# Lecture 12: Linear algebra review

## Recap

In recent lectures, we've been looking at least-squares fitting problems, all of which reqire inversion of a matrix. We know that a matrix with repeated rows or columns does not have an inverse, but there are of course more details to this. You've likely seen some of this in a previous linear algebra class, but we'll take a moment to review key concepts.

## A brief review of linear vector spaces

We began with the inverse problem of minimizing $||\mathbf{Gm} - \mathbf{d}||_2$, in which our primary concern is the matrix $\mathbf{G}$. The matrix $\mathbf{G}$ told us about the model but had no information about the data. We will use the language of linear algebra to discuss how $\mathbf{G}$ is a transformation of the vector $\mathbf{m}$ into the vector $\mathbf{d}$.

By the time we are done with this review, we will have the tools to determine which combination of model parameters are most important to fit data, and which data contribute most to misfit. Notions of resolution in model parameters and data will be forthcoming. These same tools will help to understand empirical orthogonal functions, which we will cover later.

*Linear vector spaces*

You probably already have a feeling for what a vector space is simply be considering three-dimensional physical space. The nifty thing about vector spaces is that the allow us to "see" abstract relations in geometrical terms. It is worth remembering what a physicist thinks of a "vector". To a physicist, a vector is a quantity with a magnitude and a direction that combines with other vectors according to a specific rule. Velocity, force, and displacement are examples of quantities conveniently expressed as vectors. A sum of two vectors is obtained by putting the second vector's tail to the first vector's head (Figure 1). If we change the coordinate system (for example, by rotation, as we did earlier when we considered variance ellipses) from which we are observing a vector, a physicist would say that the essential vector quantity stays the same, but the expression of the vector in the new coordinate system would be different.

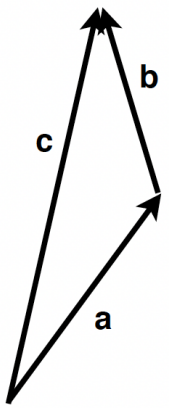


Figure 1: Example summed vectors.

We are about to embark on a review of linear vector spaces, in which such a change of coordinate systems, which we will call a transformation, results in a "new" vector. Rotation, reflection, and rescaling are ways of changing a coordinate system, and will be described as transformations.

Another kind of transformation is projection. An example is the projection of a two-dimensional vector onto the horizontal axis. The formalism of linear vector spaces allows compact description of these transformations, which we will use to understand the matrix $\mathbf{G}$.

*Definition of a linear vector space*

A **linear vector space** consists of:

1. A field $F$ of scalars (for example, real numbers).
2. A set $V$ of entities called vectors.
3. An operation called vector addition such that
   a. Addition is commutative and associative: $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$, $\mathbf{x}+(\mathbf{y}+\mathbf{z})=(\mathbf{x}+\mathbf{y})+\mathbf{z}$.
   b. There exists a null vector $\mathbf{0}$ such that $\mathbf{x}+\mathbf{0}=\mathbf{x}$.
4. An operation called scalar multiplication such that for $\alpha$ in $F$, $\mathbf{x}$ in $V$, then $\alpha\mathbf{x}\in V$ and
   a. Scalar multiplication is associative, and distributive with respect to scalar addition: $\alpha(\beta\mathbf{x})=(\alpha\beta)\mathbf{x}$, $(\alpha+\beta)\mathbf{x}=\alpha\mathbf{x}+\beta\mathbf{x}$
   b. Scalar multiplication is distributive with respect to vector addition: $\alpha(\mathbf{x}+\mathbf{y})=\alpha\mathbf{x}+\beta\mathbf{y}$
   c. $1\cdot\mathbf{x}=\mathbf{x}$.

*Examples of vector spaces*

The space $R^N$ of ordered sequences of $N$ real numbers is an example of a linear vector space. The data vector $\mathbf{d}$ is an element of $R^N$, and $\mathbf{m}$ is an element of $R^M$. Addition is defined as

$$\mathbf{x}+\mathbf{y}=(x_1+y_1,x_2+y_2,\ldots,x_N+y_N) \tag{1}$$

Scalar multiplication is defined as

$$\alpha\mathbf{x}=(\alpha x_1,\alpha x_2,\ldots,\alpha x_N) \tag{2}$$

$R^3$ is the Cartesian representation of three-dimensional space. $R$ is the set of real numbers, and is a perfectly valid vector space.

*Linear combination, linear independence, linear dependence*

A **linear combination** of vectors $\mathbf{f}_1,\mathbf{f}_2,\ldots,\mathbf{f}_K$ is a vector of the form

$$\mathbf{g}=\sum_{i=1}^{K}\alpha_i\mathbf{f}_i. \tag{3}$$

If for some choice of $\alpha_i$ (where not all the $\alpha_i$ are zero) $\mathbf{g}=0$, then the vectors $\mathbf{f}_i$ are **linearly dependent**. When a set of vectors is linearly dependent it is possible to express one of the vectors as a linear combination of the others. The vectors $\mathbf{f}1,\mathbf{f}_2,\ldots,\mathbf{f}_K$ are **linearly independent** if $\mathbf{g}=0$ if and only if $\alpha_i=0$.

*Subspace, basis, dimension, rank*

The set of all linear combinations formed from a fixed collection of vectors is a **subspace** of the original space. The fixed vectors are said to **span** the subspace. A **basis** for a vector space is a linearly independent set of vectors that spans the space. If the number of vectors in the space is finite then the space is **finite-dimensional**. For a given finite-dimensional space, there may be more than one basis, but all such bases contain the same number of vectors. This number is called the **dimension** of the space.

Consider these ideas in physical space. Valid bases are created by arbitrary rotations, reflections, and scalings of the Cartesian coordinate system. However, it is intuitively obvious that 3 basis vectors are required in any case.

The **row space** of an $N \times M$ matrix $\mathbf{G}$ is that subspace of $R^M$ spanned by the rows of $\mathbf{G}$. The **column space** of $\mathbf{G}$ is that subspace of $R^N$ spanned by the columns of $\mathbf{G}$. The row and column spaces of $\mathbf{G}$ have the same dimension. This integer is called the **rank** of $\mathbf{G}$. If the rank of $\mathbf{G}$ is equal to the smaller of $N$ and $M$, then $\mathbf{G}$ is said to be full rank.

*Matrices and linear transformations*

A **linear transformation** takes a vector from one vector space to another. In particular, we are concerned with the transformation of a vector from $R^M$ to $R^N$. The $N \times M$ matrix $\mathbf{G}$ defines such a transformation. The **null space** of $\mathbf{G}$ is the subspace of $R^M$ consisting of all solutions to

$$\mathbf{Gm} = \mathbf{0}. \tag{4}$$

The set of all vectors in $R^N$ that are the result of $\mathbf{Gm}$ for any $\mathbf{m}$ is called the **range space**. The **dimension theorem** is stated as follows: If $K$ is the rank of $\mathbf{G}$, then the null space of $\mathbf{G}$ has the dimension $M - K$.

*Inner products, norms, and orthogonality*

The inner product of vectors $\mathbf{x}$ and $\mathbf{y}$ is the scalar quantity denoted in general as $(\mathbf{x}, \mathbf{y})$. In $R^N$, a sensible inner product is

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} x_i y_i = \mathbf{x}^T \mathbf{y}. \tag{5}$$

Given an inner product, a norm can be defined as

$$||\mathbf{x}|| = (\mathbf{x}, \mathbf{x})^{1/2}, \tag{6}$$

which is interpreted as the length of the vector. Given the inner product (5), the norm would be

$$||\mathbf{x}|| = (\mathbf{x}^T \mathbf{x})^{1/2}, \tag{7}$$

which is recognized as the $L_2$ norm. A vector space which has an inner product defined for every pair of vectors is an **inner product space**. (Formally, an inner product space is a **Hilbert space** if the space is complete. If you want to explore this further, you can look up what complete means.)

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be orthogonal if

$$(\mathbf{x}, \mathbf{y}) = 0 \tag{8}$$

A set of vectors is orthogonal if every pair of vectors in the set is orthogonal.

If each vector in an orthogonal set satisfies

$$\mathbf{x} = 1, \tag{9}$$

then the set is said to be **orthonormal**. The notion of an inner product is important as it leads to expressions for the length of a vector, and the angle between vectors.

*Orthogonal matrices*

Consider the orthonormal basis

$$\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N\} \tag{10}$$

where

$$(\mathbf{p}_i, \mathbf{p}_j) = \delta_{ij}. \tag{11}$$

Now consider the matrix $\mathbf{P}$ whose columns are the vectors in the orthonormal basis (10):

$$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N]\,. \tag{12}$$

It follows that

$$\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}, \tag{13}$$

and

$$\mathbf{P}^{-1} = \mathbf{P}^T. \tag{14}$$

A matrix with the properties (13-14) is said to be an **orthogonal matrix**.

Some properties of orthogonal matrices:

1. Both the rows and columns of an orthogonal matrix form an orthonormal set.
2. If $\mathbf{P}$ is orthogonal then $|\text{det}\mathbf{P}| = 1$.
3. If $\mathbf{P}$ and $\mathbf{Q}$ are orthogonal, then so is $\mathbf{PQ}$.
4. If $\mathbf{P}$ is orthogonal, then for all vectors $\mathbf{x}$, $\mathbf{y}$, we have $(\mathbf{Px}, \mathbf{Py}) = (\mathbf{x}, \mathbf{y})$ and $||\mathbf{Px}|| = ||\mathbf{x}||$; interpreted geometrically this means that $\mathbf{P}$ preserves both angles and lengths when considered as a linear transformation.

Linear transformations by orthogonal matrices behave like rotations or reflections. There is no change in length of the transformed vector.

*Exercises:*

In class we considered a series of exercises to look at matrices. For each example, we asked, What is the rank of each of these matrices? Are the rows independent? Are the columns independent

1.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0.01 \end{bmatrix} \tag{15}$$

*This is a $3 \times 2$ matrix, so in principle it could be rank 2 at most. The second and third rows are not independent, but the columns are independent, which means that it is rank 2.*

2.
$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0.01 & 0 \end{bmatrix} \tag{16}$$

*This is a* $3\times3$ *matrix, so in principle could be rank 3. The rows and columns are independent, which means that it is rank 3.*

3.
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0.01 & 0.01 \end{bmatrix} \tag{17}$$

*This is a* $3 \times 3$ *matrix, so in principle could be rank 3. However the last 2 rows are not independent, and the last 2 columns are not independent, meaning that the matrix is rank 2.*

4.
$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{18}$$

*This is a* $3 \times 4$ *matrix, so in principle could be rank 3. However the last 2 rows are not independent, the first 2 columns are not independent, and the last 2 columns are not independent. The matrix is rank 2.*

We can test any of these using software function. For example, in Matlab

```
% The rank of the matrix A is
rank(A)

% If a square matrix is not full rank, it won't invert
% in which case the inverse will produce warning messages
inv(A)
```

**Eigensystems**

Now let's take our matrix fundamentals and think about eigensystems. An eigensystem is defined by the equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{19}$$

where $\mathbf{A}$ is a square matrix, $\mathbf{x}$ is a vector, and $\lambda$ is a scalar. In other words, the transformation $\mathbf{A}\mathbf{x}$ results in a simple scaling of $\mathbf{x}$. Given a **normal matrix A** (a class of matrix that includes symmetric and orthogonal matrices), we can always find a set of $\lambda$'s (eigenvalues) and a corresponding set of $\mathbf{x}$'s (eigenvectors).

The eigenvectors are equivalent to modes of physical systems. Consider the transverse oscillations of beads on a string (Figure 2). The two beads have mass $m$, and are separated by flexible strings of length $l$ when at equilibrium. Suppose displacements $x_n$ of the beads are so small that the tension $T$ in the strings can be taken to be constant. The angle of each string to the horizontal is $\theta_n$ as illustrated in the figure. The equation of motion for the displacement $x_1$ of the first bead is

$$m\frac{d^2x_1}{dt^2} = -T\sin\theta_1 + T\sin\theta_2. \tag{20}$$

Under the assumption that the displacements are small, $\sin\theta_n$ may be approximated as $\tan\theta_n$, so

$$m\frac{d^2x_1}{dt^2} = -T\frac{x_1}{l} + T\frac{x_2 - x_1}{l}. \tag{21}$$

Rearranging produces

$$m\frac{d^2x_1}{dt^2} = \frac{T}{l}\left(x_2 - 2x_1\right). \tag{22}$$

Similarly, the equation of motion for the second bead is

$$m\frac{d^2x_2}{dt^2} = \frac{T}{l}\left(x_1 - 2x_2\right). \tag{23}$$

Proceed by assuming a solution of the form

$$x_n = X_n e^{i\omega t} \tag{24}$$

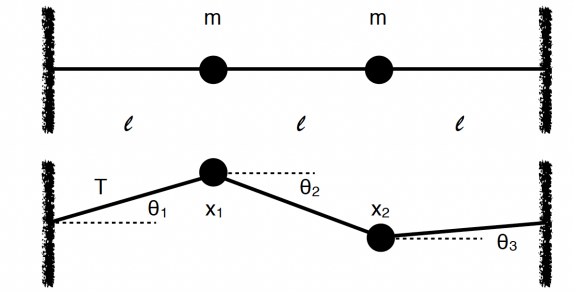


Figure 2: Beads on a string at equilibrium (top) and displaced (bottom).

Then the pair of equations to be solved are

$$(2-\lambda)X_1 - X_2 = 0 \tag{25}$$
$$-X_1 + (2-\lambda)X_2 = 0 \tag{26}$$

where $\lambda = \omega^2 ml/T$ are eigenvalues. For a nontrivial solution to this problem, we must have

$$\det\begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix} = 0 \tag{27}$$

So

$$(2-\lambda)^2 - 1 = 0 \tag{28}$$

which has the two solutions: $\lambda = 1, 3$. For $\lambda = 1$,

$$X_1 - X_2 = 0 \tag{29}$$

which says simply that $X_1 = X_2$. Writing this solution as a normalized vector

$$\mathbf{x} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{30}$$

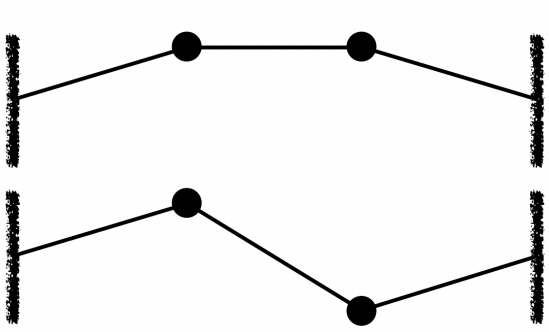

Figure 3: Modes of a two-beaded string. Mode 1 (top) has both beads moving in phase, and mode 2 (bottom) has the beads out of phase.

which is the eigenvector for the eigenvalue $\lambda = 1$. For $\lambda = 3$,

$$-X_1 - X_2 = 0 \tag{31}$$

with the normalized solution

$$\mathbf{x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{32}$$

The two solutions, commonly referred to as modes, are shown in Figure 3.

Given an initial specification of bead positions in terms of modes we can predict the evolution of the system. While equations of motion of the beads (22-23) are coupled, the equations for the modes are uncoupled. That is, the modes evolve independently of each other, and the evolution of the system is a linear combination of the two modes.

Note that the equation we solve was in the form (19) with

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \tag{33}$$

So what we were doing in solving the the problem of beads on a string, was precisely the solution of an eigensystem. In linear vector space language, the modes are the most convenient coordinates. The eigensystem can be written

$$\mathbf{AP} = \mathbf{PD} \tag{34}$$

where $\mathbf{D}$ is diagonal with the eigenvalues along the diagonal, and the columns of the orthogonal matrix $\mathbf{P}$ are the eigenvectors.

A few standard relations are

$$\mathbf{D} = \mathbf{P}^T\mathbf{AP} \tag{35}$$

$$\mathbf{A} = \mathbf{PDP}^T \tag{36}$$

$$\mathbf{A}^{-1} = \mathbf{PD}^{-1}\mathbf{P}^T \tag{37}$$

Given the eigenvalue decomposition of a matrix, (37) gives an easy way of determining invertibility simply by determining whether any of the eigenvalues are zero. The ratio of the smallest to largest eigenvalue, referred to as the condition number, is an indication of the stability of the inversion to numerical error.