Advanced Concepts in Linear Algebra:
An exploration of special types of matrices and their applications to sums of powers

By Alexander Jones
American University
Spring 2009
Special Thanks to Dan Kalman

Part I: A Brief Review of Basic Concepts

One of the most useful tools in mathematics is the linear algebraic concept of the matrix. A matrix, put simply, is a rectangular array of values. Because this paper will focus primarily on special types of matrices and their applications, it will first be useful to identify a standard notation for our matrix, which we will call A.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ a_{m1} & & & a_{mn} \end{bmatrix}$$

In this case each *a* can be any value stored within the matrix. Notice that we use *i* and *j* to represent the row and column of the entry, respectively, with the row *i* always written first. We can also refer to the dimension of a matrix, generally $m \times n$, which means a matrix has *m* rows and *n* columns. For example, a 3×2 matrix would have three rows and two columns:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

As with any useful mathematical tool, matrices have operations associated with them, the most basic of which are matrix addition and matrix multiplication. Matrix addition is simply the addition of corresponding entries in matrices of the same dimensions. For example, if we take two 3×2 matrices, *A* and *B* and add them using matrix addition, we get:

$$A + B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \\ a_{31} + b_{31} & a_{32} + b_{32} \end{bmatrix}.$$

Notice that the resulting matrix after matrix addition is the same dimension as the two matrices being added. It is also useful to note that when multiplying a matrix by a scalar value, which is to say any value with magnitude but without direction, each entry is simply multiplied by said scalar. For example, if we multiply the matrix *A* by some scalar *c* we get:

$$cA = c\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \\ ca_{31} & ca_{32} \end{bmatrix}$$

We can also multiply two matrices together through a process known as matrix multiplication. For matrix multiplication to be applicable, we must have the same number of rows in one matrix as columns in the other matrix. For example, let *A* be a 2×3 matrix and *B* be a 3×2 matrix. Then multiplying A and B through matrix multiplication, we have:

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

We can see that this yields a 2×2 matrix, and in general matrix multiplication yields a matrix with the number of rows in the first matrix by the number of columns in the second matrix. You will notice that if *A* and *B* happened to be square matrices, the resulting matrix would be a square matrix of the same dimensions. It can easily be shown that matrix multiplication, in general, is not commutative.

Through matrix multiplication we understand a matrix *A* is a linear mapping. In particular we will be most concerned with square matrices which map a vector from $\mathbf{R}^n$ (the n-dimensional set of real numbers) to $\mathbf{R}^n$. This will be further illustrated later.

Let us quickly exemplify matrix multiplication in terms of powers of square matrices, while at the same time introducing a very simple but special type of matrix known as the nilpotent matrix *N*, which contains all zero entries except 1's along its superdiagonal (the values just above the diagonal). Let us examine powers of *N* through matrix multiplication.

$$N = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$N^2 = NN = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$N^{n-1} = NN\ldots N = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

We can see that each successive power of N the 1's move one diagonal higher, until in the *n-1* case only the top corner contains a 1. In the n case we get the 0 matrix. This will be important later.

Another useful (and for the purposes of this paper, necessary) matrix operation is known as the determinant. A determinant is only applicable for square matrices, and it yields a single number as a result. This number may be thought of us an expansion (or contraction) factor of the linear transformation being applied by the matrix. If the determinant of a matrix is 2, for example, it means that the measure (length, area, volume, etc. depending on what dimensions we are dealing with) of the new vector, after the linear transformation has been applied, will be twice as great as the original vector. We will be primarily concerned with calculating determinants, which can be a difficult task. For two-dimensional matrices, the formula is relatively simple. For a matrix A,

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{1}$$

For higher dimensional matrices the formula becomes increasingly complicated and is often best left to computer software. However, we can use a method known as expansion by cofactors, which works as follows. (We will work with a 4×4 matrix but the application to higher dimensions should be apparent.) We choose any row or column of the matrix. In this case let us choose the first row. We then write each term of that row with alternating signs and multiply each by a subdeterminant. To find the subdeterminant we mentally draw a line through the row and column of our original matrix which contain the entry in question. We then put together the remaining rows and columns to form a matrix one dimension smaller. This is illustrated below.

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{14}\begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix}$$

We have now broken down our original determinant into individual terms multiplied by the determinants of smaller matrices. We repeat this process with the smaller determinants until we arrive at an equation featuring the determinants of 2×2 matrices, which we can calculate directly using formula (1). We attempt to choose rows and columns featuring as many zero terms as possible, since obviously those zero terms will make the subdeterminants inconsequential. While we usually leave the larger determinants to computer software, it will be necessary to understand the concept of expansion by cofactors later in this paper.

Now armed with some basic operations we can proceed to discuss a few special types of matrices, how they relate to one another, as well as their applications. In this paper we will examine companion matrices and the Jordan canonical form of matrices, specifically the Jordan canonical form of companion matrices. We will then use those concepts as well as a type of generalized Vandermonde matrix to eventually derive a formula to find sums of powers. While we will consider this power series solution to be the end point of this paper, we should not ignore the significance of the results used to reach that point. I should say in fact that the true goal of this paper is simply to examine some interesting results in linear algebra while using the power series to tie them all together in the end.

Part II: Eigenvectors, Eigenvalues, Characteristic Polynomials, Minimal Polynomials, and Linear Independence

We agreed in section I that a square matrix A maps a vector from $\mathbf{R}^n$ to $\mathbf{R}^n$. In most cases this results in a change of direction for the vector. However, for many a matrix *A* there exists a certain number of nonzero vectors whose direction, when A is applied to them, remains unchanged. We call these vectors eigenvectors, and they satisfy the equation:

$$Ax = \lambda x$$

where λ is some scalar value, which we refer to as an eigenvalue. Any non-zero vector x that satisfies this equation is referred to as an eigenvector, and the result of applying the linear transformation is simply that the vector x scaled up or down by the factor λ.

It becomes necessary at this point to establish a polynomial associated with a square matrix, which we refer to as the characteristic polynomial of the matrix. We can find a formula for the characteristic polynomial as follows. Assume λ is an eigenvalue of A. Then there must exist an eigenvector x such that

$$Ax = \lambda x$$

or

$$(A - \lambda I)x = 0$$

where I is the identity matrix. Because we know by definition that x is non-zero, the determinant of $(\lambda I - A)$ must equal 0, as shown in reference [1]. This yields the equation for the characteristic polynomial:

$$\det(A - \lambda I) = 0$$

This will produce a polynomial in terms of λ set equal to 0, which when solved will produce the eigenvalues of the matrix A as roots of the characteristic polynomial.

Now given a matrix A we are able to find the eigenvalues λ associated with it by using its characteristic polynomial. Once an eigenvalue is known, the equation $Ax = \lambda x$ or $(\lambda I - A)x = 0$ will be easily solvable for x in order to find the eigenvector(s) associated with the eigenvalue λ. Note that a matrix may have more than one eigenvalue associated with it and each eigenvalue has its own eigenvector(s) associated with it.

At this point it becomes prudent to discuss the idea of linear dependence. A set of vectors is considered linearly dependent if any of the vectors can be written as a combination of any of the other vectors. To show this, let $v_1, v_2, \ldots, v_n$ be vectors. Then to be linearly dependent we must have

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0$$

where at least some value(s) of c_i must be nonzero. Therefore if this is not true, that is if all of the values of c_i must be equal to 0, then the vectors $v_1, v_2, \ldots, v_n$ are linearly independent.

Another noteworthy concept is that of the minimal polynomial of a matrix. We established above that the characteristic polynomial $p(\lambda)$ is a polynomial associated with a square matrix A which is annihilated by the eigenvalues of the matrix. By The Cayley-Hamilton Theorem described in reference [2], the matrix A itself will also be a root of its characteristic polynomial. If $p(A)$ is the smallest degree polynomial for which this is true, we say that $p(A)$ is the minimal polynomial of the matrix A. This is not true for matrices in general, and one way to check whether the characteristic polynomial is indeed the minimal polynomial is to show that the $n\text{-}1$ powers of the matrix A (including A^0 which is the identity matrix I) are linearly independent. We will use this method in the next section when discussing Companion Matrices.

Part III: Companion Matrices

Let $p(t)$ be a polynomial in the general form such that

$$p(t) = t^n + c_{n-1}t^{n-1} + \ldots + c_1 t + c_0 .$$

We have a special type of matrix associated with $p(t)$ known as its companion matrix. The companion matrix takes the following form:

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{n-1} \end{bmatrix}$$

As you can see, the companion matrix has 1-entries along the superdiagonal and 0-entries everywhere else until the bottom row. In the bottom row the negatives of the coefficients from the polynomial $p(t)$ are simply listed in order from the constant term in ascending order up to the coefficient corresponding to the n-1 degree term of the polynomial.

From here we are going to prove that the characteristic polynomial of a companion matrix is basically the same as the polynomial $p(\lambda)$. We begin with our proposition that given the above polynomial and matrix, for $n \geq 2$:

$$\det(C - \lambda I) = (-1)^n \left(c_0 + c_1\lambda + \ldots + c_{n-1}\lambda^{n-1} + \lambda^n\right) = (-1)^n p(\lambda)$$

We will proceed with our proof through induction, beginning with the $n=2$ case. Then for $p(t) = c_0 + c_1 t + t^2$:

$$C_p = \begin{bmatrix} 0 & 1 \\ -c_0 & -c_1 \end{bmatrix}$$

So,

$$\det(C_p - \lambda I) = \det\left(\begin{bmatrix} -\lambda & 1 \\ -c_0 & -\lambda - c_1 \end{bmatrix}\right)$$

$$= c_0 + c_1\lambda + \lambda^2 = (-1)^2\left(c_0 + c_1\lambda + \lambda^2\right) = (-1)^2\, p(\lambda)$$

So the proposition holds for *n=2*.

Now we suppose the proposition is true for *n=k*, that is for:

$$q(t) = c_0 + c_1 t + \cdots + c_{k-1} t^{k-1} + t^k$$

and

$$C_q = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{k-1} \end{bmatrix},$$

$$\det(C_q - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -\lambda - c_{k-1} \end{vmatrix}$$

$$= (-1)^k\left(c_0 + c_1 t + \cdots + c_{k-1} t^{k-1} + t^k\right) = (-1)^k\, q(\lambda)\,.$$

Assuming this is true for the n=k case we now examine a polynomial of degree *k+1* where:

$$p(t) = c_0 + c_1 t + \cdots + c_k t^k + t^{k+1}$$

and

$$C_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_k \end{bmatrix}.$$

$$\det(C_p - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -\lambda - c_k \end{vmatrix}$$

Then, expanding by cofactors down the first column, we find:

$$\det(C_p - \lambda I) = (-\lambda)\det\begin{bmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_1 & -c_2 & -c_3 & \cdots & -\lambda - c_k \end{bmatrix} + (-1)^{k+1} c_0$$

Notice that the $k \times k$ subdeterminant matrix is of the same form as $(C_q - \lambda I)$ where

$$q(t) = c_1 + c_2 t + \cdots + c_k t^{k-1} + t^k$$

So by the induction assumption we know that

$$\det(C_q - \lambda I) = (-1)^k \left(c_1 + c_2 t + \cdots + c_k t^{k-1} + t^k\right) = (-1)^k q(\lambda)$$

Thus,

$$\det(C_p - \lambda I) = (-1)^{k+1} c_0 + (-\lambda)(-1)^k q(\lambda)$$
$$= (-1)^{k+1}\left[c_0 + \lambda(c_1 + \ldots + c_k \lambda^{k-1} + \lambda^k)\right] = (-1)^{k+1} p(\lambda)$$

So the proposition holds for $n=k+1$ when it holds for $n=k$. By the principle of induction, then, the proposition holds for all $n \geq 2$.◊

It should now be apparent that the characteristic polynomial for a companion matrix is very special in that it requires no calculation and so one can easily associate a polynomial to a companion matrix and vice versa. Now what about the minimal polynomial of a companion matrix *C*? It turns out that in the case of companion matrices the characteristic polynomial is always equal to the minimal polynomial. Let us examine this concept further.

Say we have a companion matrix as described previously, so that its characteristic polynomial looks like this:

$$\det(C - \lambda I) = (-1)^n \left(c_0 + c_1 \lambda + \ldots + c_{n-1} \lambda^{n-1} + \lambda^n\right)$$

The Cayley-Hamilton Theorem states that a matrix, in this case C, is the root of its own characteristic polynomial. That is:

$$c_0 + c_1 C + \ldots + c_{n-1} C^{n-1} + C^n = 0$$

This shows immediately that the matrix *C* does in fact annihilate the polynomial *p(C),* which means that either *p(C)* is the minimal polynomial or there exists a lower degree polynomial for which *C* is also a root. Suppose this is true; then the set $\{C^0, C^1, C^2, \ldots, C^{n-1}\}$ must be linearly dependent, because if

$$a_0 C^0 + a_1 C^1 + a_2 C^2 + \ldots + a_{n-1} C^{n-1} = 0$$

then that means

$$a_0 C^0 + a_1 C^1 + a_2 C^2 + \ldots + a_{n-1} C^{n-1} - a_i C^i = -a_i C^i$$

which would show that one of the elements of the set can indeed be written as a linear combination of other elements.

So for this set to be linearly dependent means there must exist some coefficients $a_0, a_1, \ldots, a_{n-1}$, where $a_i \neq 0$ for at least some *i*, such that

$$a_0 C^0 + a_1 C^1 + a_2 C^2 + \ldots + a_{n-1} C^{n-1} = 0\,.$$

Let us examine this again with the powers of *C* represented in matrix form:

$$a_0 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} + a_1 \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ * & * & \cdots & * & * \end{bmatrix} + a_2 \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ * & * & \cdots & * & * \\ * & * & \cdots & * & * \end{bmatrix} + \ldots$$
$$+ a_{n-1} \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \\ * & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & * \\ * & * & \cdots & * & * \\ * & * & \cdots & * & * \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We leave some entries marked as $*$ indicating that those entries may obtain any value, because, as it turns out, those entries are insignificant. Let us examine the equation one more time except only looking at the top rows of each matrix:

$$a_0 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix} + a_1 \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \end{bmatrix} + \ldots$$
$$+ a_{n-1} \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

It should be obvious that because C^0 is the only matrix with a non-zero entry in the 1st column of row 1, a_0 must equal 0 in order to achieve a 0 in the 1st column of row 1 of the zero matrix. Similarly since C^1 will be the only matrix with a 0 in the 2nd column of row 1, a_1 must be 0. Tracing this line of reasoning all the way to a_{n-1} it should be apparent that $a_i = 0$ for all *i*. However this is contrary to our requirement for linear dependence, which means $\{C^0, C^1, C^2, \ldots, C^{n-1}\}$ must be linearly independent. Since the *(n-1)*th powers of *C* are linearly independent, then there can be no polynomial of degree less than *n* for which *C* is a root. Therefore *p(C)* must be the minimal polynomial for which *C* annihilates *p*.◊

We have now shown that the characteristic polynomial of a companion matrix is easy to find and that this characteristic polynomial must also be equal to the minimal polynomial. We now turn our attention to an interesting result regarding eigenvectors and companion matrices. Let us suppose that an eigenvector *v* for a companion matrix *C* appears as follows:

$$v = \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \vdots \\ \lambda^{n-1} \end{bmatrix}$$

for some eigenvalue λ. For this to be true, *v* must satisfy the equation

$$Cv = \lambda v\,.$$

So we check.

$$Cv = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{n-1} \end{bmatrix} \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \vdots \\ \lambda^{n-1} \end{bmatrix} = \begin{bmatrix} \lambda \\ \lambda^2 \\ \lambda^3 \\ \vdots \\ -c_0 - c_1\lambda - \ldots - c_{n-1}\lambda^{n-1} \end{bmatrix}$$

Notice the bottom row is equal to $-\left(p(\lambda) - \lambda^n\right) = -p(\lambda) + \lambda^n$, and, since λ by definition is a root of *p(λ)*, $p(\lambda) = 0$. So continuing the above equation…

$$\begin{bmatrix} \lambda \\ \lambda^2 \\ \lambda^3 \\ \vdots \\ -c_0 - c_1\lambda - \ldots - c_{n-1}\lambda^{n-1} \end{bmatrix} = \begin{bmatrix} \lambda \\ \lambda^2 \\ \lambda^3 \\ \vdots \\ \lambda^n \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \vdots \\ \lambda^{n-1} \end{bmatrix} = \lambda v\,.$$

And we see that our proposed eigenvector does indeed satisfy the necessary equation.

We have now shown that in addition to an easily calculable characteristic polynomial, which happens to equal the minimal polynomial, companion matrices also have easy to calculate eigenvectors once we find the eigenvalues, which are simply the roots of the characteristic polynomial. Therefore we see that in the special case of companion matrices we can determine both the characteristic and minimal polynomials, as well as the eigenvalues and eigenvectors, with minimal calculations. These special properties will come into play shortly. For now, however, we move on to examine the Jordan Canonical Form.

Part IV: Jordan Canonical Form

By this point it may be apparent that many of the operations we have discussed in earlier sections can be time consuming to compute by hand when dealing with a matrix in its generalized form. It is obvious that the more zero-entries in a matrix the easier it is to work with. However, equally obvious should be the restrictions of confining ourselves to those specific matrices which are filled with zero-entries. In this section we discuss a pleasant compromise. It turns out that while we cannot change a matrix per se, we can alter its form to accommodate our computational needs. In fact we can "diagonalize" some matrices so that their only non-zero entries occur along the diagonal. This makes things much simpler. We are particularly interested in the simplicity of the powers of a diagonal matrix, which by applying the matrix multiplication as described earlier, should be apparent:

$$\begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & t_n \end{bmatrix}^n = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & t_n \end{bmatrix}\begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & t_n \end{bmatrix}\cdots = \begin{bmatrix} t_1{}^n & 0 & \cdots & 0 \\ 0 & t_2{}^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & t_n{}^n \end{bmatrix}$$

It is also noteworthy that the determinant of a diagonal matrix is simply the product of its diagonal values, or $t_1 t_2 \ldots t_n$. In some cases we can diagonalize a matrix through a change of basis. However, not all matrices are diagonalizable. It turns out in fact that a matrix is diagonalizable if it has *n* distinct eigenvalues. We are more concerned with a more general form of diagonalization which can be applied to any matrix. This is known as the Jordan canonical form, and it appears as a compilation of blocks known as Jordan blocks, which appear as follows:

$$\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}$$

Each Jordan block takes this form, with the particular eigenvalue repeating down the diagonal, 1's on the superdiagonal, and 0's elsewhere. A Jordan block may be as small as a single entry (in the case of an eigenvalue which occurs only once) or its dimension may be as high as the number of repetitions of the eigenvalue. The Jordan canonical form is then the diagonal combination of these Jordan blocks.

It turns out that algebraically the Jordan canonical form of a matrix *A*, which we will call *J*, is related to *A* by an equation of the form:

$$A = PJP^{-1}$$

or

$$J = P^{-1}AP$$

where *A* is our original matrix, *P* is what we call a "transition matrix" composed of the eigenvectors of *A*, and P^{-1} is the inverse of *P* (the inverse of any matrix *P* is the matrix P^{-1} such that $PP^{-1} = P^{-1}P = I$). The proof of this equation can be found in nearly any linear algebra textbook including many of the sources listed in the reference section. In the context of this paper, though, we are more concerned with converting a matrix to its Jordan canonical form. To demonstrate this let us work through a step-by-step example for a 4×4 matrix and in the process we shall aim to develop an algorithm that can be applied to a matrix of any size. First, let us choose a matrix:

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

Step 1: We need to know the eigenvalues, so we compute the characteristic polynomial.

$$\det(A - \lambda I) = \begin{vmatrix} -1-\lambda & -3 & 3 & 0 \\ -2 & -1-\lambda & 0 & -3 \\ -2 & 0 & -1-\lambda & -3 \\ 0 & 2 & -2 & -1-\lambda \end{vmatrix} = (-1-\lambda)^4$$

This example is already in a beautifully factored form, so we know in this case there is only one eigenvalue, which is -1. Notice that because we have $(-1-\lambda)^4$ the eigenvalue -1 repeats 4 times. This is known as the algebraic multiplicity of the eigenvalue and it tells us how many times total the eigenvalue must appear in the Jordan form.

Step 2: Next we must find the linearly independent corresponding eigenvectors to each eigenvalue. In this case we have only one eigenvalue, so we work with that. Recall that we said an eigenvector must satisfy the equation $(A - \lambda I)x = 0$, so

$$(A-\lambda I)=(A-(-1)I)=\begin{bmatrix} -1+1 & -3 & 3 & 0 \\ -2 & -1+1 & 0 & -3 \\ -2 & 0 & -1+1 & -3 \\ 0 & 2 & -2 & -1+1 \end{bmatrix}=\begin{bmatrix} 0 & -3 & 3 & 0 \\ -2 & 0 & 0 & -3 \\ -2 & 0 & 0 & -3 \\ 0 & 2 & -2 & 0 \end{bmatrix}$$

So we have

$$\begin{bmatrix} 0 & -3 & 3 & 0 \\ -2 & 0 & 0 & -3 \\ -2 & 0 & 0 & -3 \\ 0 & 2 & -2 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}=0$$

and so we set about row reducing this matrix, which results in the matrix

$$\begin{bmatrix} 2 & 0 & 0 & 3 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which gives us a system:

$$\begin{bmatrix} 2 & 0 & 0 & 3 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Therefore we know the following equations hold:

$$\begin{aligned} 2x_1+3x_4 &= 0 \\ x_2-x_3 &= 0 \end{aligned}$$

or

$$\begin{aligned} 2x_1 &= -3x_4 \\ x_2 &= x_3 \end{aligned}$$

Thus

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

And so we see we get two distinct eigenvectors:

$$v_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \\ -2 \end{bmatrix}; u_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

If we plug these into the above equation we can indeed confirm that these are eigenvectors. However, we have only two independent eigenvectors and we are going to need a 4×4 transition matrix, which means four total basis vectors. Note that the number of unique eigenvectors associated with an eigenvalue is known as its geometric multiplicity and it tells us how many Jordan blocks in the Jordan form matrix. So in this case we know we need two Jordan blocks in *J*.

Step 3: We must find two more basis vectors. We call these generalized eigenvectors, and they must satisfy the equation:

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

which, for each successive generalized eigenvector, appears as follows, described in reference [2]:

$$(A - \lambda I) v_{k+1} = v_k$$

So we repeat the same process as above except this time instead of setting our row reduction equal to 0 we set it equal to v_k. We also add v_k as a row so that the resulting generalized eigenvector will be orthogonal to our original eigenvector. So, beginning first with v_1

$$\begin{bmatrix} 0 & -3 & 3 & 0 \\ -2 & 0 & 0 & -3 \\ -2 & 0 & 0 & -3 \\ 0 & 2 & -2 & 0 \\ 3 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ -2 \\ 0 \end{bmatrix}$$

So this time when we row reduce our matrix we augment it with v_1.

$$\left[\begin{array}{cccc|c} 0 & -3 & 3 & 0 & 3 \\ -2 & 0 & 0 & -3 & 0 \\ -2 & 0 & 0 & -3 & 0 \\ 0 & 2 & -2 & 0 & -2 \\ 3 & 0 & 0 & -2 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Which yields the system

$$\begin{aligned} &x_1 = 0 \\ &1x_2 - 1x_3 = -1 \\ &x_4 = 0 \end{aligned}$$

And from that we can see our generalized eigenvector

$$v_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \end{bmatrix}$$

We now attempt to find a v_3 by augmenting with v_2:

$$\left[\begin{array}{cccc|c} 0 & -3 & 3 & 0 & 0 \\ -2 & 0 & 0 & -3 & 1 \\ -2 & 0 & 0 & -3 & 2 \\ 0 & 2 & -2 & 0 & 0 \\ 3 & 0 & 0 & -2 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

But this yields $0 = 1$ which is an inconsistent solution, so we know there are no more generalized eigenvectors to be found from v, but we still need one more eigenvector, so we move on to find a u_2 and repeat the same steps as with v_2 by adding a row to ensure the eigenvector is orthogonal to u_1 and augmenting with u_1:

$$\left[\begin{array}{cccc|c} 0 & -3 & 3 & 0 & 0 \\ -2 & 0 & 0 & -3 & 1 \\ -2 & 0 & 0 & -3 & 1 \\ 0 & 2 & -2 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 2 & 0 & 0 & 3 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

which yields the system

$$
\begin{aligned}
2x_1 + 3x_4 &= -1 \\
x_2 &= 0 \\
x_3 &= 0
\end{aligned}
$$

And we can see the eigenvector

$$u_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}.$$

At this point we have all four eigenvectors required to form a 4×4 matrix, so we can stop looking.

Step 4: We now form our transition matrix by combining our four eigenvectors, making sure that the generalized eigenvector(s) follow their corresponding basic eigenvectors, like so:

$$P = \begin{bmatrix} v_1 & v_2 & u_1 & u_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & -1 \end{bmatrix}$$

Step 5: We now find P^{-1} by augmenting P with the identity matrix and row reducing:

$$[P|I] = \left[\begin{array}{cccc|cccc} 3 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1/5 & 0 & 0 & 1/5 \\ 0 & 1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 & 2/5 & 0 & 0 & 3/5 \end{array}\right] = [I|P^{-1}]$$

Step 6: Finally we can use our transition matrix and its inverse to find our Jordan form.

$$J = P^{-1}AP = \begin{bmatrix} 1/5 & 0 & 0 & 1/5 \\ 0 & -1 & 1 & 0 \\ 0 & 2 & -1 & 0 \\ 2/5 & 0 & 0 & 3/5 \end{bmatrix} \begin{bmatrix} -1 & -3 & 3 & 0 \\ -2 & -1 & 0 & -3 \\ -2 & 0 & -1 & -3 \\ 0 & 2 & -2 & -1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

As we can see J is composed of two separate Jordan blocks of the form described above, as predicted by the geometric multiplicity. We should also note that if we had no repeating eigenvalues, that is n distinct eigenvalues, since each one must have a unique associated eigenvector, there would be n Jordan blocks, which, as should be apparent, would result in a diagonal matrix. Thus we see that a diagonal matrix is merely a special case of the Jordan canonical form.

Another noteworthy fact about Jordan blocks is this: the smallest integer k such that $(A-\lambda I)^k = 0$ is called the index of λ and is also the size of the largest Jordan block for λ. Now, recall that we established previously that the characteristic polynomial of a companion matrix is always equal to the minimal polynomial. This means the degree of λ in the characteristic polynomial must be the smallest degree for which λ is indeed a root. This, in turn, tells us that the index k always equals the algebraic multiplicity n in the case of companion matrices, which means that the largest Jordan block for the eigenvalue of a companion matrix is equal to the total number of appearances of that eigenvalue in the Jordan matrix. Therefore we can conclude that there is exactly one Jordan block for each eigenvalue in the Jordan canonical form matrix of a companion matrix. This makes the Jordan form of a companion matrix extremely easy to calculate.

We will now apply the foregoing ideas to derive a nifty formula to solve sums of powers.

Part V: Sums of Powers using Matrix Methods

A difference equation determines a sequence that is defined recursively; for example, let us examine the sums of powers sequence, that is

$$\{s_n^r\} = \sum_{k=0}^{n} k^r$$

We can see that this sequence can be recursively defined by

$$s_{n+1}^r = s_n^r + (n+1)^r$$

or, we can use a lag operator L to define the sequence in terms of a function. We define L such that

$$La_n = a_{n+1}$$

which can be further expressed as

$$p(L)\{a_n\} = 0$$

In reference [3] a method is then derived such that we can see:

$$(L-1)^{r+2} a_n = 0 \tag{2}$$

We can also create a characteristic polynomial for an order *k* which expresses a term of the sequence $\{a_n\}$ as a function of the preceding *k* terms and *n*, as shown:

$$a_{n+k} + c_{k-1}a_{n+k-1} + \ldots + c_0 a_n = 0$$

We see that this characteristic polynomial fits easily into one of our handy companion matrices, like so:

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{k-1} \end{bmatrix}$$

We can also represent our *k* entries of the sequence $\{a_n\}$ as a vector:

$$v_n = \begin{bmatrix} a_n \\ a_{n+1} \\ a_{n+2} \\ \vdots \\ a_{n+k-1} \end{bmatrix}$$

which can be thought of as a sample window containing the desired *k* entries from the infinite column:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \end{bmatrix}$$

We can now see that the effect of multiplying Cv_n is that the *a* entries in our window shift one spot up in our infinite column, in the same way we demonstrated the eigenvectors for a companion matrix previously. We can represent this algebraically by saying,

$$v_{n+1} = Cv_n$$

or

$$v_n = C^n v_0$$

since it should be apparent that v_0 is the starting window, and each successive time we apply C^n the window is shifted down by one value.

Now let us assume we are looking for one particular term in the sequence, a_n. Then we want only the n^{th} term in the vector $v_n = C^n v_0$, so we can obtain that by creating a formula for a_n as follows:

$$a_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{k-1} \end{bmatrix}^n \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{bmatrix}.$$

This way multiplying the row matrix on the left will eliminate all but the first row of the power of the companion matrix, which will in turn eliminate all but the desired term in the column vector. This is a problem, however, since the powers of companion matrices can be difficult to calculate directly. Fortunately, we know that every matrix is similar to a matrix in Jordan canonical form, so we can rewrite C as:

$$C^n = PJP^{-1}PJP^{-1}\ldots PJP^{-1} = PJ^nP^{-1}$$

since we know $P^{-1}P = I$. Also, because C is a companion matrix, we established previously that each of its eigenvalues has exactly one Jordan block associated with it, which makes J easy to represent as soon as we know the eigenvalues of C. Note from equation (2) that the characteristic polynomial can be expressed as $(t-1)^{r+2}$, since we know $a_n = 1$ because the n^{th} degree term of our characteristic polynomial has no other coefficient in front. This tells us we have a single eigenvalue of -1 with an algebraic multiplicity of $(r+2)$. Thus J is easy to formulate:

$$J = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

which we can see is easily rewritten as the sum of the identity matrix plus a nilpotent matrix:

$$J = I + N = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

which can be summarized with the equation (see reference [3]):

$$J^n = (I+N)^n = \sum_{j=0}^{r+1} \binom{n}{j} N^j$$

$$\text{so } J^n = (I+N)^n = \begin{bmatrix} 1 & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{r+1} \\ 0 & 1 & \binom{n}{1} & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & \binom{n}{2} \\ \vdots & \vdots & \vdots & \ddots & \binom{n}{1} \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

So we now know how to calculate J^n. We still need to know our transition matrix P and its inverse. This is a case of the generalized Vandermonde matrix described in source [4] and P looks like this:

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{r+1}{0} & \binom{r+1}{1} & \binom{r+1}{2} & \cdots & 1 \end{bmatrix}$$

which is a special case of the general form:

$$M(\lambda) = \left(\binom{i-1}{j-1} \lambda^{i-j} \right)_{ij} .$$

One interesting property of $M(\lambda)$ as described in source [5] is that $M(\lambda)^{-1} = M(-\lambda)$, which in this case means

$$P^{-1} = M(-1) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pm\binom{r+1}{0} & \mp\binom{r+1}{1} & \pm\binom{r+1}{2} & \cdots & 1 \end{bmatrix}$$

or basically the same as *P* except with negative signs in front of each odd subdiagonal.

We now know:

$$s_n^r = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix} P J^n P^{-1} \begin{bmatrix} s_0^r \\ s_1^r \\ s_2^r \\ \vdots \\ s_{r+1}^r \end{bmatrix}$$

However, because the row matrix out front will eliminate all but the first row of *P* and the first row of *P* also happens to be $\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$ we get all but the first row of J^n eliminated as well. Therefore we can simplify the above equation to:

$$s_n^r = \begin{bmatrix} \binom{n}{0} & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{r+1} \end{bmatrix} P^{-1} \begin{bmatrix} s_0^r \\ s_1^r \\ s_2^r \\ \vdots \\ s_{r+1}^r \end{bmatrix}.$$

Now we can also replace the initial terms of $\{s_n^r\}$ with the initial terms of $\{n^r\}$ by observing that

$$\begin{bmatrix} s_0^r \\ s_1^r \\ s_2^r \\ \vdots \\ s_{r+1}^r \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 0^r \\ 1^r \\ 2^r \\ \vdots \\ (r+1)^r \end{bmatrix}.$$

So substituting we now see:

$$s_n^r = \begin{bmatrix} \binom{n}{0} & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{r+1} \end{bmatrix} P^{-1} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 0^r \\ 1^r \\ 2^r \\ \vdots \\ (r+1)^r \end{bmatrix}$$

Let us first examine

$$P^{-1} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pm\binom{r+1}{0} & \mp\binom{r+1}{1} & \pm\binom{r+1}{2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$
= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \mp\binom{r+1}{1} & \pm\binom{r+1}{2} & \cdots & 1 \end{bmatrix} = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & P^* \end{array}\right]
$$

where P^* is the submatrix as shown, and we partition for simplicity's sake.

Now we can plug in our partitioned matrix, and express our other matrices in partitioned form for ease of calculations, as so:

$$
s_n^r = \left[\begin{array}{c|cccc} \binom{n}{0} & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{r+1} \end{array}\right] \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & P^* \end{array}\right] \left[\begin{array}{c} 0 \\ \hline 1^r \\ 2^r \\ \vdots \\ (r+1)^r \end{array}\right]
$$

$$
= \begin{bmatrix} \binom{n}{1} & \binom{n}{2} & \binom{n}{3} & \cdots & \binom{n}{r+1} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pm 1 & \mp\binom{r}{1} & \pm\binom{r}{2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1^r \\ 2^r \\ 3^r \\ \vdots \\ (r+1)^r \end{bmatrix}
$$

which gives us a calculable equation for any sequence of sums of powers. For example, let us plug in $r = 4$. Then

$$s_n^4 = \begin{bmatrix} \binom{n}{1} & \binom{n}{2} & \binom{n}{3} & \binom{n}{4} & \binom{n}{5} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ -1 & 3 & -3 & 1 & 0 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 16 \\ 81 \\ 256 \\ 625 \end{bmatrix}$$

$$= \binom{n}{1} + 15\binom{n}{2} + 50\binom{n}{3} + 60\binom{n}{4} + 24\binom{n}{5}.$$

And so we have reached our conclusion; by combining and applying the linear algebra tools we have discussed, plus bringing in a couple additional tools which we did not have the time to discuss at length (but which are expanded upon in the sources listed) we are able to reduce the sum of any power series to a relatively easy to solve equation of binomial coefficients.

Part VI: References

[1] David C. Lay. *Linear Algebra and its Applications*, 3rd ed. Pearson, 2006.

[2] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA: SIAM, 2000.

[3] Dan Kalman. "Sums of Powers by Matrix Methods". 1988.

[4] Dan Kalman. "The Generalized Vandermonde Matrix." *Mathematics Magazine*. 57.1 (1984): 15-21.

[5] Dan Kalman. "Polynomial Translation Groups." *Mathematics Magazine*. 56.1 (1983): 23-25.