Linear Algebra Overview

CSE 6040 – Computing for Data Analysis

Sep. 19, 2014

In this lecture, we will introduce linear algebra in the context of data analysis. We will cover vectors, matrices, and norms as the mathematical tools to process data. This lecture is partly based on the following courses:
http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/
http://vision.stanford.edu/teaching/cs131_fall1314/
http://cs229.stanford.edu/section/cs229-linalg.pdf
http://see.stanford.edu/see/courseInfo.aspx?coll=17005383-19c6-49ed-9497-2ba8bfcfe5f6

Contents

1 Vectors

A finite-dimensional vector is represented as an array of numbers (scalars), say $\begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$, in a three-dimensional space. More generally,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1}$$

is a vector in an n-dimensional space $\mathbb{R}^n$, where $x_1, \cdots, x_n$ are real numbers ($x_1, \cdots, x_n \in \mathbb{R}$). The n elements are called entries, or coordinates, of a vector. A vector can be used to describe data. For example, if 4 coordinates are needed to describe something, then we say we describe it in 4 dimensions, or in a 4-dimensional space, or 4-dimensional vector space; if 10 coordinates are needed to describe something, then we say we describe it in 10 dimensions.

If we found later that we don't really need 10 coordinates to describe the data, we can consider transforming the 10-dimensional vectors to, say, 2-dimensional vectors, called *dimension reduction*.

By default, all the vectors we consider are column vectors with notation x; for row vectors, we use the notation x^T (reads "x transpose").

Multiple vectors can be arranged side by side, creating a matrix. However, we do not emphasize the role of a matrix in describing data for this lecture. The utility that describes a data point is still a vector.

2 Matrices

A matrix is a rectangular array, for example,

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 4 & 0 \end{bmatrix} \tag{2}$$

is a 2×3 matrix. By $A \in \mathbb{R}^{m \times n}$, we denote a matrix with m rows and n columns where the entries of A are real numbers. We use A_{ij} to denote the entry of A at the i-th row and the j-th column. So the above matrix A can be written as:

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}. \tag{3}$$

We denote the the n columns of matrix $A \in \mathbb{R}^{m \times n}$ as $A(:,1), A(:,2), \cdots, A(:,n)$, and the m rows as $A(1,:), A(2,:), \cdots, A(m,:)$. A vector $x \in \mathbb{R}^n$ is often regarded as an $n \times 1$ matrix.

Scalar multiplication on a matrix αA produces a matrix of the same size, is defined as multiplying each entry of A by the scalar α:

$$(\alpha A)_{ij} = \alpha A_{ij} \tag{4}$$

The key to understanding a matrix is *matrix-vector multiplication*. We generalize this notion and introduce matrix-matrix multiplication first.

2.1 Matrix-matrix multiplication

The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ is defined as a matrix $C = AB \in \mathbb{R}^{m \times p}$, where

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}. \tag{5}$$

That is to say, to compute the (i,j)-th entry of C, we need to multiply the corresponding entries of $A(i,:)$ and $B(:,j)$, and add up these n scalar products. Therefore, matrix multiplication is only defined for A and B where the number of columns of A is the same as the number of rows of B.

Like the associative and distributive laws of scalar products, matrix multiplication has similar properties, which is easy but cumbersome to verify:

- Associative law: $(AB)C = A(BC)$;
- Distributive law: $A(B + C) = AB + AC, \quad (A + B)C = AC + BC$.

However, unlike scalar products, matrix multiplication does not have the commutative property in general, i.e. $AB \neq BA$.

2.1.1 Vector-vector multiplication

Vector can be seen as a special type of matrix, i.e. a matrix with only one column or only one row. One type of vector-vector multiplication is to multiply a row vector and a column vector, called inner product (or dot product). Based on the definition (5), the inner product of $x^T \in \mathbb{R}^{1 \times n}$ and $y \in \mathbb{R}^{n \times 1}$ can be written as:

$$x^T y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \tag{6}$$

Thus, viewing the matrix A as m row vectors and the matrix B as p column vectors, their product can be written in the form of inner products:

$$C = \begin{bmatrix} A(1,:)B(:,1) & A(1,:)B(:,2) & \cdots & A(1,:)B(:,p) \\ A(2,:)B(:,1) & A(2,:)B(:,2) & \cdots & A(2,:)B(:,p) \\ \vdots & \vdots & \ddots & \vdots \\ A(m,:)B(:,1) & A(m,:)B(:,2) & \cdots & A(m,:)B(:,p) \end{bmatrix}. \tag{7}$$

Theorem 1. *Let $\|v\|$ be the (Euclidean) length of a vector $v \in \mathbb{R}^n$: $\|v\| = \sqrt{v_1^2 + \cdots + v_n^2}$ (more precisely, 2-norm, to be defined later), then $x^T y = \|x\|\|y\| \cos(x, y)$.*

Remark 1. *Note that $\|x\|^2 = x_1^2 + \cdots + x_n^2 = x^T x$.*

Note that $cos(x, y) = 1$ when $x = y$, and $cos(x, y) = 0$ when $x \perp y$. Thus, when $\|x\| = \|y\|$, $x^T y$ can serve as the similarity between the vectors x and y (consider two vectors in a plane). When $x^T y = 0$, we say that the vectors x and y are *orthogonal*.

2.1.2 Matrix-vector multiplication

Matrix-vector multiplication is a special case of matrix-matrix multiplication where the second operand has only one column. Multiplying a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $x \in \mathbb{R}^n$ produces a vector $b \in \mathbb{R}^m$. Because matrix-vector multiplication is so important, we show an example before proceeding to further discussions. For the following A and x,

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 4 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} \tag{8}$$

their product b is

$$b = Ax = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + 0 \times 0 + (-2) \times (-2) \\ 3 \times 1 + 4 \times 0 + 0 \times (-2) \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} \tag{9}$$

The i-th entry of b is just the inner product of the i-th row of A, i.e. $A(i, :)$, and the vector x.

For matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, viewing B as p column vectors, the multiplication of A and B can be seen as multiplying A and p vectors separately. Thus matrix-matrix multiplication is just a bunch of matrix-vector multiplication.

2.2 Matrix-vector multiplication as a linear mapping

Often we want to transform the representation of a data point from one vector to another. For example, to visualize a data set represented in a high-dimensional space, we have to transform the data points into a two- or three-dimensional space; sometimes we want to rotate the data points in a vector space (when do we need this?). Because matrix-vector multiplication Ax for $A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n$ produces a vector $b \in \mathbb{R}^m$ as the result, we can view this operation as a mapping from the n-dimensional space to the m-dimensional space:

$$f : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Ax. \tag{10}$$

Let us look at some examples of such a mapping where $n = m$.

2.2.1 Diagonal matrix and identity matrix

Sometimes we want to scale each dimension, which is equivalent to multiplying a *diagonal matrix* D and a vector:

$$\begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} d_1 x_1 \\ d_2 x_2 \end{bmatrix}. \tag{11}$$

We often write $D = \mathrm{diag}(d_1, d_2)$. A special case appears when $d_1 = d_2 = d$ ($d \neq 0$):

$$\begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = d \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{12}$$

Here the direction of the vector is unchanged. Because

$$\begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} = d \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{13}$$

we have

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{14}$$

The matrix $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ where all the entries on the diagonal are 1, and 0 otherwise, is called an *identity matrix*. Applying an identity matrix on a vector returns the same vector.

2.2.2 Rotation matrix and orthogonal matrix

Sometimes we want to rotate a vector in a space while keeping its length unchanged. We can achieve this by multiplying a *rotation matrix* A and the vector:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \end{bmatrix}. \tag{15}$$

The above matrix A is parameterized by θ. This operation rotates the input vector $[x_1, x_2]^T$ counter-clockwise by an angle θ, yielding the result vector. We can show that the input and result vectors have the same length.

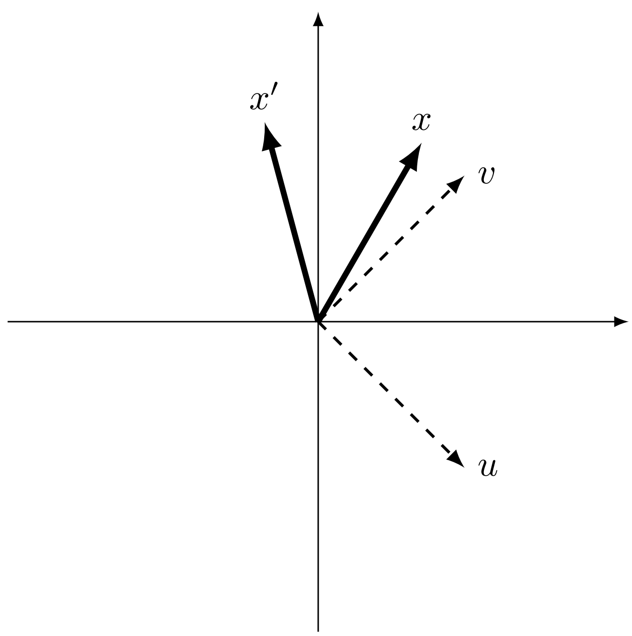

Figure 1: Illustration for rotation matrix with $\theta = 45°$.

Note that due to $\sin^2\theta + \cos^2\theta = 1$, the matrix A has the following properties:

1. All the columns have length 1.
2. Any two columns are orthogonal.
3. All the rows have length 1.
4. Any two rows are orthogonal.

More generally, a square matrix $Q \in \mathbb{R}^{n\times n}$ satisfying the above properties 1-2 is called an *orthogonal matrix* (properties 3-4 hold if properties 1-2 are satisfied).

To give the orthogonal matrix a definition in a symbolic way, we introduce the transpose of a matrix first. The *transpose* of a matrix $A \in \mathbb{R}^{m\times n}$, denoted as $A^T \in \mathbb{R}^{n\times m}$, is obtained by flipping the matrix A so that

$(A^T)_{ij} = A_{ji}$. The transpose of a vector is a special case of this definition. A handy and frequently-used (and easy-to-prove) property of the transpose operation is

$$(AB)^T = B^T A^T. \tag{16}$$

Then an orthogonal matrix is defined as a matrix Q satisfying $Q^T Q = QQ^T = I$ (recall I is the identity matrix). Check for yourself that $Q^T Q = I$ is equivalent to the above properties 1-2, and $QQ^T = I$ is equivalent to the above properties 3-4. The following is an important property for an orthogonal matrix:

Theorem 2. *An orthogonal matrix does not change the length of a vector, that is,* $\|Qx\| = \|x\|$.

Proof. $\|Qx\|^2 = (Qx)^T(Qx) = x^T Q^T Qx = x^T(Q^T Q)x = x^T x = \|x\|^2$. □

In other words, the effect of an orthogonal matrix on a vector is to rotate the vector in the n-dimensional space.

2.2.3 Another example: Difference matrix

Now we show another example that motivates our discussion in later sections. Consider a 3×3 matrix A,

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \tag{17}$$

When A applies on a vector $x = [x_1, x_2, x_3]^T \in \mathbb{R}^3$, we have

$$Ax = \begin{bmatrix} 1 \cdot x_1 + 0 \cdot x_2 & +0 \cdot x_3 \\ (-1) \cdot x_1 + 1 \cdot x_2 & +0 \cdot x_3 \\ 0 \cdot x_1 + (-1) \cdot x_2 & +1 \cdot x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}. \tag{18}$$

We can see that the effect of A is to take element-wise difference of x while keeping x_1 unchanged (imagine an implicit $x_0 = 0$). Unlike in the previous examples, here we do not emphasize the geometric interpretation of the transformation induced by A, but focus on its manipulation of the individual entries of x.

As a summary of Section 2.2, we generalize the mapping $f : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Ax$ that we have studied so far. As we have seen in the above examples, matrix-vector multiplication performs a transformation from a vector to another vector. Due to the distributive law of matrix multiplication, we easily have

$$\begin{aligned} A(x + y) &= Ax + Ay, \\ A(\alpha x) &= \alpha(Ax). \end{aligned} \tag{19}$$

More generally, any operator f that satisfies these two conditions,

$$\begin{aligned} f(x + y) &= f(x) + f(y), \\ f(\alpha x) &= \alpha f(x), \end{aligned} \tag{20}$$

is called a *linear mapping*, and it performs a *linear transformation*. An important fact is that the set of linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$ and the set of matrices in $\mathbb{R}^{n \times m}$ have a one-to-one correspondence. Therefore, a matrix can be viewed as being equivalent to a linear operator in finite-dimensional spaces.

2.3 Matrix-vector multiplication as a linear combination of matrix columns

Recall the difference matrix A in Section 2.2.3. We can rewrite the matrix-vector multiplication as the following:

$$b = Ax = \begin{bmatrix} 1 \cdot x_1 + 0 \cdot x_2 & +0 \cdot x_3 \\ (-1) \cdot x_1 + 1 \cdot x_2 & +0 \cdot x_3 \\ 0 \cdot x_1 + (-1) \cdot x_2 & +1 \cdot x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{21}$$

Note that x_1, x_2, x_3 are scalars, and it is more common to put them to the left of a vector/matrix to express scalar multiplication on a vector/matrix. This way, we can view matrix-vector product as **a linear combination of the columns of the matrix A, where the linear coefficients are the entries in the vector x**. In general, Ax can be written as

$$Ax = x_1 A(:,1) + x_2 A(:,2) + \cdots + x_n A(:,n) \tag{22}$$

This is perhaps the most fundamental property in the entire domain of linear algebra.

Imagine the linear mapping induced by a matrix $A \in \mathbb{R}^{3\times 3}$ as a blackbox, with an input vector $x \in \mathbb{R}^3$ and an output vector $b \in \mathbb{R}^3$. We have considered the case where x is known. What if the output b is known but we want to solve for the input x? The formulation $Ax = b$ where b is given is called a *linear system of equations*, or simply a *linear system*, which is a fundamental task in linear algebra. With the "linear combination" view, the goal of solving a linear system is

to find linear coefficients $x_1, \cdots, x_n$ of the expansion of b in terms of the columns of A.

Example 1. *Consider the polynomial fitting problem: We want to fit a sequence $b_1, \cdots, b_m$ using a polynomial of degree $\leq n-1$.*

2.3.1 Range and subspace

The concept of linear subspace is important to understand many data analysis problems. First, we define the range of a matrix in the "textbook" way: The *range* of a matrix $A \in \mathbb{R}^{m\times n}$ is the set of vectors that can be expressed as Ax for some x, i.e. $\text{range}(A) = \{Ax \mid x \in \mathbb{R}^n\} \subset \mathbb{R}^m$. But what is this thing?

Viewing matrix-vector multiplication as a linear combination of columns, range(A) is simply the set of all the vectors in $\mathbb{R}^m$ we can reach by linearly combining the columns of A.

For example, we denote the first two columns of A as S:

$$S = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}. \tag{23}$$

Then range(S) contains all the vectors that can be written as a linear combination of two vectors,

$$x_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \quad \forall x_1, x_2 \in \mathbb{R}. \tag{24}$$

By plotting the vectors constructed this way, we can see that range(S) is a 2-dimensional plane in the 3-dimensional space. Any vector (or data point) b in this plane can be represented as a pair of real numbers (x_1, x_2) that are used as linear coefficients to combine the two columns of S and produce b. We say that this plane is a *linear subspace*, or simply *subspace*, **spanned** by the two columns of S. The two columns of S are called a *basis* of the 2-dimensional subspace.

Suppose a data set is represented in the 3-dimensional space (each data point is a vector in $\mathbb{R}^3$), but we found that all the actual data points lie on a plane or close to a plane. In this case, a common task in data analysis is to find a basis (two linearly independent vectors, to be defined soon) of this plane and transform the original representation into vectors in $\mathbb{R}^2$, called *dimension reduction*.

2.3.2 Linear independence and the matrix inverse

Again, consider the difference matrix A in Section 2.2.3. Suppose we want to solve $Ax = b$ where $b = [0,0,0]^T$ is given, that is,

$$Ax = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{25}$$

We clearly see that the only solution is $x_1 = x_2 = x_3 = 0$, that is, $x = [0,0,0]^T$. A set of vectors $v^{(1)}, \cdots, v^{(n)} \in \mathbb{R}^m$ is said to be *linearly independent* if $x_1 v^{(1)} + \cdots + x_n v^{(n)} = \mathbf{0}$ implies $x_1 = \cdots = x_n = 0$, where $\mathbf{0} \in \mathbb{R}^m$ is the all-zero vector.

To understand the concept of linear independence, we plot the 3 column vectors of A, denoted as $a^{(1)}, a^{(2)}, a^{(3)}$, and we observe that none of these vectors lie in the plane spanned by the other two vectors. In other words, none of these vectors can be represented as a linear combination of the other two vectors. **Therefore,** $a^{(1)}, a^{(2)}, a^{(3)}$ **are linearly independent**. Otherwise, according to the definition, there must exist x_1, x_2, x_3 and at least one of them, say, $x_1 \neq 0$, such that $x_1 a^{(1)} + x_2 a^{(2)} + x_3 a^{(3)} = \mathbf{0}$. This means that

$$a^{(1)} = -\frac{x_2}{x_1} a^{(2)} - \frac{x_3}{x_1} a^{(3)}, \tag{26}$$

which is contradictory with the fact that $a^{(1)}$ cannot be represented as a linear combination of $a^{(2)}$ and $a^{(3)}$.

Now we observe from (25) that

$$\begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \Longleftrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_1 + b_2 \\ b_1 + b_2 + b_3 \end{bmatrix}, \tag{27}$$

which can be further written as

$$x = b_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + b_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{28}$$

We notice that we have expressed x as a linear combination of three vectors with entries in b as coefficients! Written in a matrix form, x becomes a matrix-vector product:

$$x = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} b \equiv Bb. \tag{29}$$

Note that the effect of the matrix B on the vector b is to compute cumulative sums of the entries in b. Thus we can name B as the *summation matrix*.

The difference matrix A and the summation matrix B perform opposite transformations. In particular, for any vector $x \in \mathbb{R}^3$, after A transforms x to the difference of its entries, resulting in Ax, B can transform Ax back to the original vector x, that is to say, $BAx = x$. Then we know the matrix BA is just the identity matrix. In general, if we have $BA = I$ for $A, B \in \mathbb{R}^{n \times n}$, B is called the *inverse matrix* of A denoted as A^{-1}, and A is also the inverse matrix of B.

Remark 2. *For an orthogonal matrix Q, its inverse Q^{-1} is equal to its transpose Q^T.*

Consider a slightly different matrix C, and its matrix-vector product:

$$C = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \quad Cx = \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}. \tag{30}$$

We can name C a *cyclic difference matrix*. What about its inverse?

We repeat the same idea that is used to analyze A. Suppose we want to solve $Cx = b$ where $b = [0,0,0]^T$ is given, that is,

$$\begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{31}$$

Here, not only is $[0,0,0]^T$ a solution, but also $[1,1,1]^T$, $[2,2,2]^T$, and so on. This means that $c^{(3)}$ can be represented as a linear combination of $c^{(1)}$ and $c^{(2)}$, where $c^{(1)}, c^{(2)}, c^{(3)}$ are the column vectors of C:

$$c^{(3)} = (-1) \cdot c^{(1)} + (-1) \cdot c^{(2)}. \tag{32}$$

You can verify this yourself. By definition, we know that $c^{(1)}, c^{(2)}, c^{(3)}$ are *not* linearly independent. By plotting these three vectors in $\mathbb{R}^3$, we can clearly see that they lie in the same plane.

Now we have a difficulty: When we try to solve $Cx = b$ for a general b, we must require $b_1 + b_2 + b_3 = 0$ to have a solution of x (see (31)). Otherwise, no solution of x exists. Indeed, considering the range of C, range(C) is a two-dimensional plane instead of the entire $\mathbb{R}^3$. This means that for a vector b not lying on that plane, we cannot find any linear combination of the columns of C that produces b.

This phenomenon – when $Cx = \mathbf{0}$ has more than one solution, the solution for $Cx = b$ does not necessarily exist for every b – is a general theorem in linear algebra.

The above discussion implies that there is no inverse matrix of C, that is, there is no such matrix D that $DC = I$. This conclusion is natural: When multiple vectors in $\mathbb{R}^3$ are mapped to the same vector $b \in$ range(C) by the linear mapping induced by C, we have no idea how to map the vector b back.

You may hear a lot about the *rank* of a matrix. Now it should not seem mysterious at all: The rank of a matrix is the number of dimensions of the linear subspace spanned by its columns, i.e. the range. The difference matrix A has rank 3 (full rank, invertible, nonsingular), while the cyclic difference matrix C has rank 2 (not invertible, singular).

3 Norms

In data analysis, we often need to quantify the distance between two vectors, for example, to compare the difference between actual home prices $b = [b_1, \cdots, b_n]^T$ and predicted home prices $\hat{b} = [\hat{b}_1, \cdots, \hat{b}_n]^T$ by some model. We need to determine if the error vector $[b_1 - \hat{b}_1, \cdots, b_n - \hat{b_n}]$ is "large" or "small".

Norm can be seen as a generalized notion of Euclidean length. A *norm* is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following three properties:

1. $\|x\| \geq 0$; and $\|x\| = 0$ if and only if $x = \mathbf{0}$.
2. $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).
3. $\|\alpha x\| = \alpha \|x\|$ $(\alpha \in \mathbb{R})$.

Several commonly used norms are:

- 1-norm: $\|x\|_1 = |x_1| + \cdots + |x_n|$
- 2-norm: $\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$ (we will refer to Euclidean length as "Euclidean norm" or 2-norm from now on)
- Infitity-norm: $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

The 2-norm of a vector can be natural extended to a norm of a matrix, called *Frobenius norm*:

- Frobenius norm: $\|A\|_F = \sqrt{\sum_{i,j=1}^n A_{ij}^2}$

The Frobenius norm can evaluate the difference between two matrices of the same size ($\|A - B\|_F$).