Some Basic Linear Algebra

Gidon Eshel

In this section of the notes I want to define, and give many examples of, some basic ideas in linear algebra.

The first idea is the **nullspace** $\mathcal{N}(\mathbf{A})$ of an $M \times N$ $\mathbf{A}$ (M rows by N columns). The nullspace is the collection of all N-vectors $\mathbf{n} \in \mathcal{R}^N$, $\{\mathbf{n}\}$, that are mapped by $\mathbf{A}$ onto the zero M-vector (by J-vector we mean a vector with J elements):

$$\mathbf{A}\mathbf{n} = \mathbf{0} \in \mathcal{R}^M.$$

Examples:

- for

$$\mathbf{A} = \begin{pmatrix} 1 & -2 \\ 0 & 0 \\ -1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & -2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

  so $\alpha \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ with any arbitrary α will be mapped by $\mathbf{A}$ to $\mathbf{0}$. Above, we have clarified the situation by reducing $\mathbf{A}$ to an upper-diagonal matrix $\mathbf{U}$ with nonzero elements only on or above the main diagonal. To reduce $\mathbf{A}$ to $\mathbf{U}$ we have used **Gaussian elimination**, the systematic application of elementary operations to the rows so as to zero out the sub-diagonal elements. In the above example, we first exchanged the 2nd and 3rd rows, and then added the 1st row to the 2nd one to get $\mathbf{U}$ (the right-most matrix). The element $a_{11} = 1$ is boxed in order to emphasize that it is a **pivot** - a nonzero diagonal element of $\mathbf{U}$. The fact that $\mathbf{U}$ has only 1 pivot means that its **rank**, and thus that of $\mathbf{A}$'s too, is $r = 1$. Because in this case $\mathbf{n} \in \mathcal{R}^2$, $\mathbf{A}$'s highest possible rank is 2, which means that the nullspace dimension is 1. Thus we are looking for a spanning set (in this case a one-member set) for the nullspace, which satisfies

$$\mathbf{A}\mathbf{n} = \mathbf{U}\mathbf{n} = \mathbf{0}$$

or

$$\begin{pmatrix} 1 & -2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \alpha = 2\beta.$$

Therefore,

$$\mathbf{n} = \gamma \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

for any arbitrary γ is a valid basis vector for spanning $\mathbf{A}$'s nullspace. It is traditional, but not fundamentally necessary, to normalize basis vectors to unit norm,

$$\hat{\mathbf{n}} = \frac{\mathbf{n}}{\|\mathbf{n}\|} \equiv \frac{\mathbf{n}}{\sqrt{\mathbf{n}^T\mathbf{n}}}$$

(where $\|\mathbf{n}\|$ is $\mathbf{n}$'s L_2-norm). To complete the tradition, unit basis vectors are often adorned with a hat as above.

To check our result,

$$\mathbf{An} = \begin{pmatrix} 1 & -2 \\ 0 & 0 \\ -1 & 2 \end{pmatrix} \alpha \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} 1 & -2 \\ 0 & 0 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} =$$

$$= \alpha \left[2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix} \right] = \alpha \begin{pmatrix} 2-2 \\ 0+0 \\ -2+2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \in \mathcal{R}^3,$$

as necessary.

- $$\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 4 & 7 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \\ 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \rightarrow$$

$$\rightarrow \begin{pmatrix} \boxed{1} & 1 & 2 \\ 0 & \boxed{1} & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \equiv \mathbf{U}.$$

Because now $\mathbf{n} \equiv (\ \alpha\ \beta\ \gamma\ )^T \in \mathcal{R}^3$ and $r = 2$ (corresponding to the 2 boxed nonzero pivots in $\mathbf{U}$), the nullspace is still 1-dimensional. From the 2nd row

$$\beta = -\gamma,$$

and from the 1st row

$$\alpha - \gamma + 2\gamma = 0 \quad \longrightarrow \quad \alpha = -\gamma.$$

We next set $\gamma = 1$, and obtain the spanning vector for the 1D nullspace,

$$\mathbf{n} = \mu \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

for any arbitrary μ.

To check, as usual, we compute

$$\mathbf{An} = \mu \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 4 & 7 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} = \mu \begin{pmatrix} -1-1+2 \\ -1-2+3 \\ -2-3+5 \\ -3-4+7 \end{pmatrix} = \mu \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

as required.

- $$\begin{pmatrix} 1 & 1 & -1 & 3 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 1 & 4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & -1 & 3 \\ 1 & 2 & 0 & 3 \\ 0 & 1 & 3 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & -1 & 3 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 3 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} \boxed{1} & 1 & -1 & 3 \\ 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & \boxed{2} & -2 \end{pmatrix}.$$

Since there are 3 pivots, $r = 3$, so there is only 1 left dimension for the nullspace, i.e., it is spanned by a single $\mathbf{n} \in \mathcal{R}^4$. Assuming $\mathbf{n} = (\ \alpha\ \beta\ \ \gamma\ \delta\ )^T$ and given that we have only 3 constraints (corresponding to $r = 3$ from the 3 nonzero pivots), we can set 1 of $\mathbf{n}$'s 4 elements at will. Let's set $\delta = 1$, so that

$$\begin{pmatrix} 1 & 1 & -1 & 3 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 2 & -2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

From the 3rd row, $2\gamma - 2 = 0 \quad \rightarrow \quad \gamma = 1$, from the 2nd row $\beta = -\gamma = -1$, and then from the 1st row $\alpha = 1 + 1 - 3 = -1$, so

$$\mathbf{n} = \mu \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} .$$

To check, again,

$$\mathbf{An} = \mu \begin{pmatrix} 1 & 1 & -1 & 3 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \mu \begin{pmatrix} -1-1-1+3 \\ -1-2+0+3 \\ -2-3+1+4 \end{pmatrix} = \mu \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

as required of a nullspace basis vector.

-
$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 2 & -2 \\ 1 & 0 & 2 & -1 \\ 1 & -1 & 4 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 1 & 2 & -2 \\ 1 & 0 & 2 & -1 \\ 0 & -1 & 2 & 0 \end{pmatrix} \rightarrow$$
$$\rightarrow \begin{pmatrix} 2 & 1 & 2 & -2 \\ 0 & -\frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 1 & 2 & -2 \\ 0 & -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

so $r = 2$, and, since $\mathbf{n} = (\ \alpha\ \beta\ \gamma\ \delta\ )^T \in \mathcal{R}^4$, the nullspace is 2-dimensional. From the 2nd row, $\beta = 2\gamma$, and from the 1st row,

$$\begin{aligned} 2\alpha + \beta + 2\gamma - 2\delta &= 0 \\ 2\alpha + 4\gamma - 2\delta &= 0 \\ \alpha + 2\gamma &= \delta \end{aligned} .$$

Therefore, the vector we are looking for looks like

$$\mathbf{n} = \begin{pmatrix} \alpha \\ 2\gamma \\ \gamma \\ \alpha + 2\gamma \end{pmatrix} \rightarrow \begin{cases} \text{set } \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \mathbf{n}_1 = \mu \begin{pmatrix} 0 \\ 2 \\ 1 \\ 2 \end{pmatrix} \\ \\ \text{set } \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \mathbf{n}_2 = \nu \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \end{cases}$$

for arbitrary μ and ν.

Check:

$$\mathbf{A}\mathbf{n}_1 = \mu \begin{pmatrix} 2 & 1 & 2 & -2 \\ 1 & 0 & 2 & -1 \\ 1 & -1 & 4 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 1 \\ 2 \end{pmatrix} = \mu \begin{pmatrix} 2+2-4 \\ 2-2 \\ -2+4-2 \end{pmatrix} = \mu \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\mathbf{A}\mathbf{n}_2 = \nu \begin{pmatrix} 2 & 1 & 2 & -2 \\ 1 & 0 & 2 & -1 \\ 1 & -1 & 4 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \nu \begin{pmatrix} 2-2 \\ 1-1 \\ 1-1 \end{pmatrix} = \nu \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Now let's (re)examine the ODE example

$$\frac{d\mathbf{x}}{dt} \equiv \frac{d}{dt} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 9 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

and try the solution

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} e^{\lambda t}.$$

Substituting this solution and its derivative in the ODE leads to

$$\begin{array}{rcrcl} \lambda a e^{\lambda t} & = & a e^{\lambda t} & + & b e^{\lambda t} \\ \lambda b e^{\lambda t} & = & 9 a e^{\lambda t} & + & b e^{\lambda t} \end{array}.$$

Because $e^{\lambda t} \neq 0$ always, we can divide by it, so

$$\begin{array}{rcrcl} \lambda a & = & a & + & b \\ \lambda b & = & 9a & + & b \end{array} \quad \rightarrow \quad \lambda \mathbf{u} = \mathbf{A}\mathbf{u}$$

where $\mathbf{u} \equiv (\ a\ b\ )^T$. Rearranging, we get

$$(\mathbf{A} - \lambda \mathbf{I})\,\mathbf{u} = \mathbf{0}$$

where $\mathbf{I}$, the identity matrix (diagonal matrix with the the diagonal elements being ones and all off-diagonal elements being zero), is needed to keep the equation dimensionally sensible. Thus we are looking for a set of N scalars $\lambda_i, i = 1 \cdots N$ such that the matrix $\mathbf{A} - \lambda\mathbf{I}$ is singular, i.e., has a non-trivial nullspace. For that to be the case, its **determinant** must vanish. For the 2×2 case,

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$$

(where the as here correspond to $\mathbf{A}$'s elements, not the coefficient). In this particular case,

$$\det\left[\begin{pmatrix} 1-\lambda & 1 \\ 9 & 1-\lambda \end{pmatrix}\right] = 0.$$

This leads to the celebrated **characteristic polynomial**, in this case

$$(1-\lambda)^2 - 9 = 0 \quad \rightarrow \quad \lambda^2 - 2\lambda - 8 = 0,$$

with roots

$$\lambda_{1,2} = \frac{2 \pm \sqrt{4+32}}{2} = \frac{2 \pm 6}{2} = \left\{ \begin{array}{rcr} \lambda_1 & = & 4 \\ \lambda_2 & = & -2 \end{array} \right.$$

which are $\mathbf{A}$'s **eigenvalues**. Next we need to find the corresponding **eigenvectors** $\mathbf{u}_{1,2}$ which satisfy individually $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ with $\mathbf{u} = (\ a\ b\ )^T$ as before.

For $\underline{\lambda_1 = 4}$,

$$\begin{pmatrix} 1 & 1 \\ 9 & 1 \end{pmatrix} \mathbf{u}_1 = 4\mathbf{u}_1$$

or, equivalently

$$\begin{pmatrix} -3 & 1 \\ 9 & -3 \end{pmatrix} \mathbf{u}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} -3 & 1 \\ 0 & 0 \end{pmatrix} \mathbf{u}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \rightarrow \quad \mathbf{u}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

times any constant, which we take to be one.

For $\underline{\lambda_2 = -2}$, we get

$$\begin{pmatrix} 1 & 1 \\ 9 & 1 \end{pmatrix} \mathbf{u}_2 = -2\mathbf{u}_2$$

or, equivalently

$$\begin{pmatrix} 3 & 1 \\ 9 & 3 \end{pmatrix} \mathbf{u}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix} \mathbf{u}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \rightarrow \quad \mathbf{u}_2 = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

times any constant, which we take again to be one.

Before we proceed, let's test our findings regarding the eigenvalue/eigenvector pairs. If each satisfies the eigen equation $\mathbf{Au}_i = \lambda_i \mathbf{u}_i$ individually, they must satisfy together

$$\mathbf{A} \begin{pmatrix} | & | \\ \mathbf{u}_1 & \mathbf{u}_2 \\ | & | \end{pmatrix} = \begin{pmatrix} | & | \\ \mathbf{u}_1 & \mathbf{u}_2 \\ | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

which in this particular case is

$$\begin{pmatrix} 1 & 1 \\ 9 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & -3 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$$

or

$$\begin{pmatrix} 4 & -2 \\ 12 & 6 \end{pmatrix} = \begin{pmatrix} 4 & -2 \\ 12 & 6 \end{pmatrix}$$

which vindicates our results of $\lambda_{1,2}$ and $\mathbf{u}_{1,2}$.

With these substantiated results, the solution of the original ODE is

$$\mathbf{x}(t) \equiv \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = a \begin{pmatrix} 1 \\ 3 \end{pmatrix} e^{4t} + b \begin{pmatrix} 1 \\ -3 \end{pmatrix} e^{-2t}.$$

To complete the solution, we need **initial conditions**, given symbolically as $\mathbf{x}_o = (\ x_o\ y_o\ )^T$ at $t = 0$. At that time, the exponentials are both one, so the equation at $t = 0$ reads

$$\begin{pmatrix} x(t=0) \\ y(t=0) \end{pmatrix} = \begin{pmatrix} x_o \\ y_o \end{pmatrix} = \begin{pmatrix} a+b \\ 3a-3b \end{pmatrix} =$$

which, after some trivial algebra, leads to

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \frac{1}{2}x_o & + & \frac{1}{6}y_o \\ \frac{1}{2}x_o & - & \frac{1}{6}y_o \end{pmatrix}$$

which leads to our complete solution

$$\mathbf{x}(t) = \left(\frac{1}{2}x_o + \frac{1}{6}y_o\right)\begin{pmatrix} 1 \\ 3 \end{pmatrix} e^{4t} + \left(\frac{1}{2}x_o - \frac{1}{6}y_o\right)\begin{pmatrix} 1 \\ -3 \end{pmatrix} e^{-2t}.$$

Notice that for large t, the solution is dominated by the 1st right-hand term,

$$\mathbf{x}\big|_{t \gg t_o} \approx \left(\frac{1}{2}x_o + \frac{1}{6}y_o\right)\begin{pmatrix} 1 \\ 3 \end{pmatrix} e^{4t},$$

which is called, consequently, the **gravest mode** or, less pompously, the **leading mode**. Beyond, pomp, though, what is important is that the solution becomes dominated by the mode with the largest (real part of the) eigenvalue; this is a general lesson!

Is $\mathbf{A}$ a ***normal*** matrix? Normal matrices satisfy

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T.$$

In this case,

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 1 & 9 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 9 & 1 \end{pmatrix} = \begin{pmatrix} 82 & 10 \\ 10 & 2 \end{pmatrix}$$

and

$$\mathbf{A}\mathbf{A}^T = \begin{pmatrix} 1 & 1 \\ 9 & 1 \end{pmatrix}\begin{pmatrix} 1 & 9 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 10 \\ 10 & 82 \end{pmatrix}$$

so, no, $\mathbf{A}$ is non-normal. A consequence of that is that its eigenvectors are not mutually orthogonal,

$$\mathbf{u}_1^T\mathbf{u}_2 = \begin{pmatrix} 1 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ -3 \end{pmatrix} = -8 \neq 0$$

which means that even if the system were asymptotically stable, some transient growth is still possible because of projection of information from one eigenvector to the other.

Let's look at an **example in which these issues become clearer**. For clarity, it is more helpful to consider a discrete-time linear model of the form

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i,$$

like, e.g., the age-structured Spotted Owl model we've encountered before. To get things rolling, let's consider a specific example with

$$\mathbf{A} = \frac{1}{10}\begin{pmatrix} 1 & 1 \\ -24 & -9 \end{pmatrix} \qquad \text{and} \qquad \mathbf{x}(t) \equiv \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}.$$

We start, as always, by eigen-analyzing $\mathbf{A}$, first solving for the λ_i using

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{u} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \rightarrow \qquad \frac{1}{10}\begin{vmatrix} 1-10\lambda & 1 \\ -24 & -9-10\lambda \end{vmatrix} = 0$$

where the last (right-most) zero is a scalar, and $|\cdot|$ is a notation often used instead of $\det(\cdot)$. This leads to the characteristic polynomial

$$-(1-10\lambda)(9+10\lambda) + 24 = 0$$

$$100\lambda^2 + 80\lambda + 15 = 0$$

with roots

$$\lambda_{1,2} = \frac{-80 \pm \sqrt{6400-6000}}{200} = \frac{-80 \pm 20}{200} \qquad \rightarrow \qquad \left\{ \begin{array}{lcl} \lambda_1 & = & -3/10 \\ \lambda_2 & = & -5/10 \end{array} \right.$$

i.e., the system is asymptotically stable because both eigenvalues satisfy $|\lambda_i| < 1$.

Next the eigenvectors,

$$\frac{1}{10}\begin{pmatrix} 1 & 1 \\ -24 & -9 \end{pmatrix}\mathbf{u}_1 = -\frac{3}{10}\mathbf{u}_1 \rightarrow \begin{pmatrix} 4 & 1 \\ -24 & -6 \end{pmatrix}\mathbf{u}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 4 & 1 \\ 0 & 0 \end{pmatrix}\mathbf{u}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\frac{1}{10}\begin{pmatrix} 1 & 1 \\ -24 & -9 \end{pmatrix}\mathbf{u}_2 = -\frac{5}{10}\mathbf{u}_2 \rightarrow \begin{pmatrix} 6 & 1 \\ -24 & -4 \end{pmatrix}\mathbf{u}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 6 & 1 \\ 0 & 0 \end{pmatrix}\mathbf{u}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which we solve for,

$$\mathbf{u}_1^{(r)} = \begin{pmatrix} 1 \\ -4 \end{pmatrix} \qquad \text{and} \qquad \mathbf{u}_2^{(r)} = \begin{pmatrix} 1 \\ -6 \end{pmatrix}$$

(where the superscript stands for 'raw'), normalize,

$$\mathbf{u}_1 = \frac{\mathbf{u}_1^{(r)}}{\|\mathbf{u}_1^{(r)}\|} \qquad \text{and} \qquad \mathbf{u}_2 = \frac{\mathbf{u}_2^{(r)}}{\|\mathbf{u}_2^{(r)}\|},$$

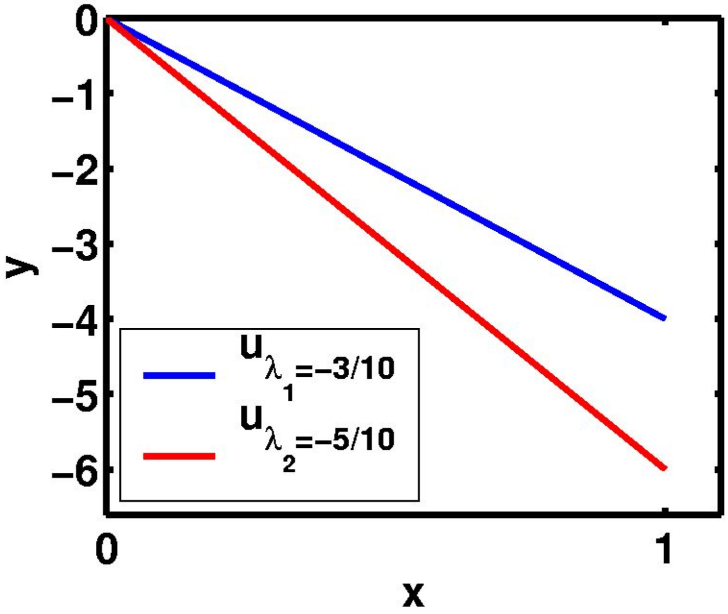

Figure 1: The 2 eigenvectors of the system discussed in the text.

and place in a single matrix $\mathbf{E} = (\ \mathbf{u}_1\ \mathbf{u}_2\ )$ of eigenvectors

$$\mathbf{AE} = \mathbf{ED} \quad \rightarrow \quad \underbrace{\frac{1}{10}\begin{pmatrix} 1 & 1 \\ -24 & -9 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} 1 & 1 \\ -4 & -6 \end{pmatrix}}_{\text{raw } \mathbf{E}} = \underbrace{\begin{pmatrix} 1 & 1 \\ -4 & -6 \end{pmatrix}}_{\text{raw } \mathbf{E}} \underbrace{\frac{1}{10}\begin{pmatrix} -3 & 0 \\ 0 & -5 \end{pmatrix}}_{\mathbf{D}}$$

which you can verify holds because

$$10\mathbf{AE}^{(r)} = 10\mathbf{E}^{(r)}\mathbf{D} = \begin{pmatrix} -3 & -5 \\ 12 & 30 \end{pmatrix}.$$

Now let's examine the orthogonality of the eigenvectors, shown in Fig. 1,

$$(\ 1\ \ -4\ )\begin{pmatrix} 1 \\ -6 \end{pmatrix} = 25 \neq 0$$

so, no, the eigenvectors are not orthogonal. In fact, using some trivial yet slightly tedious algebra, we can show that $\mathbf{u}_1$ and $\mathbf{u}_2$ are nearly parallel, with an angle $\theta \approx 0.08^\circ$ between them. (That this is so can be also inferred from the fact that $\mathbf{u}_1^T\mathbf{u}_2 = 25 \approx \sqrt{(\mathbf{u}_1^T\mathbf{u}_1)(\mathbf{u}_2^T\mathbf{u}_2)} \simeq 25.08$.) That the eigenvectors, while possibly linearly independent and forming a complete spanning set, are not mutually orthogonal is a remarkable general property of matrices for which $\mathbf{A}^T\mathbf{A} \neq \mathbf{A}\mathbf{A}^T$, with potentially profound implications for transient (i.e., non-asymptotic) stability, as shown in Fig. 2.

To better understand the issue of non-normality, note that we can use the decomposition of $\mathbf{A}$,

$$\mathbf{AE} = \mathbf{ED}$$

and the inverse of $\mathbf{E}$ to rewrite the model as

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i = \mathbf{EDE}^{-1}\mathbf{x}_i.$$

(If you are not sure what the inverse is or how to find it, please refer to any linear algebra book, or, if you prefer, the Appendix.)

Defining $\mathbf{y}_i \equiv [\ y_i(1)\ y_i(2)\ ]^T \equiv \mathbf{E}^{-1}\mathbf{x}_i$, we can rewrite our 2×2 discrete-time model

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i = \mathbf{EDE}^{-1}\mathbf{x}_i$$

as

$$\mathbf{x}_{i+1} = \mathbf{ED}\mathbf{y}_i.$$

Schematically, the above equation is

$$\begin{aligned}
\mathbf{x}_{i+1} &= \begin{pmatrix} | & | \\ \mathbf{u}_1 & \mathbf{u}_2 \\ | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} | \\ \mathbf{y}_i \\ | \end{pmatrix} \\
&= \begin{pmatrix} | & | \\ \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 \\ | & | \end{pmatrix} \begin{pmatrix} | \\ \mathbf{y}_i \\ | \end{pmatrix} \\
&= y_i(1)\lambda_1 \begin{pmatrix} | \\ \mathbf{u}_1 \\ | \end{pmatrix} + y_i(2)\lambda_2 \begin{pmatrix} | \\ \mathbf{u}_2 \\ | \end{pmatrix} \\
&= \alpha \begin{pmatrix} | \\ \mathbf{u}_1 \\ | \end{pmatrix} + \beta \begin{pmatrix} | \\ \mathbf{u}_2 \\ | \end{pmatrix}
\end{aligned}$$

where the last step defines $\alpha \equiv y_i(1)\lambda_1$ and $\beta \equiv y_i(2)\lambda_2$. This representation of the state vector emphasizes the fact that the solution at any given time is a linear combination of the eigenvectors. The effect of the non-normality of $\mathbf{A}$ is to emphasize one of the eigenvectors at the expense of the other by dictating that either $\alpha \ll \beta$ or $\beta \ll \alpha$. This effect in our particular example is shown in Fig. 2. Panel a, where the 2 element of the state vector are shown as a function of time, simply shows that this is in fact the case, namely that despite our $\mathbf{A}$ being asymptotically stable, some transient growth is realized. A different representation of essentially the same phenomenon is shown in panel b, where $\|\mathbf{x}_t\|$ is displayed. Panels c and d delve a bit deeper into this phenomenon by showing where exactly the growth arises from. Together, these 2 panels show that at the beginning of the march forward in time, $\alpha \approx \beta/2$ so that the coefficient of $\mathbf{u}_1$ in the solution is only roughly one-half that of $\mathbf{u}_2$'s coefficient. The effect of this disparity may be understood with a more extreme example. Consider a hypothetical problem whose eigenvectors are

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{u}_1 = \begin{pmatrix} 1 \\ 1.01 \end{pmatrix}.$$

Clearly this is a perfectly fine $\mathcal{R}^2$-spanning complete set; since the 2 are linearly independent, any 2-vector can be represented as their linear combination. However, suppose the vector we wish to represent as the linear combination of those eigenvectors is one that's nearly orthogonal to both,

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1.01 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

From the top row, we infer that $\beta = 1 - \alpha$ while from multiplying the bottom row by 100 we infer that $-100 = 101\alpha + 100\beta$. Substituting the former into the latter yields $\alpha = -200$ and thus $\beta = 201$, so that

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} = -200 \begin{pmatrix} 1 \\ 1.01 \end{pmatrix} + 201 \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

What this example shows is that while 2 linearly independent vectors that are very nearly parallel do indeed span their corresponding space, in this case $\mathcal{R}^2$, they do so very un-parsimoniously, in that they require very large opposite sign coefficients.

Going back to our linear 2×2 discrete-time model in which $\mathbf{u}_1$ and $\mathbf{u}_2$ are also very nearly parallel, we see that if at a given time the state vector is nearly orthogonal to the 2 (nearly parallel) eigenvectors, it will require very large coefficients to form the eigenvector expansion of the state. In that case, the small difference

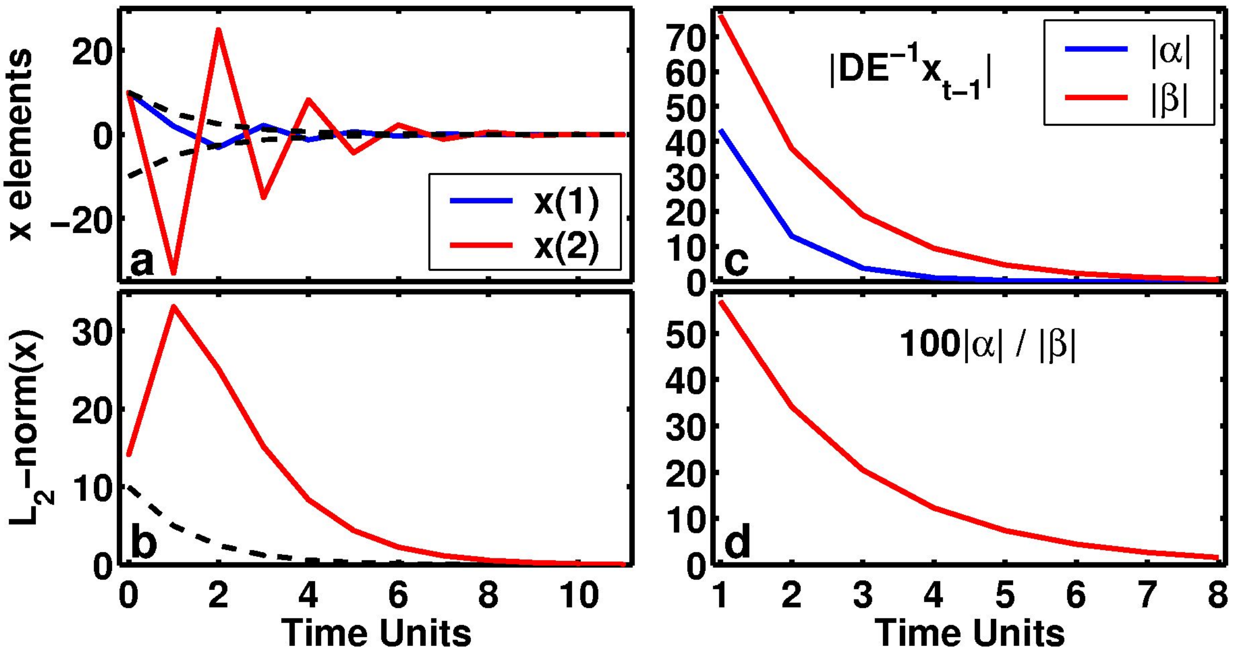

Figure 2: Transient perturbation growth in the asymptotically stable non-normal system of linear difference equations discussed in the text. In panel a both age groups (the 2 elements of the state vector $\mathbf{x}(t)$) are plotted in red and blue. The initial state is $\mathbf{x} = (\ 10\ 10\ )^T$. The dashed lines show the expected geometrical decay given the least damped mode, $\lambda_2 = -5/10$, $10|\lambda|^t$, $t = 0, 1 \cdots 11$. Panel b shows the norm of the state vector, $\|\mathbf{x}(t)\|$, as well as its expected geometric decay based on asymptotic theory. Panel c shows the absolute values of the 2 elements of $\mathbf{DE}^{-1}\mathbf{x}_{t-1}$ (α and β in the text) as a function of time index t, while panel d shows their ratio, $100|\alpha|/\beta$.

between the eigenvalues is magnified many times over by their multiplication by the very large coefficients. **In summary**, the near-parallelism of the eigenvectors requires very large coefficients. On multiplication by the disparate eigenvalues, those coefficients that correspond to trailing modes (small eigenvalues) nearly vanish, clearing the stage for the leading-mode coefficient to dominate, and amplify the solution norm. If you find this explanation less than fully satisfying, do one of two: if you are familiar with the SVD, go right ahead. If not, go back to the course home page, and familiarize yourself with SVD first, then proceed.

Analyzing Non-Normality Using the SVD

To start, let's obtain the SVD representation of

$$\mathbf{A} = \frac{1}{10}\begin{pmatrix} 1 & 1 \\ -24 & -9 \end{pmatrix}$$

using

$$\begin{aligned}
\mathbf{A}\mathbf{A}^T &= \tfrac{1}{100}\begin{pmatrix} 2 & -33 \\ -33 & 657 \end{pmatrix} \\
\mathbf{A}^T\mathbf{A} &= \tfrac{1}{100}\begin{pmatrix} 577 & 217 \\ 217 & 82 \end{pmatrix}.
\end{aligned}$$

Since you already know a bunch about eigen analysis, let's proceed in an abbreviated fashion. The 2 characteristic polynomials are

$$\frac{1}{100}\left[(2-100\lambda)(657-100\lambda)-33^2\right]=0$$

$$\frac{1}{100}\left[(577-100\lambda)(82-100\lambda)-217^2\right]=0,$$

and their roots yield the same eigenvalues of both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$,

$$\begin{aligned}
\lambda_1 &= 6.587 \\
\lambda_2 &= 0.003.
\end{aligned}$$

The factor of ∼2000 separating the 2 eigenvalues implicitly already tells our story, as it is clear that one mode will completely dominate. Let's call $\mathbf{U}$ the matrix containing the normalized (to unit L_2-norm) eigenvectors of $\mathbf{A}\mathbf{A}^T$, and $\mathbf{V}$ the one containing those of $\mathbf{A}^T\mathbf{A}$,

$$\begin{aligned}
\mathbf{U} &= \begin{pmatrix} -0.050 & 0.999 \\ 0.999 & 0.050 \end{pmatrix} \\
\mathbf{V} &= \begin{pmatrix} -0.936 & -0.352 \\ -0.352 & 0.936 \end{pmatrix},
\end{aligned}$$

which together yield the SVD representation of $\mathbf{A}$,

$$\mathbf{A} = \underbrace{\begin{pmatrix} -0.050 & 0.999 \\ 0.999 & 0.050 \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} 2.566 & 0 \\ 0 & 0.058 \end{pmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{pmatrix} -0.936 & -0.352 \\ -0.352 & 0.936 \end{pmatrix}}_{\mathbf{V}^T}$$

where $2.566=\sqrt{6.587}$ and $0.058=\sqrt{0.003}$, and the fact that one of the singular values is larger than one clearly indicates the potential growth due to the non-normality of $\mathbf{A}$. Using this representation, we rewrite the model as

$$\begin{aligned}
\mathbf{x}_{t+1} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}_t \\
&= \begin{pmatrix} | & | \\ \mathbf{u}_1 & \mathbf{u}_2 \\ | & | \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \text{---} & \mathbf{v}_1^T & \text{---} \\ \text{---} & \mathbf{v}_2^T & \text{---} \end{pmatrix} \begin{pmatrix} | \\ \mathbf{x}_t \\ | \end{pmatrix} \\
&= \begin{pmatrix} | & | \\ \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 \\ | & | \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T\mathbf{x}_t \\ \\ \mathbf{v}_2^T\mathbf{x}_t \end{pmatrix} \\
&= \sigma_1\mathbf{v}_1^T\mathbf{x}_t \begin{pmatrix} | \\ \mathbf{u}_1 \\ | \end{pmatrix} + \sigma_2\mathbf{v}_2^T\mathbf{x}_t \begin{pmatrix} | \\ \mathbf{u}_2 \\ | \end{pmatrix} \\
&\approx \sigma_1\mathbf{v}_1^T\mathbf{x}_t \begin{pmatrix} | \\ \mathbf{u}_1 \\ | \end{pmatrix}
\end{aligned}$$

where the final approximation is due to $\sigma_1 > \sigma_2 > \cdots > \sigma_N$ in general, and $\sigma_1 \approx 44\sigma_2$ in our particular example. Note that while this expansion of the solution (the state vector at $t+1$) in the left singular vectors $\{\mathbf{u}_i\}$ is similar in appearance to its previous expansion in $\mathbf{A}$'s eigenvectors, it has a clear distinction that now we need not define $\mathbf{y}_i$ as before, and the disparity in the representation of each of the modes in the solution is therefore much more transparent. It includes, as before, the 2 key ingredients: the necessity for very large coefficients that follow from the near-parallelism of the spanning vectors (in this case the left singular vectors $\mathbf{u}_i$), and the disparity of the scalars that multiply the coefficients, in the SVD case the singular values σ_i.

As a final aside, let's look briefly at the use of the SVD to compute **the norm of a matrix**. This is not as widely agreed upon as the norm of a vector, where a simple "norm" with specification almost always means the L_2-norm. For matrices, the norm and appears to be related to the context. For data matrices (often in the context of inverse calculations) the matrix norm is simply a two-dimensional generalization of the L_2 norm of a vector,

$$\|\mathbf{A}\| \equiv \sqrt{\sum_i \sum_j \mathbf{A}_{ij}^2}.$$

However, in many dynamical systems context, it is often more natural to define the norm as the maximum growth potential presented by the matrix, in which case

$$\|\mathbf{A}\| \equiv \max_i(\sigma_i),$$

where σ_i are the singular values of the matrix. Thus, for example, in the discrete time model discussed earlier,

$$\|\mathbf{E}\| \approx 1.4$$

while

$$\|\mathbf{E}^{-1}\| \approx 17.7,$$

which shows exactly where the potential growth is hidden.

Appendix I: The Inverse of a matrix

A matrix inverse is to that matrix what $1/a$ is to the scalar a, i.e., it's the multi-dimensional generalization of the division by 1. For a scalar, clearly

$$a\frac{1}{a} = 1.$$

Similarly, for a matrix $\mathbf{A}$

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I},$$

where

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \vdots & \vdots & \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

is the $N \times N$ identity matrix which plays for $\mathcal{R}^N$ the same role as the number 1 plays for the real line $\mathcal{R}^1$.

So how do we actually compute the inverse? It is actually a costly and flaky calculation which in practice should be avoided if at all possible. However, the basic idea is simple, and can be approached from a number of angles. Let's denote

$$\mathbf{E}^{-1} \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

so that, from the conditions satisfied by the inverse,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -4 & -6 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -4 & -6 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Solving the latter of the 2, the 1st row yields

$$a + c = 1 \rightarrow c = 1 - a$$

and

$$b + d = 0 \rightarrow d = -b$$

which we substitute into the 2nd row to get

$$-4a = 6c = 6 - 6a \rightarrow a = 3 \rightarrow c = -2$$

and

$$-4b - 6d = 1 \rightarrow -4b + 6b = 1 \rightarrow b = \frac{1}{2} \rightarrow d = -\frac{1}{2},$$

so that

$$\mathbf{E}^{-1} = \begin{pmatrix} 3 & \frac{1}{2} \\ -2 & -\frac{1}{2} \end{pmatrix}.$$

To check this, compute

$$\mathbf{E}\mathbf{E}^{-1} = \begin{pmatrix} 1 & 1 \\ -4 & -6 \end{pmatrix} \begin{pmatrix} 3 & \frac{1}{2} \\ -2 & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{E}^{-1}\mathbf{E} = \begin{pmatrix} 3 & \frac{1}{2} \\ -2 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -4 & -6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

which shows that the above $\mathbf{E}^{-1}$ is indeed $\mathbf{E}$'s inverse.

Just for your information, note that the inverse of a 2 × 2 matrix

$$\mathbf{G} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is given by

$$\mathbf{G}^{-1} = \frac{1}{\det \mathbf{A}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

which, for our matrix, yields

$$\mathbf{E}^{-1} = \frac{1}{-6+4} \begin{pmatrix} -6 & -1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 3 & \frac{1}{2} \\ -2 & -\frac{1}{2} \end{pmatrix}$$

as required.

Finally, for any matrix $\mathbf{A}$, you can append an identity matrix of the appropriate dimension to the right of $\mathbf{A}$,

$$(\mathbf{A} \mid \mathbf{I})$$

and try to reduce $\mathbf{A}$'s elements (to the left of the separator) into $\mathbf{I}$. If you can do that (while, of course, treating the rows to the right of the separator the same way you treat $\mathbf{A}$'s rows) then once the left part due to $\mathbf{A}$ has been reduced to $\mathbf{I}$, the part due to the original $\mathbf{I}$ to the right of the separator will be $\mathbf{A}^{-1}$. An example will clarify matters, and let's use the above $\mathbf{E}$ as our example. We start by constructing

$$\left(\begin{array}{rr|rr} 1 & 1 & 1 & 0 \\ -4 & -6 & 0 & 1 \end{array}\right) \to \left(\begin{array}{rr|rr} 1 & 1 & 1 & 0 \\ 0 & -2 & 4 & 1 \end{array}\right) \to$$

$$\to \left(\begin{array}{rr|rr} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & -\frac{1}{2} \end{array}\right) \to \left(\begin{array}{rr|rr} 1 & 0 & 3 & \frac{1}{2} \\ 0 & 1 & -2 & -\frac{1}{2} \end{array}\right) = \left(\mathbf{I} \mid \mathbf{E}^{-1}\right)$$