1. Lectures 1, 2: Basic concepts of linear algebra

1.1. Introductory remarks

A matrix is as much a bunch of numbers in a table
as a human is a bunch of molecules in a bag.

The list of the least useful things in linear algebra:

- ◇ Determinant
- ◇ Cramer's rule
- ◇ Calculating the matrix inverse

The list of the most useful things in linear algebra:

- ◇ Eigenvalues and eigenvectors
- ◇ Fredholm alternative
- ◇ Gaussian elimination
- ◇ SVD (Singular Value Decomposition)

Some basic examples

We illustrate the need to learn the language of linear algebra with several practical examples coming from science and engineering.

- ◇ A matrix is a linear function that maps vectors to vectors: $Ax = y$. Here A is $m \times n$, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$.
- ◇ A matrix rotates and stretches vectors.

Example 1. Eigenvalues and eigenvectors.

A square matrix has special vectors for which it acts as a number: $As = \lambda s$. These vectors and numbers are called eigenvectors and eigenvalues and are very important properties of a matrix. Recall, how we find them.

Since $As - \lambda s = (A - \lambda I)\, s = 0$ is a homogeneous system, it has nontrivial solutions only if $\det(A - \lambda I) = 0$. This gives an n-th order algebraic equation to find λ, and in general there will be n complex roots of this equation. For each root, one finds eigenvectors s.

For example, suppose

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Then

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{vmatrix} = \begin{vmatrix} 0 & -1-(2-\lambda)^2 & 2-\lambda \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{vmatrix} =$$

$$\left[(\lambda-2)\left(1+(\lambda-2)^2\right) - \lambda + 2\right] = (\lambda-2)\left[(\lambda-2)^2 - 2\right] = 0.$$

Therefore $\lambda_1 = 2$ and $\lambda_{2,3} = 2 \pm \sqrt{2}$. For each one, there is an eigenvector. For example, for $\lambda_1 = 2$, we get

$$(A - 2I)\, s_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} s_1 = 0.$$

Therefore $s_1 = [1\ 0\ -1]^T$. The other eigenvectors are also not difficult to find: $s_{2,3} = \left[1\ \mp\sqrt{2}\ 1\right]^T$.

Example 2. Image analysis.

Now we take an image, SVD it and look at approximate images. Example of Einstein image svd'ed in Matlab.

```
%% SVD of Einstein. A. Kasimov. Skoltech. MM-EAS F-2019
% press Cmd-Enter to run the script

clear all; close all; clf

m = 312; n = 223;  % image size in pixels
AE = imresize(double(rgb2gray(imread('einstein.tif'))),[m n]); %
    load and resize image of size mxn

subplot(1,2,1), pcolor(flipud(AE)), shading interp, colormap(gray),
     axis equal
set(gca,'Xtick',[],'Ytick',[])
title('Full rank approximation of Einstein');
%pause;

[U S V] = svd(AE);
for k = 1:2:31  % rank-k approximation of Einstein
     AEk = U(:,1:k)*S(1:k,1:k)*V(:,1:k)';
     subplot(1,2,1), pcolor(flipud(AEk)), shading interp, colormap(
        gray), axis equal
     title(['Rank ',int2str(k),' approximation of Einstein']);
     pause(0.5);  % pause for 0.5 seconds.
end

% plot the singular values
subplot(1,2,2), loglog(diag(S),'.','MarkerSize',20), axis on, grid
    on
title 'Singular values'
ylabel '\sigma'
xlabel 'n'
```

The third example comes from analysis of data.

Example 3. Least squares approximation.

Suppose we have a lot of data of the form (x_i, y_i) that could come from experiment, simulation, or elsewhere. We expect that these data depend on each other in some unknown form, $y = f(x)$, and want to approximate this dependence somehow. For example, by a straight line, $y = ax + d$. We may be lucky that all the data fall on the line, in which case $y_i = ax_i + d$ for all i. Typically

this will not be the case. The question is: How do we choose the line (i.e., a and d) so that it approximates the data best? The answer depends of course on the measure of error. If we choose to minimize the square error $\sum_i (y_i - ax_i - d)^2$, then we will solve the least-squares problem. It is done as follows. Write down the "wish" equation:

$$\begin{aligned} ax_1 + d &= y_1 \\ ax_2 + d &= y_2 \\ &\dots \\ ax_n + d &= y_n, \end{aligned}$$

which is asking for too much – to determine two parameters a and b to satisfy n equations, where n is typically very large. This system in general has no solution. However, we can find an approximation that gives the least square error between the line and the data. It is solved as follows (we will explain why later in the course).

Form a matrix A and a vector b:

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{bmatrix}, \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}.$$

Then the wish equation can be written as

$$Au = b$$

where $u = [a\ d]^T$, and its least-squares solution is found by forming a normal equation

$$A^T A u = A^T b.$$

This equation has a unique solution if the columns of A are linearly independent (why is this true?). And the solution is

$$u = \begin{bmatrix} a \\ d \end{bmatrix} = \left(A^T A\right)^{-1} A^T b.$$

Note $A^T A$ is just a 2×2 matrix.

Example 4. Matrices from approximations of derivatives.

Consider a boundary value problem of the form (1D Poisson equation)

$$\begin{aligned} u'' &= f(x), \quad 0 < x < 1, \\ u(0) &= a, u(1) = b. \end{aligned}$$

We want to solve it on a grid, $x = ih$, where $h = 1/N$ is the grid size, and $i = 1, 2, \dots, N$ is index of the grid point. Then

$$\begin{aligned} u' &\approx \frac{u_{i+1} - u_i}{h} \approx \frac{u_i - u_{i-1}}{h} \\ u'' &\approx \frac{u'_{i+1} - u'_i}{h} \approx \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} \end{aligned}$$

and the differential equation can be approximated as

$$-u_{i+1} + 2u_i - u_{i-1} = -h^2 f(x_i), \quad i = 1, \dots, N$$

with properly taken care boundary conditions. Note that this system can be written in matrix form as $Ax = b$ with matrix A given by

$$A = \begin{bmatrix} 2 & -1 & 0 & & 0 \\ -1 & 2 & -1 & 0 & \\ 0 & -1 & 2 & -1 & 0 \\ & & & & -1 \\ 0 & & 0 & -1 & 2 \end{bmatrix}.$$

This is second difference matrix, has many nice properties that come from its symmetry. We will learn more about such matrices later.

1.2. **Main concepts**

- x is a column vector in $\mathbb{R}^n$, A is an $m \times n$ matrix (m rows and n columns) in $\mathbb{R}^{m\times n}$
- $b = Ax$ is a column vector in $\mathbb{R}^m$ defined as
$$b_i = \sum_{j=1}^{n} a_{ij}x_j, \quad i = 1, \ldots, m,$$
where a_{ij} is the element of A in i-th row and j-th column.
- A is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$: it takes all vectors $x \in \mathbb{R}^n$ and produces some vectors $b \in \mathbb{R}^m$, maybe all in $\mathbb{R}^m$, maybe not, depending on properties of A.
- Linear means:
$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay$$
for any vectors x, y from $\mathbb{R}^n$ and scalars α, β.
- Ax is a combination of columns of A:
$$b = Ax = \sum_{j=1}^{n} a_j x_j$$
where a_j is the j-th column of A.
- Example:
$$\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + 5 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 13 \\ 6 \end{bmatrix}.$$
- $Ax = b$ can be interpreted as a matrix A acting on vector x to produce vector b, but also as x acting on columns of A to produce b.
- A matrix A times a matrix B, $AB = C$, is a new matrix C obtained by multiplying matrix A with columns of B. That is, if
$$B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \ldots & b_n \\ | & | & & | \end{bmatrix},$$
then
$$C = \begin{bmatrix} | & | & & | \\ c_1 & c_2 & \ldots & c_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \ldots & Ab_n \\ | & | & & | \end{bmatrix}$$

◇ Example. If A is just a column vector u in $\mathbb{R}^m$ and B is a row vector $v \in \mathbb{R}^n$, then we obtain the outer product

$$uv^T = \begin{bmatrix} \\ u \\ \\ \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} = \begin{bmatrix} | & | & & | \\ v_1 u & v_2 u & \dots & v_n u \\ | & | & & | \end{bmatrix} = \begin{bmatrix} - & vu_1 & - \\ \dots & \dots & \dots \\ - & vu_n & - \end{bmatrix}$$

which is a rank-1 matrix in $\mathbb{R}^{m\times n}$, as each column is a multiple of vector u, and each row is a multiple of vector v. When such a matrix acts on a vector $x \in \mathbb{R}^n$, we obtain

$$\left(uv^T\right) x = u\left(v^T x\right) = \alpha u, \quad \alpha\text{–scalar},$$

a vector in $\mathbb{R}^m$ in the direction of u. That is uv^T projects any vector $x \in \mathbb{R}^n$ to a single direction along u in $\mathbb{R}^m$.

◇ Example:

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 4 & 5 \end{bmatrix} = \begin{bmatrix} 4\cdot 2 & 5\cdot 2 \\ 4\cdot 1 & 5\cdot 1 \end{bmatrix} = \begin{bmatrix} 2\cdot 4 & 2\cdot 5 \\ 1\cdot 4 & 1\cdot 5 \end{bmatrix} = \begin{bmatrix} 8 & 10 \\ 4 & 5 \end{bmatrix}.$$

◇ Example – very important. We can multiply two matrices A and B in three different ways:

- 1) the usual way of forming a scalar product of a row_i of A with a col_j of B to get the element ij, c_{ij}, of $C = AB$;
- 2) multiplying A with each column of B to form a column of C; and
- 3) by forming a combination of rows of B with coefficients from A. The second method uses a combination of columns of A to calculate each column of C, while the third uses the combination of rows of B. Each one is useful in its own ways.

$$\underbrace{\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 4 & 3 \\ 5 & 1 \end{bmatrix}}_{B} = \begin{bmatrix} 2\cdot 4+1\cdot 5 & 3\cdot 2+1\cdot 1 \\ -1\cdot 4+2\cdot 5 & -1\cdot 3+2\cdot 1 \end{bmatrix} = \begin{bmatrix} 13 & 7 \\ 6 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \underbrace{\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \end{bmatrix}}_{Ab_1}, & \underbrace{\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix}}_{Ab_2} \end{bmatrix} = \begin{bmatrix} 13 & 7 \\ 6 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2\cdot row_1(B) + 1\cdot row_2(B) \\ -1\cdot row_1(B) + 2\cdot row_2(B) \end{bmatrix} = \begin{bmatrix} 2\begin{bmatrix} 4 & 3 \end{bmatrix} + 1\begin{bmatrix} 5 & 1 \end{bmatrix} \\ -1\begin{bmatrix} 4 & 3 \end{bmatrix} + 2\begin{bmatrix} 5 & 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 13 & 7 \\ 6 & -1 \end{bmatrix}.$$

Therefore, two ways of seeing $AB = C$:

- Every column of $C = AB$ is a combination of columns of A.
- Every row of $C = AB$ is a combination of rows of B.

◇ Permutation matrix P is a matrix such that PA is the same as A, but with rows of A exchanged. P itself is a permutation of rows of the identity matrix. For example

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

permutes rows 1 and 2 of A: $PA = P \begin{bmatrix} & a_1 & \\ \dots & a_2 & \dots \\ & \dots & \end{bmatrix} = \begin{bmatrix} & a_2 & \\ \dots & a_1 & \dots \\ & \dots & \end{bmatrix}$, because $row_1(PA) = 0 \cdot row_1(A) + 1 \cdot row_2(A) + 0 \cdot row_3(A) + ... = row_2(A)$ and $row_2(PA) = 1 \cdot row_1(A) + 0 \cdot row_2(A) + 0 \cdot row_3(A) + ... = row_1(A)$.

Note that multiplying from the right permutes columns of A. This can be directly verified, but is also seen from $AP = \left(P^T A^T\right)^T$, and $P^T A^T = PA^T$ is a permutation of columns of A.

- ◇ The range of a matrix A is the set of all vectors that can be formed by acting with A on all vectors x. As we have seen, Ax is a combination of columns of A, therefore the range of A is all the possible combinations of columns of A – the column space of A, denoted as $C(A)$ or $range(A)$. The range of A is spanned by the columns of A.
- ◇ The nullspace of A is the set of all vectors x such that $Ax = 0$.
- ◇ The left nullspace is the nullspace of A^T, and the row space of A is the space spanned by the rows of A, in other words it is a column space of A^T.
- ◇ Rank. The column rank of A is the dimension of its column space, the row rank is the dimension of the row space.

Theorem 5. *Row rank = column rank*

Proof. Skip for now □

Theorem 6. *A matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ (tall matrix) has full rank iff it maps no two distinct vectors to the same vector.*

Proof. Skip for now □

- ◇ **Four fundamental subspaces of a matrix**. The column space $C(A)$, the row space $R(A)$, the nullspace $N(A)$, and the left nullspace $N\left(A^T\right)$ form the four fundamental subspaces of A.

  It is important to note the following facts:
  - $Ax_0 = 0$ for $x_0 \in N(A)$ implies that all x_0 are perpendicular to rows of A. Therefore

$$N(A) \perp R(A). \tag{1.1}$$

  - By similar reasoning

$$N\left(A^T\right) \perp C(A). \tag{1.2}$$

  - If the column space is spanned by r vectors, i.e. $dim(C(A)) = rank(A) = r$, then the left nullspace is spanned by the remaining $m - r$ vectors of $\mathbb{R}^m$, where the columns of A live. That is

$$dim(C(A)) + dim\left(N\left(A^T\right)\right) = m.$$

  - Similarly,

$$dim(R(A)) + dim(N(A)) = n.$$

- ◇ The inverse of a square matrix A, $m \times m$, is a matrix A^{-1} such that $A^{-1}A = I$, the identity matrix. A square matrix of full rank has a unique inverse, it is a non-singular matrix.

Theorem 7. *For $A \in \mathbb{R}^m$, the following statements are equivalent:*

(1) A *has an inverse* A^{-1}
(2) $rank\,(A) = m$
(3) $range\,(A) = \mathbb{R}^m$
(4) $N\,(A) = \{0\}$
(5) 0 *is not an eigenvalue of* A
(6) 0 *is not a singular value of* A
(7) $\det(A) \neq 0$

1.3. **Norms of vectors and matrices.**

◇ Norms are measures of size. The measure of a vector is its length. What is a measure of a matrix? A reasonable answer to this question is: measure the maximum possible length a vector Ax can have for all x of unit length. That is, the size of A measures how much it can stretch a vector,

$$\|A\| = \max_{x \in \mathbb{R}^n} \frac{(Ax)^T Ax}{x^T x} = \max_{x:\, x^T x = 1} (Ax)^T Ax$$

◇ Other norms besides the usual Euclidian norm can be useful. The general definition of a norm is:
- A norm is a function $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ such that for any vectors x and y and scalars α, the following conditions are satisfied:
  * $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$
  * $\|x + y\| \leq \|x\| + \|y\|$
  * $\|\alpha x\| = |\alpha|\,\|x\|$.

◇ Useful examples of norms:
- $\|x\|_1 = \sum_{i=1}^m |x_i|$
- $\|x\|_2 = \left(\sum_{i=1}^m |x_i|^2\right)^{1/2} = \sqrt{x^T x}$
- $\|x\|_\infty = \max\limits_{1 \leq i \leq m} |x_i|$
- $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$, $1 \leq p \leq \infty$ Why is $p \geq 1$

◇ Geometric interpretations of these norms.

◇ Matrix norms induced by the above vector norms simply follow from the definition given above

$$\|A\|_{(m,n)} = \sup_{x \in \mathbb{R}^n, \|x\| = 1} \|Ax\|_{(m)},$$

where we are measuring the size of an $m \times n$ matrix A, such that $x \in \mathbb{R}^n$ while $Ax \in \mathbb{R}^m$.

◇ Use *eigshow*(A) in Matlab to see how A acts on a unit vector. Note the alignment of x and Ax. Do this with $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$.

1.4. **Matrix factorizations.** ..... Essential matrix factorizations include:

- $A = LU$ – LU-factorization into a product of lower triangular and upper triangular matrices of the form

$$L = \begin{bmatrix} 1 & & & \\ l_{21} & 1 & & 0 \\ l_{31} & l_{32} & \ddots & \\ & & & \\ l_{n1} & \cdots & & 1 \end{bmatrix}, \quad U = \begin{bmatrix} u_{11} & u_{12} & \cdots & \\ & u_{22} & u_{23} & \cdots \\ & & \ddots & \\ & 0 & & \\ & \cdots & & u_{nn} \end{bmatrix}$$

- $A = S\Lambda S^{-1}$ – the eigenvalue decomposition, columns of S are eigenvectors s_i and $\Lambda = diag\,(\lambda_i)$ is a diagonal matrix with eigenvalues on the diagonal, $As_i = \lambda_i s_i$.
- $A = U\Sigma V^T$ – singular value decomposition
- $A = QR$ – orthogonal Q times upper triangular R.

1.5. Gaussian elimination and $A = LU$

Example 8. Gaussian elimination. Suppose we need to solve a linear system

$$x_1 + x_2 + x_3 = 1 \tag{1.3}$$

$$x_1 + 2x_2 + 4x_3 = -1 \tag{1.4}$$

$$x_1 + 3x_2 + 9x_3 = 1. \tag{1.5}$$

It can be written as $Ax = b$ with

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

The Gaussian elimination proceeds by subtracting first equation from the second and third to get

$$x_1 + x_2 + x_3 = 1 \tag{1.6}$$

$$x_2 + 3x_3 = -2 \tag{1.7}$$

$$2x_2 + 8x_3 = 0. \tag{1.8}$$

Then we subtract twice the second from the third to get

$$x_1 + x_2 + x_3 = 1 \tag{1.9}$$

$$x_2 + 3x_3 = -2 \tag{1.10}$$

$$2x_3 = 4. \tag{1.11}$$

Now we move from bottom up. First $x_3 = 2$, then $x_2 = -2 - 3x_3 = -8$, finally $x_1 = 1 - x_2 - x_3 = 7$. Problem solved.

Things to note:

(1) The final system can be written as $Ux = b_1$, where U is upper triangular:

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

(2) The elimination process is a linear combination of rows of the augmented matrix $[A\,b]$. We know this can be accomplished by a matrix multiplication. What is that elimination matrix?

To find the elimination matrix, note that the subtraction from row 2 of row 1 is accomplished by the matrix

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which leaves rows 1 and 3 of A unchanged, while row 2 of A is replaced by row 2 - row 1. Similarly, multiplication with

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

replaces row 3 with its difference with row 1.

Then we also subtracted 2*row 2 from row 3, and that is accomplished by

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}.$$

Therefore, the three steps of elimination can be accomplished by multiplying the augmented matrix with the product $E_3E_2E_1$. However, notice the following useful fact: E_2E_1 subtracts row 1 from the row 3 of E_1 to yield

$$E_2E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

Now $E_3\,(E_2E_1)$ subtracts 2*row 2 of this matrix from row 3:

$$E_3E_2E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}.$$

Therefore, the Gaussian elimination produces a system

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix},$$

which is the same as before, as it should be.

Notice now that the elimination matrix is lower triangular with ones on the main diagonal. Furthermore, moving the elimination matrix to the right, we find that $A = LU$, where

$$L = (E_3E_2E_1)^{-1} = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

Therefore, we can write the factorization $A = LU$ as

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

With such a factorization available, the solution of $Ax = b$ proceeds in two steps. Let $LUx = b$ and denote Ux as y. Then:

First solve

$$Ly = b \quad \leftarrow \text{forward elimination to find } y.$$

Then, with y known, solve for x

$$Ux = y \quad \leftarrow \text{back substitution to find } x.$$

This example can be generalized to any system that has a unique solution. Note however, that sometimes one may have to multiply the matrix A by a permutation matrix P, when a zero pivot appears, in order to move equations with non-zero pivots up.