LINEAR EQUATION SOLVERS*

BY

F. J. MURRAY

Columbia University

1. There are two types of devices for the solution of simultaneous linear equations which have been developed. Suppose the given system of equations is

$$\sum_{j=1}^{n} a_{ij}x_j = b_i , \qquad i = 1, 2, \cdots , n. \tag{1}$$

In one type the b_i are fed into the machine in such a way as to drive the unknowns x_j to their correct values. In the second type, the x_j are not driven by power supplied from the constant inputs but reach an equilibrium situation corresponding to the solution by a process of adjustment. We are concerned in this article with the operating conditions for this last type of machine when the adjusting process is determined by a linear operator with constant coefficients.

We suppose that each a_{ij} can run independently through a real range which contains the origin. Thus, the determinant may be zero and any matrix can be represented by a suitable choice of scale for the unknowns and the b_i . Adjusting machines which are stable even when the determinant is zero may be designed. For instance, a block diagram is given in the author's book[1] for such a machine. Another example is the set-up described by Goldberg and Brown[2] which will insure stability when a certain type of feedback is used.

However, in each case the coefficient network is duplicated. In the present article, we point out that if an adjusting type of machine is to operate successfully whenever the determinant A is not zero, then the square of the determinant must enter the indicial equation of the equations of motion for the machine. This necessary condition for successful operation rules out any linear feedback which does not involve using the a_{ij} twice. This result generalizes certain aspects of the necessity argument indicated in Goldberg and Brown.

In Secs. 2 and 3 below, we describe precisely the type of machine we are concerned with. These machines may function continuously or in discrete steps. In Secs. 4 and 5 we obtain necessary and sufficient conditions that the machine should operate successfully in all cases where a solution is uniquely determined. These conditions are analogous to stability conditions for a linear network. In the case of a continuous machine, this analogy is readily established; in the case of a discrete step machine, these operational conditions are obtained by considering certain parts of the theory of linear difference equations. In Sec. 6 we prove the mathematical theorem upon which our result is based. It is shown in Sec. 7 that an adjusting machine with a linear feedback network, which is independent of the coefficients of the equations, will not always operate successfully. Section 8 contains the mathematical basis for Sec. 5 which is concerned with discrete step machines.

---

*Received May 28, 1948.

[1]F. J. Murray, *Mathematical machines*, Kings Crown Press, New York, 1947 p. 92 (1st ed.), pp. III 20-21 (2nd ed.).

[2]E. A. Goldberg and G. W. Brown, J. Appl. Phys. 19, 339-344 (1948).

2. A mathematical machine can be regarded as a combination of computing components, each of which performs a specific mathematical operation. Each component has various inputs and a specific output. A combination of components can be used to evaluate a formula or function of the inputs. In a machine for solving a system of linear equations

$$\sum_j a_{ij}x_j + b_i = 0 \tag{1}$$

the coefficients a_{ij} and the constants b_i are inputs whose values do not vary during the operation of the machine.

In an adjusting type machine for solving linear equations, there are variables to represent the x_i . Each of these is associated with a unit of the following character. Each unit has an input X_i and an output x_i . The output x_i is suitable for use as an input in the computing components of the machine and the input X_i must correspond to an output of computing components. The relation between the input X_i and the output x_i is an operational one, $L_i(x_i) = X_i$. Various possibilities for L_i are discussed below; L_i may depend on i, it will always be linear.

The adjusting type device functions as follows. While the x_i units are inactivated, the a_{ij} and the b_i are entered into the machine. Presumably at this point the values of the x_i do not constitute a solution of the given system of equations (1). Now, however, the x_i units are activated. Various combinations of components compute the errors in each equation

$$\epsilon_i = \sum_{j=1}^{n} a_{ij}x_j + b_i \tag{2}$$

and these, in turn, are used to compute the X_i as some function $f(\epsilon_1 , \cdots , \epsilon_n)$ of the errors. The inputs X_i of the x_i units cause the latter to vary so as to approach the solution of the system (1). Since as we have pointed out in Sec. 1, for this purpose certain f values exist which are linear combinations of the ϵ_i , we will assume that f is linear in the ϵ_i :

$$X_j = \sum_i k_{ij}\epsilon_j \tag{3}$$

where the k_{ij} are constants in the sense that they do not change during the process in which the x_i are adjusted to the correct value.

3. We have not specified the nature of the operators L_i ,

$$L_i(x_i) = X_i , \tag{4}$$

or the method of functioning for the components. In general, there are two ways or manners in which a machine of this type may operate: (A) The adjusting process proceeds continuously. Each component has continuous inputs and output and the L_i are differential operators. We suppose the coefficients in each L_i are constants and that the coefficient of the highest derivative is 1. (B) The adjusting process proceeds in discrete steps. L_i is a linear difference operator with constant coefficients but the components may have discrete inputs or continuous inputs. Again the coefficient of the highest order difference in L_i is 1.

The two examples cited in the introduction operate in the (A) manner. An electronic digital computer programmed to operate in the sequence indicated by Eqs. (2), (3),

(4) would operate like (B). A non-essential generalization permits us to include the manually adjusted type of equation solver such as those described in the author's book *Mathematical machines* (2nd edition, pp. III 16-20; 1st ed., pp. 87-91). A full cycle of the adjusting process in these machines corresponds to one step in the sense of (B) above. However, the adjusting equations (4) are to be replaced by a set of relations in the form

$$\begin{aligned} L_1(x_1) &= X_1 , \\ L_2(x_1 , x_2) &= X_2 , \\ &\cdots \\ L_n(x_1 , \cdots , x_n) &= X_n , \end{aligned} \tag{4'}$$

where each L is linear in the x_i. It will be seen that the generalization represented by (4′) does not affect our argument. One may mention that L_1 may be taken simply as the differencing operation Δ. In general L_i contains x_i only in the form Δx_i so that the first equation determines Δx_1, the second Δx_2 and so forth.

We are interested then in the type of machine represented by the sequence of equations (2), (3), (4) or (4′). The obvious problem that appears here is concerned with the choice of the k_{ij} in (3) and the L_i in (4) so that the machine will work, i.e. so that the machine will adjust itself to a solution. As we have pointed out, sufficient conditions for adequate operation in all circumstances are known, but these require that the k_{ij} depend on the a_{ij}. This means that the a_{ij} must be used in computing ϵ_i and also again in forming the X_i.

Our objective is to establish that this double use of the a_{ij} is necessary if the machine is to function in all cases in which there is a solution. We suppose that the a_{ij} are permitted to assume independently all values in an interval which includes zero, say, for instance, from minus 1 to plus 1. If this is true then, by suitable choice of scale for the equations and the unknowns, any system of equations can be represented in the machine.

The above statement seems to neglect the case in which the a_{ij} appear digitally or as decimal fractions. However, when we are given a device in which the a_{ij} vary discretely in the components, we can regard each component as replaced by a continuous component of perfect accuracy and apply our argument to it. Now suppose in the idealized machine, we find a region of non-operation or instability. Then, in general, the original discrete machine will have permissible values for the a_{ij} which fall in this region and, of course, it will also be unstable.

Let us end this section by pointing out certain conclusions concerning the above mathematical setup which one may reach from the assumption that the machine will operate successfully as an adjusting device.

1) If the machine operates in the manner (A) and the L_i are differential operators such that

$$L_i(x_i) = x_i^{(t)} + l_{1i}x_i^{(t-1)} + \cdots + l_{ti}x_i ,$$

then $l_{ti} = 0$.

2) If the machine operates in manner (B) and

$$L_i(x_i) = \Delta^t x_i + l_{1i}\Delta^{t-1}x_i + \cdots + l_{ti}x_i ,$$

then $l_{ti} = 0$.

From Eqs. (3) and (4), in either case, we get

$$L_i(x_i) = \sum_j k_{ij}\epsilon_j . \tag{5}$$

Now suppose we are solving a system of equations in which x_i is not zero. Then after an adequate time interval in a successfully operating machine all the ϵ_i will be small, x_i will be close to its true value, all the differencing operators Δ^k and differentiating operators $(d/dt)^k$ will yield a small result, and these equations will be inconsistent with the assumption that $l_{ti} \neq 0$. The case (4′) is readily treated on an inductive basis.

4. Consider the Eqs. (5) in the case (A):

$$\begin{aligned} L_i(x_i) &= x_i^{(t)} + \cdots + l_{t-1,i}x_i^1 \\ &= \sum_{j=1}^{n} k_{ij}\epsilon_j \\ &= \sum_{j=1}^{n} k_{ij}\left(\sum_{\alpha=1}^{n} a_{j\alpha}x_\alpha + b_j\right). \end{aligned} \tag{5A}$$

Equation (5A) can be written in the form

$$L_i(x_i) - \sum_\alpha \left(\sum_j k_{ij}a_{j\alpha}\right)x_\alpha = \sum_j k_{ij}b_j = (\text{say})\ B_i .$$

The operators L_i can be treated as numbers, and we may employ the process by which Cramer's Rule is usually established to eliminate all but one unknown x. Consequently each x satisfies a differential equation in the form

$$\nabla x_i = C_i \tag{6}$$

where ∇ is a differential operator,

$$\nabla x_i = x_i^{(m)} + D_1x_i^{(m-1)} + \cdots + D_mx_i ,$$

and each C_i is a constant which of course depends on the b_i. The coefficients D do not depend on i. Clearly D_m is the determinant of the matrix with elements $\sum_j k_{ij}a_{j\alpha}$. (This follows from the fact established in the previous section that l_{ti} is zero.) Thus

$$D_m = KA, \tag{7}$$

where K and A are, respectively, the determinants of the matrices with elements k_{ij} and a_{ij}.

We proceed next to obtain the condition mentioned in Sec. 1 for successful operation in case (A). This condition is concerned with the algebraic equation

$$\mu^m + D_1\mu^{m-1} + \cdots + D_m = 0, \tag{8}$$

which is usually referred to as the indicial equation of the homogeneous differential equation $\nabla x_i = 0$.

We first obtain the general solution of $\nabla x_i = 0$. Suppose $\mu_1, \cdots, \mu_{s_1}$ are the real roots of (8) and that μ_j has multiplicity r_j. Also suppose that $\alpha_1 + i\beta_1, \alpha_1 - i\beta_1, \cdots, \alpha_{s_2} + i\beta_{s_2}, \alpha_{s_2} - i\beta_{s_2}$ are the complex roots of (8) and suppose that $\alpha_j + i\beta_j$ has multiplicity u_j. Then it is well-known that the general solution x_i^* of $\nabla x_i = 0$ can be written in the form

$$x_i^* = \sum_{k=1}^{s_1} \sum_{j=1}^{n_k} M_{j,k} t^{j-1} \exp(\mu_k t) + \sum_{k=1}^{s_2} \sum_{j=1}^{n_j} t^{j-1} \exp(\alpha_k t)(P_{j,k} \cos \beta_k t + Q_{j,k} \sin \beta_k t). \quad (9A)$$

On the other hand if s_0 is such that D_{m-s_0} is the D with highest subscript which does not vanish, then a particular solution of (6) is obtainable in the form $X_{s_i} t^{s_0}$ and the general solution of (6) can be written in the form

$$x_i = X_{s_i} t^{s_0} + x_i^*. \quad (10)$$

Now suppose that the system of equations has a unique solution in which no unknown Y_i is zero. For the machine to function correctly under these circumstances, each x_i must approach the correct constant value Y_i as t approaches infinity. Now one can show that an expression of the type (10) will approach a constant value as t approaches infinity if, and only if, $s_0 = 0$ and all the μ_k and α_k are negative, i.e. if (8) is stable.

Thus we have shown that:

An adjusting machine of type (A) *will operate successfully if, and only if, all the roots of the indicial equation* (8) *lie in that half of the complex plane for which the real part of a number is negative, whenever the system of equations is non-singular.*

5. We now wish to go through the analogous discussion for case (B). In this case, the L_i are difference operators

$$L_i(x_i) = \Delta^t x_i + l_{1,i} \Delta^{t-1} x_i + \cdots + l_{t-1,i} \Delta x_i ,$$

and we obtain

$$L_i(x_i) = \sum_{j=1}^{n} k_{ij} \epsilon_j . \quad (5B)$$

As before, we also obtain by elimination

$$\nabla x_i = C_i , \quad (6B)$$

where

$$\nabla x_i = \Delta^m x_i + D_1 \Delta^{m-1} x_i + \cdots + D_m x_i .$$

The statements concerning D_j and D_m are the same as before and, in particular, (7) holds.

The customary method of handling a homogeneous difference equation $\nabla x_i = 0$ with constant coefficients is to look for solutions in the form $x_i(t) = a^t$. For these we have $\Delta x_i = a^{t+1} - a^t = (a - 1)a^t = (a - 1)x_i(t) = \mu x_i(t)$ for $\mu = a - 1$. Thus if we permit μ to have this meaning, we can obtain the indicial equation

$$\mu^m + D_1 \mu^{m-1} + \cdots + D_m = 0. \quad (8B)$$

If we had two machines, one of which operated in manner (A) and the other in manner (B) with, however, the L_i and X_i being such that the coefficients L_{ij} and k_{ij} were the same in each case, then the Eq. (8) would also be the same for each case.

Corresponding to every distinct solution of (8B) we have a solution a^t, where $a = \mu + 1$. For multiple roots the same technique as that of ordinary differential equations permits one to find r linearly independent solutions if the root has multiplicity r. For instance, if μ_0 has multiplicity 3 then there is a polynomial ϕ such that

$$\mu^m + D_1\mu^{m-1} + \cdots + D_m = (\mu - \mu_0)^3\phi(\mu)$$

and

$$\nabla(\mu + 1)^t = (\mu - \mu_0)^3\phi(\mu)(\mu + 1)^t.$$

From this we can infer that

$$\nabla(\mu_0 + 1)^t = 0$$

$$\frac{\partial}{\partial\mu}[\nabla(\mu + 1)^t]_{\mu=\mu_0} = 0,$$

$$\frac{\partial^2}{\partial\mu^2}[\nabla(\mu + 1)^t]_{\mu=\mu_0} = 0.$$

Since Δ and $\partial/\partial\mu$ are commutative operations, we can infer that

$$\nabla\left(\frac{\partial}{\partial\mu}[(\mu + 1)^t]_{\mu=\mu_0}\right) = 0$$

and

$$\nabla\left(\frac{\partial^2}{\partial\mu^2}[(\mu + 1)^t]_{\mu=\mu_0}\right) = 0.$$

Thus $(\mu_0 + 1)^t$, $t(\mu_0 + 1)^{t-1}$ and $t(t - 1)(\mu_0 + 1)^{t-2}$ are solutions of $\nabla x_i = 0$ when μ_0 is a triple root of the indicial equation. Since linear combinations of solutions are also solutions $(\mu_0 + 1)^t$, $t(\mu_0 + 1)^t$ and $t^2(\mu_0 + 1)^t$ are equivalent to the given set if $\mu_0 + 1$ does not equal zero. If $\mu_0 + 1 = 0$, then $x_i = 1$ is a solution; if $\mu_0 = -1$ is a root of multiplicity 3 of (8), then, of course, the analogous equation for a has zero as a triple root and thus $x_i = 1$, t and t^2 are solutions. If μ_0 is a complex root and the D_i are real then $\bar{\mu}_0$ is also a root and solutions are obtained by taking the real and imaginary parts of the above solutions.

It is, of course, characteristic of the theory of linear difference equations that it parallels the theory of linear differential equations except that constants are replaced by periodic functions in the solutions. (There is another departure indicated below.) However, we need only consider the variable t to have discrete values $0, 1, 2, \cdots$; then periodic functions need not enter our discussion. For this situation we give a sequence of lemmas which yield results analogous to those we have used in the differential equation case. These are proved in Sec. 8 below.

Lemma A. *Let*

$$\nabla x = 0 \tag{6B$'$}$$

be a linear difference equation with constant coefficients. Let the variable t take on only the values 0, 1, 2, $\cdots$ *so that a solution of* (6B′) *is an infinite sequence*

$$\{x(0), x(1), x(2), \cdots\}.$$

There are m linearly independent real solutions of (6B′).

COROLLARY. *For the solutions obtained in Lemma A, we can suppose that the first determinant is not zero.*

LEMMA B. *Under the hypothesis of Lemma A, every solution of* (6B') *is a linear combination of the m linearly independent solutions obtained in Lemma A.*

Lemma B specifies the solution of the homogeneous equation $\nabla x_i = 0$; we can readily show that this is in a form analogous to (9A). For to each solution μ of (8) we can find a principal value of the logarithm such that

$$1 + \mu = \exp(\alpha' + i\beta')$$

or if $1 + \mu$ is real and positive, we can write simply

$$1 + \mu = \exp(\mu').$$

If s_0 is again such that D_{m-s_0} is the D with highest subscript which is not zero, then a particular solution of (6B) is readily obtained in the form

$$x_i = X_{s_0 i} \frac{t!}{s_0!(t - s_0)!}.$$

The argument used in the previous section now shows that if the machine is to function satisfactorily, then when the solution is unique one must have all μ' and α' negative. However, we are now one step removed from Eq. (8) and the condition for satisfactory operation of the machine is now that all the roots of (8) should be in the unit circle with center at $\mu = -1$. Thus we have proved:

An adjusting machine which operates in the manner (B) *will operate successfully if, and only if, all the roots μ of the indicial equation* (8) *lie in the unit circle with center at $\mu = -1$, whenever the determinant of a system* (A) *is not zero.*

6. We have established two facts which hold for adjusting machines quite irrespective of whether they operate in manner (A) or (B).

a) *The last coefficient D_m of the indicial equation is in the form KA where A is the determinant of the given system of equations.*

b) *If any root μ of Eq.* (8) *has a positive real part when the given system of equations is non-singular, then the machine will not operate successfully for this system.*

We now wish to establish the following theorem upon which our conclusions are based.

THEOREM. *Let Eq.* (8) *be the indicial equation of the Eq.* (6) *obtained by eliminating all but one x_i from the Eqs.* (4). *Furthermore, let us suppose that if the system of equations* (1) *is non-singular, then none of the roots μ of the indicial equation* (8) *has positive real parts. Then D is divisible by A^2, where A is the determinant of the system of equations* (1).

PROOF. Let us suppose that our system of equations has been chosen in the following way. Let $A(\lambda)$ be the characteristic determinant of the matrix and suppose we choose a set of a_{ij} so that all the latent roots, i.e. the roots of $A(\lambda) = 0$, are real and distinct. A symmetric real matrix of this type is readily found. Furthermore, for real a_{ij} one can show that for small changes in the a_{ij} the roots remain real and distinct. This is readily seen from the graph of $A(\lambda)$ which will have $n - 1$ maxima and minima whose ordinates alternate in sign if, and only if, $A(\lambda) = 0$ has n distinct roots. Thus we have an open region in the n^2 dimensional a_{ij} space in which the characteristic roots are real and distinct.

Let us, however, for the moment regard all the a_{ij} as fixed but consider all systems of equations in the form

$$\begin{aligned}(a_{11} - \lambda)x_1 + \cdots + a_{1n}x_n + b_1 &= 0,\\ \cdots\\ a_{n1}x_1 + \cdots + (a_{nn} - \lambda)x_n + b_n &= 0.\end{aligned} \tag{1.λ}$$

Equation (8) can now be written in the form

$$\mu^m + D_1(\lambda)\mu^{m-1} + \cdots + D_m(\lambda) = 0,$$

where now the $D_k(\lambda)$ are polynomials in a_{ij} and λ. We know that $D_m(\lambda) = KA(\lambda)$. ($K$ is constant, but only in the sense that it does not change during the operation of the machine, i.e. in the sense a_{ij} and λ are constants.)

Now suppose λ' is a root of $A(\lambda) = 0$. Then since $D_m(\lambda) = KA(\lambda)$, $\mu = 0$ is a root of the indicial equation for the system of equations (1.λ). Let p be the last integer such that $D_{m-p}(\lambda')$ does not equal zero. Since D_{m-k} is a polynomial in λ, $D_{m-k}(\lambda) = (\lambda - \lambda')\bar{D}_{m-k}(\lambda)$ for $k < p$ where $\bar{D}_{m-k}(\lambda)$ is also a polynomial in λ.

Now suppose $\bar{D}_m(\lambda') \neq 0$. Let us make the usual construction of the Newton polygon in order to obtain a series expansion[3] of the roots μ of the indicial equation in powers of $\lambda - \lambda'$:

$$\mu = c_1(\lambda - \lambda')^e + c_2(\lambda - \lambda')^{2e} + \cdots .$$

We readily find that (0, 1) and (p, 0) are vertices of the polygon. Let $\theta_1, \cdots, \theta_p$ denote the p roots of the equation

$$\bar{D}_m(\lambda') + \theta^p D_{m-p}(\lambda') = 0.$$

We have for each $k = 1, \cdots, p$ a solution in the form

$$\mu_k = \theta_k(\lambda - \lambda')^{1/p} + \cdots .$$

Now if $p = 1$, θ_1 is real and $\theta_1(\lambda - \lambda')$ changes sign as λ varies through λ'. This would yield a μ with a positive real part when $A(\lambda)$ is not zero. But this violates our hypothesis. For $p = 2$, we can choose λ so that both $\theta_1(\lambda - \lambda')^{1/2}$ and $\theta_2(\lambda - \lambda')^{1/2}$ are real and one will be positive. This will again destroy the stability. For $p = 3, 4, \cdots$ we can take $\lambda - \lambda'$ positive and since at least one θ_k has a positive real part, the real part of μ will be positive, and again we have contradicted the hypothesis.

Thus the assumption that $\bar{D}_m(\lambda') \neq 0$ contradicts the hypothesis that the roots have negative real parts when A is not zero. Hence $\bar{D}_m(\lambda') = 0$ and D_m has a factor $(\lambda - \lambda')^2$.

Now let us divide $D_m(\lambda)$ by $A^2(\lambda)$, considering them as polynomials in λ and the a_{ij}. Since the coefficient of λ^{2n} in $A^2(\lambda)$ is one, we find that there are two polynomials Q and R such that

[3]Cf., for instance, K. W. S. Hensel and G. Landsberg, *Theorie der Algebraischen Funktionen Einer Variabeln*, Teubner, Leipzig, 1902, pp. 39-52.

$$D_m(\lambda) - Q(\lambda)A^2(\lambda) = R(\lambda),$$

where $R(\lambda)$ has degree less than $2n$ in λ.

This holds for all matrices $\{a_{ij}\}$ in the $\{a_{ij}\}$ region we decided upon, within which all the roots λ' are real and distinct. However, in this region for each choice of $\{a_{ij}\}$, $R(\lambda)$ is divisible by $(\lambda - \lambda')^2$ for every root λ' of $A(\lambda) = 0$, and since the degree of $R(\lambda)$ is less than $2n$, $R(\lambda)$ must be identically zero in λ for $\{a_{ij}\}$ fixed. Since this holds for every matrix $\{a_{ij}\}$ of the region, R must be zero as a polynomial in a_{ij} and λ. This establishes the theorem.

7. The above necessary criterion for successful operation permits us to show that a simple adjusting machine, with a linear feedback in which the a_{ij} are not used, cannot function successfully for all systems for which the determinant is not zero. For here $D_m = KA$ where K does not depend on the a_{ij} and thus D_m is not divisible by A^2.

8. Lemma A. *Let*

$$\nabla x = 0 \tag{6B$'$}$$

be a linear difference equation with constant coefficients. Let the variable t take on only the values 0, 1, 2, $\cdots$ *so that a solution of* (6B′) *is an infinite sequence*

$$\{x(0), x(1), x(2), \cdots\}.$$

There are m linearly independent real solutions of (6B′).

Let us first ignore the restriction that the solutions be real. The above argument indicates how m solutions can be obtained by means of the indicial equation. We must establish then the linear independence of the m solutions obtained:

$$\begin{array}{ll} \{x^{(1)}(0), & x^{(1)}(1), \cdots\}, \\ \{x^{(2)}(0), & x^{(2)}(1), \cdots\}, \\ \cdots & \\ \{x^{(m)}(0), & x^{(m)}(1), \cdots\}. \end{array}$$

Regarding these solutions as an infinite matrix, we must show that the rank is m. If we do this, a finite submatrix must have rank m and have linearly independent rows. Consequently, the full matrix has linearly independent rows.

If the roots of (8) are all distinct then, of course, the determinant of the first m columns is the well-known Vandermonde determinant

$$\begin{vmatrix} 1 & \mu_1 + 1 & \cdots & (\mu_1 + 1)^{m-1} \\ 1 & \mu_2 + 1 & \cdots & (\mu_2 + 1)^{m-1} \\ \cdots & & & \\ 1 & \mu_m + 1 & \cdots & (\mu_m + 1)^{m-1} \end{vmatrix} = \prod_{i>j} (\mu_i - \mu_j),$$

which is not zero.

The case of multiple roots is treated by induction. We suppose that μ_0 is a triple root and that the first three solutions correspond to μ_0. In addition, for the moment we suppose that all other roots are distinct. Then the first three rows of the matrix are

$$\left\{\begin{matrix} 1 & \mu_0 + 1 & (\mu_0 + 1)^2 & (\mu_0 + 1)^3 & (\mu_0 + 1)^4 & \cdots \\ 0 & 1 & 2(\mu_0 + 1) & 3(\mu_0 + 1)^2 & 4(\mu_0 + 1)^3 & \cdots \\ 0 & 0 & 2\cdot 1 & 3\cdot 2(\mu_0 + 1) & 4\cdot 3(\mu_0 + 1)^2 & \cdots \end{matrix}\right\}.$$

In the distinct root case, we have

$$\begin{vmatrix} 1 & \mu_0 + 1 & \cdots & (\mu_0 + 1)^{m-1} \\ 1 & \mu_1 + 1 & \cdots & (\mu_1 + 1)^{m-1} \\ 1 & \mu_2 + 1 & \cdots & (\mu_2 + 1)^{m-1} \\ \cdots & & & \end{vmatrix}$$

$$= (\mu_0 - \mu_1)(\mu_0 - \mu_2)(\mu_1 - \mu_2) \prod (\mu_0 - \mu_i) \prod (\mu_1 - \mu_i) \prod (\mu_2 - \mu_i) \prod (\mu_i - \mu_j)$$

where i, j do not equal 0, 1, 2. If we let $\mu_1 = \mu_0 + h$, the elements in the second row of the determinant become

$$(\mu_1 + 1)^j = (\mu_0 + 1)^j + h[j(\mu_0 + 1)^{j-1}] + 1/2\, h^2[j(j-1)(\mu_0 + 1)^{j-2}] + \cdots + h^j. \tag{11}$$

Now we can subtract the first row of the determinant from this second row, divide both sides by h and take the limit as h approaches zero. Then we obtain

$$\begin{vmatrix} 1 & \mu_0 + 1 & (\mu_0 + 1)^2 & \cdots & (\mu_0 + 1)^{m-1} \\ 0 & 1 & 2(\mu_0 + 1) & \cdots & (m-1)(\mu_0 + 1)^{m-2} \\ 1 & \mu_2 + 1 & (\mu_2 + 1)^2 & \cdots & (\mu_2 + 1)^{m-1} \\ \cdots & & & & \end{vmatrix}$$

$$= (\mu_0 - \mu_2)^2 \prod (\mu_0 - \mu_i)^2 \prod (\mu_2 - \mu_i)^2 \prod (\mu_i - \mu_j).$$

We next let $\mu_2 = \mu_0 + h$ and the third row is in the form (11). Then we subtract the first row and h times the second row from this third row, divide both sides by h^2, and take the limit as h approaches zero. Thus we get

$$\frac{1}{2}\begin{vmatrix} 1 & \mu_0 + 1 & (\mu_0 + 1)^2 & \cdots & (\mu_0 + 1)^{m-1} \\ 0 & 1 & 2(\mu_0 + 1) & \cdots & (m-1)(\mu_0 + 1)^{m-2} \\ 0 & 0 & 2\cdot 1 & \cdots & (m-1)(m-2)(\mu_0 + 1)^{m-3} \\ \cdots & & & & \end{vmatrix}$$

$$= \prod (\mu_0 - \mu_i)^3 \prod (\mu_i - \mu_j).$$

Since the determinant has been expressed as a product of differences of distinct numbers, it is not zero. It should be clear how the cases of higher multiplicity are treated and also how one could successively deal with the case of two multiple roots, three multiple roots and so forth.

Thus, in every case we have that the solutions previously obtained are linearly independent. If we have complex roots, we write the determinant in the form

$$\begin{vmatrix} 1 & \mu+1 & \cdots & (\mu+1)^{m-1} \\ 1 & \bar{\mu}+1 & \cdots & (\bar{\mu}+1)^{m-1} \\ \cdots & & & \end{vmatrix}$$

i.e. every line containing a complex number is followed by the corresponding line for the complex conjugate. Now a determinant in the form

$$\begin{vmatrix} A_1 & A_2 & A_3 & A_4 \\ \bar{A}_1 & \bar{A}_2 & \bar{A}_3 & \bar{A}_4 \\ B_1 & B_2 & B_3 & B_4 \\ \bar{B}_1 & \bar{B}_2 & \bar{B}_3 & \bar{B}_4 \end{vmatrix}$$

can be converted into one whose first row consists of the real parts of the A_i and the second row the imaginary parts, and whose third and fourth rows depend in the corresponding manner on the B_i, by multiplying on the left by the determinant

$$\begin{vmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{-i}{2} & \frac{i}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{-i}{2} & \frac{i}{2} \end{vmatrix} = -1/4.$$

Similarly, the determinant

$$\begin{vmatrix} 1 & \mu+1 & \cdots & (\mu+1)^{m-1} \\ 1 & \bar{\mu}+1 & \cdots & (\bar{\mu}+1)^{m-1} \\ \cdots & & & \end{vmatrix}$$

can be converted into the determinant of the real solutions we wish to use by multiplying by a non-zero determinant. Thus the lemma is established.

COROLLARY. *For the solutions obtained in Lemma A, we can suppose that the first determinant is not zero.*

LEMMA B. *Under the hypothesis of Lemma A, every solution*

$$\{x(0),\ x(1),\ \cdots\}$$

of (6B′) *is a linear combination of the m linearly independent solutions obtained in Lemma A.*

PROOF. From (6B′) we can infer that there are constants c_i such that for p greater than or equal to m

$$x(p) = \sum_{j=1}^{m} c_j x(p - j)$$

since the coefficient of $\Delta^m x$ is 1. We consider the matrix,

$$\begin{Bmatrix} x_1(0) & x_1(1) & \cdots \\ \cdots & & \\ x_m(0) & x_m(1) & \cdots \\ x(0) & x(1) & \cdots \end{Bmatrix}.$$

The above result shows that the $(m + 1)$-th column of this matrix is a linear combination of the first m columns. The $(m + 2)$-th column is a linear combination of 2, 3, $\cdots$, $m + 1$ columns, but since the $(m +)$-th column is a linear combination of the columns 1 to m it follows that the $(m + 2)$-th column is also. The above result can be used to show inductively that every column is a linear combination of the first m columns and this implies that the rank of the matrix is less than or equal to m. However, the preceding corollary now permits us to infer that the rank is m.

For p greater than m, let us consider the finite matrix obtained by ignoring the columns for which t is greater than p. This finite matrix is of rank m. From the fact that the determinant of order m in the upper left-hand column is not zero, we can infer that the last row is a linear combination of the first m rows with coefficients which are determined solely by the first m columns, i.e. we have

$$x(t) = \sum_{j=1}^{n} c_j x_j(t) \tag{12}$$

for $t = 0, 1, \cdots, p$. But since the first m of these equations are adequate to determine c_i the latter do not depend on p. Since p can be arbitrarily large, (12) holds for every p.