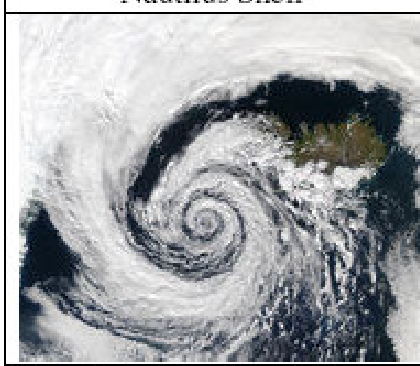
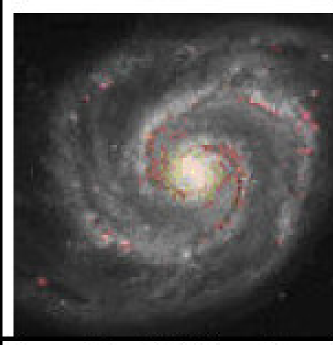
FITTING A LOGARITHMIC SPIRAL TO EMPIRICAL DATA WITH DISPLACED ORIGIN

SK Mishra
Dept. of Economics
NEHU, Shillong
India

Introduction: Nature produces amazingly varied geometrical patterns (Gielis, 2003). In particular, logarithmic spirals are abundantly observed in nature. Gastropods/cephalopods (such as nautilus, cowie, grove snail, thatcher, etc.) in the mollusca phylum have spiral shells, mostly exhibiting logarithmic spirals vividly. Spider webs show a similar pattern. The low-pressure area over Iceland and the Whirlpool Galaxy resemble logarithmic spirals. Many materials develop spiral cracks either due to imposed torsion (twist), as in the spiral fracture of the tibia, or due to geometric constraints, as in the fracture of pipes. Spiral cracks may, however, arise in situations where no obvious twisting is applied; the symmetry is broken spontaneously (Néda et al., 2002). Fonseca (1989) found that rank size pattern of the cities of USA approximately follows a logarithmic spiral.

		
Nautilus Shell*	A Thatcher Shell**	Grove Snail's Shell**
		
Low pressure area over Iceland*	Cowie Shell (cross section)**	The Whirlpool Galaxy*
Courtesy: Wikipedia (*); Xah Special Plane Curves Seashell (**)		

The Mathematical Representation: In the Cartesian coordinate system a logarithmic spiral (variously named as Bernoulli's spiral, Descartes' spiral, equiangular spiral, spira mirabilis, or growth spiral) is described by two parametric equations, viz.

$$x = r\cos(\theta + 360k) = r\cos(\theta)$$
$$y = r\sin(\theta + 360k) = r\sin(\theta) \qquad \ldots (1)$$

where, $0^0 \le \theta < 360^0$; $r = (x^2 + y^2)^{0.5}$; k is a non-negative integer; θ = arctan(y/x) for x $\neq 0$, otherwise $\theta = 90^0$ for (x, y) = (0, > 0) and 270^0 for (x, y) = (0, < 0), while for (x, y) = (0, 0), the angle, θ, is undefined.

In the polar coordinate system a logarithmic spiral is described by the relationship

$$r = \mathbf{a}\exp(\mathbf{b}(\theta + 360k)) \qquad \ldots (2)$$

where, **a** is a positive constant and θ and k are specified as in the relationship (1) above. In view of the relationship (2), the parametric equations of a logarithmic spiral may also be rewritten as

$$x = \mathbf{a}\exp(\mathbf{b}(\theta + 360k))\cos(\theta + 360k) = \mathbf{a}\exp(\mathbf{b}((\theta + 360k))\cos(\theta)$$
$$y = \mathbf{a}\exp(\mathbf{b}(\theta + 360k))\sin(\theta + 360k) = \mathbf{a}\exp(\mathbf{b}((\theta + 360k))\sin(\theta) \qquad \ldots (2a)$$

The sign of **b** in (2) determines whether the spiral is left or right handed. A negative value of **b** makes a spiral go clock-wise as in case of the Whirlpool Galaxy or the low pressure area over Iceland as shown in the figures shown above. On the other hand, a positive value of **b** makes a spiral going anti-clock wise as in the nautilus or cowie shell. When **b** is zero, the spiral degenerates into a circle.

An Empirical Viewpoint : In fitting spiral or conical curves in empirical data some important studies have been made. Among those, Kanatani (1994), Werman and Geyzel (1995), Ho et al. (1996) and Ferris (2000) may be relevant in the present context.

Let there be a set of (empirically obtained) points $Z = (z_1, z_2, \ldots, z_n) : n \ge 10$ (say) and any $z_i = (x_i, y_i)$. Let an inspection of the pattern that these points suggest or a conjecture regarding the law governing the generation of these points indicate that they resemble the trace of a logarithmic spiral. Then there may arise a need to investigate the law generating such a spiral or, to begin with, fit a logarithmic spiral to the empirical data.

The usual procedure of curve-fitting fails miserably in fitting a spiral to empirical data. The author tried with several algorithms available for non-linear regression and non-linear optimization, but unsuccessfully. The main reason for the failure of these algorithms is easily discernible. A spiral is a periodic function for which f (θ) = f (θ + 360k) for any non-negative integer, k. Periodicity also results into multiple values of f (θ) for any given θ. As these algorithms are not designed for tackling such a situation, a good many genuine values of f (θ) are taken for errors by the procedure adopted by them. Failure of the available statistical software packages also in fitting the spiral led the author to develop a new algorithm to fit an Archimedean spiral to the empirical data (Mishra, 2004).

The Shift in Origin: The difficulties in fitting a spiral to data become much more intensified when $z_i = (x_i, y_i)$ are not measured from their origin (0, 0). Once such a shift occurs, the center of the spiral is not known in general. Unless the true (0, 0) or the center of the spiral is known, many mathematical properties of the spiral for fitting it to data cannot be exploited. Ferris (2000) has discussed this problem in some detail.

The Objective: We intend in this paper to devise a method to fit a logarithmic spiral to empirical data measured with a displaced origin. The method would also be tested on numerical data.

The Method: We begin with the recognition of the fact that $z'_i = (x'_i, y'_i)$ are measured from different origin than the center of the spiral, true (0, 0). Let $z_i = (x_i, y_i)$ be the points measured from true (0, 0) such that $z'_i = z_i + c_z$ or $(x'_i, y'_i) = (x_i + c_x, y_i + c_y)$. Here c_x is a constant by which value the measured x' has shifted from the true x and c_y is a constant by which value the measured y' has shifted from the true y. Thus, if we subtract (c_x, c_y) from (x'_i, y'_i), we get the true values, (x_i, y_i), with reference to the center of the spiral $(0,0)$. The values of (c_x, c_y) are unknown and have to be estimated. Once they are obtained, we translate (x'_i, y'_i) into (x_i, y_i). Then, we find out **a** and **b** (the parameters of the spiral) in r = **a** exp(**b**(θ+ 360k)).

Unfortunately, a closed form of such translation and estimation of (c_x, c_y), **a** and **b** is mostly intractable. Further, a small error in estimation of (c_x, c_y) affects **a** and **b** greatly and quite unpredictably.

We choose arbitrary values of (c'_x, c'_y) each within a pre-specified range (based on the inspection of the graphical presentation of the spiral obtained from the data on (x'_i, y'_i)). We define a measure of fit, R^2 = 1- (var(error)/var(r)), where var(error) is the statistical variance of error and var(r) is the statistical variance of r (the radii) given by $r_i = [(x'_i - c'_x)^2 + (y'_i - c'_y)^2]^{0.5}\ ;\forall\ i = 1, 2, ..., n.$

We identify the quadrant of location of each point, $((x'_i - c'_x), (y'_i - c'_y))\ ;\forall\ i = 1, 2, ..., n.$ Depending on the signs of $((x'_i - c'_x), (y'_i - c'_y))$, the quadrant index, q_i, is either 1, 2, 3 or 4, which identifies the location of a point in a particular quadrant. We also define the iso-periodical index, $\kappa_i = k$, of a point $((x'_i - c_x), (y'_i - c_y))$ if $r_i = a\exp(b(\theta_i + 2\pi k))$ for any non-negative integer $k = (0, 1, 2, ...)$ and $\theta_i = \tan^{-1}((y'_i - c_x)/(x'_i - c_y))$.

We arrange r_i (and along with it the associated $((x'_i - c'_x), (y'_i - c'_y))$ and q_i) in an ascending order such that $r_i \le r_{i+1}\ \forall\ i = 1, 2, ..., n-1.$ With an anti-clock movement from the first to the fourth quadrant, the value of q_i increases with the increasing value of r_i. However, with a further increase in the value of r_i, the value of q_i drops down from 4 to 1 which means that we have entered into the first quadrant and so on. From this fact, we identify if the angle, $t = \theta + 2\pi k$ and so on. This process linearizes the relationship between the radius, r, and the angle, t. More explicitly,

$$t_i = \tan^{-1}(y_i' / x_i') + [\text{int}(q_i/2)\pi + 2\pi\kappa_i] \quad \ldots(3)$$

Next, we run a linear regression of $\log_e(r)$ on t to obtain $\log_e(\hat{a})$ and $\hat{b}$ for the model

$$\log_e(r) = \log_e(a) + bt + u \quad \ldots(4)$$

for which we solve the normal equations (5a) and 5(b) simultaneously. Here u is the random disturbance term.

$$\sum_{i=1}^{n} \log_e(r_i) = n\log_e(a) + b\sum_{i=1}^{n} t_i \quad \ldots(5a)$$

$$\sum_{i=1}^{n} \log_e(r_i)t_i = \log_e(a)\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^2 \quad \ldots(5b)$$

Once the values of $\log_e(\hat{a})$ and $\hat{b}$ are obtained, the estimated values of the random disturbances, $\hat{u}_i$, are available from

$$\hat{u}_i = \log_e(r_i) - (\log_e(\hat{a}) + bt_i) \text{ for all } i = 1, 2, ..., n. \quad \ldots(6)$$

Now,

$$R^2 = 1 - \text{var}(\hat{u})/\text{var}(\log_e(r)) \quad \ldots(7)$$

where, $\text{var}(\hat{u}) = (1/n)\sum_{i=1}^{n} u_i^2$ and $\text{var}(\log_e(r)) = (1/n)\sum_{i=1}^{n} (\log_e(r_i))^2 - (1/n)\sum_{i=1}^{n} \log_e(r_i)$

We have to choose (c_x', c_y') such that R^2 is maximized.

Implicit Assumptions: We assume that the points (x', y') are measured without errors and pattern-disturbing approximations. When errors of measurement are present so as to disturb the pattern of the spiral arms, the method may falter. In this line, research is needed so as to incorporate the errors of measurement in (x', y')

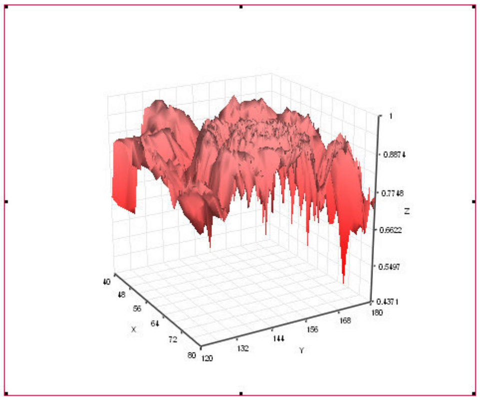
The Issues in Optimization: The surface of R^2 for arbitrarily chosen (c_x', c_y') are highly non-linear, and multi-modal with sharp ridges. For instance, some surfaces for different ranges and different choices of (c_x', c_y') are given in the 3-d graphs below.

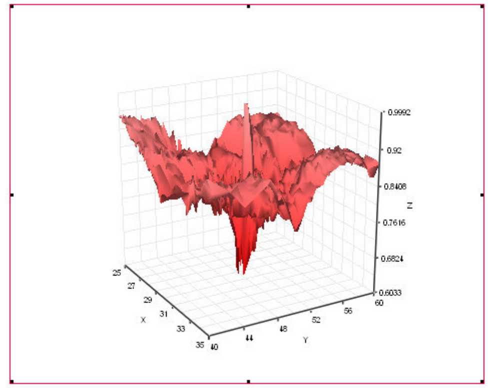
In view of this, we use the Box's algorithm (Box, 1965) for non-linear optimization. Our model is

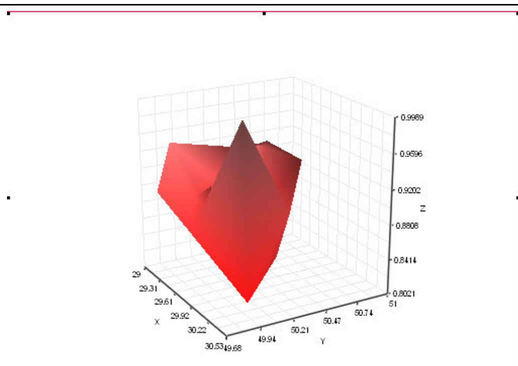
Maximize $R^2 = f(c_x, c_y \,|\, (x_i', y_i'); i = 1, 2, ..., n)$

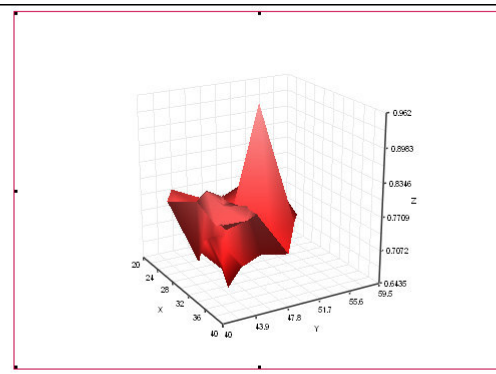
Subject to $\left]\begin{matrix} g_x \le c_x \le h_x \\ g_y \le c_y \le h_y \end{matrix}\right[$ where g is the given lower and h is the given upper limits on c.

We iteratively try with numerous random starting values and by sequential search maximize R^2.









Experimental Findings: We generated 30 angles randomly between 0^0 and 1000^0 (for k=0 to 2). From this we generated 30 points of z=(x_i, y_i) with the parameters a and b, and origins of x and y were shifted by adding c_x and c_y as given in the table below. R^2 was optimized by the 'Box algorithm', through using multiple points and trials with randomly shrinking ranges. The estimated parameters are given below in the table.

Results of experimentation on estimation of the Parameters of a Logarithmic Spiral												
Parameters								Estimated Parameters				
c_x	c_y	g_x	h_x	g_y	h_y	a	b	$\hat{c}_x$	$\hat{c}_y$	$\hat{R}^2$	$\hat{a}$	$\hat{b}$
10	20	0	20	10	35	0.5	0.16	10.000220	19.999900	1.00	0.499947	0.160008
5	7	0	12	2	18	0.7	0.08	5.000019	6.999934	1.00	0.700014	0.079998
4	12	1	10	2	20	1.00	0.08	4.000159	11.999870	1.00	0.999980	0.080001
13	10	1	20	1	20	1.10	0.04	12.999940	10.000074	1.00	1.099997	0.040001
16	6	1	25	2	17	1.20	0.20	16.000433	5.999645	1.00	1.199928	0.200004

Conclusion: It appears that our method is successful in estimating the parameters of a logarithmic spiral. However, it may be noted that the range specification is very important. We have observed that for unduly large range (h-g) the efficiency of the method is reduced. It is suggested that before judging the range, the graphical presentation of the spiral should be observed and location of center guessed. The algorithm allows for a large range, but not for any unduly large range.

We have assumed that the spiral has been shifted into the 1^{st} quadrant (c_x and c_y are positive) and the value of b is positive. In case the value of b is negative (the spiral expands clock-wise), one may use the mirror image of the spiral to convert it into leftwards expanding spiral and then use the algorithm. To shift the spiral from other quadrants to the 1^{st} quadrant, one may use shift parameters (c_x and c_y). The algorithm and the computer program assumes that there are no errors of measurement in x and y. This is a serious limitation, but we have presently not addressed to this issue.

Notes:

- *The author is still perfecting the method and plans to try with other methods of optimization as well. Those interested in perfecting the method may make experiments with it and introduce changes into the method. The FORTRAN program developed by the author may be obtained on request.*
- *The author is thankful to Dr. Juri Persits (an independent researcher of Berlin, Germany) for motivating the former to develop interest in fitting a logarithmic spiral to empirical data.*
- *Contact :* mishrasknehu@hotmail.com

References

Box, M.J.: "A new method of constrained optimization and a comparison with other methods", *Computer Journal*, 8, pp. 42-52, 1965.

Ferris, T.L.J.: "Matching observed spiral form curves to equations of spirals in 2-D images", in *Applications of Electromagnetic Phenomena in Electrical and Mechanical Systems* : The First Japanese-Australian Joint Seminar, 16-17 March 2000, Adelaide, Australia. University of South Australia, 2000.

Fonseca, J.W.: "Urban rank size hierarchy: a mathematical interpretation" *Monograph (ISBN 1877751 16 2),* Institute of Mathematical Geography, Ohio State Univ. Columbus, Ohio. http://www.zanesville.ohiou.edu/geography/urbanrank/index.htm 1989.

Gielis, J.: "A generic geometric transformation that unifies a wide range of natural and abstract shapes", *American Journal of Botany,* 90(3): pp. 333–338. 2003.

Ho, C.T., Chen, L.W.: "A high-speed algorithm for elliptical object detection", *IEEE Transactions on Image Processing*, 5, 3, pp.547-550, 1996.

Kanatani, K.: "Statistical bias of conic fitting and renormalization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 3, pp.320-325, 1994.

Mishra, S.K.: "An algorithm for fitting an Archimedean spiral to empirical data", *Working Paper Series, Social Science Research Network*; available at SSRN: http://ssrn.com/abstract=531542 , 2004.

Mukhopadhyay, U.: "Logarithmic spiral - a splendid curve", *Resonance*, Nov. 2004; pp. 39-45, 2004.

Néda, Z, Leung, K, Józsa, L, and Ravasz, M : "Spiral cracks in drying precipitates" *Physical Review Letters*, 88(9), pp. 095502: 1-4, 2002.

Werman, M., Geyzel, Z.: "Fitting a second degree curve in the presence of error", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, 2, pp.207-211, 1995.
