Systems of Linear Equations

Finding Solutions

Definition 1. A system of linear equations is a collection of m hyperplanes in $\mathbb{R}^n$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &= b_m. \end{aligned} \tag{$\mathcal{S}$}$$

In matrix notation, we have

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

In a more compact notation, we write

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

It is easy to check that

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + ... + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

and so

$$\mathbf{b} \in \operatorname{Span}(\mathbf{c}_1(A), \mathbf{c}_2(A), ..., \mathbf{c}_n(A)).$$

Question 2. Given a linear system $A\mathbf{x} = \mathbf{b}$:
(1) Does there exists a solution vector $\mathbf{x}$?
(2) If a solution exists is it unique?

Definition 3. If the linear system $A\mathbf{x} = \mathbf{b}$ has a solution, then it is called a consistent linear system.

Theorem 4. The following statements are equivalent.
(1) $A\mathbf{x} = \mathbf{b}$ is consistent with solution $\mathbf{x} \in \mathbb{R}^n$.
(2) $\mathbf{x} \in \bigcap_{i=1}^{m} H(\mathbf{y}_i, \mathbf{a}_i)$ where $\mathbf{a}_i = (a_{i1}, a_{i2}, ..., a_{in})$ and $\mathbf{a}_i \cdot \mathbf{y}_i = b_i$.
(3) $\mathbf{b} \in \operatorname{Span}(\mathbf{c}_1(A), A\mathbf{c}_2(A), ..., \mathbf{c}_n(A))$.

Definition 5. We define the solution set of the system $\mathcal{S}$ to be the set

$$\mathrm{Sol}(\mathcal{S}) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}.$$

We say that two systems $\mathcal{S}$ and $\mathcal{S}'$ are equivalent (written $\mathcal{S} \sim \mathcal{S}'$) if $\mathrm{Sol}(\mathcal{S}) = \mathrm{Sol}(\mathcal{S}')$

Theorem 5. Let $E \in \mathcal{M}_{m\times m}$ be a product of elementary matrices. Then $A\mathbf{x} = \mathbf{b}$ and $EA\mathbf{x} = E\mathbf{b}$ are equivalent.

Proof. It suffices to check the three elementary cases. The only one of interest is the third one. If $\mathbf{u} = (u_1, u_2, ..., u_n)$ is a solution to $a_{i1}x_1+a_{i2}x_2+...+a_{in}x_n = b_i$, then it is also a solution to $(a_{i1}+ca_{k1})x_1+(a_{i2}+ca_{k2})x_2+...+(a_{i1}+ca_{k1})x_n = b_i + cb_k$. Indeed,

$$\begin{aligned}
& (a_{i1} + ca_{k1})u_1 + (a_{i2} + ca_{k2})u_2 + ... + (a_{i1} + ca_{k1})u_n \\
= \; & (a_{i1} + ca_{k1})u_1 + (a_{i2} + ca_{k2})u_2 + ... + (a_{i1} + ca_{k1})u_n \\
= \; & (a_{i1}u_1 + a_{i2}u_2 + ... + a_{in}u_n) + c(a_{k1}u_1 + a_{k2}u_2 + ... + a_{kn}u_n) \\
= \; & b_i + cb_k.
\end{aligned}$$

This proves that $\mathrm{Sol}(\mathcal{S}) \subseteq \mathrm{Sol}(E\mathcal{S})$. Conversely, if $\mathbf{u}$ is a solution to $EA\mathbf{x} = E\mathbf{b}$. That is, $EA\mathbf{u} = E\mathbf{b}$. Since E is a product of invertible matrices, it too is invertible so that $A\mathbf{u} = \mathbf{b}$. It follows that $\mathrm{Sol}(E\mathcal{S}) \subseteq \mathrm{Sol}(\mathcal{S})$.

Definition 6. Let $A \in \mathcal{M}_{m\times n}$. The leading entry of $\mathbf{r}_i(A)$ is the first nonzero entry of the row. That is, if $a_{i1}, a_{i2}, ..., a_{ij-1} = 0$ and $a_{ij} \neq 0$, then a_{ij} is the leading entry of $\mathbf{r}_i(A)$. In this case, we write $l(i) = j$.

Example 7. Let

$$A = \begin{bmatrix} 0 & 1 & 3 & -\frac{1}{2} & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -1 & 0 & \frac{1}{2} & -7 \\ 0 & 0 & 0 & 8 & 4 \end{bmatrix}.$$

Then $l(1) = 2$, $l(2)$ is undefined, $l(3) = 1$ and $l(4) = 4$.

Definition 7. A matrix $A \in \mathcal{M}_{m\times n}$ is said to be in reduced row echelon form ($rref$) if
(1) All zero rows are at the bottom of the matrix. Say $\mathbf{r}_i(A) \neq \mathbf{0}$ for all $i \in \{1, 2, ..., r\}$ and $\mathbf{r}_i(A) = \mathbf{0}$ for all $i \in \{r+1, r+2, ..., n\}$.
(2) As the row number increases, the leading entries move to the right. That is, $l(1) < l(2) < ... < l(r)$.
(3) All leading entries are 1. That is, if $l(i) = j$, then $a_{ij} = 1$.
(4) If a column contains a leading entry, then every other entry in the column is 0. That is, if $l(i) = j$, then $a_{kj} = 0$ for all $k \neq i$.

Theorem 8. Every matrix is row equivalent to a unique matrix in reduced row echelon form. Therefore, to find the solutions of $A\mathbf{x} = \mathbf{b}$, it suffices to solve the system $A_{rref}\mathbf{x} = \mathbf{b}_{rref}$.

Definition 9. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{y}$. The line segment with end points $\mathbf{x}, \mathbf{y}$ is the set

$$S(\mathbf{x}, \mathbf{y}) = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = (1-t)\mathbf{x} + t\mathbf{y}\}.$$

Remark 10. Since there are infinitely many points in the interval $[0, 1]$, the set $S(\mathbf{x}, \mathbf{y})$ must be infinite (in fact uncountable).

Theorem 11. Exactly one of the following statements holds:
(1) $A\mathbf{x} = \mathbf{b}$ has no solution.
(2) $A\mathbf{x} = \mathbf{b}$ has a unique solution.
(3) $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions.

Proof. It is easy to show that (1) can hold and so can (2). Moreover, they cannot happen at the same time. Now suppose that $\mathbf{x} \neq \mathbf{y}$ are two distinct solutions to the system $A\mathbf{x} = \mathbf{b}$. We claim that every $\mathbf{v} \in S(\mathbf{x}, \mathbf{y})$ is a solution. Indeed, we have

$$\begin{aligned}
& A\mathbf{v} \\
= \;& A((1-t)\mathbf{x} + t\mathbf{y}) \\
= \;& (1-t)A\mathbf{x} + tA\mathbf{y} \\
= \;& (1-t)\mathbf{b} + t\mathbf{b} \\
= \;& \mathbf{b} - t\mathbf{b} + t\mathbf{b} \\
= \;& \mathbf{b}.
\end{aligned}$$

Therefore, if (1) and (2) are false, then (3) must hold.

Definition 12. A linear system $A\mathbf{x} = \mathbf{b}$ is called homogeneous if $\mathbf{b} = \mathbf{0}$.

Theorem 13. Let $\mathcal{H}$ be the homogeneous linear system $A\mathbf{x} = \mathbf{0}$.
(1) Either $\mathcal{H}$ has a unique solution or infinitely many solutions.
(2) If $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k \in \mathrm{Sol}(\mathcal{H})$, then $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k) \subseteq \mathrm{Sol}(\mathcal{H})$.

Proof.
(1) Notice that $\mathbf{x} = \mathbf{0}$ is always a solution (called the trivial solution) to the homogeneous system $A\mathbf{x} = \mathbf{0}$. That is, statement (1) in Theorem 11 is false and therefore, either (2) or (3) holds.
(2) Exercise.

Theorem 14. Let $\mathcal{S}$ be the linear system $A\mathbf{x} = \mathbf{b}$ and let $\mathbf{x}_p$ be a solution. If $\mathcal{H}$ is the associated homogeneous linear system $A\mathbf{x} = \mathbf{0}$, then

$$\mathrm{Sol}(\mathcal{S}) = \{\mathbf{v} + \mathbf{x}_p : \mathbf{v} \in \mathrm{Sol}(\mathcal{H})\}.$$

Proof. ($\subseteq$) Choose any $\mathbf{y} \in \mathrm{Sol}(\mathcal{S})$. Then $A\mathbf{x}_p = \mathbf{b}$ and $A\mathbf{y} = \mathbf{b}$. It follows that $A\mathbf{x}_p = A\mathbf{y}$ and so $A(\mathbf{y} - \mathbf{x}_p) = \mathbf{0}$. Therefore $\mathbf{y} - \mathbf{x}_p \in \mathrm{Sol}(\mathcal{H})$ and setting $\mathbf{v} = \mathbf{y} - \mathbf{x}_p$, we find that $\mathbf{y} = \mathbf{v} + \mathbf{x}_p$.

($\supseteq$) Choose any $\mathbf{y} \in \{\mathbf{v} + \mathbf{x}_p : \mathbf{v} \in \mathrm{Sol}(\mathcal{H})\}$. Then $\mathbf{y} = \mathbf{v} + \mathbf{x}_p$ for some $\mathbf{v} \in \mathrm{Sol}(\mathcal{H})$. We have $A\mathbf{y} = A(\mathbf{v} + \mathbf{x}_p) = A\mathbf{v} + A\mathbf{x}_p = \mathbf{0} + \mathbf{b} = \mathbf{b}$. It follows that $\mathbf{y} \in \mathrm{Sol}(\mathcal{S})$ as needed.

Definition 15. Let $A \in \mathcal{M}_{m\times n}$ be a matrix in $rref$.
(1) The rank of A is the natural number $\mathrm{rank}(A) = r$ (See Definition 7(1)). That is, $\mathrm{rank}(A)$ is the number of nonzero rows.
(2) If $l(i) = j$, then the entry $a_{ij} = 1$ is called a pivot of the matrix A.
(3) We write $[A \mid \mathbf{b}]$ to denote the augmented matrix associated with the linear system $A\mathbf{x} = \mathbf{b}$.

Facts 16. Let $A \in \mathcal{M}_{m\times n}$ be a matrix in $rref$. Then the number of pivots is the same as $\mathrm{rank}(A)$. Moreover, $\mathrm{rank}(A) \le n$ and $\mathrm{rank}(A) \le m$.

Proof. A pivot is the same as a leading entry of a row. Every nonzero row has exactly one leading entry. Therefore, there are $\mathrm{rank}(A)$ many pivots.

Theorem 17. Let $A \in \mathcal{M}_{m\times n}$ and let $A\mathbf{x} = \mathbf{b}$ be a linear system. Assume that this system is in $rref$.
(1) If $\mathrm{rank}(A) < \mathrm{rank}[A \mid \mathbf{b}]$, then $A\mathbf{x} = \mathbf{b}$ has no solution.
(2) If $\mathrm{rank}(A) = \mathrm{rank}[A \mid \mathbf{b}] = n$, then $A\mathbf{x} = \mathbf{b}$ has exactly one solution.
(3) If $\mathrm{rank}(A) = \mathrm{rank}[A \mid \mathbf{b}] < n$, then $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions

Proof.
(1) If $\mathrm{rank}(A) = r < \mathrm{rank}[A \mid \mathbf{b}]$, then $\mathbf{r}_{r+1}[A \mid \mathbf{b}] = (0, 0, ..., 0 \mid b_{r+1})$ where $b_{r+1} \neq 0$. If $x = (x_1, x_2, ..., x_n)$ is a solution to $A\mathbf{x} = \mathbf{b}$, then $0x_1 + 0x_2 + ... + 0x_n = b_{r+1}$ which is absurd.

(2) Since there are n pivots in the n-column matrix A, the Pigeon Hole Principle says that the reduced system must be of the form

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Therefore, the unique solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

(3) Suppose that $\text{rank}(A) = \text{rank}[A \mid \mathbf{b}] = r < n$. Now the form is

$$\begin{bmatrix} 1 & & & & c_{1,r+1} & c_{1,r+2} & \cdots & c_{1n} \\ & 1 & & & c_{2,r+1} & c_{2,r+2} & \cdots & c_{2n} \\ & & \ddots & & \vdots & & & \\ & & & 1 & c_{r,r+1} & c_{r,r+2} & & c_{rn} \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ x_{r+1} \\ x_{r+2} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The $n - r$ variables $x_{r+1}, x_{r+2}, ..., x_n$ are free and the r variables $x_1, x_2, ..., x_r$ are dependent. Indeed,

$$\begin{aligned} x_1 &= b_1 - c_{1,r+1}x_{r+1} - c_{1,r+2}x_{r+2} - ... - c_{1n}x_n \\ x_2 &= b_2 - c_{2,r+1}x_{r+1} - c_{2,r+2}x_{r+2} - ... - c_{2n}x_n \\ &\vdots \\ x_r &= b_r - c_{r,r+1}x_{r+1} - c_{r,r+2}x_{r+2} - ... - c_{rn}x_n \end{aligned}$$