Generalized eigenvector

From Wikipedia, the free encyclopedia

In linear algebra, for a matrix *A*, there may not always exist a full set of linearly independent eigenvectors that form a complete basis – a matrix may not be diagonalizable. This happens when the algebraic multiplicity of at least one eigenvalue λ is greater than its geometric multiplicity (the nullity of the matrix $(A - \lambda I)$, or the dimension of its nullspace). In such cases, a **generalized eigenvector** of *A* is a nonzero vector **v**, which is associated with λ having algebraic multiplicity $k \geq 1$, satisfying

$$(A - \lambda I)^k \mathbf{v} = \mathbf{0}.$$

The set of all generalized eigenvectors for a given λ, together with the zero vector, form the **generalized eigenspace** for λ.

Ordinary eigenvectors and eigenspaces are obtained for k=1.

Contents

For defective matrices

Generalized eigenvectors are needed to form a complete basis of a defective matrix, which is a matrix in which there are fewer linearly independent eigenvectors than eigenvalues (counting multiplicity). Over an algebraically closed field, the generalized eigenvectors *do* allow choosing a complete basis, as follows from the Jordan form of a matrix.

In particular, suppose that an eigenvalue λ of a matrix A has an algebraic multiplicity m but fewer corresponding eigenvectors. We form a sequence of m eigenvectors and generalized eigenvectors $x_1, x_2, \ldots, x_m$ that are linearly independent and satisfy

$$(A - \lambda I)x_k = \alpha_{k,1}x_1 + \cdots + \alpha_{k,k-1}x_{k-1}$$

for some coefficients $\alpha_{k,1}, \ldots, \alpha_{k,k-1}$, for $k = 1, \ldots, m$. It follows that

$$(A - \lambda I)^k x_k = 0.$$

The vectors $x_1, x_2, \ldots, x_m$ can always be chosen, but are not uniquely determined by the above relations. If the geometric multiplicity (dimension of the eigenspace) of λ is p, one can choose the first p vectors to be eigenvectors, but the remaining $m - p$ vectors are only generalized eigenvectors.

Examples

Example 1

Suppose

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Then there is one eigenvalue λ=1 with an algebraic multiplicity m of 2.

There are several ways to see that there will be one generalized eigenvector necessary. Easiest is to notice that this matrix is in Jordan normal form, but is not diagonal, meaning that this is not a diagonalizable matrix. Since there is 1 superdiagonal entry, there will be one generalized eigenvector (or you could note that the vector space is of dimension 2, so there can be only one generalized eigenvector). Alternatively, you could compute the dimension of the nullspace of $A - I$ to be p=1, and thus there are m-p=1 generalized eigenvectors.

Computing the ordinary eigenvector $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is left to the reader (see the eigenvector page for examples).

Using this eigenvector, we compute the generalized eigenvector v_2 by solving

$$(A - \lambda I)v_2 = v_1.$$

Writing out the values:

$$\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

This simplifies to

$$\begin{aligned} v_{21} + v_{22} - v_{21} &= 1 \\ v_{22} - v_{22} &= 0. \end{aligned}$$

This simplifies to

$$v_{22} = 1.$$

And v_{21} has no restrictions and thus can be any scalar. So the generalized eigenvector is $v_2 = \begin{bmatrix} * \\ 1 \end{bmatrix}$, where the * indicates that any value is fine. Usually picking 0 is easiest.

Example 2

The matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 6 & 3 & 2 & 0 & 0 \\ 10 & 6 & 3 & 2 & 0 \\ 15 & 10 & 6 & 3 & 2 \end{bmatrix}$$

has *eigenvalues* of 1 and 2 with *algebraic multiplicities* of 2 and 3, but *geometric multiplicities* of 1 and 1.

The *generalized eigenspaces* of A are calculated below.

$$(A - 1I)\begin{bmatrix} 0 \\ 1 \\ -3 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 6 & 3 & 1 & 0 & 0 \\ 10 & 6 & 3 & 1 & 0 \\ 15 & 10 & 6 & 3 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ -3 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$(A - 1I)\begin{bmatrix} 1 \\ -15 \\ 30 \\ -1 \\ -45 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 6 & 3 & 1 & 0 & 0 \\ 10 & 6 & 3 & 1 & 0 \\ 15 & 10 & 6 & 3 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -15 \\ 30 \\ -1 \\ -45 \end{bmatrix} = 3\begin{bmatrix} 0 \\ 1 \\ -3 \\ 3 \\ -1 \end{bmatrix}$$

$$(A - 2I)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 3 & -1 & 0 & 0 & 0 \\ 6 & 3 & 0 & 0 & 0 \\ 10 & 6 & 3 & 0 & 0 \\ 15 & 10 & 6 & 3 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$(A - 2I)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 3 & -1 & 0 & 0 & 0 \\ 6 & 3 & 0 & 0 & 0 \\ 10 & 6 & 3 & 0 & 0 \\ 15 & 10 & 6 & 3 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = 3\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$(A-2I)\begin{bmatrix}0\\0\\1\\-2\\0\end{bmatrix} = \begin{bmatrix}-1&0&0&0&0\\3&-1&0&0&0\\6&3&0&0&0\\10&6&3&0&0\\15&10&6&3&0\end{bmatrix}\begin{bmatrix}0\\0\\1\\-2\\0\end{bmatrix} = 3\begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}$$

This results in a basis for each of the *generalized eigenspaces* of A. Together they span the space of all 5 dimensional column vectors.

$$\left\{\begin{bmatrix}0\\1\\-3\\3\\-1\end{bmatrix}\begin{bmatrix}1\\-15\\30\\-1\\-45\end{bmatrix}\right\}, \left\{\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\\-2\\0\end{bmatrix}\right\}$$

The *Jordan Canonical Form* is obtained.

$$T = \begin{bmatrix}0&0&0&1&0\\3&0&0&-15&0\\-9&0&0&30&1\\9&0&3&-1&-2\\-3&9&0&-45&0\end{bmatrix} \quad J = \begin{bmatrix}1&1&0&0&0\\0&1&0&0&0\\0&0&2&1&0\\0&0&0&2&1\\0&0&0&0&2\end{bmatrix}$$

where

$$AT = TJ$$

Other meanings of the term

- The usage of generalized eigenfunction differs from this; it is part of the theory of rigged Hilbert spaces, so that for a linear operator on a function space this may be something different.

- One can also use the term *generalized eigenvector* for an eigenvector of the *generalized eigenvalue problem*

$$Av = \lambda Bv.$$

The Nullity of $(A - \lambda I)^k$

Introduction

In this section it is shown, when λ is an *eigenvalue* of a matrix A with *algebraic multiplicity* k, then the *null space* of $(A - \lambda I)^k$ has dimension k.

Existence of Eigenvalues

Consider a **nxn** matrix **A**. The *determinant* of **A** has the fundamental properties of being *n linear* and *alternating*. Additionally **det(I) = 1**, for **I** the **nxn** identity matrix. From the determinant's definition it can be seen that for a *triangular* matrix **T = (t_{ij})** that

$\mathbf{det(T) = \prod(t_{ii})}$.

There are three *elementary row operations*, *scalar multiplication*, *interchange* of two rows, and the *addition* of a *scalar multiple* of one row to another. Multiplication of a row of **A** by $\boldsymbol{\alpha}$ results in a new matrix whose determinant is $\boldsymbol{\alpha}$ **det(A)**. Interchange of two rows changes the *sign* of the determinant, and the addition of a scalar multiple of one row to another does not affect the determinant.

The following simple theorem holds, but requires a little proof.

Theorem:
The equation $\mathbf{A\ x = 0}$ has a solution $\mathbf{x \neq 0}$, if and only if $\mathbf{det(A) = 0}$.

proof:
Given the equation $\mathbf{A\ x = 0}$ attempt to solve it using the *elementary row operations* of *addition* of a *scalar multiple* of one row to another and row *interchanges* only, until an equivalent equation $\mathbf{U\ x = 0}$ has been reached, with **U** an upper triangular matrix.
Since $\mathbf{det(U) = \pm det(A)}$ and $\mathbf{det(U) = \prod(u_{ii})}$
we have that $\mathbf{det(A) = 0}$ if and only if at least one $\mathbf{u_{ii} = 0}$. The back substitution procedure as performed after *Gaussian Elimination* will allow placing at least one non zero element in **x** when there is a $\mathbf{u_{ii} = 0}$. When all $\mathbf{u_{ii} \neq 0}$ back substitution will require $\mathbf{x = 0}$.

Theorem:
The equation $\mathbf{A\ x = \lambda\ x}$ has a solution $\mathbf{x \neq 0}$, if and only if $\mathbf{det(\ \lambda\ I - A) = 0}$.

proof:
The equation $\mathbf{A\ x = \lambda\ x}$ is equivalent to $\mathbf{(\ \lambda\ I - A)\ x = 0}$.

Constructive proof of Schur's triangular form

The proof of the main result of this section will rely on the *similarity transformation* as stated and proven next.

Theorem: *Schur Transformation to Triangular Form Theorem*

For any $\mathbf{n \times n}$ matrix **A**, there exists a *triangular* matrix **T** and a *unitary* matrix **Q**, such that $\mathbf{A\ Q\ =\ Q\ T}$. (The transformations are not unique, but are related.)

Proof:

Let $\boldsymbol{\lambda_1}$, be an *eigenvalue* of the $\mathbf{n \times n}$ matrix **A** and **x** be an associated *eigenvector*, so that $\mathbf{A\ x = \lambda_1\ x}$.
Normalize the *length* of **x** so that $\mathbf{|x|\ =\ 1}$.

$$\text{For } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, \text{ construct a } \textit{unitary} \text{ matrix } \mathbf{Q} = \begin{bmatrix} x_1 & q_{1\,2} & q_{1\,3} & \cdots & q_{1\,n} \\ x_2 & q_{2\,2} & q_{2\,3} & \cdots & q_{2\,n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ x_n & q_{n\,2} & q_{n\,3} & \cdots & q_{n\,n} \end{bmatrix}$$

Q should have **x** as its first column and have its columns an *orthonormal basis* for $\mathbf{C^n}$.
Now, $\mathbf{A\ Q = Q\ U_1}$, with $\mathbf{U_1}$ of the form:

$$U_1 = \begin{bmatrix} \lambda_1 & u_{1\,2} \; u_{1\,3} \; \cdots \; u_{1\,n} \\ 0 & \\ \vdots & U_0 \\ 0 & \end{bmatrix}$$

Let the *induction hypothesis* be that the theorem holds for all **(n-1)×(n-1)** matrices.
From the construction, so far, it holds for **n = 2**.

Choose a unitary $\mathbf{Q_0}$, so that $\mathbf{U_0\,Q_0 = Q_0\,U_2}$, with $\mathbf{U_2}$ of the *upper triangular* form:

Define $\mathbf{Q_1}$ by:

$$Q_1 = \begin{bmatrix} 1 & 0 \; 0 \; \cdots \; 0 \\ 0 & \\ \vdots & Q_0 \\ 0 & \end{bmatrix}$$

Now:

$$U_1 Q_1 = \begin{bmatrix} \lambda_1 & u_{1\,2} \; u_{1\,3} \; \cdots \; u_{1\,n} \\ 0 & \\ \vdots & U_0 \\ 0 & \end{bmatrix} \begin{bmatrix} 1 & 0 \; 0 \; \cdots \; 0 \\ 0 & \\ \vdots & Q_0 \\ 0 & \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & z_{12} \; z_{13} \cdots z_{1n} \\ 0 & \\ \cdot & \\ \cdot & U_0 Q_0 \\ \cdot & \\ 0 & \end{bmatrix} = \begin{bmatrix} \lambda_1 & z_{12} \; z_{13} \cdots z_{1n} \\ 0 & \\ \cdot & \\ \cdot & Q_0 U_2 \\ \cdot & \\ 0 & \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \quad 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & Q_0 \\ \cdot & \\ 0 & \end{bmatrix} \begin{bmatrix} \lambda_1 & z_{12} \; z_{13} \cdots z_{1n} \\ 0 & \\ \cdot & \\ \cdot & U_2 \\ \cdot & \\ 0 & \end{bmatrix}$$

Summarizing,

$$U_1 Q_1 = Q_1 U_3$$

with:

$$U_3 = \begin{bmatrix} \lambda_1 & z_{12} & z_{13} & \cdots & z_{1n} \\ 0 & \lambda_2 & z_{23} & \cdots & z_{2n} \\ \cdot & 0 & \cdot & \cdots & \cdot \\ \cdot & \cdot & 0 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdot & \cdots 0 & \lambda_n \end{bmatrix}$$

Now, $\mathbf{A\,Q = Q\,U_1}$ and $\mathbf{U_1\,Q_1 = Q_1\,U_3}$, where $\mathbf{Q}$ and $\mathbf{Q_1}$ are *unitary* and $\mathbf{U_3}$ is *upper triangular*. Thus $\mathbf{A\,Q\,Q_1 = Q\,Q_1\,U_3}$. Since the product of two unitary matrices is unitary, the proof is done.

Nullity Theorem's Proof

Since from $\mathbf{A\,Q = Q\,U}$, one gets $\mathbf{A = Q\,U\,Q^T}$. It is easy to see $(\mathbf{x\,I - A}) = \mathbf{Q}\,(\mathbf{x\,I - U})\,\mathbf{Q^T}$ and hence $\mathbf{det(x\,I - A)} = \mathbf{det(x\,I - U)}$. So the characteristic polynomial of **A** is the same as that for **U** and is given by $\mathbf{p(x)} = (x - \lambda_1)(x - \lambda_2) \cdot \ldots \cdot (x - \lambda_n)$. (**Q** unitary)

Observe, the construction used in the proof above, allows choosing any order for the eigenvalues of **A** that will end up as the diagonal elements of the upper triangular matrix **U** obtained. The *algebraic mutiplicity* of an eigenvalue is the count of the number of times it occurs on the diagonal.

Now. it can be supposed for a given eigenvalue **λ**, of algebraic multiplicity **k**, that **U** has been contrived so that **λ** occurs as the first **k** diagonal elements.

$$\mathbf{U} = \begin{bmatrix} \lambda & z_{1\,2} & z_{1\,3} & \cdot & \cdot & \cdot & \cdot & z_{1\,n} \\ 0 & \lambda & z_{2\,3} & \cdot & \cdot & \cdot & \cdot & z_{2\,n} \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & \lambda & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \lambda_{k+1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & \lambda_n \end{bmatrix}$$

Place **(U − λ I)** in *block form* as below.

$$\mathbf{U} - \lambda\,\mathbf{I} = \left[\begin{array}{cccc|cccc} 0 & z_{1\,2} & z_{1\,3} & \cdot & \cdot & \cdot & \cdot & z_{1\,n} \\ 0 & 0 & z_{2\,3} & \cdot & \cdot & \cdot & \cdot & z_{2\,n} \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & 0 & \cdot & \cdot & \cdot & z_{k\,n} \\ \hline 0 & 0 & \cdot & 0 & \beta_{k+1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & 0 & 0 & \cdot & 0 & \beta_n \end{array}\right]$$

The lower left block has only elements of *zero*.
The $\boldsymbol{\beta_i = \lambda_i - \lambda \neq 0}$ for **i = k+1, ..., n**. It is easy to verify the following.

$$(\mathbf{U} - \lambda\,\mathbf{I}) = \left[\begin{array}{c|c} \mathbf{B} & \begin{matrix} \cdot & \cdot & \cdot & z_{1\,n} \\ \cdot & \cdot & \cdot & z_{2\,n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & z_{k\,n} \end{matrix} \\ \hline \begin{matrix} 0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & 0 \end{matrix} & \mathbf{T} \end{array}\right]$$

$$(U - \lambda I)^k = \left[\begin{array}{c|cccc} & \cdot & & \cdot & \cdot\, y_{1\,n} \\ \mathbf{B}^k & \cdot & & \cdot & \cdot\, y_{2\,n} \\ & \cdot & \cdot & \cdot & \cdot \\ & \cdot & & \cdot & \cdot\, y_{k\,n} \\ \hline \begin{matrix} 0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & 0 \end{matrix} & & & \mathbf{T}^k & \end{array}\right]$$

Where **B** is the **kxk** sub triangular matrix, with all elements on or below the diagonal equal to **0**, and **T** is the **(n-k)x(n-k)** upper triangular matrix, taken from the blocks of $(\mathbf{U} - \boldsymbol{\lambda}\, \mathbf{I})$, as shown below.

$$\mathbf{B} = \begin{bmatrix} 0 & z_{1\,2} & z_{1\,3} & \cdot \\ 0 & 0 & z_{2\,3} & \cdot \\ \cdot & 0 & \cdot & \cdot \\ 0 & \cdot & 0 & 0 \end{bmatrix} \text{ and } \mathbf{T} = \begin{bmatrix} \beta_{k+1} & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & \beta_n \end{bmatrix}$$

Now, almost trivially!

$$\mathbf{B}^k = \begin{bmatrix} 0 & 0 & 0 & \cdot & 0 \\ 0 & 0 & 0 & \cdot & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & \cdot & 0 \end{bmatrix} \text{ and } \mathbf{T}^k = \begin{bmatrix} \beta_{k+1}{}^k & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & \beta_n{}^k \end{bmatrix}$$

That is $\mathbf{B}^\mathbf{k}$ has only elements of **0** and $\mathbf{T}^\mathbf{k}$ is triangular with all non zero diagonal elements. Just observe that if a column vector $v = \langle v_1, v_2, ..., v_k \rangle^T$, is mutiplied by **B**, then after the first multiplication the last, **k'**th, component is zero. After the second multiplication the second to last, **k-1'**th component is zero, also, and so on.

The conclusion that $(\mathbf{U} - \boldsymbol{\lambda}\, \mathbf{I})^\mathbf{k}$ has *rank* **n-k** and *nullity* **k** follows.

It is only left to observe, since $(\mathbf{A} - \boldsymbol{\lambda}\, \mathbf{I})^\mathbf{k} = Q\,(U - \lambda\, I)^k\, Q^T$**,**
that $(A - \lambda\, I)^k$ has *rank* **n-k** and *nullity* **k**, also. A *unitary*, or any other similarity transformation by a non-singular matrix preserves rank.

The main result is now proven.

Theorem:
If $\boldsymbol{\lambda}$ is an *eigenvalue* of a matrix **A** with *algebraic multiplicity* **k**,
then the *null space* of $(\mathbf{A} - \boldsymbol{\lambda}\, \mathbf{I})^\mathbf{k}$ has dimension **k**.

An important observation is that raising the power of $(\mathbf{A} - \boldsymbol{\lambda}\, \mathbf{I})$ above **k** will not affect the *rank* and *nullity* any further.

Motivation of the Procedure

Introduction

In the section *Existence of Eigenvalues* it was shown that when a **nxn** matrix $\mathbf{A}$, has an *eigenvalue* $\boldsymbol{\lambda}$, of *algebraic multiplicity* $\mathbf{k}$, then the *null space* of $\mathbf{(A - \lambda\, I)^k}$, has dimension $\mathbf{k}$.

The *Generalized Eigenspace* of $\mathbf{A, \lambda}$ will be defined to be the *null space* of $\mathbf{(A - \lambda\, I)^k}$. Many authors prefer to call this the *kernel* of $\mathbf{(A - \lambda\, I)^k}$.

Notice that if a **nxn** matrix has *eigenvalues* $\boldsymbol{\lambda_1, \lambda_2, ..., \lambda_r}$ with *algebraic multiplicities* $\mathbf{k_1, k_2, ..., k_r}$, then $\mathbf{k_1 + k_2 + ... + k_r = n}$.

It will turn out that any two *generalized eigenspaces* of $\mathbf{A}$, associated with different *eigenvalues*, will have a trivial intersection of $\mathbf{\{0\}}$. From this it follows that the *generalized eigenspaces* of $\mathbf{A}$ combined span $\mathbf{C^n}$, the set of all $\mathbf{n}$ dimensional column vectors of complex numbers.

The motivation for using a recursive procedure starting with the *eigenvectors* of $\mathbf{A}$ and solving for a basis of the *generalized eigenspace* of $\mathbf{A, \lambda}$ using the matix $\mathbf{(A - \lambda\, I)}$, will be expounded on.

Notation

Some notation is introduced to help abbreviate statements.

- $\mathbf{C^n}$ is the vector space of all $\mathbf{n}$ dimensional *column* vectors of *complex numbers.*
- The *Null Space* of $\mathbf{A}$, $\mathbf{N(A) = \{x: A\, x = 0\}}$.
- $\mathbf{V \subseteq W}$ will mean $\mathbf{V}$ is a *subset* of $\mathbf{W}$.
- $\mathbf{V \subset W}$ will mean $\mathbf{V}$ is a *proper subset* of $\mathbf{W}$.
- $\mathbf{A(V) = \{y: y = A\, x,}$ for some $\mathbf{x \in V\}}$.
- $\mathbf{W \setminus V}$ will mean $\mathbf{\{ x: x \in W}$ and $\mathbf{x}$ is not in $\mathbf{V\}}$.
- The *Range* of $\mathbf{A}$ is $\mathbf{A(C^n)}$ and will be denoted by $\mathbf{R(A)}$.
- $\mathbf{dim(V)}$ will stand for the *dimension* of $\mathbf{V}$.
- $\mathbf{\{0\}}$ will stand for the *trivial subspace* of $\mathbf{C^n}$.

Preliminary Observations

Throughout this discussion it is assumed that $\mathbf{A}$ is a **nxn** matrix of complex numbers.

Since $\mathbf{A^m\, x = A\, (A^{m-1}\, x)}$, the inclusions

$$\mathbf{N(A) \subseteq N(A^2) \subseteq ... \subseteq}\ N(A^{m-1}) \subseteq N(A^m),$$

are obvious. Since $\mathbf{A^m\, x = A^{m-1}(A\, x)}$, the inclusions

$$\mathbf{R(A) \supseteq R(A^2) \supseteq ... \supseteq}\ R(A^{m-1}) \supseteq R(A^m),$$

are clear too.

Theorem:
When the more trivial case $\mathbf{N(A^2)} = \mathbf{N(A)}$, does not hold,
there exists $\mathbf{k \geq 2}$, such that the inclusions,
$\mathbf{N(A) \subset N(A^2) \subset \ldots \subset}\ N(A^{k-1}) \subset N(A^k) = N(A^{k+1}) = \ldots$,
and
$\mathbf{R(A) \supset R(A^2) \supset \ldots \supset}\ R(A^{k-1}) \supset R(A^k) = R(A^{k+1}) = \ldots$,
are proper.

proof:

$\mathbf{0 \leq \dim(R(A^{m+1})) \leq \dim(R(A^m))}$ so eventually $\mathbf{\dim(R(A^{m+1})) = \dim(R(A^m))}$,
for some $\mathbf{m}$. From the inclusion $\mathbf{R(A^{m+1}) \subseteq R(A^m)}$ it is seen that a basis for
$\mathbf{R(A^{m+1})}$ is a basis for $\mathbf{R(A^m)}$ too. That is $\mathbf{R(A^{m+1}) = R(A^m)}$.
Since $\mathbf{R(A^{m+1}) = A(R(A^m))}$, when $\mathbf{R(A^{m+1}) = R(A^m)}$, it will be
$\mathbf{R(A^{m+2}) = A(R(A^{m+1})) = A(R(A^m))} = R(A^{m+1})$.
By the *rank nullity theorem*, it will also be the case that
$\mathbf{\dim(N(A^{m+2})) = \dim(N(A^{m+1})) = \dim(N(A^m))}$, for the same $\mathbf{m}$.
From the inclusions $\mathbf{N(A^{m+2}) \subseteq N(A^{m+1}) \subseteq N(A^m)}$,
it is clear that a basis for $\mathbf{N(A^{m+2})}$ is also a basis for $\mathbf{N(A^{m+1})}$ and $\mathbf{N(A^m)}$.
So $\mathbf{N(A^{m+2})} = \mathbf{N(A^{m+1})} = \mathbf{N(A^m)}$.
Now, $\mathbf{k}$ is the first $\mathbf{m}$ for which this happens.

Since certain expressions will occur many times in the following,
some more notation will be introduced.

- $\mathbf{A_{\lambda,\,k} = (A - \lambda\, I)^k}$
- $\mathbf{N_{\lambda,\,k} = N((A - \lambda\, I)^k)} = N(A_{\lambda,\,k})$
- $\mathbf{R_{\lambda,\,k} = R((A - \lambda\, I)^k)} = R(A_{\lambda,\,k})$

From the inclusions $\mathbf{N_{\lambda,\,1}} \subset N_{\lambda,\,2} \subset \ldots \subset N_{\lambda,\,k-1} \subset N_{\lambda,\,k} = N_{\lambda,\,k+1} = \ldots$,
$\mathbf{N_{\lambda,\,k} \setminus \{0\} = \cup}\ (N_{\lambda,\,m} \setminus N_{\lambda,\,m-1})$, for $\mathbf{m = 1, \ldots, k}$ and $\mathbf{N_{\lambda,\,0} = \{0\}}$, follows.

When $\boldsymbol{\lambda}$ is an eigenvalue of $\mathbf{A}$, in the statement above, $\mathbf{k}$ will not exceed the
algebraic multiplicity of $\boldsymbol{\lambda}$, and can be less. In fact when $\mathbf{k}$ would only be $\mathbf{1}$ is
when there is a full set of linearly independent eigenvectors. Let's consider when $\mathbf{k \geq 2}$.

Now, $\mathbf{x \in N_{\lambda,\,m} \setminus N_{\lambda,\,m-1}}$, if and only if $\mathbf{A_{\lambda,\,m}\, x = 0,}$ and $\mathbf{A_{\lambda,\,m-1}\, x \neq 0}$. Make the observation
that $\mathbf{A_{\lambda,\,m}\, x = 0,}$ and $\mathbf{A_{\lambda,\,m-1}\, x \neq 0}$,
if and only if $\mathbf{A_{\lambda,\,m-1}\, A_{\lambda,\,1}\, x = 0,}$ and $\mathbf{A_{\lambda,\,m-2}\, A_{\lambda,\,1}\, x \neq 0}$.

So, $\mathbf{x \in N_{\lambda,\,m} \setminus N_{\lambda,\,m-1}}$, if and only if $\mathbf{A_{\lambda,\,1}\, x \in N_{\lambda,\,m-1} \setminus N_{\lambda,\,m-2}}$.

Recursive Procedure

Consider a matrix $\mathbf{A}$, with an *eigenvalue* $\boldsymbol{\lambda}$ of *algebraic multiplicity* $\mathbf{k \geq 2}$,
such that there are not $\mathbf{k}$ *linearly independent eigenvectors* associated with $\boldsymbol{\lambda}$.

It is desired to extend the *eigenvectors* to a *basis* for $\mathbf{N_{\lambda,\,k}}$. That is a *basis* for
the *generalized eigenvectors* associated with $\boldsymbol{\lambda}$.

There exists some $\mathbf{2 \le r \le k}$, such that

$\mathbf{N_{\lambda,\, 1}} \subset N_{\lambda,\, 2} \subset \dots \subset N_{\lambda,\, r\text{-}1} \subset N_{\lambda,\, r} = N_{\lambda,\, r+1} = \dots,$
$\mathbf{N_{\lambda,\, r} \setminus \{0\} = \cup}\ (N_{\lambda,\, m} \setminus N_{\lambda,\, m\text{-}1})$**,** for $\mathbf{m = 1, \dots, r}$ and $\mathbf{N_{\lambda,\, 0} = \{0\}}$, .

The *eigenvectors* are $\mathbf{N_{\lambda,\, 1} \setminus \{0\}}$, so let $\mathbf{x_1, \dots, x_{r_1}}$ be a basis for $\mathbf{N_{\lambda,\, 1} \setminus \{0\}}$.

Note that each $\mathbf{N_{\lambda,\, m}}$ is a *subspace* and so a *basis* for $\mathbf{N_{\lambda,\, m\text{-}1}}$
can be extended to a *basis* for $\mathbf{N_{\lambda,\, m}}$.

Because of this we can expect to find some $\mathbf{r_2 = dim(N_{\lambda,\, 2}) - dim(N_{\lambda,\, 1})}$

linearly independent vectors

$\mathbf{x_{r_1+1}, \dots, x_{r_1+r_2}}$ such that $\mathbf{x_1, \dots, x_{r_1}}$, $\mathbf{x_{r_1+1}, \dots, x_{r_1+r_2}}$

is a *basis* for $\mathbf{N_{\lambda,\, 2}}$

Now, $\mathbf{x \in N_{\lambda,\, 2} \setminus N_{\lambda,\, 1}}$, if and only if $\mathbf{A_{\lambda,\, 1}\, x \in N_{\lambda,\, 1} \setminus \{0\}}$.

Thus we can expect that for each $\mathbf{x \in \{x_{r_1+1}, \dots, x_{r_1+r_2}\}}$,

$\mathbf{A_{\lambda,\, 1}\, x = \alpha_1\, x_1 + \dots + \alpha_{r_1}\, x_{r_1}}$,

for some $\mathbf{\alpha_1, \dots, \alpha_{r_1}}$, depending on $\mathbf{x}$.

Suppose we have reached the stage in the construction so that **m-1** sets,

$\mathbf{\{x_1, \dots, x_{r_1}\}}$, $\mathbf{\{x_{r_1+1}, \dots, x_{r_1+r_2}\}}$, ..., $\mathbf{\{x_{r_1+ \dots + r_{m-2}+1}, \dots, x_{r_1+ \dots + r_{m-1}}\}}$

such that

$\mathbf{x_1, \dots, x_{r_1}}$, $\mathbf{x_{r_1+1}, \dots, x_{r_1+r_2}}$, ..., $\mathbf{x_{r_1+ \dots + r_{m-2}+1}, \dots, x_{r_1+ \dots + r_{m-1}}}$

is a *basis* for $\mathbf{N_{\lambda,\, m\text{-}1}}$, have been found.

We can expect to find some $\mathbf{r_m = dim(N_{\lambda,\, m}) - dim(N_{\lambda,\, m\text{-}1})}$

linearly independent vectors

$\mathbf{x_{r_1+ \dots + r_{m-1}+1}, \dots, x_{r_1+ \dots + r_m}}$ such that

$\mathbf{x_1, \dots, x_{r_1}}$, $\mathbf{x_{r_1+1}, \dots, x_{r_1+r_2}}$, ..., $\mathbf{x_{r_1+ \dots + r_{m-1}+1}, \dots, x_{r_1+ \dots + r_m}}$

is a *basis* for $\mathbf{N_{\lambda,\, m}}$

Again, $\mathbf{x \in N_{\lambda,\, m} \setminus N_{\lambda,\, m\text{-}1}}$, if and only if $\mathbf{A_{\lambda,\, 1}\, x \in N_{\lambda,\, m\text{-}1} \setminus N_{\lambda,\, m\text{-}2}}$.

Thus we can expect that for each $\mathbf{x \in \{x_{r_1+ \dots + r_{m-1}+1}, \dots, x_{r_1+ \dots + r_m}\}}$,

$\mathbf{A_{\lambda,\, 1}\, x = \alpha_1\, x_1 + \dots + \alpha_{r_1+ \dots + r_{m-1}}\, x_{r_1+ \dots + r_{m-1}}}$,

for some $\mathbf{\alpha_1, \dots, \alpha_{r_1+ \dots + r_{m-1}}}$, depending on $\mathbf{x}$.

Some of the $\mathbf{\{\alpha_{r_1+ \dots + r_{m-2}+1}, \dots, \alpha_{r_1+ \dots + r_{m-1}}\}}$, will be **non zero**,

since $\mathbf{A_{\lambda,\, 1}\, x}$ must lie in $\mathbf{N_{\lambda,\, m\text{-}1} \setminus N_{\lambda,\, m\text{-}2}}$.

The procedure is continued until $\mathbf{m = r}$.

The α_i are not truly arbitrary and must be chosen, accordingly, so that sums $\alpha_1 x_1 + \alpha_2 x_2 + ...$ are in the range of $A_{\lambda, 1}$.

Generalized Eigenspace Decomposition

As was stated in the Introduction, if a **nxn** matrix has *eigenvalues* $\lambda_1, \lambda_2, ..., \lambda_r$ with *algebraic multiplicities* $k_1, k_2, ..., k_r$, then $k_1 + k_2 + ... + k_r = n$.

When V_1 and V_2 are two *subspaces*, satisfying $V_1 \cap V_2 = \{0\}$, their *direct sum*, $\oplus$, is defined and notated by

- $V_1 \oplus V_2 = \{v_1 + v_2 : v_1 \in V_1 \text{ and } v_2 \in V_2\}$.

$V_1 \oplus V_2$ is also a *subspace* and $\dim(V_1 \oplus V_2) = \dim(V_1) + \dim(V_2)$.

Since $\dim(N_{\lambda_i, k_i}) = k_i$, for $i = 1, 2, ..., r$, after it is shown that

$N_{\lambda_i, k_i} \cap N_{\lambda_j, k_j} = \{0\}$, for $i \neq j$,

we have the main result.

Theorem: ***Generalized Eigenspace Decomposition Theorem***

$C^n = N_{\lambda_1, k_1} \oplus N_{\lambda_2, k_2} \oplus ... \oplus N_{\lambda_r, k_r}$.

This follows easily after we prove the theorem below.

Theorem:
Let λ be an *eigenvalue* of A and $\beta \neq \lambda$. Then
$A_{\beta, r}(N_{\lambda, m} \setminus N_{\lambda, m-1}) = N_{\lambda, m} \setminus N_{\lambda, m-1}$,
for any positive integers m and r.

proof:
If $x \in N_{\lambda, 1} \setminus \{0\}$, $A_{\lambda, 1} x = (A - \lambda I)x = 0$,
then $A x = \lambda x$ and $A_{\beta, 1} x = (A - \beta I)x = (\lambda - \beta)x$.
So $A_{\beta, 1} x \in N_{\lambda, 1} \setminus \{0\}$ and $A_{\beta, 1} (\lambda - \beta)^{-1} x = x$.
It holds $A_{\beta, 1} (N_{\lambda, 1} \setminus \{0\}) = N_{\lambda, 1} \setminus \{0\}$. Now, $x \in N_{\lambda, m} \setminus N_{\lambda, m-1}$, if and only if
$A_{\lambda, m} x = (A - \lambda I)A_{\lambda, m-1} x = 0$, and $A_{\lambda, m-1} x \neq 0$.
In the case, $x \in N_{\lambda, m} \setminus N_{\lambda, m-1}$,
$A_{\lambda, m-1} x \in N_{\lambda, 1} \setminus 0$, and $A_{\beta, 1} A_{\lambda, m-1} x = (\lambda - \beta) A_{\lambda, m-1} x \neq 0$.
The *operators* $A_{\beta, 1}$ and $A_{\lambda, m-1}$ commute.
Thus $A_{\lambda, m}(A_{\beta, 1} x) = 0$ and $A_{\lambda, m-1}(A_{\beta, 1} x) \neq 0$,
which means $A_{\beta, 1} x \in N_{\lambda, m} \setminus N_{\lambda, m-1}$.

Now, let our *induction hypothesis* be,
$A_{\beta, 1}(N_{\lambda, m} \setminus N_{\lambda, m-1}) = N_{\lambda, m} \setminus N_{\lambda, m-1}$, .
The relation $A_{\beta, 1} x = (\lambda - \beta) x + A_{\lambda, 1} x$ **holds.**

For $\mathbf{y \in N_{\lambda, m+1} \setminus N_{\lambda, m}}$, let $\mathbf{x = (\lambda - \beta)^{-1} y + z}$.
Then $\mathbf{A_{\beta, 1} x} = \mathbf{y} + (\lambda - \beta)^{-1} A_{\lambda, 1} y + (\lambda - \beta) z + A_{\lambda, 1} z$
$= \mathbf{y} + (\lambda - \beta)^{-1} A_{\lambda, 1} y + A_{\beta, 1} z$.
Now, $\mathbf{A_{\lambda, 1} y} \in N_{\lambda, m} \setminus N_{\lambda, m-1}$ and, by the
induction hypothesis, there exists $\mathbf{z} \in N_{\lambda, m} \setminus N_{\lambda, m-1}$ that solves
$\mathbf{A_{\beta, 1} z} = -(\lambda - \beta)^{-1} A_{\lambda, 1} y$.
It follows $\mathbf{x \in N_{\lambda, m+1} \setminus N_{\lambda, m}}$ and solves $\mathbf{A_{\beta, 1} x = y}$.
So $\mathbf{A_{\beta, 1}(N_{\lambda, m+1} \setminus N_{\lambda, m})} = N_{\lambda, m+1} \setminus N_{\lambda, m}$, .

Repeatedly applying $\mathbf{A_{\beta, r} = A_{\beta, 1} A_{\beta, r-1}}$ finishes the proof.

¶

In fact, from the theorem just proved, for $\mathbf{i \neq j}$,

$\mathbf{A_{\lambda_i, k_i}(N_{\lambda_j, k_j})} = N_{\lambda_j, k_j}$.

Now, suppose that $\mathbf{N_{\lambda_i, k_i}} \cap N_{\lambda_j, k_j} \neq \{0\}$, for some $\mathbf{i \neq j}$.

Choose $\mathbf{x \in N_{\lambda_i, k_i}} \cap N_{\lambda_j, k_j} \neq 0$.

Since $\mathbf{x \in N_{\lambda_i, k_i}}$, it follows $\mathbf{A_{\lambda_i, k_i} x = 0}$.

Since $\mathbf{x \in N_{\lambda_j, k_j}}$, it follows $\mathbf{A_{\lambda_i, k_i} x \neq 0}$,
because $\mathbf{A_{\lambda_i, k_i}}$ preserves dimension on $\mathbf{N_{\lambda_j, k_j}}$.

So it must be $\mathbf{N_{\lambda_i, k_i}} \cap N_{\lambda_j, k_j} = \{0\}$, for $\mathbf{i \neq j}$.

This concludes the proof of the *Generalized Eigenspace Decomposition Theorem*.

Powers of a Matrix

using generalized eigenvectors

Assume **A** is a **nxn** matrix with *eigenvalues* $\mathbf{\lambda_1, \lambda_2, ..., \lambda_r}$
of *algebraic multiplicities* $\mathbf{k_1, k_2, ..., k_r}$.

For notational convenience $\mathbf{A_{\lambda, 0} = I}$.

Note that $\mathbf{A_{\beta, 1} = (\lambda - \beta)I + A_{\lambda, 1}}$. and apply the *binomial theorem*.

$$\mathbf{A_{\beta, s} = ((\lambda - \beta)I + A_{\lambda, 1})^s = \sum_{m=0}^{s} \binom{s}{m} (\lambda - \beta)^{s-m} A_{\lambda, m}}$$

When $\boldsymbol{\lambda}$ is an *eigenvalue* of *algebraic multiplicity* **k**, and $\mathbf{x \in N_{\lambda, k}}$,
then $\mathbf{A_{\lambda, m} x = 0}$, for $\mathbf{m \geq k}$, so in this case:

$$\mathbf{A}_{\beta,\, s}\,\mathbf{x} = \sum_{m=0}^{\min(s,\, k-1)} \binom{s}{m} (\lambda - \beta)^{s-m} \mathbf{A}_{\lambda,\, m}\,\mathbf{x}$$

Since $\mathbf{C}^n = N_{\lambda_1,\, k_1} \oplus N_{\lambda_2,\, k_2} \oplus \ldots \oplus N_{\lambda_r,\, k_r}$,

any $\mathbf{x}$ in $\mathbf{C}^n$ can be expressed as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_r$,

with each $\mathbf{x}_i \in \mathbf{N}_{\lambda_i,\, k_i}$. Hence:

$$\mathbf{A}_{\beta,\, s}\,\mathbf{x} = \sum_{i=1}^{r} \sum_{m=0}^{\min(s,\, k_i-1)} \binom{s}{m} (\lambda_i - \beta)^{s-m} \mathbf{A}_{\lambda_i,\, m}\,\mathbf{x}_i.$$

The *columns* of $\mathbf{A}_{\beta,\, s}$ are obtained by letting $\mathbf{x}$ vary across the *standard basis* vectors.

The case $\mathbf{A}_{0,\, s}$ is the power $\mathbf{A}^s$ of $\mathbf{A}$.

the minimal polynomial of a matrix

Assume **A** is a **nxn** matrix with *eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_r$
of *algebraic multiplicities* $k_1, k_2, \ldots, k_r$.

For each **i** define $\alpha(\lambda_i)$, the *null index* of λ_i, to be the
smallest positive integer α such that $\mathbf{N}_{\lambda_i,\, \alpha} = N_{\lambda_i,\, k_i}$.

It is often the case that $\alpha(\lambda_i) < k_i$.

Then $\mathbf{p(x)} = \prod (\mathbf{x} - \lambda_i)^{\alpha(\lambda_i)}$ is the *minimal polynomial* for **A**.

To see this note $\mathbf{p(A)} = \prod \mathbf{A}_{\lambda_i,\alpha(\lambda_i)}$ and the factors can be commuted in any order.

So $\mathbf{p(A)}\,(\mathbf{N}_{\lambda_j,\, k_j}) = \{0\}$, because $\mathbf{A}_{\lambda_j,\alpha(\lambda_j)}\,(\mathbf{N}_{\lambda_j,\, k_j}) = \{0\}$. Being that

$\mathbf{C}^n = N_{\lambda_1,\, k_1} \oplus N_{\lambda_2,\, k_2} \oplus \ldots \oplus N_{\lambda_r,\, k_r}$, it is clear $\mathbf{p(A)} = \mathbf{0}$.

Now $\mathbf{p(x)}$ can not be of less degree because $\mathbf{A}_{\beta,\, 1}\,(\mathbf{N}_{\lambda_j,\, k_j}) = \mathbf{N}_{\lambda_j,\, k_j}$,

when $\beta \neq \lambda_j$, and so $\mathbf{A}_{\lambda_j,\alpha(\lambda_j)}$ must be a factor of $\mathbf{p(A)}$, for each **j**.

using confluent Vandermonde matrices

An alternative strategy is to use the *characteristic polynomial* of matrix **A**.

Let $\mathbf{p(x)} = a_0 + a_1\,x + a_2\,x^2 + \ldots + a_{n-1}\,x^{n-1} + x^n$

be the *characteristic polynomial* of **A**.

The *minimal polynomial* of **A** can be substituted for **p(x)** in this discussion, if it is known,
and different, to reduce the degree **n** and the multiplicities of the eigenvalues.

Then $\mathbf{p(A) = 0}$ and $\mathbf{A^n = -(a_0 I + a_1 A + a_2 A^2 + ... + a_{n-1} A^{n-1})}$.

So $\mathbf{A^{n+m} = b_{m,0} I + b_{m,1} A + b_{m,2} A^2 + ... + b_{m,n-1} A^{n-1}}$,

where the $\mathbf{b_{m,0}, b_{m,1}, b_{m,2}, ..., b_{m,n-1}}$, satisfy the recurrence relation

$\mathbf{b_{m,0} = -a_0 b_{m-1,n-1}}$,
$\mathbf{b_{m,1} = b_{m-1,0} - a_1 b_{m-1,n-1}}$,
$\mathbf{b_{m,2} = b_{m-1,1} - a_2 b_{m-1,n-1}}$,
...,
$\mathbf{b_{m,n-1} = b_{m-1,n-2} - a_{n-1} b_{m-1,n-1}}$

with $\mathbf{b_{0,0} = b_{0,1} = b_{0,2} = ... = b_{0,n-2} = 0}$, and $\mathbf{b_{0,n-1} = 1}$.

This alone will reduce the number of multiplications needed to calculate a higher power of **A** by a factor of $\mathbf{n^2}$, as compared to simply multiplying $\mathbf{A^{n+m}}$ by **A**.

In fact the $\mathbf{b_{m,0}, b_{m,1}, b_{m,2}, ..., b_{m,n-1}}$, can be calculated by a formula.

Consider first when **A** has *distinct eigenvalues* $\lambda_1, \lambda_2, ..., \lambda_n$.
Since $\mathbf{p(\lambda_i) = 0}$, for each **i**, the λ_i satisfy the recurrence relation also. So:

$$\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdot & \cdot & \cdot & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdot & \cdot & \cdot & \lambda_2^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \lambda_n & \lambda_n^2 & \cdot & \cdot & \cdot & \lambda_n^{n-1} \end{bmatrix} \begin{bmatrix} b_{m,0} \\ b_{m,1} \\ \cdot \\ \cdot \\ \cdot \\ b_{m,n-1} \end{bmatrix} = \begin{bmatrix} \lambda_1^{n+m} \\ \lambda_2^{n+m} \\ \cdot \\ \cdot \\ \cdot \\ \lambda_n^{n+m} \end{bmatrix}$$

The matrix **V** in the equation is the well studied *Vandermonde's*,
for which formulas for it's determinant and inverse are known.

$$\det(V(\lambda_1, \lambda_2, ..., \lambda_n)) = \prod_{1 \leq i < j \leq n} (\lambda_j - \lambda_i)$$

In the case that $\lambda_2 = \lambda_1$, consider instead when λ_1 is near λ_2, and subtract row **1** from row **2**, which does not affect the determinant.

$$\left[\begin{array}{cccccc|c} 1 & \lambda_1 & \lambda_1^2 & \cdot\ \cdot\ \cdot & \lambda_1^{n-1} & & \lambda_1^{n+m} \\ 0 & \lambda_2-\lambda_1 & \lambda_2^2-\lambda_1^2 & \cdot\ \cdot\ \cdot & \lambda_2^{n-1}-\lambda_1^{n-1} & & \lambda_2^{n+m}-\lambda_1^{n+m} \\ \cdot & \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot & & \cdot \\ 1 & \lambda_n & \lambda_n^2 & \cdot\ \cdot\ \cdot & \lambda_n^{n-1} & & \lambda_n^{n+m} \end{array}\right]$$

After dividing the second row by $(\lambda_2 - \lambda_1)$ the determinant will be affected by the removal of this factor and still be non-zero.

$$\left[\begin{array}{ccccc|c} 0 & \dfrac{\lambda_2-\lambda_1}{(\lambda_2-\lambda_1)} & \dfrac{\lambda_2^2-\lambda_1^2}{(\lambda_2-\lambda_1)} & \cdot\ \cdot\ \cdot & \dfrac{\lambda_2^{n-1}-\lambda_1^{n-1}}{(\lambda_2-\lambda_1)} & \dfrac{\lambda_2^{n+m}-\lambda_1^{n+m}}{(\lambda_2-\lambda_1)} \end{array}\right]$$

Taking the limit as $\lambda_1 \rightarrow \lambda_2$, the new system has the second row *differentiated*.

$$\left[\begin{array}{ccccc|c} 1 & \lambda_2 & \lambda_2^2 & \cdot\ \cdot\ \cdot & \lambda_2^{n-1} & \lambda_2^{n+m} \\ 0 & 1 & 2\lambda_2 & \cdot\ \cdot\ \cdot & (n-1)\lambda_2^{n-2} & (n+m)\lambda_2^{n+m-1} \\ 1 & \lambda_3 & \lambda_3^2 & \cdot\ \cdot\ \cdot & \lambda_3^{n-1} & \lambda_3^{n+m} \\ \cdot & \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot & \cdot \\ 1 & \lambda_n & \lambda_n^2 & \cdot\ \cdot\ \cdot & \lambda_n^{n-1} & \lambda_n^{n+m} \end{array}\right]$$

The new system has determinant:

$$\det(V(\lambda_2, \ldots, \lambda_n)) = \prod_{3 \leq j \leq n} (\lambda_j - \lambda_2)^2 \prod_{3 \leq i < j \leq n} (\lambda_j - \lambda_i)$$

In the case that $\lambda_3 = \lambda_2$, also, consider like before when λ_2 is near λ_3, and subtract row **1** from row **3**, which does not affect the determinant. Next divide row three by $(\lambda_3 - \lambda_2)$ and then subtract row **2** from the new row **3** and follow by dividing the resulting row **3** by $(\lambda_3 - \lambda_2)$ again. This will affect the determinant by removing a factor of $(\lambda_3 - \lambda_2)^2$.

Each element of row **3** is now of the form

$$((f(\lambda_3) - f(\lambda_2)) / (\lambda_3 - \lambda_2) - f'(\lambda_2)) / (\lambda_3 - \lambda_2)$$

and

$$((f(\lambda_3) - f(\lambda_2)) / (\lambda_3 - \lambda_2) - f'(\lambda_2)) / (\lambda_3 - \lambda_2) \rightarrow ½ f''(\lambda_3) \quad \text{as } \lambda_2 \rightarrow \lambda_3.$$

The effect is to differentiate twice and multiply by one half.

1	λ_3	λ_3^2	.	.	.	λ_3^{n-1}	λ_3^{n+m}
0	1	$2\lambda_3$	.	.	.	$(n-1)\,\lambda_3^{n-2}$	$(n+m)\,\lambda_3^{n+m-1}$
0	0	1	$3\lambda_3$	.	.	$½(n-1)(n-2)\,\lambda_3^{n-3}$	$½(n+m)(n+m-1)\,\lambda_3^{n+m-2}$
1	λ_4	λ_4^2	.	.	.	λ_4^{n-1}	λ_4^{n+m}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
1	λ_n	λ_n^2	.	.	.	λ_n^{n-1}	λ_n^{n+m}

The new system has determinant:

$$\det(V(\lambda_3, ..., \lambda_n)) = \prod_{4 \leq j \leq n} (\lambda_j - \lambda_3)^3 \prod_{4 \leq i < j \leq n} (\lambda_j - \lambda_i)$$

If it were that the multiplicity of the eigenvalue was even higher, then the next row would be differentiated three times and mutiplied by **1/3!**. The progression is $\mathbf{1/s!\ f^{(s)}}$, with the constant coming from the coefficients of the derivatives in the *Taylor* expansion. This being done for each *eigenvalue* of *algebraic multiplicity* greater than **1**.

example

The matrix $$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 6 & 3 & 2 & 0 & 0 \\ 10 & 6 & 3 & 2 & 0 \\ 15 & 10 & 6 & 3 & 2 \end{bmatrix}$$

has *characteristic polynomial* $\mathbf{p(x) = (x - 1)^2(x - 2)^3}$.

The $\mathbf{b_{m,0}, b_{m,1}, b_{m,2}, b_{m,3}, b_{m,4},}$ for which

$$\mathbf{A^{5+m} = b_{m,0}\, I + b_{m,1}\, A + b_{m,2}\, A^2 + b_{m,3}\, A^3 + b_{m,4}\, A^4},$$

satisfy the *confluent* Vandermonde system next.

$$\begin{bmatrix} 1 & 1 & 1^2 & 1^3 & 1^4 \\ 0 & 1 & 2\cdot 1 & 3\cdot 1^2 & 4\cdot 1^3 \\ 1 & 2 & 2^2 & 2^3 & 2^4 \\ 0 & 1 & 2\cdot 2 & 3\cdot 2^2 & 4\cdot 2^3 \\ 0 & 0 & 1 & 3\cdot 2 & 6\cdot 2^2 \end{bmatrix} \begin{bmatrix} b_{m,0} \\ b_{m,1} \\ b_{m,2} \\ b_{m,3} \\ b_{m,4} \end{bmatrix} = \begin{bmatrix} 1^{5+m} \\ (5+m)\cdot 1^{5+m-1} \\ 2^{5+m} \\ (5+m)\cdot 2^{5+m-1} \\ \tfrac{1}{2}(5+m)(5+m-1)\cdot 2^{5+m-2} \end{bmatrix}$$

$$\begin{bmatrix} b_{m,0} \\ b_{m,1} \\ b_{m,2} \\ b_{m,3} \\ b_{m,4} \end{bmatrix} = \begin{bmatrix} -16 & -8 & 17 & -10 & 4 \\ 48 & 20 & -48 & 29 & -12 \\ -48 & -18 & 48 & -30 & 13 \\ 20 & 7 & -20 & 13 & -6 \\ -3 & -1 & 3 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ (5+m) \\ 32\cdot 2^m \\ 16(5+m)\cdot 2^m \\ 4(5+m)(5+m-1)\cdot 2^m \end{bmatrix}$$

using difference equations

Returning to the recurrence relation for $\mathbf{b_{m,0}, b_{m,1}, b_{m,2}, ..., b_{m,n-1}}$,

$\mathbf{b_{m,0} = -a_0\, b_{m-1,n-1}}$,

$\mathbf{b_{m,1} = b_{m-1,0} - a_1\, b_{m-1,n-1}}$,

$\mathbf{b_{m,2} = b_{m-1,1} - a_2\, b_{m-1,n-1}}$,

...,

$\mathbf{b_{m,n-1} = b_{m-1,n-2} - a_{n-1}\, b_{m-1,n-1}}$

with $\mathbf{b_{0,0} = b_{0,1} = b_{0,2} = ... = b_{0,n-2} = 0}$, and $\mathbf{b_{0,n-1} = 1}$.

Upon substituting the first relation into the second,

$\mathbf{b_{m,1} = -a_0\, b_{m-2,n-1} - a_1\, b_{m-1,n-1}}$,

and now this one into the next $\mathbf{b_{m,2} = b_{m-1,1} - a_2\, b_{m-1,n-1}}$,

$\mathbf{b_{m,2}} = -a_0\, b_{m-3,n-1} - a_1\, b_{m-2,n-1} - a_2\, b_{m-1,n-1}$,

..., and so on, the following difference equation is found.

$\mathbf{b_{m,n-1}} =$

$-a_0\, b_{m-n,n-1} - a_1\, b_{m-n+1,n-1} - a_2\, b_{m-n+2,n-1} - ... - a_{n-2}\, b_{m-2,n-1} - a_{n-1}\, b_{m-1,n-1}$

with $\mathbf{b_{0,n-1} = b_{1,n-1} = b_{2,n-1} = ... = b_{n-2,n-1} = 0}$, and $\mathbf{b_{n-1,n-1} = 1}$.

See the subsection on *linear difference equations* for more explanation.

Chains of generalized eigenvectors

Some notation and results from previous sections are restated.

- **A** is a **nxn** matrix of complex numbers.
- $\mathbf{A_{\lambda,k} = (A - \lambda I)^k}$
- $\mathbf{N_{\lambda,k} = N((A - \lambda I)^k)} = N(A_{\lambda,k})$
- For $\mathbf{V_1 \cap V_2 = \{0\}}$, $\mathbf{V_1 \oplus V_2 = \{v_1 + v_2 : v_1 \in V_1}$ **and** $v_2 \in V_2\}$.

Assume $\mathbf{A}$ has *eigenvalues* $\lambda_1, \lambda_2, ..., \lambda_r$
of *algebraic multiplicities* $k_1, k_2, ..., k_r$.

For each $\mathbf{i}$ define $\alpha(\lambda_i)$, the *null index* of λ_i, to be the smallest positive integer α such that $N_{\lambda_i, \alpha} = N_{\lambda_i, k_i}$.

It is always the case that $\alpha(\lambda_i) \leq k_i$.

When $\alpha(\lambda) \geq 2$,

$N_{\lambda, 1} \subset N_{\lambda, 2} \subset ... \subset N_{\lambda, \alpha-1} \subset N_{\lambda, \alpha} = N_{\lambda, \alpha+1} = ...,$

$N_{\lambda, \alpha} \setminus \{0\} = \cup \ (N_{\lambda, m} \setminus N_{\lambda, m-1})$, for $m = 1, ..., \alpha$ and $N_{\lambda, 0} = \{0\}$.

$x \in N_{\lambda, m} \setminus N_{\lambda, m-1}$, if and only if $A_{\lambda, 1}\, x \in N_{\lambda, m-1} \setminus N_{\lambda, m-2}$

Define a ***chain*** of *generalized eigenvectors* to be a set
$\{ x_1, x_2, ..., x_m \}$ such that $x_1 \in N_{\lambda, m} \setminus N_{\lambda, m-1}$, and $x_{i+1} = A_{\lambda, 1}\, x_i$.

Then $x_m \neq 0$ and $A_{\lambda, 1}\, x_m = 0$.

When $x_1 \in N_{\lambda, 1} \setminus \{0\}$, $\{x_1\}$ can be, for the sake of not requiring extra terminology, considered *trivially* a *chain*.

When a *disjoint* collection of *chains* combined form a *basis set* for $N_{\lambda, \alpha(\lambda)}$, they are often referred to as *Jordan chains* and are the vectors used for the columns of a *transformation* matrix in the *Jordan canonical form*.

When a *disjoint* collection of *chains* that combined form a *basis set*, is needed that satisfy $\beta_{i+1} x_{i+1} = A_{\lambda, 1}\, x_i$, for some scalars β_i, *chains* as already defined can be scaled for this purpose.

What will be proven here is that such a *disjoint* collection of *chains* can always be constructed.

Before the proof is started, recall a few facts about *direct sums*.

When the notation $V_1 \oplus V_2$ is used, it is assumed $V_1 \cap V_2 = \{0\}$.

For $x = v_1 + v_2$ with $v_1 \in V_1$ **and** $v_2 \in V_2$, then $x = 0$,

if and only if $v_1 = v_2 = 0$.

In the discussion below
$\delta_i = \dim(N_{\lambda, i}) - \dim(N_{\lambda, i-1})$, with $\delta_1 = \dim(N_{\lambda, 1})$.

First consider when $N_{\lambda, 2} \setminus N_{\lambda, 1} \neq \{0\}$ **, Then a *basis* for** $N_{\lambda, 1}$ can be *extended* to a *basis* for $N_{\lambda, 2}$. If $\delta_2 = 1$, then there exists $x_1 \in N_{\lambda, 2} \setminus N_{\lambda, 1}$, such that $N_{\lambda, 2} = N_{\lambda, 1} \oplus \operatorname{span}\{x_1\}$. Let $x_2 = A_{\lambda, 1}\, x_1$. Then $x_2 \in N_{\lambda, 1} \setminus \{0\}$, with x_1 and x_2 *linearly independent*. If $\dim(N_{\lambda, 2}) = 2$, since $\{x_1, x_2\}$ is a *chain* we are through. Otherwise x_1, x_2 can be extended to a *basis* $x_1, x_2, ..., x_{\delta_1}$ for $N_{\lambda, 2}$. The sets $\{x_1, x_2\}, \{x_3\}, ..., \{x_{\delta_1}\}$ form a *disjoint* collection of *chains*. In the case that $\delta_2 > 1$, then there exist *linearly independent* $\mathbf{x_1, x_2, ..., x_{\delta_2} \in N_{\lambda, 2} \setminus N_{\lambda, 1}}$, such that

$\mathbf{N_{\lambda, 2} = N_{\lambda, 1} \oplus span\{x_1, x_2, ..., x_{\delta_2}\}}$. Let $\mathbf{y_i = A_{\lambda, 1} x_i}$.

Then $\mathbf{y_i \in N_{\lambda, 1} \setminus \{0\}}$, for $\mathbf{i = 1, 2, ..., \delta_2}$. To see the $\mathbf{y_1, y_2, ..., y_{\delta_2}}$

are *linearly independent*, assume that for some $\boldsymbol{\beta_1, \beta_2, ..., \beta_{\delta_2}}$,

that $\boldsymbol{\beta_1 y_1 + \beta_2 y_2 + ... + \beta_{\delta_2} y_{\delta_2} = 0}$, Then for $\boldsymbol{x = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{\delta_2} x_{\delta_2}}$,

$\mathbf{x \in N_{\lambda, 1}}$, and $\mathbf{x \in span\{x_1, x_2, ..., x_{\delta_2}\}}$, which implies that $\mathbf{x = 0}$, and

$\boldsymbol{\beta_1 = \beta_2 = ... = \beta_{\delta_2} = 0}$. Since $\mathbf{span\{y_1, y_2, ..., y_{\delta_2}\} \subseteq N_{\lambda, 1}}$, the vectors

$\mathbf{x_1, x_2, ..., x_{\delta_2}, y_1, y_2, ..., y_{\delta_2}}$ are a *linearly independent* set.

If $\boldsymbol{\delta_2} = \delta_1$, then the sets $\mathbf{\{x_1, y_1\}, \{x_2, y_2\}, ..., \{x_{\delta_2}, y_{\delta_2}\}}$ form a

disjoint collection of *chains* that when combined are a *basis set* for $\mathbf{N_{\lambda, 2}}$.

If $\boldsymbol{\delta_1} > \delta_2$, then $\mathbf{x_1, x_2, ..., x_{\delta_2}, y_1, y_2, ..., y_{\delta_2}}$ can be extended to a *basis*

for $\mathbf{N_{\lambda, 2}}$ by some vectors $\mathbf{x_{\delta_2+1}, ..., x_{\delta_1}}$ in $\mathbf{N_{\lambda, 1}}$, so that

$\mathbf{\{x_1, y_1\}, \{x_2, y_2\}, ..., \{x_{\delta_2}, y_{\delta_2}\}, \{x_{\delta_2+1}\}, ..., \{x_{\delta_1}\}}$

forms a *disjoint* collection of *chains*.

To reduce redundancy, in the next paragraph, when $\boldsymbol{\delta = 1}$ the notation $\mathbf{x_1, x_2, ..., x_\delta}$ will be understood simply to mean just $\mathbf{x_1}$ and when $\boldsymbol{\delta = 2}$ to mean $\mathbf{x_1, x_2}$.

So far it has been shown that, if *linearly independent*

$\mathbf{x_1, x_2, ..., x_{\delta_2} \in N_{\lambda, 2} \setminus N_{\lambda, 1}}$, are chosen, such that

$\mathbf{N_{\lambda, 2} = N_{\lambda, 1} \oplus span\{x_1, x_2, ..., x_{\delta_2}\}}$, then there exists a *disjoint*

collection of *chains* with each of the $\mathbf{x_1, x_2, ..., x_{\delta_2}}$ being the first member or *top*

of one of the *chains*. Furthermore, this collection of *vectors*, when combined,

forms a *basis* for $\mathbf{N_{\lambda, 2}}$.

Now, let the *induction hypothesis* be that, if *linearly independent*

$\mathbf{x_1, x_2, ..., x_{\delta_m} \in N_{\lambda, m} \setminus N_{\lambda, m-1}}$, are chosen, such that

$\mathbf{N_{\lambda, m} = N_{\lambda, m-1} \oplus span\{x_1, x_2, ..., x_{\delta_m}\}}$, then there exists a *disjoint*

collection of *chains* with each of the $\mathbf{x_1, x_2, ..., x_{\delta_m}}$ being the first member or *top*

of one of the *chains*. Furthermore, this collection of *vectors*, when combined,

forms a *basis* for $\mathbf{N_{\lambda, m}}$.

Consider $\mathbf{m < \boldsymbol{\alpha}(\boldsymbol{\lambda})}$. A *basis* for $\mathbf{N_{\lambda, m}}$ can always be *extended* to a *basis* for

$\mathbf{N_{\lambda, m+1}}$. So *linearly independent* $\mathbf{x_1, x_2, ..., x_{\delta_{m+1}} \in N_{\lambda, m+1} \setminus N_{\lambda, m}}$, such that

$\mathbf{N_{\lambda, m+1} = N_{\lambda, m} \oplus span\{x_1, x_2, ..., x_{\delta_{m+1}}\}}$, can be chosen. Let $\mathbf{y_i = A_{\lambda, 1} x_i}$.

Then $\mathbf{y_i \in N_{\lambda, m} \setminus N_{\lambda, m-1}}$, for $\mathbf{i = 1, 2, ..., \delta_{m+1}}$. To see the $\mathbf{y_1, y_2, ..., y_{\delta_{m+1}}}$

are *linearly independent*, assume that for some $\boldsymbol{\beta_1, \beta_2, ..., \beta_{\delta_{m+1}}}$, that $\beta_1 y_1 + \beta_2 y_2 + ... + \beta_{\delta_{m+1}} y_{\delta_{m+1}} = 0$, Then for

$x = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{\delta_{m+1}} x_{\delta_{m+1}}$, $x \in N_{\lambda,1}$, and

$x \in \mathrm{span}\{x_1, x_2, ..., x_{\delta_{m+1}}\}$, which implies that $x = 0$, and

$\beta_1 = \beta_2 = ... = \beta_{\delta_{m+1}} = 0$. In addition, $\mathrm{span}\{y_1, y_2, ..., y_{\delta_{m+1}}\} \cap N_{\lambda, m-1} = \{0\}$.

To see this assume that for some $\beta_1, \beta_2, ..., \beta_{\delta_{m+1}}$,

that $\beta_1 y_1 + \beta_2 y_2 + ... + \beta_{\delta_{m+1}} y_{\delta_{m+1}} \in N_{\lambda, m-1}$ Then for

$x = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{\delta_{m+1}} x_{\delta_{m+1}}$, $x \in N_{\lambda, m}$, and

$x \in \mathrm{span}\{x_1, x_2, ..., x_{\delta_{m+1}}\}$, which implies that $x = 0$, and

$\beta_1 = \beta_2 = ... = \beta_{\delta_{m+1}} = 0$. The proof is nearly done.

At this point suppose that $b_1, b_2, ..., b_{d_{m-1}}$ is any *basis* for $N_{\lambda, m-1}$.

Then $B = \mathrm{span}\{b_1, b_2, ..., b_{d_{m-1}}\} \oplus \mathrm{span}\{y_1, y_2, ..., y_{\delta_{m+1}}\}$

is a *subspace* of $N_{\lambda, m}$. If $B \neq N_{\lambda, m}$, then

$b_1, b_2, ..., b_{d_{m-1}}$, $y_1, y_2, ..., y_{\delta_{m+1}}$ can be *extended* to a *basis* for $N_{\lambda, m}$,

by some set of vectors $z_1, z_2, ..., z_{(\delta_m - \delta_{m+1})}$, in which case

$N_{\lambda, m} = N_{\lambda, m-1} \oplus \mathrm{span}\{y_1, y_2, ..., y_{\delta_{m+1}}\} \oplus \mathrm{span}\{z_1, z_2, ..., z_{(\delta_m - \delta_{m+1})}\}$.

If $\delta_m = \delta_{m+1}$, then

$N_{\lambda, m} = N_{\lambda, m-1} \oplus \mathrm{span}\{y_1, y_2, ..., y_{\delta_{m+1}}\}$

or if $\delta_m > \delta_{m+1}$, then

$N_{\lambda, m} = N_{\lambda, m-1} \oplus \mathrm{span}\{z_1, z_2, ..., z_{(\delta_m - \delta_{m+1})}, y_1, y_2, ..., y_{\delta_{m+1}}\}$

In either case apply the *induction hypothesis* to get that there exists a *disjoint* collection of *chains* with each of the $y_1, y_2, ..., y_{\delta_{m+1}}$ being the first member or *top* of one of the *chains*. Furthermore, this collection of *vectors*, when combined, forms a *basis* for $N_{\lambda, m}$. Now, $y_i = A_{\lambda, 1} x_i$, for $i = 1, 2, ..., \delta_{m+1}$, so each of the *chains* beginning with y_i can be extended upwards into $N_{\lambda, m+1} \setminus N_{\lambda, m}$ to a *chain* beginning with x_i. Since $N_{\lambda, m+1} = N_{\lambda, m} \oplus \mathrm{span}\{x_1, x_2, ..., x_{\delta_{m+1}}\}$, the *combined vectors* of the *new chains* form a *basis* for $N_{\lambda, m+1}$.

Differential equations *y′= Ay*

Let $\mathbf{A}$ be a $\mathbf{n \times n}$ matrix of complex numbers and λ an *eigenvalue* of $\mathbf{A}$, with *associated eigenvector* $\mathbf{x}$. Suppose $\mathbf{y(t)}$ is a $\mathbf{n}$ *dimensional vector valued* function*, sufficiently smooth, so that* $y'(t)$ *is continuous. The restriction that* $y(t)$ *be* smooth *can be relaxed somewhat, but is not the main focus of this discussion.*

The solutions to the equation $y'(t) = Ay(t)$ are sought. The first observation is that $\mathbf{y(t) = e^{\lambda t} x}$ will be a solution. When $\mathbf{A}$ does not have $\mathbf{n}$ *linearly independent* *eigenvectors*, solutions of this kind will not provide the total of **n** needed for a *fundamental basis set*.

In view of the existence of *chains* of *generalized eigenvectors* seek a solution of the form '$y(t) = e^{\lambda t} x_1 + t' e^{\lambda t} \mathbf{x_2}$, then
$\mathbf{y'(t)} =' \lambda e^{\lambda t} x_1 + e^{\lambda t} x_2 + \lambda t e^{\lambda t} x_2 =' e^{\lambda t}(\lambda \mathbf{x_1} + \mathbf{x_2})'''' + t e^{\lambda t}(\lambda x_2)$
and
$\mathbf{Ay(t) = e^{\lambda t} A x_1 + t e^{\lambda t} A x_2}$.

In view of this, $\mathbf{y(t)}$ will be a solution to $\mathbf{y'(t) = Ay(t)}$, when '$A x_1 = \lambda x_1 + x_2$' and $\mathbf{A x_2 = \lambda x_2}$. That is when $\mathbf{(A - \lambda I)x_1 = x_2}$ and $\mathbf{(A - \lambda I)x_2 = 0}$. Equivalently, when $\mathbf{\{x_1, x_2\}}$ is a *chain* of *generalized eigenvectors*.

Continuing with this reasoning seek a solution of the form
$\mathbf{y(t) = e^{\lambda t} x_1 + t e^{\lambda t} x_2} + t^2 e^{\lambda t} x_3$ **,** then
$\mathbf{y'(t)} = \lambda e^{\lambda t} x_1 + e^{\lambda t} x_2 + \lambda t e^{\lambda t} x_2 + 2 t e^{\lambda t} x_3 + \lambda t^2 e^{\lambda t} x_3$
$= e^{\lambda t}(\lambda x_1 + x_2) + t e^{\lambda t}(\lambda x_2 + 2 x_3) + t^2 e^{\lambda t}(\lambda x_3)$' and
$\mathbf{Ay(t) = e^{\lambda t} A x_1 + t e^{\lambda t} A x_2} + t^2 e^{\lambda t} A x_3$ **.**

Like before, $\mathbf{y(t)}$ will be a solution to $\mathbf{y'(t) = Ay(t)}$, when '$A x_1 = \lambda x_1 + x_2$' , '$A x_2 = \lambda \mathbf{x_2 + 2 x_3}$' , and $\mathbf{A x_3 = \lambda x_3}$. That is when $\mathbf{(A - \lambda I)x_1 = x_2}$, $\mathbf{(A - \lambda I)x_2 = 2 x_3}$, and $\mathbf{(A - \lambda I)x_3 = 0}$. Since it will hold $\mathbf{(A - \lambda I)(2 x_3) = 0}$, also, equivalently, when $\mathbf{\{x_1, x_2, 2 x_3\}}$ is a *chain* of *generalized eigenvectors*.

More generally, to find the progression, seek a solution of the form
$\mathbf{y(t) = e^{\lambda t} x_1 + t e^{\lambda t} x_2} + t^2 e^{\lambda t} x_3 + t^3 e^{\lambda t} x_4 + ... + t^{m-2} e^{\lambda t} x_{m-1} + t^{m-1} e^{\lambda t} x_m$ **,**
then
$\mathbf{y'(t)} = \lambda e^{\lambda t} x_1 + e^{\lambda t} x_2 + \lambda t e^{\lambda t} x_2' + 2 t e^{\lambda t} x_3' + \lambda t^2 e^{\lambda t} x_3' + 3 t^2 e^{\lambda t} x_4' + \lambda t^3 e^{\lambda t} x_4'$
$+ ...' + (m-2)t^{m-3} e^{\lambda t} x_{m-1}' + \lambda t^{m-2} e^{\lambda t} x_{m-1}' + (m-1)t^{m-2} e^{\lambda t} x_m' + \lambda t^{m-1} e^{\lambda t} x_m$
$=' e^{\lambda t}(\lambda x_1 + x_2) + t e^{\lambda t}(\lambda x_2 + 2 x_3) + t^2 e^{\lambda t}(\lambda x_3 + 3 x_4) + t^3 e^{\lambda t}(\lambda x_4 + 4 x_5)$
$+ ...$
$+ t^{m-3} e^{\lambda t}(\lambda x_{m-2} + (m-2) x_{m-1}) + t^{m-2} e^{\lambda t}(\lambda x_{m-1} + (m-1) x_m) + t^{m-1} e^{\lambda t}(\lambda x_m)$'
and
$\mathbf{Ay(t)} =$
$\mathbf{e^{\lambda t} A x_1 + t e^{\lambda t} A x_2} + t^2 e^{\lambda t} A x_3 + t^3 e^{\lambda t} A x_4 + ... + t^{m-2} e^{\lambda t} A x_{m-1} + t^{m-1} e^{\lambda t} A x_m$ **.**

Again, $\mathbf{y(t)}$ will be a solution to $\mathbf{y'(t) = Ay(t)}$, when
'$A x_1 = \lambda x_1 + x_2$' , $\mathbf{A x_2 = \lambda x_2 + 2 x_3}$, $\mathbf{A x_3 = \lambda x_3 + 3 x_4}$, $\mathbf{A x_4 = \lambda x_4 + 4 x_5}$,
$\mathbf{A x_{m-2} = \lambda x_{m-2} + (m-2) x_{m-1}}$, $\mathbf{A x_{m-1} = \lambda x_{m-1} + (m-1) x_m}$,

and $\mathbf{A\, x_m = \lambda\, x_m}$.

That is when

$\mathbf{(A - \lambda\, I)x_1 = x_2}$, $\mathbf{(A - \lambda\, I)x_2 = 2\, x_3}$, $\mathbf{(A - \lambda\, I)x_3 = 3\, x_4}$, $\mathbf{(A - \lambda\, I)x_4 = 4\, x_5}$,

...,

$\mathbf{(A - \lambda\, I)x_{m-2} = (m-2)\, x_{m-1}}$, $\mathbf{(A - \lambda\, I)x_{m-1} = (m-1)\, x_m}$, and

$\mathbf{(A - \lambda\, I)x_m = 0}$.

Since it will hold $\mathbf{(A - \lambda\, I)((m-1)!\, x_3) = 0}$, also, equivalently, when

$\{\mathbf{x_1, 1!\, x_2, 2!\, x_3, 3!\, x_4, ...,}\ (m-2)!\, x_{m-1}, (m-1)!\, x_m\}$

is a *chain* of *generalized eigenvectors*.

Now, the *basis set* for all solutions will be found through a *disjoint collection of chains of generalized eigenvectors* of the matrix **A**.

Assume **A** has *eigenvalues* $\mathbf{\lambda_1, \lambda_2, ..., \lambda_r}$
of *algebraic multiplicities* $\mathbf{k_1, k_2, ..., k_r}$.

For a given *eigenvalue* $\mathbf{\lambda_i}$ there is a *collection* of **s**, with **s** depending on **i**,
disjoint chains of *generalized eigenvectors*
$C_{i,1} = \{^1z_1, {}^1z_2, ..., {}^1z_{j1}\}$, $C_{i,2} = \{^2z_1, {}^2z_2, ..., {}^2z_{j2}\}$, ..., $C_{i,js(i)} = \{^sz_1, {}^sz_2, ..., {}^sz_{js}\}$,
that when *combined* form a *basis set* for $\mathbf{N_{\lambda_i, k_i}}$. The total number of *vectors* in this set will be $\mathbf{j1 + j2 + ... + js = k_i}$. Sets in this collection may have only one or two members so in this discussion understand the notation $\mathbf{\{^\beta z_1, {}^\beta z_2, ..., {}^\beta z_{j\beta}\}}$ will mean $\mathbf{\{^\beta z_1\}}$ when $\mathbf{j\beta = 1}$, and $\mathbf{\{^\beta z_1, {}^\beta z_2\}}$ when $\mathbf{j\beta = 2}$, and so forth.

Being that this *notation* is cumbersome with many *indices*, in the next paragraphs any particular $\boldsymbol{C_{i,\beta}}$, when more explanation is not needed, may just be notated as $C = \{z_1, z_2, ..., z_j\}$.

For each such of these *chain* sets, $\boldsymbol{C} = \mathbf{\{z_1, z_2, ..., z_j\}}$
the sets $\mathbf{\{x_j\}}$, $\mathbf{\{x_{j-1}, x_j\}}$, $\mathbf{\{x_{j-2}, x_{j-1}, x_j\}}$, **...,** $\mathbf{\{z_2, z_3, ..., z_j\}}$, $\mathbf{\{z_1, z_2, ..., z_j\}}$
are also *chains*. This notation being understood to mean when
$\boldsymbol{C} = \mathbf{\{z_1\}}$ just $\mathbf{\{z_1\}}$, when $\boldsymbol{C} = \mathbf{\{z_1, z_2\}}$ just $\mathbf{\{z_2\}, \{z_1, z_2\}}$ and when
$\boldsymbol{C} = \mathbf{\{z_1, z_2, z_2\}}$ just $\mathbf{\{z_3\}, \{z_2, z_3\}, \{z_1, z_2, z_3\}}$, and so on.

The conclusion of the top of the discussion was that
$\mathbf{y(t) = e^{\lambda t}x_1}$, is a solution when $\mathbf{\{x_1\}}$ is a *chain*.
$\mathbf{y(t) = e^{\lambda t}x_1 + t\, e^{\lambda t}x_2}$, is a solution when $\mathbf{\{x_1, 1!\, x_2\}}$ is a *chain*.
$\mathbf{y(t) = e^{\lambda t}x_1 + t\, e^{\lambda t}x_2} + t^2\, e^{\lambda t}x_3$, is a solution when $\mathbf{\{x_1, 1!\, x_2, , 2!\, x_3\}}$ is a *chain*.
The progression continues to
$\mathbf{y(t) = e^{\lambda t}x_1 + t\, e^{\lambda t}x_2} + t^2\, e^{\lambda t}x_3 + t^3\, e^{\lambda t}x_4 + ... + t^{m-2}\, e^{\lambda t}x_{m-1} + t^{m-1}\, e^{\lambda t}x_m$,

is a solution when $\{x_1, 1!\,x_2, 2!\,x_3, 3!\,x_4, ..., (m-2)!\,x_{m-1}, (m-1)!\,x_m\}$,

is a *chain* of *generalized eigenvectors*.

In light of the preceding calculations, all that must be done is to provide the proper *scaling* for each of the *chains* arising from the set $C = \{z_1, z_2, ..., z_j\}$.

The progression for the *solutions* is given by

$y(t) = e^{\lambda t} z_j$, for *chain* $\{z_j\}$

$y(t) = e^{\lambda t} z_{j-1} + (1/1!)\, t\, e^{\lambda t} z_j$, for *chain* $\{z_{j-1}, 1!(1/1!)\, z_j\}$

$y(t) = e^{\lambda t} z_{j-2} + (1/1!)\, t\, e^{\lambda t} z_{j-1} + (1/2!)\, t^2 e^{\lambda t} z_j$,

for *chain* $\{z_{j-2}, 1!(1/1!)\, z_{j-1}, 2!(1/2!)\, z_j\}$

$y(t) = e^{\lambda t} z_{j-3} + (1/1!)\, t\, e^{\lambda t} z_{j-2} + (1/2!)\, t^2 e^{\lambda t} z_{j-1} + (1/3!)\, t^3 e^{\lambda t} z_j$,

for *chain* $\{z_{j-3}, 1!(1/1!)\, z_{j-2}, 2!(1/2!)\, z_{j-1}, 3!(1/3!)\, z_j\}$,

and so on until,

$y(t) = e^{\lambda t} z_1 + (1/1!)\, t\, e^{\lambda t} z_2 + (1/2!)\, t^2 e^{\lambda t} z_3 + ... + (1/(j-1)!)\, t^{j-1} e^{\lambda t} z_j$,

for the *chain* of *generalized eigenvectors*,

$\{z_1, 1!(1/1!)\, z_2, 2!(1/2!)\, z_3, ..., (j-2)!(1/(j-2)!)\, x_{j-1}, (j-1)!(1/(j-1)!)\, z_j\}$.

What is left to show is that when all the *solutions* constructed from the *chain sets*, as described, are considered, they form a *fundamental set* of *solutions*. To do this it has to be shown that there are **n** of them and that they are *linearly independent*.

Reiterating, for a given *eigenvalue* λ_i there is a *collection* of **s**, with **s** depending on **i**,

disjoint chains of *generalized eigenvectors*

$C_{i,1} = \{{}^1z_1, {}^1z_2, ..., {}^1z_{j1(i)}\}$, $C_{i,2} = \{{}^2z_1, {}^2z_2, ..., {}^2z_{j2(i)}\}$,

..., $C_{i,js(i)} = \{{}^{s(i)}z_1, {}^{s(i)}z_2, ..., {}^{s(i)}z_{js(i)}\}$,

that when *combined* form a *basis set* for N_{λ_i, k_i}. The total number of *vectors* in this set will be $j1(i) + j2(i) + ... + js(i) = k_i$.

Thus the total number of all such *basis vectors* and so *solutions* is

$k_1 + k_2 + ... + k_r = n$.

Each solution is one of the forms $y(t) = e^{\lambda t} x_1$, $y(t) = e^{\lambda t} x_1 + t\, e^{\lambda t} x_2$,

$y(t) = e^{\lambda t} x_1 + t\, e^{\lambda t} x_2 + t^2 e^{\lambda t} x_3$, $y(t) = e^{\lambda t} x_1 + t\, e^{\lambda t} x_2 + t^2 e^{\lambda t} x_3 + ...$

Now each *basis vector* v_j, for $j = 1, 2, ..., n$; of the *combined* set of *generalized eigenvectors*, occurs as x_1 in one of the expressions immediately above ***precisely once***. That is, for each **j**, there is one $y_j(t) = e^{\lambda t} v_j + ...$

Since $y_j(0) = e^{\lambda 0} v_j = v_j$, the set of *solutions* are *linearly independent* at $t = 0$.

Revisiting the powers of a matrix

As a notational convenience $\mathbf{A_{\lambda,\,0} = I}$.

Note that $\mathbf{A} = \lambda\,\mathbf{I} + \mathbf{A_{\lambda,\,1}}$. and apply the *binomial theorem.*

$$\mathbf{A^s} = (\lambda\,\mathbf{I} + \mathbf{A_{\lambda,\,1}})^{\mathbf{s}} = \sum_{\mathbf{r=0}}^{\mathbf{s}} \binom{\mathbf{s}}{\mathbf{r}} \lambda^{\mathbf{s-r}} \mathbf{A_{\lambda,\,r}}$$

Assume $\boldsymbol{\lambda}$ is an *eigenvalue* of $\mathbf{A}$, and let $\{\mathbf{x_1, x_2, ..., x_m}\}$

be a ***chain*** of *generalized eigenvectors* such that $\mathbf{x_1 \in N_{\lambda,\,m} \setminus N_{\lambda,\,m-1}}$,

$\mathbf{x_{i+1} = A_{\lambda,\,1}\,x_i}$,. $\mathbf{x_m \neq 0}$, and $\mathbf{A_{\lambda,\,1}\,x_m = 0}$.

Then $\mathbf{x_{r+1} = A_{\lambda,\,r}\,x_1}$, for $\mathbf{r = 0, 1, ..., m\text{-}1}$.

$$\mathbf{A^s\,x_1} = \sum_{\mathbf{r=0}}^{\mathbf{s}} \binom{\mathbf{s}}{\mathbf{r}} \lambda^{\mathbf{s-r}} \mathbf{A_{\lambda,\,r}\,x_1} = \sum_{\mathbf{r=0}}^{\mathbf{s}} \binom{\mathbf{s}}{\mathbf{r}} \lambda^{\mathbf{s-r}}\,\mathbf{x_{r+1}}$$

So for $\mathbf{s \leq m-1}$

$$\mathbf{A^s\,x_1} = \sum_{\mathbf{r=0}}^{\mathbf{s}} \binom{\mathbf{s}}{\mathbf{r}} \lambda^{\mathbf{s-r}}\,\mathbf{x_{r+1}}$$

and for $\mathbf{s \geq m-1}$, since $\mathbf{A_{\lambda,\,m}\,x_1 = 0}$,

$$\mathbf{A^s\,x_1} = \sum_{\mathbf{r=0}}^{\mathbf{m-1}} \binom{\mathbf{s}}{\mathbf{r}} \lambda^{\mathbf{s-r}}\,\mathbf{x_{r+1}}\,.$$

Ordinary linear difference equations

Ordinary *linear difference equations* are equations of the sort:

$y_n = a\,y_{n-1} + b$

$y_n = a\,y_{n-1} + b\,y_{n-2} + c$

or more generally,

$y_n = a_m y_{n-1} + a_{m-1} y_{n-2} + ... + a_2 y_{n-m+1} + a_1 y_{n-m} + a_0$

with initial conditions

$y_0,\ y_1,\ y_2,\ ...,\ y_{m-2},\ y_{m-1}$.

A case with $\mathbf{a_1 = 0}$ can be excluded, since it represents an equation of less degree.

They have a characteristic polynomial

$p(x) = x^m - a_m x^{m-1} - a_{m-1} x^{m-2} - ... - a_2 x - a_1$.

To solve a *difference equation* it is first observed, if $\mathbf{y_n}$ and $\mathbf{z_n}$ are both solutions, then $\mathbf{(y_n - z_n)}$ is a solution of the *homogeneous* equation:

$y_n = a_m y_{n-1} + a_{m-1} y_{n-2} + ... + a_2 y_{n-m+1} + a_1 y_{n-m}.$

So a *particular* solution to the *difference equation* must be found together with all solutions of the *homogeneous* equation to get the *general solution* for the *difference equation*. Another observation to make is that, if $\mathbf{y_n}$ is a solution to the *inhomogeneous* equation, then

$\mathbf{z_n = y_{n+1} - y_n}$

is also a solution to the *homogeneous* equation.

So all solutions of the *homogeneous* equation will be found first.

When $\boldsymbol{\beta}$ is a root of $\mathbf{p(x) = 0}$, then it is easily seen

$\mathbf{y_n = \boldsymbol{\beta}^n}$ is a solution to the *homogeneous* equation since

$y_n - a_m y_{n-1} - a_{m-1} y_{n-2} - ... - a_2 y_{n-m+1} - a_1 y_{n-m},$

becomes upon the substitution $\mathbf{y_n = \boldsymbol{\beta}^n}$,

$\beta^n - a_m\beta^{n-1} - a_{m-1}\beta^{n-2} - ... - a_2\beta^{n-m+1} - a_1\beta^{n-m}$

$= \beta^{n-m}(\beta^m - a_m\beta^{m-1} - a_{m-1}\beta^{m-2} - ... - a_2\beta - a_1)$

$= \beta^{n-m}p(\beta) = 0.$

When $\boldsymbol{\beta}$ is a repeated root of $\mathbf{p(x) = 0}$, then

$\mathbf{y_n = n\boldsymbol{\beta}^{n-1}}$ is a solution to the *homogeneous* equation since

$n\beta^{n-1} - a_m(n-1)\beta^{n-2} - a_{m-1}(n-2)\beta^{n-3} - ... - a_2(n-m+1)\beta^{n-m} - a_1(n-m)\beta^{n-m-1}$

$= (n-m)\beta^{n-m-1}(\beta^m - a_m\beta^{m-1} - a_{m-1}\beta^{m-2} - ... - a_2\beta - a_1)$

$+ \beta^{n-m-1}(m\beta^{m-1} - (m-1)a_m\beta^{m-2} - (m-2)a_{m-1}\beta^{m-3} - ... - 2a_3\beta - a_2)$

$== (n-m)\beta^{n-m-1}p(\beta) + \beta^{n-m-1}p'(\beta) == 0.$

After reaching this point in the calculation the *mystery* is solved. Just notice when $\boldsymbol{\beta}$ is a root of $\mathbf{p(x) = 0}$ with *mutiplicity* $\mathbf{k}$, then for $\mathbf{s = 1, 2, ..., k-1}$

$d^s(\beta^{n-m}p(\beta))/d\beta^s = 0.$

Referring this back to the original equation

$\beta^n - a_m\beta^{n-1} - a_{m-1}\beta^{n-2} - ... - a_2\beta^{n-m+1} - a_1\beta^{n-m}$

it is seen that

$y_n = d^s(\beta^n)/d\beta^s$

are solutions to the *homogeneous* equation. For example, if $\boldsymbol{\beta}$ is a root of *multiplicity* $\mathbf{3}$, then $\mathbf{y_n = n(n-1)\boldsymbol{\beta}^{n-2}}$ is a solution. In any case this gives $\mathbf{m}$ *linearly independent* solutions to the *homogeneous* equation.

To look for a *particular solution* first consider the simpliest equation.

$y_n = a\, y_{n-1} + b.$

It has a *particular* solution $\mathbf{y_{p,n}}$ given by

$y_{p,0} = 0,\ y_{p,1} = b,\ y_{p,2} = (1 + a)b,\ \ldots,\ y_{p,n} = (1 + a + a^2 + \ldots + a^{n-1})b,\ \ldots,\ .$

It's *homogeneous* equation $\mathbf{y_n = a\, y_{n-1}}$ has solutions $\mathbf{y_n = a^n y_0}$.

So $\mathbf{z_n = y_{n+1} - y_n = a^n b}$

can be *telescoped* to get

$y_n = (y_n - y_{n-1}) + (y_{n-1} - y_{n-2}) + \ldots + (y_2 - y_1) + (y_1 - y_0) + y_0$

$= z_{n-1} + z_{n-2} + \ldots + z_1 + z_0 + y_0$

$= (1 + a + a^2 + \ldots + a^{n-1})b$ **,**

the *particular* solution with $\mathbf{y_0 = 0}$.

Now, returning to the general problem, the equation

$y_n = a_m y_{n-1} + a_{m-1} y_{n-2} + \ldots + a_2 y_{n-m+1} + a_1 y_{n-m} + a_0.$

When $\mathbf{y_{p,n}}$ is a *particular* solution with $\mathbf{y_{p,0} = 0}$, then

$\mathbf{z_n = y_{p,n+1} - y_{p,n}}$

is a solution to the *homogeneous* equation with $\mathbf{z_0 = y_{p,1}}$.

So $\mathbf{z_n = y_{p,n+1} - y_{p,n}}$

can be *telescoped* to get

$y_{p,n} = (y_{p,n} - y_{p,n-1}) + (y_{p,n-1} - y_{p,n-2}) + \ldots + (y_{p,2} - y_{p,1}) + (y_{p,1} - y_{p,0}) + y_{p,0}$

$= z_{n-1} + z_{n-2} + \ldots + z_1 + z_0$

Considering

$y_{p,m} = a_m y_{p,m-1} + a_{m-1} y_{p,m-2} + \ldots + a_2 y_{p,1} + a_1 y_{p,0} + a_0.$

and rewriting the equation in the $\mathbf{z_i}$

$z_{m-1} + z_{m-2} + \ldots + z_1 + z_0$

$= (a_m)\,(z_{m-2} + z_{m-3} + \ldots + z_1 + z_0) + (a_{m-1})\,(z_{m-3} + z_{m-4} + \ldots + z_1 + z_0)$

$+ (a_{m-2})\,(z_{m-4} + z_{m-5} + \ldots + z_1 + z_0)$

$+ \cdots$

$+ (a_3)\,(z_1 + z_0) + (a_2)\,(z_0) + (a_0)$

and

z_{m-1}

$= (a_m - 1)\, z_{m-2} + (a_m + a_{m-1} - 1)\, z_{m-3} + (a_m + a_{m-1} + a_{m-2} - 1)\, z_{m-4}$

$+ \cdots$

$+ (a_m + a_{m-1} + \ldots + a_4 + a_3 - 1)\, z_1 + (a_m + a_{m-1} + \ldots + a_3 + a_2 - 1)\, z_0$

$+ (a_0).$

Since a solution of the *homogeneous* equation can be found for any *initial conditions*

$z_0,\ z_1,\ z_2,\ \ldots,\ z_{m-2},\ z_{m-1}.$

reasoning *conversely* find such $\mathbf{z_i}$ satisfying the equation,

just before and define $\mathbf{y_{p,n}}$ by the relation

$y_{p,0} = 0,\ \ y_{p,n} = z_{n-1} + z_{n-2} + \ldots + z_1 + z_0$

One choice is, for example, $z_{m-1} = a_0$, $z_0 = z_1 = z_2 = ... = z_{m-2} = 0$.
This solution solves the problem for all *initial values* equal to *zero*.

The *general solution* to the *inhomogeneous* equation is given by
$\mathbf{y_n = y_{p,n} + \gamma_1\, w(1)_n} + \gamma_2\, w(2)_n + ... + \gamma_{m-1}\, w(m-1)_n + \gamma_m\, w(m)_n$
where
$\mathbf{w(1)_n,\ w(2)_n,\ ...,\ w(m-1)_n, w(m)_n}$
are a *basis* for the *homogeneous* equation, and
$\boldsymbol{\gamma_1,\ \gamma_2,\ ...,\ \gamma_{m-1},\ \gamma_m}$
are *scalars*.

example

$y_n = 8\, y_{n-1} - 25\, y_{n-2} + 38\, y_{n-3} - 28\, y_{n-4} + 8\, y_{n-5} + 1$
with initial conditions
$y_0 = 0$, $y_1 = 0$, $y_2 = 0$, $y_3 = 0$, **and** $y_4 = 0$.

The *characteristic polynomial* for the equation is
$p(x) = x^5 - 8x^4 + 25x^3 - 38x^2 + 28x - 8 = (x-1)^2(x-2)^3$.

The *homogeneous* equation has *independent* solutions
$w1_n = 1^n = 1$, $w2_n = n\cdot 1^{n-1} = n$, and
$w3_n = 2^n$, $w4_n = n\cdot 2^{n-1}$, $w5_n = n(n-1)\cdot 2^{n-2}$.
The solution to the *homogeneous* equation
$z_n = -3\, w1_n - w2_n + 3\, w3_n - 2\, w4_n + ½\, w5_n$
satisfies the *initial conditions*
$z_4 = 1$, $z_0 = z_1 = z_2 = z_3 = 0$.
A *particular solution* can be found by
$\mathbf{y_{p,0} = 0}$, $\mathbf{y_{p,n}} = z_{n-1} + z_{n-2} + ... + z_1 + z_0$.

Calculating sums:
$\sum w1 = w1_{n-1} + w1_{n-2} + ... + w1_1 + w1_0 = n$.
$\sum w2 = w2_{n-1} + w2_{n-2} + ... + w2_1 + w2_0 = (n-1)n / 2$.
$\sum w3 = w3_{n-1} + w3_{n-2} + ... + w3_1 + w3_0 = 2^n - 1$.
Sums of these kinds are found by differentiating $\mathbf{(x^n - 1) / (x - 1)}$.
$\sum w4 = w4_{n-1} + w4_{n-2} + ... + w4_1 + w4_0 = (n-2)2^{n-1} + 1$.
$\sum w5 = w5_{n-1} + w5_{n-2} + ... + w5_1 + w5_0 = (n^2 - 5n + 8)2^{n-2} - 2$.

Now,
$y_{p,n} = -3 \sum w1_n - \sum w2_n + 3 \sum w3_n - 2 \sum w4_n + ½ \sum w5_n$
solves the *initial value problem* of this example.

At this point it is worthwhile to notice that all the *terms* that are combinations of *scalar multiples* of *basis elements* can be removed. These are any multiples of

$1,\ n,\ 2^n,\ n\cdot 2^{n-1},$ and $\mathbf{n^2 \cdot 2^{n-2}}$**.**

So instead the *particular* solution next, may be preferred.

$y_{p,n} = -\tfrac{1}{2} n^2$.

This solution has *non zero* initial values, which must be taken into account.

$y_0 = 0,\ y_1 = -1/2,\ y_2 = -2,\ y_3 = -9/2,$ **and** $y_4 = -8.$

References

- Axler, Sheldon (1997). *Linear Algebra Done Right* (2nd ed.). Springer. ISBN 978-0-387-98258-8.