Index to Volume 69

Authors

Andersen, K.M., *A Characterization of Polynomials*, 137–142

Ash, J. Marshall, *The Limit of* $x^{x^{\cdot^{\cdot^{\cdot^{x}}}}}$ *as* x *Tends to Zero*, 207–209

Bailey, D.F., *Counting Arrangements of* 1*'s and* −1*'s*, 128–131

Beauregard, Raymond A. and Suryanarayan, E.R., *Proof Without Words: Parametric Representation of Primitive Pythagorean Triples*, 189

Bechhoefer, John, *The Birth of Period 3, Revisited*, 115–118

Benjamin, Arthur T., see McCarthy, Clifford A.

Beyer, W.A.; Louck, J.D.; and Zeilberger, D., *Math Bite: A Generalization of a Curiosity that Feynman Remembered All His Life*, 43–44

Blom, Gunnar; Englund, Jan-Eric; and Sandell, Dennis, *General Russian Roulette*, 293–297

Bloom, David M., *Probabilities of Clumps in a Binary Sequence (and How to Evaluate Them Without Knowing a Lot)*, 366–372

Bloy, Greg, *Goldbach's Problem in Matrix Rings*, 136–137

Bradley, Michael J., *Building Home Plate: Field of Dreams or Reality?*, 44–45

Burgiel, H.; Franzblau, D.S.; and Gutschera, K.R., *The Mystery of the Linked Triangles*, 94–102

Byrne, Philip J. and Hesse, Robert, *A Markov Chain Analysis of the Game of Jai Alai*, 279–283

Chapman, Robin, *A Polynomial Taking Integer Values*, 121

Chilaka, James O. *Proof Without Words: Decomposing the Combination* $\binom{kn}{2}$, 127

——, *Proof Without Words:* $\sum_{k=1}^{n} k(k+1)(k+2) = n(n+1)(n+2)(n+3)/4$, 63

——, *Proof Without Words: An Alternating Series*, 355–356

DeTemple, Duane and Harold, Sonia, A Round-Up of Square Problems, 15–27

Devlin, Keith, *Good-bye Descartes?*, 344–349

Drucker, Daniel and Locke, Phil, *A Natural Classification of Curves and Surfaces With Reflection Properties*, 249–256

Englund, Jan-Eric, see Blom, Gunnar

Falkowski, Bernd-Jürgen, *On the Convergence of Hillam's Iteration Scheme*, 299–302

Farris, Frank A., *Wheels on Wheels on Wheels—Surprising Symmetry*, 185–189

Fjelstad, Paul, *Extending the Pythagorean Theorem to Other Geometries*, 222–223

Francis, Richard L., *Math Bite: Recitation of Large Primes*, 260

Franzblau, D.S., see Burgiel, H.

Gadbois, Steve, *Poker With Wild Cards—A Paradox?*, 283–285

Gass, Saul I., *On Copying a Compact Disk to Cassette Tape: An Integer-Programming Approach*, 57–61

Goldfeather, Jack, *Using Quadratic Forms to Correct Orientation Errors in Tracking*, 110–114

Golomb, Solomon W., *A Symmetry Criterion for Conjugacy in Finite Groups*, 373–375

Gordon, William B., *Period Three Trajectories of the Logistic Map*, 118–120

Gutschera, K.R., see Burgiel, H.

Harold, Sonia, see DeTemple, Duane

Harper, James D., *The Golden Ratio is Less Than* $\pi^2/6$, 266

Hesse, Robert, see Byrne, Philip J.

Hirschhorn, M.D., *A Proof in the Spirit of Zeilberger of an Amazing Identity of Ramanujan*, 267–269

Hogan, Guy T., *More on the Converse of Lagrange's Theorem*, 375–376

Hungerbühler, Norbert, *Proof of a Conjecture of Lewis Carroll*, 182–184

Jovanović, Milan V. and Jungić, Veselin M., *Algebraic Set Operations, Multifunctions, and Indefinite Integrals*, 350–354

Jungić, Veselin M., see Jovanović, Milan V.

Jungnickel, Dieter, *On the Order of a Product in a Finite Abelian Group*, 53–57

Kaminsky, Kenneth, *Professor Fogelfroe*, 62, 142, 209, 303, 383

Klamkin, Murray S. and Kung, Sidney H.,

Ceva's and Menelaus' Theorems and Their Converses via Centroids, 49–51

Knudsen, Finn F. and Skau, Ivar, *On the Asymptotic Solution of a Card-Matching Problem*, 190–197

Kortram, R.A., *Simple Proofs for* $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$ *and* $\sin x = x \prod_{k=1}^{\infty} \left(1 - \frac{x^2}{k^2\pi^2}\right)$, 122–125

Krause, Eugene F., *Maximizing the Product of Summands; Minimizing the Sum of Factors*, 270–278

Krusemeyer, Mark, *A Parenthetical Note (to a Paper of Guy)*, 257–260

Kung, Sidney H., *Proof Without Words: The Difference-Product Identities*, 278

——, *Proof Without Words: The Sum-Product Identities*, 269

——, see Klamkin, Murray S.

Lazebnik, Felix, *On Systems of Linear Diophantine Equations*, 261–266

Ledder, Glenn, *A Simple Introduction to Integral Equations*, 172–181

Locke, Phil, see Drucker, Daniel

Lofaro, Thomas, see Newton, Tyre

Louck, J.D., see Beyer, W.A.

McCarthy, Clifford A. and Benjamin, Arthur T., *Determinants of the Tournaments*, 133–135

Mackiw, George, *Finite Groups of* 2×2 *Integer Matrices*, 356–361

Meijer, A.R., *Groups, Factoring, and Cryptography*, 103–109

Meyers, Leroy F., *Update on William Wernick's "Triangle Constructions with Three Located Points,"* 46–49

Meyerson, Mark D., *The* x^x *Spindle*, 198–206

Milanfar, Peyman, *A Persian Folk Method of Figuring Interest*, 376

Mulcahy, Colm, *Plotting and Scheming with Wavelets*, 323–343

Muldoon, Martin E. and Ungar, Abraham A., *Beyond Sin and Cos*, 3–14

Nelsen, Roger B., *Proof Without Words: Bernoulli's Inequality (Two Proofs)*, 197

——, *Proof without Words: Five Means—and Their Means*, 64–65

Newton, Tyre and Lofaro, Thomas, *On Using Flows to Visualize Functions of a Complex Variable*, 28–34

Noy, Marc, *A Short Solution of a Problem in Combinatorial Geometry*, 52–53

Papanicolaou, Vassilis G., *On the Asymptotic Stability of a Class of Linear Difference Equations*, 34–43

Pinkernell, Guido M., *Identities on Point-Line Figures in the Euclidean Plane*, 377–382

Ramnath, Sarnath and Scully, Daniel, *Moving Card i to Position j with Perfect Shuffles*, 361–365

Reiter, Ashley, see Reiter, Harold

Reiter, Harold and Reiter, Ashley, *The Space of Closed Subsets of a Convergent Sequence*, 217–221

Robertson, John P., *Magic Squares of Squares*, 289–293

Sakmar, I.A., *Chanson Sans Paroles*, 62

Sandefur, James T., *The Baseball-Card Collector's Query*, 243–248

Sandell, Dennis, see Blom, Gunnar

Schaer, J., *Parallels on the Sphere*, 298

Schaumberger, N., *Power Mean for Zero Exponent*, 216

Scully, Daniel, see Ramnath, Sarnath

Skau, Ivar, see Knudsen, Finn F.

Spain, P.G., *The Fermat Point of a Triangle*, 131–133

Stangl, Walter D. *Counting Squares in* $\mathbb{Z}_n$, 285–289

Staring, Mike, *The Pythagorean Proposition: A Proof by Means of Calculus*, 45–46

Stein, S.K., *Inverse Problems for Central Forces*, 83–93

Suryanarayan, E.R., see Beauregard, Raymond A.

Tan, Lin, *The Group of Rational Points on the Unit Circle*, 163–171

Ungar, Abraham A., see Muldoon, Martin E.

Wong, Roman W., *On the Convergence of the Sequence of Powers of a* 2×2 *Matrix*, 210–215

Yuefeng, Feng, *Proof Without Words: Jordan's Inequality* $\frac{2x}{\pi} \leq \sin x \leq x$, 126

Zeilberger, D., see Beyer, W.A.

TITLES

Algebraic Set Operations, Multifunctions, and Indefinite Integrals, *by Milan V. Jo-*

vanović and Veselin M. Jungić, 350–354
Asymptotic Solution of a Card-Matching Problem, On the, *by Finn F. Knudsen and Ivar Skau*, 190–197
Asymptotic Stability of a Class of Linear Difference, Equations, On the, *by Vassilis G. Papanicolaou*, 34–43
Baseball-Card Collector's Query, The, *by James T. Sandefur*, 243–248
Beyond Sin and Cos, *by Martin E. Muldoon and Abraham A. Ungar*, 3–14
Birth of Period 3, Revisited, The, *by John Bechhoefer*, 115–118
Building Home Plate: Field of Dreams or Reality?, *by Michael J. Bradley*, 44–45
Ceva's and Menelaus' Theorems and Their Converses via Centroids, *by Murray S. Klamkin and Sidney H. Kung*, 49–51
Chanson Sans Paroles, *by I.A. Sakmar*, 62
Characterization of Polynomials, A, *by K.M. Andersen*, 137–142
Convergence of Hillam's Iteration Scheme, On the, *by Bernd-Jürgen Falkowski*, 299–302
Convergence of the Sequence of Powers of a 2×2 Matrix, On the, *by Roman W. Wong*, 210–215
Converse of Lagrange's Theorem, More on the, *by Guy T. Hogan*, 375–376
Copying a Compact Disk to Cassette Tape: An Integer-Programming Approach, On, *by Saul I. Gass*, 57–61
Counting Arrangements of 1's and -1's, *by D.F. Bailey*, 128–131
Counting Squares in $\mathbb{Z}_n$, *by Walter D. Stangl*, 285–289
Determinants of the Tournaments, *by Clifford A. McCarthy and Arthur T. Benjamin*, 133–135
Extending the Pythagorean Theorem to Other Geometries, *by Paul Fjelstad*, 222–223
Fermat Point of a Triangle, The, *by P.G. Spain*, 131–133
Finite Groups of 2×2 Integer Matrices, *by George Mackiw*, 356–361
General Russian Roulette, *by Gunnar Blom, Jan-Eric Englund, and Dennis Sandell*, 293–297
Goldbach's Problem in Matrix Rings, *by Greg Bloy*, 136–137
Golden Ratio is Less Than $\pi^2/6$, The, *by James D. Harper*, 266
Good-bye Descartes?, *by Keith Devlin*, 344–349
Group of Rational Points on the Unit Circle, The, *by Lin Tan*, 163–171
Groups, Factoring, and Cryptography, *by A.R. Meijer*, 103–109
Identities on Point-Line Figures in the Euclidean Plane, *by Guido M. Pinkernell*, 377–382
Inverse Problems for Central Forces, *by S.K. Stein*, 83–93
Limit of $x^{x^{\cdot^{\cdot^{\cdot^{x}}}}}$ as x Tends to Zero, The, *by J. Marshall Ash*, 207–209
Magic Squares of Squares, *by John P. Robertson*, 289–293
Markov Chain Analysis of the Game of Jai Alai, A, *by Philip J. Byrne and Robert Hesse*, 279–283
Math Bite: A Generalization of a Curiosity that Feynman Remembered All His Life, *by W.A. Beyer, J.D. Louck, and D. Zeilberger*, 43–44
Math Bite: Recitation of Large Primes, *by Richard L. Francis*, 260
Maximizing the Product of Summands; Minimizing the Sum of Factors, *by Eugene F. Krause*, 270–278
Moving Card i to Position j with Perfect Shuffles, *by Sarnath Ramnath and Daniel Scully*, 361–365
Mystery of the Linked Triangles, The, *by H. Burgiel, D.S. Franzblau, and K.R. Gutschera*, 94–102
Natural Classification of Curves and Surfaces With Reflection Properties, A, *by Daniel Drucker and Phil Locke*, 249–256
Order of a Product in a Finite Abelian Group, On the, *by Dieter Jungnickel*, 53–57
Parallels on the Sphere, *by J. Schaer*, 298
Parenthetical Note (to a Paper of Guy), A, *by Mark Krusemeyer*, 257–260
Period Three Trajectories of the Logistic Map, *by William B. Gordon*, 118–120
Persian Folk Method of Figuring Interest, A, *by Peyman Milanfar*, 376
Plotting and Scheming with Wavelets, *by Colm Mulcahy*, 323–343
Poker With Wild Cards—A Paradox?, *by Steve Gadbois*, 283–285
Polynomial Taking Integer Values, A, *by Robin Chapman*, 121
Power Mean for Zero Exponent, *by*

N. Schaumberger, 216

Probabilities of Clumps in a Binary Sequence (and How to Evaluate Them Without Knowing a Lot), *by David M. Bloom*, 366–372

Professor Fogelfroe, *by Kenneth Kaminsky*, 62, 142, 209, 303, 383

Proof in the Spirit of Zeilberger of an Amazing Identity of Ramanujan, A, *by M.D. Hirschhorn*, 267–269

Proof of a Conjecture of Lewis Carroll, *by Norbert Hungerbühler*, 182–184

Proof Without Words: $\sum_{k=1}^{n} k(k+1)(k+2) = n(n+1)(n+2)(n+3)/4$, *by James O. Chilaka*, 63

Proof Without Words: An Alternating Series, *by James O. Chilaka*, 355–356

Proof Without Words: Bernoulli's Inequality (two proofs), *by Roger B. Nelsen*, 197

Proof Without Words: Decomposing the Combination $\binom{kn}{2}$, *by James O. Chilaka*, 127

Proof Without Words: Five Means—and Their Means, *by Roger B. Nelsen*, 64–65

Proof Without Words: Jordan's Inequality $\frac{2x}{\pi} \leq \sin x \leq x$, *by Feng Yuefeng*, 126

Proof Without Words: Parametric Representation of Primitive Pythagorean Triples, *by Raymond A. Beauregard and E.R. Suryanarayan*, 189

Proof Without Words: The Difference-Product Identities, *by Sidney H. Kung*, 278

Proof Without Words: The Sum-Product Identities, *by Sidney H. Kung*, 269

Pythagorean Proposition, The: A Proof by Means of Calculus, *by Mike Staring*, 45–46

Round-Up of Square Problems, A, *by Duane DeTemple and Sonia Harold*, 15–27

Short Solution of a Problem in Combinatorial Geometry, A, *by Marc Noy*, 52–53

Simple Introduction to Integral Equations, A, *by Glenn Ledder*, 172–181

Simple Proofs for $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$ and $\sin x = x\prod_{k=1}^{\infty}\left(1 - \frac{x^2}{k^2\pi^2}\right)$, *by R.A. Kortram*, 122–125

Space of Closed Subsets of a Convergent Sequence, The, *by Harold Reiter and Ashley Reiter*, 217–221

Symmetry Criterion for Conjugacy in Finite Groups, A, *by Solomon W. Golomb*, 373–375

Systems of Linear Diophantine Equations, On, *by Felix Lazebnik*, 261–266

Update on William Wernick's "Triangle Constructions with Three Located Points," *by Leroy F. Meyers*, 46–49

Using Flows to Visualize Functions of a Complex Variable, On, *by Tyre Newton and Thomas Lofaro*, 28–34

Using Quadratic Forms to Correct Orientation Errors in Tracking, *by Jack Goldfeather*, 110–114

Wheels on Wheels on Wheels—Surprising Symmetry, *by Frank A. Farris*, 185–189

x^x Spindle, The, *by Mark D. Meyerson*, 198–206

PROBLEMS

The letters P, Q, *and* S *refer to Proposals, Quickies, and Solutions, respectively; page numbers appear in parentheses. For example, P1489(66) refers to Proposal 1489, which appears on page 66.*

February: P1489–1493; Q844–846; S1464–1468

April: P1494–1498; Q847–849; S1469–1473

June: P1499–1503; Q850–852; S1474–1478

October: P1504–1508; Q853–855; S1479–1483

December: P1509–1513; Q856–858; S1484–1488

Anchorage Math Solutions Group, S1474(225)

Andreoli, Michael H., S1476(227)

Bencze, Mihály, P1497(143)

——, Q852(225)

Binz, J.C., P1498(144)

Byrd, Stan; Smith, Ronald L.; and Muchlis, Ahmad, S1468(72)

Callan, David, P1489(66)

——, P1509(384)

Chao, Wu Wei, P1491(66)

——, P1499(224)

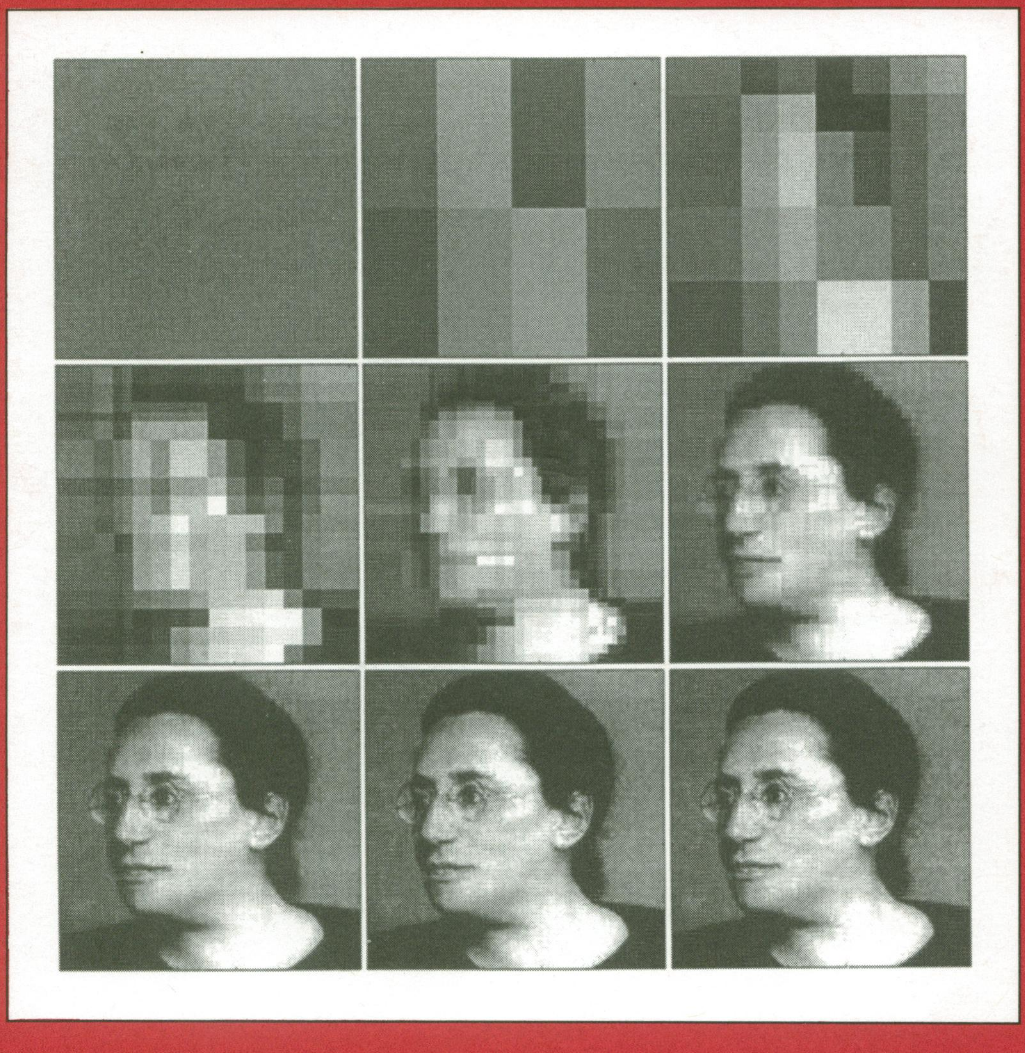
——, P1506(304)

——, Q857(385)

Chapman, Robin, S1466(70)
——, S1472(149)
Christensen, Jens Peter Reus and Larsen, Mogens Esrom, Q855(305)
Con Amore Problem Group, see Plank, Donald
Deutsch, Emeric, P1502(224)
—— and Gessel, Ira M., P1494(143)
Doster, David, S1477(227)
Doucette, Robert L., S1475(226)
Dove, Kevin L., see Sumner, John S.
Eggleton, Roger B. and Leong, Thomas, S1465(69)
Fischer, Ismor, Q847(144)
Foster, L.L., S1484(386)
Freidkin, Evgenii S., P1490(66)
Fukuta, Jiro, P1493(67)
Gessel, Ira M., see Deutsch, Emeric
Golomb, Michael, P1492(67)
Grossman, Jerrold W. and Mellendorf, Stephen, Q856(385)
Harminc, Matúš and Soták, Roman, P1501(224)
Holshouser, Arthur L. and Klein, Benjamin G., P1512(385)
Horrigan, Frank A., S1486(388)
Hoyt, John P., Q846(68)
Just, Erwin, Q844(67)
——, P1504(304)
Kandall, Geoffrey A. and Kappus, Hans, S1469(144)
Kappus, Hans, see Kandall, Geoffrey A.
Karmakar, S.B. and Klamkin, Murray S., Q853(305)
Klamkin, Murray S., P1496(143)
——, Q851(225)
——, Q853(305)
——, Q858(385)
—— and Rousseau, Cecil C., P1505(304)
——, see Karmakar, S.B.
Klein, Benjamin G. see Holshouser, Arthur L.
Koh, Sung Eun, S1467(71)
Kotkowski, Bogdan, S1482(308)
Kutsenok, Victor, S1481(308)
Kwong, Harris, S1470(146)
Larsen, Mogens Esrom, see Christensen, Jens Peter Reus
Larson, Loren C., P1513(385)
Lau, Kee-Wai, S1488(390)
Laugwitz, Detlef, P1510(384)
Leong, Thomas, see Eggleton, Roger B.
Lord, Nick, P1503(225)
Lossers, O.P., see Yang, Yongzhi
Mellendorf, Stephen, see Grossman, Jerrold W.
Morris, Howard, P1507(304)
Muchlis, Ahmad, see Byrd, Stan
Plank, Donald and Con Amore Problem Group, S1478(228)
Rembis, F.C., S1480(307)
Rousseau, Cecil, C., see Klamkin, Murray S.
Sard, Eugene, Q854(305)
Schaumberger, Norman, Q844(67)
Shapiro, Daniel B., Q849(144)
Sinefakopoulos, Achilleas, P1495(143)
——, S1464(68)
Smith, Ronald L., see Byrd, Stan
Soták, Roman, see Harminc, Matúš
Stahl, Saul, P1500(224)
——, P1508(305)
Sumner, John S. and Dove, Kevin L., S1485(386)
Taylor, Catherine, S1483(309)
Wardlaw, William P., Q845(68)
Wee, Hoe Teck and Woltermann, Michael, S1487(389)
White, Homer, Q848(144)
——, Q850(225)
Woltermann, Michael, see Wee, Hoe Teck
Yang, Yongzhi and Lossers, O.P., S1473(149)
Young, Anne L., S1471(148)
Zhu, David, S1479(305)

Vol. 69, No. 5 December 1996

MATHEMATICS MAGAZINE



- Plotting and Scheming with Wavelets
- Good-bye Descartes?

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 69, pp. 78–79, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Send new manuscripts to Paul Zorn, Editor, Department of Mathematics, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

Cover illustration: Eight wavelet-compressed versions (using normalized Haar wavelets) of an image of Emmy Noether; the last picture is the "original." Image k uses 4^{k-1} pieces of data. Prepared by Colm Mulcahy, using *Matlab*.

AUTHORS

Keith Devlin is Dean of the School of Science at Saint Mary's College of California, in Moraga, California, a Senior Researcher at the Center for the Study of Language and Information at Stanford University, and a Consulting Research Professor in the Department of Computing and Information Science at the University of Pittsburgh. This assortment of institutions reflects his interests in science education, mathematics, logic, and applications of mathematics to language and information. He is the current editor of the MAA newsletter *FOCUS* and the author of *Devlin's Angle*, a monthly column on MAA Online.

A native of England, Dr. Devlin obtained his Ph.D. in mathematical logic at the University of Bristol in 1971. Since 1987 he has lived in the United States.

Colm Mulcahy earned his B.Sc. and M.Sc. degrees from University College Dublin, in Ireland. In 1985 he received a Ph.D. from Cornell University, having pursued research in the abstract theory of ordered fields and higher level forms under the guidance of Alex F. T. W. Rosenberg.

Since 1988 he has been in the department of mathematics at Spelman College, where his interests have broadened to include computer-aided geometric design, image processing, computer graphics and wavelets, and algebraic codology.

This paper was written while the author took a break from trying to prove that the real part of his Erdős number is $-1/2$.

Vol. 69, No. 5 December 1996

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, the Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

Second class postage paid at Washington, D.C. and additional mailing offices.

Postmaster: Send address changes to Mathematics Magazine, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036-1385.
PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Plotting and Scheming with Wavelets

Dedicated with great affection to Alex F. T. W. Rosenberg,
on the occasion of his 70th birthday

COLM MULCAHY
Spelman College
Atlanta, GA 30314

1. Introduction

Wavelets are acquiring a visibility and popularity that may soon be on the scale first enjoyed by fractals a few years back. Like fractals, wavelets have attractive and novel features, both as mathematical entities and in numerous applications. They are often touted as worthwhile alternatives to classical Fourier analysis, which works best when applied to periodic data: wavelet methods make no such assumptions. However, the mathematics of wavelets can seem intractable to the novice. Indeed, most introductions to wavelets assume that the reader is already well versed in Fourier techniques.

Our main goal is simple: to convince the reader that at their most basic level, wavelets are fun, easy, and ideal for livening up dull conversations. We demonstrate how elementary linear algebra makes accessible this exciting and relatively new area at the border of pure and applied mathematics.

In **Plotting**, we explore several ways of visually representing data, with the help of *Matlab* software. In **Scheming**, we discuss a simple wavelet-based compression technique, whose generalizations are being used today in signal and image processing, as well as in computer graphics and animation. The basic technique uses only addition, subtraction, and division by two! Only later, in **Wavelets**, do we come clean and reveal what wavelets are, while unveiling the multiresolution setting implicit in any such analysis.

In **Averaging and Differencing with Matrices**, which may be read independently of **Wavelets**, we provide a matrix formulation of the compression scheme. In **Wavelets on the World Wide Web** we mention a natural form of progressive image transmission that lends itself to use by the emerging generation of web browsers (such wavelet-enhanced software is already on the market).

In **Wavelet Details**, we attempt to put everything in context, while hinting at the more sophisticated mathematics that must be mastered if one wishes to delve deeper into the subject. Finally, in **Closing Remarks**, we mention some other common applications of wavelets.

Along the way we find ourselves trying out an adaptive plotting technique for ordinary functions of one variable that differs from those currently employed by many of today's popular computer algebra packages. While this technique, as described here, is limited in its usefulness, it can be modified to produce acceptable results.

We were much inspired by Stollnitz, DeRose, and Salesin's fine wavelets primers (**[15, 16]**), which, along with **[17]**, **[18]**, **[19]**, we recommend heartily to beginners who desire more details. More general surveys can be found in [7] and [11]. Although the wavelets we discuss here had their origins in work of Haar early in this century, the subject proper really gathered momentum only in the last decade. The historical development of wavelets is quite complex, as the main concepts arose independently in several different fields. We do not cite the numerous groundbreaking papers in these fields, leaving that to the books and surveys listed in the bibliography. It's a fascinating story, combining ideas first studied by electrical engineers, physicists, and seismologists, as well as pure mathematicians. For an especially readable account of how it all happened, we recommend Barbara Burke Hubbard's *The World According to Wavelets* [10], a remarkable book which also goes into greater detail about wavelet applications than we do. A more mathematically concise version of this story can be found in Jawerth and Sweldens' survey paper [11].

2. Plotting

We begin by reviewing standard ways of plotting discrete data sets, in particular, sampled functions of the form $y=f(x)$, and two-dimensional digital images. The limitations inherent in attempts to plot functions by uniform sampling will lead us, in the next section, to suggest a wavelet-based scheme to work around this difficulty. The need for adaptive plotting techniques will become obvious. The real purpose of this section is to drum up support for some sort of data compression.

Suppose we have a finite set of planar data points (x, y), which might be samples of a function $y=f(x)$. A common method of displaying these data is to plot the individual points and then join adjacent points with line segments; this is precisely what happens when many computer algebra packages graph functions. Graphing with *Matlab*'s `plot` command, for instance, requires us to pick the x-values to be used. Plotting $y=\sin(15x)$ and $y=\sin(90x)$ this way, on the interval $[0,1]$, using 32 equally spaced x-values, yields the pictures in FIGURE 1. The true nature of $y=\sin(15x)$ can be safely inferred from the first plot, as increasing the number of points sampled verifies. The second plot is another story, however.

FIGURE 1(b) suggests a function whose oscillations exhibit a pulsing pattern, although, symbolically, we expect a horizontally telescoped version of the preceding graph. The apparent pulsing behavior is an artifact of sampling uniformly at 32 points: $\sin(90x)$ has frequency $\frac{90}{2\pi}\approx 14.3239$, which is just under half the sampling frequency.

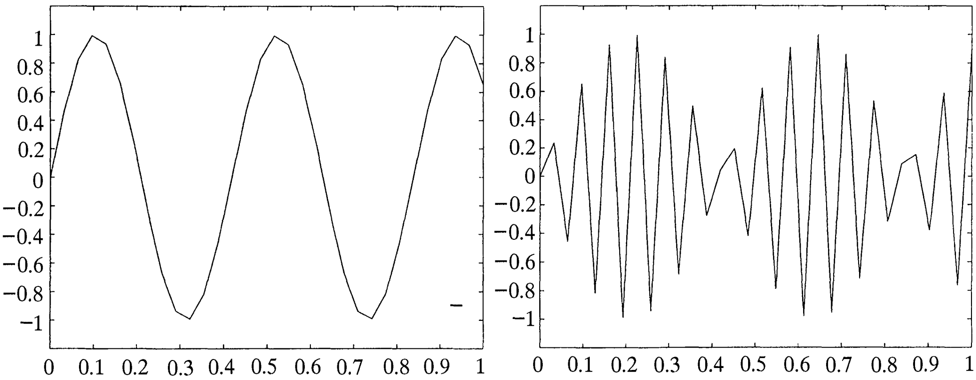


FIGURE 1

Plots of $y=\sin(15x)$ and $y=\sin(90x)$, using 32 uniformly sampled points

While we have just enough information to determine or reconstruct the function [**8**, p. 340], the graphical anomaly (known as *aliasing*) is not too surprising. Using a few more points yields the anticipated 6-fold repetition of the pattern seen in FIGURE 1(a).

FIGURE 1 makes one thing abundantly clear: using uniformly-spaced sample points isn't the smartest approach, unless we are prepared to use a lot of points. Much better pictures are obtained from *Matlab*'s `fplot` command, and from the corresponding *Maple* or *Mathematica* commands. These commands produce adaptive plots, with points clustered where the function seems to exhibit great variation. These adaptive plotting routines examine angles between connecting line segments in provisional internally-generated plots based on uniform sampling. Having identified regions of great variation, they subdivide certain intervals further before producing a visible plot. (For details, see [**12**, p. 216], [**9**, pp. 303–304], [**22**, pp. 579–584].) In the next section we will illustrate how wavelets give rise to an adaptive plotting scheme that does not require us, or the computer, to consider angles first while peeking at default plots.

For the sorts of (differentiable) functions considered so far, intuition correctly suggests that, on the one hand, if we continually replot a function, sampling more and more frequently (uniformly or otherwise), we get a sequence of pictures that converges to the true graph. On the other hand, no matter what scale we view (or print) at, there comes a stage past which it is impossible to detect the use of additional sampled points. Just how many points need we plot to give the illusion of a correct graph? The answer depends very much on the amount of variation the function possesses over the interval in question, as well as on the size of the picture we are going to look at, as the next examples make clear.

Some functions are beyond redemption from the point of view of plotting and displaying at any reasonable scale. A function like $y = \sin(\frac{1}{x})$, which has infinitely many extrema on $(0, 1]$, is going to give this or any other plotting routine a run for its money. The (algebraically) innocent-looking function $y = \sin(e^{2x+9})$ achieves so many extrema (a staggering $\lfloor e^{11}/\pi - 1/2 \rfloor - \lceil e^{9}/\pi - 1/2 \rceil + 1 = 19058 - 2579 + 1 = 16480$, to be precise) in the interval $[0, 1]$, that the plot in FIGURE 2(a), which is based on uniform sampling at 32 points, is totally misleading. Worse still, the more points we plot (even if we plot adaptively) the denser the pictures appear, on account of the nonzero thickness of depicted line segments. FIGURE 2(b) shows what we get if we sample uniformly at 256 points; increase this number further, and the plots start to fill up with connecting line segments. Sadly, given the natural

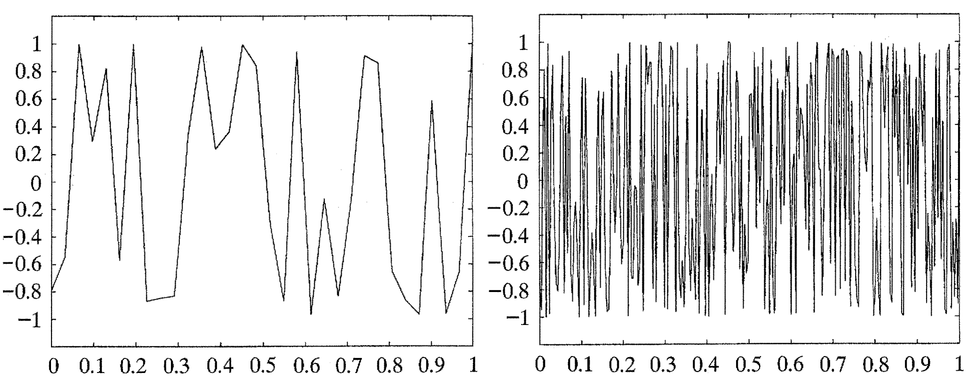


FIGURE 2

Plots of $y = \sin(e^{2x+9})$ using 32 and 256 points, respectively physical limitations of plotters and printers, there is little hope in this life of getting an accurate graph of $y = \sin(e^{2x+9})$ on [0, 1], and we hereby admit defeat.

Of course, linear interpolation of sampled points is just one way of plotting: instead of joining the points with line segments, we could use the y-values as step levels for a staircase effect. FIGURES 3(a) and 3(b) illustrate step function alternatives to FIGURE 1(a) and FIGURE 2(a) respectively, namely $y = \sin(15x)$ and $y = \sin(e^{2x+9})$ using the same uniformly sampled points.

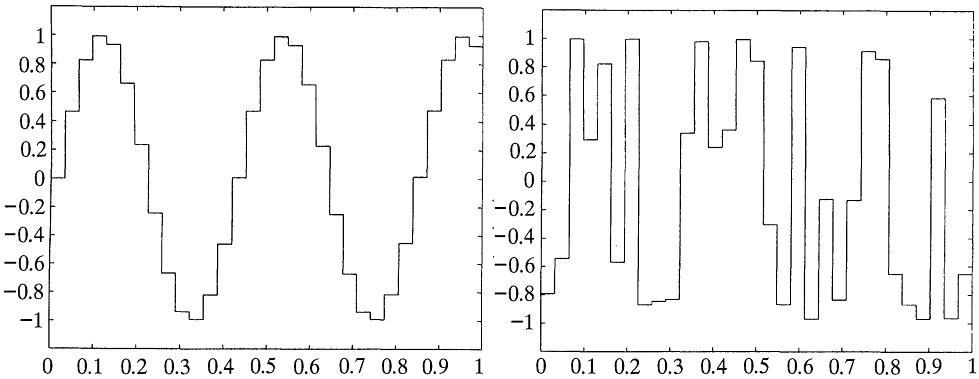


FIGURE 3

Stairs plots of $y = \sin(15x)$ and $y = \sin(e^{2x+9})$ using 32 uniformly sampled points

Purists may balk at the vertical lines connecting the steps in these pictures, which were generated using *Matlab*'s `stairs` command, but for our purposes these uninvited guests are quite harmless. While these plots leave a lot to be desired, just as the linearly interpolated plots earlier did, the staircase method will lead to better and better approximations of the true graph when more points are used, although a lot more are needed to get away from the jaggies and obtain a continuous effect. (That continuous functions can be approximated on [0, 1] to arbitrary precision by piecewise linear functions, or by step functions, is a simple consequence of their being uniformly continuous on compact intervals [**2**, 24.4, 24.5].)

There are also questions of data storage and transmission. These become particularly crucial when we explore higher-dimensional analogues of data points in the plane, such as digital images.

The images of Emmy Noether in FIGURE 4 are derived from two-dimensional arrays of *pixels*—numbers that represent gray levels ranging from black (minimum number) to white (maximum number). These can be thought of as data points (x, y, z), where z measures the gray level at position (x, y): we draw a two-dimensional array of small squares, each shaded a constant gray level z according to its position (x, y) in the array. What we really have here are two-dimensional step functions—viewed from above—where the steps are shaded according to their height. (Color images can be dealt with by decomposing into red, green, and blue components, and treating each of these like grayscales.)

FIGURE (4a) is composed of 256×256 pixels; so it is derived from a matrix of $256^2 = 65536$ pieces of data, each representing a gray level. To produce FIGURE 4(b) we extracted a 64×64 submatrix from the original 256×256 matrix; the submatrix shows the region around the eyes. This second image requires $64^2 = 4096$ pieces of data to store. Due to the lower resolution it is noticeably more "blocky;" we can

FIGURE 4
Emmy Noether—in person and up close

explicitly see the steps that make it up. Both images use $256 = 2^8$ levels of gray, and so are called *8-bit images*.

Clearly, it requires a lot of data to represent an image in this way, and that leads to practical problems. For one thing, a standard 1.44MB high-density floppy disc can only accommodate a handful of large, high-quality, color images. Furthermore, image files are time-consuming to transmit, as anybody who has viewed pictures on the World Wide Web can attest. In the images of Emmy Noether, there are regions of little or no variation. Our goal is to take advantage of these somehow, and come up with a more economical way to store the matrices that represent the images.

3. Scheming

Here we get down to business and describe an elementary wavelet scheme for transforming, and ultimately compressing, digital data. Whether these data represent samples of a function, a matrix of gray levels, or something else entirely, has no bearing on the scheme itself. While wavelets are behind the ideas presented, we defer any further mention of the "W" word until the next section. Readers who wish to duplicate the results and pictures found here can proceed directly to **Averaging and Differencing with Matrices** upon reading this section.

After we describe the basics of the scheme, and look at some examples, we explain what we really mean by compression. A key ingredient is the standard technique for storing large sparse matrices in terms of their nonzero entries—values and locations only—rather than in matrix form.

As motivation, we first consider the images in Figure 5, which use only two shades of gray. How much information is required to store the first one? If we assume that we have black unless specified otherwise, we need only say where the white is, so it seems reasonable to claim that two pieces of information suffice. If the image is a $4 = 2^2$ pixel image, we could store the facts that pixels $(1, 2)$ and $(2, 1)$ are white. But what if the image is, say, a $65536 = 256^2$ pixel image, which just happens to be composed of large black and white blocks? We will show how to use two pieces of

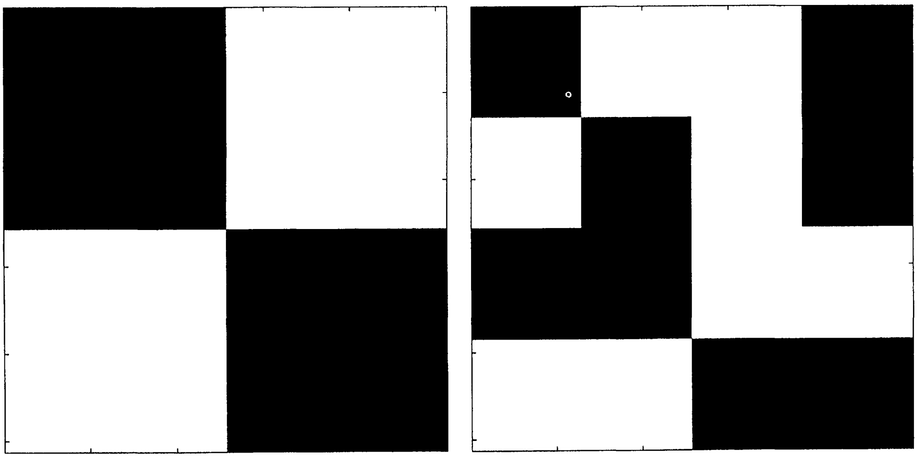


FIGURE 5
How many pieces of information are needed to store these simple images?

information to store the image in this case too; the principle works regardless of the actual resolution.

Next consider the more complex image in FIGURE 5(b). If we treat this as a $16 = 4^2$ pixel image, we see eight white and eight black blocks, so we could argue that eight pieces of information suffice to specify this arrangement. But we can do better if we also use the fact that several of the white blocks are adjacent to each other. We will see presently that only five pieces of information are needed, even if the image is at a greater resolution than is apparent. (For a hint as to why five might be enough, consider the top left quarter of this array as a copy of the first arrangement.)

We now move on to our main goal: describing how to transform arrays of data to a form in which regions of "low activity" in the original become easy to locate in the transformed version. Since matrices consist of neatly stacked rows of numbers, we begin with strings of data. Our method will have immediate application to plotting $y = f(x)$ type functions, as we can identify uniformly sampled functions with data strings.

Consider a string of eight data. This could, for instance, be uniform samples of a function, or a row of an 8×8 pixel image. In order to avoid fractions below, we use these specially cooked-up numbers:

$$64 \quad 48 \quad 16 \quad 32 \quad 56 \quad 56 \quad 48 \quad 24$$

We process these in several stages, in a manner commonly referred to as *averaging and differencing*, which we will explain in a moment. Successive rows of the table show the starting, intermediate, and final results.

64	48	16	32	56	56	48	24
56	24	56	36	**8**	**−8**	**0**	**12**
40	46	**16**	**10**	**8**	**−8**	**0**	**12**
43	**−3**	**16**	**10**	**8**	**−8**	**0**	**12**

The first row is our original data string, which we can think of as four pairs of numbers. The first four numbers in the second row are the averages of those pairs. Similarly, the first two numbers in the third row are the averages of those four averages, taken two at a time, and the first entry in the fourth and last row is the average of the preceding two computed averages.

The remaining numbers, shown in bold, measure deviations from the various averages. The first four bold entries, in the second half of the second row, are the result of subtracting the first four averages from the first elements of the pairs that gave rise to them: subtracting 56, 24, 56, 36 from 64, 16, 56, 48, element by element, yields **8, − 8, 0, 12**. These are called *detail coefficients*; they are repeated in each subsequent row of the table. The third and fourth entries in the third row are obtained by subtracting the first and second entries in that row from the first elements of the pairs that start row two: subtracting 40, 46 from 56, 56, element by element, yields **16, 10**. These two new detail coefficients are also repeated in each subsequent row of the table. Finally, the second entry in the last row, **−3**, is the detail coefficient obtained by subtracting the overall average, 43, from the 40 that starts row three.

It is not hard to see that the last average computed is also the overall average of the original eight numbers. This has no effect on the shape of (any plot of) these data: it merely anchors the data vertically. The seven detail coefficients are what really determines the shape.

We have transformed our original string of eight numbers into a new string of eight numbers. The transformation process is, moreover, reversible: we can work back from any row in the table to the previous row—and hence to the first row—by means of appropriate additions and subtractions. In other words, we have lost nothing by transforming our string. *What have we gained? The opportunity to fiddle with the "mostly detail" version!* If we alter the transformed version, by taking advantage of regions of low activity, and use this doctored version to work back up the table, we obtain an approximation to the original data. If we are lucky, this approximation may be visually close to the original.

Our string has one detail coefficient of 0, due to the adjacent 56's in the original string; this is one region of low activity. The next smallest detail coefficient (in magnitude) is the **−3**. Let's reset that to zero, putting 43, **0, 16, 10, 8, − 8, 0, 12** in the last row of a blank table, and work our way back up by adding and subtracting as indicated above. The completed table looks like this:

67	51	19	35	53	53	45	21
59	27	53	33	**8**	**−8**	**0**	**12**
43	43	**16**	**10**	**8**	**−8**	**0**	**12**
43	**0**	**16**	**10**	**8**	**−8**	**0**	**12**

The first row in this table is our approximation to the original data. In FIGURE 6(a) we plot the original and the approximation, the latter using dashed lines; for reasons which will be clear later, we have plotted the string as y-values against eight equally spaced x-values in $[0, 1]$. While the differences are discernible, many observers would be hard-pressed to distinguish the plots if seen one at a time.

In FIGURE 6(b) we plot the original against the approximation (59 59 27 27 53 53 45 21), obtained by the above procedure after dropping two more detail coefficients, namely the −8 and the 8. Considering how few data (only five numbers) this approximation is based on, it's surprisingly good.

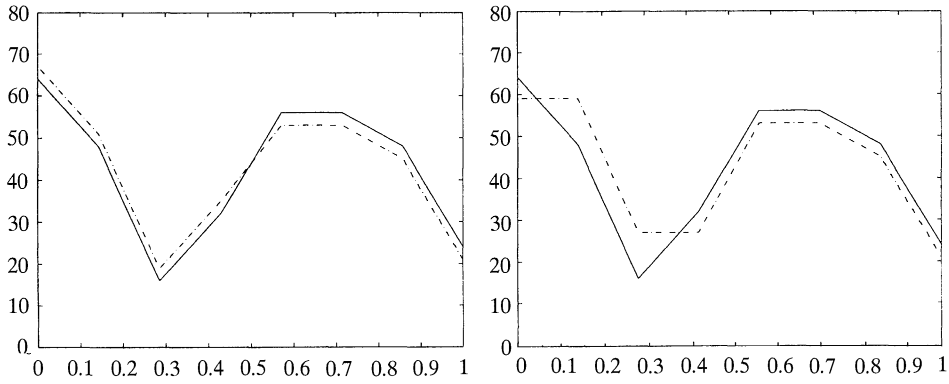


FIGURE 6
Eight pieces of data versus approximations based on six and four detail coefficients, respectively

Before we go on, we note that the process can be generalized to strings of any length. We can always pad at the end, say with zeros, until a string has length equal to a power of two.

To appreciate the full potential of this scheme, we must think big. Starting with a string of length $256 = 2^8$, eight applications of averaging and differencing yield a string with one overall average and 255 detail coefficients. We can then fiddle with this and work back to an approximation of the original.

In general the compression scheme works like this: Start with a data string, and a fixed nonnegative *threshold* value ε. Transform the string as above, and decree that any detail coefficient whose magnitude is less than or equal to ε will be reset to zero. Hopefully, this leads to a relatively sparse string (one with a high proportion of zeros), which is thus compressible when it comes to storage. This process is called *lossless compression* when no information is lost (e.g., if $\varepsilon = 0$); otherwise it's referred to as *lossy compression* (in which case $\varepsilon > 0$). In the former case we can get our original string back. In the latter we can build an approximation of it based on the altered version of the transformed string. The surprise is that we can throw out a sizable proportion of the detail coefficients, and still get decent results.

Let's try this for $y = e^{-10x}\sin(100x)$ on $[0, 1]$, which has a large region of relatively low activity. The plots in FIGURE 7 are based on 32 and 256 uniformly sampled points.

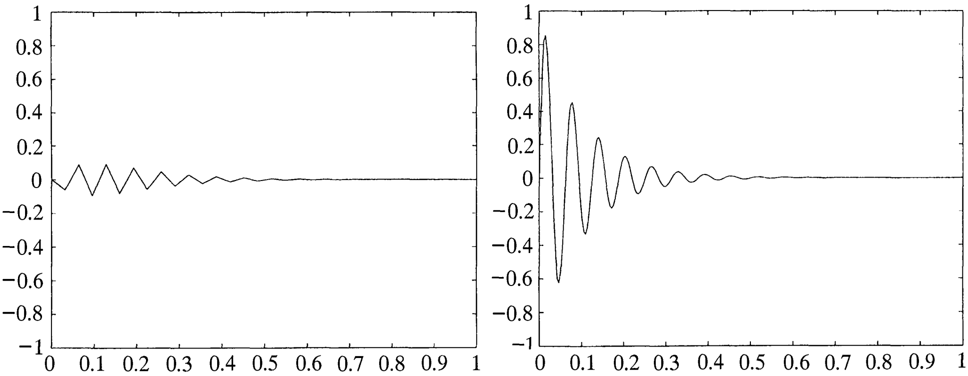


FIGURE 7
Plots of $y = e^{-10x}\sin(100x)$ using 32 and 256 uniformly sampled points, respectively

As FIGURE 7(b) illustrates, half of the points plotted are essentially wasted. Consider the string of 256 y-values used to derive this plot, which range from -0.6246 to 0.8548. After eight rounds of averaging and differencing, we get a transformed string which ranges from -0.2766 to 0.4660. Dumping all detail coefficients less than or equal to 0.04 in magnitude, we get an altered transformed string with 32 nonzero entries. From this sparse string we build an approximation of the original string, which is plotted in FIGURE 8(a). Despite its limitations, this does a better job than FIGURE 7(a) of conveying the flavor of the actual graph. FIGURE 8(b) shows the even better picture obtained when we reduce the cut-off threshold to 0.01, in which case the altered transformed string has 70 nonzero values.

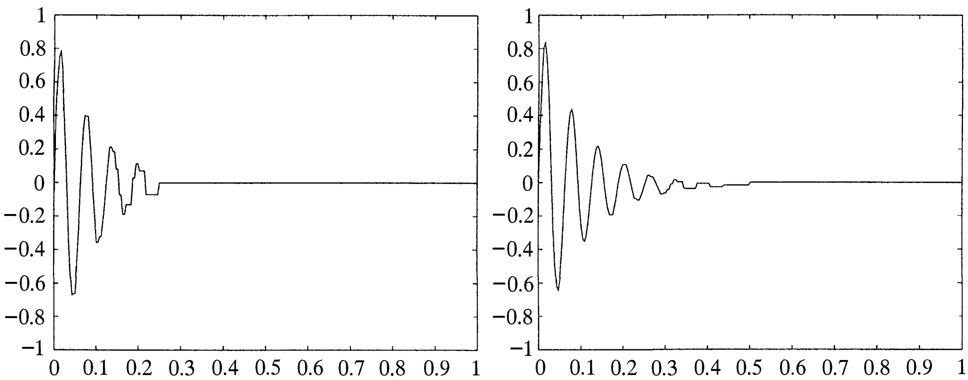


FIGURE 8

Approximations to $y = e^{-10x}\sin(100x)$ using 32 and 70 detail coefficients, respectively

The reason why such low thresholds (relative to the range of values) give good results here, using few detail coefficients, is that this function's pulse is rather weak in half of the interval of interest. Using 70 detail coefficients out of 256 gives a "compression ratio" of around 3.5 : 1.

There is a subtlety here worth highlighting: the plots in FIGURE 8 were generated from just 32 and 70 nonzero numbers, respectively, in sparse strings of length 256 (the doctored transformed strings). However, the plots themselves used all 256 (mostly nonzero) numbers obtained from those strings by reversing our averaging and differencing process. The lossy compression comes into play once we note that it takes significantly less space to store sparse strings of length 256—with only 32 or 70 nonzero entries—than arbitrary strings of length 256.

The approximation technique just outlined has shortcomings as an adaptive plotting scheme—shortcomings that were apparent as early as our first efforts in FIGURE 6. Most obviously, modest-sized data sets such as those we have been considering lead to thresholded strings of data that produce unacceptably jagged plots. This is because thresholding often yields data strings with constant stretches (horizontal steps) followed by dramatic leaps or drops (steep segments). Perhaps surprisingly, regions of lower activity produce the worst "jaggies." A less obvious problem, which FIGURE 6(a) illustrates, is that the range of y-values in the approximation may exceed the range of the original y-values. In **Wavelet Details** we will mention smoother schemes that largely avoid these problems.

We explored the above transformation technique in some detail because we can repeat it for image data sets with almost no extra work. What's more, we get better results, since realistic images consist of much larger data sets, in which steps have to be quite extreme to produce visible blockiness (the higher-dimensional analogue of jaggedness).

We could simply concatenate the rows to obtain one long string, but then we wouldn't be able to exploit natural correlations between adjacent rows of real-world image matrices. Instead, we treat each row as a string and process as above, obtaining an intermediate matrix, and then apply exactly the same transformations to the columns of this matrix to obtain a final row- and column-transformed matrix.

Specifically, to apply the scheme to a 256×256 matrix, we do the averaging and differencing eight times on each row separately, and then eight times on the columns of the resulting matrix. Averaging and differencing columns can also be achieved by transposing the row-transformed matrix, doing row transformations to the result of that transposition, and transposing back. The final result is a new 256×256 matrix, with one overall average pixel value in the top left hand corner, and an awful lot of detail elements. Regions of little variation in the original image manifest themselves as numerous small or zero elements in the transformed matrix, and the thresholding principle described earlier above can be used to effect lossy image compression.

First, let's go back to the simple images in FIGURE 5. Suppose both are 256×256 pixel images, composed of 128×128 and 64×64 monochromatic sub-blocks, respectively. If black pixels match up with matrix entries of 0, and white ones with 1, then performing eight row and then eight column transformations on the matrices corresponding to the images, we obtain matrices that are extremely sparse. The only nonzero entries are bunched up in these 4×4 submatrices in their respective upper left-hand corners:

$$\begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \end{pmatrix}.$$

Thus the first transformed matrix has only two nonzero entries—whereas the second has five. Storing these matrices efficiently leads to a form of lossless compression. The original images can be reconstructed exactly from these smaller data sets.

Now we move on to lossy compression. Applying our thresholding scheme to images with only a few gray levels, such as those in FIGURE 5, is guaranteed to produce poor results, because the averaging process introduces numbers which, if altered and transformed back to image form, correspond to gray levels that were not originally present.

Consider the 8-bit image Noetherian image in FIGURE 4(a), which contains a great deal of black; in fact, black accounts for 20% of the pixels. When we apply eight row and eight column transformations, we obtain a matrix 30% of whose entries are zero; an increase that can be attributed to the other areas of little variation in the original. For appropriate choices of ε—depending on the range of numbers in the matrix used to represent the gray levels of the original image—we get the compressed images in FIGURE 9. Note the concentration of small blocks near the hairline and collar line, and in the facial features, illustrating the adaptiveness of this scheme. The extreme blockiness of these images is due to the nature of averaging and differencing, which is equivalent to working with certain step functions, as we will see in the next section.

The first image uses 6558 out of $256^2 = 65536$ (actually 65535) coefficients, and the second only 1320. In a sense, we could claim compression ratios of 10:1 and 50:1,

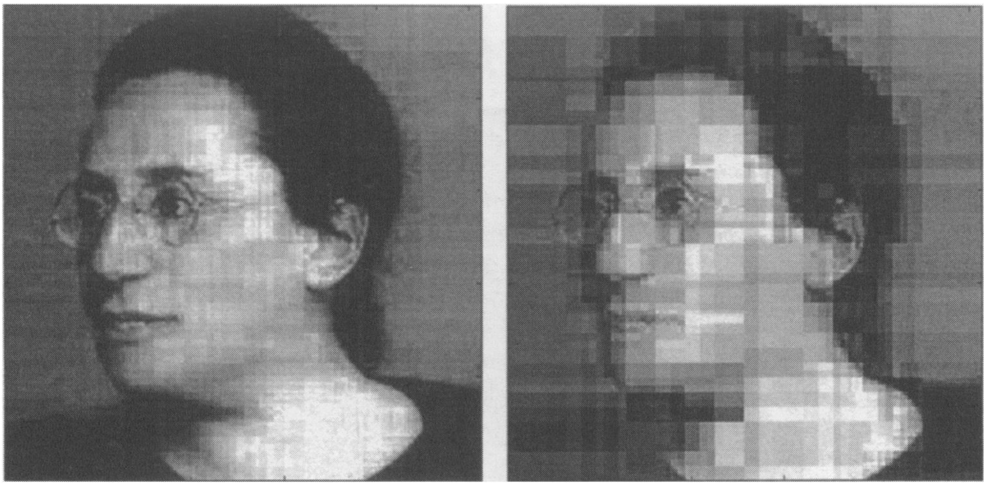


FIGURE 9
Noetherian compression—using 10% and 2% of the detail coefficients, respectively

respectively, but in view of the fact that the original matrix has only 45870 nonzero elements, a more realistic claim might be ratios of 7:1 and 35:1, respectively. Indeed, ratios very close to these turn up when we check how many bytes *Matlab* needs to store the sparse forms of these matrices, whether within a *Matlab* session workspace or in external data files. However we compute compression ratios, it's impressive that the images are recognizable at all, considering how little information was used to generate them.

A modification of the above approach, known as *normalization*, that will likely seem unmotivated for now, yields significantly better results: In the "averaging and differencing" process, divide by $\sqrt{2}$ instead of 2 (so that a pair a and b is processed to yield $(a+b)/\sqrt{2}$ and $(a-b)/\sqrt{2}$). Perhaps unexpectedly, this leads to compressed images that are more acceptable to the human eye than those above. FIGURE 10(a) shows

FIGURE 10
Normalized compression—using 2% and 1% of the detail coefficients, respectively

the normalized compressed version of FIGURE 9(b); both images use 2% of the coefficients. FIGURE 10(b) shows a normalized compressed image that is visually comparable to—if not better than—FIGURE 9(b), but uses only 1% of the coefficients.

We freely admit that compression ratio computation is a rather delicate matter, but the compression schemes we have outlined are surely worthwhile, no matter how one computes these ratios.

We remark that for random matrices, whose entries are interpreted as representing gray levels, there is no hope of any compression. The transformed versions tend to have no nonzero entries to speak of, and thresholding leads to approximations which look unacceptably non-random.

We summarize the central idea of the compression scheme: *Data that exhibit some sort of structure can be efficiently stored in equivalent form as sparse matrices; specifically, in "transformed and sparse" form for lossless compression, and in "transformed, thresholded, and sparse" form for lossy compression.* To view the data, or an approximation of it, one simply "expands" to non-sparse form and applies the inverse transformation.

(At this point, readers may skip to **Averaging and Differencing with Matrices** if they wish. There we describe one matrix multiplication implementation of the compression scheme just discussed.)

4. Wavelets

Our principal aim here is to put our earlier discussions on a firmer mathematical foundation, and to acquaint the reader with some of the standard concepts and notations used in the general study of wavelets. What *are* wavelets, anyway? Before we try to answer this question, we present an alternative vector space description of our discrete, 8-member data sets.

First we identify data strings with a certain class of step functions. A string of length k is identified with the step function on $[0,1]$ which (potentially) changes at $k-1$ equally spaced x-values and uses the string entries as its step heights. For instance, the string of y-values arising from uniformly sampling $\sin(15x)$ 32 times in $[0,1]$ is identified with the step function plotted in FIGURE 3(a). These step functions can in turn can be thought of as linear combinations of dyadically dilated and translated unit step functions on $[0,1)$. We now explain this in some detail.

Consider the *Haar scaling function*:

$$\phi(x) := \begin{cases} 1 & \text{on } [0,1) \\ 0 & \text{elsewhere.} \end{cases}$$

Note that ϕ satisfies a *scaling equation* of the form $\phi(x) = \sum_{i\in\mathbb{Z}} c_i \phi(2x-i)$, where in our case the only nonzero c_i's are $c_0 = c_1 = 1$, i.e., $\phi(x) = \phi(2x) + \phi(2x-1)$.

For each $0 \le i \le 2^3 - 1$, we get an induced (dyadically) dilated and translated scaling function

$$\phi_i^3(x) = \phi(2^3x - i).$$

These eight functions form a basis for the vector space $\mathscr{V}^3$ of piecewise constant functions on $[0,1)$ with possible breaks at $\frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \ldots, \frac{7}{8}$. Note that ϕ_0^3 is 1 on $[0,\frac{1}{8})$ only, ϕ_1^3 is 1 on $[\frac{1}{8},\frac{2}{8})$ only, ϕ_2^3 is 1 on $[\frac{2}{8},\frac{3}{8})$ only, and so on. FIGURE 11 shows three of these basis functions together with a typical element of $\mathscr{V}^3$. Actually, the last plot in FIGURE 11 shows the rather special element

$$64\phi_0^3 + 48\phi_1^3 + 16\phi_2^3 + 32\phi_3^3 + 56\phi_4^3 + 56\phi_5^3 + 48\phi_6^3 + 24\phi_7^3 \in \mathscr{V}^3,$$

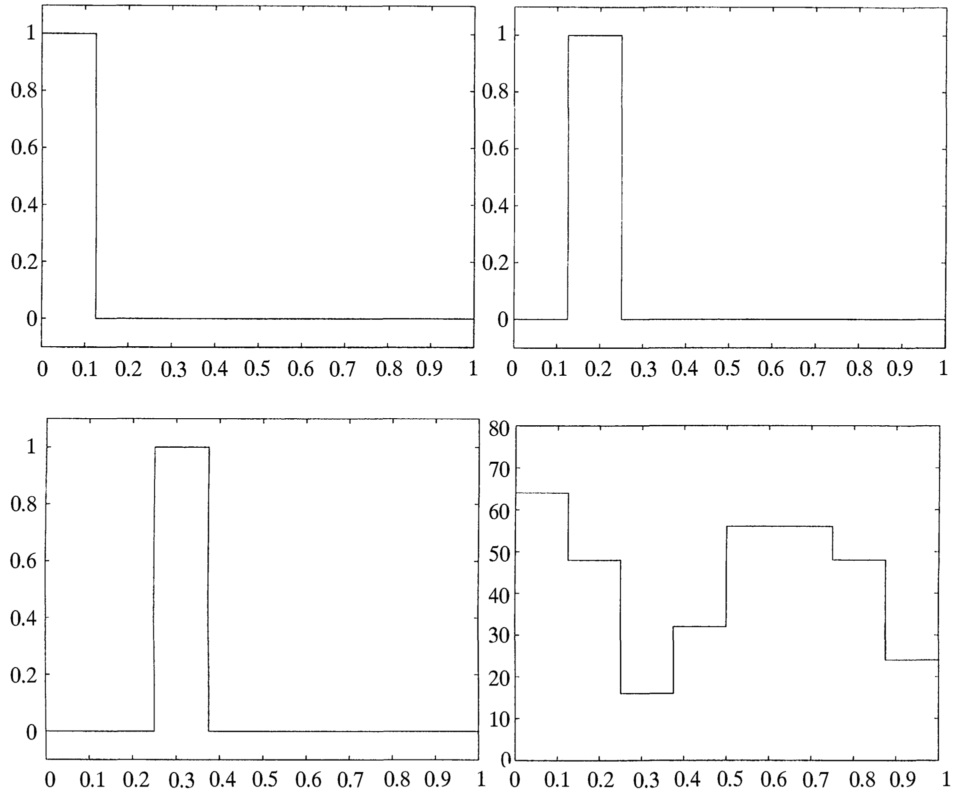


FIGURE 11

The first three of the eight basis functions ϕ_i^3 $(0 \le i \le 7)$ and an element of $\mathscr{V}^3$

which is just another way of thinking of our earlier data string

$$64 \quad 48 \quad 16 \quad 32 \quad 56 \quad 56 \quad 48 \quad 24.$$

In contrast to the piecewise linear plot in Figure 6, we now have a step function representation of our data string. Similarly, any string of length eight can be identified with an element of $\mathscr{V}^3$. We can describe the averaging and differencing scheme from the last section in terms of this version of data strings, but first we need some more vector spaces. As above, the four functions ϕ_i^2 defined by

$$\phi_i^2(x) := \phi(2^2x - i),$$

for $0 \le i \le 2^2 - 1$, form a basis for the vector space $\mathscr{V}^2$ of piecewise constant functions on $[0,1)$ with possible breaks at $\frac{1}{4}, \frac{2}{4}, \frac{3}{4}$; the two functions ϕ_i^1 defined by

$$\phi_i^1(x) := \phi(2^1x - i),$$

for $0 \le i \le 2^1 - 1$, form a basis for the vector space $\mathscr{V}^1$ of piecewise constant functions on $[0,1)$ with a possible break at $\frac{1}{2}$; and $\phi_0^0 := \phi$ itself is a basis for the vector space $\mathscr{V}^0$ of constant functions on $[0,1)$. Note that $\mathscr{V}^0 \subset \mathscr{V}^1 \subset \mathscr{V}^2 \subset \mathscr{V}^3$.

We can identify the various averages derived in **Scheming** with elements of these new vector spaces, by treating these averages as lower-resolution versions of the original string. Specifically, we match up $56, 24, 56, 36$ with $56\phi_0^2 + 24\phi_1^2 + 56\phi_2^2 + 36\phi_3^2$, then $40, 46$ with $40\phi_0^1 + 46\phi_1^1$, and finally 43 with $43\phi_0^0 = 43\phi$.

It only remains to find a new interpretation for the detail coefficients. This is where wavelets finally enter the picture—fasten your seatbelts! Consider the inner product

$$\langle f, g \rangle := \int_0^1 f(t) g(t)\, dt$$

defined on $\mathscr{V}^3$; two functions are orthogonal if and only if their product on $[0,1]$ encloses equal areas on each side of the horizontal axis. For each $j = 0, 1, 2$, we define the *wavelet space* $\mathscr{W}^j$ to be the orthogonal complement of $\mathscr{V}^j$ in $\mathscr{V}^{j+1}$, so that we get the (orthogonal) direct sum decomposition:

$$\mathscr{V}^{j+1} = \mathscr{V}^j \oplus \mathscr{W}^j.$$

Certainly $\mathscr{W}^0 \subset \mathscr{W}^1 \subset \mathscr{W}^2$; in fact we have

$$\mathscr{V}^3 = \mathscr{V}^2 \oplus \mathscr{W}^2 = \mathscr{V}^1 \oplus \mathscr{W}^1 \oplus \mathscr{W}^2 = \mathscr{V}^0 \oplus \mathscr{W}^0 \oplus \mathscr{W}^1 \oplus \mathscr{W}^2.$$

Each $\mathscr{W}^j$ has a natural basis $\{\chi_i^j\colon\ 0 \le i \le 2^j - 1\}$ which we will describe in a moment. Expressing step functions in $\mathscr{V}^3$ in terms of these new bases brings us to the various detail coefficients we encountered before, which will henceforth be known as *wavelet coefficients*.

The *mother Haar wavelet* is defined by

$$\chi(x) := \begin{cases} 1 & \text{on } [0, \frac{1}{2}) \\ -1 & \text{on } [\frac{1}{2}, 1) \\ 0 & \text{elsewhere.} \end{cases}$$

(Equivalently, we could have defined $\chi(x) := \phi(2x) - \phi(2x - 1)$.) Notice that $\{\chi\}$ is a basis for $\mathscr{W}^0$ since χ is clearly orthogonal to ϕ. The four functions

$$\chi_i^2(x) := \chi(2^2 x - i),$$

for $0 \le i \le 2^2 - 1$, form a basis for $\mathscr{W}^2$, because, on the one hand, they are orthogonal to the corresponding functions ϕ_i^2 $(0 \le i \le 3)$ which form a basis for the subspace $\mathscr{V}^2$ of $\mathscr{V}^3$, and, on the other hand, they are visibly orthogonal to each other. (See FIGURE 12.)

Similarly, the two functions χ_i^1 defined by

$$\chi_i^1(x) := \chi(2^1 x - i),$$

for $0 \le i \le 2^1 - 1$, form a basis for $\mathscr{W}^1$.

In present notation, the three steps in the averaging and differencing transformation in the preceding section correspond to the following chain of identities:

$$\begin{aligned} &64\phi_0^3 + 48\phi_1^3 + 16\phi_2^3 + 32\phi_3^3 + 56\phi_4^3 + 56\phi_5^3 + 48\phi_6^3 + 24\phi_7^3 \\ &\quad = 56\phi_0^2 + 24\phi_1^2 + 56\phi_2^2 + 36\phi_3^2 + \mathbf{8}\chi_0^2 - \mathbf{8}\chi_1^2 + \mathbf{0}\chi_2^2 + \mathbf{12}\chi_3^2 \\ &\quad = 40\phi_0^1 + 46\phi_1^1 + \mathbf{16}\chi_0^1 + \mathbf{10}\chi_1^1 + \mathbf{8}\chi_0^2 - \mathbf{8}\chi_1^2 + \mathbf{0}\chi_2^2 + \mathbf{12}\chi_3^2 \\ &\quad = 43\phi_0^0 - \mathbf{3}\chi_0^0 + \mathbf{16}\chi_0^1 + \mathbf{10}\chi_1^1 + \mathbf{8}\chi_0^2 - \mathbf{8}\chi_1^2 + \mathbf{0}\chi_2^2 + \mathbf{12}\chi_3^2. \end{aligned}$$

The final, fully-transformed version, consists of one overall average and seven wavelet coefficients; this is simply a decomposition with respect to a very special basis.

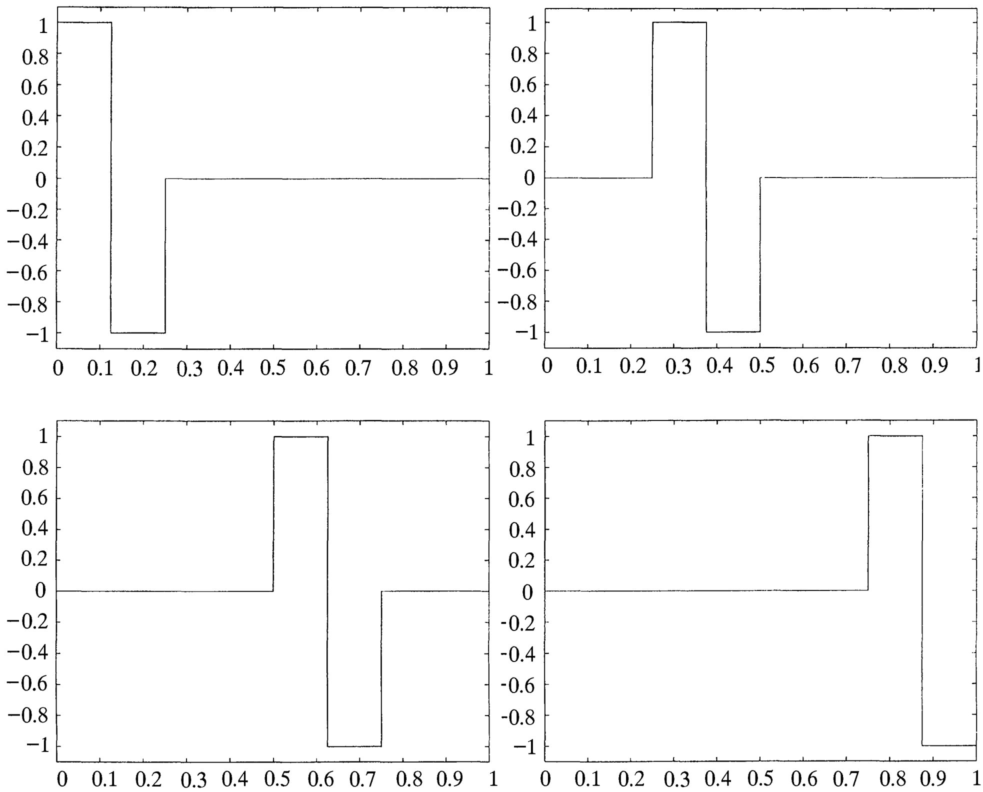


FIGURE 12
The four wavelets χ_i^2 $(0 \leq i \leq 3)$, which form a basis for $\mathscr{W}^2$

Our earlier dumping of the smallest detail coefficients, to effect a good approximation to the original data, boils down to setting some of the wavelet coefficients to zero. In our first compression example, we approximated

$$\begin{aligned} &64\phi_0^3 + 48\phi_1^3 + 16\phi_2^3 + 32\phi_3^3 + 56\phi_4^3 + 56\phi_5^3 + 48\phi_6^3 + 24\phi_6^3 \\ &= 43\phi_0^0 - \mathbf{3}\chi_0^0 + \mathbf{16}\chi_0^1 + \mathbf{10}\chi_1^1 + \mathbf{8}\chi_0^2 - \mathbf{8}\chi_1^2 + \mathbf{0}\chi_2^2 + \mathbf{12}\chi_3^2 \end{aligned}$$

by the element $43\phi_0^0 + \mathbf{0}\chi_0^0 + \mathbf{16}\chi_0^1 + \mathbf{10}\chi_1^1 + \mathbf{8}\chi_0^2 - \mathbf{8}\chi_1^2 + \mathbf{0}\chi_2^2 + \mathbf{12}\chi_3^2$. These are illustrated in the stairs plots in FIGURE 13. As in the zig-zag plots in FIGURE 6(a), the two data strings are difficult to tell apart visually.

These ideas can be extended in the obvious way: For each nonnegative integer j, let $\mathscr{V}^j$ be the vector space of piecewise constant functions on $[0, 1)$ with possible breaks at $1/2^j, 2/2^j, 3/2^j, \ldots, 1 - 1/2^j$. Then the 2^j functions ϕ_i^j defined by $\phi_i^j(x) := \phi(2^j x - i)$, $0 \leq i \leq 2^j - 1$, form a basis for $\mathscr{V}^j$. We thus get an infinite ascending chain[1] of vector spaces $\mathscr{V}^0 \subset \mathscr{V}^1 \subset \mathscr{V}^2 \subset \ldots \subset \mathscr{V}^j \subset \mathscr{V}^{j+1} \subset \ldots$, each of which is an inner product space with respect to the inner product $\langle f, g \rangle := \int_0^1 f(t) g(t)\, dt$. The wavelet space $\mathscr{W}^j$ is then defined to be the orthogonal complement of $\mathscr{V}^j$ in $\mathscr{V}^{j+1}$. The functions

$$\chi_i^j(x) := \chi(2^j x - i),$$

[1] Emmy Noether's presence in these pages might prompt one to ask whether this chain stops! Ideally, no, but in practice, yes: for sampled signals there is a limit to the resolution that can be attained.

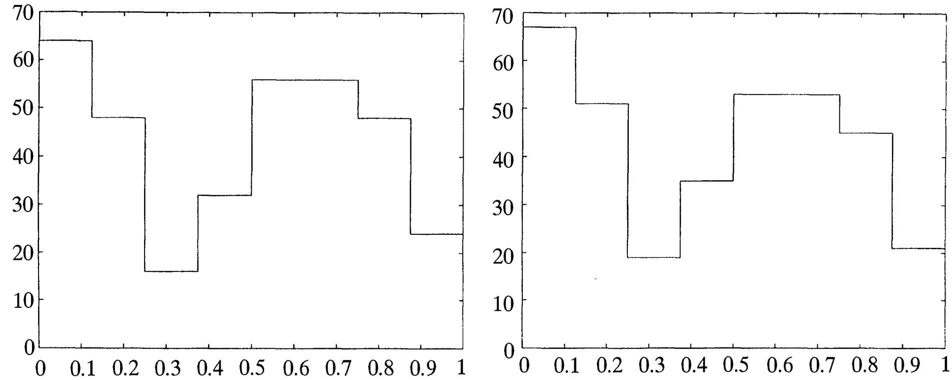


FIGURE 13
Spot the difference—a step function and an approximation of it

for $0 \leq i \leq 2^j - 1$, form a basis for $\mathscr{W}^j$. This gives rise to another infinite ascending chain of vector spaces $\mathscr{W}^0 \subset \mathscr{W}^1 \subset \mathscr{W}^2 \subset \ldots \subset \mathscr{W}^{j-1} \subset \mathscr{W}^j \subset \ldots$, and for any j we have

$$\begin{aligned} \mathscr{V}^j &= \mathscr{V}^{j-1} \oplus \mathscr{W}^{j-1} = \mathscr{V}^{j-2} \oplus \mathscr{W}^{j-2} \oplus \mathscr{W}^{j-1} = \cdots \\ &= \mathscr{V}^0 \oplus \mathscr{W}^0 \oplus \mathscr{W}^1 \oplus \ldots \oplus \mathscr{W}^{j-2} \oplus \mathscr{W}^{j-1}. \end{aligned}$$

Working with strings of length 256 (such as when approximating plots of functions sampled uniformly at 2^8 points) is thus equivalent to working in the larger space $\mathscr{V}^8$ and using the identity:

$$\mathscr{V}^8 = \mathscr{V}^0 \oplus \mathscr{W}^0 \oplus \mathscr{W}^1 \oplus \ldots \oplus \mathscr{W}^6 \oplus \mathscr{W}^7.$$

There are two-dimensional analogs of these ideas, based on products of dilated and translated versions of univariate scaling functions and mother wavelets, which provide a theoretical framework for the digital image representation and compression ideas from the last section. Details can be found in **[15, 16, 17, 8, 7, 11]**.

5. Averaging and Differencing with Matrices

Here we give a natural matrix formulation of the averaging and differencing technique explained in **Scheming**. We provide enough details to allow the curious reader to use a standard computer algebra package, such as *Matlab*, to reproduce the pictures in this article. The *Matlab M-files* we used are available from `http://www.spelman.edu/~colm`. Matrix multiplication is not necessarily the most efficient approach here; for large data sets there are better ways to effect the transformations.

Let A_1, A_2, and A_3, respectively, denote the following matrices:

$$\begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The three-stage transformation from (64 48 16 32 56 56 48 24) to (43 **−3 16 10 8 −8 0 12**) can be thought of in terms of these matrix equations:

$$\begin{aligned} (56\ \ 24\ \ 56\ \ 36\ \ \mathbf{8}\ \ \mathbf{-8}\ \ \mathbf{0}\ \ \mathbf{12}) &= (64\ \ 48\ \ 16\ \ 32\ \ 56\ \ 56\ \ 48\ \ 24)\,A_1, \\ (40\ \ 46\ \ \mathbf{16}\ \ \mathbf{10}\ \ \mathbf{8}\ \ \mathbf{-8}\ \ \mathbf{0}\ \ \mathbf{12}) &= (56\ \ 24\ \ 56\ \ 36\ \ \mathbf{8}\ \ \mathbf{-8}\ \ \mathbf{0}\ \ 12)\,A_2, \\ (43\ \ \mathbf{-3}\ \ \mathbf{16}\ \ \mathbf{10}\ \ \mathbf{8}\ \ \mathbf{-8}\ \ \mathbf{0}\ \ \mathbf{12}) &= (40\ \ 46\ \ \mathbf{16}\ \ \mathbf{10}\ \ \mathbf{8}\ \ \mathbf{-8}\ \ \mathbf{0}\ \ \mathbf{12})\,A_3, \end{aligned}$$

or, equivalently, this single equation:

$$(43\ \ \mathbf{-3}\ \ \mathbf{16}\ \ \mathbf{10}\ \ \mathbf{8}\ \ \mathbf{-8}\ \ \mathbf{0}\ \ \mathbf{12}) = (64\ \ 48\ \ 16\ \ 32\ \ 56\ \ 56\ \ 48\ \ 24)\,A_1 A_2 A_3.$$

So it all boils down to linear algebra! Since the columns of the A_i's are evidently orthogonal to each other with respect to the standard dot product, each of these matrices is invertible. The inverses are even easy to write down—after all they simply reverse the three averaging and differencing steps. In any case, we can recover the original string from the transformed version by the operation:

$$(64\ \ 48\ \ 16\ \ 32\ \ 56\ \ 56\ \ 48\ \ 24) = (43\ \ \mathbf{-3}\ \ \mathbf{16}\ \ \mathbf{10}\ \ \mathbf{8}\ \ \mathbf{-8}\ \ \mathbf{0}\ \ \mathbf{12})\,A_3^{-1}A_2^{-1}A_1^{-1}.$$

It is a routine matter to construct the corresponding $2^r \times 2^r$ matrices $A_1, A_2, \ldots, A_r$ needed to work with strings of length 2^r, and to write down the corresponding equations. For simplicity we write W in place of the product $A_1 A_2 \ldots A_r$ from now on. As mentioned earlier, there is no loss of generality in assuming that each string's length is a power of 2.

For two-dimensional image matrices, we do the same row transformations to each row, followed by corresponding column transformations. The beauty of the string transformation approach is that the equations relating the "before" and "after" strings are valid applied to an image matrix and its row-transformed form. If P is a $2^r \times 2^r$ image matrix then the equations $Q = PW$ and $P = QW^{-1}$ express the relationships between P and its row-transformed image Q. To handle column transformations, we repeat the steps above with a few transposes (denoted by $'$) thrown in. Putting everything together gives the following equations, which express the relationship between the original P and the *row-and-column-transformed* image T:

$$T = ((PW)'W)' = W'PW \quad \text{and} \quad P = ((T')W^{-1})'W^{-1} = (W^{-1})'TW^{-1}.$$

One smart shortcut we can take is to replace all of the $\pm 1/2$'s in the matrices A_j with $\pm 1/\sqrt{2}$'s: this is equivalent to the non-intuitive "averaging" mentioned at the end of the last section. The columns of each matrix A_j then form an orthonormal set. Consequently the same is true of the matrix W, which speeds up the reconstruction process, since the matrix inverses are simply transposes. There is more than mere speed at stake here: as we already saw in FIGURE 10, this normalization also leads to compressed images that are more acceptable to the human eye. (In the language and notation introduced in **Wavelets**, this is equivalent to normalizing the Haar scaling and wavelets functions, so that we use

$$\phi_i^j(x) = 2^{j/2}\phi(2^j x - i) \quad \text{and} \quad \chi_i^j(x) = 2^{j/2}\chi(2^j x - i),$$

(for $0 \leq i \leq 2^j - 1$) as bases for $\mathscr{V}^j$ and $\mathscr{W}^j$, respectively.)

In matrix terms, the image compression scheme works like this: Start with P, and compute $T = W'PW$, which (we hope) will be somewhat sparse. Choose a threshold value ε, and replace by zero any entries of T whose absolute value is less than or equal to ε. Denote the resulting doctored matrix by D; this is sparse by design, and thus easier to store and transmit than P. To reconstruct an image from D, compute $R = (W^{-1})'DW^{-1}$.

Lossless compression is the case where $D = T$ (e.g., if $\varepsilon = 0$) so that $R = P$. Otherwise we have lossy compression, in which case the goal is to pick ε carefully, so as to balance the conflicting requirements of storage (the more zeros in D, the better) and visual acceptability of the reconstruction R.

6. Wavelets on the World Wide Web

In the case of real-time image retrieval, such as grabbing images on the World Wide Web, the compression technique we have discussed allows for a type of progressive image transmission: When an image P is requested electronically, a wavelet-encoded version T is brought out of storage, and bits of information about it are sent "over the wires," starting with the overall average and the larger wavelet coefficients, and working down to the smallest wavelet coefficients.

As this information is received by the user, it is used to display a reconstruction of P, starting with a very crude approximation of the image that, rapidly updated and refined, looks noticeably better as more wavelet coefficients are used. For instance, the images in FIGURES 9 and 10 could form stages in a progressive transmission. Eventually (assuming the user has deemed this picture worth waiting for) all of the wavelet coefficients will have been transmitted and a perfect copy of P displayed. If the user loses interest or patience along the way, she can easily halt the process and move on to some more pressing task, such as learning Fourier analysis.

7. Wavelet Details

In attempting to make this introduction to wavelets as easy and painless as possible, we may have suggested that the subject is neither deep nor profound: nothing could be further from the truth. Here we try to put the Haar wavelets, which were used in image processing as far back as the 1970s **[14]**, in context, and hint at the recent generalizations which have generated so much interest in the mathematical community and elsewhere.

A wide variety of wavelets is available to decompose, analyze, and synthesize both discrete and continuous data. In general, a wavelet is any function whose dilations and translations form a Riesz basis for the function space $\mathscr{L}^2(\mathbb{R})$ (the set of square integrable functions on the real line). For simplicity, we ignore normalization considerations. We also assume that all functions are real-valued.

Most wavelets are derived from a corresponding *scaling function*, namely a function ϕ satisfying a *scaling equation* $\phi(x) = \sum_{i \in \mathbb{Z}} c_i \phi(2x - i)$. Given such a function, we define $\mathscr{V}^0$ to be the closure of the linear span of the set of integer translates $\phi_i^0(x) := \phi(x-i)$, $i \in \mathbb{Z}$, of $\phi(x)$, and then for each $j \in \mathbb{Z}$ take $\mathscr{V}^j$ to be the closure of the linear span of the set of dilated and translated functions $\phi_i^j(x) := \phi(2^j x - i)$, $i \in \mathbb{Z}$. A *multiresolution analysis* (MRA) is said to exist when the induced doubly infinite collection of vector spaces $\ldots \subset \mathscr{V}^{-2} \subset \mathscr{V}^{-1} \subset \mathscr{V}^0 \subset \mathscr{V}^1 \subset \mathscr{V}^2 \subset \ldots$ satisfies three criteria:

1. $f(x) \in \mathscr{V}^j$ if and only if $f(2^j x) \in \mathscr{V}^0$, $\forall j \in \mathbb{Z}$
2. $\bigcap_{i \in \mathbb{Z}} \mathscr{V}^i = \{0\}$
3. $\overline{\bigcup_{i \in \mathbb{Z}} \mathscr{V}^i} = \mathscr{L}^2(\mathbb{R})$.

Once a MRA is in place, it is an easy matter to define the corresponding *mother wavelet*:

$$\chi(x) := \sum_{i \in \mathbb{Z}} (-1)^i c_{1-i} \phi(2x - i),$$

where $\phi(x) = \sum_{i \in \mathbb{Z}} c_i \phi(2x - i)$. This wavelet turns out to have zero integral over the whole real line.

One way to generalize the Haar scaling function (which is a first order B-spline) and wavelet is as follows: for any $k \in N$, the kth order B-spline (which can be thought of as the convolution of the Haar scaling function with itself $k-1$ times) satisfies the scaling equation $\phi(x) = \sum_{i=0}^{k} 2^{-k+1} \binom{k}{i} \phi(2x - i)$. This yields an MRA, and hence a wavelet in the manner just described **[6**, Chapter 5], **[18]**. These *spline wavelets* are compactly supported and have $k-2$ continuous derivatives, but only in the Haar case do we get orthogonality between members of the induced family of translated and dilated functions.

While one does not always insist on orthogonality for such basis functions, it is generally considered desirable for wavelets to have compact support, or at least rapid decay, in sharp contrast to the behavior of the sines and cosines which play a central role in Fourier analysis. This renders wavelets ideal for representing non-periodic functions, especially those with spikes or discontinuities. For one thing, fewer basis elements and coefficients are needed to represent such a function when compared with the classical Fourier series expansion.

There are three things to try to juggle here: smoothness, support, and orthogonality. Sadly, we can't have everything: there are no infinitely differentiable orthonormal wavelets which have exponential decay (never mind compact support) **[6**, Chapter 5]; so some sort of compromise is in order.

The spline wavelet construction above can be modified to yield the so-called Battle–Lemarie wavelets, which have exponential decay, are $k-2$ times continuously differentiable *and* orthonormal. In 1988, Daubechies made a breakthrough with the construction of compactly supported, orthonormal wavelets with any desired finite degree of smoothness. Her simplest non-trivial example is continuous, and is derived from a continuous scaling function ϕ, which satisfies

$$\phi(x) = \frac{1+\sqrt{3}}{4}\phi(2x) + \frac{3+\sqrt{3}}{4}\phi(2x-1) + \frac{3-\sqrt{3}}{4}\phi(2x-2) + \frac{1-\sqrt{3}}{4}\phi(2x-3).$$

There are no closed form expressions for these functions: they are studied by means of a careful analysis that starts with taking the Fourier transform of the scaling equation **[6**, Chapter 6]. (For fixed x, we *can* solve for $\phi(x)$ as the limit of the sequence $\Phi_j(x)$

defined recursively via: $\Phi_j(x) = \sum_{i \in \mathbb{Z}} c_i \Phi_{j-1}(2x - i)$, where Φ_0 is the Haar scaling function.) Despite our ingrained instincts, which suggest that we try to "solve" any equation set in front of us, in general there is no need to get our hands on the scaling functions or wavelets themselves; in many ways they are best explored using the numbers c_i alone.

These more sophisticated, continuous wavelets produce smoother, more satisfactory compressed images than the ones that we obtained **[15, 16, 17]**. Here lies the real potential for progressive image transmission, and perhaps adaptive plotting, too.

For full mathematical treatments, the reader could start with **[6]**, **[5]**, or **[13]**. Books covering applications as well as theory include **[3]** and **[1]**. A gentler survey of the field can be found in **[10]**.

8. Closing Remarks

A major advantage of wavelet over Fourier methods, which we have not touched on at all, is that with wavelets it is possible simultaneously to localize in space (or time) and frequency. Wavelets capture detail at different scales at the same time: the plots of the damped sine curve in FIGURE 8 and the compressed images of Emmy Noether illustrate how the wavelet details take advantage of the changing nature of the data variation over different regions. See **[6]**, **[5]**, or **[13]** for further details.

Glassner's *Principles of Digital Image Synthesis* is an excellent resource, full of helpful pictures, for wavelet basics as they relate to graphics, that also discusses some of the connections with Fourier methods **[8**, Chapter 6**]**. Strang and Nguyen **[20]** treat wavelets from a signal processing perspective.

For an account of a recent adoption of wavelets as a standard for image compression, see **[4]** or **[20]**. Another common use of wavelets is to the *denoising* of digital data. There, unlike in the compression we considered, one discards detail coefficients *larger* than a certain threshold (see **[7]**, **[20]**). There are also wavelet applications to video compression **[20]**; to medicine (tomography, MRI images, mammography, radiography, and neural networks) **[1]**; to audio and speech signals **[21]**, **[20]**; and to partial differential operators and equations **[3]**, **[20]**.

An excellent World Wide Web resource for wavelets is The Wavelet Digest at the University of South Carolina, `http://www.math.sc.edu/~wavelet/`.

Acknowledgement. This research was funded in part by the W. K. Kellogg Foundation, through Spelman College's Center for Scientific Applications of Mathematics (CSAM). The author would also like to thank CSAM Director Sylvia Bozeman, who first stimulated his interest in wavelets.

Thanks to: Tony DeRose and colleagues at the University of Washington for showing the author (and the world) that wavelet basics can be understood and appreciated without a solid background in Fourier methods; Björn Jawerth of the University of South Carolina for an inspiring workshop at SIAM-SEAS in Charleston in March 1995; Alain Fournier *et al.* for the superb wavelets course at SIGGRAPH '95; and Rhonda Hughes of Bryn Mawr College for helpful conversations and clarifications.

Matthew Wright of the University of Southampton dispensed sound advice about *Matlab* and other matters. Shun Yan Cheung of Emory University provided invaluable LaTeX assistance. The Editor and the referees made numerous helpful suggestions, which led to a better and more readable article.

REFERENCES

1. Akram Aldroubi and Michael Unser (editors), *Wavelets in Medicine and Biology*, CRC Pr., Boca Raton, FL, 1996.
2. Robert G. Bartle, *The Elements of Real Analysis*, 2nd edition, John Wiley & Sons, New York, NY, 1976.
3. John J. Benedetto and Michael Frazier (editors), *Wavelets; Mathematics and Applications*, CRC Pr., Boca Raton, FL, 1996.

4. Christopher M. Brislawn, Fingerprints go digital, *AMS Notices* 42 (1995), 1278–1283.
5. Charles Chui, *An Introduction to Wavelets*, Academic Pr., San Diego, CA, 1992.
6. Ingrid Daubechies, *Ten Lectures on Wavelets*, CBMS 61, SIAM Press, Philadelphia, PA, 1992.
7. Ron DeVore, Björn Jawerth, and Bradley Lucier, Image compression through wavelet transform coding, *IEEE Trans. Information Theory* 38, 2 (1992), 719–746.
8. Andrew S. Glassner, *Principles of Digital Image Synthesis*, Morgan Kaufmann, San Francisco, CA, 1995.
9. André Heck, *Introduction to Maple*, Springer-Verlag, New York, NY, 1993.
10. Barbara Burke Hubbard, *The World According to Wavelets*, A. K. Peters, Wellesley, MA, 1996.
11. Björn Jawerth and Wim Sweldens, An overview of wavelet based multiresolution analyses, *SIAM Reviews* 36, 3 (1994), 377–412.
12. *Matlab Reference Guide*, The Math Works, Natick, MA, 1992.
13. Yves Meyer, *Wavelets: Algorithms and Applications*, SIAM Pr., Philadelphia, PA, 1993.
14. William Pratt, *Digital Image Processing*, Wiley-Interscience, New York, NY, 1978.
15. Eric Stollnitz, Tony DeRose, and David Salesin, Wavelets for computer graphics: a primer, Part 1, *IEEE Computer Graphics And Applications* Vol. 15, No. 3 (1995), 76–84.
16. Eric Stollnitz, Tony DeRose, and David Salesin, Wavelets for computer graphics: a primer, Part 2, *IEEE Computer Graphics And Applications* Vol. 15, No. 4 (1995), 75–85.
17. Eric Stollnitz, Tony DeRose, and David Salesin, *Wavelets for Computer Graphics*, Morgan Kaufmann, San Francisco, CA, 1996.
18. Gilbert Strang, Wavelets and dilation equations: a brief introduction, *SIAM Reviews* 31, 4 (1989), 614–627.
19. Gilbert Strang, Wavelets, *American Scientist* 82 (1994), 250–255.
20. Gilbert Strang and Truoung Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.
21. Mladen Victor Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A. K. Peters, Wellesley, MA, 1994.
22. Tom Wickham-Jones, *Mathematica Graphics*, TELOS/Springer-Verlag, Santa Clara, CA, 1994.

Good-bye Descartes?

KEITH DEVLIN
Saint Mary's College of California
Moraga, CA 94575

Happy birthday, Descartes

This year is the 400th anniversary of the birth of René Descartes. To future historians, this might well be seen as the period when the Cartesian domination of science came to an end. Having long underpinned Mankind's attempts to comprehend the physical world, mathematics has hitherto failed to achieve comparable success in our attempts to understand the human world of people and minds. The lofty goals of artificial intelligence, cognitive science, and mathematical linguistics that were prevalent in the 1950s and 1960s (and even as late as the 1970s) have now given way to a realization that the 'soft' world of people and societies is almost certainly not amenable to a precise, predictive, mathematical analysis to anything like the same degree as is the 'hard' world of the physical universe.

In the days when physics and chemistry were the fundamental sciences that underpinned society, mathematics occupied a premier position. It was sometimes referred to as the 'Queen of the Sciences.' But today, in the Age of Information, psychology, sociology, and communication science—the human sciences—occupy at least an equal position in the pecking order, and the mathematical sciences—and mathematics itself—each has to adjust to being just one of a number of ways of understanding how minds work, how people communicate, and how societies function. These days, mathematics frequently finds itself blended in with other disciplines, giving rise to a fascinating new way of using mathematics—what in a forthcoming book [3] I refer to as 'soft mathematics.'

And with this change, we are moving away from the Cartesian view of the world that has been characteristic of scientific investigation for the past three hundred years.

The rise of Cartesian science

To begin at the beginning, what is nowadays often referred to as 'Cartesian science' (or just plain 'science') has its most identifiable beginning with the ancient Greek philosopher and mathematician Thales around 600 B.C., with the Pythagoreans a hundred years later, and with Plato and Aristotle around 350 B.C. To Plato and others we owe the notion that mathematics provides the key to understanding the physical world. The success of mathematics in astronomy and later the study of the physical word in general was dramatic. As a result, it is hardly surprising that mathematics came to occupy a pivotal role in what is generally known today as the 'scientific method.'

The modern scientific method, based on observation, mathematical measurement and description, and logical analysis, owes much to the three individuals Galileo Galilei, Francis Bacon, and Descartes. In the words of Galileo, "The great book of nature can be read only by those who know the language in which it is written, and this language is mathematics." In a similar vein, Descartes wrote that he "neither allows for nor hopes for principles in physics other than those that lie hidden in geometry or in abstract mathematics, for in this way all phenomena of nature will yield to explanation, and a deduction of them can be given."

I think, therefore I am

For Galileo, the role of the scientist was focused on measurement and the discovery of descriptive, quantitative formulas, rather than the formulation of casual explanations obtained by philosophical reflection, which had been typical of earlier work in 'science.' In many ways, Galileo and Bacon were each early forerunners of the 'no-nonsense, down-to-earth, practical scientist' of the twentieth century.

Descartes was in many ways an early forerunner of today's 'applied philosopher.' A delicate individual throughout his life, he was born to a noble family on March 31, 1596, at La Haye, near Tours in France. He received his early education at Jesuit College in La Fleche, leaving school in 1612. Early success at the Paris gambling tables might have indicated the keen mathematical mind that was later to emerge, but it was to soldiering that he turned first, enlisting in the cause of Prince Maurice of Orange at Breda in Holland in 1916.

On November 10, 1619, he reported having three vivid dreams that persuaded him to turn from being a soldier to the more peaceful life of a philosopher. The legacy he left to Mankind as a result of that career switch was swiftly established. For it was during the ensuing two years that he both created analytic geometry and proposed the idea that scientific truth be established not by dialectic reasoning but by rational deduction based on experiment and observational evidence.

Between 1619 and 1621, Descartes moved between Paris and Rome, and it was during this period that he met Cardinal Richelieu, later to become his patron. He lived in Holland from 1628 to 1648, and it was there that he wrote his work *Le Monde*. In 1634, when *Le Monde* was completed, Galileo's enforced public rejection of the Copernican system persuaded Descartes to abandon publication, and he made arrangements to have it published after his death. However, in June of 1637, with the approval of Cardinal Richelieu, he started to publish major parts of his work as the series *Essais Philosophiques*.

After serving as tutor to Princess Elizabeth of Holland for several years, Descartes spent the last year of his life in Sweden, at the invitation of Queen Christine, who had heard of his reputation and desired to be instructed by him. He died on February 11, 1650, the victim of a combination of his delicate health and the cold Swedish winter.

Though he believed that all science could be reduced to mathematics, Descartes made use of very little mathematics in his own work, and his only substantial contribution to mathematics was his famous *La Géomètrie*, in which he created analytic (or 'Cartesian') geometry. This work was included as an appendix to the volume *Discourse on the method of properly guiding the reason in the search of truth in the sciences*, the first of the *Essais Philosophiques*. Though the publication of this volume helped establish the modern scientific approach to knowledge, Descartes' own ultimate interest was elsewhere, namely the nature of human thought and what it is to know something—an interest reflected in his oft-repeated remark "I think, therefore I am."

The science of mind

Descartes believed that his method, the method of science and mathematics, could be applied to the inner world of the mind as well as to the outer world of the physical universe. He wrote, "The long concatenations of simple and easy reasoning which geometricians use in achieving their most difficult demonstrations gave me occasion to imagine that all matters which may enter the human mind were interrelated in the same fashion."

In large part because of the enormous influence Descartes had on the development of modern science—Newton, in particular, was influenced by him—Descartes' views have led to numerous attempts to develop 'mathematical sciences' of language and reasoning modeled on physics, attempts that continue to this day. The belief is that, once we have identified the right features—the equivalents of the length, mass, velocity, acceleration, force, momentum, inertia, and so forth of physics—we can develop a mathematical theory of language and/or reasoning that is every bit as rigorous and precise as physics. In such a science of mind, as much as in physics, mathematics will be both "maidservant and queen."

It is within the Cartesian tradition that modern logic tries to capture in mathematics the patterns of reasoning, and to some extent the patterns of language required to formulate a logical argument. Key to such an approach is the assumption that the thinking mind can be studied in isolation, free from context. However, it was only after Descartes that this approach became the dominant one. Prior to the seventeenth century, logic was regarded largely as an aspect of rhetoric—a study of how one person's argument could convince another. That was certainly the way Aristotle regarded logic. Plato disagreed, condemning the use of rhetoric as "making the worse arguments appear the better," but it was Aristotle's view that predominated. And it continued to do so until Descartes advocated Plato's context-free, 'isolationist' approach in the seventeenth century. For Descartes, the only knowledge worth pursuing was that which could be expressed by eternal, context-free, precise rules that captured general patterns.

Underlying the Cartesian approach to the study of reasoning is Descartes' view that the mind and the brain-body are separate entities. For Descartes, the mind was an abstract entity that resides in the physical brain, and mathematics can be used to explain the workings of that abstract mind. 'Dualism' is the name given to this fundamental separation of mind from body. For the student of language and reasoning who works in the dualist tradition, there are two distinct domains, the subjective, internal world of the mind, and the external world, an objective reality made up of things that bear properties and stand in relations to one another. It is assumed that there are objective facts about the external world that do not depend on the interpretation—or even the existence—of any person. We make our way in the world by acquiring information about those things and constructing an internal representation or 'mental model' of the external world. Thinking is a process of manipulating those internal representations. Cognition is based on the manipulations of the internal representations. Language is a system of symbols that are composed into patterns that stand for things in the world.

One of the major puzzles that arise from the dualist view of the world is the so-called 'mind-body problem': how can our abstract, internal thoughts and intentions about action cause the physical motion of our bodies?

So deeply rooted has Descartes' dualist view become in present-day science—and indeed in much of our present-day world view—that until very recently, not only was it widely believed that it was only a matter of time before familiar-looking, mathematical sciences of reasoning, language, and communication are developed, but any theory—of cognition, language, society, or whatever—that does not fit the expectations of Cartesian science runs the risk of being dismissed, at least by scientists, as 'not completely respectable.'

However, despite its extensive and pervasive acceptance, in recent times a number of philosophers have seriously challenged Cartesian dualism—Husserl, Heidegger, Ricoeur, Gadamer, Merleau-Ponty, Sartre, Mead, Dewey, Habermas, Wittgenstein, Dreyfus, and others. So too have a number of biologists and neuroscientists, among them Maturana, Varela, and Damasio. Within the last decade or so, some leading figures in the computer world have also begun to question the Cartesian view on which much of computer science is based: Winograd and Flores with their 1987 book *Understanding Computers and Cognition* **[11]**, Lucy Suchman with her book *Plans and Situated Action* **[8]**, which also appeared in 1987, and others.

One of the first people to try to move away from the dualist position was the German philosopher Martin Heidegger. In his book *Being and Time* **[5]**, published in 1927, Heidegger investigated the subject known as 'phenomenology,' introduced earlier by Husserl, which seeks to understand the foundations of everyday experience and action. Phenomenology challenges some of our basic assumptions about ourselves and the world. According to Heidegger, it is wrong to adopt a simple objective stance, where the primary reality is an objective physical world, and it is likewise wrong to take a simple subjective stance, where your thoughts and feelings are the primary reality. Rather, neither can exist without the other, and you have to consider both together, as a single whole. In your normal, everyday life, says Heidegger, you do not adopt a detached, 'rational' view of what you do; you simply act. If you think about your actions in a detached, rational way at all, you do so 'after the event,' perhaps because something 'went wrong' and you decide to reflect on what you did. Since this is the way we actually experience the world, moment-to-moment, Heidegger insists, the detached Cartesian view is misleading and, far from leading to a deep understanding of our existence and our actions, will in fact prevent us from achieving an adequate understanding.

For instance, we approach every situation from a prior context that inescapably shapes and prejudices the way we encounter and react to that situation. Because this is how things always are, because we are never in the position of a completely detached observer with no prior experiences—we are never a clean slate if you like—we should not regard prejudice as a condition that leads us to interpret the world falsely. Our prior experiences are a necessary condition for us to interpret the world at all. Interpretation is always relative to prior experiences. Trying to strip away all context is an investigative strategy that can lead to a way of understanding ourselves and the world that may, on occasion, be useful. However, we should not confuse this investigative strategy with the way things 'really are.'

Coming from a very different intellectual background, the biologist Humberto Maturana argues that the dualist view of cognition obscures its complex biological nature, and in so doing creates a misleading view of thought and communication. In their 1980 book *Autopoiesis and Cognition* **[6]**, Maturana and his student Francesco Varela describe living systems (such as organisms) in terms of 'autopoiesis,' a technical notion introduced by Maturana to describe the way the different parts of a living system interact to produce what we call life. Rather than view the system as 'acquiring information' by forming an internal representation, they argue, we should concentrate on the ongoing changes to the system brought about by constant interaction with the environment. Communication between two such systems should not be regarded as a 'transmission of information' but a form of coupling between them.

For Maturana, it is misleading to think of a single, isolated 'state' of an autopoietic system. You have to consider both its environment and its history. In particular, the mind cannot be understood in isolation from the body, a point discussed further in the book *The Embodied Mind* **[10]**, written by Varela, Thompson, and Rosch, and published in 1991.

The year 1991 also saw the appearance of the book *Consciousness Explained* **[2]**, by the philosopher Daniel Dennett, in which he argues against the view of consciousness as a so-called 'Cartesian theater' in which an 'inner self' observes external events played out before us in the mind like the action on a stage. Dennett presented arguments to show that the phenomenon of consciousness can only be understood by looking at the way the components of a complex system interact with each other over a period of time.

Another recent 'attack' on Cartesian dualism comes from the neurologist Antonio Damasio. In his 1994 book *Descartes' Error* **[1]**, he argues that the emotions play a crucial role in human reasoning. While he acknowledges that allowing the emotions to interfere with our reasoning can lead to irrational behavior, Damasio presents evidence to show that a complete absence of emotion can likewise lead to irrational behavior. Damasio's evidence comes from case studies of patients for whom brain damage—either by physical accident, stroke, or disease—has impaired their emotions but has left intact their ability to perform 'logical reasoning,' as may readily be verified by using standard tests of logical reasoning skill. Take away the emotions and the result is a person who, while able to conduct an intelligent conversation and score highly on standard IQ tests, is not at all rational in his or her behavior. Such people will often act in ways highly detrimental to their own well being. Damasio's evidence shows that, when taken to its extreme, the Cartesian idea of a 'coolly rational person' who reasons in a manner unaffected by emotions is an oxymoron. Truly emotionless thought leads to behavior that by anyone else's standards is quite clearly 'irrational.'

And there is more of the same from other sources. It is all relatively new, and almost all controversial. Science never provides 'right' answers. At most a scientific theory might gain universal or almost universal acceptance among the scientific community as 'the best explanation available at the time.' With science in the making, controversy is far more common than agreement. In the case of investigations into human rationality, so deeply is the dualist view ingrained in the psyche of twentieth century Western Man that any theory that challenges that view will have a hard time of it. But for all that we may rail against it in much the same way that our ancestors could not accept that the earth was not flat, the evidence continues to mount that the answers to the age old questions concerning the nature of thought, communication, and action will not be found until we go beyond the boundaries imposed by the legacy of Descartes.

Time to leave the Omega

The contemporary philosopher Stephen Toulmin, in his book *Cosmopolis* **[9]**, likens the course of post-seventeenth century human thought to the Greek letter Omega, Ω. He writes:

> The formal doctrines that underpinned human thought and practice from 1700 on followed a trajectory with the shape of an Omega, i.e. "Ω." After 300 years we are back close to our starting point. Natural scientists no longer separate the "observer" from the "world observed," as they did in the heyday of classical physics. ... Descartes' *foundational* ambitions are discredited, taking philosophy back to [that of an earlier era]. (Page 167, emphasis in the original.)

The Cartesian approach—with its pinnacle role for mathematics—was extremely successful. It led to all of today's science and technology. These days, natural science is often referred to as 'Cartesian science.' Its success motivated attempts to adopt the same approach to the study of mind and language. For instance, the linguist Noam

Chomsky used the term 'Cartesian linguistics' to refer to the mathematically-based analysis of language he developed in the 1950s. And Descartes' philosophy lay behind three decades of immense efforts to develop artificial intelligence. But the very success of Cartesian science has led us around the loop of the Omega. Our inability to develop mathematically-based, Cartesian theories of mind and language and to endow machines with 'intelligence' (see Dreyfus [4]) has forced us to abandon the Cartesian approach and go back to the view advocated by Aristotle. If we want to understand reasoning and communication, we cannot consider them in isolation. We have to consider the context—the context where a person reasons and the context where two people communicate. And that means that mathematics cannot go it alone. At the very least, we have to consider the mental and social contexts, which means that the methods of sociology and psychology will be required. And maybe we have to consider the physical context as well, the fact that the brain is a physical organ in our bodies, requiring the contribution of the neuroscientists. In any event, it is time to leave the Omega. We need to say goodbye to Descartes.

This is hardly a new cry. The same suggestion was made by the mathematician Blaise Pascal while the ink on Descartes' page was barely dry. The following words, taken from Pascal's book *Pensées*, published in 1670, provide an excellent way to close and bid a fond farewell to Descartes. Perhaps.

> The difference between the mathematical mind and the perceptive mind: the reason that mathematicians are not perceptive is that they do not see what is before them, and that, accustomed to the exact and plain principles of mathematics, and not reasoning till they have well inspected and arranged their principles, they are lost in matters of perception where the principles do not allow for such arrangement. ... These principles are so fine and so numerous that a very delicate and very clear sense is needed to perceive them, and to judge rightly and justly when they are perceived, without for the most part being able to demonstrate them in order as in mathematics; because the principles are not known to us in the same way, and because it would be an endless matter to undertake it. We must see the matter at once, at one glance, and not by a process of reasoning, at least to a certain degree. ... Mathematicians wish to treat matters of perception mathematically, and make themselves ridiculous ... the mind ... does it tacitly, naturally, and without technical rules.

REFERENCES

1. Damasio, A. *Descartes' Error: Emotion, Reason, and the Human Brain*, Grosset/Putnam, New York, NY, 1994.
2. Dennett, D. *Consciousness Explained*, Little, Brown, Boston, MA, 1991.
3. Devlin, K. *Goodbye Descartes: The End of Logic and the Search for a New Cosmology of Mind*, John Wiley and Sons, New York, NY, 1997.
4. Dreyfus, H. *What Computers Still Can't Do*, MIT Press, Cambridge, MA, 1993.
5. Heidegger, M. *Being and Time*, Harper and Row, New York, (English translation, 1962).
6. Maturana, H. and Varela, F. *Autopoiesis and Cognition*, D. Reidel, Boston, MA, 1980.
7. Pascal, B. *Pensées*, Paris, France, 1670.
8. Suchman, L. *Plans and Situated Actions: The Problem of Human Machine Communication*, Cambridge University Press, Cambridge, U.K., 1987.
9. Toulmin, S. *Cosmopolis: The Hidden Agenda of Modernity*, Free Press, New York, NY, 1990.
10. Varela, F., Thompson, E. and Rosch, E. *The Embodied Mind*, MIT Press, Cambridge, MA, 1993.
11. Winograd, T. and Flores, F. *Understanding Computers and Cognition: A New Foundation for Design*, Addison-Wesley, Reading, MA, 1987.

NOTES

Algebraic Set Operations, Multifunctions, and Indefinite Integrals

MILAN V. JOVANOVIĆ
Matematički Institut
Kneza Mihaila 35
Beograd, Yugoslavia

VESELIN M. JUNGIĆ
Simon Fraser University
Burnaby, B.C., Canada

The fact that an indefinite integral is a set of functions is often ignored, perhaps because of the apparent simplicity of the situation. However, if we regard

$$\int \frac{dx}{x} \quad \text{or} \quad \int \frac{\cos x}{\sin x}\, dx$$

as functions, we can easily develop fallacious proofs of such "identities" as $0 = 1$.

In this note we introduce a semigroup operation on the set of all nonempty subsets of a vector space. Then we indicate how the indefinite integral can be viewed as a set-valued function (or *multifunction*) and how this point of view avoids the fallacies mentioned above. Finally, we show how the multifunction given by the indefinite integral induces a linear function on the space of continuous functions.

Algebraic set operations Let X be a vector space over the real numbers, and let $P(X)$ denote the family of all nonempty subsets of X. We define addition and scalar multiplication on the family $P(X)$ by

$$A + B = \{a + b : a \in A,\ b \in B\}$$

and

$$\alpha A = \{\alpha a : a \in A\},$$

where $A, B \in P(X)$ and $\alpha \in \mathbb{R}$. In particular, $A - B = A + (-1)B$. These are called *algebraic set operations*. Notice that $(P(X), +)$ is not a group if $X \neq \{0\}$. Indeed, $\{0\}$ is the neutral element in $(P(X), +)$, and for every $A \in P(X)$

$$A + X = X,$$

so X has no inverse element. The operation $+$ is associative and commutative. The following properties of the operations hold

$$\alpha(\beta A) = (\alpha\beta) A \tag{1}$$

$$\alpha(A + B) = \alpha A + \alpha B \tag{2}$$

$$1 A = A. \tag{3}$$

The inclusion

$$(\alpha + \beta) A \subseteq \alpha A + \beta A \tag{4}$$

holds, but the opposite inclusion need not hold. (Setting $X = \mathbb{R}$, $A = \{-1, 1\}$, and $\alpha = \beta = \frac{1}{2}$ gives a counterexample.) Other properties of algebraic set operations include the following, where $A, B, C \in P(X)$ and $\alpha \in \mathbb{R}$:

$$0 \in A - A \tag{5}$$

$$(0 \in A \text{ and } A + B \subseteq C) \Rightarrow B \subseteq C \tag{6}$$

$$\alpha \neq 0 \Rightarrow (A \subseteq B \Leftrightarrow \alpha A \subseteq \alpha B) \tag{7}$$

$$A + B \subseteq C \Rightarrow B \subseteq C - A \tag{8}$$

$$A = B \Rightarrow A + C = B + C. \tag{9}$$

The converse to (8) does not hold, as shown by the example $A = B = X$ and $C = \{0\}$. For $X = \mathbb{R}$, $A = [0, 1]$, $B = \{1\}$, and $C = [1, 2]$ we have $A + B = C$ and $B \neq C - A$. Thus, in general, $A + B = C$ does not imply $B = C - A$.

Some formulae, that do not hold in the general case, do hold for convex sets. A set $A \in P(X)$ is *convex* if for every α, $\beta \in \mathbb{R}$ such that $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta = 1$,

$$\alpha A + \beta A \subseteq A.$$

The converse implication to (9) need not hold in general (e.g., $A \neq X$ and $C = X$). However, if X is a normed vector space, B is closed and convex, and C is bounded, then (see, e.g., [**2**, Lemma 1])

$$A + C \subseteq B + C \Rightarrow A \subseteq B.$$

Let A be a convex set, $\alpha > 0$ and $\beta > 0$; then

$$\frac{\alpha}{\alpha + \beta} A + \frac{\beta}{\alpha + \beta} A \subseteq A.$$

From (7) and (2) we get $\alpha A + \beta A \subseteq (\alpha + \beta) A$, and by (4),

$$\alpha A + \beta A = (\alpha + \beta) A.$$

In particular, if A is convex then

$$A + A = 2A.$$

A subset C of X is a *subspace* if for all $\alpha, \beta \in \mathbb{R}$

$$\alpha C + \beta C \subseteq C.$$

Now, for fixed $x \in X$, the subset $\{x\} + C$ is called an *affine subspace* (*flat*) *parallel to* C. A flat is a convex set. If C is a subspace of X and L is a flat parallel to C, the following algebraic properties are easily proved:

$$C + C = C \tag{10}$$

$$L - L = C \tag{11}$$

$$\alpha \neq 0 \Rightarrow \alpha C = C \tag{12}$$

$$C - C = 0 \tag{13}$$

$$A \subseteq C \Rightarrow A + C = C \tag{14}$$

$$A \subseteq C \Rightarrow A + L = L. \tag{15}$$

For more on algebraic operations with convex sets, see **[3]**.

Indefinite integrals Let $I \subseteq \mathbb{R}$ be an interval, $C(I)$ the vector space of all continuous real functions on I, $C^1(I)$ the subspace of all continuously differentiable functions, and C the subspace of all constant functions. A differentiable function φ is a *primitive function* of f if $\varphi' = f$ holds. The set of all primitive functions of f is called the *indefinite integral* of f, and denoted by

$$\int f = \{\varphi : \varphi' = f\}.$$

Let $f, g \in C(I)$ and $\alpha \in \mathbb{R}$; then

$$\int f \neq \phi \tag{16}$$

$$C = \int f - \int f \tag{17}$$

$$\int f + g = \int f + \int g \tag{18}$$

$$0 \cdot \int f \subseteq \int 0 \cdot f \tag{19}$$

$$\varphi \in \int f \Leftrightarrow \int f = \{\varphi\} + C \tag{20}$$

$$\alpha \neq 0 \Rightarrow \int \alpha f = \alpha \int f \tag{21}$$

$$A \subseteq C \Rightarrow \int f + A = \int f \tag{22}$$

$$f \in C^1(I) \Rightarrow \int f' = \{f\} + C \tag{23}$$

$$u, v \in C^1(I) \Rightarrow \int uv' = \{uv\} - \int u'v. \tag{24}$$

We prove the properties (18) and (24); the others have similar proofs. For (18), let $\varphi \in \int f$ and $\psi \in \int g$; then $\int f + \int g = (\{\varphi\} + C) + (\{\psi\} + C)$. By the commutative and associative laws and (10), we have

$$\int f + \int g = \{\varphi + \psi\} + C.$$

Since $(\varphi + \psi)' = f + g$, (20) shows that

$$\int f + \int g = \int (f + g).$$

To deduce (24), the formula for integration by parts, let $\varphi \in \int uv'$. Then $\varphi' = (uv)' - u'v$. Since $\int (uv)' - u'v = \{uv\} - \int u'v$ we have $\varphi \in \{uv\} - \int u'v$, so $\int uv' \subseteq \{uv\} - \int u'v$. Conversely, if $\varphi \in \{uv\} - \int u'v$ there exists a function $\psi \in \int u'v$ such that $\varphi = uv - \psi$. Since $\varphi' = (uv)' - u'v = uv'$, we have $\varphi \in \int uv'$, so $\{uv\} - \int u'v \subseteq \int uv'$.

Note that, by (20), $\int f$ is a flat, so (11) implies (17).

Example 1. Let $I = (0, \pi)$ and for $x \in I$, $f(x) = \cos x / \sin x$. Let $J = \int f$. Using integration by parts, where $u(x) = 1/\sin x$ and $v(x) = \sin x$, we get $J = 1 + J$. Failure to notice that an indefinite integral is a set leads to the fallacious conclusion that $0 = 1$. However, from (24) we have $J = \{uv\} + J$ where $uv \in C$. Therefore, by (22), $J = J$.

A mistake can also be made in calculating integrals when incorrect set formulae are used. For example, from the "equality"

$$J = u(x)v(x) - J$$

one might conclude that $2J = u(x)v(x)$, which is also incorrect.

Example 2. Let $I = \mathbb{R}$, $f(x) = e^x \sin x$, and $g(x) = e^x(\cos x - \sin x)$. Using (24) we get

$$\int f = \{g\} - \int f.$$

Clearly, $\{g\} \neq 2\int f$. However, $\int f = \{g\} - \int f$ implies, by (9), that

$$\int f + \int f = \{g\} + \left(\int f - \int f\right).$$

Therefore

$$2\int f = \{g\} + C,$$

and, by (7) and (12),

$$\int f = \left\{\frac{g}{2}\right\} + C.$$

The antiderivative multifunction Let X and Y be Banach spaces. A multivalued function (or simply a multifunction) $F\colon X \to P(Y)$ is called *convex* if its graph

$$\operatorname{gr} F = \{(x, y) \in X \times Y \colon y \in F(x)\}$$

is a convex set. This is equivalent to tl ͻ condition that

$$\alpha F(x_1) + \beta F(x_2) \subseteq F(\alpha x_1 + \beta x_2)$$

for all $x_1, x_2 \in X$, and all $\alpha \geq 0$, $\beta \geq 0$ with $\alpha + \beta = 1$.

We say that F is *closed* on X if $\operatorname{gr} F$ is a closed set in the product topology on $X \times Y$. This is equivalent to the condition that $x_k \to x$, $y_k \to y$, $x_k \in X$, and $y_k \in F(x_k)$ imply $y \in F(x)$.

Le Van Hot [**1**, Theorem 2] has proved that if X and Y are Banach spaces and F: $X \to P(Y)$ is a convex closed multifunction such that $\operatorname{dom}(F) = X$ and $F(x_0)$ is bounded for some $x_0 \in X$, then there exists a unique linear single-valued function T: $X \to Y$ such that

$$F(x) = F(0) + T(x). \tag{25}$$

Without the assumption that $F(x_0)$ is bounded for some $x_0 \in X$, the conclusion of Le Van Hot's Theorem is not true. Consider, for example, the multifunction F: $X \to P(Y)$, given by $F(f) = \int f$, where $X = Y = C([0,1])$. Note that $C([0,1])$ is a Banach space with

$$\|f\| = \max\{|f(x)|: x \in [0,1]\}.$$

By (18) and (21), F is a convex function. Also, F is closed by the uniform convergence and differentiation theorem [**4**, Theorem 7.17]. By (16), we have

$$\operatorname{dom}(F) = \{f \in X: F(f) \neq \phi\} = C([0,1]).$$

However, $F(f)$ is unbounded for each $f \in C([0,1])$. In this case the formula (25) becomes

$$\int f = \int 0 + \{T(f)\} \tag{26}$$

or, equivalently,

$$\int f = \{T(f)\} + C.$$

If we let T: $C([0,1]) \to C([0,1])$ be the linear function given by

$$T(f)(x) = \varphi(x) - \varphi(c)$$

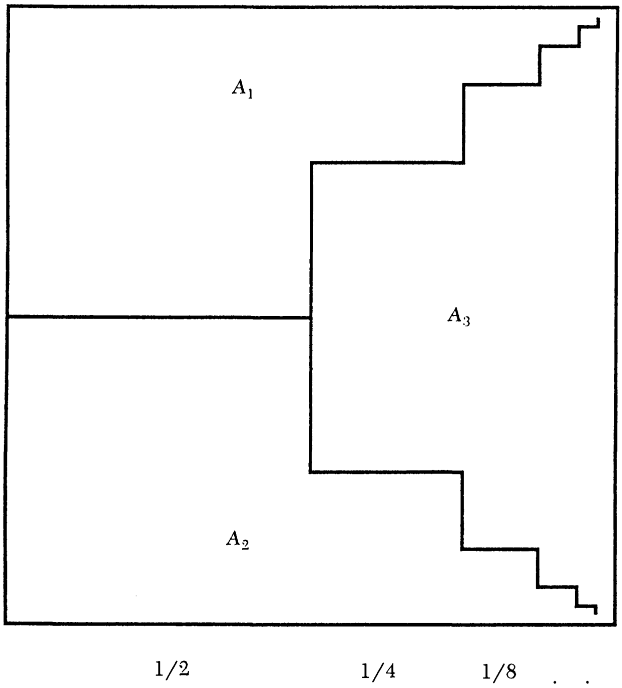
where $\varphi \in \int f$ and c is any number in $[0,1]$, then (26) holds. However, T is not unique.

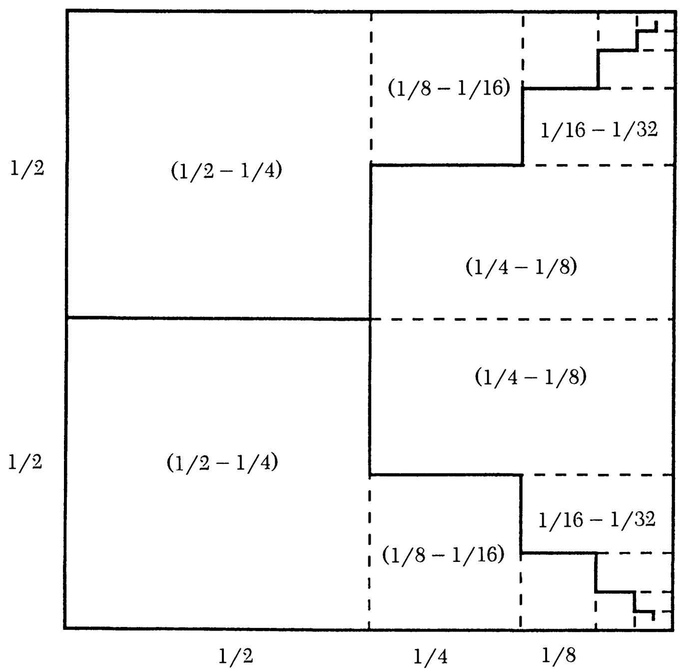
REFERENCES

1. Le Van Hot, *On the open mapping principle and convex multivalued mappings*, Acta. Univ. Carol.-Math. Phys. 26 (1985), 53–59.
2. H. Rådström, *An embedding theorem for spaces of convex sets*, Proc. Amer. Math. Soc. 3 (1952), 165–169.
3. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
4. W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, NY, 1976.

Proof without Words: An Alternating Series

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \frac{1}{32} - \frac{1}{64} + \cdots = \frac{1}{3}$$





$$A_1 = \frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \frac{1}{32} - \frac{1}{64} + \cdots$$

$$A_1 = A_2 = A_3.$$

$$A_1 + A_2 + A_3 = 1.$$

$$\therefore \quad A_1 = \frac{1}{3}.$$

—JAMES O. CHILAKA
LONG ISLAND UNIVERSITY
BROOKVILLE, NY 11548

Finite Groups of 2 × 2 Integer Matrices

GEORGE MACKIW
Loyola College in Maryland
Baltimore, MD 21210

Introduction The story behind this article begins in a classroom, with a presentation intended to show that the dihedral group D_6 of symmetries of the hexagon can be realized as a group of invertible 2×2 matrices with real number entries. Two matrices that can be used to generate this group are

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} -1 & -1 \\ 0 & 1 \end{pmatrix};$$

R has multiplicative order six and F has order two. There is geometric motivation for this choice of generators. As in FIGURE 1, picture a regular hexagon centered at the origin; highlight two of its adjacent radii (v_1 and v_2 in FIGURE 1). Regard these radii as vectors, to form a basis for $\mathbb{R}^2$. Relative to this basis, the matrix R (for "rotation") represents a counterclockwise rotation through 60°, while F (for "flip") corresponds to a reflection of the hexagon through the y-axis.

The set of matrices $\{F^iR^j | i = 0, 1; j = 0, 1, \ldots, 5\}$ forms a group isomorphic to D_6. Familiar relations, such as $FRF = R^{-1}$, can either be checked by multiplying matrices

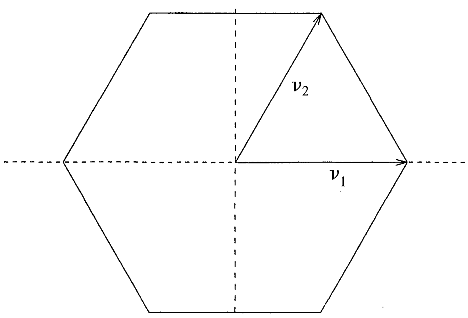


FIGURE 1

$$A_1 = \frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \frac{1}{32} - \frac{1}{64} + \cdots$$

$$A_1 = A_2 = A_3.$$

$$A_1 + A_2 + A_3 = 1.$$

$$\therefore \quad A_1 = \frac{1}{3}.$$

—James O. Chilaka
Long Island University
Brookville, NY 11548

Finite Groups of 2 × 2 Integer Matrices

GEORGE MACKIW
Loyola College in Maryland
Baltimore, MD 21210

Introduction The story behind this article begins in a classroom, with a presentation intended to show that the dihedral group D_6 of symmetries of the hexagon can be realized as a group of invertible 2×2 matrices with real number entries. Two matrices that can be used to generate this group are

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} -1 & -1 \\ 0 & 1 \end{pmatrix};$$

R has multiplicative order six and F has order two. There is geometric motivation for this choice of generators. As in Figure 1, picture a regular hexagon centered at the origin; highlight two of its adjacent radii (v_1 and v_2 in Figure 1). Regard these radii as vectors, to form a basis for $\mathbb{R}^2$. Relative to this basis, the matrix R (for "rotation") represents a counterclockwise rotation through 60°, while F (for "flip") corresponds to a reflection of the hexagon through the y-axis.

The set of matrices $\{F^i R^j | i = 0, 1;\ j = 0, 1, \ldots, 5\}$ forms a group isomorphic to D_6. Familiar relations, such as $FRF = R^{-1}$, can either be checked by multiplying matrices

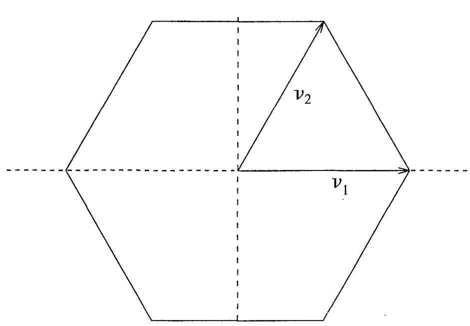


FIGURE 1

or interpreted geometrically. An interesting and attractive feature of this representation of a non-abelian group of order 12 is that all of the matrices have *integer* entries.

Seeing this, a student wondered whether the alternating group A_4, another non-abelian group of order 12, could also be written using integer matrices of size two. I suspected that the answer to this question was well-known, though, sadly, at that moment not by me. Some instinct suggested to me that no such representation was possible, but this was far from proof. To save face, I pointed out that a similar question could be posed for D_4, the group of symmetries of the square. Indeed, elementary arguments show that D_4 can be represented using 2×2 integer matrices. Can the quaternion group, the other non-abelian group of order 8, also be written this way? Better yet, what are *all* the finite groups that can be realized using two by two integer matrices?

Some exploration in the library soon revealed that the possibilities for groups admitting such presentations can be narrowed quite quickly—provided one knows some basic results in the theory of group representations and about degrees of primitive roots of unity over the rationals **[3]**. There remained, then, the challenge of answering the question using only elementary means—say, those available after one semester each of linear and abstract algebra. What follows is an attempt to meet this challenge; an interesting mix of group theory and linear algebra appear along the way.

For any finite group G admitting a matrix representation of the type at hand, the subgroup G^+ of integer matrices of determinant 1 will play a fundamental role. The finite group $SL(2,3)$ of 2×2 matrices of determinant 1 with entries in $\mathbb{Z}_3$, the field with three elements, will prove equally important. In fact, we will show that any such G^+ must be isomorphic to a subgroup of $SL(2,3)$. We will use elementary techniques to find all of the subgroups of $SL(2,3)$, a non-abelian group of order 24. In the process, we will find all possible candidates for a G^+. Once G^+ is known, the structure of the full group G will be easy to determine.

Elements of finite order in $GL(2,\mathbb{Z})$ We denote by $GL(2,\mathbb{Z})$ the group of invertible 2×2 integer matrices whose inverses also have integer entries. We seek to classify the *finite* subgroups of $GL(2,\mathbb{Z})$. If both a matrix A and its inverse have integer entries, then, necessarily, $\det A = \pm 1$, since $\det A^{-1} = 1/(\det A)$. The subset $SL(2,\mathbb{Z})$ of matrices of determinant 1 is a normal subgroup of index two in $GL(2,\mathbb{Z})$.

If a matrix $A \in GL(2,\mathbb{Z})$ has order n, then $A^n = I$ (the identity matrix), so the eigenvalues of A must be nth roots of unity. We claim that such an A must be diagonalizable. If not, then A must have a repeated eigenvalue, say λ. Let v be an eigenvector of A with eigenvalue λ, and choose any vector w so that $\{v, w\}$ is a basis for the complex vector space $\mathbb{C}^2$. Relative to this basis, the matrix of the linear transformation determined by A is of the form $\begin{pmatrix} \lambda & a \\ 0 & b \end{pmatrix}$, for some complex numbers a and b, with $a \neq 0$. Because the characteristic polynomial of A is $(x-\lambda)^2$, we see that $b = \lambda$. Direct computation of powers shows that the matrix $\begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}$, which is similar to A over $\mathbb{C}$, has infinite order. But A has finite order, so we have a contradiction. (A shorter but less elementary proof can by given by appealing to the Jordan canonical form.)

One consequence of diagonalizability is that if A has order 2, and $\det A = 1$, then A must be the matrix $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. In other words, $SL(2,\mathbb{Z})$ has a unique element of order 2. Suppose that A has order greater than 2. Since 1 and -1 are the only complex roots of unity which are also real and $A^2 \neq I$, at least one eigenvalue, λ, of A is not real. Moreover, since the characteristic polynomial of A has integer (and therefore real) coefficients, the eigenvalues of A must be complex conjugates λ and $\overline{\lambda}$, with $\lambda\overline{\lambda} = 1$. But the product of the eigenvalues of a matrix is its determinant, so $\det A = 1$. Thus every element in $GL(2,\mathbb{Z})$ of order greater than 2 has determinant 1.

Reduction mod 3: a mapping into $SL(2,3)$ Our goal is to classify finite subgroups G of $GL(2,\mathbb{Z})$. For any such G, G^+, the subset of elements of determinant 1 in G, is a subgroup of G, with index either 1 or 2. Since G^+ is a finite subset of $SL(2,\mathbb{Z})$, it is tempting to reduce the elements of G^+ mod p, for various primes p. The groups $SL(2,p)$, for p prime, are finite counterparts of $SL(2,\mathbb{Z})$; each consists of 2×2 matrices of determinant 1 over $\mathbb{Z}_p$, the integers mod p. The natural projection from $\mathbb{Z}$ to $\mathbb{Z}_p$ extends to a homomorphism from $SL(2,\mathbb{Z})$ into $SL(2,p)$; it will prove useful to examine the image of G^+ under such a mapping. Indeed, the case $p=3$ provides a wealth of information.

Suppose that the matrix A, $A \neq I$, is in the kernel of the mapping $G^+ \to SL(2,3)$. Since $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, the unique matrix of order two, is not congruent to the identity mod 3, A must have order greater than 2. Also $\operatorname{tr}(A)$, the trace of A, must be an integer with $\operatorname{tr} A \equiv 2 \pmod 3$. But the eigenvalues of A are ω and $\overline{\omega}$, where ω is a (non-real) nth root of unity, so $|\operatorname{tr}(A)| = |\omega + \overline{\omega}| < |\omega| + |\overline{\omega}| = 2$. The only possibility, therefore, is $\operatorname{tr}(A) = -1$, and it follows that A has the form $A = \begin{pmatrix} a & b \\ c & -1-a \end{pmatrix}$, for some integers a, b, and c. Now, $b \equiv c \equiv 0 \pmod 3$ since A is in the kernel of the mapping, and so bc must be divisible by 9. Because A is in G^+, $-a(1+a) - bc = 1$. This relation, taken mod 9, yields $a^2 + a + 1 \equiv 0 \pmod 9$; a direct check shows that no such integer a exists. We have established the following result.

THEOREM 1. *Let G be a finite subgroup of $GL(2,\mathbb{Z})$ and let $G^+ = G \cap SL(2,\mathbb{Z})$. Then the mapping from G^+ to $SL(2,3)$ is an injective homomorphism.*

Thus G^+ is isomorphic to a subgroup of $SL(2,3)$, so the latter group merits a closer look.

The order of $SL(2,3)$ We will compute the order of $SL(2,p)$ for any prime p, and then specialize to $p=3$. Clearly, $SL(2,p)$ is a subgroup of $GL(2,p)$, the full group of invertible 2×2 matrices with entries in $\mathbb{Z}_p$. For any prime p, the orders of $GL(2,p)$ and $SL(2,p)$ are related by $|SL(2,p)| = |GL(2,p)|/(p-1)$. This can be seen by applying the fundamental theorem of group homomorphisms to the mapping $\phi\colon GL(2,p) \to Z_p^*$, given by $\phi(A) = \det(A) \bmod p$, where Z_p^* is the multiplicative group of non-zero elements of $\mathbb{Z}_p$ (Z_p^* has order $p-1$). The kernel of ϕ is $SL(2,p)$.

The order of $GL(2,p)$ can be found by a direct count. A matrix in this group can have any of the (p^2-1) non-zero vectors in Z_p^2 as its first column; the second column can be any vector other than one of the p multiples of the first column—a total of p^2-p choices. This shows that $|GL(2,p)| = (p^2-1)(p^2-p)$; therefore $|SL(2,p)| = p(p^2-1)$. In particular, $SL(2,3)$ has order 24.

$SL(2,3)$ and its subgroups We now proceed to find the subgroups of this group.

LEMMA.

(1) $SL(2,3)$ *contains a unique element of order* 2.

(2) $T = \left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \middle| a \in Z_3 \right\}$ *is a subgroup of order* 3. *Its normalizer*, $N(T)$, *is a cyclic group of order six.*

(3) $SL(2,3)$ *contains a subgroup of order* 8 *isomorphic to the quaternion group.*

Proof. The argument used above to show that $SL(2,\mathbb{Z})$ has a unique element of order two can be used here to establish (1).

In (2), T is clearly a subgroup of order 3. Direct computation shows that elements of $N(T)$ must be of the form $\begin{pmatrix} b & a \\ 0 & b \end{pmatrix}$, where $a \in \mathbb{Z}_3$ and b is either 1 or -1. The matrix $\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}$ has order six and generates $N(T)$.

For (3), let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}$. Direct calculation shows that A and B have order 4, $A^2 = B^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ (the unique element of order two), and $BAB^{-1} = A^{-1}$. Thus A and B generate a quaternion group of order 8.

Let T be defined as in the Lemma. In any finite group, the number of conjugates of a subgroup is the index in the group of the normalizer of the subgroup (for example, see [**4**, p. 52]). Since $N(T)$ has index 4 in $SL(2,3)$, the subgroup T has four distinct conjugates $T_1, \dots, T_4$ in $SL(2,3)$. The normalizers of these four conjugates of T yield four distinct cyclic subgroups of order 6: $S_i = N(T_i)$, $i = 1, \dots, 4$. Each S_i contains the unique element of order two and a single subgroup of order three. Thus, if $i \neq j$, $|S_i \cap S_j| = 2$.

These four subgroups of order six thus account for 18 elements of $SL(2,3)$: 8 elements of order 6, 8 elements of order 3, the single element of order 2, and the identity. The quaternion subgroup from the Lemma above contributes 6 elements of order four. We have now enumerated all 24 of the elements of $SL(2,3)$. In particular, $SL(2,3)$ contains no elements of order 8 or 12. We can now describe the subgroup structure of $SL(2,3)$.

THEOREM 2. *$SL(2,3)$ contains*

(1) *no subgroup of order* 12;
(2) *a unique subgroup of order* 8 (*isomorphic to the quaternion group*);
(3) *no non-abelian subgroup of order* 6;
(4) *cyclic subgroups of orders* 3, 4, *and* 6;
(5) *no subgroup isomorphic to Klein's four group*[1];
(6) *a unique subgroup of order* 2.

Proof. Let α be the unique element of order two in $SL(2,3)$. Suppose there were a subgroup H of order 12. Since H has even order, H must contain α [**4**, p. 17, Ex. 2.18]. Since H has index 2, it must contain the square of any element in $SL(2,3)$. If A is any element of order 3, then A is a square since $A = A^4 = (A^2)^2$. Thus, H must contain all eight elements of order 3. Since α commutes with elements of order 3, multiplying them by α produces 8 more elements of order 6 in H. This places at least seventeen elements in H—a contradiction.

To establish (2), recall that $SL(2,3)$ contains only one element of order 2, no element of order 8, and 6 elements of order 4. Thus, any subgroup of order 8 must contain the six elements of order 4 that generate the quaternion subgroup of the Lemma. Assertions (3), (5) and (6) follow from the fact that $SL(2,3)$ contains only one element of order 2. We have established (4) above.

Observe that our analysis of subgroup structure did not require use of the Sylow theorems.

[1]Named after the mathematician Felix Klein, this is the non-cyclic group of order four and is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$.

The finite subgroups of $GL(2,\mathbb{Z})$ There are only two non-cyclic subgroups of $SL(2,3)$: the quaternion subgroup of order 8 and $SL(2,3)$ itself. If G is a finite subgroup of $GL(2,\mathbb{Z})$, we have seen that G^+ is isomorphic to a subgroup of $SL(2,3)$. We now show that G^+ must be cyclic.

Suppose, instead, that G^+ is isomorphic to the quaternion group of order 8. To derive a contradiction, we reduce G^+ mod 2, producing a homomorphism ϕ: $G^+ \to SL(2,2)$. Since $SL(2,2)$ has order 6, the kernel of ϕ must contain an element, A, of order 4. As we observed earlier, the eigenvalues of A are i and $-i$ (two of the complex fourth roots of unity). Thus, A has trace zero, and so must be of the form $\begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ for some integers a and b. Now, $b \equiv c \equiv 0 \pmod 2$ since A is in the kernel of the mapping, so bc is divisible by 4. Since A has determinant 1, $-a^2 - bc = 1$. It follows that $a^2 \equiv -1 \pmod 4$. This is impossible, since the square of every odd integer is congruent to 1 (mod 4).

The same argument rules out the possibility that G^+ is isomorphic to $SL(2,3)$, since such a G^+ would have a subgroup isomorphic to the quaternion group of order 8. Theorem 8 says, therefore, that G^+ must be isomorphic to one of the groups

$$C_1, C_2, C_3, C_4, \text{ or } C_6,$$

where C_i denotes the cyclic group of order i.

The structure of G itself now follows readily. Our earlier discussion shows that, among elements of finite order in $GL(2,\mathbb{Z})$, only elements of order two have determinant -1. If $G^+ \neq G$, then G^+ has index 2 in G. Let x be an element of G that is not in G^+. Then all the elements of the coset G^+x must have order 2, since they are matrices of determinant -1. In particular, if y is a generator of the cyclic group G^+, then yx must have order 2. Thus, $(yx)(yx) = 1$ and $xyx^{-1} = y^{-1}$; in other words, conjugating by x inverts G^+. This means that G must then be isomorphic to one of the dihedral groups

$$D_1, D_2, D_3, D_4, \text{ or } D_6.$$

Since all the groups C_i and D_i above are subgroups of one of the dihedral groups D_4 or D_6, and since (as noted at the outset) both D_4 and D_6 can be written using integer matrices, we can summarize our results as follows.

THEOREM 3. *A finite group G can be represented as a group of invertible 2×2 integer matrices if and only if G is isomorphic to a subgroup of D_4 or D_6.*

Conclusion A more economical presentation could be achieved by using the Sylow theorems in analyzing $SL(2,3)$, and by noting that the minimum polynomial of an element of finite order n in $GL(2,\mathbb{Z})$ must be divisible by the minimal polynomial over the rationals of a primitive nth root of unity. A famous theorem due to Gauss asserts that the degree of a primitive nth root of unity over the rationals is $\phi(n)$, where ϕ is Euler's totient function. In our situation, $\phi(n) = 1$ or 2; this forces $n = 1$, 2, 3, 4, or 6.

The results above are related to a geometric result called the *crystallographic restriction*, which arises in classifying symmetry groups of crystals (see e.g., [**1**, p. 151]). This restriction says that the only rotations admitted by lattices in dimensions 2 or 3 are those through angles $2\pi/n$, where $n = 1$, 2, 3, 4, or 6. Indeed, given a matrix A of the type under consideration, of order $n \geq 3$, we have seen that the eigenvalues of A are precisely $e^{i\theta}$ and $e^{-i\theta}$, with $\theta = 2\pi m/n$ and m and n relatively prime. But the rotation matrix $R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ has exactly the same two distinct eigenvalues. Thus A and R are similar over the complex numbers, and hence also over the real numbers ([2, p. 158]); i.e., $CA = RC$ for some invertible real matrix C. The columns of C can be viewed as the basis of a two dimensional lattice L. Since A has integer entries, the relation $RC = CA$ shows that rotating lattice vectors through angle θ produces vectors that are *integer* linear combinations of a basis of L. So the lattice L admits a rotational symmetry and the crystallographic restriction can be invoked to reveal the possible values of n.

Acknowledgment I would like to thank the referee and the editor for many helpful comments.

REFERENCES

1. M. A. Armstrong, *Groups and Symmetry*, Springer-Verlag, New York, NY, 1988.
2. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
3. M. Newman, *Integral Matrices*, Academic Press, New York, NY, 1972.
4. J. J. Rotman, *An Introduction to the Theory of Groups*, 3rd ed., Allyn and Bacon, Boston, MA, 1984.

Moving Card *i* to Position *j* with Perfect Shuffles

SARNATH RAMNATH
Dept. of Computer Science

DANIEL SCULLY
St. Cloud State University
St. Cloud, MN 56301-4489

To perform a perfect riffle shuffle, or faro shuffle, on a deck of $2n$ cards, you cut the deck into two stacks of n cards and interlace them perfectly. This can be done in two ways. If the shuffle leaves the top card on top, it is called an *out* shuffle. If the shuffle moves the top card into the second position, it is called an *in* shuffle.

Perfect shuffles have been of great interest to a wide variety of people for a long time. We have seen references to books on card cheating that described the perfect shuffle back in the eighteenth century. Magicians use perfect shuffles in card tricks (see Marlo [7] and [8]), and computer scientists use them in parallel processing (see Stone [12] and Chen, et al. [3]).

For the mathematician, perfect shuffles provide a deep and complex structure from a very simple and natural setting. Mathematics literature on the perfect shuffle ranges from the recreational and nontechnical in Gardner [5], Ball and Coxeter [2], Adler [1], Herstein and Kaplansky [6], and Rosenthal [11] to the very sophisticated work of Diaconis, Graham, and Kantor [4] where the group generated by the in and out shuffles is determined. Generalizations of the perfect shuffle provide more grist for the mathematical mill in Morris and Hartwig [10], and Medvedoff and Morrison [9].

Moving cards to desired positions through perfect shuffles is of interest to magicians and card cheaters because perfect shuffles appear to be random but are not. It has long been known, and easily proved [4], that the top card can be moved to position j (the top card is in position 0) through a sequence of in and out shuffles determined by the base-two representation of j. Reading the base two digits from left to right, simply perform a shuffle for each digit: an in shuffle for a 1 and an out shuffle for a 0. The distinct eigenvalues. Thus A and R are similar over the complex numbers, and hence also over the real numbers ([**2**, p. 158]); i.e., $CA = RC$ for some invertible real matrix C. The columns of C can be viewed as the basis of a two dimensional lattice L. Since A has integer entries, the relation $RC = CA$ shows that rotating lattice vectors through angle θ produces vectors that are *integer* linear combinations of a basis of L. So the lattice L admits a rotational symmetry and the crystallographic restriction can be invoked to reveal the possible values of n.

Acknowledgment I would like to thank the referee and the editor for many helpful comments.

REFERENCES

1. M. A. Armstrong, *Groups and Symmetry*, Springer-Verlag, New York, NY, 1988.
2. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
3. M. Newman, *Integral Matrices*, Academic Press, New York, NY, 1972.
4. J. J. Rotman, *An Introduction to the Theory of Groups*, 3rd ed., Allyn and Bacon, Boston, MA, 1984.

Moving Card *i* to Position *j* with Perfect Shuffles

SARNATH RAMNATH
Dept. of Computer Science

DANIEL SCULLY
St. Cloud State University
St. Cloud, MN 56301-4489

To perform a perfect riffle shuffle, or faro shuffle, on a deck of $2n$ cards, you cut the deck into two stacks of n cards and interlace them perfectly. This can be done in two ways. If the shuffle leaves the top card on top, it is called an *out* shuffle. If the shuffle moves the top card into the second position, it is called an *in* shuffle.

Perfect shuffles have been of great interest to a wide variety of people for a long time. We have seen references to books on card cheating that described the perfect shuffle back in the eighteenth century. Magicians use perfect shuffles in card tricks (see Marlo [**7**] and [**8**]), and computer scientists use them in parallel processing (see Stone [**12**] and Chen, et al. [**3**]).

For the mathematician, perfect shuffles provide a deep and complex structure from a very simple and natural setting. Mathematics literature on the perfect shuffle ranges from the recreational and nontechnical in Gardner [**5**], Ball and Coxeter [**2**], Adler [**1**], Herstein and Kaplansky [**6**], and Rosenthal [**11**] to the very sophisticated work of Diaconis, Graham, and Kantor [**4**] where the group generated by the in and out shuffles is determined. Generalizations of the perfect shuffle provide more grist for the mathematical mill in Morris and Hartwig [**10**], and Medvedoff and Morrison [**9**].

Moving cards to desired positions through perfect shuffles is of interest to magicians and card cheaters because perfect shuffles appear to be random but are not. It has long been known, and easily proved [**4**], that the top card can be moved to position j (the top card is in position 0) through a sequence of in and out shuffles determined by the base-two representation of j. Reading the base two digits from left to right, simply perform a shuffle for each digit: an in shuffle for a 1 and an out shuffle for a 0. The reverse problem of bringing any card to the top is probably of greater interest to the magician and is, in Martin Gardner's words, "much harder to analyze." Still quoting from Gardner's *Mathematical Carnival*: "Some attempts at efficient algorithms, combining shuffles of different types, have been proposed, but the problem is far from satisfactorily disposed of."

We present a procedure for determining the shortest possible sequence of perfect in and out shuffles that moves a card from position i to position j in a deck of $2n$ cards. The procedure is easy and efficient (of order $\log n$). Gardner's problem is solved as a special case by choosing the jth position to be the top of the deck ($j = 0$).

Label the $2n$ positions in the deck 0 through $2n - 1$ consecutively, with 0 representing the top position. It is easy to see that the out and in shuffles move a card in position x to positions $\mathcal{O}(x)$ and $I(x)$, respectively, where

$$\mathcal{O}(x) = \begin{cases} 2x \bmod 2n & \text{if } 0 \le x < n \\ (2x+1) \bmod 2n & \text{if } n \le x < 2n \end{cases}$$

and

$$I(x) = \begin{cases} (2x+1) \bmod 2n & \text{if } 0 \le x < n \\ 2x \bmod 2n & \text{if } n \le x < 2n. \end{cases}$$

Let $D(x) = 2x$ and $E(x) = 2x + 1$, without any modding. The effects of the functions D and E on the base-two representation of x is clear. If the binary expansion of x is $x = x_1 x_2 \dots x_k$ (all such expansions will be binary in this paper) then $D(x) = x_1 x_2 \dots x_k 0$ and $E(x) = x_1 x_2 \dots x_k 1$.

We construct a binary tree with root 0 that reflects all possible sequences of compositions of D and E applied to 0. Moving down the tree, a step to the left indicates the application of D, and a step to the right indicates E. Thus, as FIGURE 1 shows, 0 is sent to 4 by applying E followed by two D's.

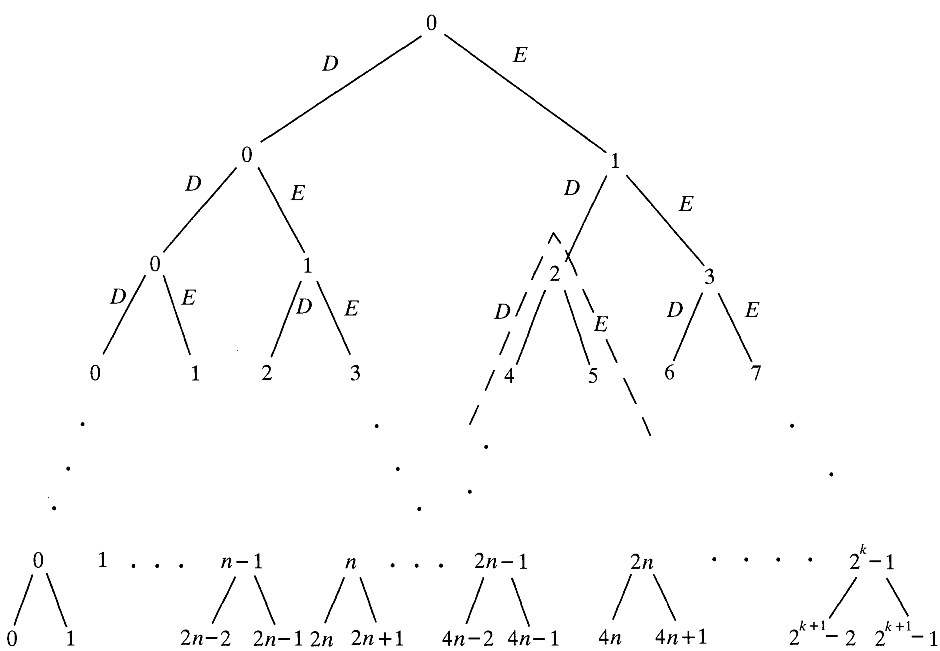


FIGURE 1

It is clear that the integers that appear on the row of the tree that is t steps from the root are the consecutive integers 0 through $2^t - 1$ since the t-digit base-two representations of these numbers generated by D and E are $00\ldots0$ through $11\ldots1$. If j appears on the t-th level of this tree, then the path from the root labeled 0 to that vertex labeled j is revealed by the t-digit base-two representation of j. If $j = j_1 j_2 \ldots j_t$, then the sequence $F_1, F_2, \ldots, F_t$ moves 0 to j, where $F_k = D$ if $j_k = 0$ and $F_k = E$ if $j_k = 1$.

Similarly generated binary trees with different roots appear as subtrees of this one. FIGURE 1 illustrates how the tree with a root of 2 is such a subtree. On the subtree rooted at i, the t-th level down from the root consists of the consecutive integers $D^t(i) = 2^t i$ through $E^t(i) = 2^t(i+1) - 1$. If j appears on the t-th level down from the root of this subtree, then the path from the root labeled i to that vertex labeled j is revealed by the t-digit base-two representation of $j - D^t(i) = j - 2^t i$ as above.

The introduction of modular arithmetic does not complicate matters much. Let $D_m(x) = 2x \bmod 2n$ and $E_m(x) = (2x+1) \bmod 2n$. The analogous tree using these functions is very similar to the previous one. FIGURE 2 illustrates how the integers along a row of this tree or one of its subtrees are consecutive from left to right, but $2n-1$ is followed by 0, 1, 2 etc. This follows immediately from the fact that the $\bmod 2n$ function from $\mathbb{Z}$ to $\mathbb{Z}_{2n}$ is a homomorphism.

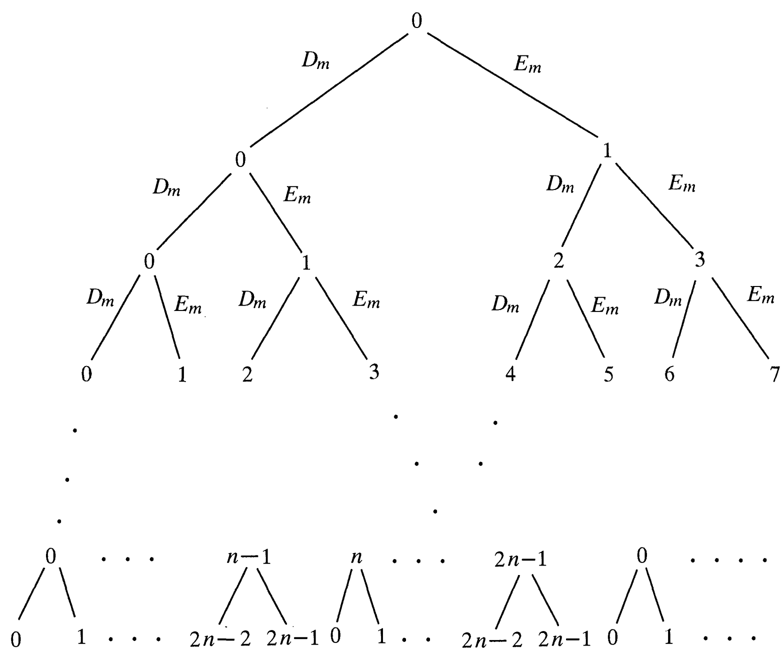


FIGURE 2

Let $i, j \in \mathbb{Z}_{2n}$. We can find the shortest sequence of D_m's and E_m's that sends i to j by simply constructing the subtree rooted at i until j first appears. The path from the root i to j determines the sequence of D_m's and E_m's.

Constructing such a tree could be time-consuming if $2n$ is large. It is of order n. But it isn't necessary to construct the whole subtree. First, construct the two outermost paths of the subtree ($D_m(i), D_m^2(i), \ldots, D_m^t(i)$ and $E_m(i), E_m^2(i), \ldots, E_m^t(i)$) until j falls in the interval $[D_m^t(i), E_m^t(i)]$. (In $\mathbb{Z}_{2n}$, $[a, b] = \{a, a+1, \ldots, b\}$ if $a \le b$, and $[a, b] = \{a, a+1, \ldots, 2n-1, 0, 1, \ldots, b\}$ if $b < a$.) The integer j first appears on that level t of the subtree since the integers that appear along each level are consecutive. This is sure to happen in fewer than $\log_2(2n)+1$ steps since, for any t larger than $\log_2(2n)$, the width of the t-th level of the subtree exceeds $2n$. Scanning the vertices on the t-th level of this subtree from left to right, j first appears $(j-D_m^t(i)) \bmod 2n$ vertices to the right of the first vertex. Therefore, the t-digit base-two representation of $(j-D_m^t(i)) \bmod 2n$ indicates the sequence of D_m's and E_m's needed to get from i to j. A 0 indicates D_m and a 1 indicates E_m.

Example. Find the shortest possible sequence of compositions of the functions D_m and E_m that takes $i=6$ to $j=47$ in $\mathbb{Z}_{52}$.

Solution: We construct the outermost paths of the subtree rooted at 6.

$$6 \xrightarrow{D_m} 12 \xrightarrow{D_m} 24 \xrightarrow{D_m} 48 \xrightarrow{D_m} 44 \ (=96 \bmod 52)$$

$$6 \xrightarrow{E_m} 13 \xrightarrow{E_m} 27 \xrightarrow{E_m} 3 \xrightarrow{E_m} 7$$

Our number 47 falls in the interval $[44,7]$ in $\mathbb{Z}_{52}$, and we had to apply the functions four times to get an interval that contains it. The difference $(j-D_m^4(i)) \bmod 52 = (47-44) \bmod 52 = 3 = 0011$ as a four-digit base two numeral, so the sequence D_m, D_m, E_m, E_m does the trick. Checking, we see

$$6 \xrightarrow{D_m} 12 \xrightarrow{D_m} 24 \xrightarrow{E_m} 49 \xrightarrow{E_m} 47\,.$$

We are now a short step from solving our problem for card shuffling. Observe that

$$\mathscr{O}(x) = \begin{cases} D_m(x) & \text{if } 0 \le x < n \\ E_m(x) & \text{if } n \le x < 2n \end{cases}$$

and

$$I(x) = \begin{cases} E_m(x) & \text{if } 0 \le x < n \\ D_m(x) & \text{if } n \le x < 2n. \end{cases}$$

So, $D_m(x)=\mathscr{O}(x)$ in the top half of the deck and $D_m(x)=I(x)$ in the bottom half. The reverse is true for E_m.

Working from the example above, the shortest sequence of in and out shuffles that will move card 6 to position 47 in a deck of 52 cards is $\mathscr{O}, \mathscr{O}, I, \mathscr{O}$, since

$$6 \xrightarrow[\mathscr{O}]{D_m} 12 \xrightarrow[\mathscr{O}]{D_m} 24 \xrightarrow[I]{E_m} 49 \xrightarrow[\mathscr{O}]{E_m} 47\,.$$

In the first three shuffles $D_m=\mathscr{O}$ and $E_m=I$, since the cards of interest are in the top half of the deck. But, in the last shuffle E_m translates to $\mathscr{O}$ since the card, 49, is in the bottom half of the deck.

This proves and illustrates the following theorem.

THEOREM. *Label the positions in a deck of $2n$ cards 0 through $2n-1$ consecutively with 0 representing the top position. To determine the shortest possible sequence* *of perfect in and out shuffles that will move a card in position i to position j, proceed as follows*:

1. *Calculate the sequences* $D_m(i), D_m^2(i), \dots, D_m^t(i)$ *and* $E_m(i), E_m^2(i), \dots, E_m^t(i)$ *until* $j \in [D_m^t(i), E_m^t(i)]$.
2. *Let* $s = (j - D_m^t(i)) \bmod 2n$ *and write s as a t-digit base two numeral* $s_1 s_2 \dots s_t$.
3. *Reading the digits* $s_1 s_2 \dots s_t$ *from left to right, apply consecutively* D_m *to i if* s_k *is* 0 *and* E_m *if* s_k *is* 1.
4. *Make the translation* $D_m = \mathcal{O}$ *if* D_m *is being applied to an integer in* $[0, n-1]$ *and* $D_m = I$ *if applied to an integer in* $[n, 2n-1]$. *Similarly,* $E_m = I$ *if in* $[0, n-1]$ *and* $E_m = \mathcal{O}$ *in* $[n, 2n-1]$. *The resulting sequence of in and out shuffles* (I*'s and* $\mathcal{O}$*'s*) *moves the card in position i to position j in a minimum number of perfect shuffles.*

This procedure is general. It applies to a deck of any even number of cards, and it can be used to move any card in such a deck to any other position. It is easy and efficient to apply. It is of order $\log n$. It can be used to move any card to the top of the deck by simply letting $j = 0$ in the theorem above.

REFERENCES

1. I. Adler, Make up your own card tricks, *J. of Recreational Math*, Vol. 6, Spring 1973, 87–91.
2. W. W. R. Ball and H. M. S. Coxeter, *Mathematical Recreations and Essays*, 12th edition, University of Toronto Press, Buffalo, NY, 1974.
3. P. Y. Chen, D. H. Lawrie, P. C. Yew, D. A. Podera, Interconnection networks using shuffles, *Computer*, December (1981), 55–64.
4. P. Diaconis, R. L. Graham, W. M. Kantor, The mathematics of perfect shuffles, *Adv. Appl. Math.*, 4 (1983), 175–196.
5. M. Gardner, *Mathematical Carnival*, MAA, Washington, DC, 1989.
6. I. N. Herstein and I. Kaplansky, *Matters Mathematical*, Harper & Row, New York, NY, 1974.
7. E. Marlo, *Faro Shuffle*, Ireland Magic Co., Chicago, IL, 1958.
8. ______, *Faro Notes*, Ireland Magic Co., Chicago, IL, 1958.
9. S. Medvedoff and K. Morrison, Groups of perfect shuffles, this MAGAZINE, Vol. 60 (1987), 3–14.
10. S. B. Morris and R. E. Hartwig, The generalized faro shuffle, *Discrete Math.*, 15 (1976), 333–346.
11. J. W. Rosenthal, Card shuffling, this MAGAZINE, Vol. 54 (1981), 64–67.
12. H. S. Stone, Parallel processing with the perfect shuffle, *IEEE Trans. Comput.* 2 (1971), 153–161.

Probabilities of Clumps in a Binary Sequence (and How to Evaluate Them Without Knowing a Lot)

DAVID M. BLOOM
Brooklyn College of CUNY
Brooklyn, NY 11210

1. Introduction When I was growing up in the 1940s and early '50s, my father, though a non-mathematician, encouraged my already strong interest in mathematics by bringing home books for me: problem and puzzle books, Hogben's *Mathematics for the Million*, Kasner and Newman's *Mathematics and the Imagination*, and others. In December 1985, when my son Eric and I visited him to celebrate his 80th birthday, we found that Dad hadn't changed his ways. He had picked up a copy of one of Martin Gardner's books (namely **[1]**), thinking that Eric and/or I might find in it items of interest. (Martin Gardner needs no introduction to most readers. He wrote regularly about mathematics for *Scientific American* magazine for many years and has written many books—some of them published by the MAA—about mathematical puzzles, curiosities, etc.)

What happened next was this: Eric (then 10) took a look at **[1]**, found in it the assertion (p. 124) that in an ordinary shuffled deck of 52 cards

> "there will almost always be a clump of six or seven [consecutive] cards of the same color," (0)

took out a deck of cards and did the experiment, obtained no such "clump," and came to me for an explanation. *Question*: Did Eric witness an extremely unlikely occurrence, or was **[1]** wrong? That is,

> In the case $(m, k, t) = (26, 26, 6)$, what is the probability that, in a random string of m red and k black objects, some t consecutive objects have the same color? (1)

Essentially the same problem, in a different guise, came to my attention more recently. In December 1992, I had to give a class test and a final exam in a required course for non-majors. To make it harder for a student to copy, I wrote two versions ("odd" and "even") of each exam. At the class test, I gave out the two versions alternately according to where the students had chosen to sit. Afterward, upon marking "o" or "e" (16 odds, 15 evens) next to each name on my alphabetically arranged roster, I was surprised to find that no three consecutive names had had the same version of the test. Even more surprising, the same thing happened at the final exam: Out of 15 "odds" and 17 "evens," no three alphabetically consecutive names had the same version. *Question*: Did I witness an extremely unlikely pair of occurrences, or was my surprise unwarranted? That is,

> In the cases $(m, k, t) = (16, 15, 3)$ and $(15, 17, 3)$, what is the probability that, in a random string of m "odd" and k "even" objects, no t consecutive objects have the same parity? (2)

Clearly, (1) and (2) are different cases of the same problem. What follows is a discussion of some elementary ways to solve it. In particular, we exhibit two (equivalent) recurrences for the probabilities, both of which can be proved by straightforward counting arguments. I found one of these recurrences purely by trial and error (how I did so is described in §3). Subsequently, David M. Jackson of the University of Waterloo showed me a simpler one, which we exhibit (with proof) in §5. In §2, a non-recurrence method is discussed briefly.

2. A false start, and an answer to question (1) What is the probability of a t-clump (some t or more consecutive cards of the same color) in the shuffled deck? When first trying to answer this question, I had a silly mental lapse. I reasoned as if the colors of successive cards were independent (as of course they are not!); i.e., as if the problem were to find the probability $P_t(n)$ that in n consecutive coin-tosses some t-clump occurs. The latter problem is easier than (1). Indeed, there are just two mutually exclusive ways that a t-clump can appear among n tosses: Either (i) a t-clump occurs among the first $n-1$ tosses, or (ii) the last t tosses form a clump, its type (heads or tails) is opposite to that of the $(n-t)$-th toss (unless $n-t=0$), and no t-clump occurs among the first $n-t$ tosses. Thus,

$$P_t(n) = P_t(n-1) + 2^{-t}(1 - P_t(n-t))$$

when $n > t$ (and we have $P_t(n) = 0$ when $n < t$, $P_t(t) = 2 \cdot 2^{-t}$). This recurrence for $P_t(n)$ was easily incorporated into a computer program and produced the value $P_6(52) = .5595\ldots$, which certainly would call into question Gardner's "...almost always... ." So I told Eric, back then in 1985. Three or four years later, I realized that I'd solved the wrong problem!

(Before addressing the *right* problem, note the intuitive likelihood that the correct probability of a 6-clump in the shuffled deck is even smaller than the value of $P_6(52)$ obtained above. If the first coin-toss is heads, the second has a 50% chance of being heads also; but if the first card is red, the probability that the next card is also red is only 25/51.)

OK, what next? It occurred to me to try the well-known *principle of inclusion-exclusion*, which states that if $A_1, \ldots, A_n$ are events and P denotes probability, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = B_1 - B_2 + B_3 - \cdots + (-1)^{n-1} B_n \tag{3}$$

where

$$B_1 = \sum_i P(A_i); \quad B_2 = \sum_{i<j} P(A_i \cap A_j); \quad B_3 = \sum_{i<j<k} P(A_i \cap A_j \cap A_k); \quad \text{etc.}$$

If G is the probability that a 6-clump occurs in a randomly shuffled 52-card deck, then G will equal the quantity (3) if we let n be a sufficiently large integer and then define A_i to be the event that some 6-clump *begins* with the i-th card in the deck; i.e., that cards $i, i+1, \ldots, i+5$ have the same color *and* this color is opposite to that of card $i-1$ if $i > 1$. Thus, e.g., $P(A_i) = 0$ if $i > 47$, $P(A_i \cap A_j) = 0$ if $i < j < i+6$, etc.; and $B_k \neq 0$ only for $1 \leq k \leq 8$. I won't make you wade through the calculations. Suffice it to say that two days' work with a hand calculator (a computer wasn't needed!) produced the bounds

$$.4640 < G < .4644, \tag{4}$$

indicating that a 6-clump won't even appear in the shuffled deck *half the time*—a result wholly incompatible with statement (0). Not sure that I myself hadn't erred, I tried a "random" simulation by computer. Among 2000 simulated "shuffled decks," only about 45% had 6-clumps, a figure roughly 1.3 standard deviations from the value (4) (but in the same ballpark). **[1]** was wrong, after all.

3. A better method The foregoing method, though viable for the particular parameters $(m, k, t) = (26, 26, 6)$ (using the notation of (1)), is far too cumbersome for the general case, which calls for a general recurrence. Yet I had already tried, and failed, to find such a recurrence, using a table obtained by brute-force enumeration for $t = 3$, the smallest nontrivial value of the clump-length t. To fix notation, let $C_t(m,k)$ denote the number of strings of m indistinguishable objects of Type A and k indistinguishable objects of Type B (say 1's and 0's) in which *no* t-clump (run of length t) occurs. The following table (5) gives values of $C_3(m, k)$ for small m, k.

VALUES OF $C_3(m, k)$

$m \backslash k$	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1	1	1	0	0	0	0	0	0	0	0	0	0
1	1	2	3	2	1	0	0	0	0	0	0	0	0
2	1	3	6	7	6	3	1	0	0	0	0	0	0
3	0	2	7	14	18	16	10	4	1	0	0	0	0
4	0	1	6	18	34	45	43	30	15	5	1	0	0
5	0	0	3	16	45	84	113	114	87	50	21	6	1
6	0	0	1	10	43	113	208	285	300	246	157	77	
7	0	0	0	4	30	114	285	518	720	786	683		
8	0	0	0	1	15	87	300	720					
9	0	0	0	0	5	50	246	786					
10	0	0	0	0	1	21	157	683					

(5)

Because the roles of Type A and Type B are interchangeable, the matrix (5) is symmetric. *Challenge*: Can you find a recurrence that generates it? (If you'd like, stop and do some trial-and-error before reading further. It may take a while.) My own attempt had left me stumped.

But the events of December 1992 (my "odd" and "even" tests) brought the problem back to my attention, and I took another look at Table (5). Let's now examine it together. The rows for $m = 0, 1, 2$ are familiar: They are the sequences of coefficients in the expansion of $(1 + x + x^2)^n$ (the trinomial coefficients) for $n = 1, 2, 3$. In particular, each entry in rows $m = 1$ and $m = 2$ of Table (5) (say in column k) is the sum of the three entries from the preceding row in columns $k, k-1, k-2$. However, for $m \geq 3$ this pattern fails; in fact, the rows no longer have left-right symmetry. So let's ask: *By how much* does the pattern fail? Let $D_3(m, k)$ be the answer to that question; that is,

$$D_3(m,k) = \sum_{j=0}^{2} C_3(m-1, k-j) - C_3(m,k). \tag{6}$$

Let's tabulate D_3:

VALUES OF $D_3(m, k)$

$m \backslash k$	0	1	2	3	4	5	6	7	8	9	10
0	−1	−1	−1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	1	2	3	2	1	0	0	0	0	0	0
4	0	1	3	5	5	3	1	0	0	0	0
5	0	1	4	9	13	13	9	4	1	0	0
6	0	0	2	9	21	32	34	26	14	5	1
7	0	0	1	7	24	52	79	88	73	45	20

This time, the rows have left-right symmetry all the way through $m = 5$, with nonsymmetry beginning at $m = 6$, whereas in (5) the nonsymmetry began at $m = 3$. This suggests comparing row 6 of the D_3 matrix with row 3 of the C_3 matrix:

k	0	1	2	3	4	5	6	7	8	9	10
$C_3(3, k)$	0	2	7	14	18	16	10	4	1	0	0
$D_3(6, k)$	0	0	2	9	21	32	34	26	14	5	1

Aha! Do you see the pattern? If not, let's make it easier by shifting the D_3 row $1\frac{1}{2}$ spaces to the left:

(C_3 row)	0	2	7	14	18	16	10	4 1	0
(D_3 row)	2	9	21	32	34	26	14	5	1

and now the scheme is as evident as in Pascal's triangle: $D_3(6, k) = C_3(3, k-1) + C_3(3, k-2)$. A check of other such pairs of rows (row m of D_3 versus row $m-3$ of C_3) reveals a similar pattern:

$$D_3(m, k) = C_3(m-3, k-1) + C_3(m-3, k-2) \tag{7}$$

with *two exceptions*: when $0 \le k \le 2$ and $m = 0$ or 3, the left side of (7) minus the right side equals -1 or $+1$, respectively. Thus, the correct recurrence for C_3 (in view of (6)) appears to be

$$C_3(m, k) = \sum_{i=0}^{2} C_3(m-1, k-i) - \sum_{i=1}^{2} C_3(m-3, k-i) + e(m, k)$$
$$e(m, k) = \left\{ \begin{array}{rl} 1, & \text{if } m = 0 \text{ and } 0 \le k \le 2 \\ -1, & \text{if } m = 3 \text{ and } 0 \le k \le 2 \\ 0, & \text{in all other cases} \end{array} \right\}. \tag{8}$$

Generalizing (8) to arbitrary values of t in place of $t = 3$, a natural guess was that, for all positive integers t and all integers m, k,

$$C_t(m, k) = \sum_{i=0}^{t-1} C_t(m-1, k-i) - \sum_{i=1}^{t-1} C_t(m-t, k-i) + e_t(m, k)$$
$$e_t(m, k) = \left\{ \begin{array}{rl} 1, & \text{if } m = 0 \text{ and } 0 \le k < t \\ -1, & \text{if } m = t \text{ and } 0 \le k < t \\ 0, & \text{in all other cases} \end{array} \right\}, \tag{9}$$

and numerical data (brute-force enumeration again) seemed to confirm it. At this point, I found proof easier than discovery and soon had an elementary combinatorial proof of (9). The proof is available to the reader on request, but will not be given here; instead, we shall exhibit in §5 a simpler such proof of the equivalent Jackson recurrence. At any rate, next on my agenda was to use (9) to obtain numerical results, including answers to the specific questions (1) and (2) posed in §1.

4. Numerical results For a random sequence of m objects of one type and k of another, the probability that a t-clump occurs is clearly

$$P_t(m, k) = 1 - C_t(m, k) \Big/ \binom{m+k}{k}.$$

Using (9), a program to compute $C_t(m, k)$, and hence $P_t(m, k)$, was written, tested, and run for various values of the parameters. One result was

$$P_6(26, 26) = .46424\ldots,$$

agreeing with (4) (at which the programmer felt great relief). Two other results were

$$P_5(26, 26) = .77307\ldots; \qquad P_4(26, 26) = .97396\ldots,$$

so that the phrase "almost always" in assertion (0) still seems exaggerated even with $t = 5$ in place of Gardner's $t = 6$. (For $t = 4$, the phrase seems more appropriate.)

As for my "odd" and "even" tests in December '92: The probability of no 3-clump on the class test was

$$C_3(16, 15) \bigg/ \binom{31}{15} = .0042342\ldots, \tag{10}$$

and the probability of no 3-clump at the final exam was

$$C_3(15, 17) \bigg/ \binom{32}{15} = .0028779\ldots. \tag{11}$$

If the two distributions were independent,* we would conclude that the probability of a 3-clump occurring on *neither* exam was the product of the numbers (10) and (11), namely

$$.00001218\ldots,$$

less than 1 out of 82,000. (And yet it happened. A nonmathematical friend to whom I reported the event—and the odds against its occurrence—reacted thus: "So I *could* win the lottery!")

Postscript. It is easy to find the expected *number* $E = E(m, k, t)$ of noncontiguous t-clumps in a sequence of m 1's and k 0's. Indeed, the probability that the i-th term of the sequence *begins* such a run is $([m]_t + [k]_t)/[m + k]_t$ if $i = 1$ and is $(k[m]_t + m[k]_t)/[m + k]_t(m + k - t)$ if $1 \le i - 1 \le m + k - t$, where $[x]_n$ denotes the product $x(x - 1)\ldots(x - (n - 1))$. (Compare with the discussion following display (3).) Summing over all i, since expectation is additive, we get

$$E(m, k, t) = \{(k + 1)[m]_t + (m + 1)[k]_t\}/[m + k]_t \qquad (1 \le t \le m + k).$$

In particular, $E(26, 26, 6) = .610\ldots$. Since this number clearly must exceed the probability of at least *one* 6-clump in the shuffled deck, we don't even need the actual probability to see that statement (0) is much too strong. Similarly, with respect to our question (2), $E(16, 15, 3) + E(15, 17, 3) = 7.55\ldots$, so that the extraordinariness of "no clumps" seems evident even before we have found the recurrence (8) or (9).

For n (fair) coin tosses, the formula for the expected number of t-clumps (obtained similarly) is even simpler:

$$E(n, t) = (n + 2 - t)/2^t. \tag{12}$$

*Approximate independence, at least, seems likely to me; I had instructed "friends" to sit apart, and the two exams were held in different rooms with different seat layouts.

It has been shown (e.g., in [4]) that if L_n is the length of the *longest* run in a sequence of n tosses, then $E(L_n) \sim \log_2 n$ as $n \to \infty$. (12) makes the latter intuitively plausible. For example, if $n = 2^{1000}$ ($\log_2 n = 1000$), then $E(n, 997)$ is extremely close to 8 and $E(n, 1003)$ is extremely close to 1/8, making it seem very likely that $997 \le L_n \le 1002$. Viewed in this light, it is not surprising that the variance of L_n is nearly constant when n is large, a fact that Schilling's award-winning article [7] calls "remarkable" (as indeed it seems when first encountered). [7] and [4] give more precise expressions for $E(L_n)$ and $\mathrm{Var}(L_n)$.

5. A more efficient recurrence The number of terms on the right side of (9) increases with t. In [6], Jackson gave a partial proof, using the theory of combinatorial generating functions (as developed, e.g., in [5]) of the following alternate recurrence for $C_t(m, k)$, in which the right side has only six terms no matter how large t is:

$$C_t(m,k) = \begin{cases} C_t(m-1,k) + C_t(m,k-1) - C_t(m-t,k-1) - C_t(m-1,k-t) \\ \quad + C_t(m-t,k-t) + e_t^*(m,k) \end{cases}$$

$$e_t^*(m,k) = \left\{ \begin{array}{ll} 1, & \text{if } (m,k) = (0,0) \text{ or } (t,t) \\ -1, & \text{if } (m,k) = (0,t) \text{ or } (t,0) \\ 0, & \text{otherwise} \end{array} \right\}. \tag{13}$$

A referee of this article has pointed out that (13) can in fact be derived without generating functions, as follows:

For fixed t, we call a sequence of 1's and 0's *good* if it contains no t-clump. A good sequence of m ones and k zeros will be denoted by $S(m, k)$; an $S(m, k)$ beginning with the digit i (= 0 or 1) will be denoted by $S_i(m, k)$; and $x^{(t)}$ (where $x = 0$ or 1) will denote the sequence $(x, x, \ldots, x)$ (t terms). Also, let $[A, B]$ denote the sequence consisting of the sequence A followed by the sequence B.

By inspection, (13) holds if $(m, k) = (0, 0)$ or $(0, t)$ or $(t, 0)$, so we assume (m, k) is not one of those three pairs. Since $(m, k) \ne (0, 0)$, every $S(m, k)$ has the form

$$[1, S(m-1,k)] \quad \text{or} \quad [0, S(m,k-1)]. \tag{14}$$

Conversely, since $(m, k) \ne (0, t)$ or $(t, 0)$, a sequence (14) is an $S(m, k)$ if and only if it is not of the form

$$\left[1^{(t)}, 0, S(m-t,k-1)\right] \quad \text{or} \quad \left[0^{(t)}, 1, S(m-1,k-t)\right]. \tag{15}$$

Next, if $(m, k) \ne (t, t)$ then a sequence (15) also has the form (14) if and only if it is not of the form

$$\left[1^{(t)}, 0^{(t)}, S_1(m-t,k-t)\right] \quad \text{or} \quad \left[0^{(t)}, 1^{(t)}, S_0(m-t,k-t)\right]; \tag{16}$$

the excluded sequences (16) all have the form (15); and their number is $C_t(m - t, k - t)$. If instead $(m, k) = (t, t)$, there are exactly two excluded sequences, namely $[1^{(t)}, 0^{(t)}]$ and $[0^{(t)}, 1^{(t)}]$, and $2 = C_t(0, 0) + 1$. In either case, the number of sequences (15) not of the form (14) is $C_t(m - t, k - t) + e_t^*(m, k)$, so that

$$\text{no. of } S(m,k)\text{'s} = (\text{no. of sequences } (14)) - (\text{no. of sequences } (15)) \\ + C_t(m-t,k-t) + e_t^*(m,k),$$

which is precisely (13).

6. Some exercises We asserted (§1) that the recurrences (9) and (13) are equivalent; a nice exercise for the student is to prove this assertion algebraically, without reference to combinatorics. Here are two more such exercises:

I. For fixed t, if $d_n = \Sigma_k C_t(n-k, k)$ (the n-th northeast-to-southwest diagonal sum in the matrix C_t), then

$$d_n = 2^n \ (0 \le n < t); \quad d_n = \sum_{j=1}^{t-1} d_{n-j} \ (n \ge t). \tag{17}$$

For example, when $t = 3$ then $d_n = d_{n-1} + d_{n-2}$ $(n \ge 3)$, and in fact the d's are twice the Fibonacci numbers: $d_n = 2F_{n+1}$ when $n \ge 1$ (see table (5)). (17) can be proved either combinatorially or by induction.

II. For fixed t, if $r_m = \Sigma_k C_t(m, k)$ (the m-th row sum in C_t), then

$$r_m = t^{m+1} \ (0 \le m < t); \quad r_m = (t-1) \cdot \sum_{j=1}^{t-1} r_{m-j} \ (m \ge t). \tag{18}$$

This can be proved by induction using (9); I haven't found a combinatorial argument.

Either (17) or (18) can be used to check the matrix C_t, after constructing C_t from (9) or (13).

7. Related problems Space does not permit a comprehensive listing here of the literature on clump-related problems, but a few quite recent references (called to my attention by a knowledgeable referee) deserve brief mention. Godbole [2] obtains an explicit formula (as a sum) for the probability that, in the first n terms of a sequence of m 1's and k 0's, no run of t consecutive 1's occurs. (I know of no such explicit formula for "consecutive 1's *or* consecutive 0's," which was the problem addressed herein.) Sequences whose successive terms are independent (i.e., no parameters m, k) are easier to deal with, and several articles have done so in considerable generality. In particular, two papers in [3] treat "random n-letter words formed from an r-letter alphabet" ($r \ge 2$): Suman [3, 119–130] obtains three formulas (involving sums) for the probability that no t-clump occurs in such a "word", and Chryssaphinou, et al. [3, 231–241] study the waiting time until at least one of a given set of patterns occurs ("at least one t-clump" would be a special case). For readers wishing to pursue such matters further, the aforementioned articles also contain useful bibliographies; in addition, [3] contains recent articles on other clump-related topics.

Editor's Note. After this paper was accepted, it was pointed out that a recursion for the probability of clumps has been obtained by E. F. Schuster in [3], pp. 91–111. His recurrence is more complicated in that the terms of his recurrence must themselves be obtained from a different recurrence. Schuster presents a table of the probabilities that no t-clump occurs in a sequence of m 1's and n 0's, up to $m + n = 50$. Just short of what's needed for a deck of cards!

REFERENCES

1. Martin Gardner, *aha! Gotcha*, Freeman, New York, NY, 1982.
2. A. P. Godbole, On hypergeometric and related distributions of order k, *Commun. Statist.: Theory and Methods* 19 (1990), 1291–1301.
3. A. P. Godbole and S. Papastavridis (eds.), *Runs and Patterns in Probability: Selected Papers*, Kluwer Academic, Boston, MA, 1994.
4. L. Gordon, M. F. Schilling, and M. S. Waterman, An extreme value theory for long head runs, *Probability Theory and Related Fields* 72 (1986), 279–287.
5. I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, Wiley-Interscience, New York, NY, 1983.
6. D. M. Jackson, personal communication, 3/12/93.
7. M. F. Schilling, The longest run of heads, *Coll. Math. J.* 21 (1990), 196–207.

A Symmetry Criterion for Conjugacy in Finite Groups

SOLOMON W. GOLOMB
University of Southern California
Los Angeles, CA 90089-2565

A standard technique in finite group theory is to partition the elements in some natural way. The two most fruitful partitions of a group are into *cosets* and into *conjugacy classes*. Lagrange's theorem and factor groups are two consequences of partitioning a group into cosets. The class equation and the Sylow theorems are two consequences of partitioning a group into conjugacy classes (for more details, see, e.g., **[1]**.) In this note we give a simple criterion for conjugacy.

DEFINITION. *Two elements a and b in a group G are said to be* conjugate *if for some element g in G, $b = g^{-1}ag$.*

The definition has several simple consequences.

1. Conjugacy is an equivalence relation on G, with respect to which the elements of G are partitioned into *conjugacy classes*.
2. In a commutative group, each element is in a conjugacy class by itself.
3. In a non-commutative group, the identity element is in a conjugacy class by itself (since $g^{-1}eg = e$ for all g), and, more generally, a group element c is in a conjugacy class by itself if and only if c commutes with every element of the group.
4. If G is finite, the size of every conjugacy class in G divides the order of G (the number of elements in G).

To determine whether given elements a and b in G are conjugate, it is certainly *sufficient* to calculate $g^{-1}ag$ for all g in G, and see whether any of these equals b. To determine the conjugacy classes in G, it is *sufficient* to calculate $g^{-1}ag$ for all g and a in G. The purpose of this note is to show that these calculations are unnecessary if the *group table* (or *multiplication table*) of G is already available. Although this criterion for conjugacy is extremely simple to state and to prove, it seems to have eluded generations of writers of textbooks on modern algebra in general and group theory in particular.

CRITERION. *Two elements a and b of a finite group G are conjugate if and only if they can be found symmetrically situated relative to the main diagonal of the group table.*

We restate this criterion as follows:

THEOREM 1. *Distinct elements a and b in G are conjugate if and only if there are elements u and v in G such that $a = uv$ and $b = vu$. (Thus, a is in row u and column v, while b is in row v and column u of the group table, for some elements u and v in G.)*

Proof. The products uv and vu are always conjugate elements in G, because $u^{-1}(uv)u = (u^{-1}u)(vu) = vu$.

Conversely, given conjugate elements a and b in G, there is an element g in G with $b = g^{-1}ag$. Let $u = g$ and $v = g^{-1}a$. Then $uv = g(g^{-1}a) = a$, while $vu = (g^{-1}a)g = b$. This completes the proof.

It is useful to observe that the element a of the group G occurs in every row g in column $g^{-1}a$, in the group table of G.

FIGURE 1 shows the multiplication table of D_3, the *dihedral group of the equilateral triangle*, which is isomorphic to S_3, the group of all permutations on three symbols. (The three vertices of the equilateral triangle are permuted in all possible ways by the elements of D_3; hence the isomorphism with S_3.) This is the smallest noncommutative group.

D_3	I	R_1	R_2	A	B	C
I	I	R_1	R_2	A	B	C
R_1	R_1	R_2	I	B	C	A
R_2	R_2	I	R_1	C	A	B
A	A	C	B	I	R_2	R_1
B	B	A	C	R_1	I	R_2
C	C	B	A	R_2	R_1	I

FIGURE 1

The group table for D_3, with the symmetrically situated pairs (B,C), (A,B), and (A,C) highlighted. By Theorem 1, the conjugate classes in D_3 are easily seen to be {I}, $\{R_1, R_2\}$, and $\{A, B, C\}$.

In FIGURE 1, I is the identity element; R_1 and R_2 are rotations by 120° and 240°, respectively; and A, B, and C are 180° reflections in each of the three axes of the triangle. We note that R_1 and R_2 are symmetrically situated *three times* in the table; that each of the pairs (A, B), (A,C), and (B,C) are found *twice* in symmetric locations in the table; and that I is symmetric only to itself. Thus, the conjugacy classes in D_3 are {I}, $\{R_1, R_2\}$, and $\{A, B, C\}$.

Each pair of conjugate elements a, b, with $a \neq b$, appears at least twice in symmetric positions in the multiplication table for G. The precise multiplicity of occurrence of pairs a and b of conjugate elements in symmetric positions in the multiplication table of G is given in the following theorem.

THEOREM 2. *If a and b are conjugate elements of the finite group G, with $a \neq b$, then there are $n/k = r$ pairs of elements $\{u_i, v_i\}$ in $G \times G$ such that $u_i v_i = a$ and $v_i u_i = b$, where n is the order of G, and k is the size of the conjugacy class in G to which a and b belong.*

Proof. Let a and b belong to a k-element conjugacy class K of the n-element group G, and let $C(a)$ be the subset of elements of G such that $h^{-1}ah = a$ for h in $C(a)$. It is immediate that $C(a)$ is a subgroup of G. Let g_1 be any element of G such that $g_1^{-1}ag_1 = b$. Then every element g_i of the right coset $C(a)g_1$ of $C(a)$ also gives $g_i^{-1}ag_i = b$, and the right cosets of $C(a)$ are in one-to-one correspondence with the elements of K. Thus, the order r of $C(a)$ is n/k. Let the elements of $C(a)$ be $\{h_1, h_2, \ldots, h_r\}$. Then the elements of the right coset $C(a)g_1$ are $\{g_1, g_2, \ldots, g_r\}$ with $g_i = h_i g_1$ for $i = 1, 2, \ldots, r$. Note that

$$g_i^{-1}ag_i = (h_i g_1)^{-1}a(h_i g_1) = g_1^{-1}(h_i^{-1}ah_i)g_1 = g_1^{-1}ag_1 = b$$

for each g_i, $i = 1, 2, \ldots, r$. Let $u_i = g_i$ and $v_i = g_i^{-1}a$. Then $u_i v_i = a$ and $v_i u_i = b$ for each i, where $u_i = g_i$ runs through the r distinct elements of $C(a)g_1$. Conversely, suppose there are any two elements u and v in g with $uv = a$ and $vu = b$. Then $v = u^{-1}a$, and $vu = u^{-1}au = b$; so u is an element g in the coset $C(a)g_1$.

Note. In FIGURE 1, we observed that each of the pairs (A, B), (A, C), (B, C) occurs *twice* in symmetric positions relative to the main diagonal, while the pair (R_1, R_2) occurs *three times*. In view of Theorem 2, this corresponds to the fact that $\{A, B, C\}$ is a conjugacy class with 3 elements, and $6/3 = 2$; while $\{R_1, R_2\}$ is a conjugacy class of 2 elements, and $6/2 = 3$.

REFERENCE

1. Jacobson, N., *Basic Algebra I*, W. H. Freeman, San Francisco, CA, 1974.

More on the Converse of Lagrange's Theorem

GUY T. HOGAN
Norfolk State University
Norfolk, VA 23504

It is certainly true, and clear, beyond peradventure, that the Theorem of Lagrange, which says that the order of a subgroup of a finite group divides the order of the group, is one of the most basic results in the theory of finite groups. See Herstein [**4**, p. 66], or Birkhoff and MacLane [**1**, p. 111]. Indeed, it may be claimed that this was the result that started the "arithmetization" of the theory.

Recently there appeared in this MAGAZINE [**2**, p. 23] and later in [**3**, p. 139] a simple argument, based on the properties of cosets, showing that A_4 (the alternating group on four symbols) has no subgroup of order 6. This, of course, means that the natural converse of Lagrange's Theorem is false, a fact known for almost 200 years [**2**, **3**]. What we offer here is another simple proof of the same result, using nothing more sophisticated than element orders, and Lagrange's Theorem itself.

Using the same notation as in [**2**], we write A_4 as

$$A_4 = \{(1), (12)(34), (13)(24), (14)(23), (123), (132),$$
$$(124), (142), (134), (143), (234), (243)\}.$$

Note that the first four elements listed form a subgroup, V, the Klein four-group, in which the product of any two of the three involutions (elements of order 2) is the third one. It is also worth pointing out that the next eight elements are all of order 3. Suppose there exists a subgroup H of A_4, $|H| = 6$. Since 6 is even, there exists an element, b, of order 2 in H. And since there are eight elements of order 3 in A_4, at least two of them must belong to H. Let t be one of them. Now tbt^{-1} belongs to H and has order 2. If $tbt^{-1} = b$, then t and b commute, and tb would have order 6. But there are no elements of order 6 in A_4. Hence, $tbt^{-1} = c$ is a second element of order 2 in H, so that bc belongs to H as well. Finally, since all three involutions belong to H, it follows that V is a subgroup of H, contradicting Lagrange's Theorem.

for each g_i, $i = 1, 2, \dots, r$. Let $u_i = g_i$ and $v_i = g_i^{-1}a$. Then $u_i v_i = a$ and $v_i u_i = b$ for each i, where $u_i = g_i$ runs through the r distinct elements of $C(a)g_1$. Conversely, suppose there are any two elements u and v in g with $uv = a$ and $vu = b$. Then $v = u^{-1}a$, and $vu = u^{-1}au = b$; so u is an element g in the coset $C(a)g_1$.

Note. In FIGURE 1, we observed that each of the pairs (A, B), (A, C), (B, C) occurs *twice* in symmetric positions relative to the main diagonal, while the pair (R_1, R_2) occurs *three times*. In view of Theorem 2, this corresponds to the fact that $\{A, B, C\}$ is a conjugacy class with 3 elements, and $6/3 = 2$; while $\{R_1, R_2\}$ is a conjugacy class of 2 elements, and $6/2 = 3$.

REFERENCE

1. Jacobson, N., *Basic Algebra I*, W. H. Freeman, San Francisco, CA, 1974.

More on the Converse of Lagrange's Theorem

GUY T. HOGAN
Norfolk State University
Norfolk, VA 23504

It is certainly true, and clear, beyond peradventure, that the Theorem of Lagrange, which says that the order of a subgroup of a finite group divides the order of the group, is one of the most basic results in the theory of finite groups. See Herstein [**4**, p. 66], or Birkhoff and MacLane [**1**, p. 111]. Indeed, it may be claimed that this was the result that started the "arithmetization" of the theory.

Recently there appeared in this MAGAZINE [**2**, p. 23] and later in [**3**, p. 139] a simple argument, based on the properties of cosets, showing that A_4 (the alternating group on four symbols) has no subgroup of order 6. This, of course, means that the natural converse of Lagrange's Theorem is false, a fact known for almost 200 years [**2**, **3**]. What we offer here is another simple proof of the same result, using nothing more sophisticated than element orders, and Lagrange's Theorem itself.

Using the same notation as in [**2**], we write A_4 as

$$A_4 = \{(1), (12)(34), (13)(24), (14)(23), (123), (132), (124), (142), (134), (143), (234), (243)\}.$$

Note that the first four elements listed form a subgroup, V, the Klein four-group, in which the product of any two of the three involutions (elements of order 2) is the third one. It is also worth pointing out that the next eight elements are all of order 3. Suppose there exists a subgroup H of A_4, $|H| = 6$. Since 6 is even, there exists an element, b, of order 2 in H. And since there are eight elements of order 3 in A_4, at least two of them must belong to H. Let t be one of them. Now tbt^{-1} belongs to H and has order 2. If $tbt^{-1} = b$, then t and b commute, and tb would have order 6. But there are no elements of order 6 in A_4. Hence, $tbt^{-1} = c$ is a second element of order 2 in H, so that bc belongs to H as well. Finally, since all three involutions belong to H, it follows that V is a subgroup of H, contradicting Lagrange's Theorem.

REFERENCES

1. G. Birkhoff and S. MacLane, *A Brief Survey of Modern Algebra*, 2nd edition, Macmillan Publishing Co., New York, NY, 1965.
2. J. A. Gallian, *On the Converse of Lagrange's Theorem*, this MAGAZINE, **63** (1993), 23.
3. J. A. Gallian, *Contemporary Abstract Algebra*, 3rd edition, D. C. Heath and Company, Lexington, MA, 1994.
4. I. N. Herstein, *Abstract Algebra*, 2nd edition, Macmillan Publishing Co., New York, NY, 1990.

A Persian Folk Method of Figuring Interest

PEYMAN MILANFAR
SRI International
333 Ravenswood Ave.
Menlo Park, CA 94025

I recently learned a very quick and effective way of estimating monthly payments on a loan. My father showed me the method, having learned it himself from my grandfather, who was a merchant in nineteenth century Iran. While its origins remain a mystery, the method is still in use among merchants all around Iran, and perhaps elsewhere.

My father used the formula:

$$\text{Monthly payment} = \frac{1}{\text{Number of months}}(\text{Principal} + \text{Interest});$$

he calculated the interest as

$$\text{Interest} = \tfrac{1}{2}\,\text{Principal} \times \text{Number of years} \times \text{Annual interest rate}.$$

The *exact* formula, assuming interest accrued monthly, can be found in any basic finance textbook:

$$C = \frac{r(1+r)^N P}{(1+r)^N - 1}, \tag{1}$$

where C is the (exact) monthly payment, r is the monthly interest rate (1/12 the annual interest rate), N is the total number of months, and P is the principal. With this notation, the folk formula becomes

$$C_f = \frac{1}{N}\left(P + \tfrac{1}{2}PNr\right). \tag{2}$$

In many cases, C_f is a surprisingly good approximation to C. As an example, for a 4-year auto loan of \$10,000 at an annual rate of 7% compounded monthly, the exact formula gives monthly payments of \$239.46 while the folk estimate gives \$237.50.

To see why the approximation works, we regard C as a function of r, with all other quantities held fixed. (The singularity in (1) at $r = 0$ can be cancelled out.) A straightforward calculation shows that the first order Maclaurin polynomial for $C(r)$ has the form

$$C(r) \approx \frac{1}{N}\left(P + \frac{1}{2}P(N+1)r\right), \tag{3}$$

which closely resembles the definition of C_f. For a fixed P, when r is sufficiently small and N sufficiently large, the difference between (2) and (3) is small.

REFERENCES

1. G. Birkhoff and S. MacLane, *A Brief Survey of Modern Algebra*, 2nd edition, Macmillan Publishing Co., New York, NY, 1965.
2. J. A. Gallian, *On the Converse of Lagrange's Theorem*, this MAGAZINE, **63** (1993), 23.
3. J. A. Gallian, *Contemporary Abstract Algebra*, 3rd edition, D. C. Heath and Company, Lexington, MA, 1994.
4. I. N. Herstein, *Abstract Algebra*, 2nd edition, Macmillan Publishing Co., New York, NY, 1990.

A Persian Folk Method of Figuring Interest

PEYMAN MILANFAR
SRI International
333 Ravenswood Ave.
Menlo Park, CA 94025

I recently learned a very quick and effective way of estimating monthly payments on a loan. My father showed me the method, having learned it himself from my grandfather, who was a merchant in nineteenth century Iran. While its origins remain a mystery, the method is still in use among merchants all around Iran, and perhaps elsewhere.

My father used the formula:

$$\text{Monthly payment} = \frac{1}{\text{Number of months}}(\text{Principal} + \text{Interest});$$

he calculated the interest as

$$\text{Interest} = \tfrac{1}{2}\,\text{Principal} \times \text{Number of years} \times \text{Annual interest rate}.$$

The *exact* formula, assuming interest accrued monthly, can be found in any basic finance textbook:

$$C = \frac{r(1+r)^N P}{(1+r)^N - 1}, \tag{1}$$

where C is the (exact) monthly payment, r is the monthly interest rate (1/12 the annual interest rate), N is the total number of months, and P is the principal. With this notation, the folk formula becomes

$$C_f = \frac{1}{N}\left(P + \tfrac{1}{2}PNr\right). \tag{2}$$

In many cases, C_f is a surprisingly good approximation to C. As an example, for a 4-year auto loan of \$10,000 at an annual rate of 7% compounded monthly, the exact formula gives monthly payments of \$239.46 while the folk estimate gives \$237.50.

To see why the approximation works, we regard C as a function of r, with all other quantities held fixed. (The singularity in (1) at $r = 0$ can be cancelled out.) A straightforward calculation shows that the first order Maclaurin polynomial for $C(r)$ has the form

$$C(r) \approx \frac{1}{N}\left(P + \frac{1}{2}P(N+1)r\right), \tag{3}$$

which closely resembles the definition of C_f. For a fixed P, when r is sufficiently small and N sufficiently large, the difference between (2) and (3) is small.

Identities on Point-Line Figures in the Euclidean Plane

GUIDO M. PINKERNELL
University of Wales, College of Cardiff
Cardiff CF2 4YH Wales, UK

Introduction In [3] Larry Hoehn gave a proof of a theorem that is known as the Theorem of Pratt–Kasapi [1]. This note collects some ideas on how to generalize the theorem so that it holds not only for the pentagram but also for many other figures of the Euclidean plane and their duals.

THEOREM 1. *In a pentagram* $A_1 A_2 A_3 A_4 A_5$ *with* B_1, B_2, B_3, B_4, B_5 *as the points of intersection of its sides* (FIGURE 1), *the Pratt identity holds*:

$$\frac{A_1B_1}{B_1A_2}\,\frac{A_2B_2}{B_2A_3}\,\frac{A_3B_3}{B_3A_4}\,\frac{A_4B_4}{B_4A_5}\,\frac{A_5B_5}{B_5A_1}=1.$$

Proof. In any triangle ABC the sine rule

$$\frac{\sin\alpha}{\sin\beta}=\frac{a}{b} \tag{1}$$

holds. Apply this to the five "tips" of the star-shaped pentagram, then multiply and simplify the five equations while observing that two adjacent triangles or "star-tips" have equal angles at their common vertex.

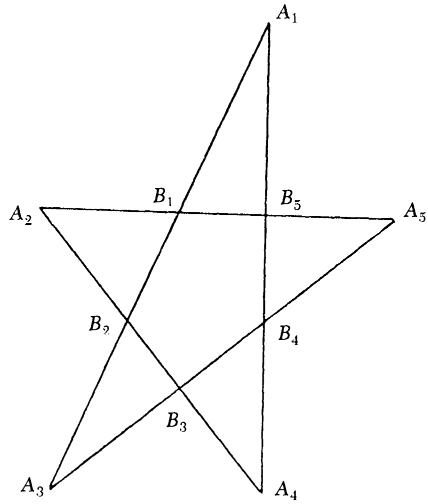
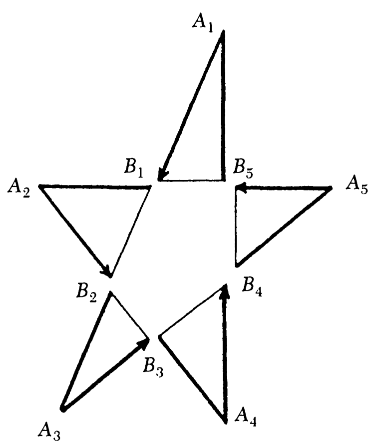


FIGURE 1

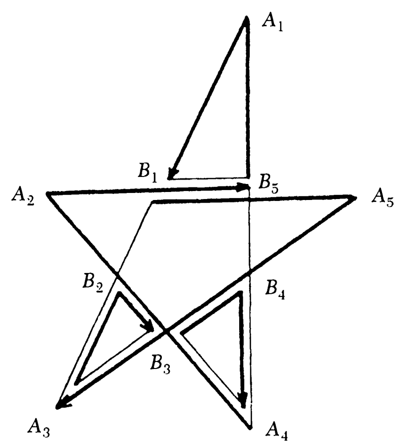
From the idea of the proof one can easily derive a more general, quite straightforward method to prove similar theorems on a pentagram and other figures. Corresponding to the five "tips" $A_{i+1}B_{i+1}B_i$, $(i = 1,\ldots,5)$ of the pentagram where it is understood here and elsewhere that $A_{5+i} = A_i$, one has to find a closed chain of triangles where two adjacent triangles have two common lines as sides. Then the point of intersection of the sides is a common vertex and the angles of each triangle at this point are either vertically opposite, adjacent, or the same angle counted twice. In all cases the sines of the angles are equal, and by multiplying and simplifying the equations (1) for all triangles of the chain we obtain an identity again. Any identity obtained this way by a triangle chain we will call a *Pratt identity*.

It is easy to formulate the identity straightaway once a triangle chain is found: For each triangle, (1) selects two sides, now considered as segments. All these segments form a closed polygon that exactly follows the ratios of the resulting Pratt identity. FIGURE 2 demonstrates the use of a triangle chain for proving Theorem 1; FIGURE 3 shows the result of a different chain in a pentagram.



$$\frac{A_1B_1}{B_1A_2}\,\frac{A_2B_2}{B_2A_3}\,\frac{A_3B_3}{B_3A_4}\,\frac{A_4B_4}{B_4A_5}\,\frac{A_5B_5}{B_5A_1} = 1$$

FIGURE 2



$$\frac{A_1B_1}{B_1A_5}\,\frac{A_5A_3}{A_3B_2}\,\frac{B_2B_3}{B_3B_4}\,\frac{B_4A_4}{A_4A_2}\,\frac{A_2B_5}{B_5A_1} = 1$$

FIGURE 3

Pratt identities on point-line figures The definition of a triangle chain does not depend on the specific order of the points on the lines. From FIGURE 4 it becomes apparent that, when changing the order—here lines AC and DE meeting in B—$\sin\alpha$ still equals $\sin\beta$ while the triangles ABD and CBE in each case have two common lines as sides, as required in the definition. Hence the corresponding Pratt identity remains invariant. For the point-line structure of a pentagram $A_1A_2A_3A_4A_5$, which is given by

ten points	A_i, B_i	$i = 1,\ldots,5$	
and five lines	$a_i = A_iA_{i+2}$	$i = 1,\ldots,5$	(2)
such that	$B_i = a_i \cap a_{i+1}$,	$i = 1,\ldots,5$,	

this argument allows us to state the following generalization of Theorem 1. A point-line figure is said to be *isomorphic* to that of a pentagram if it meets the conditions (2).

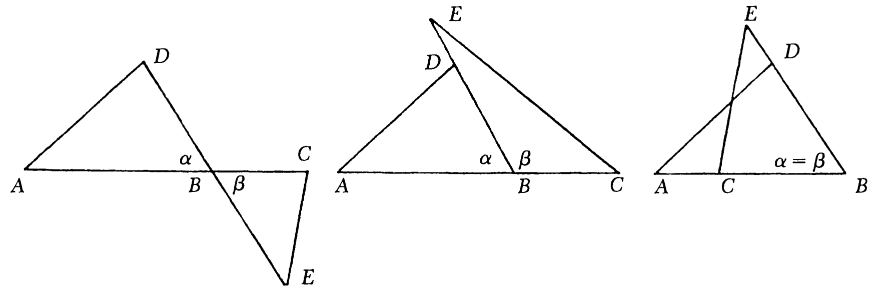
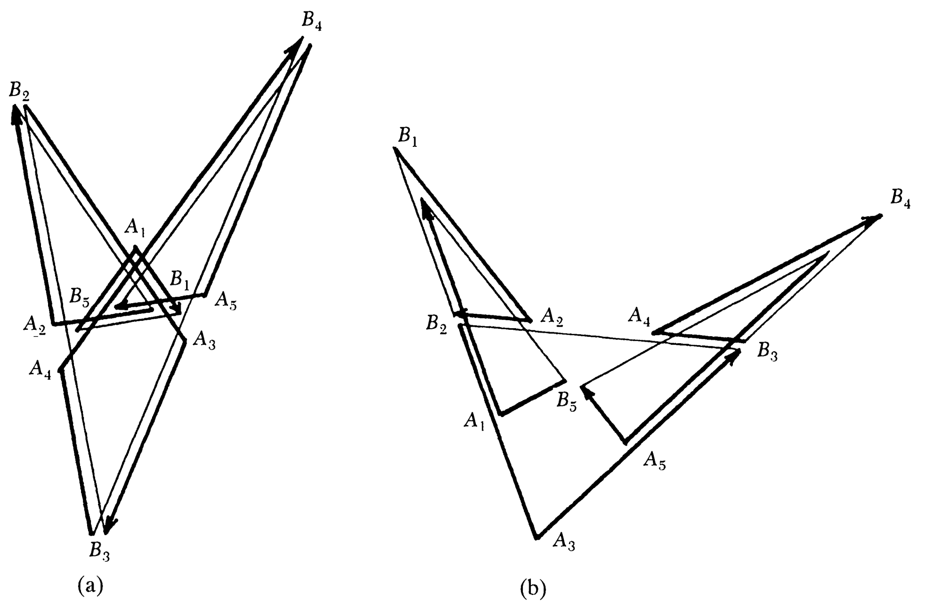


FIGURE 4

THEOREM 2. *The Pratt identity of Theorem 1 holds on any point-line figure isomorphic to that of a pentagram.*

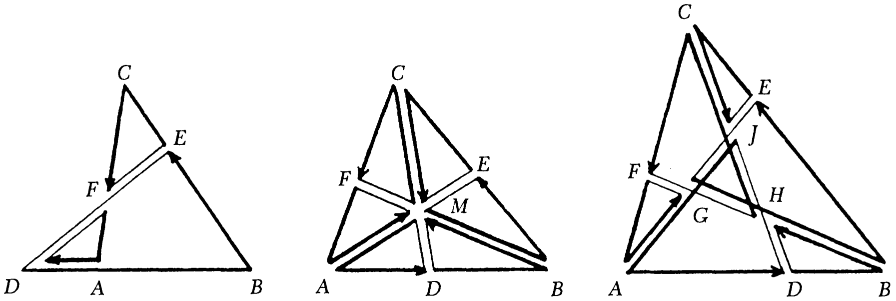
Some examples are given in FIGURE 5(a) and FIGURE 5(b).



$$\frac{A_1B_1}{B_1A_2}\frac{A_2B_2}{B_2A_3}\frac{A_3B_3}{B_3A_4}\frac{A_4B_4}{B_4A_5}\frac{A_5B_5}{B_5A_1}=1$$

FIGURE 5

Triangle chains in other figures It is obvious that triangle chains can be found in any star-shaped n-gon, like a pentagram, heptagram, etc. One has only to take the "tips" of the figure as the triangles. But on the other hand, constructing a simple closed triangle chain without having defined a point-line structure beforehand can lead to Pratt identities on many other figures, some of which might be unfamiliar. FIGURE 6(a) and FIGURE 6(b) show the well-known theorems of Menelaus and Ceva, FIGURE 6(c) could be described as Ceva without concurrent transversals, and FIGURE 7 is a solution of a problem by H. Gülicher **[2]**.



$$\frac{AD}{DB}\frac{BE}{EC}\frac{CF}{FA} = -1$$

(a)

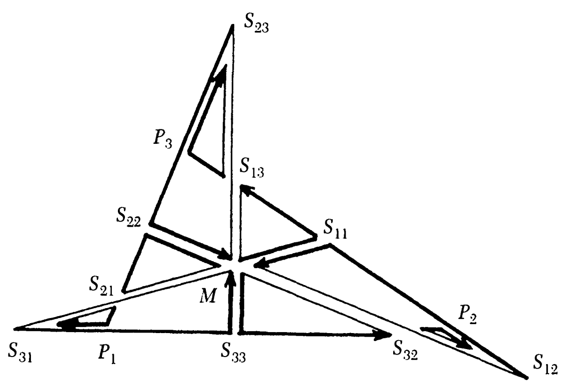
$$\frac{AD}{DB}\frac{BE}{EC}\frac{CF}{FA} = 1$$

(b)

$$\frac{AD}{DB}\frac{BH}{HC}\frac{CF}{FA}\frac{AG}{GB}\frac{BE}{EC}\frac{CJ}{JA} = -1$$

(c)

FIGURE 6

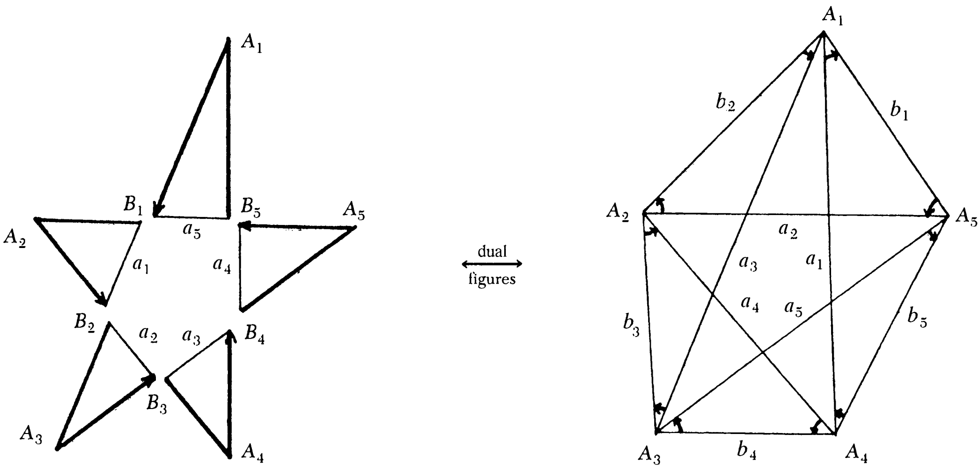


$$\frac{P_1S_{31}}{S_{31}S_{33}}\frac{S_{33}M}{MS_{11}}\frac{S_{11}S_{13}}{S_{13}P_3}\frac{P_3S_{23}}{S_{23}S_{22}}\frac{S_{22}M}{MS_{33}}\frac{S_{33}S_{32}}{S_{32}P_2}\frac{P_2S_{12}}{S_{12}S_{11}}\frac{S_{11}M}{MS_{11}}\frac{S_{11}M}{MS_{22}}\frac{S_{22}S_{21}}{S_{21}P_1} = 1$$

FIGURE 7

One usually decides whether such an identity equals 1 or -1 by observing whether it contains ordered segments in opposite directions on the various lines. Thus we are led to the concept of an ordered point-line figure. We deliberately ignored this concept at the beginning, when we used the sine rule, which does not take account of signs.

Identities on dual figures The symmetry of the sine rule (1) leads to a second group of identities, which holds on the dual of a given figure. When converting every point P into a line p and vice versa without changing the incidences, a triangle ABC of a triangle chain becomes a triangle formed by the three lines a, b, c. And instead of two common sides, two adjacent triangles now have two common points, i.e., a common side considered as segment. (See FIGURE 8.) Hence after multiplying and simplifying the equations (1) of the triangle chain, the right-hand side of each (1) will have vanished and an identity is left consisting only of sines. Interestingly enough, the actual formulation of this new Pratt identity is similar to the original: One simply has to replace "A_iB_j" by "$\sin\angle a_ib_j$," i.e., the sine of the angles in which the lines a_i, b_j



A_iB_j = length of segment A_1B_j

$$\frac{A_1B_1}{B_1A_2}\,\frac{A_2B_2}{B_2A_3}\,\frac{A_3B_3}{B_3A_4}\,\frac{A_4B_4}{B_4A_5}\,\frac{A_5B_5}{B_5A_1}=1$$

a_ib_j = sine of angle at $a_i \cap b_j$

$$\frac{a_1b_1}{b_1a_2}\,\frac{a_2b_2}{b_2a_3}\,\frac{a_3b_3}{b_3a_4}\,\frac{a_4b_4}{b_4a_5}\,\frac{a_5b_5}{b_5a_1}=1$$

FIGURE 8

intersect. What we have exercised on the pentagram can obviously be generalized to any figure where there exists a Pratt identity. Hence

THEOREM 3. *Given any Pratt identity on a point-line figure of the Euclidean plane, involving lengths of segments A_iB_j, there exists a corresponding identity on the dual structure involving sines of angles $\angle a_ib_j$.*

Appendix

To prove Pratt–Kasapi I was originally looking for a different proof via Ceva's theorem. I came across something else that is quite nice and should be mentioned here:

THEOREM 4. *Let M be any point inside the inner pentagon $B_1B_2B_3B_4B_5$ of a convex pentagram with the notation of Theorem 1, and let C_i be the point of intersection of the ray MB_i and the side A_iA_{i+1}, $(i = 1,\dots,5)$ (FIGURE 9). Then*

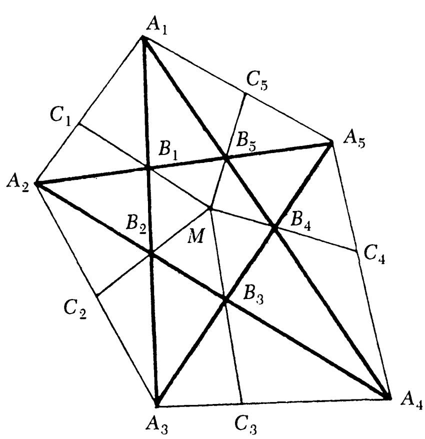


FIGURE 9

$$\frac{A_1C_1}{C_1A_2}\frac{A_2C_2}{C_2A_3}\frac{A_3C_3}{C_3A_4}\frac{A_4C_4}{C_4A_5}\frac{A_5C_5}{C_5A_1}=1. \tag{3}$$

This is a corollary of the following theorem (which is more an instruction to construct a point from given points):

THEOREM 5. *Let C_1, C_2, C_3, C_4 be points on the sides of a convex pentagon $A_1A_2A_3A_4A_5$. If C_5 is constructed as described below, then the equation* (3) *holds.*

Construction of C_5. Let M be any point inside $A_1A_2A_3A_4A_5$. For each of the five triangles $A_iA_{i+1}M$ we will construct—one after another—the situation of Ceva's theorem:

In triangle A_1A_2M let D_1 be a point on the side A_2M. Then define $B_1 = A_1D_1 \cap MC_1$, $D_5 = A_2B_1 \cap MA_1$.

In triangle A_2A_3M define $B_2 = A_3D_1 \cap MC_2$, $D_2 = A_2B_2 \cap MA_3$ (FIGURE 10).

Similarly construct D_3, D_4 in the triangles A_3A_4M, A_4A_5M.

In triangle A_5A_1M define $B_5 = A_1D_4 \cap A_5D_5$. Then $C_5 = A_5A_1 \cap MB_5$ is the point we were looking for.

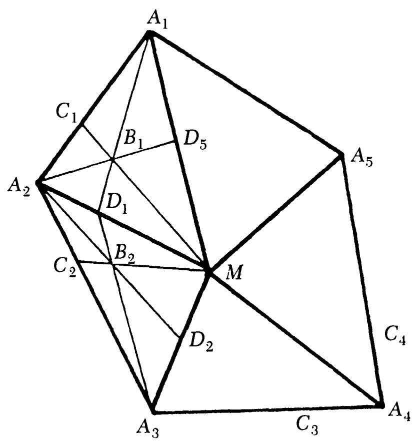


FIGURE 10

For a proof apply Ceva's theorem on each of the triangles $A_iA_{i+1}M$ and multiply and simplify the corresponding equations.

Theorem 4 then follows when A_i, A_{i+2}, D_i are collinear.

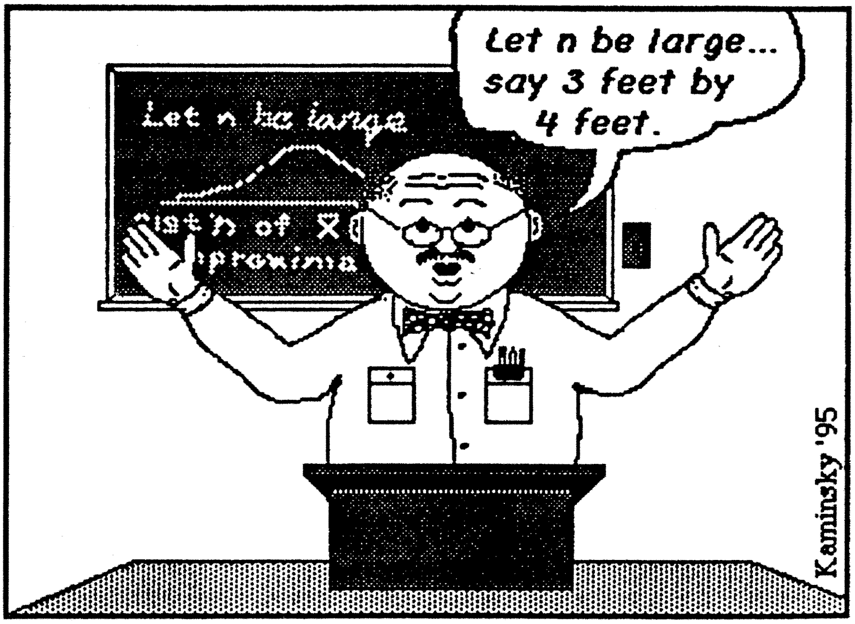
The author is grateful to the referees and to Dr. J. F. Rigby of the University of Wales College of Cardiff for their various helpful comments.

REFERENCES

1. S. C. Chou, *Mechanical Geometry Theorem Proving*, D. Reidel Publishing Co., Dordrecht, 1987.
2. H. Gülicher, Gemeinsamer Schnittpunkt von Transversalen, *Math. Semesterber.* 40 (1993), 91–92.
3. L. Hoehn, A Menelaus-Type Theorem for the Pentagram, this MAGAZINE 66 (1993), 121–123.

Professor Fogelfroe

Professor F. Fogelfroe is Professor of Mathematics at ValuPak™ University, in Margo's Forehead, Minnesota.



Professor Fogelfroe has his own litmus test to see whether or not his students are paying attention.

—Kenneth Kaminsky
Augsburg College
Minneapolis, MN 55454

PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

ZE-LI DOU, KEN RICHARDSON, and SUSAN G. STAPLES, *Assistant Editors*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by May 1, 1997.

1509. *Proposed by David Callan, University of Wisconsin, Madison, Wisconsin.*

Let A be a real $n \times n$ matrix satisfying (i) each row sums to 1; (ii) each entry immediately above the main diagonal is $1/2$; (iii) all other entries above the main diagonal are 0. Prove that the permanent of A is $1/2^{n-1}$.

(The permanent of a matrix is $\sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}$. Thus, it is similar in form to the determinant: $\sum_{\sigma \in S_n} (-1)^{\operatorname{sgn}(\sigma)} \prod_{i=1}^{n} a_{i,\sigma(i)}$.)

1510. *Proposed by Detlef Laugwitz, Technische Hochschule Darmstadt, Darmstadt, Germany.*

Find the largest positive number c such that for every positive integer n, there is at most one perfect square in the set $\{1 + k^2 n \colon 2 \le k \le c\sqrt{n}\}$.

1511. *Proposed by "Ruby Rose Z"L," Pacific Lutheran University, Tacoma, Washington.*

Let $P_1, \dots, P_7$ be 7 points in the plane. Consider the 35 convex polygons $\mathscr{P}_i$ formed by selecting 4 of the 7 points and taking their convex hull. Prove that:

(i) Among any 4 of the polygons $\mathscr{P}_i$ one can always find 3 that have a point in common.

(ii) There are 3 points in the plane such that every polygon contains at least one of the 3 points.

(iii) There are configurations of 7 points for which there do not exist 2 points such that every polygon contains at least one of the 2 points.

We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution.

Solutions should be written in a style appropriate for this MAGAZINE. *Each solution should begin on a separate sheet containing the solver's name and full address.*

Solutions and new proposals should be mailed to George T. Gilbert, Problems Editor, Department of Mathematics, Box 298900, Texas Christian University, Fort Worth, TX 76129, or mailed electronically (ideally as a LaTeX file) to g.gilbert@tcu.edu. Readers who use e-mail should also provide an e-mail address.

1512. *Proposed by Arthur L. Holshouser, Charlotte, North Carolina, and Benjamin G. Klein, Davidson College, Davidson, North Carolina.*

Let R be a commutative ring such that $x^3 = x$ for every $x \in R$. For $x, y \in R$, let $F(x, y) = xy - x^2y - xy^2 - x^2y^2$. If $F(a, b) = a$ and $F(b, c) = b$, prove that $F(a, c) = a$.

1513. *Proposed by Loren C. Larson, St. Olaf College, Northfield, Minnesota.*

Can every set of $4n$ points in the plane, no three of which are collinear, be evenly quartered by two mutually perpendicular lines?

(The original, continuous version of this question appeared in Hugo Steinhaus' *One Hundred Problems in Elementary Mathematics*, Dover, 1979, 26.)

Quickies

Answers to the Quickies are on page 391.

Q856. *Proposed by Jerrold W. Grossman and Stephen Mellendorf, Oakland University, Rochester, Michigan.*

A football league with $2n$ teams draws up the first two weeks of its schedule such that each team plays one game each week. There is no restriction on two teams meeting more than once. Assume that each team in a given game has a fifty percent probability of winning and that the results of the games are independent.

(i) Determine the probability that each team's record is 1-1 at the end of two weeks (as a function of the schedule).

(ii) Assume that each week's schedule is a random matching. Determine the probability that each team's record is 1-1 at the end of two weeks.

Q857. *Proposed by Wu Wei Chao, He Nan Normal University, Xin Xiang City, He Nan Province, China.*

Let $0 < a_1 \leq a_2 \leq \cdots \leq a_n$ and $0 < b_1 \leq b_2 \leq \cdots \leq b_n$ be given, with $\Sigma_{i=1}^n a_i \geq \Sigma_{i=1}^n b_i$. In addition, assume there exists k with $1 \leq k \leq n$ so that $b_i \leq a_i$ if $i \leq k$ and $b_i \geq a_i$ if $i > k$. Prove that

$$\prod_{i=1}^n a_i \geq \prod_{i=1}^n b_i.$$

Q858. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Show that the Diophantine equation

$$x^2 + 27y^2z^2(y+z)^2 = 4(y^2 + yz + z^2)^3$$

has an infinite number of integral solutions (x, y, z) with x, y, and z relatively prime and $xyz \neq 0$.

Solutions

An Euler's Phi-Function Congruence — December 1995

1484. *Proposed by Lenny Jones, Shippensburg University, Shippensburg, Pennsylvania.*

Let $\sigma(n)$ be the sum of the positive divisors of the positive integer n and let $\phi(n)$ be Euler's totient function. For an arbitrary positive integer k, find all positive integers n that satisfy

$$n^k\sigma(n) \equiv 2 \pmod{\phi(n)}.$$

Solution by L. L. Foster, Northridge, California.

For a given $k \geq 1$, the congruence is satisfied when $n = 1$, a prime, 4, or $2q$, where q is any odd prime having the property that $(q-1)/2$ divides $3\cdot 2^k - 1$.

The congruence is easily seen to hold for all k when $n = 1$, any prime, 4, or 6. For a given k, suppose that n is another integer such that the congruence holds. Write $n = \prod_{i=1}^{s} p_i^{\alpha_i}$, where the p_i are distinct primes and the α_i are positive integers. Then

$$\prod_{i=1}^{s} p_i^{\alpha_i k}\sigma(n) \equiv 2 \left(\operatorname{mod} \prod_{i=1}^{s} p_i^{\alpha_i - 1}(p_i - 1)\right).$$

If $\alpha_i > 1$, then $p_i^{\alpha_i - 1}$ divides 2, hence $p_i = 2$ and $\alpha_i = 2$. In this case, $n = 4m$, where m is odd. If $m > 1$, then $\phi(n) \equiv 0 \pmod 4$ and 4 divides 2, a contradiction. We have the same contradiction if we suppose that n has two (distinct) odd prime factors p_i and p_j. For then 4 divides $(p_i + 1)(p_j + 1)$ which divides $\sigma(n)$, and 4 divides $(p_i - 1)(p_j - 1)$ which divides $\phi(n)$. Hence we need only consider $n = 2q$, where $q > 3$ is an odd prime. It follows that $\phi(n) = q - 1$ and

$$n^k\sigma(n) = 3\cdot 2^k q^k(q+1) \equiv 2 \pmod{q-1}.$$

Equivalently, $3\cdot 2^k - 1 \equiv 0 \pmod{(q-1)/2}$. It is not difficult to prove that such an odd prime q must be of the form $12m - 1$. The examples $q = 71$ and $q = 83$ show this is not sufficient.

Also solved by Robin Chapman (United Kingdom), John Christopher, Con Amore Problem Group (Denmark), D. Kipp Johnson, O. P. Lossers (The Netherlands), Heinz-Jürgen Seiffert (Germany), Western Maryland College Problems Group, and the proposer.

Integrality of the Arithmetic-Geometric Mean Ratio — December 1995

1485. *Proposed by Yasutoshi Nomura, Hyogo University of Teacher Education, Hyogo, Japan.*

Let $n > 1$ be a natural number and consider the statement Q_n:

There exist positive integers $x_1, x_2, \ldots, x_n$ for which the arithmetic-geometric mean quotient $\dfrac{x_1^n + \cdots + x_n^n}{nx_1 \cdots x_n}$ is an integer greater than 1.

(a) Show that Q_2 is false.
(b) Show that Q_n is true for even $n > 2$ or for prime n congruent to 5 modulo 6.
(c)* Find another n for which Q_n is false or an infinite family for which it is true.

Solution by John S. Sumner and Kevin L. Dove, University of Tampa, Tampa, Florida.

(a) Suppose $(x_1^2+x_2^2)/(2x_1x_2)=k$ for some integer k. By factoring and canceling, we may assume that x_1 and x_2 are relatively prime. Then $x_1^2+x_2^2=2kx_1x_2$, so that x_1 divides x_2^2 and x_2 divides x_1^2. Hence $x_1=x_2=1$, proving that $k=1$.

(b) If n is even, then $(n-1)^{n-1}\equiv -1 \pmod n$. For $1\le i\le n-1$, let $x_i=1$, and let $x_n=n-1$. Then

$$\frac{x_1^n+\cdots+x_n^n}{nx_1\dots x_n}=\frac{1+(n-1)^{n-1}}{n}$$

is an integer greater than 1, since $n>2$.

Similarly suppose n is a prime congruent to 5 (mod 6). Let $x_{n-1}=n-1$, let $x_n=n^2-3n+3$, and otherwise let $x_i=1$. Since $n^2-3n+3=(n-1)(n-2)+1$, it follows that $n-1$, n, and n^2-3n+3 are pairwise relatively prime. It is straightforward to show that

$$(n-1)^5\equiv -n+2 \pmod{n^2-3n+3},$$

and then that

$$(n-1)^6\equiv 1 \pmod{n^2-3n+3}.$$

Using these observations and the fact that $3^n\equiv 3 \pmod n$, it is clear that

$$\frac{x_1^n+\cdots+x_n^n}{nx_1\dots x_n}$$

is an integer exceeding 1.

(c) Let p be an odd prime and $k>1$ a natural number. We show that Q_n is true for $n=p^k$.

For $1\le i\le n-1$, let $x_i=1$, and let $x_n=p^{k-1}+p^{k-2}+\cdots+p+1$. Define m so that $p^{k-1}+p^{k-2}+\cdots+p=pm$. The binomial theorem yields

$$(p^{k-1}+p^{k-2}+\cdots+p+1)^{p^k}=\sum_{i=0}^{p^k}\binom{p^k}{i}p^im^i.$$

If $k\le i\le p^k$, then clearly p^k divides $\binom{p^k}{i}p^i$. The power of p dividing $i!$ for $1\le i$ is

$$\left\lfloor\frac{i}{p}\right\rfloor+\left\lfloor\frac{i}{p^2}\right\rfloor+\cdots\le\frac{i}{p}+\frac{i}{p^2}+\cdots=\frac{i}{p-1}<i.$$

Thus, for $1\le i<k$, we see that

$$\binom{p^k}{i}=\frac{p^k(p^k-1)\cdots(p^k-i+1)}{i!}$$

is divisible by p^{k-i}. Hence

$$(p^{k-1}+p^{k-2}+\cdots+p+1)^{p^k}\equiv 1 \pmod{p^k}.$$

Using this and the fact that $p^k - 1 = (p-1)(p^{k-1} + \cdots + p + 1)$, it is clear that $(x_1^n + \cdots + x_n^n)/(nx_1 \ldots x_n)$ is an integer exceeding 1.

Comment. The editors of this section will consider *significant* progress on part (c) for publication during their term as editors.

Parts (a) and (b) also solved by Robin Chapman (United Kingdom), Con Amore Problem Group (Denmark), Gerald A. Heuer, D. Kipp Johnson, Yanir and Zalman Rubinstein (Israel), Michael Vowe (Switzerland), and the proposer. Part (a) was also solved by Can A. Minh (student).

A Remainder for a Logarithmic Series **December 1995**

1486. *Proposed by Paul Bracken, University of Waterloo, Waterloo, Ontario, Canada.*

For $-1 < x$, $x \neq 0$, define the sequence $\theta_n(x)$ by

$$\log(1+x) = x - \frac{x^2}{2} + \cdots + (-1)^{n-1}\frac{\theta_n(x)x^n}{n}.$$

Show that the sequence (θ_n) is monotonic in n and find its limit.

Solution by Frank A. Horrigan, Raytheon Electronic Systems, Tewksbury, Massachusetts.

By repeated division or other means,

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \cdots + (-1)^{n-1}\frac{t^{n-1}}{1+t}.$$

Integrating term-by-term, for all $x > -1$,

$$\log(1+x) = \int_0^x \frac{dt}{1+t} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n-1}\int_0^x \frac{t^{n-1}}{1+t}\,dt.$$

Excluding the trivial case $x = 0$,

$$\theta_n(x) = \frac{n}{x^n}\int_0^x \frac{t^{n-1}}{1+t}\,dt = \int_0^1 \frac{ds}{1+xs^{1/n}}$$

for all $n > 0$.

As n approaches ∞, $s^{1/n}$ approaches 1 for $0 < s \le 1$, hence

$$\lim_{n\to\infty}\theta_n(x) = \frac{1}{1+x}.$$

To show monotonicity, consider the difference of successive terms,

$$\theta_{n+1}(x) - \theta_n(x) = \int_0^1 \left(\frac{1}{1+xs^{1/(n+1)}} - \frac{1}{1+xs^{1/n}}\right)ds$$

$$= x\int_0^1 \frac{(s^{1/n} - s^{1/(n+1)})}{(1+xs^{1/(n+1)})(1+xs^{1/n})}\,ds.$$

Since the integrand is negative for $0 < s < 1$ and $-1 < x$, the integral is negative.

Therefore $(\theta_n(x))$ is monotonic in n, depending on the sign of the variable x:

$$\theta_n(x) < \theta_{n+1}(x) < \frac{1}{1+x} \text{ for } -1 < x < 0,$$

$$\theta_n(x) > \theta_{n+1}(x) > \frac{1}{1+x} \text{ for } 0 < x.$$

Also solved by David M. Bloom, Robin Chapman (United Kingdom), Con Amore Problem Group (Denmark), Hans Kappus (Switzerland), Benjamin G. Klein, Kee-Wai Lau (Hong Kong), Nick Lord (England), O. P. Lossers (The Netherlands), Roger Pinkham, Heinz-Jürgen Seiffert (Germany), Kenneth Schilling, Michael Vowe (Switzerland), and the proposer. There was one incorrect solution.

Concurrency in Tangent Circles **December 1995**

1487. *Proposed by Edward Kitchen, Santa Monica, California.*

Given circles $\mathscr{C}$ and $\mathscr{C}'$ with centers O and O', and circles $\mathscr{C}_1$ and $\mathscr{C}_2$ externally tangent to $\mathscr{C}$ at points M_1 and M_2 and internally tangent to $\mathscr{C}'$ at points N_1 and N_2, prove that the lines M_1N_1, M_2N_2, and OO' are concurrent.

I. *Solution by Hoe Teck Wee, Lengkok Bahru, Singapore.*

The result is trivial if M_i, N_i, O, and O' are collinear, for $i = 1$ or 2. Hence, we may assume that M_i, N_i, O, and O' are not collinear, for $i = 1$ and 2. Then, let P_i denote the point of intersection between the lines OO' and M_iN_i, and let O_i denote the center of the circle $\mathscr{C}_i$. Also, let r, r', and r_i denote the radii of $\mathscr{C}$, $\mathscr{C}'$, and $\mathscr{C}_i$, respectively. In terms of directed line segments, we have

$$\frac{O_iM_i}{OM_i} = -\frac{r_i}{r} \quad \text{and} \quad \frac{O'N_i}{O_iN_i} = \frac{r'}{r_i}.$$

Applying Menelaus' Theorem to $\Delta OO'O_i$, we have

$$\frac{OP_i}{O'P_i} \cdot \frac{O'N_i}{O_iN_i} \cdot \frac{O_iM_i}{OM_i} = 1,$$

hence

$$\frac{OP_i}{O'P_i} = -\frac{r}{r'}.$$

Therefore, $P_1 = P_2$, so M_1N_1, M_2N_2, and OO' are concurrent.

II. *Solution by Michael Woltermann, Washington and Jefferson College, Washington, Pennsylvania.*

Let C'' with center O'' be externally tangent to C at M and internally tangent to C' at N. Then M is between O and O'', and N is between O' and O''. If O, O', and O'' are not collinear, then ray $\overrightarrow{NM}$ intersects segment OO' at a point P. Applying the law of sines to ΔOMP and $\Delta O'NP$,

$$\frac{OP}{\sin\angle OMP} = \frac{OM}{\sin\angle OPM} \quad \text{and} \quad \frac{O'P}{\sin\angle O'NP} = \frac{O'N}{\sin\angle O'PN}.$$

Since $O''M = O''N$, we have $\angle OMP \cong \angle O''MN \cong \angle O'NP$, while $\angle OPM$ and $\angle O'PN$ are supplementary. Thus

$$\frac{OP}{O'P} = \frac{OM}{O'N}.$$

If O, O' and O'' are collinear, the line MN contains segment OO'. Since the point P that divides segment OO' into segments proportional to the radii of C and C' is uniquely determined, it follows that lines M_1N_1, M_2N_2, and OO' are concurrent at P.

Also solved by Anchorage Math Solutions Group, Francisco Bellot Rosado and María Ascensión López (Spain), Robin Chapman (United Kingdom), Con Amore Problem Group (Denmark), Hans Kappus (Switzerland), Victor Kutsenok, Neela Lakshmanan, Jianyuan Liu (China), Kwan Sze Ming (Hong Kong), Jose Heber Nieto (Venezuela), David Zhu, and the proposer.

A Product and Sum Inequality — December 1995

1488. *Proposed by Heinz-Jürgen Seiffert, Berlin, Germany.*

Let n be a positive integer. Show that if $0 < x_1 \le x_2 \le \cdots \le x_n$, then

$$\left(\prod_{i=1}^{n}(1+x_i)\right)\left(\sum_{j=0}^{n}\prod_{k=1}^{j}\frac{1}{x_k}\right) \ge 2^n(n+1),$$

with equality if, and only if, $x_1 = x_2 = \cdots = x_n = 1$. (Empty products are understood to be unity.)

Solution by Kee-Wai Lau, Hong Kong.

From $1 + x_i \ge 2\sqrt{x_i}$, it follows that $\prod_{i=1}^{n}(1+x_i) \ge 2^n\sqrt{x_1x_2\cdots x_n}$. Thus we need only prove that

$$\sqrt{x_1x_2\cdots x_n}\left(\sum_{j=0}^{n}\prod_{k=1}^{j}\frac{1}{x_k}\right) \ge n+1.$$

The left-hand side of this expression equals

$$\sqrt{x_1x_2\cdots x_n} + \sqrt{\frac{x_2x_3\cdots x_n}{x_1}} + \sqrt{\frac{x_3x_4\cdots x_n}{x_1x_2}} + \cdots + \sqrt{\frac{x_n}{x_1x_2\cdots x_{n-1}}} + \frac{1}{\sqrt{x_1x_2\cdots x_n}}.$$

By the arithmetic mean-geometric mean inequality, the last expression is greater than or equal to

$$(n+1)\left[(x_1x_2\cdots x_n)\left(\frac{x_2x_3\cdots x_n}{x_1}\right)\left(\frac{x_3x_4\cdots x_n}{x_1x_2}\right) \times \cdots \left(\frac{x_n}{x_1x_2\cdots x_{n-1}}\right)\left(\frac{1}{x_1x_2\cdots x_n}\right)\right]^{1/(2n+2)}$$

$$= (n+1)\prod_{k=1}^{\lfloor n/2\rfloor}\left(\frac{x_{n+1-k}}{x_k}\right)^{(n+1-2k)/(2n+2)}$$

Since $x_n \ge x_{n-1} \ge \cdots \ge x_1$, or more generally if $x_{n+1-k} \ge x_k$ for $k \le \lfloor n/2\rfloor$, the terms in the last product are greater than or equal to 1. The claimed inequality follows. Retracing the steps in the proof above, it is easy to see that equality holds if and only if $x_1 = x_2 = \cdots = x_n = 1$.

Also solved by J. C. Binz (Switzerland), David Callan, Robin Chapman (United Kingdom), Con Amore Problem Group (Denmark), Qais Haider Darwish (Oman), Tim Flood, Jennifer Hoornstra, Jianyuan Liu (China), O. P. Lossers (The Netherlands), Kwan Sze Ming (Hong Kong), Can A. Minh (student), Kenneth Schilling, Achilleas Sinefakopoulos (student, Greece), Michael Vowe (Switzerland), Robert J. Wagner, Hoe Teck Wee (Singapore), Western Maryland College Problems Group, David Zhu, and the proposer.

Answers

Solutions to the Quickies on page 385.

A856. (i) Form a graph whose vertices are the $2n$ teams, and place an edge between two vertices if they play each other during the first two weeks. (If rematches occur, there are two parallel edges.) Clearly every vertex has degree two, so the graph is a disjoint union of cycles. Once the outcome of one game within a cycle has been determined, there is a unique set of outcomes of the rest of the games within that cycle that will result in each team in the cycle having a 1-1 record. Therefore, the probability of all $2n$ teams ending up at 1-1 after two weeks is $(1/2)^{2n-c}$, where c is the number of cycles.

(ii) For every team to end up with a 1-1 record, the winners in the second week must be exactly the losers from the first week. We may specify the matchings and outcomes for the second week by first choosing the n winning teams and then pairing each winning team with a losing team. From this, we see that the probability that each team's record is 1-1 at the end of two weeks is $1/\binom{2n}{n}$.

Comment. The method of solution shows that for every two-week schedule, if one assumes each team has a chance of winning in each game, it is possible for all $2n$ teams to end up with a 1-1 record.

A857. Suppose that $a_1, \ldots, a_n, b_1, \ldots, b_n$ give a counterexample to the claim. Let $a'_i = a_k$ and $b'_i = b_i a_k / a_i$ for $i = 1, \ldots, n$. Then $a'_i - b'_i = (a_i - b_i)a_k/a_i \geq a_i - b_i$. Permuting the i's if necessary, we see that $a'_1, \ldots, a'_n, b'_1, \ldots, b'_n$ also give a counterexample to the claim. Applying the arithmetic-geometric mean inequality to $b'_1, \ldots, b'_n$ yields $\prod_{i=1}^n b'_i \leq ((1/n)\sum_{i=1}^n b'_i)^n \leq ((1/n)\sum_{i=1}^n a'_i)^n = \prod_{i=1}^n a'_i$, a contradiction.

A858. *I.* Since $4(y^2 + yz + z^2)^3 - 27y^2z^2(y+z)^2 = (y-z)^2(y+2z)^2(z+2y)^2$, the general solution is immediate.

II. Provided by the Editors. The roots of $f(t) = t^3 - (y^2 + yz + z^2)t + yz(y+z)$ are y, z, and $-y-z$. Therefore, the discriminant of this cubic satisfies

$$4(y^2 + yz + z^2)^3 - 27y^2z^2(y+z)^2 = (y-z)^2(y+2z)^2(2y+z)^2.$$

Thus every integral pair (y, z) with $yz(y-z)(y+2z)(2y+z) \neq 0$ gives rise to two integral solutions (x, y, z) to the given equation with $xyz \neq 0$. In particular, we may take y and z to be distinct, relatively prime, positive integers.

Correction

Q841, December 1995. The word "nonhomogeneous" in the problem statement should read "homogeneous."

REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of mathematics literature. Readers are invited to suggest items for review to the editors.

Prime time, *New York Times Magazine* (6 October 1996) 31. News track: Largest prime, *Communications of the Association for Computing Machinery* 39 (11) (November 1996) 10. Holden, Constance (ed.), Random samples: Grassroots search for primes ... , *Science* 273 (9 August 1996) 743. Gillmore, Dan, Computer scientists make a prime discovery that's too long to print here, *Newark (NJ) Star-Ledger* (4 September 1996) 5.

A new largest known Mersenne prime (and largest known prime) has been identified: $M_{34} = 2^{1,257,787} - 1$. Like recent predecessors, this Mersenne prime was found by David Slowinski and co-workers at Cray Research. But why leave it to Cray? Throughout the U.S. and the world, there are millions of microcomputers like yours—idle just about all day (and night). Why not donate your computer's unused potential to mathematics? You too—well, your Pentium-processor microcomputer—can now join "The Great Internet Mersenne Prime Search"! More than 400 volunteers are working with the project, at `http://ourworld.compuserve.com/homepages/justforfun/prime.htm` . In fact, the project was 90% of the way through checking the very same number when the announcement came from Cray Research.

Calinger, Ronald (ed.), *Vita Mathematica: Historical Research and Integration with Teaching*, MAA, 1996; xii + 358 pp, \$34.95(P). ISBN 0-88385-097-4. Bos, Henk J.M., *Lectures in the History of Mathematics*, American Mathematical Society, 1993; x + 197 pp, \$82. ISBN 0-8218-9001-8.

The first of these two books celebrates and advocates the idea that the history of mathematics should be incorporated in the teaching of mathematics. Three essays discuss historiography and the use of sources, the bulk of the book is devoted to specific historical studies, and a dozen articles deal with integration of history into mathematics teaching, including the origins and teaching of calculus. The second book collects essays by a single researcher in the history of mathematics, who presents historical studies but also considers mathematics in a larger historical context. Particularly engaging is his last essay, "Mathematics and its social context ... ," which takes up the big questions about the relations between mathematics and society, the mathematical needs of adult life, and historical "laws" about the development of mathematics.

Kolata, Gina, Paul Erdős, 83, a wayfarer in math's vanguard, is dead, *New York Times* (24 September 1996) A1, B8. Pearson, Richard, Paul Erdős: An eccentric titan of mathematical theory, dies, *Washington Post* (24 September 1996). Krauthammer, Charles, Paul Erdős, sweet genius, *Washington Post* (27 September 1996).

Paul Erdős (1913–1996), with whom hundreds of mathematicians wrote joint papers, died of a heart attack while in Warsaw at a mathematics conference.

Hively, Will, Math against tyranny, *Discover* 17 (11) (November 1996) 74–85. Ramirez, Anthony, Why the election is like baseball, *New York Times* (3 November 1996), Section 4, 4. Natapoff, A., A mathematical one-man one-vote rationale for Madisonian presidential voting based on maximum individual voting power, *Public Choice* 88 (1996) 259ff.

Just about when this MAGAZINE is delivered, the electoral college will elect the president of the U.S. Voters in November chose electors pledged to candidates, with each state getting as many electoral votes as seats in Congress (plus three for the District of Columbia). In some past elections, the electoral college did not choose the winner of the popular vote. So, is the electoral college a good idea? Yes, says Alan Natapoff, a physicist at MIT, who has modeled the electoral process and proved a theorem: Individual voting power—the probability that the person's vote will decide the election—is higher under the current electoral system than under direct national election, unless the gap in voter preference between the candidates is razor thin. "The Madisonian system, by requiring candidates to win states on the way to winning the nation, has forced majorities to win the consent of minorities, checked the violence of factions, and held the country together." The connection with baseball? In the World Series, the most runs doesn't necessarily win the series—they "must be grouped in a way that wins games, just as popular votes must be grouped in a way that wins states."

Hildebrandt, Stefan, and Anthony Tromba, *The Parsimonious Universe: Shape and Form in the Natural World*, Springer-Verlag, 1996; xiv + 330 pp, $32. ISBN 0–387–97991–3.

Coffee-table books featuring mathematics appear rarely. This issue of this MAGAZINE may reach you just in time to order holiday gifts. The less mathematically inclined your friends may be, the more appropriate is this book, as it investigates natural forms without using mathematical formulas and symbols—even those who profess to detest mathematics can be seduced by geometry. This book is a revised and enlarged version of *Mathematics and Optimal Form* (W.H. Freeman, 1984). It takes the reader from Maupertuis's principle ("Nature always minimizes action"), through astronomical theories, the Steiner problem, stable equilibria, soap films and isoperimetric problems, and on to optimization in nature. It is very richly illustrated with drawings, figures, photographs, and reproductions of paintings.

TUG'95: Questions and answers with Prof. Donald E. Knuth, TUGBOAT*: The Communications of the TEX Users Group* 17 (1) (March 1996) 7–22. Knuth comments on code, *Byte* 21 (9) (September 1996) 60. Erickson, Jonathan, Letters, we get letters ... , *Dr. Dobb's Journal* 21 (9) (September 1996) 6.

"Computer programs are the most complicated things that humans have ever created." So says Donald Knuth, world-famous computer scientist. He retired recently at age 57 to work full-time on completing his *magnum opus*, the seven volumes of *The Art of Computer Programming*, which he started in 1962 and suspended for many years to create the TEX document preparation system and the METAFONT font design system. He estimates that it will take 20 more years to finish the remaining four volumes, at a rate of 256 pages per year. Knuth was the recipient of the Inamori Foundation's 1996 Kyoto Prize in the category of Advanced Technology; along with a suitable plaque came $460,000.

Passell, Peter, 2 theorists of real-life problems get Nobel, *New York Times* (9 October 1996) (National Edition) C1, C4.

James A. Mirrlees (Cambridge University) and William Vickrey (Columbia University) were awarded the 1996 Nobel Prize in Economic Science. Vickrey, who died suddenly three days after the award was announced, was the inventor of the Vickrey auction, in which the highest bidder wins but pays only the second-highest bid. Such an auction encourages the bidders to bid the maximum amounts that they are willing to pay.

sci.math FAQ [Frequently asked questions for the Usenet newsgroup sci.math]. http://daisy.uwaterloo.ca/~alopez-o/math-faq/math-faq.html .

Here is a quick and handy source for answers to questions that often pop up in the minds of students. Here are some examples: What is 0^0 and why? Why is it true that $0.9999\ldots = 1$? What are the details of the attempt to legislate the value of π to be 3? How are the digits of π computed? Who has won the Fields Medal? Why is there no Nobel prize in mathematics? Who is Bourbaki? What are the 23 Hilbert problems? What is the largest known Mersenne prime? It's much easier to access this source, or point students to it, than to figure out where to look for printed information, make a trip to the library, photocopy the material, etc. The collaborators who issue these little essays are looking for volunteers who would expand this online mini-encyclopedia.

Dubrovsky, Vladimir, Nesting puzzles, Part I: Moving oriental towers, *Quantum* 6 (3) (January–February 1996) 53–57, 49–51; Part II: Chinese rings produce a Chinese monster, 6 (4) (March–April 1996) 61–65, 58–59.

Beginning with the Tower of Hanoi and Chinese rings puzzles, this pair of articles explores contemporary versions and extensions of such puzzles. It goes on to note their connection with dragon curves, which are formed by folding a strip of paper in half multiple times (see Nikolay Vasilyev and Victor Gutenmacher, Dragon curves, *Quantum* 6 (1) (September–October 1995) 5–10, 60.

Donahue, Bill, Jugglers now juggle numbers to compute new tricks for ancient art, *New York Times* (Ntl. Ed.) (16 April 1996) B5, B10.

This article describes the mathematization of juggling. A juggling pattern can be described by a sequence of integers, which both represent the heights of the throws and measure the times between successive throws of a ball. Using this "site swap" notation, jugglers can design new patterns, watch them be enacted by a computer program, and efficiently remember them. For mathematical details, see Joe Buhler *et al.*, Juggling drops and descents, *American Mathematical Monthly* 101 (6) (June-July 1994) 507–519.

Naik, Gautam, In sunlight and cells, science seeks answers to high-tech puzzles, *Wall Street Journal* (16 January 1996) A1, A8.

Wall Street has discovered genetic algorithms. This cliché-ridden article ("back to Darwin," "reckless and random ways of nature," "cold, digital domain of silicon-based technology," etc.) cites the development of "T-cells" of computer code that seek out potential virus-containing code and of an evolving system of self-reproducing systems of rectangles, and of an electronic system of drawing portraits of suspects. But those applications aren't where the money is—in a short throwaway paragraph, the author mentions companies that use genetic algorithms to farm out computer-service jobs and pick stocks for a pension fund.

Wakeling, Edward (ed.), *Lewis Carroll's Games and Puzzles*, Dover, 1991; 128 pp, $4.95 (P). ISBN 0-486-26922-1. *Rediscovered Lewis Carroll Puzzles*, Dover, 1995; xiii + 79 pp, $4.95 (P). ISBN 0-486-28861-7.

These are two volumes of games and puzzles from the writings of Lewis Carroll, including ones extracted from previously unpublished letters, papers, and diary entries. Not all of the puzzles and games may be original with Carroll, but all were used by him to entertain his colleagues and young friends. Compiler Wakeling has included his own solutions.

NEWS AND LETTERS

Paul Erdős, 1913–1996

Paul Erdős, the world's best traveled mathematician, died on September 20, 1996 in Poland. He was 83. For mathematics this was the end of an era. Even among mathematicians, who often do not appear to be quite of this world, Erdős stood out as *sui generis*. One of the most prolific of twentieth-century mathematicians, at the time of his death he had published over 1500 mathematical papers, with more on the way. Though his mathematical productivity did not match that of the legendary Euler, with his 70+ volumes of published works, it did invite comparisons.

Erdős started life as a child prodigy, encouraged by his parents who were both mathematics teachers. He discovered negative numbers at the age of four, and by the age of eighteen he had proved a significant theorem in number theory, a field where he made many important discoveries, including contributions to an elementary proof of the prime number theorem. In later years, while maintaining an interest in number theory, he shifted to combinatorics. When asked about this he replied that the remaining open problems in number theory were too hard. One area of combinatorics where he made some of his most profound and interesting contributions is Ramsey theory. Another achievement in combinatorics is his well-known theorem (with Anning) which states that if an infinite number of points in the plane are all separated by integer distances, then all the points lie on a straight line.

Clearly Erdős will be missed as a productive mathematician. But he will also be missed as a source of good-natured anecdotes and stories. Mathematicians have often been accused of not being interested in the history of their subject, only in stories of mathematicians. Erdős provided them with an endless stimulus for stories: an Erdős number (the least distance between a mathematician and Erdős measured by a chain of coauthorships); the language Erdese (English pronounced as if it were Hungarian); the special vocabulary ("epsilon" for "child," "poison" for "alcohol," "to leave" for "to no longer do mathematics"); the "Book" (God's list of all the most elegant proofs in mathematics), and so on. Also, as the most traveled of mathematicians—possibly the most traveled of scientists of any kind—he served mathematics as a twentieth-century Marin Mersenne, a one-man clearing-house for information on the status of problems in his fields.

Erdős's incessant traveling around the world led to a great number of coauthors—over 400—writing in many languages, and this led Leo Moser to compose the following limerick:

A conjecture both deep and
  profound
Is whether a circle is round.
In a paper of Erdős
Written in Kurdish
A counterexample is found.

After that, Erdős tried to publish a paper in Kurdish but could not find a journal.

What has been neglected in some recent articles on Erdős is mention of his basic kindness: his entertaining "epsilons" with tricks and sleight of hand, his visiting the families of mathematicians who had recently died, his loans and outright gifts to promising students. Probably one of the explanations for his extraordinary mathematical career was his almost child-like curiosity, always about mathematics but about other things as well. His first question when he arrived on my own campus was: "What was the temperature in this valley during the Ice Age?" (I didn't know.)

Erdős was acutely aware that life—especially one's productive life as a mathematician—is finite. He joked about his first two-and-a-half billion years in mathematics (when he was born scientists thought the earth was two billion years old; they later revised the figure upward to four-and-a-half billion years!). But for years he described himself as old. He would on occasion add letters to the end of his name like degrees: P.G.O.M. (poor great old man), L.D. (living dead—a titled added at age 60), A.D. (archeological discovery—added at age 65), L.D. (legally dead—at age 70), and so on. At a memorial symposium to honor George Pólya, Erdős said: "In the *Arabian Nights*, they say 'May the King live forever.' In Pólya's case, we can say, 'May his theorems live forever.'" May Erdős's theorems live forever. And may all of his proofs turn out to be in the "Book."

G.L. Alexanderson
Santa Clara University

Acknowledgments

Along with our associate editors, the following referees have assisted the MAGAZINE *during the past year. We thank them for their time and care.*

Adams, William, *University of Maryland, College Park, College Park, MD*

Andrews, Peter, *Eastern Illinois University, Charleston, IL*

Askey, Richard, *University of Wisconsin, Madison, Madison, WI*

Axler, Sheldon, *Michigan State University, East Lansing, MI*

Barbeau, Edward, *University of Toronto, Toronto, Ontario, Canada*

Barcellos, Anthony, *Davis, CA*

Barksdale, Jr., James, *Western Kentucky University, Bowling Green, KY*

Barria, Jose, *Santa Clara University, Santa Clara, CA*

Baxley, John, *Wake Forest University, Winston-Salem, NC*

Beals, Richard, *Yale University, New Haven, CT*

Beezer, Robert, *University of Puget Sound, Tacoma, WA*

Beineke, Lowell, *Indiana-Purdue University at Fort Wayne, Fort Wayne, IN*

Berndt, Bruce, *University of Illinois, Urbana, IL*

Borwein, Peter, *Dalhousie University, Halifax, Nova Scotia, Canada*

Bridger, Mark, *Northeastern University, Boston, MA*

Brualdi, Richard, *University of Wisconsin, Madison, Madison, WI*

Brunson, Barry, *Western Kentucky University, Bowling Green, KY*

Campbell, Douglas, *Brigham Young University, Provo, UT*

Chakerian, G.D., *University of California, Davis, Davis, CA*

Chihara, Laura, *St. Olaf College, Northfield, MN*

Choike, James, *Oklahoma State University, Stillwater, OK*

Conrey, J. Brian, *Oklahoma State University, Stillwater, OK*

Crannell, Annalisa, *Franklin and Marshall College, Lancaster, PA*

Dean, Alice, *Skidmore College, Saratoga Springs, NY*

Devaney, Robert, *Boston University, Boston, MA*

Dobbs, David, *University of Tennessee, Knoxville, TN*

Drucker, Thomas, *Dickinson College, Carlisle, PA*

Dudley, Underwood, *DePauw University, Greencastle, IN*

Edgar, Hugh M.W., *San Jose State University, San Jose, CA*

Firey, William, *Oregon State University, Corvallis, OR*

Fisher, Evan, *Lafayette College, Easton, PA*

Fisher, J. Chris, *University of Regina, Regina, Saskatchewan, Canada*

Flanders, Harley, *University of Michigan, Ann Arbor, MI*

Foster, Lorraine, *California State University, Northridge, Northridge, CA*

Fraser, Craig, *University of Toronto, Victoria College, Toronto, Ontario, Canada*

Gallian, Joseph, *University of Minnesota, Duluth, Duluth, MN*

Geissinger, Ladnor, *Institute for Academic Technology, Durham, NC*

Gibbs, Richard, *Fort Lewis College, Durango, CO*

Gillman, Leonard, *Austin, TX*

Gluchoff, Alan, *Villanova University, Villanova, PA*

Gold, Bonnie, *Wabash College, Crawfordsville, IN*

Grabiner, Judith, *Pitzer College, Claremont, CA*

Grimaldi, Ralph, *Rose-Hulman Institute of Technology, Terre Haute, IN*

Grosshans, Frank, *Baltimore, MD*

Growney, Joanne, *Bloomsburg University of Pennsylvania, Bloomsburg, PA*

Grunbaum, Branko, *University of Washington, Seattle, WA*

Gulick, Denny, *University of Maryland, College Park, College Park, MD*

Guy, Richard, *University of Calgary, Calgary, Alberta, Canada*

Halmos, Paul, *Santa Clara University, Santa Clara, CA*

Hanson, Bruce, *St. Olaf College, Northfield, MN*

Harborth, Heiko, *Braunschweig, Germany*

Hatcher, Rhonda, *Texas Christian University, Fort Worth, TX*

Haunsperger, Deanna, *Carleton College, Northfield, MN*

Hayashi, Elmer, *Wake Forest University, Winston-Salem, NC*

Holzsager, Richard, *American University, Washington, DC*

Honsberger, Ross, *University of Waterloo, Waterloo, Ontario, Canada*

Horn, Roger, *University of Utah, Salt Lake City, UT*

Humke, Paul, *St. Olaf College, Northfield, MN*

Hungerford, Thomas, *Cleveland State University, Cleveland, OH*

Hutchinson, Joan, *Macalester College, St. Paul, MN*

Jacobson, Bernard, *Franklin and Marshall College, Lancaster, PA*

Jamison, Robert, *Clemson University, Clemson, SC*

Janke, Steven, *Colorado College, Colorado Springs, CO*

Johnson, Charles, *College of William and Mary, Williamsburg, VA*

Johnsonbaugh, Richard, *DePaul University, Chicago, IL*

Johnston, Elgin, *Iowa State University, Ames, IA*

Kedlaya, Kiran, *Princeton University, Princeton, NJ*

Kiltinen, John, *Northern Michigan University, Marquette, MI*

Kimberling, Clark, *University of Evansville, Evansville, IN*

Kitchen, Edward, *Santa Monica, CA*

Kleiner, Israel, *York University, North York, Ontario, Canada*

Krusemeyer, Mark, *Carleton College, Northfield, MN*

Kuzmanovich, James, *Wake Forest Univer-*

sity, Winston-Salem, NC

Lagarias, Jeffrey, *Murray Hill, NJ*

Larson, Dean, *Gonzaga University, Spokane, WA*

Larson, Loren, *St. Olaf College, Northfield, MN*

Lautzenheiser, Roger, *Rose-Hulman Institute of Technology, Terre Haute, IN*

Liu, Andrew, *University of Alberta, Edmonton, Alberta, Canada*

LoBello, Anthony, *Allegheny College, Meadville, PA*

Long, Calvin, *Washington State University, Pullman, WA*

McCleary, John, *Vassar College, Poughkeepsie, NY*

McDermot, Richard, *Allegheny College, Meadville, PA*

Merrill, Kathy, *Colorado College, Colorado Springs, CO*

Nelsen, Roger, *Lewis and Clark College, Portland, OR*

Nunemacher, Jeffrey, *Ohio Wesleyan University, Delaware, OH*

Otero, Daniel, *Xavier University, Cincinnati, OH*

Patterson, Sam, *Carleton College, Northfield, MN*

Pedersen, Jean, *Santa Clara University, Santa Clara, CA*

Reid, James, *Wesleyan University, Middletown, CT*

Reznick, Bruce, *University of Illinois, Urbana, IL*

Richman, Fred, *Florida Atlantic University, Boca Raton, FL*

Richmond, Thomas, *Western Kentucky University, Bowling Green, KY*

Rosenstein, George, *Franklin and Marshall College, Lancaster, PA*

Roy, Ranjan, *Beloit College, Beloit, WI*

Sandefur, James, *Georgetown University, Washington, DC*

Schaer, Jonathan, *University of Calgary, Calgary, Alberta, Canada*

Schattschneider, Doris, *Moravian College, Bethlehem, PA*

Schaumberger, Norman, *Hofstra University, Hempstead, NY*

Senechal, Marjorie, *Smith College, Northampton, MA*

Shahriari, Shahriar, *Pomona College, Claremont, CA*

Shapiro, Lou, *Howard University, Washington, DC*

Sherman, Gary, *Rose-Hulman Institute of Technology, Terre Haute, IN*

Smith, Bryan, *University of Puget Sound, Tacoma, WA*

Solheid, Ernie, *California State University, Fullerton, Fullerton, CA*

Straffin, Philip, *Beloit College, Beloit, WI*

Strang, Gilbert, *MIT, Cambridge, MA*

Strogatz, Steven, *Cornell University, Ithaca, NY*

Tan, Lin, *West Chester University of Pennsylvania, West Chester, PA*

Tanis, Elliot, *Hope College, Holland, MI*

Townsend, Douglas, *Indiana-Purdue University at Fort Wayne, Fort Wayne, IN*

Tucker, Alan, *SUNY at Stony Brook, Stony Brook, NY*

Van Brummelen, Glen, *The King's University College, Edmonton, Alberta, Canada*

Velleman, Daniel, *Amherst College, Amherst, MA*

Wakin, Shirley, *Southern Connecticut State University, New Haven, CT*

Watkins, John, *Colorado College, Colorado Springs, CO*

Wetzel, John, *University of Illinois, Urbana, IL*

Zeilberger, Doron, *Temple University, Philadelphia, PA*

UNITED STATES POSTAL SERVICE™

Statement of Ownership, Management, and Circulation

(Required by 39 USC 3685)

1. Publication Title: MATHEMATICS MAGAZINE
2. Publication Number: 0 0 2 5 - 5 7 0 X
3. Filing Date: November 6, 1996
4. Issue Frequency: bimonthly except July and August
5. Number of Issues Published Annually: 5
6. Annual Subscription Price: $22
7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*: MAA, 1529 Eighteenth St., NW, Washington, DC 20036-1385

Contact Person: Harry Waldman

Telephone: 202-387-5200

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*: Same

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*: The Mathematical Association of America, 1529 Eighteenth St., NW Washington, DC 20036-1385

Editor *(Name and complete mailing address)*: Paul Zorn, St. Olaf College, Northfield, MN 55057

Managing Editor *(Name and complete mailing address)*: Harry Waldman, MAA, 1529 18th St., NW, Washington, DC 20036

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
The Mathematical Association of America	1529 Eighteenth St.,NW, Washington, DC 20036

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box ☒ None

Full Name	Complete Mailing Address

12. Tax Status *(For completion by nonprofit organizations authorized to mail at special rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526**, September 1995 *(See Instructions on Reverse)*

13. Publication Title: MATHEMATICS MAGAZINE

14. Issue Date for Circulation Data Below: JUNE 1996

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*		15,875	13,450
b. Paid and/or Requested Circulation	(1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales *(Not mailed)*	0	0
	(2) Paid or Requested Mail Subscriptions *(Include advertiser's proof copies and exchange copies)*	15,025	12,150
c. Total Paid and/or Requested Circulation *(Sum of 15b(1) and 15b(2))*		15,025	12,150
d. Free Distribution by Mail *(Samples, complimentary, and other free)*		400	450
e. Free Distribution Outside the Mail *(Carriers or other means)*		0	0
f. Total Free Distribution *(Sum of 15d and 15e)*		400	450
g. Total Distribution *(Sum of 15c and 15f)*		15,425	12,600
h. Copies not Distributed	(1) Office Use, Leftovers, Spoiled	450	850
	(2) Returns from News Agents	0	0
i. Total *(Sum of 15g, 15h(1), and 15h(2))*		15,875	13,450
Percent Paid and/or Requested Circulation *(15c / 15g x 100)*		97%	96%

16. Publication of Statement of Ownership
☒ Publication required. Will be printed in the DECEMBER issue of this publication.
☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Donald J. Albers, Dir. of Publications — Date: 11/6/96

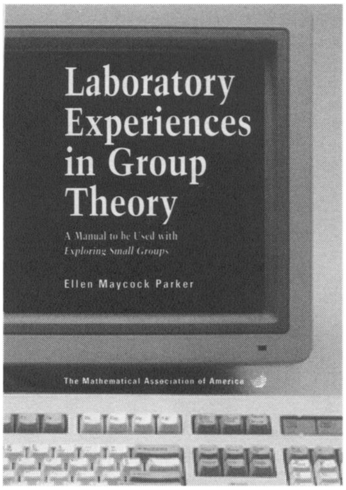
I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including multiple damages and civil penalties).

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. If the publication had second-class authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
5. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
6. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of second-class authorization.

PS Form **3526**, September 1995 *(Reverse)*



THE MATHEMATICAL ASSOCIATION OF AMERICA

Laboratory Experiences in Group Theory

A Manual to be Used with *Exploring Small Groups*

Ellen Maycock Parker

Series: Classroom Resource Materials

A lab manual with software for introductory courses in group theory or abstract algebra

Laboratory Experiences in Group Theory is a workbook of 15 laboratories designed to be used with the software *Exploring Small Groups* as a supplement to the regular textbook in an introductory course in group theory or abstract algebra. Written in a step-by-step manner, the laboratories encourage students to discover the basic concepts of group theory and to make conjectures from examples that are easily generated by the software. The labs can be assigned as homework or can be used in a structured laboratory setting. Since the software is user-friendly and the laboratories are complete, students and faculty should have no difficulty in using the labs without training.

Most students find that the laboratories provide an enjoyable alternative to the "theorem-proof-example" format of a standard abstract algebra course. At the end of the semester, one student wrote in his evaluation of the course:

I am truly grateful for the laboratory component...Work on the computer helped to make the abstract theory more concrete... One of the best things about the labs was that we formed our own conjectures about the patterns we saw...I believe that the progression of (1) lab, (2) conjecture, (3) class discussion, and (4) proof was highly beneficial in gaining understanding of the abstract material of the course.

Table of Contents: 1. Groups and Geometry; 2. Cayley Tables; 3. Cyclic Groups and Cyclic Subgroups; 4. Subgroups and Subgroup Lattices; 5. The Center and Commutator Subgroups; 6. Quotient Groups; 7. Direct Products; 8. The Unitary Groups; 9. Composition Series; 10. Introduction to Endomorphisms; 11. The Inner Automorphisms of a Group; 12. The Kernel of an Endomorphism; 13. The Class Equation; 14. Conjugate Subgroups; 15. The Sylow Theorems; Appendix A. Table Generation Menu of *Exploring Small Groups (ESG)*; Appendix B. Sample Library of *ESG*; Appendix C. Group Library of *ESG*; Appendix D. Group Properties Menu

Exploring Small Groups, the software packaged with this lab manual, is on a 3½" DD PC compatible disk. This is a DOS program that can be run in Windows. The software was developed by Ladnor Geissinger, University of North Carolina at Chapel Hill.

112 pp., Paperbound, 1996
ISBN 0-88385-705-7
List: $22.00 MAA Member: $16.00
Catalog Code: LABEJR

ORDER FROM:
THE MATHEMATICAL ASSOCIATION OF AMERICA
PO Box 91112, Washington, DC 20090-1112
1-800-331-1622 (301) 617-7800 FAX (301) 206-9789

Membership Code: ____ ____ ____ ____ ____ ____

Name ____________

Address ____________

City ____________

State ________ Zip ________

QTY.	CATALOG CODE	PRICE	AMOUNT
____	LABE/JR	____	____
____	____	____	____
		TOTAL	____

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ____________ Expires ___/___

Signature ____________

CONTENTS

ARTICLES

323 Plotting and Scheming with Wavelets, *by Colm Mulcahy*

344 Good-bye Descartes?, *by Keith Devlin*

NOTES

350 Algebraic Set Operations, Multifunctions, and Indefinite Integrals, *by Milan V. Jovanović and Veselin M. Jungić*

355 Proof without Words: An Alternating Series, *by James O. Chilaka*

356 Finite Groups of 2×2 Integer Matrices, *by George Mackiw*

361 Moving Card *i* to Position *j* with Perfect Shuffles, *by Sarnath Ramnath and Daniel Scully*

366 Probabilities of Clumps in a Binary Sequence (and How to Evaluate Them Without Knowing a Lot), *by David M. Bloom*

373 A Symmetry Criterion for Conjugacy in Finite Groups, *by Solomon W. Golomb*

375 More on the Converse of Lagange's Theorem, *by Guy T. Hogan*

376 A Persian Folk Method of Figuring Interest, *by Peyman Milanfar*

377 Identities on Point-Line Figures in the Euclidean Plane, *by Guido M. Pinkernell*

383 Professor Fogelfroe, *by Kenneth Kaminsky*

PROBLEMS

384 Proposals 1509–1513

385 Quickies 856–858

386 Solutions 1484–1488

391 Answers 856–858

REVIEWS

392

NEWS AND LETTERS

395 Paul Erdős, 1913–1996

396 Acknowledgments

399 Index to Volume 69

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, D.C. 20036