The Art of Polynomial Interpolation

The Art of Polynomial Interpolation

STUART MURPHY

Contents

Introduction

The inspiration for this text grew out of a simple question that emerged over a number of years of teaching math to Middle School, High School and College students.

Practically speaking, what is the origin of a particular polynomial?

So much time is spent analyzing, factoring, simplifying and graphing polynomials that it is easy to lose sight of the fact that polynomials have a wealth of practical uses. Exploring the techniques of interpolating data allows us to view the development and birth of a polynomial. This text is focused on laying a foundation for understanding and applying several common forms of polynomial interpolation. The principal goals of the text are:

1. Breakdown the process of developing polynomials to demonstrate and give the student a feel for the process and meaning of developing estimates of the trend (s) a collection of data may represent.
2. Introduce basic matrix algebra to assist students with understanding the process without getting bogged down in purely manual calculations. Some manual calculations have been included, however, to assist with understanding the concept.
3. Assist students in building a basic foundation allowing them to add additional techniques, of which there are many, not covered in this text.

What this text is not:

<u>It is not a comprehensive survey of interpolation techniques.</u>

The techniques presented are ones the author believes will provide a basic understanding of polynomial interpolation that students can build upon. There are many flavors and sub flavors of interpolation and I encourage students who are interested to check them out.

<u>It is not a lesson in using interpolation apps:</u>

Quite the opposite. By engaging in exercising calculations, the student is better equipped to understand how and why these techniques work.

What is polynomial interpolation?

<u>We experience information in discreet ways.</u>

Typically, it comes from measurements or observations. However, what we often want to do is look at a continuous process the data represents; all at once or at least at any point we choose. While we cannot represent a continuous process with a single number we can do so with an equation. Graphically this equation could be a single point (not usually that interesting). A straight line (degree one polynomial), a curved line (degree two polynomial) that we often call a quadratic equation or parabola; or some higher degree that graphically, often begins to look like a wave repeating itself.

When dealing with data the specific numbers always represent a snapshot. For example, if we measure rainfall and wind speed each day for a year, we have a collection of data points that compare rainfall to wind speed. We might ask if there is a relationship between wind speed and rainfall. For example, hurricane force winds are usually accompanied by heavy rainfall. It would be nice to develop an equation that can predict rainfall when high winds are expected. Normally someone analyzing this data would plot the points on the x, y coordinate plane. In this text, the sample data used to illustrate the various interpolation methods will be plotted in this way.

Can the math stand alone? Most certainly not. The challenge for someone utilizing interpolation techniques is to apply expertise and experience to determine the most appropriate polynomial structure. In other words, is the model most likely to accurately (or at least reasonably) produce useful estimated values? This is what I mean by the “Art” of polynomial interpolation.

Interpolation uses a known set of independent and dependent values to estimate other dependent values, typically along a continuous line represented by a polynomial. Technically if you use the model to identify additional data points outside of the range of the given points this is known as extrapolation. Our focus will be on interpolation within the given range.

Adaptations of the techniques we explore have been used in pre-computer times to generate tables of trig or log values used in applications such as navigation. Nowadays they are adapted for use by computers and calculators and they are an important part of the tool kit researchers use to predict future events such as emerging storms tracks, climate change, political elections, changing demographics, spread of disease, and so forth.

We will explore five Interpolation techniques: Elimination (Substitution), Newton's Divided Difference, Splines, Least Squares and Taylor Series.

Techniques

A Brief Explanation of the Techniques Presented in This Text

A) Elimination (Substitution) (or solving a linear system of n equations with n unknowns)

Essentially this technique utilizes a process known to high school Algebra One students: Elimination (or substitution). This allows for solving a set of n-unknown variables in a set of n-equations.

B) Newton's Divided Difference

Newton's Divided Difference interpolation has many applications. Historically it and similar techniques have been used to develop trigonometric and logarithmic tables.

An important advantage is that if you start out with a handful of known points plotted on a coordinate plane you can decide on an appropriate degree polynomial that would be representative of the general trend. A benefit is that any new given points can easily be added one at a time thus increasing the degree of the polynomial for each new point added, without having to start over.

C) Splines

Splines interpolation is a great technique to employ if there are certain discreet points that modify the nature of the trend. For example, the trajectory of a rocket launch could be broken into segments: Launch to Stage One separation, interval to Stage Two separation, a major scheduled course adjustment and so on. Each of the resulting intervals could be represented by a separate polynomial. Spline interpolation creates just such separate polynomials while at the same time recognizing the continuity inherent in the event and building that into the resulting set of equations that collectively represent the event. A variation of this would be a single spline developed in a sub-interval of the domain that is of particular interest.

D) Least Squares Regression

Polynomial Least Squares regression is useful for fitting a polynomial such as a quadratic equation to many data points ensuring that each point influences the resulting polynomial in such a manner that the resulting polynomial is considered a best fit for that set of data.

E) Taylor Series Polynomial

A technique that employs use of the Taylor Series to develop a polynomial that approximates the actual function at and near a given domain value. It does not require a set of data points. The major limitation is that it works for a limited class of functions.

Note there are plenty of applications that will provide the desired results very quickly. However, this textbook is meant to assist students with an understanding of the computations and reasons for them. Included are the manual calculations with explanation as well as basic Matrix commands that students can use to mirror the manual calculations.

Let's look at a simple example.

Assumption: The faster a car is driven the lower the fuel efficiency.

Sample Vehicle Fuel Efficiency Measurements

X (Miles per Hour) MPH	Y (Miles per Gallon) MPG
45	43
55	42
65	38
75	32

Long Description

Chart of Sample Vehicle Speed

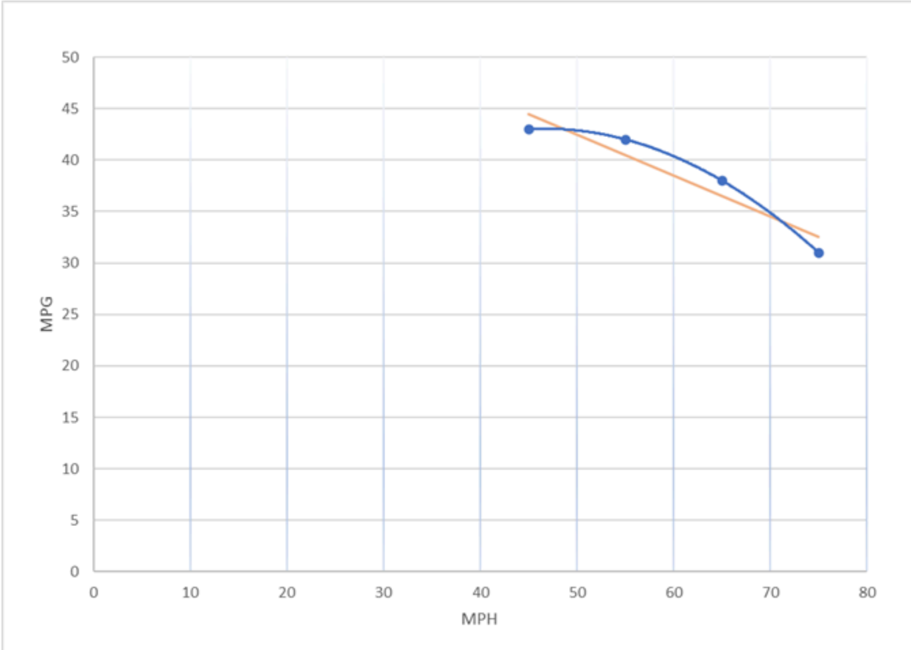


Figure 1 – Comparing Linear to Quadratic Interpolation Methods

Long Description

Figure 1 - Comparing Linear to Quadratic Interpolation Methods

The above plot suggests two likely scenarios. The question is: Which more accurately represents what is really happening?

Linear: Pick two points that seem reasonable and draw a straight line (red) through them.

Quadratic: Someone else looking at the data might conclude the curved line (blue) is more reasonable and accurate.

Visually we would conclude that the quadratic is mathematically a better fit because the curve is significantly closer to the given data points. However, it is important to remember that while this is true, an automotive expert applying expertise and experience may conclude that in fact the linear interpolation is more meaningful or that more data points are needed. We want to keep in mind that the "Art" component is what has to be applied to determine what degree polynomial and which technique will provide the best approximation.

Chapter One - Elimination (Substitution) Interpolation

A common method for solving the resulting system of equations is using linear algebra and matrix math. However, neither are necessary to illustrate this technique and apply to a practical problem. We will use elimination to solve the example below. While I think it is important students experience how basic algebra works for interpolation, they will quickly see that unless the numbers are small and simple this particular technique quickly becomes unwieldy for large values generated during the process.

For example:

Sample Vehicle Fuel Efficiency Measurements

X (Miles per Hour) MPH	Y (Miles per Gallon) MPG
45	43
55	42
65	38
75	32

Long Description

Sample Vehicle Fuel Efficiency Measurement

Apply expertise and experience to create a polynomial that will reasonably predict the fuel efficiency of the particular vehicle used to gather the above data.

Step one: Deciding that a quadratic equation looks like the best fit, we select the first, second and fourth points to construct a second degree (quadratic) polynomial.

Step Two: Even though the result will be a quadratic equation we are able to use straightforward linear techniques of elimination and substitution. The reason for this is that we are not trying to find x and y. The three points we selected already give us those. Instead, we are trying to create the quadratic in standard form by solving for the unknown constants a, b and c:

$$ax^2 + bx + c = y$$

Step Three: Lets create three quadratic equations with the same three unknowns a, b, c and replacing x, y in each with the actual data point values.

Eq. one: $a(45)^2 + b(45) + c = 43$ -----> $2025a + 45b + c = 43$

Eq. two: $a(55)^2 + b(55) + c = 42$ -----> $3025a + 55b + c = 42$

Eq. three: $a(75)^2 + b(75) + c = 32$ -----> $5625a + 75b + c = 32$

Step Four: The elimination process:

45(Eq. two) – 55(Eq one):

[$136125a + 2475b + 45c = 1890$] - [$111375a + 2475b + 55c = 2365$]

Eq. four: $24750a - 10c = -475$ **b is eliminated**

55(Eq. three) – 75(Eq. two):

[$309375a + 4125b + 55c = 1760$] - [$226875a + 4125b + 75c = 3150$]

Eq. five: $82500a - 20c = -1390$ **b is eliminated**

Conduct elimination on the resulting two equations with two unknowns to eliminate c.

2(Eq. four) – Eq. five:

$$49500a - 20c = -950$$
$$82500a - 20c = -1390$$

____

$$-33000a = 440$$

Eq. six: $a = -0.013$ **c is eliminated**

Plugging the resulting value of a into Eq. 4 allows us to solve for c:

$$24750(-0.013) - 10c = -475$$
$$c = 14.508$$

Step Five: Substitute a and c into any of the original equations to find b:

$$2025(-0.013) + 45b + 14.508 = 43$$
$$b = 1.233$$

Our interpolated polynomial is:

$$P(x) = -0.013x^2 + 1.233x + 14.508$$

For students looking for a less manual process here is the setup using matrix math to run the calculations in a spreadsheet.

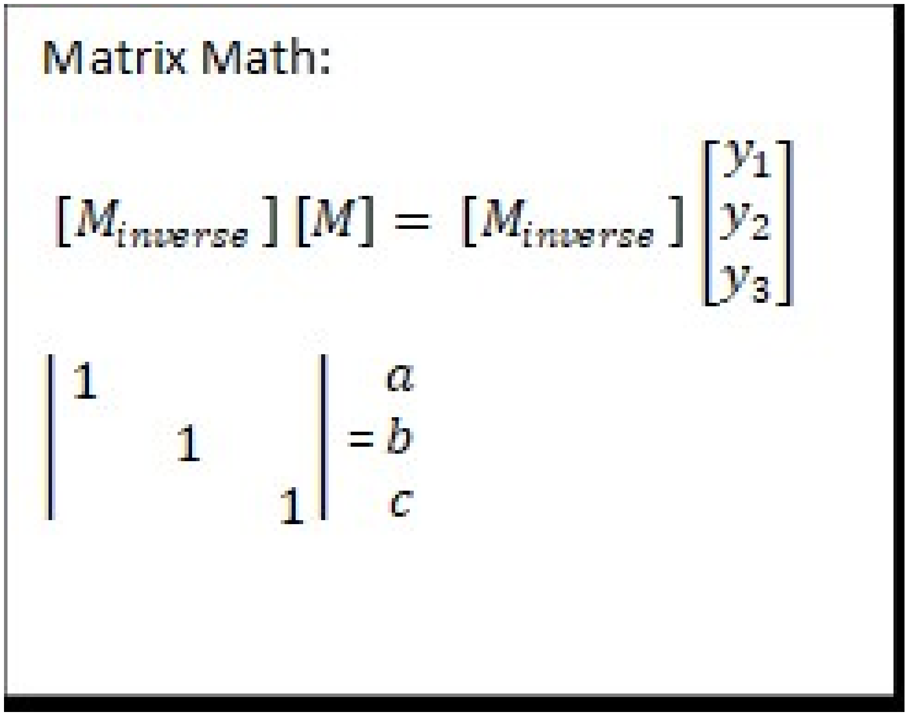
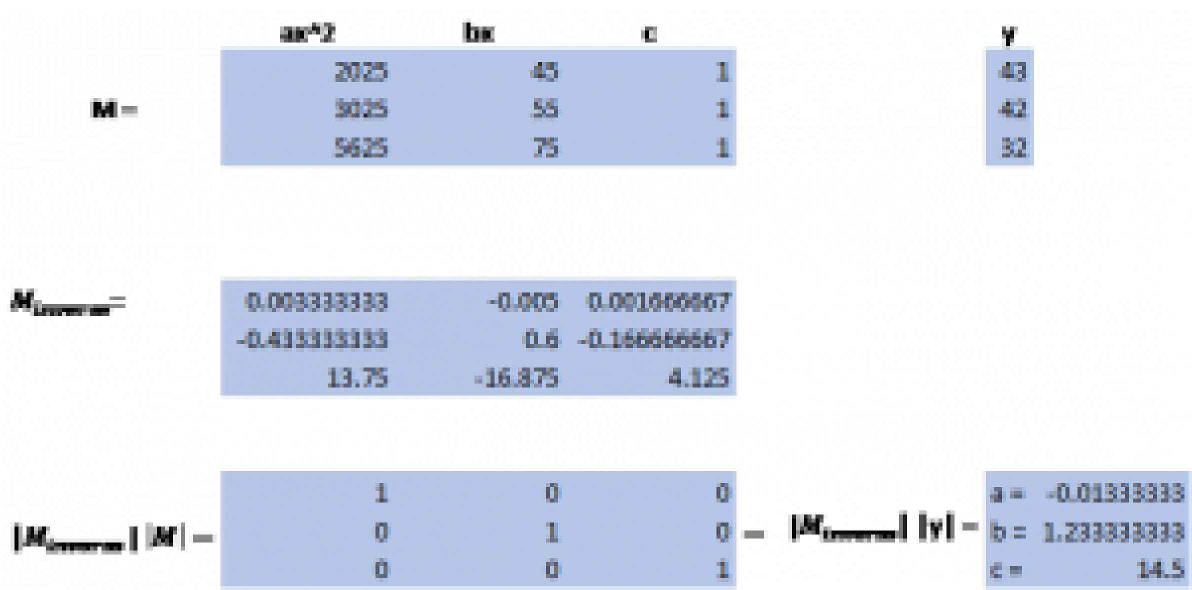


Figure 1.1 The Matrix Math formula



	ax^2	bx	c	y
M =	2025	45	1	43
	3025	55	1	42
	5625	75	1	32

$M_{inverse}$=	0.003333333	-0.005	0.001666667
	-0.433333333	0.6	-0.166666667
	13.75	-16.875	4.125

	1	0	0		a =	-0.01333333
$\lvert M_{inverse}\rvert \lvert M\rvert$ =	0	1	0	= $\lvert M_{inverse}\rvert$ $\lvert y\rvert$ =	b =	1.233333333
	0	0	1		c =	14.5

Figure 1.2 – Setup of Solution in Matrix Notation

Long Description

Figure 1.2 - Setup of Solution in Matrix Notation

Let's look between 45 and 55 at $x = 50$ and see how well our polynomial estimates a reasonable value:

$$P(50) = -0.013x(50)^2 + 1.233(50) + 14.508$$

It is recommended that the original points also be plugged into the equation as a check.

$P(50) = 43.7$ Plotting on our graph shows that this is indeed a very good estimate.

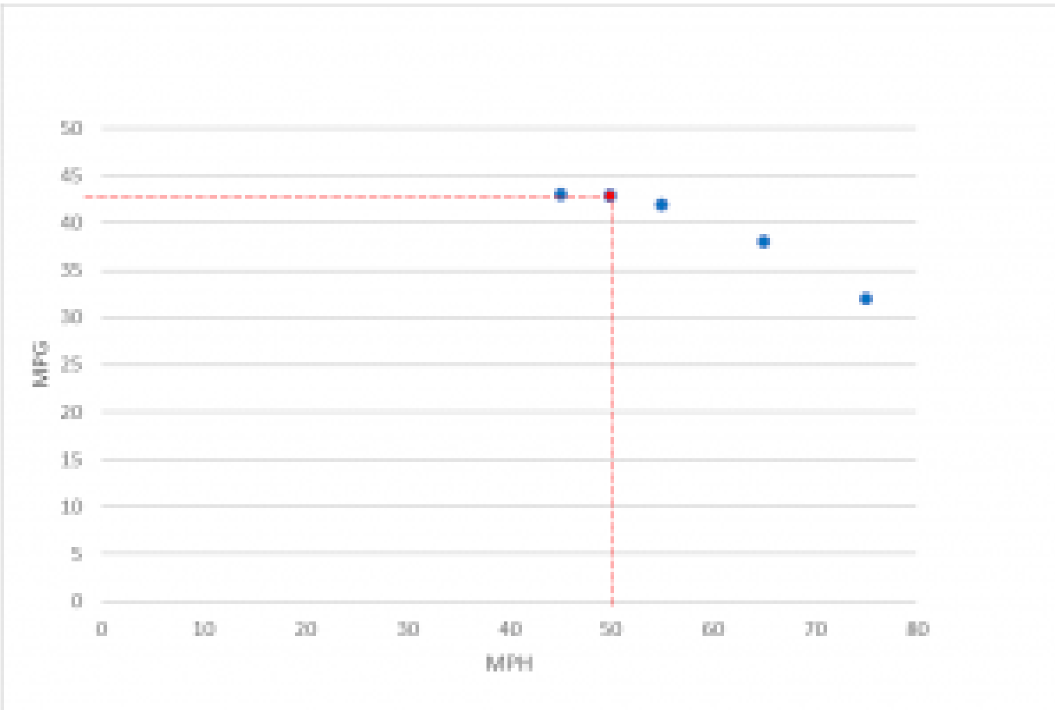


Figure 1.3 – Line graph displaying the results of the Quadratic Interpolation

Long Description

Figure 1.3 - Line graph displaying the results of the Quadratic Interpolation

Chapter One – Practice Exercises

1a)

The owner of the ABC Children's Party Company has offered a limited menu of pricing options depending on the number of children attending the party. The available prices are included in the table below:

ABC Children's Party Company

Maximum children attending the party	Cost per Child	Total Cost of Party
10	$37	$370
25	$28	$700
50	$22	$1100
100	$15	$1500

Long Description

ABC Children's Party Company

The prices cover the cost plus acceptable profit and have worked well in the past. To improve the companies competitiveness, the owner would like to offer more flexible pricing that is specific to the actual number of children. She would like to develop a cubic (3rd degree) polynomial that will generate the unit price when she inputs the expected number of children attending the party. To develop this polynomial the student must use the algebraic technique of substitution (elimination) discussed in this chapter.
(Solution Given)

1b)

This exercise offers practice in using basic matrix commands either manually or in a spreadsheet program to solve n-equations in n-unknowns.
(Solution given for 2nd to 5th row of data)

Given the following data points, develop a polynomial that will interpolate any value of p(x) on the given interval, for the bracketed points. It will result in a third-degree polynomial:

Exercise 1b Sample Data Point

x	y or f(x)
-4	12
[-1.75]	[-2]
[1]	[-3.7]
[3.3]	[-1.4]
[6.9]	[4]
7	3.9
9.1	6

Long Description

Exercise 1b Sample Data Point

Tables are provided to assist students

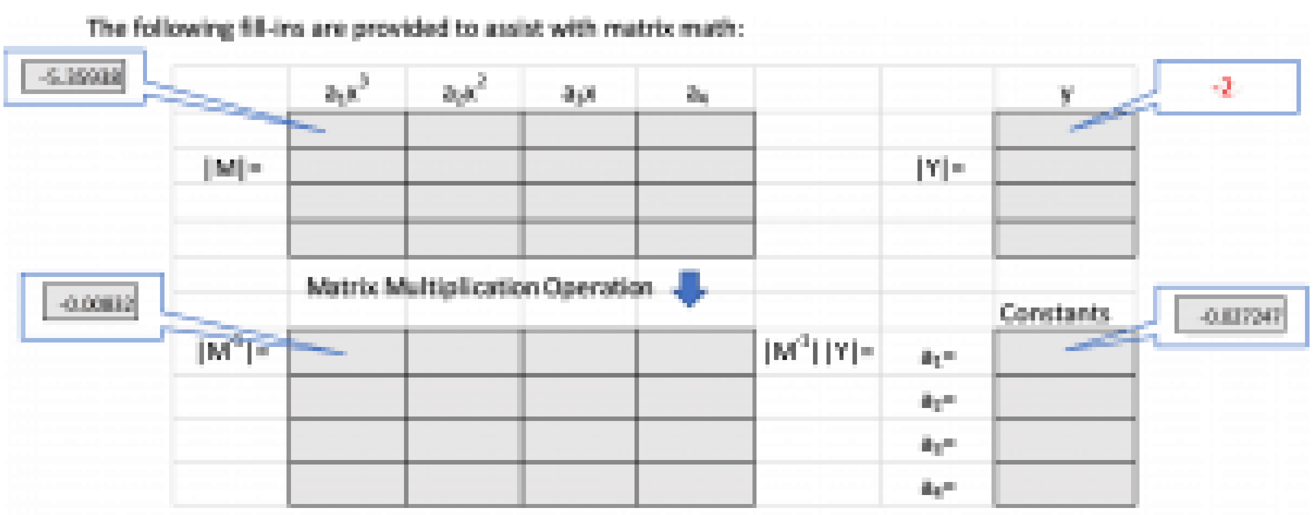


Figure 1.4 Guide for students

Long Description

Tables to Assist Atudents

1c)

Select any three data points from the above table and develop a 2nd degree (quadratic) Polynomial.

Chapter Two - Newton's Divided Difference Interpolation

A quick word regarding Divided Difference. The title might suggest that derivatives are involved, and in a way that would be correct. The good news is that knowledge of derivatives is not necessary for this technique. However, students should be familiar with the concept of slope, slope-intercept form and how slope is calculated since the process utilizes the change in the dependent variable (commonly known as y or f(x)) divided by the change in the independent variable (commonly known as x).

Students may have already encountered Divided Difference technique in high school algebra when asked to analyze a set of data to determine the non-linear (usually quadratic) equation that produced the dependent variable, as the following example illustrates.

Example

Given the following set of x values, determine the quadratic (2nd degree polynomial) that correctly produces the corresponding y values. Show in standard form:

Sample Data

x	y
-2	25.2
-1	11.3
0	2
1	-2.7
2	-2.8

Long Description

Sample Data

Solution

This simplified use of Newton's Divided Difference works because one of the x values is zero and there is a uniform distance of one between each x value.

x	y	1st divided difference	2nd divided difference
-2	25.2		
		$\frac{25.2-11.3}{-1}=-13.9$	
-1	11.3		$\frac{-13.9-(-9.3)}{-2}=2.3$
		$\frac{11.3-2}{-1}=-9.3$	
0	2		$\frac{-9.3-(-4.7)}{-2}=2.3$
		$\frac{2-(-2.7)}{-1}=-4.7$	
1	-2.7		$\frac{-4.7-(-0.1)}{-2}=2.3$
		$\frac{-2.7-(-2.8)}{-1}=-0.1$	
2	-2.8		

Figure 2.1 Simplified use of Newton's Divided Difference

Long Description

Figure 2.1 Simplified use of Newton's Divided Difference

Since the 2nd divided differences are all the same this tells us that there is a quadratic solution with $a = 2.3$

By plugging in the x,y values (0,2) we can easily solve for c as follows:

$$2.3(0)^2 + b(0) + c = 2$$

Or simply $c = 2$. Now that we know a and c we plug those in using one of the other points such as (1,-2.7) and solve for b as follows: $2.3(1)^2 + b(1) + 2 = -2.7$ which simplifies to $b = -7$

Resulting in the solution equation of $y = 2.3x^2 - 7x + 2$ which works for all given points and approximates everything in between.

Newton's Divided Difference Interpolation generalizes the above process. The given points no longer have to be in any particular order and the x values do not have to be spaced at uniform intervals; offering a welcome flexible technique.

The Generalized Process

Using Newton's Divided Difference approach, let's develop a polynomial that takes a limited number of data points (think points plotted on the coordinate plane) and fit them to a polynomial that is continuous across the interval.

This method is an iterative process that allows us to begin with one point. We can then add additional data points at our discretion, especially if we believe they will produce a better, more representative, polynomial.

Each time we add a point the resulting polynomial increases by a degree resulting in a polynomial of degree one less than the number of points included in the interpolation process.

(x_1, y_1): Constant Function: $f_0(x) = C$

(x_1, y_1), (x_2, y_2): Linear Function: f_1(x) = a_1x + C

(x_1, y_1), (x_2, y_2), (x_3, y_3): Quadratic Function: $f_2(x) = a_1 x^2 + a_2 x + C$

.

.

.

$(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots\ldots (x_n, y_n)$ resulting in an $n-1$ degree Function:

$$f_{(n-1)}(x) = a_1 x^{(n-1)} + a_2 x^{(n-2)} + \ldots a_n x + C$$

The following example illustrates the iterative process and demonstrates its validity.

I) The Constant Solution

The Constant Solution

x	f(x)
-2	3

Long Description

The Constant Solution

$f_0(x) = 3$ the constant solution

II) The Linear Solution: By adding a second point we move to a straight-line solution

The Linear Solution

x	f(x)
-2	3
-1	-4

Long Description

The Linear Solution

This is accomplished by preserving the constant solution $f_0(x) = 3$ while adding a linear component that works for all points on the straight line passing through both given points as follows.

$f_1(x) = f_0(x) + a_1(x - (-2))$ This added component will not alter the solution for $f(-2)$ while introducing the appropriate linear structure (degree one polynomial).

Solving for $f(x) = -4$ ensuring f(x) will satisfy both points and everything on the line passing through the two given points.

$-4 = 3 + a_1(x + 2)$
$-4 = 3 + a_1(-1 + 2)$
$-4 = 3 + a_1(1)$
$a_1 = -7$

Thus $f_1(x) = 3 + -7(x - (-2))$

Simplifying $f_1(x) = -7x - 11$ since this is valid slope intercept form, we have a linear solution

Checking 1st point $f(-2) = -7(-2) - 11 = 3$ it checks

Checking 2nd point $f(-1) = -7(-1) - 11 = -4$ it checks

III) The Quadratic Solution – 2nd degree polynomial

The Quadratic Solution

x	f(x)
-2	3
-1	-4
3	6

Long Description

The Quadratic Solution

Adding a third point, allows for the development of a quadratic (2nd degree) equation. We repeat the process with the same goal:

preserving the constant solution at the first point and the linear solution for first two points. The newly added third point will be satisfied by the previous linear solution plus the added quadratic component.

$f_2(x) = f_1(x) + a_2(x - x_1)(x - x_2)$

this component (in red) ensures this new solution works for previous points as well as establishing a valid quadratic form.

Remember $f_1(x) = -7 - 11$

$$f_2(x) = f_1(x) + a_2(x - (-2))(x - (-1))$$

$$f_2(3) = 6 = -7(3) - 11 + a_2(3 + 2)(3 + 1)$$

Solving for the constant: $a_2 = \frac{38}{20}$

Plug in and simplify

$$f_2(x) = -7x - 11 + \frac{30}{20}(x - (-2))(x - (-1))$$

$$f_2(x) = \frac{38}{20}x^2 - \frac{26}{20}x - \frac{144}{20}$$

As a check we will plug in our three given values of x to verify it produces the corresponding given y values.

Check One

$$f_2(-2) = \frac{38}{20}(-2)^2 - \frac{26}{20}(-2) - \frac{144}{20}$$

$$f_2(-2) = 3$$

CheckTwo

$$f_2(-1) = \frac{38}{20}(-1)^2 - \frac{26}{20}(-1) - \frac{144}{20}$$
$$f_2(-2) = -4$$

Check Three

$$f_2(3) = \frac{38}{20}(3)^2 - \frac{26}{20}(3) - \frac{144}{20}$$
$$f_2(-2) = 6$$

We have engaged in an iterative process. Utilizing generalized notation for the above we conducted three iterations, with an additional point added at each iteration.

Single point: (x_1, y_1):

$f_0(x1) = b1$ Constant Solution

Second Point Added: (x_2, y_2):

$f_1(x_2) = b_1 + b_2(x_2 - x_1)$ solving for $b_2 = (f_1(x_2) - b_1)/(x_2 - x_1)$ linear

Third Point Added: (x_3, y_3):

$f_2(x_3) = f_1(x_3) + b_3(x_3 - x_1)(x_3 - x_2)$

solve for $b_3 = (f_2(x_3) - f_1(x_3))/((x_3 - x_1)(x_3 - x_2))$ quadratic

Each new iteration builds upon and preserves the previous solutions.

In general, the solution polynomial can continue to be increased one degree at a time solving for each new variable as long as additional points become available. This results in the following general form:

$$fn - 1(x) = b_1 + b_2(x - x_1) + b_3(x - x_1)(x - x_2) + \ldots b(n+1)(x - x_1)(x - x_2)\ldots + (x - x_n)$$

Normally it is best to select the lowest order polynomial that is reasonable. Higher order polynomials can introduce unwanted error.

The table approach below offers a convenient methodology for manually calculating the constants. It lends itself to adding additional points as needed without having to start over.

The following **Table Methodology** illustrates and simplifies the above process.

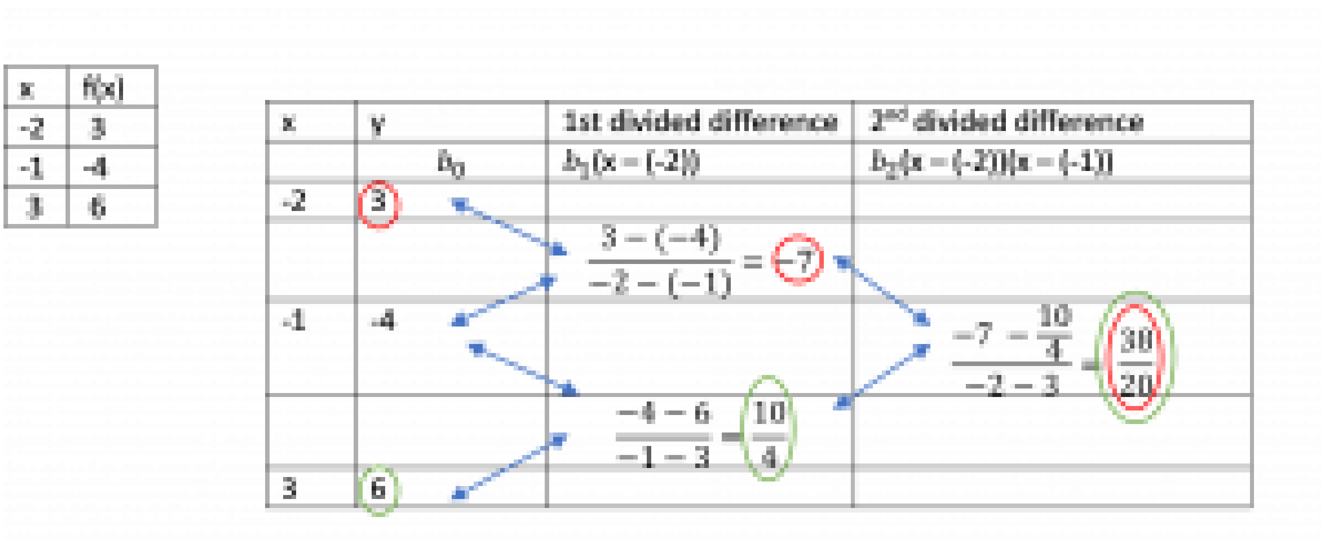


Figure 2.2 Table Methodology

Long Description

Figure 2.2 Table Methodology

Starting at the right-hand column we backtrack diagonally left and up (circled in red). Backtracking left and downward would have produced the same simplified equation (circled in green)

This produces the following results:

$$f(x) = \frac{38}{20}(x-(-2))(x-(-1)) + (-7(x-(-2)) + 3$$

Simplifying: $$f_2(x) = \frac{38}{20}x^2 - \frac{26}{20}x - \frac{144}{20}$$

This satisfies the three given points as well as any interpolated points between the minimum and maximum value of x. Because it is a continuous function, it also produces extrapolated points beyond the range. These extrapolated points may or may not be valid for any particular situation being analyzed. That is part of the "Art" of interpolation which relies on the experience and expertise of the one studying a particular phenomenon.

The Sin function – An interesting example

One of the neat things we can use interpolation for is to create a polynomial that provides reasonable estimates of the sin (or cos) of an angle for any given measure. In fact, the numbers we will use are small and simple that even the Elimination (Substitution) approach will easily produce the desired result.

The Sine function illustrated on the coordinate plane

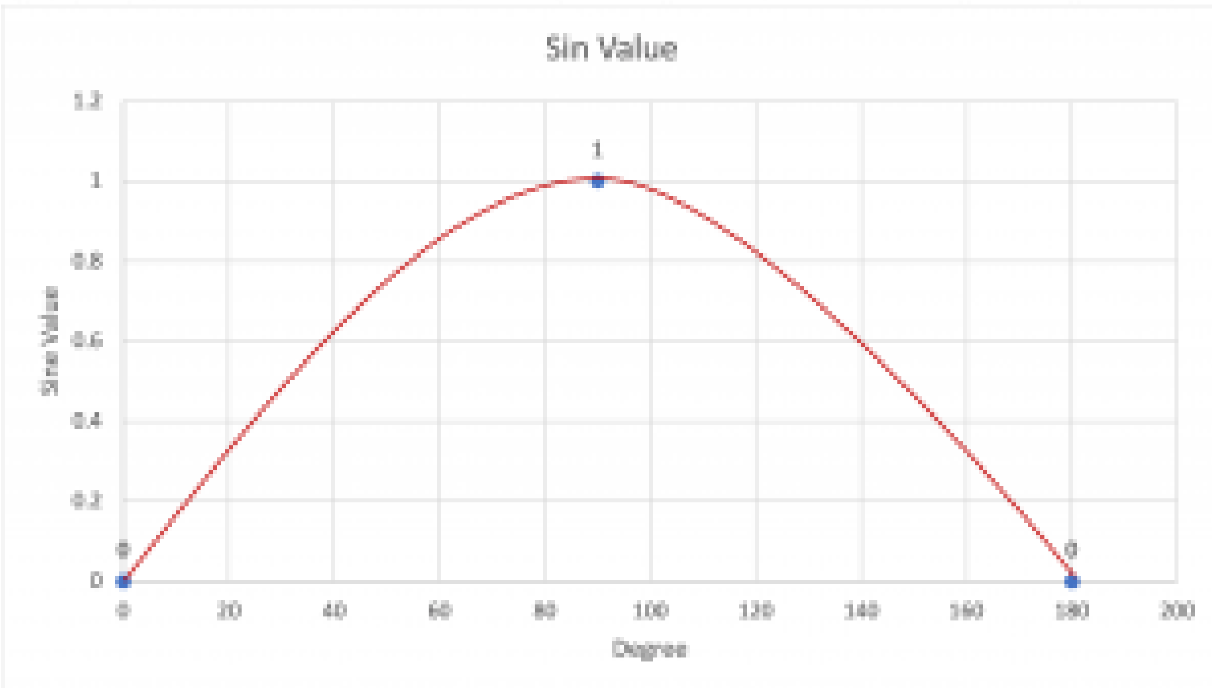


Figure 2.3 Sine Function Graph

Long Description

Figure 2.3 Sine Function Graph

Note: we will use Radians as they are more manageable.

Degree	Radian	Sin Value
0	0	0
90	π/2	1
180	π	0

Plug into the Newton Divided Difference Table

	Radian x	Sin Value f(x)	1st divided difference	2nd divided difference
		b_0	$b_1(x-x_0)$	$b_2(x-x_0)*(x-x_1)$
X_0	0	0		
			$\frac{0-1}{0-\frac{\pi}{2}}=\frac{2}{\pi}$	
X_1	π/2	1		$\frac{\frac{2}{\pi}-\frac{-2}{\pi}}{0-\pi}=\frac{-4}{\pi^2}$
			$\frac{1-0}{\frac{\pi}{2}-\pi}=-\frac{2}{\pi}$	
X_2	π	0		

Figure 2.4 Estimating sin wave – Newton's Divided Difference Table

Long Description

Figure 2.4 Estimating sin wave - Newton's Divided Difference Table

$$f(x)=-\frac{4}{\pi^2}(x-0)(x-\frac{\pi}{2})+\frac{2}{\pi}(x-0)+0$$

Simplifying the resulting equation produces: $f(x)=\frac{-4}{\pi^2}x^2+\frac{4}{\pi}x$

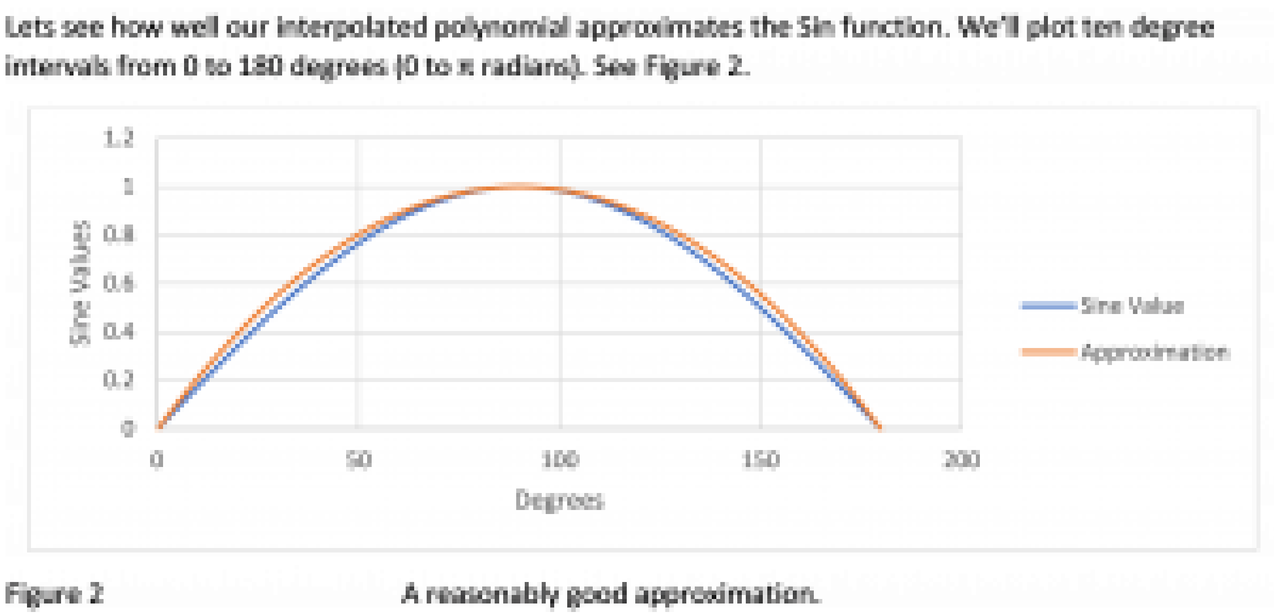

Lets see how well our interpolated polynomial approximates the Sin function. We'll plot ten degree intervals from 0 to 180 degrees (0 to π radians). See Figure 2.

Figure 2 A reasonably good approximation.

Figure 2.5 An approximation of sin value

Long Description

Figure 2.5 Sine Function Approximation

Chapter Two – Practice Exercises

2a)

While the owner in exercise 1a) was happy with the results of using elimination/substitution, she was curious to see if the results would differ using Newton's Divided Difference (NDD) interpolation. You have decided to assist her by generating a cubic polynomial using NDD. (Solution given) The data is:

ABC Children's Party Company

Maximum children attending the party	Cost per Child	Total Cost of Party
10	$37	$370
25	$28	$700
50	$22	$1100
100	$15	$1500

Long Description

ABC Children's Party Company

2b)

Using the same seven data points from the previous chapter select three data points and plug into the grid below to produce a quadratic solution. Simplify the resulting polynomial and put in standard form. Note solution given for the three bracketed points.
(Solution given)

Seven Data Points

x	y or f(x)
-6.2	-8
[-3]	[-7]
-1.5	-2.2
[1]	[0.7]
3.5	3
4.25	5
[7.9]	[11]

Long Description

Seven Data Points

Exercise 2b Answer Grid

-	x	f(x)	1st divided difference	2nd divided difference
-	-	b_0	$b_1(x - x_0)$	$b_2(x - x_0)(x - x_1)$
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-

Long Description

Exercise 2b answer grid

2c)

Add an additional data point and develop a 3rd degree (cubic) polynomial. Compare this to the solution from 2a) and decide whether or not it improves the interpolation. Note student answers may vary.

Chapter Three - Quadratic Spline Interpolation

This technique offers several advantages over other techniques. It produces a smooth curve over the interval being studied while at the same time offering a distinct polynomial for each subinterval (known as Splines). Secondly it eliminates some of the problems inherent in trying fit a single higher order polynomial which can actually produce misleading estimates by being too precise.

One disadvantage that we quickly discover is that the resulting set of polynomials can be taxing to solve manually using techniques such as elimination/substitution, Gauss-Jordan reduction or Cramer's rule. Fortunately, many applications including most spreadsheet programs allow us to solve the resulting system, easily producing the family of equations.

Let's begin with a simple case that the student can choose to solve manually to can gain an understanding of the process. The matrix operations are shown as well.

Spline Example

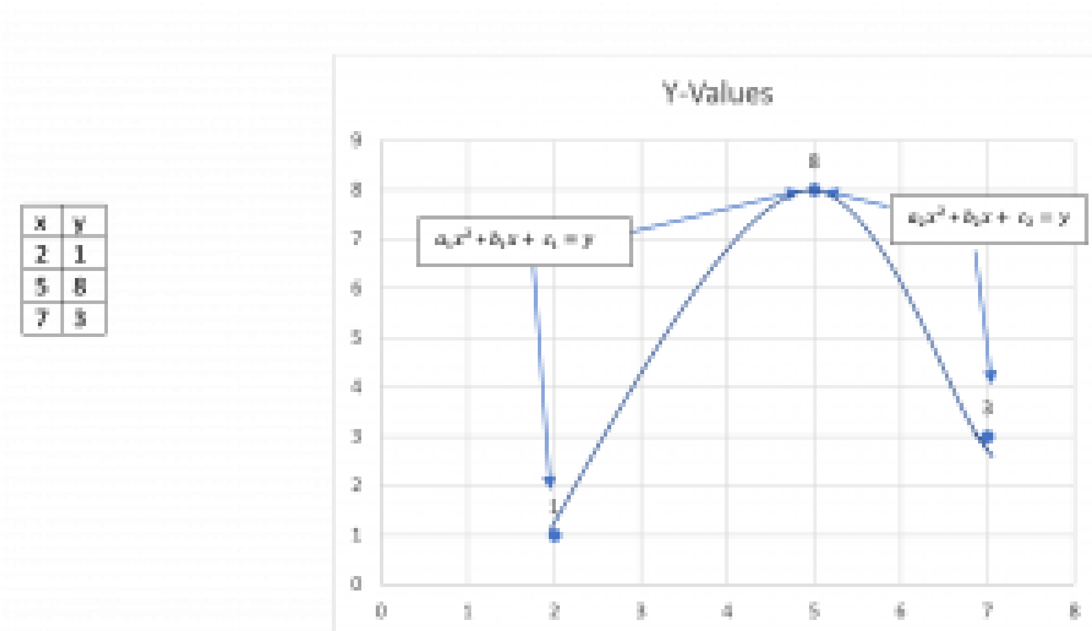


Figure 3.1 Spline Example

Long Description

Figure 3.1 Spline Example

Instead of one equation we could have an equation representing the interval [2,5] and a second equation [5,7]. The key is that the point in the middle contributes to both equations creating a connection that ensures a smooth handoff from the first to the second equation. The general form is:

$$a_1x^2 + b_1x + c_1 = y \qquad 2 \le x \le 5$$
$$a_2x^2 + b_2x + c_2 = y \qquad 5 \le x \le 7$$

Since we want to solve for the six constants in a proper linear fashion, we need four more equations. To find them we employ the connection at $x = 5$. Since each equation satisfies two endpoints this allows us to double the number of equations as follows:

$$a_1 2^2 + b_1 2 + c_1 = 1$$
$$a_1 5^2 + b_1 5 + c_1 = 8$$
$$a_2 5^2 + b_2 5 + c_2 = 8$$
$$a_2 7^2 + b_2 7 + c_2 = 3$$

We now have four equations. The fifth equation we can develop at the point (5,8) known as an internal knot. Note the two endpoints are sometimes referred to as external knots.

If we take the derivative of the two equations at $y = 8$, we know they have to be equal because the slope has to be the same at that point. We can set them equal to each other and simplify. This results in:

$$2a_1x + b_1 = 2a_2x + b_2$$

Rearrange

$$2a_1x + b_1 - 2a_2x - b_2 = 0$$

We now have five of the six equations

$$a_1 2^2 + b_1 2 + c_1 = 1$$
$$a_1 5^2 + b_1 5 + c_1 = 8$$
$$a_2 5^2 + b_2 5 + c_2 = 8$$
$$a_2 7^2 + b_2 7 + c_2 = 3$$
$$2a_1 5 + b_1 - 2a_2 5 - b_2 = 0$$

$a_1 = 0$ **This is the sixth equation; see explanation below.**

The sixth equation is based on the assumption that the line leaving the endpoint is a straight line. The quadratic component zeros out thus our sixth equation is simply $a_1 = 0$. The other endpoint would have produced $a_2 = 0$ which would have worked equally well. We'll use these six equations and solve with matrix operations.

Constants Displayed in Matrix Form

a_1	b_1	c_1	a_2	b_2	c_2	y
4	2	1	0	0	0	1
25	5	1	0	0	0	8
0	0	0	25	5	1	8
0	0	0	49	7	1	3
10	1	0	-2	-10	0	0
1	0	0	0	0	0	0

For illustrative purposes a detailed flow of the matrix operation is offered below:

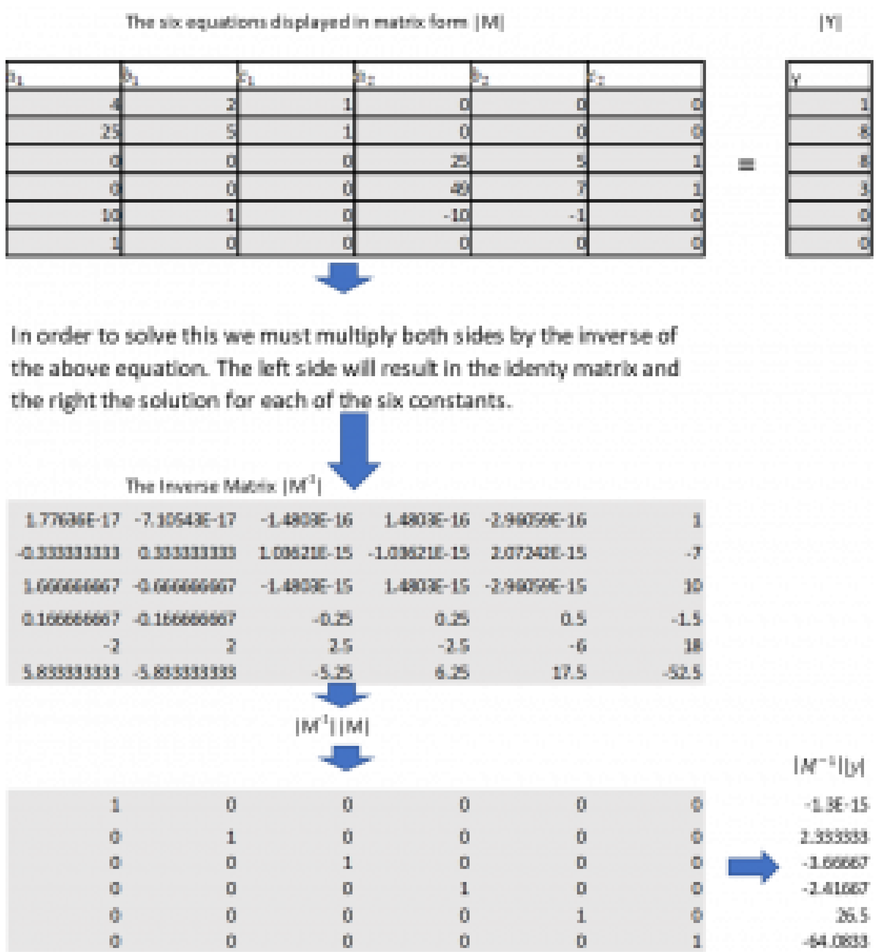


Figure 3.2 Matrix Operation Flow

Long Description

Figure 3.2 Matrix Operation Flow

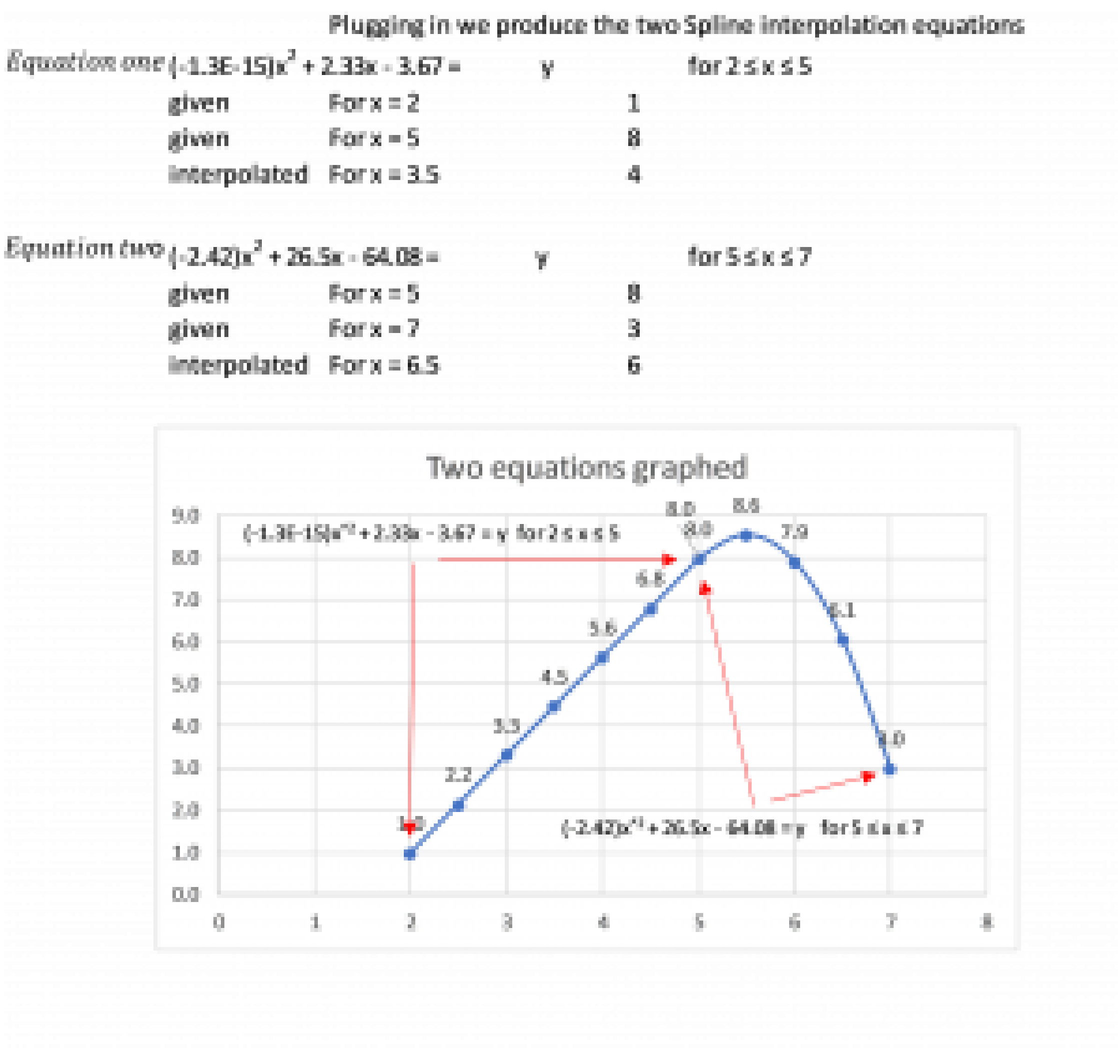


Figure 3.3 Two Spline Interpolation Equations

Long Description

Figure 3.3 Two Spline Interpolation Equations

Example – Space Launch Data

The following combines a general explanation of the technique along with a specific example. We will use the following launch data for the Saturn 5 rocket. Note this data was pulled from readily available data for several launches and in fact does not represent any one launch. The data tracks a hypothetical Saturn 5 from launch until third stage shutdown shortly before entering earth orbit.

Space Launch Data

x (time in minutes)	**y (velocity in 1000ft per second)**
0	1
1	2
2.5	9
3	9.2
4	10
5	12
6	14.5
7	17
8	20
8.75	23
9	23.5
10	24
11	25.5
11.25	25.9
11.5	25.9

Selected Interval Points (knots)

x	**y**	**Interval**
1	2	start of first interval
2.5	9	1st stage separation
8.75	23	2nd stage separation
11.25	25.9	3rd stage shutdown

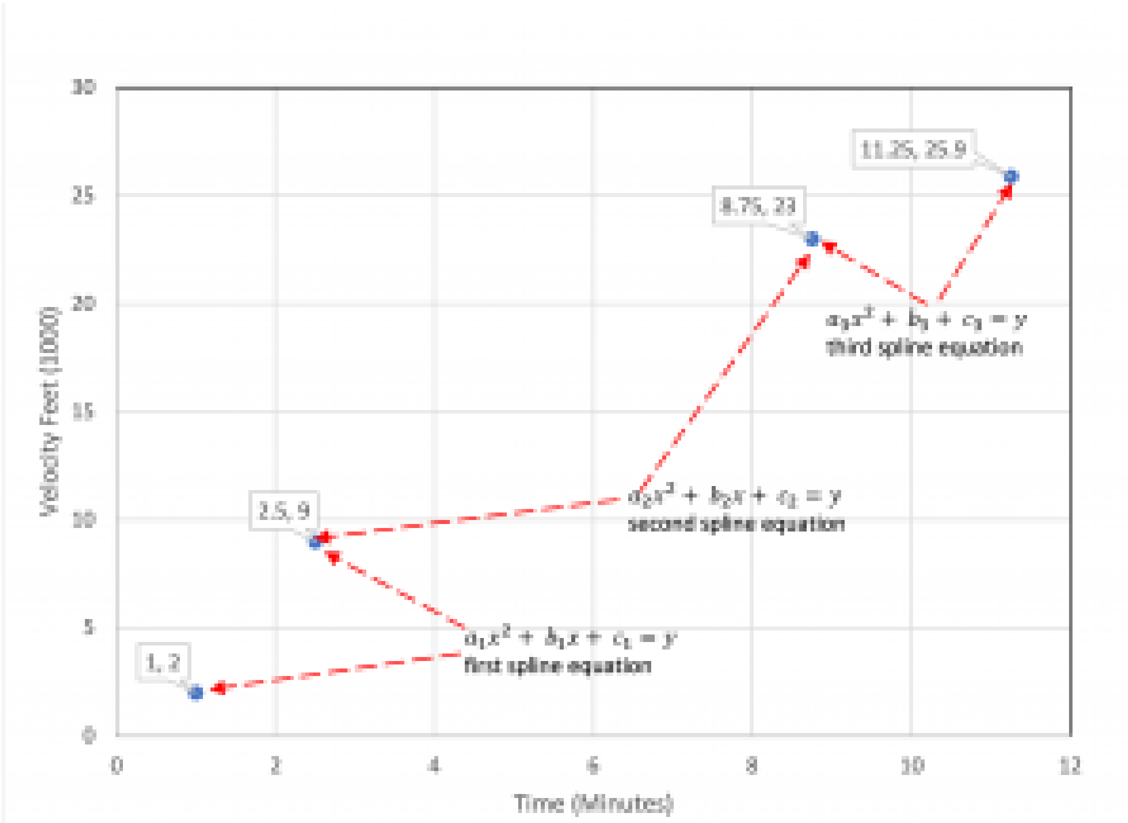


Figure 3.4 Plot of Launch-Knots Identified

Long Description

Figure 3.4 Plot of Launch-Knots Identified

Since we have selected $n = 4$ data points we create $n-1 = 3$ quadratic spline equations each with three unknowns:

$$a_1x^2 + b_1x + c_1 = y \qquad 1 \le x \le 2.5$$
$$a_2x^2 + b_2x + c_2 = y \qquad 2.5 \le x \le 8.75$$
$$a_3x^2 + b_3 + c_3 = y \qquad 8.75 \le x \le 11.25$$

We want to solve for the $n = 9$ unknowns. However, with only three equations we need to create six additional equations in order to apply one of the standard techniques for solving n equations in n unknowns.

Notice that each equation is a solution for two of the knots as shown in figure 1. This allows us to split each spline equation into two equations providing a total of n= 6 equations as follows:

$$a_1(1)^2 + b_1(1) + c_1 = 2 \qquad (x1, y1) = (1, 2)$$
$$a_1(2.5)^2 + b_1(2.5) + c_1 = 9 \qquad (x2, y2) = (2.5, 9)$$
$$a_2(2.5)^2 + b_2(2.5) + c_2 = 9 \qquad (x2, y2) = (2.5, 9)$$
$$a_2(8.75)^2 + b_2(8.75) + c_2 = 23 \qquad (x3, y3) = (8.75, 23)$$
$$a_3(8.75)^2 + b_3(8.75) + c_3 = 23 \qquad (x3, y3) = (8.75, 23)$$
$$a_3(11.25)^2 + b_3(11.25) + c_3 = 25.9 \quad (x4, y4) = (11.25, 25.9)$$

We are getting closer. We will now create two more equations using basic knowledge of the derivative and the fact that two pairs of equations are solutions for the two interior knots. This works because the first derivative of each equation in a pair will have the same slope at the common data point (knot).

This is not a course in calculus so I will simply show the first derivatives for each pair to obtain our additional equations.

$$\frac{d}{dx}[a_1x^2 + b_1x + c_1] = \frac{d}{dx}[a_2x^2 + b_2x + c_2] \qquad \text{at} \quad x = 2.5$$
$$2a_1x + b_1 = 2a_2x + b_2$$
$$2a_1(2.5) + b_1 - 2a_2(2.5) - b_2 = 0 \ \textbf{the seventh equation}$$

$$\frac{d}{dx}[a_2x^2 + b_2x + c_2] = \frac{d}{dx}[a_3x^2 + b_3x + c_3] \qquad \text{at} \quad x = 8.75$$
$$2a_2x + b_2 = 2a_3x + b_3$$
$$2a_2(8.75) + b_2 - 2a_3(8.75) - b_3 = 0 \ \textbf{the eighth equation}$$

For our ninth equation we recognize that at each endpoint the resulting line extending beyond the interval is a straight line. Since this eliminates the quadratic component, we can simply make our ninth equation be:

a_1 = 0

We now have our nine equations with nine unknowns. Figure 4 below includes the nine equations.

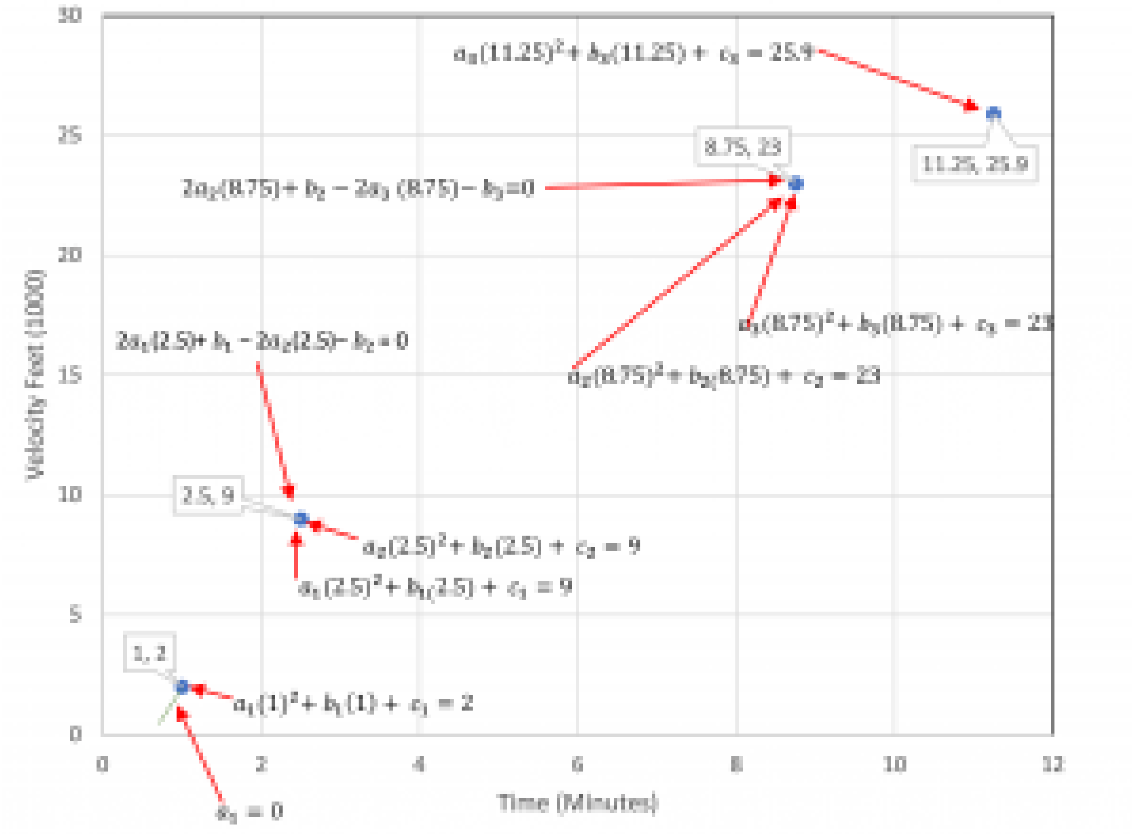


Figure 3.5 The Nine Equations

Long Description

Figure 3.5 The Nine Equations

Gathering the equations and squaring the quadratic variables results in following nine equations with nine unknowns. The x variables are replaced with the x-value from the related knot.

$$a_1(1) + b_1(1) + c_1 = 2$$

$$a_1(6.25) + b_1(2.5) + c_1 = 9$$

$$a_2(6.26) + b_2(2.5) + c_2 = 9$$

$$a_2(76.56) + b_2(8.75) + c_2 = 23$$

$$a_3(76.56) + b_3(8.75) + c_3 = 23$$

$$a_3(126.56) + b_3(11.25) + c_3 = 25.9$$

$$a_1(5) + b_1(1) - a_2(5) - b_2(1) = 0$$

$$a_2(17.5) + b_2(1) - a_3(17.5) - b_3(1) = 0$$
$$a_1(1) = 0$$

It would be a cumbersome task to solve the above system by hand. Instead, we will put the data in matrix form and solve.

Nine Equations Solved with Matrix Math

1	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**a_1**	**2**
6.25	2.5	1	0	0	0	0	0	0	b_1	9
0	0	0	6.25	2.5	1	0	0	0	c_1	9
0	0	0	76.56	8.75	1	0	0	0	a_2	23
0	0	0	0	0	0	76.56	8.75	1	b_2	23
0	0	0	0	0	0	126.56	11.25	1	c_1	25.9
5	1	0	-5	-1	0	0	0	0	a_3	0
0	0	0	17.5	1	0	-17.5	-1	0	b_3	0
1	0	0	0	0	0	0	0	0	c_3	0

Plug the above into a spreadsheet and apply matrix operations as follows:

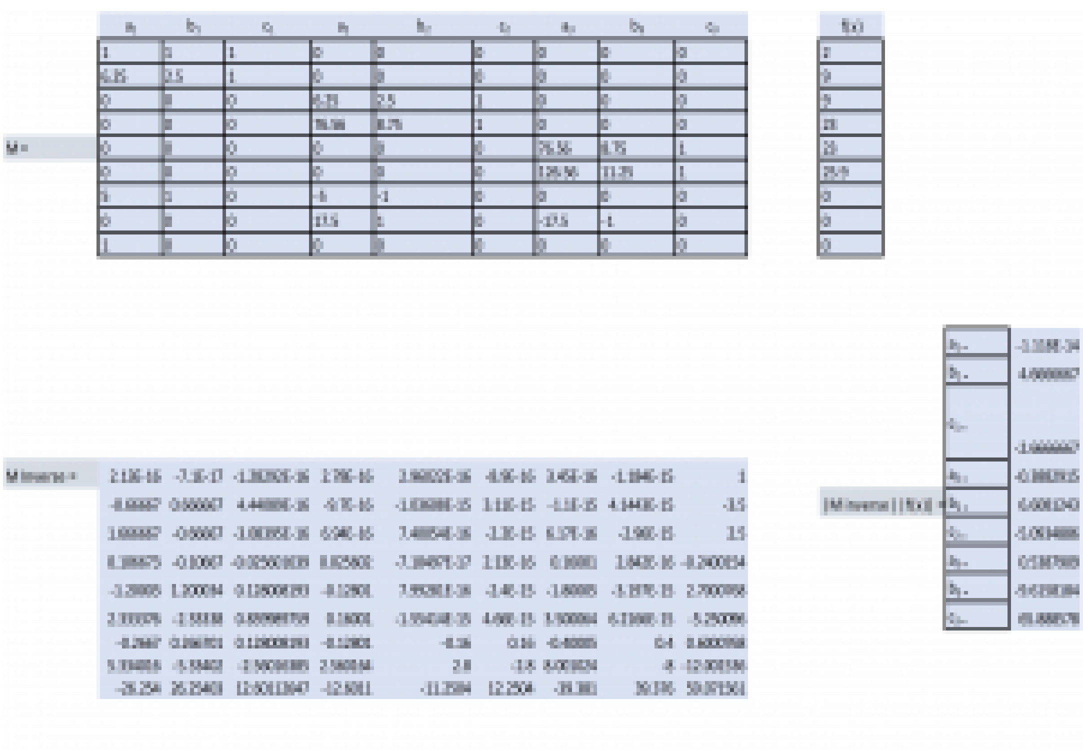


Figure 3.6 Nine Equations Solved with Matrix Math

Long Description

Figure 3.6 Nine Equations Solved with Matrix Math

The calculations produced three polynomials for the interval $1 \leq x \leq 11.25$

$$(-1.118 * 10^{-14})x^2 + 4.67x - 2.67 = y \qquad 1 \leq x \leq 2.5$$
$$-0.39x^2 + 6.61x - 5.09 = y \qquad 2.5 \leq x \leq 8.75$$
$$0.539x^2 - 9.616c + 65.889 = y \qquad 8.75 \leq x \leq 11.25$$

These equations produce reasonable estimates for the overall flight pattern as shown in figure 3.7.

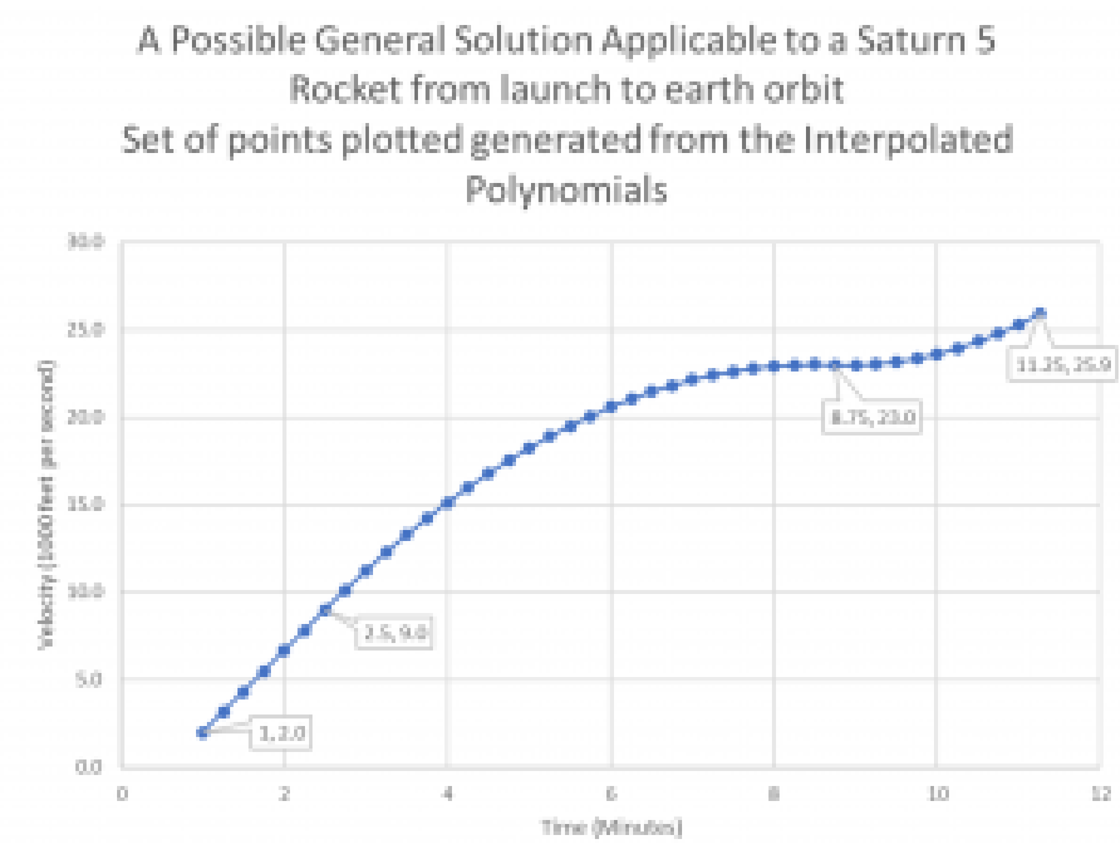


Figure 3.7 A solution for a space launch

Long Description

Figure 3.7 Saturn 5 Rocket Possible Solution

Direct Method Cubic Interpolation

Cubic interpolation takes us to the next level and is a common method for developing an equation that approximates f(x) for a particular value of x as well the neighborhood on either side made up of the four closest given data points. It is well suited if we want to interpolate for a particular interval of x. This will not provide a family of polynomials that satisfy the domain of the function. Rather it provides that single cubic polynomial that gives us a good picture of what is happening at and near a particular point of interest. This approach allows us to setup and solve a single cubic equation. The principal limitation is that it is not valid for the entire domain of x only the four closest points. Since we often only want to look at a limited range the benefits of a significant reduction in algebraic manipulation outweighs the limitation.

We will use our table of data from the previous example.

Saturn 5 Rocket Launch Data

x (time in minutes)	y (velocity in 1000 ft per second
0	1
1	2
2.5	9
3	9.2
4	10
5	12
6	14.5
7	17
8	20
8.75	23
9	23.5
10	24
11	25.5
11.25	25.9
11.5	25.9

Let's say we want to estimate the velocity when $x = 5.85$ minutes. We check the points on either side to determine the four closest values to 5.85 (shown in red).

Closest Values to x=5.85

Checking Distances	Four Data Points
5.85 - 3 = 2.85	-
5.85 - 4 = 1.85	(4,10)
5.85 - 5 = 0.85	(5,12)
6 - 5.85 = 0.15	(6,14.5)
7 - 5.85 = 1.15	(7,17)
8 - 5.85 = 2.15	-

Other Data Points From Example

x (time in minutes)	y (velocity in 1000 ft per second)
4	10
5	12
6	14.5
7	17

Utilizing the standard form for a cubic polynomial allows us to quickly set up four equations with four unknowns. Remember we are not finding x and y we already know those. Rather our unknowns are the constants a,b,c,d.

$$a(4)^3 + b(4)^2 + c(4) + d = 10$$
$$a(5)^3 + b(5)^2 + c(5) + d = 12$$
$$a(6)^3 + b(6)^2 + c(6) + d = 14.5$$
$$a(7)^3 + b(7)^2 + c(7) + d = 17$$

Using high school algebra (elimination/substitution), Gauss Jordan reduction or some other method, solve for the four unknowns. Below shows the setup using Matrix math to solve the cubic polynomial in a spreadsheet program.

	a	b	c	d		y	
	64	16	4	1		10	
M=	125	25	5	1	y=	12	
	216	36	6	1		14.5	
	343	49	7	1		17	
M^{-1}=	-0.16667	0.5	-0.5	0.166667		a=	-0.08333
	3	-8.5	8	-2.5	|M-1| |y|=	b=	1.5
	-17.8333	47	-41.5	12.33333		b=	-6.41667
	35	-84	70	-20		d=	17

Figure 3.8 Solved Cubic Polynomial via Spreadsheet Program

Long Description

Figure 3.8 Solved Cubic Polynomial via Spreadsheet Program

Resulting Equation

$-0.08333x^3 + 1.5x^2 - 6.41667x + 17 = y$ **for the interval** $4 \leq x \leq 7$

Let's see how well our cubic polynomial fits when plotted against all the given points plus x = 5.85

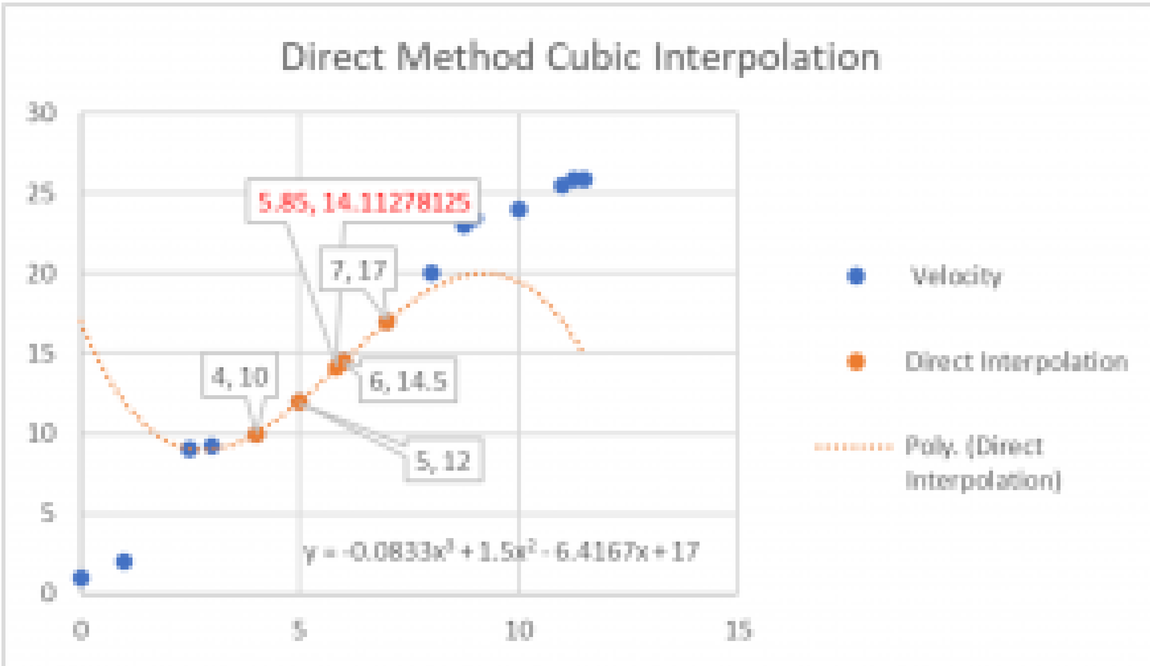


Figure 3.9 Graph of Cubic Interpolation

Long Description

Figure 3.9 Direct Method Cubic Interpolation

Notice that the solution provides the best estimate in the neighborhood of the closest points.

Chapter Three - Practice Exercises

3a)

Using the data from the Saturn launch example in chapter three calculate the family of quadratic splines for the following different Selected Interval Data Points (knots) and compare to the example.

Saturn 5 Rocket Launch Data

x (time in minutes)	y (velocity in 1000 ft per second
0	1
1	2
2.5	9
3	9.2
4	10
5	12
6	14.5
7	17
8	20
8.75	23
9	23.5
10	24
11	25.5
11.25	25.9
11.5	25.9

Selected Interval Points (knots)

x	y	Interval
1	2	start of first interval
2.5	9	1st stage separation
8.75	23	2nd stage separation
11.25	25.9	3rd stage shutdown

3b)

Using the table in 3a) for time = 7.5, conduct a Direct Method Cubic Interpolation. Show the resulting polynomial in standard form and graph the solution manually or with your favorite graphing tool.
(Solution given)

Chapter Four - Least Squares Regression

This technique is often used when many points of data are involved and the analyst would like the resulting polynomial to be influenced by all the identified points. The degree of the Interpolated polynomial should be selected ahead of time based on the expertise of the analyst. As a general rule of thumb, the lowest degree polynomial that appears to fit is the better choice. So, one might fit a quadratic or cubic solution to a large number of points which could run to dozens or even hundreds of points. The result will always be considered mathematically a best fit to the data.

To gain an understanding of the underlying principle and process we will begin with a simple data set consisting of five points.

Scenario

A helium balloon that gathers meteorological data is released. For each mile it rises, the distance it travels downrange is also recorded. The data is recorded in the following table.

Altitude and Downrange

Altitude - x miles	**Downrange - y miles**
1	2
2	3
3	5
4	5
5	4

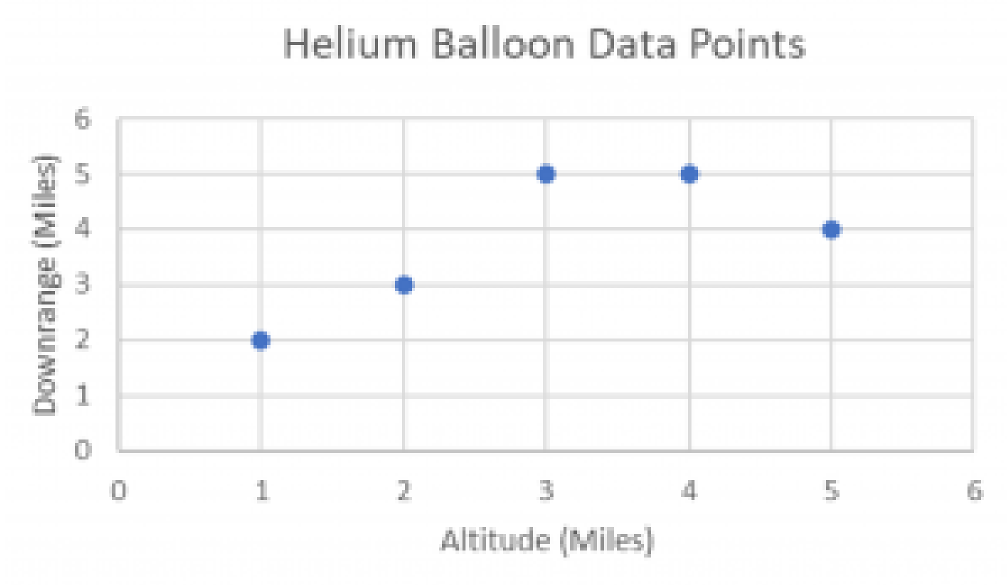


Figure 4.1 Data points for a Helium Balloon

Long Description

Figure 4.1 Helium Balloon Data Points

Let's begin with the simplest model – the straight line. We want to find a best fit linear equation that minimizes the sum of the distances between the actual and interpolated values of y for a given value of x.

1) A generalized linear equation $y = ax + b$ will serve as our starting point.

2) It is easy to see that with a little rearranging we have an equation that lends itself to finding that minimum distance mentioned above: y - (ax + b) = 0

We will square this equation so that resulting differences in distance are always positive as we are not interested in the direction of the difference but the sum of the differences.

Since we want the sum of these squared equations, we have the following for this example:

$$[y_1 - (ax_1 + b)]^2$$
$$+[y_2 - (ax_2 + b)]^2$$
$$+[y_3 - (ax_3 + b)]^2$$
$$+[y_4 - (ax_4 + b)]^2$$
$$+[y_5 - (ax_5 + b)]^2$$

Interestingly by squaring these equations we will obtain a quadratic equation which will be useful in finding a linear solution. In fact, it will allow us to create two partial derivative equations for each of the constants we are trying to solve for. In this case a, b. This will result in two linear equations in two unknowns which we can solve using elimination/

substitution or more advanced techniques such as matrix computations. And because they are upward facing quadratics, we minimize each equation be setting them to zero.

1) $$\frac{d}{da} = -2\sum_{i=1}^{5}[y_i - (ax_i + b)]x_i = 0$$

2) $$\frac{d}{db} = -2\sum_{i=1}^{5}[y_i - (ax_i + b)] = 0$$

Next, we simplify each equation by distributing the summation notation. And, since they are equal to zero, we simply divide out the -2. We now have two equations in two unknowns a,b.

Simplify 1) $$\sum_{i=1}^{5} x_i y_i - a\sum_{i=1}^{5} x_i^2 - b\sum_{i=1}^{5} x_i = 0$$

Simplify 2) $$\sum_{i=1}^{5} y_i - a\sum_{i=1}^{5} x_i - b\sum_{i=1}^{5} 1 = 0$$

We now have two equations in two unknowns a, b. Let's calculate the various sums and plug in.

$$\sum_{i=1}^{5} x_i y_i = 63$$

$$\sum_{i=1}^{5} x_i^2 = 55$$

$$\sum_{i=1}^{5} x_i = 15$$

$$\sum_{i=1}^{5} y_i = 19$$

$$\sum_{i=1}^{5} 1 = 5$$

I) Plug in to set up the two equations as follows:

One: $63 - a55 - b15 = 0$

Two: $19 - a15 - b5 = 0$

II) Rearrange:

One: 55a + 15b = 63

Two: 15a + 5b = 19

III) Apply substitution/elimination to solve for a, b

$$a = \frac{3}{5} = 0.6$$

$$b = 2$$

We now have a polynomial that can interpolate values in the interval [1,5]

$$y = \frac{3}{5}x + 2 \text{ or } y = 0.6x + 2$$

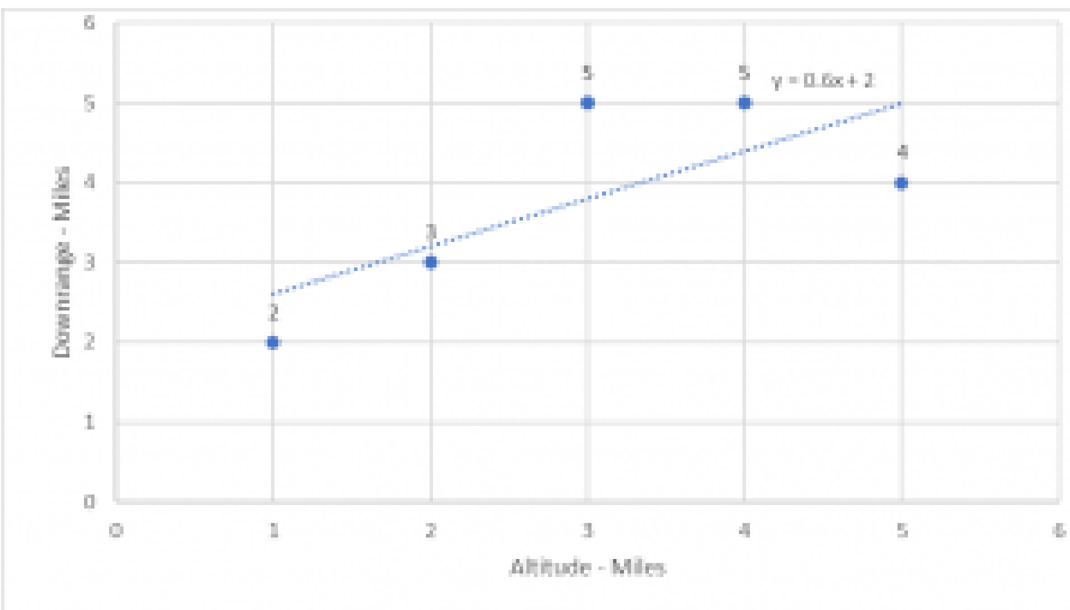


Figure 4.2 Graph of Linear Solution

Long Description

Figure 4.2 Graph of Linear Solution

As we can see, the linear solution offers an estimate that is closer to some of the given points than others. Can we do better by generating a curved line? (2nd degree polynomial)

The Quadratic Solution

The challenge is to expand on the above technique and apply it to develop the best fit quadratic equation.

In the linear, our goal was to solve two equations in two unknowns. Now we want to solve three equations in three unknowns. The unknowns are the constants of our quadratic equation in standard form:

Rearranging the standard form, we develop the Least Squares Summation equation:

$$E = \sum_{i=1}^{5} [y_1 - (ax_i^2 + bx_i + c)]^2$$

Now we take partial derivatives with respect to each of the three constants a, b, c as follows:

a –> $-2\sum_{i=1}^{5} [y_1 - (ax_i^2 + bx_i + c)]x_i^2 = 0$

b –> $-2\sum_{i=1}^{5} [y_1 - (ax_i^2 + bx_i + c)]x_i = 0$

c –> $-2\sum_{i=1}^{5} [y_1 - (ax_i^2 + bx_i + c)] = 0$

Simplify by dividing out the -2 and distributing the summation notation

$$\frac{d}{da} = \sum_{i=1}^{5} x_i^2 y_i - a\sum_{i=1}^{5} x_i^4 - b\sum_{i=1}^{5} x_i^3 - c\sum_{i=1}^{5} x_i^2 = 0$$

$$\frac{d}{db} = \sum_{i=1}^{5} x_i y_i - a\sum_{i=1}^{5} x_i^3 - b\sum_{i=1}^{5} x_i^2 - c\sum_{i=1}^{5} x_i = 0$$

$$\frac{d}{dc} = \sum_{i=1}^{5} y_i - a\sum_{i=1}^{5} x_i^2 - b\sum_{i=1}^{5} x_i - c5$$

Let's calculate the additional sums needed. We already calculated some of the sums for the linear equation. These are:

$$\sum_{i=1}^{5} x_i y_i = 63$$

$$\sum_{i=1}^{5} x_i^2 = 55$$

$$\sum_{i=1}^{5} x_i = 15$$

$$\sum_{i=1}^{5} y_i = 19$$

$$\sum_{i=1}^{5} 1 = 5$$

Additional sums:

$$\sum_{i=1}^{5} x_i^2 y_i = 239$$

$$\sum_{i=1}^{5} x_i^4 = 979$$

$$\sum_{i=1}^{5} x_i^3 = 225$$

Plugging in shows the three equations in three unknowns:

$$239 - a(979) - b(225) - c(55) = 0$$
$$63 - a(225) - b(55) - c(15) = 0$$
$$19 - a(55) - b(15) - c(5) = 0$$

Rearranging

$$979a + 225b + 55c = 239$$
$$225a + 55b + 15c = 63$$
$$55a + 15b + 5c = 19$$

Solving manually or using spreadsheet software the following equation is obtained:

$$y = -0.4286x^2 + 3.1714x - 1$$

This is the interpolation polynomial that generates a curved line (parabola) that is the best fit for the five given data points and it estimates y values for any other point within interval.

Matrix Operations simplify the calculations

Note: multiplying the transpose by the matrix produces the summation in n-equations with n-unknowns. This holds true no matter how many data points are involved.

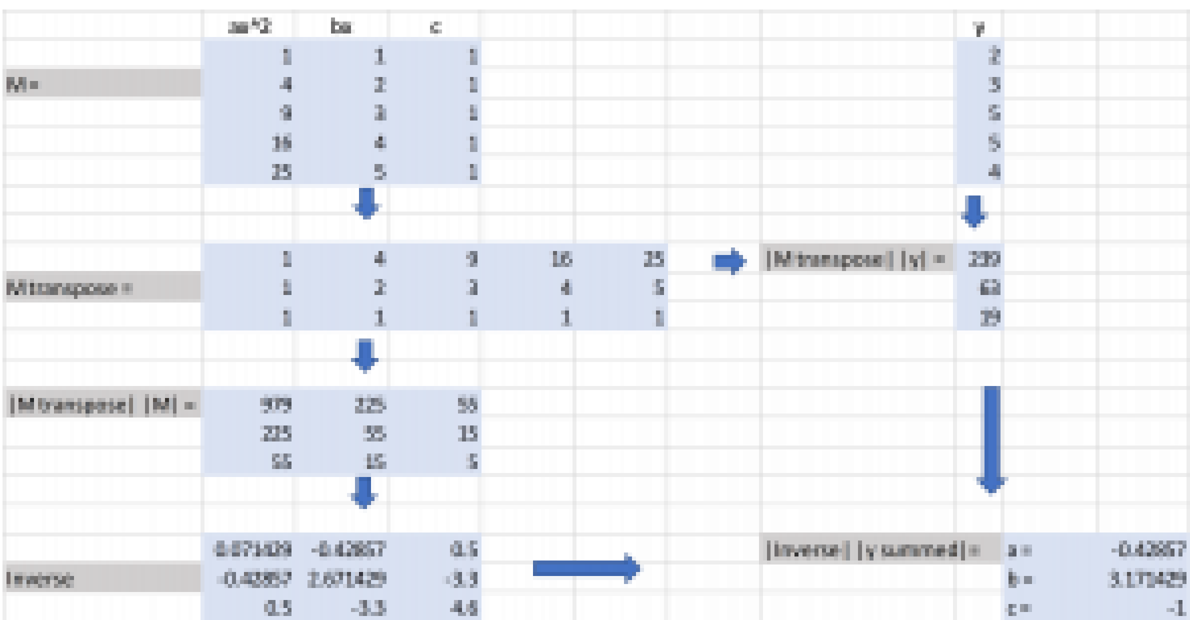


Figure 4.3 Matrix Operations Simplifying Calculations

Long Description

Figure 4.3 Matrix Operations Simplifying Calculations

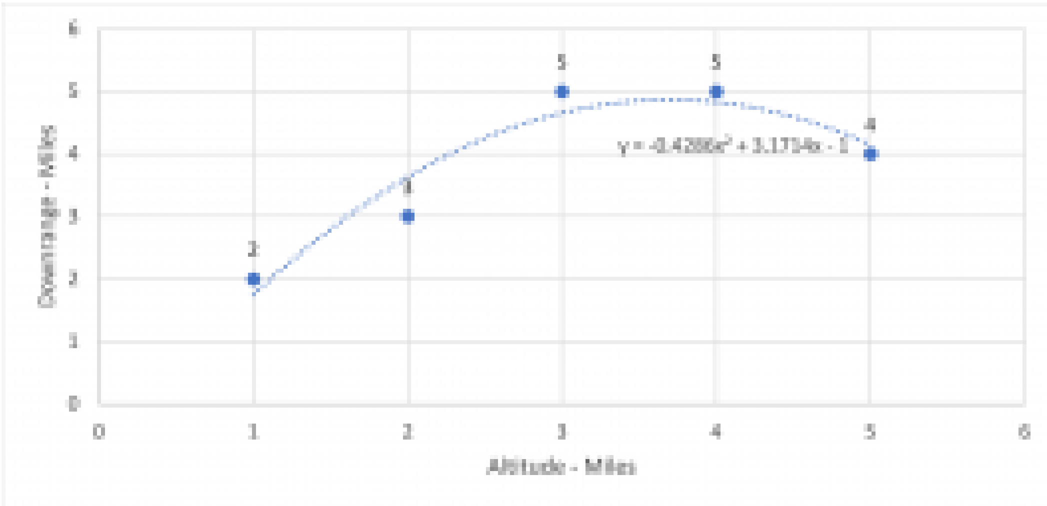


Figure 4.4 Graph of Quadratic Solution

Long Description

Figure 4.4 Graph of Quadratic Solution

Visually, the quadratic is a better fit than the linear solution.
In the next section we'll show how to measure the goodness of the fit quantitatively.

Chapter Four – Practice Exercise

4a)

Use weekly closing data for the Dow Jones Industrial Average and run a Least Squares Regression to produce a 3rd degree (cubic) interpolation polynomial. Plot the data on a chart for a visual representation. Solution given uses data from January 2020 through July 2021, during height of the COVID-19 pandemic.
(Solution given)

Chapter Five - Measuring the Least Squares Fit/ Exponential Least Squares Regression

How Well Does the Linear Polynomial Fit the Data?

It is natural and useful to ask: How good a predictor is the resulting polynomial for the given values of x. In other words, how close do the predicted values of y come to the actual values of y for a particular value of x.

Let's look at the chart for the linear regression we calculated (red dotted line) in Chapter Four. The length of red vertical lines between the actual and predicted values tells us how good the fit is. The smaller the red lines (closer), the better the fit.

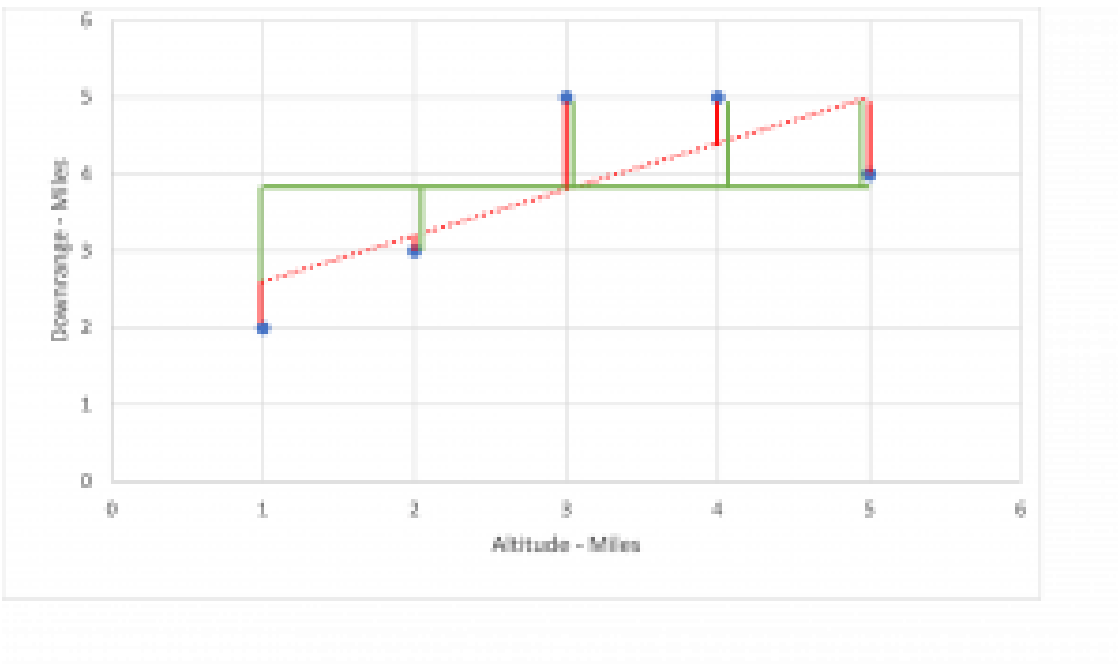


Figure 5.1 The Linear Fit

Long Description

Figure 5.1 The Linear Fit

However, simply measuring each distance and adding them together presents some problems. We want to eliminate direction because the negatives and positives tend to cancel each other out. An easy way to do this is measure each distance and then square the result. Hence the name Least Squared Regression.

Next, we need a baseline or something to compare our summed squared regression. It turns out a horizontal line passing through the mean of the y values offers us a worst-case scenario. In other words, the distance between the given y and the horizontal line is essentially no fit. So we add the given y values and divide by 5 (number of data points in this example). $Y_{mean} = \frac{2+3+5+5+4}{5} = 3.8$ Shown in green above. The closer the predicted value is from the actual value and the farther it is from the mean value, the better our prediction.

Using the data above we will conduct a R^2 (Squared Regression) analysis to gauge numerically how well the linear and quadratic polynomials fit the data.

Squared Regression Analysis

x	y	**Generated y values** $Y_* = 0.6x + 2$	**Difference between actual and generated squared:** $(y - y_*)^2$
1	2	2.6	0.36
2	3	3.2	0.04
3	5	3.8	1.44
4	5	4.4	0.36
5	4	5	1
-	-	-	$Sum = 3.2$

However, to put this in perspective we need to add a column and calculate the sum of the squared distance between the actual values of y and the mean value of y.

Squared Regression Analysis with Total Differences

x	y	**Generated y values** $Y_* = 0.6x + 2$	**Difference between actual and generated squared:** $(y - y_*)^2$	**Total Squared difference between actual and mean.** $(y - y_(mean))^2$
1	2	2.6	0.36	3.24
2	3	3.2	0.04	0.64
3	5	3.8	1.44	1.44
4	5	4.4	0.36	1.44
5	4	5	1	0.04
-	-	-	$Sum = 3.2$	$Sum = 6.8$

By taking the ratio of the sum of our squared error to the sum of the No-Fit values and subtracting from one we get a number (percent) that tells us how good our fit is in terms that is understandable.

$$R^2 = 1 - \frac{3.2}{6.8}$$
$$R^2 = 0.53 = 53\%$$

The R^2 value of 53% suggests that this may not be the best fit.

Let's calculate R^2 for the quadratic fit to see if it is a better fit.

Squared Regression Analysis with Different Generated y Values

x	y	Generated y values $y = -0.4286x^2 + 3.1714x - 1$	Difference between actual and generated squared: $(y - y_*)^2$	Total Squared difference between actual and mean. $(y - y_{mean})^2$
1	2	1.7428	0.06615184	3.24
2	3	3.6284	0.39488656	0.64
3	5	4.6568	0.11778624	1.44
4	5	4.828	0.029584	1.44
5	4	4.142	0.020164	0.04
-	-	-	$Sum = 0.62857264$	$Sum = 6.8$

$$R^2 = 1 - \frac{0.62857264}{6.8}$$
$$R^2 = 0.908 = 90.8\%$$

The quadratic is a better fit than the straight line. However, part of the "Art" of interpolation means the analyst still has to decide which is more meaningful and representative of the situation being analyzed.

Exponential Least Squares Regression

An important interpolation is one involving exponential polynomials. It has many applications in finance, biochemistry, and radioactive decay.

We will focus on the standard form using the constant e. This is known as the natural number or Euler's number value. Its importance lies in the fact that it represents the fundamental rate of growth shared by continually growing processes. One example is continuous compounding of money in a savings account.

The form of the polynomial is $y = Ae^{rx}$

In this, we can think of r as the rate and A we can think of as both the y intercept and demonstrating whether it is growth (positive value) or decay (negative value).

Graphically it looks like (A and r are both set to 1):

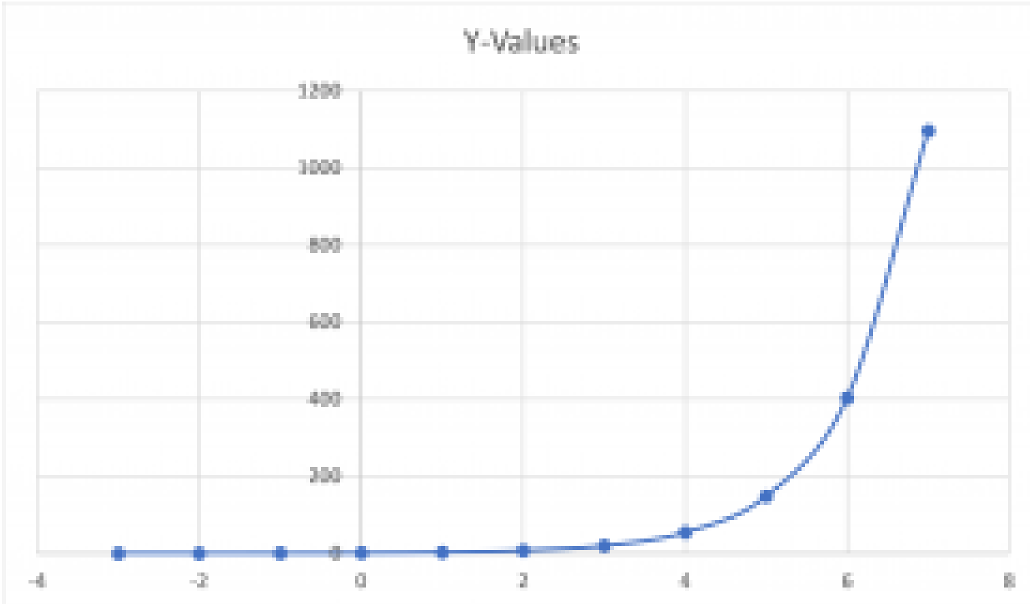


Figure 5.2 Exponential Growth

Long Description

Figure 5.2 Exponential Growth.

$y = Ae^{rx}$ does not lend itself to directly calculating an interpolative polynomial. This is due in part because standard deviation does not apply to this type of continuous and ever accelerating growth.

Since we already know how to deal with standard polynomials that can be solved used linear techniques such as matrix arithmetic, our goal is to eliminate e. Solve for r and A then plug the results back into the original polynomial.

Since we are dealing with the natural number e, we can convert the above to a linear function by taking the natural log of both sides as follows:

$$lny = ln(Ae^{rx})$$
$$lny = lnA + rx$$

When we rearrange, we have a linear equation in slope intercept form:

$$lny = rx + lnA$$

Let's use the following sample set of data points and use Matrix math to develop the interpolated data:

Interpolated Data

x	actual y	lny
-1	0.4	-0.916
0	1.1	0.095
1	2.62	0.963
2	8.1	2.092
3	24.03	3.179
4	57.9	4.059

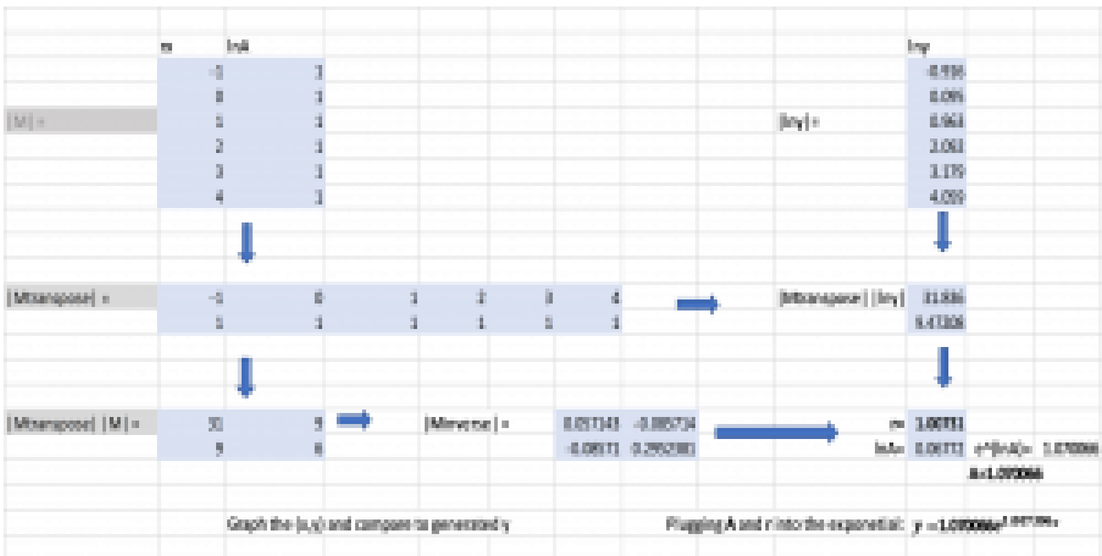


Figure 5.3 Matrix Math Solution

Long Description

Figure 5.3 Matrix Math Solution

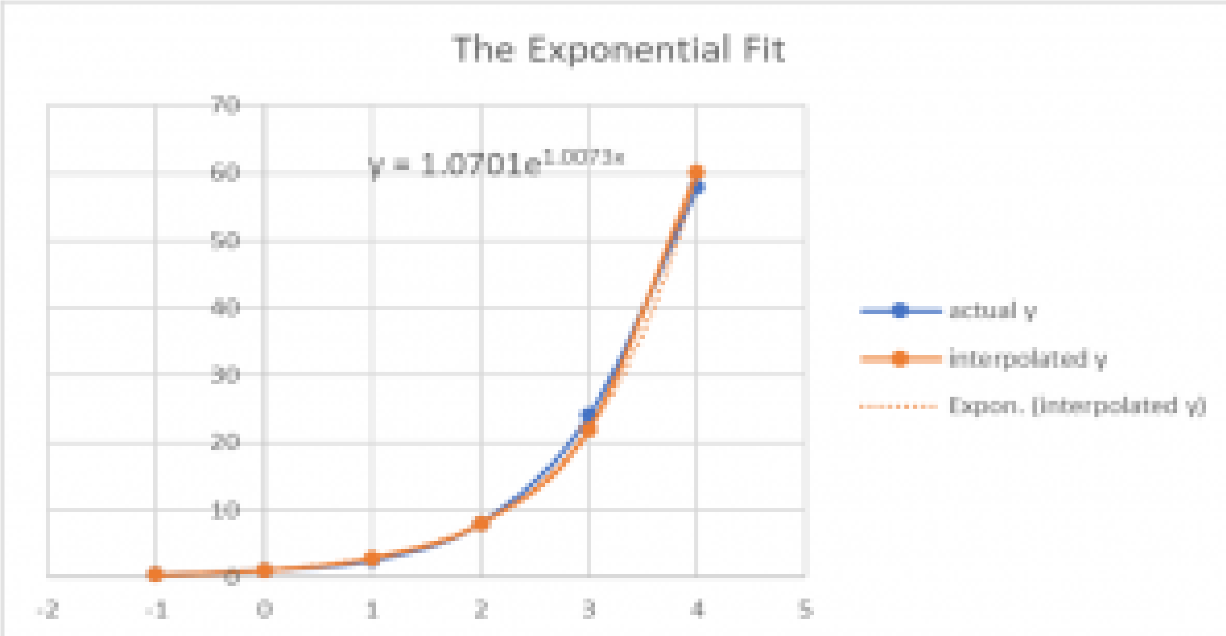


Figure 5.4 Graph of a line of fit for exponential function

Long Description

Figure 5.4 The Exponential Fit

This resulted in a very good fit.

Chapter Five - Practice Exercise

5a)

Measure the accuracy of the Fit from Exercise 4a, i.e. find R^2.
(Solution given)

Chapter Six - Approximation with Taylor Series

While this text is not about calculus, I believe it is important for students to become familiar with approximation using Taylor Series. References to derivatives are necessary but the actual derivatives in the examples will be given.

A way to think about Taylor Series polynomials is that they are simply a polynomial of any degree you wish to use that approximates a function being studied. Similar to Newton's divided difference we start with the simplest approximation, the constant.

Let's call our approximation $P(x)$. We will let $x = a$ be a particular point on the x-axis that will be the center of our approximation. The approximation improves the closer the value of x is to a. The function we are approximating is f(x).

For a straight line at a particular point, we can say an approximation polynomial is $p(a) = f(a)$.

Suppose we choose a point $x = a$, the graph might look something like:

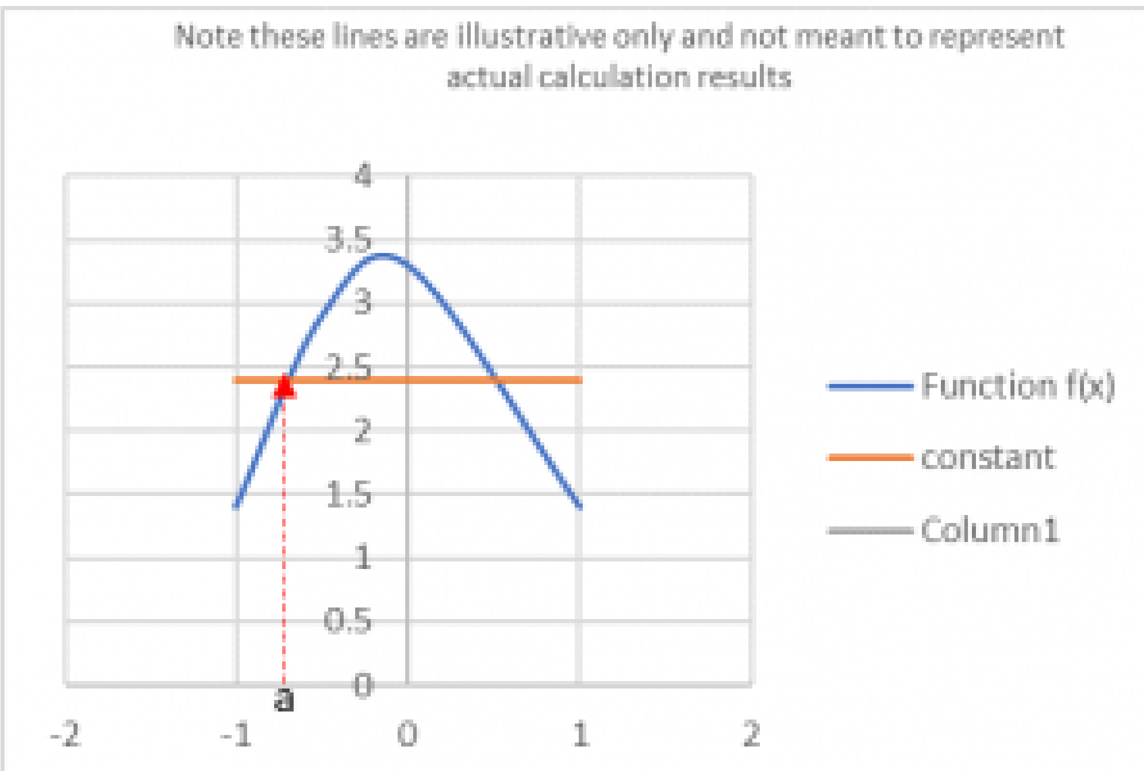


Figure 6.1 The Horizontal Straight Line Estimator

Long Description

Figure 6.1 The Horizontal Straight Line Estimator

At $x = a$, the horizontal line is an excellent approximation.

We can say that $p(a) = f(a)$ which is a constant.

Clearly, once we move away from a in either direction it turns out the constant does not serve us very well.

Our next step is adding a linear component while still retaining the constant. Which means we now have a polynomial that allows us to adjust the slope of the line. Let's try $p(a) = b_0 f(a) + b_1 f^{(1)}(a)(x)$ where $f^{(1)}$ is the first derivative of the function.

By adding the linear component, we can think of $b_1 f^{(1)}(a)$ as the slope. This improves our approximation:

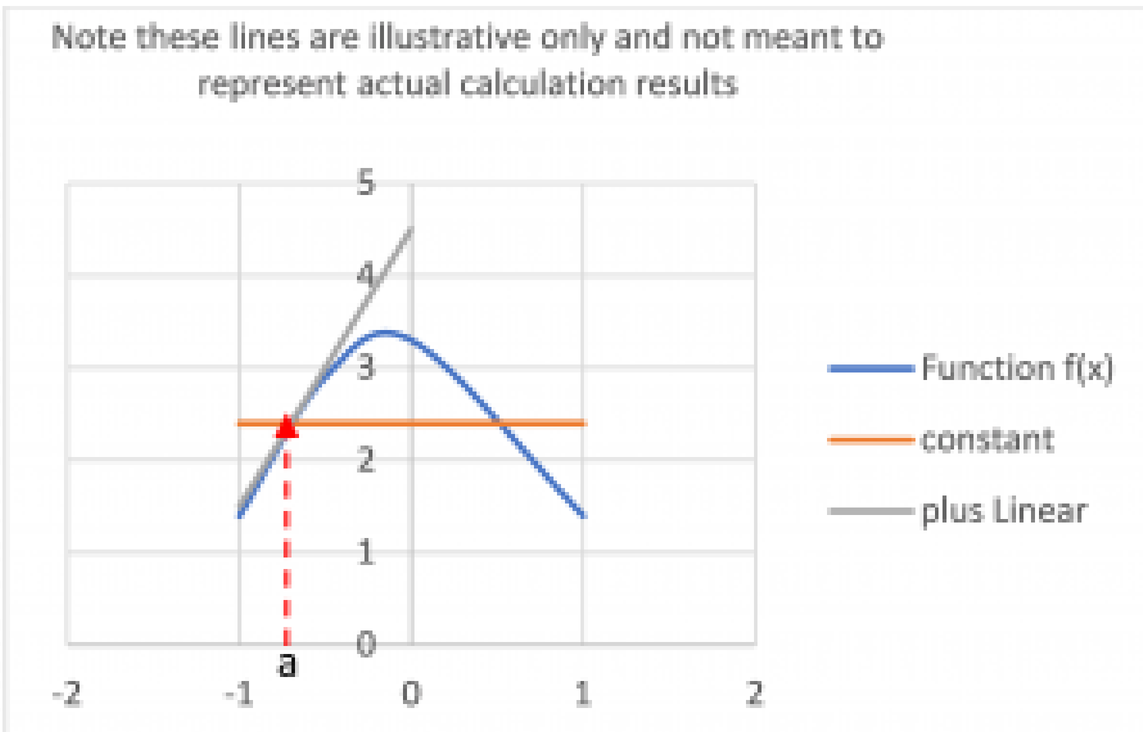


Figure 6.2 The Linear Solution

Long Description

Figure 6.2 The Linear Solution

By adding the linear component, we can see how the picture improves at the point $x = a$ because we now have a line tangent (representing the slope at $x = a$). Definitely an improvement over the constant as our approximation is pretty good as long as we stay near a.

So far, we have brought to bear a constant value and the slope. Because Taylor series allows us to add higher degree terms to our polynomial, we can now bring to bear the effect of concavity to the approximation. Think of concavity as adding curviness to what so far has been a straight line.

Let's add a quadratic (second degree) and cubic (third degree) component to our polynomial. These will introduce the curviness by adjusting the line at any given x value up or down. Figure 3 also illustrates the effect of higher order polynomials.

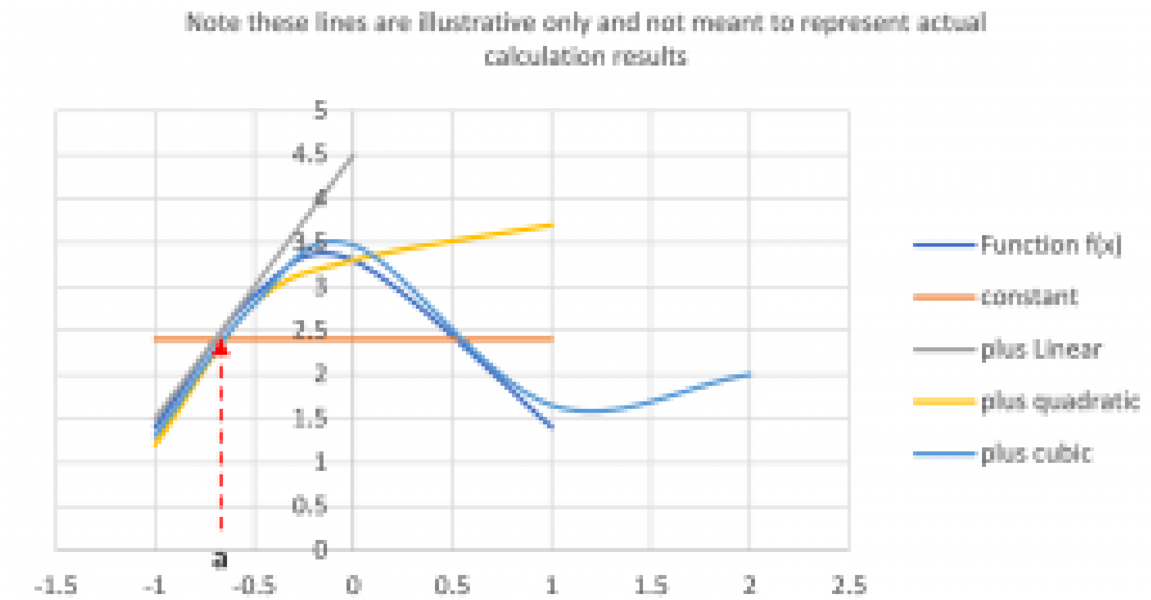


Figure 6.3 Effect of Higher Order Polynomials

Long Description

Figure 6.3 Effect of Higher Order Polynomials

We can see that as each higher-level component is added the approximation improves the farther we travel from the point x = a.

Quadratic: $b_0 f(a) + b_1 f^{(1)}(a)x + b_2 f^{(2)}(a)x^2$

Cubic: $b_0 f(a) + b_1 f^{(1)}(a)x + b_2 f^{(2)}(a)x^2 + b_3 f^{(3)}(a)x^3$

We could continue this indefinitely:

$$b_0 f(a) + b_1 f^{(1)}(a)x + b_2 f^{(2)}(a)x^2 + b_3 f^{(3)}(a)x^3 \cdots b_n f^{(n)} x^n$$

From here we will develop the general form of the Taylor series employing basic algebra.

This is done iteratively by solving one constant at a time. We set $a = o$ since in fact all Taylor polynomials either start with $x = 0$ or include the adjustment, $(x - a)$ so that in effect the center will always equal zero.

We will solve for a fourth-degree polynomial. This will be enough to demonstrate the general pattern of the Taylor series. To solve for each constant, we replace each of the $b_n f^{(n)}$ with c_n as follows:

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + c_4(x-a)^4$$

$$f(a) = c_0 + c_1(a-a) + c_2(a-a)^2 + c_3(a-a)^3 + c_4(a-a)^4$$

since $a - a = zero$

$$c_0 = f(a)$$

next we take first derivative of both sides

$$f^{(1)}(a) = c_1 + 2c_2(a-a) + 3c_3(a-a)^2 + 4c_4(a-a)^3$$

Again $a - a = zero$ so we are left with $f^{(1)}(a) = c_1$

$$c_1 = f^{(1)}(a)$$

The second derivative of both sides

$$f^{(2)}(a) = 2c_2 + 6c_3(a-a) + 12c_4(a-a)^2$$

since $a - a = zero$ we're left with $f^{(2)}(a) = 2c_2$

$$c_2 = f^{(2)}(a)/2$$

The third derivative of both sides

$$f^{(3)}(a) = 6c_3 + 24c_4(a-a)$$

since $a - a = zero$ we're left with $f_{(3)}(a) = 6c^2$

$$c_3 = f^{(3)}(a)/6$$

The fourth derivative

$$f^{(4)}(a) = 24c_4$$

$$c_4 = f^{(4)}(a)/24$$

Plugging in the solution for the four constants produces the general form:

$$\frac{f(a)}{1} + \frac{f^{(1)}(a)}{1}(x-a) + \frac{f^{(2)}(a)}{2}(x-a)^2 + \frac{f^{(3)}(a)}{6}(x-a)^3 + \frac{f^{(4)}(a)}{24}(x-a)^4$$

Normally we don't show the denominators when they are simply one. However, I've done so to illustrate the emerging pattern. Remember $0! = 1$ and $1! = 1$. This allows us to observe that the denominators are really successive factorials.

$$\frac{f(a)}{0!} + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \frac{f^{(4)}(a)}{4!}(x-a)^4$$

$$\cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Sin Function

Let's use an actual example to illustrate the process. Some things to remember. Taylor Series approximation only works for certain functions; typically, those that are continuous, repeatedly differential and irrational. They are also known as transcendental functions. Trig functions such as sin and cos, as well as exponential and logarithmic functions, imperfect roots, along with several other categories work well. Suppose we have been assigned a project to create our own App that will generate sin values.

We will focus on the mechanics of the process. For students who would like to delve deeper into Taylor Series there are a wealth of texts and videos available.

Step One: Select the function to be approximated. For this example, we will choose the sin function. It is well suited for Taylor Series approximation. It is continuous over the real numbers and it is repeatedly differential.

Step Two: Select an $x = a$ value that we want to center our approximation around. It turns out 0 degree is an easy value to work with as we differentiate sin.

Step Three: Repeatedly differentiate sin until the desired final degree of our Taylor Polynomial is reached. In this example we arbitrarily decided a ninth degree Taylor polynomial will produce Sin values accurate enough to meet our needs. Note we will work with radians as the angle measure.

Derivatives

$$f(0) = sin(0) = 0$$
$$f^{(1)}(0) = cos(0) = 1$$
$$f^{(2)}(0) = -sin(0) = 0$$
$$f^{(3)}(0) = -cos(0) = -1$$
$$f^{(4)}(0) = sin(0) = 0$$
$$f^{(5)}(0) = cos(0) = 1$$
$$f^{(6)}(0) = -Sin(0) = 0$$
$$f^{(7)}(0) = -cos(0) = -1$$
$$f^{(8)}(0) = sin(0) = 0$$
$$f^{(9)}(0) = cos(0) = 1$$

Step Four: Plug in our derivatives into the general form of the Taylor polynomial:

$$p(0) = \frac{0}{0!} + \frac{1}{1!}(x-a) + \frac{0}{2!}(x-a)^2 + \frac{-1}{3!}(x-a)^3 + \frac{0}{4!}(x-a)^4 + \frac{1}{5!}(x-a)^5 + \frac{0}{6!}(x-a)^6 + \frac{-1}{7!}(x-a)^7 + \frac{0}{8!}(x-a)^8 + \frac{1}{9!}(x-a)^9$$

Since every other term has zero in the numerator we can drop these and condense p(0). Further since a = 0, we can simplify the binomials.

The resulting Taylor Series polynomial is:

\large p(0) = \frac {1}{1!}(x) + \frac {-1}{3!}(x)^3 + \frac {1}{5!}(x)^5 + \frac {-1}{7!}(x)^7 + \frac {1}{9!}(x)^9

We have a relatively simple polynomial we can program into our app to produce values of sin for angles between 0 and 90 degrees . Since sin is periodic, we can program in computations that give us the reference angle for angles greater than 90 or less than 0 degrees .

Step Five: We are now ready to test p(a) for various angles between 0 and 90 degrees. Since it is easier to work with Radians, I've included a conversion for students not familiar with them. f(x) is generated from an app precise to 15 decimal positions. p(x) is our Taylor approximation.

Step 5 of Taylor Approximation

Degrees	Radians	f(x)	p(x)
0	0	0.000000000000000	0.000000000000000
18	$\frac{\pi}{10}$	0.309016994374947	0.309016994375021
22.5	$\frac{\pi}{8}$	0.382683432365090	0.382683432365947
30	$\frac{\pi}{6}$	0.500000000000000	0.500000000000000
45	$\frac{\pi}{4}$	0.707106781186547	0.707106782936867
72	$\frac{2\pi}{5}$	0.951056516295154	0.951056822327524
90	$\frac{\pi}{2}$	1.000000000000000	1.000003542584290

p(x) provides an excellent approximation out to at least six decimal places for the values of x we tested. The symmetry and reflectivity properties of the sin function will allow us to generate values less than 0^0 and greater than 90^0.

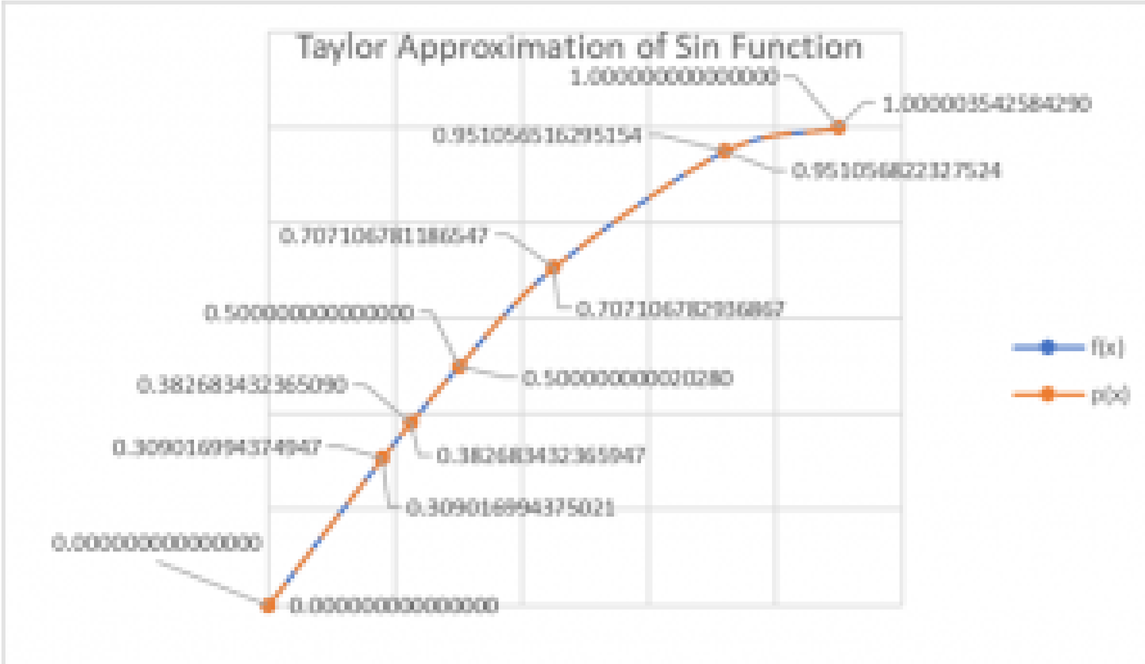


Figure 6.4 Graph of Taylor Approximation

Long Description

Figure 6.4 Taylor Approximation of Sin Function

Note: Difference slight enough that lines appear to overlap on the graph.

Chapter Six – Practice Exercise

6a)

Replicate the above example (sin) for the cos. Compare the resulting graph to the one for sin.
(Solution given)

Chapter Seven - Taylor Series Remainder Test

A formal way to test the accuracy of a Taylor polynomial approximation is to employ the Taylor Remainder test. By adding a remainder term to our Taylor polynomial approximation, we in effect convert it into an equation,

Our function $f(x) = p(x) + remainderterm$. Written more compactly we have f_n(x) = p_n(x) + r_n(x) . This remainder term becomes the difference between $f(x)$ at a particular point and $p(x)$ at that same value of x.

In the above example we ran our polynomial out to the ninth-degree term.

$r_n(x)$ actually looks like the next higher degree term:

$$r_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$ **where c is between a and x**

The question we ask is what value for c should we use. The answer in this case is to solve the remainder twice for the endpoints of the range we are interested in. In this case we want to know how accurate c will be between 0 and $\frac{\pi}{10}$.

This will provide a range of possible values between 0 and $\frac{\pi}{10}$

$$f^{(10)}(c) = -sin(c)$$

For $c = 0$ $r_9(0) = \frac{-sin(0)}{(10)!}(0-0)^{10} = 0$

For $c = \frac{\pi}{10}$ $r_9(\frac{\pi}{10}) = \frac{-sin(\frac{\pi}{10})}{(10)!}(\frac{\pi}{10} - 0)^{10} = -0.000000000000797$ we drop the negative as it's a matter of distance, not direction.

This bounds the possible error of our approximation: $0 \leq r_{10} \leq 0.000000000000797\,(7.97 * 10^{-13})$

Chapter Seven - Practice Exercise

7a)

Conduct the Taylor Remainder Test on your solution for Practice Problem 6a.
(Solution given)

Solutions to Selected Practice Exercises

Solution to Exercise One Practice Problems

Exercise 1a)

ABC Children's Party Company

Maximum children attending the party	Cost per Child	Total Cost of Party
10	$37	$370
25	$28	$700
50	$22	$1100
100	$15	$1500

The four equations in four unknowns:

$$a(10^3) + b(10^2) + c(10) + d = 37$$
$$a(25^3) + b(25^2) + c(25) + d = 28$$
$$a(50^3) + b(50^2) + c(50) + d = 22$$
$$a(100^3) + b(100^2) + c(100) + d = 15$$

Equations in Table Form

a	b	c	d	cost
1000	100	10	1	37
15625	625	25	1	28
125000	2500	50	1	22
1000000	10000	100	1	15

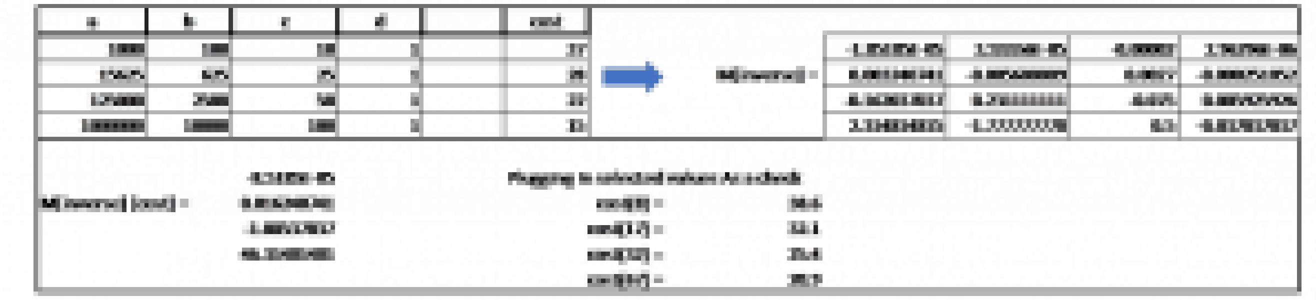


Figure 8.1 Matrix Setup for Exercise 1a

Long Description

Figure 8.1 Matrix Setup for Exercise 1a

Resulting Pricing Polynomial

$$cost = (-8.52 * 10^{-5})x^3 + (1.62 * 10^{-2})x^2 + (-1.09)x + (4.63 * 10^1)$$

	a_1x^3	a_2x^2	a_3x	a_4			y
	-5.35938	3.0625	-1.75	1			-2
|M| =	1	1	1	1		|Y| =	-3.7
	35.937	10.89	3.3	1			-1.4
	328.509	47.61	6.9	1			4
	Matrix Multiplication Operation						
							Constants
$\|M^{-1}\|$ =	-0.00832	0.026797	-0.02392	0.005443	$\|M^{-1}\|\|Y\|$ =	a_1=	-0.027247
	0.093235	-0.22644	0.14708	-0.01388		a_2=	0.38991181
	-0.27446	0.131842	0.165614	-0.023		a_3=	-0.2627393
	0.18955	1.067797	-0.28878	0.031433		a_4=	-3.7999255

Solution Polynomial: $-0.0272x^3 + 0.3899x^2 - 0.2627x - 3.7999 = p(x)$

Graph of Solution

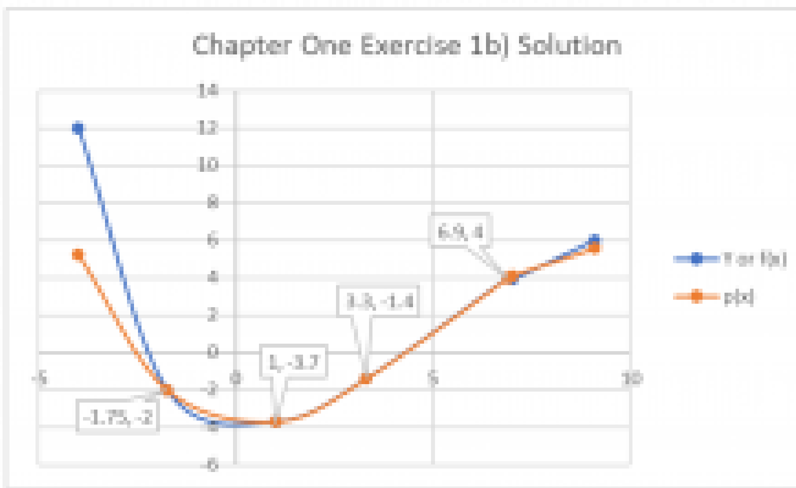


Figure 8.2 Exercise 1b

Long Description

Figure 8.2 Exercise 1b)

Solution to Exercise Two Practice Problems

2a)

Newton's Divided Difference Table is populated as follows:

Newton's Divided Difference Table

x	y	b_0	**Linear** $b_1(x-10)$	**Quadratic** $b_2(x-10)(x-25)$	**Cubic** $b_3(x-10)(x-25)(x-50)$
10	37	37	–	–	–
-	–	–	$\dfrac{37-28}{10-25}=-0.6$	–	–
25	28	28	–	$\dfrac{-0.6-(-.24)}{-40}=0.009$	–
-	–	–	$\dfrac{28-22}{25-50}=-0.24$	–	$\dfrac{0.009-0.001}{-90}=0.0000888$
50	22	22	–	$\dfrac{-0.24-(-0.14)}{-75}=0.001$	–
-	–	–	$\dfrac{22-15}{50-100}=-0.14$	–	–
100	15	15	–	–	–

Simplifies to: $-0.0000888x^3+0.016548x^2-1.0926x+46.36$

2b)

2b Table

x	y or f(x)
-6.2	-8
-3	-7
-1.5	-2.2
1	0.7
3.5	3
4.25	5
7.9	8

2b Difference Table

x	f(x)	1st Divided Difference	2nd Divided Difference
-	b_0	$b_1(x-x_0)$	$b_2(x-x_0)(x-x_1)$
-3	-7	-	-
-	-	$\frac{-7-0.7}{-3-1}=\frac{-7.7}{-4}=1.925$	-
1	0.7	-	$\frac{1.925-1.493}{-3-7.9}=\frac{0.432}{-10.9}=-0.040$
-	-	$\frac{0.7-11}{1-7.9}=\frac{-10.3}{-6.9}=1.493$	-
7.9	11	-	-

Simplifies to: $-0.040x^2 + 1.845x - 1.105$

Solution to Chapter Three Practice Exercises

Exercise 3b)

	a_1x^3	a_2x^2	a_3x	a_4		y
$\|M\|=$	216	36	6	1.0	$\|Y\|=$	14.5
	343	49	7	1.0		17
	512	64	8	1.0		20
	669.92188	76.5625	8.75	1.0		23

$\|M^{-1}\|=$	-0.1818182	0.571429	-0.666666667	0.3	$\|M\text{-}1\|\|Y\|=$	$a_1=$	0.116883
	4.3181818	-13	14.5	-5.8		$a_2=$	-2.20455
	-34.045455	97.42857	-103.8333333	40.5		$a_3=$	16.31494
	89.090909	-240	245	-93.1		$a_4=$	-29.2727

Solution $p(x) = 0.116883\,x^3 - 2.20455\,x^2 + 16.31494\,1x - 29.2727$

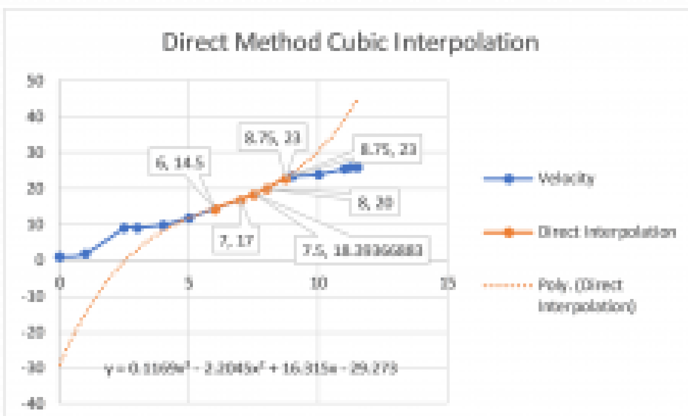


Figure 8.3 Exercise 3b

Long Description

Figure 8.3 Exercise 3b

Solution to Chapter Four Practice Exercise

4a)

The Setup:

Abbreviated List of weekly Dow Jones closing averages:

Weekly Closing Averages

Week	Actual	$a_1 x^3$	$a_2 x^2$	$a_3 x$	a_4	Interpolation
1	28,583.68	1.00	1.00	1.00	1	28,416.89149
2	28,939.67	8.00	4.00	2.00	1	28,034.20169
3	29,196.04	27.00	9.00	3.00	1	27,677.60694
4	28,722.85	64.00	16.00	4.00	1	27,346.5493
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
78	34,292.29	474,552.00	6,084.00	78.00	1	34,491.0287
79	34,577.37	493,039.00	6,241.00	79.00	1	34,485.14307
80	34,888.79	512,000.00	6,400.00	80.00	1	34,462.39157
81	34,511.99	531,441.00	6,561.00	81.00	1	34,422.21628
82	35,058.52	551,368.00	6.724.00	82.00	1	34,364.05925
83	35,084.53	571,787.00	6,889.00	83.00	1	34,287.36256

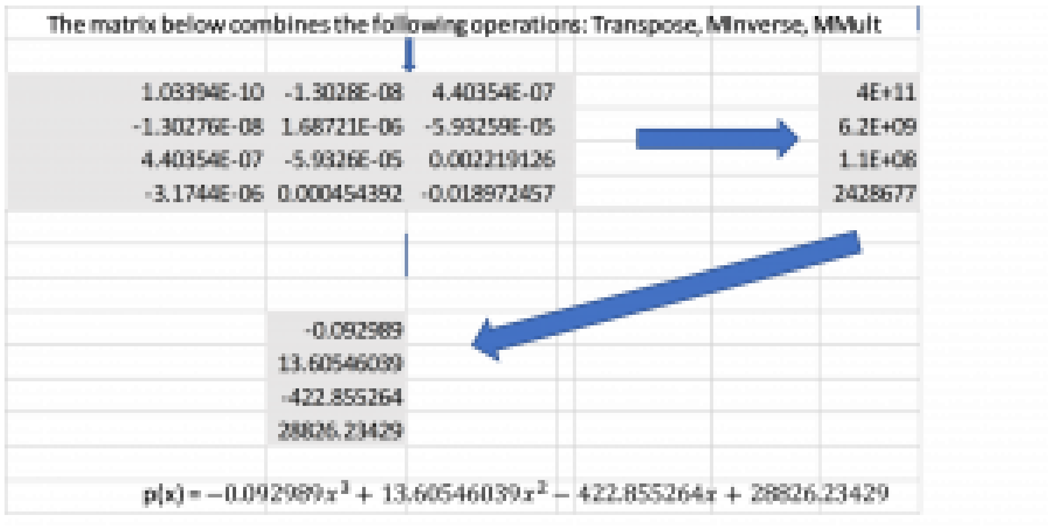


Figure 8.4 Matrix Solution 4a

Long Description

Figure 8.4 Matrix Solution 4a

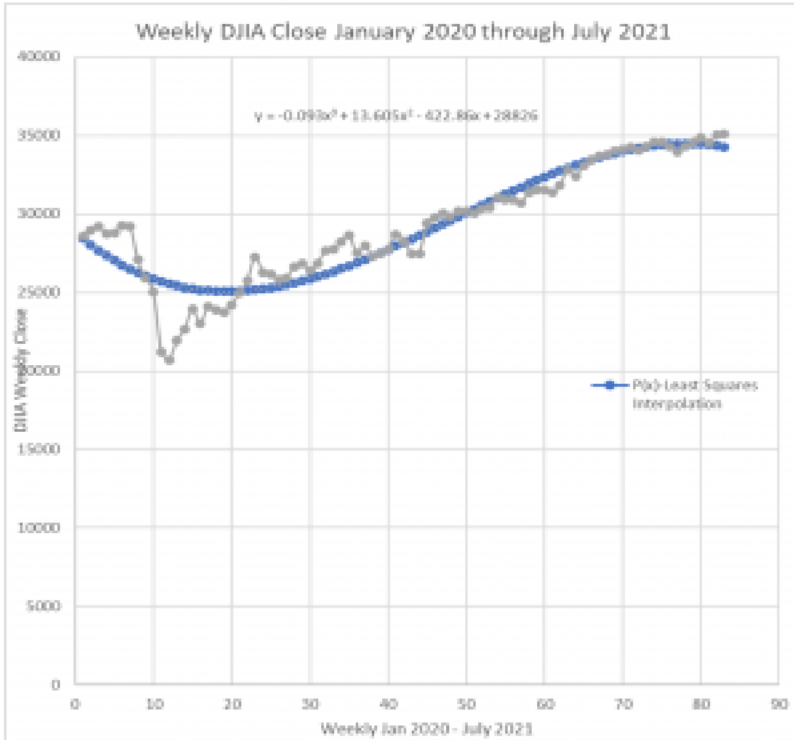


Figure 8.5 Graph of Weekly DJIA

Long Description

Figure 8.5 Weekly DJIA Close January 2020 through July 2021

Solution to Chapter Five Practice Exercises

Step One:

1a) Find the difference between each actual value and its associated value generated by the interpolative polynomial. Square the result.

1b) Find the difference between each actual value and the Mean of the actual values. Square the result.

Step Two:

2a) Sum the results from 1a

2b) Sum the results from 1b

Step Three:

Divide 2a by 2b subtracting the result from 1.

Answer: $R^2 = 0.882707285 \approxeq 88.3\%$

Solution to Chapter Six Practice Exercises

6a)

Select the Function to be approximated. Cos function centered at x=0

Derivatives of cos

$$f(0) = cos(0) = 1$$
$$f^{(1)}(0) = -sin(0) = 0$$
$$f^{(2)}(0) = -cos(0) = -1$$
$$f^{(3)}(0) = sin(0) = 0$$
$$f^{(4)}(0) = cos(0) = 1$$
$$f^{(5)}(0) = -sin(0) = 0$$
$$f^{(6)}(0) = -cos(0) = -1$$
$$f^{(7)}(0) = sin(0) = 0$$
$$f^{(8)}(0) = cos(0) = 1$$
$$f^{(9)}(0) = -sin(0) = 0$$

Plug derivatives into the general form of the Taylor polynomial:

$$p(0) = \frac{1}{0!} + \frac{0}{1!}(x-a) + \frac{-1}{2!}(x-a)^2 + \frac{0}{3!}(x-a)^3 + \frac{1}{4!}(x-a)^4 + \frac{0}{5!}(x-a)^5 + \frac{-1}{6!}(x-a)^6 + \frac{0}{7!}(x-a)^7 + \frac{1}{8!}(x-a)^8 + \frac{0}{9!}(x-a)^9$$

Every other term has zero in the numerator so we can drop these and condense p(0). Further since a = 0 we can simplify the binomials.

$$p(0) = \frac{1}{0!} + \frac{-1}{2!}(x)^2 + \frac{1}{4!}(x)^4 + \frac{-1}{6!}(x)^6 + \frac{1}{8!}(x)^8$$

f(x) is generated from an app precise to 15 decimal positions. p(x) is the Taylor approximation for Cosine.

Figure 8.6 Speech Bubble

Taylor Approximation for Cosine

Degrees	**Radians**	**f(x)**	**p(x)**
0	0	1.000000000000000	1.000000000000000
18	$\frac{\pi}{10}$	0.951056516295154	0.951056516297732
22.5	$\frac{\pi}{8}$	0.923879532511287	0.923879532535293
30	$\frac{\pi}{6}$	0.866025403784439	0.866025404210352
45	$\frac{\pi}{4}$	0.707106781186548	0.707106805683294
72	$\frac{2\pi}{5}$	0.309016994374947	0.309019668329804
90	$\frac{\pi}{2}$	0.000000000000000	0.000000000000000

Solution to Chapter Seven Practice Exercise

7a)

$r_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$ where c is between a and x

Solving the remainder twice for 0 and $\frac{\pi}{10}$

This will provide a range of possible values between 0 and $\frac{\pi}{10}$

$$f^{(9)}(c) = -sin(c)$$

for $c = 0 \quad r_8(0) = \frac{-sin(0)}{(9)!}(0-0)^9 = 0$

for $c = \frac{\pi}{10} \quad r_8(\frac{\pi}{10}) = \frac{-sin(\frac{\pi}{10})}{(9)!}(\frac{\pi}{10} - 0)^9 = -0.000000000025384$

Drop negative as it is a matter of distance not direction.

Gives us an error possibility $0 \leq r_{10} \leq 0.000000000025384 \ (2.54 * 10^{-11})$

ACKNOWLEDGEMENTS

Content Editor: Darius Rub, MS University at Buffalo
Content Editor: Charlene Cardinale, Math and Science educator
Cover Design: Denise J. Murphy-Rohr, Graphic Designer
Copy Editor: Christina Riehman-Murphy, Open & Affordable Educational Resources Librarian, Penn State Libraries

About the Author

Figure 10.1 Stuart Murphy

Stuart Murphy spent a number of years working in the insurance industry, managing and implementing health plans for commercial and government entities. During this time, he also served as a registered lobbyist.

Over the years Stu has taught middle, high school, and college level math; as well as COBOL and Assembler. Stu currently teaches middle school mathematics.

He and his wife Sharon have three children and eight grandchildren. They make their home in Pennsylvania.

Versioning History

This page will provide a record of edits and changes made to this book since its initial publication in July 2022. Whenever edits or updates are made, we make the required changes in the text and provide a record and description of those changes here. If the change is minor, the version number increases by 0.1. However, if the edits involve substantial updates, the version number goes up to the next full number.

If you find an error in this book, please contact smurph11@gmail.com or cer20@psu.edu. We will make the necessary changes, and update this Versioning History page to reflect the edits made.

Version	Date	Change	Details
1.1	July 2022	Initial Publication	–