

GeoSci 236: A Linear Algebra Mini-Primer

Gidon Eshel

491 Hinds

Dept. of the Geophysical Sciences,

5734 S. Ellis Ave., The Univ. of Chicago,

Chicago, IL 60637

(773) 702-0440, geshel@midway.uchicago.edu

August 1, 2006

Geophysical data analysis almost universally boils down to linear algebraic techniques. Unfortunately, there is no way we can do justice to this extremely important and wonderful topic in these humble notes. We will just introduce a few absolutely essential ideas. If you would like to learn more (as you should!), there is clearly no better introduction to the subject than Strang, G. *Linear Algebra and its Applications* 3rd Ed., Harcourt Brace Jovanovich, San Diego (1988) ISBN 0-15-551005-3.

1 Matrices and vectors

Matrices are the most fundamental building block of linear algebra. They arise in many, highly diverse, situations, which we will get to later. A matrix is a rectangular array of numbers, e.g.,

$$\begin{pmatrix} 1 & 1 & -4 \\ 0 & 3 & 2 \\ 5 & 11 & 24 \\ -1 & 31 & 4 \end{pmatrix}.$$

A matrix is said to be $M \times N$ (M by N) when it comprises M rows and N columns. The typical notation reserves boldfaced uppercase letters to matrix variables (e.g., \mathbf{A}). It is sometime helpful to explicitly report the matrix size as a subscript when it is first introduced, $\mathbf{A}_{M \times N}$. A vector, typically denoted by boldfaced lowercase letters, is a special case of matrix, one for which either M or N are 1 (e.g., $\mathbf{x} = \mathbf{x}_{M \times 1}$, $\mathbf{y} = \mathbf{y}_{1 \times N}$). The former (\mathbf{x}) is column vector, while \mathbf{y} is a row vector. Unless otherwise stated, we will treat vectors as column vectors.

It is useful at this point to introduce some basic matrix operations. Addition (and subtraction) are simple; for $\mathbf{C} = \mathbf{A} + \mathbf{B}$ to be defined, \mathbf{A} and \mathbf{B} must have the same dimensions. Then, \mathbf{C} ‘inherits’ these dimensions, and

$$\mathbf{C}_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij}.$$

The *transpose* of a matrix is constructed by swapping columns and rows. That is, if $\mathbf{B} = \mathbf{A}^T$ (where the superscript T denotes the transpose of a real matrix), then $\mathbf{B}_{ij} = \mathbf{A}_{ji}$, where the first and second indices index rows and columns, respectively (i.e., element ij resides in row i and column j). Note that taking the transpose changes the matrix dimension; if \mathbf{A} is $M \times N$, \mathbf{A}^T is $N \times M$, and if \mathbf{x} is $M \times 1$, \mathbf{x}^T is $1 \times M$.

The *product* of 2 matrices is much more restrictive than that of 2 scalars. To see this, let us first introduce the *inner product* of 2 vectors, which yields a scalar. Given 2 column vectors \mathbf{x} and \mathbf{y} of the same length N ,

$$a = \mathbf{x}^T \mathbf{y} \equiv \sum_{i=1}^N x_i y_i.$$

It is immediately obvious that the inner product is only defined when both participating vectors are of the same length, the first a row vector, the second a column vector. An easy way to check if a given pair of vectors can be multiplied together is to write down their dimensions in the order of the multiplication; if the inner dimensions agree (are the same), then the multiplication ‘consumes’ the inner dimension, leaving the result a 1×1 matrix, better known as a scalar. For example, if we wish to consider \mathbf{xy} with $\mathbf{x}_{N \times 1}$ and $\mathbf{y}_{N \times 1}$, we write $(N \times 1) \times (N \times 1)$. Since the inner dimensions do not match, the operation is not permitted. Conversely, $\mathbf{x}^T \mathbf{y}$ is permitted, because $(1 \times N) \times (N \times 1) \rightarrow (1 \times N) \times (N \times 1) \rightarrow (1 \times 1)$, a scalar. Note that $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$, i.e., the inner product is commutative.

The product of 2 matrices \mathbf{A} and \mathbf{B} is given by

$$\mathbf{C} = \mathbf{AB}, \quad \mathbf{C}_{ij} = \mathbf{a}_i \mathbf{b}_j,$$

where \mathbf{a}_i is \mathbf{A} ’s i th *row*, and \mathbf{b}_j is \mathbf{B} ’s j th *column*. Thus $\mathbf{C}_{M \times N} = \mathbf{A}_{M \times K} \mathbf{B}_{K \times N}$, where K is some shared inner dimension which gets ‘consumed’ by the multiplication process. To visualize this, consider

$$\mathbf{A} = \begin{pmatrix} 1 & 4 \\ 2 & -3 \\ 3 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} -1 & 2 & 1 \\ -9 & 2 & -3 \end{pmatrix},$$

whose product $\mathbf{C}_{4 \times 3} = \mathbf{AB} =$

$$= \begin{pmatrix} (1 \ 4) \begin{pmatrix} -1 \\ -9 \end{pmatrix} & (1 \ 4) \begin{pmatrix} 2 \\ 2 \end{pmatrix} & (1 \ 4) \begin{pmatrix} 1 \\ -3 \end{pmatrix} \\ (2 \ -3) \begin{pmatrix} -1 \\ -9 \end{pmatrix} & (2 \ -3) \begin{pmatrix} 2 \\ 2 \end{pmatrix} & (2 \ -3) \begin{pmatrix} 1 \\ -3 \end{pmatrix} \\ (3 \ 1) \begin{pmatrix} -1 \\ -9 \end{pmatrix} & (3 \ 1) \begin{pmatrix} 2 \\ 2 \end{pmatrix} & (3 \ 1) \begin{pmatrix} 1 \\ -3 \end{pmatrix} \\ (-1 \ 0) \begin{pmatrix} -1 \\ -9 \end{pmatrix} & (-1 \ 0) \begin{pmatrix} 2 \\ 2 \end{pmatrix} & (-1 \ 0) \begin{pmatrix} 1 \\ -3 \end{pmatrix} \end{pmatrix} = \mathbf{C} = \begin{pmatrix} -37 & 10 & -11 \\ 25 & -2 & 11 \\ -12 & 8 & 0 \\ 1 & -2 & -1 \end{pmatrix}.$$

Note that in general the matrix product is *not* commutative; $\mathbf{AB} \neq \mathbf{BA}$. In the above example, \mathbf{BA} is not even permitted [it is $(2 \times 3) \times (2 \times 3)$ which does not have the required common inner dimension; $3 \neq 2$]. For this reason, we must explicitly distinguish post-multiplication from pre-multiplication, unless there is no ambiguity.

2 Vector Spaces

I want to be as concise and informal as possible, often at the expense of rigor, completeness and generality.

For our purposes, it is sufficient to think of a vector space as the set of all vectors of a certain kind. While the vectors need not in fact be actual vectors like the ones we have just discussed (they can also be functions, matrices, etc.), it will suffice here to consider ‘vectors’ as literally column vectors. An example will clarify this; \mathfrak{R}^1 (pronounced “r-one”) is the real line (we will only consider here real matrices, whose elements are all real numbers). In \mathfrak{R}^1 , one and only one kind of inhabitants can be found: 1-vectors whose single element is any one of the real numbers from $-\infty$ to ∞ . The numerical value of $v \in \mathfrak{R}^1$ (“ v which is an element from within \mathfrak{R}^1 ”) is the distance along the real line from the origin, (0) (it is not boldfaced here, because it is one-dimensional, or a scalar). \mathfrak{R}^0 is a vector space too, but it comprises a single point, (0); not too interesting. Vector spaces must have a few simple rules: they are commutative and have a zero vector \mathbf{o} satisfying $\mathbf{v} + \mathbf{o} = \mathbf{v}$, vectors and their negative counterparts satisfy $\mathbf{v} + (-\mathbf{v}) = \mathbf{o}$, and addition and multiplication are defined, yielding vectors still within the space.

The familiar plane is \mathfrak{R}^2 , where all 2-element vectors exist. Each of these 2-vectors describe a point in the plane. That point is one end of the vector, the other end being the origin, (0,0). Thus the 2 elements are the distance from the origin to the projections of the point on the 2 coordinates. Likewise, the familiar physical space is \mathfrak{R}^3 , where each vector $\mathbf{v} = (v_1 \ v_2 \ v_3)^T$ is stretched between the origin (0,0,0) and (v_1, v_2, v_3) .

Vector spaces must be *spanned*. By this we mean a particular (non-unique) choice of N vectors from \mathfrak{R}^N which we choose to express any other arbitrary \mathfrak{R}^N vector as a linear combination of. Once the choice of these vectors is made, the vectors are collectively referred-to as a ‘basis’ for \mathfrak{R}^N , and each one of them is a basis vector. For example, for spanning \mathfrak{R}^3 (‘the physical space’), the Cartesian basis is often chosen. This basis comprises the 3 vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ (sometimes denoted $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$; the ‘hat’ denotes a basis vector)

$$\hat{\mathbf{i}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{j}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{k}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Any \mathfrak{R}^3 vector can be expressed as a linear combination of these basis vectors;

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = v_1 \hat{\mathbf{i}} + v_2 \hat{\mathbf{j}} + v_3 \hat{\mathbf{k}}.$$

Note, again, that this is *not* a unique choice for spanning \mathfrak{R}^3 ; there are infinitely many such choices. The only constrain on the choice is that in order to fully span \mathfrak{R}^3 , the 3 vectors must be *linearly independent*, that is, that *no* non-trivial α , β and γ satisfying

$$\alpha \hat{\mathbf{i}} + \beta \hat{\mathbf{j}} + \gamma \hat{\mathbf{k}} = \mathbf{0}$$

can be found. Linearly dependent vectors are mutually parallel.

The requirement for mutual linear independence of the basis vector follows from the fact that a 3-vector has 3 independent pieces of information, v_1 , v_2 and v_3 . Given these 3 degrees of freedom (much more on that later in the course), we must have corresponding 3 basis vectors to work with. If one of the basis vectors is a linear combination of other ones, e.g., if $\hat{\mathbf{j}} = \alpha \hat{\mathbf{k}}$ say, then $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ no longer represent 2 directions in \mathfrak{R}^3 , but just 1. To show how this happens, consider the choice

$$\hat{\mathbf{i}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{j}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{k}} = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = 2\hat{\mathbf{i}} + 3\hat{\mathbf{j}}.$$

With this basis set, the vector $(0 \ 0 \ v_3)$ cannot be represented. Thus this choice of a basis doesn't fully span \mathfrak{R}^3 (it does span \mathfrak{R}^2 , though, just as well as $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ alone, but for 3 vectors to span \mathfrak{R}^2 is not very impressive). Note that this is *not* because none of our basis vectors has a nonzero 3rd element; try to come up with α , β and γ so that

$$\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \gamma \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}!!$$

You cannot find such coefficients [other than $(0, 0, 0)$] no matter how hard you try, because

$$\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

To better understand these issues, it is useful to visualize the geometry of the problem, which will be made easier with

$$\hat{\mathbf{i}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{j}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{k}} = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = 2\hat{\mathbf{i}} + 3\hat{\mathbf{j}}$$

as in the previous example. The first 2 basis vectors span \mathfrak{R}^2 , as was mentioned above. In order to now add a 3rd basis vector that will complete the spanning of \mathfrak{R}^3 , we need a vector, any vector, that rises above the plane $z = 0$. Since $(2 \ 3 \ 0)^T$ does *not* do that, it doesn't help us at all; it is fully contained within the plane already successfully spanned by the previous 2 basis vectors. The same happens with

$$\hat{\mathbf{i}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \hat{\mathbf{j}} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \hat{\mathbf{k}} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

only this situation is harder to visualize because the redundancy takes place in a plane that is not parallel to any one of the usual basis vectors $[(1 \ 0 \ 0)^T, (0 \ 1 \ 0)^T \text{ and } (0 \ 0 \ 1)^T]$, but in a plane inclined with respect to all of them. (Note, however, that we can transform the coordinates, as we will learn later, so that this redundant plane becomes a fixed value of one coordinate, which

we can then eliminate from the problem, thus reducing the apparent dimensionality, 3, to the actual dimensionality, 2.)

Now let's go back to the previous example [$\hat{\mathbf{i}} = (1 \ 0 \ 0)^T$, $\hat{\mathbf{j}} = (0 \ 1 \ 0)^T$], which is easier to picture. We have realized above that in order to assist in spanning \mathfrak{R}^3 , the additional basis vector must rise above the $z = 0$ plane. However, we are still left with a choice of exactly which vector among all those satisfying this requirement we are going to add; based on this criterion, we can equally well add $(1 \ 1 \ 1)^T$, $(0 \ 0 \ 1)^T$, $(1 \ 1 \ -4)^T$, etc., all rising to $|z| > 0$, as required. Given this indeterminacy, the choice is ours; any one of these vectors will do just fine. It is often useful to choose the basis vectors so that they are mutually *orthogonal*.

At the simplest, two vectors \mathbf{u} and \mathbf{v} are said to be *orthogonal* if $\mathbf{u}^T \mathbf{v} = \mathbf{v}^T \mathbf{u} = 0$; the information that one contains is entirely absent from the other, and vice versa. With $(1 \ 0 \ 0)^T$ and $(0 \ 1 \ 0)^T$ already chosen, the vector that is orthogonal to both is $(0 \ 0 \ \alpha)^T$, with arbitrary α . To show this, denote the sought vector by $\hat{\mathbf{k}} = (a \ b \ c)^T$. To be orthogonal to both $(1 \ 0 \ 0)^T$ and $(0 \ 1 \ 0)^T$, $\hat{\mathbf{k}}$ must satisfy

$$(1 \ 0 \ 0) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = (0 \ 1 \ 0) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0,$$

which can only hold if $a = b = 0$. When these conditions are met, *any* c will satisfy the orthogonality conditions. Hence the undetermined α .

The arbitrariness of α can be alleviated by the further requirement, which is customary but not essential, to normalize basis vectors to unit norm, making them *orthonormal*. The *norm* of a vector can be defined in a number of ways; the most commonly used is the so-called L_2 -norm, defined for a general column vector $\mathbf{v} \in \mathfrak{R}^N$ by

$$\|\mathbf{v}\| \equiv \sqrt{\mathbf{v}^T \mathbf{v}} = \sqrt{\sum_{i=1}^N v_i^2}.$$

Given this definition, the normalization of the basis vectors means that

$$\hat{\mathbf{i}}^T \hat{\mathbf{i}} = \hat{\mathbf{j}}^T \hat{\mathbf{j}} = \hat{\mathbf{k}}^T \hat{\mathbf{k}} = 1,$$

i.e.,

$$(0 \ 0 \ \alpha) \begin{pmatrix} 0 \\ 0 \\ \alpha \end{pmatrix} = 1,$$

$$\alpha^2 = 1 \rightarrow \alpha = 1.$$

3 Fundamental Spaces and the Rank of A Matrix

An attribute of matrices that is very important in multivariate statistics (as well as in most situations where matrices arise) is the matrix rank. The simplest definition of a rank is the number of

independent columns or rows; the rank q of an $M \times N$ matrix satisfies $q \leq \min(M, N)$. But this is not very revealing or satisfying. To better understand what this means and why it is important, let us introduce now a new view of matrices, that of a matrix as a *map*.

When an $M \times N$ matrix operates on (pre-multiplies) a vector (as in \mathbf{Ax} , where \mathbf{A} operates on \mathbf{x}) the vector must be a column vector of the same length as \mathbf{A} 's rows; $\mathbf{A}_{M \times N} \mathbf{x}_{N \times 1}$. That is, \mathbf{x} is a vector from \mathfrak{R}^N ; $\mathbf{x} \in \mathfrak{R}^N$. But in general $M \neq N$, which is most often the case with data matrices (for reasons we will explore later). Consequently, $\mathbf{b} = \mathbf{Ax} \notin \mathfrak{R}^N$; $\mathbf{b} \in \mathfrak{R}^M$. That is, the outcome vector \mathbf{b} is *not* from \mathfrak{R}^N , but rather from \mathfrak{R}^M . Thus we say that \mathbf{A} *maps* \mathfrak{R}^N vectors to \mathfrak{R}^M ones. For example, if \mathbf{A} is 41×12 and \mathbf{x} is 12×1 , \mathbf{b} will be 41×1 ; $\mathbf{x} \in \mathfrak{R}^{12} \xrightarrow{\mathbf{A}} \mathbf{b} \in \mathfrak{R}^{41}$. Thus a matrix is associated with 2 fundamental spaces, that of its rows (the row-space, where vectors it operates on come from), and that of its columns (the column-space, where vectors that result from its operation on row-space vectors come from). That the resultant vectors \mathbf{b} are from \mathbf{A} 's column-space can be seen by noting that

$$\begin{aligned} \mathbf{Ax} &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ & \vdots & \vdots & \\ a_{M-1\ 1} & a_{M-1\ 2} & \dots & a_{M-1\ N} \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_N \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = \\ &= x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_N \mathbf{a}_N = \sum_{i=1}^N x_i \mathbf{a}_i = \mathbf{b}. \end{aligned}$$

This demonstrates visually the fact that the resulting vector \mathbf{b} is a linear combination of \mathbf{A} 's columns, i.e., \mathbf{b} is a vector from \mathbf{A} 's column-space. Note also that the weighting factors are none other than \mathbf{x} 's elements, x_i , $i = 1, 2, \dots, N$. The row-space of \mathbf{A} is best viewed as simply the column-space of \mathbf{A}^T , following trivially from the definition of the transpose.

Now comes a slightly tricky twist. We have established above that the column- and row-spaces of $\mathbf{A}_{M \times N}$ are associated with \mathfrak{R}^M and \mathfrak{R}^N , respectively. However, the pairs need not coincide! That is, the column-space may simply *be* \mathfrak{R}^M , but it doesn't *have* to be; it can be a *subspace* of \mathfrak{R}^M . Similarly, the row-space may be either the full \mathfrak{R}^N , or a subspace thereof. Let's clarify this with an example. Consider an \mathbf{A} given by

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix}.$$

Clearly this \mathbf{A} 's column-space is associated with \mathfrak{R}^4 ; but *is* it \mathfrak{R}^4 ?! Right off the bat we notice that \mathbf{A} has only 3 columns. Thus *at the most*, \mathbf{A} 's column-space is 3-dimensional (because 3 vectors can span at most a 3-dimensional space, even if they are 4-vectors). Thus we already know that \mathbf{A} 's column-space cannot possibly *be* \mathfrak{R}^4 ; it can, at most, be a 3-dimensional subspace of \mathfrak{R}^4 . This demonstrates the importance of the minimum operator we used before to describe q , \mathbf{A} 's rank [$q \leq \min(M, N)$].

Next, we want to find out if \mathbf{A} 's column-space is a 3-dimensional subspace of \mathfrak{R}^4 , or an even lower-dimensional subspace thereof. Recall that for \mathbf{A} 's column-space to be \mathfrak{R}^3 , \mathbf{A} 's 3 columns must be linearly independent. This is the case when no non-trivial α , β and γ can be found that satisfy

$$\alpha \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + \gamma \begin{pmatrix} 4 \\ 3 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}.$$

We need to solve the above system. One way to proceed is *Gaussian elimination*. We operate on \mathbf{A} with a sequence of so-called *elementary operations* (adding to various rows multiples of rows above them), to reduce \mathbf{A} 's elements below the main diagonal (the ij positions) to zero. We start by adding to rows 2-4 a copy of the 1st row, so that their left-most elements will vanish. With

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{E}_1 \mathbf{A} = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & -1 \\ 0 & 2 & -2 \\ 0 & 3 & -3 \end{pmatrix}.$$

Next, we turn to the 3rd row; With

$$\mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & -\frac{1}{3} \end{pmatrix}, \quad \mathbf{E}_2 \mathbf{E}_1 \mathbf{A} = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We have achieved our objective (all sub-diagonal elements are zero), but with a twist; both bottom rows became all zeros. This is an indication that \mathbf{A} 's rank is 2, not 3; $q_{\mathbf{A}} = 2$. In retrospect, we should have expected the *last* row to vanish; as we said before, because \mathbf{A} has only 3 columns, its rank can be *at most* 3. And since there are 4 rows, the last one must vanish. The fact that the second-to-last row vanished too is a consequence (or a demonstration, if you will) of the fact that $q_{\mathbf{A}} = 2$. And this, in turn, is true because \mathbf{A} 's first column is one fifth the sum of the other 2 columns;

$$\frac{1}{5} \left[\begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \right] = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

In other words, with

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} -5 \\ 1 \\ 1 \end{pmatrix},$$

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

which is the previously emphasized condition for \mathbf{A} 's columns being linearly dependent.

A matrix whose $q < \min(M, N)$ is said to be *rank-deficient*; if $q = \min(M, N)$, the matrix is *full-rank*. The meaning of these terms becomes clearer when \mathbf{A} is considered in the context of a set of coupled linear algebraic equations, like the set we have just solved [for $(\alpha \ \beta \ \gamma)^T$]. Then, the system $\mathbf{A}_{M \times N} \mathbf{x}_{N \times 1} = \mathbf{b}_{M \times 1}$ is a system of M equations with N unknowns. The question is whether there are enough equations (constraints) to fully determine \mathbf{x} (a related, and equally important, question is whether or not the system is *consistent*; we will get to it later). Let's consider again the same \mathbf{A} , but introduce a rhs (right hand side) vector. To emphasize the equation set, I will write it out explicitly;

$$\begin{aligned} \alpha + \beta + 4\gamma &= 8 \\ \alpha + 2\beta + 3\gamma &= 0 \\ \alpha + 3\beta + 2\gamma &= 1 \\ \alpha + 4\beta + \gamma &= 0 \end{aligned} \implies \begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ 6 \\ 5 \end{pmatrix} \equiv \mathbf{b}.$$

Recall the 2 elementary matrices that we have obtained above to reduce \mathbf{A} to an upper-triangular form (i.e., to make all sub-main-diagonal elements vanish); we called them \mathbf{E}_1 and \mathbf{E}_2 , and their product (in the appropriate order) turned out to be

$$\mathbf{E}_2 \mathbf{E}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ -\frac{1}{3} & 1 & 0 & -\frac{1}{3} \end{pmatrix}$$

and

$$\mathbf{E}_2 \mathbf{E}_1 \mathbf{A} = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

To conserve the conditions of the equations, we must operate on the rhs vector in the same way;

$$\mathbf{E}_2 \mathbf{E}_1 \mathbf{b} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ -\frac{1}{3} & 1 & 0 & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} 8 \\ 7 \\ 6 \\ 5 \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

What we have done here is to transform the original problem, $\mathbf{A}\mathbf{x} = \mathbf{b}$ to an equivalent, but readily solvable, problem, $\mathbf{U}\mathbf{x} = \mathbf{d}$, simply by premultiplying both sides by $\mathbf{E}_2 \mathbf{E}_1$,

$$\mathbf{E}_2 \mathbf{E}_1 \mathbf{A} \mathbf{x} = \mathbf{U} \mathbf{x} = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \mathbf{E}_2 \mathbf{E}_1 \mathbf{b} = \begin{pmatrix} 8 \\ -1 \\ 0 \\ 0 \end{pmatrix} \equiv \mathbf{d}.$$

Because \mathbf{A} is rank-deficient, the bottom 2 equations do not tell us anything. The 2nd equation says that $\beta - \gamma = -1$, or $\beta = \gamma - 1$. The 1st equations, $\alpha + \beta + 4\gamma = 8$, can be combined with $\beta = \gamma - 1$, yielding

$$\alpha = 9 - 5\gamma.$$

That is, *any*

$$\mathbf{x} = \begin{pmatrix} 9 - 5\gamma \\ \gamma - 1 \\ \gamma \end{pmatrix}$$

satisfies the system $\mathbf{A}\mathbf{x} = \mathbf{x}$. In other words, we can solve the equation up to a point; the value of γ is totally undetermined, and any choice will satisfy the constraints (equations) perfectly (and equally) well; \mathbf{A} is rank-deficient. It is convenient to split \mathbf{x} ,

$$\mathbf{x} = \begin{pmatrix} 9 \\ -1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} -5 \\ 1 \\ 1 \end{pmatrix} = \mathbf{x}_p + \mathbf{x}_h,$$

the *particular* and *homogeneous* parts of the solution.

The first vector, \mathbf{x}_p , solves the equations;

$$\begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 9 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 9 - 1 \\ 9 - 2 \\ 9 - 3 \\ 9 - 4 \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ 6 \\ 5 \end{pmatrix},$$

as required.

However, that's not the end of the story. The vector

$$\hat{\mathbf{n}} = \gamma \begin{pmatrix} -5 \\ 1 \\ 1 \end{pmatrix}$$

spans the null-space of \mathbf{A} 's column-space. That is, any vector parallel to $\hat{\mathbf{n}}$ (any choice of γ) is mapped by \mathbf{A} to $(0, 0, 0, 0)^T$, so we can add it to the solution without violating the constraints. For example, let's choose $\gamma = 1$, in which case

$$\begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} -5 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5 + 1 + 4 \\ -5 + 2 + 3 \\ -5 + 3 + 2 \\ -5 + 4 + 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Now let's contrast the above situation with

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 2 & 3 \\ -1 & 3 & 2 \\ 0 & 4 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \\ 3 \\ 5 \end{pmatrix}.$$

Now, with

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & -2 & 0 & 1 \end{pmatrix},$$

$$\mathbf{E}_2\mathbf{E}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & -2 & 0 & 1 \end{pmatrix},$$

$$\mathbf{E}_2\mathbf{E}_1\mathbf{A} = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & -5 \end{pmatrix} \quad \text{and} \quad \mathbf{E}_2\mathbf{E}_1\mathbf{b} = \begin{pmatrix} 7 \\ 5 \\ 0 \\ -5 \end{pmatrix}.$$

That is,

$$\mathbf{U}\mathbf{x} = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & -5 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \\ 0 \\ -5 \end{pmatrix} \equiv \mathbf{d}.$$

We now employ back-substitution; The 4th equation indicates that $\gamma = 1$. The 2nd then becomes $2\beta + 3\gamma = 5$, or $\beta = 1$. Finally, the 1st equation yields $\alpha = 2$, and

$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

Now \mathbf{x} is completely and fully determined, with no lingering ambiguity or indeterminacy. This is because *this* \mathbf{A} is full rank, $q = 3 = \min(4, 3)$. In turn, this is so because \mathbf{A} 's left-most column is independent of the other 2, and thus \mathbf{A} 's columns are all mutually independent, yielding full-rank.

The number of independent columns (or rows) of an arbitrary \mathbf{A} is a very fundamental question, one that has a number of ways to answer (all of which, of course, will lead to the same conclusion!) It is also a question that is somewhat ambiguous in practice. To see why, recall that linearly dependent vectors are parallel. In simple examples, vectors are either mutually parallel or not; there is no in-between. Noisy data matrices, on the other hand, can have columns that are almost parallel, but not quite. Then, the question is a more subtle one. The *conditioning number* of a matrix (which we will define later) is meant to answer the question, but in a somewhat subjective way. We need to know more about matrices in order to fully appreciate this dilemma.

So for now, we conclude this brief discussion by noting that the rank of an $M \times N$ matrix is the number of independent rows or columns; if $q = \min(M, N)$, \mathbf{A} is full-rank. Otherwise [$q < \min(M, N)$], \mathbf{A} is rank-deficient. Rank-deficiency implies the presence of a nontrivial null-space in the appropriate space [the smaller of (M, N)]. If $M \neq N$, the larger of (M, N) must have a non-trivial null-space, because \mathbf{A} 's highest possible rank is $\min(M, N)$.

With the rank q defined, and the distinction between rank-deficient and full-rank matrices established, we are now ready to introduce the all-important *inverse* of a square, full-rank, matrix (to be clear: only full-rank square matrices have an inverse in the sense we use here). The inverse is denoted by \mathbf{A}^{-1} , and

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

where \mathbf{I}_N , the $N \times N$ identity matrix, is the N -dimensional generalization of 1;

$$\mathbf{I}_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad \text{or}$$

$$\mathbf{I}_N = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

It leaves intact vectors it operates on, changing neither their direction nor their magnitude.