Module 4 : Solving Linear Algebraic Equations

Section 9 : Appendix A: Behavior of Solutions of Linear Difference Equations

9 Appendix A: Behavior of Solutions of Linear Difference Equations

Consider difference equation of the form

$$\mathbf{z}^{(k+1)} = \mathbf{B}\mathbf{z}^{(k)} \quad \text{--------(178)}$$

where $\mathbf{z} \in \mathbf{R}^n$ and $\mathbf{B}$ is a $n \times n$ matrix. Starting from an initial condition $\mathbf{z}^{(0)}$, we get a sequence of vectors $\left\{\mathbf{z}^{(0)}, \mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(k)}, \ldots\right\}$ such that

$$\mathbf{z}^{(k)} = \mathbf{B}^k \mathbf{z}^{(0)}$$

for any k. Equations of this type are frequently encountered in numerical analysis. We would like to analyze asymptotic behavior of equations of these type without solving them explicitly.

To begin with, let us consider scalar linear iteration scheme

$$z^{(k+1)} = \beta z^{(k)} \quad \text{--------(179)}$$

where $z^{(k)} \in R$ and β is a real scalar. It can be seen that

$$\mathbf{z}^{(k)} = (\beta)^k \mathbf{z}^{(0)} \rightarrow 0 \text{ as } k \rightarrow \infty \quad \text{--------(180)}$$

if and only if $|\beta| < 1$. To generalize this notation to a multidimensional case, consider equation of type (diffeq) where $\mathbf{z}^{(k)} \in R^n$. Taking motivation from the scalar case, we propose a solution to equation (diffeq) of type

$$\mathbf{z}^{(k)} = \lambda^k \mathbf{v} \quad \text{--------(181)}$$

where λ is a scalar and $\mathbf{v} \in R^n$ is a vector. Substituting equation (soln) in equation (diffeq), we get

$$\lambda^{k+1}\mathbf{v} = \mathbf{B}(\lambda^k \mathbf{v}) \quad \text{--------(182)}$$

$$\text{or } \lambda^k(\lambda I - \mathbf{B})\mathbf{v} = \overline{0} \quad \text{--------(183)}$$

Since we are interested in a non-trivial solution, the above equation can be reduced to

$$(\lambda I - \mathbf{B})\mathbf{v} = \overline{0} \quad \text{--------(184)}$$

where $\mathbf{v} \neq \overline{0}$. Note that the above set of equations has n equations in $(n+1)$ unknowns (λ and n elements of vector $\mathbf{v}$). Moreover, these equations are nonlinear. Thus, we need to generate an additional equation to be able to solve the above set exactly. Now, the above equation can hold only when the columns of matrix $(\lambda I - \mathbf{B})$ are linearly dependent and $\mathbf{v}$ belongs to null space of $(\lambda I - \mathbf{B})$. If columns of matrix $(\lambda I - \mathbf{B})$ are linearly dependent, matrix $(\lambda I - \mathbf{B})$ is singular and we have

$$\det(\lambda I - \mathbf{B}) = 0 \quad \text{--------(185)}$$

Note that equation (ceq) is nothing but the characteristic polynomial of matrix A and its roots are called eigenvalues of matrix A. For each eigenvalue λ_i we can find the corresponding eigen vector $\mathbf{v}^{(i)}$ such that

$$\mathbf{B}\mathbf{v}^{(i)} = \lambda_i \mathbf{v}^{(i)} \quad \text{--------(186)}$$

Thus, we get n fundamental solutions of the form $(\lambda_i)^k \mathbf{v}^{(i)}$ to equation (diffeq) and a general solution to equation (diffeq) can be expressed as linear combination of these fundamental solutions

$$\mathbf{z}^{(k)} = \alpha_1(\lambda_1)^k\mathbf{v}^{(1)} + \alpha_2(\lambda_2)^k\mathbf{v}^{(2)} + ...+\alpha_n(\lambda_n)^k\mathbf{v}^{(n)} \quad \text{--------(187)}$$

Now, at $k = 0$ this solution must satisfy the condition

$$\mathbf{z}^{(0)} = \alpha_1\mathbf{v}^{(1)} + (\alpha_2)^k\mathbf{v}^{(2)} + ...+\alpha_n\mathbf{v}^{(n)} \quad \text{--------(188)}$$

$$= \begin{bmatrix} \mathbf{v}^{(1)} & \mathbf{v}^{(2)} & .... & \mathbf{v}^{(n)} \end{bmatrix}\begin{bmatrix} \alpha_1 & \alpha_2 & .... & \alpha_n \end{bmatrix}^T \quad \text{--------(189)}$$

$$= \Psi\alpha \quad \text{--------(190)}$$

where Ψ is a $n \times n$ matrix with eigenvectors as columns and α is a $n \times 1$ vector of n coefficients. Let us consider the special case when the eigenvectors are linearly independent. Then, we can express α as

$$\alpha = \Psi^{-1}\mathbf{z}^{(0)} \quad \text{--------(191)}$$

Behavior of equation (sol) can be analyzed as $k \to \infty$. Contribution due to the i'th fundamental solution $(\lambda_i)^k\mathbf{v}^{(i)} \to \overline{0}$ if and only if $|\lambda_i| < 1$. Thus, $\mathbf{z}^{(k)} \to \overline{0}$ as $k \to \infty$ if and only if

$$|\lambda_i| < 1 \text{ for } i = 1,2,....n \quad \text{--------(192)}$$

If we define **spectral radius** of matrix A as

$$\rho(\mathbf{B}) = \max_i |\lambda_i| \quad \text{--------(193)}$$

then, the condition for convergence of iteration equation (diffeq) can be stated as

$$\rho(\mathbf{B}) < 1 \quad \text{--------(194)}$$

Equation (sol) can be further simplified as

$$\mathbf{z}^{(k)} = \begin{bmatrix} \mathbf{v}^{(1)} & \mathbf{v}^{(2)} & .... & \mathbf{v}^{(n)} \end{bmatrix}\begin{bmatrix} (\lambda_1)^k & 0 & ..... & 0 \\ 0 & (\lambda_2)^k & 0 & ... \\ .... & .... & ..... & .... \\ 0 & .... & 0 & (\lambda_n)^k \end{bmatrix}\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ ... \\ \alpha_n \end{bmatrix} \quad \text{--------(195)}$$

$$= \Psi\begin{bmatrix} (\lambda_1)^k & 0 & ..... & 0 \\ 0 & (\lambda_2)^k & 0 & ... \\ .... & .... & ..... & .... \\ 0 & .... & 0 & (\lambda_n)^k \end{bmatrix}\Psi^{-1}\mathbf{z}^{(0)} = \Psi(\Lambda)^k\Psi^{-1}\mathbf{z}^{(0)} \quad \text{--------(196)}$$

where Λ is the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & ..... & 0 \\ 0 & \lambda_2 & 0 & ... \\ .... & .... & ..... & .... \\ 0 & .... & 0 & \lambda_n \end{bmatrix} \quad \text{--------(197)}$$

Now, consider set of n equations

$$\mathbf{B}\mathbf{v}^{(i)} = \lambda_i\mathbf{v}^{(i)} \quad \text{for} \quad (i = 1,2,....n) \quad \text{--------(198)}$$

which can be rearranged as

$$\Psi = \begin{bmatrix} \mathbf{v}^{(1)} & \mathbf{v}^{(2)} & .... & \mathbf{v}^{(n)} \end{bmatrix}$$

$$\mathbf{B}\Psi = \begin{bmatrix} \lambda_1 & 0 & ..... & 0 \\ 0 & \lambda_2 & 0 & ... \\ .... & .... & ..... & .... \\ 0 & .... & 0 & \lambda_n \end{bmatrix} \Psi \qquad \text{--------(199 \&200)}$$

$$\text{or } \mathbf{B} = \Psi\Lambda\Psi^{-1}$$

Using above identity, it can be shown that

$$\mathbf{B}^k = \left(\Psi\Lambda\Psi^{-1}\right)^k = \Psi(\Lambda)^k\Psi^{-1} \qquad \text{--------(201)}$$

and the solution of equation (diffeq) reduces to

$$\mathbf{z}^{(k)} = \mathbf{B}^k\mathbf{z}^{(0)} \qquad \text{--------(202)}$$

and $\mathbf{z}^{(k)} \to \bar{0}$ as $k \to \infty$ if and only if $\rho(\mathbf{B}) < 1.$ The largest magnitude eigen value, i.e., $\rho(\mathbf{B})$ will eventually dominate and determine the rate at which $\mathbf{z}^{(k)} \to \bar{0}.$ The result proved in this section can be summarized as follows:

Theorem A sequence of vectors $\left\{\mathbf{z}^{(k)} : k = 0, 1, , 2, ...\right\}$ generated by the iteration scheme

$$\mathbf{z}^{(k+1)} = \mathbf{B}\mathbf{z}^{(k)}$$

where $\mathbf{z} \in R^n$ and $\mathbf{B} \in R^n \times R^n,$ starting from any arbitrary initial condition $\mathbf{z}^{(0)}$ will converge to limit $\mathbf{z}^* = \bar{0}$ if and only if

$$\rho(\mathbf{B}) < 1$$

Note that computation of eigenvalues is a computationally intensive task. The following theorem helps in deriving a sufficient conditions for convergence of linear iterative equations.

Theorem For a $n \times n$ matrix $\mathbf{B}$, the following inequality holds for any induced matrix norm

$$\rho(\mathbf{B}) \leq \|\mathbf{B}\| \qquad \text{--------(203)}$$

Proof Let λ_i be eigen value of $\mathbf{B}$ and $\mathbf{v}^{(i)}$ be the corresponding eigenvector. Then, we can write

$$\left\|\mathbf{B}\mathbf{v}^{(i)}\right\| = \left\|\lambda_i\mathbf{v}^{(i)}\right\| = |\lambda_i|\left\|\mathbf{v}^{(i)}\right\| \qquad \text{--------(204)}$$

for $i = 1, 2, ..., n.$ From these equations, it follows that

$$\rho(\mathbf{B}) = \max_i |\lambda_i| = \max_i \frac{\left\|\mathbf{B}\mathbf{v}^{(i)}\right\|}{\left\|\mathbf{v}^{(i)}\right\|} \qquad \text{--------(205)}$$

Using the definition of the induced matrix norm, we have

$$\frac{\|\mathbf{B}\mathbf{z}\|}{\|\mathbf{z}\|} \leq \|\mathbf{B}\| \qquad \text{--------(206)}$$

for any $\mathbf{z} \in R^n$. Thus, it follows that

--------(207)

$$\rho(\mathbf{B}) = \max_{i} \frac{\left\|\mathbf{B}\mathbf{v}^{(i)}\right\|}{\left\|\mathbf{v}^{(i)}\right\|} \leq \|\mathbf{B}\|$$

Since $\rho(\mathbf{B}) \leq \|\mathbf{B}\| < 1 \Rightarrow \rho(\mathbf{B}) < 1$, a sufficient condition for convergence of iterative scheme can be derived as follows

$$\|\mathbf{B}\| < 1 \quad \text{--------(208)}$$

The above sufficient condition is more useful from the viewpoint of computations as $\|\mathbf{B}\|_1$ and $\|\mathbf{B}\|_\infty$ can be computed quite easily. On the other hand, the spectral radius of a large matrix can be comparatively difficult to compute.

6. Induced Matrix Norms

We have already mentioned that set of all $m \times n$ matrices with real entries (or complex entries) can be viewed a linear vector space. In this section, we introduce the concept of *induced norm* of a matrix, which plays a vital role in the numerical analysis. A norm of a matrix can be interpreted as *amplification power* of the matrix. To develop a numerical measure for ill conditioning of a matrix, we first have to quantify this *amplification power* of the matrix.

Definition 51 (Induced Matrix Norm): The induced norm of a $m \times n$ matrix $\mathbf{A}$ is defined as mapping from $R^m \times R^n \rightarrow R^+$ such that

$$\|\mathbf{A}\| = \underset{\mathbf{x} \neq \overline{\mathbf{0}}}{Max} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \quad \text{-------- (96)}$$

In other words, $\|\mathbf{A}\|$ bounds the amplification power of the matrix i.e.

$$\frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}\| \text{ for all } \mathbf{x} \in R^n, \mathbf{x} \neq \overline{\mathbf{0}} \quad \text{-------- (97)}$$

The equality holds for at least one non zero vector $\mathbf{x} \in R^n$. An alternate way of defining matrix norm is as follows

$$\|\mathbf{A}\| = \underset{\|\hat{\mathbf{x}}\| = 1}{Max} \|\mathbf{A}\hat{\mathbf{x}}\| \quad \text{-------- (98)}$$

Defining $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

it is easy to see that these two definitions are equivalent. The following conditions are satisfied for any matrices $\mathbf{A}, B \in R^m \times R^n$

1. $\|\mathbf{A}\| > 0$ if $\mathbf{A} \neq [\mathbf{0}]$ and $\|[\mathbf{0}]\| = 0$
2. $\|\alpha\mathbf{A}\| = |\alpha| . \|\mathbf{A}\|$
3. $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$
4. $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \, \|\mathbf{B}\|$

The induced norms, i.e. norm of matrix induced by vector norms on R^m and R^n, can be interpreted as maximum **gain** or **amplification factor** of the matrix.

6.1 Computation of 2-norm

Now, consider **2-norm** of a matrix, which can be defined as follows

$$\|\mathbf{A}\|_2 = \underset{\mathbf{x} \neq 0}{max} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \quad \text{-------- (99)}$$

Squaring both sides

$$\|A\|_2^2 = \max_{x \neq 0} \frac{(Ax)^T(Ax)}{(x^T x)} = \max_{x \neq 0} \frac{x^T B x}{(x^T x)}$$

where $\mathbf{B} = \mathbf{A}^T\mathbf{A}$ is a symmetric and positive definite matrix. Positive definiteness of matrix $\mathbf{B}$ requires that

$$\mathbf{x}^T\mathbf{B}\mathbf{x} > 0 \text{ if } \mathbf{x} \neq \overline{0} \text{ and } \mathbf{x}^T\mathbf{B}\mathbf{x} = 0 \text{ if and only if } \mathbf{x} = \overline{0} \quad \text{-------- (100)}$$

If columns of $\mathbf{A}$ are linearly independent, then it implies that

$$\mathbf{x}^T\mathbf{B}\mathbf{x} = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) > 0 \text{ if } \mathbf{x} \neq \overline{\mathbf{0}} \quad \text{-------- (101)}$$

$$= 0 \quad \text{if } \mathbf{x} = \overline{\mathbf{0}} \quad \text{-------- (102)}$$

Now, a positive definite symmetric matrix can be diagonalized as

$$\mathbf{B} = \Psi\Lambda\Psi^T \quad \text{-------- (103)}$$

Where Ψ is matrix with eigen vectors as columns and Λ is the diagonal matrix with eigenvalues of $\mathbf{B}(= \mathbf{A}^T\mathbf{A})$ on the main diagonal. Note that in this case Ψ is unitary matrix ,i.e.,

$$\Psi\Psi^T = \mathbf{I} \text{ i.e } \Psi^T = \Psi^{-1} \quad \text{-------- (104)}$$

and eigenvectors are orthogonal. Using the fact that Ψ is unitary, we can write

$$\mathbf{x}^T\mathbf{x} = \mathbf{x}^T\Psi\Psi^T\mathbf{x} = \mathbf{y}^T\mathbf{y} \quad \text{-------- (105)}$$

This implies that

$$\frac{\mathbf{x}^T\mathbf{B}\mathbf{x}}{(\mathbf{x}^T\mathbf{x})} = \frac{\mathbf{y}^T\Lambda\mathbf{y}}{(\mathbf{y}^T\mathbf{y})} \quad \text{-------- (106)}$$

where $\mathbf{y} = \Psi^T\mathbf{x}$. Suppose eigenvalues λ_i of $\mathbf{A}^T\mathbf{A}$ are numbered such that

$$0 < \lambda_1 \leq \lambda_2 \leq \ldots\ldots \leq \lambda_n \quad \text{-------- (107)}$$

Then, we have

$$\frac{\mathbf{y}^T\Lambda\mathbf{y}}{(\mathbf{y}^T\mathbf{y})} = \frac{(\lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots\ldots + \lambda_n y_n^2)}{(y_1^2 + y_2^2 + \ldots\ldots + y_n^2)} \leq \lambda_n \quad \text{-------- (108)}$$

which implies that

$$\frac{\mathbf{y}^T\Lambda\mathbf{y}}{(\mathbf{y}^T\mathbf{y})} = \frac{\mathbf{x}^T B\mathbf{x}}{(\mathbf{x}^T\mathbf{x})} = \frac{\mathbf{x}^T(\mathbf{A}^T\mathbf{A})\mathbf{x}}{(\mathbf{x}^T\mathbf{x})} \leq \lambda_n \quad \text{-------- (109)}$$

The equality holds only at the corresponding eigenvector of $\mathbf{A}^T\mathbf{A}$, i.e.,

$$\frac{[\mathbf{v}^{(n)}]^T(\mathbf{A}^T\mathbf{A})\mathbf{v}^{(n)}}{[\mathbf{v}^{(n)}]^T\mathbf{v}^{(n)}} = \frac{[\mathbf{v}^{(n)}]^T\lambda_n\mathbf{v}^{(n)}}{[\mathbf{v}^{(n)}]^T\mathbf{v}^{(n)}} = \lambda_n \quad \text{-------- (110)}$$

Thus, 2 norm of matrix $\mathbf{A}$ can be computed as follows

$$\|A\|_2^2 = \max_{x \neq 0} \|Ax\|^2/\|x\|^2 = \lambda_{\max}(A^T A) \quad \text{-------- (111)}$$

i.e.

$$\|\mathbf{A}\|_2 = [\lambda_{max}(\mathbf{A}^T\mathbf{A})]^{1/2} \quad \text{-------- (112)}$$

where $\lambda_{max}(\mathbf{A}^T\mathbf{A})$ denotes maximum magnitude eigenvalue or the *spectral radius* of $\mathbf{A}^T\mathbf{A}$.

6.2 Other Matrix Norms

Other commonly used matrix norms are

- **1-norm**: Maximum over column sums

$$\|\mathbf{A}\|_1 = \max_{1 \le j \le n} \left[\sum_{i=1}^{n} |a_{ij}| \right] \text{ -------- } (113)$$

- $\infty -$ **norm:** Maximum over row sums

$$\|\mathbf{A}\|_\infty = \max_{1 \le i \le n} \left[\sum_{j=1}^{n} |a_{ij}| \right] \text{ -------- } (114)$$

RemarkThere are other matrix norms, such as Frobenious norm, which are not induced matrix norms. Frobenious norm is defined as

$$\|\mathbf{A}\|_F = \left[\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right]^{1/2}$$