The UMAP Journal

Vol. 33, No. 2

Subscription Rates for 2012 Calendar Year: Volume 33

Institutional Web Membership (Web Only)

Institutional Web Memberships do not provide print materials. Web memberships allow members to search our online catalog, download COMAP print materials, and reproduce them for classroom use.

(Domestic) #3030 $467 (Outside U.S.) #3030 $467

Institutional Membership (Print Only)

Institutional Memberships receive print copies of The UMAP Journal quarterly, our annual CD collection UMAP Modules, Tools for Teaching, and our organizational newsletter Consortium.

(Domestic) #3040 $312 (Outside U.S.) #3041 $351

Institutional Plus Membership (Print Plus Web)

Institutional Plus Memberships receive print copies of the quarterly issues of The UMAP Journal, our annual CD collection UMAP Modules, Tools for Teaching, our organizational newsletter Consortium, and online membership that allows members to search our online catalog, download COMAP print materials, and reproduce them for classroom use.

(Domestic) #3070 $615 (Outside U.S.) #3071 $659

For individual membership options visit
www.comap.com for more information.

To order, send a check or money order to COMAP, or call toll-free
1-800-77-COMAP (1-800-772-6627).

The UMAP Journal is published quarterly by the Consortium for Mathematics and Its Applications (COMAP), Inc., Suite 3B, 175 Middlesex Tpke., Bedford, MA, 01730, in cooperation with the American Mathematical Association of Two-Year Colleges (AMATYC), the Mathematical Association of America (MAA), the National Council of Teachers of Mathematics (NCTM), the American Statistical Association (ASA), the Society for Industrial and Applied Mathematics (SIAM), and The Institute for Operations Research and the Management Sciences (INFORMS). The Journal acquaints readers with a wide variety of professional applications of the mathematical sciences and provides a forum for the discussion of new directions in mathematical education (ISSN 0197-3622).

Periodical rate postage paid at Bedford, MA and at additional mailing offices.

Send address changes to: info@comap.com
COMAP, Inc., Suite 3B, 175 Middlesex Tpke., Bedford, MA, 01730

Vol. 33, No. 2 2012

Table of Contents

Guest Editorial

Article

UMAP Modules

Guest Editorial
Calculus and DNA

Keith Stroyan
Mathematics Dept.
University of Iowa
Iowa City, IA 52242
keith-stroyan@uiowa.edu

Introduction

I joke that I'm the only one in our Mathematical Biology seminar who was born before we had DNA. A lot of things have changed since Thomas [1951] developed his syllabus for calculus in the late 1940s; calculus isn't one of them. We compete for our undergraduates' time with amazing advances in many fields, including biology and computing. Some people are again clamoring to make statistics or combinatorics part of basic college mathematics.

Can we still do a good job in calculus if we and the students have less time for it? I think the answer is "yes," but it will require fundamental change.

I recently wrote about using student projects as a small part of the course to help improve student reasoning and understanding of calculus [Stroyan 2011]. I believe that is a worthwhile idea; but in this article, I want to ask:

What ought to be in the 21st-century calculus curriculum?

Campbell [2006] wrote an earlier essay that bears on this question. There should not be a single answer, as Bressoud [2004] and I [Stroyan 2006] have tried to point out. At a large school like my University of Iowa, we can teach separate courses in calculus for business majors, for biology majors, and for mathematics / science majors, as well as for engineers. In itself, this division is good because we could adjust the content to suit their interests and preparation. I believe that we should strive—in the first course—to really:

Convince students that the subject speaks to their interests!

The UMAP Journal 32 (2) (2011) 93–98.

Moving Beyond 1950

Winnowing out some of the 1950 details of calculus could help to improve the course in the time that we have available with our students.

"Defining" calculus too narrowly has consequences that impede progress in mathematics education. Do we really need to spend months training students to compute the usual antiderivatives in Calculus 2? Don't we know that they promptly forget those details anyway? If the goal is to "teach integration," why don't we mention the Risch [1969; 1970] algorithm and solvability of integration in finite terms in the course [Rosenlicht 1972]? The first college course need not be locked in 1950 and be overwhelmed by technical issues defined back then.

Wouldn't most students be better served by more work in setting up problems that can be solved by integration? Couldn't a more balanced use of basic mathematics and modern computing give students both sound concepts and powerful tools that they retain? Couldn't this approach go beyond Thomas's old syllabus, to teaching more advanced concepts while at the same time giving students tools to produce more reliable and difficult computations? How do we teach the concepts and computing to realize this potential?

Combining Calculus with Computing

Mathematica and Maple are important advances in mathematics, and it seems silly to ignore them in elementary courses; yet they do present teaching challenges. The complaint that allowing students to use computing just has them pushing buttons is simple-minded. The course by Davis et al. [2008] and Gray et al.'s wonderful book on differential geometry [2006] are both good starts at creative teaching of mathematics with computing. But I don't think that we have found the magic formula for combining calculus and computing. It is still hard to help students learn integration by parts and show them why it is important.

We Need a Different Foundation, Students Need Encouragement

The usual foundations of calculus (based on pointwise limits) are wrong! Is a function with derivative +1 *at a point* increasing there? Shouldn't it be (locally)? I agree with Lax et al. [1976], whose text tried to make this point long ago. A uniform notion of derivative makes the foundations of calculus "smoother" and much closer to natural intuition [Stroyan 2007; see "derivable" on pp. 1181ff in Nelson 1977]. Formulas based on the high school functions are complex analytic, so they satisfy the stronger approximation. I don't find the few standard pointwise "counterexamples" like $x^2 \sin(1/x)$ at 0 compelling (in terms of the overall contribution of calculus); they seem

more like an artifact of the pointwise over-generalization of the notion of local linearity. It is worth some of the *instructor*'s time thinking about stronger foundations for smooth functions that are closer to the intuitive notions—say, more like classical complex variables than modern real variables. This approach could encourage students to reason about the subject themselves.

It can be destructive to tell students a plausible argument that convinces them—and then say, "Of course, that's false!" and proceed with a technical argument that they don't understand. But I also believe we should *not* teach foundations of analysis in basic calculus. I also don't advocate teaching Robinson's 50-year-old solution to the foundations with infinitesimals [Robinson 1974; Keisler 1976; Nelson 1977; Lawler 2011; Goldbring 2010; Ehrlich 2012], though I think it deserves mention in the course and personally feel comfortable allowing students to try to reason informally that way. When I read a student derivation in a project [Stroyan 2011], I ask myself how I could complete the argument technically and whether or how to help the student improve their argument. Gauss [1902] didn't have "rigorous definitions" in the sense of Stewart [2011], say. For example, Gauss wrote

> A curved surface is said to possess continuous curvature at one of its points A, if the directions of all the straight lines drawn from A to points of the surface at an infinitely small distance from A are deflected infinitely little from one and the same plane passing through A.
>
> [1902, 6]

Proofs in Calculus

The usual proofs most teachers know by heart are only "simple" and convincing once you understand the formulation in function notation. This wasn't always "obvious" to smart people thinking about varying quantities [Bos 1974]. I think that it is better to encourage student reasoning at a more informal level using concepts such as Keisler's [1976] infinitesimal microscopes (or as in Davis et al. [2008], or Hughes-Hallett et al. [2009] and McCallum et al. [2009]) to help students find their own "proofs" of some of the ideas. The students' difficulties are much more basic, such as understanding what a contrapositive is, as opposed to comprehending the generality of Apostol's statement of the second half of the Fundamental Theorem of Integral Calculus [1967, 202] (a great book, but silly generality in this case). I believe that if we have—in our own minds—stronger foundations, we could better encourage student reasoning.

I also think that 21st-century calculus could make some "leaps of faith" that we don't prove at all. Lots of pure and applied things in basic calculus depend on unique solution of smooth initial value problems. It's a rich topic, so let students use and explore the consequences. Quite a while ago, Peter Lax wrote, "That today we can use computers to explore the solutions of [differential] equations is truly revolutionary; we are only beginning to glimpse the consequences" [Uhl n.d.]. William Boyce, author of a popular text on differential equations [Boyce and DiPrima 2008] wrote in 1995:

> The traditional course on differential equations that I took many years ago and taught up until the past couple of years dealt almost entirely with derivation of formulas for solutions of various kinds of differential equations,.... Most of the algebraic manipulations featured in the traditional course can now easily be relegated to a computer.... A good deal of this traditional material can be dispensed with and replaced by experiences more valuable for the student.... A decrease of time spent on symbol manipulation by hand should provide an opportunity for more emphasis on conceptual understanding,.... Details of mathematical procedures and algorithms are rapidly forgotten unless they are used frequently, but underlying concepts and ideas become part of an individual's mindset. [Uhl n.d.]

Fifteen years later, many of us are teaching as if we still don't have DNA...;-).

References

Apostol, Tom A. 1967. *Calculus.* Vol. 1. 2nd ed. Waltham, MA: Blaisdell Publishing Co.

Bos, H.J.M. 1974. Differentials, higher-order differentials and the derivative in the Leibnizian calculus. *Archive for History of Exact Sciences* 14 (1): 1–90. http://tau.ac.il/~corry/teaching/toldot/download/Bos1974.pdf .

Boyce, William E., and Richard C. DiPrima. 2008. *Elementary Differential Equations.* 9th ed. New York: Wiley.

Bressoud, David M. 2004. The changing face of calculus: First- and second-semester calculus as college courses. *MAA Focus* 24 (8) (November 2004): 14–16. http://www.maa.org/pubs/nov04.pdf .

Campbell, Paul J. 2006 Calculus is crap. *The UMAP Journal* 27 (4): 415–429. (Neither he nor I believe the provocative claim in the title!)

Davis, William J., J.J. Uhl, and Horacio Porta. 2008. Calculus and Mathematica. http://www-cm.math.uiuc.edu/WhatIsCM .

Ehrlich, Philip. 2012. The absolute arithmetic continuum and the unification of all numbers great and small. *Bulletin of Symbolic Logic* 18 (1): 1–45. (See Theorem 20 on p. 37.)

Gauss, Karl [sic] Friedrich. 1902. *General Investigations of Curved Surfaces of 1825 and 1827.* Translated with notes and a bibliography by James Caddall Morehead and Adam Miller Hiltebeitel. Princeton, NJ: C.S. Robinson and Co., University Press. 2005. Reprint. Ann Arbor, MI: Scholarly Publishing Office, University of Michigan Library. 2010. Reprint. Charleston, SC: Nabu Press. 2011. Reprint. Toronto, Canada: University of Toronto Libraries. 2012. Reprint. North Manchester, IN: Repressed Publishing LLC.

Goldbring, Isaac. 2010. Hilbert's fifth problem for local groups. *Annals of Mathematics* 172 (2): 1269–1314. (Solved for the first time with infinitesimals.)

Gray, Alfred, Elsa Abbena, and Simon Salamon. 2006. *Modern Differential Geometry of Curves and Surfaces with Mathematica.* 3rd ed. Boca Raton, FL: Chapman and Hall / CRC.

Hughes-Hallett, Deborah, Andrew M. Gleason, William G. McCallum, David O. Lomen, David Lovelock, Jeff Tecosky-Feldman, Thomas W. Tucker, Daniel E. Flath, Joseph Thrash, Karen R. Rhea, Andrew Pasquale, Sheldon P. Gordon, Douglas Quinney, and Patti Frazer Lock. 2009. *Calculus: Single Variable.* 5th ed. New York: John Wiley & Sons.

Keisler, H. Jerome. 1976. *Elementary Calculus: An Infinitesimal Approach.* Boston, MA: Prindle, Weber & Schmidt. 1986. 2nd ed. 2000. Online 2nd ed. `http://www.math.wisc.edu/~keisler/calc.html`. 2012. 3rd ed. Mineola, NY: Dover Publications. `http://www.math.wisc.edu/~keisler/calc.html`.

Lawler, Gregory F. 2011. Comments on Nelson [1977]. *Bulletin (New Series) of the American Mathematical Society* 48 (4) (October 2011): 503–506.

Lax, Peter, Samuel Burstein, and Anneli Lax. 1976. *Calculus with Applications and Computing.* Vol. 1. New York: Springer-Verlag.

McCallum, William G., Deborah Hughes-Hallett, Andrew M. Gleason, David O. Lomen, David Lovelock, Jeff Tecosky-Feldman, Thomas W. Tucker, Daniel E. Flath, Joseph Thrash, Karen R. Rhea, Andrew Pasquale, Sheldon P. Gordon , Douglas Quinney, and Patti Frazer Lock. 2009. *Calculus: Multivariable.* 5th ed. New York: John Wiley & Sons.

Nelson, Edward. 1977. Internal set theory: A new approach to nonstandard analysis. *Bulletin of the American Mathematical Society* 83 (6) (1977): 1165–1198. 2011. Reprint. *Bulletin (New Series) of the American Mathematical Society* 48 (4) (October 2011): between 506 and 507.

Robinson, Abraham. 1974. *Non-standard Analysis.* New York: American Elsevier. 1996. Revised ed. Princeton, NJ: Princeton University Press.

Risch, R. 1969. The problem of integration in finite terms. *Transactions of the American Mathematical Society* 139: 167–189.

_____. 1970. The solution of the problem of integration in finite terms. *Bulletin of the American Mathematical Society* 76 (3): 605–608.

Rosenlicht, Maxwell. 1972. Integration in finite terms. *American Mathematical Monthly* 79 (9) (November 1972): 963–972.

Stewart, James. 2011. *Calculus.* 7th ed. Pacific Grove, CA: Brooks Cole.

Stroyan, Keith D. 2006. The changing face of calculus: Engineering math at the University of Iowa. *MAA Focus* 26 (2) (February 2006): 8–9. http://www.maa.org/pubs/feb06web.pdf .

_____. 2007. Calculus with infinitesimals. In *The Strength of Nonstandard Analysis*, edited by Imme van den Berg and Vitor Neves, 369–394. New York: Springer. http://www.cima.uevora.pt/FullText/vdBerg-Neves%20(2007)%20@%20NSMBOOKIV24_11_06.pdf .

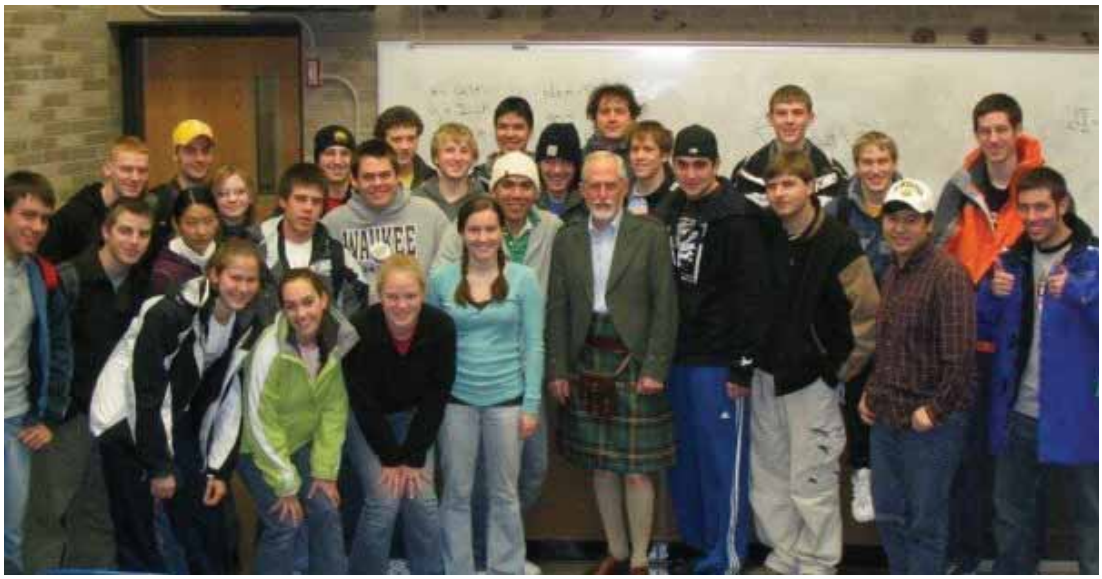
_____. 2011. Why do so many students take calculus? *Notices of the American Mathematical Society* 58 (8) (September 2011): 1122–1124. http://www.ams.org/notices/201108/rtx110801122p.pdf .

_____. n.d. A brief introduction to infinitesimal calculus. http://www.math.uiowa.edu/~stroyan/InfsmlCalculus/InfsmlCalc.htm.

Thomas, George B. 1951. *Calculus and Analytic Geometry.* Reading, MA: Addison-Wesley.

Uhl, Jerry. n.d. Jerry Uhl's favorite quotations. http://www.dougshaw.com/uhl/ .

About the Author



Keith D. Stroyan is Professor of Mathematics at the University of Iowa. He did undergraduate work in engineering and physics at Drexel, graduate work in mathematics at Caltech, and had a postdoc at the University of Wisconsin. He has had a career-long obsession with teaching, especially the use of computers in undergraduate mathematics courses, and as a result served a long sentence in calculus reform school during the 1990s. Several of his books are available on his Website http://www.math.uiowa.edu/~stroyan/ . Keith was a 2006 recipient of the Deborah and Franklin Tepper Haimo teaching award from the Mathematical Association of America.

The Mathematical Sorting Hat

Andrew Beveridge and Sean Cooke

Dept. of Mathematics, Statistics, and Computer Science
Macalester College
Saint Paul, MN 55105
abeverid@macalester.edu

Introduction

The most well-known matching problem in recent popular culture is in J.K. Rowling's *Harry Potter* series [Rowling 1998]. Each year, roughly 40 new first-year students at Hogwarts School of Witchcraft and Wizardry are assigned to one of the residential Houses (Griffindor, Ravenclaw, Hufflepuff, and Slytherin). This duty is performed by the magical Sorting Hat, which considers each student in turn and determines the best match for that student among the four Houses. The Sorting Hat uses its magical perception to assign students to houses that best match their personalities. However, we also notice that the assignments split fairly evenly, with each House receiving about 5 girls and 5 boys. Perhaps this house distribution and gender balance reflects a secondary goal of the Sorting Hat.

A similar sorting occurs annually on college campuses across the United States. The past two decades have witnessed the rise of the *first-year seminar*, a curricular strategy to help students transition from high school to college [Hunter and Linder n.d.]. A seminar is a small course that emphasizes active learning, discussion, and exchange of ideas.

First-year seminars leverage the closeness of this format to introduce the paradigms and standards of higher education. First-year seminars at some institutions also strive to develop community within the classroom, the student body, and the institution as a whole, in part with a view to retaining students. Many colleges devote considerable effort in assigning students to these seminars, trying to create an immediate sense of belonging, just as the fictional Sorting Hat achieves at Hogwarts.

This paper describes the development and implementation of an automated procedure—the *Mathematical Sorting Hat*—for assigning first-year students to seminars, adopted at Macalester College in 2009. "The Hat"

The UMAP Journal 33 (2) (2012) 99–118.

models the matching problem using a weighted bipartite graph that captures both student preferences and college constraints. We run the well-known Hungarian algorithm to find a minimum-weight perfect matching of this graph. Such a matching corresponds to an optimal assignment of students to seminars.

When run on historical data, the Hat finds student assignments that are quantitatively better than the manual assignments used in those years. In the three years since adoption, the Hat has produced high-quality matchings. The algorithm, implemented in Matlab (and compatible with the open-source Octave project), matches roughly 500 students to seminars in about 70 seconds on a 2.8 GHz Mac Pro. It saves 40 hours of labor; it also provides repeatability and customization.

We explain the modeling process of converting the applied problem into a classical optimization problem. Many colleges are faced with similar assignment problems, and we hope that this paper inspires others to consider automated methods.

We quickly survey matching problems in recent applications and then discuss the history of the First-Year Course Assignment Problem (FYCAP) at Macalester. Then we develop our weighted bipartite graph model for FYCAP, call on mathematical magic to solve it, reflect on our solution, and discuss generalizations. An **Appendix** collects implementation notes concerning repeatability, missing data, and fairness. The Matlab / Octave code for the Mathematical Sorting Hat is available at the COMAP Website.

Applications of Matching Algorithms

Matching Medical Residents

The National Resident Matching Program (NRMP) [2011] each year places 30,000 graduating medical doctors into hospital residencies. This is an example of a matching problem known as the *stable marriage problem*. The preferences of both applicants and hospitals are taken into account. The algorithm finds a matching that is *stable* in the sense that no two matched pairs would agree to switch assignments. Since 1998, the NRMP uses an applicant-proposing deferred-acceptance algorithm [Gale and Shapley 1962; Roth and Peranson 1999] to find a stable matching. The procedure gives applicant preferences a slight edge over institutional preferences. Indeed, there is an inherent asymmetry to this problem, since the parties doing the "proposing" would never do better (and may do worse) if the roles were reversed.

School Assignments

An annual matching is also involved in the assignment of 90,000 new students to public high schools in New York City [Abdulkadiroğlu et al.

2005]. The approach employs the same stable matching strategy used by the NRMP. However, the algorithm was adapted to co-exist with existing mechanisms, such as lotteries and separate offers from specialized schools. These alterations can lead to unstable matchings, which must be tolerated in a complex system with legacy procedures.

Dorm Room Allocation

The standard procedure to match college students to dorm rooms is:

- A student may choose to remain in the current room or else enter the applicant pool.
- Each applicant in the pool is randomly assigned a priority (perhaps weighted by seniority).
- Applicants choose an available room according to the priority ordering.

However, this naïve algorithm leads to an inefficiency: Many students opt to remain in their current room for fear of ending up worse. A provably superior extension was proposed by Abdulkadiroğlu and Sönmez [1999]:

- The priority queue is created as before.
- An applicant can request an occupied dorm room, at which point the current occupant moves to the top of the queue, directly in front of the requester.
- The current occupant of a requested room can relinquish it and choose a new room (possibly requesting another occupied room, which pushes that occupant to the top of the queue) or decide to remain in the current one (in which case the requester must make a different request).

Kidney Transplants

Abraham et al. [2007] designed a program to improve efficiency in matching donors to recipients in kidney transplants. Finding a compatible match for a potential recipient is very difficult. A recipient may have a willing donor who has an incompatible tissue type; the two attempt to find another recipient-donor pair for which a cross-transplant would be a better match. The algorithm also tries to connect longer chains of people, though the logistics of multiple almost-simultaneous transplants can be a limiting factor. The first cross-transplant using a match from this algorithm was performed in December 2010 [Emspak 2010].

The Hogwarts Sorting Hat

The Sorting Hat of Hogwarts appears to perform an *online weighted bipartite matching algorithm*, processing the information and making decisions as

it goes, rather than considering the whole input before making any assignments. An online algorithm is forced to make decisions that ultimately may not be globally optimal. Indeed, no online weighted bipartite matching algorithm can achieve the performance of an optimal offline algorithm (that has access to all of the data) for all inputs [Khuller et al. 1991]. However, this result should not affect our faith in the Sorting Hat at Hogwarts: A magically sentient hat is probably not bound by the same rules as a computer algorithm.

The Assignment Problem

The *Assignment Problem* (cf. [Cook et al. 1998]) closely models our seminar assignment problem. Suppose that we have n workers and n jobs to be filled. Each worker is qualified for a subset of jobs, and a worker's salary varies with the job assigned (the worker may have a higher certification, capability, or efficiency for one job vs. another). The objective is to find a matching of workers to jobs that minimizes total cost.

We recall some standard terminology from graph theory (cf. West [2000]). A *graph* $G = (V, E)$ consists of a set of vertices V along with a set of edges E consisting of pairs of vertices. When $v_1, v_2 \in V$ and $e = (v_1, v_2) \in E$, we say that v_1 and v_2 are *adjacent* and that edge e is *incident* with both v_1 and v_2. A *matching* M of G is a set of edges such that each vertex is incident with at most one edge. When a vertex v is incident with an edge in M, then we say that M *covers* v. A *maximum matching* is one of maximum size and a *perfect matching* is one that covers all vertices. A perfect matching is also a maximum matching, but the reverse statement is not always true (for example, consider a graph on an odd number of vertices).

In the Assignment Problem, the graph has one vertex for each worker and one vertex for each job. We add an edge between worker x and job y whenever x is qualified to perform job y. This results in a *bipartite graph*; the vertex set is partitioned into disjoint sets X, Y with every edge having one end in X and the other in Y.

Finally, we address the cost of a matching by making the graph a *weighted graph*: Each edge $e \in E$ is assigned a weight w_e, corresponding to the cost of that assignment (salary of the worker). The objective is to find a perfect matching of minimum weight.

This optimization problem can be solved by the *Hungarian algorithm*, developed by Kuhn by extending the work of the Hungarian mathematicians Kőnig and Egerváry [Cook et al. 1998; Evans and Minieka 1992]. This algorithm runs in time $\mathcal{O}(n^3)$. The fundamental strategy is to search for augmenting paths. Suppose that you start with a matching M on a graph G. An *M-alternating path* is a set of successive edges of G that are alternately in M and not in M. An *M-augmenting path* is an M-alternating path that begins and ends at uncovered vertices. Such an augmenting path identifies

a series of swaps that results in a larger matching.

Essentially, the Hungarian algorithm considers a series of subgraphs $G_1 \subset G_2 \subset G_3 \subset \cdots$ of the bipartite graph G. On each graph G_k, we attempt to find a perfect matching, using an augmenting path algorithm. If no perfect matching exists, we add enough edges (of minimal additional cost) to get around a current obstruction to obtain G_{k+1}, and then repeat the process. COMAP published a detailed and accessible Module [Gale 1981] on the optimal assignment algorithm that provides an economic interpretation of the procedure. (This Module considers the maximum weight assignment problem, which formulation is equivalent to the minimum weight version via multiplication by -1.)

First-Year Courses at Macalester College

Every department at Macalester offers one or more first-year seminars, called First-Year Courses (FYCs). The content of a seminar is discipline-specific; recent topics include global health, the social structure of the World Wide Web, representations of vampires, and the American West. All FYCs develop writing skills and critical thinking. The professor becomes the students' academic advisor, meeting with them before the semester begins and coordinating various social functions. Some FYCs are residential (at the request of the professor), with the students housed in close proximity to one another.

A typical year has roughly 500 first-year students and 35 FYCs. Incoming students read the seminar descriptions and rank their top four choices.

The Office of Academic Programs compiles the preferences and matches students to seminars, while observing modest constraints on the distribution of students to seminars.

Prior to 2009, the assignment was performed by hand. Each student's preferences were listed on a sheet of paper, and the sheets were organized into piles (spread on the floor!), each pile representing a seminar. Much like the augmenting paths of the Hungarian algorithm, staff members would look for chains of swaps to improve the overall profile of the current assignment. The entire process consumed 40 person-hours before the staff was satisfied with the assignment. **Figure 1** shows the historical results for these manual assignments from 2003 to 2008. On average, Academic Programs was able to place 87% of the students into their first or second choice.

The Collaboration

The project was conceived by the second author in Fall 2008 while serving as a teaching assistant for an FYC in discrete mathematics. At one of the first class meetings, the professor introduced graph theory by describ-

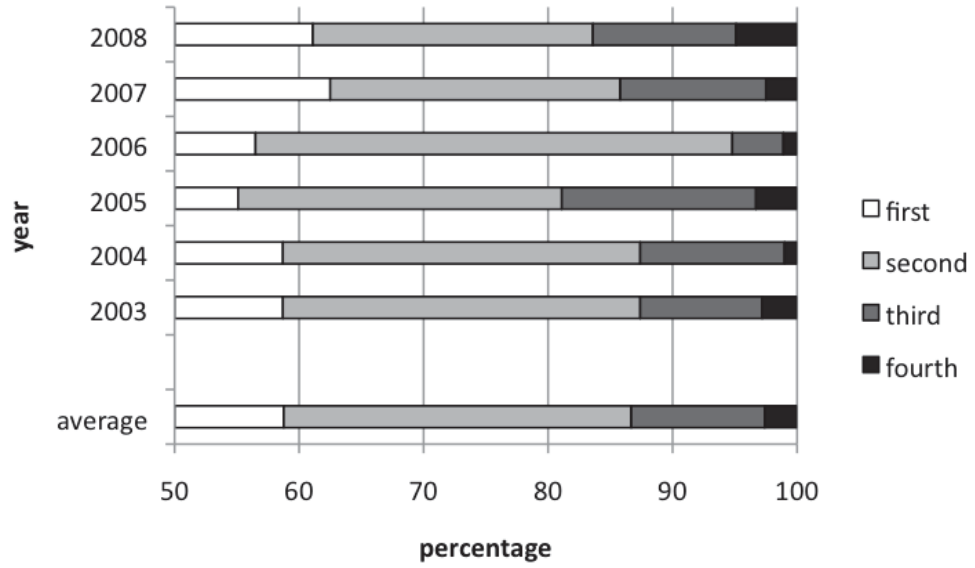


Figure 1. Distributions of manual assignments by the Office of Academic Programs, 2003–2008.

ing the connection between graph matchings and seminar assignments. With interest piqued, this second author suggested that specific topic for his project in the senior seminar on combinatorial optimization taught by the first author.

In Spring 2009, the authors scheduled a meeting with Academic Programs to discuss the project. While we hoped that an implementation by us would eventually be adopted, it was crucial to "sell the client" on such a solution. In the initial meeting, the authors described the project in broad strokes. Academic Programs was supportive and agreed to evaluate an implementation. They provided information on their constraints, methods, and objectives, as well as a historical data set from 2004 (that cohort of students had by then graduated). Using these specifications and test data, we implemented a Mathematical Sorting Hat. The Hat slightly out-performed the manual assignments from 2004 and was deemed by all a success.

In Summer 2010, encouraged by the initial results, Academic Programs agreed to collaborate on a further version of the Hat. In a series of meetings, we and they mapped out specifications and requirements. Academic Programs provided feedback on the model and further refined their constraints to reflect more accurately their needs and priorities. We and they agreed on the data format for input: a comma-separated value (CSV) format. The second author implemented the changes, creating a fully-featured Mathematical Sorting Hat.

In July 2010, the Mathematical Sorting Hat was used to create an assignment for the incoming first-year students. Simultaneously, Academic Programs developed a manual assignment. Once again, the Hat's assignment was slightly better. Convinced of the benefits of the Mathematical

Sorting Hat, Academic Programs opted to use the Hat's assignment. Since then, Academic Programs has fully embraced the Hat's assignments and no longer spends any effort on manual assignments.

The Mathematical Sorting Hat

From the college's perspective, an assignment should result in seminar enrollments that are a microcosm of the college, in terms of gender balance and international diversity. From the students' perspective, a successful assignment should place as many students as possible in seminars that they most desire.

In other words, the First-Year Course Assignment Problem (FYCAP) is a constrained optimization problem in which we try to maximize student satisfaction while respecting constraints on seminar enrollment.

We converted FYCAP into an assignment problem, mapping students to workers and seminars to jobs. This presented two main challenges:

- The bipartite graph structure had to enforce all of the college constraints.
- The edge-weighting scheme had to encode student preferences and also result in an assignment in which most students receive their first or second choice.

Figure 2 shows the the performance results of the Hat for the last three years.

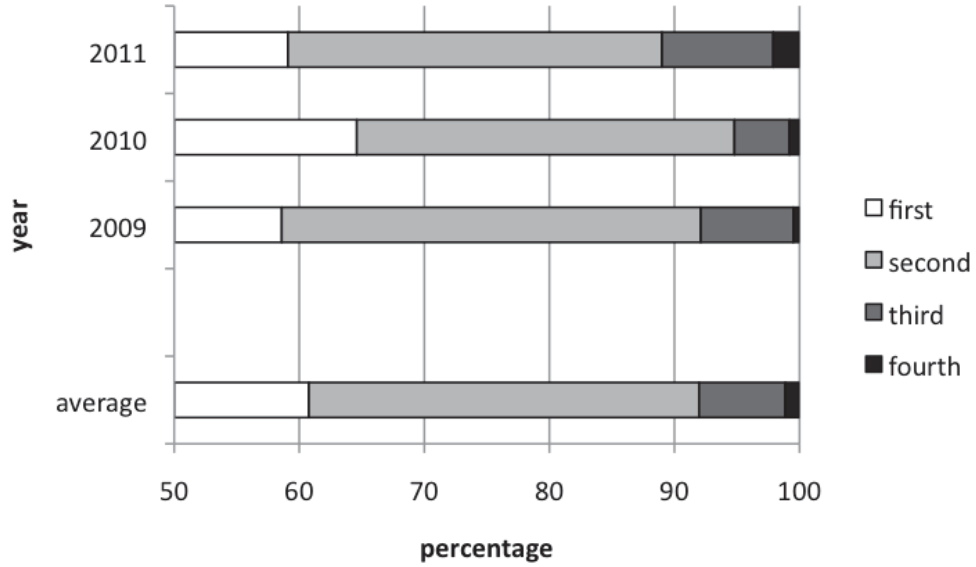


Figure 2. Distributions of assignments produced by the Mathematical Sorting Hat, 2009–2011.

These results compare favorably to those for the earlier manual assignments in **Figure 1**. The Hat placed 91% of students into their first or second

choice. The Hat's best year (2010) is comparable to the best manual year (2006). Indeed, we expect the Hat to produce high-quality assignments. We were also actually pleased to discover that its assignments were *not* significantly better than the historical manual assignments, meaning that the hours of effort put into those earlier assignments truly paid off.

Constructing the Mathematical Sorting Hat

We describe the bipartite graph $G = (V, E)$ whose vertex set V is partitioned as $V = X \cup Y$, where X consists of students and Y consists of seminar seats. Suppose that we have r students and s seminars, where each seminar has enrollment limit of t students, and $r \leq st$. The r students correspond to vertices $X = \{x_1, x_2, \ldots, x_r\}$. If $r < st$, then we add $(st - r)$ *dummy* vertices $x'_{r+1}, x'_{r+2}, \ldots, x'_{st}$ to X. For $1 \leq i \leq s$, let $Y_i = \{y_{i,1}, y_{i,2}, \ldots y_{i,t}\}$ represent the t seats in seminar i. Let $Y = \cup_{i=1}^{t} Y_i$ be the set of all seminar seat vertices.

We now have a natural mapping between FYCAP and an Assignment Problem: The students are workers and the seminar seats are jobs.

Constraints

The Office of Academic Programs provided four constraints:

- Every seminar must contain at most 16 students.
- The minimum seminar size is 10 students.
- No seminar can have a majority international enrollment (international students make up 11% of the Macalester student body).
- The final constraint is a modicum of gender balance. Having already placed restrictions on international students, we decided to enforce gender balance only for domestic students. Therefore, in each fully enrolled seminar, 4 seats are reserved for domestic women and 4 seats are reserved for domestic men. For seminars with fewer than 16 students, we only require 3 domestic women and 3 domestic men.

Edge Structure of the Graph

We enforce these four constraints via the edge structure of the graph. **Figure 3** shows the local graph structure of the Mathematical Sorting Hat. It depicts the 16 seminar seat vertices for one seminar, along with one student of each type that is interested in the seminar topic.

We divide the seminar seats into three categories, with 4 domestic female seats, 4 domestic male seats, and 8 generic seats.

Every domestic female student interested in this seminar is adjacent to the 4 domestic female vertices and the 8 generic vertices. Likewise, interested domestic males are adjacent to 4 domestic male seats and 8 generic

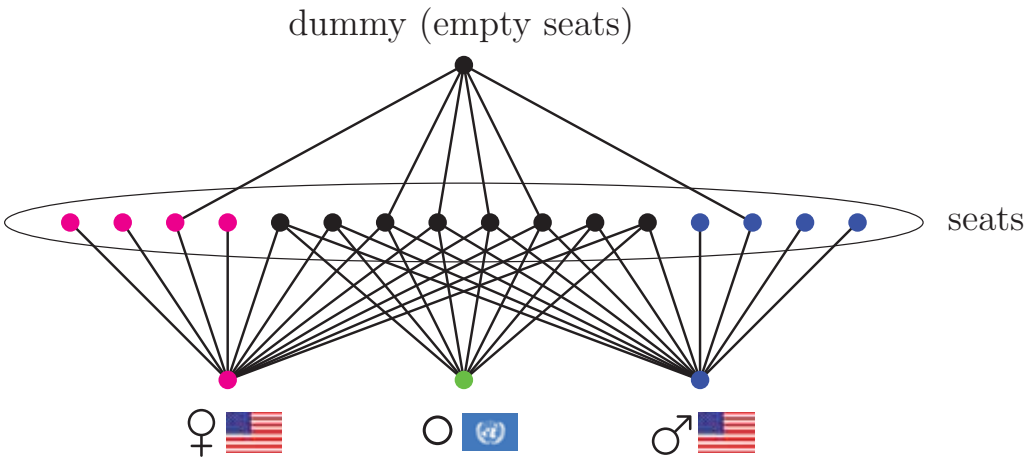


Figure 3. The edges connecting the 16 seats of a seminar to the 4 types of students: domestic female, international, domestic male, and dummy students. The edge structure ensures that the seminar assignment will adhere to the college's constraints.

seats. Finally, every interested international student is adjacent to the 8 generic vertices.

Dummy students correspond to empty seminar seats. We allocate edges from dummy students to enforce the minimum seminar size of 10. Therefore, every dummy vertex is adjacent to 1 female domestic seat, 1 male domestic seat, and 4 generic seats. This allocation allows the lower-enrolled seminars to have a gender minimum of 3 students, while fully-enrolled seminars have a gender minimum of 4 students.

Table 1 summarizes the distribution of seats to students.

Table 1.
The number of seminar seats adjacent to each student. The existence of edges depends upon the category of the seat and the characteristics of the student. The minimum enrollment per category is total seats minus dummy seats.

	total	USA male	USA female	international	dummy	minimum enrolment
USA female seats	4	4	0	0	1	3
USA male seats	4	0	4	0	1	3
generic seats	8	8	8	8	4	4
total seats	16	12	12	8	6	10

We reflect on the structure of the bipartite graph $G = (X \cup Y, E)$ that we have created. By adding dummy students to fill empty seats, we have $|X| = |Y|$. The edges incident with student vertices reflect their seminar preferences. The global structure of these student edges also enforce gender and international balance. The edges incident with dummy vertices enforce minimum seminar size. The dummy edge structure also mitigates

the gender constraint for lower enrolled seminars. The graph G always has a maximum matching (in terms of number of edges of the matching). This maximum matching is a perfect matching if and only if there is a valid assignment of students to seminar seats.

Modeling Student Preferences with Edge Weights

We use edge weights to capture student preferences. In this section, we describe how our choice of weighting scheme influences the global characteristics of the optimal assignment. Let (w_1, w_2, w_3, w_4) be the edge weights for the student choices, and let w_5 be the weight for dummy choices. These weights satisfy $w_1 \le w_2 \le w_3 \le w_4 < w_5$.

We give two simple examples to illustrate the impact of our weighting scheme.

Example. If we set $w_i = 1$ for $1 \le i \le 4$, then the algorithm will ignore student preference between acceptable seminars.

Example. If we choose $w_1 = 1$ and $w_2 = w_3 = w_4 = 100$ (or any large number), then the algorithm will maximize the number of students in their first choice and treat the other options equivalently.

For $1 \le i \le 4$, let P_i denote the percentage of students assigned to their ith choice. We define the *profile* of an assignment to be (P_1, P_2, P_3, P_4).

Using the average of the historical results in **Figure 1** as a guide, we find that Academic Programs requires that the assignment profile obey the following inequalities:

$$\begin{aligned} P_1 &\ge 59\%, \\ P_1 + P_2 &\ge 87\%, \\ P_1 + P_2 + P_3 &\ge 97\%. \end{aligned} \tag{1}$$

These conditions are the primary criterion for measuring the quality of an assignment.

The secondary criterion, specified by Academic Programs, is minimizing the number of students who get their fourth choice.

We did not directly incorporate these two criteria as hard constraints to our matching problem. Instead, we considered multiple weighting schemes, both experimentally and heuristically. A weighting was considered viable if it gave an assignment that met these two criteria. These weightings were reported back to Academic Programs.

We considered two types of weighting schemes (w_1, w_2, w_3, w_4). First, we considered the naïve weighting $(1, 2, 3, 4)$. Second, we considered geometric weightings of the form $(1, \alpha, \alpha^2, \alpha^3)$ for $\alpha > 1$. Geometric weightings have a natural interpretation, described below.

What about dummy vertices? Assignment of a dummy vertex corresponds to an unfilled seat in the seminar. The dummy vertices have no

preferences, so all of their edges are assigned the same weight w_5 We always prefer to assign a real student to a seminar, so the dummy weight should be much larger than w_4 to ensure priority to real students. For a geometric weighting, setting $w_5 = \alpha^4$ is sufficient.

Figure 4 presents the resulting assignment profiles for four different weighting schemes on the data from 2004. (The 2004 data were the initial test data provided by Academic Programs.) The results for 2004 are typical for other historical data sets.

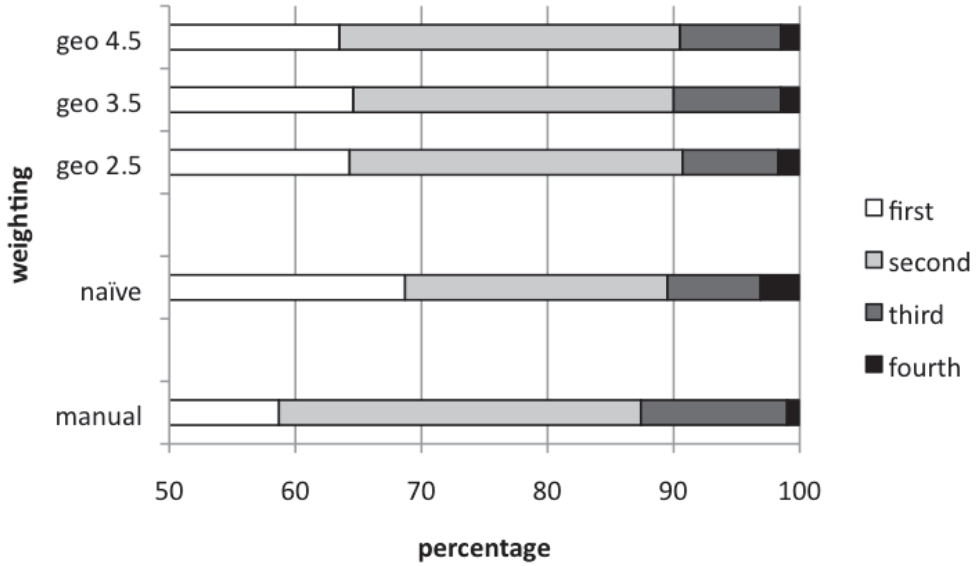


Figure 4. Profiles of assignments for four weighting schemes for the data from 2004. Geometric weightings are of the form $(1, \alpha, \alpha^2, \alpha^3)$; the naïve weighting is $(1, 2, 3, 4)$; the manual assignment is that performed by Academic Programs in 2004.

We see that all four weightings give assignments that satisfy the conditions of **(1)**. All of the geometric weightings improve upon the manual assignment performed in 2004. The naïve weighting achieves the largest percentage of first-choice placements, but at the cost of assigning far more students to their fourth choice.

To determine the best weighting, we use our secondary criterion: the number of students receiving their fourth choice. The geometric weightings with $\alpha = 3.5$ and $\alpha = 4.5$ reduce fourth-choice placements to 1.55%, while $\alpha = 2.5$ achieves only 1.75%. Larger values of α give the same ranking as $\alpha = 4.5$, suggesting that the graph structure obstructs further improvements. Deciding between $\alpha = 3.5$ and $\alpha = 4.5$ required consulting Academic Programs: They had to decide whether it was worth moving 1.16% down from first to second choice in order to move 0.57% up from third to second choice. They were pleased with the trade-off; so for our final implementation, we settled on a geometric weighting with $\alpha = 3.5$.

We now give an intuitive interpretation of a geometric weighting scheme. We start by considering integer α weights, then describe how sensitivity

analysis led us to choose half-integer weights instead.

Consider the weighting scheme (w_1, w_2, w_3, w_4). We may always take $w_1 = 1$ because multiplying all weights by a constant $c > 0$ does not affect the optimal matching. The geometric weighting scheme $(w_1, w_2, w_3, w_4) = (1, \alpha, \alpha^2, \alpha^3)$ has a natural interpretation in terms of trade-offs made by swapping students between seminars. Suppose that we have a student in her third choice but we can move this student into her second choice via a swaps that move k other students down from first choices to second choices; see **Figure 5**.

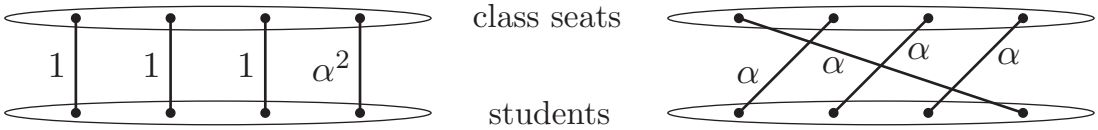


Figure 5. Two potential assignments using a geometric weighting scheme $(1, \alpha, \alpha^2, \alpha^3)$. The matching on the left has weight $3 + \alpha^2$ and the matching on the right has weight 4α. When $\alpha < 3$, the left matching has lower weight; and when $\alpha > 3$, the right matching has lower weight.

The corresponding change in the weight of the matching is

$$k(w_2 - w_1) + (w_2 - w_3) = (k+1)\alpha - k - \alpha^2 = -(\alpha - 1)(\alpha - k).$$

Therefore the swap is acceptable whenever $-(\alpha - 1)(\alpha - k) \le 0$, or equivalently, $k \le \alpha$. This inequality captures our willingness to move $k \le \alpha$ students down from first to second choices in order to move a single student up from third choice to second choice.

The advantage of a geometric weighting is its *translation invariance*: We obtain an equivalent inequality for the trade-off for moving k students down from second choices to third choices in order to move one student up from fourth choice to third choice. (The naïve weighting does not have this invariance; in fact, its penalty for fourth choices is not severe enough.) Constraints for other swaps have similar interpretations.

Figure 5 also reveals a minor disadvantage to a geometric weighting $(1, \alpha, \alpha^2, \alpha^3)$ with integer α. For $\alpha = 3$, the two matchings have equal weight; so the Hungarian algorithm could return either one as its optimal assignment. To avoid such ambiguity, we guide the algorithm to prefer the right-hand matching by slightly perturbing the weights. In our final implementation, we use $\alpha = 3.5$, which gives a consistent preference for the right-hand matching. Doing so aligns with the goal to reduce the number of students assigned to their fourth choice. A similar argument shows that we favor moving students out of their third choice under comparable circumstances.

Sensitivity Analysis

Sensitivity analysis of α reveals another reason to choose non-integer weights. We ran the Hat on a range of α values on recent data sets. The results for 2010 are presented in **Figure 6**; results for other years are similar.

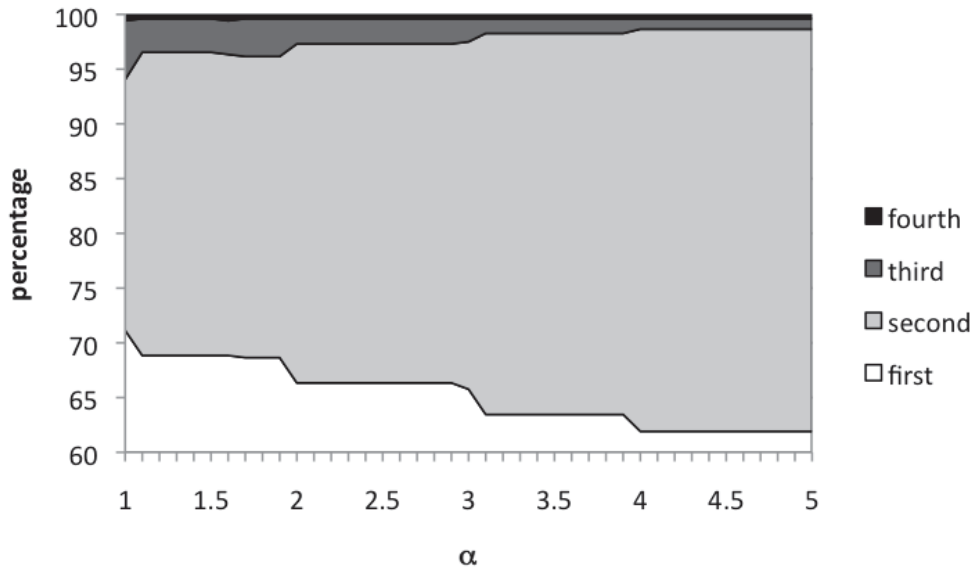


Figure 6. Sensitivity analysis for weight α on the 2010 data set.

- For every $\alpha \geq 4$, the Hat returns the same matching.
- For $\alpha \geq 2$, the optimal matching changes at integer values for α. This suggests that the optimal matchings are primarily shaped by the types of trade-offs exemplified in **Figure 5**.
- The results are slightly more unstable for $\alpha < 2$, with an additional change at $\alpha = 1.5$, probably due to beneficial third-choice to second-choice swaps.

The stability may be related to the relatively small size of the data set; for larger data sets, the stability for $\alpha \geq 2$ might decrease due to additional opportunities for more exotic beneficial swaps.

The sensitivity analysis suggests that for our data sets, half-integer values of α give the most stable results when $\alpha \geq 2$, while integer α values give the most unstable results. Therefore, we use half-integer values for α.

In working with Academic Programs, we focussed on $\alpha = 2.5$ and on $\alpha = 3.5$. Both values result in high-quality matchings. Moreover, having a choice was instrumental in making Academic Programs comfortable with trusting the Hat. Indeed, these values were chosen because their results match the historical assignment methodology. *Client satisfaction is the ultimate requirement*, and we were very happy to provide a solution that Academic Programs could embrace.

Running the Sorting Hat

We now describe how we use the Mathematical Sorting Hat code in practice, and discuss how we handle the produced assignment. The Matlab / Octave code, along with a more detailed user's manual, is available on the COMAP Website.

The Sorting Hat requires two inputs: the location of the data file, and the weight α. The code constructs the weighted bipartite graph model described in this section, with one modification. We actually create a complete bipartite graph: Edges from a student to an undesired seat are given weight α^5. The advantage to this modification is that the Hungarian algorithm then always returns an assignment. Indeed, since we now have a complete bipartite graph, there are $n!$ perfect matchings. The Hungarian algorithm will return one of minimum weight.

After the algorithm completes (in time $\mathcal{O}(n^3)$), the Hat organizes the the results into two human-readable formats: one list organized by student name, another list organized by class name. Next, we decide (manually) whether this assignment is acceptable. We say that an assignment is *student-valid* when every student is assigned to one of her preferred courses. To assess validity, the Hat produces a third output: a 5-entry vector of the form (first, second, third, fourth, other). The first entry contains the number of students assigned to their first choice; the second, third, and fourth entries are populated similarly. The final entry contains the number of students who were not assigned to any of their specified choices. This final entry is zero if and only if the assignment is student-valid.

We say that the assignment is *college-valid* if it adheres to all of the college's constraints (class size, gender balance, international balance). It is possible for an assignment to be student-valid but not college-valid (or vice versa). For example, suppose that there is a particularly "unpopular" course, and only eight students include that course among their choices. In this case, those eight students will be assigned to that course, but it will still be below the minimum enrollment of 10. When such a case occurs, there is no way to produce an assignment that is both student-valid and college-valid: We must either accept the smaller seminar enrollment, or force students into a course that they do not want. As you would expect, Macalester would choose the former option.

If there is an assignment that is both student-valid and college-valid, then the Hat will find it. For α large enough, the penalties for failing to meet both criteria drive the Hungarian algorithm to find this assignment. In the past three years, the Sorting Hat has always returned a student-valid assignment. However, one year it returned an assignment that was not college-valid. In particular, one seminar did not satisfy the gender balance requirement because not enough men were interested in the course. In this case, Academic Programs allowed this exception to their gender balance constraint (since placing students in unwanted courses is considered unac-

ceptable). In the future, Macalester will continue to bend its constraints in situations like this. However, if we ever fail to find a student-valid assignment, Academic Programs will resolve this by increasing the enrollment cap in a seminar to create a seat for the unmatched student.

In conclusion, we reflect on the necessity of fourth-choice preferences. Indeed, it is natural to wonder whether an assignment exists in which every student gets one of their top three choices. If such an assignment exists that is also college-valid, then the Hat will find it for large enough α. Historically, this has not yet happened. It is easy to assess the impact of the college constraints on these fourth-place assignments: We simply run the Hat on the choice data, altered so that all fourth-choice preferences are removed. If the resulting assignment is student-valid, it will be of lower weight than the assignment from using all four choices. However, this lower-weight assignment will certainly be even further from conforming with the college's constraints. So it is "better" for students and "worse" for the college. Nevertheless, it is useful to perform an assignment using only the first three choices, since the result gives us a way to measure the impact of the college's constraints on student satisfaction.

Conclusions

We have described the modeling of the First-Year Course Assignment Problem at Macalester College into a classical Assignment Problem. We developed a weighted graph that captures student preferences and enforce college constraints. Through experimentation on historical data, we developed an edge-weighting scheme that places 91% of the students in their first or second choice. Academic Programs has been thrilled with these results, both in terms of time saved and in the quality of the assignments.

In the future, we hope to produce even higher-quality results; doing so will require collecting higher-fidelity preference data. For example, rather than a simple numerical ranking, we could use a scale that would allow distinction between strong and weak preferences between a student's first and second choices.

The ratio of seminar seats to constraints was favorable enough that we could encode the constraints via the edge structure of the graph. Other schools would have to develop a graph structure that reflects their specific priorities; too many constraints may lead to no feasible solution. If the constraints are too complex to model in a bipartite graph, a school could consider using a binary integer program to perform the assignment (cf. Bertsimas and Tsitsiklis [1997]).

Appendix: Implementation Notes

We address some practical issues in using the Mathematical Sorting Hat to assign students to seminars.

A successful application of the Hat returns a minimum-weight assignment. However, there is no guarantee that this assignment is the unique minimizer. (In an extreme case, if 64 students gave identical rankings to 4 available courses, then every matching would be minimal.) We believe that most real-world data will have multiple optimal assignments. Therefore, we must address factors that influence the optimal assignment returned by the Hat.

Dependence on Student Ordering

The most important observation is that the Hungarian algorithm gives an advantage to students who are processed first. Indeed, in each iteration, the assignment algorithm makes partial assignments and then tries to improve upon them via alternating paths. A student processed early will be placed into her first choice; she will be moved from it only if there is an augmenting path that leads to a matching of lower weight. On the other hand, a student desiring a seat in an already-full seminar will be assigned to it only if there is augmenting path that improves the current assignment by removing another student from it.

To mitigate this advantage, we randomize the order of the student data. However, we do so only once, so as to be able to compare the results of repeated runs with tweaked constraints.

Performing Multiple Runs

We always perform multiple runs on the data. Early runs are used to validate the input data; and we typically perform multiple runs on the final data with tweaked constraints, so as to observe their impact.

For example, if we achieve a particularly good assignment (say 95% in their first or second choice), Academic Programs may be interested in seeing the impact of greater gender-balance constraints. If the resulting trade-off is not too bad, they may decide that the new assignment is preferable. At the other extreme, they may want to comprehend the negative effect of a particular constraint; in that case, we remove the constraint to get a base-line comparison.

Data Validation

Our code requires that all entries are populated. During validation, we check that every student has indicated both gender and international status

and has ranked four courses. We perform this validation manually, since our 8-column CSV data format is easy to scan for missing entries. Students who rank fewer than four courses are at an advantage, since a student can be assigned only to a seminar that the student ranked. Data omissions and errors are reported to Academic Programs, which gives instructions for resolution.

Unusual Students

There may be students who must be treated exceptionally. For example, a student may be equivocal about whether they should be regarded as international, or a transgendered student may not feel that they fit into one of the current designated gender categories. There may be a student who must be assigned to a particular seminar, or one who does not have the prerequisites for some of choices.

Missing Student Preferences

Every year, a handful of students do not submit preferences by the deadline. These students are removed from the data set, and the assignment is performed without them. When those students finally contact Academic Programs, they are constrained to choose from the seminars that still have unfilled seats after the matching.

Is the Hat Strategy-Proof?

We know very little about how the students rank their choices. From the student perspective, first-year seminar assignment is a game in which they are competing against all other students. Is this game *strategy-proof*, meaning that the students have no incentive to lie about their true preferences? A student who ranks a very popular course as her fourth choice has virtually no chance of getting that course. Could a student take advantage of this feature to increase the chance of being assigned to the first choice? Or would this strategy have a greater chance of leaving her unassigned and forced to choose from the remaining unfilled seminars? This is an interesting open question, whose answer might depend on the weighting scheme.

Appendix: Availability of the Software

The Mathematical Sorting Hat is written in Matlab. It is also compatible with Octave, an open-source implementation of Matlab. The Hat consists of four `m`-files and requires a library freely available on the MathWorks.com website.

The Mathematical Sorting Hat files and a manual are available at the supplements page for *The UMAP Journal*:

http://www.comap.com/product/periodicals/supplements.html

References

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2005. The New York City high school match. *American Economic Review* 95 (2) (May 2005): 364–367. http://www.duke.edu/~aa88/articles/nycAERPP.pdf.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 1999. House allocation with existing tenants. *Journal of Economic Theory* 88 (2) (October 1999): 233–260. https://www2.bc.edu/~sonmezt/AbdulkadirogluSonmez-JET99.pdf.

Abraham, D., A. Blum, and T. Sandholm. 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *ACM Conference on Electronic Commerce* (EC), San Diego, CA, June 11–15, 2007: 295–304. http://www.cs.cmu.edu/~sandholm/kidneyExchange.EC07.withGrantInfo.pdf.

Bertsimas, D., and J.N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Nashua, NH: Athena Scientific.

Cook, W., W. Cunningham, W. Pulleyblank, and A. Schrijver. 1998. *Combinatorial Optimization*. New York: John Wiley & Sons.

Emspak, J. 2010. First kidney transplants using new matching algorithm. http://www.ibtimes.com/articles/92162/20101214/first-kidney-transplants-using-new-matching-algorithm.htm.

Evans, J., and E. Minieka. 1992. *Optimization Algorithms for Networks and Graphs*. 2nd ed. Boca Raton, FL: CRC Press.

Gale, D. 1981. The optimal assignment problem. UMAP Modules in Undergraduate Mathematics and Its Applications: Module 317. 1982. Reprinted in *UMAP Modules: Tools for Teaching 1981*, 175–212. Boston, MA: Birkhäuser.

Gale, D., and L. Shapley. 1962. College admissions and the stability of marriage. *American Mathematical Monthly* 69 (1) (January): 9–15. http://www.econ.ucsb.edu/~tedb/Courses/Ec100C/galeshapley.pdf.

Hunter, Marty Stuart, and Carrie W. Linder. n.d. College seminars for first-year students—types of first-year seminars, course objectives and content, pedagogy and staffing, instructor development. http://education.stateuniversity.com/pages/1862/College-Seminars-First-Year-Students.html.

Khuller, Samir, Stephen G. Mitchell, and Vijay V. Vazirani. 1991. On-line algorithms for weighted bipartite matching and stable marriages. In *Lecture Notes in Computer Science* 510, edited by Javier Albert, Burkhard Monien, and Mario Artalejo, 728–738. New York: Springer. 1994. Reprinted in *Theoretical Computer Science* 127 (2): 255–267. http://www.cs.umd.edu/users/samir/grant/icalp91.pdf .

National Resident Matching Program. 2011. National resident matching program. http://www.nrmp.org/res_match/index.html .

Roth, A.E., and E. Peranson. 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review* 89 (4) (September 1999): 748–780. http://www.uibk.ac.at/economics/bbl/lit_se/papierews06_07/roth_pearson(1999)_.pdf .

Rowling, J.K. 1998. *Harry Potter and the Sorcerer's Stone*. New York: Scholastic.

West, D.B. 2000. *Introduction to Graph Theory*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Acknowledgments

We are grateful to Tom Halverson for introducing this application to the second author. We thank Ann Minnick and Brenda Piatz, from the Macalester Office of Academic Programs, for their openness in embracing this project and for sharing their historical data with us. Finally, we thank the anonymous referee for suggested improvements to both the exposition and the accompanying code.

About the Authors

Andrew Beveridge earned his B.A. in mathematics from Williams College (1991) and his Ph.D. in mathematics from Yale University (1997). After spending time in both academia and industry, he is currently an Assistant Professor at Macalester College. His main research area is probabilistic combinatorics, where he is particularly interested in the interplay between randomness and strategy. His hobbies include volleyball, running, and playing bass guitar.

Sean Cooke graduated from Macalester College in 2009, with majors in Mathematics and in Computer Science and a minor in Hispanic Studies. He is interested in applied mathematics, computation, and combinatorial optimization. He is currently a chef at Sodexo, and plans to pursue a quantitative M.B.A. He is also a freelance graphic designer and a musician who plays a wide array of musical styles from across the globe. He plays guitar, drum set, Latin and Brazilian percussion, and he is learning saxophone. He released his own album in December 2011.

UMAP

Modules in Undergraduate Mathematics and Its Applications

Published in cooperation with

The Society for Industrial and Applied Mathematics,

The Mathematical Association of America,

The National Council of Teachers of Mathematics,

The American Mathematical Association of Two-Year Colleges,

The Institute for Operations Research and the Management Sciences, and

The American Statistical Association.

COMAP

Module 811

Determining the Size and Shape of the Earth

Richard J. Pulskamp

Geodesy

COMAP, Inc., Suite 3B, 175 Middlesex Tpk., Bedford, MA 01730 (781) 862-7878

INTERMODULAR DESCRIPTION SHEET: UMAP Unit 811

TITLE: Determining the Size and Shape of the Earth

AUTHOR: Richard J. Pulskamp
Dept. of Mathematics and Computer Science
Xavier University
Cincinnati, OH 45207
pulskamp@xavier.edu

MATHEMATICAL FIELD: Trigonometry, calculus

APPLICATION FIELD: Geodesy

TARGET AUDIENCE: Sophomore mathematics majors

ABSTRACT: There is a long history of attempts to measure the size and shape of the Earth. During the 17th century, theoretical work predicted its shape to be that of an oblate spheroid. Later empirical work challenged this result. In the 18th century, to settle the question, measurements of arcs of meridian were made at various locations. This Module shows how, under the assumption that its shape is well-approximated as an ellipsoid of revolution, both the size and shape of the Earth can be deduced from the known lengths of arcs of meridian measured at various latitudes.

PREREQUISITES: Plane trigonometry, first-year calculus, scalar product of vectors, and the method of least squares.

RELATED UNITS: Module 562: Finding the shortest distance on the Earth's surface from here to Timbuktu, by Paul R. Patton, 1985. Reprinted in *UMAP Modules: Tools for Teaching 1985*, 73–89.

The UMAP Journal 33 (2): 119–148.

COMAP, Inc., Suite 3B, 175 Middlesex Tpke., Bedford, MA 01730
(800) 77-COMAP = (800) 772-6627, or (781) 862-7878; http://www.comap.com

Determining the Size and Shape of the Earth

Richard J. Pulskamp

Dept. of Mathematics and Computer Science
Xavier University
Cincinnati, OH 45207
pulskamp@xavier.edu

Table of Contents

Modules and Monographs in Undergraduate Mathematics and its Applications (UMAP) Project

Paul J. Campbell — Editor
Solomon Garfunkel — Executive Director, COMAP

The goal of UMAP is to develop, through a community of users and developers, a system of instructional modules in undergraduate mathematics and its applications, to be used to supplement existing courses and from which complete courses may eventually be built.

The Project was guided by a National Advisory Board of mathematicians, scientists, and educators. UMAP was funded by a grant from the National Science Foundation and now is supported by the Consortium for Mathematics and Its Applications (COMAP), Inc., a nonprofit corporation engaged in research and development in mathematics education.

1. Introduction

An intriguing problem investigated during the 17th and 18th centuries is the determination of the size and shape of the Earth. Through theoretical work, Isaac Newton had deduced its shape to be approximated by an *oblate spheroid*, that is, it bulges at the equator. However, empirical data led Jacques Cassini to claim that the Earth is instead *prolate*, or elongated toward the poles. To settle this question, expeditions to various regions of the world collected data on the length of arcs of meridians. If the shape were oblate, the length of a degree would increase from the equator to the poles; and if prolate, the length would decrease.

We show how the size and shape of the Earth can be deduced through measurement of these arcs and describe how their lengths were determined.

2. The Earth as a Sphere

Let us first idealize the Earth as a sphere. On its surface we identify four special objects: the equator, the North and South poles, and the prime meridian passing through Greenwich, England. The surface being two-dimensional, we require only two numbers to determine a point on it. Customarily, these are the latitude (ϕ) and the longitude (λ), for which we take $-90^\circ < \phi < 90^\circ$ and $-180^\circ < \lambda < 180^\circ$. The equator has latitude 0°, the North Pole has latitude 90°, and the prime meridian has longitude 0°. Latitudes north of the equator and longitudes east of the prime meridian are positive. In particular, for the contiguous United States, we have approximately that latitudes lie between 25° and 50° degrees and longitudes between 125° W and 65° W (that is, -125° and -65°).

On a sphere, a *great circle* is any circle of greatest diameter lying on its surface. Equivalently, it is the intersection of the sphere with any plane passing through the center of the sphere. Examples of great circles are the equator and any meridian. The poles are also examples of *antipodes*, that is, points at the extremity of any diameter of the sphere. Because the center and antipodes are collinear, an infinite number of great circles pass through any pair of antipodes and thus antipodes cannot determine a unique great circle. However, *for any pair of points on a sphere that are not antipodes, there is a unique great circle passing through them*.

3. Length of an Arc of a Great Circle

Let A and B be two points (not antipodes) on a sphere of radius ρ, as in **Figure 1**. In this figure, the two points lie at different latitudes and on different meridians. Because these points and the center are not collinear, they determine a unique plane and consequently must lie on a unique great circle. The arc of this same great circle joining A and B is shown in the figure. Our present task is to compute the length of this arc.

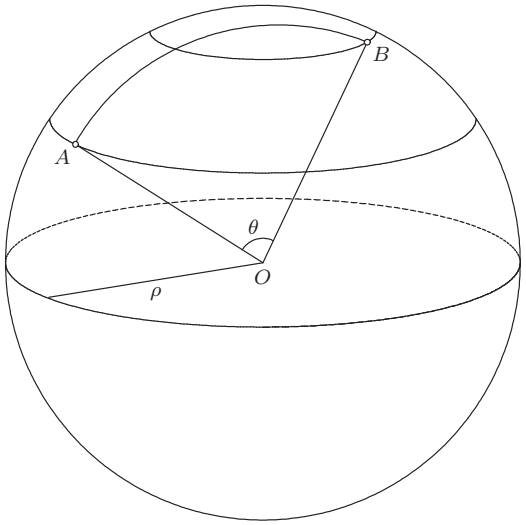


Figure 1. Great circle distance between points.

The radii from the center O to the points A and B form a central angle θ. Since the circumference of a great circle is $2\pi\rho$, it follows that the length of the arc of the great circle between A and B is

$$\frac{\theta}{2\pi} \cdot 2\pi\rho = \theta\rho,$$

where θ is expressed in radians (not degrees).

The problem now reduces to the determination of θ. Consider the radii from the center of the sphere to the points A and B as vectors $\vec{A}$ and $\vec{B}$. The scalar (or dot product) of these vectors can be expressed as

$$\vec{A} \cdot \vec{B} = |\vec{A}||\vec{B}| \cos\theta = \rho^2 \cos\theta, \tag{1}$$

where $|\vec{A}|$ and $|\vec{B}|$ denote the lengths of $\vec{A}$ and $\vec{B}$. So θ can be computed quite easily from the vectors.

If we locate the Earth in a right-handed rectangular coordinate system as shown in **Figure 2**, where the center of the Earth lies at the origin O, the positive z-axis passes through the North Pole, and the positive x-axis passes through the equator and the prime meridian, then the coordinates of the point $P = (\phi, \lambda)$ on the surface of the sphere can be expressed as

$$x = \rho\cos\phi\cos\lambda, \qquad y = \rho\cos\phi\sin\lambda, \qquad z = \rho\sin\phi.$$

To derive these, we note that the vector from the origin terminating at the point $P = (\phi, \lambda)$ on the sphere makes angle ϕ with the plane passing through the equator, and its projection onto that plane makes angle λ with the x-axis. The x- and y-coordinates are consequently the projections of the latter onto the corresponding coordinate axes.

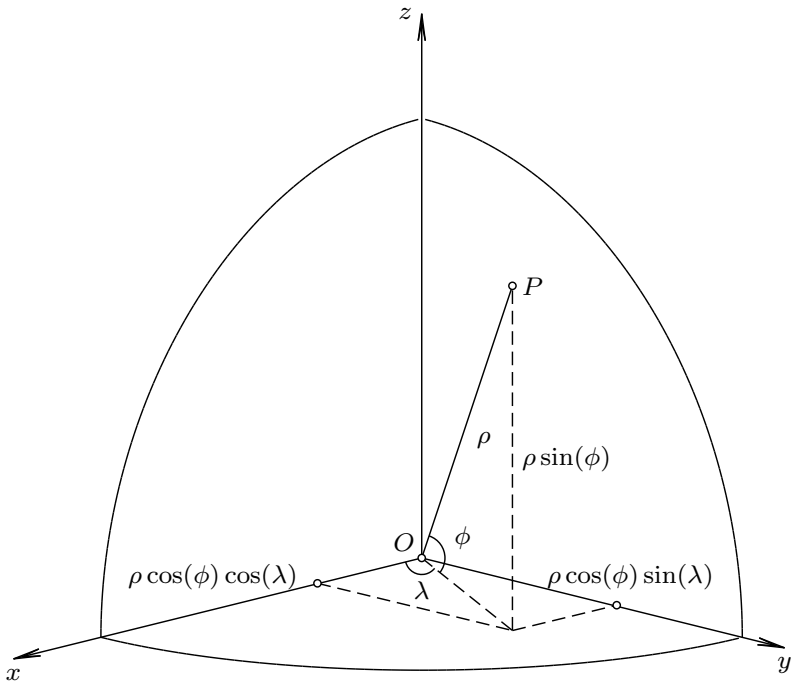


Figure 2. Conversion of latitude and longitude to rectangular coordinates.

Let now the points A and B have spherical coordinates (ϕ_1, λ_1) and (ϕ_2, λ_2). It follows from **(1)** (see **Exercise 1**) that

$$\cos\theta = \cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2) + \sin\phi_1 \sin\phi_2, \qquad \textbf{(2)}$$

and therefore the length S of the great circle distance between A and B is

$$S = \rho \arccos\big(\cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2) + \sin\phi_1 \sin\phi_2\big). \qquad \textbf{(3)}$$

However, contrary to our model, the Earth is not a sphere, but rather, as we shall see, an oblate spheroid—thicker at the equator than at the poles. As a compromise, we take its mean radius to be $\rho = 3{,}959$ miles, or 6,371 km, an approximation that is quite suitable for the temperate zones. See Patton [1985] for a fuller discussion of great-circle distance computations.

Exercises

1. Derive **(2)**.

2. The city of Alexandria in Egypt lies at $31^\circ\ 12'$ N, $29^\circ\ 55'$ E. The city of Aswan, formerly Syene, lies at $24^\circ\ 05'$ N, $32^\circ\ 56'$ E. Convert the coordinates of each to radians and determine the great circle distance separating them, in both miles and in kilometers.

3. The city of Amiens, France is located at $49^\circ\ 54'\ 46''$ N, $2^\circ\ 17'\ 39''$ E, and the location known as Malvoisine was reported to be at $48^\circ\ 31'\ 48''$ N, $2^\circ\ 25'\ 36''$ E. Repeat the previous exercise for these two locations.

4. Early Measurements of the Earth

Aristotle, writing around 350 B.C., states in his *De cælo* [On the Heavens] II.14 [Aristotle 1984] that mathematicians arrived at the figure 400,000 *stades* for the circumference of the Earth. The stade is assumed here to be the Olympic stadium of length 600 Greek feet; in turn, a Greek foot is composed of 16 dactyls (or fingers). Unfortunately, Aristotle gave no indication as to how the figure of 400,000 stades was determined.

Eratosthenes (276–196 B.C.), who headed the library at Alexandria, is reported to have observed that the cities of Syene and Alexandria lie on the same meridian and are separated by 1/50th of the meridian circle. The angle was determined by observing on the summer solstice at noon that on a sundial at Syene no shadow was cast by a pointer aimed toward the zenith (the point directly overhead). However, a pointer on a sundial set at Alexandria at that same time cast a shadow corresponding to an angle of 1/50th of a great circle.

Figure 3 illustrates his reasoning. The circle represents the meridian through Syene (S) and Alexandria (A). Light from the sun is assumed to strike the Earth in parallel. The pointer AZ at Alexandria casts a shadow AB. Because BZ is parallel to OS, by the alternate interior angle theorem, the central angle is $\theta = \angle AOS = \angle BZA$.

Since the distance between the two cities was known to be 5,000 stades, it follows that the circumference of the Earth is 250,000 stades. Eratosthenes later corrected this figure to 252,000 stades. Because $252{,}000/360 = 700$, one is led to suspect that his intent was to make the measure of 1° of arc along the meridian to be 700 stades [Heath 1932].

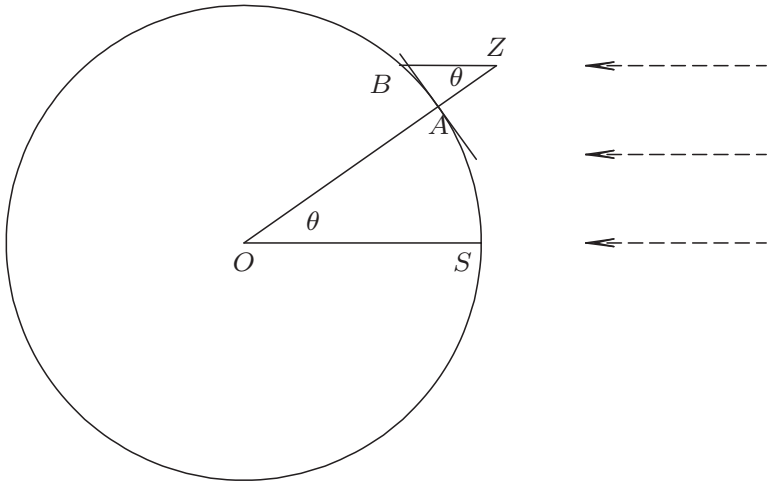


Figure 3. Eratosthenes' method of measurement.

Our knowledge of the method of Eratosthenes comes from Cleomedes (date uncertain) in his *De motu circulari corporum cælestium* [On the circular

motion of heavenly bodies] Book I, Chapter 10 [Cleomedes 2004]. He also reported a figure determined by a certain Posidonius (ca. 135–45 B.C.). Posidonius claimed that the Greek island Rhodes and the city of Alexandria lie on the same meridian and also that they are 5,000 stades apart. Moreover, he observed that the elevation of the star Canopus at these two locations differs by $1/48$th of a great circle. Therefore, the circumference of the Earth is $5,000 \times 48 = 240{,}000$ stades, so that 1° of arc is $666\frac{2}{3}$ stades.

Neither of Posidonius's claims is true. We know that the location of Alexandria is $31^\circ\ 12'$ N, $29^\circ\ 55'$ E, while Rhodes is at $36^\circ\ 10'$ N, $28^\circ\ 0'$ E; so the places are clearly not on the same meridian, nor are they as far apart as Posidonius claimed. In fact, since the two cities are separated by the Mediterranean Sea, it is difficult to conceive how a good measurement of distance could have been achieved.

Claudius Ptolemy in his *Geography* I.7 [Ptolemy 2000] asserts that 1° of arc is only 500 stades.

There is little purpose in converting stades to meters or in attempting to reconcile Posidonius, Eratosthenes, and Ptolemy, since the stade was never standardized in ancient times. What is important is to note that the methodology of Eratosthenes is quite sound. Particularly, *if each latitude and the precise distance between two points falling on the same meridian are known, then the circumference of the Earth can be determined under the assumption that it is a sphere.*

We mention two other measurements of an arc. One was undertaken in 833 at the order of Caliph Al-Mamun; it took place in the Mesopotamian plain of Sinjar in present-day northwestern Iraq [Abulfeda 1848]. Finally, Jean Fernel measured the distance from Paris to Amiens by means of an odometer, publishing the results in *Cosmothéorie* (1528) [Butterfield 1906].

5. Length of an Arc of an Ellipse

The Earth is not a sphere; but a good approximation to its shape is an ellipsoid of revolution, also known as a *spheroid*. A plane through the poles intersects the spheroid in an ellipse; so we investigate how arclength is measured along an ellipse.

Consider an ellipse centered at the origin in the xz-plane for which the length of the major radius is a, the length of the minor radius is b, and $b \le a$. In rectangular coordinates, a formula for this ellipse is

$$\frac{x^2}{a^2} + \frac{z^2}{b^2} = 1,$$

when its major axis is horizontal and its minor axis is vertical.

Substituting $x = a\cos t$ and $z = b\sin t$, a parametric representation is

$$[x, z] = [a\cos t, b\sin t], \quad \text{where } 0 \le t < 2\pi. \qquad \textbf{(4)}$$

Recall the *eccentricity* of an ellipse:

$$e = \sqrt{\frac{a^2 - b^2}{a^2}} = \sqrt{1 - \left(\frac{b}{a}\right)^2}.$$

It depends solely on the ratio b/a; so any two of the parameters a, b, and e suffice to determine an ellipse uniquely, provided that it is centered at the origin with its major axis horizontal and its minor axis vertical.

Figure 4 shows a segment of arc formed by the interval $\alpha \le t \le \beta$.

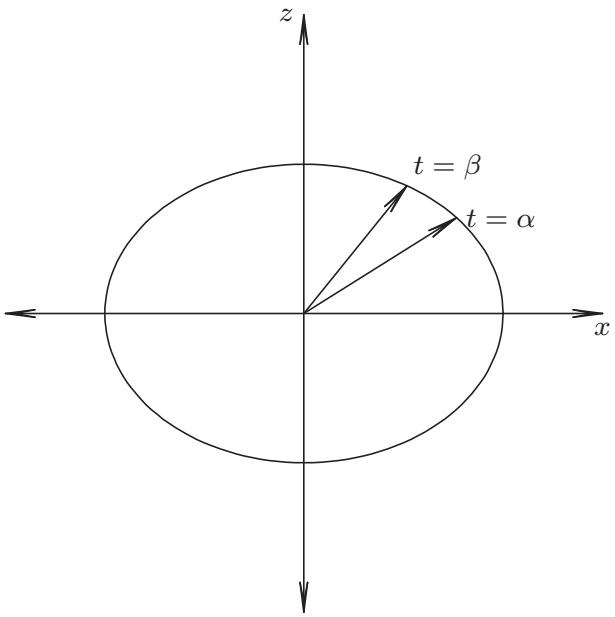


Figure 4. An ellipse.

To compute its length, we proceed as follows: By the Pythagorean theorem, the length of arc is

$$ds = \sqrt{dx^2 + dz^2} = \sqrt{a^2 \sin^2 t + b^2 \cos^2 t}\, dt.$$

Therefore, the arclength S is

$$S = \int_\alpha^\beta ds = \int_\alpha^\beta \sqrt{a^2 \sin^2 t + b^2 \cos^2 t}\, dt. \qquad \textbf{(5)}$$

If $a \ne b$, the value of this integral cannot be expressed in terms of elementary functions. We must make use of elliptic functions or a computer algebra system.

6. Arcs on the Earth

Newton's theory of gravitation predicted the shape of the Earth to be that of an oblate spheroid. Due to the rotation about its axis, centrifugal forces cause the Earth to be flattened at the poles and to bulge at the equator. Consequently, the diameter of the equator of the Earth will be greater than the diameter joining the poles. In Book Three of Newton's *Principia Mathematica*, we find Proposition XIX, Problem III: "To find the proportion of the axis of a planet to the diameter perpendicular thereto" [Newton 1729]. For the Earth, Newton concluded that the ratio of the polar diameter to the equatorial diameter is 229:230. Its *flattening* f, which is nothing more than the difference in radii as a proportion of the equatorial radius, is therefore 1/230 as computed by Newton.

A current refinement of measurements[1] provides an elliptical cross-section of the Earth with an equatorial radius of $a = 6378.1366$ km and a polar radius of $b = 6356.7519$ km. Its *flattening* $f = (a - b)/a$ is officially given as approximately $1/298.25642$. The eccentricity can be derived from f, since $e^2 = f(2 - f)$; so the eccentricity of the Earth as an ellipsoid is $e = 0.0579$.

6.1 Latitude

The geographical latitude ϕ at a point on the Earth is the angle between the normal to the surface (that is, the direction of zenith) and the equator. Let's investigate the relationship between ϕ and the parameter t.

Returning to the parametric representation of the ellipse in **(4)**, we see that

$$\vec{v} = [x(t), z(t)] = [a\cos t, b\sin t].$$

Differentiating with respect to t, we have

$$[x'(t), z'(t)] = [-a\sin t, b\cos t],$$

so that a normal at a point t on the ellipse is

$$\vec{n} = [b\cos t, a\sin t].$$

To find the angle between $\vec{n}$ and the x-axis, we form the dot product **(1)** with the unit vector $\vec{e} = [1, 0]$.

The latitude ϕ is determined by

$$\cos\phi = \frac{\vec{n}\cdot\vec{e}}{|\vec{n}||\vec{e}|} = \frac{b\cos t}{\sqrt{b^2\cos^2 t + a^2\sin^2 t}}. \qquad \textbf{(6)}$$

[1]The dimensions vary slightly depending upon the model used for the Earth. See, for example, Williams [2010].

It is easy to derive a simple relationship between ϕ and t from **(6)** by using customary trigonometric identities (**Exercise 4**):

$$\tan t = \frac{b}{a}\tan\phi = (1-f)\tan\phi. \qquad \textbf{(7)}$$

Now, $1^\circ = \pi/180$ radians. If we set $\delta = \pi/360$, intervals of the form $\phi \pm \delta$ will span roughly 1° on the arc of the ellipse. Consequently, in order to derive the length of 1° of arc on the surface, it is necessary to compute arclength using the corresponding interval for the parameter t derived from **(7)**.

Example. The arclength corresponding to 1° at latitude $\phi = 0$ is

$$\int_\alpha^\beta \sqrt{a^2\sin^2 t + b^2\cos^2 t}\,dt = 110.574,$$

where $a = 6378.1366$ km and $b = 6356.7519$ km. We also have

$$\alpha = \arctan[(1-f)\tan(-\delta)], \qquad \beta = \arctan[(1-f)\tan(\delta)].$$

Table 1 presents the length of 1° of surface arc (centered at the value of ϕ indicated) using the parameters given above. As points of reference, we offer various major cities that lie at approximately the same latitude.

Table 1.
Length of 1 degree of arc at selected angles.

Latitude	Reference Point	Length (km)
90°	North Pole	111.694
60°	Oslo, Stockholm, Anchorage	111.412
45°	Milan, Montreal	111.132
30°	Tallahassee, Houston, Shanghai	110.852
15°	Guatemala City, Dakar	110.649
0°	Quito	110.574

We observe that the length of 1° of arc at the North Pole is 1.120 km greater than 1° at the equator. We also mention in passing that the equatorial circumference is $2\pi a = 40,075$ km and the *meridianal circumference* (the perimeter of the elliptical cross-section) is 40,008 km.

Although the central angle (*geocenter latitude*) θ is near in value to ϕ, the two are equal only at the equator and at the poles. Since the ellipse is parametrized as $[x, z] = [a\cos t, b\sin t]$, it follows that the central angle θ formed by 'the line passing through the origin and a point on the ellipse' and 'the line corresponding to the x-axis' is given by

$$\tan\theta = \frac{b\sin t}{a\cos t} = \frac{b}{a}\tan t = (1-f)\tan t. \qquad \textbf{(8)}$$

Combining **(7)** and **(8)**, we find

$$\tan\theta = (1-f)^2 \tan\phi$$

and hence $\tan\theta/\tan\phi$ is always slightly less than or equal to 1 when $\phi \neq 0$, since $(1-f)^2 = 0.9933$.

Exercises

4. Derive **(7)**.

5. Using a computer algebra system such as Maple or Mathematica, verify the entries in **Table 1** using the values of a and b given in the example in this section.

7. Triangulation

Willebrord Snell (1580–1626) published in *Eratosthenes Batavus de Terræ ambitus vera quantitate* [The Dutch Eratosthenes: On the true measure of the circuit of the Earth] (1617) his results of the mapping of Holland and parts of Flanders. He determined the location of the principal cities, working from Leiden northward to Alkmaar and southward to Bergen-op-Zoom. Snell was the first to introduce the use of *triangulation* in order to avoid making direct measures of lengths.

Geodetic measurements by triangulation are made as follows (refer to **Figure 5**). To compute the length of an arc of meridian between two points A and A', one first establishes a baseline. For simplicity, let us assume that the baseline is the line segment from A to B at the bottom of the figure. The length $\overline{AB}$ is determined by direct measurement.

Next, stations are erected at various positions between A and A', say at C, C', C''. The surveyors, after establishing the stations, measure the angles of the triangles formed by them. These would be $\triangle ABC$, $\triangle BC'C$, $\triangle CC'C''$, and $\triangle C'C''A'$. Given the length $\overline{AB}$ and each angle, the length of each side is determined and hence so is $\overline{AA'}$.

As a check, a second baseline could be established at A' and the process repeated toward A.

After Snell, Richard Norwood (c. 1590–1675) measured the arc from London to York during the years 1633 to 1635 by pacing the distance and noting the difference in latitude between the cities to be $2^\circ\ 28'$; his values were published in *The Seaman's Practice* (1637). The Italian Jesuits Giovanni Riccoli (1598–1671) and Francesco Grimaldi (1613–1663) triangulated the region of Lombardy from Bologna to Modena; their analysis was published in *Geographiæ et hydrographiæ reformatæ* [On a reformed geography and hydrography] (1661). Neither of these results was deemed satisfactory, due to lack of precision in the measurements.

Jean Picard (1620–1682) undertook a precise triangulation of northern France near Paris in 1669. His unit of measure was a *toise*, equal to 6 ft or about 1.949 m. Originally, the meridian under consideration spanned Malvoisine to Sourdon but was subsequently extended to Amiens (see **Exercise 3**). Picard determined that at that latitude $1^\circ = 57,060$ toises. An extract from his report Picard [1671] is in the **Appendix**, but we omit there his determination of latitude.

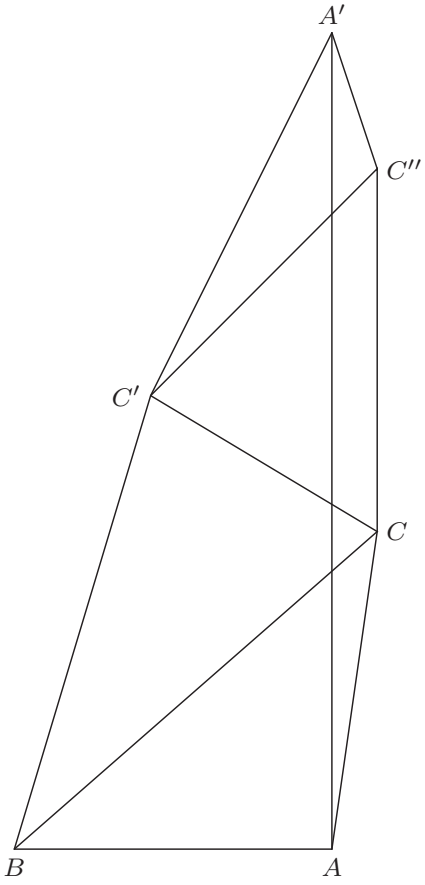


Figure 5. A simple triangulation.

In reality, adjustments are required:

1. A baseline 5 or 6 miles in length was ideally measured by laying a rule end-to-end. The length of such a rule can vary with temperature; so over great distances, the errors can become appreciable. Often, the baseline is interrupted by physical features (e.g., hills, streams, etc.) that prohibit direct measure, so that triangulation may be required here as well.
2. Over large distances, the triangles must be treated as spherical triangles. They require correction, since the sum of the interior angles of a spherical triangle exceeds 180°; the difference is known as the *spherical excess*. Legendre showed that a small spherical triangle can be replaced by a plane triangle with the same dimensions if one-third of the spherical excess is subtracted from each of the angles.[a] As triangulation proceeds, it may be discovered that these angles need further adjustment due to errors in measurement.
3. If the measurements are made at high altitudes or not in the same horizontal plane, distances must be reduced to the equivalent at sea level.

[a]Let A, B, C be the angles of a spherical triangle on a sphere of radius ρ. The spherical excess in degrees is $E = A + B + C - 180^\circ$. The relationship between the area A of the triangle and its spherical excess is given by

$$A = \frac{\pi}{180}\,\rho^2 E.$$

Exercise

6. We consider the circumstances under which the spherical excess may become a concern. Let E denote the spherical excess of a triangle of area A. Using the relation

$$A = \frac{\pi}{180}\,\rho^2 E$$

and taking the radius of the Earth to be $\rho = 3{,}959$ miles, for what area does the spherical excess exceed one second of arc? one minute of arc? What would be the approximate dimensions of an equilateral triangle for which the spherical excess is one second of arc? one minute of arc?

7. Given the following measures related to **Figure 5**, determine the length of $\overline{AA'}$ when $\overline{AB} = 10$.

Triangle	Angle	Degrees	Angle	Degrees	Angle	Degrees
$\triangle ABC$	$\angle ABC$	41.1859	$\angle BCA$	40.6840	$\angle CAB$	98.1301
$\triangle BC'C$	$\angle CBC'$	32.1148	$\angle BC'C$	75.7355	$\angle C'CB$	72.1497
$\triangle CC'C''$	$\angle CC'C''$	75.9638	$\angle C'C''C$	45.0000	$\angle C'CC''$	59.0362
$\triangle C'C''A'$	$\angle C''C'A'$	18.4349	$\angle C'A'C''$	45.0000	$\angle A'C''C'$	116.5651

8. Determining the Shape of the Earth

Suppose that we wish to determine the dimensions and the shape of the Earth under the assumption that it is an ellipsoid of revolution. We derive a formula that gives this information computed from arclengths. One observation does not suffice, since two parameters—say, the equatorial and polar radii or the equatorial radius and eccentricity—must be determined. Even in France, where travel is relatively easy and tools exist, good measurements are difficult to obtain. In addition, the angular size of France is small relative to the size of the Earth. These circumstances speak to the need for a large number of accurate and well-spaced observations. Our data will consist of the length of an arc of meridian taken at various latitudes on the Earth. Here we follow George Airy (1801–1892) [1858].

Let an ellipse be defined parametrically as earlier, that is,

$$[x, z] = [a\cos t, b\sin t], \quad \text{where } 0 \le t < 2\pi.$$

The *curvature* κ at the point $[x, z] = [a\cos t, b\sin t]$ is

$$\kappa = \frac{|x'z'' - z'x''|}{(x'^2 + z'^2)^{3/2}} = \frac{ab}{(a^2\sin^2 t + b^2\cos^2 t)^{3/2}}. \qquad \textbf{(9)}$$

Put $b = a(1 - f)$, f being the flattening. We have then (**Exercise 8**)

$$\kappa = \frac{1 - f}{a[1 + f(f - 2)\cos^2(t)]^{3/2}}. \tag{10}$$

The *radius of curvature* ρ at t is the reciprocal of κ and hence

$$\rho(t) = \frac{a[1 + f(f - 2)\cos^2(t)]^{3/2}}{1 - f}.$$

How can we simplify this expression? We know that the flattening f of the Earth is small. For this reason, terms in the Maclaurin polynomial expansion of $\rho(t)$ in terms of f of order f^2 and higher can be neglected. We then obtain (**Exercise 9**)

$$\rho(t) \approx a[1 + f(1 - 3\cos^2 t)], \tag{11}$$

which is valid as long as f is small.

Since the circumference of a circle of radius r is $2\pi r$, it follows that the length of an arc of circumference subtended by the angle dt is $r\,dt$. By analogy, the length of an arc of an ellipse ds subtended by the angle dt at the angle t is the product of the radius of curvature and the angle:

$$ds = \rho(t)\,dt.$$

In practice, we observe a length $S(\phi_1, \phi_2)$ of an arc of meridian between two latitudes $\phi_1 < \phi_2$. Letting $\Delta\phi = \phi_2 - \phi_1$, we have

$$\begin{aligned} S(\phi_1, \phi_2) &= a \int_{\phi_1}^{\phi_2} 1 + f(1 - 3\cos^2 t)\,dt \\ &= a\left[\Delta\phi - \frac{1}{2} f \Delta\phi - \frac{3}{2} f(\cos\phi_2 \sin\phi_2 - \cos\phi_1 \sin\phi_1)\right]. \end{aligned}$$

Let's put this into a more convenient form. We first observe that

$$\begin{aligned} \cos\phi_2 \sin\phi_2 - \cos\phi_1 \sin\phi_1 &= \frac{\sin 2\phi_2}{2} - \frac{\sin 2\phi_1}{2} \\ &= \frac{1}{2} \cdot 2\sin(\phi_2 - \phi_1) \cdot \cos(\phi_1 + \phi_2) \\ &= \sin\Delta\phi \cdot \cos(2\phi_m), \end{aligned}$$

where $\phi_m = (\phi_1 + \phi_2)/2$.

It follows that

$$\frac{S(\phi_1, \phi_2)}{\Delta\phi} = a - \frac{1}{2} af \left(1 + 3\cos(2\phi_m)\frac{\sin(\Delta\phi)}{\Delta\phi}\right). \tag{12}$$

Exercises

8. Derive **(10)** from **(9)** by making the indicated substitution of $b = a(1 - f)$ and using the identity $\sin^2 t + \cos^2 t = 1$.

9. Show that the Maclaurin polynomial in terms of f of degree 1 for

$$\rho(t) = \frac{a[1 + f(f - 2)\cos^2(t)]^{3/2}}{1 - f}$$

is indeed that given in **(11)**.

9. Further Arc Measurements

Plans were made about 1681 to extend the arc of Picard in both directions. Jean Dominique Cassini (1625–1712) was to supervise the southward triangulation and Phillipe de la Hire (1640–1718) the northward. Soon after its commencement, the work was interrupted due to the death of the French minister in charge.

From 1700 to 1718, under the direction of Jacques Cassini (1677–1756), son of Jean, the line was extended south to Perpignon. The length of 1° of arc south of Paris was found to be 57,097 toises. In the northern direction, completed in 1718, the triangulation proceeded to Dunkirk, where the corresponding length of 1° of arc was here found to be 56,960 toises. These data suggest that the length of 1° decreases toward the pole, that is, the shape of the Earth is that of a prolate spheroid. Consequently, Cassini's measurements, if valid, contradicted Newtonian theory.

To put the matter to rest, an expedition under Charles Marie de la Condamine (1701–1774) embarked for Peru in 1735 to measure the equatorial arc. A second expedition was sent to Lapland in 1736 under Pierre Luis Maupertuis (1698–1759) to measure the northern arc.

The results of these two very arduous expeditions confirmed that the length of 1° of arc lengthens toward the poles. Voltaire [1829, 20] critiqued the meridian measures in the fourth discourse of his *Discours en vers sur l'homme*: "De la modération en tout, dans l'étude, dans l'ambition, dans les plaisirs." [On moderation in all, in study, in ambition, in pleasures.] With regard to Maupertuis, he says:

Vous avez confirmé dans ces lieux pleins d'ennui
Ce que Newton connut sans sortir de chez lui.
Vous avez arpenté quelque faible partie
Des flancs toujours glacés de la terre aplatie.

You have confirmed in these places full of vexation
That which Newton knew without leaving his house.

You have surveyed some feeble part
Of the ever-frozen side of the flattened Earth.

(translation by the author)

10. Analysis

Let us now jump to 1830, when the astronomer George Airy examined 14 measurements of meridians. With the various units of measure converted to the English foot, we have **Table 2**, adapted from his article in the *Encyclopedia Metropolitana* [Airy 1858].

Proceeding from this table, we convert each midlatitude ϕ_m and amplitude $\Delta\phi$ into radians, and the lengths $S(\phi_1, \phi_2)$ into miles. From these new values, we compute the ratio of length to amplitude, which we denote by

$$y = \frac{S(\phi_1, \phi_2)}{\Delta\phi},$$

together with the value of the expression

$$x = -\frac{1}{2}\left[1 + 3\cos(2\phi)\,\frac{\sin(\Delta\phi)}{\Delta\phi}\right].$$

We display these new data in **Table 3**.

If the cross section of the Earth is indeed an ellipse and if the measurements over the various parts of the world were done accurately, then according to **(12)** a scatterplot of the points should display a linear relation. An examination of **Figure 6** shows some discrepancies. In particular, point 12, the measure of the Piedmont arc, is an outlier.

Airy proceeded as follows. He wrote **(12)** in the form

$$S(\phi_1, \phi_2) = b\left[1 + \frac{e}{2} - \frac{3e}{2}\cos(2\phi_m)\,\frac{\sin(\Delta\phi)}{\Delta\phi}\right]\Delta\phi, \qquad \textbf{(13)}$$

where b is the length of the semi-minor axis and $e = (a - b)/b$, where a is the length of the semimajor axis.

Airy began by choosing several pairs of observations: 7 and 11, 7 and 2, 7 and 8, etc., from which he calculated the values of b and e independently. He observed disagreement among them in the value of the eccentricity. Further experimentation of this type led to the conclusion that the discordance was greatest among the values produced in mountainous regions, Airy therefore discarded arcs 1, 2, 4, 5, 8, and 12; these are marked with a ○ in **Figure 6**. The conclusion was that the astronomical measurements required to determine latitude, which depend on finding a precise normal to the Earth's surface, are disturbed by the gravitational attraction of mountains.

Table 2.
14 meridian measurements.

	Arc	Mid-latitude			Amplitude			Length
1	Peruvian arc as calculated by Delambre	-1°	$31'$	$0''$	3°	$7'$	$3.1''$	1,131,057
2	Maupertuis' Swedish arc	66	19	37	0	57	30.4	351,832
3	French arc by Lacaille and Cassini de Thury	46	52	2	8	20	0.3	3,040,605
4	Roman arc by Boscovich	42	59	0	2	9	47	787,919
5	La Caille's arc near the Cape of Good Hope	−33	18	30	1	13	17.5	445,506
6	American arc by Mason and Dixon	39	12	0	1	28	45	538,100
7	French arc from Formentera to Dunkirk	44	51	2	12	22	12.6	4,509,402
8	Svanberg's Swedish arc	66	20	10	1	37	19.3	593,278
9	English arc from Dunnose to Burleigh Moor	52	35	45	3	57	13.1	1,442,953
10	Lambton's 1st Indian arc	12	32	21	1	34	56.4	574,368
11	Lambton's 2nd Indian arc as extended by Everest	16	8	22	15	57	40.2	5,794,599
12	Piedmontese arc by Plana and Carlini	44	57	30	1	7	31.1	414,657
13	Hanoverian arc by Gauss	52	32	17	2	0	57.4	736,426
14	Russian arc by Struve	58	17	37	3	35	5.2	1,309,742

Table 3.
Meridian measures converted.

Arc	Mid-latitude (r)	Amplitude (r)	$S(\phi_1,\phi_2)$ (miles)	$\frac{S(\phi_1,\phi_2)}{\Delta\phi}$	$-\frac{1}{2}\left[1+3\cos(2\phi)\,\frac{\sin(\Delta\phi)}{\Delta\phi}\right]$
1	−0.0265	0.0544	214	3,937	−1.9972
2	1.1576	0.0167	67	3,983	0.5163
3	0.8180	0.1454	576	3,959	−0.4026
4	0.7502	0.0378	149	3,953	−0.6055
5	−0.5813	0.0213	84	3,958	−1.0953
6	0.6842	0.0258	102	3,948	−0.8016
7	0.7828	0.2159	854	3,956	−0.5078
8	1.1578	0.0283	112	3,969	0.5166
9	0.9180	0.0690	273	3,960	−0.1072
10	0.2188	0.0276	109	3,939	−1.8584
11	0.2817	0.2786	1,097	3,940	−1.7518
12	0.7847	0.0196	79	3,999	−0.5022
13	0.9170	0.0352	139	3,964	−0.1099
14	1.0174	0.0626	248	3,965	0.1709

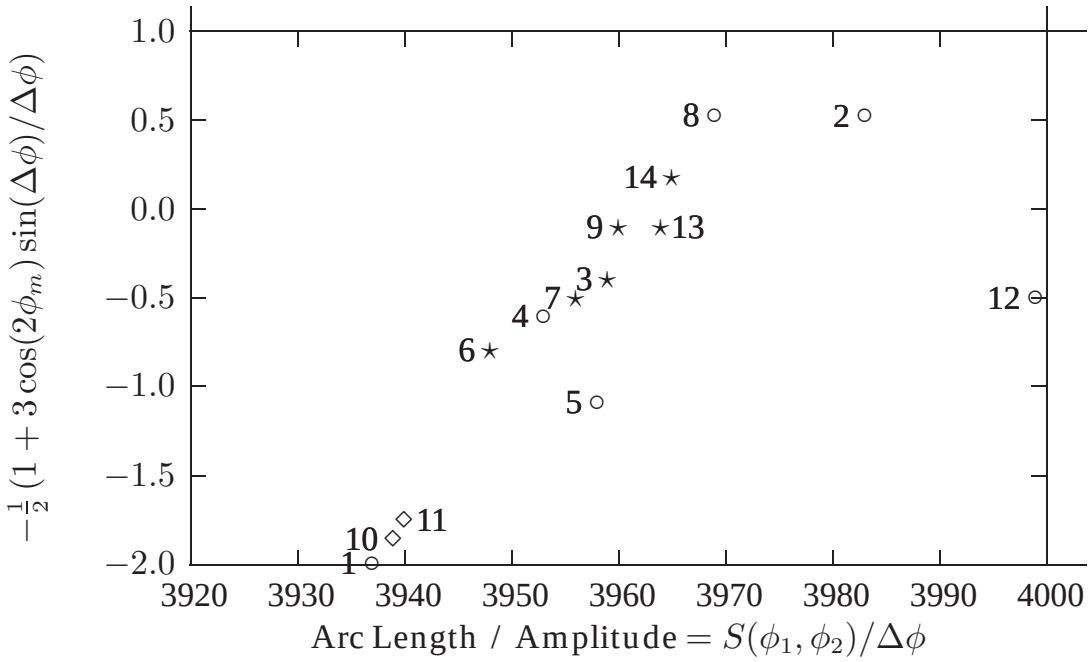


Figure 6. Scatterplot of the 14 meridian arc measures.

Airy also rejected the method of least squares, preferring instead to group the equations judiciously into two sets as shown in **Table 4**. The first set consisted of observations 3, 6, 7, 9, 13, and 14, each marked with a $\star$ in the figure; the second, of observations 10 and 11, each marked with a $\diamond$. An examination of the values in **Table 4** shows why he grouped them as he did. The coefficient of be in **(13)** is positive for the members of the first group and negative for the second. Airy summed the equations in each group, thereby creating two equations in two unknowns:

$$11577228 = 0.553917b + 0.364459be,$$
$$6368967 = 0.306192b - 0.233153be,$$

whose solution he found to be $b = 20{,}854{,}270$ ft and $e = 0.0033808$.

Forming a third set of observations with the more-modern arc measures 7, 9, 13, and 14, and again using the set consisting of 10 and 11, he obtained a second pair of equations and determined new values $b = 20{,}853{,}355$ ft and $e = 0.0033232$.

Averaging the two values yields $b = 20,853,810$ ft and $e = 0.0033520$. Since e is positive, it follows that $a > b$, namely $a = 20{,}923{,}712$ ft, so that the Earth is found to be an oblate spheroid. (In checking Airy's work, we have obtained slightly different solutions.)

The scatterplot suggests that least squares should yield good values of a and b after deleting observation 12. We have by the linear model **(12)** that $a = 3{,}965$ miles and $af = 14.3$, so that $f = 1/278$.

To learn more about the endeavor to determine the length of an arc of

Table 4.
Airy's calculations.

Number	Length (ft)	ϕ_m	$\Delta\phi_m$	$\left[\frac{1}{2} - \frac{3}{2}\cos(2\phi_m)\frac{\sin(\Delta\phi)}{\Delta\phi}\right]\Delta\phi$
3	3,040,605	0.817987339	0.145445559	0.086882519
6	538,100	0.684169067	0.025816329	0.005122389
7	4,509,402	0.782789866	0.215900137	0.106273775
9	1,442,953	0.917970464	0.069004016	0.061604120
13	736,426	0.916962052	0.035184868	0.031317081
14	1,309,742	1.017415447	0.062566175	0.073258834
Total	11,577,228		0.553917084	0.364458717
10	574,368	0.218849744	0.027616927	−0.023706950
11	5,794,599	0.281686445	0.278574911	−0.209446208
Total	6,368,967		0.306191837	−0.233153158

meridian, the interested reader may consult the references, including the recent Murdin [2008], Ferreiro [2011], Hoare [2005], Terrall [2002], Simoson [2009], and Simoson [2011].

Exercises

10. Find the least-squares line (line of best fit, or regression line) with observation 12 excluded and observe the residuals. Which two observations have large residuals, and where are they on the scatterplot? Are these also observations that Airy excluded?

11. Suppose, like Airy, you reject observations 1, 2, 4, 5, 8, and 12. Find the least-squares line using the remaining values. Determine the values of the equatorial radius and the flattening. How do these values compare to the previously determined values?

12. We investigate Airy's two systems of equations used to determine b and e.

a) Solve the system given previously in the text:

$$11577228 = 0.553917b + 0.364459be$$
$$6368967 = 0.306192b - 0.233153be$$

b) Solve the second system consisting of the equation obtained by combining observations 7, 9, 13, and 14 of **Table 4** and the second equation in part **a)** above.

c) Determine the average value of b and the average value of e. Compare these to the corresponding values obtained by Airy.

11. Solutions to the Exercises

1. Taking the dot product, we have

$$\begin{aligned}\vec{A}\cdot\vec{B} &= \rho^2[\cos\phi_1\cos\phi_2\cos\lambda_1\cos\lambda_2 + \cos\phi_1\cos\phi_2\sin\lambda_1\sin\lambda_2 \\ &\qquad + \sin\phi_1\sin\phi_2] \\ &= \rho^2[\cos\phi_1\cos\phi_2(\cos\lambda_1\cos\lambda_2 + \sin\lambda_1\sin\lambda_2) + \sin\phi_1\sin\phi_2] \\ &= \rho^2[\cos\phi_1\cos\phi_2\cos(\lambda_1-\lambda_2) + \sin\phi_1\sin\phi_2].\end{aligned}$$

But $\vec{A}\cdot\vec{B} = \rho^2\cos\theta$. Therefore,

$$\cos\theta = \cos\phi_1\cos\phi_2\cos(\lambda_1-\lambda_2) + \sin\phi_1\sin\phi_2.$$

2. We find that their latitudes and longitudes, expressed in radians, are $(\phi,\lambda) = (0.54454, 0.52214)$ and $(0.42033, 0.57480)$. Moreover, using **(2)**, we find $\cos\theta = 0.991214$, so that the central angle $\theta = 0.13266$ radians. By **(3)**, the great circle distance between the two cities is approximately 525 miles, or 845 km.

3. Their latitudes and longitudes, expressed in radians, are

$$(\phi,\lambda) = (0.87315, 0.041742), \qquad (\phi,\lambda) = (0.849248, 0.044312).$$

Moreover, $\cos\theta = 0.99971$, so that the central angle is $\theta = 0.02396$. Thus, the great circle distance between the two sites is approximately 95 miles, or 153 km.

4. We have from the development in the text

$$\cos\phi = \frac{b\cos t}{\sqrt{b^2\cos^2 t + a^2\sin^2 t}}.$$

Squaring both sides and dividing the numerator and denominator of the right hand side by $\cos^2 t$, we obtain

$$\cos^2\phi = \frac{b^2}{b^2 + a^2\tan^2 t}.$$

Taking reciprocals and making use of the identity $1+\tan^2\phi = \sec^2\phi$, we have

$$\sec^2\phi = 1 + \tan^2\phi = 1 + \frac{a^2}{b^2}\tan^2 t.$$

Therefore, cancelling the unit and extracting the square root, we obtain

$$\tan\phi = \frac{a}{b}\tan t.$$

6. An excess of $E = 1'' = \frac{1}{3600}^\circ$ corresponds to 76 square miles. An equilateral triangle has area $A = \frac{1}{4}a^2\sqrt{3}$, where a is the length of the side. Therefore, a would be roughly 13.25 miles. An excess of $E = 1' = \frac{1}{60}^\circ$ corresponds to 4,559 square miles, or 60 times as much. An equilateral triangle would be roughly 102.6 miles on each side.

8. We now consider our expression for $\rho(t)$ as a function of f. Put

$$g(f) = \frac{a\left[1 + f(f-2)\cos^2(t)\right]^{3/2}}{1-f}.$$

Note first that $g(0) = a$. Then, by the quotient rule, we have

$$g'(f) = \frac{g(f)}{1-f} - 3a\left[1 + f(f-2)\cos^2(t)\right]^{1/2}\cos^2 t.$$

Evaluating the derivative at $f = 0$, we obtain $g'(0) = a(1 - 3\cos^2 t)$. Therefore, the Maclaurin polynomial is

$$g(0) + g'(0)f = a + af(1 - 3\cos^2 t).$$

9. By successive application of the Law of Sines, we find the lengths of the line segments in the figure. All angles must first be converted to radians.

Triangle	Segment	Length	Segment	Length	Segment	Length
$\triangle ABC$	$\overline{AB}$	10.0000	$\overline{BC}$	15.1859	$\overline{AC}$	10.1015
$\triangle BC'C$	$\overline{BC}$	15.1859	$\overline{BC'}$	14.9147	$\overline{CC'}$	8.3299
$\triangle CC'C''$	$\overline{CC'}$	8.3299	$\overline{C'C''}$	10.1015	$\overline{CC''}$	11.4286
$\triangle C'C''A'$	$\overline{C'C''}$	10.1015	$\overline{A'C''}$	4.5175	$\overline{A'C'}$	12.7775

Now introduce the line segment AC', which will form $\triangle ABC'$ and $\triangle AC'A'$. By the Law of Cosines, $\overline{AC'} = 15.3862$. By the Law of Sines applied to $\triangle ABC'$, we have $\angle BC'A = 38.5007^\circ$. From that angle and from $\angle BC'A' = 170.1342^\circ$, we deduce that $\angle A'C'A = 131.63^\circ$. Finally, applying the Law of Cosines to $\triangle AC'A'$, we have $\overline{AA'} = 25.7142$.

10. In this exercise and the next, we apply the least-squares routine of Excel to the data from **Table 3**. Answers may vary slightly due to rounding. Omitting Arc 12 (Piedmont-Plana), we find $a = 3965.06$ and $af = 14.28$, so that $f = 0.03602$. Consequently, the flattening is $f = 1/278$.

Arcs 2 (Lapland-Maupertuis) and 5 (Cape of Good Hope) have large residuals, 10.6 and 8.6 respectively, as can be seen in **Table S1**. This might be expected, since in the scatterplot their positions are somewhat distant from the main trend. Airy did exclude them.

Table S1.
Results from line of best fit through data for all 14 arcs (**Exercise 10**).

Observation	Predicted Y	Residual
1	3936.53	0.47
2	3972.43	10.57
3	3959.31	−0.31
4	3956.41	−3.41
5	3949.41	8.59
6	3953.61	−5.61
7	3957.80	−1.80
8	3972.43	−3.43
9	3963.53	−3.53
10	3938.51	0.49
11	3940.04	−0.04
13	3963.49	0.51
14	3967.50	−2.50

11. Omitting the arcs as indicated, we find $a = 3963$ and $af = 13.3$, so that $f = 0.003355$. Consequently, the flattening is $f = 1/298$. Both a and f are remarkably close to the commonly accepted values.

12. Answers may vary. Since Airy reports 8 digits for b and 7 digits to the right of the decimal point for e, we do likewise. Note too that Airy's final value of b should be 20,853,812. Mathematica values courtesy of Paul Campbell.

a)

	b	e
Airy	20,854,**270**	0.0033**808**
Mathematica	20,854,**263**	0.0033**808**
Maple	20,854,**259**	0.0033**812**

b) The system of equations to be solved is

$$7998523 = 0.382655196\, b + 0.272453810\, be$$
$$6368967 = 0.306192\quad b - 0.233153\quad be$$

	b	e
Airy	20,853,3**55**	0.003323**2**
Mathematica	20,853,3**47**	0.003323**3**
Maple	20,853,3**42**	0.003323**6**

c)

	b	e
Airy	20,853,8**12**	0.003352**0**
Mathematica	20,853,8**05**	0.003352**1**
Maple	20,853,8**01**	0.003352**4**

12. Appendix

We give a small edited extract from Picard's report on his triangulation of the line from Malvoisine to Amiens. **Figure A1** reproduces the placement of the baseline and stations and identifies the stations. The unit of measure is a toise, equal to 6 feet or about 1.949 m. The baseline was a road from Villejuif to Juvisy, denoted by the line segment $\overline{AB}$, which measured 5,663 toises. Originally, the meridian under consideration spanned Malvoisine (E) to Sourdon (N) but was subsequently extended to Amiens (V). This meridian is denoted by $\overline{\alpha\beta}$ in the figure. Bear in mind that once the length of the baseline has been determined, the remainder of the process generally involves solving triangles, given the three angles and the length of one side. It is an interesting exercise to confirm the calculations of Picard.

The points O, P, Q, S, and Z in the figure were used to create a second set of triangles to verify or correct earlier measurements. The points α, β, γ, θ, and ϵ were sites used to provide final adjustments to the meridian.

Picard proceeded in several stages. His goal was to determine the length of $\overline{EN}$ as the sum of $\overline{EG}$, $\overline{GI}$, and $\overline{IN}$.

Triangles I to VI were utilized to cover the distance $\overline{GE}$ from Malvoisine to Mareüil. Consider triangle I, that is $\triangle$ABC. The side $\overline{AB}$ is known to be precisely 5,663 toises by direct measurement. Using the angular measures in **Table A1** for this triangle and applying the Law of Sines, we find

$$\overline{AC} = \sin(\angle ABC) \times \frac{\overline{AB}}{\sin(\angle ACB)} = 11{,}012.9 \text{ toises.}$$

Picard expresses this as 11,012 toises and 5 ft.

Moving next to triangle II, $\triangle$ADC in the same figure, we can now use the computed side to find the length of side $\overline{CD}$ by another application of the Law of Sines. In this manner, Picard found successively $\overline{DE}$, $\overline{DF}$, and $\overline{DG}$.

Now in $\triangle$DEG, the Law of Cosines can be applied to determine $\overline{GE}$. Indeed,

$$\begin{aligned}\overline{GE}^2 &= \overline{DG}^2 + \overline{DE}^2 - 2\overline{DG} \times \overline{DE}\cos(\angle GDE)\\ &= \left(25, 643\tfrac{2}{6}\right)^2 + \left(8, 870\tfrac{2}{6}\right)^2 - 2\left(25, 643\tfrac{2}{6}\right)\left(8, 870\tfrac{3}{6}\right)\cos(128^\circ\ 9'\ 30'').\end{aligned}$$

Thus, $\overline{GE} = 31{,}897$ toises.

Picard engaged in another triangulation utilizing the new station points O and P to produce a second determination of $\overline{GE}$. This he found to be 31,893 toises. Splitting the difference in the two computed values, he chose $\overline{GE} = 31{,}895$ toises.

The next phase evolved in a similar manner to determine the distance from Mareüil to Sourdon. Here triangles VII to XIII were utilized. However

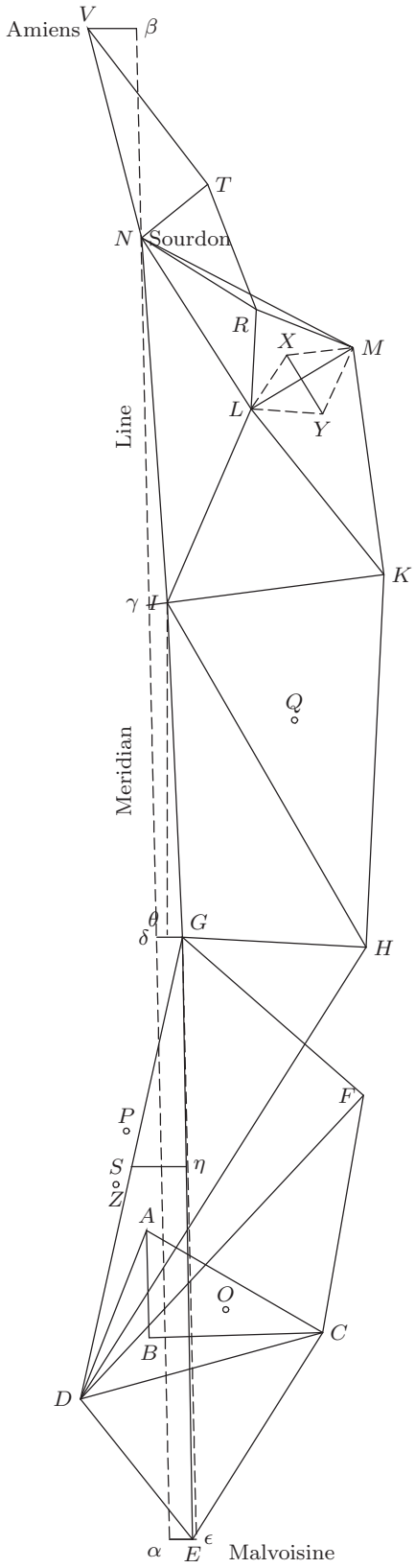


Figure A1. Picard's triangulation

A "is the middle of the Mill of Villejuif.

B "the nearest corner of the Pavilion of Juvisy[-sur-Orge].

C "the point of the Steeple of Brie-Comte-Robert.

D "the middle of the Tower of Montlhéry.

E "the height of the Pavilion of Malvoisine.

F "a Piece of wood stood intentionally at the height of the ruins of the Tower of Monjai, & enlarged with straw.

G "the middle of the Mound of Mareüil, where we have been obliged to make some fires in order to mark it.

H "the middle of the large Pavilion in oval of the Château of Dammartin.

I "the Steeple of Saint Samson of Clermont.

K "the Mill of Jonquieres near Compiegne.

L "the Steeple of Coivrel.

M "a small tree on the Mountain of Boulogne near Montdidier.

N "the Steeple of Sourdon.

O "a small forked tree on the Hill of the Griffon near Villeneuve Saint Georges.

P "the Steeple of Montmartre.

Q "the Steeple of Saint Christophle near Senlis.

R "the Steeple of S. Pierre de Montdidier.

S "a look-out Turret above the Level of the Southern Tower of Notre-Dame of Paris.

T "a tree on the Mountain of Moreüil.

V "the Steeple of Notre-Dame of Amiens.

X "the Mill at Mery

y "a point close to the Vally of S. Martin near Montdidier

Z "is the middle of the Southern face of the Building of the Observatory."

Table A1.

Picard's 13 triangles from Malvoisine to Sourdon.

Triangle	Angle	Degree	Minute	Second	Side	Toises	Feet
I	$\angle CAB$	54	4	35	AB	5,663	0
	$\angle ABC$	95	6	55	AC	11,012	5
	$\angle ACB$	30	48	30	BC	8,954	0
II	$\angle DAC$	77	25	50	AC	11,012	5
	$\angle ADC$	55	0	10	CD	13,121	3
	$\angle ACD$	47	34	0	AD	9,922	2
III	$\angle DEC$	74	9	30	CD	13,121	3
	$\angle DCE$	40	34	0	DE	8,870	3
	$\angle CDE$	65	16	30	CE	12,389	3
IV	$\angle DCF$	113	47	40	DC	13,121	3
	$\angle DFC$	33	40	0	DF	21,658	0
	$\angle FDC$	32	32	20			
V	$\angle DFG$	92	5	20	DF	21,658	0
	$\angle DGF$	57	34	0	DG	25,643	0
	$\angle GDF$	30	20	40	FG	12,963	3
VI	$\angle GDE$	128	9	30	DG	25,643	2
	$\angle DEG$	39	12	30	DE	8,870	3
	$\angle DGE$	12	38	0	GE	31,897	0
VII	$\angle FGH$	39	51	0	FG	12,963	3
	$\angle FHG$	91	46	30	GH	9,695	0
	$\angle HFG$	48	22	30			
VIII	$\angle GHI$	55	58	0	GH	9,695	0
	$\angle GIH$	27	14	0	GI	17,557	0
	$\angle IGH$	96	48	0	HI	21,037	0
IX	$\angle HIK$	65	46	0	HI	21,043	0
	$\angle HKI$	80	59	40	IK	11,678	0
	$\angle KHI$	33	14	20			
X	$\angle LIK$	58	31	50	IK	11,683	0
	$\angle IKL$	58	31	0	KL	11,188	2
	$\angle ILK$	62	57	10	IL	11,186	4
XI	$\angle LKM$	28	52	30	KL	11,188	2
	$\angle KML$	63	31	0	LM	6,036	2
XII	$\angle LMN$	60	38	0	LM	6,036	2
	$\angle MNL$	29	28	20	LN	10,691	0
XIII	$\angle ILN$	119	32	40	LN	10,691	0
					IL	11,186	4
					IN	18,905	0

Picard found it necessary to recompute $\overline{GI}$ as 17,562 toises using $\triangle$QFG. This permitted an adjustment of the value of $\overline{HI}$ to 21,043 toises. In a similar manner, he used $\triangle QIK$ to recompute $\overline{IK}$ as 11,683 toises.

Finally, the computations needed to be verified. He set a new baseline between the points X and Y with length $\overline{XY} = 3{,}902$ toises.

Extending the triangulation to Amiens is straightforward. As can be seen in **Table A2**, work gives the length of $\overline{NV} = 11,161$ toises and 4 ft.

Another set of measurements is required to identify the latitude and longitude of the points. In **Figure A1**, the dashed line $\alpha\beta$ marks the meridian. Each of the computed lengths $\overline{NV}$, $\overline{IN}$, $\overline{GI}$, and $\overline{EG}$ must be referred to this line in order to obtain its length.

Table A2.

Triangulation to Amiens.

Triangle	Angle	Degree	Minute	Second	Side	Toises	Feet
LMR	$\angle LMR$	58	21	50	LM	6,037	0
	$\angle MRL$	68	52	30	LR	5,510	3
LNR	$\angle NRL$	115	1	30	LR	5,510	3
	$\angle RNL$	27	50	30	NR	7,122	2
	$\angle NLR$	37	8	0			
NRT	$\angle NTR$	72	25	40	NR	7,122	2
	$\angle TNR$	67	21	40	NT	4,822	4
	$\angle NRT$	40	12	40			
NTV	$\angle NTV$	83	58	40	NT	4,822	4
	$\angle TNV$	70	34	30	NV	11,161	4
	$\angle NVT$	25	26	50			
$VN\beta$	$\angle VN\beta$	18	55		NV	11,161	4
					$N\beta$	10,559	3
					$V\beta$	3,617	4

Actually, Picard constructed parallels to the meridian. These are shown in **Table A3** as $\overline{N\gamma}$, $\overline{I\theta}$, and $\overline{G\epsilon}$, spanning Malvoisine to Sourdon. The addition of $\overline{N\beta}$ extends the arc of the meridian to Amiens.

Referring to **Table A4**, we see that the length of the meridian from Malvoisine to Amiens is 78,907 toises.

Picard gave the latitude of Notre Dame of Amiens as $49^\circ\ 54'\ 46''$ and that of Malvoisine $48^\circ\ 31'\ 48''$. The difference is $1^\circ\ 22'\ 55''$. However, the astronomical observations were not made at the same terminal points of the meridian, but rather 75 toises south of the Amiens terminus and 18 toises south of Malvoisine. Consequently, he made adjustments to the distance between Malvoisine and Amiens to compensate for the different locations. That is, the distance from Malvoisine to Amiens is adjusted to $78{,}907 - 75 + 18 = 78{,}850$ toises. Dividing this length by the difference in latitudes gives $1^\circ = 57{,}057$ toises, which Picard rounded to 57,060 toises.

Table A3.
Adjustment of lengths to the meridian.

Triangle	Angle	Degree	Minute	Second	Side	Toises	Feet
$IN\gamma$	$\angle\gamma NI$	2	9	10	NI	18,907	0
					$N\gamma$	18,893	3
					$I\gamma$	710	0
$GI\theta$	$\angle GI\theta$	1	9	0	IG	17,564	0
					$I\theta$ (or $\gamma\delta$)	17,560	3
					$G\theta$	352	0
$GE\epsilon$	$\angle E\gamma\epsilon$	0	26	0	GE	31,895	0
					$G\epsilon$ (or $\delta\alpha$)	31,894	0
					$E\epsilon$	241	3

Table A4.
Length of the meridian from Malvoisine to Amiens.

Interval	Segment	Toises	Feet
Malvoisine to Mareüil	$G\epsilon$	31,894	0
Mareüil to Clermont	$I\theta$	17,560	3
Clermont to Sourdon	$N\gamma$	18,893	3
Sourdon to Amiens	$N\beta$	10,559	3
Malvoisine to Amiens	$\beta\epsilon$	78,907	3

Thus, the length of 1° of arc at the latitude of Paris has been determined.

References

Abulfeda. 1848. *Géographie d'Aboulféda*. Vol. II. Paris: Imprimerie Nationale.

Airy, George Biddell. 1858. *Mathematical Tracts*.... 4th ed. Cambridge, England: MacMillan and Co.

Aristotle. 1984. *The Complete Works of Aristotle*. Vol. 1. Princeton, NJ: Princeton University Press.

Butterfield, Arthur D. 1906. *A History of the Determination of the Figure of the Earth from Arc Measurements*. Worcester, MA: Davis Press.

Cleomedes. 2004. *Cleomedes' Lectures on Astronomy: A Translation of the Heavens*. Transl. by Robert B. Todd and Alan C. Bowen. Berkeley: University of California Press.

Ferreiro, Larrie D. 2011. *Measure of the Earth: The Enlightenment Expedition That Reshaped Our World*. New York: Basic Books.

Heath, Thomas L. 1932. *Greek Astronomy*. London: J.M. Dent & Sons.

Hoare, Michael Rand. 2005. *The Quest for the True Figure of the Earth: Ideas and Expeditions in Four Centuries of Geodesy (Science, Technology and Culture, 1700–1945)*. Burlington, VT: Ashgate Pub. Co.

Murdin, Paul. 2008. *Full Meridian of Glory: Perilous Adventures in the Competition to Measure the Earth.* New York: Springer.

Newton, Isaac. 1729. *The Mathematical Principles of Natural Philosophy*. Vol. II. London.

Patton, Paul R. 1985. Module 562: Finding the shortest distance on the Earth's surface from here to Timbuktu. 1986. Reprinted in *UMAP Modules: Tools for Teaching 1985*, 73–89. Arlington, MA: COMAP, Inc.

Picard, Jean. 1671. Mesure de la terre. `http://gallica.bnf.fr/ark:/12148/btv1b7300361b/f1.image`. 1729. Reprinted. Mesure de la terre par Monsieur l'Abbé Picard. *Mémoires de l'Académie Royale des Sciences depuis 1666 jusqu'à 1699* VII Part I: 133–190.

Ptolemy, Claudius. 2000. *Ptolemy's* Geography*: An Annotated Translation of the Theoretical Chapters.* Transl. J. Lennart Berggren and Alexander Jones. Princeton, NJ: Princeton University Press.

Simoson, Andrew J. 2009. *Voltaire's Riddle: Micromégas and the Measure of All Things.* Washington, DC: Mathematical Association of America.

_____. 2011. Newton's radii, Maupertuis' arc length, and Voltaire's giant. *College Mathematics Journal* 42: 183–190.

Terrall, Mary. 2002. *The Man Who Flattened the Earth: Maupertuis and the Sciences in the Enlightenment.* Chicago, IL: University of Chicago Press.

Voltaire, François-Marie Arouet. 1829. *Œvres complètes de Voltaire.* Vol. 11. *Poésies.* Edited by Armand-Aubrée. Paris: Imprimerie de Paul Renouàrd.

Williams, David R. 2010. Earth fact sheet. `http://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html`.

Acknowledgments

The author wishes to thank the editor, two anonymous referees, and his colleague Danny Otero for many suggestions that improved this Module.

About the Author

Richard Pulskamp is Professor of Mathematics and Computer Science at Xavier University in Cincinnati, Ohio. He received his Ph.D. in mathematical statistics from the University of Cincinnati. After spending the last 10 years in administration, he has recently returned to his academic department. His research interests currently lie in the history of probability and statistics.

UMAP

Modules in Undergraduate Mathematics and Its Applications

Published in cooperation with

The Society for Industrial and Applied Mathematics,

The Mathematical Association of America,

The National Council of Teachers of Mathematics,

The American Mathematical Association of Two-Year Colleges,

The Institute for Operations Research and the Management Sciences, and

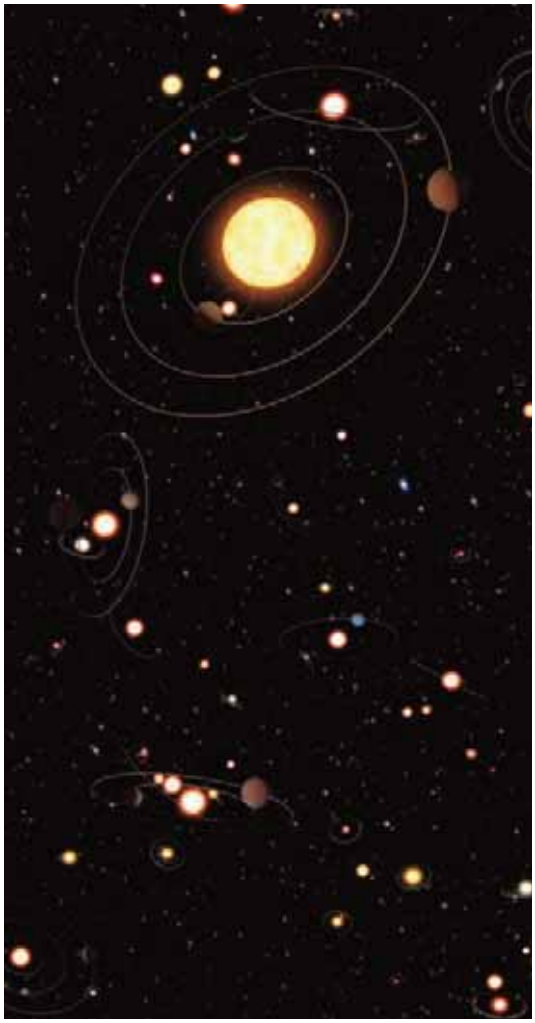
The American Statistical Association.

COMAP

Module 809

Spirograph Patterns of Star Clusters

Lisa Holden



NASA.gov Illustration

Astronomy

COMAP, Inc., Suite 3B, 175 Middlesex Tpk., Bedford, MA 01730 (781) 862-7878

INTERMODULAR DESCRIPTION SHEET: UMAP Unit 809

TITLE: Spirograph Patterns of Star Clusters

AUTHOR: Lisa Holden
Dept. of Mathematics
Northern Kentucky University
Highland Heights, KY 41099
holdenl@nku.edu

MATHEMATICAL FIELD: Differential equations

APPLICATION FIELD: Astronomy

TARGET AUDIENCE: Students in differential equations who have had multivariable calculus.

ABSTRACT: We develop and analyze a model describing the spirographic motion of a young star within its birthplace, an embedded cluster. We begin by first introducing the mathematical formalism of the classical two-body problem. We then extend this formalism for the case of a star moving through a spherically symmetric extended mass distribution. Finally, we undertake an analytical treatment of the resulting spirographic orbits and provide the necessary tools to explore the beauty of these spirographic orbits numerically.

PREREQUISITES: Courses in multivariable calculus and physics (introductory-level mechanics). Specifically: knowledge of vector dot and cross products, partial differentiation, volume integrals, vector-valued functions, line integrals, surface integrals, and the divergence theorem, as well as Newton's Second Law of Motion and Newton's Universal Law of Gravitation. Mathematica is used to explore numerical solutions of the model equations.

The UMAP Journal 33 (2): 149–184.

COMAP, Inc., Suite 3B, 175 Middlesex Tpke., Bedford, MA 01730
(800) 77-COMAP = (800) 772-6627, or (781) 862-7878; http://www.comap.com

Spirograph Patterns in Star Clusters

Lisa Holden

Dept. of Mathematics
Northern Kentucky University
Highland Heights, KY 41099
holdenl@nku.edu

Table of Contents

Modules and Monographs in Undergraduate Mathematics and its Applications (UMAP) Project

Paul J. Campbell — Editor
Solomon Garfunkel — Executive Director, COMAP

The goal of UMAP is to develop, through a community of users and developers, a system of instructional modules in undergraduate mathematics and its applications, to be used to supplement existing courses and from which complete courses may eventually be built.

The Project was guided by a National Advisory Board of mathematicians, scientists, and educators. UMAP was funded by a grant from the National Science Foundation and now is supported by the Consortium for Mathematics and Its Applications (COMAP), Inc., a nonprofit corporation engaged in research and development in mathematics education.

1. Introduction

Celestial mechanics, the study of the motion of celestial objects, has received the attention of some of the greatest mathematical minds in history, including Leibniz, Laplace, Gauss, Euler, and of course, Newton. Most introductory calculus books discuss how Newton's laws, when applied to the motion of a planet orbiting its sun, provide a system of differential equations whose solutions yield Kepler's famous elliptical orbits. That those same equations can also yield chaotic solutions when a third body is added is also generally well known by the mathematics community.

What is not as well known, however, is that those same equations also yield beautiful spirographic orbits when applied to the motion of a celestial body moving through an extended mass distribution. We explore the mathematical foundation of this class of orbits, focusing explicitly on stars moving through an embedded cluster—a group of many nascent stars swirling through the gas and dust from which they formed.

Most of the stars within our galaxy are born within cloud-like gaseous structures comprised mostly of molecular hydrogen (90%) and helium (10%) plus a trace amount of heavier gases and dust. Known as *giant molecular clouds*, these structures are $\sim$40 parsecs (pc) in diameter (where a parsec is about 200,000 times the distance between the Earth and the Sun) and typically contain a total mass of $\sim 10^5$ solar masses. While these structures have an average density of $\sim$100 gas molecules per cubic cm, they are highly nonuniform and are comprised of several distinct "clumps." In turn, each clump contains between 30 and 2,000 small ($\sim$0.1 pc) "cores," which are dense (10^4–10^5 molecules / cc). In time, these cores collapse, due to their self-gravitational attraction, and the gas that accumulates at the center eventually becomes a young star. As a result, each clump within a giant molecular cloud becomes a sort of stellar nursery (or, more formally, an *embedded cluster*) where nascent stars swirl around within the leftover gas and dust from which they formed.

Because cores have an initial rotation, the stars that form from their collapse are orbited by a thick disk of gas and dust from which planets eventually form. A central question for astronomers is what effect the embedded cluster environment has on these disks, and as such, on the formation of planetary systems. For example, a single massive star within a cluster could produce enough ultraviolet light to evaporate a significant fraction of these disks, thereby reducing the planet-building potential within the cluster. This effect depends, in part, on how the stellar members orbit within the cluster environment [Adams et al. 2006; Fatuzzo and Adams 2008; Holden et al. 2011].

Our goal is to develop and analyze a model describing the motion of a young star within its birthplace of an embedded cluster. We begin by first introducing the mathematical formalism of the classical two-body problem. We then extend this formalism for the case of a star moving through a

spherically symmetric extended mass distribution. Finally, we undertake an analytical treatment of the resulting spirographic orbits, and provide the necessary tools to explore the beauty of these spirographic orbits using Mathematica.

2. The Standard Two-Body Problem

We consider the motion of a planet orbiting about a fixed star of mass M, where the planet's mass m is negligible compared to that of its stellar companion (i.e., $m \ll M$). We note, however, that our small-mass approximation can be omitted by adopting a center-of-mass coordinate system, wherein a reduced mass $\mu = mM/(M+m)$ orbits a fixed body with total mass $M+m$ [Carroll and Ostlie 2006].

This two-body treatment can be used to study the dynamics of our solar system, since our Sun's gravitational pull on each planet overwhelms the gravitational interactions between planetary members. This approach thus represents the simplest model for how planets orbit our Sun.

Of course, interactions, albeit small, do have an effect—a point of fact that led to the discovery of Neptune. Interestingly, this effect can be modeled using the spirographic formalism developed in **Section 3**.

2.1 Standard Formalism

An object's motion is defined through the vector quantities of position $\vec{r}$, velocity $\vec{v}$, and acceleration $\vec{a}$, where the acceleration is the rate at which velocity changes with time, and velocity is the rate at which position changes with time, so that

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d^2\vec{r}}{dt^2} \; . \tag{1}$$

Newton's Second Law of Motion connects the acceleration $\vec{a}$ of an object to the sum $\vec{F}$ of the forces acting on it, as expressed by the well-known relation $\vec{F} = m\vec{a}$. For our two-body system, the only force acting on the planet results from its gravitational interaction with the star, as given by Newton's Universal Law of Gravitation:

$$\vec{F} = \frac{-GMm\vec{r}}{r^3} = \frac{-GMm}{r^2}\,\hat{r} \; , \tag{2}$$

where G is a universal constant whose value depends on the system of units, $\vec{r}$ is the displacement vector between the star and planet, and $\hat{r} = \vec{r}/r$ is the unit displacement vector, where $r = |\vec{r}|$. The planet's acceleration is therefore

$$\vec{a} = \frac{\vec{F}}{m} = \frac{-GM}{r^2}\,\hat{r}\,. \tag{3}$$

Equation **(3)** tells us that the acceleration vector is parallel to the position vector, hence the cross product $\vec{r} \times \vec{a} = \vec{0}$. As we explore in **Exercise 1**, this condition allows us to show that

$$\frac{d}{dt}(\vec{r} \times \vec{v}) = \vec{0}\,, \tag{4}$$

and conclude that the vector quantity $\vec{j} = \vec{r} \times \vec{v}$ (the *specific angular momentum*) must be constant for all time. This result implies that the position vector $\vec{r}$ must always remain orthogonal to $\vec{j}$, and as such, the planet's motion must be planar.

To keep the analysis as tractable as possible, we now make the following assumptions:

- The star has the same mass as our Sun (1 $M_\odot$) and is fixed at the origin of our coordinate system.
- The planet moves in the xy-plane, so that $\vec{r} = \langle x, y \rangle$ is its position vector, $\vec{v} = \left\langle \frac{dx}{dt}, \frac{dy}{dt} \right\rangle$ is its velocity vector, and $\vec{a} = \left\langle \frac{d^2x}{dt^2}, \frac{d^2y}{dt^2} \right\rangle$ is its acceleration vector. We note that $r = \sqrt{x^2 + y^2}$ is then the distance between the star and the planet, and $\vec{j} = j\hat{z}$.
- Distance is in units of AUs (where 1 AU, or astronomical unit, is the approximate distance between the Earth and the Sun), time is in units of years, and mass is in units of solar masses ($M_\odot$). In this unit system,

$$G = 4\pi^2 \frac{\text{AU}^3}{M_\odot \text{yr}^2}.$$

- The planet's initial conditions take the form:

$$x(0) = x_0 \qquad x'(0) = 0 \qquad y(0) = 0 \qquad y'(0) = v_0\,, \tag{5}$$

  meaning that we start the planet's motion at the point $(x_0, 0)$ with initial velocity v_0 in the vertical ($\hat{y}$) direction. These initial conditions correspond to the planet starting either at what's known as *perihelion* (position closest to the star) or *aphelion* (position farthest from the star). The orbit of the planet, as given by the solutions to **(3)**, is completely specified by the initial conditions.

The basic elements of our star-planet system are illustrated in **Figure 1**, where the fixed star is represented by the large yellow disk, and the planet (as shown in its initial location and a subsequent location) is represented by the smaller blue disks. The locations of perihelion (where the planet is

closest to the star) and aphelion (where the planet is farthest to the star) are indicated by the arrows. As discussed further below, these locations are turning points in the planet's orbit.

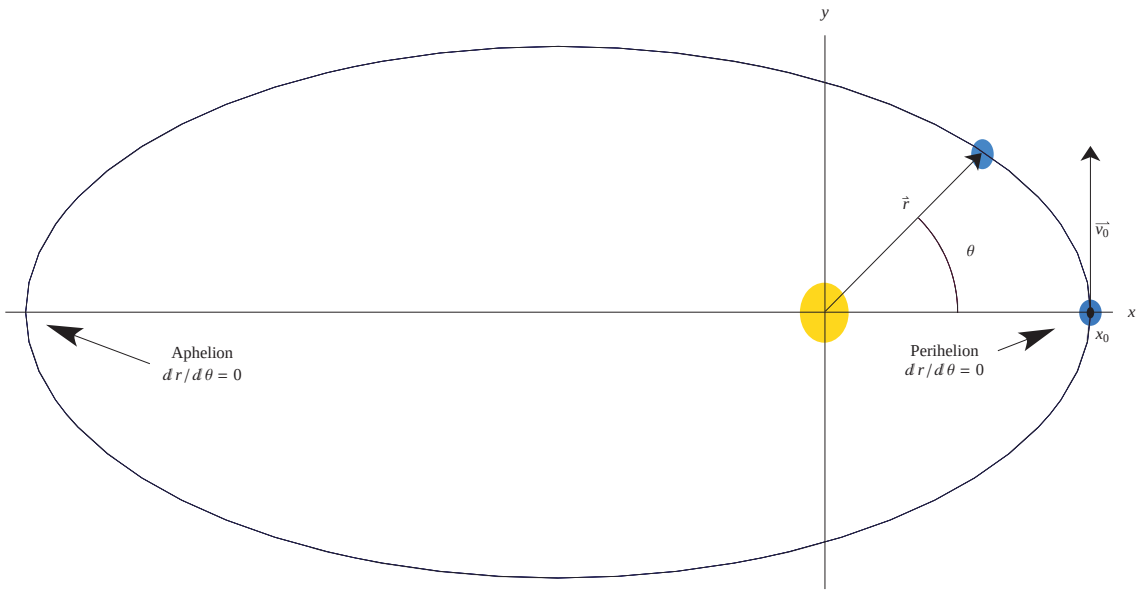


Figure 1. The basic elements of the elliptical (closed) orbit for a star-planet system. The star (large yellow disk) is fixed at the origin of our coordinate system, which is then a focus of the elliptical trajectory followed by the orbiting planet (depicted in its initial position $\vec{r}_0 = \langle x_0, 0\rangle$ and a subsequent position $\vec{r}$ by the smaller blue disks). The planet's initial velocity is assumed to have the form $\vec{v}_0 = \langle 0, v_0\rangle$, so that the ensuing orbit, as traced out by $\vec{r}$ as it sweeps out an angle θ, lies on the xy-plane. The locations of perihelion (where the planet is closest to the star) and aphelion (where the planet is farthest to the star) are indicated by arrows. Both of these locations are turning points of the orbit, in that the derivative $dr/d\theta = 0$.

As is well known, three possible trajectories exist for a celestial body interacting gravitationally with a star—hyperbolic, parabolic, and elliptical. The first two trajectories are open, and therefore represent cases for which the celestial body is able to escape the gravitational pull of the star. Clearly, planets move on elliptical (closed) orbits. For illustrative purposes, the initial conditions x_0 and v_0 required to produce the orbits of the inner planets and Halley's Comet are presented in **Table 1**, where the initial positions occur at perihelion. Two related quantities that will be discussed in **Section 2.2**—ε and j—are also listed.

2.2 Potential Formalism: Analytical Treatment of the Problem

We now consider the structure of elliptical orbits using an analytical approach, with the goal of gaining an understanding of how that structure depends on the two initial conditions that define an orbit. This analysis will be particularly useful when we turn our attention to stars orbiting within embedded clusters. We therefore take the time to develop the ideas for the familiar case of the planar elliptical orbits. Toward that end, we note that

Table 1.
Initial conditions for the inner planets.

	x_0 (AU)	v_0 (AU/yr)	ε (AU^2/yr^2)	j (AU^2/yr)
Mercury	0.307	12.4	−51.0	3.82
Venus	0.718	7.44	−27.3	5.34
Earth	0.98	6.39	−19.7	6.28
Mars	1.38	5.59	−13.0	7.72
Halley's Comet	0.586	11.5	− 1.11	6.74

the force of gravity as given in **(2)** can be expressed in terms of the gradient of a potential function Ψ (see **Exercise 2**),

$$\vec{F} = -m\vec{\nabla}\Psi \ , \tag{6}$$

where

$$\Psi = -\frac{GM}{r} \, . \tag{7}$$

The force of gravity is therefore known as a conservative force and it can be shown that the total energy per unit mass (specific energy) is conserved throughout the motion. (See **Exercise 3**.) The specific energy is given by

$$\varepsilon = \tfrac{1}{2}v^2 + \Psi \ , \tag{8}$$

with the two terms on the right-hand side of **(8)** being the kinetic energy per unit mass (*specific kinetic energy*) and the gravitational potential energy per unit mass (*specific potential energy*) of the planet, respectively. The two conserved quantities ε (*specific energy*) and j (*specific angular momentum*) provide another pair of parameters that completely specify the orbit of the planet, since their values map one-to-one with the initial conditions x_0 and v_0 at perihelion or aphelion (see **Exercise 4**):

$$\varepsilon = \frac{v_0^2}{2} - \frac{GM}{x_0}, \qquad j = x_0 v_0,$$

$$x_0 = \frac{j^2}{GM + \sqrt{(GM)^2 + 2\varepsilon j^2}}, \qquad v_0 = \frac{GM}{j} + \sqrt{\left(\frac{GM}{j}\right)^2 + 2\varepsilon} \, . \tag{9}$$

These quantities are listed in **Table 1** for the inner planets and for Halley's Comet.

Equation **(8)** then leads to an interesting alternative differential equation of motion that describes the orbit in terms of the θ (the polar angle) and r

shown in **Figure 1** [Binney and Tremaine 1987]:

$$\frac{d\theta}{dr} = \frac{1}{r}\left[\frac{2(\varepsilon - \Psi)r^2}{j^2} - 1\right]^{-1/2}. \tag{10}$$

(See **Exercise 5** for a derivation of **(10)**.) This new equation can be easily analyzed to gain insight as to the structure of the resulting orbits. For example, both perihelion and aphelion occur at turning points of the orbit, for which the quantity $dr/d\theta = 0$, as illustrated in **Figure 1**. As such, the radii of these points can be obtained by finding the zeros of the function within the square brackets of **(10)**. For the case under consideration, the turning points therefore correspond to the solutions of the quadratic

$$f(r) = 2\varepsilon r^2 + 2GMr - j^2 = 0\,, \tag{11}$$

and are given by

$$r = \frac{-GM \pm \sqrt{G^2M^2 + 2\varepsilon j^2}}{2\varepsilon}\,. \tag{12}$$

From **(12)**, we notice that ε must be less than zero if we are to obtain two positive roots for r. Hence the closed (elliptical) orbits can exist only when the total specific energy, ε, is negative. The planet's closest distance to the star (perihelion) then corresponds to the value of r obtained by using the $+$ sign, and the farthest distance (aphelion) corresponds to the value of r obtained by using the $-$ sign. The specific angular momentum is also constrained by the requirement that the argument in the square root of **(12)** not be negative, so that $j \le j_{\max} = \sqrt{-G^2M^2/2\varepsilon}$. Since the equality must correspond to circular orbits, the specific energy and specific angular momentum for a circular orbit of radius r are $\varepsilon = -GM/2r$ and $j = \sqrt{GMr}$, respectively.

In **Figure 2**, we graph the quadratic expression $f(r)$ for different values of ε and j, to illustrate how these parameters affect the elliptical orbits. If we fix ε and increase the value of j, we see that the turning points move closer together, meaning that the orbit becomes more circular. On the other hand, if we fix j and increase ε, we see that the turning point corresponding to aphelion moves farther away from the origin (star). In fact, as ε goes to zero, aphelion will approach infinity, leading to an unbound orbit.

We conclude here by pointing out that **(10)** can be used to derive the integral expression

$$\Delta\theta_{1\to 2} = \int_{r_1}^{r_2}\left[\frac{2(\varepsilon - \Psi)r^2}{j^2} - 1\right]^{-1/2}\frac{dr}{r}\,, \tag{13}$$

where $\Delta\theta_{1\to 2}$ is the angle subtended by the planet as it orbits from closest distance r_1 (perihelion) to farthest distance r_2 (aphelion). This angle must

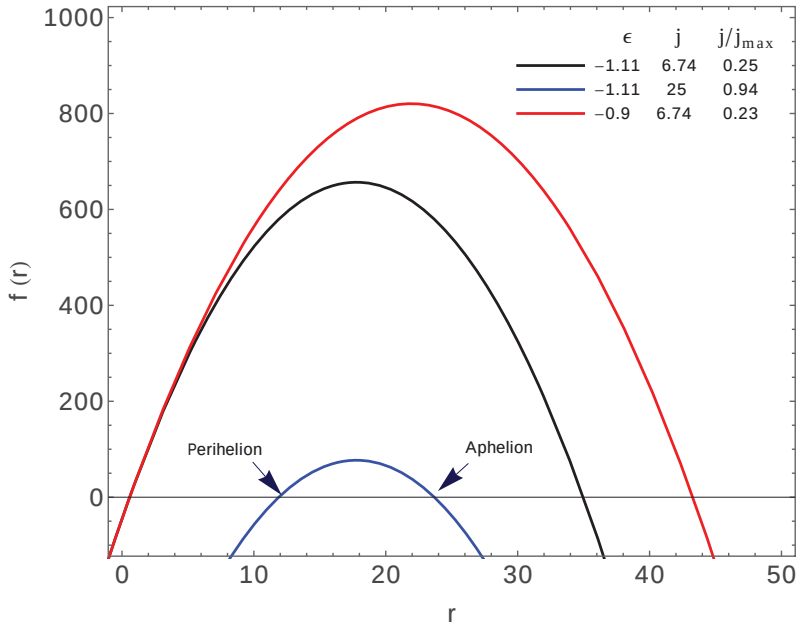


Figure 2. Quadratic expression $f(r)$ is graphed for different values of the parameters ε and j. Aphelion and perihelion correspond to the zeros of the function. Note that when ε is held constant and j is increased, the turning points move closer together indicating a more circular orbit. When the parameter j is held constant and ε is increased, the outer turning point becomes larger.

be equal to π for elliptical orbits but cannot be for spirographic orbits. Equation **(13)** will prove useful in our upcoming analysis of orbits within embedded clusters.

2.3 Numerical Analysis Using Mathematica

In **Appendix A**, we present a program in Mathematica that demonstrates how elliptical orbits depend upon initial conditions. Our program numerically solves the governing equations of a planet-star system and generates plots of the orbits resulting from the initial conditions, in terms of ε and j, as specified by the user. This program was designed only for the short time frame indicated by the specified value of $t_{\max}$; for much longer time frames, accumulating rounding errors produce unphysical solutions.

In practice, one can obtain numerical approximations to the solutions of either **(3)** or **(10)**. We choose to use the former approach, as it is easier to integrate over a complete orbit. For completeness, we note that **(3)** can be written in terms of Ψ,

$$\vec{a} = -\vec{\nabla}\Psi \ , \tag{14}$$

resulting in the coupled ordinary differential equations

$$\frac{d^2x}{dt^2} + \frac{\partial \Psi}{\partial x} = 0, \qquad \frac{d^2y}{dt^2} + \frac{\partial \Psi}{\partial y} = 0. \tag{15}$$

Three slider bars allow the user to vary the specific energy (ε), the specific angular momentum (j), and the orbital period ($t_{\max}$). The slider bar for the specific angular momentum is given in terms of $j/j_{\max}$ and runs from 0 to 1. For a fixed value of ε, the value $j/j_{\max} = 1$ corresponds to a circular orbit, while $j/j_{\max} = 0$ results in a purely radial motion. In **Figure 2**, the star is located at the origin of the graph, the motion of the orbiting body is started at perihelion, and the path of the orbiting body is the top (red) curve. For specific values of ε and $j/j_{\max}$, users can explore the time (in years) necessary to complete one orbit. (See **Exercise 6**.) Users are also encouraged to explore how the orbits change as ε is varied while holding $j/j_{\max}$ fixed (and vice versa).

Exercises

1. Show that $\dfrac{d}{dt}(\vec{r} \times \vec{v}) = \vec{0}$.

2. Show that $\vec{F} = -m\vec{\nabla}\Psi = -\dfrac{GMm\vec{r}}{r^3}$.

3. In this exercise we show that the quantity given by **(8)**, $\varepsilon = \frac{1}{2}v^2 + \Psi$ is conserved throughout the motion.

a) Beginning with $\vec{F} = m\vec{a} = -m\vec{\nabla}\Psi$, show that $\dfrac{d\vec{v}}{dt} = -\vec{\nabla}\Psi$.

b) Now integrate the vector fields over the path swept out by the tip of vector $\vec{r}$, to obtain

$$\int_C \frac{d\vec{v}}{dt} \cdot d\vec{r} = -\int_C \vec{\nabla}\Psi \cdot d\vec{r},$$

or

$$\int_\alpha^\beta \frac{d\vec{v}}{dt} \cdot \frac{d\vec{r}}{dt} dt = -\int_\alpha^\beta \vec{\nabla}\Psi \cdot \frac{d\vec{r}}{dt}\, dt,$$

where α and β are arbitrary values of time.

c) Use the Fundamental Theorem of Line Integrals to evaluate

$$-\int_\alpha^\beta \vec{\nabla}\Psi \cdot \frac{d\vec{r}}{dt}\, dt.$$

d) Show that

$$\frac{d}{dt}\left(\frac{d\vec{r}}{dt} \cdot \frac{d\vec{r}}{dt}\right) = 2\,\frac{d^2\vec{r}}{dt^2} \cdot \frac{d\vec{r}}{dt} = 2\,\frac{d\vec{v}}{dt} \cdot \frac{d\vec{r}}{dt}.$$

e) Using the result of part **(d)**, show that

$$\int_\alpha^\beta \frac{d\vec{v}}{dt}\cdot\frac{d\vec{r}}{dt}\,dt = \frac{1}{2}\int_\alpha^\beta \frac{d}{dt}\left(\frac{d\vec{r}}{dt}\cdot\frac{d\vec{r}}{dt}\right)\,dt.$$

f) Show that

$$\frac{1}{2}\int_\alpha^\beta \frac{d}{dt}\left(\frac{d\vec{r}}{dt}\cdot\frac{d\vec{r}}{dt}\right)\,dt = \frac{1}{2}\big[v^2(\beta)-v^2(\alpha)\big],$$

where v is the magnitude of $\vec{v}$.

g) Using the results of parts **(c)** and **(f)**, show that the equation given in part **(b)** can be expressed as $\frac{1}{2}v^2(\beta)+\Psi(\beta)=\frac{1}{2}v^2(\alpha)+\Psi(\alpha)$, or $\varepsilon(\beta)=\varepsilon(\alpha)$. Since β and α are arbitrary values of time, we see that the specific energy, ε, is constant for all time.

4. a) Evaluate ε and j at $t=0$ to obtain

$$\varepsilon = \frac{1}{2}\,v_0^2 - \frac{GM}{x_0}, \qquad j = x_0 v_0.$$

b) Using the results of part **(a)**, solve for v_0 and x_0 in terms of j and ε.

5. In this exercise, we derive **(10)**. Refer to **Figure 3** as we start with some preliminary observations.

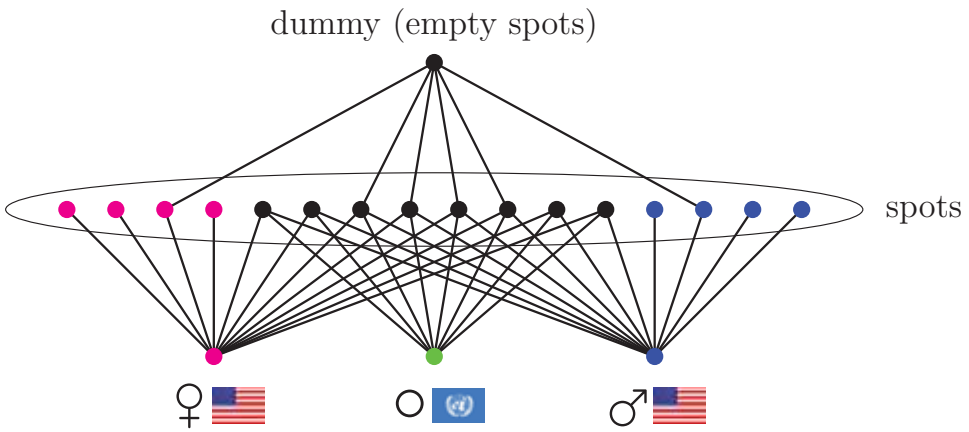


Figure 3. The orbital elements defined in **Exercise 5**. *Left panel:* The position vectors $\vec{r}_1$ and $\vec{r}_2$ indicate the location of the planet at times t_1 and t_2. The corresponding angle $\Delta\theta$ subtended by the planet's motion and the displacement vector $\vec{r}_2-\vec{r}_1$ are also shown, along with its components Δr and Δs of the displacement vector along the radial and azimuthal directions, as indicated by the unit vectors $\hat{u}_r$ and $\hat{u}_\theta$. *Right panel:* The instantaneous velocity vector $\vec{v}$ at position $\vec{r}$ is obtained in the limit $\Delta t\to 0$, along with its radial and azimuthal components v_r and v_θ.

First, let $\hat{u}_r$ denote a unit vector in the radial direction and $\hat{u}_\theta$ denote a unit vector that is perpendicular to $\hat{u}_r$. Let $\vec{r}(t_1)=\vec{r}_1$, $\vec{r}(t_2)=\vec{r}_2$, let

Δt denote $t_2 - t_1$, and let $\Delta\theta$ indicate the corresponding change in polar angle. The average velocity is given by

$$\frac{\vec{r}_2 - \vec{r}_1}{t_2 - t_1} \qquad \text{or, in component form,} \qquad \frac{\Delta r\hat{u}_r + \Delta s\hat{u}_\theta}{\Delta t}.$$

For small Δt (and thus small $\Delta\theta$),

$$|\vec{r}_1| \approx |\vec{r}_2| \approx r \qquad \text{and} \qquad \Delta s \approx r\Delta\theta \quad \text{(arclength)}.$$

Then the velocity vector is given by

$$\vec{v} = \lim_{\Delta t \to 0} \left(\frac{\Delta r\hat{u}_r}{\Delta t} + \frac{r\Delta\theta\hat{u}_\theta}{\Delta t} \right) = \frac{dr}{dt}\,\hat{u}_r + r\,\frac{d\theta}{dt}\,\hat{u}_\theta.$$

Letting v_r denote $\dfrac{dr}{dt}$ and v_θ denote $r\,\dfrac{d\theta}{dt}$, we can write $\vec{v} = v_r\hat{u}_r + v_\theta\hat{u}_\theta$ and note that $v^2 = v_r^2 + v_\theta^2$.

a) Show that $|\vec{r} \times \vec{v}| = j$ yields $\dfrac{d\theta}{dt} = \dfrac{j}{r^2}$.

b) Show that $\dfrac{dr}{dt} = \dfrac{j}{r^2}\dfrac{dr}{d\theta}$.

c) Starting with $\varepsilon = \frac{1}{2}v^2 + \Psi$, show that

$$\frac{1}{2}\left[\left(\frac{dr}{dt}\right)^2 + \left(r\,\frac{d\theta}{dt}\right)^2\right] = \varepsilon - \Psi.$$

d) Using the results of parts **(a)**, **(b)**, and **(c)**, show that

$$\frac{dr}{d\theta} = r\sqrt{\frac{2r^2(\varepsilon - \Psi)}{j^2} - 1}.$$

e) Finally, using **(d)**, show that $\dfrac{d\theta}{dr} = \dfrac{1}{r}\left[\dfrac{2(\varepsilon - \Psi)r^2}{j^2} - 1\right]^{-1/2}$.

6. Using the Mathematica program from **Appendix A** and the values from **Table 1**, estimate how many years it takes for Halley's Comet to orbit the Sun.

7. Using the Mathematica program from **Appendix A**, graph the orbits corresponding to the values given in **Figure 2** for ε and $j/j_{\max}$.

a) Graph i): $\varepsilon = -1.11$, $j/j_{\max} = 0.25$;

b) Graph ii): $\varepsilon = -1.11$, $j/j_{\max} = 0.94$;

c) Graph iii): $\varepsilon = -0.90$, $j/j_{\max} = 0.23$.

Comparing graphs i) and ii), notice that you have kept ε (and $j_{\max}$) constant and increased j, which has resulted in a more circular orbit because the turning points have moved closer together.

Comparing graphs i) and iii), notice that you have kept j constant and increased ε which has increased the value of the outer turning point.

3. Analytical Treatment

In **Section 2**, we explored the motion of a planet about a fixed star through a potential formalism that yields information about the structure of the resulting elliptical orbits. We now use the same formalism to explore the motion of a young star within its birthplace, an embedded cluster. This time, however, we'll find that *the resulting orbits are spirographic in nature rather than elliptical*. Our goal is to gain insight into the structure of the resulting spirographic orbits, and in particular, how that structure depends on the choice of the two initial conditions that define an orbit, as we did in **Section 2**.

The equations of motion given by **(15)** for the path of a planet within our solar system then provide the starting point for our current analysis. It is clear, however, that our previous potential function, $\Psi = -GM/r$, is no longer valid as we cannot neglect the gravitational pull coming from the gas and other stars in the closely packed embedded cluster. Our first task then is to derive a new gravitational potential function that adequately describes the cluster environment.

We start by discussing a particular density profile for the cluster environment. We then show that the density profile is related to its corresponding potential function through Poisson's equation. Finally, solving Poisson's equation for our new potential function allows us to carry out the same sort of analysis that was done for the motion of a planet in our solar system.

3.1 Hernquist Density Profile

The majority of embedded clusters observed within our galaxy contain between $N_* \approx 100 - 1000$ stars, with $N_* = 300$ generally taken as a typical value. These stars are contained within a radius R_c between 0.1–2.0 pc, with a clear correlation observed between R_c and N_*. Most of the gas from which these stars were born remains unused by the star formation process. In an embedded cluster, this leftover gas accounts for 70–90% of the total (gas and stellar) mass of the cluster, and extends well beyond R_c until eventually merging smoothly into the background gas.

The complex nature of embedded clusters poses a formidable challenge to our formalism. We need to specify a potential function Ψ that suitably represents the gravitational interactions between our star and all of the

other cluster members and gas, but it must also be a potential function that keeps the problem tractable. One way to do so is to assume that the gas and stellar components of a young cluster can be approximated by a smooth mass density profile that is spherically symmetric, and to then obtain the corresponding potential function using Poisson's equation.

Several spherically symmetric density functions have been proposed, which, guided by observations, have high central values that decrease monotonically with radius. We consider here one of these functions—the *Hernquist density profile*—which was first introduced to describe elliptical galaxies [Hernquist 1990]. This density profile, which has also been used to describe dark matter halos of galaxies, galaxy clusters, and the embedded clusters that form the focus of this module [see, e.g., Adams and Bloch 2005], is given by the function

$$\rho(\xi) = \frac{\rho_0}{\xi(1+\xi)^3}\,, \tag{16}$$

where $\xi = r/r_s$ is the dimensionless radius, r_s is an effective lengthscale, and ρ_0 is a scaling parameter set by specifying the mass of the environment being considered. In the context of our problem, we take the constant r_s to represent the cluster radius (so that $r_s = R_c$), and specify the value of the constant ρ_0 through the mass of the cluster (gas and stars) contained *within* radius r as found by integrating the density function over a spherical volume of radius r:

$$M_r = \int \rho(\xi)dV = \int_0^r 4\pi r'^2 \rho(\xi)\, dr' = \frac{2\pi r_s \rho_0 r^2}{(1+r/r_s)^2}\,. \tag{17}$$

In the center of the cluster, where ξ is small, $\rho(\xi) \sim \rho_0/\xi \to \infty$, but $M_r \to 0$. At the effective edge of the cluster $r = r_s$ ($\xi = 1$), we have $\rho(\xi) = \rho_0/8$ and $M_r = \pi\rho_0 r_s^3/2$. Moving out beyond this effective edge, we arrive at $\rho(\xi) \sim \rho_0/\xi^4 \to 0$ and $M_r \to 2\pi\rho_0 r_s^3$. The normalized radial profiles for the Hernquist density and the corresponding cluster mass contained within radius r, as given by **(16)** and **(17)**, are shown in **Figure 4**.

3.2 Hernquist Potential Function

We now outline how the potential function Ψ can be obtained from the Hernquist density profile, ρ, using Poisson's equation, but leave the details to be addressed through exercises.

The first step is to calculate the force $\vec{F}(\vec{r})$ acting on a unit mass located at position $\vec{r}$ resulting from its gravitational attraction to the extended mass distribution. Using Newton's Law of Gravitation, we find the small contribution to the overall force from a small volume element ΔV located at position $\vec{r}\,'$ within the cluster, as diagrammed in **Figure 5**.

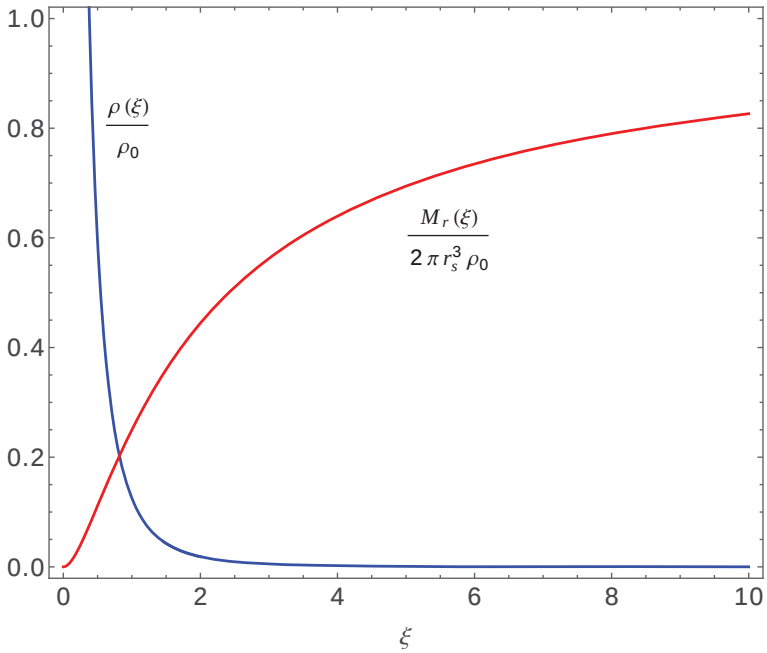


Figure 4. Normalized radial profiles for the Hernquist density (falling blue curve) and corresponding cluster mass contained within radius r (rising red curve), as given by **(16)** and **(17)**, respectively.

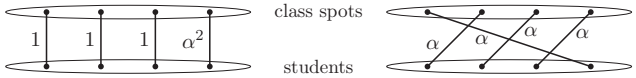


Figure 5. Sketch of the force element $\Delta\vec{F}$ resulting from the gravitational attraction between a volume element ΔV of the embedded cluster located at position $\vec{r}\,'$ and the star located at $\vec{r}$.

Denoting the mass of this small element by $\Delta M = \rho(\vec{r}\,')\Delta V$, we can then find the force element by generalizing **(2)**:

$$\Delta\vec{F}(\vec{r}) = \frac{G(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\,\rho(\vec{r}\,')\,\Delta V\,. \tag{18}$$

Summing up over the volume of the cluster, the total force acting on our unit mass is given by:

$$\vec{F}(\vec{r}) = G\int\int\int_V \frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\,\rho(\vec{r}\,')\,dV\,. \tag{19}$$

Taking the divergence (with respect to $\vec{r}$) of **(19)**, we have

$$\vec{\nabla}\cdot\vec{F}(\vec{r}) = G\int\int\int_V \vec{\nabla}_{\vec{r}}\cdot\frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\,\rho(\vec{r}\,')\,dV\,. \tag{20}$$

We can show (see **Exercise 9**) that the integrand on the right is zero if $\vec{r}\,' - \vec{r} \neq \vec{0}$. Therefore, the only contribution to the integral must come from the point $\vec{r}\,' = \vec{r}$. The integral then reduces to a volume integral computed over a very small sphere of radius h centered at the point indicated by $\vec{r}$. Summarizing, we have

$$\vec{\nabla}\cdot\vec{F}(\vec{r}) = G\int\int\int_{|\vec{r}\,'-\vec{r}|<h} \vec{\nabla}_{\vec{r}}\cdot\frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\,\rho(\vec{r}\,')\,dV\,. \tag{21}$$

Since h is small, we can assume that the density is essentially constant throughout the small sphere and equal to $\rho(\vec{r})$. Rearranging the integral from **(21)**, we get

$$\vec{\nabla}\cdot\vec{F}(\vec{r}) = G\,\rho(\vec{r})\int\int\int_{|\vec{r}\,'-\vec{r}|<h} \vec{\nabla}_{\vec{r}}\cdot\frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\,dV\,. \tag{22}$$

Noting that (see **Exercise 10**)

$$\vec{\nabla}_{\vec{r}}\cdot\frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3} = -\vec{\nabla}_{\vec{r}\,'}\cdot\frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\,, \tag{23}$$

we have

$$\vec{\nabla}\cdot\vec{F}(\vec{r}) = -G\,\rho(\vec{r})\int\int\int_{|\vec{r}\,'-\vec{r}|<h} \vec{\nabla}_{\vec{r}\,'}\cdot\frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\,dV\,. \tag{24}$$

The Divergence Theorem allows us to express the volume integral in **(24)** as a surface integral that can then be easily evaluated (see **Exercise 11**):

$$\vec{\nabla}\cdot\vec{F}(\vec{r}) = -G\,\rho(\vec{r})\int\int_{|\vec{r}\,'-\vec{r}|=h} \frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}\cdot d\vec{S}\,' = -4\pi G\rho(\vec{r})\,. \tag{25}$$

If we now define

$$\Psi(\vec{r}) = -G \int \int \int_V \frac{\rho(\vec{r}\,')}{|\vec{r}\,' - \vec{r}\,|}\, dV\,, \tag{26}$$

and notice that (see **Exercise 12**)

$$\vec{F}(\vec{r}) = -\vec{\nabla}\Psi(\vec{r})\,, \tag{27}$$

then **(25)** can be rewritten as *Poisson's equation*

$$\nabla^2\Psi(\vec{r}) = 4\pi G\rho(\vec{r})\,. \tag{28}$$

Since ρ is spherically symmetric, we express $\vec{r}$ in spherical coordinates and note that $\vec{r}$ depends only on r (and not on θ or ϕ). Equation **(28)** then reduces to the following second-order ordinary differential equation:

$$\frac{d^2\Psi}{dr^2} + \frac{2}{r}\frac{d\Psi}{dr} = 4\pi G\,\rho\,. \tag{29}$$

One can then show (see **Exercises 13** and **14**) that for the Hernquist density profile

$$\rho = \frac{\rho_0}{\dfrac{r}{r_s}\left(1 + \dfrac{r}{r_s}\right)^3},$$

the potential function

$$\Psi = -\frac{\Psi_0}{1+\xi}, \tag{30}$$

known as the *Hernquist potential*, solves **(29)**, where $\Psi_0 = 2\pi G r_s^2 \rho_0$.

3.3 Nondimensionalized Potential Formalism

With the potential function specified, we now perform an analysis similar to that presented in **Section 2.2** for planetary orbits. Since the density profile for the cluster environment is spherically symmetric, the net gravitational force exerted on the star by the cluster environment must be directed radially inward. As such, the resulting acceleration must also be directed radially inward, so that $\hat{a} = -\hat{r}$. As shown in **Section 2.1**, the motion of the star must therefore be planar. The governing equations, given by **(15)** and **(10)**, along with the potential function derived in **Section 3.2**, describe the resulting orbits.

In the planet-star system explored in **Section 2**, it is natural to use the 1 AU as the lengthscale, 1 yr as the timescale, and 1 $M_\odot$ (mass of our Sun) as the mass scale. Doing so then sets the value of the constant G, as discussed

in **Section 2.1**. In contrast, the cluster environment provides a natural lengthscale r_s, density ρ_0, and potential Ψ_0 by which to nondimensionalize the equations. We therefore define the following dimensionless quantities:

- $\xi = r/r_s$, $\bar{x} = x/r_s$ and $\bar{y} = y/r_s$;
- $\bar{t} = t/\tau_0$, where $\tau_0 = r_s/\sqrt{2\Psi_0}$ and $\Psi_0 = 2\pi G r_s^2 \rho_0$;
- $\epsilon = \varepsilon/\Psi_0$;
- $q = j^2/(2\Psi_0 r_s^2)$.

Analogous to the planet-star system case, the two dimensionless quantities ϵ and q are constants of the motion whose values completely specify the orbit within the cluster. As such, they must map one-to-one to the initial conditions, which in dimensionless form are expressed as $\bar{x}_0 = x_0/r_s$ and $\bar{v}_0 = v_0/(r_s/\tau_0) = v_0/\sqrt{2\Psi_0}$. Specifically, substituting **(30)** into **(8)** and expression $j = x_0 v_0$ into the definition of q above, it is easy to show that

$$\epsilon = \bar{v}_0^2 - \frac{1}{1+\bar{x}_0}, \qquad q = \bar{x}_0^2\,\bar{v}_0^2\,. \tag{31}$$

As elaborated below, obtaining expressions for $\bar{x}_0$ and $\bar{v}_0$ in terms of ϵ and q requires solving a cubic equation. While analytic solutions can always be obtained, we leave it to the adventurous reader to do so.

In our dimensionless system, **(10)** becomes

$$\frac{\xi\, d\theta}{d\xi} = \left[\left(\epsilon + \frac{1}{1+\xi}\right)\frac{\xi^2}{q} - 1\right]^{-1/2}\,. \tag{32}$$

As discussed in **Section 2.2**, the derivative $d\xi/d\theta$ equals zero when the star is either nearest to or farthest from the cluster center. The radial distances ξ_1 and ξ_2 of the inner and outer turning points must therefore be roots of the expression within the square bracket of **(32)**. However, while the formalism is **Section 2** yielded a quadratic equation whose roots were the distances to the turning points, here those distances are given by the positive roots of the cubic function

$$f(\xi) = \epsilon\xi^3 + (1+\epsilon)\xi^2 - q\xi - q\,. \tag{33}$$

Closed orbits require that two of the three roots be real and positive, while the third root must then be negative over the range of physically realistic parameter values.

Analytic expressions for these roots can always be obtained, and their nature can always be determined by the sign of the discriminant. However, the length and complexity of the resulting expression would not provide any meaningful insight into the problem. In addition, we note that numerical solutions to **(33)** may be extremely inaccurate for ϵ near zero. We direct

the interested reader to standard numerical analysis literature and to the analysis of Adams and Bloch [2005]; we summarize their results.

Specifically, we note that closed orbits require $-1 \leq \epsilon \leq 0$, where $\epsilon \sim -1$ corresponds to an orbit confined near to the center of the cluster, and orbits confined within $\xi = 1$ require that ϵ be less than approximately -0.375. In addition, orbits can exist only if

$$q \leq q_{\max} \equiv \frac{(1+\sqrt{1-8\epsilon}+4\epsilon)^3}{-8\epsilon(1+\sqrt{1-8\epsilon})^2}\,. \tag{34}$$

where $q = 0$ corresponds to radial motion, and $q = q_{\max}$ corresponds to circular orbits.

To gain a further understanding of how the spirographic orbits depend on the parameters ϵ and q, we plot $f(\xi)$ as a function of ξ in **Figure 6** for three sets of (ϵ, q) values.

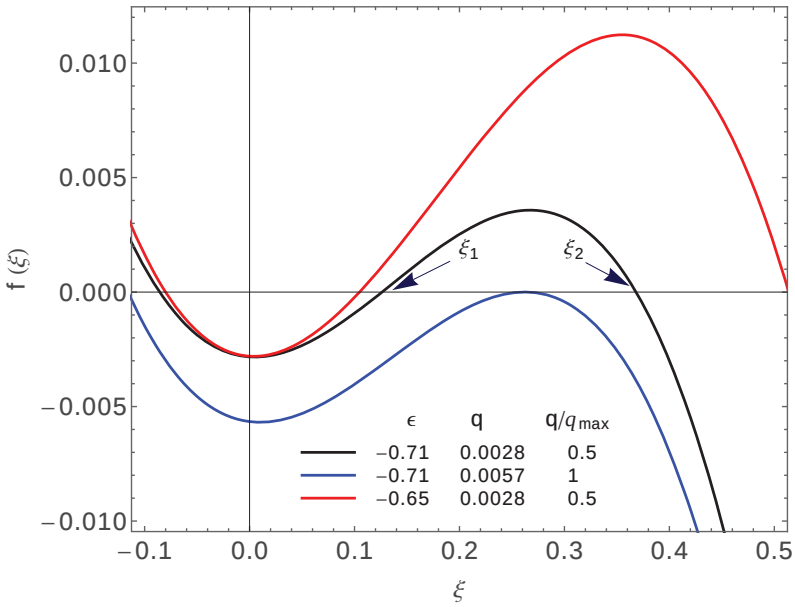


Figure 6. Cubic expression $f(\xi)$ for values of the parameters ϵ and q. The turning points ξ_1 and ξ_2 of the orbit correspond to the zeros of the function, and are shown for the case with $\epsilon = -0.71, q = 0.5q_{\max}$ (black curve in the middle). Note that when ϵ is held constant and q is increased, the turning points move closer together, and a circular orbit is obtained when $q = q_{\max}$ (bottom blue curve). When the parameter q is held constant and ϵ is increased, the outer turning point becomes larger (top red curve).

If we fix ϵ and increase the value of q, we see that the turning points move closer together, and they become circular when $q = q_{\max}$. On the other hand, if we fix q and increase ϵ, we see that the outer turning point moves farther away from the center of the cluster. In fact, as ϵ goes to zero, this outer turning point approaches infinity leading to an unbound orbit.

Finally, we note that the angle $\Delta\theta_{1\to 2}$ through which the orbit moves from closest distance ξ_1 to farthest distance ξ_2 is still given by **(13)**, which

for the Hernquist potential and the adopted dimensionalization, reduces to

$$\Delta\theta_{1\to 2} = \sqrt{q}\int_{\xi_1}^{\xi_2}\left[\epsilon\xi^2 + \frac{\xi^2}{1+\xi} - q\right]^{-1/2}\frac{d\xi}{\xi}\,. \tag{35}$$

For the parameters of interest here, the value of $\Delta\theta_{1\to 2}$ is less than π [Adams and Bloch 2005], leading to spirographic orbits rather than to elliptical orbits, as can be seen from **Figure 7**.

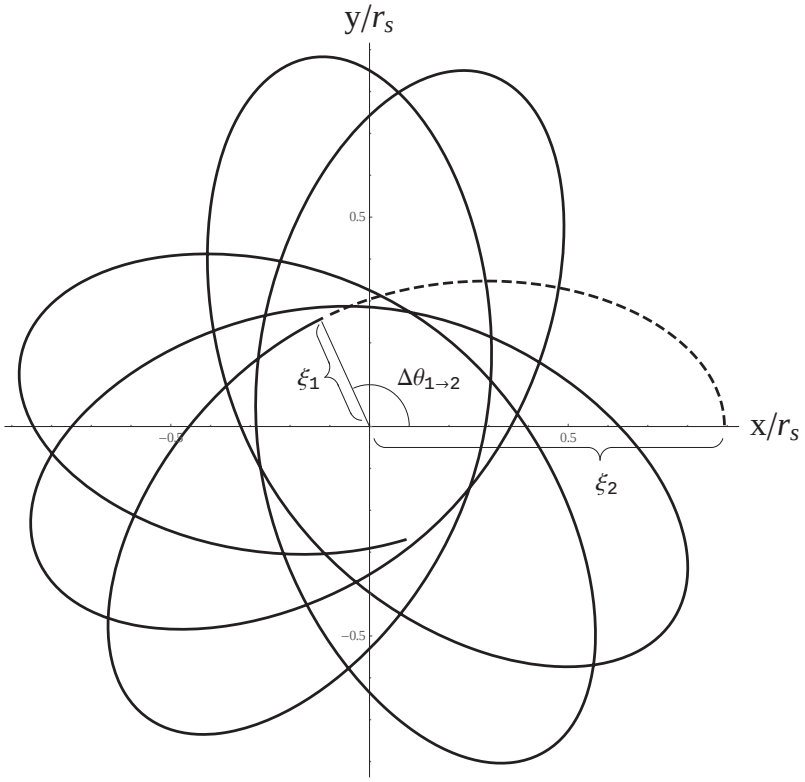


Figure 7. Spirographic orbit for dimensionless parameters $\epsilon = -0.5$ and $q/q_{\max} = 0.5$. The dashed curve depicts a half-orbit with the turning points $\xi_1 = 0.28$ and $\xi_2 = 0.89$. The angle $\Delta\theta_{1\to 2}$, through which the orbit moves from farthest distance ξ_2 to closest distance ξ_1, is less than π, leading to a spirographic orbit rather than to an elliptical orbit.

Exercises

8. Integrate

$$\int_0^r 4\pi r'^2\rho(\xi')dr' = \int_0^r 4\pi r'^2\frac{\rho_0}{\frac{r'}{r_s}(1+\frac{r'}{r_s})^3}\,dr'$$

to show that $M_r = \dfrac{2\pi r_s\rho_0 r^2}{(1+r/r_s)^2}$.

9. We show that if $\vec{r}' \neq \vec{r}$, then $\vec{\nabla}_{\vec{r}} \cdot \dfrac{(\vec{r}' - \vec{r})}{|\vec{r}' - \vec{r}|^3} = 0$, hence justifying the replacement of **(20)** by **(21)**.

a) Show that

$$\vec{\nabla}_{\vec{r}} \cdot \frac{(\vec{r}' - \vec{r})}{|\vec{r}' - \vec{r}|^3} = \frac{1}{|\vec{r}' - \vec{r}|^3}\left[\vec{\nabla}_{\vec{r}} \cdot (\vec{r}' - \vec{r})\right] + (\vec{r}' - \vec{r}) \cdot \vec{\nabla}_{\vec{r}} \left[\frac{1}{|\vec{r}' - \vec{r}|^3}\right].$$

b) Show that

$$\vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r}' - \vec{r}|^3}\right) = \frac{3(\vec{r}' - \vec{r})}{|\vec{r}' - \vec{r}|^5}.$$

c) Show that $\vec{\nabla}_{\vec{r}} \cdot (\vec{r}' - \vec{r}) = -3$.

d) Using **(b)** and **(c)**, show that if $\vec{r}' \neq \vec{r}$, then the expression in **(a)** is equal to 0.

10. Show that

$$\vec{\nabla}_{\vec{r}} \cdot \frac{(\vec{r}' - \vec{r})}{|\vec{r}' - \vec{r}|^3} = -\vec{\nabla}_{\vec{r}'} \cdot \frac{(\vec{r}' - \vec{r})}{|\vec{r}' - \vec{r}|^3}.$$

11. Show that

$$-G\,\rho(\vec{r}) \int\int_{|\vec{r}' - \vec{r}| = h} \frac{(\vec{r}' - \vec{r})}{|\vec{r}' - \vec{r}|^3} \cdot d\vec{S} = -4\pi G \rho(\vec{r}).$$

12. Given that

$$\Psi(\vec{r}) = -G \int\int\int_V \frac{\rho(\vec{r}')}{|\vec{r}' - \vec{r}|}\, dV,$$

show that $-\vec{\nabla}\Psi(\vec{r})$ produces $\vec{F}(\vec{r})$, as given by **(19)**.

13. Show that $\Psi = -\dfrac{\Psi_0}{1 + \xi}$ solves **(29)**.

14. In this exercise, we solve the nonhomogeneous Euler equation given by **(29)**. We first rewrite **(29)** in the form

$$\frac{r^2 d^2\Psi}{dr^2} + 2r\frac{d\Psi}{dr} = 4r^2\pi G\,\rho. \qquad \textbf{(36)}$$

a) Solve the homogeneous Euler equation

$$r^2\,\frac{d^2\Psi}{dr^2} + 2r\,\frac{d\Psi}{dr} = 0$$

by assuming a solution of the form $\Psi = r^n$.

b) Use the method of variation of parameters to find the particular solution of **(36)**.

c) Write down the general solution of **(36)** using parts **(a)** and **(b)**.

d) Solve for the arbitrary constants in the general solution to produce the solution $\Psi = -\dfrac{\Psi_0}{1+\xi}$, where $\Psi_0 = 2\pi G r_s^2 \rho_0$, by imposing the conditions:

$$\lim_{r\to\infty} \Psi = 0, \qquad \lim_{r\to 0} \Psi = c,$$

where c represents a nonzero constant value.

15. Substitute **(30)** into **(8)** and expression $j = x_0 v_0$ into the definition of q from **Section 3.3** to show that

$$\epsilon = \bar{v}_0^2 - \frac{1}{1+\bar{x}_0}, \qquad q = \bar{x}_0^2\, \bar{v}_0^2.$$

16. Show that **(10)** can be rewritten as **(32)** using the change of variables given in **Section 3**.

17. Show that setting the quantity inside the brackets in **(32)** equal to zero yields **(33)**.

4. Numerical Simulations

In **Appendix B**, we present a Mathematica program that demonstrates the spirographic orbits of stars within an embedded cluster, for varying initial conditions.

Our program numerically solves the governing equations of our embedded cluster system and generates plots of the orbits resulting from the initial conditions, in terms of ϵ and $q/q_{\max}$, as specified by the user. This program was designed only for the short time frame indicated by the specified value of $t_{\max}$. For much longer time frames, accumulating rounding errors produce unphysical solutions.

As with the elliptical orbits, we numerically integrate **(3)**, which when written in terms of Ψ and in dimensionless form yields

$$\frac{d^2\bar{x}}{d\bar{t}^2} - \frac{1}{2}\frac{\partial}{\partial \bar{x}}\left[\frac{1}{1+\xi}\right] = 0, \qquad \frac{d^2\bar{y}}{d\bar{t}^2} - \frac{1}{2}\frac{\partial}{\partial \bar{y}}\left[\frac{1}{1+\xi}\right] = 0. \qquad \textbf{(37)}$$

Three slider bars allow the user to vary three different (dimensionless) parameters: the two constants of the motion ϵ and q, and the orbital period

$\bar{t}_{\max}$ of the star. The slider bar for q is given in terms of $q/q_{\max}$ and as such runs from 0 to 1. For a fixed value of ϵ, $q/q_{\max} = 1$ corresponds to a circular orbit, while $q/q_{\max} = 0$ results in a purely radial motion. Note that the center of the cluster is located at the origin of the graph, the motion of the orbiting body is started at the outer turning point ξ_2, and the path of the orbiting body is drawn in blue. A particular case is shown in **Figure 8**; see also Landis et al. [n.d.].

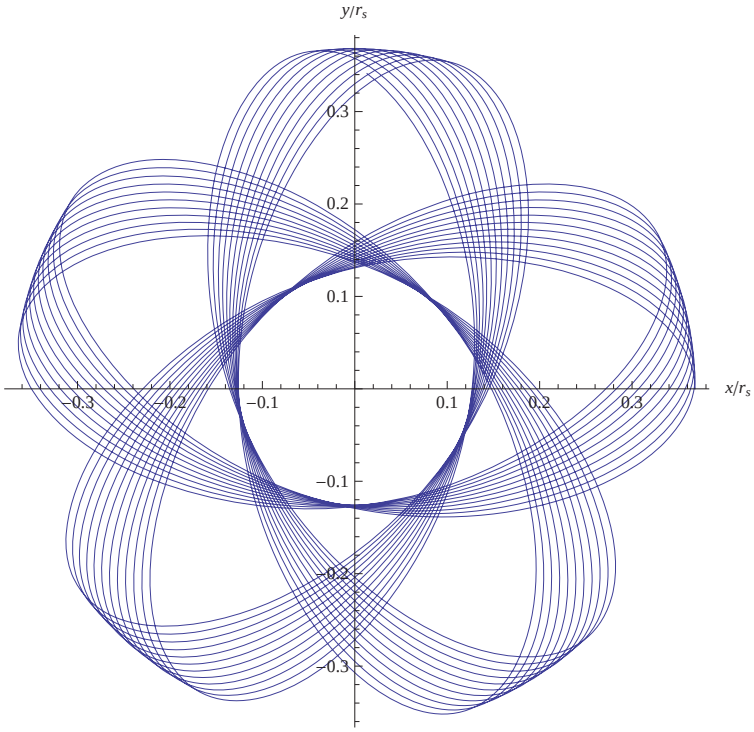


Figure 8. The spirographic orbit obtained using the Mathematica `Manipulate` for the parameters corresponding to the case shown by the black curve in **Figure 6** ($\epsilon = -0.71$, $q/q_{\max} = 0.5$).

Exercises

18. Using the dimensionless quantities presented in **Section 3.3**, show that **(15)** can be rewritten as **(37)**.

19. Using the Mathematica program from **Section 4**, graph the spirographic orbits corresponding to the values given in **Figure 6** for ϵ and $q/q_{\max}$.

a) Graph i): $\epsilon = -0.71, \;\; q/q_{\max} = 0.5$;

b) Graph ii): $\epsilon = -0.71, \;\; q/q_{\max} = 1.0$;

c) Graph iii): $\epsilon = -0.65, \;\; q/q_{\max} = 0.5$.

Comparing graphs i) and ii), notice that you have kept ϵ (and $q_{\max}$) constant and increased q to $q_{\max}$, which has resulted in a circular orbit because the turning points have coalesced.

Comparing graphs i) and iii), notice that you have kept q constant and increased ϵ, which has increased the value of the outer turning point.

5. Solutions to the Exercises

1. $\dfrac{d}{dt}(\vec{r}\times\vec{v}) = \vec{r}\,'(t)\times\vec{v}(t) + \vec{r}(t)\times\vec{v}\,'(t) = \vec{v}\times\vec{v} + \vec{r}\times\vec{a} = \vec{0}+\vec{0} = \vec{0}.$

2. Using $\Psi = -\dfrac{GM}{r}$,

$$\begin{aligned}
-m\vec{\nabla}\Psi &= -m\left\langle \frac{\partial}{\partial x}\left(\frac{-GM}{\sqrt{x^2+y^2}}\right), \frac{\partial}{\partial y}\left(\frac{-GM}{\sqrt{x^2+y^2}}\right)\right\rangle \\
&= GMm\left\langle \frac{\partial}{\partial x}\left(\frac{1}{\sqrt{x^2+y^2}}\right), \frac{\partial}{\partial y}\left(\frac{1}{\sqrt{x^2+y^2}}\right)\right\rangle \\
&= -GMm\left\langle \frac{x}{(x^2+y^2)^{\frac{3}{2}}}, \frac{y}{(x^2+y^2)^{\frac{3}{2}}}\right\rangle \\
&= -\frac{GMm}{(x^2+y^2)^{3/2}}\langle x, y\rangle = -\frac{GMm}{r^3}\vec{r}.
\end{aligned}$$

3. **a)** Since $\vec{a} = d\vec{v}/dt$, we have $m\,d\vec{v}/dt = -m\vec{\nabla}\Psi$. Since m is not zero, we conclude that $d\vec{v}/dt = -\vec{\nabla}\Psi$.

b) We told you how it is supposed to come out....

c) $-\displaystyle\int_\alpha^\beta \vec{\nabla}\Psi\cdot\frac{d\vec{r}}{dt}\,dt = -\big[\Psi(\beta)-\Psi(\alpha)\big].$

d) $\dfrac{d}{dt}\left(\dfrac{d\vec{r}}{dt}\cdot\dfrac{d\vec{r}}{dt}\right) = \dfrac{d^2\vec{r}}{dt^2}\cdot\dfrac{d\vec{r}}{dt} + \dfrac{d\vec{r}}{dt}\cdot\dfrac{d^2\vec{r}}{dt^2} = 2\,\dfrac{d^2\vec{r}}{dt^2}\cdot\dfrac{d\vec{r}}{dt} = 2\,\dfrac{d\vec{v}}{dt}\cdot\dfrac{d\vec{r}}{dt}.$

e) Use the fact that $\dfrac{d\vec{v}}{dt}\cdot\dfrac{d\vec{r}}{dt} = \dfrac{d^2\vec{r}}{dt^2}\cdot\dfrac{d\vec{r}}{dt} = \dfrac{1}{2}\dfrac{d}{dt}\left(\dfrac{d\vec{r}}{dt}\cdot\dfrac{d\vec{r}}{dt}\right).$

f) $\dfrac{1}{2}\displaystyle\int_\alpha^\beta \frac{d}{dt}\left(\frac{d\vec{r}}{dt}\cdot\frac{d\vec{r}}{dt}\right)dt = \frac{1}{2}\int_\alpha^\beta \frac{d}{dt}\big(v^2(t)\big)\,dt = \frac{1}{2}\big(v^2(\beta)-v^2(\alpha)\big).$

g) Then

$$\int_\alpha^\beta \frac{d\vec{v}}{dt}\cdot\frac{d\vec{r}}{dt}\,dt = -\int_\alpha^\beta \vec{\nabla}\Psi\cdot\frac{d\vec{r}}{dt}\,dt,$$

$$\frac{1}{2}\big(v^2(\beta)-v^2(\alpha)\big) = -\big[\Psi(\beta)-\Psi(\alpha)\big],$$

$$\frac{1}{2}\,v^2(\beta)+\Psi(\beta) = \frac{1}{2}\,v^2(\alpha)+\Psi(\alpha)\,.$$

4. a) Note that when $t = 0$, then $r = x_0$, $v = v_0$ and the angle between $\vec{v}$ and $\vec{r}$ is 90°(see **Figure 1**). Using $\Psi = -GM/r$ and **(8)**, we have $\varepsilon = \frac{1}{2}v_0^2 - \frac{GM}{x_0}$. Using the fact that $j = |\vec{r} \times \vec{v}| = |\vec{r}||\vec{v}| \sin\theta$, we have $j = x_0 v_0$.

b) Replacing x_0 with j/v_0, we have

$$\varepsilon = \frac{1}{2}v_0^2 - \frac{v_0 GM}{j} \qquad \text{or} \qquad v_0^2 - \frac{2GM}{j}v_0 - 2\varepsilon = 0.$$

Solving this quadratic, we obtain $v_0 = \frac{GM}{j} \pm \sqrt{\left(\frac{GM}{j}\right)^2 + 2\varepsilon}$.

Since the planet's initial position is at perihelion (closest point to the star) and $x_0 = j/v_0$, we pick the root with the + sign, and

$$x_0 = \frac{j^2}{GM + \sqrt{(GM)^2 + 2\varepsilon j^2}}.$$

5. a) $|\vec{r} \times \vec{v}| = j$, $|\vec{r}||\vec{v}| \sin\theta = j$, $rv\sin\theta = j$. Using $v_\theta = r\frac{d\theta}{dt}$, we have $r^2\frac{d\theta}{dt} = j$, and rearranging gives the desired result $\frac{d\theta}{dt} = \frac{j}{r^2}$.

b) Using the chain rule, we have $\frac{dr}{dt} = \frac{dr}{d\theta}\frac{d\theta}{dt}$. Using the result of **(a)**, we can replace $\frac{d\theta}{dt}$ with $\frac{j}{r^2}$ to obtain $\frac{dr}{dt} = \frac{j}{r^2}\frac{dr}{d\theta}$.

c) $\varepsilon = \frac{v^2}{2} + \Psi$, $\frac{v^2}{2} = \varepsilon - \Psi$, $\frac{1}{2}v^2 = \varepsilon - \Psi$, $\frac{1}{2}(v_r^2 + v_\theta^2) = \varepsilon - \Psi$. Using $v_\theta = r\frac{d\theta}{dt}$ and $v_r = \frac{dr}{dt}$, we obtain $\frac{1}{2}\left[\left(\frac{dr}{dt}\right)^2 + \left(r\frac{d\theta}{dt}\right)^2\right] = \varepsilon - \Psi$.

d) Starting with the result of **(c)** and replacing $\frac{d\theta}{dt}$ with $\frac{j}{r^2}$ and $\frac{dr}{dt}$ with $\frac{j}{r^2}\frac{dr}{d\theta}$, we obtain $\frac{1}{2}\left[\left(\frac{j}{r^2}\frac{dr}{d\theta}\right)^2 + \left(r\frac{j}{r^2}\right)^2\right] = \varepsilon - \Psi$. Now solving for $\frac{d\theta}{dr}$, we produce the desired result, $\frac{d\theta}{dr} = \frac{1}{r}\left[\frac{2(\varepsilon - \Psi)r^2}{j^2} - 1\right]^{-1/2}$.

e) Use the result of **(d)** and the fact that $\frac{d\theta}{dr} = \frac{1}{\frac{dr}{d\theta}}$.

6. Approximately 75 years.

7. $$\int_0^r 4\pi r'^2 \frac{\rho_0}{\frac{r'}{r_s}\left(1+\frac{r'}{r_s}\right)^3}\,dr' = 4\pi\rho_0 r_s^4 \int_0^r \frac{r'}{(r_s+r')^3}\,dr'.$$

Letting $w = r_s + r'$, the integral can be written as

$$\begin{aligned}
4\pi\rho_0 r_s^4 \int_{r_s}^{r_s+r} (w^{-2} - r_s w^{-3})\,dw &= 4\pi\rho_0 r_s^4 \left(-\frac{1}{w} + \frac{r_s}{2w^2}\right)\Bigg|_{r_s}^{r_s+r} \\
&= 4\pi\rho_0 r_s^4 \left(\frac{-2w + r_s}{2w^2}\right)\Bigg|_{r_s}^{r_s+r} \\
&= 4\pi\rho_0 r_s^4 \left[\frac{-2(r_s+r)+r_s}{2(r_s+r)^2} - \frac{(-2r_s+r_s)}{2r_s^2}\right] \\
&= 2\pi\rho_0 r_s^3 \left[\frac{r^2}{(r_s+r)^2}\right] \\
&= \frac{2\pi\rho_0 r_s r^2}{\left(1+\frac{r}{r_s}\right)^2}.
\end{aligned}$$

8. a) Use the identity $\nabla\cdot(f\vec{F}) = f\nabla\cdot\vec{F} + \vec{F}\cdot\nabla f$, where f is a scalar.

b) Letting $\vec{r} = \langle x, y, z\rangle$ and $\vec{r}' = \langle x', y', z'\rangle$, we have

$$\begin{aligned}
\vec{\nabla}_{\vec{r}}\left(\frac{1}{|\vec{r}'-\vec{r}|^3}\right) &= \vec{\nabla}_{\vec{r}}\left(\frac{1}{\left((x'-x)^2+(y'-y)^2+(z'-z)^2\right)^{3/2}}\right) \\
&= \frac{3}{\left((x'-x)^2+(y'-y)^2+(z'-z)^2\right)^{5/2}} \\
&\qquad \times \langle x'-x,\ y'-y,\ z'-z\rangle \\
&= \frac{3(\vec{r}'-\vec{r})}{|\vec{r}'-\vec{r}|^5}.
\end{aligned}$$

c) $\vec{\nabla}_{\vec{r}}\cdot(\vec{r}'-\vec{r}) = \vec{\nabla}_{\vec{r}}\cdot\langle x'-x,\ y'-y,\ z'-z\rangle = -1-1-1 = -3.$

d)

$$\vec{\nabla}_{\vec{r}}\cdot\frac{(\vec{r}'-\vec{r})}{|\vec{r}'-\vec{r}|^3} = \frac{1}{|\vec{r}'-\vec{r}|^3}\left[\vec{\nabla}_{\vec{r}}\cdot(\vec{r}'-\vec{r})\right] + (\vec{r}'-\vec{r})\cdot\vec{\nabla}_{\vec{r}}\left[\frac{1}{|\vec{r}'-\vec{r}|^3}\right]$$

$$= \frac{1}{|\vec{r}\,' - \vec{r}|^3}(-3) + (\vec{r}\,' - \vec{r}) \cdot \frac{3(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^5}$$

$$= \frac{-3}{|\vec{r}\,' - \vec{r}|^3} + \frac{3|\vec{r}\,' - \vec{r}|^2}{|\vec{r}\,' - \vec{r}|^5} = 0.$$

9.
$$\vec{\nabla}_{\vec{r}} \cdot \frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3} = \left\langle \frac{\partial}{\partial x}\left(\frac{x' - x}{\left((x' - x)^2 + (y' - y)^2 + (z' - z)^2\right)^{3/2}}\right), \right.$$
$$\frac{\partial}{\partial y}\left(\frac{y' - y}{\left((x' - x)^2 + (y' - y)^2 + (z' - z)^2\right)^{3/2}}\right),$$
$$\left. \frac{\partial}{\partial z}\left(\frac{z' - z}{\left((x' - x)^2 + (y' - y)^2 + (z' - z)^2\right)^{3/2}}\right)\right\rangle$$
$$= -\left\langle \frac{\partial}{\partial x'}\left(\frac{x' - x}{\left((x' - x)^2 + (y' - y)^2 + (z' - z)^2\right)^{3/2}}\right), \right.$$
$$\frac{\partial}{\partial y'}\left(\frac{y' - y}{\left((x' - x)^2 + (y' - y)^2 + (z' - z)^2\right)^{3/2}}\right),$$
$$\left. \frac{\partial}{\partial z'}\left(\frac{z' - z}{\left((x' - x)^2 + (y' - y)^2 + (z' - z)^2\right)^{3/2}}\right)\right\rangle$$
$$= -\vec{\nabla}_{\vec{r}\,'} \cdot \frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3}.$$

10. Centering the small sphere of radius h at the origin and letting $\vec{r}$ replace $\vec{r}\,' - \vec{r}$ allows us to rewrite

$$\int\int_{|\vec{r}\,' - \vec{r}| = h} \frac{(\vec{r}\,' - \vec{r})}{|\vec{r}\,' - \vec{r}|^3} \cdot d\vec{S}\,' \qquad \text{as} \qquad \int\int_{|\vec{r}| = h} \frac{\vec{r}}{|\vec{r}|^3} \cdot d\vec{S}\,'$$

without loss of generality. Using the parameterization

$$\vec{r} = \langle h\sin\phi\cos\theta,\ h\sin\phi\sin\theta,\ h\cos\phi\rangle, \qquad 0 \le \phi \le \pi, \quad 0 \le \theta \le 2\pi,$$

and normal vector $\vec{n} = h^2\sin\phi\,\langle\sin\phi\cos\theta,\ \sin\phi\sin\theta,\ \cos\phi\rangle$, we have

$$\int\int_{|\vec{r}| = h} \frac{\vec{r}}{|\vec{r}|^3} \cdot d\vec{S} = \int_0^{\pi}\int_0^{2\pi} \frac{1}{h^3}\langle h\sin\phi\cos\theta, h\sin\phi\sin\theta, h\cos\phi\rangle$$
$$\cdot\, h^2\sin\phi\langle\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\theta\rangle\, d\theta\, d\phi$$

$$= \int_0^{\pi} \int_0^{2\pi} \sin\phi \, d\theta \, d\phi = 2\pi \int_0^{\pi} \sin\phi \, d\phi = 4\pi.$$

Therefore, $-G\,\rho(\vec{r}) \int \int_{|\vec{r}\,'-\vec{r}\,|=h} \dfrac{(\vec{r}\,'-\vec{r})}{|\vec{r}\,'-\vec{r}\,|^3} \cdot d\vec{S} \; = -G\rho(\vec{r})(4\pi).$

11. $-\vec{\nabla}\Psi(\vec{r}) = G \int \int \int_V \rho(\vec{r}\,') \vec{\nabla}_{\vec{r}} \left(\dfrac{1}{|\vec{r}\,'-\vec{r}\,|} \right) dV$

$$= G \int \int \int_V \rho(\vec{r}\,') \vec{\nabla}_{\vec{r}} (((x'-x)^2 + (y'-y)^2 + (z'-z)^2)^{-1/2}) \, dV$$

$$= G \int \int \int_V \rho(\vec{r}\,') \left\langle \frac{\partial}{\partial x} \left(\frac{1}{\left((x'-x)^2 + (y'-y)^2 + (z'-z)^2\right)^{1/2}} \right), \right.$$

$$\frac{\partial}{\partial y} \left(\frac{1}{\left((x'-x)^2 + (y'-y)^2 + (z'-z)^2\right)^{1/2}} \right),$$

$$\left. \frac{\partial}{\partial z} \left(\frac{1}{\left((x'-x)^2 + (y'-y)^2 + (z'-z)^2\right)^{1/2}} \right) \right\rangle dV$$

$$= G \int \int \int_V \rho(\vec{r}\,') \, \frac{\langle x'-x, y'-y, z'-z \rangle}{\left((x'-x)^2 + (y'-y)^2 + (z'-z)^2\right)^{3/2}} \, dV$$

$$= G \int \int \int_V \rho(\vec{r}\,') \, \frac{1}{|(\vec{r}\,'-\vec{r})|^3} \, (\vec{r}\,'-\vec{r}) \, dV$$

12. Note that

$$\frac{d\Psi}{dr} = \frac{\Psi_0 r_s}{(r_s+r)^2}, \qquad \frac{d^2\Psi}{dr^2} = \frac{-2\Psi_0 r_s}{(r_s+r)^3}, \qquad \text{and} \qquad \rho = \frac{\rho_0 r_s^4}{r(r_s+r)^3}.$$

Plugging these quantities into **(29)** yields

$$\frac{-2\Psi_0 r_s}{(r_s+r)^3} + \frac{2}{r} \frac{\Psi_0 r_s}{(r_s+r)^2} = 4\pi G \, \frac{\rho_0 r_s^4}{r(r_s+r)^3}$$

or

$$4\pi G \, \frac{\rho_0 r_s^4}{r(r_s+r)^3} = 4\pi G \, \frac{\rho_0 r_s^4}{r(r_s+r)^3},$$

where we have used the fact that $\Psi_0 = 2\pi G r_s^2 \rho_0$.

13. a) Plugging in $\Psi = r^n$, we have $n(n-1)r^n + 2nr^n = 0$ or $n = 0, -1$. Therefore, the general solution for the homogeneous equation is $\Psi = c_1 + c_2 r^{-1}$.

b) Assume that the particular solution is of the form

$$\Psi_p = u_1(r) + u_2(r)r^{-1}.$$

Then we obtain a linear system of equations to solve for the unknowns $u_1{}'$ and $u_2{}'$:

$$-\frac{1}{r^2}\,u_2{}' = \frac{4\pi G\rho_0 r_s^4}{r(r_s+r)^3}, \qquad u_1{}' + \frac{1}{r}\,u_2{}' = 0.$$

Solving, we have $\quad u_1{}' = \dfrac{4\pi G\rho_0 r_s^4}{(r_s+r)^3}, \quad u_2{}' = \dfrac{-4\pi G\rho_0 r r_s^4}{(r_s+r)^3}.$

Integrating, we find $\quad u_1 = \dfrac{-2\pi G\rho_0 r_s^4}{(r+r_s)^2}, \quad u_2 = \dfrac{2\pi G\rho_0 r_s^4(2r+r_s)}{(r+r_s)^2}.$

Then $\Psi_p = \dfrac{2\pi G\rho_0 r_s^4}{r(r+r_s)}$.

c) $\Psi = c_1 + c_2\dfrac{1}{r} + \dfrac{2\pi G\rho_0 r_s^4}{r(r+r_s)}$.

d) Imposing $\lim\limits_{r\to\infty}\Psi = 0$ means that we must set $c_1 = 0$. Imposing $\lim\limits_{r\to 0}\Psi = c$ where c represents a nonzero constant value means that we must set $c_2 = -2\pi G\rho_0 r_s^3$, so that $\Psi = \dfrac{-2\pi G\rho_0 r_s^3}{r(r+r_s)}$, or $\Psi = -\dfrac{\Psi_0}{1+\xi}$, where $\Psi_0 = 2\pi G\rho_0 r_s^2$.

14. Since $\bar{x} = x/r_s$, we have $\bar{x}_0 = x_0/r_s$. Using the fact that

$$\frac{d}{dt} = \frac{d}{d\bar{t}}\frac{d\bar{t}}{dt} = \frac{1}{\tau_0}\frac{d}{d\bar{t}},$$

we have

$$\vec{v} = \left\langle \frac{dx}{dt}, \frac{dy}{dt} \right\rangle = \left\langle \frac{r_s}{\tau_0}\frac{d\bar{x}}{d\bar{t}}, \frac{r_s}{\tau_0}\frac{d\bar{y}}{d\bar{t}} \right\rangle = \frac{r_s}{\tau_0}\,\vec{\bar{v}}.$$

Then $v_0 = \dfrac{dy(0)}{d\bar{t}} = \dfrac{r_s}{\tau_0}\,\bar{v}_0$. Also, note that $\Psi_0 = \dfrac{r_s^2}{2\tau_0^2}$.
Substituting **(30)** into **(8)**, we have

$$\epsilon = \frac{\frac{1}{2}v_0^2 - \dfrac{\Psi_0}{1+\bar{x}_0}}{\Psi_0} = \left(\frac{v_0\tau_0}{r_s}\right)^2 - \frac{1}{1+\bar{x}_0} = \bar{v}_0^2 - \frac{1}{1+\bar{x}_0}.$$

Substituting into the expression for q, we have

$$q = \frac{j^2}{2\Psi_0 r_s^2} = \frac{(x_0 v_0)^2}{2\Psi_0 r_s^2} = \left(\frac{x_0}{r_s}\right)^2 \frac{v_0^2}{2\Psi_0} = \bar{x}_0^2 \frac{v_0^2}{2\Psi_0} = \bar{x}_0^2 \left(\frac{v_0 \tau_0}{r_s}\right)^2 = \bar{x}_0^2 \bar{v}_0^2.$$

15. Using the fact that $\dfrac{d}{dr} = \dfrac{d}{d\xi}\dfrac{d\xi}{dr} = \dfrac{d}{d\xi}\dfrac{1}{r_s}$, we have $\dfrac{d\theta}{dr} = \dfrac{1}{r_s}\dfrac{d\theta}{d\xi}$. Equation **(10)** then becomes

$$\frac{1}{r_s}\frac{d\theta}{d\xi} = \frac{1}{r}\left[\frac{2(\varepsilon - \Psi)r^2}{j^2} - 1\right]^{-1/2},$$

$$\frac{r}{r_s}\frac{d\theta}{d\xi} = \left[\frac{2(\varepsilon - \Psi)r^2}{j^2} - 1\right]^{-1/2},$$

$$\xi\frac{d\theta}{d\xi} = \left[\frac{2(\varepsilon - \Psi)r^2}{j^2} - 1\right]^{-1/2},$$

$$= \left[\frac{2\left(\Psi_0\epsilon + \frac{\Psi_0}{1+\xi}\right)r^2}{2q\Psi_0 r_s^2} - 1\right]^{-1/2}$$

$$= \left[\frac{\left(\epsilon + \frac{1}{1+\xi}\right)r^2}{q r_s^2} - 1\right]^{-1/2}$$

$$= \left[\left(\epsilon + \frac{1}{1+\xi}\right)\frac{\xi^2}{q} - 1\right]^{-1/2}.$$

16.

$$\left(\epsilon + \frac{1}{1+\xi}\right)\frac{\xi^2}{q} - 1 = 0,$$

$$\frac{\epsilon\xi^2}{q} + \xi^2(1+\xi)q - 1 = 0,$$

$$(1+\xi)\epsilon\xi^2 + \xi^2 - (1+\xi)q = 0,$$

$$\epsilon\xi^3 + \xi^2(1+\epsilon) - q\xi - q = 0.$$

Hence, $f(\xi) = \epsilon\xi^3 + \xi^2(1+\epsilon) - q\xi - q$.

17. Using $\dfrac{d^2x}{dt^2} = \dfrac{r_s}{\tau_0^2}\dfrac{d^2\bar{x}}{d\bar{t}^2}, \qquad \dfrac{d^2y}{dt^2} = \dfrac{r_s}{\tau_0^2}\dfrac{d^2\bar{y}}{d\bar{t}^2},$

equations **(15)** become

$$\frac{r_s}{\tau_0^2}\frac{d^2\bar{x}}{d\bar{t}^2}+\frac{\partial\Psi}{\partial x}=0, \qquad \frac{r_s}{\tau_0^2}\frac{d^2\bar{y}}{d\bar{t}^2}+\frac{\partial\Psi}{\partial y}=0.$$

Using $$\frac{\partial}{\partial x}=\frac{\partial}{\partial\bar{x}}\frac{\partial\bar{x}}{\partial x}=\frac{1}{r_s}\frac{\partial}{\partial\bar{x}}, \qquad \frac{\partial}{\partial y}=\frac{\partial}{\partial\bar{y}}\frac{\partial\bar{y}}{\partial y}=\frac{1}{r_s}\frac{\partial}{\partial\bar{y}},$$

these equations can be rewritten as

$$\frac{r_s}{\tau_0^2}\frac{d^2\bar{x}}{d\bar{t}^2}+\frac{1}{r_s}\frac{\partial\Psi}{\partial\bar{x}}=0, \qquad \frac{r_s}{\tau_0^2}\frac{d^2\bar{y}}{d\bar{t}^2}+\frac{1}{r_s}\frac{\partial\Psi}{\partial\bar{y}}=0.$$

Substituting in for Ψ, we have

$$\frac{r_s}{\tau_0^2}\frac{d^2\bar{x}}{d\bar{t}^2}-\frac{1}{r_s}\frac{\partial}{\partial\bar{x}}\left(\frac{\Psi_0}{1+\xi}\right)=0, \qquad \frac{r_s}{\tau_0^2}\frac{d^2\bar{y}}{d\bar{t}^2}-\frac{1}{r_s}\frac{\partial}{\partial\bar{y}}\left(\frac{\Psi_0}{1+\xi}\right)=0.$$

Finally, remembering that $\Psi_0=\dfrac{r_s^2}{2\tau_0^2}$, we have

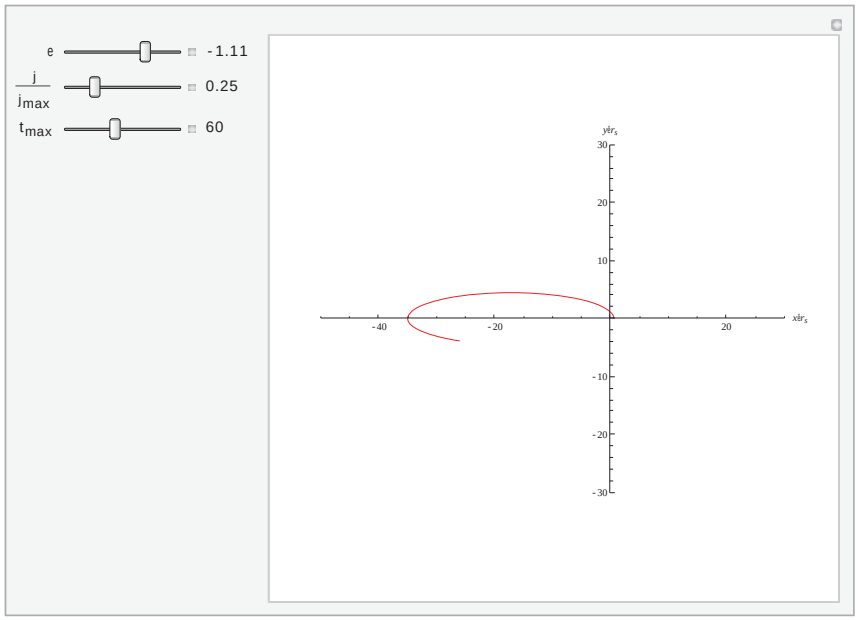
$$\frac{d^2\bar{x}}{d\bar{t}^2}-\frac{1}{2}\frac{\partial}{\partial\bar{x}}\left(\frac{1}{1+\xi}\right)=0, \qquad \frac{d^2\bar{y}}{d\bar{t}^2}-\frac{1}{2}\frac{\partial}{\partial\bar{y}}\left(\frac{1}{1+\xi}\right)=0.$$

Acknowledgments

The author would like to acknowledge the efforts of Christopher Whitaker, an undergraduate student at Northern Kentucky University. Chris's careful work in checking exercise solutions, reading the manuscript, and making figures contributed greatly to the final version of this Module. The author would also like to thank Fred Adams and Marco Fatuzzo for useful discussions.

Appendix A: Elliptic Orbits

```
Clear@x, y, v0, x0D;

func@eps_, jR_, tmax_D := Module@8e = eps, jRatio = jR, T = tmax<,

  jMax = Sqrt@-8 * Pi^4 ê e D;
  j = jMax * jRatio;

  x0 = j^2 ê H4 Pi^2 + Sqrt@16 Pi^4 + 2 e  j^2DL;
  v0 = 4 Pi^2 ê j + Sqrt@16 Pi^4 ê Hj^2L + 2 e D;

  solution = NDSolve@8x''@tD ã - 4 Pi^2 x@tD ê HSqrt@x@tD^2 + y@tD^2DL^3,
     y''@tD ã -4 Pi^2 y@tD ê HSqrt@x@tD^2 + y@tD^2DL^3, x@0D ã x0, x'@0D ã 0, y@0D ã 0, y'@0D ã v0<,
    8x@tD, y@tD<, 8t, 0, T<D;

  ParametricPlot@Evaluate@8x@tD, y@tD< ê. solutionD, 8t, 0, T<, PlotRange Ø 88-50, 30<, 8-30, 30<<,
   PlotStyle Ø Red, AxesLabel Ø 8Style@"xêr_s", ItalicD, Style@"yêr_s", ItalicD<, ImageSize Ø 8520, 520<,
   ImagePadding Ø 30D
 D

ManipulateB
 QuietüfuncDa, b, cD,
 88a, -1.11, "e"<, -2, -0.8, .1, Appearance Ø "Labeled", ImageSize Ø Tiny<,
 ::b, .25, "j/j_max">, 0.05, .99, .01, Appearance Ø "Labeled", ImageSize Ø Tiny>,
 88c, 60, "t_max"<, 1, 140, 1, Appearance Ø "Labeled", ImageSize Ø Tiny<,
 ControlPlacement Ø Left, SaveDefinitions Ø True, SynchronousUpdating Ø FalseF
```

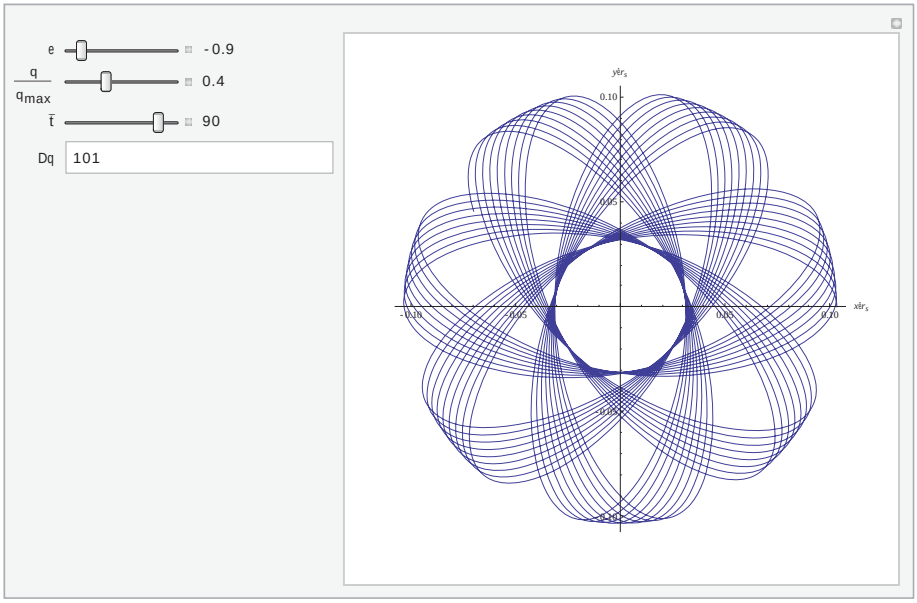


Files of the Mathematica programs of **Appendix A** and **Appendix B** are available at the supplements page for *The UMAP Journal*:

http://www.comap.com/product/periodicals/supplements.html

Appendix B: Spirographic Orbits

```
func@eps_, qR_, T_D := Module@8ep = eps, qRatio = qR, t = T<,

  qMax = H1 + Sqrt@1 - 8 * epD + 4 * epL^3 ê H-8 * ep * H1 + Sqrt@1 - 8 * epDL^2L;
  q = qMax * qRatio;

  d = 0.00001;
  x2 = xxbar0 ê. NSolve@ep ã -H1 ê H1 + xxbar0L - q ê xxbar0^2L, xxbar0D@@1DD;
  x1 = xxbar0 ê. NSolve@ep ã -H1 ê H1 + xxbar0L - q ê xxbar0^2L, xxbar0D@@2DD;
  Dq = Round@Sqrt@qD * NIntegrate@1 ê Hx * Sqrt@ep * x^2 + x^2 ê H1 + xL - qDL, 8x, x1 + d, x2 - d<D * 180 ê PiD;
  xBar0 = x2 ;
  vBar0 = Sqrt@qD ê xBar0;

  solution =
   NDSolve@8xbar''@tbarD ã -H1 ê 2L * D@-1 ê H1 + Sqrt@xbar@tbarD^2 + ybar@tbarD^2DL, xbar@tbarDD,
     ybar''@tbarD ã -H1 ê 2L * D@-1 ê H1 + Sqrt@xbar@tbarD^2 + ybar@tbarD^2DL, ybar@tbarDD,
     xbar@0D ã xBar0,
     xbar'@0D ã 0,
     ybar@0D ã 0,
     ybar'@0D ã vBar0<,
    8xbar@tbarD, ybar@tbarD<,
    8tbar, 0, t<
   D;

  ParametricPlot@Evaluate@8xbar@tbarD, ybar@tbarD< ê. solutionD, 8tbar, 0, t<, PlotRange Ø All,
   AxesLabel Ø 8Style@"xêrs", ItalicD, Style@"yêrs", ItalicD<, ImageSize Ø 8520, 520<, ImagePadding Ø 30D
 D

ManipulateB
 QuietüfuncDa, b, cD,
 88a, -.9, "e"<, -.96, -.1, .01, Appearance Ø "Labeled", ImageSize Ø Tiny<,
 ::b, .4, "q/qmax">, 0.1, 1, .01, Appearance Ø "Labeled", ImageSize Ø Tiny>,
 99c, 90, "t̄"=, 1, 100, 1, Appearance Ø "Labeled", ImageSize Ø Tiny=,
 8Dq<,
 ControlPlacement Ø Left, SaveDefinitions Ø True, SynchronousUpdating Ø FalseF
```



References

Adams, Fred C., and Anthony M. Bloch. 2005. Orbits in extended mass distributions: General results and the spirographic approximation. *The Astrophysical Journal* 629: 204–218. doi:10.1086/431455, http://iopscience.iop.org/0004-637X/629/1/204/pdf/0004-637X_629_1_204.pdf .

Adams, Fred C., Eva M. Proszkow, Marco Fatuzzo, and Philip C. Myers. 2006. Early evolution of stellar groups and clusters: Environmental effects on forming planetary systems. *The Astrophysical Journal* 641: 504–525. doi:10.1086/500393, http://iopscience.iop.org/0004-637X/641/1/504/pdf/0004-637X_641_1_504.pdf .

Binney, J., and Scott Tremaine. 1987. *Galactic Dynamics.* Princeton, NJ: Princeton University Press.

Carroll, Bradley, and Dale Ostlie. 2006. *Introduction to Modern Astrophysics.* 2nd ed. New York: Benjamin Cummings.

Fatuzzo, Marco, and Fred C. Adams. 2008. UV radiation fields produced by young embedded star clusters. *The Astrophysical Journal* 675 (2): 1361–1374. doi:10.1086/527469, http://iopscience.iop.org/0004-637X/675/2/1361/pdf/0004-637X_675_2_1361.pdf .

Hernquist, Lars. 1990. An analytical model for spherical galaxies and bulges. *The Astrophysical Journal* 356: 359–364. http://adsabs.harvard.edu/full/1990ApJ...356..359H.

Holden, Lisa, Edward Landis, Jeremy Spitzig, and Fred C. Adams. 2011. An investigation of the loss of planet-forming potential in intermediate sized young embedded star clusters. *Publication of the Astronomical Society of the Pacific* 123 (No. 899): 14–25. doi:10.1086/658081, http://www.jstor.org/stable/10.1086/658081.

Landis, Ted, Jeremy Spitzig, and Lisa Holden. n.d. Spirographic orbit of a star in extended mass distribution. Wolfram Demonstrations Project. http://demonstrations.wolfram.com/SpirographicOrbitOfAStarInExtendedMassDistribution/.

About the Author

Lisa Holden is an Assistant Professor of Mathematics at Northern Kentucky University. She received her Ph.D. in applied mathematics from Northwestern University and is currently interested in problems relating to astronomy.