GENG 8010–Part 2 - Elements of Applied Linear Algebra

Mehrdad Saif ©

University of Windsor

Winter 2023

Solutions of $Ax = y$– I

Consider

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= y_2 \\ \vdots \qquad \vdots \qquad \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= y_m \end{aligned}$$

where $a_{ij} \in \mathscr{F}$ and $y_i \in \mathscr{F}$, also $x_i \in \mathscr{F}$. The above can be written as

$$\boldsymbol{Ax} = \boldsymbol{y} \qquad \begin{matrix} \boldsymbol{A} : mxn \\ \boldsymbol{x} : nx1 \\ \boldsymbol{y} : mx1 \end{matrix}$$

Solutions of $Ax = y$– II

Questions:

1. **Existence**– Condition on $\boldsymbol{A}$ and $\boldsymbol{y}, \ni \boldsymbol{x}$ exits.
2. **Number of solutions**– Number of linearly independent $\boldsymbol{x}$'s $\ni \boldsymbol{Ax} = \boldsymbol{y}$.

Range

The **range** of a linear operator $\boldsymbol{A}$ is the set $\mathscr{R}(\boldsymbol{A})$ defined as:

$$\mathscr{R}(\boldsymbol{A}) = \{\text{all } \boldsymbol{y} \in (\mathscr{F}^m, \mathscr{F}) \text{ for which } \exists \text{ at least one } \boldsymbol{x} \in (\mathscr{F}^n, \mathscr{F}) \ni \boldsymbol{Ax} = \boldsymbol{y}\}$$

Solutions of $Ax = y$– III

Theorem 3.1.1

$\mathscr{R}(\boldsymbol{A})$ is a subspace of $(\mathscr{F}^m, \mathscr{F})$

Solutions of $Ax = y$– IV

Consider now the following definition.

Rank

Rank of a matrix $\boldsymbol{A}$, represented by $\rho(\boldsymbol{A})$, is the maximum number of linearly independent columns in $\boldsymbol{A}$ or equal to the dimension of the range space of $\boldsymbol{A}$, i.e. $\rho(\boldsymbol{A}) = \dim \mathscr{R}(\boldsymbol{A})$. Hence, for an nxm matrix

$$\rho(\boldsymbol{A}) \leq \min(n, m)$$

Theorem 3.1.2

A square matrix is **nonsingular** *or has a non-zero determinant, iff it is of full rank.*

Solutions of $Ax = y$– V

Example–consider $\boldsymbol{A}$ where

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

if you calculate $\det \boldsymbol{A}$ you will find that it is zero. Hence this matrix is **singular**.

$$\boldsymbol{A} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & -2 & 2 \\ 0 & -2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore we conclude $\rho(\boldsymbol{A}) = 2$ which also implies that the matrix has a zero determinant and is singular.

Solutions of $Ax = y$– VI

Null space and nullity

The **null (or kernel) space** of a linear operator $\boldsymbol{A}$ is the set $\mathscr{N}(\boldsymbol{A})$ defined as

$$\mathscr{N}(\boldsymbol{A}) = \{\text{all } \boldsymbol{x} \in (\mathscr{F}^n, \mathscr{F}), \ni \boldsymbol{Ax} = \boldsymbol{0}\}$$

The dimension of $\mathscr{N}(\boldsymbol{A})$ is called the **nullity** of $\boldsymbol{A}$ and is denoted by $\gamma(\boldsymbol{A}) = \dim \mathscr{N}(\boldsymbol{A})$.

Remark

If $\gamma(\boldsymbol{A}) = 0$, then the only solution to $\boldsymbol{Ax} = \boldsymbol{0}$ is $\boldsymbol{x} = \boldsymbol{0}$. If on the other hand $\gamma(\boldsymbol{A}) = k$, then there are k linearly independent solution to $\boldsymbol{Ax} = \boldsymbol{0}$ is $\boldsymbol{x} = \boldsymbol{0}$.

Solutions of $Ax = y$– VII

Lemma 3.1.3 (Fundamental Lemma)

Let $\boldsymbol{A}$ be an nxn matrix. Then the homogeneous equation

$$\boldsymbol{Ax} = \boldsymbol{0}$$

has a nonzero solution iff $\boldsymbol{A}$ is singular.

Example–Find $\rho, \mathscr{N}, \gamma$ for $\boldsymbol{A}$.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad \text{verify } \rho(\boldsymbol{A}) = 2$$

$$\boldsymbol{Ax} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Longrightarrow x_1 = 0 \,;\; x_2 = 0$$

and $\gamma(\boldsymbol{A}) = 0$, so the null space only contains the zero vector.

Solutions of $Ax = y$– VIII

Example–Consider $\boldsymbol{A} : (\Re^5, \Re) \longrightarrow (\Re^3, \Re)$ where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & -1 & 2 \\ -1 & 3 & 2 & -4 & 6 \end{bmatrix}$$

Verify that $\rho(\boldsymbol{A}) = 2$ and that the first two columns of $\boldsymbol{A}$ are linearly independent. Note

$$Row3 = -Row1 + 3Row2$$

As a result of the above

$$\begin{aligned} x_1 + x_3 + x_4 &= 0 \\ x_2 + x_3 - x_4 + 2x_5 &= 0 \end{aligned}$$

Solutions of $Ax = y$– IX

a set of solution is

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{pmatrix}$$

The above vectors are linearly independent and every $\boldsymbol{x}$ for which $\boldsymbol{Ax} = \boldsymbol{0}$ must be a linear combination of the above vectors. Therefore, $\gamma(\boldsymbol{A}) = 3$.

Solutions of $Ax = y$– X

Theorem 3.1.4

Let $\boldsymbol{A}$ be an mxn matrix, then $\rho(\boldsymbol{A}) + \gamma(\boldsymbol{A}) = n$.

Corollary 3.1.5

The number of linearly independent solutions of $\boldsymbol{Ax} = \boldsymbol{0}$ is equal to $n - \rho(\boldsymbol{A})$, where n is the number of columns in $\boldsymbol{A}$.

Solutions of $Ax = y$– XI

Theorem 3.1.6 (Solutions of $\boldsymbol{Ax} = \boldsymbol{y}$)

Consider the set of algebraic equation described by

$$\boldsymbol{Ax} = \boldsymbol{y} \qquad \boldsymbol{A} \in \Re^{mxn}; \boldsymbol{x} \in \Re^n, \boldsymbol{y} \in \Re^m$$

Let $\boldsymbol{x}_s$ be a solution and suppose $\gamma(\boldsymbol{A}) = n - \rho(\boldsymbol{A})$ is the nullity of $\boldsymbol{A}$. If $\gamma = 0$, then $\boldsymbol{x}_s$ is unique. If however $\gamma > 0$, then $\forall \alpha_i \in \mathscr{F}$

$$\boldsymbol{x} = \boldsymbol{x}_s + \alpha_1 \boldsymbol{n}_1 + \alpha_2 \boldsymbol{n}_2 + \cdots + \alpha_\gamma \boldsymbol{n}_\gamma$$

represent solutions where $\boldsymbol{n}_i \in \mathscr{N}(\boldsymbol{A}), \quad i = 1, 2, \cdots, \gamma$

Proof.

$$\boldsymbol{Ax} = \boldsymbol{Ax}_s + \cancelto{0}{\sum_{i=1}^{\gamma} \alpha_i \boldsymbol{An}_i} = \boldsymbol{y}$$

Solutions of $Ax = y$– XII

Example 3.1.7

Consider

$$\begin{pmatrix} 1 & 2 & -2 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 1 & -1 & -2 \end{pmatrix} \boldsymbol{x} = \begin{pmatrix} 5 \\ -2 \\ 2 \\ 1 \end{pmatrix}$$

verify the followings:

$$\rho(\boldsymbol{A}) = 3 \Longrightarrow \gamma(\boldsymbol{A}) = 4 - \rho(\boldsymbol{A}) = 1 \text{ and } \boldsymbol{n} = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} \in \mathscr{N}(\boldsymbol{A})$$

Solutions of $Ax = y$– XIII

Furthermore,

$$\boldsymbol{x}_s = \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \end{pmatrix}$$

is a solution. Hence, the above has infinitely many solutions and they are given by

$$\boxed{\boldsymbol{x} = \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix}} \quad \text{where } \alpha \in \Re$$

Left and right inverses– I

Left Inverse

A matrix $\boldsymbol{A} \in \Re^{mxn}$ that satisfies $\boldsymbol{A}_L\boldsymbol{A} = \boldsymbol{I}$ is called the **left inverse** of $\boldsymbol{A}$, and $\boldsymbol{A}$ is said to be **left invertible**. Note that the left inverse in this case will be of dimension nxm.

Remark

The left inverse of a matrix does not have to be unique, and a matrix that has more than one left inverse, will have infinitely many.

Left and right inverses– II

Example–consider

$$\boldsymbol{A} = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix}$$

The following two matrices are both left inverses of $\boldsymbol{A}$

$$\boldsymbol{A}_L^1 = \frac{1}{2}\begin{bmatrix} 0 & -1 & 6 \\ 0 & 1 & -4 \end{bmatrix} \qquad \boldsymbol{A}_L^2 = \frac{1}{3}\begin{bmatrix} -11 & -7 & -2 \\ 7 & 5 & 1 \end{bmatrix}$$

In both of the above cases

$$\boldsymbol{A}_L^{1,2}\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Left and right inverses– III

Recall that $\boldsymbol{A}\boldsymbol{x} = 0$ has a trivial solution, i.e. $\boldsymbol{x} = 0$ if its nullity, $\gamma(\boldsymbol{A}) = 0$ in which case that implies that $\boldsymbol{A}$ is full rank. Also knowing that $\boldsymbol{A}_L\boldsymbol{A} = \boldsymbol{I}$ means

$$\boldsymbol{A}_L\boldsymbol{A}\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = \boldsymbol{x} = \boldsymbol{0}$$

which means that given $\boldsymbol{A}_L$ exists, $\boldsymbol{A}$ would be full rank and that columns of $\boldsymbol{A}$ are linearly independent.

Remark

1. If matrix $\boldsymbol{A}$ has a left inverse, then the columns of $\boldsymbol{A}$ are linearly independent.
2. The above implies that *a matrix with a left inverse is either square or tall.*

Left and right inverses– IV

Now going back to the problem of finding the solution of

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

<u>If $\boldsymbol{A}$ is square, then</u>

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$$

and we have a unique solution for $\boldsymbol{x}$. However,

<u>If $\boldsymbol{A}$ is tall, then</u>

$$\boldsymbol{A}_L\boldsymbol{y} = \boldsymbol{A}_L(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{I}\boldsymbol{x} = \boldsymbol{x} \qquad \therefore \boxed{\boldsymbol{x} = \boldsymbol{A}_L\boldsymbol{y}}$$

Left and right inverses– V

<u>**Example**</u>–Find the solution $\boldsymbol{x}$ for

$$\begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}$$

from the previous example, recall that both of the following are the left inverses of $\boldsymbol{A}$

$$\boldsymbol{A}_L^1 = \frac{1}{2}\begin{bmatrix} 0 & -1 & 6 \\ 0 & 1 & -4 \end{bmatrix} \qquad \boldsymbol{A}_L^2 = \frac{1}{3}\begin{bmatrix} -11 & -7 & -2 \\ 7 & 5 & 1 \end{bmatrix}$$

you can verify that

$$\boxed{\boldsymbol{x} = \boldsymbol{A}_L^1\boldsymbol{y} = \boldsymbol{A}_L^2\boldsymbol{y} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}}$$

Left and right inverses– VI

Right Inverse

A matrix $\boldsymbol{A} \in \Re^{mxn}$ that satisfies $\boldsymbol{A}\boldsymbol{A}_R = \boldsymbol{I}$ is called the **right inverse** of $\boldsymbol{A}$, and $\boldsymbol{A}$ is said to be **right invertible**. Note that the right inverse in this case will be of dimension nxm.

Left and right inverses– VII

Remark

1. Consider

$$AA_R = I \implies (AA_R)^T = I \implies \underbrace{A_R^T}_{B_L}\ \underbrace{A^T}_{B} = I$$

   From the above it can be concluded that A is right invertible, then A^T must be left invertible.

2. From the above and previous discussion on left invertability of a matrix, we conclude that A is invertible if its rows are linearly independent.
3. From (2), we conclude that right invertible matrices are square or wide.

Left and right inverses– VIII

Example–Find the solution x for

$$\begin{bmatrix} -3 & 4 & 1 \\ -4 & 6 & 1 \end{bmatrix} x = \begin{bmatrix} -6 \\ -9 \end{bmatrix}$$

From the last example, we can readily write down

$$A_R^1 = \frac{1}{2}\begin{bmatrix} 0 & 0 \\ -1 & 1 \\ 6 & -4 \end{bmatrix} \qquad A_R^2 = \frac{1}{3}\begin{bmatrix} -11 & 7 \\ -7 & 5 \\ -2 & 1 \end{bmatrix}$$

$$\boxed{x = \frac{1}{2}\begin{bmatrix} 0 & 0 \\ -1 & 1 \\ 6 & -4 \end{bmatrix}\begin{bmatrix} -6 \\ -9 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 \\ -3 \\ 0 \end{bmatrix}} \; ; \; \boxed{x = \frac{1}{3}\begin{bmatrix} -11 & 7 \\ -7 & 5 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} -6 \\ -9 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}}$$

2 equations and 3 unknowns $\implies$ infinitely many other solutions.

Left and right inverses– IX

Remark

1. If A has both left and right inverse the matrix must be square.
2. For a square matrix the following statements are equivalent (i.e. if any hold, others hold)
   - A is invertible.
   - Columns of A are linearly independent.
   - Rows of A are linearly independent.
   - A has a left and right inverse and they are equal.

Least square solution of $Ax = y$ – I

Consider again the equation

$$Ax = y \qquad A \in \Re^{mxn}$$

Let us now try to determine the *"best"* solution to an **inconsistent set** as described above. To do this, let us define the *"best"* solution in the *least square* sense. That is, let us try to solve the following optimization problem

$$\begin{aligned} min_x J = \|Ax - y\|^2 &= (Ax - y)^T(Ax - y) \\ &= (x^T A^T - y^T)(Ax - y) \\ &= x^T A^T A x - x^T A^T y - y^T A x + y^T y \end{aligned}$$

to find the minimum of J take $\frac{\partial}{\partial x}$ and set equal to zero

Least square solution of $Ax = y$ – II

$$x^T A^T A + x^T A^T A - y^T A - y^T A = 0 \Longrightarrow A^T A x = A^T y$$

$$\boxed{x^o = (A^T A)^{-1} A^T y}$$

Where it is assumed that A is full rank, and hence, $A^T A$ is invertible.

Remark

1. Note that the nxm matrix $A_L = (A^T A)^{-1} A^T$ is the left inverse of A, since $A_L A = I$.
2. The least square solution to $Ax = y$ when A is tall is $x^o = (A^T A)^{-1} A^T y = A_L y$.

State estimation– I

Consider the measurement equation

$$z = Hx + v \qquad x \in \Re^n, z \in \Re^m, m \gg n, v \quad \text{is noise or error}$$

The objective is to obtain an optimum estimate of x, i.e., $\hat{x}$.

Note that the actual measurement should be $y = Hx$, however due to measurement errors or noise in the sensor collecting the data, at each sampling interval we obtain z which is corrupted by the error/noise. Hence we want

$$\hat{y} = H\hat{x}$$

State estimation– II

Define the estimation error as

$$e = z - \hat{y}$$

Obviously, the objective should be to obtain the best estimate of $\hat{x}$ so that the estimation error is minimized in some sense. Therefore define a quadratic cost function as

$$\begin{aligned} J &= \frac{1}{2} e^T e = \frac{1}{2}(z - H\hat{x})^T (z - H\hat{x}) \\ &= z^T z - z^T H\hat{x} - \hat{x}^T H^T z + \hat{x}^T H^T H\hat{x} \end{aligned}$$

The objective is now to obtain the optimal estimate $\hat{x}$ which would minimize the cost function J, and hence, minimizes the estimation error e.

State estimation– III

By taking $\frac{\partial J}{\partial \hat{x}}$ and setting it to zero, we obtain

$$\frac{\partial J}{\partial \hat{x}} = \frac{1}{2}\left(-2z^T H + 2\hat{x}^T H^T H\right) = 0 \Longrightarrow H^T z = H^T H\hat{x}$$

$$\boxed{\hat{x} = (H^T H)^{-1} H^T z}$$

and this gives

$$\boxed{\hat{y} = H(H^T H)^{-1} H^T z}$$

Remark

An estimate can be obtained with $m = n$ measurements. However, for a better estimate and for averaging to occur, we would ideally need a larger sample of measurement and $m \gg n$. Therefore, we wish for H to be tall rather than square.

State estimation– IV

Example–

The above has many useful applications. To illustrate one, suppose we have collected a bunch of data from an experiment and wish to fit a cubic polynomial to the data, i.e.,

$$y(t) = \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \alpha_4 t^3$$

This problem can be cast into the formulation. Assuming that our data is contained in a vector $\boldsymbol{z}$, we reformulate the problem as a least square problem just discussed.

State estimation– V

$$\boldsymbol{z} = \begin{bmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ 1 & t_3 & t_3^2 & t_3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 & t_m^3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix}$$

It is now possible to use least square estimation to find the optimum $\boldsymbol{\alpha}$.

In our case, consider the set of data given in the table

State estimation– VI

Time	z(t)	Time	z(t)
0	-226.1	11	366.6
1	27.17	12	328.2
2	227.6	13	313.8
3	330.7	14	323.2
4	430.8	15	326.9
5	494.1	16	358.1
6	519.2	17	460.2
7	518.9	18	587.7
8	453.4	19	797
9	424.5	20	1007
10	423.9		

Table 1: A set of noisy data

State estimation– VII

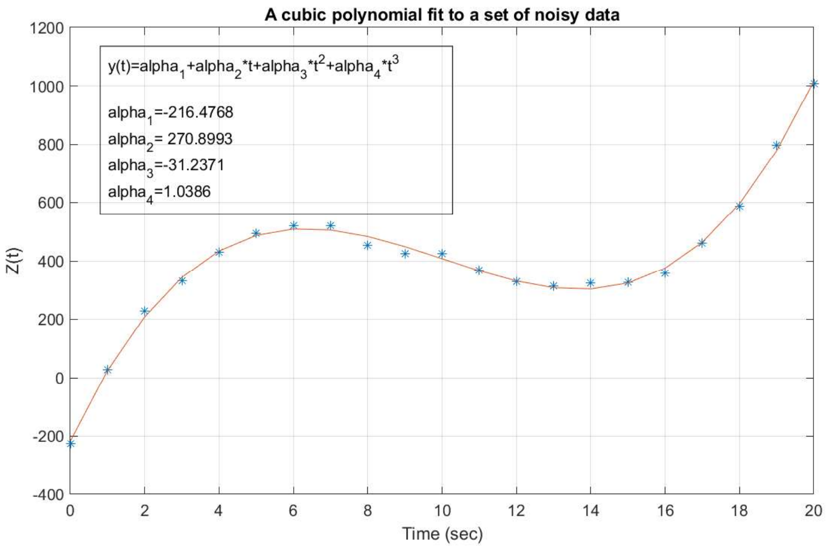
The data ranges from $t = 0$ to $t = 20$ seconds and has been sampled each second for a total of 21 data points. Therefore

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 20 & 400 & 8000 \end{bmatrix}_{21x4}$$

Now

$$\boldsymbol{\alpha} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{z} = \begin{bmatrix} -216.4768 \\ 270.8993 \\ -31.2371 \\ 1.0386 \end{bmatrix}$$

State estimation– VIII



Example–Now consider fitting a cubic polynomial to the following set of data

Time	z(t)	Time	z(t)
0	0	11	0.8085
1	0.1987	12	0.6755
2	0.3894	13	0.5155
3	0.5646	14	0.3350
4	0.7174	15	0.1411
5	0.8415	16	-0.0584
6	0.9320	17	-0.2555
7	0.9854	18	-0.4425
8	0.9996	19	-0.6119
9	0.9738	20	-0.7568
10	0.9093		

Table 2: A set of data

Again,

$$\boldsymbol{\alpha} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{z} = \begin{bmatrix} -0.0873 \\ 0.3127 \\ -0.0262 \\ 0.0004 \end{bmatrix}$$