Linear independence and systems of linear equations

Francesco Preta

July 21, 2020

Until now, we have used the theory of determinants and inverses to calculate the solution of square systems of linear equations whenever the determinant of the coefficient matrix A is nonzero. However, we haven't yet developed a theory to understand what happens whenever $\det(A) = 0$, or when A is not a square matrix. In order to do so, we need to analyze A as a linear operator.

Consider the general system of linear equations $A\mathbf{x} = \mathbf{b}$, for $A \in \mathbb{R}^{m\times n}$. Finding a solution for the system corresponds to finding a vector $\mathbf{x} \in \mathbb{R}^n$ which is mapped to a given vector $\mathbf{b} \in \mathbb{R}^m$. In this sense, there are three possibilities:

- $\mathbf{b} \in \mathrm{Im}(A)$, $\mathrm{Ker}(A) = \{0\}$, in which case there exists a left inverse A^{-1} granting a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.
- $\mathbf{b} \in \mathrm{Im}(A)$, $\mathrm{Ker}(A) \neq \{0\}$, in which case there are infinitely many solutions.
- $\mathbf{b} \notin \mathrm{Im}(A)$, in which case there are no solutions.

In our previous class we have only analyzed the first possibility for the case $m = n$. We will see today that $\det(A) \neq 0$ corresponds to $\mathrm{Ker}(A) = \{0\}$, which for square matrices is equivalent to A being bijective. In this way, every $\mathbf{b} \in \mathrm{Im}(A)$ and therefore a system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $b \in \mathbb{R}^n$.

Linear independence

Definition 1. Let $\mathcal{V}$ be a vector space and $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ a subset of vectors in $\mathcal{V}$. We say that $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are *linearly independent* if, for $\alpha_1, ..., \alpha_k \in \mathbb{R}$

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + ... + \alpha_k\mathbf{v}_k = 0 \implies (\alpha_1, ..., \alpha_k) = \mathbf{0}$$

In other word, $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are linearly independent if any nonzero linear combination of them is also nonzero.

Notice that the definition of linear independence implies that no single vector $\mathbf{v}_j \in \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is a linear combination of the other $k-1$ vectors. In particular, two vectors are linearly independent if one is not a multiple of the other. Also, if $\mathbf{0}$ is in $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, then the vectors are not independent, since any linear combination attributing a nonzero weight only to $\mathbf{0}$ would give $\mathbf{0}$ as a result.

For a given set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ of $\mathbb{R}^n$, we need to be able to determine whether they're independent or not. This corresponds to asking if we can find $(\alpha_1, ..., \alpha_k) \neq \mathbf{0}$ such that

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + ... + \alpha_k \mathbf{v}_k = 0$$

This condition corresponds to solving the homogeneous system $V\mathbf{a} = \mathbf{0}$ for

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & ... & \mathbf{v}_k \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} \alpha_1 \\ ... \\ \alpha_k \end{bmatrix}$$

Example: determine if the following vectors are linearly independent.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 2 \\ -2 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -2 \\ 8 \\ -3 \end{bmatrix}.$$

The augmented matrix of the system $V\mathbf{a} = \mathbf{0}$ is

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 2 & -2 & 0 \\ 2 & -2 & 8 & 0 \\ -1 & 0 & -3 & 0 \end{bmatrix}$$

The echelon form of this matrix is

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

so that the reduced echelon form becomes:

$$\begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since there are not rows of the type $[\mathbf{0}|b]$ for $b \neq 0$ or blocks of the type $[I|\mathbf{c}]$, there are infinitely many solutions to this systems. For this reason, there exist $(\alpha_1, \alpha_2, \alpha_3) \neq \mathbf{0}$ such that $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 = 0$, so that these vectors are not linearly independent.

Example: determine if the following vectors are linearly independent.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 2 \\ -2 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ -3 \end{bmatrix}.$$

Then the augmented matrix of the system $V\mathbf{a} = \mathbf{0}$ is

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 2 & -1 & 0 \\ 2 & -2 & 0 & 0 \\ -1 & 0 & -3 & 0 \end{bmatrix}$$

The echelon form of this matrix is

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & -8 & 0 & 0 \\ 0 & 0 & -8 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

so that the reduced echelon form becomes:

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

Since there is a block of the type $[I|\mathbf{c}]$, the system has only one solution corresponding to $\mathbf{0}$ and therefore the three vectors are linearly independent.

Another important question is the following: given a vector space $\mathcal{V}$, how many independent vectors can we find? The answer gives us the notions of *basis* and *dimension* of a vector space.

Definition 2. Let $\mathcal{V}$ be a vector space. A set $\{\mathbf{v}_1, .., \mathbf{v}_n\}$ is called a *basis* for $\mathcal{V}$ if

1. $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is a set of linearly independent vector.
2. For every $\mathbf{v} \in \mathcal{V}$ there exists $\alpha_1, ..., \alpha_n \in \mathbb{R}$ such that

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + ... + \alpha_n \mathbf{v}_n$$

In other words, $\mathcal{V} = \text{span}(\mathbf{v}_1, ..., \mathbf{v}_n)$

The *dimension* of $\mathcal{V}$ is the cardinality of its basis.

For example, for $\mathcal{V} = \mathbb{R}^n$, the canonical basis $\mathbf{e}_1, ..., \mathbf{e}_n$ constitute a basis. The dimension is n, as expected.

In this course, we will mostly encounter with finite-dimensional vector spaces. The following is an important property of finite-dimensional vector spaces:

Theorem 1. *Let $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a subset of a n-dimensional vector space. Then if $k > n$, $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is not linearly independent.*

In fact, there are at most n linearly independent vectors in a n-dimensional vector space and any group of them constitutes a basis for $\mathbb{R}^n$.

Finally, consider the following definition:

Definition 3. Let $\mathcal{V}$ be a finite dimensional vector space. A subset $\mathcal{V}' \subset \mathcal{V}$ is called a vector subspace of $\mathcal{V}$ if the following conditions are true:

- $\mathbf{0} \in \mathcal{V}'$.
- For every $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}'$ and $\alpha_1, \alpha_2 \in \mathbb{R}$, $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 \in \mathcal{V}'$.

In other words, a vector subspace is a subset of a vector space which is closed under linear combinations.

Given a set $\{\mathbf{v}_1, ..., \mathbf{v}_k\} \subset \mathbb{R}^n$, $\operatorname{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ is a vector subspace of $\mathbb{R}^n$. If $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are linearly independent, then $\operatorname{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ is a k-dimensional vector subspace of $\mathbb{R}^n$. Otherwise, $\operatorname{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ has dimension $d < k$.

The following is an important result we will make extensive use of:

Theorem 2. *Let A be a $m \times n$ matrix. Then $\ker(A)$ is a vector subspace of $\mathbb{R}^n$ and $\operatorname{Im}(A)$ is a vector subspace of $\mathbb{R}^m$ spanned by the columns of A.*

Proof. We already know that $\operatorname{Im}(A) = \operatorname{span}(\mathbf{a}_1, ..., \mathbf{a}_n)$, so that since the span of a finite number of vectors is a vector space, $\operatorname{Im}(A)$ is a vector subspace of $\mathbb{R}^m$. In order to prove that $\operatorname{Ker}(A)$ is a vector subspace of $\mathbb{R}^n$, notice that $A\mathbf{0} = \mathbf{0}$, so that $\mathbf{0} \in \operatorname{Ker}(A)$. Moreover, if $\mathbf{v}_1, \mathbf{v}_2 \in \operatorname{Ker}(A)$, then for every $\alpha, \beta \in \mathbb{R}$

$$A(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha A \mathbf{v}_1 + \beta A \mathbf{v}_2 = \mathbf{0}$$

so that $\operatorname{Ker}(A)$ is closed under linear combinations and the theorem is proved. □

The following theorem allows us to find a basis for any given vector space or subspace for which we know a set of generators.

Theorem 3. *Let $\mathcal{V}$ be a vector space, $\{\mathbf{v}_1, ..., \mathbf{v}_k\} \subset \mathcal{V}$ and $H = \operatorname{span}(v_1, ..., v_k)$. Then*

- *If $\mathbf{v}_j$ is a linear combination of the other $\mathbf{v}_i$'s, then*
$$H = \operatorname{span}(\mathbf{v}_1, ..., \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, ..., \mathbf{v}_k\}.$$
- *If $\mathbf{v}$ is a non-trivial linear combination of $\mathbf{v}_l$ with $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, then*
$$H = \operatorname{span}(\mathbf{v}_1, ..., \mathbf{v}_{l-1}, \mathbf{v}, \mathbf{v}_{l+1}, ..., \mathbf{v}_k\}.$$
- *If $H \neq \{\mathbf{0}\}$, a subset of $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is a basis for H.*

Therefore, in order to find a basis for a vector space H, we need to remove $\mathbf{0}$ and linear combinations from the vectors that span H.

Example: find a basis for $\operatorname{Ker}(A)$ and $\operatorname{Im}(A)$ for

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

Let us start with Ker(A):

$$\mathrm{Ker}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$$

that is, Ker(A) solves the system of linear equations $A\mathbf{x} = 0$. We can do so by Gaussian elimination on the augmented matrix

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 4 & 2 & -1 & 0 & 0 \\ 2 & 2 & -1 & 2 & 0 \end{bmatrix}$$

to get a reduced echelon form

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

This system has solutions of the type

$$\begin{cases} x_1 = x_4 \\ x_2 = \frac{x_3}{2} - 2x_4 \\ x_3 \text{ free} \\ x_4 \text{ free} \end{cases}$$

which can be written as

$$\mathrm{Ker}(A) = \left\{ \begin{bmatrix} \beta \\ \frac{\alpha}{2} - 2\beta \\ \alpha \\ \beta \end{bmatrix} \mid \alpha, \beta \in \mathbb{R} \right\} = \mathrm{span}\left(\begin{bmatrix} 0 \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right)$$

since the vectors $\begin{bmatrix} 0 \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \end{bmatrix}$ are linearly independent, they form a basis for Ker(A). Therefore $\dim(\mathrm{Ker}(A)) = 2$. Notice that the system $A\mathbf{x} = \mathbf{0}$ has always at least a solution equal to $\mathbf{0}$, so Ker(A) is never empty.

As for Im(A), we know that it is generated by the columns of A

$$\mathrm{Im}(A) = \mathrm{span}\left(\begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \right)$$

In order to find a basis, we need to find the solutions of $A\mathbf{x} = \mathbf{0}$ to identify the relations of linear dependence, that is, the x_1, x_2, x_3, x_4 such that

$$x_1 \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We have already found the solutions to be

$$\begin{cases} x_1 = x_4 \\ x_2 = \frac{x_3}{2} - 2x_4 \\ x_3 \text{ free} \\ x_4 \text{ free} \end{cases}$$

As long as there are free coefficients, there will be nonzero linear combinations of the columns giving a solution to the homogeneous system. Therefore, in order to find a basis for $\text{Im}(A)$, we need to eliminate all the vectors that have a free coefficient. This gives a basis as

$$\text{Im}(A) = \text{span}\left(\begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \right).$$

Notice that in the previous example we have shown the following:

Theorem 4. $\text{Im}(A)$ *is spanned by the pivot columns of* A*.*

This happens because the pivot columns correspond to non-free variables whose only feasible solution for the homogeneous system $A\mathbf{x} = \mathbf{0}$ is given by $\mathbf{0}$, once the free variables are eliminated.

Rank of matrices

The question about linear independence of the columns of A, for

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & ... & \mathbf{a}_n \end{bmatrix}$$

gives rise to the notion of *rank* of a matrix.

Definition 4. Let A be a $m \times n$ matrix. Then $\text{Rank}(A)$ is the dimension of the linear subspace of $\mathbb{R}^m$ generated by the column vectors $\{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n\}$.

Actually the following are all equivalent definitions:

Theorem 5. *Let* A *be a* $m \times n$ *matrix. Then the following are equivalent:*

1. $\text{Rank}(A) = r$.
2. *The subspace of* $\mathbb{R}^n$ *generated by the row vectors of* A *has dimension* r*.*
3. *There exists a square* $r \times r$ *submatrix of* A *with nonzero determinant and every square submatrix of higher dimension has determinant equal to zero.*

In particular, the third statement allows us to calculate the rank of a matrix.

Example:

$$\begin{bmatrix} 1 & 3 & 0 \\ 0 & 2 & -2 \\ 2 & -2 & 8 \\ -1 & 0 & -3 \end{bmatrix}$$

The submatrix $\begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$ has nonzero determinant, therefore the rank is at least 2. However, all the higher dimensional square submatrix have zero determinant, that is:

$$\begin{vmatrix} 1 & 3 & 0 \\ 0 & 2 & -2 \\ 2 & -2 & 8 \end{vmatrix} = 0, \qquad \begin{vmatrix} 1 & 3 & 0 \\ 0 & 2 & -2 \\ -1 & 0 & -3 \end{vmatrix} = 0$$

$$\begin{vmatrix} 1 & 3 & 0 \\ 2 & -2 & 8 \\ -1 & 0 & 3 \end{vmatrix} = 0, \qquad \begin{vmatrix} 0 & 2 & -2 \\ 2 & -2 & 8 \\ -1 & 0 & -3 \end{vmatrix} = 0$$

Actually the calculation can be simplified thanks to the following theorem:

Theorem 6. *Let A be a $n \times m$ matrix and A' be a r-dimensional square submatrix with nonzero determinant. Then* $\text{Rank}(A) = r$ *if for every row i and column j not already in A', the matrix obtained by adding such row and column to A' has zero determinant.*

In the previous example, for $A' = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$, it suffices to check

$$\begin{vmatrix} 1 & 3 & 0 \\ 0 & 2 & -2 \\ 2 & -2 & 8 \end{vmatrix} = 0 \qquad \text{and} \qquad \begin{vmatrix} 1 & 3 & 0 \\ 0 & 2 & -2 \\ -1 & 0 & -3 \end{vmatrix} = 0$$

. In order to understand whether a few vectors $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are linearly independent we can do the following:

1. Consider the matrix V having $\mathbf{v}_i$ as i-th column.
2. Calculate the rank of V through determinants of square submatrices.
3. A basis for $\text{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ is given by the columns included in the square submatrix of maximum rank.

The following is a very important theorem:

Theorem 7 (Rank-Kernel theorem)**.** *Let A be a $m \times n$ matrix. Then*

$$\text{Rank}(A) + \dim(\text{Ker}(A)) = n$$

Proof. Consider the homogeneous system $A\mathbf{x} = \mathbf{0}$. The general form of its solution in $\mathbb{R}^n$ can be divided into free variables and non-free variables. The number of free variables correspond to the dimension of $\text{Ker}(A)$, while the number of non-free variables correspond to the number of pivot columns. Since this last value is captured by the rank, the theorem is proved. □

Going back to our previous example,

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

we've proved that $\dim(\mathrm{Ker}(A)) = 2$ and $\mathrm{Rank}(A) = 2$, so that the Theorem is verified, as $n = 2$. The same result could have been obtained by calculating the rank directly through determinants of square submatrices.

Properties of the rank: let A be a $m \times n$ matrix.

1. $\mathrm{Rank}(A) = \mathrm{Rank}(A^T)$.
2. $\mathrm{Rank}(A) \leq \min\{m, n\}$.
3. If B is obtained by row-reduction operations on A, then $\mathrm{Rank}(B) = \mathrm{Rank}(A)$.
4. If B is obtained by column-reduction operations on A, then $\mathrm{Rank}(B) = \mathrm{Rank}(A)$.
5. If $m = n$ then $\mathrm{Rank}(A) = n$ if and only if A is invertible.

In particular, this last statement can be proven through the use of the rank-kernel Theorem. In fact, if $\mathrm{Rank}(A) = n$, then A is surjective and $\ker(A) = \{0\}$. Therefore A is bijective and can be inverted. On the other hand, if A is invertible, then it is surjective and so $\mathrm{Rank}(A) = n$.

This gives us an idea on why a square system $A\mathbf{x} = \mathbf{b}$ with nonzero determinant always has a unique solution: $\det(A) \neq 0$ is equivalent to $\mathrm{Rank}(A) = n$, so that for any $\mathbf{b} \in \mathbb{R}^n$ there exists a unique linear combination of the columns of A with coefficients $x_1, ..., x_n$ such that $\mathbf{b} = x_1\mathbf{a}_1 + ... + x_n\mathbf{a}_n$.

Linear independence in systems of linear equation

We are now ready to state the following theorem:

Theorem 8. *Consider a system of linear equation having matrix form* $A\mathbf{x} = \mathbf{b}$. *Then the system admits solutions if and only if* $\mathbf{b} \in \mathrm{span}(\mathbf{a}_1, ..., \mathbf{a}_n)$, *where* $\mathbf{a}_i$ *is the i-th column of A.*

The practical decision rule is explained by the following theorem:

Theorem 9 (Rouché-Capelli)**.** *Consider a system of linear equation* $A\mathbf{x} = \mathbf{b}$. *Then the system admits solution if and only if the coefficient matrix A and the augmented matrix* $A' = [A|\mathbf{b}]$ *have the same rank. If that is the case, the number of free variables is given by* $n - \mathrm{rank}(A)$, *where n is the total number of variables.*

In practice, we are asking for $\mathbf{b}$ to be a linear combination of the columns of A, which means that when considering the additional column vector $\mathbf{b}$, the matrix does not grow in rank.

We will now consider examples of underdetermined, square and overdetermined system. In what follows, if A is an $m \times n$ matrix, we say that A has maximum rank if $\mathrm{Rank}(A) = \min\{m, n\}$

Underdetermined systems

Let A be a $m \times n$ coefficient matrix of a system $A\mathbf{x} = \mathbf{b}$ for $m < n$. In case $\mathrm{Rank}(A) = m$, adding the column matrix $\mathbf{b}$ will not have an impact on the system, since $m \leq \mathrm{Rank}([A|\mathbf{b}]) \leq \min\{m, n+1\} = m$. By Theorem 9 the system will admit solutions. In particular, $n - m$ variables will be free, while the rest will be determined.

In case $\mathrm{Rank}(A) = r < m$, then the existence of solution depends on $\mathrm{Rank}([A|\mathbf{b}])$, which could potentially be larger than r if $\mathbf{b}$ is not a linear combination of the columns of A. In general, if solutions exist, $n - r$ variables will be free, while the rest will be determined.

Example:

$$\begin{bmatrix} 2 & -1 & 0 & 1 \\ -1 & 3 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Consider the submatrix $\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$ having nonzero determinant. Then $\mathrm{Rank}(A) = 2$, so the system will have solutions with $4 - 2 = 2$ free variables. In fact by Gaussian elimination we get

$$\begin{bmatrix} 10 & 0 & 2 & 6 & 6 \\ 0 & 5 & 2 & 1 & 1 \end{bmatrix}$$

which gives us

$$\begin{cases} x_1 = \frac{1}{5}(3 - x_3 - 3x_4) \\ x_2 = \frac{1}{5}(1 - 2x_3 - x_4) \\ x_3 \text{ free} \\ x_4 \text{ free} \end{cases}$$

Example:

$$\begin{bmatrix} 2 & -1 & 0 & 1 \\ -1 & 3 & 1 & 0 \\ 0 & 5 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Consider the submatrix $\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$ having nonzero determinant. Any 3×3

submatrix including this matrix has determinant equal to 0, that is

$$\begin{vmatrix} 2 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 5 & 2 \end{vmatrix} = 0, \qquad \begin{vmatrix} 2 & -1 & 1 \\ -1 & 3 & 0 \\ 0 & 5 & 1 \end{vmatrix} = 0$$

However, if we consider the augmented matrix $[A|\mathbf{b}]$, we can find another 3×3 submatrix given by

$$\begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 0 \\ 0 & 5 & 0 \end{bmatrix}$$

Since this matrix has nonzero determinant, the rank of the augmented matrix is greater than the rank of A and the system has no solutions.

Square systems

In case $n = m$, the coefficient matrix A is a square matrix. In this case, being maximum rank coincides with having nonzero determinant, which is exactly the case we dealt with in the last lecture. Notice that in this case

$$n = \text{Rank}([A|\mathbf{b}]) \leq \min\{n, n+1\} = n$$

so that the condition of Theorem 9 is respected.

In case $\text{Rank}(A) = r < n$, we have two possibilities depending on $\text{Rank}([A|\mathbf{b}])$. If $\text{Rank}([A|\mathbf{b}]) = r$, then the system has infinitely many solutions and $n-r$ variables will be free. Otherwise, the system will have no solution.

Example:

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

Notice that $\det(A) = 0$, but $\begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = 3 \neq 0$ so $\text{Rank}(A) = 2$. Moreover, we have that

$$\begin{vmatrix} 1 & 2 & 1 \\ 0 & 3 & 1 \\ -2 & -1 & -1 \end{vmatrix} = 0$$

so the rank of the augmented matrix is also 2 and therefore the system admits infinitely many solutions. We can find the solution set by usual Gaussian elimination. The echelon form of the augmented matrix is

$$\begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and the reduced echelon form is:

$$\begin{bmatrix} 3 & 0 & -5 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which gives the result

$$\begin{cases} x_1 = \frac{1}{3}(1+5x_3) \\ x_2 = \frac{1}{3}(1-x_3) \\ x_3 \text{ free} \end{cases}$$

Overdetermined system

In case $m > n$, the maximum rank of the coefficient matrix is n. Even in the maximum rank case, adding the column $\mathbf{b}$ could lead to an increase in the rank of the augmented matrix. This happens because there are more equations than unknowns. Ideally, every additional equation should be a linear combination of all the previous one, so that we don't add stringent requirements that cannot be satisfied by the unknowns.

Example: Consider the system

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

By taking the square submatrix $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 3 \\ -1 & 2 & 1 \end{bmatrix}$, we notice that A has maximum rank. However, we can calculate the determinant of the augmented matrix:

$$\begin{vmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 3 & 2 \\ -1 & 2 & 1 & -1 \\ 3 & -2 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 \\ 2 & 1 & -1 \\ -2 & 1 & 1 \end{vmatrix} - \begin{vmatrix} 0 & 3 & 2 \\ -1 & 1 & -1 \\ 3 & 1 & 1 \end{vmatrix} - \begin{vmatrix} 0 & 1 & 3 \\ -1 & 2 & 1 \\ 3 & -2 & 1 \end{vmatrix} = 32 \neq 0$$

Therefore there are no solution to this system of linear equations.

Right and left inverses

It is possible to generalize the notion of inverse to rectangular matrices as well.

Theorem 10. *Let A be a $m \times n$ matrix. Then*

- *There exists a right inverse Y such that $AY = I$ if and only if the rows of A are linearly independent. This is generally not unique.*

- *There exists a left inverse X such that $XA = I$ if and only if the columns of A are linearly independent. This is generally not unique.*

Notice that for wide matrices ($n > m$), columns cannot be linearly independent, therefore they only admit right inverses, while in tall matrices ($n < m$) rows cannot be linearly independent and so they only admit left inverses. In any case, a matrix A admits a (left and/or right) inverse if and only if it has maximum rank.

When it comes to the solutions of a system $A\mathbf{x} = b$, we have the following possibilities:

- If the system is under-determined ($n > m$) and A has maximum rank, then A admits a right inverse Y. Since A has maximum rank, by Theorem 9, the system admits solutions. Then, let $\mathbf{x} = Y\mathbf{b}$, we have $A\mathbf{x} = AY\mathbf{b} = \mathbf{b}$, therefore $\mathbf{x}$ is a solution to the system. Since Y is not unique, we can find other solutions by finding different right inverses.

- If the system is over-determined ($m > n$) and A has maximum rank, there is no guarantee that the system has solutions. However, if it does, it has only one solution by Theorem 9. If $\mathbf{x}$ is a solution of the system, then $XA\mathbf{x} = X\mathbf{b}$, meaning that $\mathbf{x} = X\mathbf{b}$ is the unique desired solution. On the other hand, if $AX\mathbf{b} \neq \mathbf{b}$, then there are no solutions.

Left and right inverse are generally not unique. However, a general formula is given by the following theorem.

Theorem 11. *A admits a left inverse if and only if the square matrix A^TA is invertible. In this case, a left inverse for A is given by*

$$A^\dagger = (A^TA)^{-1}A^T$$

A admits a right inverse if and only if the square matrix AA^T is invertible. In this case, a right inverse for A is given by

$$A^\dagger = A^T(AA^T)^{-1}$$

The matrices above are called *pseudo-inverses* of A.

Example: Consider

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

Since A is a tall matrix with linearly independent columns, it admits a left inverse. To find the pseudo-inverse we need the following calculations:

$$A^TA = \begin{bmatrix} 11 & 2 \\ 2 & 6 \end{bmatrix}$$

Then

$$A^\dagger = (A^TA)^{-1}A^T = \frac{1}{62}\begin{bmatrix} 6 & -2 \\ -2 & 11 \end{bmatrix}\begin{bmatrix} 1 & 3 & 1 \\ -2 & 1 & 1 \end{bmatrix} = \frac{1}{62}\begin{bmatrix} 10 & 16 & 4 \\ -24 & 5 & 9 \end{bmatrix}$$

And we can easily check that $A^\dagger A = I_2$. Moreover, for the system $A\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, we have that

$$AA^\dagger \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{62} A \begin{bmatrix} 10 \\ -24 \end{bmatrix} = \begin{bmatrix} \frac{29}{31} \\ \frac{3}{31} \\ -\frac{7}{31} \end{bmatrix} \neq \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

so that this system doesn't admit solutions. However, for $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$ we have

$$AA^\dagger \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} = A \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} = \mathbf{b}$$

so there exists a unique solution of the system given by $A^\dagger \mathbf{b} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.