Transient Growth and Why we should Care about it

Rahul Bale and Rama Govindarajan

The phrases 'transient growth' and 'non-normality' have become common parlance in fluid mechanics nowadays. We present these ideas with a simple two-dimensional system, to enable the reader to look for transient growth, as a trigger for nonlinear behaviour to set in, in a variety of situations probably having nothing to do with fluid mechanics. The article is aimed at undergraduate students of science, engineering, finance, etc., and the material is based completely on the excellent books of Trefethen and Embree, and Schmid and Henningson [1,2].

Rahul Bale is presently a graduate student of mechanical engineering at Northwestern University, USA. His research interests are in the area of fluid dynamics. He is currently working on bio-fluid dynamics.

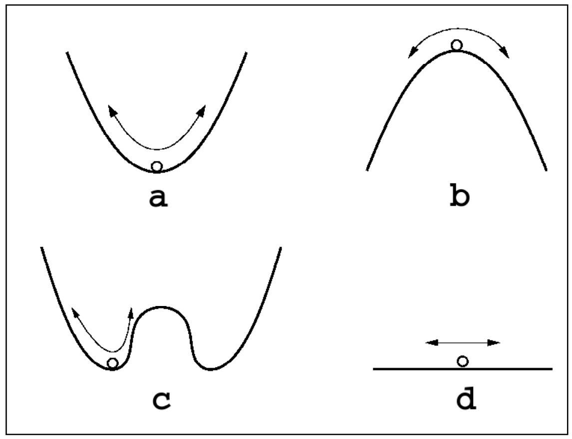
Rama Govindarajan is a Professor at JNCASR, Bangalore. Her research interest is fluid dynamics.

Introduction

In many situations, we have to ask a question about stability. A civil engineer who puts up a tall structure has to ask herself (or himself) whether the structure will stand or fall down when disturbed. A chemist has to ask whether a product of a reaction will be stable. From a child on a bicycle to a corporate giant in the fickle market, none of us is immune to the roller-coaster that the stability, or otherwise, of systems subject us to. How can we decide whether a system is stable? For this, we often need to go back to the example we were once taught, of a ball in various landscapes. A ball at the bottom of a well as shown in *Figure* 1(a) is of course in a stable equilibrium, meaning that if it is moved away from the bottom, i.e., perturbed away from its equilibrium state, it starts oscillating as shown, but no matter what the magnitude of the initial perturbation is, the oscillations will decay monotonically due to friction, and the ball will eventually settle back at the bottom.

Keywords
Flow instability, algebraic growth.

Figure 1. Different stability scenarios.



Conversely, a ball on top of a hill (*Figure*1(b)) is in an unstable equilibrium, since even the smallest perturbation will have it rolling off. A third possibility is shown in *Figure* 1(c), where the present state of the ball is stable to small perturbations but can be unstable to large perturbations which may cause the ball to cross the barrier and go to the adjacent well. In *Figure* 1(d) we have a boring situation where all positions are the same as each other and the ball does not choose between them. The case (c) is of particular interest to us.

What then is 'transient growth'? In this article, and to an increasing number of people, the term means an incredible situation when 'theory' promises us that a system is stable to small perturbations, but at short times (i.e., transiently), the 'theory' appears to be incorrect, and we have disturbances growing, sometimes by many orders of magnitude, even when we carefully ensure that no disturbance bigger than a certain tiny size is given to the system. In terms of *Figure* 1(c), we could have a situation where we have moved the ball ever so slightly from the bottom of its present well, but find the ball in the other well at a much later time! A system whose dynamics can be thus, would need to be slightly more complicated than the one shown in the figure. If we had only one plane of oscillation, a small

The present state of the ball is stable to small perturbations but can be unstable to large perturbations which may cause the ball to cross the barrier and go to the adjacent well.

perturbation would always decay due to friction, but if the ball were sitting in a three-dimensional landscape, transient growth is possible. Let us see what we mean by transient growth in the following section.

2. A Condition for Transient Growth

Let us do a simple exercise with two vectors, $\vec{X}_1$ and $\vec{X}_2$ as shown in *Figure* 2. We will allow the vectors to grow or shrink in time while pointing in the same direction always, i.e., the respective magnitudes X_1 and X_2 are functions of time, and the angle between them stays constant. If the angle between the vectors is $\pi - \phi$ radians, the magnitude R of the resultant of $\vec{X}_1$ and $\vec{X}_2$ is given by

$$R^2 = (e^{-\lambda_1 t} - Ce^{-\lambda_2 t}\cos\phi)^2 + (Ce^{-\lambda_2 t}\sin\phi)^2. \quad (1)$$

When $\phi = 90^{\text{o}}$, the vectors would be orthogonal, or normal to each other, while for any other ϕ they would be non-normal; we will use this terminology repeatedly. It is easy to see that if X_1 and X_2 were growing with time, so would the magnitude R of the resultant. In this paper we are interested in the opposite case, i.e., one where X_1 and X_2 decay exponentially, as $X_1 = e^{-\lambda_1 t}$, and $X_2 = Ce^{-\lambda_2 t}$, where λ_1 and λ_2 are positive constants. Going by our instinct, we expect R to decrease monotonically too. Surprisingly however, we can sometimes have R increasing for some time before finally decreasing. This special situation, where the individual

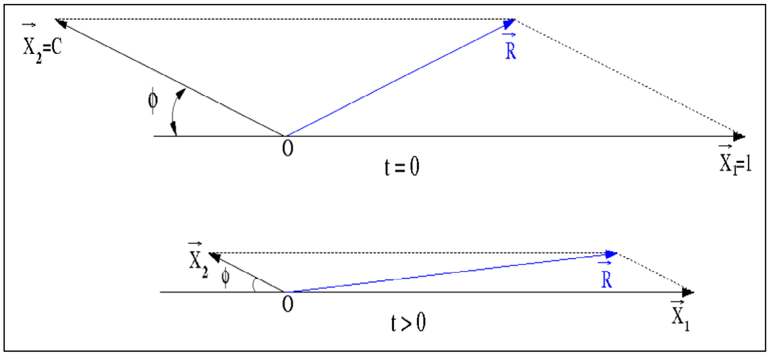


Figure 2. Transient growth of the resultant of two non-orthogonal vectors.

Given a pair of decay rates λ_1 and λ_2, the angle between the vectors must exceed a certain value to make it possible for their resultant to grow.

vectors are all monotonically shrinking but their resultant grows for some time, is called 'transient growth'. Let us see when and how such transient growth can happen. For R to grow at any time, we must have

$$\frac{\mathrm{d}R}{\mathrm{d}t} > 0. \qquad (2)$$

For the vectors in *Figure* 2, upon defining

$$y \equiv \frac{\exp\left[(\lambda_2 - \lambda_1)t\right]}{C} \equiv \frac{z}{C}, \text{ and } r \equiv \frac{\lambda_2}{\lambda_1}, \qquad (3)$$

equation (2) can be rewritten as

$$y^2 - y(1+r)\cos\phi + r < 0. \qquad (4)$$

Without loss of generality we may choose $\lambda_2 < \lambda_1$, such that $r < 1$. The quantity z lies between 0 and 1, so for $C > 0$ we must have

$$y < \frac{(1+r)}{2}\left[\cos\phi \pm \sqrt{\left(\frac{1-r}{1+r}\right)^2 - \sin^2\phi}\right] \qquad (5)$$

for transient growth. For y to be real this implies

$$\phi < \sin^{-1}\left(\frac{1-r}{1+r}\right). \qquad (6)$$

Thus, given a pair of decay rates λ_1 and λ_2, the angle between the vectors must exceed a certain value to make it possible for their resultant to grow. This condition for transient growth is equivalent to that given in Trefethen and Embree [1] and Schmid [3] in terms of the matrix condition (16), (see below). We see from the above equation that if the two decay rates are very close to each other, i.e., if $r \to 1$, the vectors have to be practically collinear to see any transient growth. On the other hand, if the two decay rates are different by many orders of magnitude such that $r \to 0$, then any non-normality (departure from orthogonality) is sufficient for some transient growth. We will return to this point later.

If equation (6) is satisfied, it still does not mean that the resultant of two such vectors will show any growth. It also matters what the initial magnitude of each vector is. If however, ϕ lies in the right range, and we are at liberty to choose the ratio C of the initial magnitudes of the vectors, we can always realise a growth in R as follows. A ϕ which satisfies equation (6) may be written as

$$\cos\phi = \frac{2(h+1)\sqrt{r}}{r+1}, \tag{7}$$

where $h > 0$, and $h + 1 \leq (1+r)/2\sqrt{r}$. The condition in (5) for transient growth can then be rewritten as

$$y < h\sqrt{r}[1 + \sqrt{1 + 2/h}\,]. \tag{8}$$

For a given pair of decaying vectors, the above condition determines the range of C for which transient growth is possible. Since all quantities as defined are positive, we have chosen the positive root in (5) to arrive at the above.

We next wish to understand how large a growth in the resultant we can possibly get. The answer lies in picking the *optimal* initial conditions, namely the best C. Denoting the magnitude of the resultant at the initial time as R_0, we define a growth parameter G for a given C as

$$G(t) = \left[\frac{R(t)}{R_0}\right]^2 \text{ and } G_{\max} = \max_t G(t). \tag{9}$$

Setting $\mathrm{d}G/\mathrm{d}C = 0$ for the optimum C at a given time, we get

$$C_{\text{optimum}} = \frac{(1+z) \pm \sqrt{(1-z)^2 + 4z\sin^2\phi}}{2\cos\phi}, \tag{10}$$

where z is evaluated at the given time. This gives, if $\mathrm{d}^2G/\mathrm{d}C^2 < 0$, the maximum value attainable by G at that time. The variation of G with C is shown in

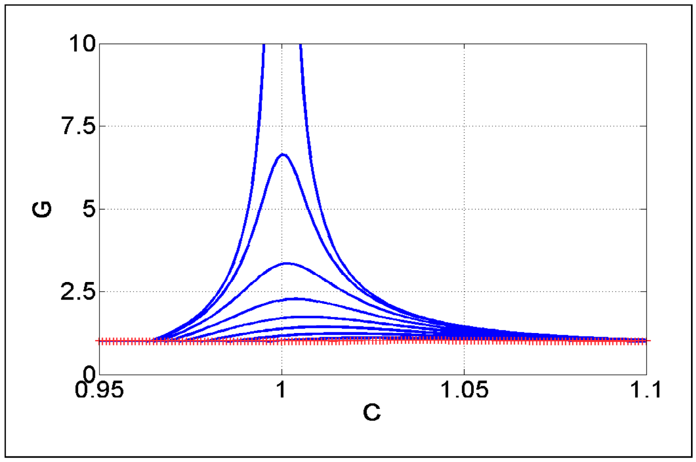


Figure 3. The maximum transient growth G_{max} as a function of the initial ratio C of the magnitudes of the two vectors. The different lines correspond to various values of the angle (180–ϕ) between them, beginning with $\phi = 0^o$ for the topmost curve, and ending with the critical value $\phi = 3.82^o$ for the bottom curve shown by the plus symbols. The curves in between are for ϕ = 0.4°, 0.9°, 1.3°, 1.7°, 2.1°, 2.6°, 3.0° and 3.4° respectively. Any $\phi > 3.82^o$, would ensure a monotonic decay of the resultant. Note that the reason G_{max} goes to infinity at small ϕ for $C \to 1$ is because the initial value of the resultant goes to zero. This is an artefact of the object we have chosen to optimise.

Figure 3 for $\lambda_1 = 8$, $\lambda_2 = 7$ and various values of ϕ ranging from 0 to the critical value $\sin^{-1}((1-r)/(1+r))$. The greater of the two roots of equation (10) gives the C_{optimum}. It may be confirmed that this condition does indeed correspond to a maximum by checking the sign of the second derivative.

The phrases *transient growth* and *non-normality* are often used interchangeably with each other. Indeed, if we set $\phi = 90^o$ then

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{-1}{R}(C^2\lambda_2 e^{-2\lambda_2 t} + \lambda_1 e^{-2\lambda_1 t}) \; , \qquad (11)$$

which is always less than zero; hence we can have no growth if the eigenvectors are orthogonal. The above exercise points out that the converse is often not true, i.e., non-normality, or ϕ being different from 90^o does not guarantee transient growth. In fact, we need ϕ to be sufficiently far away from 90^o, i.e., the system to be sufficiently non-normal, to obtain transient growth.

What does $G_{\max}$ mean for our ball in *Figure* 1(c)? Imagine that the well where the ball is sitting is a three-dimensional wonky-bowl shaped valley, with its walls differently sloped everywhere. Imagine also that in some direction the walls give way to a second valley as shown.

If there were no other valley in the vicinity, the ball may oscillate with larger amplitude for some time due to transient growth, but would eventually go back to the same old valley's bottom. However, given that there is another valley, a subset of the initial perturbations can be such that the ball goes over the hump and slides into the second valley. In other words, this happens when $G_{\max}$ is large enough for the resultant of the initial perturbation in the direction of the barrier to cross it. A typical matrix system to study this is described in the next section.

Given that there is another valley, a subset of the initial perturbations can be such that the ball goes over the hump and slides into the second valley.

Can the resultant from the optimal perturbation decay initially and then grow? We leave this section with that as food for thought for the reader.

3. Transient Growth and Stability

The previous section showed that predictions which seem intuitively obvious may not always turn out to be true. Our objective however is not just to show this. We studied the example above because it is the simplest explanation of a lot of dynamics seen in many real-life situations.

To see this better, we consider the dynamics described by the initial value problem

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mathcal{A} * X, \;\; X(0) = X_0 \;, \tag{12}$$

where $X = [X_1 X_2 X_3 .... X_n]^T$, the superscript T denoting a matrix transpose. When $\mathcal{A}$ is independent of X, the dynamical equation is linear in X. We first consider

$$\mathcal{A} = \lambda_1 \begin{bmatrix} -1 & p \\ 0 & -r \end{bmatrix}. \tag{13}$$

Note that any 2×2 matrix can be written in the above form if we align our coordinate system with one of the eigenvectors. The eigenvalues of $\mathcal{A}$ are $-\lambda_1$ and $-\lambda_2$, and with

$$p = (1-r)\cot\phi/\lambda_1, \qquad (14)$$

$\vec{X}_1$ and $\vec{X}_2$ as defined in the previous section form its eigenvectors. The solution to (12) gives their magnitudes as $X_1 = e^{-\lambda_1 t}$, and $X_2 = Ce^{-\lambda_2 t}$. We are familiar with the resultant of the system, whose square can be interpreted as the energy of the system. This is because the dynamics in many real systems, for example a simple pendulum, can be written in the form of (12), and the energy is indeed the square of the resultant of the eigenvectors. In the case of a simple pendulum this is the sum of the potential and the kinetic energy. In fluid systems without gravity, the resultant is often nothing but twice the kinetic energy $u^2+v^2+w^2$. We have $p = 0$ when $\phi = \pi/2$, i.e., when the vectors are orthogonal; $\mathcal{A}$ is then just a diagonal matrix, and the system is incapable of displaying transient growth, as we have seen. In other words, a necessary condition for transient growth is that the matrix $\mathcal{A}$ should be non-normal.

A matrix M is called normal if

$$MM^\dagger = M^\dagger M \ , \qquad (15)$$

where $M^\dagger$ is adjoint of M. Any other matrix is termed 'non-normal'. Incidentally, self-adjoint matrices form a subset of normal matrices which satisfy $M = M^\dagger$.

The adjoint of a matrix is the transpose of its complex conjugate in the simplest case. We emphasise again that since ϕ will have to satisfy equation (6) before any transient growth is possible, non-normality is not a sufficient condition for transient growth. A sufficient condition is [1]

$$\lambda_{\max}\{M + M^{\mathrm{H}}\} > 0, \qquad (16)$$

where $\lambda_{\max}$ is the largest eigenvalue of $M + M^{\mathrm{H}}$. If we apply this condition to $\mathcal{A}$ to obtain a constraint on ϕ, we get back our original condition given by (6).

Going back again to *Figure* 1(c) we consider our two vectors $\vec{X}_1$ and $\vec{X}_2$ as small perturbations in two coordinates of the ball from its position shown. The dynamics very close to the bottom of this valley is described by (17). Such a matrix equation is usually a very good approximation in real-life problems. Assuming $\mathcal{A}$ in the general case has N distinct eigenvalues, the general solution of (12) is

$$X = V_1 e^{\lambda_1 t} + V_2 e^{\lambda_2 t} + \dots , \tag{17}$$

where λ_i and V_i are the eigenvalues and the eigenvectors of the matrix $\mathcal{A}$ respectively. We look at two example matrices, which have only negative eigenvalues. We set $N = 2$ for simplicity, but our conclusions hold good for large N as well. The reader is encouraged to work out what would happen in the case of repeated eigenvalues.

$$\mathcal{A}_1 = \begin{bmatrix} -0.1 & 50 \\ 0 & -0.2 \end{bmatrix}, \quad \mathcal{A}_2 = \begin{bmatrix} -10 & 5 \\ 5 & -5 \end{bmatrix} . \tag{18}$$

The eigenvectors of $\mathcal{A}_1$ lie at an angle of 0.11^{o} to each other while those of $\mathcal{A}_2$ are 90° apart, i.e., $\mathcal{A}_1$ is non-normal and $\mathcal{A}_2$ is normal.

The growth parameter G is the ratio of the energy at time t to the initial energy, since the energy is defined here to be merely the square of the magnitude of the resultant R. Note that in different physical problems, we may need to monitor the growth of different parameters, and would have to adjust our treatment accordingly. The maximum attainable value E of the growth parameter G at each time is shown in *Figure* 4 for the two examples $\mathcal{A}_1$ and $\mathcal{A}_2$. We use this figure to go back to our stability analysis. If we stopped at a traditional linear theory, i.e., after solving the system (12), we would give ourselves the possibly erroneous belief that the ball will monotonically go back to the same steady-state. We now know however, that for a range of initial conditions, the energy of the small perturbation can grow a lot,

If we stopped at a traditional linear theory, we would give ourselves the possibly erroneous belief that the ball will monotonically go back to the same steady-state.

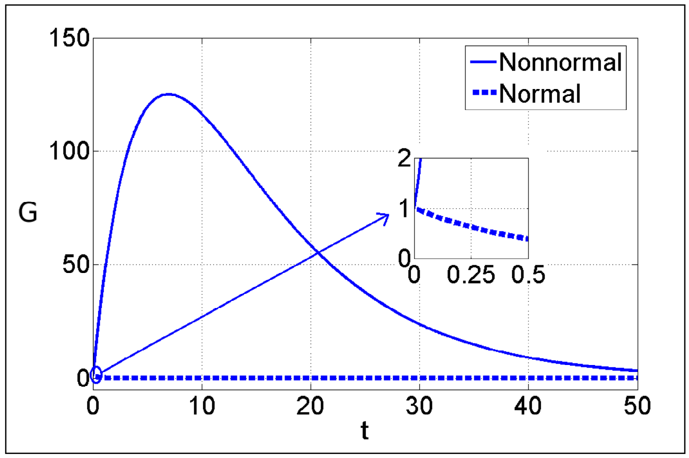


Figure 4. Evolution of optimal energy E for $\mathcal{A}_1$ and $\mathcal{A}_2$. The monotonic decay in the energy of $\mathcal{A}_2$ is seen clearly in the inset. E is the optimum value attainable by G at a given time by choosing the optimal initial condition, i.e., the best C.

sometimes by several orders of magnitude, and take the system into a totally different state. In fluid dynamics, transient growth occurs quite commonly, for example, in the flow of water through a pipe. As mentioned in the final section, the transient growth of disturbances can make this flow change its state from laminar to turbulent.

4. Sensitivity of Eigenvalues: Pseudospectra

We briefly divert our attention to the concept of 'pseudospectra' to see how a pseudospectrum helps in estimating the transient growth possible in a system. We only introduce the topic here. Read Trefethen and Embree[1] for more on it.

In fluid dynamics, transient growth occurs quite commonly, for example, in the flow of water through a pipe.

The spectrum of a matrix $\mathcal{A}$ is the set of its eigenvalues, denoted by $\sigma(\mathcal{A})$. When we perturb the matrix $\mathcal{A}$ by a small amount $\Delta\mathcal{A}$, the perturbed matrix $\mathcal{A}_{\mathrm{p}} = \mathcal{A} + \Delta\mathcal{A}$ will have a different spectrum. To quantify what we mean by perturbing by a 'small' amount, we first define the 'size' of a matrix. The size of a matrix is estimated by its norm, which may be defined in many ways, but we use here the Euclidean or $L2$ norm, also known simply as the 2-norm. The $L2$ norm $\|\cdot\|_2$ for a vector $\vec{v} = [v_1, v_2, v_3 ... v_n]$ is $\|v\|_2 = (v_1^2 + v_2^2 + v_3^2 + ... + v_n^2)^{1/2}$, which is just the magnitude of its resultant. For a square matrix

$\mathcal{M}$ this norm is the square root of the largest eigenvalue of $\mathcal{M}\mathcal{M}^\dagger$ or $\mathcal{M}^\dagger\mathcal{M}$. The $L2$ norm of $\mathcal{A}$ in (13) is thus

$$\|\mathcal{A}\|_2 = \frac{\lambda_1}{\sqrt{2}}\Big[1 + r^2 + p^2 + \Big\{(1-r^2)^2 + p^4 + 2p^2(1+r^2)\Big\}^{1/2}\Big]^{1/2}. \quad (19)$$

To simplify this discussion we assume the norm of $\mathcal{A}$ to be of order 1. (A matrix whose norm is much larger or smaller can be scaled appropriately by a constant in order to satisfy this.) We now perturb $\mathcal{A}$ by a matrix $\Delta\mathcal{A}$ with $\|\Delta\mathcal{A}\|_2 \leq \epsilon << 1$. The perturbation matrix $\Delta\mathcal{A}$ is denoted by

$$\Delta\mathcal{A} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 \\ \varepsilon_3 & \varepsilon_4 \end{bmatrix},$$

where each individual element of $\Delta\mathcal{A}$ is $\leq \epsilon$, so all the elements of the perturbed matrix $\mathcal{A}_\mathrm{p}$ are close to the corresponding elements of $\mathcal{A}$. The ϵ pseudospectrum of a matrix can be defined in several ways and we will choose the simplest. Once the maximum size ϵ of the perturbation is fixed, there are of course infinitely many ways to choose $\Delta\mathcal{A}$. Each will result in a different spectrum for $\mathcal{A}_\mathrm{p}$. The union of all such spectra for a given ϵ is called the ϵ *pseudospectrum* of $\mathcal{A}$. Our example consists of real eigenvalues only, but for a matrix with complex eigenvalues we may write

$$\sigma_\epsilon(\mathcal{A}) = \{z \in C : z \in \sigma(\mathcal{A} + \Delta\mathcal{A}) \text{ for all } \Delta\mathcal{A} \text{ with } \|\Delta\mathcal{A}\| \leq \epsilon\},$$

and follow the same procedure [1,4]. For a small ϵ we expect a small ϵ pseudospectrum, which is always the case for a normal matrix. However, when a matrix is non-normal, the spectrum of $\mathcal{A}_\mathrm{p}$ can differ greatly from that of $\mathcal{A}$, resulting in a large pseudospectrum. Let us see how this can happen. The eigenvalues of $\mathcal{A}_\mathrm{p}$ are

$$\lambda_{\rm p} = \frac{\lambda_1}{2}[-1 - r + \epsilon_4 + \epsilon_1 \;\pm$$
$$(1 - 2r + 2\epsilon_4 - 2\epsilon_1 + r^2 - 2r\epsilon_4 + 2r\epsilon_1$$
$$+\;\epsilon_4^2 - 2\epsilon_1\epsilon_4 + \epsilon_1^2 + 4\epsilon_2\epsilon_3 + 4p\epsilon_3)^{1/2}]. \quad (20)$$

where we have rescaled ϵ_i by $1/\lambda_1$. Neglecting terms of order ϵ, and remembering that p can be large, (20) can be reduced to

$$\lambda_{\rm p} = \frac{\lambda_1}{2}[-1 - r \pm (1-r)^{1/2}\sqrt{(1-r) + 4\epsilon_3 \cot\phi/\lambda_1}]. \quad (21)$$

When $\cot\phi$ is of the order of 1 or smaller, the psuedospectrum is small, but when it is large, which happens when the matrix is substantially far from normal, the psuedospectrum can be much larger in magnitude than ϵ. It is also important to note that as the two eigenvalues become close to each other, i.e., as $r \to 1$, the quantity under the square-root symbol above is of order $\epsilon^{1/2}$, which is of lower order (i.e., bigger) than ϵ. Here even for $\cot\phi \sim 1$, i.e., small p, we have a significant pseudospectrum. When the eigenvalues are equal, i.e., $\lambda_1 = \lambda_2$, an equivalent equation will be

$$\lambda_{\rm p} = \lambda_1 \left[-1 \pm \sqrt{\epsilon_3 p}\right], \quad (22)$$

which is the limit of (21) as $r \to 1$. This procedure can be repeated for any $N \times N$ matrix, and it will be seen that if all the eigenvalues are close to each other, the pseudospectrum will be of size $\sim (\epsilon p)^{1/N}$, where p is the magnitude of the off-diagonal elements. This is much larger than ϵ, so non-normality can become very important in multi-dimensional systems. Thus for any $N \times N$ matrix we can often comment on its non-normality by just looking at its pseudospectrum; if the pseudospectrum is larger than the perturbation which caused it, then the matrix is non-normal.

If the pseudospectrum is larger than the perturbation which caused it, then the matrix is non-normal.

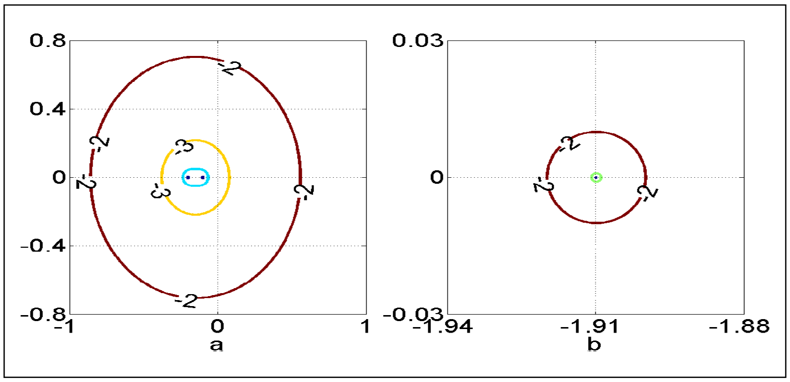


Figure 5. The contours of the pseudospectra of (a) matrix $\mathcal{A}_1$ and (b) $\mathcal{A}_2$ in $\log_{10}$ scale. Note that the scale of the second plot is hugely magnified compared to the first, i.e., the pseudo-spectrum of $\mathcal{A}_2$ is much smaller than that of $\mathcal{A}_1$.

The pseudospectra of $\mathcal{A}_1$ and $\mathcal{A}_2$ studied above are shown in *Figure* 5. The different curves represent pseudospectra for different values of ϵ up to $\epsilon = 10^{-2}$. The non-normal nature of $\mathcal{A}_1$ is evident, albeit qualitatively, in the large size of its pseudospectrum, as opposed to that of $\mathcal{A}_2$. Note in particular that beyond some value of ϵ, the eigenvalues of the perturbed matrix $\mathcal{A}_{1p}$ actually lie in the unstable half-plane. When the eigenvalues are in the unstable half-plane, i.e., the eigenvalues are positive, exponential decay is replaced by exponential growth. The maximum protrusion α_ϵ into this region gives us a lower limit of the maximum possible transient growth for a given matrix [4]. For any $\epsilon > 0$

$$G_{\max} \geq \frac{\alpha_\epsilon(\mathcal{A})}{\epsilon}. \tag{23}$$

For $\mathcal{A}_1$, for $\epsilon = 10^{-2}$ we have $\alpha_\epsilon = 0.559$, which indicates that $G_{\max} \geq 55.9$, while for $\epsilon = 10^{-3}, \alpha_\epsilon = 0.079$, indicating $G_{\max} \geq 79$. In *Figure* 4, we see that $G_{\max} = 125$, satisfying our expectations from the pseudospectra. The pseudospectra of $\mathcal{A}_2$ on the other hand do not cross into the right half plane even for $\epsilon \to 1$.

5. Non-normality in Fluid Dynamics, and the Role of Nonlinearity

Non-normality and transient growth are ubiquitous in

Reynolds showed that at low flow velocities the pipe flow remains *laminar*. As the flow velocity is increased, some disturbances begin to appear in the pipe flow, which increase as the velocity is increased, until at high velocities the entire flow appears *turbulent*.

many fields, ranging from population dynamics to economics, and the reader is encouraged to look everywhere for them. We discuss the most famous example in fluid dynamics here. In 1883 Osbourne Reynolds studied the flow of water through a pipe. He showed that at low flow velocities the pipe flow remains *laminar*. As the flow velocity is increased, some disturbances begin to appear in the pipe flow, which increase as the velocity is increased, until at high velocities the entire flow appears *turbulent*. We will refrain from a mathematically precise definition of the terms 'laminar' and 'turbulent'. For our purposes it is enough to know that laminar flow through a pipe is usually steady and predictable. Turbulent flow on the other hand is very unsteady, chaotic and vortical. At some flow velocity, laminar flow becomes unstable and if the velocity is increased progressively, the flow ultimately becomes turbulent. Reynolds found that the transition between laminar and turbulent flow happens when a non-dimensional number, which now goes by his name, attains a value of about 2000. The Reynolds number is defined for this flow by $\rho\, UD/\mu$, where ρ and μ are the density and viscosity of the fluid respectively, D is the diameter of the pipe, and U is the velocity of the fluid at the pipe centreline. For a given pipe and given fluid, increasing the flow velocity thus amounts to increasing the Reynolds number.

Reynolds found that the transition between laminar and turbulent flow happens when a non-dimensional number attains a value of about 2000.

Early theoretical results were at variance with what Reynolds, and anyone since who has seen water flowing through a transparent pipe, could see plainly. The theoreticians asked the question: Is the laminar flow stable or not, i.e., does this state figuratively correspond to the bottom of the well we saw in the beginning of this article? To find out, they perturbed the laminar solution by a small amount to see whether the disturbances would decay or grow. In 1927 Sexl (see e.g., Davey and Drazin [5]) found that laminar pipe flow seems stable *to very small perturbations* at any Reynolds number. This result was further strengthened by Davey and Drazin in 1969 who computed the complete set of eigenvalues and found them all to be damped [5].

Often, when this particular disagreement between theory and experiment takes place, it is sorted out by recourse to 'nonlinearity' (see e.g., Stuart [6]). By this we simply mean that worrying only about whether a flow is unstable when disturbed very slightly from the 'bottom of the well' may not give the correct answer. We may need to know what happens when it is moved quite a distance away. The mathematical consequence is as follows. When we apply a small perturbation X, we normally discard terms containing X^2, X^3, etc., as being too small to make a difference to the answer. The resulting dynamical equation is linear in X, and has the form of (13). When the disturbance X from the steady state is large, powers of X greater than 1 cannot be neglected, and the resulting equation is non-linear in X. Thus 'nonlinear stability analysis' usually means that we have not made the assumption that the initial perturbation is small.

Anybody with some background in dynamics knows that extremely complicated and strange behaviour may be expected from non-linear systems. The surprising thing about transient growth, and an important reason to study it, is that it provides a way by which we can perturb a linearly stable system by extremely small amounts, and still have the perturbations grow to large values, after which non-linear effects may no longer be neglected. We discuss a very simple system to give this central message, that transient growth can help make a system leave a locally 'stable' state and go to another.

Anybody with some background in dynamics knows that extremely complicated and strange behaviour may be expected from non-linear systems.

Consider

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathcal{A}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \begin{bmatrix} a_1 X_1^2 \\ a_2 X_2^2 \end{bmatrix}. \qquad (24)$$

The system has two steady states, defined by $X = 0$

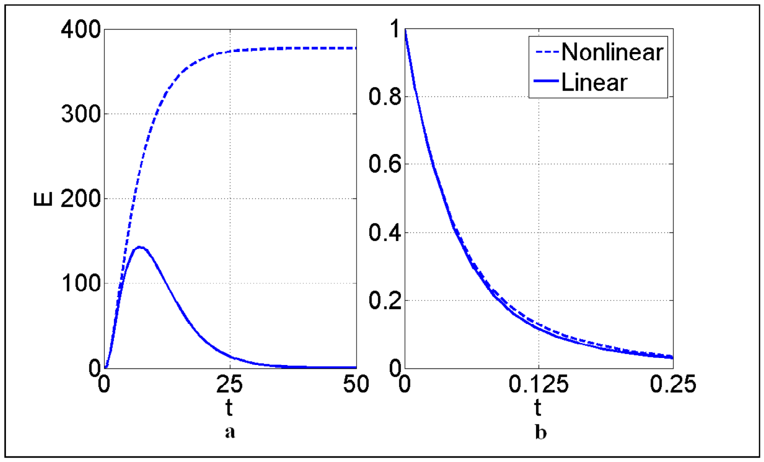


Figure 6. Solutions of equation (24) when $\mathcal{A}$ is (a) non-normal and (b) normal. The solid line shows the solution with the second term on the right-hand side set to zero, i.e., it constitutes the linear solution. The dashed line is the complete solution. The non-normal matrix contributes the initial growth, which is made use of by the nonlinear term to reach a new steady state. In the normal system however, since we have begun with a small X, both solutions show a decaying trend.

and $\mathrm{d}X/\mathrm{d}t = 0$. These two states may be thought of as describing the two wells in *Figure* 1(c). Close to the first state, we may neglect X'^2 (where $X'^2 = [a_1 X_1^2, a_2 X_2^2]$), but the second term becomes important for larger X. *Figure* 6(a) shows the linear transient growth for $\mathcal{A} = \mathcal{A}_1$ and the nonlinear kick-off by X'^2 to a new state and in *Figure* 6(b) with $\mathcal{A} = \mathcal{A}_2$ we can see that the nonlinear term X'^2 does not qualitatively change the solution. We have chosen $a_1 = 0.005121$ and $a_2 = 5$. In summary, transient growth provides a mechanism for nonlinearities to set in even when we begin with extremely small perturbations near a 'stable' state.

The simple system described above is roughly analogous to what is now thought to happen in a pipe flow. Although the transition from laminar flow to turbulence is not well understood even a century and a quarter after Reynolds, it is becoming increasingly clear to scientists working on this problem that transient growth is often an important player in the sequence of events leading to turbulence. The complicated manner by which transient growth occurs and triggers nonlinearities in pipe flow is beyond the scope of this article, but the interested reader can look up Trefethen *et al* [7] for example. We will only mention here that the governing equations (derived from the Navier–Stokes and continuity equations) can be highly non-normal, and involve a large number of eigenmodes which can interact to give large transient growth. In the examples presented here, the transient growth was relatively modest, but in situations like heated channel flow [8], transient growth may cause the disturbance energy to increase by several orders of magnitude.

Aknowledgements: We thank Peter Schmid for discussions.

Suggested Reading

[1] L N Trefethen and Mark Embree, *Spectra and Pseudospectra – The Behaviour of Nonnormal Matrices and Operators*, 1st ed, Princeton University Press, Princeton and Oxford, 2005.

[2] P J Schmid and D S Henningson, *Stability and Transition in Shear Flows*, 1st ed, Springer-Verlag, New York, 2001.

[3] P J Schmid, Nonmodal stability theroy, *Annu. Rev. Fluid Mech*. Vol.39, pp.129–162, 2007.

[4] L N Trefethen, Pseudospectra of linear operators, *SIAM Rev*., Vol.39, pp.383–406, 1997.

[5] A Davey and P G Drazin, The stability of Poiseuille flow in a pipe. *J. Fluid Mech*.,Vol.36, pp.209–218, 1969.

[6] J T Stuart, On the nonlinear mechanics of hydrodynamic stabiliy, *J. Fluid Mech*., Vol.4, pp.1–21, 1958.

[7] L N Trefethen, A E Trefethen, S C Reddy, and T A Driscoll, Hydrodynamic stability without eigenvalues, *Science*, Vol.261, pp.578–584, 1993.

[8] A Sameen and R Govindarajan, The effect of wall heating on the instablity of channel flow, *J. Fluid Mech*., Vol.577, pp.417–442, 2007.

Address for Correspondence
Rahul Bale
Department of Mechanical Engineering
Northwestern University
Illinois, USA
Email: bale.rahul@gmail.com

Rama Govindarajan
Engineering Mechanics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
Jakkur
Bangalore 560064, India.
Email: rama@jncasr.ac.in