

An investigative study on the best mathematical model for the spiral of a mollusk

Personal Code: gzk284

1. Introduction

Circular objects and optical illusions, especially those with spirals, have always astonished my mind. Being a frequent visitor of the beach, I have come across various seashells at the beach, namely Gastropods, Bivalves and Mollusks. While mollusks are really widespread, the spiral of a mollusk wallows me into the depths of infinity; yet, when I attempt to reach my encyclopedias about mollusks, most sources make little to no mention about its spiral, some making references to the Fibonacci spiral, a special kind of logarithmic spiral praised for its “aesthetical pleasing appearance” (O’Callaghan, 2014). Since then, I’ve always wondered whether this is true or not.

In my Mathematics Option Statistic classes, I learnt that a bivariate population can be modeled into a regression line that demonstrates the overall trend of data. However, we didn’t have the opportunity to delve into the specific calculations for polynomial regression since we haven’t started with calculus. When I researched about regression, I heard the rumor that “if one was given 100 data points of data, it would be most accurate to interpolate the 100 data points to a 99th order polynomial” or in other words, “the more (degrees of the polynomial) the merrier” (Kaw, 2008). I question the validity of this statement, and I would like to challenge myself by using this newly obtained skill to model the curve of the mollusk spiral. I will first use regression to create 6 models of the spiral, namely:

- 1) The **arithmetic spiral** since it is known for an arithmetic progression of radius from origin to a point on the same radial, which makes it the easiest spiral to model (Lockwood, 1967).
- 2) The **logarithmic spiral**, since a special type of logarithmic spiral, namely the Fibonacci spiral, is known for being omnipresent in many objects in nature, including seashells.
- 3) The **quadratic, cubic, quartic and quintic spirals**, since it was claimed that a higher-degree polynomial is always better, though a polynomial with degree 6 or above will be extremely difficult to compute in practicality.

Then, based on the spirals modelled with regression, I will determine which spiral is the best model by: (1) comparing their total arc lengths, and; (2) performing an F-test of overall significance (ANOVA method) with respect to the actual mollusk spiral. With these factors in mind, I will conclude which spiral best resembles my mollusk.

2. Data Collection

Since I do not have a scanner to scan my mollusks properly, I have accessed an online image of a mollusk for data collection of its spiral. Using a powerful graphing utility called GeoGebra, systematic sampling has been applied to collect points at approximately every $\frac{\pi}{6}$ radians, as outlined by **Figure 1**. In terms of raw data, appendix 3 outlines the x coordinates, y-coordinates, the radius and the angle in radians of each point when plotted on a cartesian plane.

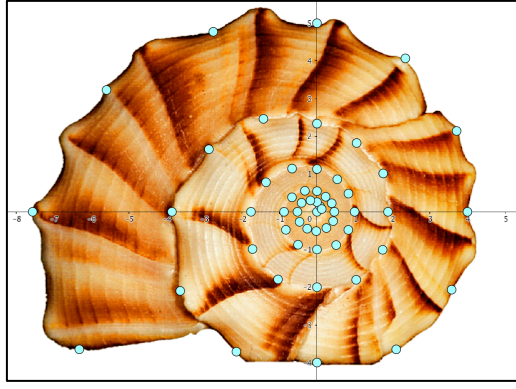


Figure 1 Mollusk Spiral (Bartzen, 2015), with points plotted at intervals of $\frac{\pi}{6}$ radians (devised by author, i.e. me).

3. Modelling the curves using regression

In my science classes, I learnt that regression or “curve-fitting” is often used by statisticians to describe trends between variables. I wonder how well this approach works for mathematical models. In the following part, I will utilize regression to calculate the coefficients of the polynomial spirals. I will use the quintic (5th degree) spiral as an example since it will be the hardest one to compute.

Consider a general quintic function $f(\theta)$, where **$f(\theta)$ is modelled radius, θ is angle in radians**, and a_0, a_1, a_2, a_3, a_4 and a_5 are **coefficients** of $\theta^0, \theta^1, \theta^2, \theta^3, \theta^4$ and θ^5 respectively:

$$f(\theta) = a_0 + a_1\theta^1 + a_2\theta^2 + a_3\theta^3 + a_4\theta^4 + a_5\theta^5 \text{ (Function 1)}$$

Firstly, I will use a least squares regression approach (Efunda, 2018); when a proposed quintic polynomial is fit onto the points, there will be 55 “deviations” for the 55 “differences” or errors between the points on $f(\theta)$ and the 55 actual data points on the mollusk. However, since the errors may be negative, to achieve positive absolute errors, I will square these errors to eliminate negative errors. In the following, the sum of all the 55 “errors have been minimized by calculus to find the quintic function; the **total error E** which we want to minimize using the least squares approach, in the context of a spiral with a parametric curve where **r is radius of point from origin and θ is angle in radians**, is:

$$E = \sum_{i=1}^{55} (D_i)^2 \quad \text{Where}$$

$$= \sum_{i=1}^{55} (r_i - f(\theta_i))^2 \text{ (Function 2)} \quad \begin{array}{l} E \text{ is total error} \\ (\theta_i, f(\theta_i)) \text{ are points on } f(\theta) \text{ corresponding to } (\theta_i, r_i) \text{ on actual mollusk} \\ D_i \text{ is distance between } (\theta_i, r_i) \text{ and } (\theta_i, f(\theta_i)) \end{array}$$

$i = 1, 2, 3, \dots, 55$

But what does this mean in the context of a quintic spiral? Indeed, when I substitute the general quintic function $f(\theta)$ (function 1) into the total error function (function 2), I obtain:

$$E = \sum_{i=1}^{55} [r_i - (a_0 + a_1\theta_i^1 + a_2\theta_i^2 + a_3\theta_i^3 + a_4\theta_i^4 + a_5\theta_i^5)]^2 \text{ (Function 3)}$$

In the following, my aim is to find $a_0, a_1, a_2, \dots, a_5$ such that E is minimized. I will perform partial differentiation with respect to $a_0, a_1, a_2, \dots, a_5$ and have each derivative set to zero for minimization. Since this is partial differentiation, I simply treated the other variables than the variable I am differentiating as constants (detailed steps given in appendix 1). Applying the chain rule, I found that:

$$\begin{aligned} \frac{\partial E}{\partial a_0} &= \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)]^2 \\ &= -2 \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \end{aligned}$$

Similarly, when partial differentiation is performed with respect to $a_1, a_2, a_2 \dots, a_5$, I obtain:

$$\begin{aligned} \frac{\partial E}{\partial a_1} &= -2 \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \theta_i^1 \\ \frac{\partial E}{\partial a_2} &= -2 \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \theta_i^2 \\ &\vdots \\ \frac{\partial E}{\partial a_5} &= -2 \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \theta_i^5 \end{aligned}$$

I then equate all the derivatives to 0, which will give me the turning points. However, as you may recall, I should be calculating the solutions (coefficients of the quintic spiral) with respect to the minima since I want to minimize the error; normally, I would perform a second derivative test to prove that it is at a minimum, but since there is only one set of solutions, this test is not necessary and will be skipped.

$$\begin{aligned} \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \theta_i^0 &= 0 \\ \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \theta_i^1 &= 0 \\ &\vdots \\ \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \theta_i^5 &= 0 \end{aligned}$$

Doesn't this look confusing? Hence, I multiplied the θ_i^x (variable at the left of the "=" sign) into each line and rearranged the equations in a manner such that all variables in the equations are positive:

$$\begin{aligned} \sum_{i=1}^{55} a_0 + \sum_{i=1}^{55} a_1 \theta_i + \sum_{i=1}^{55} a_2 \theta_i^2 + \sum_{i=1}^{55} a_3 \theta_i^3 + \sum_{i=1}^{55} a_4 \theta_i^4 + \sum_{i=1}^{55} a_5 \theta_i^5 &= \sum_{i=1}^{55} r_i \\ \sum_{i=1}^{55} a_0 \theta_i + \sum_{i=1}^{55} a_1 \theta_i^2 + \sum_{i=1}^{55} a_2 \theta_i^3 + \sum_{i=1}^{55} a_3 \theta_i^4 + \sum_{i=1}^{55} a_4 \theta_i^5 + \sum_{i=1}^{55} a_5 \theta_i^6 &= \sum_{i=1}^{55} r_i \theta_i \\ &\vdots \\ &3 \end{aligned}$$

$$\sum_{i=1}^{55} a_0 \theta_i^5 + \sum_{i=1}^{55} a_0 \theta_i^6 + \sum_{i=1}^{55} a_0 \theta_i^7 + \sum_{i=1}^{55} a_0 \theta_i^8 + \sum_{i=1}^{55} a_0 \theta_i^9 + \sum_{i=1}^{55} a_0 \theta_i^{10} = \sum_{i=1}^{55} r_i \theta_i^5$$

I still think this looks messy. What happens when I place them in matrices? Indeed, here's what I get:

$$\begin{bmatrix} 55 & \sum_{i=1}^{55} \theta_i & \sum_{i=1}^{55} \theta_i^2 & \sum_{i=1}^{55} \theta_i^3 & \sum_{i=1}^{55} \theta_i^4 & \sum_{i=1}^{55} \theta_i^5 \\ \sum_{i=1}^{55} \theta_i & \sum_{i=1}^{55} \theta_i^2 & \sum_{i=1}^{55} \theta_i^3 & \sum_{i=1}^{55} \theta_i^4 & \sum_{i=1}^{55} \theta_i^5 & \sum_{i=1}^{55} \theta_i^6 \\ \sum_{i=1}^{55} \theta_i^2 & \sum_{i=1}^{55} \theta_i^3 & \sum_{i=1}^{55} \theta_i^4 & \sum_{i=1}^{55} \theta_i^5 & \sum_{i=1}^{55} \theta_i^6 & \sum_{i=1}^{55} \theta_i^7 \\ \sum_{i=1}^{55} \theta_i^3 & \sum_{i=1}^{55} \theta_i^4 & \sum_{i=1}^{55} \theta_i^5 & \sum_{i=1}^{55} \theta_i^6 & \sum_{i=1}^{55} \theta_i^7 & \sum_{i=1}^{55} \theta_i^8 \\ \sum_{i=1}^{55} \theta_i^4 & \sum_{i=1}^{55} \theta_i^5 & \sum_{i=1}^{55} \theta_i^6 & \sum_{i=1}^{55} \theta_i^7 & \sum_{i=1}^{55} \theta_i^8 & \sum_{i=1}^{55} \theta_i^9 \\ \sum_{i=1}^{55} \theta_i^5 & \sum_{i=1}^{55} \theta_i^6 & \sum_{i=1}^{55} \theta_i^7 & \sum_{i=1}^{55} \theta_i^8 & \sum_{i=1}^{55} \theta_i^9 & \sum_{i=1}^{55} \theta_i^{10} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{55} r_i \\ \sum_{i=1}^{55} r_i \theta_i \\ \sum_{i=1}^{55} r_i \theta_i^2 \\ \sum_{i=1}^{55} r_i \theta_i^3 \\ \sum_{i=1}^{55} r_i \theta_i^4 \\ \sum_{i=1}^{55} r_i \theta_i^5 \end{bmatrix}$$

Since these matrices are finally format-wise suitable for solving simultaneous equations, I calculated the constants and placed them in the matrices below, rounded to 4 significant figures for display:

$$\begin{bmatrix} 55 & 777.5 & 1.479 \times 10^4 & 3.166 \times 10^5 & 7.226 \times 10^6 & 1.718 \times 10^8 \\ 777.5 & 1.479 \times 10^4 & 3.166 \times 10^5 & 7.226 \times 10^6 & 1.718 \times 10^8 & 4.202 \times 10^9 \\ 1.479 \times 10^4 & 3.166 \times 10^5 & 7.226 \times 10^6 & 1.718 \times 10^8 & 4.202 \times 10^9 & 1.049 \times 10^{11} \\ 3.166 \times 10^5 & 7.226 \times 10^6 & 1.718 \times 10^8 & 4.202 \times 10^9 & 1.049 \times 10^{11} & 2.660 \times 10^{12} \\ 7.226 \times 10^6 & 1.718 \times 10^8 & 4.202 \times 10^9 & 1.049 \times 10^{11} & 2.660 \times 10^{12} & 6.830 \times 10^{13} \\ 1.718 \times 10^8 & 4.202 \times 10^9 & 1.049 \times 10^{11} & 2.660 \times 10^{12} & 6.830 \times 10^{13} & 1.771 \times 10^{15} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 114.5 \\ 2415 \\ 5.569 \times 10^4 \\ 1.341 \times 10^6 \\ 3.321 \times 10^7 \\ 8.832 \times 10^8 \end{bmatrix}$$

This 6 x 6 matrix corresponds to 6 simultaneous equations, which have the 6 coefficients to be solved for. I placed these matrices in a powerful matrix calculator. The calculated coefficients are:

$$a_0 \approx -0.08308; a_1 \approx 0.3679; a_2 \approx -0.08028; a_3 \approx 0.007750; a_4 \approx -0.0003040; a_5 \approx 4.446 \times 10^{-6}$$

Hence, substituting the calculated coefficients into the general formula, the modelled quintic spiral is:

$$f(\theta) = -0.08308 + 0.3679\theta - 0.08028\theta^2 + 0.007750\theta^3 - 0.0003040\theta^4 + 4.446 \times 10^{-6}\theta^5$$

But what are the equations for the other spirals at lower orders? Using similar steps to calculate the coefficients of polynomials for degrees one to 5, I obtain the results as outlined in **table 1**.

Degree	Polynomial $f(\theta)$
1 (Arithmetic)	$f(\theta) = 0.2097\theta - 0.8827$
2 (Quadratic)	$f(\theta) = 0.01330\theta^2 - 0.1098\theta + 0.5948$
3 (Cubic)	$f(\theta) = 0.09677 + 0.1118\theta - 0.008477\theta^2 + 0.0004663\theta^3$
4 (Quartic)	$f(\theta) = 0.1804 + 0.04731\theta + 0.002012\theta^2 - 0.0001143\theta^3 + 1.027 \times 10^{-5}\theta^4$
5 (Quintic)	$f(\theta) = -0.08308 + 0.3679\theta - 0.08028\theta^2 + 0.007750\theta^3 - 0.0003040\theta^4 + 4.446 \times 10^{-6}\theta^5$

Table 1 Equations for Modelled Spirals, coefficients rounded to 4 s.f. (significant figures)

Obtaining the logarithmic model appears hard, but after doing some research I realized calculating the coefficients for a logarithmic spiral is actually similar to calculating the coefficients of a linear equation. The secret lies beneath the exponential relationship between r and θ ; that is, plotting $\ln(r)$ against θ will yield a straight line. Hence, using the above method, when I treat $\ln(r)$ as $f(\theta)$ with respect to θ , I obtain:

$$\ln[f(\theta)] = 0.1143\theta - 1.312 \text{ (Coefficients rounded to 4 s.f.)}$$

Nevertheless, what does these modelled polynomials really mean? Indeed, they are pretty meaningless without visual representation. Hence, with respect to the statement “the more the merrier”, for greatest contrast, I decided to plot the lowest and highest-order models, that is, the modelled arithmetic and quintic spiral (polar curves) in GeoGebra as outlined by **figures 2 and 3** respectively:

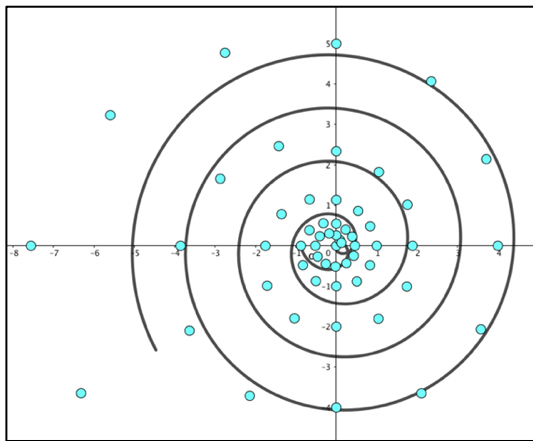


Figure 2 Modelled arithmetic spiral

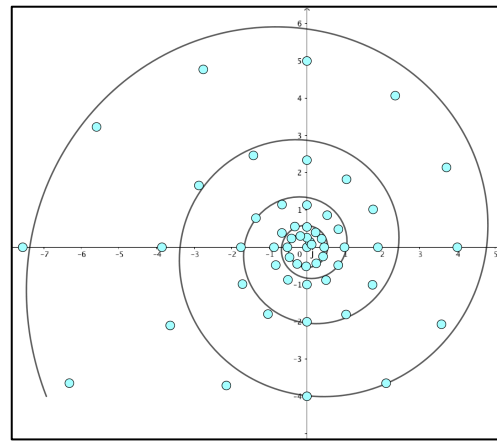


Figure 3 Modelled Quintic Spiral

Indeed, I think that the results of the arithmetic spiral do not look promising, while that of the quintic spiral looks exceptionally promising. However, at this stage, I am judging on the resemblance solely based on the looks of the graph, which is very unreliable. To satisfy my curiosity, I would like to judge the resemblance of the spirals based on their arc lengths (parts 4 and 5) and their f statistic (part 6).

4. Arc length of the actual mollusk

By nature, there are concave edges on a mollusk, but since it is near impossible to trace the edges to obtain the length, the sum of the lengths of line segments between each consecutive point is used for

approximation as outlined by **figure 4**. I admit this is not an optimal solution but I consider this the best solution for this context.

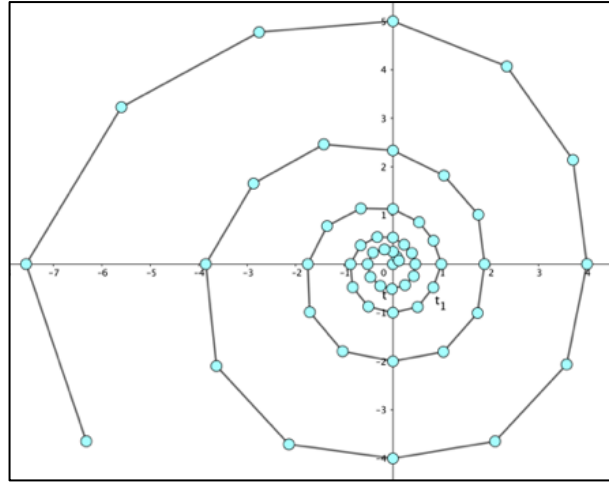


Figure 4 The total arc length is the sum of lengths of line segments between each point on a cartesian plane.

By Pythagoras theorem, the distance between **point $P_i (x_i, y_i)$ and $P_{i-1} (x_{i-1}, y_{i-1})$** on a cartesian plane is:

$$|P_i - P_{i-1}| = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$$

Consequently, the total arc length of the mollusk spiral is:

$$Length = \sum_{i=1}^{54} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} = 59.4742 \text{ units}$$

Coordinates of the plotted points and the values of individual lengths are shown on appendix 3. Using these numbers, I obtain the sum of lengths, which is 59.4742 units. Note that I did not attempt to calculate the real arc length of the mollusk (for example in centimetres), since I will be comparing the modelled spirals which are all based on the same-sized image on the same plane.

5. Arc lengths of the modelled spirals

On the other hand, since I already have the functions of the modelled spirals, I will utilise vector calculus and the first principle to calculate the arc length. Consider a **polar curve $f(\theta)$** with parametrized x and y equations, where **$m(\theta)$ is a function of the x component** and **$m'(\theta)$ is a derivative of $m(\theta)$** , **$n(\theta)$ is a function of the y component** and **$n'(\theta)$ is a derivative of $n(\theta)$** , and finally **h is a small increment in the First Principle**. Note that $\frac{dx}{d\theta}$ and $\frac{dy}{d\theta}$ are alternate forms of $m'(\theta)$ and $n'(\theta)$ respectively. With respect to vector calculus, the **tangential vector of a line segment** from $\overline{f'(\theta)}$ between $\theta + h$ and θ is:

$$\overline{f'(\theta)} = \lim_{h \rightarrow 0} \frac{\vec{f}(\theta + h) - \vec{f}(\theta)}{h} = \lim_{h \rightarrow 0} \frac{\begin{pmatrix} m(\theta+h) \\ n(\theta+h) \end{pmatrix} - \begin{pmatrix} m(\theta) \\ n(\theta) \end{pmatrix}}{h} = \lim_{h \rightarrow 0} \frac{\begin{pmatrix} m(\theta+h) - m(\theta) \\ n(\theta+h) - n(\theta) \end{pmatrix}}{h}$$

$$\text{By definition of First Principle, } \lim_{h \rightarrow 0} \frac{\begin{pmatrix} m(\theta+h) - m(\theta) \\ n(\theta+h) - n(\theta) \end{pmatrix}}{h} = \begin{pmatrix} m'(\theta) \\ n'(\theta) \end{pmatrix} = \begin{pmatrix} \frac{dx}{d\theta} \\ \frac{dy}{d\theta} \end{pmatrix}$$

Now consider the curve $f(\theta)$, on a polar plane with interval $[\theta_2, \theta_1]$. Imagine dividing it into many infinitely small subarcs, estimating the length of each subarc by the length of a straight-line segment like we do with Pythagoras theorem. Bear this in mind. In terms of vector calculus, the arc length of a segment is the infinite sum of the length of these straight-line segments, which is the integral of the magnitude of the tangential vectors $\overline{f'(\theta)}$ of the line segments; in mathematical terms:

$$\text{Arc Length} = \lim_{n \rightarrow \infty} \sum_{i=1}^n |\overline{f'(\theta)}| = \int_{\theta_1}^{\theta_2} |\overline{f'(\theta)}| d\theta = \int_{\theta_1}^{\theta_2} \sqrt{\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2} d\theta \text{ (Equation 4)}$$

Don't you think this will look complicated when I substitute $\frac{dx}{d\theta}$ and $\frac{dy}{d\theta}$? Hence, I will need to simplify this.

To start off, for a polar curve $f(\theta)$ with parametric equations $y = r\sin\theta$ and $x = r\cos\theta$, its derivatives are:

$$\frac{dx}{d\theta} = f'(\theta)\cos\theta - f(\theta)\sin\theta \text{ and } \frac{dy}{d\theta} = f'(\theta)\sin\theta + f(\theta)\cos\theta$$

When I place this back into Equation 4, the arc length of a curve for polar coordinates is therefore:

$$\text{Arc Length} = \int_{\theta_1}^{\theta_2} \sqrt{\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2} d\theta = \int_{\theta_1}^{\theta_2} \sqrt{(f'(\theta)\cos\theta - f(\theta)\sin\theta)^2 + (f'(\theta)\sin\theta + f(\theta)\cos\theta)^2} d\theta$$

I expand and simplify the variables within the square root to yield:

$$\int_{\theta_1}^{\theta_2} \sqrt{[f(\theta)]^2 + [f'(\theta)]^2} d\theta$$

This will yield the arc length of each modelled spiral. For instance, using this method for the arithmetic spiral (polynomial spiral of degree 1), the arc length is:

$$\begin{aligned} \int_0^{28.2743} \sqrt{[f(\theta)]^2 + [f'(\theta)]^2} d\theta &= \int_0^{28.2743} \sqrt{(0.2097\theta - 0.8827)^2 + (0.2097)^2} d\theta \\ &= \left[(0.5\theta - 2.10467) \sqrt{(0.0439741\theta^2 - 0.370204\theta + 0.823133)} - 0.10485 \sinh^{-1}(4.20935 - \theta) \right]_0^{28.2743} \\ &= 62.6860 \text{ units} \end{aligned}$$

Note that the upper limit is $\theta = 28.2743$ since the point is manually plotted on GeoGebra. Indeed, I think that 62.8689 units looks mildly promising compared to the actual mollusk length of 59.4742 units. But are there other modelled spirals that will yield a more promising result?

Performing similar steps on other spirals, I obtained the following as outlined in **table 2**. Note that I did not provide the indefinite integrals as many of them are too complicated to be computed on a wide range of integral calculators, including Wolfram-Alpha. Instead, I used MATLAB to calculate the definite integral (arc length) directly, and the following table shows the results and the percentage error when compared to the arc length of the actual mollusk. Refer to appendix 2 for the codes I used in MATLAB.

Spiral	Total Arc Length	Percentage Error
Mollusk (Actual)	59.4742	-
Arithmetic	62.6860	5.400%
Quadratic	73.6045	23.7587%
Cubic	58.5034	1.6323%
Quartic	58.4401	1.7387%
Quintic	58.7568	1.2062%
Logarithmic	57.7785	2.8512%

Table 2 Total arc length and percentage error (compared with arc length of actual mollusk) for all modelled spirals.

At the moment, it appears that all modelled spirals, with the exception of the quadratic spiral, has a percentage error of less than 10% when compared with the total arc length of the actual mollusk. With this in mind, I declare all these spirals a good fit for the mollusk.

However, I think that the percentage error of 23.7587% for the quadratic spiral appears very alarming, though I believe I should not jump to conclusions. But based on percentage error of total arc length alone, since quintic spiral has the least percentage error, I declare it best resembles the mollusk spiral.

6. Using the F-Test of overall significance to determine the modelled spiral with highest resemblance

After modelling the spirals and calculating percentage error of total arc length of the modelled curves compared with that of the mollusk, I realized the method is flawed since the quadratic spiral appears to have a drastically larger arc length value, with reasons that will be explained in the evaluation.

Hence, I will now employ a statistical test called F-test of overall significance (later simply referred as F-test), also known as analysis of variance (ANOVA), to find which model is the most statistically significant for the data set; or put it simple, which spiral best resembles the mollusk. I use the F-test for two reasons: firstly, it is commonly used when models have been generated using the least squares approach (Lomax, 2007); secondly, unlike the t-test which can only assess one regression coefficient, the F-test can assess multiple coefficients at the same time (Minitab Editor, 2015). At the end of the test, using all the F-statistics and p-values, I will perform hypothesis testing for each modelled spiral which can allow me to compare the modelled spirals in a more objective manner (compared with the method in part 4) and validate my claims.

However, to my disadvantage, the F-test cannot be used to validate the significance of regression lines of order one, which includes the arithmetic and logarithmic spiral (Remember: obtaining the equation for the logarithmic spiral is identical to that of the arithmetic spiral, other than the fact that I treat $\ln(r)$ as

$f(\theta)$ when I model the data). Sadly, I will have to ignore F-test for my modelled arithmetic and logarithmic spiral, but on the bright side, arc lengths can still be compared. Nevertheless, here I begin.

Before we go any further however, I believe it is crucial to understand the underlying significance of the F-Statistic. Indeed, a definition for F statistic is “a ratio of two quantities that are expected to be roughly equal under the null hypothesis, which produces an F-statistic of approximately 1” (Stone, 2016).

Doesn't this sound a bit too vague? To be specific, I found that the F statistic f^* is:

$$f^* = \frac{\text{Variations between sample means}}{\text{Variation within samples}} \text{ (Frost, 2017)}$$

Since the variation between sample means is **sum of squares for regression model (SSM)** divided by the **degree of freedom for sum of squares for model $df(M)$** , while variance within samples is **sum of squares for error (SSE)** divided by **degree of freedom for sum of squares for error $df(E)$** (Frost, 2017), then:

$$f^* = \frac{\frac{SSM}{df(M)}}{\frac{SSE}{df(E)}}$$

Where

$$df(M) = \text{order of polynomial} - 1$$

$$df(E) = \text{number of data points} - \text{order of polynomial} \\ = 55 - \text{order of polynomial}$$

Since the variation between sample means is the variance of **output (modelled radius $f(\theta)$) parameters** from the **mean data point \bar{r}** , the SSM is hence:

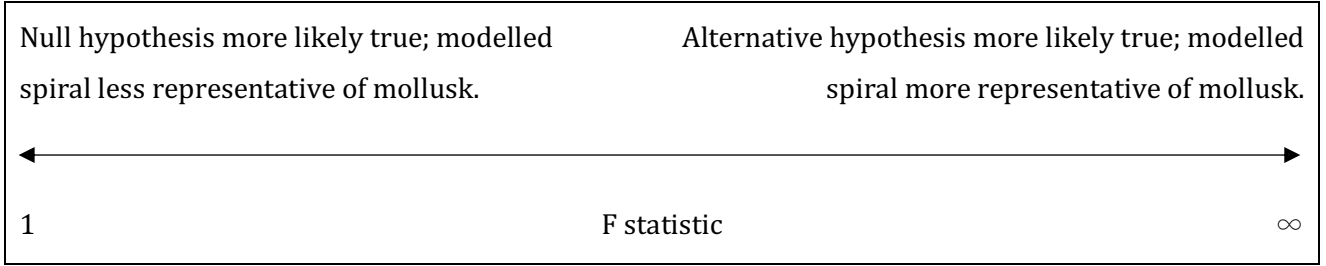
$$SSM = \sum_{i=1}^{55} (f(\theta) - \bar{r})^2$$

On the other hand, since variance within samples is the variance of **individual parameters (the actual radius r_i on mollusk)** from the **output (modelled radius $f(\theta)$) parameter**, the SSE is hence:

$$SSE = \sum_{i=1}^{55} (r_i - f(\theta))^2$$

With respect to Stone's theory, since the 2 variances in the F-statistic ratio have to be approximately equal to produce an F-statistic of 1, I will have to assume that the null hypothesis is true for calculations, and calculate a p-value that will provide evidence for me to prove or disprove the null hypothesis.

Indeed, intuitively, since I want the error to be small, I want SSE to be as small as possible as well. Hence, the larger the F-statistic, the better the model resembles the mollusk; a larger F statistic will mean that the alternative hypothesis will more likely be accepted, or in other words, the modelled spiral is more likely to be significant (Dallal, 2000).



Now I will perform hypothesis testing, which will allow me to verify my hypothesis, that is, the resemblance of my 6 modelled spirals has not occurred by chance. In hypothesis testing, a confidence level of 95% or rejection level of 5% is used since it is utilized most academic papers, so a rejection level of 5% or 0.05 will be compared with the p-value, that is: the probability of rejecting a true null hypothesis. With these premises in mind, I will now set up the hypotheses for my F-test, both in symbols and in words. The following is an example for the modelled quintic function (refer to table 1), where $a_0, a_1, a_2, \dots, a_5$ are coefficients of this model.

Null Hypothesis $H_0: a_1 = a_2 = \dots = a_5 = 0 ; \therefore H_0: p - \text{value} > 0.05$

H_0 : All coefficients of θ are 0, creating a horizontal line, that is, a linear model; hence linear model works the best and modelled spiral bears no statistical significance at all. Therefore, the p-value, which is the probability that H_1 has been accepted by chance, is greater than 0.05.

Alternative Hypothesis $H_1: \text{at least one } a_i \neq 0, 2 \leq i \leq 5 ; \therefore H_1: p - \text{value} < 0.05$

H_1 : Modelled spiral is at least statistically significant for one for its coefficients, hence the linear model is significantly reduced compared to this regression model; modelled spiral is therefore statistically significant for the entirety of whole data set.) Therefore, the p-value, which is the probability that H_1 has been accepted by chance, must be less than 0.05.

Note that I will calculate the p-value for hypothesis testing using an online p-value calculator for f-distribution (Fireman, 2016).

For instance, for the quintic spiral $f(\theta) = -0.08308 + 0.3679\theta - 0.08028\theta^2 + 0.007750\theta^3 - 0.0003040\theta^4 + 4.446 \times 10^{-6}\theta^5$, I found that $df(M) = 4, df(E) = 55 - 5 = 50, \bar{r} = 2.0815, SSM = 196.358796, SSE = 4.6002878$. Hence, the F statistic f^* for quintic spiral, to 4 s.f., is:

$$f^* = \frac{\frac{SSM}{df(M)}}{\frac{SSE}{df(E)}} = \frac{\frac{196.358796}{4}}{\frac{4.6002878}{50}} = 533.6$$

Wow, the F-statistic is certainly enormous! Nevertheless, what does this F-statistic mean? Indeed, the magnitude of the F-statistic itself is still pretty unconvincing since it does not validate my hypotheses. Therefore, I will calculate the p-value for this one-tailed F-distribution. This is because I only want to reject the null hypothesis when the F-statistic is unusually large or in the upper tail (which is when it is in the right segment) of the diagram as outlined by **figure 5**.

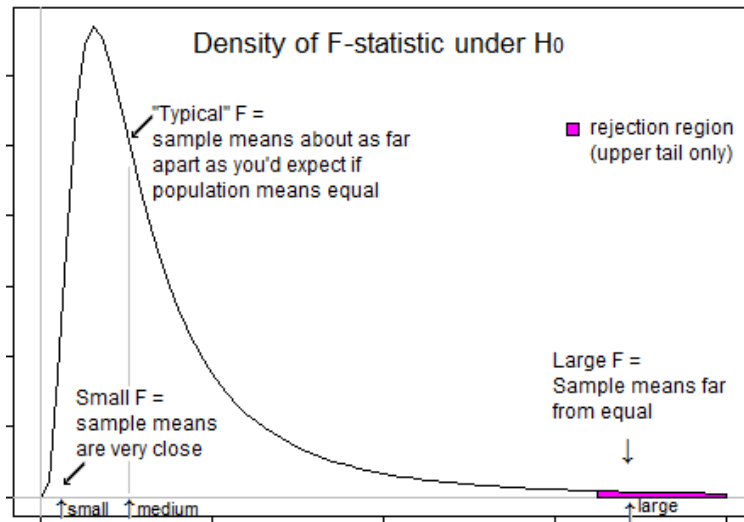


Figure 5 Probability density distribution of an F distribution (Glen_b, 2015)

The p-value for a test with a 95% confidence level or 5% rejection level for the quintic spiral is hence

$$p - value = p(F < f^*) = 2.496 \times 10^{-40}$$

Since the p-value is less than the rejection level of 0.05, there is sufficient evidence at a 95% confidence level to conclude that the alternative hypothesis H_1 should be accepted and that the regression model is significant. Repeating these steps for other modelled spirals, I obtain the results as outlined by table 3:

Modelled Spiral	F-statistic (4 s.f.)	p-value (4 s.f.)	Decision Rule
Mollusk (Actual)	-	-	-
Arithmetic	-	-	-
Quadratic	466.7	6.106×10^{-28}	Alternative hypothesis H_1 should be accepted and regression model is significant for the entirety of data set.
Cubic	969.8	6.872×10^{-42}	
Quartic	642.1	1.776×10^{-40}	
Quintic	533.6	2.496×10^{-40}	
Logarithmic	-	-	-

Table 3 Results for F-test, including the F-statistic, p-value and decision rule.

7. Conclusion

The following table, table 4, sums up some of the key findings of this study:

Modelled Spiral	Total Arc Length	Percentage Error of Total arc length	F-statistic	p-value	Decision Rule
Mollusk (Actual)	59.4742	-	-	-	-
Arithmetic	62.6860	5.4000%	-	-	-
Quadratic	73.6045	23.7587%	466.7	6.106×10^{-28}	Alternative hypothesis H_1 should be accepted
Cubic	58.5034	1.6323%	969.8	6.872×10^{-42}	
Quartic	58.4401	1.7387%	642.1	1.776×10^{-40}	
Quintic	58.7568	1.2062%	533.6	2.496×10^{-40}	
Logarithmic	57.7785	2.8512%	-	-	-

Table 4 Summary of results for study.

As mentioned in the introduction, I will decide which modelled spiral best resembles the mollusk based on their F-statistic value and their percentage error in total arc length. Based on percentage error of total arc length, I will eliminate the arithmetic, quadratic and logarithmic spiral, since their total percentage errors are significantly larger than that of the remaining 3, which are between 1 and 2%. While all modelled spirals are statistically significant, to compare between the three remaining spirals, I will utilize their F-statistic. Since I have stated that “the larger the F-statistic, the more representable the modelled spiral is of the mollusk”, with this in mind, I hereby declare the modelled cubic spiral the best spiral that resembles the mollusk.

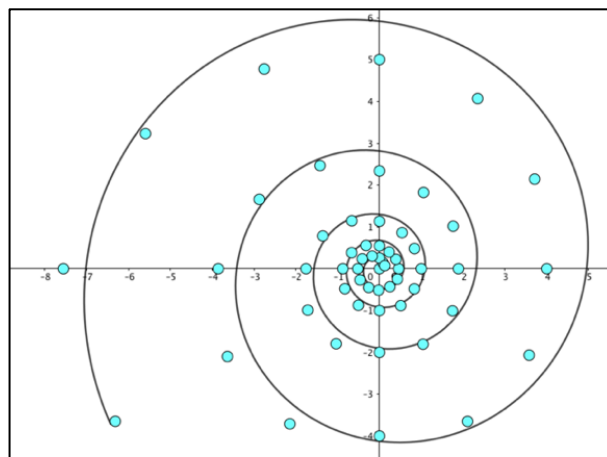


Figure 6 Cubic spiral, the modelled spiral that best resembles the mollusk.

Figure 6 shows what I get when I plot the modelled cubic spiral $f(\theta) = 0.09677 + 0.1118\theta - 0.008477\theta^2 + 0.0004663\theta^3$ a cartesian plane. Indeed, not only does the mathematics determine the resemblance of the cubic spiral to the mollusk, the visual representation of the modelled cubic spiral on the polar plot looks really promising as well!

I would also like to answer the enquiries I have proposed in the introduction. I wondered (1) whether the mollusk resembles the Fibonacci spiral, a special kind of logarithmic spiral, and (2) questioned the validity of the claim “the more the merrier”.

My answer to my first enquiry is no (or that the statement is correct only to a small extent), since the cubic spiral is the modelled spiral that best resembles the mollusk. I think this may be attributed to the scientific fact that, contrary to the belief that many objects in nature resemble the Fibonacci sequence or its “golden ratio” growth factor, spirals are not “ideal” and many are actually slight variations of approximate logarithmic spirals. In addition, although nature may resemble some of the “divine proportions” that humans have observed, every single living thing is unique. Indeed, nature, a subset of these living things, lacks the consciousness or intellect to know what is ideal for itself; for instance, another mollusk that I may pick up on the beach may yield completely different results. Hence, from a philosophical perspective, it is irrational to quantify these patterns. (Oktar, 2005).

My answer to my second enquiry is the claim is false, at least in the context of my internal assessment. While using a higher-degree polynomial will yield a curve that can almost resemble anything, the act of doing so is questionable as well. Essentially, overfitting is “the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably” (OxfordDictionaries.com, 2018) In the context of this piece, overfitting simply makes the model way too mutually exclusive for this mollusk.

8. Evaluation

Since I have answered my own enquiries and reached a conclusion regarding the modelled spiral that best resembles the mollusk, I hereby declare this a successful investigation. I am also proud to declare that I have challenged myself to venture into some statistics and calculus beyond my course. That said, there are certainly many limitations in the methodology to be improved upon.

Firstly, the percentage error in total arc length method may be unreliable since it does not account for the individual differences in arc lengths between the respective points on the actual mollusk and the modelled spiral. Performing percentage error for arc lengths between each points and averaging the value will better reflect its resemblance, though it would be tedious and time-consuming.

Secondly, the sample size, which is the number of points plotted on the seashell, may be insufficient, or in other words, the plots are spaced too far apart. With respect to the “law of large numbers” which states that performing an experiment a large number of times will result in an average close to the expected value, I believe collecting more samples (or to be specific, placing points at smaller angular intervals) will make the models more accurate.

Thirdly, when I collected data for the plots on the seashell, the points were plotted on a graphical user interface on GeoGebra, which may have resulted in great inaccuracy. Although I have already plotted

lines to align the dots with my desired angle, fixing the θ parameter of the polar coordinates may have reduced the imprecision of plotting the points.

Fourthly, the F-test does not account for the arithmetic and logarithmic spiral. While it does not severely affect this investigation as I have eliminated both spirals based on their larger percentage error, I still cannot ignore the fact that I did not fairly compare the arithmetic and logarithmic spiral with other modelled spirals, thus possibly affecting the trustworthiness of my findings.

Last but certainly not least, to a certain extent, the F-test of overall significance is not entirely significant without considering the p-values for individual coefficients of the regression models. For example, an overall significant F-test could determine that the coefficients are jointly not all equal to zero, while tests for individual coefficients could determine that all of them are individually equal to zero (Minitab Editor, 2015). One way to solve this would be to perform F-test for every individual coefficient, though this would be tedious and time-consuming as well.

9. Works Cited

- "Arc Length of Polar Curves." *Arc Length of Polar Curve*, The University of Texas at Austin, Department of Mathematics, www.ma.utexas.edu/users/m408s/m408d/CurrentWeb/LM10-4-4.php.
- "Arc-Length." *Kathleen Ollerenshaw / The University of Manchester / School of Mathematics*, The University of Manchester, School of Mathematics, 30 Jan. 2013, www.maths.manchester.ac.uk/~gajjar/1e2/ie2_tsu9.html.
- Bartzen, Jeanice. "Brown Ribbed Spiral Seashell." DeviantArt, jeanicebartzen27, 17 Jan. 2015, www.deviantart.com/jeanicebartzen27/art/Brown-Ribbed-Spiral-Seashell-507798751.
- Daccache, Rosie. "The Relationship between Nature and Mathematics That Will Surprise You - Rosie Daccache." *GreenArea.me*, Green Area, 19 Feb. 2018, greenarea.me/en/180878/relationship-nature-mathematics-will-surprise/.
- Dallal, Gerard E. "How to Read the Output From Multiple Linear Regression Analyses". *The Little Handbook of Statistical Practice*, 23 May 2012, <http://www.jerrydallal.com/lhsp/regout.htm>.
- Efunda Inc. "Least-Squares Polynomials." Engineering Fundamentals, EFunda, www.efunda.com/math/least-squares/lstsqrmdcurve.cfm.
- Fireman, Ellen. "P-Value Calculator for F Distribution." P Value Calculator, F Distribution, University of Illinois at Urbana-Champaign Department of Statistics, 2016, courses.atlas.illinois.edu/spring2016/STAT/STAT200/pf.html.
- "Getting Angles along with the Length of Lines at Points on a Spiral Using Matlab / Octave." *Stack Overflow*, 2015, stackoverflow.com/questions/27253661/getting-angles-along-with-the-length-of-lines-at-points-on-a-spiral-using-matlab.
- Glen_b. "Why Do We Use a One-Tailed Test F-Test in Analysis of Variance (ANOVA)?" Cross Validated, StackExchange, Stack Exchange Inc, 28 Oct. 2013, stats.stackexchange.com/questions/67543/why-do-we-use-a-one-tailed-test-f-test-in-analysis-of-variance-anova.
- Frost, Jim. "How F-Tests Work in Analysis of Variance (ANOVA)." Statistics By Jim, Jim Frost, 8 Oct. 2018, statisticsbyjim.com/anova/f-tests-anova/.
- Jost, Steve D. "The F-Test for Linear Regression". *Steve Jost -- CSC 423/324 -- Sections 501, 510 - Loop*, Depaul University College of Computing and Digital Media, <http://facweb.cs.depaul.edu/sjost/csc423/documents/f-test-reg.htm>.
- Kaw, Autar. "High Order Interpolation Is a Bad Idea?" The Numerical Methods Guy, Wordpress.com, 14 June 2008, autarkaw.org/2008/06/14/higher-order-interpolation-is-a-bad-idea/.
- Lomax, Richard G., and Debbie L. Hahs-Vaughn. *Statistical Concepts: a Second Course*. Routledge, 2013.
- "Matrix Calculator." *Matrix Calculator*, matrixcalc.org/en/.

“Mean Value Theorem.” *Khan Academy*, Khan Academy, www.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-1/v/mean-value-theorem-1.

O'Callaghan, Jonathan. “Universe Has a 'Golden Ratio' That Keeps Everything in Order, Researchers Claim.” Daily Mail Online, Associated Newspapers, 5 Dec. 2014, www.dailymail.co.uk/sciencetech/article-2862243/Is-space-time-shaped-like-SPIRAL-Universe-golden-ratio-keeps-order-researchers-claim.html.

Oktar, Adnan. “The Measure of Beauty Created in Nature The Golden Ratio.” Harun Yahya, Harun Yahya, 30 June 2005, www.harun-yahya.net/en/Articles/3343/the-measure-of-beauty-created.

Overfitting.” Oxford Living Dictionaries, Oxford University Press, 2018, en.oxforddictionaries.com/definition/overfitting.

“P-Value Calculator for F Distribution.” P Value Calculator, F Dist., University of Illinois at Urbana-Champaign, 2016, courses.atlas.illinois.edu/spring2016/STAT/STAT200/pf.html.

“Spirals.” A Book of Curves, by E. H. Lockwood, Cambridge University Press, 1967, pp. 173–175.

Stone et al. “Understanding Analysis of Variance (ANOVA) and the F-Test”. *The Minitab Blog*, Minitab Inc, 18 May 2016, <http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-analysis-of-variance-anova-and-the-f-test>.

Appendix 1: Using calculus to differentiate function 3

In the following, I simply treated the other variables than the variable I am differentiating as constants. For instance, if I differentiate the following function with respect to a_0 , all other coefficients will become the constant c_i ; r_i will be untouched since it is already a constant.

$$\begin{aligned} & \frac{\partial}{\partial a_0} \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)]^2 \\ &= \frac{\partial}{\partial a_0} \sum_{i=1}^{55} [r_i - c_i - a_0]^2 \\ &= -2 \sum_{i=1}^{55} [r_i - c_i - a_0] = -2 \sum_{i=1}^{55} [r_i - (a_0 + a_1 \theta_i^1 + a_2 \theta_i^2 + a_3 \theta_i^3 + a_4 \theta_i^4 + a_5 \theta_i^5)] \end{aligned}$$

Appendix 2: MATLAB Codes for obtaining arc lengths

Note: The arc length of the arithmetic spiral is not evaluated on MATLAB.

Quadratic Spiral

```
syms x
% r= 0.01330(x^2)-0.1098*x+0.5948
% r'= 0.0266*x - 0.1098
f= ((0.01330*(x^2)-0.1098*x+0.5948)^2+(0.0266*x - 0.1098 )^2 )^0.5;
int(f,x);
vpaintegral(f,x,[ 0 28.2743 ])
```

Cubic Spiral

```
syms x
% r= 0.09677+0.1118*x-0.008477*(x^2)+0.0004663*(x^3)
% r'=0.0013989*(x^2)- 0.016954*x + 0.1118
f= ((0.09677+0.1118*x-0.008477*(x^2)+0.0004663*(x^3))^2+(0.0013989*(x^2)-
0.016954*x + 0.1118)^2 )^0.5;
int(f,x);
vpaintegral(f,x,[ 0 28.2743 ])
```

Quartic Spiral

```
syms x
% r= 0.1804+0.04731*x+0.002012*(x^2)-0.0001143*(x^3)+1.027*(10^-5)*(x^4)
% r'= 0.00004108*(x^3)- 0.0003429 *(x^2) + 0.004024*x + 0.04731
f= ((0.1804+0.04731*x+0.002012*(x^2)-0.0001143*(x^3)+1.027*(10^-
5)*(x^4))^2+(0.00004108*(x^3)- 0.0003429 *(x^2) + 0.004024*x + 0.04731)^2 )^0.5;
int(f,x);
vpaintegral(f,x,[ 0 28.2743 ])
```

Quintic Spiral

```
syms x
```

```
f= (((-0.08308+0.3679*x-0.08028*(x^2)+0.007750*(x^3)-0.0003040*(x^4)+4.446*(10^-6)*(x^5))^2)+((0.3679-0.16056*x+0.02325*(x^2)-1.216*(10^-3)*(x^3)+2.223*(10^-5)*(x^4))^2))^0.5;
int(f,x);
vpaintegral(f,x,[ 0 28.2743 ])
```

Logarithmic Spiral

```
syms x
f= (((exp(0.11438*x - 1.3119)))^2+(0.0307819*exp(0.1143*x))^2)^0.5;
int(f,x);
vpaintegral(f,x,[ 0 28.2743 ])
```

Appendix 3: Raw Data Collection

These are the unrounded coordinates of the points plotted at a regular angular interval on a cartesian plane. Radius is rounded to 4 s.f.

Angle in radians θ	x	y	Radius (r)
0.0000	0.00	0.00	0.0000
0.5236	0.13	0.07	0.1476
1.0472	0.08	0.14	0.1612
1.5708	0.00	0.26	0.2600
2.0944	-0.17	0.30	0.3448
2.6180	-0.41	0.24	0.4751
3.1416	-0.52	0.00	0.5200
3.6652	-0.46	-0.27	0.5334
4.1888	-0.26	-0.45	0.5197
4.7124	-0.02	-0.51	0.5104
5.2360	0.25	-0.43	0.4974
5.7596	0.43	-0.25	0.4974
6.2832	0.46	0.00	0.4600
6.8068	0.39	0.23	0.4528
7.3304	0.23	0.40	0.4614
7.8540	0.00	0.55	0.5500
8.3776	-0.32	0.56	0.6450
8.9012	-0.67	0.38	0.7703
9.4248	-0.88	0.00	0.8800
9.9484	-0.83	-0.48	0.9588
10.4720	-0.50	-0.87	1.003
10.9956	0.00	-1.00	1.000
11.5192	0.51	-0.88	1.017
12.0428	0.83	-0.48	0.959

12.5664	1.00	0.00	1.000
13.0900	0.84	0.48	0.967
13.6136	0.54	0.86	1.015
14.1372	0.00	1.13	1.130
14.6608	-0.66	1.15	1.326
15.1844	-1.36	0.78	1.568
15.7080	-1.75	0.00	1.750
16.2316	-1.71	-0.99	1.976
16.7552	0.00	-2.00	2.000
17.2788	1.04	-1.80	2.079
17.8024	1.75	-1.01	2.021
18.3260	1.89	0.00	1.890
18.8496	1.76	1.02	2.034
19.3732	1.05	1.82	2.101
19.8968	0.00	2.34	2.340
20.4204	-1.42	2.46	2.840
20.9440	-2.87	1.66	3.315
21.4675	-3.86	0.00	3.860
21.9911	-3.64	-2.10	4.202
22.5147	-2.14	-3.71	4.283
23.0383	0.00	-4.00	4.000
23.5619	2.11	-3.65	4.216
24.0855	3.58	-2.07	4.135
24.6091	4.00	0.00	4.000
25.1327	3.71	2.14	4.283
25.6563	2.35	4.07	4.700
26.1799	0.00	5.00	5.000
26.7035	-2.76	4.77	5.511
27.2271	-5.59	3.23	6.456
27.7507	-7.56	0.00	7.560
28.2743	-6.32	-3.65	7.298