

From the Numerical to the Theoretical in Calculus

**Teaching Contemporary Mathematics
NCSSM
February 6-7, 2003
Doug Kuhlmann
Phillips Academy
Andover, MA 01810
dkuhlmann@andover.edu**

How and Why Numerical Integration Should Precede the Fundamental Theorem

The short answer is: "We measure area, we don't define it." (Jerry Uhl)

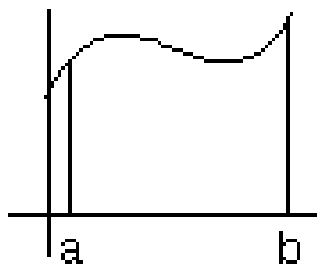
HOW TO INTRODUCE NUMERICAL INTEGRATION FIRST.

Assumptions: Students know the derivatives of the usual elementary functions, including polynomials, trig functions, exponential and logs. In addition they know the power rule for non-integer exponents, the product, quotient and chain rules. They have done a few antiderivative problems, mainly velocity/distance types.

Introducing area under a curve: L_n and R_n .

On the first day of integration I ask the basic question:

What is this area?



How can we measure this area? One approximation leads to the left-hand Riemann sum in the usual fashion:

$$L_n = f(x_0)h + f(x_1)h + \dots + f(x_{n-1})h$$

where $h = \frac{b-a}{n}$. Note that I only consider approximations using rectangles of **equal** width. The first time through calculus is not the time to generalize to things like partitions, norms of partitions, refinements, etc.

We immediately consider the right-hand Riemann sum and discuss why it is not *a priori* possible to choose which is better, the Left Hand Rule or the Right Hand Rule.

At this point the class writes a short TI-83 program, called INTEGRAL or AREA, that calculates this sum with a , b and n as user inputs. This is a good time to teach and/or review For-End loops. (A listing of the final version of the INTEGRAL program is at the end of these notes.) The first integral program using Left and Right hand rules only and assumes that the function is in Y1 is listed here.

```

INTEGRAL          • Program
:ClrHome          :For(J,1,N)
:Input "LOWER LIMIT=",A    :L+Y1*H->L
:Input "UPPER LIMIT=",B    :X+H->X
:Prompt N           :R+Y1*H->R
:(B-A)/N->H        :End
:0->L:0->R         :Disp "L, R",L,R
:A->X

```

Trapezoid and Midpoint Rules

At this point the class usually invents the trapezoid rule themselves:

$$T_n = \frac{1}{2}(L_n + R_n)$$

This leads to a modification of the INTEGRAL program..

I then suggest to the class another way, the **Midpoint** rule, which uses the y-coordinate at the middle of the subinterval for the height of the rectangles. One can compare the Trapezoid and Midpoint rules in the following way:

$$T_n = \frac{f(x_0) + f(x_1)}{2}h + \frac{f(x_1) + f(x_2)}{2}h + \dots + \frac{f(x_{n-1}) + f(x_n)}{2}h$$

and

$$M_n = f\left(\frac{x_0 + x_1}{2}\right)h + f\left(\frac{x_1 + x_2}{2}\right)h + \dots + f\left(\frac{x_{n-1} + x_n}{2}\right)h$$

We now make another, this time more subtle, change in our INTEGRAL program to allow us to calculate M_n .

Simpson's Rule without parabolas.

For L_n and R_n there was no *a priori* reason for choosing one or the other. Each seems to be equally accurate and so when we average them for the trapezoid rule, we weight each of them the same. Can we do a similar thing with the Trapezoid and Midpoint rules?

We can compare the accuracy of the Trapezoid and Midpoint rules by looking at some examples that we already know. By using our INTEGRAL program we see that as n increases, our estimates of the area under $y = \cos(x)$ between $x = 0$ and

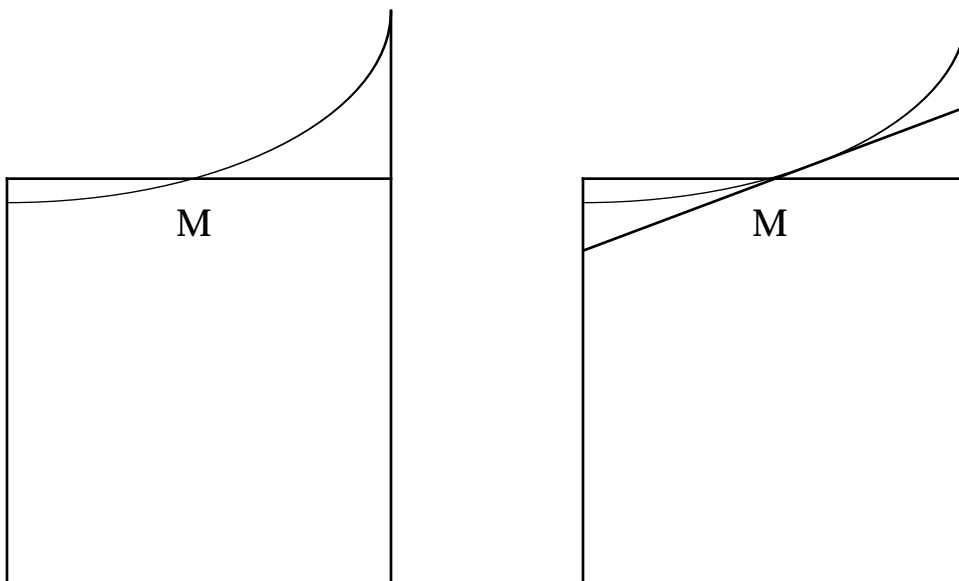
$x = \frac{\pi}{2}$ approach 1. Taking it as a fact that this area is indeed 1, let's compare the errors of the Trapezoid and Midpoint rules, i.e let's examine $T_n - 1$ and $M_n - 1$.

Here are three cases using $n = 5, 10$ and 20 with the resulting errors or deviations from 1 for the trapezoid and midpoint approximations.

N	$T_n - 1$	$M_n - 1$
5	-.008238	.004124
10	-.002057	.001029
20	-.000514	.000257

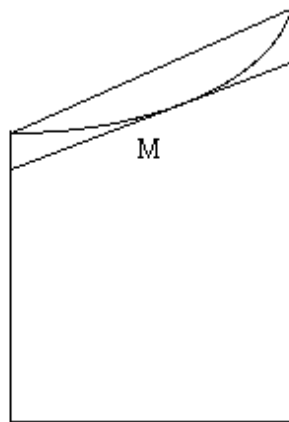
Note that the Mid-point approximation error is about half that of the Trap approximation and is of opposite sign. In this case the Trap rule is too small and the Mid rule is too big but the Mid rule is about twice as close to the actual area.

We can also compare the accuracies of T_n and M_n by the following argument. Consider one rectangle in the midpoint rule:



Keeping the point marked M fixed on the curve, rotate the top of the rectangle, extending it at the same time, until you make it tangent to the curve at M. What is the area under the trapezoid formed with this tangent line as the non-parallel top? With a little consideration, one can show that this area is the same as the area of the original rectangle. This means that the straightforward Midpoint rule, as easy to calculate as the Left- and Right-hand rules, can be thought of as the ***Tangent to the Midpoint Trapezoid Rule***.

Which is more accurate, T_n or M_n ? First answer this question: Which is a better approximation to a function, a tangent line or a secant line. The answer gives support to the argument that M_n is a better approximation. So does the following heuristic argument.



For functions that are concave up, the Midpoint rule is too small and the Trap rule is too big. (Other way round for concave down.) The picture above also suggests that the error for Trap is twice as large in magnitude as the error for Mid and is of opposite sign, a point we discovered numerically earlier.

Since M_n is closer to the true value than T_n by a factor of two and M_n and T_n are on opposite sides of the true value, that suggests we use a weighted average of the two. Consequently we *define* Simpson's rule as:

$$S_{2n} = \frac{2M_n + T_n}{3}$$

WARNING: Note the subscript. If we use n subdivisions for the Trap rule and n subdivisions for the Midpoint rule, we have effectively used $2n$ subdivisions for Simpson's rule this way.

WHY DO NUMERICAL INTEGRATION FIRST?

By now I have also introduced the integral notation $\int_a^b f(x) dx$ to represent the area. By doing several problems approximating this integral, the students come

to know that $\int_a^b f(x) dx$ is a real number, not a class of functions. Additionally, they can verify the obvious $\int_a^c f(x) dx + \int_c^b f(x) dx = \int_a^b f(x) dx$ and $\int_a^a f(x) dx = 0$.

They can even verify and think about why

$$\int_a^b f(x) dx = -\int_b^a f(x) dx \quad (\text{h is negative in one of the sums}).$$

Finally, I have the class perform this exercise:

Let $Y_1 = \cos(X)$ and evaluate the following for various values of b

$$\int_0^b \cos(x) dx$$

A little discussion allows them to see that we have created a new function: For every 'b' there is exactly one number associated with that b, namely the area from 0 to b. I then ask them to each find the value of this integral for $b = 0, .2, .4, .6, .8, 1.0, 1.2, \dots$ as many as there are students, maybe twice as many with every one doing two.

When I first did this exercise we plotted the points on the board:
(0,0) (.2, .199) (.4, .389) (.6, .565) (.8, .717) etc

These questions were asked by students as we plotted them;

1. Why are they going down? (Eventually it does)
2. Why are they negative? (Ditto)
3. Why is it sin?

If students can ask these questions on their own, I was convinced that this method of introducing the integral was valuable.

At this point we can prove the Fundamental Theorem in the usual way, having motivated it with the above and several other examples. However we can also explore antiderivatives by solving differential equations graphically via slope fields and numerically with Euler's method.

Slope Fields and Euler's Method

Suppose we wanted to find solutions to a differential equation of the form

$$y' = g(x, y).$$

Examples of such differential equations include:

$$y' = y \quad (1) \qquad y' = x \quad (2)$$

$$y' = x - y \quad (3) \qquad y' = \cos y \quad (4)$$

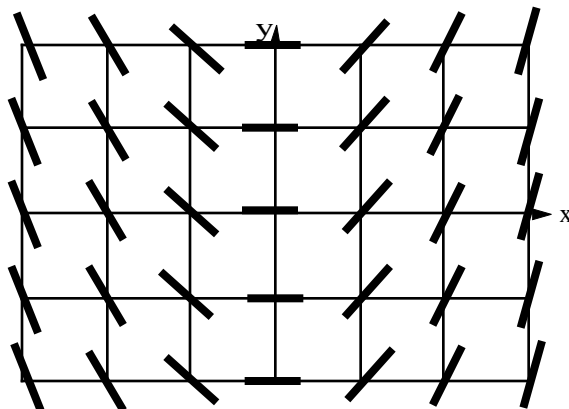
$$y' = \exp(-x^2) = e^{-x^2} \quad (5) \qquad y' = \sqrt{1 + x^3} - x \quad (6)$$

Equations (1) through (4) are solvable by analytic techniques but equations (5) and (6) are not. Nevertheless, we can get some qualitative information about the solutions of all of the above by looking at the slope fields for each.

SLOPE FIELDS

If a first-order differential equation can be put in the form $y' = g(x, y)$, then we can determine the *slope* of the solution $y = f(x)$ through any point (x, y) . Graphically, we can draw a short line of the proper slope through each of several points (x, y) in the plane. The resulting picture of short slope lines is called a **slope field**. I think of a slope field as a series of conditional statements: If a solution passes through the point (x, y) , this is the slope it would have.

Consider a simple case: $f'(x) = x$. On a piece of graph paper, here chosen very small to make it easier, select a grid of points and at each point selected draw a short segment whose slope is x , the first coordinate of the point. This is the resulting picture:



We know the solutions to this differential equation are $f(x) = \frac{x^2}{2} + C$, and if we mentally overlay these on the grid above, we can see how the solutions follow the slope field.

In practice this can be very time consuming or at least tedious, but a computer or programmable calculator can do the task for us. In the following examples, a TI-83 program drew the slope fields. The listing for this program is given at the end.

Before we proceed, let's remind ourselves of the following theorem found in almost any text on elementary differential equations.

Theorem. For a first order differential equation $y' = g(x, y)$ with initial condition $y(x_0) = y_0$, a sufficient (though not necessary) condition that a unique solution $y = f(x)$ exist is that g and $\frac{\partial g}{\partial y}$ be real, finite, single-valued, and continuous over a rectangular region of the plane containing the point (x_0, y_0) .

In simpler terms, if $g(x, y)$ is suitably *nice*, we can safely make the assumption that a unique solution exists.

In all of the cases we shall examine, the above conditions are satisfied. What this means is that given a specific point in our slope field there is exactly one solution that passes through this point.

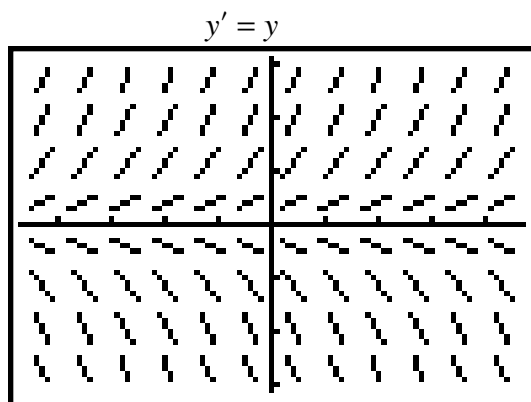
Example 1. $y' = y$

Before we sketch the slope field, notice that if $y' = y$ then $y'' = y' = y$.

Similarly, any order derivative of y is equal to y .

Below is the slope field for this equation on the viewing window

$[-4.7, 4.7] \times [-3.1, 3.1]$ (Unless otherwise stated, all viewing windows will have these dimensions.)



We easily recognize the exponential functions $f(x) = Ae^x$ as solutions and can easily visualize them on top of the slope field.

Example 2. $y' = x$ (Left to reader to expand above slope field.)

Example 3. $y' = x - y$

Note that $y'' = 1 - y' = 1 - (x - y) = 1 - x + y$.

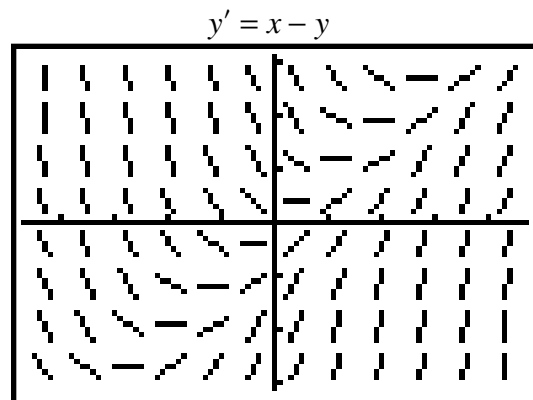
We then conclude that

$$y' > 0 \text{ if } y < x \qquad y'' > 0 \text{ if } y > x - 1$$

$$y' = 0 \text{ if } y = x \qquad y'' = 0 \text{ if } y = x - 1$$

$$y' < 0 \text{ if } y > x \qquad y'' < 0 \text{ if } y < x - 1$$

The slope field for this equation shows that any solution satisfies the above conditions.



Note that anywhere along the line $y = x$, the slope is zero.

It is easy to sketch qualitative solutions to the differential equations by following the slope fields, but here again the computer can help using numerical methods to approximate solutions.

EULER'S METHOD

The first numerical method is Euler's method, a straightforward, simple (and hence prone to inaccuracies at times) method for approximating solutions to $y' = g(x, y)$. It is analogous to using the left hand rule for approximating integrals. If we know that the point (x_0, y_0) is on our curve and we define our new x_1 to be

$$x_1 = x_0 + h$$

where h is the step-size or increment.

Then the new y is given by

$$y_1 = y_0 + g(x_0, y_0)h .$$

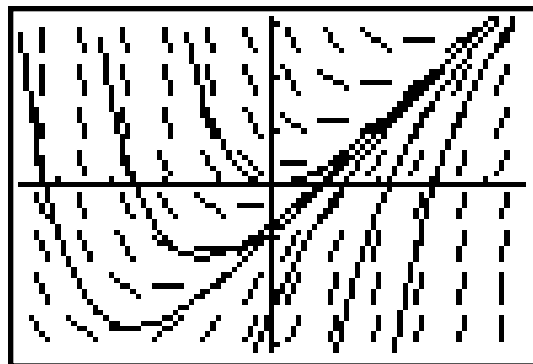
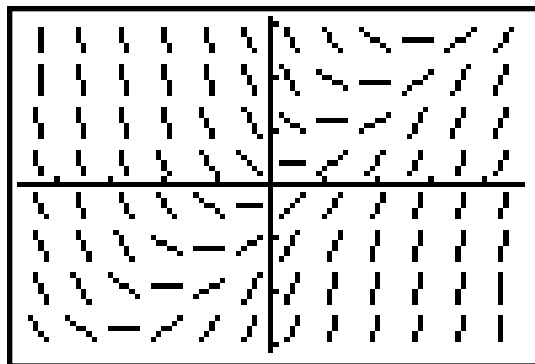
We approximate y by using the tangent line which has slope given by $g(x_0, y_0)$.

The new point is connected to the old point and the process is repeated, generating an approximate solution to the differential equation. That is, we find

$$x_2 = x_1 + h \quad \text{and} \quad y_2 = y_1 + g(x_1, y_1)h .$$

Note that we recalculated the slope at the point (x_1, y_1) . We repeat this as many times as needed.

Below is the original slope field and next to it is a slope field with several solutions sketched on it using Euler's method for $y' = x - y$



One question that immediately arises, is the existence of a possible asymptote to all of the solutions. They all seem to converge to a line. In fact, if we use analytic techniques (using integrating factors) to solve $y' = x - y$ we get the general solution to be

$$y = x - 1 + Ce^{-x}$$

It is now clear that the line $y = x - 1$ is an asymptote for *all* solutions to our equation.

Example 4. $y' = \cos y$

Using analytic techniques, namely separating variables, we get

$$\sec y \, dy = dx$$

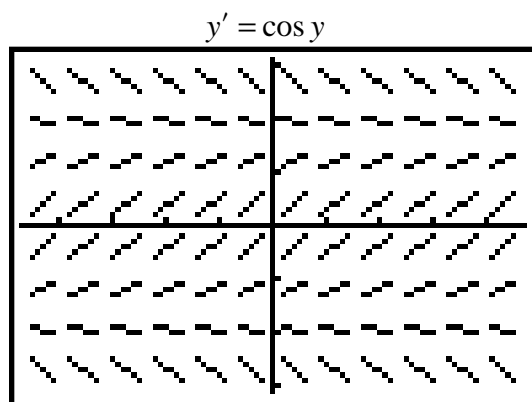
Integrating and remembering the trick for integrating secant, we get

$$\ln|\sec y + \tan y| = x + C$$

But what do these curves look like? Can you sketch one?

Question: What does the solution that goes through the point $(-4, 3)$ look like?

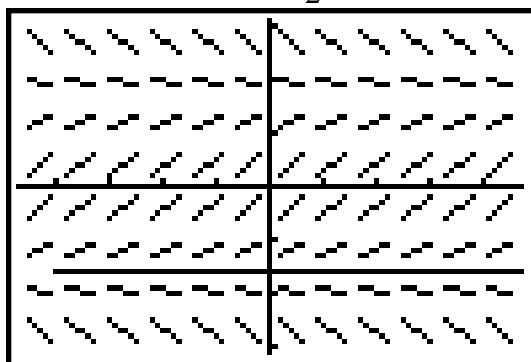
Answer: Huh? Slope fields can help. Below is the slope field for $y' = \cos y$.



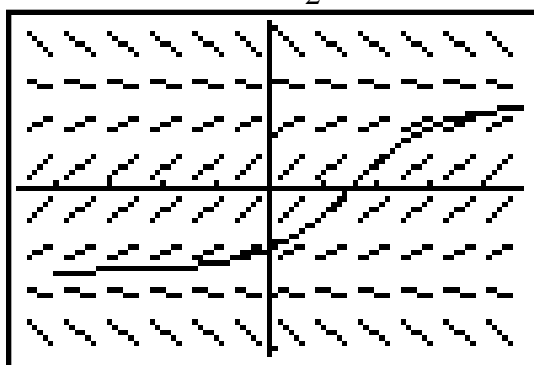
Examining it carefully, we see that the lines $y = \frac{\pi}{2}$ and $y = \frac{-\pi}{2}$ are important. Both are fixed points in the sense that if $y = \frac{\pi}{2}$ or $y = \frac{-\pi}{2}$, then $\cos y = 0$ and so there is no change in y . However, there is a very big difference between the two lines. The line $y = \frac{\pi}{2}$ is an **attractive fixed point**. If you start within a neighborhood of $y = \frac{\pi}{2}$ the solution will converge to the line: It will be "attracted" to the line. On the other hand, the line $y = \frac{-\pi}{2}$ is a **repelling fixed point**. If you start even slightly away from $y = \frac{-\pi}{2}$ the curve moves further away and eventually converges to another fixed point.

Below are Euler solutions with different starting points all close to each other.

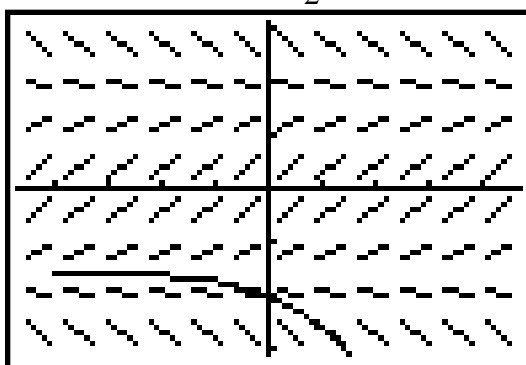
$$x_0 = -4, y_0 = \frac{-\pi}{2}$$



$$x_0 = -4, y_0 = \frac{-\pi}{2} + .01$$



$$x_0 = -4, y_0 = \frac{-\pi}{2} - .01$$



I leave it to the reader to determine what the curve through $(-4, 3)$ looks like.

Euler's Method and the Fundamental Theorem

If the derivative is a function of x alone, i.e. $\frac{dy}{dx} = f(x)$, then Euler's method can be used to illuminate the Fundamental Theorem of Calculus. Suppose we know the derivative and a starting point for a function so that we are given a point $(a, f(a))$ and the derivative $f'(x)$ and we wish to evaluate (estimate) $f(b)$ where for this argument $a < b$ (although it doesn't have to be).

If we define $h = \frac{b-a}{n}$ and then use Euler's with step size h and n steps, we get the following:

$$\begin{array}{ll}
x_0 = a & f(x_0) = f(a) \\
x_1 = x_0 + h & f(x_1) \approx f(x_0) + f'(x_0)h \\
x_2 = x_1 + h & f(x_2) \approx f(x_1) + f'(x_1)h \\
\cdot & \\
\cdot & \\
\cdot & \\
x_{n-1} = x_{n-2} + h & f(x_{n-1}) \approx f(x_{n-2}) + f'(x_{n-2})h \\
x_n = x_{n-1} + h & f(x_n) \approx f(x_{n-1}) + f'(x_{n-1})h
\end{array}$$

Note that $x_n = b$ and $f(x_n) = f(b)$.

If we substitute the next to last equation for $f(x_{n-1})$ into to the last one we get

$$f(x_n) \approx f(x_{n-2}) + f'(x_{n-2})h + f'(x_{n-1})h$$

Substitute the previous equation above and get

$$f(x_n) \approx f(x_{n-3}) + f'(x_{n-3})h + f'(x_{n-2})h + f'(x_{n-1})h. \text{ Repeating this we get}$$

$$f(x_n) \approx f(x_0) + f'(x_0)h + f'(x_1)h + \dots + f'(x_{n-1})h$$

Remembering that $f(x_0) = f(a)$ and $f(x_n) = f(b)$ we see that the last equation just says that

$$f(b) - f(a) \approx f'(x_0)h + f'(x_1)h + \dots + f'(x_{n-1})h$$

and the right hand side of this equation is just the Left-hand Riemann sum approximating the integral of $f'(x)$ and so we are led to

$$f(b) - f(a) = \int_a^b f'(x) dx$$

which is the Fundamental Theorem.

Program listings:

```
INTEGRAL          • Program
:ClrHome
:Input "LOWER LIMIT=",A
:Input "UPPER LIMIT=",B
:Prompt N
:(B-A)/N->H
:0->L:0->M:0->R
:A->X
:For(J,1,N)
: L+Y1*H->L
: X+H/2->X
: M+Y1*H->M
: X+H/2->X
: R+Y1*H->R
: End
: (L+R)/2->T
: (2M+T)/3->S
: Disp "L, R, T,M,S",L,R,T,M,S
```

```
SLOPEFLD        • Program
:Func:Fnoff
:PlotsOff
:ClrDraw
:ClrHome
:Disp "AXES ON (1)
:Disp "OR OFF (0)"
:Input A
:If A=0
:Then:AxesOff
:Else:AxesOn
:End
:Disp "HOW MANY MARKS?"
:Input "ACROSS: ",A
:Input "DOWN: ",D
:(Xmax-Xmin)/A->H
:(Ymax-Ymin)/D->V
:For(I,1,D)
:For(J,1,A)
:Xmin+(J-1)H+H/2->X
:Ymin+(I-1)V+V/2->Y
:Y1->M
:Y-.30MH->S
:Y+.30MH->Z
:X-.30H->P
:X+.30H->Q
:If abs (Z-S)>.6V
:Then
:Y+.30V->Z
:Y-.30V->S
:(Z-Y)/M+X->Q
:(S-Y)/M+X->P
:End
:Line(P,S,Q,Z)
:End:End

EULER          • Program
:Func:Fnoff
:Input "X START=",A
:Input "Y START=",B
:Input "STEP SIZE=",H
:Input "NO. STEPS=",N
:A->X:B->Y
:For(I,1,N)
:X->U:Y->V
:Y+Y/H->Y
:X+H->X
:Line(U,V,X,Y)
:End
```

Runge-Kutta techniques are analogous to Simpson's rule. Below is the code for a program that uses a fourth order Runge-Kutta technique to numerically approximate a solution to a first order differential equation $y' = g(x,y)$.

RK4GRAPH • Program

```
:Func:Fnoff
:Input "X START=" ,A
:Input "Y START=" ,B
:Input "STEP SIZE=" ,H
:Input "NO. STEPS=" ,N
:A->X:B->Y
:For( I, 1 ,N)
:X->U:Y->V
:Y1->K
:X+H/2->X
:V+H*K/2->Y
:Y1->L
:V+H*L/2->Y
:Y1->S
:X+H/2->X
:V+H*S->Y
:Y1->T
:V+(H/6)(K+2L+2S+T)->Y
:Line(U,V,X,Y):End
```

Doug Kuhlmann
Phillips Academy
Andover, MA 01810
dkuhlmann@andover.edu
6 February 2003