International Journal of Mathematical Education in Science and Technology, Vol. 38, No. 4, 15 June 2007, 529–569

Taylor & Francis
Taylor & Francis Group

Classroom notes

Constructing a line segment whose length is equal to the measure of a given angle

W. JACOB and T. J. OSLER*
Rowan University, Glassboro, NJ, USA

(*Received 14 February 2006*)

Given an angle α (in radians), can we construct (using straight edge and compass) a straight line with length α? The answer is no. If we could construct a line length of π, then we could square the circle, which we know is impossible. However, we can construct approximations to the length α with arbitrary accuracy. It is the purpose of this paper to demonstrate a simple way to make this construction.

In figure 1 we see the angle α made by the x-axis (OL_∞) and the line OL_0. The circle shown has diameter 1 and equation $r = \sin\theta$. The line OL_0 meets the circle at point P_0. Line OL_1 bisects angle α Construct line OL_2 so that it bisects angle L_1OL_∞. Continuing we construct line OL_3 so that it bisects angle L_2OL_∞, etc.

From point P_0 we construct a line perpendicular to line OL_0 meeting line OL_1 at P_1. From P_1 we construct a line perpendicular to line OL_1 meeting line OL_2 at P_2. We continue in this way forming additional points P_3, P_4, etc. We will show that the lengths of the line segments $OP_0, OP_1, OP_2, \ldots$ converge to α.

We now pause for some trigonometric manipulation. Repeated use of a familiar trigonometric identity gives us

$$\begin{aligned}
\sin\alpha &= 2\cos\frac{\alpha}{2}\sin\frac{\alpha}{2}\\
\sin\alpha &= 2^2\cos\frac{\alpha}{2}\cos\frac{\alpha}{2^2}\sin\frac{\alpha}{2^2}\\
\sin\alpha &= 2^3\cos\frac{\alpha}{2}\cos\frac{\alpha}{2^2}\cos\frac{\alpha}{2^3}\sin\frac{\alpha}{2^3}
\end{aligned}$$

Continuing this way and dividing by α we get

$$\frac{\sin\alpha}{\alpha} = \frac{\sin(\alpha/2^N)}{\alpha/2^N}\prod_{k=1}^{N}\cos\frac{\alpha}{2^k}$$

or taking the reciprocal we get

$$\frac{\alpha}{\sin\alpha} = \frac{\alpha/2^N}{\sin(\alpha/2^N)}\prod_{k=1}^{N}\sec\frac{\alpha}{2^k} \tag{1}$$

*Corresponding author. Email: Osler@rowan.edu

International Journal of Mathematical Education in Science and Technology
ISSN 0020–739X print/ISSN 1464–5211 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00207390601002849

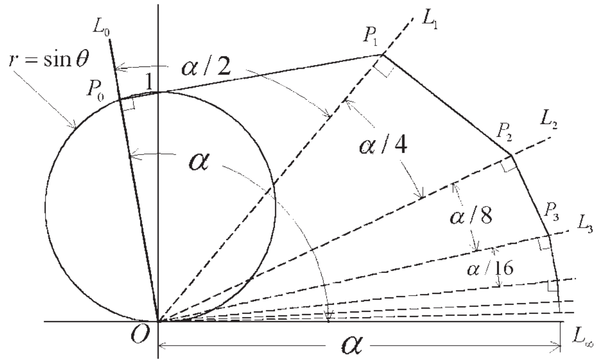

Figure 1. Constructing a line of length α.

Finally we let N tend to infinity so that (1) now takes the form

$$\alpha = \sin\alpha \prod_{k=1}^{\infty} \sec\frac{\alpha}{2^k} \tag{2}$$

Since the equation of the circle in figure 1 is $r=\sin\theta$, we see that

$$\overline{OP_0} = \sin\alpha$$

and

$$\overline{OP_1} = \overline{OP_0}\sec\left(\frac{\alpha}{2}\right) = \sin\alpha\sec\left(\frac{\alpha}{2}\right)$$

$$\overline{OP_2} = \overline{OP_1}\sec\left(\frac{\alpha}{4}\right) = \sin\alpha\sec\left(\frac{\alpha}{2}\right)\sec\left(\frac{\alpha}{4}\right)$$

etc. It is now clear from (2) that our constructions converge to α. Even though our constructions approach α as a limiting case, we cannot say that we have constructed α since limiting cases are not permitted in classical geometric construction.

This construction shown here is a simple generalization of the construction shown in [1] for the specific angle $\alpha=\pi/2$. In that paper the construction is related to Vieta's famous product

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}}\cdots$$

Reference

[1] Osler, T.J., Geometric constructions approximating π related to Vieta's product of nested radicals. To appear in *The Mathematical Spectrum*.

Generating functions for the powers of Fibonacci sequences

D. TERRANA† and H. CHEN*

Department of Mathematics, Christopher Newport University, Newport News, VA 23606

(*Received 25 April 2006*)

In this note, based on the Binet formulas and the power-reducing techniques, closed forms of generating functions for the powers of Fibonacci sequences are presented. The corresponding results are extended to some other famous sequences as well.

1. Introduction

The *Fibonacci sequence* $\{F_n\}$ [1, 4] is defined by

$$F_1 = F_2 = 1;$$

$$F_n = F_{n-1} + F_{n-2} \quad (\text{for } n \geqslant 3). \tag{1}$$

It is well known that the *generating function* for the Fibonacci sequence is given by

$$\sum_{n=1}^{\infty} F_n x^n = \frac{x}{1-x-x^2}. \tag{2}$$

In general, let the generating function of $\{F_n^k\}$ be

$$G_k(x) = \sum_{n=1}^{\infty} F_n^k x^n.$$

The following homogeneous recurrence relations

$$F_{n+3}^2 = 2F_{n+2}^2 + 2F_{n+1}^2 - F_n^2,$$
$$F_{n+4}^3 = 3F_{n+3}^3 + 6F_{n+2}^3 - 3F_{n+1}^3 - F_n^3$$

enable us to find

$$G_2(x) = \sum_{n=1}^{\infty} F_n^2 x^n = \frac{x(1-x)}{(1+x)(1-3x+x^2)},$$
$$G_3(x) = \sum_{n=1}^{\infty} F_n^3 x^n = \frac{x(1-2x-x^2)}{(1+x-x^2)(1-4x-x^2)}.$$

---

*Corresponding author. Email: hchen@cnu.edu

†Graduated from Christopher Newport University with B.S. in Mathematics last May; currently working in Wasington, D.C.

For example, the detail derivation of $G_3(x)$ is given in [1, pp. 227–228]. However, this path is hard to continue since the similar homogeneous recurrence relations for $k \geq 4$ are unknown. In 1962, Riordan [3], instead of giving a closed form, found that $G_k(x)$ satisfies the recurrence relation

$$(1 - a_k x + (-1)^k x^2) G_k(x) = 1 + kx \sum_{i=1}^{[k/2]} (-1)^i a_{ki} G_{k-2i}((-1)^i x)/i,$$

where a_{ki} have a complicated structure. Recently, Mansour [4] gave a closed form in terms of determinants of two complicated $k \times k$ matrices, which is difficult to calculate even for $k = 3, 4$.

Naturally, we ask whether there is any elementary method to find a closed form of $G_k(x)$ for any positive integer k? The purpose of this note is to give an affirmative answer. First, in section 2, we derive generating functions for $\{F_{kn}\}$ and the Lucas sequence $\{L_{kn}\}$ based on the Binet formulas. Then, in section 3, we establish closed forms for $G_k(x)$ via power-reducing techniques. Finally, in section 4, we extend our results to more general sequences including Lucas and Pell sequences. Since the method requires only a general knowledge of calculus, we expect that it will give more insight into the nature of these sequences and make them more accessible.

2. Generating functions for $\{F_{kn}\}$ and $\{L_{kn}\}$

Let the *Lucas sequence* $\{L_n\}$ be defined by

$$L_1 = 1, \quad L_2 = 3;$$

$$L_n = L_{n-1} + L_{n-2} \quad (\text{for } n \geqslant 3). \tag{3}$$

Set

$$\alpha = \frac{1+\sqrt{5}}{2}, \quad \beta = \frac{1-\sqrt{5}}{2}.$$

Based on the *Binet formulas*

$$F_n = \frac{\alpha^n - \beta^n}{\sqrt{5}}$$

and

$$L_n = \alpha^n + \beta^n,$$

we find generating functions for $\{F_{kn}\}$ and $\{L_{kn}\}$ as follows

$$\sum_{n=0}^{\infty} F_{kn} x^n = \frac{F_k\, x}{1 - L_k\, x + (-1)^k x^2}, \tag{4}$$

$$\sum_{n=0}^{\infty} L_{kn} x^n = \frac{2 - L_k\, x}{1 - L_k\, x + (-1)^k x^2}. \tag{5}$$

Indeed, in view of the fact that $\alpha\beta = -1$, we have

$$\sum_{n=0}^{\infty} F_{kn}x^n = \sum_{n=0}^{\infty} \frac{\alpha^{kn} - \beta^{kn}}{\sqrt{5}}\, x^n$$

$$= \frac{1}{\sqrt{5}}\left(\sum_{n=0}^{\infty} \alpha^{kn}\, x^n - \sum_{n=0}^{\infty} \beta^{kn}\, x^n\right)$$

$$= \frac{1}{\sqrt{5}}\left(\frac{1}{1-\alpha^k\, x} - \frac{1}{1-\beta^k\, x}\right)$$

$$= \frac{1}{\sqrt{5}}\,\frac{(\alpha^k - \beta^k)x}{1 - L_k\, x + (-1)^k\, x^2}$$

$$= \frac{F_k\, x}{1 - L_k\, x + (-1)^k x^2}.$$

Similarly, we can derive (5), the generating function for L_{kn}.

3. Closed forms of $G_k(x)$

In this section, using the generating functions (4) and (5), we show how to establish closed forms of $G_k(x)$ for any positive integer k. To motivate the idea, we begin with two special cases, which also give new proofs of the closed forms for $G_2(x)$ and $G_3(x)$.

For $k = 2$, from (4), we have

$$\sum_{n=0}^{\infty} L_{2n}x^n = \frac{2-3x}{1-3x+x^2}.$$

Combining with

$$F_n^2 = \frac{(\alpha^n - \beta^n)^2}{5}$$

$$= \frac{\alpha^{2n} - 2\alpha^n\beta^n + \beta^{2n}}{5}$$

$$= \frac{1}{5}\,(L_{2n} - 2(-1)^n),$$

we find

$$\sum_{n=1}^{\infty} F_n^2 x^n = \frac{1}{5}\left(\sum_{n=0}^{\infty} L_{2n}x^n - 2\sum_{n=0}^{\infty}(-1)^n x^n\right)$$

$$= \frac{1}{5}\left(\frac{2-3x}{1-3x+x^2} - \frac{2}{1+x}\right)$$

$$= \frac{x(1-x)}{(1+x)(1-3x+x^2)}.$$

For $k=3$, from (5), we have

$$\sum_{n=0}^{\infty} F_{3n}x^n = \frac{2x}{1-4x-x^2};\ \sum_{n=0}^{\infty} F_n(-x)^n = \frac{3x}{1+x-x^2}.$$

Again, combining with

$$\begin{aligned} F_n^3 &= \frac{(\alpha^n-\beta^n)^3}{5\sqrt{5}} \\ &= \frac{\alpha^{3n}-3\alpha^{2n}\beta^n+3\alpha^n\beta^{2n}-\beta^{3n}}{5\sqrt{5}} \\ &= \frac{1}{5}\,(F_{3n}-3(-1)^nF_n), \end{aligned}$$

we find

$$\begin{aligned} \sum_{n=1}^{\infty} F_n^3x^n &= \frac{1}{5}\left(\sum_{n=0}^{\infty} F_{3n}x^n - 3\sum_{n=0}^{\infty} F_n(-x)^n\right) \\ &= \frac{1}{5}\left(\frac{2x}{1-4x-x^2}+\frac{3x}{1+x-x^2}\right) \\ &= \frac{x(1-2x-x^2)}{(1+x-x^2)(1-4x-x^2)}. \end{aligned}$$

These two examples suggest that there exists an explicit closed form for $G_k(x)$ up on that F_n^k in terms of F_{nk} and L_{nk}. For this purpose, in general, we establish the following

Power-reducing formula

$$5^kF_n^{2k} = \sum_{i=0}^{k-1}\binom{2k}{i}(-1)^{i(n+1)}\,L_{2(k-i)n} + (-1)^{k(n+1)}\binom{2k}{k}, \tag{6}$$

$$5^kF_n^{2k+1} = \sum_{i=0}^{k}\binom{2k+1}{i}(-1)^{i(n+1)}\,F_{(2(k-i)+1)n}. \tag{7}$$

To see this, we use the binomial theorem, we have

$$5^kF_n^{2k} = (\alpha^n-\beta^n)^{2k} = \sum_{i=0}^{2k}(-1)^i\binom{2k}{i}\alpha^{in}\beta^{(2k-i)n}. \tag{8}$$

Recalling that $\alpha\beta=-1$ and

$$\binom{2k}{i} = \binom{2k}{2k-i},$$

regrouping (8) and using the Binet formula for Lucas sequences, we get (6). Similarly, we can deduce the formula (7). Finally, using the power-reducing formulas (6) and (7), we find

$$\begin{aligned}5^k G_{2k}(x) &= \sum_{i=0}^{k-1} (-1)^i \binom{2k}{i} \sum_{n=0}^{\infty} L_{2(k-i)n}\left((-1)^i x\right)^n \\ &\quad + (-1)^k \binom{2k}{k} \sum_{n=0}^{\infty} \left((-1)^k x\right)^n \\ &= \sum_{i=0}^{k-1} (-1)^i \binom{2k}{i} \frac{2 - L_{2(k-i)}(-1)^i x}{1 - L_{2(k-i)}(-1)^i x + x^2} \\ &\quad + (-1)^k \binom{2k}{k} \frac{1}{1-(-1)^k x}.\end{aligned}$$

$$\begin{aligned}5^k G_{2k+1}(x) &= \sum_{i=0}^{k} (-1)^i \binom{2k+1}{i} \sum_{n=0}^{\infty} F_{(2(k-i)+1)n}\left((-1)^i x\right)^n \\ &= \sum_{i=0}^{k} \binom{2k+1}{i} \frac{F_{2(k-i)+1} x}{1 - L_{2(k-i)+1}(-1)^i x - x^2}.\end{aligned}$$

Explicitly, we find the generating functions for G_4, G_5, G_6 as follows

$$\begin{aligned}G_4(x) &= \frac{1}{25}\left(\frac{2-7x}{1-7x+x^2} - \frac{4(2+3x)}{1+3x+x^2} + \frac{6}{1-x}\right) \\ &= \frac{x(1-4x-4x^2+x^3)}{(1-x)(1-7x+x^2)(1+3x+x^2)}, \\ G_5(x) &= \frac{1}{25}\left(\frac{5x}{1-11x-x^2} + \frac{10x}{1-x-x^2} + \frac{10x}{1+4x-x^2}\right) \\ &= \frac{x(1-7x-16x^2+7x^3+x^4)}{(1+4x-x^2)(1-x-x^2)(1-11x-x^2)}, \\ G_6(x) &= \frac{1}{125}\left(\frac{2-18x}{1-18x+x^2} + \frac{15(2-3x)}{1-3x+x^2} - \frac{6(2+7x)}{1+7x+x^2} - \frac{20}{1+x}\right) \\ &= \frac{x(1-12x-53x^2+53x^3+12x^4-x^5)}{(1-x)(1-18x+x^2)(1-3x+x^2)(1+7x+x^2)}.\end{aligned}$$

Remark: In [1] homogeneous recurrence relations are used to derive the generating function G_3. Reciprocally, $G_k(x)$ can be employed to find new homogeneous recurrence relations. For example, G_4 can be used to prove that

$$F_{n+5}^4 = 5F_{n+4}^4 + 15F_{n+3}^4 - 15F_{n+2}^4 - 5F_{n+1}^4 + F_n^4.$$

The details along this line will be studied and reported in a separate paper.

4. Some extensions

Based on the generating functions of $\{F_{kn}\}$ and $\{L_{kn}\}$ and the power-reducing techniques, we have established closed form generating functions for powers of the Fibonacci sequence. Along this line, we actually enable the establishment of closed forms of the generating functions for powers of any second-order recurrence sequences. In the following, we sketch the proofs of the closed forms of generating functions for Lucas and Pell sequences, and we leave the details and the general case for the reader to formulate.

Similar to (6) and (7), we have the following power-reducing formula:

$$L_n^{2k} = \sum_{i=0}^{k-1} \binom{2k}{i} (-1)^{in} L_{2(k-i)n} + (-1)^{kn} \binom{2k}{k},$$

$$L_n^{2k+1} = \sum_{i=0}^{k} \binom{2k+1}{i} (-1)^{in} L_{(2(k-i)+1)n}.$$

Let the generating function of $\{L_n^k\}$ be $H_k(x)$. Then,

$$H_{2k}(x) = \sum_{i=0}^{k-1} \binom{2k}{i} \frac{2 - L_{2(k-i)}(-1)^i x}{1 - L_{2(k-i)}(-1)^i x + x^2} + \binom{2k}{k} \frac{1}{1-(-1)^k x},$$

$$H_{2k+1}(x) = \sum_{i=0}^{k} \binom{2k+1}{i} \frac{2 - L_{2(k-i)+1}(-1)^i x}{1 - L_{2(k-i)+1}(-1)^i x - x^2}.$$

Explicitly,

$$H_1(x) = \frac{2-x}{1-x-x^2},$$

$$H_2(x) = \frac{4-7x-x^2}{(1+x)(1-3x+x^2)},$$

$$H_3(x) = \frac{8-13x-24x^2+x^3}{(1+x-x^2)(1-4x-x^2)},$$

$$H_4(x) = \frac{16-79x-164x^2+76x^3+x^4}{(1-x)(1+3x+x^2)(1-7x+x^2)}.$$

Next, recall that the Pell sequence is defined by $P_n = 2P_{n-1} + P_{n-2}$ with $P_1 = 1$, $P_2 = 2$ and the Pell–Lucas sequence Q_n [4] is defined by the same recurrence relation but with $Q_1 = 2$, $Q_2 = 6$. Setting

$$s = 1 + \sqrt{2}, \quad t = 1 - \sqrt{2},$$

similarly, we are able to deduce the Binet-type formulas:

$$P_n = \frac{1}{2\sqrt{2}} (s^n - t^n), \quad Q_n = s^n + t^n;$$

the generating functions of $\{P_n\}$ and $\{Q_n\}$:

$$\sum_{n=0}^{\infty} P_{kn}x^n = \frac{P_k\, x}{1 - Q_k\, x + (-1)^k x^2},$$

$$\sum_{n=0}^{\infty} Q_{kn}x^n = \frac{2 - Q_k\, x}{1 - Q_k\, x + (-1)^k x^2};$$

and the power-reducing formula:

$$2^{3k} P_n^{2k} = \sum_{i=0}^{k-1} \binom{2k}{i} (-1)^{i(n+1)}\, Q_{2(k-i)n} + (-1)^{k(n+1)} \binom{2k}{k},$$

$$2^{3k} P_n^{2k+1} = \sum_{i=0}^{k} \binom{2k+1}{i} (-1)^{i(n+1)}\, P_{(2(k-i)+1)n}.$$

Finally, setting $\mathcal{G}_k(x) = \sum_{n=0}^{\infty} P_n^k\, x^n$, we have

$$2^{3k}\mathcal{G}_{2k}(x) = \sum_{i=0}^{k-1} (-1)^i \binom{2k}{i} \frac{2 - Q_{2(k-i)}(-1)^i x}{1 - Q_{2(k-i)}(-1)^i x + x^2}$$
$$+ (-1)^k \binom{2k}{k} \frac{1}{1 - (-1)^k x}.$$

$$2^{3k}\mathcal{G}_{2k+1}(x) = \sum_{i=0}^{k} \binom{2k+1}{i} \frac{P_{2(k-i)+1}x}{1 - Q_{2(k-i)+1}(-1)^i x - x^2}.$$

In particular,

$$\mathcal{G}_1(x) = \frac{x}{1 - 2x - x^2},$$

$$\mathcal{G}_2(x) = \frac{x(1-x)}{(1+x)(1-6x+x^2)},$$

$$\mathcal{G}_3(x) = \frac{x(1-4x-x^2)}{(1+2x-x^2)(1-14x-x^2)},$$

$$\mathcal{G}_4(x) = \frac{x(1+x)(1-14x+x^2)}{(1-x)(1+6x+x^2)(1-34x-x^2)}.$$

References

[1] Koshy, T., 2001, *Fibonacci and Lucas Numbers with Applications* (New York: John Wiley).

[2] Mansour, T., 2004, A formula for the generating functions of powers of Horsdam's sequence. *Australasian Journal of Combinatorics*, **30**, 207–212.

[3] Riordan, J., 1962, Generating functions for powers of Fibonacci numbers. *Duke Mathematics Journal*, **29**, 5–12.

[4] Weisstein E., Fibonacci number, mathworld—A wolfram web resource, available at http://mathworld.wolfram.com/FibonacciNumber.html

[5] Weisstein E., Pell numbers, mathworld — A wolfram web resource, available at http://mathworld.wolfram.com/PellNumber.html

Evaluation of mean and variance integrals without integration

A. H. JOARDER and M. H. OMAR*
King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

(*Received 3 May 2006*)

The mean and variance of some continuous distributions, in particular the exponentially decreasing probability distribution and the normal distribution, are considered. Since they involve integration by parts, many students do not feel comfortable. In this note, a technique is demonstrated for deriving mean and variance through differential calculus. The general nature of the technique has potential for wider applications.

1. Introduction

In service courses in engineering statistics, sometimes we need to evaluate the mean and variance of some continuous distributions, especially the exponentially decreasing probability density function and the normal probability density function. Because these involve integration by parts and/or the use of L'Hospital's rule, many students find this difficult. In addition, instructors also face a teaching digression. In this note, we assume that the probability density function (pdf) in question has at least one continuous parameter. Since any probability density function integrates to unity, we call this integral a density identity (DI) in the parameters of the distribution. We derive the mean and variance integrals by repeatedly differentiating the DI with respect to the parameters.

The probability density function of an exponential random variable X, say the lifetime of a battery, is given by

$$f(x) = \lambda e^{-\lambda x}, \quad 0 < x < \infty, \quad 0 < \lambda < \infty. \tag{1}$$

For the motivation of this distribution see [1, p. 184]. Suppose that we want to know the expected lifetime of the batteries and also the variance of the lifetimes of the batteries. The mean and variance integrals are given by

$$\begin{aligned} E(X) &= \int_0^\infty x f(x) dx \\ &= \int_0^\infty x(\lambda e^{-\lambda x}) dx \end{aligned} \tag{2}$$

and

$$\begin{aligned} V(X) &= \int_0^\infty (x - E(X))^2 f(x) dx \\ &= \int_0^\infty x^2 f(x) dx - (E(X))^2 \end{aligned} \tag{3}$$

respectively.

*Corresponding author. Email: omarmh@kfupm.edu.sa

The mean and variance integrals given by (2) and (3) involve integration by parts, which many students find difficult. Since the pdf in (1) integrates to 1,

$$\int_0^\infty f(x)dx = 1$$

$$\text{i.e.} \quad \int_0^\infty \lambda e^{-\lambda x} dx = 1. \tag{4}$$

The above integral will be called the density identity, and will be used to evaluate the mean and variance integrals (2) and (3). We repeatedly differentiate the density identity with respect to λ to get new identities that are in turn exploited to evaluate the mean and variance integrals.

2. Method

The method is described for the exponential and normal distributions. The following lemma is obvious.

Lemma 2.1: Let x and λ be nonnegative. Then we have

(i) $\dfrac{d}{d\lambda}(e^{-\lambda x}) = -xe^{-\lambda x}$,

(ii) $\dfrac{d}{d\lambda}(\lambda e^{-\lambda x}) = (-\lambda x + 1)e^{-\lambda x}$.

Example 2.1: Let X have the exponential pdf given by (1). Then we have

(i) $\int_0^\infty x(\lambda e^{-\lambda x})dx = \dfrac{1}{\lambda}$,

(ii) $\int_0^\infty x^2(\lambda e^{-\lambda x})dx = \dfrac{2}{\lambda^2}$.

Proof: (i) Differentiating the density identity (4) with respect to λ (see Lemma 2.1 (ii)), we have

$$\int_0^\infty (-\lambda x + 1)e^{-\lambda x} dx = 0$$

which simplifies to

$$\begin{aligned}\int_0^\infty \lambda x e^{-\lambda x} dx &= \int_0^\infty e^{-\lambda x} dx \\ &= \frac{1}{\lambda} \text{ by (4)},\end{aligned}$$

which is part (i).

That is, $E(X) = 1/\lambda$, which is the mean of the exponential distribution.

(ii) Differentiating the identity (i) in this example again (See Lemma 2.1 (ii)) with respect to λ, we have

$$\int_0^\infty x\big((-\lambda x + 1)e^{-\lambda x}\big)dx = \frac{-1}{\lambda^2}.$$

Then by Example 2.1 (i), we have

$$\int_0^\infty x^2\lambda e^{-\lambda x}dx = \int_0^\infty xe^{-\lambda x}dx + \frac{1}{\lambda^2}$$
$$= \frac{1}{\lambda^2} + \frac{1}{\lambda^2}.$$
$$\text{i.e.}\quad E(X^2) = \frac{2}{\lambda^2}.$$

Given these results, the variance integral (3) is given by

$$V(X) = E(X^2) - (E(X))^2$$
$$\text{i.e.}\quad \sigma^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Example 2.2: Let X have the normal distribution $N(\mu, 1)$ with pdf

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}, \qquad -\infty < x < \infty, \ -\infty < \mu < \infty$$

Then $E(X) = \mu$.

Solution: The density identity is given by

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}dx = 1$$

Differentiating both sides of the above identity with respect to μ, we have

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}\left(-(2)\frac{1}{2}(x-\mu)(-1)\right)dx = 0$$
$$\text{i.e.}\quad \int_{-\infty}^\infty (x-\mu)f(x)dx = 0$$
$$\text{or,}\ E(X-\mu) = 0.$$

Example 2.3: For a general normal distribution $N(\mu, \sigma^2)$ with pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}, \qquad -\infty < x < \infty, \ -\infty < \mu < \infty, \ 0 < \sigma < \infty$$

find the mean and variance.

Solution: The density identity is given by

$$\int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}dx = 1 \tag{5}$$

Differentiating both sides of the above identity with respect to μ, we have

$$\int_{-\infty}^\infty \left(\frac{x-\mu}{\sigma^2}\right)f(x)dx = 0$$
$$\text{or,}\ E(X-\mu) = 0$$
$$\text{that is,}\ E(X) = \mu.$$

Again differentiating both sides of the density identity (5) with respect to σ we have

$$\int_{-\infty}^{\infty}\left(\frac{(x-\mu)^2-\sigma^2}{\sigma^3}\right)f(x)dx = 0$$

$$\text{or, } E((X-\mu)^2-\sigma^2) = 0$$

$$\text{that is, } V(X) = \sigma^2.$$

Thus, the mean and variance of a normal distribution $N(\mu,\ \sigma^2)$ are given by $E(X)=\mu$ and $V(X)=\sigma^2$ respectively. In particular, when X has the normal distribution $N\ (0,\ \sigma^2)$ with pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/(2\sigma^2)}, \quad -\infty < x < \infty,\ 0 < \sigma < \infty$$

it follows from Example 2.3 that $E(X)=0$ and $V(X)=\sigma^2$.

3. An application

Example 3.1: Let the continuous random variable Y denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 millimetres. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of Y can be modelled by a probability density function

$$f(y) = 20e^{-20(y-12.5)}, \quad y \geq 12.5 \tag{6}$$

([2], p. 59). If parts with diameter larger than 12.6 millimetres are scrapped, the probability that a part is scrapped is given by

$$\begin{aligned} P(Y > 12.60) &= \int_{12.6}^{\infty} f(y)dy \\ &= e^{250}e^{-252} \\ &\approx 0.135. \end{aligned}$$

Thus for a sample of size n, approximately $0.135n$ parts should be scrapped. Note that the number of scrapped parts with larger diameters holes will have a binomial distribution $B(n,p)$ where $p \approx 0.135$ is the probability that a part with larger diameters ($Y > 12.60$) will be scrapped.

Suppose that we want to know the expected value and variance of the diameters of holes drilled in a sheet metal component. The mean is given by

$$\begin{aligned} E(Y) &= \int_{12.5}^{\infty} yf(y)dy \\ &= \int_{12.5}^{\infty} y(20e^{-20(y-12.5)})dy \end{aligned} \tag{7}$$

Letting $y = x + 12.5$, we have

$$E(Y) = \int_0^\infty (x + 12.5)(20e^{-20x})dx$$
$$= \int_0^\infty x(20e^{-20x})dx + 12.5\int_0^\infty (20e^{-20x})dx$$
$$= \int_0^\infty x(20e^{-20x})dx + 12.5.$$

Then by Lemma 2.1, we have

$$E(Y) = \frac{1}{20} + 12.5 = 12.55.$$

The variance is given by

$$V(Y) = \int_0^\infty y^2 f(y)dy - 12.55^2 \tag{8}$$

$$\text{But } E(Y^2) = \int_{12.5}^\infty y^2 f(y)dy$$
$$= \int_{12.5}^\infty y^2(20e^{-20(y-12.5)})dy$$
$$= \int_0^\infty (x + 12.5)^2(20e^{-20x})dx. \tag{9}$$

By Example 2.1, the integral in (9) is evaluated as

$$\int_0^\infty x^2(20e^{-20x})dx + 2(12.5)\int_0^\infty x(20e^{-20x})dx + (12.5)^2\int_0^\infty (20e^{-20x})dx$$
$$= \frac{2}{20^2} + 2(12.5)\frac{1}{20} + 12.5^2(1)$$
$$= 157.505$$

Therefore the variance in (8) is given by $V(Y) = 157.505 - 12.55^2 = 0.0025$ ([2], pp. 59–62).

We remark that the method discussed here is easily applied to most other continuous distributions. If a pdf does not explicitly have a continuous parameter, we can formally insert it into the pdf and apply the technique discussed.

Acknowledgement

The authors gratefully thank King Fahd University of Petroleum & Minerals, Saudi Arabia for providing excellent research facilities.

References

[1] Scheaffer, R.L. and McClave, J.T., 1995, *Probability and Statistics for Engineering* (Duxbury Press).
[2] Montgomery, D.C., Runger, G.C. and Hubele, N.F., 1998, *Engineering Statistics* (New York: John Wiley).

A note on the visibility in the $[1, N] \times [1, N]$ integer domain

G. D. KIM* and J. ENGELHARDT
Southern Oregon University, Ashland, USA

(*Received 25 April 2006*)

A k-dimensional integer point is called visible if the line segment joining the point and the origin contains no proper integer points. This note proposes an explicit formula that represents the number of visible points on the two-dimensional $[1, N] \times [1, N]$ integer domain. Simulations and theoretical work are presented.

1. Introduction

A k-dimensional vector point is called an integer point if all the components of the point are integers. Let A be a k-dimensional integer point. The point A is called "visible" if there is no proper integer point on the straight segment $\overline{AO}$ joining A and the origin O. As the term "visible" indicates, there exists a sensible geometric interpretation. For an illustration, suppose that we are standing at the origin in the two-dimensional integer domain and are observing other integer points. It is not very difficult to imagine that some integer points are visible and some are invisible from where you stand. For instance, in figure 1 the point (2,1) is visible because there is no proper integer point blocking the view of (2,1) from where we stand. The point (2,2), however, is invisible because the proper integer point (1,1) blocks the view of (2,2) from where we stand. For another example, consider the two points (4,2) and (4,3). The point (4,2) is an invisible point because the proper integer point (2,1), which is on the line segment between (4,2) and the origin (0,0), blocks the view of (4,2) from the origin. The point (4,3), however, is visible because it does not have any proper integer point blocking the view of it from the origin. As we shall see in Lemma 1, a point is visible if the components of the point are relative primes. In this article we consider how many visible points exist in the two-dimensional $[1, N] \times [1, N]$ integer lattice domain and also consider how to represent the number of visible points.

*Corresponding author. Email: kimd@sou.edu

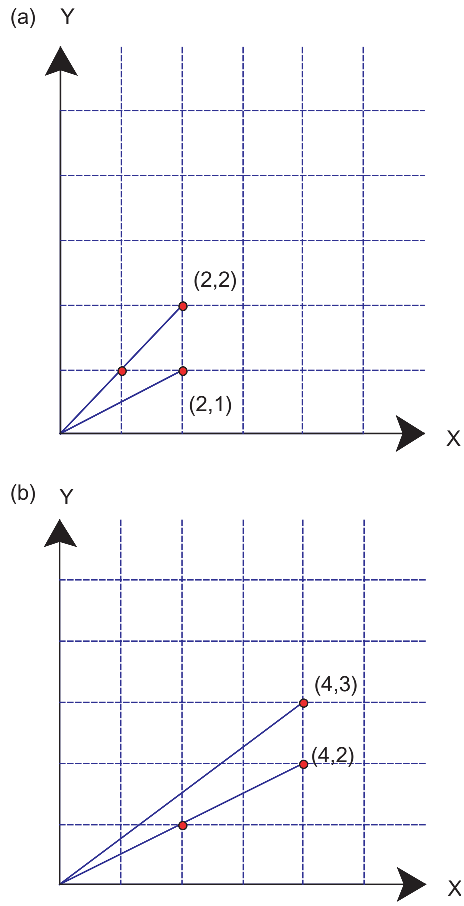

Figure 1. (a) and (b), (2,1) and (4,3) are visible, (2,2) and (4,2) are invisible.

2. Main

The two visible points (2,1) and (4,3) given in the introduction give us an idea about which points might be visible. A common feature of (2,1) and (4,3) is that the components of each point are relatively prime, whereas each of the invisible points (2,2) and (4,2) has a proper integer point between itself and the origin. Lemma 1 verifies that for an integer point to be visible from the origin it is necessary and sufficient that the components of the point are relatively prime.

Lemma 1: *A non-zero k-dimensional integer point* $A = (a_1, a_2, \ldots, a_k)$ *is visible if and only if the components* $a_1, a_2, \ldots, a_k$ *are relatively prime.*

Proof: We prove necessity first. Let us suppose that A is visible. From the definition of visibility, there exists no proper integer that divides all the components in A. This means that all the components are relatively prime. Now in order to prove sufficiency, suppose that the components $a_1, a_2, \ldots, a_k$ are relatively prime. This implies that there is no non-trivial common divisor of the components, which implies that the point A must be visible. $\square$

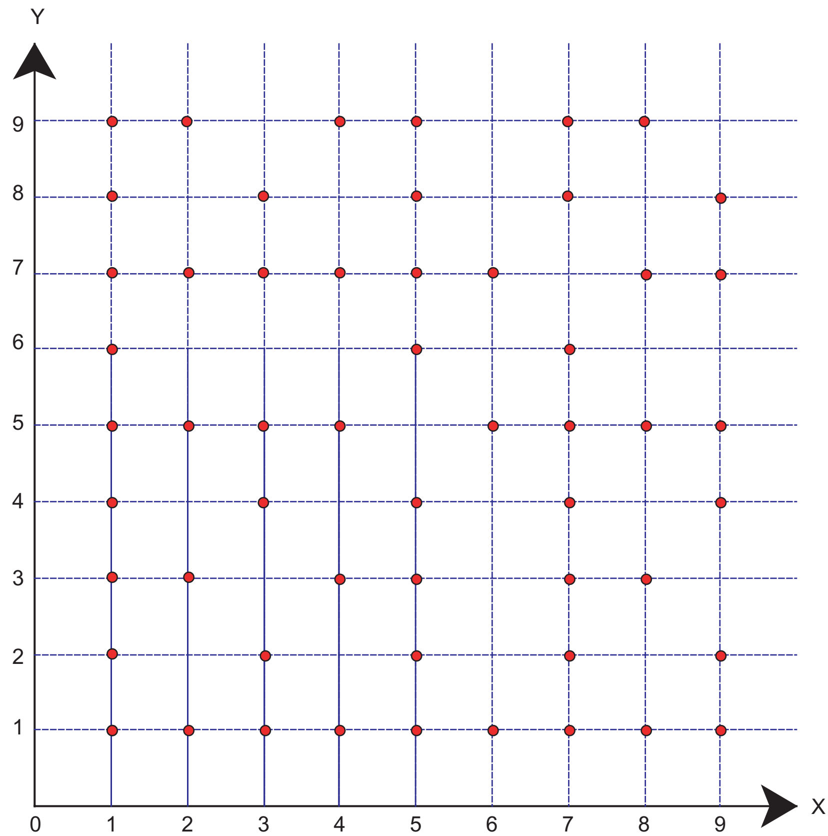

Figure 2. Visible points in the $[1, N] \times [1, N]$ integer lattice plane.

Lemma 1 tells us that for a point to be visible it is necessary and sufficient to have all the components of the point relatively prime, which means that the greatest common divisor of the components is 1. Let V_N represent the number of visible points in the integer lattice domain $[1, N] \times [1, N]$, and P_N the proportion of visible points in the integer lattice domain. Figure 2 illustrates the visible points in the $[1, 9] \times [1, 9]$ integer lattice domain, and table 1 summarizes V_N and P_N for $N = 1, 2, \ldots, 9$. Note that P_N is obtained by V_N/N^2 because there exist N^2 integer points in the $[1, N] \times [1, N]$ integer domain.

We now consider how many visible points exist in the $[1, N] \times [1, N]$ integer lattice domain. Lemma 2 presents a nice mathematical representation of V_N.

Lemma 2: *The number of visible points V_N in the $[1, N] \times [1, N]$ integer lattice domain is*

$$V_N = 2 \cdot \sum_{n=1}^{N} n \prod_{i=1}^{m_n} \left(1 - \frac{1}{p_{i_n}}\right)$$

where $p_1, p_2, \ldots, p_{m_n}$ are the distinct prime divisors of n.

Table 1. Tabulation of N^2, V_N, and P_N for $N = 1, 2, \ldots, 9$.

N	N^2, the number of integer points	V_N, the number of visible points	P_N, the proportion of visible points
1	1	1	1.00
2	4	3	0.75
3	9	7	0.78
4	16	11	0.69
5	25	19	0.76
6	36	23	0.64
7	49	35	0.71
8	64	43	0.67
9	81	55	0.68

Table 2. Euler $\phi(N)$-function and the visible points V_N for $N = 1, 2, \ldots, 9$.

N	Prime Divisors less than N	$\phi(N)$	$V_N = 2\sum_{n=1}^{N} \phi(n)$
1	NA	1	1
2	2	1	$1 + 2 \cdot 1 = 3$
3	3	2	$3 + 2 \cdot 2 = 7$
4	2	2	$7 + 2 \cdot 2 = 11$
5	5	4	$11 + 2 \cdot 4 = 19$
6	2,3	2	$19 + 2 \cdot 2 = 23$
7	7	6	$23 + 2 \cdot 6 = 35$
8	2	4	$35 + 2 \cdot 4 = 43$
9	3	6	$43 + 2 \cdot 6 = 55$

Proof: The Euler function $\phi\ (n)$ of a positive integer n is defined to be the number of integers relatively prime to n among the integers less than n. Note that, from number theory, the Euler function can be represented as follows:

$$\phi\ (n) = n \prod_{i=1}^{m} \left(1 - \frac{1}{p_i}\right)$$

where $p_1, p_2, \ldots, p_m$ are all distinct prime divisors of n. It follows from Lemma 1, therefore, that

$$V_N = 2 \cdot \sum_{n=1}^{N} \phi\ (n) = 2 \cdot \sum_{n=1}^{N} n \prod_{i=1}^{m_n} \left(1 - \frac{1}{p_{i_n}}\right). \qquad \square$$

Table 2 demonstrates how V_N is obtained through the formula in Lemma 2, for $N = 1, 2, \ldots, 9$. Note that the results are identical to the visible points V_N provided in table 1.

3. Discussion

1. Does V_N grow indefinitely as N increases? This question is not difficult to answer. If we consider the integer points of $(1, p)$ type where p is a prime integer, these are all visible points because the integer points of this type cannot have a proper

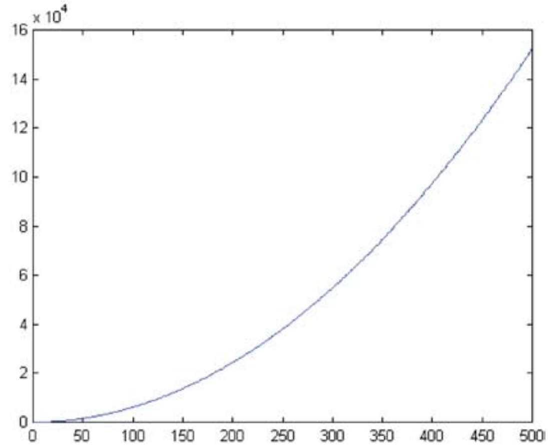

Figure 3. V_N, the number of visible points in the $[1, N] \times [1, N]$ integer domain for $N = 1, 2, \ldots, 500$.

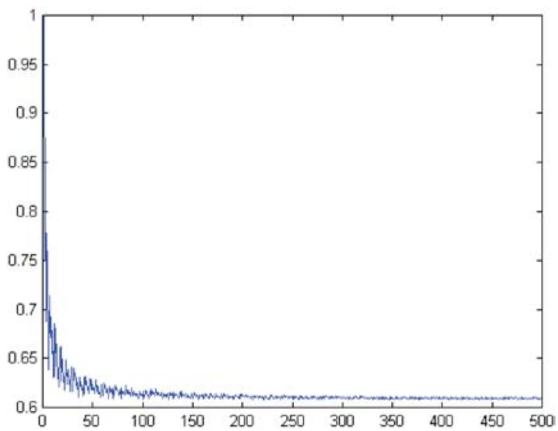

Figure 4. The proportion P_N of visible points in $[1, N] \times [1, N]$-plane.

integer point. Since there exist infinitely many prime integers, we can conclude that there are an infinite number of visible points. Figure 3 illustrates the pattern of how V_N grows as N increases to 500.

2. The proportion P_N of visible points in the $[1, N] \times [1, N]$ integer domain is easily obtained by the ratio of the number of visible points V_N by N^2 to the total number of integer points in $[1, N] \times [1, N]$. It is

$$P_N = \frac{V_N}{N^2} = \frac{2}{N^2} \cdot \sum_{n=1}^{N} \prod_{i=1}^{m_n} \left(1 - \frac{1}{p_{i_n}}\right)$$

where m_n is the number of distinct prime divisors of n. Figure 4 illustrates the proportion of visible points in the integer domain $[1, 500]^2$. The graph illustrates that the proportion of visible points approaches 0.6 quickly after around 50. For example, if we use the formula obtained in Lemma 2, we find $V_{500} = 152231$. The proportion of visible points is obtained as $P_{500} = (152231/500^2) = 0.60892$.

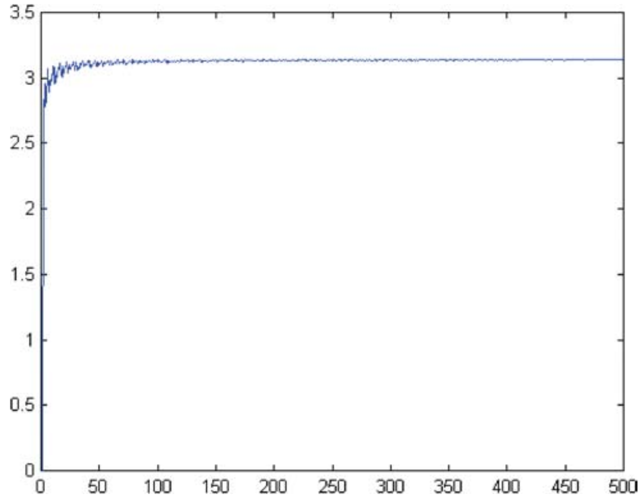

Figure 5. For N ranging from 1 to 500, $\sqrt{6/P_N}$ as an approximate of π.

3. Figure 4 illustrates that P_N, the proportion of visible points, approaches 0.6. The limit of P_N, as N increases to infinity, is not immediately clear from the graph. It is also a very difficult task to take the limit of the P_N-formula considered earlier. The algebraic structure of $P_N = (2/N^2) \cdot \sum_{n=1}^{N} \prod_{i=1}^{m_n}(1 - (1/p_{i_n}))$ requires us to know how to represent the prime divisors of all positive integers, which is not a very well cultivated area in mathematics. In fact, the proportion of visible points in the infinite domain $[1, \infty) \times [1, \infty)$ has been studied independently of the structure of P_N, and it was proven to be $6/\pi^2$ (see 1). This quantity is approximately 0.607927, which matches the illustration for P_N in figure 4.

4. The visible points bring up an interesting educational lesson to us. It is a general consensus of many people that π is a mystic number, which occurs in many unexpected areas in life. This time, visible points offer us another way to consider π. The discussions made in sections 2 and 3 above lead us to an equation that relates P_N to π as follows:

$$\lim_{N\to\infty} P_N = \frac{6}{\pi^2}$$

where $P_N = (2/N^2) \cdot \sum_{n=1}^{N} \prod_{i=1}^{m_n}(1 - (1/p_{i_n}))$ and $p_1, p_2, \ldots, p_{m_n}$ are the distinct prime divisors of n. By solving this equation for π, we get

$$\pi = \lim_{N\to\infty} \sqrt{\frac{6}{P_N}}.$$

Figure 5 shows a graph illustrating $\sqrt{6/P_N}$ as an approximation to π for $N = 1, 2, \ldots, 500$. As it shows, for N greater than 10 or so, the approximations are very good.

Reference

[1] Jones, G. and Jones, M., 1998, *Elementary Number Theory* (London: Springer), pp. 170–174.

Solving quartics

R. G. KULKARNI*
Bharat Electronics, Bangalore, India

(*Received 5 June 2006*)

A technique is presented, which is different from the well-known Ferrari's method, to solve a general quartic equation. Formulae for the four roots of quartic are derived. A numerical example verifies the formulae obtained.

1. Introduction

The quest to solve polynomial equations is not new. Before 2000 BC, Greeks, Hindus, and Babylonians knew the solution to quadratics in one form or the other. However, the cubic challenged mathematicians for more than three thousand years after the quadratic was solved. In the 11th century, Omar Khayyam [1] solved the cubic geometrically, by intersecting parabolas and circles, and Scipione del Ferro [1, 2] found the solution to the cubic in 1515, but did not publish it. In 1535 Tartaglia [1, 3] obtained formulae to solve cubic equations, while Cardano [3, 4] published the solution to the general cubic using complex numbers, at a time when the use of complex numbers was considered absurd. At around the same time, Cardano's friend Ferrari [1, 4] obtained formulae for the roots of general quartic equation.

In this note we describe a method to solve the general quartic equation, which is different from the well-known Ferrari's method. The method decomposes the general quartic equation into two quadratic polynomials as factors in a novel fashion, eventually leading to its solution.

2. Polynomial decomposition method

Consider the general quartic equation in x whose solution is sought:

$$x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0 \tag{1}$$

where a_0, a_1, a_2, and a_3 are real and independent coefficients. Let us construct another quartic equation in the form as shown below.

$$\frac{[(x^2 + b_1x + b_0)^2 - p^2(x^2 + c_0)^2]}{(1 - p^2)} = 0, \quad p \neq 1 \tag{2}$$

where b_0, b_1, c_0, and p are unknowns to be determined. Observe that the above constructed quartic equation (2) can be decomposed into two factors as given below.

$$\left\{\frac{[(x^2 + b_1x + b_0) - p(x^2 + c_0)]}{(1 - p)}\right\}\left\{\frac{[(x^2 + b_1x + b_0) + p(x^2 + c_0)]}{(1 + p)}\right\} = 0. \tag{3}$$

*Email: rgkulkarni@ieee.org

Therefore if the given quartic equation (1) can be converted into the constructed quartic equation (2), then equation (1) can be factorized into two factors as indicated in (3). This means that the following relation must be satisfied.

$$x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = \frac{[(x^2 + b_1x + b_0)^2 - p^2(x^2 + c_0)^2]}{(1 - p^2)} \tag{4}$$

Expanding and rearranging the right-hand side of equation (4), in descending powers of x, we can rewrite (4) as:

$$\begin{aligned} x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = x^4 &+ \left[\frac{2b_1}{(1 - p^2)}\right]x^3 + \left[\frac{(b_1^2 + 2b_0 - 2c_0p^2)}{(1 - p^2)}\right]x^2 \\ &+ \left[\frac{2b_0b_1}{(1 - p^2)}\right]x + \left[\frac{(b_0^2 - c_0^2p^2)}{(1 - p^2)}\right] \end{aligned} \tag{5}$$

Equating the coefficients of x^0, x, x^2, and x^3 on the left-hand side to that on the right-hand side results in the following four equations in the unknowns, b_0, b_1, c_0, and p.

$$\frac{2b_1}{(1 - p^2)} = a_3 \tag{6}$$

$$\frac{(b_1^2 + 2b_0 - 2c_0p^2)}{(1 - p^2)} = a_2 \tag{7}$$

$$\frac{2b_0b_1}{(1 - p^2)} = a_1 \tag{8}$$

$$\frac{(b_0^2 - c_0^2p^2)}{(1 - p^2)} = a_0 \tag{9}$$

To determine the four unknowns, b_0, b_1, c_0, and p, we solve the four equations (6)–(9). Dividing equations (7)–(9), by equation (6), we obtain the following expressions (when $a_3 \neq 0$). The special cases [when one or more coefficients in (1) are zero] are discussed separately in the next section.

$$b_0 = \frac{a_1}{a_3} \tag{10}$$

$$\frac{(b_0^2 - c_0^2p^2)}{(2b_1)} = \frac{a_0}{a_3} \tag{11}$$

$$\frac{(b_1^2 + 2b_0 - 2c_0p^2)}{(2b_1)} = \frac{a_2}{a_3} \tag{12}$$

Equation (6) is rearranged to obtain the following expression for p^2:

$$p^2 = \frac{(a_3 - 2b_1)}{a_3} \tag{13}$$

We note that (10) gives b_0, and (13) expresses p^2 in terms of b_1. Eliminating b_0 and p^2 from equations (11) and (12), using equations (10) and (13), respectively, we obtain the following two expressions in b_1 and c_0.

$$a_3 c_0^2(2b_1 - a_3) = 2a_0 a_3 b_1 - a_1^2 \tag{14}$$

$$2c_0(2b_1 - a_3) = 2a_2 b_1 - 2a_1 - a_3 b_1^2 \tag{15}$$

Note that both equations, (14) and (15), contain a term, $c_0(2b_1 - a_3)$, on their left-hand side. Therefore dividing equation (14) by (15) eliminates the term, $c_0(2b_1 - a_3)$, and after some rearrangement we get the following expression for c_0 in terms of b_1.

$$c_0 = \frac{2(2a_0 a_3 b_1 - a_1^2)}{[a_3(2a_2 b_1 - 2a_1 - a_3 b_1^2)]} \tag{16}$$

We eliminate c_0 from equations (15) and (16) to obtain the following expression involving b_1:

$$a_3^3 b_1^4 - 4a_2 a_3^2 b_1^3 + 4(a_1 a_3^2 + a_2^2 a_3 - 4a_0 a_3) b_1^2 + 8(a_0 a_3^2 + a_1^2 - a_1 a_2 a_3) b_1 = 0 \tag{17}$$

Note that in equation (17), b_1 emerges as a factor; however, $b_1 = 0$, is not part of our solution to the general quartic equation, since it makes the value of p^2 as unity [see equation (13)], and as a result there is division by zero in equations (2) to (9). Therefore factoring out b_1 from equation (17), we get the following cubic equation in b_1:

$$b_1^3 - \left(\frac{4a_2}{a_3}\right) b_1^2 + \left[\frac{4(a_1 a_3 + a_2^2 - 4a_0)}{a_3^2}\right] b_1 + \left[\frac{8(a_0 a_3^2 + a_1^2 - a_1 a_2 a_3)}{a_3^3}\right] = 0 \tag{18}$$

Solving the above cubic equation by Cardan's (or Cardano's) method, b_1 is determined. The remaining unknowns, p^2 and c_0, are found from equations (13) and (15), respectively. Note that if $b_1 = a_3/2$, then from (13) we obtain the value of p^2 as zero. In this case the given quartic equation (1) has repeated roots, as can be seen from equation (2). It is a worthwhile exercise for an interested reader to take an example of a quartic equation with repeated roots, and check the values of b_1 and p^2.

Since all unknowns have been determined, we can now represent the given quartic equation (1) in the form of our constructed quartic equation (2) and decompose the given quartic equation into two factors as shown in (3). Equating each of the factors in (3) to zero, we obtain the two quadratic equations as below.

$$x^2 + \left[\frac{b_1}{(1-p)}\right] x + \left[\frac{(b_0 - c_0 p)}{(1-p)}\right] = 0 \tag{19}$$

$$x^2 + \left[\frac{b_1}{(1+p)}\right] x + \left[\frac{(b_0 + c_0 p)}{(1+p)}\right] = 0 \tag{20}$$

Solving equations (19) and (20), we get all four roots of the given quartic equation. Let x_1 and x_2 be roots of quadratic equation (19); and, x_3 and x_4 be roots of quadratic equation (20). Then the roots, x_1, x_2, x_3, and x_4, are given by:

$$
\begin{aligned}
x_1 &= \frac{\{-b_1 + [b_1^2 - 4(b_0 - c_0p)(1-p)]^{1/2}\}}{[2(1-p)]} \\
x_2 &= \frac{\{-b_1 - [b_1^2 - 4(b_0 - c_0p)(1-p)]^{1/2}\}}{[2(1-p)]} \\
x_3 &= \frac{\{-b_1 + [b_1^2 - 4(b_0 + c_0p)(1+p)]^{1/2}\}}{[2(1+p)]} \\
x_4 &= \frac{\{-b_1 - [b_1^2 - 4(b_0 + c_0p)(1+p)]^{1/2}\}}{[2(1+p)]}
\end{aligned}
\tag{21}
$$

We have solved the general quartic equation and determined its four roots using the proposed polynomial decomposition method. In section 4, a numerical example is given to validate the formulae (21).

3. Special cases of quartic equation

Several interesting situations arise when one or more coefficients in the quartic equation (1) become zero. For example, when $a_3 = 0$, we cannot perform the division of equations (7)–(9) by equation (6) (as done earlier) since it amounts to division by zero in equations (10) to (12). Instead we change the variable in the quartic equation (1) to u by the transformation $x = u + d$, where d is a real non-zero number. The resulting quartic equation is then expressed as:

$$u^4 + A_3u^3 + A_2u^2 + A_1u + A_0 = 0$$

Notice that in the above equation, $A_3 \neq 0$, and we can proceed to solve the above quartic equation in u, with coefficients, A_0, A_1, A_2, and A_3, using the proposed method. When the roots of the above equation are determined as, say, u_1, u_2, u_3, and u_4, we obtain the roots of the given quartic equation (1) as:

$$
\begin{aligned}
x_1 &= u_1 + d, \\
x_2 &= u_2 + d, \\
x_3 &= u_3 + d, \\
x_4 &= u_4 + d.
\end{aligned}
$$

We invite the reader to solve such a quartic equation. In addition to a_3, if a_1 is also zero, the quartic (1) is reduced to the following bi-quadratic equation:

$$x^4 + a_2x^2 + a_0 = 0,$$

which can be converted into a quadratic equation by the variable transformation, $y = x^2$, as below:

$$y^2 + a_2y + a_0 = 0.$$

The two roots of the above quadratic, say y_1 and y_2, yield four roots of quartic (1) as:

$$x_1 = (y_1)^{1/2}, x_2 = -(y_1)^{1/2}, x_3 = (y_2)^{1/2}, x_4 = -(y_2)^{1/2}.$$

Thus when a_1 is zero along with a_3, there is no need to use the proposed method. When a_3 is non-zero, the given quartic can be solved by the proposed method for any combination of remaining coefficients being zero. However when $a_0 = 0$, one root of the quartic equation lies on origin, and the quartic can be reduced to a cubic equation as shown below.

$$x^3 + a_3x^2 + a_2x + a_1 = 0.$$

The cubic can be solved by Cardan's method to obtain the remaining three roots.

4. Numerical example

Consider a numerical example of a quartic equation, whose roots are irrational and complex in order to represent the most general case. The quartic equation chosen is:

$$x^4 + 2.0533927x^3 - 2.8917903x^2 + 7.6758959x + 29.5803989 = 0$$

First b_0 is determined from (10), and then b_1 is determined from cubic equation (18). Note that one real root of (18) is sufficient to determine the remaining unknowns, c_0 and p, which are determined from equations, (15) and (13) respectively. Thus the values of b_0, b_1, c_0, and p are evaluated as given below:

$$b_0 = 3.738154, b_1 = -13.44884, c_0 = 5.336053, \text{ and } p = 3.754882.$$

Using these values the four roots are evaluated from the expression set (21) as:

$$x_1 = -2.236067, x_2 = -2.645752, x_3 = 1.414213 + j1.732051, \text{ and}$$
$$x_4 = 1.414213 - j1.732051$$

5. Conclusions

An elegant polynomial decomposition technique to solve general quartic equation is presented. It is different from the well-known Ferrari's method; the formulae for the four roots of the general quartic are obtained, and verified by solving a numerical example.

Acknowledgements

The author thanks the management of Bharat Electronics for supporting this work.

References

[1] http://www.vimagic.de/hope/, History of polynomial equations.
[2] http://en.wikipedia.org/wiki/Cubic equation
[3] http://mathforum.org/dr.math/faq/faq.cubic.equations.html
[4] http://en.wikipedia.org/wiki/Quartic equation

A converse of Fermat's Little Theorem

P. S. BRUCKMAN*
Sointula, BC, Canada

(*Received 12 June 2006*)

As the name of the paper implies, a converse of Fermat's Little Theorem (FLT) is stated and proved. FLT states the following: if p is any prime, and x any integer, then $x^p \equiv x \pmod{p}$. There is already a well-known converse of FLT, known as Lehmer's Theorem, which is as follows: if x is an integer coprime with m, such that $x^{m-1} \equiv 1 \pmod{m}$, and if there exists no integer $e < m-1$ such that $x^e \equiv 1 \pmod{m}$, then m is prime. The new converse in question states the following: if p is any prime and $x^p \equiv x \pmod{p}$, where x is known only to be algebraic, then x must be an integer (mod p).

1. Introduction

In this paper, as the title implies, we present and prove a converse of Fermat's Little Theorem, as the latter is called in the literature. Another converse is known and has been proved, notably Lehmer's Theorem, by which name it is known. As is well known, Fermat's Little Theorem (FLT) states the following:

If p is any prime, and x any integer, then $x^p \equiv x \pmod{p}$.

A slightly weaker version of FLT states the following:

If p is any prime and if x is relatively prime to p, then $x^{p-1} \equiv 1 \pmod{p}$. Here is Lehmer's Theorem:
If x is an integer coprime with m, such that $x^{m-1} \equiv 1 \pmod{m}$, and if there exists no integer $e < m-1$ such that $x^e \equiv 1 \pmod{m}$, then m is prime.

If the added hypothesis involving the exponent e in Lehmer's Theorem is removed, the conclusion is false, leading to the fascinating concept of pseudoprimes, primality testing, witnesses, etc.

The new converse that we have in mind (CFLT) is the following:

Theorem (Converse of FLT)
Given a prime p and an algebraic number x such that $x^p \equiv x \pmod{p}$, then x must be an ordinary residue (mod p).

This is not as trivial a problem as may appear at first glance. For example, we might have, *a priori*, that $x = a + b\theta$, where $\theta^2 \equiv r \pmod{p}$, and a, b and r are ordinary residues (mod p), with b not congruent to 0 (mod p), but $(r/p) = -1$. The converse theorem states that the congruence $x^p \equiv x \pmod{p}$ cannot be satisfied in this case.

*Email: pbruckman@hotmail.com

2. Proof of CFLT

Consider the polynomial product $P(z) = (z-1)(z-2)\cdots(z-p)$, which may be expanded to the polynomial:

$$P(z) = z^p - \sigma(1)z^{p-1} + \sigma(2)z^{p-2} - \cdots + (-1)^{p-1}\sigma(p-1)z + (-1)^p\sigma(p) \qquad (*)$$

In this expression, $\sigma(m)$ represents the elementary symmetric function of order m of the quantities $1, 2, \ldots, p$. The following results were proved in Problem 1111 published in *The IIME Journal*, 2005, **12**(3), 151 (whose solution appeared in *The IIME Journal*, 2006, **12**(4), 241):

$$\sigma(m) \equiv 0(\mathrm{mod}\, p), m = 1, 2, \ldots, p-2, p; \sigma(p-1) \equiv -1(\mathrm{mod}\, p) \qquad (**)$$

Substituting these results into $(*)$, we see that $P(z) \equiv z^p + (-1)^p\, z \equiv z^p - z$ (mod p).

Note that this is also true for $p = 2$, since $z \equiv -z$ (mod 2). However, we know from the given congruence of CFLT that $x^p - x \equiv 0$ (mod p). Therefore, $P(x) \equiv 0$ (mod p), which implies (by the original definition of P) that $x \in \{1, 2, \ldots, p\}$. Equivalently, $x \in \{1, 2, \ldots, p-1\}$ or any other complete residue system (mod p). □

Case study projects for college mathematics courses based on a particular function of two variables

Y. SHI*
Bloomsburg University of Pennsylvania, Bloomsburg, PA, USA

(*Received 21 December 2005*)

Based on a sequence of number pairs, a recent paper (Mauch, E. and Shi, Y., 2005, Using a sequence of number pairs as an example in teaching mathematics, *Mathematics and Computer Education*, **39**(3), 198–205) presented some interesting examples that can be used in teaching high school and college mathematics classes such as algebra, geometry, calculus, and linear algebra. In this paper, this study is generalized further to develop a few interesting case study proposals that can be used for student projects in college mathematics courses such as real functions, analytic geometry, and complex variables. In addition to using them in individual courses, these studies may also be combined to offer seminars or workshops to college mathematics students. Projects like these are likely to promote student interest and get students more involved in the learning process, and therefore make the learning process more effective.

*Email: yshi@bloomu.edu

1. Introduction

A recent paper [1] by Mauch and Shi discussed some applications of a sequence of number pairs in teaching high school and college mathematics. That sequence is often used in elementary or middle school mathematics classes to illustrate the concept of 'patterns'. Based on that sequence, [1] generated a few interesting examples that may be used in teaching classes such as algebra, geometry, calculus, and linear algebra.

The sequence discussed in [1] is:

$$(2,3), \quad (5,7), \quad (12,17), \quad (29,41)\ \ldots \tag{1}$$

Here is the pattern of this sequence:

1. It starts with $(2,3)$.
2. The first number in the second pair is the sum of the two numbers in the first pair: $5=2+3$, and the second number in the second pair equals twice the first number in the first pair plus the second number in the first pair: $7=2(2)+3$.
3. The third pair is generated based on the second pair in a similar way: $12=5+7$ and $17=2(5)+7$.
4. In general, let (a_n, b_n) be the nth pair, then the next pair (a_{n+1}, b_{n+1}) will be $a_{n+1}=a_n+b_n$ and $b_{n+1}=2a_n+b_n$.

Note that the first pair of this sequence (2, 3) satisfies the equation $2a^2+1=b^2$ (that is, $2(2)^2+1=(3)^2$) and the second pair (5, 7) satisfies $2a^2-1=b^2$ (that is, $2(5)^2-1=(7)^2$). And then the third pair satisfies $2a^2+1=b^2$, and the fourth pair satisfies $2a^2-1=b^2$. This pattern of 'jumping back and forth' keeps going on and is stated as a theorem in [1]:

Property of sequence (1) (see Theorem 1 of [1])

1. If $(a_n,\ b_n)$ satisfies $2{a_n}^2+1={b_n}^2$, then $(a_{n+1},\ b_{n+1})$ will satisfy $2{a_{n+1}}^2-1={b_{n+1}}^2$.
2. If $(a_n,\ b_n)$ satisfies $2{a_n}^2-1={b_n}^2$, then $(a_{n+1},\ b_{n+1})$ will satisfy $2{a_{n+1}}^2+1={b_{n+1}}^2$.

Paper [1] then pointed out that the pairs in this sequence are in fact points jumping back and forth between two hyperbolas

$$2x^2+1=y^2 \tag{2}$$

and

$$2x^2-1=y^2 \tag{3}$$

The equations of these two hyperbolas can be also written respectively as

$$-\frac{x^2}{(1/\sqrt{2})^2}+\frac{y^2}{1^2}=1 \tag{4}$$

and

$$\frac{x^2}{(1/\sqrt{2})^2}-\frac{y^2}{1^2}=1 \tag{5}$$

The vertices and the foci of these two hyperbolas are respectively $(0, \pm 1)$ and $(0, \pm c)$, and $(\pm 1/\sqrt{2}, 0)$ and $(\pm c, 0)$, with $c = \sqrt{(1/\sqrt{2})^2 + 1^2} = \sqrt{3/2}$. Equations of asymptotic lines for both hyperbolas are $y = \pm\sqrt{2}x$.

The sequence (1) is then generalized in [1] into a function from R^2 to R^2:

$$f(x, y) = (x + y, 2x + y) \tag{6}$$

This function has an inverse function

$$f^{-1}(x, y) = (y - x, 2x - y) \tag{7}$$

What is interesting here is that the function $f(x, y)$ is one-to-one and onto not only from R^2 to R^2, but also from the hyperbola (2) to hyperbola (3) and from the hyperbola (3) to hyperbola (2).Therefore, starting from any point (x, y) on hyperbola (2) (or hyperbola (3)), the points in the sequence $(x, y), f(x, y), f(f(x, y)), \ldots$ will jump back and forth between these two hyperbolas. The same is true for the inverse function f^{-1}.

Paper [1] discussed a few possible ways of using the above observations to form interesting examples in teaching high school or college mathematics courses, such as algebra, geometry, calculus, and linear algebra. In this paper, we further generalize this example to develop a few interesting case study projects for courses such as real functions, analytic geometry, complex variables, and so on.

2. Case study proposals

In this section, we present a few proposals for possible student case study projects for college mathematics courses such as real functions, analytic geometry, and complex variables. For related subjects in these courses see [2–7].

2.1. *Case study one: where does $f^n(x, y)$ approach when $n \to \infty$?*

Consider the function $f(x, y) = (x + y, 2x + y)$, which is an invertible function from R^2 to R^2. For a point (x, y), let us use the notations $f^2(x, y)$ for $f(f(x, y))$, $f^3(x, y)$ for $f(f^2(x, y))$, and in general $f^n(x, y)$ for $f(f^{n-1}(x, y))$. One interesting question for students to consider is: starting with an arbitrary point (x, y) on hyperbola (2), where does the point $f^n(x, y)$ approach when n becomes bigger and bigger?

We display the two hyperbolas (2) and (3) in figure 1.

The instructor may suggest that students examine the following four cases for the starting point.

1. The starting point is the vertex $(0, 1)$ or in the first quadrant of the x–y plane. In this case it should be easy for students to see that for all n the point $f^n(x, y)$, while jumping back and forth between the two hyperbolas, will always stay in the first quadrant. When $n \to \infty$, $f^n(x, y)$ will approach (∞, ∞).
2. The starting point is the vertex $(0, -1)$ or in the third quadrant of the x–y plane. Similar to case 1, while jumping between hyperbolas (2) and (3), the point $f^n(x, y)$ will always stay in the third quadrant and goes to $(-\infty, -\infty)$ when $n \to \infty$.

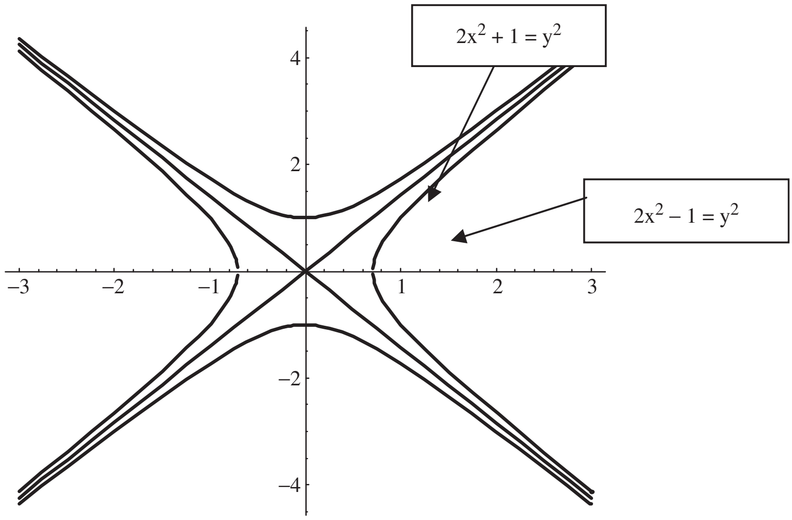

Figure 1. Hyperbolas (2) and (3).

3. The starting point (x, y) locates in the second quadrant, that is, $x < 0$ and $y > 0$. What happens in this case? The instructor may first encourage students to generate and observe some numerical results. For example, if we start with $(x, y) = (-12, 17)$. Then (x, y) is a point in the second quadrant. Students can easily compute that $f(x, y) = (5, -7)$, $f^2(x, y) = (-2, 3)$, $f^3(x, y) = (1, -1)$ and $f^4(x, y) = (0, 1)$. Once $f^4(x, y)$ is computed, from case 1 students can tell that for all $n > 4$, $f^n(x, y)$ will stay in the first quadrant and $f^n(x, y)$ approaches (∞, ∞) when $n \to \infty$.

This numerical example seems to suggest that when the starting point (x, y) is in the second quadrant, then when n gets bigger the point $f^n(x, y)$ will eventually get into the first quadrant of x–y plane. This in fact is true, and can be stated as the following theorem.

Theorem 1: Starting with any point (x, y) that is on hyperbola (2) and in the second quadrant of the x–y plane, the point $f^n(x, y)$ will eventually get into and stay in the first quadrant when n is large enough, and approach (∞, ∞) when $n \to \infty$.

With appropriate guidance from the instructor, students should be able to reach a proof to this theorem. Since the starting point (x, y) is on hyperbola (2), for all even integers n the point $f^n(x, y)$ will also be on hyperbola (2). Hence the instructor may encourage students to concentrate on the points $(x, y), f^2(x, y), f^4(x, y), \ldots$. It should be easy for students to find out that

$$f^2(x, y) = f(f(x, y)) = (3x + 2y, 4x + 3y).$$

Now let $u = 3x + 2y$, $v = 4x + 3y$, and also note that $x < 0$, $y > 0$ and

$$2x^2 + 1 = y^2. \tag{8}$$

Equation (8) implies that

$$y > \sqrt{2}|x|$$

and hence

$$3y > 3\sqrt{2}|x| > 4|x|. \tag{9}$$

This implies that

$$v = 4x + 3y > 0. \tag{10}$$

Now students should examine the value of $u = 3x + 2y$. Note that (8) indicates that $y > 1$. Hence if $|x| < 2/3$ then $2y > 3|x|$, which further implies that $u > 0$ and thus $f^2(x, y)$ will be in the first quadrant.

Therefore, when $|x| < 2/3$, that is, when $-2/3 < x < 0$, the point $f^2(x, y)$ will get into the first quadrant, and by case 1 we know that the theorem holds true.

What happens if $|x| \geq 2/3$? Students should note that (10) still holds, that is, the value of v is still positive. If u happens to be non-negative, then by case 1 the result of the theorem holds. If $u < 0$, then it must be that $2y < 3|x|$ and $|u| = 3|x| - 2y$. Since $y > \sqrt{2}|x|$ we have

$$|u| = 3|x| - 2y < (3 - 2\sqrt{2})|x| < \frac{1}{4}|x| \tag{11}$$

This shows that although $f^2(x, y) = (u,v)$ is still in the second quadrant, it moves towards the vertex (0, 1) since $|u| < (1/4)|x|$.

With this students should be able to complete the proof. Just see what would happen to $f^4(x, y)$. Either the point $f^4(x, y)$ gets into the first quadrant (or happens to be the vertex (0, 1)), or it stays in the second quadrant but much closer to the vertex (0, 1). More precisely, the absolute value of its x-coordinate will become less than $(1/4)|u| < (1/16)|x|$. The pattern keeps going on and when n (an even integer) is large enough, the point $f^n(x, y)$ will either be in the first quadrant (or on the vertex (0, 1)), or the absolute value if its x-coordinate becomes less than 2/3 so that the next point $f^{n+2}(x, y)$ will get into the first quadrant. Therefore the result of the theorem is also true when $|x| \geq 2/3$, and the proof is completed.

4. The starting point (x, y) locates in the fourth quadrant, that is, $x > 0$ and $y < 0$. Students should be able to similarly prove that in this case $f^n(x, y)$ will eventually get into and stay in the third quadrant, and approaches $(-\infty, -\infty)$ when $n \to \infty$. The instructor may want to encourage students to complete this proof independently.

A similar question may be asked about hyperbola (3), that is, starting with an arbitrary point (x, y) on hyperbola (3), where does the point $f^n(x, y)$ approach when $n \to \infty$? Perhaps students now can guess the answer out: if (x, y) is on the right branch (that is, on the right side of the y-axis) of the hyperbola, then $f^n(x, y)$ will eventually get into the first quadrant and approach (∞, ∞) when $n \to \infty$; if (x, y) is on the left branch of the hyperbola, then $f^n(x, y)$ will eventually get into the third quadrant and approach $(-\infty, -\infty)$ when $n \to \infty$.

In order to prove this result, students may tend to follow a procedure similar to the above arguments in case 3. However, the instructor may want to encourage them to try to make a short cut by simply observing the location of the point $f(x, y) = (x + y, 2x + y)$. For instance, suppose the starting point (x, y) is on the right branch, that is, $x > 0$ and $2x^2 - 1 = y^2$, then it is easy to see that $\sqrt{2}x > |y|$ and hence

$$2x + y > \sqrt{2}x - |y| > 0 \tag{12}$$

This implies that $f(x, y)$ is on the upper branch (that is, above the x-axis) of hyperbola (2). The previous arguments for case 1 and case 3 then imply the result here.

At this point, students should be able to make another interesting observation: we knew before that $f(x, y)$ maps hyperbola (2) one-to-one and onto (3) and (3) to (2), now we know that $f(x, y)$ in fact does this between the upper branch of (2) and the right branch of (3), and between the lower branch of (2) and the left branch of (3).

The same questions may be considered for points on the asymptotic lines $y = \pm\sqrt{2}x$ of these two hyperbolas. If a point (x, y) is on the line $y = \sqrt{2}x$, then $(x, y) = (x, \sqrt{2}x)$ and

$$\begin{aligned} f(x, y) &= (x + y, 2x + y) \\ &= \left(x + \sqrt{2}x, 2x + \sqrt{2}x\right) \\ &= \left(x + \sqrt{2}x, \sqrt{2}\left(x + \sqrt{2}x\right)\right) \end{aligned} \tag{13}$$

Hence $f(x, y)$ is also on the line $y = \sqrt{2}x$. Not only that, (13) also implies that

$$\begin{aligned} f(x, y) &= (1 + \sqrt{2})(x, \sqrt{2}x) \\ &= (1 + \sqrt{2})(x, y) \end{aligned} \tag{14}$$

Therefore, starting with any point (x, y) on the first quadrant portion of the asymptotic line $y = \sqrt{2}x$, the point $f^n(x, y)$ will always be on that line, in the first quadrant, and approaches (∞, ∞) when $n \to \infty$. A similar observation can be made for the third quadrant portion of that line.

If a point (x, y) is on the other asymptotic line $y = -\sqrt{2}x$, then similarly students can see that $f(x, y)$ is also on that line, and $f(x, y) = (1 - \sqrt{2})(x, y)$. Hence, starting with any point $(x, y) \neq (0, 0)$ on that line, the point $f^n(x, y)$ will always stay on that line, jumping back and forth between the second and the fourth quadrants, and approaching $(0, 0)$ when $n \to \infty$.

2.2. *Case study two: what happens on the hyperbolas $2x^2 + r = y^2$ and $2x^2 - r = y^2$?*

The instructor may encourage students to consider a more general situation. Let r be an arbitrary positive number, and consider the pair of hyperbolas

$$2x^2 + r = y^2 \tag{15}$$

and

$$2x^2 - r = y^2 \tag{16}$$

It is easy for students to see that hyperbolas (2) and (3) give a special case of (15) and (16). If we let r be zero instead of a positive number, then (15) and (16) merge into one equation

$$2x^2 = y^2 \tag{17}$$

or equivalently

$$y = \pm\sqrt{2}xy \tag{18}$$

An interesting fact for students to observe is that (18) gives the asymptotic lines of (15) and (16) for all $r > 0$.

It can be a good project for students to observe and prove that properties of the function $f(x, y) = (x + y, 2x + y)$ discussed previously with hyperbolas (2) and (3) all remain valid with (15) and (16). For instance, if (x, y) is on (15), then $f(x, y)$ will be on (16), and the converse is also true. To prove this, let $(u, v) = f(x, y)$, that is, $u = x + y$ and $v = 2x + y$. If (x, y) is on (15), then $2x^2 + r = y^2$. This implies that

$$\begin{aligned} 2u^2 - v^2 &= 2(x + y)^2 - (2x + y)^2 \\ &= (2x^2 + 4xy + 2y^2) - (4x^2 + 4xy + y^2) \\ &= y^2 - 2x^2 \\ &= r \end{aligned} \tag{19}$$

This shows that $2u^2 - r = v^2$, that is (u, v) is on the hyperbola (16). The converse can be similarly proved by students.

Similarly, under appropriate guidance of the instructor, students should be able to show that other properties of $f(x, y)$ with (2) and (3) all hold with hyperbolas (15) and (16) as well.

2.3. *Case study three: further generalizations*

Students may also try to examine further generalizations of the function $f(x, y) = (x + y, 2x + y)$. One possibility is to consider the function $F(x, y) = (x + y, sx + y)$ with s being an arbitrary number. Clearly, $f(x, y)$ is a special case of $F(x, y)$ when $s = 2$. In that special case, $f(x, y)$ has interesting properties with the hyperbola $2x^2 + r = y^2$ for all $r > 0$ as discussed before. What about an arbitrary value for s? Does $F(x, y) = (x + y, sx + y)$ has similar properties with the curve $sx^2 + r = y^2$? The instructor may encourage students to consider the following cases.

1. $s = 0$. In this case, $F(x, y) = (x + y, y)$ and the equation $sx^2 + r = y^2$ becomes $y = \pm\sqrt{r}$ which represents two horizontal lines. It should be easy for students to see that in this case if (x, y) lies on the line $y = \sqrt{r}$ then so does the point $F^n(x, y)$ for all n, and $F^n(x, y)$ approaches $(\infty, \sqrt{r})$ when $n \to \infty$. Similarly, if (x, y) is on the line $y = -\sqrt{r}$, then $F^n(x, y)$ will always stay on that line and approaches $(-\infty, -\sqrt{r})$ $ when $n \to \infty$.
2. $s \neq 0$. It can be a good exercise for students to use the method of mathematical induction to prove the following theorem.

Theorem 2: If (x, y) satisfies the equation $sx^2+r=y^2$ with $s\neq 0$ and $r>0$, then $F(x,y)$ satisfies $sx^2+(1-s)\ r=y^2$, $F^2(x, y)$ satisfies $sx^2+(1-s)^2r=y^2$, and in general $F^n(x, y)$ satisfies the equation

$$sx^2+(1-s)^n r=y^2 \tag{20}$$

To prove this, let (x, y) be on the curve $sx^2+r=y^2$. Let $(u, v)=F(x, y)=(x+y, sx+y)$, then observe that

$$\begin{aligned} su^2-v^2 &= s(x+y)^2-(sx+y)^2 \\ &= (sx^2+2sxy+sy^2)-(s^2x^2+2sxy+y^2) \\ &= (s-1)(y^2-sx^2) \\ &= (s-1)r \end{aligned} \tag{21}$$

This implies that $su^2+(1-s)r=v^2$, and hence $F(x,y)$ satisfies the equation $sx^2+(1-s)r=y^2$.

To use the mathematical induction, assume that $(w,z)=F^k(x, y)$ is on the curve $sx^2+(1-s)^k\ r=y^2$, that is, (w, z) satisfies the equation $sw^2+(1-s)^kr=z^2$. Let $(u, v)=F^{k+1}(x, y)=F(w, z)=(w+z, sw+z)$. Similar to steps in (21) students can see that

$$su^2-v^2=(1-s)(sw^2-z^2)=-(1-s)^{k+1}r$$

which implies that $su^2+(1-s)^{k+1}r=v^2$, and by mathematical induction the theorem is proved.

The instructor may encourage students to further consider the following special cases for s.

(a) $s=2$. In this case, equation (20) becomes (15) for all even n, and (16) for all odd n. This leads to the special pattern of 'jumping back and forth' between the two hyperbolas (15) and (16), as discussed in the previous subsection.
(b) $s=1$. In this case, the function $F(x, y)$ becomes $F(x, y)=(x+y, x+y)$. Hence starting with any point (x, y), the point $F^n(x, y)$ will be on the line $y=x$ for all n, which shows another special case of Theorem 2.
(c) When $0<|1-s|<1$, that is, when $0<s<1$ or $1<s<2$, the curve $sx^2+(1-s)^nr=y^2$ gets closer and closer to its asymptotic lines $y=\pm\sqrt{s}x$ in the sense that the value of $(1-s)^nr$ approaches zero when $n\to\infty$. While when $s>2$, the curve $sx^2+(1-s)^nr=y^2$ gets farther and farther away from its asymptotic lines in the sense that the value of $(1-s)^nr$ approaches infinity when $n\to\infty$.
(d) It is particularly interesting to examine the case when $s<0$. In this case the curves $sx^2+(1-s)^nr=y^2$ are no longer hyperbolas but ellipses for all n. Therefore, starting with any point (x, y) on the ellipse $sx^2+r=y^2$, the points $F^n(x, y)$ will travel through the ellipses $sx^2+(1-s)^nr=y^2$ for $n=1, 2,\ldots$. Not only that, the value of $1-s$ is always bigger than one for all negative s. Hence the ellipse $sx^2+(1-s)^nr=y^2$ gets farther and farther away from the origin (0,0) when $n\to\infty$.

(e) Finally, students may also take a look at the inverse function $F^{-1}(x, y)$ when $s \neq 1$. It should be easy for them to find the formula of $F^{-1}(x, y)$, which is

$$F^{-1}(x,y) = \frac{1}{1-s}(x-y, y-sx). \tag{22}$$

Clearly, this is a generalization of the inverse function formula given in (7), which is for the special case when $s=2$. It is also a good exercise for students to prove the following Theorem 3.

Theorem 3: Starting with a point (x, y) on the curve $sx^2+r=y^2$ where $s\neq 1$ and $r > 0$, for all $n=1, 2, \ldots,$ the point $F^{-n}(x, y)$ will be on the curve

$$sx^2 + \frac{1}{(1-s)^n} r = y^2.$$

Students may also notice that when $s < 0$, the ellipse

$$sx^2 + \frac{1}{(1-s)^n} r = y^2$$

approaches (0, 0) when $n \to \infty$. Hence, an immediate corollary of Theorem 3 is that if (x, y) is a point on the ellipse $sx^2+r=y^2$ then the point $F^{-n}(x, y) \to (0, 0)$ when $n \to \infty$.

Finally, the instructor may mention to students that every point (a, b) except for (0, 0) is on some ellipse of the form $sx^2+r=y^2$ for some $s < 0$ and $r > 0$. For example, (a, b) satisfies the equation of the ellipse

$$-x^2 + (a^2 + b^2) = y^2 \tag{23}$$

which has the form of $sx^2+r=y^2$ with $s=-1 < 0$ and $r=a^2+b^2 > 0$. Therefore the function $F(x, y)=(x+y, -x+y)$ will make $|F^n(a, b)| \to \infty$ when $n \to \infty$, and the inverse function $F^{-1}(x, y)=\frac{1}{2}(x-y, y+x)$ will make $|F^{-n}(a, b)| \to 0$ when $n \to \infty$. Meanwhile, the point (a, b) is also on the ellipse

$$-2x^2 + (2a^2 + b^2) = y^2 \tag{24}$$

Therefore, $F(x, y)=(x+y, -2x+y)$ and its inverse $F^{-1}(x, y)=\frac{1}{3}(x-y, y+2x)$ will display a similar pattern with (a, b).

2.4. *Case study four: view F(x, y) as a complex function*

In a complex variable course, the function $F(x, y)=(x+y, sx+y)$ may be viewed as a complex function $w=F(z)$ with $z=x+yi$ and $w=u+vi$, where $u=x+y$ and $v=sx+y$. The function $w=F(z)$ represents a mapping from the complex plane C onto itself. As discussed before, the function $F(z)$ maps the curve $sx^2+r=y^2$ onto $sx^2+(1-s)r=y^2$, and so on. The function $F(z)$ is also invertible when $s\neq 1$, with

$$F^{-1}(z) = \frac{1}{1-s}(x-y) + \frac{1}{1-s}(y-sx)i$$

Now an interesting question for students to consider is whether $F(z)$ is an analytic function or not. By the Cauchy–Riemann Theorem, $F(z)$ is analytic if and only if the partial derivatives satisfy the Cauchy-Riemann equations

$$u_x = v_y \text{ and } u_y = -v_x.$$

For this function, we have that $u_x = 1$, $u_y = 1$, $v_x = s$, and $v_y = 1$. Thus students can easily see that $F(z)$ is analytic if and only if $s = -1$. When $s = -1$, the function $F(z) = (x + y) + (-x + y)i$ is analytic and its derivative $F'(z) = u_x + v_x \quad i = 1 - i$ for all z.

Also note that in this case $F(z)$ can be written as

$$F(z) = (1 - i)z. \tag{25}$$

Hence the inverse function

$$F^{-1}(z) = \frac{1}{1 - i}z = \frac{1 + i}{2}z \tag{26}$$

Students can easily verify that (26) is consistent with the formula (22) for the case when $s = -1$. In fact,

$$\frac{1 + i}{2}z = \frac{1 + i}{2}(x + yi) = \frac{1}{2}[(x - y) + (x + y)i] = \frac{1}{2}(x - y, x + y)$$

which matches (22) with $s = -1$.

Since $F^{-1}(z) = ((1 + i)/2)z$ it is also analytic and $[F^{-1}(z)]' = (1 + i)/2$. The instructor may also ask students to verify that $F^{-1}(z)$ does satisfy the Cauchy–Riemann equation, and Cauchy–Riemann derivative formula $u_x + v_x \quad i = \frac{1}{2} + \frac{1}{2}i$ yields the same formula for the derivative of $F^{-1}(z)$.

It is also interesting for students to note that, when $s = -1$, the Theorem 2 in section 2.3 leads to a special case about *circles*. In fact, when $s = -1$ the equation $sx^2 + r = y^2$ becomes the equation of a circle $x^2 + y^2 = r$ with $r > 0$. Therefore, this special case of Theorem 2 can be stated as the following corollary.

Corollary of Theorem 2: For any $z = x + yi$ on the circle $x^2 + y^2 = r$ with $r > 0$, $F(z) = (x + y) + (-x + y)i$ is on the circle $x^2 + y^2 = r$, $F^2(z)$ is on the circle $x^2 + y^2 = 4r$, and in general $F^n(z)$ is on the circle $x^2 + y^2 = 2^n r$.

Since $F^{-1}(z)$ exists, $F(z)$ yields a one-to-one and onto mapping from the circle $x^2 + y^2 = r$ to the circle $x^2 + y^2 = 2r$, and in general, $F^n(z)$ is a one-to-one and onto mapping from the circle $x^2 + y^2 = r$ to the circle $x^2 + y^2 = 2^n r$. Figure 2 displays this mapping when $r = 1$, starting with the unit circlen $x^2 + y^2 = 1$.

Finally, the instructor may mention to students that, since $1 - i = \sqrt{2}e^{(-\pi/4)i}$, the function $F(z) = (1 - i)z$ can also be written as

$$F(z) = \sqrt{2}e^{(-\pi/4)i}z \tag{27}$$

Therefore, $F(z)$ may be viewed as a transformation of complex vectors in the complex plane. When applying to a complex vector z, this transformation first

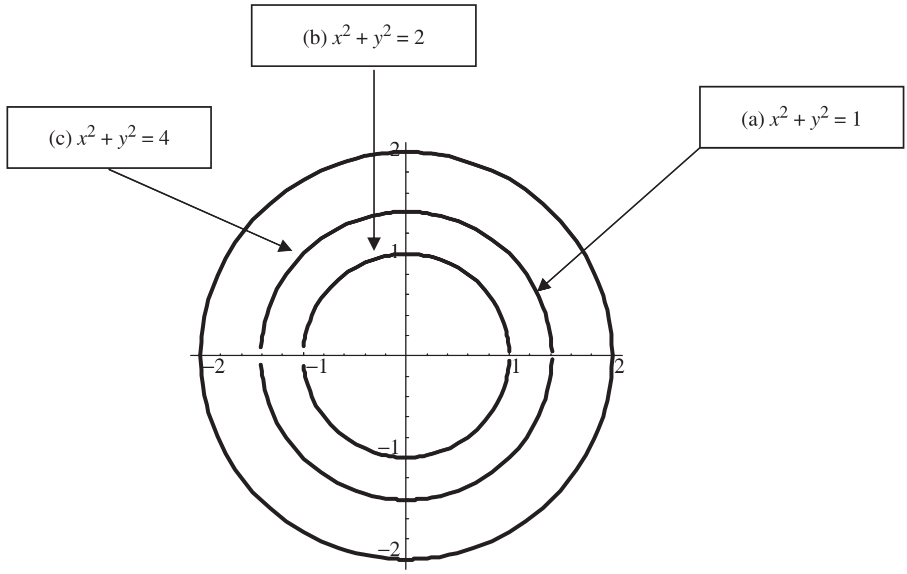

Figure 2. $F(z)$ maps circle (a) on to circle (b), and circle (b) on to circle (c), ...

Table 1. Pattern of $F(z)$.

Z_0	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
(1, 0)	(1, −1)	(0, −2)	(−2, −2)	(−4, 0)	(−4, 4)	(0, 8)	(8, 8)	(16, 0)

extends the length of z by a factor $\sqrt{2}$, and then rotates it with an angle $-\pi/4$. Thus if we start at some point on the positive x-axis, say $z_0 = (r, 0)$ for some $r > 0$, and let $z_n = F(z_{n-1})$ for $n = 1, 2, \ldots,$ then z_1 will sit on the line $y = -x$ in the fourth quadrant with $|z_1| = \sqrt{2}r$, z_2 will sit on the negative y-axis with $|z_2| = 2r$, z_3 will be on the line $y = x$ in the third quadrant with $|z_3| = 2\sqrt{2}r$, z_4 will be on the negative x-axis with $|z_4| = 4r$, and the pattern keeps going on. Table 1 displays this pattern with the first nine vectors in this sequence, starting with $z_0 = (1, 0)$.

3. Conclusions

In this paper, we have presented a few interesting case study proposals that can be used as student projects in college mathematics classes such as real functions, analytic geometry, and complex variables. In addition to using them in individual courses, these studies may also be combined to offer seminars or workshops to college mathematics students. Projects like these are likely to promote student interest and get students more involved in the learning process, and therefore make the learning process more effective.

References

[1] Mauch, E. and Shi, Y., 2005, Using a sequence of number pairs as an example in teaching mathematics. *Mathematics and Computer Education*, **39**(3), 198–205.
[2] Churchill, R. and Brown, J., 1990, *Complex Variables and Applications*, 5th edn (New York: McGraw-Hill).
[3] Goldberg, R., 1976, *Methods of Real Analysis*, 2nd edn (New York: John Wiley & Sons).
[4] Kay, D., 2001, *College Geometry, a Discovery Approach*, 2nd edn (New York: Addison-Wesley).
[5] Marsden, J. and Tromba, A., 1988, *Vector Calculus*, 3rd edn (New York: W.H.Freeman and Company).
[6] O'Daffer, P. and Clemens, S., 1992, *Geometry: an Investigative Approach*, 2nd edn (New York: Addison-Wesley).
[7] Tan, S.T., 1999, *Applied Calculus*, 4th edn (New York: Brooks/Cole).

A simple proof of the generalized Ceva theorem by the principle of equilibrium

Y.-K. MAN*
The Hong Kong Institute of Education, Hong Kong

(*Received 7 September 2005*)

In this note, a simple proof of the Generalized Ceva Theorem in plane geometry is presented. The approach is based on the principle of equilibrium in mechanics.

1. Introduction

A cevian is a line segment which joins a vertex of a triangle with a point on the opposite side (or its extension). It is named in honour of the Italian mathematician Giovanni Ceva (1647–1734), who discovered the necessary and sufficient condition for the concurrence of three cevians in 1678. This result states:

Theorem 1 (The Ceva Theorem)**:** *In figure* 1, *the cevians AD*, *BE*, *CF are concurrent if and only if*

$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=1.$$

*Email: ykman@ied.edu.hk

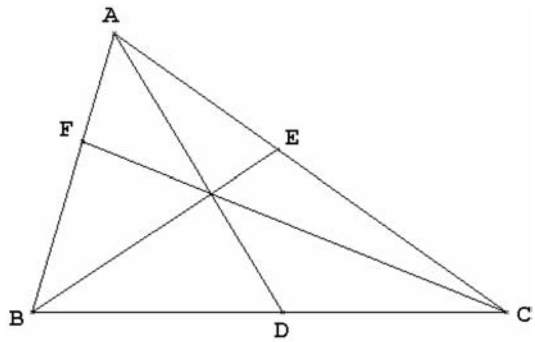

Figure 1. The cevians AD, BE and CF are concurrent at a point inside $\triangle ABC$.

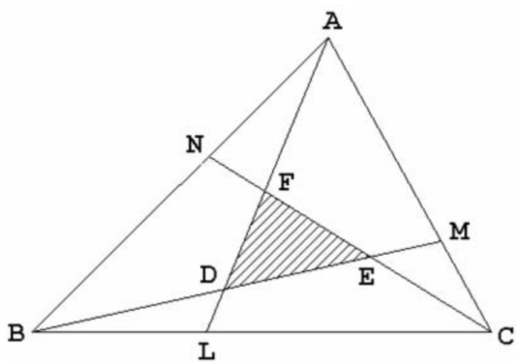

Figure 2. The area of $\triangle DEF$ can be determined by means of the generalized Ceva theorem.

A proof of this classic result can be found in [1]. In [2], a generalized version of the Ceva Theorem is described by Coxeter. In the next section, we present a simple proof of this result, based on the principle of equilibrium in mechanics, which is different from the approach adopted by Coxeter.

2. A simple proof of the generalized Ceva theorem

Theorem 2 (The Generalized Ceva Theorem)**:** *In figure 2, if the points L, M, N divide the sides of ΔABC such that $BL : LC = p : 1$, $CM : MA = q : 1$ and $AN : NB = r : 1$, then the area of ΔDEF formed by AL, BM and CN is equal to*

$$\frac{(pqr-1)^2}{(p+pq+1)(q+qr+1)(r+pr+1)} \times S_{ABC},$$

where S_{ABC} denotes the area of ΔABC.

Proof: First, we suspend masses q, qr and 1 at A, B and C respectively. Then, N is the centre of gravity of AB, with a mass of $q+qr$. Also, M is the centre of gravity of AC, with a mass of $q+1$. The centre of gravity of ABC must lie on both CN and BM, which means at E. For equilibrium, the moments of the masses at B and M about E must be equal. So, we have $qr \times BE = (q+1) \times EM$, which means $BE : EM = (q+1) : qr$. Hence,

$$\begin{aligned}\text{area of } \Delta BEC &= \frac{q+1}{q+qr+1} \times \text{area of } \Delta BMC = \frac{q+1}{q+qr+1} \times \frac{q}{q+1} \times S_{ABC} \\ &= \frac{q}{q+qr+1} \times S_{ABC}\end{aligned}$$

Next, we suspend masses r, pr and 1 at B, C and A respectively. By the same arguments, we obtain

$$\begin{aligned}\text{area of } \Delta AFC &= \frac{r+1}{r+pr+1} \times \text{area of } \Delta ANC \\ &= \frac{r+1}{r+pr+1} \times \frac{r}{r+1} \times S_{ABC} = \frac{r}{r+pr+1} \times S_{ABC}\end{aligned}$$

Third, we suspend masses p, pq and 1 at C, A and B respectively. Then,

$$\begin{aligned}\text{area of } \Delta ABD &= \frac{p+1}{p+pq+1} \times \text{area of } \Delta ABL \\ &= \frac{p+1}{p+pq+1} \times \frac{p}{p+1} \times S_{ABC} = \frac{p}{p+pq+1} \times S_{ABC}\end{aligned}$$

Hence,

$$\begin{aligned}\text{area of } \Delta DEF &= \left(1 - \frac{q}{q+qr+1} - \frac{r}{r+pr+1} - \frac{p}{p+pq+1}\right) \times S_{ABC} \\ &= \frac{(pqr)^2 - 2pqr + 1}{(p+pq+1)(q+qr+1)(r+pr+1)} \times S_{ABC} \\ &= \frac{(pqr-1)^2}{(p+pq+1)(q+qr+1)(r+pr+1)} \times S_{ABC}.\end{aligned}$$

When $pqr = 1$, it implies that the area of ΔDEF is equal to zero and the cevians are concurrent. Obviously, the converse is also true. In other words, the Ceva theorem is simply a special case of the above theorem. By using the principle of equilibrium in mechanics, we can see that this classic result can be proved with surprising ease.

References

[1] Smith, J.T., 2000, *Methods of Geometry* (New York: John Wiley & Sons).
[2] Coxeter, H.S.M., 1969, *Introduction to Geometry* (New York: John Wiley & Sons).