Linear Algebra
Week 5

G-07

24 X 2024

1 LU and LUP Decomposition

We are going to look at two more decompositions of a matrix, the LU and LUP decompositions, in this week. These decompositions are closely related to Gauss elimination and particularly come in handy when we want to solve linear systems of equations for multiple vectors $\mathbf{b}$ in $A\mathbf{x} = \mathbf{b}$, because the decomposition of the matrix A takes only $O(m^3)$ time but once we have the decomposition we can solve for any $\mathbf{b}$ in $O(m^2)$ time (when a solution exists).

A way to link Gauss elimination with LU or LUP decomposition is to remember that the matrices L and P define the relation between our initial matrix A and the upper triangular matrix U we get from Gauss elimination. So the decomposition is just a way to formalize Gauss elimination.

1.1 LU Decomposition

LU decomposition is the mild case of LUP decomposition. If we are lucky and do not have to perform any *row exchanges* as we perform Gauss elimination, then an LU decomposition of the matrix A exists. Such a Gauss elimination would look like:

$$\underbrace{\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -c_{32} & 1\end{bmatrix}}_{\substack{\text{subtract } c_{32}\cdot\text{(row 2)}\\ \text{from (row 3)}}}\underbrace{\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -c_{31} & 0 & 1\end{bmatrix}}_{\substack{\text{subtract } c_{31}\cdot\text{(row 1)}\\ \text{from (row 3)}}}\underbrace{\begin{bmatrix}1 & 0 & 0\\ -c_{21} & 1 & 0\\ 0 & 0 & 1\end{bmatrix}}_{\substack{\text{subtract } c_{21}\cdot\text{(row 1)}\\ \text{from (row 2)}}} A = U.$$

But we want a compact representation of the eliminations. So let's multiply the elimination matrices to represent all eliminations in one matrix. Remember, we can do this because the matrix multiplication is associative.

$$\begin{bmatrix}1 & 0 & 0\\ -c_{21} & 1 & 0\\ c_{32}c_{21} - c_{31} & -c_{32} & 1\end{bmatrix} A = U.$$

This matrix seems ugly. And we are trying to get an equation in form of $A = LU$ so it is on the wrong side of the equation. We can multiply both sides with the inverse, which is a beautiful lower triangular matrix:

$$\begin{bmatrix}1 & 0 & 0\\ -c_{21} & 1 & 0\\ c_{32}c_{21} - c_{31} & -c_{32} & 1\end{bmatrix}^{-1} = \begin{bmatrix}1 & 0 & 0\\ c_{21} & 1 & 0\\ c_{31} & c_{32} & 1\end{bmatrix}$$

In the end we are left with:

$$A = \underbrace{\begin{bmatrix}1 & 0 & 0\\ c_{21} & 1 & 0\\ c_{31} & c_{32} & 1\end{bmatrix}}_{L} U.$$

So we have our matrix L!

Before we proceed with the formal theorem that also defines the matrix L for us, we can read some features of L from our little experiment with 3×3 matrices. These features are valid for L of any size and also in LUP decomposition:

- L is lower triangular. (Hence the name L)
- L has 1's on the diagonal.
- The matrix L represents the row subtractions we perform during the Gauss elimination.
- In particular, the element L_{ij} represents the multiple that we have used for the row subtraction $row_j - c_{ij} \times row_j$ throughout the elimination steps. (The indices are considered after permutation in case of LUP)

Now we can take a look at the LU decomposition in a formal way as defined in the lecture notes:

Theorem 3.13 (LU decomposition). *Let A be an $m \times m$ matrix on which Gauss elimination as in Table 3.6 succeeds without row exchanges, resulting in an upper triangular matrix U. Let c_{ij} (computed in Line 18) be the multiple of row j that we subtract from row $i > j$ when we eliminate in column j. Then $A = LU$ where*

$$L = \begin{bmatrix} 1 & & & \\ c_{21} & 1 & & \\ \vdots & & \ddots & \\ c_{m1} & \cdots & c_{m,m-1} & 1 \end{bmatrix}.$$

More formally, $L = [\ell_{ij}]_{i=1,j=1}^{m\ \ m}$ where

$$\ell_{ij} = \begin{cases} 0 & \text{if } i < j \\ 1 & \text{if } i = j \\ c_{ij} & \text{if } i > j \end{cases}.$$

I usually do not include the long proofs in the weekly notes but the proof of this theorem offers you a good opportunity to understand the LU decomposition. Even though you can theoretically do LU decompositions in practice by memorizing the theorem above without its proof, I recommend taking a look at it.

On Proof of Theorem 3.13

Let's think of a matrix, which is in the middle of Gauss elimination. We are done with the first $i-1$ rows. This means we have obtained the first $i-1$ rows of the matrix U already using row operations. Now we want to subtract a multiple of row $i-1$ from the row i to get a 0 below the diagonal on row i. The picture makes everything more clear:

$$
\begin{array}{r|cccccc|l}
 & u_{11} & \cdots & & & & & \leftarrow \text{finalized (in } U) \\
 & 0 & u_{22} & \cdots & & & & \leftarrow \text{finalized (in } U) \\
 & 0 & 0 & \ddots & & & & \vdots \\
\text{row } j & 0 & 0 & \cdots & \mathbf{u_{jj}} & \cdots & u_{jm} & \leftarrow \text{finalized (in } U) \\
\vdots & & & & & & & \\
\text{row } i & 0 & 0 & \cdots & \star_{ij} & \cdots & \star_{im} & \leftarrow \text{now subtract } c_{ij} \cdot (\text{row } j)
\end{array}
$$

The first $i-1$ rows are finalized in U, what do we do to get the row i in U? Well first of all, we subtracted a multiple of each row before $i-1$ to get a zero at the corresponding column of the row i. Now we also have to subtract a multiple of the row $i-1$ to get a 0 below the diagonal on row i. This can be summarized as:

$$
\begin{array}{cccll}
 & & & (\text{row } i) \text{ in } A & \text{initially} \\
- & c_{i1} & \cdot & (\text{row } 1) \text{ in } U & \text{step } 1 \\
- & c_{i2} & \cdot & (\text{row } 2) \text{ in } U & \text{step } 2 \\
\vdots & & & & \\
- & c_{i,i-1} & \cdot & (\text{row } i-1) \text{ in } U & \text{step } i-1 \\
= & & & (\text{row } i) \text{ in } U & \text{finalized.}
\end{array}
$$

If you want to have the i-th row of A using U, then you solve this system for the i-th row of A. You subtract a bunch of things from the i-th row of A to get the i-th row of U, so if you want to calculate the i-th row of A using the i-th row of U, you add those terms to the both sides of the equation. If we solve this equation for the first term, (row i) in A, we see that this row is a linear combination of the first i rows of U:

$$
(\text{row } i) \text{ of } A = \underbrace{\begin{bmatrix} c_{i1} & c_{i2} & \cdots & c_{i,i-1} & 1 & 0 & \cdots & 0 \end{bmatrix}}_{\text{row vector}} U.
$$

Doing this operation for each row of A gives us our matrix L:

$$
A = \underbrace{\begin{bmatrix} 1 & & & \\ c_{21} & 1 & & \\ \vdots & & \ddots & \\ c_{m1} & \cdots & c_{m,m-1} & 1 \end{bmatrix}}_{L} U.
$$

□

Now that we have our matrix decomposition $A = LU$ we can use this to solve linear systems of equations. See how to do this in the next section with $PA = LU$ decomposition.
In the beginning of this section, we assumed that we do not need any row exchanges in our matrix while performing Gauss elimination. This assumption is necessary to be able to decompose A as LU without the help of a permutation matrix P. Here is an example from lecture notes:

Question 5.1. Consider the following matrix A for an arbitrary number a. Can we decompose it into two matrices L and U as described above? If not why?

$$\underbrace{\begin{bmatrix} 0 & 1 \\ 1 & a \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}}_{U}.$$

1.2 Permutations

We need permutation matrices to make $PA = LU$ decomposition. But what exactly are permutations and permutation matrices? Here I will provide the definitions and two important lemmas out of the lecture notes:

Definition 3.14 (Permutation). A permutation of $[m] = 1, 2, \ldots, m$ is a bijective function $\pi : [m] \mapsto [m]$.

Definition 3.15 (Permutation matrix). *Let $\pi : [m] \to [m]$ be a permutation. The* permutation matrix *associated with π is the $m \times m$ matrix $P = [p_{ij}]_{i=1,j=1}^{m \ \ m}$ with*

$$p_{ij} = \begin{cases} 1 & \text{if } j = \pi(i) \\ 0 & \text{otherwise} \end{cases} .$$

Vice versa, we also say that π is associated with P.

There are two important properties of permutation matrices that you should know about, in order to be able to define and understand the $PA = LU$ decomposition:

Lemma 3.16. Let P be a permutation matrix. Then $P^{-1} = P^\top$.

Lemma 3.17. Let P, P' be $m \times m$ permutation matrices with associated permutations π, π'. Then PP' is a permutation matrix as well, associated with the permutation $\pi' \circ \pi$.

You can either convince yourself that these facts actually hold true by considering some examples, or you can take a look at the proofs in the lecture notes, which is as always recommended.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ x_5 \\ x_4 \\ x_1 \end{bmatrix}$$

A permutation matrix and its effect on a vector

1.3 LU with Permutations: $PA = LU$

Now we can proceed with the general case of the $A = LU$ decomposition: $PA = LU$, or LUP decomposition:

Theorem 3.18 (LUP decomposition). *Let A be an $m \times m$ matrix with linearly independent columns, $m \geq 1$. There exist three $m \times m$ matrices P, L, U such that*

$$PA = LU,$$

where P is a permutation matrix, L a lower triangular matrix with 1's on the diagonal, and U an upper triangular matrix with nonzero diagonal entries.

The existence of such matrices $P, L,$ and U relies mainly on one fact: We perform row exchanges and row subtractions during Gauss elimination. Afterwards we summarize these row exchanges in the permutation matrix P, and the row subtractions in the matrix L. However, we usually perform these operations in a mixed order, as in the example:

$$U = E_{m-1}P_{m-1}E_{m-2}P_{m-2}\dots E_1P_1A$$

But it turns out that we can move all permutations (row exchanges) to the beginning, and then go on with the Gauss elimination. We can write the above equation as:

$$U = E'_{m-1}E'_{m-2}\dots E'_1P_{m-1}P_{m-2}\dots P_1A$$

Then we multiply the matrices together to get $PA = LU$.

The proof of this fact is rather long, but if you want to see a good use case of proofs by induction over the size of the matrices, then you can take a look at the proof in lecture notes.

1.4 *How to solve $Ax = b$ using PALU?*

We multiply with $P^{-1} = P^{\top}$ (see Lemma 3.17 for this formula) to obtain $A = P^{\top}LU$, hence we can write $A\mathbf{x} = \mathbf{b}$ as

$$P^{\top}\underbrace{L\,\underbrace{U\mathbf{x}}_{\mathbf{y}}}_{\mathbf{z}} = \mathbf{b}.$$

We first solve $P^{\top}\mathbf{z} = \mathbf{b}$ for $\mathbf{z}$. In fact, there's nothing to solve: $\mathbf{z} = P\mathbf{b}$, so we permute $\mathbf{b}$ in the same way we have permuted the rows of A. After this, we proceed as before: we solve $L\mathbf{y} = \mathbf{z}$ for $\mathbf{y}$ using forward substitution and then solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, with back substitution. All this can be done in $O(m^2)$ time, so again much faster than solving $A\mathbf{x} = \mathbf{b}$ from scratch using Gauss elimination.

For the case where we do not have a permutation matrix, we skip the step with $P^{\top}\mathbf{z} = \mathbf{b}$

2 Gauss-Jordan Elimination

Our version of Gauss elimination can fail and do not take us to a solution for all matrices. Now we want to have a form, in which you can transform any matrix.

2.1 RREF

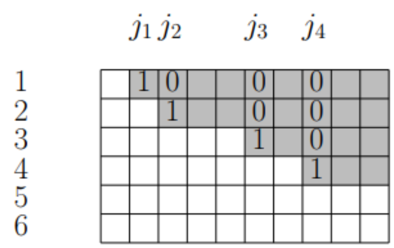
Let's begin with the definition of the reduced row echelon form:

Definition 3.19 (Row echelon and reduced row echelon form). Let $R = [r_{ij}]_{i=1,j=1}^{m,n}$ be an m×n matrix. R is in row echelon form (REF) if the following holds: There exist $r \leq m$ column indices $1 \leq j_1 < j_2 < \cdots < j_r \leq n$ such that the following two statements hold:

(i) For $i = 1, 2, \ldots, r$, we have $r_{ij_i} = 1$.

(ii) For all i, j, we have $r_{ij} = 0$ whenever $i > r$ or $j < j_i$ or $j = j_k$for some $k > i$.

If $r = m$, R is in reduced row echelon form (RREF). If we want to describe the shape of R precisely, we say that R is in $REF(j_1, j_2, \ldots, j_r)$ or $RREF(j_1, j_2, \ldots, j_m)$.



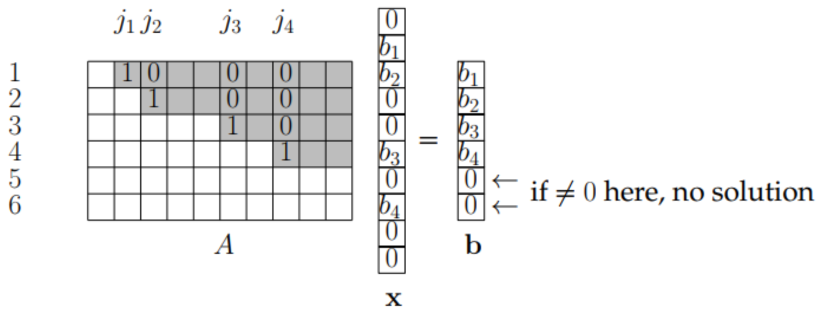
Example matrix in REF from the lecture notes

A matrix in REF or RREF let's us read important features of a matrix just by looking at the matrix. One of them is:

Observation 3.20. A matrix R in $REF(j_1, j_2, \ldots, j_r)$ has rank r.

2.2 Direct Solution

When a matrix A is in REF, solving $A\mathbf{x} = \mathbf{b}$ is easy! Even easier than back substitution with an upper triangular matrix:



Even though this image explains a lot if you think about it, here is a mathematical formalization of the possible cases:

Suppose that A is an $m \times n$ matrix in $\mathrm{REF}(j_1, j_2, \ldots, j_r)$ There are two cases: if $b_i \neq 0$ for some $i > r$, there is no solution. This is because the i-th row of A is zero for all $i > r$, so the i-th entry of $A\mathbf{x}$ is zero for $i > r$, no matter what $\mathbf{x}$ is.

If $b_i = 0$ for all $i > r$, there is a solution: we define $\mathbf{x}$ by

$$x_j = \begin{cases} b_i, & \text{if } j = j_i \\ 0, & \text{otherwise.} \end{cases}$$

Note that the given solution $\mathbf{x}$ here is not always unique. This is the 'easiest' solution, called *the canonical solution*. There are systems where you have infinitely many solutions. In the demonstration above, we have a matrix with more columns than rows, *i.e.* more variables than equations. This system has indeed infinitely many solutions, if the entries that are not supposed to be 0 -marked in the image- are not 0. Why can those entries of $\mathbf{b}$ have to be 0 for us to have a solution? Because the corresponding rows of A are 0 rows and we can not expect to multiply 0 a bunch of zeros with a set of coefficients (in this case the entries of the vector $\mathbf{x}$) and expect to get some nonzero result.
For the sake of completeness, here is the above image in pure mathematical language. Note that the j_i-th column of A is $\mathbf{e}_i$ the standard unit vector.

$$A\mathbf{x} = \sum_{i=1}^{r} b_i \mathbf{e}_i = \sum_{i=1}^{m} b_i \mathbf{e}_i = \mathbf{b}.$$

2.3 Elimination

The best way to understand how we do Gauss-Jordan elimination is to see one in practice. There is an example in the lecture notes that goes through each step of the elimination. After reading that example and eliminating one or two matrices by yourself you should have a better understanding of how to perform the elimination. In the following there is a brief overview of the row operations and the steps of a Gauss-Jordan elimination.

Row Operations

- row subtraction
- row exchange
- multiplying a row by a scalar (NEW!)

For a particular row i (the first i-1 rows are finalized)

- If the pivot -first nonzero element- on row i is not zero
  - Divide the row by the pivot (or multiply by $\frac{1}{pivot}$)
  - Now that our pivot is equal to 1, eliminate the entries **above** (NEW!) **and below** our pivot!
  - We have the standard unit vector $\mathbf{e}_i$ on the corresponding column of our matrix, we are ready to move to the next row.
- Else: If the pivot is 0
  - continue with the next column the row (NEW!) (it means we cannot make a downward step at this column)

The (NEW!) points are the differences compared to the Gauss-Elimination.

There are a few theorems lemmas on Gauss Jordan elimination in the lecture notes, that might come in handy in proofs.

Theorem 3.21 (Gauss-Jordan elimination). Let A be an $m \times n$ matrix. There exists an invertible $m \times m$ matrix M such that $R_0 = MA$ is in REF.

Lemma 3.22. Let A be an $m \times n$ matrix, M an invertible $m \times m$ matrix, and $R_0 = MA$ in $REF(j_1, j_2, \dots, j_r)$. Then A has independent columns $j_1, j_2, \dots, j_r$.

Good intuition from the lecture notes:

> It is worthwhile to understand what Gauss-Jordan elimination does on an invertible $m \times m$ matrix A. In this case, all columns are independent, so by the previous Lemma, the resulting $m \times m$ matrix R_0 is in $\mathrm{REF}(1, 2, \dots, m)$, and hence equal to the identity matrix I. So we have $R_0 = I = MA$, and this means that the matrix M in Theorem 3.21 is actually the inverse of A.

2.4 Runtime

It is always good practice to know the runtime of the algorithm you are working with. You can find further runtime analysis in the lecture notes.

Theorem 3.23. *Let A be an $m \times n$ matrix of rank r, and let $\mathbf{b} \in \mathbb{R}^m$.*

(i) Using Gauss-Jordan elimination, A can be transformed into $R_0 = MA$ in REF as given by Theorem 3.21 in time $O(rmn + mn)$.

(ii) By simultaneously transforming the $m \times m$ identity matrix using the same row operations, $M = MI$ can be computed in additional time $O(rm^2 + m^2)$.

(iii) Given M, the system $A\mathbf{x} = \mathbf{b}$ can be solved in time $O(m^2)$.

2.5 Computing the CR Decomposition

Remember the CR decomposition from two weeks ago? Now it's back! It turns out that out matrix R is the same in both the CR decomposition and REF of a matrix. There is a nice proof for this rather unexpected fact in the lecture notes, here is the theorem:

Theorem 3.24. *Let A be an $m \times n$ matrix and let $A = CR$ as in Theorem 2.23. Let $R_0 = MA$ in $\mathrm{REF}(j_1, j_2, \dots, j_r)$ be the result of Gauss-Jordan elimination on A, see Theorem 3.21.*

Then R results from R_0 by removing the zero rows at the end (if there are any); in particular, R is in $\mathrm{RREF}(j_1, j_2, \dots, j_r)$, and C is the submatrix of A with columns $j_1, j_2, \dots, j_r$.

This theorem in particular implies that the result R_0 of the Gauss-Jordan elimination is unique!

3 Hints

1. Solved In Class

2. Solved In Class

3. **a)** Similar with last week's hand in, check the solution of it! **b)** you should use unit vectors $\mathbf{e}_i$ to represent the pattern from the first subtask in general. Then just check if A times your inverse gives you I.

4. As the hint in the assignment sheet suggests, use CR decomposition. You now know how to calculate the matrix R.

5. No hints.

6. Think of w_1, w_2, w_3 and v_1, v_2, v_3 as the columns of two matrices W and V. What is the relation between W and V? What can you say about the matrix W?

7. You do not have to prove the statements from Exercise 6 on Assignment 4 again, use them as the question suggests. Use your findings from previous subtasks for some subtasks.

mkilic

♠