

Getting Students Excited About Learning Mathematics

Jen-Mei Chang

Department of Mathematics and Statistics
California State University, Long Beach
jchang9@csulb.edu

Introduction

It wasn't so long ago when mathematics students would diligently transcribe the words of wisdom of their professors without a hint of doubt. The popular phrase "just do it" has become an easy way out when things are not so easy to explain.

In many the Asian countries, mathematics students are not encouraged to ask questions. In fact, questions are considered impolite and a waste of time as they interrupt the class flow and decrease the amount of materials to be covered. Growing up in this culture, I was never curious enough to find out what powerful things mathematics could do in the real world. I never had an intrinsic motivation to learn mathematics, which I believe contributed to my mediocre performance on math tests. Little did I know that a single mathematics course later in life would alter my point of view so drastically and shape my teaching philosophy in such a profound way.

To me, the teaching and learning mathematics go hand-in-hand and always happen simultaneously. A teacher is not a walking library, which passively delivers information upon request; instead, a teacher plays the vital role of encouraging innovative thinking as well as stimulating curiosity. This can be done, for example, through the use of open discussions and hands-on projects.

What I like to do often times in the classrooms is to use various real-world problems to motivate the learning of certain concepts to get students excited about what these ideas could do before diving into the formality. And having a research area that is somewhat industrial-oriented, I have the privilege of bringing in a lot of interesting applications for this purpose.

For example: (Items 2 and 3 are not necessarily for high school.)

- 1) The idea of image sharpening can be used to motivate the learning of matrix multiplications.
- 2) The ideas behind Google's PageRank can be used to provide a practical use of eigenvalues and eigenvectors.
- 3) The ideas of image compression can be used to motivate the learning of matrix factoring such as singular value decomposition.

- 4) And the idea of handwritten digit classification can be used to motivate the study of tangent approximation.

In particular, I will use the rest of this article to provide a short illustration how one may use a real-world problem to motivate the learning of “tangent” approximation. I will do so by uncovering its uses at different grade levels as a way to increase students’ interest in learning the concept.

The take-home message for students as well as teachers is to challenge oneself in finding something that intrigues one in mathematics and see how far one can go with that idea. Hopefully, by doing so, students learn to discover more wonderful ideas in mathematics on their own.

An Illustrative Example

Some students first see the formal concept of tangent (or tangent line) in a high school geometry class where the tangent line (or simply the tangent) to a curve is the straight line that intersects the curve at one and only one point, (See Figure 1.)

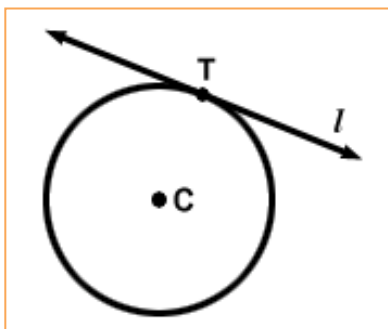
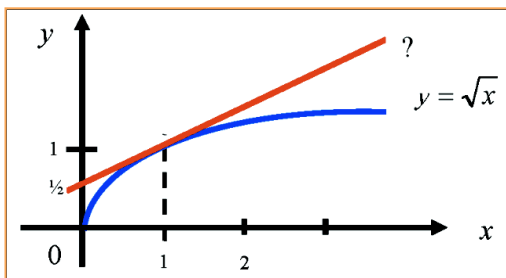


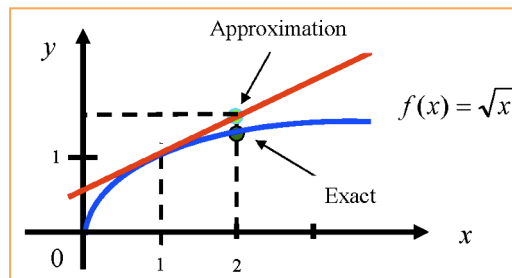
Figure 1. An illustration of tangent line to a curve at the point of tangency.

Other students first see the formal concept of tangent in a more algebraic setting where they are routinely being asked to find the equation of tangent lines without really seeing the purpose of doing so. As this point, it is important to convey the notion of tangent approximation through various non-polynomial functions, such as $y = \sqrt{x}$ as shown in Figure 2(a).

By obtaining the equation of the tangent line at $x = 1$, which is a first degree polynomial equation $y = f(x) = \frac{1}{2}x + \frac{1}{2}$, one can use it to approximate values such as $\sqrt{2}$ that are otherwise impractical to calculate by hand. With this, $\sqrt{2} \approx \frac{1}{2}(2) + \frac{1}{2} = 1.5$ is shown in Figure 2(b). Although the approximation is somewhat crude, considering that the actual value of $\sqrt{2}$ is about 1.414, it more or less provides a good starting point for further investigation.



(a)



(b)

Figure 2

In order to decrease the error of such polynomial estimation, one may consider using higher-order polynomials such as quadratic or cubic polynomials, therefore arriving at the notion of Taylor approximation.

At the point of approximation, $x = a$, we have ...

$$\begin{aligned}
 f(x) &\approx \sum_{n=1}^N \frac{f^{(n)}(a)}{n!} (x-a)^n \\
 &= f(a) + \frac{f'(a)}{1!} (x-a) + \frac{f''(a)}{2!} (x-a)^2 + \dots + \frac{f^{(N)}(a)}{N!} (x-a)^N.
 \end{aligned}$$

Notice that the truncation $f(x) \approx f(a) + f'(a)(x-a)$ resembles the equation of the tangent line discussed previously. Hence, for $f(x) = \sqrt{x}$ with $a = 1$, we can approximate \sqrt{x} using the series

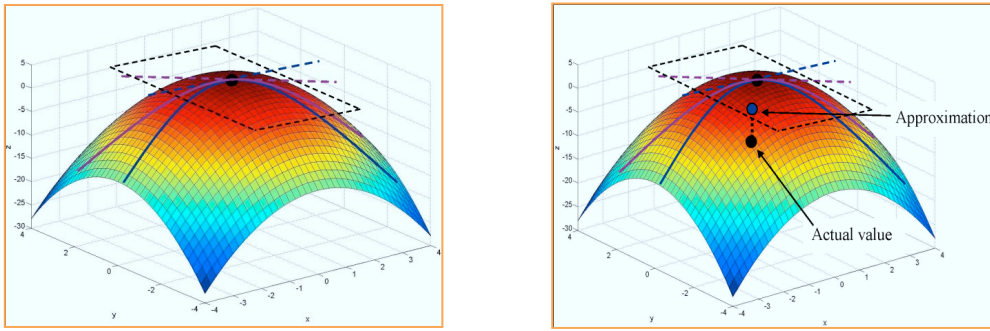
$$\sqrt{x} \approx 1 + \frac{1}{2}(x-1) - \frac{1}{8}(x-1)^2 + \frac{1}{16}(x-1)^3 - \frac{5}{128}(x-1)^4 + \frac{7}{2567}(x-1)^5 - + \dots$$

For $x = 2$,

$$\sqrt{2} \approx 1 + \frac{1}{2} - \frac{1}{8} + \frac{1}{16} - \frac{5}{128} + \frac{7}{2567} - \frac{21}{1024} + \frac{33}{2048} - \frac{429}{32768} + \dots$$

One may choose to take as many or as few terms as needed depending on the desired accuracy. In fact, most of Texas Instruments (TI) calculators such as TI-89 Titanium and TI-Nspire use this series to calculate non-polynomial function values such as this one.

The idea of tangent approximations can be generalized in higher dimensions. A function in three variables whose graph, $z = f(x, y)$, is a surface in 3-dimensional space. The notion of tangent is no longer a line in this setting. Instead, there are infinitely many lines that are tangent to a point $(x, y) = (a, b)$ on the surface, as depicted in Figure 3(a).



(a)
(b)

Figure 3. (a) An illustration of a tangent plane to a function at a given point. (b) Tangent plane can be used as a linear approximation to a function/surface for points near the point of tangency.

Together, these lines form a tangent plane $Ax + By + Cz = D$ to the surface at (a, b) . The tangent plane gives a measure of linear approximation to the function and can be used to calculate function values for points relatively near (a, b) as shown in Figure 3(b).

As expected, the absolute error of approximation at (a, b) , $\left| f(a, b) - \frac{D - Aa - Bb}{C} \right|$, increases as we move away from the point of tangency. As long as we stay relatively local, the tangent plane captures the variation afforded by the function and provides a convenient way to predict behaviors of the neighboring points.

Although our intuition fails us beyond three dimensions, the idea of linear approximation can be similarly extended to as many dimensions as we wish. In the high-dimensional analog, a surface is called a *manifold* if it looks like a plane locally along with a way to move from one patch to the next. A manifold is *differentiable* if we can do calculus on it. The “tangent” is formally called *tangent space* in higher dimensions.

The most common differentiable manifolds we often encounter are the spheres, $S^2: x^2 + y^2 = r^2$ – any small portion on the sphere looks like a plane and the latitudes and longitudes provide a recipe to relate points and give position.

Now, let us consider using the method of linear approximation to a real-world problem.

- Have you ever thought about how an ATM recognizes the dollar amount on a handwritten check?
- What about how the postal office sorts mails according to zip codes?

These two problems are considered under the umbrella of *handwritten digit recognition/classification*. You must wonder how these problems would have anything to do with manifold and tangent space. Before we proceed, a little background in the mathematics of digital images is in order.

A digital image of size $M \times N$ can be represented using a matrix of size $M \times N$ where each cell of the matrix contains a value between 2^0 and 2^k that gives different shades of gray and k is the number of *bits* a computer affords. For example, an 8-bit ($k = 8$) machine has 256 (0 – 255) shades of gray where a value of 0 means the pixel is completely dark while a value of 255 means the pixel is completely white.

Figure 4 shows how a simple black-and-white image is represented by a corresponding matrix of the respective size.

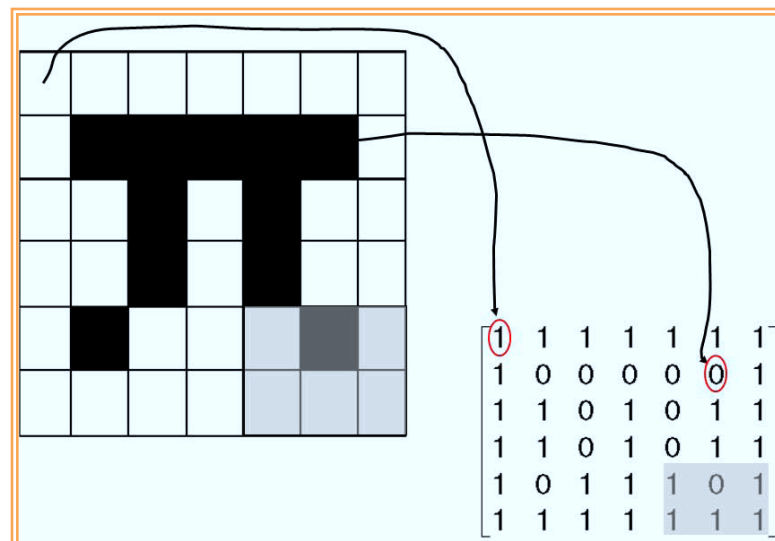


Figure 4. An illustration how an 8-bit image is represented by a matrix. In this example, black pixels are represented numerically by 0 while the white pixels are represented by 1.

Furthermore, we can concatenate an image matrix by columns so it can be viewed as a vector in MN -dimensional space. Precisely, the second column of the image matrix is moved underneath the first, the third column is then moved below the previous two, etc., as shown in Figure 5. This way, a 6×7 image corresponds to a vector in \mathbb{R}^{42} , a 42-dimensional space.

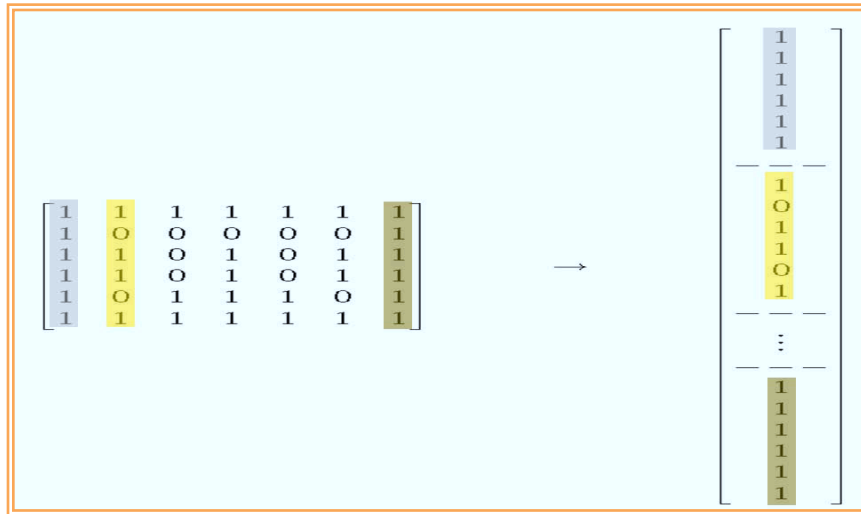


Figure 5. A matrix of size $M \times N$ can be viewed as a vector in MN -dimensional space after column concatenation. In particular, $M = 7$ and $N = 6$ in this figure.

Thus, each monochrome digital handwritten image, in Figure 6, is a member of \mathbb{R}^d , where d represents the resolution of the images. Notice that $d = 32^2 = 1024$ in Figure 6.

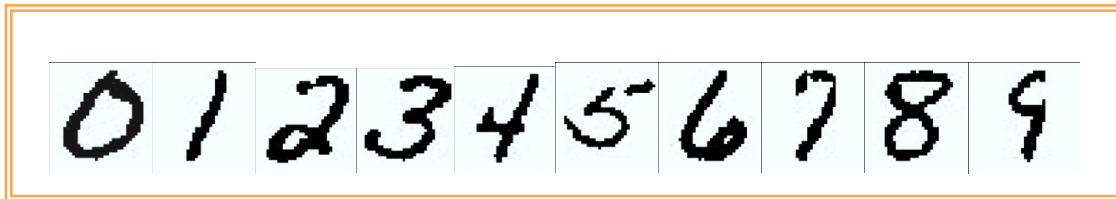


Figure 6. Example digits. Each digital image is of 32×32 , therefore corresponding to a vector in \mathbb{R}^{1024} .

The problem of handwritten digit classification is the art of classifying an unseen digit using knowledge gained from a collection of labeled one, i.e., ones that we know the identity of. For simplicity, say we want to identify the unknown digit in Figure 7(a) given the known digits in Figure 7(b).

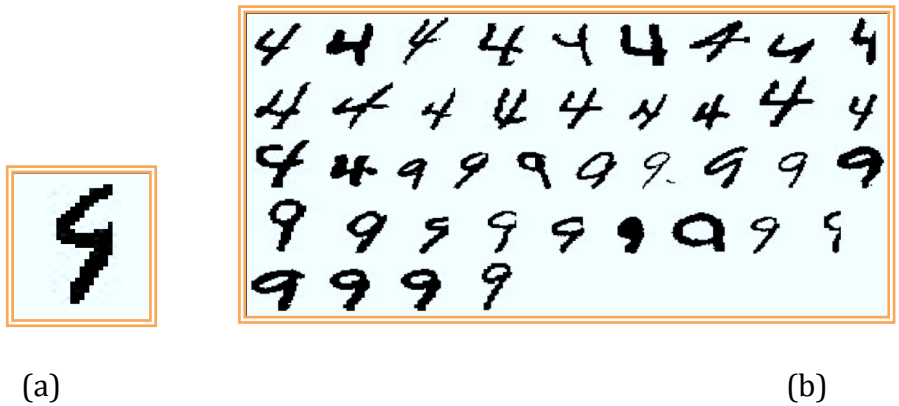


Figure 7. (a) A probe digit. (b) Gallery digits.

The goal is to somehow learn the characteristics exhibited in the gallery patterns that are class-specific and can be used easily for assigning membership of the probe.

Geometrically, if we imagine all the 4's (from the training) live on some underlying manifold and all the 9's (from the training) live on another high-dimensional manifold, as depicted in Figure 8(a), then the straight-line distances between the probe and two random points in the gallery are shown as the dashed lines in Figure 8(a).

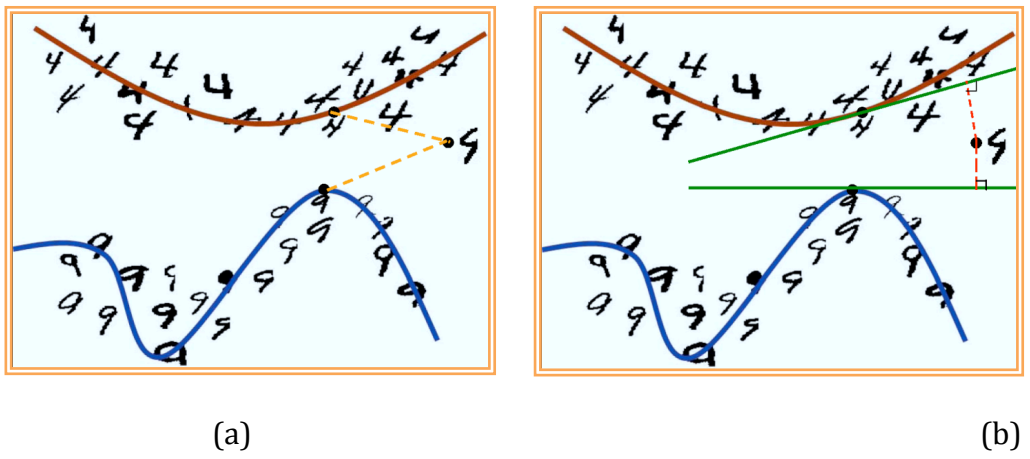


Figure 8. The two curves passing through the digits are the digit manifolds from the training digits. (a) The Euclidean (straight-line) distance between the probe and the gallery digits are depicted as the dashed line. (b) The tangent distances between the selected gallery digits and the probe.

This straight-line distance is typically known as the Euclidean distance. Based on this distance measure, the probe would be labeled as the digit 4 since its distance to the selected digit 4 is less than the distance to the selected digit 9.

On the other hand, we can create a tangent space at each of the selected gallery patterns and calculate the tangent distance between the probe and these gallery points. As illustrated in Figure 8(b), the tangent distance between the probe and the selected digit 4 is the shortest path between the probe and all the points on the tangent space. Measuring distances this way, the probe would be classified as digit 9, which turns out to be the correct label. For a more detailed discussion, readers are referred to [1].

It is not hard to see that comparing a point to a space of points would work better than comparing it to just a single point since a space of points exhibit much more variation than a single point. This is why the tangent distance works better than the Euclidean distance for this problem.

Concluding Remarks

An important point to make here is that the given example is not to impress you with all the fancy things tangent space/distance can do; rather, the idea of taking a simple concept like the “tangent approximation” and seeing how far it can go is what we are really after.

In general, students tend to learn mathematics better when there is a concrete application to relate abstract concepts to. It is easier to engage students in the problem-solving processing when they are motivated by real-world problems that are particularly personal to them.

You might think that the ability to come up with real-world applications is somewhat limited to people who have backgrounds in industrial training. In fact, if we would just pause and try to think outside of the box for a while before we sit down and plan the day's lesson, there are numerous other ways to motivate the learning of mathematics that work for our individualized teaching styles. I challenge you to find something that works for you.

References

P. Simard, Y. Cun, J. Denker, and B. Victorri. Transformation invariance in pattern recognition - tangent distance and tangent propagation. *IJIST*, **11**:181—194, 2001.