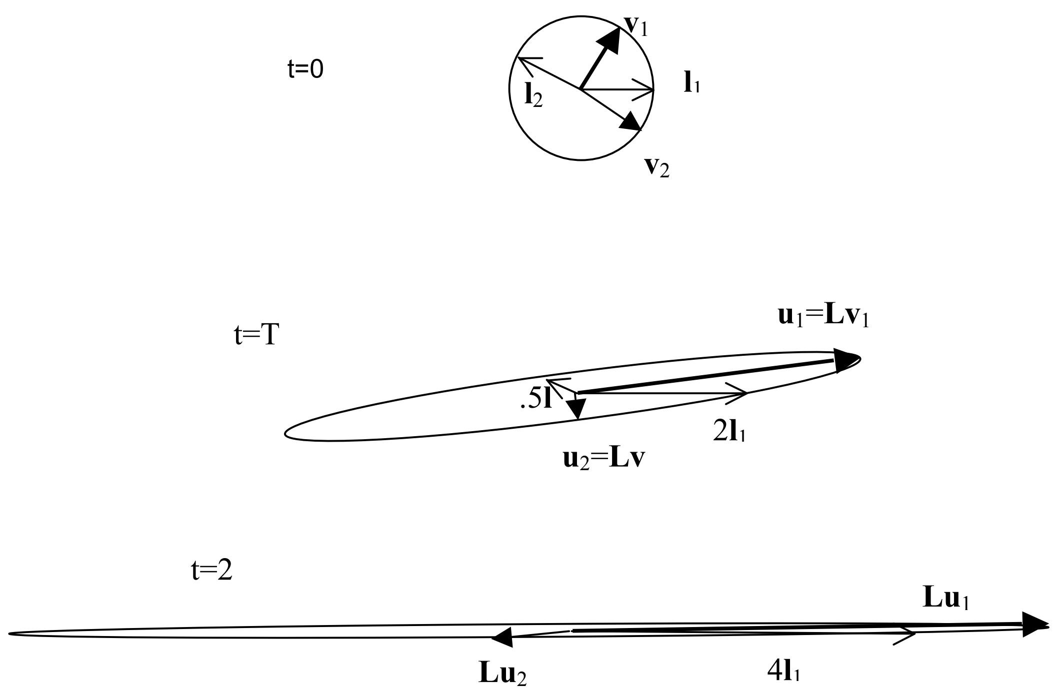
Let us now choose as the tangent linear model another matrix $\mathbf{L} = \begin{bmatrix} 2 & 30 \\ 0 & 0.5 \end{bmatrix}$, with the same eigenvalues 2 and 0.5, i.e., with eigenvectors (LVs) that still grow at a rate of 2/T and 0.5/T respectively. However, now the angle between the first and the second LV is 177°, the LVs are almost antiparallel. In this case, the first singular vector grows by a factor of over 30 during the optimization period, but beyond the optimization period it essentially continues evolving like the leading LV.

These results do not depend on the fact that one LV grows and the other decays. As a third example, we choose $\mathbf{L} = \begin{bmatrix} 2 & 3 \\ 0 & 1.5 \end{bmatrix}$ with two growing LVs with rates 2/T and 1.5/T. The LVs are almost parallel, with an angle of 170°, and the leading SV grows during the optimization period by a factor of 3.83. Applying the TLM again to the evolved SVs we obtain that at time 2T the leading SV has grown by a factor of 2.9 and its angle with respect to the leading LV is 1°. Because it is not decaying, the second LV is also part of the attractor, but only those perturbations that are exactly parallel to it will remain parallel, all others will move towards the first LV.

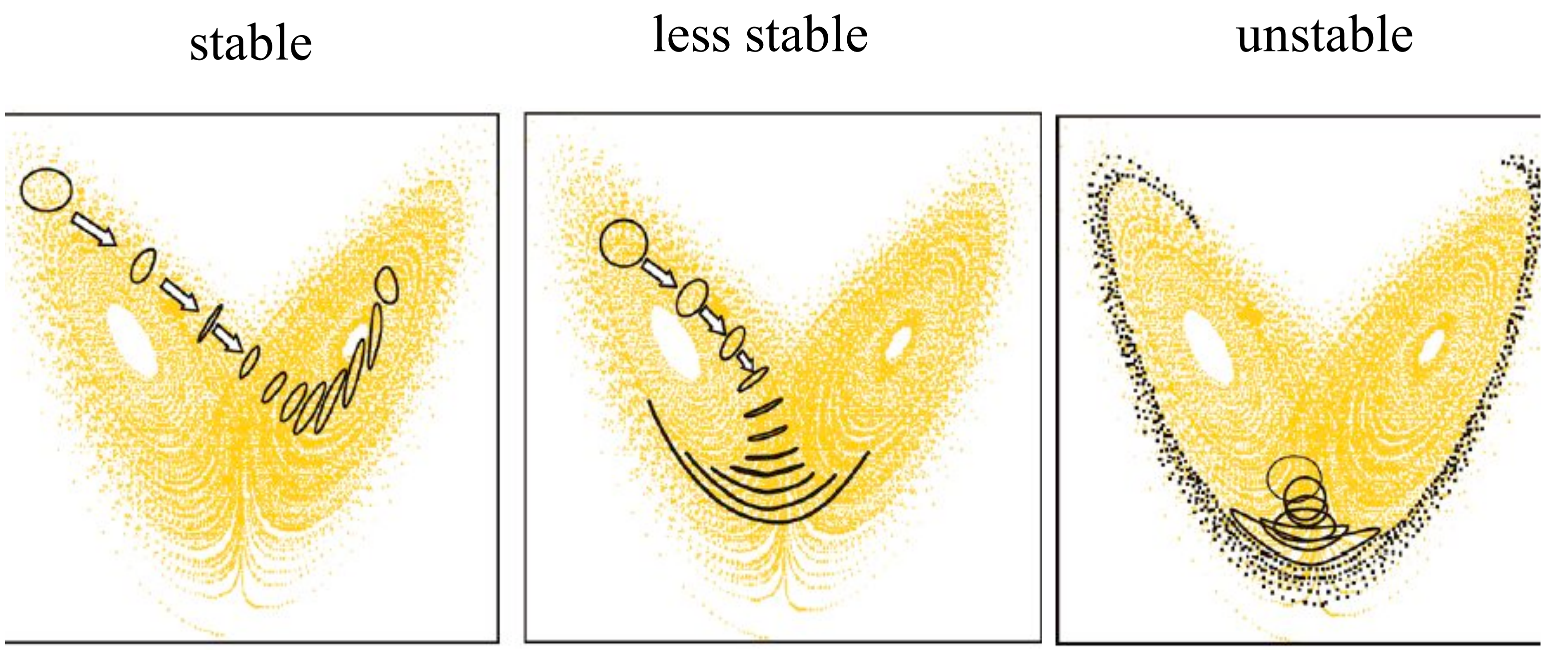
These examples illustrate the fact that the fast growth of the singular vectors during the optimization period depends on the lack of orthogonality between LVs. *A very fast "super-growth" of singular vectors is associated with the presence of almost parallel LVs*, and it takes place when the initial SV, which is not in the attractor, rotates back towards the attractor. At the end of the optimization period, the leading SV tends to be much closer to the attractor, more parallel to the leading LV. The second (trailing) SV is also moving towards the leading LVs.

Finally, we point out that this introductory discussion is appropriate for relatively low dimensional systems. For extremely high dimensional systems like the atmosphere, there may be multiple sets of Lyapunov exponents corresponding to different types of instabilities. For example, as pointed out by Toth and Kalnay (1993), convective instabilities have very fast growth but small amplitudes, whereas baroclinic instabilities have slower growth but much larger amplitudes, and each of these can lead to different types of Lyapunov vectors. If we are interested in the predictability characteristics associated with baroclinic instabilities, then the analysis of growth rates of infinitesimally small Lyapunov vectors over infinitely long times may not be appropriate for the problem (Lorenz, 1996). In that case, it may be better to consider the finite amplitude, finite time extension of Lyapunov vectors introduced by Toth and Kalnay (1993, 1997) as bred vectors (BVs). Bred vectors are discussed in section 6.5.1.

Fig. 6.8: Schematic of the evolution of the two non-orthogonal Lyapunov vectors (thin arrows $\mathbf{l}_1$ and $\mathbf{l}_2$), and the corresponding two initial Singular Vectors (thick arrows $\mathbf{v}_1(0)$ and $\mathbf{v}_2(0)$), optimized for the interval (0,T), for the tangent linear model $\mathbf{L} = \begin{bmatrix} 2 & 3 \\ 0 & 0.5 \end{bmatrix}$ with eigenvalues 2 and 0.5. a) Time t=0, showing the initial SVs $\mathbf{v}_1(0)$ and $\mathbf{v}_2(0)$, as well as the LVs $\mathbf{l}_1$ and $\mathbf{l}_2$. b) Time t=T, evolved SVs, $\mathbf{u}_1(T) = \mathbf{L}\mathbf{v}_1(0), \mathbf{u}_2(T) = \mathbf{L}\mathbf{v}_2(0)$ at the end of the optimization period; the LVs have grown by a factor of 2 and 0.5 respectively, whereas the leading SV has grown by 3.63. The second evolved SV has grown by 0.275, and is still orthogonal to the first SV. c) Time t=2T. Beyond the optimization period T, the evolved SVs $\mathbf{u}_1(t+2T) = \mathbf{L}\mathbf{u}_1(t+T), \mathbf{u}_2(2T) = \mathbf{L}\mathbf{u}_2(T)$ are not orthogonal and they approach the leading LV with similar growth rates.



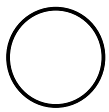
Predictability depends on the initial conditions (Palmer, 2002):



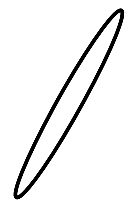
A “ball” of perturbed initial conditions is followed with time.
Errors in the initial conditions that are unstable (with “errors of the day”) grow much faster than if they are stable

Discuss SVs, LVs and BVs in this figure

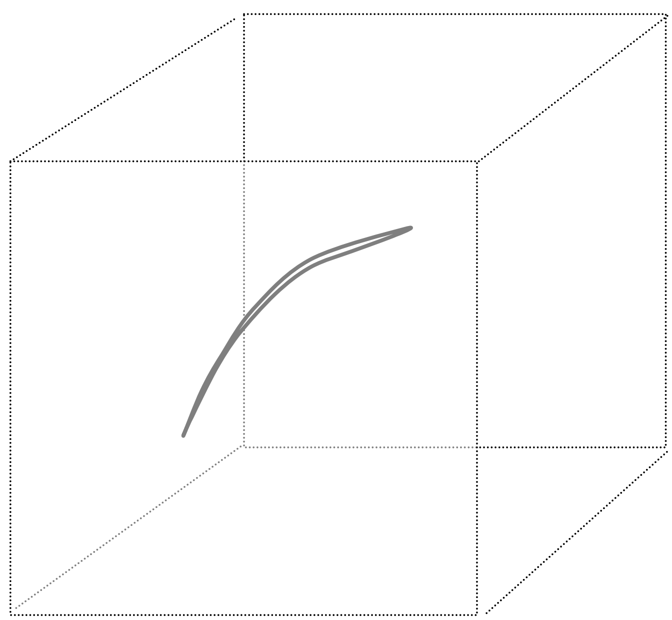
a) Initial volume: a small hypersphere



b) Linear phase: a hyper ellipsoid



c) Nonlinear phase: folding needs to take place in order for the solution to stay within the bounds



d) Asymptotic evolution to a strange attractor of zero volume and fractal structure. All predictability is lost