## Catenary "Best Fit"

## Angela Wade, Miami University

Catenary curves are often seen in architecture and in suspension bridges. In the following article, the process of "fitting" a curve to the St. Louis arch is presented using dynamic geometry software (DGS) andTI-nspire graphing calculators.

## Introduction

The St. Louis Arch is the most wellknown example of a catenary curve that opens downward. However, to the general eye, and to most high school students, the St. Louis Arch looks like a parabola. Yet, catenary curves and parabolic curves are very different (see Fig 1), and the catenary curve is rarely discussed in high school mathematics classes.


Fig 1 Catenary curve, $f(x)$, (solid) and parabolic curve, $g(x)$, (dotted) on the same graph

A catenary curve is a natural curve formed by holding the ends of a uniformly dense chain from equal height (Weisstein, 2008a). They are often seen in architecture and in suspension bridges. How can a student deduce whether or not a curve in a photograph is parabolic or catenary? High school students can address this question by searching for a "best fit" curve for a photograph.

We will explore the process of "fitting"
curves, particularly catenary curves, using dynamic geometry software (DGS) and the $T I-n s p i r e ~ a s ~ o u r ~ t o o l s . ~ F i r s t, ~ w e ~ w i l l ~ e x p l o r e ~$ how students can find a good fit catenary curve with DGS through the 'eye-ball' method. Then we will discuss error calculation with calculator-based regression using a linear example. Next, using a parabolic example, we will investigate the method of using a function's inverse to find a "best fit" curve with the help of the TI-nspire. We will assume the calculator software does find the "best fit" equation. Finally, I will apply all of the above to find a "best fit" catenary curve for points plotted on a photograph of the St. Louis Arch.

## The Catenary Curve

A catenary curve has the equation $y=\frac{a}{2}\left(e^{\frac{x}{a}}+e^{\frac{-x}{a}}\right)=a \cdot \cosh \left(\frac{x}{a}\right)$ where $a$ is a constant and cosh is the hyperbolic cosine function (MathWorld, 2008a). A catenary curve opening up is plotted with DGS with constant $a=1$ can be seen in Fig 2. A catenary curve opening down with constant $a=-1$ can also be seen in Fig 2.


Fig 2 Catenary with $a=1$ (left) and $a=-1$ (right)

## "Fitting the Curve" with Dynamic Geometry Software

In high school, students might look at a photograph and guess whether or not a curve is catenary or parabolic. Looking at a picture of the St. Louis Arch students might guess it is a catenary curve that opens down. To test this conjecture students can copy the picture into a DGS and superimpose points along the Arch (Fig 3). Once these points are in place, the picture can be deleted, or hidden, and students can alter the equation $y=-\left(\frac{a}{2}\right)\left(e^{\frac{x}{a}}+e^{\frac{-x}{a}}\right)+b$ (where $a$ changes the width of the curve and $b$ changes the $y$-value of the curve) until the curve "fits" using an eyeball approach.


Fig 3 Plotted points on the St. Louis Arch
Using DGS, two distance sliders can be constructed, one for the value of $a$ and one for the value of $b$. Students can start with a catenary curve that opens down, with $a \approx 1$ and $\mathrm{b}=0$, that is $y=-\left(\frac{1}{2}\right)\left(e^{\frac{x}{1}}+e^{\frac{-x}{1}}\right)+0$ which is shown in Fig 4. Students then widen the curve by changing the value of $a$ to a larger number using the appropriate slider. Students continue to change the values of $a$ and $b$ until the curve looks like it fits well This approach produces a variety of different "fit" curves; however, the question remains: how can one find the "best fit" curve?


Fig 4 The starting point, $a \approx 1, b=0$


Fig 5 An "eyeball" good fit curve, $a=2.01$, $b=10.98$

## Error Calculations with Calculator-Based Regression

In order to find a "best fit" curve, it is important to understand error calculation and calculator-based regression. We will explore error calculation using linear data to demonstrate how the TI-nspire calculates a regression line. Students typically struggle to find a good fit line when confronted with data such as the ordered pairs (1, 0.9), (2, 2.1), (3.2, 2.9) (3.9, 4.1), (5.1, 5.1), (5.9, 6.2), (7, 6.9), (8.1, 8), (9.2, 8.9). Thus, to find a good fit line we first can record the above points in a TI-nspire Lists and Spreadsheet page by creating lists $x$ and
$y$ in columns A and B respectively (Fig 6). Then, we can create a scatter plot of the points in a Graph and Geometry page.

The TI-nspire can perform a linear regression on the data in the Lists and Spreadsheet page. In this example the TI-nspire "fits" the line $y=.982656 x+.054155$ to the data, which is very close to the line $y=x$, and this is graphed onto the scatter plot in Fig 6. However, what makes this line "fit" and how does the calculator determine that this line is the "best fit"? First, we will look at the concept of residuals, and then we will explore two measures of fit, coefficient of correlation ( $r$ ) and coefficient of determination $\left(r^{2}\right)$ which can be used to judge which, of multiple lines, produce a better fit.



Fig 6 (Top) TI-nspire $x$ and $y$ lists; (Bottom) Plotted data and linear regression line

## Residuals

A residual is "the difference between the true and fitted $y$ values for the points in the sample" (Dielman, 2001), that is the vertical distance between a plotted or true point to the regression or fitted line. The closer the residual is to 0 , the better fit the point. In the previous example, using the point (3.2, 2.9), the $x$-value 3.2 can be plugged into the regression line the TI-nspire cal-
culated to find the predicted (or fitted) $y$ value:

$$
0.982656 \cdot 3.2+0.054155=3.19866
$$

To find the residual the true $y$-value is subtracted from the predicted $y$-value:

$$
2.9-3.19866=-0.298656 .
$$

Thus, the residual for the point $(3.2,2.9)$ is -0.298656 . Fig 7 highlights a graph of residuals for the first three data points using the fit line $y=x$.


Fig 7 A graph of residuals for the first three data points using the fit line $y=x$

## Coefficient of Correlation (r)

Given the data above, we want to find the line that minimizes the residuals, and "least squares is the most common method for fitting a regression line to data" (Moore, McCabe, Duckworth, \& Sclove, 2003). The TI-nspire calculates linear regressions using the least squares method. We will use this commonly accepted method to find the "best fit" equation for the catenary curve in Fig 3.

The least square regression line is the "best fit" "line that makes the sum of the squares of the vertical distances of the data points from the line as small as possible" (Moore et al., 2003). That is the line $y=m x+b$ where $m$ (slope) and $b$ ( $y$-intercept) are chosen such that $\sum_{i=1}^{n}\left(y_{i}-\widehat{y}_{i}\right)^{2}$ is minimized where $y$ is the truei ${ }^{1}$ or actual value of $y$ from the data set, and $\hat{y}$ is the predicted value of $y$ from the fitted line (Moore et al., 2003).


The coefficient of correlation $(r)$ is one value the calculator computes to describe how well a line fits a data sample. Correlation measures the direction and strength of the straight-line (linear) relationship between two quantitative variables. The coefficient of correlation has a value between -1 and 1 , where -1 denotes a perfect negative relationship between variables, and 1 denotes a perfect positive relationship. A perfect fit means that the regression line passes through all true or plotted points (Moore et al., 2003).

## Coefficient of Determination ( $\mathbf{r}^{2}$ )

The second measure of fit is the squared correlation or the coefficient of determination $\left(r^{2}\right)$. We square the coefficient of correlation so as to more heavily weigh the outliers while minimizing the error in points closest to the predicted value of the regression line. The coefficient of determination is calculated through the formula:

$$
r^{2}=\frac{\sum_{i=1}^{n}\left(\widehat{y}_{i}-\bar{y}\right)^{2}}{\sum_{i=1}^{n}\left(y_{i}-\bar{y}\right)^{2}}
$$

where $\hat{y}$ is the value of $y$ from the fitted line, $y$ is the true or actual value of $y$ from the data set, and $\bar{y}$ is the sample mean (Dielman, 2001). A perfect coefficient of determination is when $r^{2}=1$. This makes sense because any number divided by itself is 1 , and thus when $\hat{y}=y$ the numerator and the denominator are the same, producing a value of 1 .

Although the calculator automatically calculates the coefficient of determination every time it performs a linear regression, we can compute our own $r^{2}$ value using a TI-nspire Lists and Spreadsheet page. Returning to the Lists $x$ and $y$ in Fig 6, students can use these lists to calculate the numerator and denominator of the coefficient of determination equation. First, we
will calculate $y$ in column C , let us label this column yhat. Above, the TI-nspire calculated the line $y=.982656 x+.054155$ as the "best fit" line. Since $y$ is the value of $y$ from the fitted line, we will type the equation $=.982656 x+.054155$ into the formula bar of column C. In the equation $x$ refers to the list $x$ in column A (Fig 8). Next we can compute the sample mean in column D, let us label this $y b a r$. We can type the formula $=$ mean $(\mathrm{y})$ into the formula bar where $y$ denotes the list $y$ in column B (Fig 8).

| 1.11 .2 |  | deg auto real |  | - |
| :---: | :---: | :---: | :---: | :---: |
| ${ }^{\text {a }}$ | ${ }^{\text {E }} \mathrm{y}$ | ${ }^{\text {y }}$ hat | U | - |
| - |  | =0.982656 |  |  |
| 1 | 0.9 | 1.03681 |  |  |
| 2 | 2.1 | 2.01947 |  |  |
| 3.2 | 2.9 | 3.19865 |  |  |
| 3.9 | 4.1 | 3.88651 |  |  |
| 5.1 | 5.1 | 5.0657 |  | $\square$ |
| c yhat:=0. | . $982656 \cdot \times[$ | [] $]+0.05415$ |  |  |


| 1.1 1.2 |  | DEG AUTO REAL |  | 目 |
| :---: | :---: | :---: | :---: | :---: |
| ${ }^{\text {a }}$ | $\square^{y}$ | $\square_{\text {yhat }}$ | $\square_{\text {ybar }}$ |  |
| * |  | =0.982065 | =mean(y) |  |
| 1 | 0.9 | 1.03622 | 5.01111 |  |
| 2 | 2.1 | 2.01829 | 5.01111 |  |
| 3.2 | 2.9 | 3.19676 | 5.01111 |  |
| 3.9 | 4.1 | 3.88421 | 5.01111 |  |
| 5.1 | 5.1 | 5.06269 | 5.01111 |  |
| D ${ }^{\text {ybar: }}$-m | mean( y ) |  |  |  |

Fig 8 (Top) The lists $x, y$, and yhat; (Bottom) Finding sample mean in column D

Next we will find the numerator and denominator of the coefficient of determination formula by first subtracting yhat from ybar (Fig 9). We will use the formula $=(c 1-d 1)^{2}$ where $c$ and $d$ refer to columns C and D , and then this formula can be filled down to all nine data points. This is column E in the Lists and Spreadsheets page. We label this column num. Next we subtract $y$ from ybar in column F, labeled den. We use the formula $(b 1-d 1)^{2}$, where $b$ and $d$ refer to columns B and D. Once again the formula can be filled down for each data point (Fig 9).


Fig 9 Column E (num) and column F (den).
Finally, students can create a new Calculator page to sum the numerator and denominator. Students type in the following

$$
\frac{\text { sum }(\text { num })}{\operatorname{sum}(\text { den })}
$$

and the TI-nspire will compute a value of 0.994615 (Fig 10). This $r^{2}$ value is extremely close to the value that the linear regression produced above on lists $x$ and $y$ (Fig 10).

|  | 1.1 |  | EG AUTO REAL | EAL |
| :---: | :---: | :---: | :---: | :---: |
|  | ${ }^{\text {C }}$ y hat | ${ }^{\text {D }}$ ybar | ${ }^{\text {E }}$ num | $\mathrm{F}_{\text {den }}$ |
| - | =0.982065 | = mean(y) |  |  |
| 1 | 1.03622 | 5.01111 | 15.7998 | 16.9012 |
| 2 | 2.01829 | 5.01111 | 8.95701 | 8.47457 |
| 3 | 3.19676 | 5.01111 | 3.29186 | 4.45679 |
| 4 | 3.88421 | 5.01111 | 1.26991 | 0.830123 |
| 5 | 5.06269 | 5.01111 | 0.00266 | 0.007901 V |
| F1 ${ }^{\text {F }}=(b 1-d 1)^{2}$ | $1=(b 1-d 1)^{2}$ |  |  |  |



Fig 10 (Top) $\mathrm{r}^{2}$ computed with Lists and Spreadsheet page and Calculator page; (Bottom) The $r^{2}$ produced by linear regression

## Using the TI-nspire for a "Best Fit" Parabolic Function

Linearizing Data: A parabolic example: "Catenary Regression" does not exist in the TI-nspire. Thus, we will use linear regression and inverse functions to build our own regression for Catenary data. To demonstrate how to use a function's inverse to
find a "best fit" curve, we will first explore a parabolic example.

Given the following points $(1,1),(2,4)$, $(3,9)(4,16),(5,25),(6,36),(7,49),(8$, $64),(9,81)$ students create a list $x$ and a list $y$ in a new TI-nspire Lists and Spreadsheet page. In the same page, the TI-nspire can perform a quadratic regression on the data to produce the curve $y=x^{2}$. Finally the students can plot this data and the regression line in a TI-nspire Graphs and Geometry page (Fig 11). However, suppose the only tool available was a linear regression. Then the data must be linearized to use the linear regression function. This means an inverse function needs to be calculated, that is a function of $x$ must be solved to be a function of $y$. The inverse of a quadratic model is calculated through the following steps:

$$
\begin{aligned}
& y=x^{2} \\
& \sqrt{y}=\sqrt{x^{2}} \\
& x=\sqrt{y}
\end{aligned}
$$

Thus, to linearize the data points, the square root of the values in list $y$ must be calculated. This column is labeled inverse in the TI-nspire Lists and Spreadsheet page (Fig 11). Students can then use the TInspire to perform a linear regression on the new data, using lists X and inverse.

Next, the linearized data can be plotted in a new Graphs and Geometry page using the columns ( x , inverse). Students can then plot the linear regression line on the same graph. The above data produces a linear regression line of $y=x$ (Fig 12). In order to find the quadratic regression curve, the original steps used to find the inverse of $y=x^{2}$ must be reversed onto the linear regression line. Originally, the square root function was applied to the quadratic model to undo the right side of the equation (that is the $x$ side). Thus, to reverse this process, the right side of the equation (the $x$ side) must be squared.

$$
\begin{aligned}
& y=x \\
& y=x^{2}
\end{aligned}
$$



Fig 11 (Top) Quadratic regression on lists $x$ and y ; (Bottom) Lists $\mathrm{x}, \mathrm{y}$, and inverse

As it can be seen, by reverting the inverse process the curve $y=x^{2}$ is produced, which is the same equation as the TI-nspire quadratic regression function calculated. Indeed, the line $y=x$ is not just a good fit, it is a perfect fit because the $r^{2}$ value is 1 .


Fig 12 Linear regression line $y=x$ graphed superimposed on scatter plot of (x,inverse)

## Linearizing Data to Construct a "Best Fit" Catenary Curve

The concept of using inverse functions and error calculations can be applied to find the "best fit" catenary curve. We describe the calculation of the catenary curve's inverse below.

First, remember the catenary curve's equation $\cosh (x)$ is equivalent to $y=\left(\frac{1}{2}\right)\left(e^{x}+e^{-x}\right)$, for $a=1$

Next, we solve for $x$ in terms of $y$.

$$
\begin{aligned}
2 y & =e^{x}+e^{-x} \\
0 & =e^{x}-2 y+e^{-x} \\
e^{x}(0) & =e^{x}\left(e^{x}-2 y+e^{-x}\right) \\
0 & =e^{2 x}-2 y e^{x}+1
\end{aligned}
$$

The above equation is a quadratic in $e^{x}$, thus, using the quadratic equation, we solve for $e^{x}$.

$$
\begin{aligned}
e^{x} & =\frac{-(2 y) \pm \sqrt{(-2 y)^{2}-4(1)(1)}}{2(1)} \\
& =\frac{2 y \pm \sqrt{4 y^{2}-4}}{2} \\
& =\frac{2 y \pm 2 \sqrt{y^{2}-1}}{2} \\
& =y \pm \sqrt{y^{2}-1}
\end{aligned}
$$

Since $e^{x}>0, e^{x}=y+\sqrt{y^{2}-1}$ or $x=\ln \left(y+\sqrt{y^{2}-1}\right)$.

Thus $\cosh ^{-1}(x)=\ln \left(x+\sqrt{x^{2}-1}\right)$ for all real $x>1$ (Weisstein, 2008b; Bridgewater, 2008). Since the natural $\log$ must be positive, it is only possible to find the "best fit" equation for half of a catenary curve. However, a catenary curve is symmetrical and thus this will be sufficient. Also, by the nature of a catenary curve opening down (for example the St. Louis Arch), that is catenary curves with the parent equation $y=-\frac{a}{2}\left(e^{\frac{x}{a}}+e^{\frac{-x}{a}}\right)$, the inverse cannot be calculated. This is because the inverse equation of a catenary curve opening down is $\ln \left(-x+\sqrt{x^{2}-1}\right)$ for all real $x>1$, but this is not valid because a natural log cannot be negative and $x>x^{2}-1$. In short, catenary curves opening down must be converted to catenary curves opening up; therefore, the St. Louis Arch must be rotated 180 degrees. For convenience, we will locate the lowest point of the Arch at $(0,1)$.

With the St. Louis Arch is its new position, students once again plot points along the "upside-down" photograph, and then the photograph can be hidden. I plotted twenty points for this example. We then use DGS to display the coordinates of the
plotted points (See Fig 13).


Fig 13 The St. Louis Arch rotated 180 degrees with plotted points.
Next, students can transfer the ordered pairs into a new TI-nspire Lists and Spreadsheet page. I used the TI-nspire CAS Computer Software for convenience. Students label the $x$-values, xorignal, in column A, and the $y$-values, yoriginal, in column B. Similar to the quadratic example above, the data needs to be linearized in order to perform a regression. We recall from above that the inverse of the catenary equation is $y=\ln \left(x+\sqrt{x^{2}-1}\right)$. Therefore, the $y$-values in column yoriginal need to be transformed. This can be done in the same Spreadsheet page of the TI-nspire, in a column that I labeled yinverse. The formula $=\ln \left(b 1+\sqrt{b 1^{2}}-1\right)$ is typed into the first row of column C, where $b$ represents column B. The students fill down the formula for all twenty rows or points (Fig 14). Next, students perform a linear regression on the lists (xoriginal, yinverse) in the Spreadsheet page. The scatter plots of lists (xoriginal, yoriginal) and (xoriginal, yinverse) are plotted in Fig 14 in a Graphs and Geometry page. In this example, the TI-nspire calculated $y=.61608189680621$ $x+.084672410614995$ as the equation of the "best fit" line (Fig 15).


Fig 14 (Left) Lists xoriginal, yoriginal and yinverse; (Right) The scatter plots (xoriginal, yoriginal) and (xoriginal, yinverse).

Now, we want to find the "best fit" catenary curve for the original points. This process begins by reversing the original steps used to find the inverse of a catenary curve. Thus,


Fig 15 Regression line $y=.61608 x+.08467$ graphed on lists (xoriginal, yinverse).

Now, we want to find the "best fit" catenary curve for the original points. This process begins by reversing the original steps used to find the inverse of a catenary curve. Thus, the regression line must be placed wherever there is an $x$ in the parent catenary equation. Remember the parent equation for a catenary curve is $y=\frac{1}{2}\left(e^{x}+e^{-x}\right)$ for $a=1$. Note, the equation for $a=1$ can be used because the center of the St. Louis Arch is located at $(0,1)$. Plugging in the linear regression equation for $x$ produces the equation $y=\frac{1}{2} e^{0.61608189680621 x+.084672410619995}+$ $\frac{1}{2} e^{-(0.61608189980621 x+0.084672410614995)}$. Letting $\quad f 1(x)=.61608189680621 x+$ .084672410614995 in a new TI-nspire Graphs and Geometry page, students can write the following equation in the formula bar of the page as $y=\frac{1}{2} e^{f(x)}+e^{-(f 1(x))}$ (See Fig 16).


Fig 16 The "best fit" catenary curve.

## A Symmetry Note

It is important to note that this regression formula only fits the positive $x$-values of the St. Louis Arch data due to the nature
of the catenary curve's inverse. The positive half of the equation must be reflected over the $y$-axis to produce the full Arch. The "best fit" equation we just found is a full catenary curve, however, it is not a curve centered at $(0,1)$ anymore. This is because the inverse method of finding a "best fit" curve does not alter the value of a mentioned in the 'eye-ball' approach, but instead the inverse method changes the exponent of $e$. The $y$ intercept of the linear regression line moves the curve slightly away from its center ( 0 , 1) to accommodate the hand plotted order pairs. Therefore, the curve should be used as the "best fit" for only positive $x$ values, and a reflection should be used to find the "best fit" for the negative $x$ values.

## Error Calculations for the "Best Fit" Catenary Curve

The TI-nspire calculates the coefficient of determination $\left(r^{2}\right)$ when performing regressions. In the St. Louis Arch example, the TI-nspire computed an $r^{2}$ value of 0.997795 for the linear regression above. This suggests a very "good fit" (recall an $r^{2}$ value close to 1 is a good fit). There is no formula for the calculation of an $r^{2}$ value for a catenary curve. However, by linearizing the data and applying a linear regression, students can deduce through the use of the $r^{2}$ value of the linear regression that the curve above is a very "good fit."

## A Quadratic Note

Returning to the idea that many students might think a good fit curve is a quadratic curve; we can perform a quadratic regression on the lists xoriginal and yoriginal in the same Lists and Spreadsheets page of the TInspire. This regression produces a coefficient of determination ( $r^{2}$ ) value of 0.996405 . Although this is a good fit, the above catenary curve is a better fit, thus showing a quadratic equation is not the best option for the St. Louis Arch curve.

## Conclusion

The catenary curve is a fascinating curve for students and practiced mathematicians to explore. The TI-nspire does not have catnary regression, but one can find a good fit equation for a photograph by first plotting points along the picture with DGS. Next, using the TI-nspire, one can linearize the ordered pairs using the catenary's inverse function, plot the linear regression line, and then plug the regression line into the catenary equation for $x$. Finally the mathematician can use his or her knowledge of linear error calculations to deduce whether the catenary curve is a good fit. In short, for students, the catenary curve is another equation to learn how to manipulate, and for practiced mathematicians, the catenary curve opens the door to the use of inverse functions, linear regression and technology in order to find a "best fit" equation. $\Omega$

## Works Cited

Bridgewater, D. (2008). The Inverse Wyperbolic Functions. Retrieved November 3, 2008, from Don Bridgewater Resource Site: http://dbridgewater.info/Notes/TranscendentalsInvHyp.htm

Dielman, T. (2001). Applied Regression Analysis for Business and Economics Third Edition. Pacific Grove: Brooks/Cole.

Moore, D., McCabe, G., Duckworth, W., \& Sclove, S. (2003). The Practice of Business Statistics: Using Data for Decisions. New York: W. H. Freeman and Company.

Weisstein, E. (2008a). Catenary. Retrieved November 3, 2008, from Math-World--A Wolfram Web Resource: http:// mathworld.wolfram.com/Catenary.html

Weisstein, E. (2008b). Inverse Hyperbolic Cosine. Retrieved November 3, 2008, from MathWorld--A Wolfram Web Resource: http://mathworld.wolfram.com/InverseHyperbolicCosine.html


Angela Wade is third year student at Miami University. She is majoring in Integrated Mathematics Education, and she hopes to teach Calculus at a high school when she graduates.

## Think About It!

Why Students Shouldn't Listen to Loud Music
"Right after the destructive sound, astrocytes (one type of glial cell) begin to move into the synaptic nest, cupping their cellular hands around synapses tightly as if to shield them from the noise. ... the astrocytes are responding to the overstimulation of these acoustic neurons and sopping up the excitatory neurotransmitter, which in excess kills neurons from overstimulation. At the same time the astrocytes are releasing neurotrophic factors in an effort to prevent the neurons from dying. Failing against the damaging roar that causes these synapses to fire beyond normal limits, the synapses and nerve endings begin to wither and die."

Fields, R. D. (2009). The other brain, 102. Simon \& Schuster, NY.

