Lecture 2: Linear Algebra: Solving $Ax = b$ via Gaussian Elimination

Lecturer: *Sundar Vishwanathan*
COMPUTER SCIENCE & ENGINEERING INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

1 Formulation

A linear optimization problem can be formulated as

$$\begin{aligned} \max\ c^T x \\ Ax \leq b \end{aligned} \tag{1}$$

where, A is an $m \times n$ matrix, c a $n \times 1$ vector, b a $m \times 1$ vector and x a $n \times 1$ vector. We are given as input c, A, and b. As output, among all x that satisfies $Ax \leq b$, we wish to find one which maximises $c^T x$.

Our first goal is to understand the set $\{x : Ax \leq b\}$. We begin with a simpler set $\{x : Ax = b\}$. Before we begin to understand what this set looks like, we recall procedures we have learnt for finding at least one solution or recognising instances which do not have solutions.

2 Solving $Ax = b$

$Ax = b$ is short hand notation for the following set of equalities given below.

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + \cdots + A_{2n}x_n &= b_2 \\ &\vdots \\ A_{m1}x_1 + A_{m2}x_2 + A_{m3}x_3 + \cdots + A_{mn}x_n &= b_m \end{aligned}$$

We can solve such a system of equations using Gaussian Elimination. Here is an example.

EXAMPLE 1

$$2x + 7y = 13 \tag{I}$$
$$x + 3y = 4 \tag{II}$$

Replacing II by $-\frac{1}{2}\cdot$I + II gives

$$\begin{aligned} -\frac{7}{2}y + 3y &= -\frac{13}{2} + 4 \\ y &= 5 \end{aligned} \tag{2}$$

Now one can solve for x using the first equation.

3 Gaussian Elimination

Gaussian Elimination starts with a set of equations. There are two phases. Both phases are iterative. We discuss the first phase which is the crux. In each step of the first phase these set of equations are replaced with an equivalent set of equations. By equivalent, we mean that the solution sets of the two sets are the same. We also make sure that the new set is simpler to analyse.

Consider the set of equations.

$$\begin{aligned}
A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1n}x_n &= b_1 \\
A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + \cdots + A_{2n}x_n &= b_2 \\
&\vdots \\
A_{m1}x_1 + A_{m2}x_2 + A_{m3}x_3 + \cdots + A_{mn}x_n &= b_m
\end{aligned}$$

Assume that $A_{11} \neq 0$. If this is not so, we exchange the first row with some other row which has a non-zero first co-ordinate. Our first goal is to get an equivalent set of equations such that except for the first row, all others have their first co-ordinate zero. To do this, for each of the equations except the first, multiply the first equation by a suitable constant and subtract this from the respective equation.

Once this is done, the first co-ordinate of every equation except the first is zero. We will now ignore the first equation and the first variable and work with the other equations.

In this process of transforming the set of equations two operations are used.

1. Exchanging two rows.
2. Replacing row_j with $\alpha \cdot row_i + row_j$, where α is some constant.

It is possible that while using these two operations the co-efficients of some x_i [for example x_2] in all rows except one are zeroed out.

$$\begin{array}{lr}
A_{11}x_1 + A_{12}x_2 & +A_{13}x_3 + \cdots + A_{1n}x_n = b_1 \\
0 + 0 & +A_{23}x_3 + \cdots + A_{2n}x_n = b_2' \\
0 + 0 & +A_{33}x_3 + \cdots + A_{3n}x_n = b_3' \\
 & \vdots \\
0 + 0 & +A_{m3}x_3 + \cdots + A_{mn}x_n = b_m'
\end{array}$$

In such a case, we identify some $A_{ij} \neq 0$ with smallest j [for example $A_{23} \neq 0$] and repeat the process.

Finally, the set of equations look like:

$$\begin{array}{lr}
A_{11}x_1 + A_{12}x_2 & + \cdots + A_{1n}x_n = b_1 \\
0 + 0 & + \cdots + A_{2i_1}x_{i_1} + \cdots + A_{2n}x_n = b_2'' \\
 & \vdots \\
0 + 0 & + \cdots + 0 + A_{ki_{k-1}}x_{i_{k-1}} + \cdots + A_{kn}x_n = b_k'' \\
 & \vdots \\
0 + 0 & + \cdots + 0 + 0 + 0 + \cdots + 0 = b_l'' \\
\vdots & \\
0 + 0 & + \cdots + 0 + 0 + 0 + \cdots + 0 = b_m''
\end{array}$$

Observations: Once we are done with the first phase we note:

1. All zero rows occur after the non-zero rows.
2. If, from the top, the first t rows are nonzero, and the first non-zero entry of the ith row $i = 1, \ldots, t$, is at the k_ith column, then $k_i > k_{i-1}$, $i = 2, \ldots, t$. That is, the first non-zero entries in the rows appear later and later from top to bottom.

3.1 Existence of a solution for $Ax = b$

$Ax = b$ does not have a solution if in a particular row i, all coefficients $A_{ij} = 0$, but $b_i \neq 0$. This is both necessary and sufficient condition for non-existence of a solution for $Ax = b$ as we shall see later.

Once we have the matrix in this form, it is easy to get solutions to the set of equations, if one exists. This is the second phase. Except for $x_1, x_{i_1}, \ldots, x_{i_{k-1}}$ set any values to the other variables. Now solve for the variables $x_1, x_{i_1}, \ldots, x_{i_{k-1}}$ in the reverse order.

Clearly, if it has n non-zero rows then the system has only one solution. Otherwise $\{x : Ax = b\}$, in general, has many solutions.

3.2 Why is this procedure correct?

The procedure being correct means that the values of the variables obtained by the procedure indeed satisfy the original set of equations. So to prove the correctness of the procedure, we have to prove that the solution set does not change on applying the operations of Gaussian Elimination. Clearly exchanging two rows does not change the solution set.

THEOREM 1 *Given a set of equations, suppose A_j is replaced by $\alpha \cdot A_i + A_j$ then the solution set does not change.*

PROOF: Let A_i be $A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n = b_i$ and A_j be $A_{j1}x_1 + A_{j2}x_2 + \cdots + A_{jn}x_n = b_j$. Consider the old and the new set of equations. Note that they differ only in the jth equation. This proof consists of 2 parts.

Part 1: *If the set of solutions satisfies original set of equations then it satisfies the new set.* Let $(x'_1, x'_2, \ldots, x'_n)$ be the solution to the original set. Then $(x'_1, x'_2, \ldots, x'_n)$ specifically satisfies A_i and A_j. So,

$$A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n = b_i$$
$$A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n = b_j$$

that is,

$$A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n - b_i = 0$$
$$A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n - b_j = 0$$

So,

$$\begin{aligned} &\alpha(A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n - b_i) + (A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n - b_j) \\ &= \alpha \cdot 0 + 0 \\ &= 0 \end{aligned}$$

Therefore,

$$\alpha(A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n) + (A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n) = \alpha \cdot b_i + b_j$$

Hence $< x'_1, x'_2, \ldots, x'_n >$ satisfies $\alpha \cdot A_i + A_j$. Since the other equations in both the sets are identical, $< x'_1, x'_2, \ldots, x'_n >$ satisfies the new set.

Part 2: *If the set of solutions satisfies new set of equations then it satisfies the original set.* Let $< x'_1, x'_2, \ldots, x_n >$ be the solution to the new set. Then $< x'_1, x'_2, \ldots, x'_n >$ specifically satisfies $\alpha \cdot A_i + A_j$ and A_i. So,

$$\alpha(A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n) + (A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n) = \alpha \cdot b_i + b_j$$
$$A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n = b_i$$

that is,

$$\begin{aligned} \alpha(A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n - b_i) + (A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n - b_j) &= 0 \\ A_{i1}x'_1 + A_{i2}x'_2 + \cdots + A_{in}x'_n - b_i &= 0 \end{aligned}$$

So,

$$\begin{aligned} \alpha \cdot 0 + (A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n - b_j) &= 0 \\ \Rightarrow A_{j1}x'_1 + A_{j2}x'_2 + \cdots + A_{jn}x'_n &= b_j \end{aligned}$$

Hence $< x'_1, x'_2, \ldots, x'_n >$ satisfies A_j. □

4 Understanding $Ax = b$ geometrically

Another way of looking at $Ax = b$ is through geometry. The operation of adding a constant times another equation to an equation rotates one of the hyperplanes in $\mathbb{R}^n$ about the region of intersection. Fig. **??** illustrates this as the case of rotation of lines (1-dimensional hyperplane) in a x-y plane (i.e. $\mathbb{R}^2$). Here the solid lines are from example 1 and one of the dotted lines is obtained by rotating the line corresponding to eq. II to the one corresponding to eq. 2. The other dotted line can be obtained by similarly manipulating eq. I lecture24inGeometrical way of looking at $Ax = b$lec2:fig1

Then Gaussian Elimination can be described as rotating the hyperplanes about the region of intersection such that at each step, we make one of the hyperplanes parallel to one axis. We end with a question. Why does adding a multiple of one equality to a second rotate the second?