MODULE 8

Topics: Null space, range, column space, row space and rank of a matrix

Definition: Let $L : V_1 \to V_2$ be a linear operator. The null space $\mathcal{N}(L)$ of L is the subspace of V_1 defined by

$$\mathcal{N}(L) = \{x \in V_1 : Lx = 0\}$$

Note: The null space of L is sometimes called the kernel of L.

Examples:

i) $Lx \equiv Ax \equiv \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x = 0$ then $\mathcal{N}(A) = \text{span}\{(1, -1)\} \in R_2$.

ii) Lf defined by $(Lf)(t) \equiv f''(t)$ for $f \in C^2[a, b]$ then $\mathcal{N}(L) = \text{span}\{1, t\}$.

iii) $L : C^0[-1, 1] \to R$ defined by

$$Lf \equiv \int_{-1}^{1} f(s)ds$$

then $\mathcal{N}(L)$ contains the subspace of all odd continuous functions on $[-1, 1]$ plus many other functions such as $f(t) = t^2 - 1/3$.

We shall now restrict ourselves to $m \times n$ real matrices. We note that always $0 \in \mathcal{N}(A)$. If this is the only vector in $\mathcal{N}(A)$, i.e., if $\mathcal{N}(A) = \{0\}$ then the null space is the trivial null space with dimension 0.

We also know from

$$Ax = \sum_{j=1}^{n} x_j A_j$$

that $R(A) = \text{span}\{A_1, \ldots, A_n\} \in R_m$. The range of A is often called the column space of A and the dimension of this space is called the rank of A, i.e.,

$$r(A) = \text{rank}(A) = \dim R(A) = \dim \text{column space of } A.$$

We note that $r(A) < \min\{m, n\}$.

Example: Let x and y be two column vectors in R_n. Then the $n \times n$ matrix

$$x \bullet y^T = (y_1\vec{x} \bullet y_2\vec{x} \cdots y_n\vec{x})$$

is a matrix with rank 1 since every column is a multiple of $\vec{x}$.

Theorem: *Let A be an $m \times n$ matrix. Then*

$$\dim \mathcal{N}(A) + \text{rank}(A) = n.$$

Proof: Let $\{y_1, \ldots, y_r\}$ be a basis of $R(A)$. Let $\{x_1, \ldots, x_r\}$ be the vectors which satisfy

$$Ax_j = y_j \qquad \text{for } j = 1, \ldots, r.$$

Let $\{z_1, \ldots, z_p\}$ be a basis of $\mathcal{N}(A)$. Then the vectors $\{x_1, \ldots, x_j,\ z_1, \ldots, z_p\}$ are linearly independent because if

$$\sum_{j=1}^{r} \alpha_j x_j + \sum_{j=1}^{p} \beta_j z_j = 0$$

then

$$A \left(\sum_{j=1}^{r} \alpha_j x_j + \sum_{j=1}^{p} \beta_j z_j \right) = \sum_{j=1}^{r} \alpha_j y_j = 0$$

which implies that $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$. But then the linear independence of the $\{z_j\}$ implies that the $\{\beta_j\}$ also must vanish. Finally, let x be arbitrary in R_n. Then $Ax = \sum_{j=1}^{r} \gamma_j y_j$ for some $\{\gamma_j\}$. This implies that $A\left(x - \sum_{j=1}^{r} \gamma_j x_j\right) = 0$ so that $x - \sum_{j=1}^{r} \gamma_j x_j \in \mathcal{N}(A)$, i.e.,

$$x - \sum_{j=1}^{r} \gamma_j x_j = \sum_{j=1}^{p} \beta_j z_j.$$

Hence the linearly independent vectors $\{x_1, \ldots, x_r,\ z_1, \ldots, z_p\}$ span R_n and

$$r + p \equiv \text{rank}(A) + \dim \mathcal{N}(A) = n.$$

It follows immediately that if A is an $m \times n$ matrix and $m < n$ then $\dim \mathcal{N}(A) \geq 1$ because $\text{rank}(A) \leq \min\{m, n\}$. In particular, this implies that $Ax = 0$ has a non-zero solution so that such a matrix cannot have an inverse.

So far we have looked at the columns of A as n column vectors in R_m. Likewise, the m rows of A define a set of m vectors in R_n. What can we say about the number of linearly independent rows of A?

We recall from the homework of Module 2 that if $\langle x, y \rangle$ denotes the dot product then

$$\langle Ax, y \rangle = \langle x, A^T y \rangle$$

for $x \in R_n$ and $y \in R_m$. Next, let $\{y_1, y_2, \ldots, y_r\}$ be a basis of $R(A)$ and apply the Gram-Schmidt orthogonalization process to the vectors

$$\{y_1, y_2, \ldots, y_r, \hat{e}_1, \hat{e}_2, \ldots, \hat{e}_m\}$$

then the first r orthogonal vectors will be a basis of $R(A)$ and the remaining $m-r$ vectors $\{Y_1, Y_2, \ldots, Y_{m-r}\}$ will be orthogonal to $R(A)$. Since $A^T Y_j \in R_n$ it it follows from

$$\langle A^T Y_j, A^T Y_j \rangle = \langle Y_j, A(A^T Y_j) \rangle = 0$$

that $A^T Y_j = 0$ so that dim $\mathcal{N}(A^T) \geq (m-r)$. Finally, we observe that if $Ax \neq 0$ then $\langle A^T(Ax), x \rangle > 0$ so that Ax cannot belong to $\mathcal{N}(A^T)$. Hence dim $\mathcal{N}(A^T) = m - r$ so that rank(A^T) = number of linearly independent rows of $A = m - (m-r) = r$. In other words, an $m \times n$ matrix has as many independent rows as columns.

Finally, we observe that if we add to any row of A a linear combination of the remaining rows we do not change the number of independent rows. Hence we can apply Gaussian elimination to the rows of A and read off the number of independent rows of A from the final form of A where all elements below the diagonal are zero.

Implications for the solution of the linear system

$$Ax = b$$

where A is an $m \times n$ matrix.

1) We shall assume that $b \in R(A)$.
    i) If the columns of A are linearly independent then $Ax = b$ has a unique solution regardless of the size of the system. In this case the inverse mapping exists for every element $y \in R(A)$.
    ii) If the columns of A are linearly dependent then dim $\mathcal{N}(A) \geq 1$ and there are infinitely many solutions. One can then constrain the solution by asking, for example, for the minimum norm solution.
    iii) If $m \geq n$ the columns of A may or may not be linearly dependent. If $m < n$ then the columns of A must be linearly dependent

iv) If $\text{rank}(A) = m$ then $b \in R(A)$.

2) Regardless of the size of the system, if $b \notin R(A)$ there cannot be a solution. If $b \notin R(A)$ then Gaussian elimination will lead to inconsistent equations.

Two points of view for finding an approximate solution of $Ax = b$ when $b \notin R(A)$.

I. The "Least Squares Solution":

When the system $Ax = b$ is inconsistent then for any $x \in R_n$ the residual, defined as

$$r(x) \equiv b - Ax,$$

cannot be zero. In this case it is common to try to minimize the residual (in some sense) over all $x \in R_n$ (or possibly over some specially chosen set of "admissible" $x \in R_n$). We shall consider here only the case of minimizing a norm of the residual which is obtained from an inner product. This means we need to find the minimum of the function f defined by

$$f(x) \equiv \langle r(x), r(x) \rangle = \langle b - Ax, b - Ax \rangle.$$

Let us assume now that we are dealing with real valued vectors. Then f is a function of the n real variables $x_1, \ldots, x_n$, and calculus tells us that a necessary condition for the minimum is that

$$\nabla f(x) = 0.$$

We find that

$$\frac{\partial f}{\partial x_j} \equiv \langle -A_j, b - Ax \rangle + \langle b - Ax, A_j \rangle = 0.$$

Since in a real vector space the inner product is symmetric it follows that x must be a solution of

$$\langle A_j, Ax \rangle = \langle A_j, b \rangle \qquad \text{for } j = 1, \ldots, n.$$

If the inner product is the dot product on R_n then these n equations can be written in matrix form as

$$A^T Ax = A^T b$$

If the $n \times n$ matrix $A^T A$ has rank n then $\dim \mathcal{N}(A^T A) = 0$ and $(A^T A)^{-1}$ exists so that

$$x = (A^T A)^{-1} Ab.$$

This is the least squares solution of $Ax = b$ in Euclidean n-space. If A and hence A^T are square and have rank n then A^T is invertible and x solves $Ax = b$.

II. We know that we can solve $Ax = b'$ for any $b' \in R(A)$ since Gaussian elimination will give the answer. One may now pose the problem:

Find the solution x of $Ax = b'$ where b' is the vector in $R(A)$ which is "closest" in norm to b. As we saw in module 4 the vector b' is the orthogonal projection of b onto $\text{span}\{A_1, \ldots, A_n\}$. Thus

$$b' = \sum_{j=1}^{n} \alpha_j A_j = A\vec{\alpha}$$

where $\vec{\alpha}$ is computed from

$$\mathcal{A}\vec{\alpha} = d$$

with $\mathcal{A}_{ij} = \langle A_j, A_i \rangle$ and $d_i = \langle b, A_i \rangle$. It follows that $\mathcal{A}$ and d can be written in matrix notation as

$$\mathcal{A} = A^T A, \qquad d = A^T b$$

so that by inspection the solution of

$$Ax = b' = A\alpha = A(A^T A)^{-1} A^T b$$

is

$$x = (A^T A)^{-1} A^T b$$

provided A has rank n. Hence the least squares solution is the exact solution of the "closest" linear system for which there is an exact solution.

Module 8 - Homework

1) Let $V_1 = \{u : u \in C^0[-1,1]\}$

   $V_2 = C^0[-1,1]$

   Define
$$(Lu)(t) = \int_{-1}^{t} su(s)ds.$$
   Show that L is linear and find $\mathcal{N}(L)$. Show that the range of L is not all of V_2.

2) Let
$$A = \begin{pmatrix} 1 & 5 & 9 & 13 & 6 \\ 2 & 6 & 10 & 14 & 8 \\ 3 & 7 & 11 & 15 & 10 \\ 4 & 8 & 12 & 16 & 12 \end{pmatrix}.$$
   What is the rank of A?

   Find an orthogonal (with respect to the dot product) basis of the null space and range of A.

3) Let A be an $m \times n$ matrix. Assume that its columns are linearly independent.
   i) Show that in this case $n \leq m$.
   ii) Show that one can find an $n \times m$ matrix B such that
$$BA = I_n \qquad \text{where } I_n \text{ is the } n \times n \text{ identity matrix.}$$

4) Suppose the cost $C(t)$ of a process grows quadratically with time, i.e.,
$$C(t) = a_0 + a_1 t + a_2 t^2$$
   Company records contain the following data:

time taken	measured cost
.1	.911
.2	.84
.3	.788
.4	.76
.5	.747
.6	.77

   What would be your estimate of the cost of the process if it takes one unit of time?