Sums of numerical powers in discrete mathematics: Archimedes sums squares in the sand

David Pengelley*

First, what is a discrete sum of numerical powers?

These appeared early in mathematics from an ancient desire to know what we today call 'closed formulas' for sums like $\sum_{i=1}^{n} i^2 = 1 + 4 + 9 + \cdots + n^2$, or $\sum_{i=1}^{n} i^3 = 1 + 8 + 27 + \cdots + n^3$, or most simply, just for the sum $\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n$. By a closed formula we mean an expression that does not have the arbitrarily long sum indicated by $+ \cdots +$ that we see in the open-ended sums above, and therefore may be much easier to calculate, understand, and work with. We will see examples shortly.

They are called 'discrete' sums because the items being added together have individual, separated, effect on the result, unlike what happens when we might think of a planar region as made up of adding together infinitely many infinitely thin parallel slices, where no single slice seems to have any effect at all on the total area as the slices move continuously across to trace out the region. The term 'discrete' is thus in contrast to 'continuous'. And our three discrete sums above are called 'sums of powers' because in each case the numbers being added are made from the natural numbers by a pattern based on a certain fixed power (exponent), like 2 (for a sum of squares i^2), 3 (a sum of cubes i^3), or just the power 1, which then sums just the natural numbers i themselves. In general, $\sum_{i=1}^{n} i^k$ is called a sum of powers whenever k is a fixed natural number.

The reading and activities in this project will engage you in rich and diverse ways with many of the concepts, methods, notations, and ways of thinking in modern discrete mathematics, and later also in combinatorics, all within the context of studying sums of numerical powers through the writings of those who discovered new mathematics. You will study and apply summation notation, reindexing, multiple connections to geometry, including to areas of regions with curved sides, inequalities, telescoping sums, making conjectures from experimental evidence. And you will seek patterns, developing your mathematical intuition and judgement. You will carry out and contrast various ways of proving statements, from using geometry to algebra to proof by generalizable example or by mathematical induction, work with systems of linear equations, and see the interplay between the continuous and the discrete via connections to differential and integral calculus.

Before we go further mathematically, a little history: The discovery of closed formulas for discrete sums of numerical powers, motivated by application to approximations for solving area and volume problems, is probably the most extensive thread in the entire development of discrete mathematics, spanning the vast period from ancient Pythagorean interest during the sixth century B.C.E. in patterns of dots to the work of Leonhard Euler in the eighteenth century on a general formula for discrete summations of almost any form. Initial sources from classical Greek, Indian, and Arabic traditions include Archimedes' determination in the third century B.C.E. of the closed formula for a sum of squares, then Nicomachus (first century C.E.), Āryabhaṭa (499 C.E.), al-Karajī (c. 1000 C.E.) on sums of cubes, and al-Haytham (965–1039) on the sum of fourth powers.

*Mathematical Sciences; Dept. 3MB, Box 30001; New Mexico State University; Las Cruces, NM 88003; davidp@nmsu.edu.

In the seventeenth century Pierre de Fermat (1601–1665) claimed that he could use the "figurate numbers" to solve this, "perhaps the most beautiful problem in all of arithmetic". Fermat's work was followed shortly by Blaise Pascal's extensive treatise (circa 1654) on this topic, which produced the first explicit recursive formula for sums of powers in any arithmetic progression using binomial coefficients, and which can be proved by mathematical induction. Pascal writes "I will teach how to calculate not only the sum of squares and of cubes, but also the sum of the fourth powers and those of higher powers up to infinity".

In the early eighteenth century, in his treatise on the beginnings of probability theory, Jakob Bernoulli (1654–1705) conjectured the general pattern in the coefficients of the closed-form polynomial solution to the summation problem. The connection between probability theory and sums of powers is via the combinatorial counting numbers. Here he introduces the all-important Bernoulli numbers into mathematics. Leonhard Euler's subsequent development of his general 'summation formula' around 1730 provided the first proof of Bernoulli's conjecture. This Euler-Maclaurin summation formula also led to numerous other applications, including Euler's spectacular solution of the famous Basel problem, showing that $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$. Paradoxically, although Euler's summation formula almost always diverges, the method can provide arbitrarily accurate approximations to the correct answer, with Euler saying that he will sum the terms just "until it begins to diverge," and "From this it is clear that, although the series ... diverges, it nevertheless produces a true sum." This is a feature of so-called 'asymptotic series,' an important and subtle modern applied concept with crucial use in physics today. For a complete exposition of this entire story, see [7, ch. 1].

This project focuses largely on just the first arc of this epic, namely the understanding of sums of squares. Subsequent projects continue the story.

Sums of powers in the service of areas and volumes

The story of discrete sums of numerical powers winds through mathematics and several human cultures for millenia. In classical Greek mathematics during the several centuries before the Common Era the analysis of discrete sums was part of one of the greatest achievements of antiquity, a systematic means of approximating continuous objects in a way that actually led, astonishingly, not just to approximations, but to the exact determination of areas and volumes of geometric objects with curved sides. In this project we will see Archimedes of Syracuse (c. 287–212 B.C.E.), the greatest mathematician of antiquity, analyzing sums of squares to the purpose of finding areas and volumes. The theme of summing powers to find areas and volumes continued for almost two millenia and eventually played a key role in the development of the differential and integral calculus in the seventeenth century. Finally, the formulas discovered for sums of powers were themselves generalized to reveal a wonderful interplay, a dance, between continuous and discrete, that we began to understand only through the eighteenth century work of Leonhard Euler, and which places each partner in the dance on equal footing. This is fortuitous, perhaps even fitting, today, when digital computers cause us to revert to a discrete way of viewing things much more than we otherwise might. We will see ways in which the discrete perspective is often best understood with help from the continuous, as well as vice-versa.

We will discuss shortly exactly why ancient mathematicians like Archimedes needed to know formulas for sums of powers. But first, you may wonder what we even mean by having a 'formula' for one of these sums. Aren't they already given by formulas? Yes, they are, but the formulas are arbitrarily long, since they depend on the size of n. Just from a purely computational point of view, how quickly could you (or even your calculator) calculate $\sum_{i=1}^{1{,}000{,}000} i$, and how accurately? Give it a try. Couldn't there be problems of loss of significant digits on a calculator or computer? And without calculating the answer, could you even get any idea of how big it is? In short, our description of these sums, although formulaic, is, to put it mildly, very hard to use.

Suppose, however, that someone told you that perhaps $\sum_{i=1}^{n} i = \frac{n^2}{2}$. If this equality were true, it would be incredibly useful; would surely answer the questions raised above; and make actual computation almost trivial. For instance, even in our heads we could then compute immediately that $\sum_{i=1}^{1{,}000{,}000} i = \frac{1{,}000{,}000^2}{2} = 500,000,000,000$. If this were true, we would say that we have a 'closed formula' $\frac{n^2}{2}$ for the sum of the first n natural numbers. Recall that by a closed formula we simply mean that we don't have to perform an arbitrarily long sequence of additions (an open process with a $\cdots$) to find the answer, provided we allow ourselves standard arithmetic operations like addition, subtraction, multiplication, division, and exponentiation. So first:

Exercise 1. Explore the closed formula $\frac{n^2}{2}$ suggested for $\sum_{i=1}^{n} i$. Discuss how good or bad it is for some small values of n, and as n grows. In fact, from some simple experiments and data you can even conjecture how to fix it to be correct! A conjecture is an educated guess based on reasoning from what you already know, such as observing possible patterns. Do that right now. Congratulations: but are you sure that the new closed formula you have conjectured really always works, meaning for all infinitely many possible values of n? In fact why should we even have reason to hope in the first place that there exists any closed formula at all for a sum like $\sum_{i=1}^{n} i$? (We don't expect you to answer that question!) Here we see that while it is tempting to think that discrete mathematics is somehow finite, we often have to deal with infinity, since to find a closed formula means to find one that works for all infinitely many n simultaneously.

Now let us briefly see exactly why such formulas were so important to the mathematics of antiquity.

By the time of Euclid's writing of the *Elements* around 300 B.C.E., a powerful theoretical method had been perfected for finding areas and volumes of objects with curved sides. It is today called 'the method of exhaustion', and it involves approximating curved objects by polygons (in the plane) or polyhedra (in three dimensions), i.e., by straight-sided objects whose areas and volumes were already known. The method could then often be used, amazingly, to confirm the exact, not just approximate, area or volume of the curved object, if the straight-sided objects could be made to 'exhaust' the curve-sided one. For details, see [8]. Archimedes carried this to a high art, proving many of the area and volume results students learn today in school. His achievements were without equal until the seventeenth century, and included areas bounded by parabolas and spirals, and volumes of surfaces of revolution of conics (i.e., of parabolas, ellipses, and hyperbolas).

It should be no surprise that sums of squares of numbers played an important role in Archimedes accomplishing these things. We know that parabolas, ellipses, and hyperbolas are the three types of what are called 'conic sections' (or simply conics), because they are the curves cut from a cone by a planar section, and that each type can be interpreted as the locus of points satisfying a specific equation in the sectioning plane using planar cartesian coordinates. We also know that the cartesian equations of these conic sections are all second degree polynomial equations, i.e., involving only squares and first and zeroth powers of the variables. For instance, we know that one very simple parabola in the plane has equation $y = x^2$. Archimedes wrote in two treatises specifically about finding the area bounded by a parabola and any line cutting across it. We will not display here exactly what he did or how he did it [8], but we will explore just enough of the mathematics to see how analysis of the area of a region bounded by a parabola could have led him to need a closed formula for the sum of squares $\sum_{i=1}^{n} i^2$. In fact the importance of the connection between a sum of powers and an area is more than just historical in nature: we will see that it has direct import also for our understanding of formulas for discrete sums of powers.

Exercise 2. Draw for yourself the region bounded below by the x-axis, on the left by the y-axis ($x = 0$), on the right by the vertical line $x = 1$, and above by the parabola $y = x^2$. Notice that since it is sitting inside half a unit square, its area must be less than one half. Can we calculate it? If you have studied calculus, you will recall that we can approximate numerically the area of this curve-sided parabolic region, by inscribing or circumscribing the region with abutting rectangles.[1] So cut the horizontal interval $0 \le x \le 1$ into n equal subintervals (the number n of subintervals is arbitrary, and could be as large as we wish). For each subinterval build a rectangle with its base on that subinterval along the x-axis. Choose the top of each rectangle in either one of two ways: all to be just high enough to be at or above the parabola on their respective subintervals, thereby barely circumscribing the entire region with rectangles, or all to be just low enough to fall at or below the parabola on their respective subintervals, thereby barely inscribing the entire region with rectangles. Now work out detailed mathematical expressions for the sums of the areas of the separate circumscribed and inscribed rectangular approximations, and note how they compare with the area of the curved-sided region. In the end, and after a bit of extra algebraic manipulation of the summations, the upshot should be that you produce this:

$$\frac{\sum_{i=1}^{n} i^2}{n^3} - \frac{1}{n} < \text{Area of the curve-sided parabolic region} < \frac{\sum_{i=1}^{n} i^2}{n^3}$$

Now on the one hand this bodes excellently for finding the area of the parabolic region, because its two bounding expressions are the same except for a difference of $1/n$. In other words, we have the numerical value of the area we seek trapped between numbers whose difference becomes negligible as n grows. So by taking n as large as we wish, we can trap the area as closely as we please; our chosen geometric rectangles did a mighty good job. In fact this shows that numerically, the area of the parabolic region is actually being approached on the nose by the algebraic expression $\frac{\sum_{i=1}^{n} i^2}{n^3}$ as n grows larger and larger. If we could only calculate the actual number that is being approached thereby, we would know the area of the region. But there is the rub; we are stuck because we have little understanding of the value of the sum of the first n squares, especially in comparison with the denominator n^3. Now we see exactly why Archimedes might have wanted a closed formula for the sum of squares![2]

Archimedes sums natural numbers: the Pythagorean tale of triangular numbers

In 216 B.C.E., the Sicilian city of Syracuse allied itself with Carthage during the second Punic war, and thus was attacked by Rome, portending what would ultimately happen to the entire Hellenic world. During a long siege, soldiers of the Roman general Marcellus were terrified by the ingenious war machines defending the city, invented by the Syracusan Archimedes. These included catapults to hurl great stones, as well as ropes, pulleys, and hooks to raise and smash Marcellus's ships, and perhaps even burning mirrors setting fire to their sails. Finally though, probably through betrayal, Roman soldiers entered the city in 212 B.C.E., with orders from Marcellus to capture Archimedes alive. Plutarch relates that "as fate would have it, he was intent on working out some problem with a diagram and, having fixed his mind and his eyes alike on his investigation, he never noticed the incursion of the Romans nor the capture of the city. And when a soldier came up to him suddenly

[1] If you are tempted to point out that the area can be found using antidifferentiation and the fundamental theorem of calculus, you are right, but you are cheating on what Archimedes knew by almost two-thousand years.

[2] In fact Archimedes' geometric approximation method was quite different from ours. He inscribed triangles inside the region between the parabola and a sectioning line. But it still ultimately led him to need to know good bounds for a sum of squares.

and bade him follow to Marcellus, he refused to do so until he had worked out his problem to a demonstration; whereat the soldier was so enraged that he drew his sword and slew him" [5, p. 97].

Despite the great success of Archimedes' military inventions, Plutarch says that "He would not deign to leave behind him any commentary or writing on such subjects; but, repudiating as sordid and ignoble the whole trade of engineering, and every sort of art that lends itself to mere use and profit, he placed his whole affection and ambition in those purer speculations where there can be no reference to the vulgar needs of life" [5, p. 100]. Perhaps the best indication of what Archimedes truly loved most is his request that his tombstone include a cylinder circumscribing a sphere, accompanied by the inscription of his remarkable theorem that the sphere is exactly two-thirds of the circumscribing cylinder in both surface area and volume!

Archimedes was the greatest mathematician of antiquity, and one of the top handful of all time; his achievements seem astounding even today. The son of an astronomer, he spent most of his life in Syracuse on the island of Sicily, in today's southern Italy, except for a likely period in Alexandria studying with successors of Euclid. In addition to his mathematical achievements, and contrasting with the view expressed by Plutarch, his reputation during his lifetime derived from an impressive array of mechanical inventions, from the water snail (a screw for raising irrigation water) to compound pulleys, and his fearful war instruments. Referring to his principle of the lever, Archimedes boasted, "Give me a place to stand on, and I will move the earth." When King Hieron of Syracuse heard of this and asked Archimedes to demonstrate his principle, he demonstrated the efficacy of his pulley systems by single-handedly pulling a three-masted schooner laden with passengers and freight [5]. One of his most famous, but possibly apocryphal, exploits was to determine for the king whether a goldsmith had fraudulently alloyed a supposed gold crown with cheaper metal. He is purported to have realized, while in a public bath, the principle that his floating body displaced exactly its weight in water, and realizing he could use this to solve the problem, rushed home naked through the streets shouting "Eureka! Eureka!" (I have found it).

The treatises of Archimedes contain a wide array of area, volume, and center of gravity determinations, including the equivalent of many of the best-known formulas taught in high school today. Archimedes also laid the mathematical foundation for the fields of statics and hydrodynamics and their interplay with geometry, and frequently used intricate balancing arguments. A fascinating treatise on a different topic is *The Sandreckoner*, in which he numbered the grains of sand needed to fill the universe, by developing an effective system for dealing with large numbers. Even though he calculated in the end that only 10^{63} grains would be needed, his system could actually calculate with numbers as enormous as $\left(\left(10^8\right)^{10^8}\right)^{10^8}$. Archimedes even modeled the universe with a mechanical planetarium incorporating the motions of the sun, the moon, and the "five stars which are called the wanderers" (i.e., the known planets) [3].

We are ready to start reading what Archimedes wrote about sums of powers. At the beginning of his treatise titled *On Conoids and Spheroids*, Archimedes states two inequalities that he will use in his later proofs, and whose own proof is implicitly embedded in some of his later results [1, v. 11, p. 455][2, p. 119]. He writes, translated from the ancient Greek:

If any number of magnitudes be given, which exceed one another by an equal amount equal to the least, and also other magnitudes, equal in number to the former, but each equal in quantity to the greatest, all the magnitudes each of which is equal to the greatest will be less than the duplicate of all those exceeding one another by an equal amount and more than the duplicate of these minus the greatest. [2, p. 119]

Before trying to understand these inequalities, let us consider a related statement of equality rather than inequality, that we will write in the same style as Archimedes. Although Archimedes only actually explicitly stated the inequalities above, his proof shows that he fully understood the following equality, how to prove it, and that the two inequalities would follow immediately from it. We'll call this *Archimedes' equality.*

If any number of magnitudes be given, which exceed one another by an equal amount equal to the least, and also other magnitudes, equal in number to the former, but each equal in quantity to the greatest, all the magnitudes each of which is equal to the greatest, plus the greatest, will be the duplicate of all those exceeding one another by an equal amount.

In fact this equality was understood already by the Pythagoreans, a mysterious group led by Pythagoras in ancient Greece around the sixth century B.C.E. For them number was the substance of all things, and geometric patterns seen in dots or pebbles in the sand showed them relationships between numbers. Soon you too will see the geometric connection here, embodied in what is called a 'triangular number'.

Let us try to understand the text of Archimedes' equality, and connect it to more modern ways of thinking and expressing things. In classical Greek mathematics "magnitude" had a potentially broad meaning, but here we will interpret it just to mean a line segment of some finite length.

Exercise 3. Draw a picture of the given magnitudes, setting them side by side with one end of each all in a row, and illustrate with marks on the line segments the "which" condition stated about just how they exceed one another. Also draw a picture of the "other" magnitudes given, and then draw a picture of what the text claims is the duplicate of what. Then find a way to see visually that this is true by drawing the duplicate and displaying a geometric way to rearrange the set of duplicated magnitudes to show what is claimed. You may wish first to do some small examples, then a general picture with explanation.

Exercise 4. What kinds of things are actually 'equal' here by virtue of the words 'will be' of the claim? Is it lengths, or line segments, or numbers, or some combination of these? In what sense is 'will be' true? In what sense is 'equal' the case?

Exercise 5. Explain why there are infinitely many possibilities for the particular set of line segments satisfying the "which" condition of the claim. In fact the possibilites are infinite in at least two quite different ways; explain these. Do you consider that your picture provides a proof of the claim for all of the infinitely many possibilities? In other words, does it provide a completely convincing argument that is general enough to convince anyone that the claimed result is true for all the possibilities? Why?

Exercise 6. Now also use your picture to explain and prove the two inequalities that Archimedes wrote at the beginning. Hint: Notice that his final words 'the duplicate of these minus the greatest' could be interpreted in English in two different ways; only one will give a true inequality.

Exercise 7. Interpreting each line segment as having a length which is some numerical multiple of a fixed unit length for measurement, and supposing there are n lines in general, write Archimedes' equality from above as a modern numerical equality involving lengths represented by numbers. Do this with $+\cdots+$ notation, and then with summation notation. Manipulate your equality algebraically to solve for a closed form expression for $\sum_{i=1}^{n} i$. In other words, obtain an equality of the form $\sum_{i=1}^{n} i =$ "some closed formula" (one that has no arbitrarily long sum indicated by $+\cdots+$). Today we would say that you have found a polynomial formula for the sum of the first n natural numbers, no matter what n is. What is the polynomial, written in standard polynomial notaton? What degree is it? What is its leading coefficient? What is its constant term? Is it the same as the formula you conjectured in the first exercise to fix the suggested $\frac{n^2}{2}$?

Exercise 8. Interpret Archimedes' two inequality statements as algebraic inequalities involving $\sum_{i=1}^{n} i$ (this may involve some reindexing), and prove these algebraically from your closed formula for $\sum_{i=1}^{n} i$.

Exercise 9. Now redraw your original picture proof of Archimedes' equality from a different point of view. First let's simplify things a little by decreeing that the smallest of the line segments in the series will be chosen as our unit of length for measurement. A line segment has length but no width. Give each of them unit width in your original picture, i.e., fatten each line segment into a rectangle. Thus turn your proof into a statement about equality of two areas. Give a verbal explanation of how the rectangles fit together to make equal areas, and enhance your picture with shadings for the rectangles demonstrating how things match up. Also explain how your picture shows that the sum $\sum_{i=1}^{n} i$ is actually the area of a certain right triangle. What is its length and width? Explain how these are related to the algebraic form of the closed formula you found for the sum. This connection is why $\sum_{i=1}^{n} i$ is called a triangular number.

Exercise 10. Let's find another proof of your equality, apparently purely algebraically. Consider the expression

$$\sum_{i=1}^{n} \left[(i+1)^2 - i^2 \right].$$

Rewrite it two completely different ways, as follows.

On the one hand, notice that it is a 'telescoping sum', i.e., that all but two terms cancel out; to see this, first work out some examples for small n, then make a general calculation using $+\cdots+$ notation, and finally use summation notation and reindexing (Hint: you could split the sum into two sums, then reindex one with $j = i+1$, carefully adjusting the endpoints of the summation, and then see what you have).

On the other hand, you could expand $(i+1)^2$ with the binomial theorem, then simplify and discover that you can solve for $\sum_{i=1}^{n} i$. Voilà! This is wonderful, but it all seems like a rabbit pulled out of a hat. So let's see next if we can visualize some of it geometrically.

Exercise 11. Draw a picture of two squares, one inside the other, sharing their bottom left corner, so that the area of the L-shaped region between them is $(i+1)^2 - i^2$. An L-shaped region like this is called a 'gnomon'[3]. Now draw a picture of nested such squares, with successive non-overlapping gnomons inbetween them, for i ranging from 1 to n, so that all the gnomons together have area equaling the telescoping expression we started with above, and that fit together to make almost one large square. This would have been very familiar to classical Greek mathematicians, so now imagine you are one: Start completely over just with this geometric picture, and from it, explain how to deduce your formula for $\sum_{i=1}^{n} i$ (Hint: Each gnomon can be thought of as two equal rectangles joined by a small square).

[3] In ancient Babylonia, gnomon probably meant an upright stick whose shadow was used to tell time (a sundial). In Pythagoras' time gnomon meant a carpenter's square; do you see the connection? And in the time of Euclid and Archimedes, it meant what was left of a parallelogram when a smaller similar parallelogram was removed from one corner. [6, p. 31]

Exercise 12. (For those who have studied mathematical induction) Suppose you had a guess for your equality above, $\sum_{i=1}^{n} i =$ "some closed formula", but you weren't sure by any means, geometrical or otherwise, whether it is true or not for all natural numbers n. Verify the truth of your formula for all such n using the method of mathematical induction.

We have seen several ways to find and/or prove a closed formula for the sum of the first n natural numbers, interpreted a lot geometrically, and seen that the sum is related to the area of a triangle. Let us now move on to Archimedes' claim about a sum of squares.

Archimedes sums squares in the sand

Also in his treatise *On Conoids and Spheroids*, and as Proposition 10 in his treatise *On Spirals* [1, v. 11, p. 456][2, p. 122], Archimedes states and proves his claim about a sum of squares:

If a series of any number of lines be given, which exceed one another by an equal amount, and the difference be equal to the least, and if other lines be given equal in number to these, and in quantity to the greatest, the squares on the lines equal to the greatest, plus the square on the greatest and the rectangle contained by the least and the sum of all those exceeding one another by an equal amount will be the triplicate of all the squares on the lines exceeding one another by an equal amount. [2, p. 122]

By 'line' Archimedes means a line segment. By the 'square on a line' Archimedes means a square whose side length is the length of the line, and by the 'rectangle contained by two lines' he means a rectangle with its two perpendicular sides given by their lengths.

Exercise 13. As before, draw a picture of the two geometric constructs that Archimedes is saying 'will be' each other. In your picture, go ahead and make the simplifying choice that the smallest line represents the unit of length. Archimedes' claim will be that two constructs made of rectangles have the same total areas.

Exercise 14. Archimedes provides a quite elaborate proof of his claim, with no picture. Student Kathe Kanim recently found in Archimedes' rather algebraic proof the insight and inspiration for a geometric 'picture proof', as displayed in Figure 1 [4]. It seems that she may have rediscovered the picture that was in Archimedes' mind as he was drawing in the sand more than two thousand years ago. Mark up the displayed picture to explain why it proves what Archimedes claims. Notice that it only proves the claim for a particular number of 'lines'. How many? The picture nonetheless should convince you that the claim is true for any number of lines. Explain why it does that. Could you draw the picture necessary to prove it for nine lines? nineteen thousand lines? Explain why you are sure of that. The fact that you are convinced means that this has the nature of a 'proof by generalizable example', which was a common method of proof in mathematics until perhaps one or two hundred years ago. However, as mathematics became more highly developed, elaborate proofs by this method became less acceptable, because they rely on an intuitive sense that the example generalizes, but intuition can lead one astray or not be the same to everyone when things are complicated. Thus today we require proofs that do not rely just on the reader's acceptance of the intuitive generalization of an example.

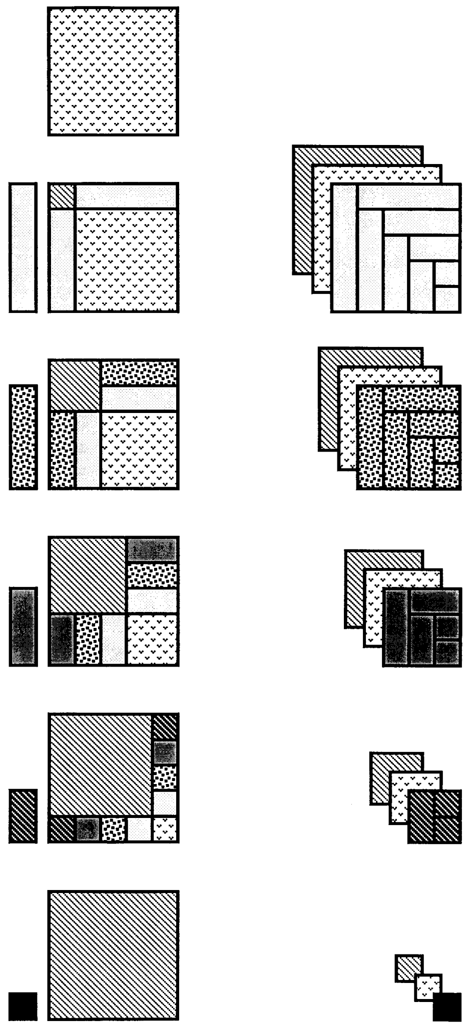

Figure 1: Sum of squares

Exercise 15. From Archimedes' claim deduce an equality involving $\sum_{i=1}^{n} i^2$, and therefrom deduce a closed formula for a sum of squares. Feel free to use the formula you already proved earlier for $\sum_{i=1}^{n} i$. Is your new closed formula a polynomial? What degree is it? What is its leading coefficient? What is its constant term?

Exercise 16. Just as with the sum of natural numbers, Archimedes had a pair of inequalities related to the triplicate of a sum of squares. He used these in his proofs about areas and volumes, where his trapping arguments only required inequalities. Conjecture his inequalities by modeling them directly on the inequalities he gave earlier for a sum of natural numbers and his statement about a sum of squares. The inequalities should involve the triplicate of a sum of squares and 'the squares on the line equal to the greatest'. Then prove your two inequalities geometrically from the picture proof. Also prove your inequalities algebraically from your new formula for a sum of squares.

Exercise 17. Prove your formula for a sum of squares by generalizing algebraically the telescoping sum approach used earlier for a sum of natural numbers. Notice that you will again use your knowledge of the formula for a sum of natural numbers. For the sum of natural numbers you interpreted this telescoping approach geometrically using nested squares and gnomons. Discuss what would be involved geometrically to do this for your telescoping proof for a sum of squares.

Exercise 18. (For those who have studied mathematical induction) Verify the truth of your formula for $\sum_{i=1}^{n} i^2$ for all natural numbers n using the method of mathematical induction.

Exercise 19. Recall that we saw that with a closed formula for arbitrary sums of squares in hand, we (and Archimedes) might be able to calculate the area of a region bounded on one side by a parabola. In particular, we determined that the area of the region that we studied underneath the parabola will be the number approached by $\frac{\sum_{i=1}^{n} i^2}{n^3}$ as n grows larger and larger. Now that you know a formula for a sum of squares, you can calculate this. Figure out exactly what that number being approached actually is, on the nose. Congratulations, you have computed the precise area of this curve-sided region. For those of you who have learned calculus, you know that this can also be obtained by antidifferentiation and the fundamental theorem of calculus; but to do so would be cheating by almost 2,000 years, wouldn't it, since you have just done it à la Archimedes, but the calculus approach was not discovered until the middle of the 17th century.

Thinking towards higher powers

In later episodes of our story, we will see formulas for sums of cubes and of fourth powers appear in India and the medieval Arabic world, then more general approaches emerge in the mathematics of renaissance Europe, combine with calculus and combinatorics, and move into the heart of modern mathematics.

For the moment, let us ask a few questions about how our discoveries so far may generalize to cubes, fourth, and higher powers of the natural numbers:

Exercise 20. Do the geometric proofs for sums of natural numbers and of squares look promising for generalizing to sums of higher powers? Why or why not?

Exercise 21. What about the telescoping sum approach? Discuss how this could be carried to higher powers. What will the challenges and difficulties be? For instance, suppose you wish to find a formula for the sum of fourth powers. What will you need already to know in order to use the telescoping sum approach for fourth powers? Thinking about implementing this approach in general, what does it tell you about the question of whether there is always a polynomial formula for sums of powers of any exponent? Does it also tell you about the degree of such a polynomial? Anything else?

Exercise 22. Mathematical induction seemed to work well for proving the formulas for sums of natural numbers and sums of squares. What would be impossible about trying to use it at this point for a sum of cubes?

Exercise 23. By some reindexing and/or algebraic manipulation of the inequalities of Archimedes that you already have, fill in the $\boxed{??}$ below to rewrite Archimedes' pairs of inequalities for sums of natural numbers and sums of squares as two double inequalities for all n:

$$\boxed{??} < \sum_{i=1}^{n} i < \boxed{??} \quad \text{and} \quad \boxed{??} < \sum_{i=1}^{n} i^2 < \boxed{??},$$

where each $\boxed{??}$ should be a simple polynomial-like expression. Depending on how you decide to approach this, you actually may not end up with exactly the same inequalities as other people do. Nonetheless, if you are consistent in your approach, you should immediately see a conjecture that generalizes yours to inequalities surrounding $\sum_{i=1}^{n} i^k$, where k could now be any positive power. State it.

Prove your conjecture as follows, by again appealing to the interplay between continuous and discrete, and allowing yourself to use the differential and integral calculus from the later seventeenth century. Recall how we approximated the area under the parabola $y = x^2$ between $x = 0$ and $x = 1$ using n abutting rectangles. Do the same thing for the so-called 'higher parabolas' $y = x^k$, but now use the interval between $x = 0$ and $x = n$ (or maybe $n - 1$ or $n + 1$, depending on your particular inequalites). Inscribe and circumscribe n (or $n - 1$) rectangles, and calculate the comparisons you get between two sums of powers and the area under the curve. Also calculate the area under the curve using calculus (the 2,000 year cheat!). Reindex the sums and inequalities as necessary to prove your conjecture.

In the few decades just prior to the invention of the fundamental theorem of calculus, various mathematicians worked to find areas under curves like $y = x^k$. In particular, Gilles Persone de Roberval (1602–1675) claimed that he could do so using inequalities like those you conjectured surrounding $\sum_{i=1}^{n} i^k$, which however he perhaps deduced by other, entirely algebraic, means; in other words, his route was the opposite of yours, deducing the area under the curve from his inequalities, rather than the inequalities from the area [7]. We, however, can view Roberval's inequalities as an aid to us in discrete mathematics, imported from the continuous, i.e., from the calculus.

Exercise 24. You are in a position to make some conjectures about formulas for sums of higher powers. Based on the formulas you know for $\sum_{i=1}^{n} i$ and $\sum_{i=1}^{n} i^2$, as well as your analysis about generalizing the telescope approach, and Roberval's inequalities, make some conjectures about a formula for the sum $\sum_{i=1}^{n} i^k$ of the k-th powers of the first n natural numbers. Will it be given by a polynomial in the variable n? What will be its degree? What will be its leading coefficient? its constant term? its term right after the leading term? and the next one? Give some reasons for your conjectures. Of course a conjecture is something we would like to prove or disprove, and that is more difficult.

Exercise 25. Suppose we have conjectured that a sum of cubes is given by a formula that is a fourth degree polynomial: $\sum_{i=1}^{n} i^3 = a_4 n^4 + a_3 n^3 + a_2 n^2 + a_1 n + a_0$, and we suspect that $a_4 = \frac{1}{4}$ and $a_0 = 0$. Then we could pick 3 small values for n, calculate both sides of the equality for each, thus obtaining 3 linear equations in 3 unknowns, and then see if these enable us to determine uniquely what the coefficients a_j would have to be. Then we would have a guess for a formula, in fact the only possible guess, and could try proving it by mathematical induction. Do this.

References

[1] Archimedes, *The Works of Archimedes, Including the Method*, T.L. Heath (ed.), Cambridge University Press, 1897, 1912, reprinted by Dover, New York, 1990, and in Great Books of the Western World (ed. R. Hutchins), vol. 11, Encyclopaedia Britannica, Chicago, 1952.

[2] E.J. Dijksterhuis, *Archimedes*, Princeton University Press, Princeton, New Jersey, 1987.

[3] C.C. Gillispie, F.L. Holmes (eds.), *Dictionary of Scientific Biography*, Scribner, New York, 1970.

[4] K. Kanim, *Proof without words: How did Archimedes sum squares in the sand?,* Mathematics Magazine **74** (2001), 314–315.

[5] V.J. Katz, *A History of Mathematics: An Introduction* (second ed.), Addison-Wesley, New York, 1998.

[6] Kline, Morris, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, 1972.

[7] A. Knoebel, R. Laubenbacher, J. Lodder, D. Pengelley, *Mathematical Masterpieces: Further Chronicles by the Explorers,* Springer Verlag, New York, 2007.

[8] R. Laubenbacher, D. Pengelley, *Mathematical Expeditions: Chronicles by the Explorers*, Springer Verlag, New York, 1999.

Notes to the instructor

This project is for students of introductory discrete mathematics or calculus (Riemann sums). It is based on Archimedes' writing on sums of squares in the service of finding areas and volumes. The content is part of a larger curricular sequence of guided primary source readings from Archimedes to Euler on the topic of sums of powers in [7, ch. 1].

This project has no formal prerequisite other than basic algebra. The project is quite flexible, and the instructor can pick and choose from various activities offered. The full project can be completed within two class weeks or less, and for a shorter project the instructor may choose selectively. Students can work productively in groups on this project, with group or individual writeups. One excellent challenge the project provides is interpreting Archimedes' verbal descriptions of adding up many magnitudes, and manipulative rods or graph paper have been encouraged productively by some instructors.

The goal of the project is for students to learn many basic notations, techniques, and skills in the context of an historically and mathematically authentic big motivating problem with multiple connections to other mathematics. Hopefully this will be much more effective and rewarding than simply being asked to learn various skills for no immediately apparent application. Many of the techniques first introduced in a discrete mathematics or calculus course arise naturally as needed in this project, like reindexing summation notation, working with algebraic inequalities, and telescoping sums. Instead of separately covering various such topics and techniques, that class time can simply be spent on the project, and students will learn those things in the process. The project has the integrating theme throughout of the application of sums of powers to Riemann sums and area calculation, and vice versa.

Most of the ideas and knowledge in the project are acquired through guided discovery exercises, a number of which are open-ended, so the instructor should work through all the details before assigning any student work, and select carefully from the big picture if exercises are omitted. Students may need substantial guidance with some parts.

The project asks students to interpret and convert verbal descriptions into modern mathematical formulations, conjecture from patterns they generate, develop their mathematical intuition and judgement, and try proving their conjectures, i.e., putting students in the creative driver seat. The setting of sums of powers in the context of primary sources allows a richness of questions and interpretations, especially includes deep connections to geometry and the two-way interplay with calculus, as well as basic algebra and linear algebra, and a richness of proof techniques, including natural comparison of the efficacy of various proof methods.