IVY KIDRON

POLYNOMIAL APPROXIMATION OF FUNCTIONS: HISTORICAL PERSPECTIVE AND NEW TOOLS[1]

ABSTRACT. This paper examines the effect of applying symbolic computation and graphics to enhance students' ability to move from a visual interpretation of mathematical concepts to formal reasoning. The mathematics topics involved, Approximation and Interpolation, were taught according to their historical development, and the students tried to follow the thinking process of the creators of the theory. They used a Computer Algebra System to manipulate algebraic expressions and generate a wide variety of dynamic graphics; thus 21st century technology was applied in order to "walk" with the students from the period of Euler in 1748 to the period of Runge in 1901. We describe some situations in which the combination of dynamic graphics, algorithms, and historical perspective enabled the students to improve their understanding of concepts such as limit, convergence, and the quality of approximation.

KEY WORDS: animation, approximation, CAS (Computer Algebra System), historical perspective, interpolation, limit

1. INTRODUCTION

Euler's book *Introductio in Analysin Infinitorum* (1748) was translated into English for the first time in 1988. In the introduction, the translator notes that the impetus for the translation was a remark by André Weil in 1979, when Weil spoke at the University of Rochester on the life and work of Leonard Euler. Weil said that he was trying to convince the mathematical community that students of mathematics would profit much more from studying Euler's *Introductio in Analysin Infinitorum* than using the available modern textbooks.

I followed Weil's advice. In order to teach the central ideas in analysis, ideas that are rather abstract, I decided to follow the intuitive thinking of the founders of the theory. This was done by using original sources that were enlivened and enhanced with laboratory experiments in a CAS (computer algebra system) environment. This paper is based on a research study (Kidron, 1999) that examined such an approach to teaching analysis at the high-school level. Usually at this level, students are only taught topics in analysis that were developed in the time of Newton and Leibniz. However, by making use of a computer algebra system, *Mathematica*

International Journal of Computers for Mathematical Learning **8:** 299–331, 2003.

(Wolfram, 1999), it was possible to extend the coverage to introduce topics from more recent periods of mathematics, focusing on two important subjects: approximation and interpolation.

The linkage between CAS and the historical perspective becomes clear if we recall that physics was the main driving force in the development of analysis. In fact, the old masters developed methods that solved "real" problems. They wanted to calculate explicitly, and thus they developed well-defined algorithms. Nowadays, students can test those algorithms by translating them into CAS programs. The founders of the theory pointed out the importance of analyzing the error that occurs in applying a numerical process. By analyzing the error, students can be introduced to the topics of continuous calculus.

I believe that the combination of dynamic graphics, algorithms, and historical perspective may lead to a more stimulating way of learning analysis by means of numerical processes. My research has focused on examining the extent to which this combination actually helped the students in the transition from their visual, intuitive interpretation of mathematical concepts to formal reasoning.

This paper deals specifically with the conceptual understanding of the convergence process obtained by approximating a function by means of polynomials. Key concepts in analysis such as limit and infinite sum are closely related to approximation theory, and therefore I tried to clarify the limit concept for students by means of polynomial approximations of functions. A motivation for this is the fact that there is general agreement in the research literature about the difficulties experienced by students in learning some of the key concepts in analysis, especially those like the limit concept that are related to infinite processes (Courant and Robbins, 1941; Davis and Vinner, 1986; Cornu, 1981, 1991; Tall, 1992; Cottril and Dubinsky, 1996).

Whereas most of the studies in the literature deal with the conceptual difficulties encountered in the notion of a limit of a sequence of numbers, this study analyzes the students' perceptions of the limit of a sequence of functions. It also investigates students' conceptual understanding of the quality of polynomial approximation. In addition, I analyze the role played by Mathematica in enabling the students to "walk the same paths" as the founders of mathematical theory from Euler (1748) to Runge (1901).[2]

The research literature regarding CAS technology reports favourably on the effects of the software in stimulating students to explore on their own (Breuer and Zwas, 1993), in shaping students' understanding by providing (sometimes unexpected) feedback (Dreyfus and Hillel, 1998), and in enabling students to develop visual intuitions before the formal statement of a theory (Tall, 2000a). The literature also provides some warnings concerning the use of CAS (for example, Tall, 1993; Artigue, 2001). The complexity of the "instrumentation process" (i.e. how the tool becomes an effective instrument of mathematical thinking for the learner) involved in learning in a CAS environment is discussed in Artigue (2001). Nevertheless, despite the complexity and the obstacles, this research was undertaken with the belief that using a CAS system to investigate the works of the "old masters" could indeed be rewarding.

One particular use of CAS in this study is to highlight the different approaches of the founders of the theory. For example, there is a dialectic between the modes of thought employed by Euler and those used by Cauchy. Thus, in the laboratory the students were given the opportunity to calculate "with Euler" and to visualize "with Cauchy". On the one hand, we wanted the students to use Mathematica's symbolic capabilities to perform Euler and Lagrange manipulations of algebraic expressions. On the other hand, we wanted the students to use Mathematica's dynamic graphics to visualize properties like convergence that were the subject of Cauchy's analysis. In this way, we aimed to create an environment that supports reasoning about formal calculation and reasoning by continuity.

2. HISTORICAL PERSPECTIVE, NEW TOOLS

In this section, I present the didactic design in two parts, devoted to the historical and technological dimensions, respectively.

2.1. *The Intuitive Ideas of the Founders of the theory of Analysis and the Implementation of Their Ideas in the Laboratory*

This historical review is divided into parts: the first is devoted to Approximation theory and the second to Interpolation theory.

2.1.1. *Approximation Theory: The works of Euler, Lagrange and Cauchy*

In the preface to his *Introductio in Analysin Infinitorum* (1748), Euler hints at the obstacles in learning Analysis, which according to him stem from a lack of knowledge of Algebra.

The following is the English translation by Blanton (Euler, 1988): "Often I have considered the fact that most of the difficulties which block the progress of students trying to learn analysis stem from this: that although they understand little of ordinary algebra, still they attempt this more subtle art". We also quote Euler's reason for applying an algebraic approach to subjects that are usually discussed in analysis: "in order

Leonhard Euler (1707–1783)

Among his mathematical achievements
The intuitive idea to express non-polynomial functions as polynomials with "an infinite number of terms".
Euler's approach is via algebra: Expansion of given functions in power series by the method of undetermined coefficients.

Our application in the laboratory
First, we used an analytical approach, then following Euler's experimental thinking, we used his algebraic approach to represent infinite sums. We made a graphical representation of the results.

Figure 1. Application of Euler's ideas in the laboratory.

PRÆFATIO.

SÆpenumero animadverti, maximam difficultatum partem, quas Matheſeos cultores in addiſcenda Analyſi infinitorum offendere ſolent, inde oriri, quod, Algebra communi vix apprehenſa, animum ad illam ſubliniorem artem appellant; quo fit, ut non ſolum quaſi in limine ſubſiſtant, ſed etiam perverſas ideas illius infiniti, cujus notio in ſubſidium vocatur, ſibi forment. Quanquam autem Analyſis infinitorum non perfectam Algebræ communis, omniumque artificiorum adhuc inventorum cognitionem requirit; tamen plurimæ extant quæſtiones, quarum evolutio diſcentium animos ad ſubliniorem ſcientiam præparare valet, quæ tamen in communibus Algebræ elementis, vel omittuntur, vel non ſatis accurate pertractantur. Hanc ob rem non dubito, quin ea, quæ in his libris congeſſi, hunc defectum abunde ſupplere queant. Non ſolum enim operam dedi, ut eas res, quas Analyſis infini-

Figure 2. Euler's Preface to *Introductio in Analysin Infinitorum*.

that the transition from finite analysis to analysis of the infinite might be rendered easier".

In the laboratory, the author (who was also the teacher) used two different approaches to approximate a given function by polynomials: analytical and algebraic. These approaches are demonstrated in Brown et al. (1991).

The analytical approach introduced the notion of *order of contact*. We examined two functions whose formulas are different but whose plots are similar when approaching x = 0. For example, Figure 3 shows the graphs of $f(x) = \frac{1+2x}{1-x-x^2}$ (the dashed curve) and $g(x) = 1 + 3x + 4x^2 + 7x^3$. Both curves pass through the point (0, 1) and the functions have the same first, second, and third derivatives at $x = 0$.

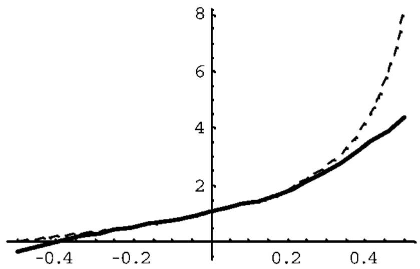


Figure 3. f(x) (the dashed curve) and g(x) for $-0.5 \leq x \leq 0.5$.

The notion of order of contact was defined:

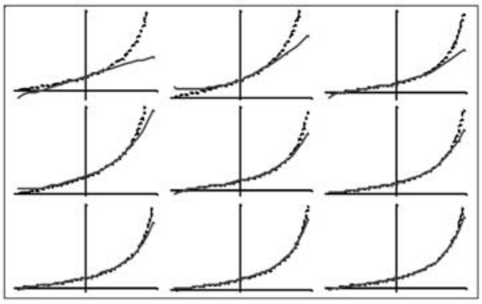
> Two curves, $y = f(x)$ and $y = g(x)$ have an order of contact n at $x = 0$ if $f(0) = g(0), f'(0) = g'(0), f''(0) = g''(0), \ldots, f^{(n)}(0) = g^{(n)}(0)$.

In the example in Figure 3, the order of contact of f(x) and g(x) at $x = 0$ is 3.

As an application, the students were required to find the polynomials $P_n(x)$ of degree n (for different given values of n) that have the highest possible order of contact with a given function at $x = 0$. The students were asked to use Mathematica to solve the relevant system of equations.

In *the algebraic approach*, Taylor polynomials were introduced by the intuitive idea of Euler to express non-polynomial functions as polynomials with "an infinite number of terms". Mathematica was used to *follow the original text of* Euler (1748/1988). The students were asked to follow Euler's "experimental" thinking and to use his algebraic approach to represent infinite sums. Euler wrote, "Since the nature of polynomial functions is very well understood, if other functions can be expressed by different powers of Z in such a way that they are put in the form $A + BZ + CZ^2 + DZ^3 + \ldots$ then they seem to be in the best form for the mind to grasp their nature, even though the number of terms is infinite". The problem posed in the lab was as follows: Let f(x) be a real function that is a quotient of two polynomials $f(x) = P(x)/Q(x)$. We seek a "polynomial with an infinite number of terms" such that $\frac{P(x)}{Q(x)} = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots$. The students were asked to work on the example, $f(x) = \frac{1+2x}{1-x-x^2}$, which is Euler's example also. According to Euler's instructions, there is an infinite series such that $\frac{1+2x}{1-x-x^2} = A + Bx + Cx^2 + Dx^3 + Ex^4 + \ldots$ and the coefficients A, B, C, D, ..., which satisfy equality, are found by multiplying both sides by $(1 - x - x^2)$ and comparing the coefficients. As a result,

$\frac{1+2x}{1-x-x^2} = 1 + 3x + 4x^2 + 7x^3 + 11x^4 + \ldots$ "which can be continued as long as desired with no trouble", according to Euler. Mathematica can help the students perform the required calculations. The coefficients are identical with the coefficients of the polynomials obtained analytically; both approaches yield the coefficients of the Taylor series. The students were asked to extract patterns from the calculations, exactly as Euler did (Euler, 1988, p. 52). In addition, they were requested to make a graphical representation of the results. The infinite sum was represented by a "dynamic" plot: the animation of the static plots represented in the analytical approach. Figure 4 shows a dynamic plot that illustrates that in a given interval, the higher the degree n of the approximating polynomial, the closer are the function $f(x) = \frac{1+2x}{1-x-x^2}$ and the approximating $P_n(x)$ polynomial of degree n.



f(x) (the dashed curve) and $P_n(x)$ for n = 1, 2, 3 from left to right.

f(x) (the dashed curve) and $P_n(x)$ for n = 4, 5, 6 from left to right.

f(x) (the dashed curve) and $P_n(x)$ for n = 7, 8, 9 from left to right

Figure 4. "Dynamic" (animated) plot of f(x) and the approximating polynomials.

The aim of using an animation was to enable the students to see the dynamic process in one picture. For examples and further information, see Example 1 of Kidron (2002).

Joseph Louis Lagrange (1736–1813)

Among his mathematical achievements
Error Analysis. The Lagrange remainder
$R_n(x) = f(x) - P_n(x)$
for any x_0 in the domain of f(x):
$R_n(x_0) = \frac{f^{(n+1)}(c)x_0^{n+1}}{(n+1)!}$
for some c between 0 and x_0

Our application in the laboratory
3-dimensional pictures of the upper bound of the absolute value of the error. Animation of those 3-dimensional pictures to illustrate that
$\lim_{n\to\infty} R_n(x) = 0$
$\lim_{x\to 0} R_n(x) = 0$

Figure 5. Application of Lagrange's ideas in the laboratory.

In Lagrange's *Oeuvres* (1884, p. 85), in the "Leçons sur le calcul des fonctions" (leçon neuvieme), Lagrange points out the importance of evaluating the remainder:

toute fonction f (x + i) se développe dans la série $f(x) + if'(x) + \frac{i^2}{2}f''(x) + \frac{i^3}{2\cdot 3}f'''(x) + \ldots$ lorsqu'elle va naturellement à l' infini, à moins que les fonctions dérivées de f(x) ne deviennent nulles, ... tant que ce développement ne sert qu'à la génération des fonctions dérivées, il est indifférent que la série aille a l'infini ou non ... mais si on veut l'employer pour avoir la valeur de la fonction dans les cas particuliers ... il est important d'avoir un moyen d'évaluer le reste de la série qu'on néglige ou du moins de trouver les limites de l'erreur qu'on commet en négligeant ce reste.

I translate here freely Lagrange's words: "every function f(x + i) is expanded in the series $f(x) + if'(x) + \frac{i^2}{2}f''(x) + \frac{i^3}{2\cdot 3}f'''(x) + \ldots$, which continues naturally to infinity unless the derivative functions of f(x) are zero, ... as long as the use of this expansion is to generate the derivative functions, it makes no difference if the series is continued to infinity or not ... but if we want to use the expansion in order to calculate the value of the function in particular cases ... it is important to have a way to evaluate the remainder of the series that we neglect, or at least to find the limits of the error one commits by neglecting the rest".

In the laboratory, the expansion of sin (x) around x = 0 up to degree 5 was called $P_5(x)$. The error $(f(x) - P_5(x))$, the remainder of Lagrange, is $\frac{f^{(6)}(c)x^6}{6!}$ for some c between 0 and the current x value. The absolute value of the error as a function of x and c with $-\pi \le x \le \pi$, $-\pi \le c \le \pi$ was plotted. The c value in $-\pi \le c \le \pi$ that corresponds to the exact error is an unknown number; therefore, the students were requested to look at all pairs (x, c) in the ranges $-\pi \le x \le \pi$, $-\pi \le c \le \pi$.

The following 3-dimensional plot (Figure 6) represents the error (in fact, an upper estimate of the absolute value of the error) as a function of the two variables x and c. (In this specific plot the upper estimate of the error occurs at $x = \pi$ and $c = \frac{\pi}{2}$.)

By animating these 3-dimensional plots, the teacher illustrated that the upper estimate of the error decreases when the degree of the approximating polynomial is increased. This animation of Lagrange's remainder (the approximated function is sin(x) and the animation is on the degree of the approximating polynomial) is demonstrated in (Kidron, 2002, Example 2). The animation illustrates that $\lim_{n\to\infty} R_n(x) = 0$ for $-\pi \le x \le \pi$ and $-\pi \le c \le \pi$. ($R_n(x) = \frac{f^{(n+1)}(c)x^{n+1}}{(n+1)!}$ for some c between 0 and x).

The students were asked the following question: "Suppose that the degree of the approximating polynomial is fixed; could we obtain an animation of Lagrange's remainder with the *domain* as a *variable*?" The teacher demonstrated how to use Mathematica as a programming language

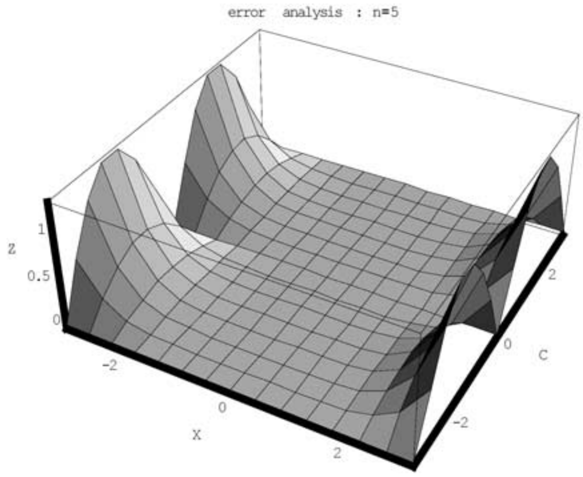


Figure 6. The error as a function of x and c.

to obtain the dynamic graphical output. The purpose of this was to demonstrate that for values of x approaching 0, the upper estimate of the error decreases. The animation (Kidron, 2002, Example 3) illustrates $\lim_{x\to 0} R_n(x) = 0$.

Lagrange pointed out the importance of analyzing the remainder $R_n(x)$ and he established an expression for it, for the practical purposes of applications to mechanics (Lagrange, 1884, p. 85). Cauchy (1821) further investigated the notion of convergence; he stressed that to obtain a convergent series the remainder must approach 0 (Kline, 1972).

Augustin-Louis Cauchy (1789–1857)

Among his mathematical achievements
The formal definition of the limit concept.

Our application in the laboratory
We visualized the formal statements
$\lim_{n\to\infty} R_n(x) = 0$
$\lim_{x\to 0} R_n(x) = 0$

Figure 7. Application of Cauchy's ideas in the laboratory.

Euler and Lagrange employed algebraic methods. Cauchy's approach was different. In the introduction to his *Cours d'Analyse de l'Ecole Royale Polytechnique* (1821), we read:

Quant aux méthodes j'ai cherché à leur donner toute la rigueur qu'on exige en géométrie, de manière à ne jamais recourir aux raisons tirées de la généralité de l'algèbre. Les raisons de cette espèce . . . tendent a faire attribuer aux formules algébriques une étendue indéfinie, tandis que, dans la réalité, la plupart de ces formules subsistent uniquement sous certaines conditions, et pour certaines valeurs des quantités qu'elles renferment. En déterminant ces conditions et ces valeurs, et en fixant d'une manière précise le sens des notations dont je me sers, je fais disparaître toute incertitude.

I translate freely Cauchy's words: "As for the methods, I tried to give them all the rigor requested in geometry, so as to never resort to reasons drawn from the generalization of the algebra. Such reasons . . . tend to attribute to the algebraic formulas an indefinite validity while, in reality, most of the algebraic formulas are valid just under certain conditions and for certain values of the quantities they contain. By determining these conditions and these values, and by fixing in a precise manner the meaning of the notations that I use, any uncertainty will be removed".

We can see here an allusion to the way Euler treated infinite series like finite sums, a way which led to inconsistencies. Euler and Lagrange considered the remainder as an expression. "The subjects of Cauchy's new analysis in 1821 were not expressions but variable quantities . . . With Cauchy, analytical conclusions were drawn not from expressions but from concepts or properties like convergence" (Laugwitz, 1994). Cauchy is talking about the "form versus function" dialectic – formal identities versus continuous variation. Thus, we notice the distinction between the modes of thought employed by Euler (reasoning about formal calculation) and those used by Cauchy (reasoning by continuity).

In Cauchy's programme, the limit concept is taken as the one on which all the others, such as convergence, derivative and integral, are based (Kleiner, 1991). In the first pages of Cauchy's analysis course, the notion of limit is introduced as follows:

Lorsque les valeurs successivement attribuées a une même variable s'approchent indéfiniment d'une valeur fixe, de manière a finir par en différer aussi peu que l'on voudra, cette dernière est appelée la limite de toutes les autres.

I translate here Cauchy's definition: "When the different values successively attributed to the same variable are getting indefinitely close to a fixed value, in a way that they will differ from it as little as we want, this fixed value is called the limit of all the others".

Cauchy's definition suggests continuous motion – an intuitive idea. After Cauchy, a significant remaining task was to give a precise "algebraic" definition of the limit concept to replace Cauchy's intuitive "kinematic" conception. This was achieved by Weierstrass with his "static" definition of limit in terms of inequalities involving ε and δ (Kleiner, 1991).

In the laboratory, the teacher demonstrated visual representations of the limit concept implied by the expressions $\lim_{x\to 0} R_n(x) = 0$, $\lim_{n\to\infty} R_n(x) = 0$

(where $R_n(x)$ is Lagrange's remainder). This was done by using animations, so as to visualize the "kinematic" viewpoint of Cauchy. The students were asked to describe what they see and to relate it to the syntax of the commands which generate the animations. Importantly, the syntax of the animation commands has a direct relationship with the process described by the definition of limit given by Weierstrass; thus the aim of the animation is to help the students to visualize the process described in the formal definition.

Interpolation Theory: The Works of Lagrange, Cauchy, Chebyshev, and Runge

The lessons on polynomial interpolation followed on from the discussion of approximation using Taylor polynomials. The Taylor polynomial approximation is limited to the points close to a specific point, and the next task was to find a polynomial that provides a "relatively accurate approximation" over an entire interval. The ideas of Lagrange, Cauchy, Chebyshev, and Runge were applied in the laboratory to explain to the students the meaning of a "relatively accurate approximation".

The interpolation problem was formulated in the following way: given the values of a function f(x) at the n + 1 points $x_0, x_1, x_2, \ldots, x_n$, find a polynomial of degree n at most, which coincides with the function f(x) at these points that is, $P_n(x_i) = f(x_i)$, $(i = 0, 1, 2, \ldots, n)$. This problem has a unique solution. In 1795, Lagrange presented his representation of the polynomial, the *Lagrange interpolation formula*, as it appears in Stieltjes (1882):

$$P_n(x) = \sum_{i=0}^{n} \frac{\varphi(x)}{(x - x_i)\varphi'(x_i)} f(x_i), \quad \text{where} \quad \varphi(x) = (x - x_0)(x - x_1)(x - x_2)\ldots(x - x_n)$$ and $\varphi'(x)$ is the first derivative of $\varphi(x)$.

By replacing $\varphi'(x_i) = (x_i - x_0)(x_i - x_1)(x_i - x_2)\ldots(x_i - x_{i-1})(x_i - x_{i+1})\ldots(x_i - x_n)$ in $P_n(x)$, we obtain an alternate form of the Lagrange interpolation formula: $P_n(x) = \sum_{i=0}^{n} \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)} f(x_i)$. This is the form that usually appears in modern textbooks (Davis, 1975).

In the laboratory, the students were introduced to two different approaches to solve the problem of finding the interpolating polynomial through n + 1 points.

In the first approach, the students were asked to write the conditions that must be satisfied in order that the polynomial $P_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$ passes through the n + 1 interpolation points

$((x_i, f(x_i))$ for $i = 0, 1, 2, \ldots, n$. The conditions led to a system of n + 1 equations with n + 1 variables $a_0, a_1, a_2, \ldots, a_n$. Mathematica was used to solve the system of equations for a given number of sample points and a given function f(x).

The students were also requested to prove, in the particular case of $n = 2$, that there is a unique polynomial of degree n that agrees with f at n + 1 points.

The second approach used in the lab is described in Cauchy's *Cours d'Analyse* (1821, pp. 86–89): For x_0, the students were required to construct a polynomial p(x) such that $p(x_0) = f(x_0)$ and $p(x_1) = p(x_2) = p(x_3) = \ldots p(x_n) = 0$. For x_1, the students were required to construct similarly a polynomial q(x) such that $q(x_1) = f(x_1)$ and $q(x_0) = q(x_2) = q(x_3) = \ldots q(x_n) = 0$. The process was repeated with the help of Mathematica for the numerical values of all the points $(x_i, f(x_i))$, $i = 0, 1, 2, \ldots, n$. Then the students added the obtained polynomials $p(x) + q(x) + \ldots$ and could check that the sum of these polynomials is equal to the interpolating polynomial obtained in the first approach. Mathematica helped to generalize the result by using the symbols $x_0, x_1, x_2, \ldots, x_n$ instead of numerical values. And Mathematica was used to plot the function and the Lagrange interpolating polynomial which they had generated. Readers can see in Kidron (2002, Example 6 – Lagrange's Interpolation) the way that students simulated the classical method for finding the polynomial in Mathematica and also the "generalized" method.

The next task was to address the question: how well does the interpolating polynomial approximate the function? Kolmogorov and Yushkevich (1998) report that the first person to estimate the remainder term $R_n(x) = f(x) - P_n(x)$ in the Lagrange interpolation formula was Cauchy in 1840. In the laboratory the students were given Stieltjes' proof (1882) of Cauchy's remainder formula: if $a \leq x_0 < x_1 < \ldots < x_n \leq b$ then

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\lambda)}{(n+1)!}(x - x_0)(x - x_1)\ldots(x - x_n),$$

where $\min(x, x_0, x_1, \ldots, x_n) < \lambda < \max(x, x_0, x_1, \ldots, x_n)$, and λ depends upon $x, x_0, x_1, \ldots, x_n$ and f.

By investigating the quality of an approximation, the students were introduced to Chebyshev's work.

In his *Oeuvres*, Chebyshev (1856, Volume 1, p. 239) points out the fruitfulness of the results obtained as a consequence of making theory and practice closely related. In Interpolation theory, both for the purposes of theory and of practice, it is important to accomplish as much as possible

Pafnuty Lvovich Chebyshev (1821–1894)

Among his mathematical achievements
The quality of an approximation.
The best polynomial approximation (non-constructive existence theorem); Chebyshev polynomials.

Our application in the laboratory
Laboratory activities precede the introduction of theory: building visual pictures in order to understand the formal statement of the theory. Application of Chebyshev's theorem in the lab. Graphical representations of the properties of Chebyshev polynomials.

Figure 8. Application of Chebyshev's ideas in the laboratory.

with a polynomial of a fixed degree. The question is: How do we choose the n + 1 interpolation points such that the approximating polynomial of degree at most n, which passes through these points, will give the best approximation? Can we speak about "the best approximation"?

In the laboratory, the students followed Chebyshev's ideas in order to answer this question. The first task was to define what is "the best approximation". The notion of closeness of approximation over an interval was clarified by taking the maximum deviation between the function and its approximating polynomial. Different approximating polynomials $P_n(x)$ that pass through (n + 1) interpolation points correspond to different choices of these (n + 1) points. The students were told that they have to look for a polynomial $P_n(x)$ for which the maximum deviation in a given interval [a, b] is the smallest of all the maximum deviations of the different interpolating polynomials. Thus, among the interpolating polynomials of degree at most n, $P_n(x) = a_0 + a_1x + \ldots + a_{n-1}x^{n-1} + a_nx^n$ was chosen such that $\max\limits_{a \le x \le b} |f(x) - P_n(x)|$ would be minimized. Then, the following question was asked:

> *Is there a special choice of interpolation points such that the maximum deviation can be minimized?*

The answer to this question is given in the formal statement of the theorem known as Chebyshev's theorem (Davis, 1975, p. 149). This theorem enables one to identify whether a given polynomial is the best approximation polynomial for a given function in a given interval.

The formal statement of Chebyshev's theorem was preceded by laboratory sessions whose purpose was to help students build visual pictures that would help them to understand the formal statement.

Although Chebyshev's theorem enables us to identify whether a given polynomial is the best approximation polynomial for a given function in a given interval, it does not provide an answer to the question: how can the best approximation polynomial be computed numerically?

In the lab, the students applied Chebyshev's criterion in specific cases. Kolmogorov and Yushkevich (1998) point out that in specific problems Chebyshev always sought a procedure that would enable him to construct the solution, rather than merely to verify that a given function is a solution. Thus, the students were requested to consider again Cauchy's remainder formula and to observe that the error has two factors, one that we have no control over, namely $\frac{f^{(n+1)}(\lambda)}{(n+1)!}$, and the other $|(x - x_0)(x - x_1) \ldots (x - x_n)|$. Then the problem under consideration is how to choose the interpolation points such that $|(x - x_0)(x - x_1) \ldots (x - x_n)|$ will be minimized.

This choice does not necessarily lead to the best approximation polynomial, but in many cases these "optimal" interpolations points give good results.

That was the motivation for introducing next the Chebyshev polynomials. The polynomials were defined for $-1 \leq x \leq 1$ as $T_n(x) = 2^{-n+1} \cos(n \cos^{-1} x)$.

The following properties of the Chebyshev polynomials were demonstrated graphically (Figure 9):

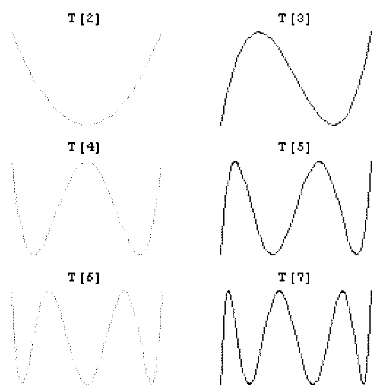


Figure 9. The first Chebyshev polynomials.

$T_n(x)$ has n different zeros in $-1 \leq x \leq 1$. $T_n(x)$ has (n + 1) alternate maxima and minima in this interval. The Chebyshev polynomials possess the "equal ripple property" (as termed by Breuer and Zwas, 1993, p. 154) in the sense that the alternate maxima and minima are of the same size.

Chebyshev's theorem was applied to demonstrate the following important result: if we choose the n + 1 interpolation points x_0, x_1, $\ldots$, x_n to be the zeros of the Chebyshev polynomial $T_{n+1}(x) = (x - x_0)(x - x_1) \ldots (x - x_n)$ then $|(x - x_0)(x - x_1) \ldots (x - x_n)|$ will be minimized for $-1 \leq x \leq 1$. It was shown how this follows from the fact that the Chebyshev polynomials minimize the deviation from 0 on [–1, 1].

The zeros $x_k = \cos(\frac{(k+1/2)\pi}{n})$, $(k = 0, 1, 2, \ldots, n-1)$ of the Chebyshev polynomial $T_n(x)$ were obtained in the laboratory directly by solving the equation $\cos(n \cos^{-1} x) = 0$.

Carle David Runge (1856–1927)

Among his mathematical achievements
Runge's example:
The decisive influence of the choice of the interpolation points upon the quality of an approximation.

Our application in the laboratory
Approximation of functions with equidistant interpolation points and non-equidistant interpolation points.
The following question was asked: If the number of interpolation points increases, is the approximation necessarily better?

Figure 10. Application of Runge's example in the laboratory.

In previous examples, the students noticed that some interpolating polynomials of degree n through equidistant points approached better the given function f(x) for larger values of n. The students might come to expect that, no matter what f(x) represents, with interpolating polynomials of degree n through equidistant points, the error would become small for sufficiently large values of n. Runge's example shows that this is not always true, as was demonstrated in the laboratory as follows: The $f(x) = \frac{1}{1+x^2}$ function was defined on the interval [–5, 5]. The students considered the sequence of interpolating polynomials $P_n(x)$ for the equidistant points $x_0, x_1, x_2, \ldots, x_n$ such that $x_k = -5 + (\frac{10}{n})k$, $(k = 0, \ldots, n)$.

Figure 11 shows f(x) (the black curve) and $P_{12}(x)$ (the dashed black curve) and $P_{24}(x)$ (the gray curve). $P_{12}(x)$ and $P_{24}(x)$ are the interpolating polynomials of degree 12 and 24, respectively. For $|x| > 3.63$, the polynomial $P_{24}(x)$ (through 25 equally spaced points) does not approach f(x) better than $P_{12}(x)$ (through 13 equally spaced points).

The sequence of polynomials $P_n(x)$ was set up, coinciding with the function at (n + 1) points of the interval. The students erroneously thought that $P_n(x)$ would converge to f(x) as $n \to \infty$. The teacher demonstrated by animation how the sequence of interpolating polynomials with equally-spaced interpolation points does *not* converge to f(x) for increasing values of n when $|x| > 3.63$ (Kidron, 2002, Example 4).

Another choice of the interpolation points gave a different result. The interpolation points were chosen this time to be the alternate maxima and minima of the Chebyshev polynomials (Schwarz, 1989, p. 99). Figure 12 shows f(x) (the black curve), $P^*_{12}(x)$ (the dashed black curve), and $P^*_{24}(x)$ (the gray curve).

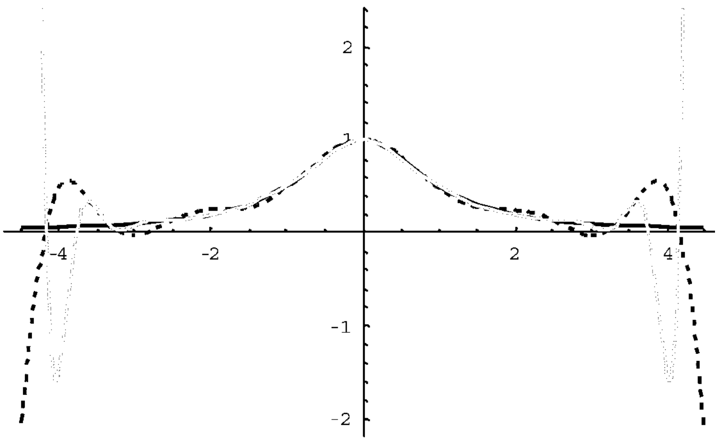


Figure 11. $f(x) = \frac{1}{1+x^2}$ (black) and the interpolating polynomials $P_{12}(x)$ (the dashed black curve) and $P_{24}(x)$ (the gray curve).

$P^*_{12}(x)$ and $P^*_{24}(x)$ are the interpolating polynomials of degrees 12 and 24, respectively, with the new choice of interpolation points. It is difficult to distinguish between f(x) and $P_{24}(x)$ (the gray curve).

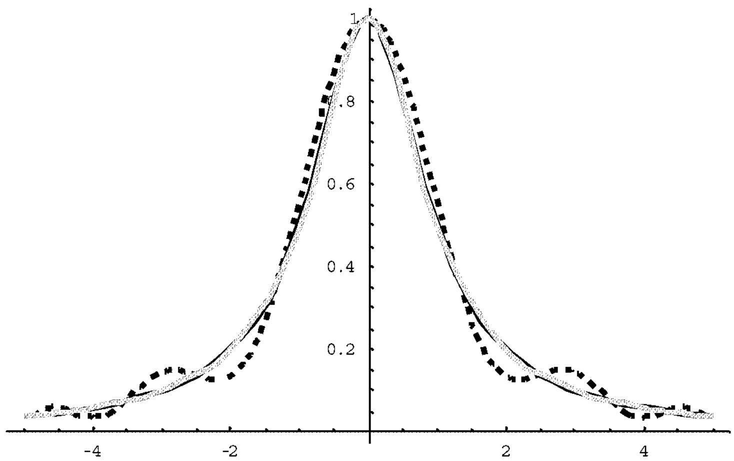


Figure 12. $f(x) = \frac{1}{1+x^2}$ (black) and the interpolating polynomials $P^*_{12}(x)$ (the dashed black curve) and $P^*_{24}(x)$ (the gray curve).

The animation (Kidron, 2002, Example 5) demonstrates how the interpolating polynomials, with this new choice of interpolation points, approach the function f(x) on the interval [–5, 5] for increasing values of n. Runge's example thus underlines the importance of analytical considerations in performing numerical processes.

2.2. *The New Tools: The Role Played by Mathematica in Enabling the Students to "Walk the Same Paths" as the Founders of the Theory*

In this section, we describe the different categories of use of Mathematica and examine the respective part the different categories played in our didactical design.

Following Euler's Algorithmic Thinking

One proposal in this study was to utilize Mathematica's symbolic computation capabilities to allow students to do the calculations while reading Euler (1988, pp. 50–54) directly. Reading Lagrange or Euler has been, until now, difficult for all but the best students. In his *Introductio in Analysin Infinitorum* (1748), Euler elevated symbol-manipulation to an art. His "algebraic analysis" accepted as an article of faith that what is true for polynomials, is true for power series (Kleiner, 1991). Solving the algebraic equations that stem from Euler's procedure is almost impractical without a good Computer Algebra System.

The students were requested to translate Euler's algorithmic thinking into Mathematica commands. We wanted the students to take an active part in the development of their mathematical knowledge. Euler's style is didactical: He explained his methods and shared with the readers of his *Introductio* the "why" of his reasoning. The CAS can perform Euler's algebraic ideas but the students had to make Euler's reasoning very explicit for Mathematica to perform according to his ideas.

Generalizing Results: Writing Results in a Symbolic Way

We were also interested in using Mathematica's symbolic computation to enable students to generalize their results and to write them symbolically. For example, Euler described his algebraic approach when the different examples are written with symbols. Later, he assigned values to the different letters in order to describe a numerical example. In the laboratory, we prefered to begin with the numerical example. Then, Mathematica was used to shift from an example with numbers to an example with symbols: the students had just to substitute symbols instead of numbers into the Mathematica commands they had already written.

Gaining Insight about Infinite Computations

Mathematica's symbolic computation capabilities were also used to gain insight about infinite processes. Some of the Mathematica commands used in following Euler's "development of functions in infinite series" enable one to get the impression of symbolic computation of an infinite number of terms done immediately, requiring "no time"; Mathematica does this by developing a function into infinite power series at once, and usage of the big O notation enhances the impression.

Supporting Alternative Representations of the Same Mathematical Concept

Mathematica's symbolic computation capabilities were also used to support alternative representations of the same mathematical concept, which may help students to become more confident with the possibility of examining a problem from different aspects. We have already mentioned the analytical and algebraic approaches to the Taylor polynomial approximation. As another example, we can cite the two approaches to the problem of finding the Lagrange interpolating polynomial through n + 1 sample points: (1) the modern algebraic approach that led to a system of n + 1 equations, and (2) the approach described in Cauchy's *Cours d' Analyse* (1821). Mathematica enables applying these two different approaches to obtain the same expression for the interpolating polynomial.

Visualizing the Process of Convergence

Dynamic graphics was used to illustrate the variable quantities that were the subjects of Cauchy's "new analysis"; in Euler's book, there are no pictures, nor visual representations of the functions. In the laboratory, Mathematica's dynamic graphical capabilities were used, for example to represent the infinite process of the different approximating polynomials approaching a given function. In this case the animation creates an illusion of completing an ongoing, infinite process.

Animations were thus used to pave the way towards understanding central concepts that involve infinite processes. One proposition in this study was to use Mathematica's dynamic graphical capabilities to help students to *visualize and analyze* the dynamic process of convergence.

Algorithmic Reasoning Required to Generate Dynamic Pictures

Using animations demonstrated by the teacher, the students were required to describe what they see in a dynamic picture. They were also asked to understand the syntax of the animation and to translate visual pictures to analytical language. The students were required also to construct their own animations, and algorithmic reasoning is required to construct an animation command. Different functions may have different analytical properties and in order to visualize their behavior the values of the variable of the animation must be assigned carefully.

Using Mathematica's Programming Features in Interacting with Graphics
The students were asked to interact with the dynamic graphics, changing parameters, choosing different functions, and changing the commands. For example, they were asked to test the convergence of the Lagrange Remainder (displayed as a 3-D animation) for different functions. The possibility of interacting with the software and changing parameters was used to help students realize that a *process* is being represented and not just a specific case. Students were also asked to interact with the graphics in order to build visual pictures that we believed would help them to understand the formal mathematical statements. For example, in the laboratory sessions that preceded the formal statement of Chebyshev's theorem, the students were asked to approximate a given function with a given number of interpolation points. They were asked to explore different possibilities of distribution of the interpolation points (equidistant or different choices of non-equidistant interpolation points). They had also to calculate the different interpolating polynomials. Then they were asked to plot the graphs of the interpolating polynomials and the given function, and to observe the way the different polynomials oscillate around the given function.

3. STUDENTS' REASONING IN THE LABORATORY SESSIONS

High-school students (age 16–17, N = 84) were the participants in the research. The author taught the students mathematics six hours a week; two hours in the PC lab were devoted to the topics of Approximation and Interpolation. The other four hours were devoted to standard subjects in Analysis, Algebra, and Trigonometry.

Each year, one 11th grade class participated in the experimental course during the entire academic year. The course was given four times, so that altogether four classes (N = 84) participated in the experiment. The laboratory consisted of 20 PCs, each equipped with Mathematica and a hardware system (called Classnet) that permits transmitting the content of the screen of one computer to all the computers in the classroom.

In this section, I will try to describe what happened in the class by inviting the reader to get a closer look at what students actually did as they followed the masters. Some episodes describing students' interactions and discussions during the lab sessions, and some findings from a written test, will be presented in order to investigate the two following questions:

1. The students "followed" the masters, equipped with a CAS. To what extent did they come to understand the formal statements of the mathematical theory?
2. As the students experimented with the CAS, did they develop visual intuitions that supported the formal theory?

3.1. *To What Extent Did the Students Come to Understand the Formal Statements of the Mathematical Theory?*

3.1.1. *Episodes Relating to the Interrelationship between Euler's Algebraic Approach and the Analytical Approach*

Students' questions and remarks during the sessions showed that the interrelationship between Euler's algebraic approach and the analytical approach offered new paths to gain a deeper understanding of the mathematical theory. The five episodes that follow demonstrate that when the two points of view, the algebraic and the analytical, were applied with Mathematica, the students were helped in the transition from Euler's intuitive ideas to the formal statement of the theory.

Episode 1

Hanna: "*We got the result* $\frac{1+2x}{1-x-x^2} = 1 + 3x + 4x^2 + 7x^3 + 11x^4 + \ldots$ *in the analytical way when we approximated the function at* $x = 0$ *using the order of contact. Euler is doing approximation for the entire line. How could this be possible? Where in Euler's procedure is it mentioned that the approximation is around* $x = 0$*?*"

In Euler's development of functions in infinite series the series are $A + BZ + CZ^2 + DZ^3 + \ldots$ and in Euler (1988, pp. 50–54) series like $A + B(Z - Z_0) + C(Z - Z_0)^2 + D(Z - Z_0)^3 + \ldots$, with $Z_0 \neq 0$ are not mentioned. Hanna's question relates to the notion of *interval of convergence*. In Mathematica's command "Series", which generates the result of Euler's procedure, it is mentioned of course that the expansion is around $x = 0$. The student's question introduced the possibility of expanding a function in a power series in $(x - x_0)$, where x_0 is a fixed number. The students were interested in substituting another value, for example, 2 instead of 0 in the Mathematica command. They obtained the expansion of a function in power series in (x – 2), where

$$f(x) = f(2) + f'(2)(x - 2) + \frac{f''(2)}{2}(x - 2)^2 + \frac{f^{(3)}(2)}{6}(x - 2)^3 + \ldots$$

The same results, for a given function f(x), could be obtained using the notion of order of contact at $x = 2$. The students used Mathematica to perform the computations, as well as to obtain the graphs of the approximating polynomials to f(x) in the neighborhood of x = 2.

Episode 2

The following discussion underscored the importance of the notion of the interval of convergence. The students were asked if the expansion:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots$$

holds for every x. They answered, "*yes, except for x = 1 for which the denominator is 0*". Then, they were asked to substitute x = 2

Guy: $-1 = \frac{1}{1-2} = 1 + 2 + 2^2 + 2^3 + \dots$ *Why not?*

In the historical development of infinite series, such situations compelled mathematicians to attempt the difficult task of establishing a rigorous foundation underlying the Calculus (Eves, 1981). In our PC lab, a first reaction was the following:

Michael: *We need a graph of* $f(x) = \frac{1}{1-x}$ *and the approximating polynomials. Looking at the graph, we could see where the equality is true.*

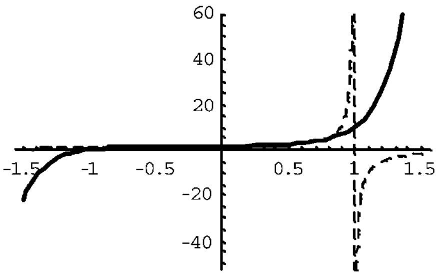


Figure 13. $f(x) = \frac{1}{1-x}$ (the dashed curve) and $P_9(x) = 1 + x + x^2 + x^3 + x^4 + \dots + x^9$ for $-1.5 \leq x \leq 1.5$.

Episode 3

Looking at the graph in Figure 13, the students noticed that the approximation breaks down on the left and on the right. Miri and Guy tried to overcome this "breaking down" at x = –1 by taking more and more terms of the expansion. The graphical and numerical capabilities of Mathematica were used in their argument.

Guy: *The higher the degree of the approximating polynomial, the bigger is the interval in which f(x) and the polynomial coincided; the order of contact is bigger.*

Miri: *We took polynomials of bigger and bigger degrees and looked at the graph of f(x) and the approximating polynomials. Even the Taylor polynomial of degree 72 breaks down at x = –1!*

Episode 4

Some students expanded $f(x) = \frac{1}{1+x^2}$ into the series

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \ldots$$

They thought that the equality is true for every x, since there is no x for which the denominator of f(x) is 0 (they worked with real numbers). They looked at the graph of f(x) and the approximating polynomials, and they were surprised to see that even in this example "the approximation breaks down" for $|x| \geq 1$. (One year later, the students studied complex numbers as part of an algebra course, and then they could learn the full explanation for why $\frac{1}{1+x^2}$ cannot be represented by the series $1 - x^2 + x^4 - x^6 + \ldots$ at the point $x = 3$, for example.)

Episode 5

The two modes of thought, analytic and algebraic, enabled the students to generalize their results. The two approaches produce the coefficients of the Taylor series in terms of the function and its derivatives in the following way: A function f is expanded in a power series at $x = 0$: $f(x) = a_0 + a_1x + +a_2x^2 + a_3x^3 + \ldots + a_nx^n + \ldots$ Assuming that the first n derivatives exist and that the polynomial of degree n $P_n(x) = a_0 + a_1x + +a_2x^2 + a_3x^3 + \ldots + a_nx^n$ has an order of contact n with f(x) at $x = 0$, the students were asked to find $a_0, a_1, a_2, a_3, \ldots, a_n$ in terms of the function and its derivatives. The students began with specific cases ($n = 3, n = 4, \ldots$) and then they generalized for n.

$$f(0) = P_n(0) = a_0;\ f'(0) = P'_n(0) = a_1;\ f''(0) = P''_n(0) = 2a_2;\ f^{(3)}(0) = P_n^{(3)}(0) = 6a^3; \ldots$$

Continuing this procedure, they found the general formula $a_n = \frac{f^{(n)}(0)}{n!}$.

After the students obtained the formula $f(x) = f(0) + xf'(0) + \frac{x^2f''(0)}{2!} + \frac{x^3f^{(3)}(0)}{6} + \ldots$ they realized that they were not just working with a specific case, but rather with a process – a general property:

Guy: *I looked at the animation that represents the different polynomials that approximate* $f(x) = \frac{1+2x}{1-x-x^2}$. *You just have to substitute another function instead of f(x) and you get directly its expansion in power series and the animation representing the different polynomials that approximate the new function.*

Episodes Relating to Error Analysis: Lagrange's Remainder Formula and Cauchy's Remainder Formula

The founders of the theory pointed out the importance of analyzing the error that occurs in applying the numerical process of approximation. In this section, we will follow the students' reasoning in the context of error analysis.

When approximating a function f(x) by means of Taylor polynomials $P_n(x)$, the error $f(x) - P_n(x)$ for a given value x_0 of x is $\frac{f^{(n+1)}(c) \cdot x_0^{n+1}}{(n+1)!}$ (Lagrange's remainder) for some c value between 0 and x_0. I noticed that the students were surprised that there exists a general expression for the error $f(x) - P_n(x)$, and they expressed some difficulties about the unknown c.

The Unknown c in Lagrange's Remainder Formula (Approximation Theory): Episode 6

Hava: "*What is the value of c? I want to know the value of the error at a given point. Why did I need an upper estimate of the error?*"

Hava observed the 3-dimensional graphics that represent the upper estimate of the absolute value of the error as a function of the two variables x and c and asked:

Did the value of the error at a given point x_0 and a given n $(f(x_0) - P_n(x))$ *equal* $\frac{f^{(n+1)}(c) \cdot x_0^{n+1}}{(n+1)!}$ *for a value of c that gives the maximal error? You can describe the error at x_0 as a function of different values of c ($0 < c < x_0$). But the error at x_0 is a single numerical value. What is the value of c that gives this numerical value? Is it the value of c for which* $\frac{f^{(n+1)}(c) \cdot x_0^{n+1}}{(n+1)!}$ *has its maximum value? I look at the error at a given point not in an interval.*

Michaël: "*No! It is not necessary the value of c for which the error gets its maximum value*".

The students observed the animation where the degree of the Taylor polynomial is varied (and the domain is fixed).

Itaï: "*If the degree will be very big, could it be that there will be no error in this fixed domain? I mean is there a degree of the polynomial for which there is no error?*"

Hava: "*There is no error when the degree of the approximating polynomial is infinite. The question is what is infinite? Infinite is not something that can be defined. 'The problem with Taylor' is that the formula says that there will be always an error*".

The Unknown λ in Cauchy's Remainder Formula (Interpolation Theory): Episode 7

Cauchy's formula is

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\lambda)}{(n+1)!}(x - x_0)(x - x_1)\ldots(x - x_n), \qquad (a \leq x_0 < x_1 < \ldots < x_n \leq b),$$

where $x_0, x_1, \ldots, x_n$ are the n + 1 interpolation points in [a, b] and $P_n(x)$ is the Lagrange interpolating polynomial of degree n.

In the lab sessions the students began with cases of functions that are polynomials of degree n + 1, so that $f^{(n+1)}(\lambda)$ is a constant. In these cases, the students had no problem with the unknown λ; the students concluded that the best approximating polynomial passes through interpolation points $x_0, x_1, \ldots, x_n$, such that $(x - x_0)(x - x_1)\ldots(x - x_n)$ gets its minimal value The difficulties arose with other functions that are not polynomials of degree n + 1, so that $f^{(n+1)}(\lambda)$ is not a constant.

Hava: "*For a given differentiable function f and for n + 1 given interpolation points* $x_0, x_1, \ldots, x_n$*, we obtain λ that depends on f and on the n + 1 points* $x_0, x_1, \ldots, x_n$*. As a consequence,* $\frac{f^{(n+1)}(\lambda)}{(n+1)!}$ *is one single numerical value for* this λ*. My question is why, for a given f and given* $x_0, x_1, \ldots, x_n$ *the plot of* $f(x) - P_n(x)$ *is not expected to behave exactly as* $(x - x_0)(x - x_1)\ldots(x - x_n)$"?

Hava's question enables us to get a deeper understanding of the meaning of the sentence λ *depends on x*, on f, and on the n + 1 points $x_0, x_1, \ldots, x_n$. Even in the case where $x_0, x_1, \ldots, x_n$ are given, $\frac{f^{(n+1)}(\lambda)}{(n+1)!}$ has not a single numerical value. It has different values as a function of x and as a function of f. It depends on both!

Testing Cauchy's Remainder Formula

The following test (Figure 14) was used to check the students' ability to use Mathematica's numerical and graphical capabilities to estimate the maximal error that is created when we replace a given function f(x) with the Interpolating Polynomial $P_n(x)$.

Were the students able to use Cauchy's remainder formula (with the help of Mathematica) in order to give a numerical value to the maximal error? In contrast with Chebyshev's theorem, which was taught in an experimental way (as we will see later in Episode 8), Cauchy's remainder formula was taught in an analytical way (it was proved by means of Rolle's theorem) and afterward the students applied it in the lab. In the previous episodes we observed the students' difficulties with the unknown λ. The following test examines the students' understanding of the notion 'an upper estimate of the error' and their ability to overcome difficulty with the unknown λ in Cauchy's remainder formula. Moreover, the students were required to use this formula and Mathematica to give a *numerical estimate* for the maximal error.

Question:
Let $f(x) = \sin(x)$ be defined for $0 \le x \le \frac{\pi}{2}$.
The x-values of the interpolation points are $x_0 = 0$; $x_1 = \frac{\pi}{10}$; $x_2 = \frac{\pi}{5}$; $\ldots$; $x_5 = \frac{\pi}{2}$ and $P_5(x)$ is the interpolating polynomial that passes through these points.
What is the maximal error created when we compute $P_5(\bar{x})$ instead of $\sin(\bar{x})$ for $0 \le \bar{x} \le \frac{\pi}{2}$? In order to answer the question:
1. Use Cauchy's remainder formula.
2. Use the graph of the error function $P_5(x) - \sin(x)$ for $0 \le x \le \frac{\pi}{2}$. Compare your results.

Figure 14. Test 2.

This test was given each year for two years, and two classes (N = 40) attempted it. The results of the two classes were almost identical, so we will discuss the results of the two classes together.

Most of the students (87.7%) were able to use the software in order to get the interpolating polynomial that approximates the given function.

Moreover, 80.3% were able to use the graphical capabilities of Mathematica to find the maximal error expressed by the difference between the function and the interpolating polynomial.

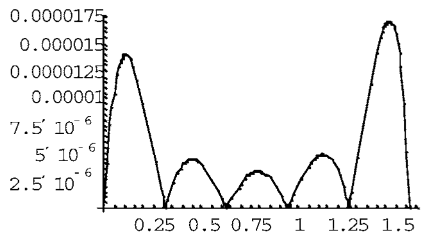


Figure 15. The graph of the error function $P_5(x) - \sin(x)$ for $0 \le x \le \frac{\pi}{2}$.

The students wrote that the maximal error is approximately 0.0000175. A smaller number of students (67.5%) succeeded in overcoming the difficulty with the unknown λ in Cauchy's remainder formula: these students found that in the specific example, $f(x) = \sin(x)$, the different derivatives are bounded by 1 and the term $\frac{f^{(n+1)}(\lambda)}{(n+1)!}$ in Cauchy's remainder satisfies the inequality $\frac{f^{(n+1)}(\lambda)}{(n+1)!} \le \frac{1}{(n+1)!}$; therefore, the estimate of the maximal error is given by $\frac{1}{6!}x(x - \frac{\pi}{10})(x - \frac{\pi}{5}) \ldots (x - \frac{\pi}{2})$. Then, using graphical or analytical considerations, the students gave a numerical estimate for the upper bound for the maximal error. Some students generated the plot of the function (or the absolute value of the function) $e(x) = \frac{1}{6!}x(x - \frac{\pi}{10})(x - \frac{\pi}{5}) \ldots (x - \frac{\pi}{2})$

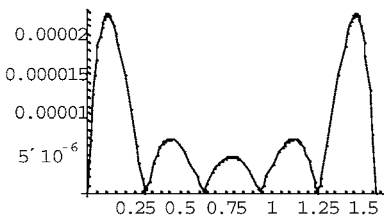


Figure 16. The graph of $\frac{1}{6!}x(x - \frac{\pi}{10})(x - \frac{\pi}{5}) \ldots (x - \frac{\pi}{2})$.

and wrote that the maximal error is approximately 0.00002. Other students applied analytical considerations: they were helped by Mathematica to solve the equation $e'(x) = 0$ and obtained the five solutions corresponding to the five extrema. They computed the value of e for the solution $x = 1.46506$ and obtained 0.0000225671.

Only those students (67.5%), who made analytical considerations (estimating $f^{(n+1)}(\lambda)$ without knowing λ), were able to apply the theory in the laboratory and to use the graphical capabilities of Mathematica in order to assign a numerical value to the maximal error given by Cauchy's remainder formula. We observed that the students (32.5%) who did not overcome the difficulty with the unknownλ by applying analytical considerations also did not use the graphical capacities of Mathematica to find the upper estimate of the error.

Although no-one did, it is worth remarking that the students could have worked in a way similar to the way the upper estimate of Lagrange's Remainder formula was obtained, using 3-dimensional graphics. The error $(f(x) - P_5(x))$, that is Cauchy's remainder formula, is $(\frac{f^{(6)}(\lambda)}{6!})x(x - \frac{\pi}{10})(x - \frac{\pi}{5}) \ldots (x - \frac{\pi}{2})$ for some λ value between 0 and $\frac{\pi}{2}$. Thus the students could have plotted the following 3-dimensional graphics (Figure 17) that represents the absolute value of the error as a function of x and λ with $0 \le x \le \frac{\pi}{2}$, $0 \le \lambda \le \frac{\pi}{2}$. The λ value in $0 \le \lambda \le \frac{\pi}{2}$ that corresponds to the exact error is unknown; therefore, they could have looked at all pairs (x, λ) such that $0 \le x \le \frac{\pi}{2}$, $0 \le \lambda \le \frac{\pi}{2}$ and obtained 0.00002 as an upper estimate of the error (Figure 17).

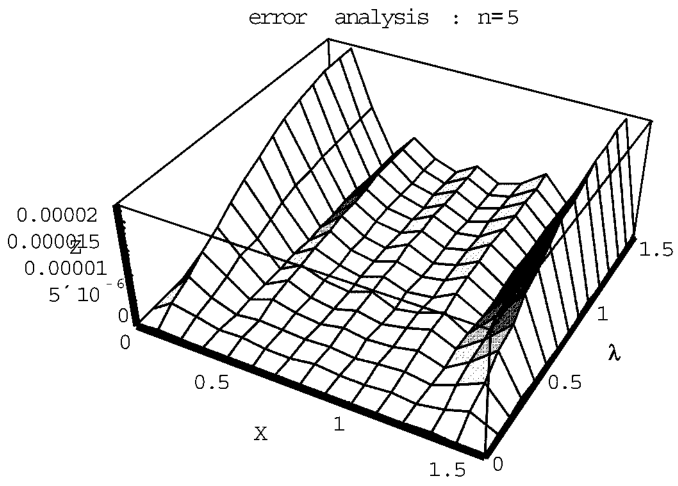


Figure 17. The error as a function of x and λ.

3.2. *Did the Students Develop Visual Intuitions that Supported the Formal Theory?*

Visualization of the Chebyshev Theorem: Episode 8

As a consequence of experimenting in the lab, it was evident that the students could visualize theory before a formal statement was given, for example in the laboratory sessions that dealt with the quality of approximation where the formal statement of the Chebyshev theorem was given afterwards.

In trying to find the best polynomial approximation, the students investigated whether there is a special choice of interpolation points and they were led in an experimental way to the best polynomial approximation (cf. Breuer and Zwas, 1993).

In the laboratory, the students were asked to approximate $f(x) = \sin(x)$ in the interval [–1, 1], based on 9 interpolation points. As a first choice, the students worked with equidistant interpolation points. They plotted the remainder $R_n(x) = f(x) - P_n(x)$ and noticed that the maximum deviations occurred at the ends of the interval. So, they decided to take more points at the ends of the interval, and fewer at the center. They observed the way the polynomials oscillate around f(x) and reached the conclusion that the polynomial that oscillates around f(x) in a way that smears the error uniformly will minimize the deviation. For example, by plotting the graph of the absolute value of the error (Figure 18), Michael suggested:

the biggest error is at the ends of the interval . . . if we will choose more points at the ends and less points at the center, the graph will be balanced. In the next plot [Figure 19] *the biggest error is* 1.5×10^{-8}.

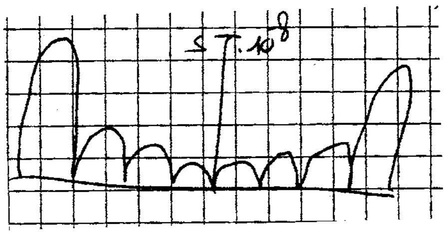


Figure 18. sin(x) – p(x) for –1 < x < 1 (equidistant interpolation points).

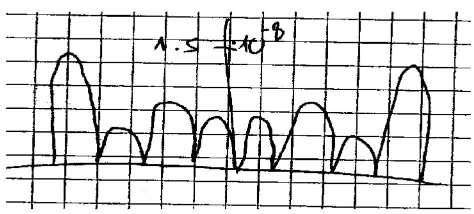


Figure 19. sin(x) – p(x) for –1 < x < 1 (non-equidistant interpolation points).

Michael concluded: "*The graph of the error for the optimal polynomial will have the same maximal height for all the tops*". (Figure 20)

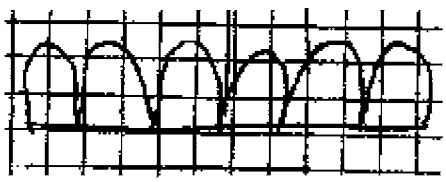
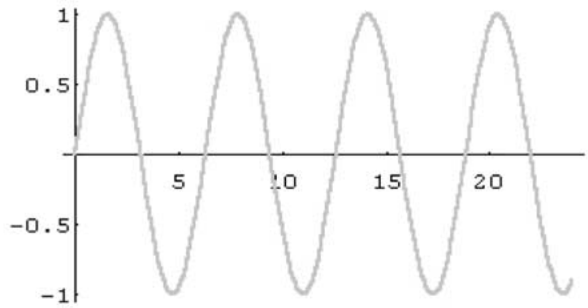


Figure 20. The graph of the error for the *optimal* polynomial.

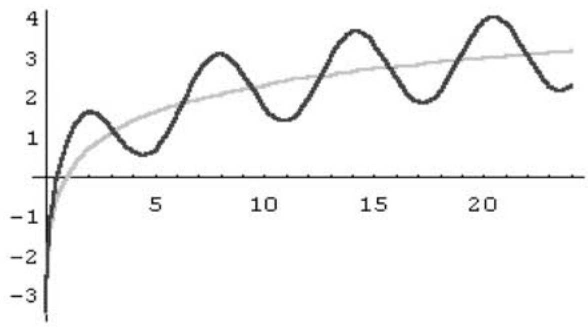
Another student gave a similar answer:

Giora: "*When the interpolation points are equidistant, we see in the graph that the error at the ends of the interval is bigger than the error at the middle of the interval. Our aim is to change the distribution of the interpolation points so that they will be denser toward the ends of the interval in such a way that the graph of the error will have the same height in the whole domain*".

The following figure (Figure 21) represents the pictures that other students proposed for the error f(x) – p(x) where p(x) is the best interpolating polynomial, in particular regarding the way that p(x) oscillates around f(x). Note that these figures were developed *before* the students were given the statement of Chebyshev's theorem.



(a)



(b)

Figure 21. The “equal ripple” property of the error: alternate maxima and minima of the same size.

In Figure 21(a) we see the graph of the remainder $R_n(x) = f(x) - P_n(x)$; this has a minimum-maximum absolute value that is spread uniformly on the interval. In (b), we see how $P_n(x)$ oscillates around f(x) in a way that distributes the error uniformly.

By plotting these graphs, the students demonstrated that they understood the importance of the “equal ripple” property (alternate maxima and minima of the same size) of the error for the best approximation. The students built visual pictures and tried to translate those pictures into mathematical expressions, and this helped them to understand the formal statement of the Chebyshev theorem. They called $P_n(x)$ the best approximating polynomial to the function f(x) in the interval [a, b] and they translated the description “points with alternate maxima and minima deviation of the same size” into the expression $f(x_j) - P_n(x_j) = (-1)^j \hat{e}$ for $j = 0, 1, 2, \ldots$ and $\hat{e} = \max_{a \leq x \leq b} |f(x) - P_n(x)|$. They could not answer the question about how many points x_j with alternate maximum and minimum deviations there are (the answer is n + 2). Some students thought there are n such points, and others thought n + 1. As a consequence of one student's question, "how do you know if the first deviation is positive or negative?", they modified their expression $f(x_j) - P_n(x_j) = (-1)^j\hat{e}$ to $f(x_j) - P_n(x_j) = \alpha(-1)^j\hat{e}$ with $\alpha = \pm 1$.

4. DISCUSSION AND CONCLUSIONS

In this section we examine how our didactical choices influenced students' mathematical reasoning. We observed the students' ability to use tools provided by a CAS, Mathematica, in order to implement the ideas of the founders of the theory of approximation and interpolation. The students had to translate the mathematicians' reasoning into explicit Mathematica commands to perform their ideas, and this historical approach was introduced as another representation in addition to modern representations.

In Kidron (2001) we analyzed the students' ability to ascertain that the algebraic and analytical approaches are two different representations of the same subject. The comparison of Euler's historical algebraic approach and the analytical approach enabled the students to ask important questions about the intuitive idea of developing a function in infinite series. Questions about the interval of convergence (described in Episodes 1, 2, and 3) indicated that Eulers' *Introductio in Analysin infinitorum* did not contain this concept. Realizing this was the beginning of the students' transition from intuitive ideas to the formal statement of the theory. However, we could not give complete answers to the students' questions, since a complete understanding of the formal theory must come later, in university-level mathematics. In the visual representations, there were hints about results that will be proved later. The CAS helped introduce the concept of interval of convergence, and the graphical representations could be used to find where "the approximation approximately breaks down on the left and the right" (Episodes 3 and 4). The theoretical justification for this will come at the appropriate time when the students study a course on "complex variables". In the meantime, it is evident they were motivated to learn more about the mathematical theory. Episode 5 demonstrates that there is a possibility to interact with the software, to change parameters, to help students to generalize their results, so that they can appreciate that a process is represented and not just a specific case.

Tall (2000b) points out that computer environments are particularly valuable in encouraging experimentation that helps, before any formal theory is developed, to give a sense of a given phenomenon and to suggest what kinds of properties are involved. This was the case with the laboratory sessions dealing with the quality of approximation that preceded the formal statement of Chebyshev's theorem (Episode 8). Experimenting in the lab, the class had to find the nodes that minimized the error in the Lagrange interpolation, and from this they developed the visual intuitions necessary to understand the formal statement of the theorem. Although the "cancelling out" of the deviations from the graph had a powerful visual impact, the purpose of the task was really to minimize the error given by Cauchy's formula $f(x) - P_n(x) = \frac{f^{(n+1)}(\lambda)}{(n+1)!}(x - x_0)(x - x_1)\ldots(x - x_n)$. Therefore, we were interested in observing the way that the students managed with the unknown λ in Cauchy's formula. In Episodes 6 and 7, we observed the students' difficulties with the unknowns c (in Lagrange's remainder formula) and λ (in Cauchy's remainder formula). The remainder formula tells us that there "exists a c (respectively, a λ) such that ...". It also tells us that c is located between 0 and x (respectively, $\min(x, x_0, x_1, \ldots, x_n) < \lambda < \max((x, x_0, x_1, \ldots, x_n))$, and that is all it has to say. The students understood that they were concerned not only with the existence of the c or the λ, but also that they had to give numerical answers for the error analysis. This was a source of difficulties. Moreover, the students had to understand the importance of estimating the error even if the c (or the λ), which gives the exact error, remains unknown.

In Episodes 6 and 7, we realized that the students became sensitive to very important questions and that they could get some ideas about possible answers to these questions. In the test that deals with Cauchy's remainder formula, we investigated the students' ability to apply the theory in the laboratory. More specifically, we investigated the students' ability to use the knowledge about the *existence* of the *unknown* λ in order to *calculate* Cauchy's remainder. In contrast to Chebyshev's theorem, Cauchy's remainder was introduced before the lab sessions, and we examined how the students applied Cauchy's formula in the lab. The test was designed in such a way that the students were forced to apply both numerical and graphical considerations, and to think analytically (in order to estimate an upper bound for $f^{(n+1)}(\lambda)$ without any calculation); the results showed the students' ability to use Mathematica's numerical and graphical capabilities, but we observed that the ability to use these capabilities did not help much those students who had not succeeded in performing the analytical reasoning.

By experimenting in the lab, visual intuitive ideas are formed. It is evident that intuition is important, but it must be accompanied by careful control, based on analytical reasoning. As Tall (2000b) mentions, "students learn by building up mental images in ways that are consistent with what they do and what they observe using the technology". The students experimented in the lab using different cases of approximating a function in a given interval by means of interpolating polynomials through equidistant points. They noticed that in their examples the interpolating polynomials better approached the given function for bigger n values. Therefore, they could have erroneously concluded that "the more equidistant points that are used, the better is the approximation". In order to overcome such misleading insights, the students should be given different, well-chosen examples (sometimes with conflicting images), in order to help them build *new* "mental images that will also be consistent with what they do and observe using the technology". Applying Runge's historical example in the laboratory exemplifies how to overcome misleading insights. The students compared different choices of interpolation points using dynamic graphics. Applying different discrete methods enabled the students to better understand the continuous methods.

The history of mathematics provides examples of intuitive ideas that are confusing because they were not established rigorously. Nevertheless, those intuitive ideas paved the way for establishing the foundation of the formal theory. Euler's intuitive idea in 1748 to express functions as polynomials with "an infinite number of terms" is such an example. The history of mathematics shows that Euler's use of the infinite was eventually proved to be consistent. Weierstrass' approximation theorem of 1886 (i.e., a continuous function is, on a closed interval, equal to a uniformly convergent series of polynomials) makes precise the analytical expressions of Euler, reviving his polynomials of infinite degree (Laugwitz, 1994).

In a technology-equipped calculus laboratory we attempted to learn from the historical process: intuitive ideas were formed and the students gained informal insight into the mathematical theory. The history of mathematics whispers into the students' ears the message: *Your mental pictures are valuable. They may need to be refined, but you can build on them.*

ACKNOWLEDGMENTS

Most of the material in this paper is based upon my Ph.D. thesis, which was supervised by Professor Michael Maschler of the Hebrew University in Jerusalem. His ideas and guidance were very important and encouraging.

I also thank Professor Tommy Dreyfus of Tel Aviv University who carefully checked the original version of the article and improved it in structure, content, and style.

NOTES

[1] This paper was prepared whilst the author was a postdoctoral fellow at the Department of Science Teaching, Weizmann Institute of Science, Rehovot, Israel.
[2] My approach of using primary sources is not new; see Arcavi et al (1987), Van Maanen (1997), Arcavi and Bruckheimer (2000).

REFERENCES

Arcavi, A., Bruckheimer, M. and Ben-Zvi, R. (1987). History of mathematics for teachers: the case of irrational numbers. *For the Learning of Mathematics* 7: 2.

Arcavi, A. and Bruckheimer, M. (2000). Didactical uses of primary sources from the history of mathematics. *Themes in Education* 1(1): 55–74.

Artigue, M. (2001). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *Paper presented at CAME 2001*, Freudenthal Institute, University of Utrecht. http://ltsn.mathstore.ac.uk/came/events/freudenthal/index.html

Breuer, S. and Zwas, G. (1993). *Numerical Mathematics – A Laboratory Approach*. Cambridge: Cambridge University Press.

Brown, D.P., Porta, H. and Uhl, J. (1991). *Calculus & Mathematica.* Reading, MA: Addison-Wesley Publishing Company.

Cauchy, A.L. (1821). *Cours d' Analyse: Analyse Algebrique*. Paris: De Bure.

Chebyshev, P.L. (1856). Sur la Construction des Cartes Geographiques. *Oeuvres de P.L. Tchebycheff Publiées par les Soins de mm.A. Markoff et N. Sonin*, Vol. 1. New York: Chelsea Publishing Company.

Cornu, B. (1981). Apprentissage de la notion de limite: Modèles spontanés et modèles propres. *Proceedings of the 5th International Conference for the Psychology of Mathematics Education* (pp. 322–329). Grenoble, France.

Cornu, B. (1991). Limits. In D. Tall (Ed.), *Advanced Mathematical Thinking* (pp. 153–166). Dordrecht: Kluwer Academic Publishers.

Cottrill, J. and Dubinsky, E. (1996). Understanding the limit concept: Beginning with a coordinated scheme. *The Journal of Mathematical Behavior* 15: 167–192.

Courant, R. and Robbins, H. (1941). Revised by Stewart, I. 1996. *What is Mathematics?* Oxford: Oxford University Press

Davis, P.J. (1975). *Interpolation and Approximation.* New York: Dover Publications.

Davis, R.B. and Vinner, S. (1986). The notion of limit: Some seemingly unavoidable misconception stages. *The Journal of Mathematical Behavior* 5: 281–303.

Dreyfus, T. and Hillel, J. (1998). Reconstruction of meanings for function approximation. *International Journal of Computers for Mathematical Learning* 3: 93–112.

Euler, L. (1748). *Introductio in Analysin Infinitorum*. Tomus Primus, Lausannæ, Apud Marcum-Michaelem Bousquet & Socios.

Euler, L. (1988). *Introduction to Analysis of the Infinite*, Vol. I (English translation by John Blanton). New York: Springer-Verlag.

Eves, H. (1981). *Great Moments in Mathematics*, Vol. 2 (after 1650). Washington: MAA.

Kidron, I. (1999). Approximation and Interpolation, Calculus Laboratory, From visual interpretation to formal reasoning. Unpublished Ph.D. Thesis. Hebrew University, Jerusalem, Israel.

Kidron, I. (2001). Teaching Euler's algebraic methods in a Calculus laboratory. *Proceedings of the 12th ICMI Conference*, Vol. 2 (pp. 368–376). Melbourne, Australia.

Kidron, I. (2002). http://stwww.weizmann.ac.il/g-math/Approximation-examples.html

Kleiner, I. (1991). Rigor and proof in mathematics: A historical perspective. *Mathematics Magazine* 64: 291–314.

Kline, M. (1972). *Mathematical Thought from Ancient to Modern Times*, Vol. 2. Oxford: Oxford University Press.

Kolmogorov, A.N. and Yushkevich A.P. (1998). *Mathematics of the 19th Century*. Birkhaüser Verlag, Basel.

Lagrange, J.L. (1884). *Oeuvres*. publiées par les soins de M.J.A Serrat, Vol. X, Paris, Gauthier-Villars.

Laugwitz, D. (1994). Real-variable analysis from Cauchy to non-standard analysis. In I. Grattan-Guinness (Ed.), *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, Vol. 1 (pp. 318–330). Routledge.

Schwarz, H.R. (1989). *Numerical Analysis, a Comprehensive Introduction*. New York: Wiley.

Stieltjes, T.J. (1882). *Oeuvres Complètes*, Vol. 1, publiées par les soins de la Société Mathématique d' Amsterdam, Groningen 1914–1918.

Tall, D. (1992). The transition to advanced mathematical thinking: Functions, limits, infinity and proof. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 495–511). New York: Macmillan.

Tall, D. (1993). Real mathematics, rational computers and complex people. *Proceedings of the Fifth Annual International Conference on Technology in College Mathematics Teaching* (pp. 243–258).

Tall, D. (2000a). Technology and versatile thinking in mathematical development. In Michael O.J. Thomas (Ed.), *Proceedings of Time 2000* (pp. 33–50). Auckland, New Zealand.

Tall, D. (2000b). Cognitive development in advanced mathematics using technology. *Mathematics Education Research Journal* 12(3): 210–230.

Van Maanen, J. (1997). New maths may profit from old methods. *For the Learning of Mathematics* 17(2): 39–46.

Wolfram, S. (1999). *The Mathematica Book*, 4th edition. New York: Cambridge University Press.

Department of Applied Mathematics
Jerusalem College of Technology
Jerusalem, Israel
E-mail: ivy@jct.ac.il or ivy.kidron@weizmann.ac.il