Elimination and row operations

COMS 3251 Fall 2022 (Daniel Hsu)

1 Systems of linear equations

Solving systems of linear equations is a ubiquitous problem in many areas. A system of m linear equations in n unknowns looks like the following:

$$\begin{cases} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n = b_1 \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n = b_2 \\ \vdots \\ A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n = b_m. \end{cases} \tag{1}$$

The $A_{i,j}$'s are the coefficients (stand-ins for some scalars), the b_i's are the right-hand side values (also stand-ins for more scalars), and $x_1, \ldots, x_n$ are the *unknowns* (a.k.a. *variables*). If the right-hand side values are all 0, then we say the system is *homogeneous*.

"Solving" a system means finding values for the variables $x_1, \ldots, x_n$ so that all m equations hold, or declaring that no such values exist. A system could have multiple solutions; the *solution set* is the set of all solutions. If the solution set is empty, the system is *inconsistent*. Otherwise, it is *consistent*.

Example. A system of 3 linear equations in the 4 unknowns (x_1, x_2, x_3, x_4):

$$\begin{cases} x_1 + 2x_2 + 3x_3 + 4x_4 = 3 \\ 2x_1 + 4x_2 + 4x_3 + 4x_4 = 2 \\ 3x_1 + 6x_2 + 5x_3 + 4x_4 = 1. \end{cases}$$

Asking whether this system of linear equations is consistent is the same as asking whether $(3, 2, 1)$ is in $\operatorname{span}(\{(1,2,3), (2,4,6), (3,4,5), (4,4,4)\})$. ∎

The system (1) can be equivalently written in *matrix form*:

$$\underbrace{\begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}}_{\mathbf{b}},$$

i.e.,

$$A\mathbf{x} = \mathbf{b}.$$

Here, $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ is the *coefficient (left-hand side) matrix*, $\mathbf{b} \in \mathbb{R}^m$ is the *right-hand side vector*, and $\mathbf{x} = (x_1, \dots, x_n)$ is the *vector of unknowns*. Each row of A, along with the corresponding right-hand side value in $\mathbf{b}$, specifies a linear equation in the system.

Note that if $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, so the m-vector $\mathbf{a}_j$ contains the coefficients for variable x_j in the m linear equations, then asking whether $A\mathbf{x} = \mathbf{b}$ is consistent is the same as asking whether $\mathbf{b}$ is in $\mathsf{CS}(A)$. And a solution $\mathbf{x} = (x_1, \dots, x_n)$ (assuming one exists) specifies how to linearly combine $\mathbf{a}_1, \dots, \mathbf{a}_n$ to produce $\mathbf{b}$:

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

One last common way of writing a system of linear equations is to omit the vector of unknowns all together, and to put A and $\mathbf{b}$ together in a single matrix, called an *augmented matrix* $[A \mid \mathbf{b}]$:

$$\left[\begin{array}{cccc|c} A_{1,1} & A_{1,2} & \cdots & A_{1,n} & b_1 \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} & b_m \end{array}\right].$$

A vertical line is used to separate the coefficient matrix from the right-hand side vector.

Continuing the previous example. The system in matrix form:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

As a linear combination of 3-vectors:

$$x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + x_4 \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

As an augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 2 & 4 & 4 & 4 & 2 \\ 3 & 6 & 5 & 4 & 1 \end{array}\right].$$

2 Elimination

2.1 Basic idea of Elimination

Elimination is an ancient algorithm for determining if a system of linear equations is consistent, and also for finding a solution when one exists.

The basic idea of Elimination is to transform the system into another one that is somewhat easier to solve:

$$A\mathbf{x} = \mathbf{b} \quad \longrightarrow \quad R\mathbf{x} = \mathbf{c}.$$

This transformation is called a *reduction* from the problem of solving $A\mathbf{x} = \mathbf{b}$ to the problem of solving $R\mathbf{x} = \mathbf{c}$. The reduction is solution-preserving: the new system has the same solution set as the original system. In particular, it will be the case that $R\mathbf{x} = \mathbf{c}$ is consistent if and only if $A\mathbf{x} = \mathbf{b}$ is consistent.

2.2 Elementary row operations

The steps of Elimination are usually described as *(elementary) row operations*; each step can be viewed as modifying a row (or two rows) of A, and the corresponding entries of $\mathbf{b}$. (Or, equivalently, the row operations are applied to the augmented matrix $[A \mid \mathbf{b}]$.) Here are the possible row operations:

1. Swap row i and row j.

   (To "undo" this operation, swap rows i and j again.)

2. Multiply row i by a non-zero scalar.

   (To "undo" this operation, divide row i by the same non-zero scalar.)

3. Subtract a multiple of row i from row j (for $i \neq j$).

   (To "undo" this operation, add the same multiple of row i to row j.)

It turns out to be important that every row operation has a corresponding "undo" operation of the same "type"; we will return to this point later.

The goal is to apply these row operations until the resulting coefficient matrix is in reduced row echelon form (RREF). It is a bit tedious to write down the precise algorithm for determining the sequence of row operations. So, instead, we'll just illustrate it by example.

Continuing the previous example. Subtract 2 times row 1 from row 2:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 2 & 4 & 4 & 4 & 2 \\ 3 & 6 & 5 & 4 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & -2 & -4 & -4 \\ 3 & 6 & 5 & 4 & 1 \end{array}\right].$$

Subtract 3 times row 1 from row 3:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & -2 & -4 & -4 \\ 3 & 6 & 5 & 4 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & -2 & -4 & -4 \\ 0 & 0 & -4 & -8 & -8 \end{array}\right].$$

Multiply row 2 by $-1/2$:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & -2 & -4 & -4 \\ 0 & 0 & -4 & -8 & -8 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & -4 & -8 & -8 \end{array}\right].$$

Subtract -4 times row 2 from row 3:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & -4 & -8 & -8 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Subtract 3 times row 2 from row 1:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & 0 & -2 & -3 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Note that the final coefficient matrix is in RREF. ∎

2.3 Solving a system with coefficient matrix in RREF

Once the coefficient matrix is in RREF, there is a simple rule for determining if the system is consistent, and also for finding a solution if one exists.

Theorem 1. *A system of linear equations $R\mathbf{x} = \mathbf{c}$ with coefficient matrix R in RREF is consistent if and only if every all-zeros row of R has a 0 as the corresponding right-hand side value in $\mathbf{c}$.*

Proof. There are exactly two possible situations.

Case 1: The coefficient matrix R has an all-zeros row where the corresponding right-hand side value in $\mathbf{c}$ is non-zero.

In this case, there is a linear equation in the system of the form

$$0\,x_1 + \cdots + 0\,x_n \;=\; c_i$$

for some $c_i \neq 0$. That equation cannot be satisfied by any values of $(x_1, \ldots, x_n)$, so the system is inconsistent.

Case 2: Every all-zeros row in the coefficient matrix R has 0 as the corresponding right-hand side value in $\mathbf{c}$.

In this case, we can drop (i.e., ignore) these all-zeros rows (if any), as they correspond to the trivial equation

$$0\,x_1 + \cdots + 0\,x_n \;=\; 0,$$

which is always satisfied. What remains is a system in which the coefficient matrix is in RREF without any all-zeros rows. So each row has a first non-zero entry, called the pivot, with value 1; and the column in which a pivot occurs appears has only a single non-zero component.

The columns containing a pivot correspond to a subset of variables, called *pivot variables* (a.k.a. *basic variables*). The remaining variables are called *free variables* (a.k.a. *non-basic variables*). Each pivot variable has a non-zero coefficient (equal to 1) in exactly one equation, and for that equation, the only other variables with non-zero coefficients are free variables. Therefore, for any assignment of values to the free variables, we can uniquely determine a value for each pivot variable so that every equation in the system is satisfied. This shows that the system is consistent. □

The second case in the proof of Theorem 1 describes how one might find a solution to the system of linear equations $R\mathbf{x} = \mathbf{c}$ (with R in RREF) when it is consistent. One simple possibility is the following:

- Assign $x_j = 0$ for every free variable x_j.
- Assign to each pivot variable x_j the right-hand side value of the equation in which x_j appears. That is, if pivot variable x_j appears in the ith equation, then assign $x_j = c_i$.

But this is only one possibility; there could be others! We will discuss later a method for characterizing all solutions to the system in a succinct way.

To summarize: the system of linear equations can either be inconsistent (Case 1 in the proof of Theorem 1), or it can be consistent (Case 2). And if it is consistent, it turns out it can either have exactly one solution (no free variables), or it can have infinitely-many solutions (at least one free variable).

Continuing the previous example. The augmented matrix for the final system is

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & -2 & -3 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

The only all-zeros row in the coefficient matrix has a 0 for its corresponding right-hand side value. So this system is consistent. The pivot variables are x_1 and x_3 ($P = \{1, 3\}$), and the free variables are x_2 and x_4 ($F = \{2, 4\}$). The non-trivial equations are:

$$\begin{cases} x_1 + 2x_2 \qquad - 2x_4 = -3 \\ \qquad\qquad x_3 + 2x_4 = 2. \end{cases}$$

Suppose we assign $x_2 = 0$ and $x_4 = 0$:

$$\begin{cases} x_1 + \cancelto{0}{2x_2} \qquad - \cancelto{0}{2x_4} = -3 \\ \qquad\qquad x_3 + \cancelto{0}{2x_4} = 2. \end{cases}$$

Then to satisfy the first equation, we should set $x_1 = -3$; and to satisfy the second equation, we should set $x_3 = 2$. So one solution to the system is $(x_1, x_2, x_3, x_4) = (-3, 0, 2, 0)$. ∎

Another example. Consider the same (original) coefficient matrix as in the previous example, but now instead a different right-hand side vector: $(1, -2, 1)$. Here is the augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 1 \\ 2 & 4 & 4 & 4 & -2 \\ 3 & 6 & 5 & 4 & 1 \end{array}\right].$$

Since the coefficient matrix is the same as in the previous example, we can apply the same sequence of row operations to bring the coefficient matrix into RREF. So we just have to apply this same sequence of row operations to the right-hand side vector $(1, -2, 1)$:

$$\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ -4 \\ 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ -4 \\ -2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} \longrightarrow \begin{bmatrix} -5 \\ 2 \\ 6 \end{bmatrix}.$$

The resulting augmented matrix is

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & -2 & -5 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 6 \end{array}\right].$$

In this case, the all-zeros row in the coefficient matrix has a non-zero corresponding right-hand side value of 6, so the system is inconsistent. ∎

3 Row operations and inverses

The correctness of reduction performed by Elimination relies crucially on its row operations and the fact that each has an associated "undo" operation.

Elimination applies a sequence of row operations to change the augmented matrix for the original system to an augmented matrix in which the coefficient matrix is in RREF. Each step modifies one or two of the rows of this augmented matrix. We can equivalently view each step as applying the transformation to every column of the augmented matrix simultaneously.

Continuing the previous example. Applying the operation "subtract 2 times row 1 from row 2" to each column of the augmented matrix:

- First column:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

- Second column:

$$\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 \\ 0 \\ 6 \end{bmatrix}$$

- Third column:

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix}$$

- Etc.

It turns out that this operation is equivalent to multiplying the column by the following 3×3 matrix:

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

To see this, observe its behavior on a generic 3-vector $\mathbf{a} = (a_1, a_2, a_3)$:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 - 2a_1 \\ a_3 \end{bmatrix}.$$

We have precisely subtracted 2 times the first component from the second component; the other components remain the same. ∎

In general, the matrix corresponding to a row operation for a system of m linear equations is obtained by starting with the $m \times m$ identity matrix, and then applying the row operation to this identity matrix. These matrices corresponding to row operations are called *elementary matrices*.

Examples.

- "Subtract 3 times row 1 from row 3":
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}.$$
- "Multiply row 2 by $-1/2$":
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$
- "Swap row 2 and row 3":
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$ ∎

We say an operator $T \colon \mathbb{R}^m \to \mathbb{R}^m$ is <u>*invertible*</u> if there exists another operator, denoted by $T^{-1} \colon \mathbb{R}^m \to \mathbb{R}^m$ and called the <u>*inverse of T*</u>, that satisfies

$$T^{-1}(T(\mathbf{v})) \ = \ T(T^{-1}(\mathbf{v})) \ = \ \mathbf{v} \quad \text{for all } m\text{-vectors } \mathbf{v}.$$

The relationship is symmetric: T is the inverse of T^{-1}.

Proposition 1. *Any row operation for a system of m linear equations is specified by applying an operator $T \colon \mathbb{R}^m \to \mathbb{R}^m$ to every column of the augmented matrix for the system. This operator T is linear and invertible.*

Proof. The first claim follows from the fact that row operations perform the same operation to each column of an augmented matrix. The claim that the operator is linear follows from the fact that the operator is specified by matrix-vector multiplication with an elementary matrix. Finally, the claim that the operator is invertible comes from the fact that every row operation has a corresponding "undo" operation. □

If the invertible operator T is linear, then it turns out its inverse is also linear. (We already knew this was true for the operators corresponding to row operations, since the "undo" operator for a row operation is also a row operation.)

Proposition 2. *If $T\colon \mathbb{R}^m \to \mathbb{R}^m$ is an invertible linear operator with inverse transformation $T^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$, then T^{-1} is linear.*

Proof. Consider any m-vectors $\mathbf{u}$ and $\mathbf{v}$ and any scalar c. Let $\mathbf{x} = T^{-1}(\mathbf{u})$ and $\mathbf{y} = T^{-1}(\mathbf{v})$. Now consider $T^{-1}(c\mathbf{u} + \mathbf{v})$:

$$\begin{aligned} T^{-1}(c\,\mathbf{u} + \mathbf{v}) &= T^{-1}(c\,T(\mathbf{x}) + T(\mathbf{y})) \\ &= T^{-1}(T(c\,\mathbf{x} + \mathbf{y})) \\ &= c\,\mathbf{x} + \mathbf{y} \\ &= c\,T^{-1}(\mathbf{u}) + T^{-1}(\mathbf{v}). \end{aligned}$$

Above, the second equality follows by linearity of T. So T^{-1} is linear. □

An $m \times m$ matrix B is *invertible* if the linear operator $T\colon \mathbb{R}^m \to \mathbb{R}^m$ given by $T(\mathbf{x}) = B\mathbf{x}$ is invertible. We denote its inverse by B^{-1}, and it satisfies $B^{-1}B = BB^{-1} = I$. A matrix that is invertible is also called *non-singular*; a matrix that is not invertible is called *singular*.

Theorem 2. *If a square matrix is invertible, then its columns are linearly independent.*

Proof. Let $\mathbf{b}_1, \dots, \mathbf{b}_m$ denote the columns of an $m \times m$ matrix B. We'll prove the contrapositive of the claim.

Suppose $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ is a set of m linearly dependent vectors. This means that there are scalars $c_1, \dots, c_m \in \mathbb{R}$, not all equal to 0, such that

$$c_1\,\mathbf{b}_1 + \cdots + c_m\,\mathbf{b}_m \;=\; \mathbf{0}.$$

We'll now show that two different m-vectors $\mathbf{x}$ and $\mathbf{y}$ have $B\mathbf{x} = B\mathbf{y}$, so the transformation given by B is not invertible. Let $\mathbf{x} = \mathbf{0}$ and $\mathbf{y} = (c_1, \dots, c_m)$. Clearly, $B\mathbf{x} = \mathbf{0}$ and

$$B\mathbf{y} \;=\; c_1\,\mathbf{b}_1 + \cdots + c_m\,\mathbf{b}_m \;=\; \mathbf{0},$$

and yet $\mathbf{x} \neq \mathbf{y}$ since not all c_i's are equal to 0.

Now suppose instead that the columns are not distinct, i.e., $\mathbf{b}_i = \mathbf{b}_j$ for some $i \neq j$. Then $\mathbf{x} = \mathbf{0}$ and $\mathbf{y} = \mathbf{e}_i - \mathbf{e}_j \neq \mathbf{0}$ satisfy $B\mathbf{x} = B\mathbf{y} = \mathbf{0}$ yet $\mathbf{x} \neq \mathbf{y}$. (Here, $\mathbf{e}_k$ is the m-vector whose kth component is equal to 1 and all other components are equal to 0.) □

4 Correctness of Elimination

Suppose the row operations carried out by Elimination correspond to the elementary matrices $E_1, E_2, \ldots, E_k$. Then Elimination produces a sequence of systems of linear equations:

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ &\downarrow \\ (E_1 A)\mathbf{x} &= E_1\mathbf{b} \\ &\downarrow \\ (E_2 E_1 A)\mathbf{x} &= E_2 E_1\mathbf{b} \\ &\downarrow \\ &\vdots \\ &\downarrow \\ (E_k \cdots E_2 E_1 A)\mathbf{x} &= E_k \cdots E_2 E_1\mathbf{b}. \end{aligned}$$

The final system has coefficient matrix $E_k \cdots E_2 E_1 A$ and right-hand side vector $E_k \cdots E_2 E_1 \mathbf{b}$. The overall transformation performed by Elimination is

$$E = E_k \cdots E_2 E_1,$$

which is the operator that first applies E_1, then applies E_2, and so on (i.e., a composition of linear operators). So the final system has coefficient matrix EA and right-hand side vector $E\mathbf{b}$:

$$(EA)\mathbf{x} = E\mathbf{b}.$$

So, if $\mathbf{x}$ is a solution to the original system $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x}$ is a solution to the final system $(EA)\mathbf{x} = E\mathbf{b}$, since $(EA)\mathbf{x} = E(A\mathbf{x})$.

Furthermore, if $\mathbf{x}$ is a solution to the final system $(EA)\mathbf{x} = E\mathbf{b}$, then it is also a solution to the original system $A\mathbf{x} = \mathbf{b}$. The reason for this is that E is also invertible, and its inverse is

$$E^{-1} = (E_k \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

Notice that we compose the inverses of $E_1, E_2, \ldots, E_k$ in the reverse of order in which they appear in E.[1] Starting from $(EA)\mathbf{x} = E\mathbf{b}$, and applying E^{-1} to both sides of the equation, we obtain $A\mathbf{x} = \mathbf{b}$.

We have thus justified the Elimination algorithm for solving a system of linear equations. Elimination reduces the problem of solving $A\mathbf{x} = \mathbf{b}$ to the problem of solving $R\mathbf{x} = \mathbf{c}$, where the coefficient matrix $R = EA$ is in RREF, and $\mathbf{c} = E\mathbf{b}$. The reduction is solution-preserving because E is invertible. Solving $R\mathbf{x} = \mathbf{c}$ is handled by the rule detailed in Theorem 1.

We note that Elimination is closely connected to the CR factorization, as both lead to the same RREF matrix.

Theorem 3. *Let A be a matrix, and let E denote the invertible linear transformation performed by Elimination to bring A to the matrix EA in RREF. Let R be obtained from EA by dropping any all-zeros rows. Then the CR factorization of A is $A = CR$ with the same matrix R and with C being the first d columns of E^{-1}, where $d = \operatorname{rank}(A)$.*

An important consequence of Theorem 3 is that the RREF R matrix obtained by any sequence of row operations is uniquely determined by A.

5 Inverse computation

Is it possible to determine if a given square matrix is invertible? By Theorem 2, a necessary condition for being invertible is to have linearly independent columns. It turns out this is also a sufficient condition.

Theorem 4. *If the columns of a square matrix are linearly independent, then the matrix is invertible.*

Proof. The CR factorization of an $n \times n$ matrix A with linearly independent columns must result in $C = A$ (all columns of A are included) and $R = I$ (in particular, $\operatorname{rank}(A) = n$). But by Theorem 3, $C = E^{-1}$, where E is the invertible linear transformation performed by Elimination for a system with A as the coefficient matrix. Since $A = C = E^{-1}$ and E is invertible, the matrix A is also invertible. □

[1] This should intuitively make sense: Since E_k is the last operator applied in E, it is the first operator we have to "undo" in the inverse E^{-1}.

The proof of Theorem 4 also suggests the *Gauss-Jordan algorithm* for computing the inverse of a given matrix (if it exists):

- Use Elimination to transform the given matrix A to RREF.
- If the RREF matrix R is the identity matrix, then A is invertible, and A^{-1} is given by the transformation carried out by Elimination, E.

A schematic trick for keeping track of E is to perform Elimination on the augmented matrix $[A \mid I]$. Elimination transforms $[A \mid I]$ to $[R \mid E]$; if $R = I$, then A is invertible, and $E = A^{-1}$.

Example. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 0 & 4 \\ 3 & 1 & 5 \end{bmatrix}.$$

Applying Elimination on the augmented matrix $[A \mid I]$:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 3 & 1 & 0 & 0 \\ 2 & 0 & 4 & 0 & 1 & 0 \\ 3 & 1 & 5 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -\frac{1}{2} & 1 \\ 0 & 1 & 0 & \frac{1}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{array}\right].$$

So A is invertible, and its inverse is given to the right of the vertical line in the final augmented matrix. ∎

Theorem 5 (Invertibility Theorem). *Let A be an $n \times n$ matrix. The following are equivalent.*

1. *A is invertible (a.k.a. non-singular).*
2. *A has rank n.*
3. *For any n-vector $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution for $\mathbf{x}$.*
4. *$A\mathbf{x} = \mathbf{0}$ has $\mathbf{x} = \mathbf{0}$ as its unique solution.*
5. *There is a product E of elementary matrices such that $EA = I$.*
6. *A is equal to a product of elementary matrices.*

A Correspondence between Elimination and CR factorization

Proof of Theorem 3. Let E denote the invertible linear transformation carried out by Elimination to bring the $m \times n$ matrix $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ to RREF. So the matrix EA is in RREF; note that it may have some all-zeros rows appearing after all non-zero rows:

$$EA = \begin{bmatrix} R \\ O \end{bmatrix},$$

where R is a $d \times n$ matrix in RREF with no all-zeros rows, and O is the $(m-d) \times n$ matrix containing $m - d$ all-zeros rows.[2]

Let C be the $m \times d$ matrix containing the first d columns of E^{-1}, and let F be the $m \times (m-d)$ matrix containing the last $m-d$ columns of E^{-1}. So

$$A = E^{-1} \begin{bmatrix} R \\ O \end{bmatrix} = \begin{bmatrix} C & F \end{bmatrix} \begin{bmatrix} R \\ O \end{bmatrix} = CR.$$

(Here, we are using block-wise matrix multiplication.) Let $C = [\mathbf{c}_1, \dots, \mathbf{c}_d]$ and $R = [\mathbf{r}_1, \dots, \mathbf{r}_n]$. Observe that $\mathbf{a}_k = C\mathbf{r}_k$.

- If the kth column of R contains a pivot in the ith row, then

  $$\mathbf{a}_k = C\mathbf{r}_k = \mathbf{c}_i.$$

  The second equality is due to the RREF of R: the only non-zero component of $\mathbf{r}_k$ is the ith component, which must have value 1. Thus the matrix C contains a subset of d columns of A. Moreover, C contains columns of E^{-1}, and the columns of E^{-1} are linearly independent by Theorem 2. By the Removal Theorem, the columns of C are linearly independent columns of A.

- If the kth column of R does not contain a pivot, and R contains i pivots among its first $k-1$ columns, then

  $$\mathbf{a}_k = C\mathbf{r}_k \in \operatorname{span}(\{\mathbf{c}_1, \dots, \mathbf{c}_i\}).$$

[2] The only reason we need the “swap”-type row operations in Elimination is to make sure any all-zeros rows appear at the bottom in EA. This is only needed because we have defined RREF to require this, which in turn is only used to make some arguments like the current one a tiny bit simpler.

> The inclusion of $C\mathbf{r}_k$ in $\operatorname{span}(\{\mathbf{c}_1, \dots, \mathbf{c}_i\}$ is due to the RREF of R: since there are only i pivots among the first $k-1$ columns of R, the last $d-i$ components of $\mathbf{r}_k$ must be equal to 0. So while $\{\mathbf{c}_1, \dots, \mathbf{c}_i\}$ is linearly independent, the set $\{\mathbf{c}_1, \dots, \mathbf{c}_i, \mathbf{a}_k\}$ is not.

We have just described the CR factorization of the matrix A, which expresses $A = CR$, where C is a matrix containing linearly independent columns of A, and R is a matrix in RREF without any all-zeros rows. $\square$