Gibbs Phenomenon and its Applications in Science and Engineering

Josué Njock Libii

Engineering Department
Indiana University-Purdue University Fort Wayne
Fort Wayne, Indiana, 46805-1499
Libii@engr.ipfw.edu

Abstract

Gibbs phenomenon arises in many applications. In this article, the author first discusses a brief history of this phenomenon and several of its applications in science and engineering. Then, using the Fourier series of a square-wave function and computer software in a classroom exercise, he illustrates how Gibbs phenomenon can be used to illustrate to undergraduate students the concept of nonuniform convergence of successive partial sums over the interval from 0 to \( \pi \).

1. Introduction

Gibbs Phenomenon is intimately related to the study of Fourier series. When a periodic function \( f(x) \) with a jump discontinuity is represented using a Fourier series, for example, it is observed that calculating values of that function using a truncated series leads to results that oscillate near the discontinuity [12]. As one includes more and more terms into the series, the oscillations persist but they move closer and closer to the discontinuity itself. Indeed, it is found that the series representation yields an overshoot at the jump, a value that is consistently larger in magnitude than that of the actual function at the jump. No matter how many terms one adds to the series, that overshoot does not disappear. Thus, partial sums that approximate \( f(x) \) do not approach \( f(x) \) uniformly over an interval that contains a point where the function is discontinuous [23]. This behavior, which appears in many practical applications, is known as Gibbs Phenomenon; it is a common example that is used to illustrate how nonuniform convergence can arise [3]. Detailed proofs and demonstrations of Gibbs phenomenon using square waves are found in the literature [2, 3, 22] and on web sites [1, 24, 25, 26].

2. History and applications

This phenomenon is not new. Indeed, an important body of knowledge already exists that relates the history and theory of Gibbs phenomenon to various applications in science and engineering. Such work can be used to enhance the education of engineering students. Gibbs phenomenon was
first observed by Wilbraham in 1848 but his analysis of it was limited [14]. It was rediscovered in a different series by Gibbs, the American mathematical physicist, who gave this unusual behavior a more precise mathematical analysis [13]. But it was Bôcher who demonstrated that this phenomenon was a general property of the behavior of partial sums of Fourier series in the vicinity of jump discontinuities [15]. This behavior was named after Wilbraham and Gibbs. However, in most references, it is commonly referred to as Gibbs phenomenon, although in their review of this phenomenon, Hewitt and Hewitt [1] referred to it as the Gibbs-Wilbraham phenomenon.

The exact details of the history of Gibbs phenomenon are very controversial, however. Lanczos[2] reports one version. Gottlieb and Shu [1] report that they were unable to find evidence to support the account given by Lanczos. Despite this controversy, however, Gibbs phenomenon is not simply a historical observation; its practical importance has been consistent in applied science and engineering. During the development of radar in Great Britain, for example, engineers used the sawtooth function to give x-coordinates on their oscilloscopes regularly. In doing so, they encountered Gibbs phenomenon [12]. Today, this phenomenon routinely appears in the processing of digital signals [5]; particularly, in the design of digital filters by electrical engineers. A digital filter that is well designed must account for this phenomenon [17]. Other applications include techniques for numerical analysis and computation [1,4], vibration and stability of complex beams[16], pseudo-spectral time-domain analysis [18], and cubic-spline interpolation of discontinuous functions [19]. In physical optics, for example, when beam irradiance of a top-hat beam in an aperture is represented as a plot of the beam irradiance cross-sectional surface past an aperture, it displays Gibbs phenomenon. Another example is the zip-zero filling in magnetic resonance (MR) imaging. In MR imaging, when one expects pixels that are smaller than the actual resolution of the image, it is common to perform the so-called zip-zero filling interpolation. In this technique, zeroes are substituted for unmeasured data points that are expected to be very small, in order to increase the size of the sample matrix prior to the application of Fourier transforms. Thus, a signal intensity profile across the skull of a patient will show Gibbs phenomenon; this happens because the measurement process is designed to decompose recorded intensity profiles into their Fourier harmonics and the intensity of the signal that is used in the measurement changes from finite to infinitesimal values at the boundary between the brain and the skull. This yields high-frequency oscillations at the edges. In general, this phenomenon is observed in MR imaging whenever there are instantaneous transitions, tissue boundaries, or tissue discontinuities. In practice, these oscillations are suppressed by filtering the images. A final example is that, from a computational point of view, Gibbs phenomenon can be viewed as an issue of recovering local information from global information, or, specifically, as one of recovering point values of a function from its expansion coefficients [1].

The rest of the paper is organized in the following manner: first, we discuss Fourier series, relate them to Gibbs Phenomenon, and introduce the square-wave function; then, we contrast ordinary convergence to uniform convergence; next, we discuss the nonuniform convergence of the square-wave function; finally, we discuss two different ways to demonstrate Gibbs Phenomenon graphically: by using the plots of successive partial sums themselves and by using plots of the
magnitudes of their local peaks.

3. Fourier series

Trigonometric series of the form

\[ f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \]

are important in many problems in physics, engineering and biology. Here, \(a_0, a_n, b_n\), are coefficients to be determined. Such series and periodic functions that can be represented by them arise naturally in many practical applications. Yamashiro and Grudins [20] used Fourier series to model the flow of air in the lungs; many electric sources of energy generate waveforms that are periodic; sweep generators used to control electron beams in cathode-ray oscilloscopes produce triangular waves; nonlinearities that arise in circuits that are supposed to be linear create periodic functions; and when a nonfiltered electronic rectifier is driven by a sinusoidal source, it produces sine waves that are rectified [21]. Fourier series are used in a wide variety of other fields including electromagnetic wave theory [6], heat transfer [9], acoustics [23], mechanical and structural vibrations [11], and frequency analysis of signals [5].

In contrast to Taylor series, which can only be used to represent functions that have many derivatives, trigonometric series can be used to represent functions that are continuous as well as those that are discontinuous. In general, the partial sum of the series approximates the function at each point. For each point where the function itself is continuous, the closeness of the approximation generated by the series improves as one adds more and more terms to the partial sum. However, when a periodic function \(f(x)\) with a discontinuity is represented using a Fourier series, it is observed that calculating values of that function using a truncated series leads to results that oscillate near the discontinuity [23]. As one includes more and more terms into the series, the oscillations persist but they move closer and closer to the discontinuity itself. Indeed, it is observed that the series representation yields an overshoot at the jump, a value that is consistently larger in magnitude than that of the actual function at the jump [12]. No matter how many terms one adds to the series, that overshoot does not disappear. Thus, partial sums that approximate \(f(x)\) do not approach \(f(x)\) uniformly over an interval that contains a point where the function is discontinuous.

Gibbs phenomenon arises naturally in the study of Fourier series. A square wave function is shown in equation (1). It is a simple example that has been used historically in books of applied mathematics to illustrate Gibbs phenomenon analytically [2, 3, 7, 22].

\[ S(x) = \begin{cases} 
-1, & -\pi < x < 0 \\
+1, & 0 < x < \pi 
\end{cases} \] ......................(1)

Nowadays, however, the availability of software that can evaluate and plot functions easily makes
it very convenient to show Gibbs Phenomenon in a way that makes it accessible to all students. The results can even be animated. Indeed, Gottlieb and Shu [1], who have studied Gibbs phenomenon extensively, display an effective animation of this phenomenon on their web site. However, in that animation, students observe the results of a completed process and it is difficult for the novices among them to learn from the details that led to the final results they see on the animation. Our experience indicates that, after students have plotted partial sums on their own, such an animation becomes a teaching tool that displays a dynamic synthesis of Gibbs Phenomenon. The procedures discussed in this article can be used to fill this gap by showing how software can be utilized interactively to demonstrate the nonuniform convergence of partial sums that give rise to Gibbs phenomenon. The methods shown here can be used in undergraduate courses in mathematics, physics, and engineering.

4. Ordinary convergence vs. uniform convergence

If the partial sums $S_n$ of an infinite series of numbers satisfy

$$\lim_{n \to \infty} S_n = S, n = 1, 2, 3, \ldots$$

then, the series is said to converge to the sum $S$. If the limit of $S_n$ does not exist, then the series is said to diverge. Thus, for any preassigned positive number $\epsilon$, no matter how small, one can find a number $N$ such that

$$|S - S_n| < \epsilon, n > N$$

This defines ordinary convergence [6].

On the other hand, a sequence $\{S_n(x)\}$ converges uniformly to $S(x)$ in a given interval $[a, b]$, if for each positive $\epsilon$, there is a number $N$ that is independent of $x$, such that, for all $n > N$, one has

$$|S_n(x) - S(x)| < \epsilon, a \leq x \leq b$$

This concept of uniform convergence, defined for sequences, is extended to series because an infinite series is defined to be the limit of the sequence of its partial sums. Accordingly, we let $\sum u_n(x)$ denote a series of functions that are defined in a given interval $[a, b]$, with partial sums $S_n(x)$ given by

$$S_n(x) = u_1(x) + u_2(x) + u_3(x) + \ldots + u_n(x) \ldots \ldots \ldots \ldots \ldots (2)$$

Then, if the sequence of partial sums converges uniformly to a function $S(x)$, the series $\sum u_n(x)$ is said to converge uniformly. Otherwise, the series is not uniformly convergent [6]. For teaching purposes, it is helpful to illustrate these concepts and their usefulness by using specific examples.
5. Nonuniform convergence of the square-wave function

Consider the Fourier series of \( f(x) \) as shown below.

\[
f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} \left( a_n \cos nx + b_n \sin nx \right); -\pi < x < \pi \quad \text{.........(3)}
\]

where

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx \quad \text{...............}(3a)
\]

\[
b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx \quad \text{...............}(3b)
\]

We use the square wave function given in equation (1) to illustrate convergence that is not uniform.

If we extend \( S(x) \) so that it becomes periodic with period \( 2\pi \), then, this function exhibits a jump discontinuity of +2 at \( x = -\pi \), a second one of -2 at \( x = 0 \), and a third one of -2 again at \( x = \pi \), and so on.

\( S(x) \) being an odd function, its Fourier series consists solely of sine terms given by

\[
S(x) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2n-1)x}{(2n-1)} \quad \text{...............(3c)}
\]

In order to visualize the behavior of \( S(x) \) clearly and compare it to the original function, it is helpful to plot the Fourier series given in (3c) together with the original function given in (1). But the Fourier series has an infinite number of terms and one cannot plot them all. In practice, one needs to decide how many terms to keep in the sum. Consequently, it is necessary to replace the infinite sum of Equation (3c) with \( S_N(x) \), a partial sum of \( N \) terms, given by

\[
S_N(x) = \frac{4}{\pi} \sum_{n=1}^{N} \frac{\sin(2n-1)x}{(2n-1)} \quad \text{...............(3d)}
\]

\[
= \frac{4}{\pi} \left( \frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \frac{\sin 7x}{7} + \ldots + \frac{\sin(2N-1)x}{2N-1} \right)
\]

Since we know that this series converges, one would expect the final value of \( S_N(x) \) at the edge of the discontinuity to be either -1 or +1, depending upon whether one is to the left or to the right of a discontinuity, or to approach these numbers asymptotically. However, such is not the case \([3, 7, \text{and } 8]\). There are several ways to show that this assertion is true.

One way uses complex notation for Fourier series. In this case, the form of the series that is
given in Eq. (3) is not as convenient to use as the complex form corresponding to it. The latter is given by

\[ f(x) = \sum_{-\infty}^{\infty} c_n e^{inx}, \ldots\ldots\ldots\ldots(4) \]

where

\[ c_0 = \frac{a_0}{2}, \text{and} \]

\[ c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\xi) e^{inx} d\xi = \frac{1}{2} (a_n - ib_n), \ldots\ldots(4a) \]

The partial sums of equation (4) can be written as [3]

\[ S_N(x) = \sum_{n=-N}^{N} c_n e^{inx} = \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(x,\xi) f(\xi) d\xi, \ldots\ldots(5) \]

where

\[ K_N(x,\xi) = \sum_{n=-N}^{N} e^{i\eta(x-\xi)} = \frac{\sin(N + \frac{1}{2})(x - \xi)}{\sin\frac{1}{2}(x - \xi)} \ldots\ldots(5a) \]

When one particularizes equations (5) and (5a) to the square wave given in Equation (1) and focuses attention to the neighborhood of the discontinuity at, say, \( x = 0 \), it is found that the maximum value of \( S_N(x) \) is reached at \( x = B/(N+1/2) \), where \( N \) can be made arbitrarily large, and the maximum value of \( S_N(x) \) at that point is given by [3,7,8]

\[ \left( S_N(x) \right)_{\text{max}} = \frac{2}{\pi} \int_{0}^{\pi} \frac{\sin\eta}{\eta} d\eta \approx 1.179, \ldots\ldots(6) \]

This result was obtained by first writing the integrand of eq.(6) as a power series and then integrating the resulting series term by term [12]. Equation (6) shows that the maximum value of the partial sum \( S_N(x) \) of the infinite series given in equation (2) is greater than 1 near the discontinuity (at \( x = 0^* \)), no matter how many terms one chooses to include in the partial sum \( S_N(x) \). But “1” is the maximum value of the original function that is being represented by Fourier series in the interval \( 0 < x < \pi \). It follows, therefore, that, near the discontinuity, the maximum difference between the value of the partial sum and the function itself, sometimes called the overshoot, the bump, or the error, remains finite as \( N \to \infty \). And it is approximately 0.18 in this case. Thus, the condition

\[ \left| S_n(x) - f(x) \right| < \varepsilon, n > N \]
is not satisfied near the discontinuity and, hence, is not satisfied everywhere in the interval \(0 < x < \pi\). Therefore, convergence is not uniform over the specified interval.

In our teaching, we have found two other methods that can be used to demonstrate nonuniform convergence using Gibbs Phenomenon. The first to be presented here is a graphical illustration of the shapes of the curves that represent successive partial sums. The second is the comparison of how the magnitudes of the local peaks achieved by these curves decrease as a function of the number of terms retained in the partial sums. These two methods are particularly effective in teaching sophomore and junior classes. Students understand them quickly because these methods utilize basic materials from math courses that the students would have taken in the first two years of their university education.

6. Graphical illustration: plots of partial sums

We can show plots of successive approximations of the series given by equation (2a) by letting \(N = 1, 2, 3, \ldots, 100\). In other words, we plot the sequence of partial sums \(S_1, S_2, S_3, \ldots, S_{100}\). We limit our plot window to the interval \(0 < x < \pi\) because this scale allows us to see the fine oscillations of the plotted function more easily than if we plotted the whole period \((-\pi < x < \pi)\). First, we plot only one term of the series, then the first two, then the first three, and so on. We continue the process this way, adding one term at a time, and watching for Gibbs phenomenon in every new plot until significant changes in the plots can no longer be detected, or until we reach the limits of the resolution of the device being used to generate the plots. We generated fifty, or a hundred plots, corresponding to as many different partial sums. However, in order to economize the use of space, only a representative sample of the whole set is actually displayed at any given time. The resulting graphs can be sorted and arranged to show the variation of the peak values of \(S_N(x)\). After looking at many of these plots, it can be seen in that the peak values of the partial sums near the edges of the discontinuities do not change very much after a while and that they do remain larger than 1.18, no matter how many of the forty terms one includes into the partial sums.

The plots discussed above and the computations on which they are based can be generated using current versions of Maple, Mathematica (Fig.1), and MATLAB. We computed the first fifty terms and Gibbs phenomenon was clearly shown in all the fifty plots. We repeated this process with the next fifty terms and obtained similar results. Our plots are not displayed herein because there are many web sites that display similar plots and they do so in color [24, 25, 26, 27].

7. Magnitudes of the local peaks of partial sums

Here, we first determine the location of the peaks that are closest to the discontinuities; then, we evaluate the functions that represent the successive partial sums at those peaks; finally, we plot the magnitudes of the peaks so obtained as a function of the number of terms in the partial sums.

The peaks and troughs were located by taking the derivatives of the partial sums, \(S_N(x)\), given by Equation (2a), setting them to zero and solving for the corresponding values of \(x\). If we use
primed quantities to denote derivatives with respect to x, then, taking first derivatives gives

\[ S_N'(x) = \frac{4}{\pi} \sum_{i=1}^{N} \cos(2n-1)x \] ..........................(7)

\[ = \frac{4}{\pi} \left( \cos x + \cos 3x + \cos 5x + \cos 7x + \ldots + \cos (2N-1)x \right) \]

Using the trigonometric identity \( \cos(A) + \cos(B) = 2 \cos\left(\frac{A-B}{2}\right)\cos\left(\frac{A+B}{2}\right) \), one can find the values of x at which Equation (7) goes to zero. Doing so repeatedly and collecting terms, one can see that the results follow a simple pattern. This detailed work was carried out for each of the first seven partial sums and mathematical induction was used thereafter. This process established the location of peaks and troughs for a partial sum that is made up of any arbitrary number of terms. For an arbitrary integer, N,

\[ S_N'(x^{(N)}_i) = 0, \quad at \]

\[ x^{(N)}_i = \frac{\pi}{2N}, \frac{2\pi}{2N}, \frac{3\pi}{2N}, \frac{4\pi}{2N}, \frac{5\pi}{2N}, \ldots, \frac{(2N-1)\pi}{2N} \] ..........................(8)

Thus, in the interval \( 0 < x < \pi \), the partial sum \( S_N(x) \) has either a local maximum or a local minimum at \((2N-1)\) different points. Their locations are given by Equation (8). The first peak in that interval is located at \( x^{(N)}_1 = (\pi / 2N) \), while the last peak is located at \( x^{(N)}_{2n-1} = [(2N-1)\pi /2N] \). As N becomes larger and larger, the location of the first peak moves to the left and approaches zero while that for the last peak moves to the right and approaches \( \pi \). The value of the partial sum, \( S_N(x) \), at any extremum (peak or trough) is determined by substituting the appropriate expression for x, chosen from Equation (8), into Equation (3c). Doing so for the first peak (the one closest to \( x = 0 \)), we get

\[ S_N\left(\frac{\pi}{2N}\right) = \frac{4}{\pi} \sum_{i=1}^{N} \sin\left[\frac{(2n-1)\pi}{2N}\right] \] ..........................(9)

\[ = \frac{4}{\pi} \left( \sin\frac{\pi}{2N} + \sin\frac{3\pi}{2N} + \sin\frac{5\pi}{2N} + \sin\frac{7\pi}{2N} + \ldots + \sin\frac{(2N-1)\pi}{2N} \right) \]

It can be seen that the expression given in equation (9) implies that of Equation (6). The former has been evaluated for \( N = 1, 2, \ldots, 100 \). The values so obtained are then studied to see how their magnitudes behave as one includes more and more terms into the partial sums. It was observed that these magnitudes did not change very much; indeed, they did remain larger than 1.18, no matter how many terms we included into the partial sums. Similarly, we can study how the magnitudes of the first through fifth peaks changed with the addition of new terms to the partial sums. Here, however, the resulting curves were so close to each other that, without excellent
resolution of the plotting software, they may appear indistinguishable. To help separate these curves, somewhat, one needs to change the scale of the graph by computing and plotting the overshoots only. An overshoot of a peak is defined as the amount by which its magnitude exceeds unity. Thus, the first peak has the largest overshoot. That overshoot decreases as one adds more terms to the partial sum. However, it does not fall below 0.18, even after one hundred terms are utilized. The second peak is smaller. The third peak is smaller yet, and so on [7]. It was also observed that the closer the peak is to the discontinuity, the larger its magnitude and the slower its rate of decrease. Thus, the first peak decreases very slowly, the second peak decreases faster than the first, the third peak faster than the second, and so on. This process continues until one reaches the middle of the interval, where, because of symmetry, the rate of decrease must be greatest. It was observed that points in the middle of the interval reach a magnitude of unity first and this effect moves slowly toward the edges. However, the points at the very edge of the discontinuity never come arbitrarily close to unity [7].

8. Conclusion

With appropriate computer software, it is possible to do a demonstration of Gibbs phenomenon interactively in class, to assign it as homework, or to use it as a small, hands-on, class project. By using software to do repetitive numerical evaluations of the partial sums involved and then plotting the results graphically on a computer screen, for example, students can create self-paced demonstrations of Gibbs phenomenon for themselves. They then see what it is, how it arises, and, at the same time, illustrate to themselves the meaning of the concept of nonuniform convergence of a function over a specified interval. Obviously, web sites currently exist that allow students to observe this behavior interactively by using results that were generated by someone else [25]. Nevertheless, the two graphical techniques discussed here are easy to implement because they are accessible to students as soon as they learn how to plot trigonometric functions and to take their derivatives. Mathematical subtleties are kept to a minimum and the results are easy to interpret. These techniques have been tested in our classrooms and labs and they have been found to be effective learning tools for our students. We believe that they will be useful to many others as well.

9. References


Fig. 1. A sample plot generated with Mathematica [26]

**Biographical Information**

JOSUE NJOCK LIBII,
Associate Professor of Mechanical Engineering,
BSCE, MSE, PhD.(Applied Mechanics), University of Michigan, Ann Arbor, Michigan