Summer 2021

The Fermat-Torricelli Point and Cauchy's Method of Gradient Descent

Kenneth M. Monks
kmmonks@gmail.com

The Fermat-Torricelli Point and Cauchy's Method of Gradient Descent

Kenneth M Monks *

July 19, 2021

Marin Mersenne (1588–1648) was the one-man "Europe-Wide Web" of seventeenth century mathematics[1]. Communication at great distances was inconvenient in this era, to say the least; even the telegraph was still roughly 200 years into the future. Say you were a lawyer/jurist in Toulouse, France, making good money but bored with your job, doing mathematics on the side for challenge and stimulation. There wasn't exactly a thriving mathematics community in Toulouse (which lies more than 400 miles to the south of the active community of mathematicians working in Paris) for you to share your thoughts with, but you could mail them to Mersenne! He would then copy, record, and distribute the works sent to him. Now, say you were a physicist in Florence, putting mathematics to good use in continuing the legacy of his mentor Galileo Galilei, and you received some of these letters from Mersenne. Well, you could reply with questions or news of progress back to Mersenne, or maybe you could even meet with him to talk and exchange materials in person on one of his journeys around the continent.

Our hypothetical people described above are anything but speculative; Pierre de Fermat (1601–1665) was exactly that bored, geographically-isolated lawyer, and Evangelista Torricelli (1608–1647) was that Florentine physicist[2] who was lucky enough to not just have been a frequent correspondent of Mersenne's, but also to have been visited by Mersenne on his trip to Rome in 1644.

Here we state just one of the many problems posed to the world by Fermat. It was communicated to Torricelli via Mersenne (likely through a letter early in 1644), as it appears in [Henry and Tannery, 1891, Vol 5, 127].[3]

> Given three points, find another point, from which one draws three straight lines to the three given points, such that these distances are minimal; that is, together their sum is less

*Department of Mathematics, Front Range Community College – Boulder County Campus, Longmont, CO 80537; kenneth.monks@frontrange.edu.

[1]In 1611, he joined the Roman Catholic Order of the Minims, a group that considered humility to be the chief virtue, and thus considering themselves as the least of all, the *minimi*. They devoted themselves to austere lifestyles focused on prayer and scholarship. Mersenne certainly took the latter to heart, having communicated and/or worked on mathematics with an extensive list of contemporary mathematicians, including Galileo Galilei (1564–1642), René Descartes (1596–1650), Blaise Pascal (1623–1662), Christiaan Huygens (1629–1695), Gilles Roberval (1602–1675), and John Pell (1611–1685), as well as the first two namesakes of this paper! For more details, see [O'Connor and Robertson, 2005]. Could he have been the *maximi minimi*?

[2]Though his contributions to mathematics were substantial, he is today perhaps most widely known for his experiments regarding pressure and vacuum, as well as the invention of the barometer.

[3]All translations of Torricelli excerpts in this project, unless otherwise noted, were prepared by the project author, 2021.

than the sum of any other three which could be drawn from any other point to the three given points.

Task 1 In order to visualize what is being asked here a bit, draw a diagram with three points labelled A,B, and C, and a fourth point between them labelled F, representing the point we seek that will minimize the sum of the distances. Also draw the line segments AF, BF, and CF.

This problem was actually the third in a sequence of problems that had been posed as challenges by Fermat which Torricelli had been working on. In the next paragraph, Torricelli lamented his current lack of progress.

These three problems[4] ... are from ... Fermat .... None of these have been proven by me, and I believe the proof remains in the hands of the author.

The interaction characterized above exemplified Fermat's unique personal style with regard to mathematics collaboration; he had a reputation for coming up with results in secret and then sending the result out into the mathematical community with no indication of how one might have come upon it, as a puzzle for the world to solve! Historian of mathematics Victor Katz commented on this [Katz, 1998, page 433]: "In many cases it is not known what, if any, proofs Fermat constructed nor is there always a systematic account of certain parts of his work. Fermat often tantalized his correspondents with hints of his new methods for solving certain problems. He would sometimes provide outlines of these methods, but his promises to fill in gaps 'when leisure permits' frequently remained unfulfilled."

Torricelli excitedly wrote to his friend and fellow Italian mathematician Michelangelo Ricci (1619–1682) on November 7, 1646 (see [Henry and Tannery, 1891, Vol 5, pp 141]).

I don't know how in these past days I solved a problem of Fermat... I will try to send the proof.

Task 2 (a) What is the longest amount of time you have ever been stuck on a single mathematics problem before finally solving it? How did it feel when you did?

[4]The first two of the problems suggested by Fermat that Torricelli tried had nothing to do with multivariable calculus and thus are not studied in this project, but we state them here to quell your curiosity! The first problem was to find a Pythagorean triple, that is, positive integers a, b, c for which $a^2 + b^2 = c^2$, for which the quantities c, $a + b$, and $b + c$ are all perfect squares. Torricelli's lack of progress on this problem was avenged by a fellow Italian almost three hundred years later, when it was solved by number theorist Michele Cipolla (1880–1947) of Palermo, Sicily. The second problem asked if numbers of the form $2^{2^n} + 1$ for natural numbers n are always prime. It turned out that they are not always prime, when they are we today call them *Fermat primes* in honor of this proposition.

(b) Notice the dates of Torricelli's letters. For roughly how long was he stuck on this problem from Fermat before he solved it?

In the margin of the paper containing the third problem, Torricelli wrote the following comment (see [Henry and Tannery, 1891, pg 127, fn 2]).

> This was then proven by me in three different ways and the proof was published in Florence, Rome, Pisa, Bologna, and in France so that others cannot boast.[5]

1 The Geometric Median

1.1 General Comments on the Geometric Median Problem

Torricelli had good reason to be excited about solving this problem, as did Fermat for having asked it in the first place! This problem can be seen as one of the first questions ever asked in the branch of mathematics known as *operational research*, the study of using mathematics to make better decisions in problems of government and industry. Imagine a shipping company has three clients and wants to locate a warehouse in a manner that makes it most efficient to serve their clients; in this scenario, choosing the point that minimizes the sum of the three distances from the warehouse to the clients might make a lot of sense.[6] What perhaps makes less sense is the restriction of there being only three clients.

This brings us to the more general version of Fermat's question: the problem of finding a *geometric median* of a set of n points, where n is a positive natural number. More specifically, if we are given a set of n points in $\mathbb{R}^2$,

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n),$$

then the geometric median of that set is the point (x, y) that minimizes the sum of distances from (x, y) to (x_i, y_i) for $i \in \{1, 2, \ldots, n\}$. If $n = 3$, then the geometric median is called the *Fermat-Torricelli Point* of the triangle that has those three points as vertices, in honor of the problem's history.

Why is the above-defined point called the "geometric median"? Well, in a sense, it generalizes the ordinary median[7] that you may have seen as a measure of central tendency in an introductory statistics course, as the median is the point that minimizes the sum of distances to all the points in a data set. We won't formally prove this fact, [8] as it is a bit of a detour from our main mission here,

[5] Do you think Torricelli was excited to have solved Fermat's problem?

[6] At the author's home institution, where the Great Plains meet the Rocky Mountains, one could frame it more adorably as the location mama prairie dog should dig her hole to minimize the sum of the distances to her three pups' holes.

[7] And much like the ordinary median, it does not have to be one of the actual points in your data set. It can lie between some of the given points!

[8] The key idea is this: if x is the median of your data set, then making an infinitesimally small change in x will increase its distance to half of the points while decreasing its distance from the other half, each by that same infinitesimal amount, and thus we have a net change of zero with regards to the sum of the distances. However, it takes a bit of fiddly casework to flesh this out into a formal proof, hence our decision to be satisfied with a so-called "proof by example", which does not count as a real proof these days, but still provides the reader with some heuristic intuition.

but let us at least see an example of this.

Task 3 (a) Calculate the median of the data set $2, 3, 5, 6, 6$.

(b) Explain why the quantity $|x - a|$ represents the distance between two real numbers x and a on the real number line. (Essentially this is a one-dimensional distance formula.)

(c) Write down the function $f(x)$ that represents the sum of distances on the number line from a point x to each of $2, 3, 5, 6$ and 6.

(d) Graph $f(x)$ using a graphing utility. Where is the absolute minimum of this graph? How does it compare to the median you computed in part (a) of this task?

After that layover in our one-dimensional airport, let us now get back on our two-dimensional plane.

Task 4 In this task, we rewrite the geometric median problem using a bit more symbolic algebra and standard multivariable calculus mindset.

(a) What is the formula for the distance between two points in the plane, (a, b) and (c, d)?

(b) Use that formula to write down an algebraic expression for the function $f(x, y)$ that represents the sum of the distances from a point (x, y) to each of $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Note that each of the x_i and y_i represent fixed numerical values, so f is really just a function of the two variables x and y.

(c) Notice that $f(x, y)$ can be represented as a *surface*: a two-dimensional graph given by $z = f(x, y)$ living in $\mathbb{R}^3$. What is the standard algorithm from your multivariable calculus course for finding extrema on a surface?

1.2 Torricelli's 3-point Geometric Median Solution

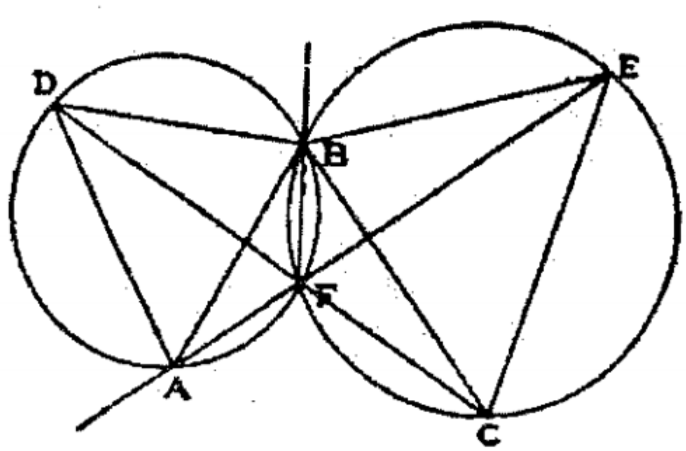
Torricelli, working in the mid 1600s, did not yet have the well-developed calculus toolbox that would exist even 100 years later[9]. Thus, his method(s) of solution to this problem were heavily based in more classical geometry. Let us consider his solution and accompanying diagram from December 1646[10], as found in [Loria and Vassura, 1919, 96].[11]

> ... having built the equilateral triangles ADB and BEC, construct a circle about each of them, two circles whose circumferences meet each other in F...

[9] After all, it was Fermat himself that was developing the technique that would grow into the algorithm for optimization taught in calculus courses today, and as mentioned previously, he was not particularly forthcoming with regard to his methods!

[10] Note that here we only give Torricelli's solution and not his justification for why his point F minimizes the sum of the distances. However, this is not because Torricelli shared Fermat's preference for withholding such details, but rather so that we do not stray too far from our focus on techniques from multivariable calculus. To read his full argument, see [Loria and Vassura, 1919, 90–97].

[11] The translation of this excerpt from Torricelli was prepared by Daniel E. Otero, Xavier University, 2021.



Let us now analyze both Torricelli's approach and the modern multivariable calculus approach in an example and then compare the results.

Task 5 Consider a triangle with vertices at $(0,0)$, $(2,2)$, and $(0,3)$. We wish to find the Fermat-Torricelli point of this triangle, or equivalently, the geometric median of these three points.

(a) What is the equation of a circle, centered at (h,k) with radius r? Write it down so that it is fresh in your head; we will use it many times in the following parts of this exercise!

(b) Conveniently, it doesn't matter which two of the three sides of our triangle we choose to perform Torricelli's construction on. Pick whatever your favorite two sides are to label as AB and BC. We will perform his construction via the following steps.

(i) Find the coordinates of the points D and E, to complete the desired equilateral triangles. To accomplish this, write down the equations of two circles, one centered at A and one centered at B, both with radius AB. The circles will intersect in two points; decide which of the two points corresponds to the point D shown in Torricelli's diagram, and label it so. The same process will find point E for you, *mutatis mutandis*[12].

(ii) Now, as Torricelli said, "having built the equilateral triangles ADB and BEC, construct a circle about each of them". To find the equation of the circle containing points A, B, and D, it is probably easiest to just write down a generic equation of a circle and then plug in the points A,B, and D, one at a time, to get a system of three equations in three unknowns (h, k, and r). Carry out this procedure, enlisting help from a computer algebra system if so desired. Then repeat to perform the analogous process for points B, C, and E to find the second circumcircle.

(iii) Find the intersection points of your two circumcircle equations. One of course, should be the point B, but the other should be the Fermat-Torricelli point, since Torricelli promised that the circles' "circumferences meet each other in F".

[12]This Latin phrase is often used in situations like this, where one is performs an analogous process to what has already been done, but must make some necessary changes to implement the process in the new case.

(iv) Use your favorite graphing utility to graph all of the objects you constructed above; make sure things look like they should! There is no better way to catch an algebra error than to see it produce a highly wonky diagram.

(c) Pick any other point inside the triangle ABC (yes, literally any one you want) and call it G. Calculate the sum of the distances AG, BG and CG. How should this compare to the sum of the distances AF, BF and CF? Calculate that sum as well and verify that they compare as they should. (Here is an alternate option! Students familiar with the system Geogebra may wish to instead perform the actual ruler-and-compass construction that builds Torricelli's point as well as point G, in which case one can quickly produce many examples of sums of distances which compare as they should to the sum of AF, BF and CF.)

Having found the Fermat-Torricelli point using Torricelli's technique, we now consider his result from the perspective of multivariable calculus.

Task 6

(a) Write down the formula for $f(x, y)$ from Task 4, this time instantiated with the points A, B, and C from Task 5.

(b) Explain why the point F found in Task 5 should be the absolute minimum of the function f.

(c) Calculate the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

(d) Recall that if f is differentiable and has a minimum at a point, all of the partial derivatives must equal zero at that point[13]. Set both of the partials equal to zero and describe what would be involved if one were to solve the resulting system of equations for x and y. Do not actually carry this out; the point here is to observe how difficult it is to solve that system of equations, and in particular how much harder the algebra would be than it was to carry out Torricelli's geometric method.

(e) Fortunately, it is not necessary to solve for x and y from scratch, because we already know the values from Torricelli's method. Take the point F that you found in Task 5 and evaluate the two partial derivatives at that point. Verify that both partials equal zero there (thus confirming that Torricelli's geometric method did in fact find a critical point on the graph of f).

1.3 Fagnano's 4-point Geometric Median Solution

Surprisingly, it turns out that finding the geometric median of four points is easier than finding it for three! In the 1775 paper *Problemata quaedam ad methodum maximorum et minimorum spectantia*, or in English, *Some problems making use of the method of maxima and minima* [Fagnano, 1775, 295], Giovanni Fagnano[14] (1715–1797) proved that, in his own words, "The solution is simple geometry."

[13] In honor of Fermat's extensive work with respect to finding maxima and minima, this fact is often called *Fermat's Theorem* or *Fermat's Stationary Point Theorem*. For more on this in the single-variable setting, see the author's Primary Source Project *Fermat's Method for Finding Maxima and Minima*, available at `digitalcommons.ursinus.edu/triumphs_calculus/11/`.

[14] Giovanni Fagnano was a priest from Sinigaglia (then in the Papal States, now Senigallia, Italy) who worked in geometry and calculus. In addition to his work on geometric medians, he also calculated some of the integrals now commonly taught in calculus courses, including $\int \tan(x)dx = -\ln(\cos(x)) + C$ and $\int \cot(x)dx = \ln(\sin(x)) + C$.

Let's look at his statement of the solution below![15]

In the quadrilateral $AFED$, draw the diagonals AE, FD, and their intersection point C. Point C is to be investigated.

He then proceeded to prove that his point C minimizes the sum of distances to the vertices of the quadrilateral by showing that the sum of the distances from any other point c would be greater. To do this, he repeatedly used the result known today as the *triangle inequality*. This rather intuitive fact says that in any triangle, the sum of any two side lengths must always be greater than the third side. One can think of this as a special case of the planar geometry principle that *the shortest distance between two points is a straight line*, so the line segment from some point A to another point B will always be shorter than the sum of the distances of A to C and then C to B for any point C not on the line segment AB.

Let us now read about Fagnano's construction of this point C. Note that he meant something quite different than what a modern reader is used to when he wrote an expression in parentheses. The reader may wish to simply ignore the parentheticals on the first reading, which we will then explore and clarify in the task following the exerpt.

For any other point c taken inside the figure, draw the line segments cA, cF, cE and cD. Notice that $FD(FC+CD) < Fc+cD$; similarly, $AE(AC+CE) < cA+cE$, therefore also $FD(FC+CD)+AE(AC+CE) < Fc+cD+cA+cE$. Q.E.D.[16]

We now analyze Fagnano's argument.

Task 7

(a) Draw the quadrilateral $AFED$ that Fagnano described, with both points C and c labelled, and draw all of the line segments he described.

(b) In modern notation, when one writes an expression of the form $x(y+z)$, what does it typically mean? Why was this certainly not what Fagnano meant when he wrote the expressions $FD(FC+CD)$ and $AE(AC+CE)$? Instead, what was he communicating by those parentheticals?

(c) How many times did Fagnano apply the triangle inequality in his argument? In each case, identify which triangle it was applied on.

(d) Why did Fagnano's applications of the triangle inequality let him conclude that the sum of the distances from c to the vertices were larger than the sum of the distances from C to the vertices?

[15] All translations of Fagnano excerpts in this project, unless otherwise noted, were prepared by the project author, 2021.

[16] Fagnano's letters "Q.E.D." stood for "Quod Erat Demonstrandum", which translates to "what was to be shown". It was left in the translation because this Latin initialism is still sometimes used verbatim in English writing today, especially in the context of a mathematical or philosophical argument.

(e) Why does Fagnano's argument imply that the point C is the geometric median of the points A, D, E and F?

We now pause for a bit of reflection on Fagnano's argument.

Task 8 (a) What goes wrong in Fagnano's argument if the quadrilateral is not convex (i.e., if one vertex is contained inside the triangle generated by the other three)? Draw a diagram that illustrates this situation. What would the geometric median be in this case?

(b) Set up a function $f(x, y)$ that represents the sum of the distances from a generic point (x, y) to each of the four vertices, labelled with coordinates $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$ and take the partial derivatives with respect to both x and y. How difficult would it be to find the absolute minimum of this function using calculus as opposed to what Fagnano did? Do not actually carry out all the steps, but give a description of what would be involved.

2 Cauchy's Method of Gradient Descent

In mathematics and its applications, it is often the case that one can solve a problem with an exact answer for simpler instances of that problem, but for more complicated instances, one must settle for a method that provides only an approximate answer rather than an exact one. A classic example of this phenomenon is the problem of finding roots of polynomials.

Task 9 (a) Given a polynomial of degree 2, i.e., a function of the form $f(x) = ax^2 + bx + c$ for real numbers a, b, c, what is the guaranteed method by which one can find exact values of its roots?

(b) It is provably impossible to have an analogous formula for a generic polynomial of degree 5 or greater.[17] However, one can still obtain numerical approximations to the roots of any polynomial equation, no matter the degree.

(i) Consider the polynomial

$$f(x) = x^5 + x^4 - x^3 + 3x^2 - 2x + 1.$$

Calculate the values of $f(-3)$ and $f(-2)$. What theorem from your first course in calculus implies that it must have a real root for some real number $c \in (-3, -2)$? (If you don't recall, perhaps consult a calculus textbook and see what major theorem would apply to this situation.)

(ii) Use two iterations of Newton's Method to approximate this root, starting from the guess $x = -2$. Compare your result with a decimal approximation obtained via a computer algebra system.

[17] The nonexistence of a general quintic formula was established by the great Norwegian mathematician Niels Abel (1802–1829) in 1826. The original work is available here: [Abel, 1826, 66–87]. A bit of background in abstract algebra is needed to understand the rather heavy argument, but Janet Barnett's primary source project *The Roots of Early Group Theory in the Works of Lagrange*, available at `https://digitalcommons.ursinus.edu/triumphs_abstract/2`, can provide the student with some insight!

We have considered above an instance of a problem (polynomial root finding) where we have an elegant solution to simpler cases of the problem, but as the problem grows in complexity, we must instead settle for an approximate solution. It turns out that the problem of finding the geometric median of a set of n points follows exactly this behavior.

Task 10

(a) For $n = 1$, why do we have an exact solution?

(b) For $n = 2$, why do we have an exact solution?

(c) For $n = 3$, why do we have an exact solution?

(d) For $n = 4$, why do we have an exact solution?

It turns out that for $n \geq 5$, there is provably no general formula that provides an exact answer for the geometric median of a set of n points.[18] Thus, we now turn to a method that will provide a numerical approximation to the answer instead of an exact answer. The method is today referred to as *Cauchy's method of gradient descent* or simply *gradient descent*, which has become the backbone of many machine learning/AI algorithms. To see what this method is exactly, we visit Cauchy's surprisingly short (two-and-a-half pages!) paper *Methode générale pour la résolution des systèmes d'équations simultanées* (or in English, *General method for solving systems of simultaneous equations*) [Cauchy, 1847] in which he first formulated this method.[19]

It should be noted that Cauchy himself was not looking for algorithms to solve the geometric median problem. Rather, he explained his motivation in the concluding paragraph of the paper.

> One is able to draw from the principles exposed here a very advantageous part for the determination of the orbit of a star, by applying them ... to the finite equations which represent the movement of this star, and by taking for unknowns the same elements of the orbit. Therefore, ... one will be able ... to obtain a very great precision in the results of the calculation.

Task 11 It may come as a surprise to the reader that as recently as the mid-nineteenth century, mathematicians were still working on better ways to predict positions of celestial bodies, seeing as we had been applying mathematics to this problem for almost as long as we have been doing mathematics itself![20] To consider the context for Cauchy's work here, do a bit of research to find each of the years in which Uranus, Neptune, and Pluto were first discovered. How do these years relate to the publication date of *Methode générale pour la résolution des systèmes d'équations simultanées*?

[18] For a proof of this fact, see [Bajaj, 1984, 12]. Also, if you are wondering if $n = 5$ here has anything to do with the difficulties mentioned regarding degree 5 polynomials in footnote 17, it sure does!

[19] All translations of Cauchy in this project were prepared by Richard J. Pulskamp, Xavier University, 2010.

[20] For example, there are records dating back to the eighth century BCE showing how mathematics was used in this manner by the Babylonians. For more details, see [van Brummelen, 2009, 14–18].

Let us now visit his gradient descent algorithm itself. Be aware that his notation $D_x u$ is just another way to write $\partial u/\partial x$, as it is more commonly written in calculus courses today, and similarly for the derivatives with respect to y and z. Also be aware that Cauchy used subtly different characters to represent entirely different objects: x, $\mathtt{x}$, and $\mathtt{X}$ all have different meanings in the passage below, so pay close attention to the fonts![21] To help process this a bit, here is a guide to help you disentangle these:

- The lowercase italicized letters $x, y, z, \ldots$ represent independent variables for our function $u = f(x, y, z, \ldots)$.
- The lowercase unitalicized symbol $\mathtt{x,y,z,\ldots}$ represent specific numerical values chosen for our variables $x, y, z, \ldots$. The unitalicized lowercase symbol $\mathtt{u}$ is the specific numerical value that f outputs when these numbers $\mathtt{x,y,z,\ldots}$ are plugged in. That is,

$$\mathtt{u} = f(\mathtt{x}, \mathtt{y}, \mathtt{z}, \ldots).$$

- The unitalicized uppercase letters represent the corresponding partial derivative of u evaluated at that values given by the lowercase symbols. So, $\mathtt{X} = \frac{\partial u}{\partial x}(\mathtt{x}, \mathtt{y}, \mathtt{z}, \ldots)$, $\mathtt{Y} = \frac{\partial u}{\partial y}(\mathtt{x}, \mathtt{y}, \mathtt{z}, \ldots)$, and so on.

Let first

$$u = f(x, y, z, \ldots)$$

be a function of the many variables $x, y, z, \ldots$ which never becomes negative and remains continuous, at least between certain limits. In order to find the values of $x, y, z, \ldots$ which will verify the equation

$$u = 0, \tag{1}$$

it will suffice to make the function u decrease indefinitely, until it vanishes. Now let $\mathtt{x}, \mathtt{y}, \mathtt{z}, \ldots$ be particular values attributed to the variables $x, y, z, \ldots$; $\mathtt{u}$ the value corresponding to u; $\mathtt{X}, \mathtt{Y}, \mathtt{Z}, \ldots$ the values corresponding to $D_x u, D_y u, D_z u, \ldots$, and $\alpha, \beta, \gamma, \ldots$ some very small increments attributed to the particular values $\mathtt{x}, \mathtt{y}, \mathtt{z}, \ldots$ When one will put

$$x = \mathtt{x} + \alpha, y = \mathtt{y} + \beta, z = \mathtt{z} + \gamma, \ldots,$$

one will have sensibly

$$u = f(\mathtt{x} + \alpha, \mathtt{y} + \beta, \mathtt{z} + \gamma, \ldots) = \mathtt{u} + \alpha\mathtt{X} + \beta\mathtt{Y} + \gamma\mathtt{Z} + \cdots \tag{2}$$

We imagine now that, θ being a positive quantity, one takes

$$\alpha = -\theta\mathtt{X}, \beta = -\theta\mathtt{Y}, \gamma = -\theta\mathtt{Z}, \ldots$$

[21] Note the author has slightly edited the translation with regards to placement of punctuation for readability. However, we retain Cauchy's penchant for similar symbols in different fonts for authenticity's sake!

Formula (2) will give sensibly

$$f(\mathtt{x}-\theta\mathtt{X},\mathtt{y}-\theta\mathtt{Y},\mathtt{z}-\theta\mathtt{Z},\ldots)=\mathtt{u}+\alpha\mathtt{X}+\beta\mathtt{Y}+\gamma\mathtt{Z}+\cdots=\mathtt{u}-\theta\left(\mathtt{X}^2+\mathtt{Y}^2+\mathtt{Z}^2+\cdots\right) \quad (3)$$

It is easy to conclude from it that the value Θ of u, determined by the formula

$$\Theta=f(\mathtt{x}-\theta\mathtt{X},\mathtt{y}-\theta\mathtt{Y},\mathtt{z}-\theta\mathtt{Z},\ldots), \quad (4)$$

will become inferior to $\mathtt{u}$, if θ is sufficiently small.

...

If the new value of u is not a *minimum*, one will be able to deduce, by operating always in the same manner, a third still smaller value; and, by continuing thus, one will find successively some values of $\mathtt{u}$ more and more small, which will converge toward a minimum value of u.

Admittedly, this is a lot to process! Let us start by working through his method in an easy-to-understand (at least as compared to the geometric median problem) example.

Task 12

(a) Let us begin with the relatively simple surface $u=f(x,y)=x^2+y^2$. Where does this surface have a minimum value?

(b) Suppose we weren't aware of the location of that minimum, and we guessed at randomly the "particular values"

$$\mathtt{x}=1,\mathtt{y}=2$$

from which we can then apply Cauchy's method. With the choices for $\mathtt{x}$ and $\mathtt{y}$ having been made, calculate the corresponding values of the following quantities:

(i) "$\mathtt{u}$ the value corresponding to u"

(ii) "$\mathtt{X},\mathtt{Y},\mathtt{Z},\ldots$ the values corresponding to $D_xu, D_yu, D_zu,\ldots$ "

(iii) Somewhat arbitrarily, choose $\alpha=0.1$ and $\beta=0.1$, and verify the truth of Cauchy's Formula (2).

(iv) While Formula (2) is true for any small α and β, Cauchy gave us a smarter way to choose these increments! (How these increments were chosen was the very clever crux of the paper.) Use $\theta=0.1$ and calculate the corresponding α and β. Now that we have made this choice, how do these values relate to the gradient of the surface f? (Do you see why the method was named *gradient descent*?)

(v) Use these new increments to calculate the corresponding value of Θ. Does it satisfy Cauchy's claim that it became "inferior to u"?

(vi) Repeat Cauchy's method, taking $x=\mathtt{x}-\theta\mathsf{X}$ and $y=\mathtt{y}-\theta\mathsf{Y}$ as the new starting points. Remember to recalculate the partial derivatives using the new x and y coordinates. What new location does it bring you to?

(vii) Iterate this method a few more times. Do you observe the behavior promised in Cauchy's claim that "by continuing thus, one will find successively some values of $\mathtt{u}$ more and more small, which will converge toward a minimum value of u"? (You may use a computer algebra system to speed up the tedium of doing lots of similar calculations over and over.) Graph the surface along with the sequence of points that you found via Cauchy's gradient descent method.

Note that Cauchy's method specifically talks about finding a zero of a function u, whereas for our geometric median problem we are more interested in finding a minimum value than a zero (as it is pretty clear that if $n > 1$, no point is going to have a sum of distances to the n given points equal to zero). However, these two problems, finding a zero and finding a min, are more similar than they may seem. Suppose we wanted to solve the equation

$$u = c,$$

where c was the absolute minimum of the function u, then one could simply rename the function $u - c$ as u, and now Cauchy's method will still find the point we wish to find, having translated the function down by c units, thus transforming the minimum into a zero.

At this point, we are ready to return to the geometric median problem!

Task 13 Consider the set of points $(0, 0), (1, 3), (0, 2), (3, 0), (4, 1)$. Write down the function $f(x, y)$ that represents the sum of the distances from a point (x, y) to each of those five points. Apply Cauchy's gradient descent method to the function $u = f(x, y)$ to find a sequence of points[22] that approximate the geometric median of the five given points. Graph the original five points along with your sequence of points given by Cauchy's method. (By all means use a computer algebra system here to help with the mess of plugging x and y coordinates into very messy partial derivatives over and over again.)

3 Conclusion

In conclusion, we have two big takeaways regarding the interplay between geometry and calculus. For one, if a geometry solution to an optimization problem exists, it is usually *far* nicer than the corresponding calculus solution to the same problem (as we saw in Torricelli and Fagnano's work). Furthermore, the geometric solution often runs out of steam at a certain level of complexity, and one can then switch to calculus to solve the problem or at least give an approximation to the solution, as we saw via Cauchy's method applied to the 5-point geometric median problem. Indeed, Bajaj has shown that this cannot yield to a nice geometric solution in [Bajaj, 1984, 12].

Task 14

(a) Can you think of another mathematics problem that has a simple geometric solution for easy cases but needs calculus for harder cases?

(b) Do a little bit of research and find some other problems where Cauchy's method of gradient descent is used to find approximate solutions in a field of interest to you.

We also have a takeaway about problem solving; being stuck for a while is a very normal state of being! Whether it was a mathematician stumped by one of Fermat's challenges or the astronomy community, stumped for thousands of years with regards to the problem of classifying all planets in our solar system, it is the most normal of conditions with regards to progress in human endeavors. Here we read a quote from an interview with the great mathematician Andrew Wiles (1953–) as quoted in [Orlin, 2017]:

[22] You might wonder how many points should be in this sequence. Well, that's up to you! How accurate an approximation to the minimum do you think is warranted?

"What you have to handle when you start doing mathematics as an older child or as an adult is accepting the state of being stuck. People don't get used to that. They find it very stressful. Even people who are very good at mathematics sometimes find this hard to get used to."

Task 15 The story of Andrew Wiles relates very directly to our topic in this project! Read the NOVA interview with Wiles posted here:

`https://www.pbs.org/wgbh/nova/proof/wiles.html`

– Who originally posed the problem being discussed in this interview?
– How long was the mathematical community collectively on this problem?
– How long was Wiles personally stuck on it before eventually solving it?
– Recall your answer to Task 2(a). How does this compare to Wiles' length of time? How long does it compare to the longest length of time you personally have ever been stuck working towards a goal outside of mathematics?

References

Niels Abel. Démonstration de l'impossibilité de la résolution algébrique des équations générales qui passent le quatrième degré. *Oeuvres Complètes de Niels Henrik Abel*, I, 1826.

Chanderjit Bajaj. The algebraic degree of geometric optimization problems. *Department of Computer Science Technical Reports*, Paper 416, 1984. Available at `docs.lib.purdue.edu/cstech/416`.

Augustin-Louis Cauchy. Methode générale pour la résolution des systèmes d'équations simultanées. *Comptes Rendus hebdomadaires des séances de l'Académie des Sciences*, 25:536–538, 1847. Translation by Richard J. Pulskamp, Xavier University, 2010. Available at www.cs.xu.edu/math/Sources/Cauchy.

Giovanni Fagnano. Problemata quaedam ad methodum maximorum et minimorum spectantia. *Nova Acta Eruditorum*, pages 281–303, 1775.

Paul Henry and Paul Tannery. *Oeuvres de Fermat.* Paris: Gauthier-Villars et fils, 1891.

Victor J Katz. *A History of Mathematics: An Introduction, Second Edition.* Addison-Wesley, 1998.

Gino Loria and Giuseppe Vassura. *Opere di Evangelista Torricelli.* Faenza, 1919. Volume I. Part II.

J J O'Connor and E F Robertson. Marin Mersenne Biography - Mactutor History of Mathematics. 2005. Available at `https://mathshistory.st-andrews.ac.uk/Biographies/Mersenne`.

Ben Orlin. The State of Being Stuck - Math with Bad Drawings. 2017. Available at `https://mathwithbaddrawings.com/2017/09/20/the-state-of-being-stuck`.

Glen van Brummelen. *The Mathematics of the Heavens and the Earth: The early history of trigonometry.* Princeton University Press, 2009. ISBN 978-0-691-12973.

Notes to Instructors

PSP Content: Topics and Goals

This Primary Source Project (PSP) is intended to introduce a multivariable calculus student to paradigms of multivariable optimization based on geometry rather than on calculus, as well as to introduce Cauchy's method of gradient descent. There are four key competencies that come up in this project:

- Partial derivatives
- Critical points of surfaces
- Multivariable optimization
- Gradients of surfaces

Student Prerequisites

In this project, we assume the student has already been exposed to the standard framework for finding maxima and minima of surfaces based on finding simultaneous zeros of partial derivatives. We also assume the student knows what a gradient is, although there is no requirement that the student must have seen the method of gradient descent. The project discusses the method from ground zero, and in fact could be used to introduce the topic.

The other prerequisite is access to, and some basic proficiency with, a computer algebra system. The project asks the student to perform many calculations (mostly calculating derivatives and solving systems of equations) that are extremely tedious do by hand. Any standard one will do: Mathematica, Maple, Sage (conveniently available for free use in a web browser at `www.cocalc.com`), etc.

PSP Design, and Task Commentary

This PSP is quite modular and as such can be implemented in a number of ways. The geometric approaches to the geometric median problem still tell quite a satisfying story and enrich a student's understanding of multivariable calculus optimization methods even if one entirely skips the section on Cauchy's method of gradient descent. Conversely, one can cover Cauchy's method of gradient descent and how it can be used to find approximate solutions to the geometric median problem without the student having ever seen Torricelli or Fagnano's work. So, there are loosely three sensible implementations of this PSP:

1. **Geometry only.** Complete the introduction and Section 1 but cut Section 2.
2. **Calculus only.** Complete up to and including Section 1.1 to introduce the problem, then skip Section 1.2, and move directly to Section 2.
3. **Both!** (Of course, this will give the richest experience if you have time for it in your class!) Do the whole PSP from start to finish.

Note that the author purposefully swept under the rug all possible issues of convergence (like getting stuck at a local minimum that isn't the one you're hoping to find) which can arise in the context of implementing Cauchy's gradient descent method, feeling that these are beyond the scope of a typical multivariable calculus course. If the instructor wishes to give the advanced student a bit deeper perspective into the method, the author recommends introducing them to the Cauchy-Schwarz inequality and seeing if the student can determine how it relates to Cauchy's method.

Suggestions for Classroom Implementation

The primary source in which Cauchy describes his algorithm is pretty heavy going, so showing the students a plot in the instructor's favorite 3D graphing utility might help them absorb it more easily. The instructor also might want to demonstrate for the class how to do the evaluation of the partial derivatives in a favorite computer algebra system, depending on how tech-savvy the students are; applying Cauchy's gradient descent method in Task 13 is most certainly not intended to be done by hand!

Copies of these PSPs are available at the TRIUMPHS website (the URL is given at the very end of the document). The author is happy to provide LaTeX code for this project. It was created in Overleaf, which makes it convenient to copy and share projects and can allow instructors to adapt this project in whole or in part as they like for their course.

Sample Implementation Schedule (based on a 50 minute class period)

This PSP will certainly take three class periods at minimum. The author recommends the following:

- **Class 1.**
  - **Prework.** Assign the reading and tasks of the introduction and Section 1.1 for a class prep assignment for the first session.
  - **Opening discussion.** Having students share some responses to Task 2 will almost certainly be a fun warmup to get in the right headspace for tackling this PSP! Then, have a student share a solution to Task 4, especially part (b). It is essential that all students have the right formula for $f(x, y)$ before proceeding to Section 1.2.
  - **Remainder of the session.** For the rest of class 1, students can work through Section 1.2 in small groups with instructor guidance. Any unfinished tasks in Section 1.2 can be homework.

- **Class 2.**
  - **Prework.** Students can read Section 1.3 and try to complete Task 7.
  - **Opening discussion.** Talking through Fagnano's surprisingly simple, clean argument could be a great opening discussion.
  - **Remainder of the session.** For the rest of class 1, students can work through the rest of Section 1.3 in small groups with instructor guidance. Task 8 (a) is actually somewhat tricky, though the answer is very clean! Any unfinished tasks in Section 1.3 can be completed for homework. There may be extra time to get started on Section 2 if a group is really cruising.

- **Class 3.**
  - **Prework.** Tasks 9, 10, and 11 along with the corresponding readings of Section 2 can be assigned for the second class session's prep. In particular, Tasks 9 and 10 should be review of ideas from Calculus 1 and a review of ideas from Section 1 of this very PSP, respectively.
  - **Remainder of the session.** Reading Cauchy's method of gradient descent and Tasks 12 through the end can be done together during the second class session.
- **Final Assignment.** Any unfinished parts of the PSP could then be assigned for homework.

Connections to other Primary Source Projects

There are three additional projects based on primary sources for the multivariable calculus classroom. The PSP author name of each is given, along with approximate implementation time. Classroom-ready versions of these projects in PDF form can be downloaded from the website `https://digitalcommons.ursinus.edu/triumphs_calculus`.

- Braess' Paradox in City Planning: An Application of Multivariable Optimization, Kenneth M Monks (2 class periods)
- Stained Glass, Windmills and the Edge of the Universe: An Exploration of Green's Theorem, Abe Edwards (4 class periods)
- The Radius of Curvature According to Christiaan Huygens, Jerry Lodder (6 class periods)

Recommendations for Further Reading

Michael Sean Mahoney's 1994 book *The Mathematical Career of Pierre de Fermat, 1601-1665* (2nd ed., Princeton University Press), makes great reading for the curious student who wants to find out more about Fermat's mathematics and interactions with Mersenne and others across Europe. For the student wanting more practical applications of gradient descent (and similar methods) with a focus on engineering systems, as well as how to implement it in a programming language, there is a recent text, *Algorithms for Optimization* (MIT Press, 2019), by Mykel J. Kochenderfer and Tim A. Wheeler, that does exactly that, with Julia code included.

Acknowledgments

The development of this project has been partially supported by the Transforming Instruction in Undergraduate Mathematics via Primary Historical Sources (TRIUMPHS) Project with funding from the National Science Foundation's Improving Undergraduate STEM Education Program under Grants No. 1523494, 1523561, 1523747, 1523753, 1523898, 1524065, and 1524098. Any opinions, findings, and conclusions or recommendations expressed in this project are those of the author and do not necessarily reflect the views of the National Science Foundation. The author would like to thank the TRIUMPHS PIs, authors, and Advisory Board for the very helpful feedback throughout the writing of this project.

For more information about the NSF-funded project TRansforming Instruction in Undergraduate Mathematics via Primary Historical Sources (TRIUMPHS), visit
http://blogs.ursinus.edu/triumphs/.