The UMAP Journal

Vol. 19, No. 3

关注数学模型
获取更多资讯

Subscription Rates for 1998 Calendar Year: Volume 19

Individuals subscribe to *The UMAP Journal* through COMAP's MEMBERSHIP PLUS. This subscription includes quarterly issues of *The UMAP Journal*, our annual collection *UMAP Modules: Tools for Teaching*, our organizational newsletter *Consortium*, a 10% discount on COMAP materials, and a choice of free materials from our extensive list of products.

(Domestic)   #MP9920 $64

Institutions can subscribe to the *Journal* through either Institutional Membership or a Library Subscription. Institutional Members receive two copies of each of the quarterly issues of *The UMAP Journal*, our annual collection *UMAP Modules: Tools for Teaching,* and our organizational newsletter *Consortium.* They also receive a 10% discount on COMAP materials and a choice of free materials from our extensive list of products.

(Domestic)   #UJ9940 $165

The Library Subscription includes quarterly issues of *The UMAP Journal* and our annual collection *UMAP Modules: Tools for Teaching.*

(Domestic)   #UJ9930 $140

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

关注数学模型
获取更多资讯

Table of Contents

Publisher's Editorial

A Time For Reflection

Solomon A. Garfunkel
Executive Director
COMAP, Inc.
57 Bedford St., Suite 210
Lexington, MA 02173
`s.garfunkel@mail.comap.com`

Don't worry, this is not yet another Y2K (Year 2000) sentimental musing of an aging mathematics educator. It's just that as I (an aging mathematics educator) write this, we are reviewing the blue lines for the *Mathematics: Modeling Our World* (ARISE) Course 3 high school textbook. The publication of this text represents the culmination of six years effort (seven, if you count writing the proposal). While we are still working on Course 4, we can quite clearly see the light at the end of this particular tunnel. But more importantly, it is time to take a clear look at where we stand, and here I mean the real we.

By the end of next year, all of the U.S. national K–12 comprehensive curriculum projects will be well out and published. Soon we will begin to see a new generation of students, those who have taken one of Everyday Math, Investigations, Connected Math, Math in Context, ARISE, Core-Plus, or IMP. These students will be our entering undergraduate students. And what will they look like? How different will they be?

They'll be better! They'll have handled a graphing calculator from 4th grade on. They'll know what residuals mean and how to look at messy data. They'll never ever ask, "What's this good for?" They won't be afraid to attack a problem because they don't know how to solve it before they start. They will be mathematical modelers, looking for new means of attack. And yes, they will have solid symbol manipulation skills. I firmly believe that these students will challenge us in much the same way that computer science students challenged their faculties some 25 to 30 years ago, when the world moved from mainframes to PCs.

As the title of this editorial suggests, I have been reflecting on COMAP's next steps. The creation of a complete high school curriculum is a major step in the implementation of reform. Where are the necessary next steps? The common wisdom says that the next major effort must be directed towards

in-service education for the K–12 teachers, to better prepare them to present students with this incarnation of change. Certainly, there is much that needs to be done in this arena, but I believe that the important answers lie elsewhere. They lie in the graduate and undergraduate math classrooms.

We cannot always be in this cycle of change the curriculum, then change the teachers to teach the new curriculum. The truest truism of education is that people teach as they were taught. We need to change the way we are taught. This means at all levels, including how we are taught as undergraduates and, yes, even as graduate students. For now we are in a transitional phase— teachers teaching curricula they did not take. That will soon change. People decide to teach mathematics because they enjoy it and are good at math in school. That math will soon be an ARISE-type curriculum.

So, it is time to work seriously on the undergraduate curriculum, to create classroom experiences that will continue the excitement of the K–12 experiences and build a core of ambassadors for mathematics. These will be our future teachers and our future users of mathematics. Just as it is long past time to think of mathematics in a layer-cake approach—algebra, geometry, trigonometry—it is also long past time to think of the style of teaching mathematics as being grade-level dependent.

With the *A Nation At Risk* report [1983], we convinced Congress (and I suspect ourselves) that we had the finest undergraduate educational system in the world, and all we needed to do was to "fix" K–12. The truth was and is much subtler. Undergraduate mathematics education needs work. It needs new courses and pedagogies that reflect the best aspects of reform.

When we finally begin to see mathematics education as one common enterprise, interconnected at all levels—with graduate school affecting elementary education and the elementary schools determining how our youngest students do or do not go on to further study—then we will have completed this cycle of reform.

And then it will be time to start over again.

Reference

United States National Commission on Excellence in Education. 1983. *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: Superintendent of Documents, U.S. Government Printing Office.

About the Author

Sol Garfunkel received his Ph.D. in mathematical logic from the University of Wisconsin in 1967. He was at Cornell University and at the University of Connecticut at Storrs for eleven years and has dedicated the last 20 years to

research and development efforts in mathematics education. He has been the Executive Director of COMAP since its inception in 1980.

He has directed a wide variety of projects, including UMAP (Undergraduate Mathematics and Its Applications Project), which led to the founding of this *Journal*, and HiMAP (High School Mathematics and Its Applications Project), both funded by the NSF. For Annenberg/CPB, he directed three telecourse projects: *For All Practical Purposes* (in which he appeared as the on-camera host), *Against All Odds: Inside Statistics*, and *In Simplest Terms: College Algebra*. He is currently co-director of the Applications Reform in Secondary Education (ARISE) project, a comprehensive curriculum development project for secondary school mathematics.
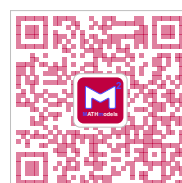
Modeling Forum

Results of the 1998 Mathematical Contest in Modeling

Frank Giordano, MCM Director
COMAP, Inc.
57 Bedford St., Suite 210
Lexington, MA 02173
f.giordano@mail.comap.com

Introduction

A total of 472 teams of undergraduates, from 246 institutions from 8 countries, spent the second weekend in February working on applied mathematics problems. They were part of the twelfth Mathematical Contest in Modeling (MCM). On Friday morning, the MCM faculty advisor opened a packet and presented each team of three students with a choice of one of two problems. After a weekend of hard work, typed solution papers were mailed to COMAP on Monday. Nine of the top papers appear in this issue of *The UMAP Journal*.

Results and winning papers from the first thirteen contests were published in special issues of *Mathematical Modeling* (1985–1987) and *The UMAP Journal* (1985–1997). The 1994 volume of *Tools for Teaching*, commemorating the tenth anniversary of the contest, contains all of the 20 problems used in the first ten years of the contest and a winning paper for each. Limited quantities of that volume and of the special MCM issues of the *Journal* for the last few years are available from COMAP.

Problem A: The Scanner Problem

Introduction

Industrial and medical diagnostic machines known as Magnetic Resonance Imagers (MRI) scan a three-dimensional object, such as a brain, and deliver

关注数学模型
获取更多资讯

their results in the form of a three-dimensional array of pixels. Each pixel consists of one number, indicating a color or a shade of gray that encodes a measure of water concentration in a small region of the scanned object at the location of the pixel. For instance, 0 can picture high water concentration in black (ventricles, blood vessels), 128 can picture a medium water concentration in gray (brain nuclei and gray matter), and 255 can picture a low water density in white (lipid-rich white matter consisting of myelinated axons). Such MRI scanners also include facilities to picture on a screen any horizontal or vertical slice through the three-dimensional array (slices are parallel to any of the three Cartesian coordinate axes).

Algorithms for picturing slices through oblique planes, however, are proprietary. Current algorithms

- are limited in terms of the angles and parameter options available,

- are implemented only on heavily used dedicated workstations,

- lack input capabilities for marking points in the picture before slicing, and

- tend to blur and "feather out" sharp boundaries between the original pixels.

A more faithful, flexible algorithm implemented on a personal computer would be useful

- for planning minimally invasive treatments;

- for calibrating the MRI machines;

- for investigating structures oriented obliquely in space, such as post-mortem tissue sections in animal research;

- for enabling cross sections at any angle through a brain atlas consisting of black-and-white line drawings.

To design such an algorithm, one can access the values and locations of the pixels but not the initial data gathered by the scanner.

Problem

Design and test an algorithm that produces sections of three-dimensional arrays by planes in any orientation in space, preserving the original gray-scale values as closely as possible.

Data Sets

The typical data set consists of a three-dimensional array A of numbers $A(i, j, k)$, where $A(i, j, k)$ is the density of the object at the location $(x, y, z)_{i,j,k}$. Typically, $A(i, j, k)$ can range from 0 through 255. In most applications, the data

set is quite large. Teams should design data sets to test and demonstrate their algorithms. The data sets should reflect conditions likely to be of diagnostic interest. Teams should also characterize data sets that limit the effectiveness of their algorithms.

Summary

The algorithm must produce a picture of the slice of the three-dimensional array by a plane in space. The plane can have any orientation and any location in space. (The plane can miss some or all data points.) The result of the algorithm should be a model of the density of the scanned object over the selected plane.

Problem B: The Grade Inflation Problem

Background

Some college administrators are concerned about the grading at A Better Class (ABC) College. On average, the faculty at ABC have been giving out high grades (the average grade now given out is an A−), and it is impossible to distinguish between the good and the mediocre students. The terms of a very generous scholarship only allow the top 10% of the students to be funded, so a class ranking is required.

The dean had the thought of comparing each student to the other students in each class, and using this information to build up a ranking. For example, if a student obtains an A in a class in which all students obtain an A, then this student is only "average" in this class. On the other hand, if a student obtains the only A in a class, then that student is clearly "above average." Combining information from several classes might allow students to be placed in deciles (top 10%, next 10%, etc.) across the college.

Problem

Assuming that the grades given out are (A+, A, A−, B+, . . .), can the dean's idea be made to work?

Assuming that the grades given out are only (A, B, C, . . .), can the dean's idea be made to work?

Can any other schemes produce a desired ranking?

A concern is that the grade in a single class could change many students' deciles. Is this possible?

Data Sets

Teams should design data sets to test and demonstrate their algorithms. Teams should characterize data sets that limit the effectiveness of their algorithms.

The Results

The solution papers were coded at COMAP headquarters so that names and affiliations of the authors would be unknown to the judges. Each paper was then read preliminarily by two "triage" judges at Southern Connecticut State University (Problem A) or at Carroll College (Montana) (Problem B). At the triage stage, the summary and overall organization are the basis for judging a paper. If the judges' scores diverged for a paper, the judges conferred; if they still did not agree on a score, a third judge evaluated the paper.

Final judging took place at Harvey Mudd College, Claremont, California. The judges classified the papers as follows:

	Outstanding	Meritorious	Honorable Mention	Successful Participation	Total
Scanner	4	31	47	106	189
Grade Inflation	3	48	69	163	283
	7	79	116	269	472

The seven papers that the judges designated as Outstanding appear in this special issue of *The UMAP Journal*, together with commentaries. We list those teams and the Meritorious teams (and advisors) below; the list of all participating schools, advisors, and results is in the **Appendix**.

Outstanding Teams

Institution and Advisor   **Team Members**

Scanner Papers

"A Method for Taking Cross Sections of
    Three-Dimensional Gridded Data"
Eastern Oregon University   Kelly Slater Cline
LaGrande, OR   Kacee Jay Giger
Norris Preyer   Timothy O'Conner

"A Model for Arbitrary Plane Imaging, or the
    Brain in Pain Falls Mainly on the Plane"

Harvey Mudd College	Jeff Miller
Claremont, CA	Dylan Helliwell
Michael Moody	Thaddeus Ladd

"A Tricubic Interpolation Algorithm
    for MRI Image Cross Sections"

Macalester College	Paul Cantrell
St. Paul, MN	Nicholas Wenninger
Karla V. Ballman	Tamás Németh-Csőri

"MRI Slice Picturing"

Tsinghua University	Ni Jiang
Beijing, China	Chen Jun
Ye Jun	Li Ling

Grade Inflation Papers

"Alternatives to the Grade Point Average
    for Ranking Students"

Duke University	Jeffrey A. Mermin
Durham, NC	W. Garrett Mitchener
Greg Lawler	John A. Thacker

"A Case for Stricter Grading"

Harvey Mudd College	Aaron F. Archer
Claremont, CA	Andrew D. Hutchings
Michael Moody	Brian Johnson

"Grade Inflation: A Systematic Approach
    to Fair Achievement Indexing"

Stetson University	Amanda M. Richardson
Deland, FL	Jeff P. Fay
Erich Friedman	Matthew Galati

Meritorious Teams

Scanner Papers (31 teams)
California Polytechnic State Univ., San Luis Obispo, CA (two teams) (Thomas O'Neil)
East China Univ. of Science and Technology, Shanghai, China (Yuanhong Lu)
Fudan University, Shanghai, China (Xi Zhou)
Harvey Mudd College, Claremont, CA (Ran Libeskind-Hadas)
Lawrence Technological Univ., Southfield, MI (Ruth G. Favro)
Macalester College, St. Paul, MN (Susan Fox)

N.C. School of Science and Mathematics, Durham, NC (two teams) (Dot Doyle)
Nankai University, Tianjin, China (XingWei Zhou)
Nat'l. Univ. of Defence Technology, Changsha, HuNan, China (Cheng LiZhi)
Nat'l. Univ. of Defence Technology, Changsha, HuNan, China (Wu Yu)
Rose-Hulman Institute of Technology, Terre Haute, IN (Aaron D. Klebanoff)
Seattle Pacific University, Seattle, WA (Steven D. Johnson)
South China Univ. of Technology, Guangzhou, Guangdong, China (Xie Lejun)
Southeast University, JiangSu, Nanjing, China (Zhou Jian Hua)
Southeast University, JiangSu, Nanjing, China (Wu Hua Hui)
Tsinghua University, Beijing, China (Hu Zhiming)
University of Alaska Fairbanks, Fairbanks, AK (John P. Lambert)
University of Colorado–Boulder, Boulder, CO (Anne Dougherty)
University of Massachusetts–Lowell, Lowell, MA (Lou Rossi)
University of Missouri–Rolla, Rolla, MO (Michael G. Hilgers)
University of Puget Sound, Tacoma, WA (Robert A. Beezer)
Univ. of Science and Technology of China, Hefei, Anhui, China (Rong Zhang)
University College–Cork, Cork, Ireland (J.B. Twomey)
Western Washington University, Bellingham, WA (Sebastian Schreiber)
Worcester Polytechnic Inst., Worcester, MA (Bogdan Vernescu)
Xi'an Jiaotung University Xi'an, Shaanxi, China (He Xiaoliang)
Xi'an Jiaotong University, Xi'an, Shaanxi, China (Zhou Yicang)
XiDian University Xi'an, Shaanxi, China (Liu Hongwei)
Youngstown State University, Youngstown, OH (Thomas Smotzer)

Grade Inflation Papers (48 teams)
Benedictine College, Atchison, KS (Jo Ann Fellin, OSB)
Bucknell University, Lewisburg, PA (Sally Koutsoliotas)
Colby College, Waterville, ME (Jan Holly)
College of William and Mary, Williamsburg, VA (Larry Leemis)
Colorado College, Colorado Springs, CO (Barry A. Balof)
David Lipscomb Institute, Nashville, TN (Mark A. Miller)
E. China Univ. of Sci. and Tech., Shanghai, China (Xiwen Lu)
Eastern Mennonite University, Harrisonburg, VA (John Horst)
Grinnell College, Grinnell, IA (Marc Chamberland)
Gustavus Adolphus College, St. Peter, MN (Gary Hatfield)
Harvey Mudd College, Claremont, CA (Ran Libeskind-Hadas)
Humboldt State Univ., Arcata, CA (Roland Lamberson)
Johns Hopkins University, Batlimore, MD (Daniel Q. Naiman)
Lafayette College, Easton, PA (Thomas Hill)
Lawrence Technological Univ., Southfield, MI (Howard Whitston)
Loyola College–Maryland, Baltimore, MD (Timothy J. McNeese)
Messiah College, Grantham, PA (Douglas C. Phillippy)
Mt. St. Mary's College, Emmitsburg, MD (John August)
N.C. School of Science and Mathematics, Durham, NC (John Kolena)
Natl. Univ. of Defence Technology, Changsha, HuNan, China (Wu MengDa)
Nazareth College, Rochester, NY (Kelly M. Fuller)
Nebraska Wesleyan University, Lincoln, NE (P. Gavin LaRose)
Pomona College, Claremont, CA (Richard Elderkin)
Rose-Hulman Institute of Technology, Terre Haute, IN (Aaron D. Klebanoff)

Saint Mary's College, Notre Dame, IN (Joanne Snow)
Salisbury State University, Salisbury, MD (Steven M. Hetzler)
Shanghai Normal University, Shanghai, China (Shenghuan Guo)
Southeast University, JiangSu, Nanjing, China (Shen Yu Jiang)
Southern Connecticut State University, New Haven, CT (Ross B. Gingrich)
Trinity University, San Antonio, TX (Diane G. Saphire)
Tsinghua University, Beijing, China (Ye Jun)
Tsinghua University, Beijing, China (Hu Zhiming)
U.S. Military Academy, West Point, NY (Kellie Simon)
United States Air Force Academy, USAF Academy, CO (Harry N. Newton)
United States Air Force Academy, USAF Academy, CO (Mark Parker)
Univ. of Science and Technology of China, Hefei, Anhui, China (Yi Shi)
Univ. of Wisconsin–Stevens Point, Stevens Point, WI (Nathan Wetzel)
University of Alaska Fairbanks, Fairbanks, AK (John P. Lambert)
University of Dayton, Dayton, OH (J.M. O'Hare)
University of Puget Sound, Tacoma, WA (Perry Fizzano)
University of Toronto, Toronto, Ontario, Canada (James G.C. Templeton)
Valparaiso University, Valparaiso, IN (Rick Gillman)
Wake Forest University, Winston-Salem, NC (Edward Allen)
Western Carolina University, Cullowhee, NC (Jeff A. Graham)
Western Carolina University, Cullowhee, NC (Scott Sportsman)
Western Connecticut State Univ., Danbury, CT (Judith A. Grandahl)
Worcester Polytechnic Inst., Worcester, MA (Arthur C. Heinricher)
Xidian University, Xi'an, Shaanxi, China (Mao Yongcai)
Youngstown State University, Youngstown, OH (Paul Mullins)

Awards and Contributions

Each participating MCM advisor and team member received a certificate signed by the Contest Director and the appropriate Head Judge.

INFORMS, the Institute for Operations Research and the Management Sciences, gave a cash award and a three-year membership to each member of the teams from Macalester College (Scanner Problem) and Stetson University (Grade Inflation Problem). Moreover, INFORMS gave free one-year memberships to all members of Meritorious and Honorable Mention teams.

The Society for Industrial and Applied Mathematics (SIAM) designated one Outstanding team from each problem as a SIAM Winner. The teams were from Macalester College (Scanner Problem) and Harvey Mudd College (Grade Inflation Problem). The Harvey Mudd team presented its results at a at a special Minisymposium of the SIAM Annual Meeting in Toronto in July. Each of the three Harvey Mudd team members was awarded a $300 cash prize. Their school was given a framed, hand-lettered certificate in gold leaf.

The Mathematical Association of America (MAA) designated one Outstanding team from each problem as an MAA Winner. The teams were from Eastern Oregon University (Scanner Problem) and Duke University (Grade Inflation

Problem). Both teams presented their solutions at a special session of the MAA Mathfest in Toronto in July. Each team member was presented a certificate by MAA President-Elect Tom Banchoff.

Judging

Director
Frank R. Giordano, COMAP, Lexington, MA

Associate Directors
David C. Arney, Dept. of Mathematical Sciences, U.S. Military Academy,
    West Point, NY
Robert L. Borrelli, Mathematics Dept., Harvey Mudd College,
    Claremont, CA

Scanner Problem

Head Judge
Marvin S. Keener, Executive Vice-President, Oklahoma State University,
    Stillwater, OK

Associate Judges
Kelly Black, Mathematics Dept., University of New Hampshire,
    Durham, NH
Paul Boisen, Defense Dept., Ft. Meade, MD
Courtney Coleman, Mathematics Dept., Harvey Mudd College,
    Claremont, CA
Patrick Driscoll, Dept. of Mathematical Sciences, U.S. Military Academy,
    West Point, NY (INFORMS)
William Fox, Dept. of Mathematical Sciences, U.S. Military Academy,
    West Point, NY
Debbie Levinson, Dept. of Mathematics, Colorado College,
    Colorado Springs, CO (SIAM)
Mark Levinson, Edmonds, WA (SIAM)
Jack Robertson, Head, Mathematics and Computer Science, Georgia College
    and State University, Milledgeville, GA (MAA)
Theresa M. Sandifer, Southern Connecticut State University, New Haven, CT
John L. Scharf, Carroll College, Helena, MT
Lee Seitelman, Glastonbury, CT

Grade Inflation Problem

Head Judge
Maynard Thompson, Mathematics Dept., University of Indiana,
    Bloomington, IN

Associate Judges
Karen Bolinger, Dept. of Mathematics, Clarion University of Pennsylvania,
    Clarion, PA
James Case, Baltimore, Maryland
Doug Faires, Dept. of Mathematics and Statistics, Youngstown State
    University, Youngstown, OH
Jerry Griggs, University of South Carolina, Columbia, SC (SIAM)
Mario Juncosa, RAND Corporation, Santa Monica, CA
John Kobza, Industrial and Systems Engineering, Virginia Polytechnic
    Institute and State University, Blacksburg, VA (INFORMS)
Mario Martelli, Dept. of Mathematics, California State University,
    Fullerton, CA
Vijay Mehrotra, Onward Inc., Mountain View, CA (INFORMS)
Veena Mendiratta, Lucent Technologies, Naperville, IL
Don Miller, Dept. of Mathematics, St. Mary's College, Notre Dame, IN
Catherine Roberts, Northern Arizona University, Flagstaff, AZ (SIAM)
Kathleen M. Shannon, Salisbury State University, Salisbury, MD (MAA)
Robert M. Tardiff, Dept. of Mathematical Sciences,
    Salisbury State University, Salisbury, MD
Michael Tortorella, Lucent Technologies, Holmdel, NJ
Marie Vanisko, Carroll College, Helena, MT
Daniel Zwillinger, Zwillinger & Associates, Arlington, MA

Triage Session

Scanner Problem

Head Triage Judge
Theresa M. Sandifer, Southern Connecticut State University, New Haven, CT

Associate Judges
Therese L. Bennett, Southern Connecticut State University, New Haven, CT
Ross B. Gingrich, Southern Connecticut State University, New Haven, CT
Cynthia B. Gubitose, Western Connecticut State University, Danbury, CT
C. Edward Sandifer, Western Connecticut State University, Danbury, CT

Grade Inflation Problem
(all were from Mathematics Dept., Carroll College, Helena, MT)

Head Triage Judge
Marie Vanisko

Sources of the Problems

The Scanner Problem was contributed by Yves Nievergelt, Mathematics Dept., Eastern Washington University. The Grade Inflation Problem was contributed by Dan Zwillinger, Zwillinger & Associates, Arlington, MA.

Acknowledgments

Cautions

To the reader of research journals:

Usually a published paper has been presented to an audience, shown to colleagues, rewritten, checked by referees, revised, and edited by a journal editor. Each of the student papers here is the result of undergraduates working on a problem over a weekend; allowing substantial revision by the authors could give a false impression of accomplishment. So these papers are essentially au naturel. Light editing has taken place: minor errors have been corrected, wording has been altered for clarity or economy, and style has been adjusted to that of *The UMAP Journal*. Please peruse these student efforts in that context.

To the potential MCM Advisor:

It might be overpowering to encounter such output from a weekend of work by a small team of undergraduates, but these solution papers are highly atypical. A team that prepares and participates will have an enriching learning experience, independent of what any other team does.

Appendix: Successful Participants

KEY:

P = Successful Participation
H = Honorable Mention
M = Meritorious
O = Outstanding (published in this special issue)
A = Scanner Problem
B = Grade Inflation Problem

INSTITUTION	CITY	ADVISOR	A	B
ALABAMA				
Huntingdon College	Montgomery	Sid Stubbs		P
University of Alabama	Huntsville	Claudio H. Morales		P
ALASKA				
Univ. of Alaska	Fairbanks	John P. Lambert	M	M
ARIZONA				
Northern Ariz. Univ.	Flagstaff	Terence R. Blows		H
University of Arizona	Tucson	Bruce J. Bayly		H
CALIFORNIA				
Calif. Inst. of Tech.	Pasadena	Richard M. Wilson	P	
Calif. Poly. State Univ.	San Luis Obispo	Thomas O'Neil	M,M	
Calif. State Univ.	Bakersfield	John Dirkse		P,P
Harvey Mudd College	Claremont	Michael Moody	O	
		Ran Libeskind-Hadas	M	O,M
Humboldt State Univ.	Arcata	Jeffrey B. Haag		H
		Roland Lamberson		M
L.A. Pierce College	Woodland Hills	Bob Martinez		P
Loyola Marymount U.	Los Angeles	Thomas M. Zachariah		P,P
Occidental College	Los Angeles	Ron Buckmire	H	
Pepperdine Univ.	Malibu	Bradley W. Brock		H,H
Pomona College	Claremont	Richard Elderkin		M
Sonoma State Univ.	Rohnert Park	Sunil K. Tiwari		P
Univ. of Redlands	Redlands	Steve Morics	P	
COLORADO				
Colorado College	Colorado Springs	Barry A. Balof		M,P
Fort Lewis College	Durango	Dick Walker		P
Mesa State College	Grand Junction	Edward Bonan-Hamada		P

INSTITUTION	CITY	ADVISOR	A	B
U.S. Air Force Academy	USAF Academy	Steven F. Baker	P	
		Harry N. Newton		M
		Mark Parker	P	M
Univ. of Colorado	Boulder	Anne Dougherty	M	
		Bengt Fornberg	P	
Univ. of South. Colorado	Pueblo	Bruce N. Lundberg		P
CONNECTICUT				
Connecticut College	New London	Kathy McKeon		H
Southern Conn. State Univ.	New Haven	Ross B. Gingrich		M
		Theresa Bennett	P	
U.S. Coast Guard Academy	New London	Janet A. McLeavey		P
Western Conn. State Univ.	Danbury	Judith A. Grandahl		M
		Paul Hines		H
		C. Edward Sandifer		P
DISTRICT OF COLUMBIA				
Georgetown University	Washington	Andrew Vogt	H	P
FLORIDA				
Florida Inst. of Technology	Melbourne	Gary W. Howell		P,P
Florida Southern College	Lakeland	William G. Albrecht		P
		Charles B. Pate		P
		Allen Wuertz	P	
Jacksonville University	Jacksonville	Paul R. Simony	P	
		Robert A. Hollister	P	P
Stetson University	Deland	Erich Friedman		O
GEORGIA				
Agnes Scott College	Decatur	Robert A. Leslie		H
Georgia College & State Univ.	Milledgeville	Craig Turner	P	
State Univ. of West Georgia	Carrollton	Scott Gordon		P
		Everett D. McCoy	P	
IDAHO				
Boise State University	Boise	Alan R. Hausrath	P	
ILLINOIS				
Greenville College	Greenville	Galen R. Peters	P	
Illinois Wesleyan University	Bloomington	Zahia Drici	P	
Northern Illinois University	Dekalb	Hamid Bellout		P
Wheaton College	Wheaton	Paul Isihara		H,P

INSTITUTION	CITY	ADVISOR	A	B
INDIANA				
Ball State University	Muncie	Fred Gylys-Colwell		P
Earlham College	Richmond	Mic Jackson		P
		Charlie Peck	H	
		Tekla Lewin		H
Indiana University	Bloomington	Larry Moss	P	H
	South Bend	Morteza Shafii-Mousavi	H	
Rose-Hulman Inst. of Tech.	Terre Haute	Frank Young	H	
		Aaron D. Klebanoff	M	M
Saint Mary's College	Notre Dame	Joanne Snow		M,H
Valparaiso University	Valparaiso	Rick Gillman		M,P
IOWA				
Drake University	Des Moines	Luz M. De Alba		P
		Alexander F. Kleiner		H
Graceland College	Lamoni	Steve K. Murdock		P
Grinnell College	Grinnell	Marc Chamberland	P	M
Iowa State University	Ames	Stephen J. Willson		P
Luther College	Decorah	Reginald D. Laursen		P
Simpson College	Indianola	Rick Spellerberg		H
		M.E. "Murphy" Waggoner	P	
Univ. of Northern Iowa	Cedar Falls	Gregory M. Dotseth		H
		Timothy L. Hardy		P
KANSAS				
Baker University	Baldwin City	Bob Fraga	P	P
Benedictine College	Atchison	Jo Ann Fellin, OSB		M
Bethel College	North Newton	Monica Meissen	P	
KENTUCKY				
Asbury College	Wilmore	Kenneth P. Rietz		H
Bellarmine College	Louisville	John A. Oppelt	P	
Brescia College	Owensboro	Chris A. Tiahrt		P
LOUISIANA				
McNeese State University	Lake Charles	Karen Aucoin		H
Northwestern State Univ.	Natchitoches	Lisa R. Galminas		P
MAINE				
Bowdoin College	Brunswick	Helen Moore		P
Colby College	Waterville	Jan Holly		M,P

INSTITUTION	CITY	ADVISOR	A	B
MARYLAND				
Goucher College	Baltimore	David Horn	H	
		Robert E. Lewand		P
Hood College	Frederick	John Boon, Jr.		P
Johns Hopkins University	Batlimore	Daniel Q. Naiman		M
Loyola College–Maryland	Baltimore	Dipa Choudhury	H,H	
		Timothy J. McNeese		M
Mt. St. Mary's College	Emmitsburg	John August		M
		Theresa A. Francis		P
Salisbury State University	Salisbury	Steven M. Hetzler		M
St. Mary's Coll. of Md.	St. Mary's City	James Tanton	P	P
MASSACHUSETTS				
Bentley College	Waltham	Lucia Kimball		P
Boston College	Chestnut Hill	Paul R. Thie	P	
Boston University	Boston	Glen Hall		P
Harvard University	Cambridge	Curtis McMullen		P
Salem State College	Salem	Joyce Anderson		P
Simon's Rock College	Great Barrington	Allen B. Altman	P	H
		Michael Bergman	P	
Smith College	Northampton	Ruth Haas	P	
Univ. of Massachusetts	Amherst	Edward A. Connors	H	
	Lowell	J. "Kiwi" Graham-Eagle	P	
		Lou Rossi	M	
Western New England Coll.	Springfield	Lorna Hanes		P
Williams College	Williamstown	Stewart Johnson	P	P
Worcester Polytechnic Inst.	Worcester	Arthur C. Heinricher		M
		Bogdan Vernescu	M	
MICHIGAN				
Albion College	Albion	Scott Dillery	P	P
		David Seely		P
Calvin College	Grand Rapids	Thomas L. Jager	H	
Eastern Michigan Univ.	Ypsilanti	Christopher E. Hee	H	P
Hillsdale College	Hillsdale	John P. Boardman	P	P
Lawrence Tech. Univ.	Southfield	Ruth G. Favro	M	
		Scott Schneider		P
		Howard Whitston		M
Michigan State University	E. Lansing	C.R. MacCluer		P
MINNESOTA				
Gustavus Adolphus Coll.	St. Peter	Gary Hatfield		M

INSTITUTION	CITY	ADVISOR	A	B
Macalester College	St. Paul	Karla V. Ballman	O	
		Susan Fox	M	
		Daniel Kaplan		P
Univ. of Minnesota	Duluth	Zhuangyi Liu		P
	Morris	Peh Ng		P,P
Winona State University	Winona	Steven Leonhardi		P
MISSOURI				
Central Missouri State Univ.	Warrensburg	L. Vincent Edmondson		P
Northwest Missouri State U.	Maryville	Russell Euler	P	P
Truman State University	Kirksville	Steve Smith	P	P
Univ. of Missouri	Rolla	Michael G. Hilgers	M,H	
MONTANTA				
Carroll College	Helena	Terence J. Mullen		P
		Jack Oberweiser	P	
		Phil Rose	H	
		Anthony M. Szpilka		P
NEBRASKA				
Hastings College	Hastings	David B. Cooke		H
Nebraska Wesleyan Univ.	Lincoln	P. Gavin LaRose		M,P
NEVADA				
Sierra Nevada College	Incline Village	Elizabeth Carter		P
NEW JERSEY				
Camden County College	Blackwood	Allison Sutton	P	
New Jersey Inst. of Tech.	Newark	John Bechtold	H	
NEW MEXICO				
New Mexico State Univ.	Las Cruces	Joseph Lakey	P	
NEW YORK				
Buffalo State College	Buffalo	Robin Sanders		H
Great Neck South HS	Great Neck	Robert Silverstone	P	
Ithaca College	Ithaca	James E. Conklin	P	
		John C. Maceli		P
Nassau Community Coll.	Garden City	Abraham S. Mantell	P	
Nazareth College	Rochester	Kelly M. Fuller		M,H
Niagara University	Niagara	Steven L. Siegel	P	
Pace University	Pleasantville	Robert Cicenia		P

INSTITUTION	CITY	ADVISOR	A	B
St. Bonaventure Univerity	St. Bonaventure	Francis C. Leary		H
		Albert G. White	P	
SUNY Geneseo	Geneseo	Chris Leary		P
U.S. Military Academy	West Point	Chuck Mitchell	P	
		James S. Rolf	H	
		Kellie Simon		M
		Charles C. Tappert	H	
Wells College	Aurora	Carol C. Shilepsky		P
Westchester Comm. College	Valhalla	Rowan Lindley		P
		Sheela Whelan		P
NORTH CAROLINA				
Appalachian State University	Boone	Holly P. Hirst	P	P
Duke University	Durham	Greg Lawler		O
N.C. School of Sci. & Math.	Durham	Dot Doyle	M,M	
		John Kolena		M
Salem College	Winston-Salem	Debbie L. Harrell		H
		Paula G. Young		P
Univ. of North Carolina	Chapel Hill	Douglas G. Kelly		H
		Jon W. Tolle	P	
	Pembroke	Raymond E. Lee		P
Wake Forest University	Winston-Salem	Edward Allen		M
		Stephen B. Robinson		H
Western Carolina University	Cullowhee	Jeff A. Graham		M
		Scott Sportsman		M
		Kurt Vandervoort	P	
NORTH DAKOTA				
Univ. North Dakota	Williston	Wanda M. Meyer	P	
OHIO				
College of Wooster	Wooster	Reuben Settergren		P
Hiram College	Hiram	Larry Becker		P,P
		Brad Gubser	P	P
Marietta College	Marietta	Tom LaFramboise	P	P
Miami University	Oxford	Douglas E. Ward		P
Ohio University	Athens	David N. Keck	P	
University of Dayton	Dayton	J.M. O'Hare		M
		Ralph C. Steinlage		H,P
Xavier University	Cincinnati	Richard J. Pulskamp		P

INSTITUTION	CITY	ADVISOR	A	B
Youngstown State University	Youngstown	Stephen Hanzely	P	
		Paul Mullins		M
		Thomas Smotzer	M	H
OKLAHOMA				
Oklahoma State University	Stillwater	John E. Wolfe	P	P
Southeastern Okla. State Univ.	Durant	John M. McArthur		P
		Karla Oty		P
Southern Nazarene University	Bethany	Philip Crow		P
OREGON				
Eastern Oregon State College	LaGrande	David Allen		H
		Norris Preyer	O,H	
		Jenny Woodworth		P
Southern Oregon University	Ashland	Kemble R. Yates	P	
PENNSYLVANIA				
Allegheny College	Meadville	David L. Housman		P
Bucknell University	Lewisburg	Sally Koutsoliotas	P	M
Chatham College	Pittsburgh	Eric Rawdon		P
Gettysburg College	Gettysburg	James P. Fink	H	P
Lafayette College	Easton	Thomas Hill		M
Messiah College	Grantham	Douglas S. Phillippy		M
		Lamarr C. Widmer	H	
Penn State Berks-Lehigh Valley	Reading	L. Miller-Van Wieren	P	
		D.M. Van Wieren		P
Shippensburg University	Shippensburg	Doug Ensley		P
		Gene Fiorini	P	
Susquehanna University	Selinsgrove	Kenneth A. Brakke		P
Westminster College	New Wilmington	Barbara Faires		H
RHODE ISLAND				
Rhode Island College	Providence	D.L. Abrahamson	P	
SOUTH CAROLINA				
Charleston Southern Univ.	Charleston	Stan Perrine		P
Coastal Carolina University	Conway	Ioana Mihaila	P	
Univ. of South Carolina	Aiken	Nieves A. McNulty	P	
SOUTH DAKOTA				
Northern State University	Aberdeen	A.S. Elkhader		H

INSTITUTION	CITY	ADVISOR	A	B
TENNESSEE				
Austin Peay State University	Clarksville	Mark C. Ginn	H	
Christian Brothers University	Memphis	Cathy W. Carter		P,P
David Lipscomb University	Nashville	Gary C. Hall	P	
		Mark A. Miller		M
TEXAS				
Abilene Christian University	Abilene	David Hendricks	P	P
Angelo State University	San Angelo	Andrew B. Wallace		P
Baylor University	Waco	Ronald B. Morgan	P	
Trinity University	San Antonio	Diane G. Saphire	P	M
University of Dallas	Irving	Richard P. Olenick	P	
		Edward P. Wilson	P	
University of Houston	Houston	Barbara Lee Keyfitz		H
University of Texas	Austin	Mike Oehrtman	P	H
	Richardson	Ali Hooshyar	P	
		T. Constantinescu		P
VERMONT				
Johnson State College	Johnson	Glenn D. Sproul		P,P
VIRGINIA				
College of William & Mary	Williamsburg	Larry Leemis		M
Eastern Mennonite University	Harrisonburg	John Horst		M,H
Randolph-Macon Woman's Coll.	Lynchburg	Eric Chandler		P
Thos. Jefferson HS for Sci.& Tech.	Alexandria	John Dell	H,P	
University of Richmond	Richmond	Kathy W. Hoke	P	
Virginia Western Comm. College	Roanoke	Ruth Sherman	P	P
WASHINGTON				
Pacific Lutheran University	Tacoma	Rachid Benkhalti	P	
Seattle Pacific University	Seattle	Steven D. Johnson	M	
University of Puget Sound	Tacoma	Robert A. Beezer	M	H
		Perry Fizzano		M,P
Western Washington University	Bellingham	Sebastian Schreiber	M	P
		Saim Ural		P,P
WISCONSIN				
Beloit College	Beloit	Philip D. Straffin		H,H
Carroll College	Waukesha	John Symms		P
		William Welch		P
Edgewood College	Madison	Ken Jewell		P
		Steven Post	P	

INSTITUTION	CITY	ADVISOR	A	B
Northcentral Technical College	Wausau	Frank J. Fernandes		P
		Robert J. Henning	P	P
St. Norbert College	De Pere	John A. Frohliger	P	
Univ. of Wisconsin	Eau Claire	Carl Schoen	P	
	Platteville	Sherrie Nicol		P
	Stevens Point	Nathan Wetzel		M
UW Colleges–Marathon County	Wausau	Fe Evangelista		H
		Paul A. Martin		P
Wisconsin Lutheran College	Milwaukee	M.C. Papenfuss		P

AUSTRALIA

Univ. of Southern Queensland	Toowoomba, QLD	C.J. Harman	H	
		Tony Roberts		H

CANADA

Univ. of Western Ontario	London, Ontario	Peter H. Poole	H	
University of Alberta	Edmonton, Alberta	Joseph So	P	
University of Calgary	Calgary, Alberta	D.R. Westbrook		H
University of Saskatchewan	Saskatoon, SK	James A. Brooke	H	
		Raj Srinivasan	H	
		Tom Steele	P	
University of Toronto	Toronto, Ontario	N.A. Derzko		P,P
		J.G.C. Templeton		M
York University	Toronto, Ontario	Neal Madras	P	H

CHINA

Anhui Inst. of Mech. & Elec. Eng.	Wuhu, Anhui	Wang Chuanyu		P
		Wang Geng		P
Anhui University	Hefei, Anhui	Wu Fuchao		P
		Yang Shangjun	H	
Beijing Institute of Technology	Beijing	Bao Zhu Guo	P	P
		Xiao Di Cui		P
Beijing Normal University	Beijing	Laifu Liu	P	P
		Wenyi Zeng		P,P
Beijing U. of Aero. & Astro.	Beijing	Li Wei guo	H	
Beijing Union University	Beijing	Ren KaiLong		P
		Zeng Qingli		P
Beijing Univ. of Chem. Tech.	Beijing	Liu Damin	P	
		Shi Xiaoding	P	
		Zhao Baoyuan		P
Central South Univ. of Tech.	Changsha, Hunan	Han Xuli		H,P
Central-south Institute of Tech.	Hengyang, Hunan	Li Xianyi		H

关注数学模型
获取更多资讯

INSTITUTION	CITY	ADVISOR	A	B
Central-south Inst. of Tech.	Hengyang, Hunan	Liu Yachun		P
China U. of Mining & Tech.	Xuzhou, Jiangsu	Zhang Xingyong		H
		Zhou Shengwu		H
Chongqing University	Chongqing	Fu Li		H
		Gong Qu	H	
		He Zhongshi	P	
		Liu Qiongsen	P	
Dalian Univ. of Technology	Dalian, Liaoning	He Mingfeng		H,P
		Yu Hongquan	P	
		Zhao Lizhong	P	
E. China Univ. of Sci. & Tech.	Shanghai	Nianci Shao	H	
		Xiwen Lu		M,H
		Yuanhong Lu	M	
East China Normal Univ.	Shanghai	Lin Wuzhong		P
Exp'l HS, Beijing Normal U.	Beijing	Han Leqing		P,P
		Math Chair		P
First Middle School	Jiading, Shanghai	Chengan		P,P
Fudan University	Shanghai	Jin Liu		P
		Xi Zhou	M	P
		Zhijie Cai		P
Harbin Inst. of Tech.	Harbin, Heilongjiang	Shang Shouting	H	P
		Wang Yong	H	H
Hebei Institute of Tech.	Tangshan, Hebei	Liu Baoxiang		P
		Liu Chunfeng		P
		Lu Zhenyu	P	
Hefei University of Tech.	Hefei, Anhui	Xueqiao Du	H	
		Yonghua Hu	P	
		Yongwu Zhou		P
		Youdu Huang	H	
Jilin Institute of Technology	Changchun, Jilin	Sun Changchun	P	
		Wang Xiuyu		P
		Xu Yunhui	P	
		Lu Xian Rui		P
		Shi ShaoYun		P
		Yin Jing Xue	P	
		Fang Peichen	P	
		Zhang Kuiyuan		P
Jinan University	Guangzhou, Guangdong	Shiqi Ye	P	
		Suohai Fan		H
Lanzhou Railway Institute	Lanzhou, Gansu	Bai Lihua		H
		He Shanglu	H	
		Li Yongan	P	
		Zhang Jianxun		H

INSTITUTION	CITY	ADVISOR	A	B
N.W. Polytech. Univ.	Xian, Shaanxi	Peng Guohua	P	
		Rong Haiwu		H
		Wang MingYu		P
		Zhang Shenggui		H
Nankai University	Tianjin	Bin Wang		H
		Jishou Ruan	H	
		XingWei Zhou	M	
Nanyang Model HS	Shanghai	Tuqing Cao		P
Natl. Univ. of Defence Tech.	Changsha, Hunan	Cheng LiZhi	M	
		Wu MengDa		M
		Wu Yu	M	
Peking University	Beijing	Jian-hua Wu	P	
		Lei Gong-yan		H,P
		Zhuoqun Xu	P	
Qufu Normal University	Qufu, Shandong	Yuzhong Zhang		P
Shandong University	Jinan, Shandong	Cui Yuquan		H
		Long Heping	P	
		Piming Ma		P
		Zhengyuan Ma	P	
Shanghai Jiaotong Univ.	Shanghai	Li Shidong		H
		Song Baorui	P	
		Sun Zhuling	P	
		Zhou Gang		P
Shanghai Normal Univ.	Shanghai	Shenghuan Guo	H	M
South China Univ. of Tech.	Guang Zhou, Guangdong	Chang Zhihua		P
		Fu Hongzhuo	H	
		Hao Zhifeng		H
		Xie Lejun	M	
Southeast University	JiangSu, Nanjing	Nie Chang hai		P
		Shen Yu jiang		M
		Wu Hua hui	M	
		Zhou Jian hua	M	
Southwest Jiaotong Univ.	Chendu, Sichuan	Deng Ping		H
		Li Tianrui	P	
		Yuan Jian	H	
		Zhao Lianwen		P
Tsinghua University	Beijing	Hu Zhiming	M	M
		Ye Jun	O	M
Univ. of Elec. Sci. & Tech.	Chengdu	Xu Quanzhi	H	
		Zhong Erjie		P

INSTITUTION	CITY	ADVISOR	A	B
Univ. of Sci. & Tech. of China	Hefei, Anhui	Chaoyang Zhu		H
		Rong Zhang	M	
		Shizhuo Ji	H	
		Yi Shi		M
Xi'an Jiaotong University	Xi'an,Shaanxi	Zhou Yicang	M	
		He Xiaoliang	M	H
Xidian University	Xi'an, Shaanxi	Hu Yupu		H
		Liu Hongwei	M	
		Mao Yongcai		M
Zhejiang University	Hangzhou, Zhejiang	Qifan Yang	H	H
		Shu Ping Chen	H	H
Zhengzhou Electr. Pwr Coll.	Zhengzhou, Henan	Liang Haijiang		P
		Wang Jiade	H	
ZhengZhou Univ. of Tech.	Zhengzhou, Henan	Wang Jinling	P	
		Wang Shubin		P
		Zhang Xinyu	P	
ZhongShan University	Guangzhou, Guangdong	She Wei Long	P	
		Tang Mengxi	H	
		Wang Yuan Shi		H
		Zhang Lei	P	

FINLAND
    Paivola College · Tarttila · Bill Shaw · H

FINLAND				
Paivola College	Tarttila	Bill Shaw	H	

HONG KONG

Hong Kong Baptist Univ.	Kowloon Tong, Kowloon	Chong Sze Tong		P
		Wai Chee Shiu		H

IRELAND

Trinity College Dublin	Dublin	T.G. Murphy	H	
		James C. Sexton	P	
University College, Cork	Cork	Patrick Fitzpatrick		H
		Finbarr O'Sullivan		P
		Gareth Thomas	P	
		J. B. Twomey	M	
University College Dublin	Dublin	Ted Cox	P	
		Maria Meehan	H	
University College Galway	Galway	Martin Meere	H	
		Michael P. Tuite		H

LITHUANIA

Vilnius University	Vilnius	Ricardas Kudzma		P

A Method for Taking Cross Sections of Three-Dimensional Gridded Data

Kelly Slater Cline
Kacee Jay Giger
Timothy O'Conner
Eastern Oregon University
LaGrande, OR 97850

Advisor: Norris Preyer

Summary

Effective three-dimensional magnetic resonance imaging (MRI) requires an accurate method for taking planar cross sections. However, if an oblique cross section is taken, the plane may not intersect any known data points. Thus, a method is needed to interpolate water density between data points.

Interpolation assumes continuity of density, but there are discontinuities in the human body at the borders of different types of tissue. Most interpolation methods try to smooth these sharp borders, blurring the data and possibly destroying useful information.

To capture qualitatively the key difficulties of this problem, we created a sequence of simulated biological data sets, such as a brain and an arm, each with some specific defect. Our data sets are cubic arrays with 100 elements on each side, for a total of one million elements, specifying water density at each point with an integer in the range $[0, 255]$. In each data set, we use differentiable functions to describe several tissue types with discontinuities between them.

To analyze these data, we created a group of algorithms, implemented in C++, and compared their effectiveness in generating accurate cross sections. We used local interpolation techniques, because the data are not continuous on a global level. Our final algorithm searches for discontinuities between tissues. If it finds one at a point, it preserves sharp edges by assigning to that point the water density of the nearest data point. If there is no discontinuity, the algorithm does a polynomial fit in three dimensions to the nearest 64 data points and interpolates the water density.

关注数学模型
获取更多资讯

We measured the accuracy of the algorithms by finding the mean absolute difference between the interpolated water density and the actual water density at each point in the cross sections. Our final algorithm has an error 16% lower than a simple closest-point technique, 17% lower than a continuous linear interpolation, and 22% lower than a continuous polynomial interpolation without discontinuity detection.

Assumptions

- An MRI scan is an equally spaced grid of data. We take it to be a $100 \times 100 \times 100$ array.

- Each element of the array is an integer ranging from 0 to 255, representing the water density at that point.

- The resolution of the cross section to be taken is equal to the resolution of the data set. (If the array elements have a spacing of one micron, then the cross section should have the same spacing.)

- We recognize that all methods of interpolation assume continuity between the data points. Thus, we assume that the water density in living tissue can be represented as continuous differentiable functions with discontinuities between tissues.

Simulated Data Sets

Since we could find few existing data sets of three-dimensional arrays, we constructed simulated data sets. While real biological organs are extraordinarily complicated, a set of simulated organs should be able to represent qualitatively the kind of problems that an MRI scan is typically used to investigate. Although actual MRI data have much greater resolution, the characteristics that we are looking for should be the same: tumors, fractures, or general anomalies. Generally, these areas will have different water densities than surrounding tissue, generating discontinuities. We created the following mock organs with imperfections:

1. **Globules**: A continuous, repeating spherical pattern, with a density peak at the center (**Figure 1**).

2. **Arm**: Smooth tissue with two bones, one containing a small spherical hole (**Figure 2**).

3. **Generic organ**: A round shape filled with several discontinuous regions (**Figure 3**).

4. **Brain**: A dense skull, periodically varying gray matter, and a small area of different density in one lobe (**Figure 4**).
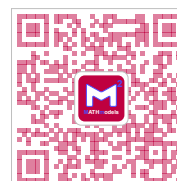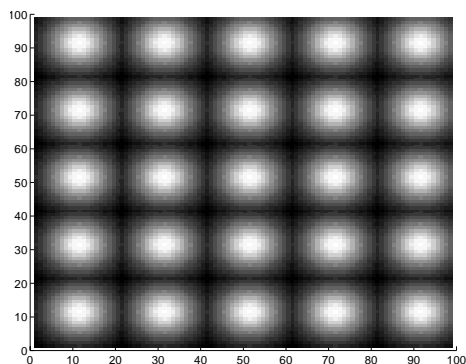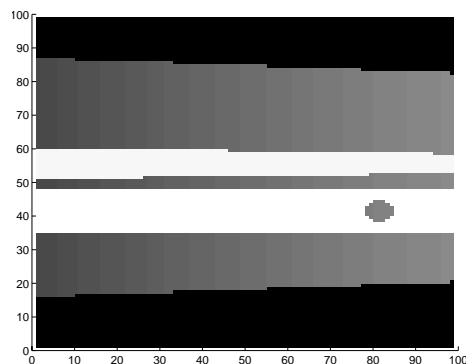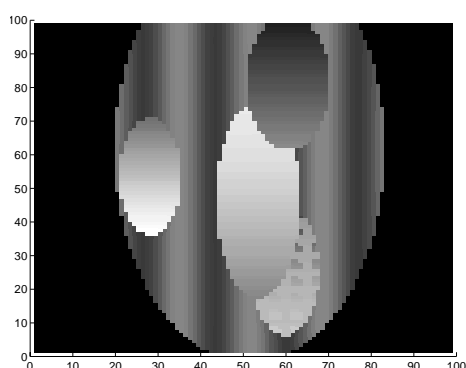
Figure 1. Globules.

Figure 2. Arm.

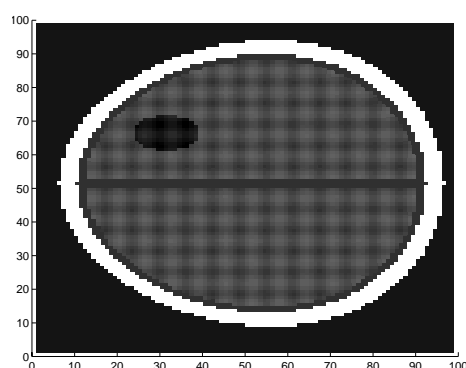**Figure 3.** Generic organ.

Figure 4. Brain.

Coordinate Systems and Definitions

The existing array of data imposes a Cartesian coordinate system on the problem. If there are n data points in each direction, then the coordinates range from $(0, 0, 0)$ to (n, n, n). We define a cross section by picking a point in this coordinate system (x_0, y_0, z_0) and two angles (θ, ϕ) representing the angles that the plane makes with the positive x-axis and the positive y-axis. This point becomes the origin of our plane, with the new x-axis being the projection of the x-axis onto this plane in the z-direction, so the unit vector is

$$\hat{x}' = \hat{x} \cos \theta + \hat{z} \sin \theta.$$

We can solve for the unit vector $\hat{y}'$ if we require it to be orthogonal to $\hat{x}'$, to make an angle ϕ with the unit vector $\hat{y}$, and to be of unit length, so

$$\hat{y}' = -\hat{x} \sin \phi \sin \theta + \hat{y} \cos \phi + \hat{z} \sin \phi \cos \theta.$$

Thus, we can convert from the (x', y') coordinate system back to the array system as:

$$x = x_0 + x' \cos\theta - y' \sin\phi \sin\theta$$
$$y = y_0 \qquad\qquad + y' \cos\phi$$
$$z = z_0 + x' \sin\theta + y' \sin\phi \cos\theta$$

We call the known points (x, y, z) *data points* and the unknown points (x', y') *plane points*.

Interpolation Algorithms

The plane points do not generally match existing data points. We know the water density surrounding each plane point, so we must interpolate to estimate plane point density.

There are two major classes of interpolation techniques:

- global methods, which use every data point in the set to estimate the density at each plane point, and

- local methods, which only use a small subset of the data points.

Because interpolation methods assume continuity, global methods are inappropriate to this problem. We know that the organs are only piecewise continuous and differentiable with discontinuities between tissues. Thus, all of our algorithms use local interpolation techniques.

Proximity

This algorithm assigns to the plane point the density of the data point that is closest to the plane point. This method seems naive, but it should preserve sharp edges without blurring. It looks at each point in the plane (x', y'), calculates x, y, and z in the original array, and rounds them to integer values (X, Y, Z), thus giving the closest data point.

Density Mean

This method uses more information to estimate the water density at each point. We can visualize every plane point as being inside a cube, with data points at the corners. To estimate the value inside, we take the arithmetic mean of the density from the surrounding eight points. Despite the use of more information, this method blurs the edges of discontinuities.

Trilinear Interpolation

This algorithm uses the same eight points as the density mean method but does a weighted average of the density (ρ) values. This method assumes that the slope $d\rho/dx$ is constant between the data points. With a low-resolution dataset, this approach will create inaccuracies; but as the resolution increases, the slope will appear to be more constant, since any differentiable function appears linear when examined on a small enough scale. The formula for linear interpolation [Press 1988, 104–105] is

$$\rho(x') = \sum_{i=1..2} (1 - T_i)\rho_i,$$

where $T_i = |x' - x_i|$.

The weight that we give to the value ρ_i is equivalent to the distance to the opposite point. Here, every pair of consecutive data points has a distance of 1 unit between them, so the distance to the opposite point is $1 - T_i$. For trilinear interpolation, we extend this sum over all the points, so that

$$\rho(x', y', z') = \sum_{i=1..2} \sum_{j=1..2} \sum_{k=1..2} (1 - T_i)(1 - U_j)(1 - V_k)\rho_{ijk},$$

where $T_i = |x' - x_i|$, $U_j = |y' - y_j|$, and $V_k = |z' - z_k|$.

Polynomial Interpolation

To estimate better the water density function, the polynomial interpolation method uses even more data. We expand the surrounding cube of eight points in every direction to make a cube with four points on each side, getting the 64 nearest points. Polynomials can fit differentiable functions better because they have more derivatives and can incorporate larger trends in the function. Recall that two points determine a unique line, three points determine a unique quadratic, and four points determine a unique cubic. Making use of this, we can develop a method for fitting functions in three dimensions. By doing a sequence of fits in the x, y, and z directions, we can synthesize these into a density estimate for a point in space. Thus, we break the problem into a series of one-dimensional interpolations.

Let (x, y, z) be the plane point, and let the data points be $(x_{1..4}, y_{1..4}, z_{1..4})$. First, we fix x_1 and y_1, fit the four points $(x_1, y_1, z_{1..4})$ to a cubic, and interpolate the density at (x_1, y_1, z). We increment to x_1, y_2 and do the same until we have the densities at the points (x_1, y_1, z), (x_1, y_2, z), (x_1, y_3, z), (x_1, y_4, z). Then we fit a polynomial in the y-direction to these four points and interpolate the density at (x_1, y, z). We repeat this whole process to find (x_2, y, z), (x_3, y, z), and (x_4, y, z), and then perform one last polynomial fit to these points to interpolate the density at (x, y, z).

There are many techniques for doing polynomial fits. We used the Lagrange formula [Acton 1990, 96] because it is the least computationally intensive:

$$\rho(x') = \frac{(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)}\,\rho_1 + \frac{(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)}\,\rho_2$$
$$+ \frac{(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)}\,\rho_3 + \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)}\,\rho_4,$$

where ρ_i is the density at x_i.

This method will blur edges but should do very well over regions described by differentiable functions.

Hybrid Algorithms

All of the above methods have strengths and weaknesses. The methods that are strongest on the differentiable regions (trilinear, polynomial) are weakest at discontinuities, because they try to smooth out the sharp borders. The method that most closely preserves discontinuities (proximity) is weakest at identifying smooth trends in the functions.

To capitalize on the strengths of both approaches, we created a hybrid algorithm. Before interpolating between a group of points, the hybrid looks for discontinuities within them; it uses the proximity method if it finds any and a continuous method otherwise. This hybrid algorithm locates discontinuities by measuring the difference in density $(\Delta\rho)$ between each pair of extreme opposite points surrounding the plane point. If there is a discontinuity, then $\Delta\rho$ will be large and we use the proximity method. If not, $\Delta\rho$ will be small and we use a continuous method, either trilinear or polynomial. To distinguish between the two cases, we set the threshold value of $\Delta\rho_0$. Thus, the hybrid algorithm allows us to use each method where it is strongest.

Testing and Results

Because we have defined precise water density functions for each of our four simulated data sets, we can compare the interpolation value with the actual density and find the residual. To measure the accuracy of a cross section, we calculate the mean absolute residual over all of the plane points (i.e., the average of the absolute values of the residuals).

To compare the algorithms, we took 12 cross sections through each simulated data set, at different angles, points, and discontinuity threshold levels. We generally selected points near the center of the data so as to generate a large planar region.

Table 1 shows the mean absolute residual for each algorithm applied to each data set and averaged over all data sets. We discuss the results on each of the data sets (columns in **Table 1**) in turn.

Table 1.

Mean absolute residuals for interpolation methods applied to the data sets.

Algorithm	Data Set				
	Globules	Arm	Generic organ	Brain	Combined
Proximity	1.82	0.97	0.89	3.25	1.73
Density mean	1.39	1.89	1.34	5.52	2.53
Trilinear interpolation	1.54	1.27	0.70	3.49	1.75
Polynomial interpolation	1.55	1.48	0.75	3.67	1.86
Hybrid trilinear interpolation					
($\Delta \rho_0 = 20$)	1.55	1.01	0.59	3.49	1.49
($\Delta \rho_0 = 30$)	1.54	1.01	0.59	2.82	1.49
($\Delta \rho_0 = 40$)	1.54	1.01	0.61	2.82	1.50
Hybrid polynomial interpolation					
($\Delta \rho_0 = 20$)	1.66	0.99	0.71	3.14	1.62
($\Delta \rho_0 = 30$)	1.63	0.99	0.61	2.86	1.52
($\Delta \rho_0 = 40$)	1.61	0.99	0.53	2.56	1.45

Globules

On average, the mean method most accurately generated oblique planes, with both the trilinear and polynomial interpolations providing cross sections of similar accuracy. For this data set, the hybrid algorithms did worse than the purely continuous methods (**Figure 5**); the proximity method did poorest of all, probably because there are no edges in the globule data sets and continuity is never broken. The trilinear and polynomial interpolations may also have had trouble with the peaks in the center of each sphere.
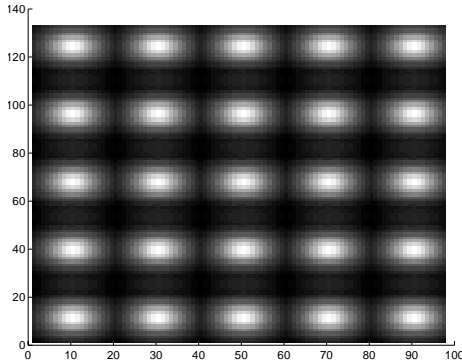
Figure 5. Globules: Polynomial hybrid method, with $\theta = 45°$, $\phi = 0°$, $(45, 45, 50)$.
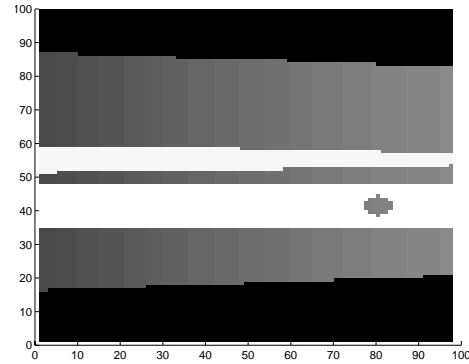
Figure 6. Arm: Polynomial hybrid method, with $\theta = 10°$, $\phi = 0°$, $(40, 80, 50)$.

Arm

Here we found the situation almost totally reversed. The proximity method and the hybrid algorithms (**Figure 6**) all performed significantly better than the purely continuous methods, with the mean method doing particularly badly.

This is quite reasonable, because our arm has many very sharp edges as we go from bone to muscle. The mean method should fail on any discontinuities, and this is what we see. The discontinuity detection seems to be working, because the hybrid algorithms perform noticeably better than the trilinear and polynomial methods.
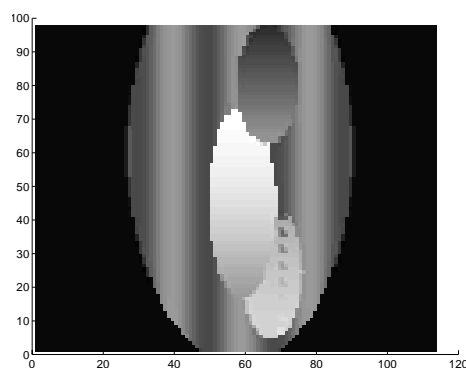

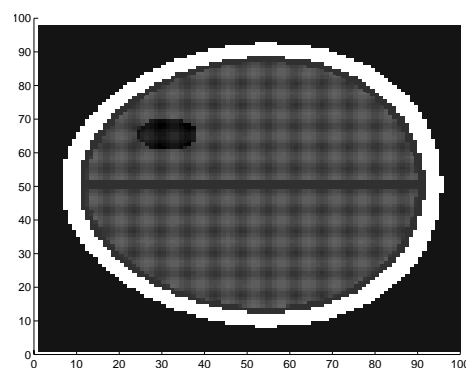**Figure 7.** Generic organ: Polynomial hybrid method, with $\theta = 0°$, $\phi = 30°$, $(50, 50, 50)$.

Figure 8.   Brain, polynomial hybrid method, $\theta = 5°$, $\phi = 0°$, $(50, 50, 50)$.

Generic Organ

We found that both of the hybrid formulas, trilinear and polynomial, produced equally favorable results (**Figure 7**). Once again, the arithmetic mean method performed poorly, followed by the proximity, polynomial, and trilinear methods, whose results were comparable. We suspect that this is the case because for this data set we used smooth functions to generate each of the different tissue types.

Brain

Due to the general smoothness of each lobe and the sharp contrast of the skull, the hybrid polynomial method with a high value of $\Delta\rho_0$ produced the most accurate results. The proximity method produced very accurate results, surpassing the trilinear and polynomial methods but falling short of the hybrid methods (**Figures 8–10**). The most inaccurate results were produced by the arithmetic mean method, and **Figure 11** clearly shows how it fails by trying to smooth the edges of the skull.

Residual Plots

Another way to examine the algorithms is to plot the residuals, which allows us to see exactly where the a method breaks down. In **Figures 12–14**, large positive or negative residuals stand out in white or gray.
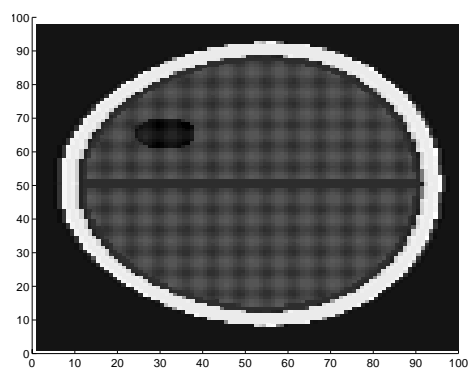
Figure 9. Brain, polynomial method, $\theta = 0°$, $\phi = 30°$, $(50, 50, 50)$.

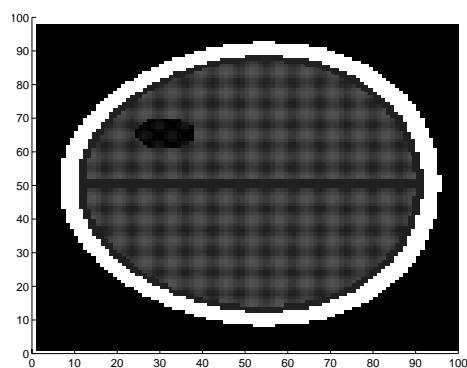**Figure 10.** Brain, proximity method, $\theta = 5°$, $\phi = 0°$, $(50, 50, 50)$.

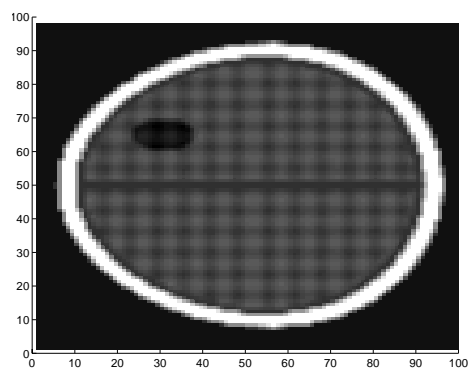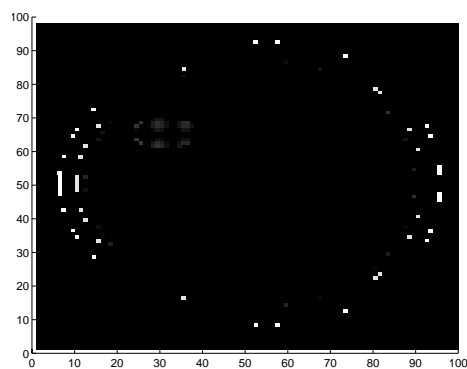**Figure 11.** Brain, density mean, $\theta = 0°$, $\phi = 30°$, $(50, 50, 50)$.

Figure 12. Brain, proximity residuals, $\theta = 5°$, $\phi = 0°$, $(50, 50, 50)$.

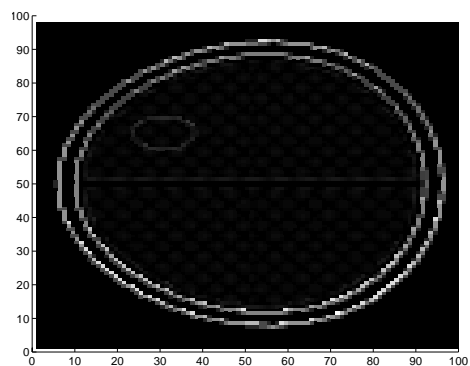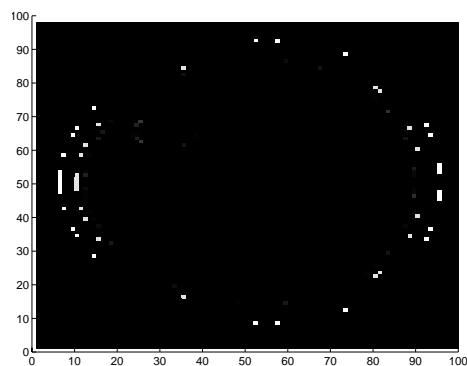**Figure 13.** Brain, density mean, residuals, $\theta = 0°$, $\phi = 30°$, $(50, 50, 50)$.

Figure 14. Brain, polynomial hybrid method, residuals, $\theta = 5°$, $\phi = 0°$, $(50, 50, 50)$.

The proximity method was generally accurate over most of the brain, except for the dark area in one of the lobes and a few areas where the sharp discontinuity of the skull also produced errors (**Figure 12**).

The arithmetic mean algorithm clearly shows errors around all the edges (**Figure 13**).

Finally, the hybrid polynomial algorithm has inaccuracies near the edge of the skull, but it handles the dark area very well (**Figure 14**).

Overall Results

- Averaging over all four data sets (see the last column of **Table 1**), the most accurate algorithm for generating oblique planes through arrays of three-dimensional data is the hybrid polynomial algorithm (with $\Delta\rho_0 = 40$), which produced an average error 16% less than the proximity method.

- The hybrid trilinear algorithm (with $\Delta\rho_0 = 30$) did almost as well, generating an average error 14% less than the proximity method.

- The proximity algorithm produced reasonably accurate planes, with average error 7% less than the continuous polynomial algorithm and about 1% better than the continuous trilinear algorithm.

- The trilinear algorithm performed better than the polynomial method, but neither handled discontinuities well enough to produce results as good as the hybrid methods.

- The arithmetic mean algorithm produced particularly poor results, because of the large errors that it makes around a discontinuity of any kind.

Strengths and Weaknesses

The hybrid polynomial algorithm is flexible and could easily be expanded to handle data arrays of any size. It can take a slice through any point at any angle. Moreover, it is effective at interpolating through smooth regions but still preserves the sharpness of edges in the original data. Even if the closest point is on the wrong side of the discontinuity, the image is still qualitatively correct: It shows a sharp edge. The threshold constant $\Delta\rho_0$ allows the user to choose how much smoothing is done.

The hybrid algorithm will miss small discontinuities in the data, and it does not look for nondifferentiable points. If there are cusps, the algorithm will probably not notice them and attempt to smooth through these points, even though to do so is inappropriate.

Future Work

The next step is to implement better methods for interpolating in the continuous regions, using larger sets of points. Cubic and quartic splines might be effective, as might other types of polynomials, or perhaps a method of rational function interpolation.

The discontinuity detection algorithm could be improved by expanding it to look for cusps and nondifferentiability. The discontinuities could also be used to perform automatic tissue typing; the algorithm might then be able automatically to output images showing just brain gray matter, or showing just tumor tissue. Even with the current software, a more detailed investigation of the dynamics of $\Delta\rho_0$ would be very useful.

Most important, the algorithm needs thorough testing against actual MRI data.

References

Acton, Forman S. 1990. *Numerical Methods That Work*. Washington DC: Mathematical Association of America.

Garcia, Alejandro L. 1994. *Numerical Methods for Physics*. Englewood Cliffs, NJ: Prentice-Hall.

Hornak, Joseph P. 1997. The Basics of MRI. `http://www.cis.rit.edu/htbooks/mri/` . (7 February 1998).

Lancaster, Peter, and Kestutis Salkauskas. 1986. *Curve and Surface Fitting*. London: Academic Press.

Press, William H., et al. 1988. *Numerical Recipes in* C. New York: Cambridge University Press.

Summit, Steve. *C Programming FAQs*. 1996. New York: Addison-Wesley.

A Model for Arbitrary Plane Imaging, or the Brain in Pain Falls Mainly on the Plane

Jeff Miller
Dylan Helliwell
Thaddeus Ladd
Harvey Mudd College
1250 N. Dartmouth Ave.
Claremont, CA 91711
{ jmiller, dhelliwe, tladd } @math.hmc.edu

Advisor: Michael Moody

Summary

We present an algorithm for imaging arbitrary oblique slices of a three-dimensional density function, based on a rectilinear array of uniformly sampled MRI data.

We

- develop a linear interpolation scheme to determine densities of points in the image plane,

- incorporate a discrete convolution filter to compensate for unwanted blurring caused by the interpolation, and

- provide an edge-detecting component based on finite differencing.

The resulting algorithm is sufficiently fast for use on personal computers and allows control of parameters by the user.

We exhibit the results of testing the algorithm on simulated MRI scans of a typical human brain and on contrived data structures designed to test the limitations of the model. Filtering distortions and inaccurate modeling due to interpolation appear in certain extreme scenarios. Nonetheless, we find that our algorithm is suitable for use in real-world medical imaging.

Constructing the Model

Our model consists of four main parts:

- First, we develop a technique for positioning a plane anywhere in $\mathbb{R}^3$.

- Then we interpolate data from the region in $\mathbb{R}^3$ that contains data onto the plane.

- Next, we use a sharpening technique to remove extra blur caused by the interpolation.

- Finally, we construct a difference array and use it to create a line drawing representing edges in the image.

Assumptions

- **Density variations in the source object are reasonably well behaved and continuous.** Discontinuities such as sharp edges will be approximated in the model but only if they are isolated on a scale of several array elements. Similarly, erratic behavior and wild fluctuations can be accurately modeled only if they exist on a scale of several pixels. The model should image the source of the data array, not the array itself; but the accuracy of the oblique slice images depends on the accuracy of the data in the array.

- **The data array represents isotropically spaced samples.** The array $A(i, j, k)$ contains discretized samples from a continuous three-dimensional space, for which we use coordinates (x, y, z). The component $A(i, j, k)$ represents a density f at some point (x_i, y_j, z_k). We assume that the source was uniformly sampled, so that

$$x_i = i\delta x, \qquad y_j = j\delta y, \qquad z_k = k\delta z,$$

where δx, δy, and δz are constant distances between samples (typically close to 1 mm). We also assume that these distances are all equal (if not, then we rescale the coordinate system to compensate).

- **The destination computing platform is a typical contemporary personal computer.** Thus, input data arrays may not be larger than the memory of a typical PC. We assume arrays of up to $256 \times 256 \times 256$ integer elements (16 MB in size) each in $[0, 255]$. Also, we gauge computing time by typical PC processor speeds.

Plane-Array Intersection

To represent the slice of the object on the computer screen, we establish a mapping between the three-space of the object and the plane of the monitor.

We represent an arbitrary plane in $\mathbb{R}^3$ in terms of the angles that it makes with the xy-plane and the z-axis, together with a displacement of the origin. The map $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$T(u, v) = R \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$$

transforms a point in $\mathbb{R}^2$ into a point on the plane, where the point (x_0, y_0, z_0) is a displacement of the origin and R is the rotation matrix

$$R = \begin{pmatrix} \cos\phi\cos\theta & -\sin\theta & \sin\phi\cos\theta \\ \cos\phi\sin\theta & \cos\theta & \sin\phi\sin\theta \\ -\sin\phi & 0 & \cos\phi \end{pmatrix}.$$

The angles ϕ and θ are the polar and azimuthal angles in spherical coordinates for the vector normal to the plane.

Interpolation

To represent the image, we seek a regularly spaced array of discretized points (u_p, v_q) corresponding to the pixels of a computer monitor. Since the points $T(u_p, v_q)$ need not coincide with the points (x_i, y_j, z_k) for the data, we need to be able to approximate density values at arbitrary points in $\mathbb{R}^3$. Thus, we interpolate the data from nearby points whose values are given by A.

With a slight abuse of notation, let $g(x, y, z)$ be the gray-scale value of the image at (x, y, z), so that $g\big(T(u, v)\big) = g(u, v)$ and $g(x_i, y_j, z_k) = A(i, j, k)$.

From the numerous techniques for interpolation, we seek an algorithm that will smoothly approximate the density without being computationally intractable.

Nearest-Neighbor Approach

Let (x^*, y^*, z^*) be the point for which we want to know the density. This point is contained in a cubic cell, of size $\delta x \times \delta y \times \delta z$, that has corners of known density given by the array A. From these eight corners, we simply find the point (x_a, y_b, z_c) that is closest to (x^*, y^*, z^*) and set $g(x^*, y^*, z^*) = A(a, b, c)$.

3-D Linear Interpolation

We also develop a technique that we call *3-D linear interpolation*. For this method, we hope to find a smooth continuation of the data within the cubic cell, starting from the density values of the corners of the cubic cell containing (x^*, y^*, z^*). We base our approach on solving the Laplace equation

$$\Delta g = \frac{\partial^2 g}{\partial x_1^2} + \cdots + \frac{\partial^2 g}{\partial x_n^2} = 0$$

successively in one, two, and three dimensions.

We choose two adjacent corners of the cell and solve the one-dimensional Laplace equation, using the densities at the corners as boundary values. This gives the smoothest function between the two corners—a straight line. We then solve the two-dimensional Laplace equation on the faces of the cubic cell, using the straight lines as boundary conditions. Finally, we fill the cube with the three-dimensional solution to the Laplace equation, using as boundary conditions the values on the faces.

The details are not difficult. Denote the points in the cubic cell as in **Figure 1**. The values along edges are constructed by simple linear interpolation; for example, the values along the lower left edge of the cube in **Figure 1** are given by

$$g(x^*, y_j, z_k) = g(x_i, y_j, z_k) + \frac{g(x_{i+1}, y_j, z_k) - g(x_i, y_j, z_k)}{x_{i+1} - x_i} (x^* - x_i)$$

$$= A(i, j, k) + \frac{A(i+1, j, k) - A(i, j, k)}{x_{i+1} - x_i} (x^* - x_i).$$
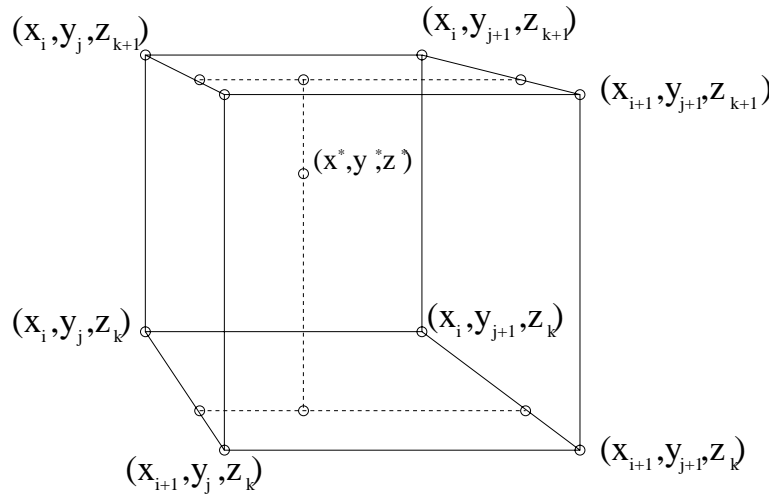

Figure 1. A cubic cell demonstrating the notation for the 3-D interpolation scheme.

Similarly, we find

- values $g(x^*, y_{j+1}, z_k)$ along the lower right edge in terms of $A(i, j+1, k)$ and $A(i+1, j+1, k)$,

- values $g(x^*, y_j, z_{k+1})$ along the upper left edge in terms of $A(i, j, k+1)$ and $A(i+1, j, k+1)$, and

- values $g(x^*, y_{j+1}, z_{k+1})$ along the upper right edge in terms of $A(i, j+1, k+1)$ and $A(i+1, j+1, k+1)$.

We continue to use linear interpolation to get the value $g(x^*, y^*, z_k)$ on the bottom face in terms of the value $g(x^*, y_j, z_k)$ on the lower left edge and the

value $g(x^*, y_{j+1}, z_k)$ on the lower right edge, as well as the value $g(x^*, y^*, z_{k+1})$ on the upper face in terms of the value $g(x^*, y_j, z_{k+1})$ on the upper left edge and the value $g(x^*, y_{j+1}, z_{k+1})$ on the upper right edge.

As a last step, we use linear interpolation yet again to obtain the value $g(x^*, y^*, z^*)$ in terms of the value $g(x^*, y^*, z_k)$ on the lower face and the value $g(x^*, y^*, z_{k+1})$ on the upper face. The result is a unique value for $g(x^*, y^*, z^*)$, in terms of the eight closest corners, which does not depend on the order of interpolation. [EDITOR'S NOTE: We omit the authors' proof of this fact, which they arrive at by explicit calculation of $g(x^*, y^*, z^*)$ and the observation that the result is symmetric in x, y, and z.] In addition,

$$\frac{\partial^2 g}{\partial x^2}(x^*, y^*, z^*) = \frac{\partial^2 g}{\partial y^2}(x^*, y^*, z^*) = \frac{\partial^2 g}{\partial z^2}(x^*, y^*, z^*) = 0,$$

which means that Laplace equation is satisfied in the cubic cell. The uniqueness theorem for the Laplace equation with Dirichlet boundary conditions implies that only one solution may be obtained by this method.

Other Techniques

We considered a three-dimensional spline, in which a cubic interpolation is chosen to make first derivatives continuous. Unfortunately, the complexity and the computing time of this technique made it intractable. We also looked at spatially weighted averaging techniques. However, in applying this method to a simple two-dimensional case, we found gross discrepancies with the true image (a simple linear ramp was altered to look like a series of wavy steps).

Ultimately, we found:

- The nearest-neighbor technique is the most useful for cursory image analysis, and

- 3-D linear interpolation is the most efficient method for a more realistic image.

Image Sharpening

Interpolation inevitably blurs, or *low-pass filters*, the actual image f. Hence, we add a stage to the algorithm that sharpens, or *high-pass filters*, the recorded image g. We considered various techniques for sharpening.

Revert from Boundaries

One approach to sharpen is to detect the location of edges or boundaries in the image and then revert to a nearest-neighbor pixel determination near those locations. We discovered that this approach has the adverse affect of increasing graininess and pixelation.

Point-Spread Function

Another approach, following Andrews and Hunt [1977], is to assume that the recorded image is the actual image convolved with a *point-spread function* (PSF), denoted $h(x, y)$. Thus,

$$g(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x - \xi, y - \eta) f(\xi, \eta) \, d\xi d\eta. \qquad \textbf{(1)}$$

The actual image is then obtained by deconvolution with a discrete Fourier transform. The PSF may be calculated a priori or measured a posteriori.

Alternatively, the discretized nature of the data and the linear nature of the interpolation procedure lead us to recast **(1)** into the matrix equation

$$g(u_p, v_q) = \sum_{m=1}^{N} \sum_{n=1}^{N} a(p, m) b(q, n) f(u_m, v_n),$$

where $a(p, m)$ is an $N \times N$ matrix that blurs the columns of the digitized plane image and $b(q, n)$ is an $N \times N$ matrix that blurs the rows. The "blurring" matrices may be approximated with components near unity on the leading diagonals and components equal to some small "mixing" parameter on the adjacent off-diagonals. The image f may then be restored by inverting these matrices.

Ultimately, we deemed this and the Fourier PSF approach to be too computationally expensive.

Convolution Filter

Our favored technique, following Rosenfeld and Kak [1982], is to use a convolution filter. This approach rests on the assumption that the blurring occurred as a diffusion process. If the actual image f is an initial condition to the diffusion equation

$$\kappa \nabla^2 g = \frac{\partial g}{\partial t},$$

then by expanding a time-dependent $g(u, v; t)$ about a small value τ of time we obtain

$$f(u, v) = g(u, v; 0) = g(u, v; \tau) - \tau \frac{dg}{dt}(u, v; \tau) + O\left(\tau^2\right)$$
$$= g - \kappa \tau \nabla^2 g + O\left(\tau^2\right).$$

Thus, f may be restored by subtracting the Laplacian of g from g. This technique, commonly called *unsharp masking*, is especially appealing in our model; since we chose an interpolation scheme that forces interpolated regions of the image g to satisfy Laplace's equation, $\nabla^2 g = 0$.

In practice, the Laplacian is approximated using finite differences. Define

$$\Delta_u g(u_p, v_q) = g(u_p, v_q) - g(u_{p-1}, v_q),$$
$$\Delta_v g(u_p, v_q) = g(u_p, v_q) - g(u_p, v_{q-1}).$$

Higher-order difference operators are defined by repeated first differencing, as in

$$\Delta_u^2 g(u_p, v_q) = \Delta_u g(u_{p+1}, v_q) - \Delta_u g(u_p, v_q),$$

leading to

$$\nabla^2 g = \Delta_u^2 g(u_p, v_q) + \Delta_v^2 g(u_p, v_q) \tag{2}$$
$$= g(u_{p+1}, v_q) + g(u_{p-1}, v_q) + g(u_p, v_{q+1}) + g(u_p, v_{q-1}) - 4g(u_p, v_q).$$

Applying the Laplacian operator to an entire matrix may be viewed as a discrete analog of convolution. That is, we can find the Laplacian of a component $g(u_p, u_q)$ of the image matrix by multiplying component-wise each value in the 3×3 neighborhood around the component with the "mask" matrix

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

and then summing all of the components of the resulting 3×3 matrix.

In light of **(2)**, we wish to convolve with the mask

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \alpha \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\alpha & 0 \\ -\alpha & 1+4\alpha & -\alpha \\ 0 & -\alpha & 0 \end{pmatrix},$$

which subtracts the Laplacian times a control parameter α (analogous to $\kappa\tau$) from the function itself.

Natural extensions of this technique are to use higher-order approximations of the Laplacian operator (thus convolving with a larger mask) or to enhance the mask with some sort of high-pass filter. Considerations of time and computational complexity prevented our development of such techniques.

Edge Detection

To make boundaries more visible in our images, it is useful to detect edges and generate corresponding line drawings. To do this, we use a variation of the finite differences method already discussed. At each point (x_i, y_j, z_k), we evaluate

$$\begin{aligned} \Delta_{2x}(i, j, k) &= A(i-1, j, k) - A(i+1, j, k) \\ \Delta_{2y}(i, j, k) &= A(i, j-1, k) - A(i, j+1, k) \\ \Delta_{2z}(i, j, k) &= A(i, j, k-1) - A(i, j, k+1). \end{aligned}$$

This set of definitions centers the difference around the point in question.

We construct a new array Γ, where

$$\Gamma(i, j, k) = \max\{\Delta_{2x}(i, j, k),\ \Delta_{2y}(i, j, k),\ \Delta_{2z}(i, j, k)\},$$

which gives a measure of how fast the values of A are changing through a given point.

This technique has many strong points:

- The values for Γ are easy to compute.

- The method does not bias the type of edge encountered (straight, curved, diagonal, etc.).

- The values of Γ remain within the range of the original grayscale range [Rosenfeld and Kak 1982].

We then apply our interpolation techniques to a plane passing though the box of data given by Γ rather than by A. Once we have this new two-dimensional image, we convert it into a binary image by applying a threshold condition: Every point with an interpolated value above the threshold value is made black, while every point with an interpolated value below the threshold value is made white. This converts regions where the difference values are high into black regions against a white background. Since edges will exhibit large differences in density, the black regions will thus represent edges.

Analysis of the Model

We implemented our plane-imaging algorithm on a Unix graphics workstation and analyzed the algorithm's behavior on several different data sets. One data set is a simulated MRI scan of a human brain, an example of a data volume that the model would expect to receive in a real-world medical imaging environment. We created several other contrived data sets to test our algorithms on known structures and thereby expose limitations of the model.

Computer Implementation

We built our model as an interactive graphical application using the C++ language and the OpenGL 3-D graphics library. We designed this program to possess many of the features that a plane-imaging system used in an actual medical situation would contain. It takes as input an arbitrarily-sized 3-D block of byte values (0–255) representing the density array A. The program presents two display windows to the user:

- The first window is a view of the (x, y, z) coordinate space, showing wire-frame representations of both the input data array and the projection plane.

- The second window shows the scanned image that lies on the plane as generated by our plane-imaging algorithm.

The user can use the keyboard and mouse to move the imaging plane to different positions (x_0, y_0, z_0) and angles (ϕ, θ) within A, viewing in real time how the projected image changes. We used this program to generate all of the figures in this paper.

The coded algorithms for interpolating in A and for creating the sharpened image in the projection plane are all straightforward translations of the mathematical expositions given earlier. We imparted some extra intelligence to the imaging algorithm so that it can determine in which part of the (u, v) plane the source data lie (see **Appendix**).

The user can control all parameters of the model, including the sharpening control factor and the edge-detection threshold. Different interpolation techniques may be selected, and the sharpening and edge-detecting filters may be toggled as well. The program executes quickly enough; but since we paid little attention to creating optimized algorithms, there is much much potential for speeding up operations such as the sharpening filter.

Results on Brain MRI Data

The principal test data set for our model is a simulated human brain MRI volume containing 181 slices, each 181×217 pixels. These data are the output of a highly realistic MRI computer simulation [Cocosko et al. 1997] and are thus likely to reflect data from an actual MRI scanner that would be of diagnostic interest to a user.

Exploration of Obliquely Oriented Structures

Figures 2–5 show sample output of our plane-imaging algorithm on the brain MRI data set, for several different plane orientations, both orthogonal and oblique. Viewing such output dynamically on a graphical computer display would allow a doctor to explore structures in the brain that lie on planes in any possible position and orientation.

Fine Structure

The brain MRI data set demonstrates the merits of applying the sharpening filter. **Figure 6** displays two brain images; the one on the left has been sharpened and the one on the right has not. The sharpened image lacks the overall blurriness of the unsharpened image. It also displays more clearly the fine structure in the brain that would be of interest to a surgeon planning a minimally invasive procedure.
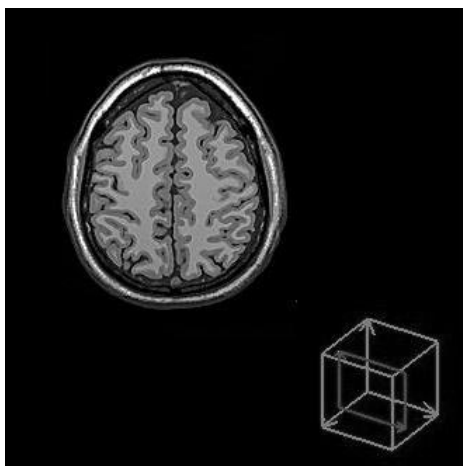
Figure 2. Image centered at $(91, 108, 120)$ with $(\phi, \theta) = (0, 0)$. The imaging plane is orthogonal to the z-axis.
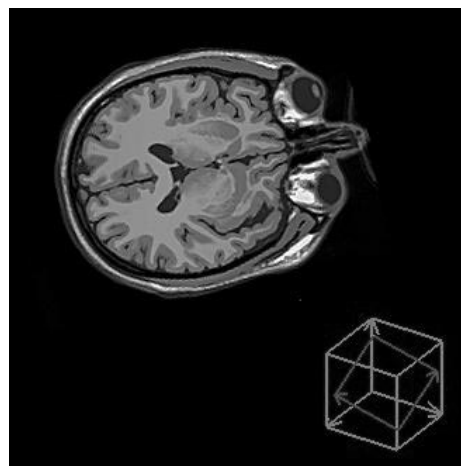

Figure 3. Image centered at $(91, 108, 120)$ with $(\phi, \theta) = (35, 75)$. The imaging plane lies obliquely within the data volume.
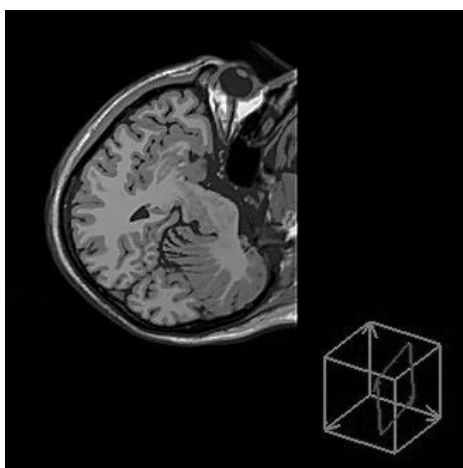

Figure 4. Image centered at $(110, 100, 70)$ with $(\phi, \theta) = (130, -30)$.
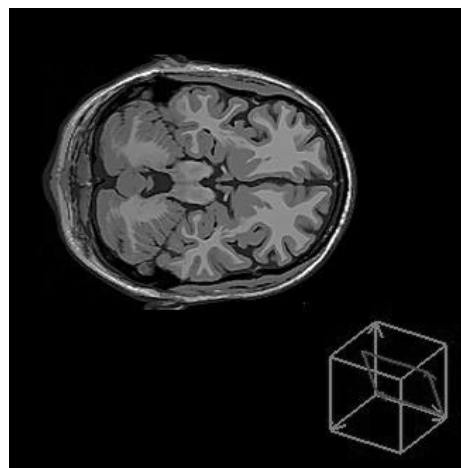

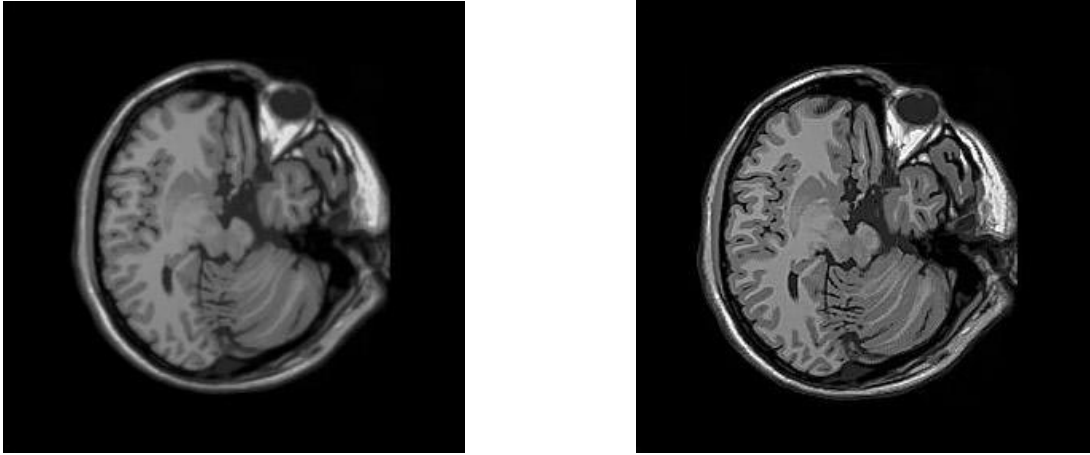**Figure 5.** Image centered at $(100, 100, 48)$ with $(\phi, \theta) = (-20, 90)$.

Figure 6. Comparison of two oblique plane scans of the brain MRI data. The image on the right has been passed through the sharpening filter, while the image on the left has not. The imaging plane is centered at $(85, 118, 58)$ with $(\phi, \theta) = (30, 0)$.

Line Drawings

The edge-detection algorithm that we described can be used to generate black-and-white line drawings of the kind found in an anatomical atlas (see **Figure 7**). Such drawings are useful for seeing clear boundaries of structures in the brain.
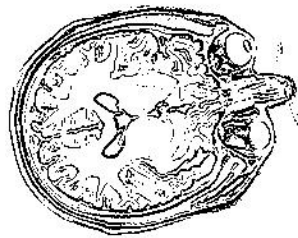

Figure 7. Black-and-white line drawing of an oblique plane through the brain MRI data. Lines were drawn using edge detection. The imaging plane is the same as in **Figure 3**.

Results on Contrived Data Sets

A question that arises about our model is whether or not the sharpening filter is removing the blurriness introduced by the interpolation algorithm or just the blurriness inherent in the data array. To attempt to answer this question, we examined the operation of our model on a data set with perfectly discrete boundaries—a data set consisting of "slabs," that is, parallel, evenly spaced planes of some nominal depth containing maximum intensity pixels. **Figure 8** shows the close-up results of passing an imaging plane through this volume at an angle of $35°$ to the slabs (an arbitrary choice). Comparing the clarity of the

slab edges in the two different images demonstrates that our sharpening filter performs well in removing effects of interpolation, provided that boundaries cross the image plane at a sufficiently large angle. When the imaging plane is at a small angle to the slabs, as in **Figure 9**, we see blurred edges even in a sharpened image—an inevitable consequence of our 3-D interpolation scheme.
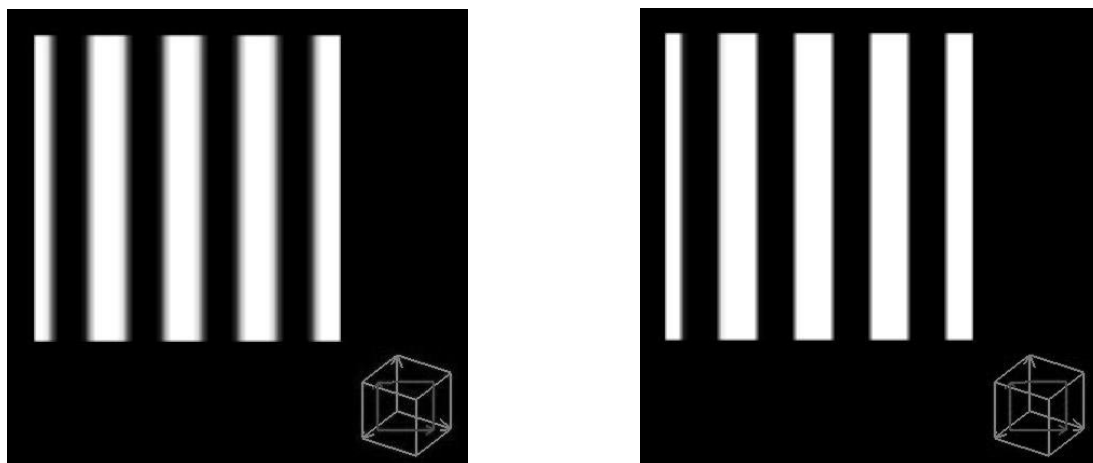

Figure 8. Comparison of two oblique plane scans of the "slabs" data. The angle of incidence between the imaging plane and the slab visible is $35°$. The image on the right has been passed through the sharpening filter, while the image on the left has not; resulting images are close-up views.

The values in the data set can be viewed as the discrete extreme of behaviors that our model can expect to encounter. To examine the opposite continuous extreme, we created a data set whose intensity does not vary in a single xy-plane but instead varies as a cubic function of z. **Figure 10** shows the projected image from a plane passing through the data volume at angle of $35°$ to the xy-plane. The interpolation algorithm in our model linearizes much of the nonlinearly varying data, and subsequently the sharpening filter introduces distortion.

Limitations of the Model

Our model has several limitations:

- Arrays of non-uniformly or anisotropically sampled data are not considered.

- Objects with planar edges parallel or nearly parallel to the projection plane are imaged inaccurately.

- Interpolation and smoothing generate minor distortions that, given more time and work, may be alleviated by the use of higher order schemes. Using additional contrived data structures as algorithm input may further illuminate the causes of such distortions.
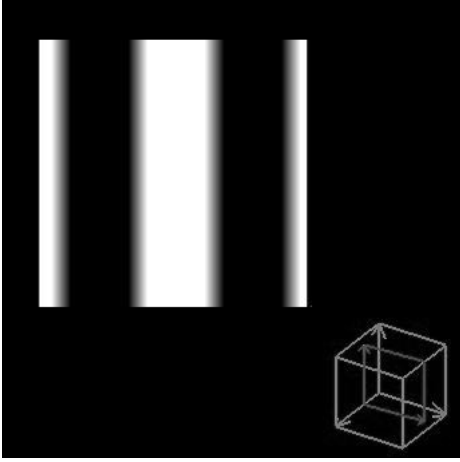
Figure 9. A nearly parallel oblique plane scan of the slabs data (angle of incidence is 5°) with sharpening filter applied. Edge ramping effects are visible.
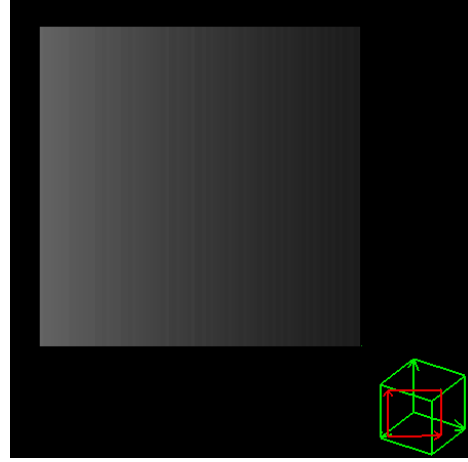

Figure 10. Oblique plane scan of a continuously varying data volume. Some linearization effects of the sharpening filter are visible.

Conclusions

Our computer implementation provides convincing pictures that illustrate the ability of our model to depict density variations in oblique planes. Our algorithm surpasses most existing algorithms by allowing any planar orientation, by generating images quickly, and by including sharpening and edge-detecting filters. The results of testing our model on simulated brain MRI images demonstrates its applicability to real-world medical imaging.

Appendix: Bounds of the Imaging Plane

For computational optimization of our plane-imaging algorithm, we derive bounds for the region of intersection of the plane and the box of data. This allows our algorithm to iterate over the smallest (u, v) region necessary to ensure that all of the intersected data have been obtained.

The transformation $T(u, v)$ gives the three equations

$$
\begin{aligned}
x &= \cos\phi\cos\theta\, u &-& \sin\theta\, v &+& x_0 \\
y &= \cos\phi\sin\theta\, u &+& \cos\theta\, v &+& y_0 \\
z &= -\sin\phi\, u & & &+& z_0.
\end{aligned}
$$

We can rewrite these to get

$$
\begin{aligned}
x - x_0 &= \cos\phi\cos\theta\, u &-& \sin\theta\, v \\
y - y_0 &= \cos\phi\sin\theta\, u &+& \cos\theta\, v \\
z - z_0 &= -\sin\phi\, u & &\,.
\end{aligned}
$$

We want to know when the plane crosses an edge of the box of data in $\mathbb{R}^3$. This

edge will be constant in two of the three variables (x, y, z). We can plug these two values into the appropriate equations above and solve for u and v.

For instance, suppose we want to know at what point (u, v) the plane intersects one of the edges of the data box that is parallel to the z-axis. In this case, we know what x and y are since we know the size and position of the box in $\mathbb{R}^3$. Using the equations for $x - x_0$ and $y - y_0$, we have:

$$\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = \begin{pmatrix} \cos\phi\cos\theta & -\sin\theta \\ \cos\phi\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

Notice that if $\phi = \pi n + \frac{\pi}{2} n$, where $n \in \mathbb{Z}$, this transformation does not have an inverse. In this case, the plane is parallel to the z-axis, and hence any vertical edge of the data box. If $\phi \neq \pi n + \frac{\pi}{2}$, then we can invert the above transformation to obtain

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{\cos\phi} \begin{pmatrix} \cos\theta & \sin\theta \\ -\cos\phi\sin\theta & \cos\phi\cos\theta \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}.$$

We can perform similar operations for the other two directions in $\mathbb{R}^3$ with similar restrictions on ϕ and θ. We thus obtain 12 points in the uv-plane, or fewer if the plane is parallel to one of the axes. With these 12 values, we simply choose the largest and smallest values for u and v to get a rectangle that tightly bounds the intersection of the plane and the data box.

References

Andrews, H.C., and B.R. Hunt. 1977. *Digital Image Restoration.* Englewood Cliffs, NJ: Prentice-Hall.

Cocosco, Chris A., Vasken Kollokian, Remi K.-S. Kwan, and Alan C. Evans. 1997. BrainWeb: Simulated Brain Database. `http://www.bic.mni.mcgill.ca/brainweb/`.

Rosenfeld, Azriel, and Avinash C. Kak. 1982. *Digital Picture Processing.* 2 vols. San Diego, CA: Academic Press.

Russ, John C. 1995. *The Image Processing Handbook.* Boca Raton, FL: CRC Press.

A Tricubic Interpolation Algorithm for MRI Image Cross Sections

Paul Cantrell
Nick Weininger
Tamás Németh-Csőri
Macalester College
St. Paul, MN 55105

Advisor: Karla V. Ballman

Introduction

We designed and implemented a program capable of:

- taking in a large three-dimensional array of one-byte grayscale voxels (volume "pixels"), the output from an MRI machine;

- slicing through that array along an arbitrary plane;

- and using interpolation to produce an image of the cross section described by the plane.

We allow the user to select the plane of cross section by specifying three points that should be in the plane, or by specifying one point and two angles. We account for the possibility of voxels of unequal size in different dimensions, but presume they are evenly spaced in each dimension. We then use a tricubic interpolation algorithm to produce a cross-sectional image. This method is our extension of bicubic interpolation, an algorithm used widely with two-dimensional images. We chose the tricubic method because it offers an optimal balance of accuracy and computational speed. Finally, we allow the user to "stain," or color, important portions of the data.

We tested the program on simple geometric figures to verify its correctness. We then tested it on actual MRI image slices of four brains, with very satisfactory results. We found that important image features were preserved well and that image staining was useful in visualization. The interpolation algorithm runs in linear time; it produces an image from a $256 \times 256 \times 256$ data volume in a few seconds.

Finally, we constructed several data sets that point out the limitations of our algorithm and of the problem itself. These limitations concern the behavior of our algorithm in areas of maximal uncertainty, farthest from the sample points.

Design Considerations

Typical Uses of MRI Images

As described in Rodriguez [1995], MRI scans of various parts of the body are used to diagnose a wide range of disorders. One of the most common uses is the detection of abnormal bodies in the brain, such as tumors, cysts, and hematomas. Because of variations in appearance of both healthy tissues and tumors, it is critical that the sharpness or fuzziness of boundaries, as well as the general shape and brightness of regions, should be preserved when taking cross sections.

An analysis program should rely on intuitive spatial understanding and also provide a straightforward way for a user to specify a volume for highlighting in subsequent cross sections.

Characteristics of Image Data

Data Size

The usual data size for one MRI slice is 256×256 grayscale pixels. In order to get data covering an entire 3-D object, multiple slices are required. The time taken for each scanning slice is dependent on a parameter to the scanning process called *repetition time*; a typical slice might take several minutes to scan, though multiple slices may sometimes be scanned simultaneously [Hornak 1997]. Since the amount of time that patients can spend immobile in the machine is limited, the number of slices that can be taken is small compared to the slice resolution. The database for our real-world test data [Johnson and Becker 1997] typically took 25–60 slices to scan an entire brain.

This means that the actual volume of space represented by each voxel is likely not to be a cube. Instead, it will be a rectangular prism, significantly longer in one dimension than in the other two; the algorithm will need to take this fact into account so as not to produce distorted output. Furthermore, if a voxel is much longer along one axis than along the others, much more interpolation in that dimension will be required, so images taken in planes parallel to that axis may be especially inaccurate.

Data Artifacts

Many different types of artifacts may be present in MRI image data; some are results of incorrect operation or configuration of the machine, while others

are products of the physical properties of the scanning process [Ballinger 1997; Hornak 1997].

Since most of these types of artifacts reflect problems with the machine's configuration that may produce misleading images, it is important that they be preserved in cross section, so that the MRI operator can see them and recalibrate the machine appropriately.

Sampling Characteristics

The manifestation of all of these image characteristics in data is fundamentally tied to the properties of discrete sampling. We can classify data in which discrete sample points describe a continuous function (as our data do) as either *undersampled*, *oversampled*, or *critically sampled* (see **Figure 1**), depending on how the sampling resolution corresponds to the actual detail in the image.
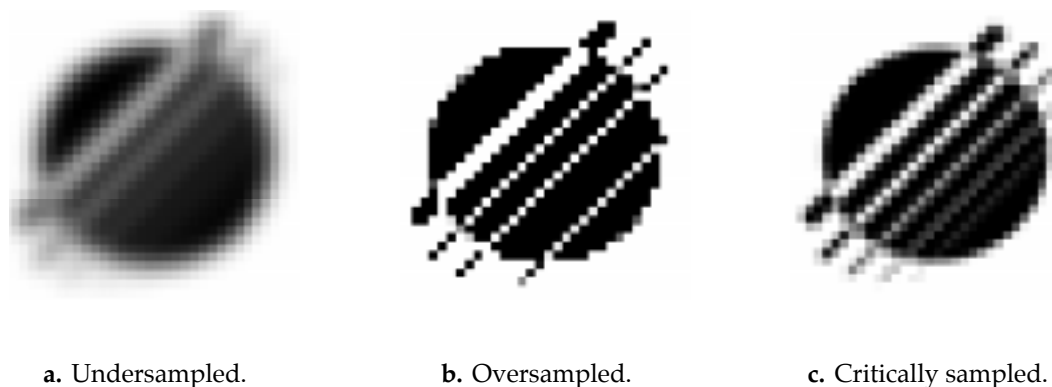

a. Undersampled.   **b.** Oversampled.   **c.** Critically sampled.

Figure 1. Typical data sampling characteristics for images.

- Oversampled data: The sample grid is finer than the image detail. Such images tend to look very blurry, and neighboring grid points tend to vary only slightly and contain essentially redundant information. This high level of detail lends these images to accurate interpolation and enhancement.

- Undersampled data: The image contains detail finer than the sample grid and there is little correlation between neighboring pixels, especially at the edges of objects in the image. If the actual sample area for each pixel is smaller than the sample area that the pixel represents, the image may be characterized by jagged edges and sharp contrasts. Such images make interpolation and enhancement a matter of heuristics and guesswork.

- Critically sampled data lies at the border of undersampling and oversampling, and MRI data fall into this category. As with oversampled data, the edges of boundaries tend to be unaliased (smooth), and the image may even appear slightly blurry; however, as with undersampled data, the detail at the pixel level is important, and interpolation possibilities are limited.

Interpolation Algorithms

Our input data come as a set of image values taken at discrete points, but the cross sections that we want to take may not pass exactly through any of these points. Therefore, we need a way to estimate image values at arbitrary points based on the image values at the sample points. That is, based on our array of samples $A_{i,j,k}$, we want an interpolating function $f : \quad^3 \rightarrow \quad$ such that

$$f(i, j, k) = A_{i,j,k}$$

when i, j, and k are integers, and such that f takes on reasonable values for nonintegral i, j, k. (This stipulation that the interpolating function match the sample points is reasonable, as MRI images tend to be very clean and have a high signal-to-noise ratio.)

In choosing an interpolating function, we had to make a trade-off between accuracy of image production and running time, limited by the typically critically sampled nature of MRI data. We chose a cubic method, which we found to be surprisingly fast and quite accurate for actual MRI data.

Tricubic Interpolation

Cubic interpolation is a special case of Lagrange interpolation, which is a simple method of finding the unique polynomial of degree $(n-1)$ that passes through n data points [Mnuk 1997].

We consider first the one-dimensional case. Cubic interpolation begins with the four sample points closest to the target point x—its two nearest neighbors on either side, $\lfloor x \rfloor$, $\lfloor x \rfloor - 1$, $\lceil x \rceil$, and $\lceil x \rceil + 1$—and fits a cubic function $p : \quad \rightarrow \quad$ to them; $p(x)$ gives the interpolated value at x (see **Figure 2**). Note that the particular f described by these four points around x gives the values only for the region between the middle two. Thus, the function f that interpolates the whole image is a piecewise composite of many different cubic functions.
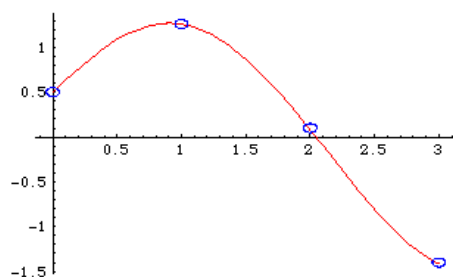

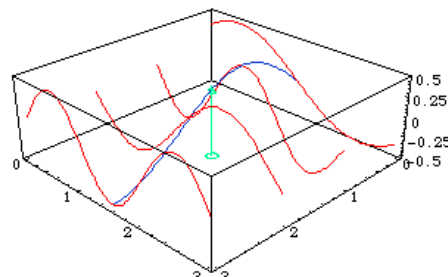**Figure 2.** One-dimensional cubic interpolation.

Figure 3. Two-dimensional cubic interpolation.

This procedure generalizes nicely to multiple dimensions. It does not require, as one might expect, the construction of an elaborate multivariate poly-

nomial or the solution of a large system of equations; in fact, it is sufficient to perform the interpolation in each dimension consecutively.

It is perhaps easiest to visualize this process in two dimensions first [Makivic 1996]. As shown in **Figure 3**, we separate the 16 points surrounding the target point into four lines of four points each. We do a one-dimensional cubic interpolation along each of these lines and evaluate the resulting cubics at points along a perpendicular line containing the target point. We then use *these* four evaluated points to interpolate another cubic that we can evaluate at the target point.

We can then extend this to three dimensions in the obvious way: Split the 64 points into four planes of 16 points each. In each of these planes, perform the two-dimensional process to get four interpolation points along a line through the target point. Finally, perform an interpolation to get a function value for our target point. This requires a total of 21 one-dimensional interpolations, five for each plane plus the final one. The process is illustrated in **Figure 4**. It runs in time linear in the total number of voxels in the volume.
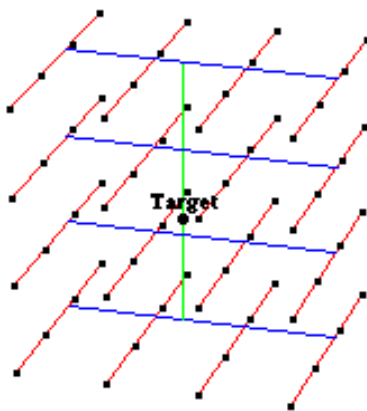

Figure 4. Schematic of three-dimensional cubic interpolation.

One key question arises: Does it matter how we choose the planes, and the lines within each plane? It turns out that the final value obtained at the target point is independent of the order in which the dimensions are chosen for the interpolation. [EDITOR'S NOTE: We omit the authors' proof.]

Cubic interpolation is particularly appropriate to critically sampled data. It relies on some correlation and continuity between neighboring points, producing smooth curves that fit slightly smoothed object edges very well without introducing artificial detail not present in the original image. However, it does not oversmooth or mangle detail past the single-pixel level, and it introduces minimal artifacts. Although cubic interpolation performs poorly on undersampled or jagged data, it is appropriate for MRI data.

Furthermore, because it involves only simple arithmetic and is linear in the size of the data set, bicubic interpolation is fast enough to produce typical MRI images in close to real time without a high-end workstation. The results of a two-dimensional bicubic enlargement are in **Figure 5a**.
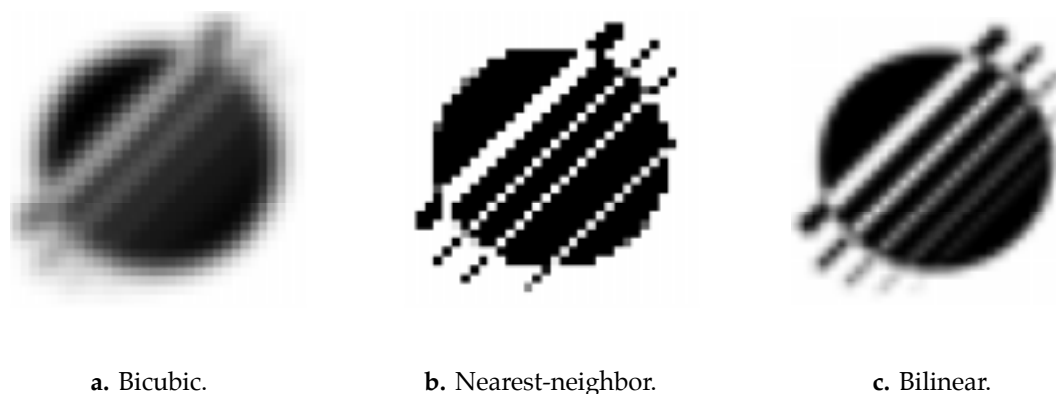
a. Bicubic.   **b.** Nearest-neighbor.   **c.** Bilinear.

Figure 5. Enlargements by various interpolation algorithms.

Alternative Interpolation Methods

We considered and rejected several alternative interpolation methods.

Nearest-Neighbor Interpolation

In nearest-neighbor interpolation, the image value at an arbitrary point is the value of the nearest sample point. This is a very fast algorithm—it requires only a single rounding operator for each dimension.

Nearest-neighbor interpolation is often appropriate to undersampled data because it preserves the jaggedness of such data and does not attempt presume that image fades smoothly across the sharp edges. It would probably be the most appropriate method for cross sections of black-and-white line art medical diagrams.

However, for these same reasons, it performs very poorly on critically sampled data, and its rounding can actually locally distort image proportions where a smoother interpolation would preserve them (**Figure 5b**).

Linear Interpolation

Linear interpolation takes the image value at a point to be the average of the image values of all its known neighbors, weighted by how far away they are. This is also a very simple and fast algorithm. However, it is little more than a blurring of the nearest-neighbor method and shares many of its problems. In particular, it leaves object edges jagged and uneven, even when they are not aliased, and tends to blur excessively (**Figure 5c**).

Convolution-Based Methods

There is a wide variety of much more intricate interpolation methods based on the convolution of the image matrix, including Fourier-based methods, CMB interpolation, and Wiener enhancement.

Although they perform extraordinarily well, for several reasons they are inappropriate to the task at hand. These methods are primarily targeted at oversampled data and tend to act more as de-blurring algorithms than interpolators. Furthermore, since they require taking the convolution, they are not linear in the data set size and tend to be quite slow (Mahan [1996] describes a run of dozens of hours to enhance a small image of Saturn).

Since MRI data tend not to be particularly blurry, and since enlargement is not our goal, these computationally expensive algorithms are simply overkill. Furthermore, especially in critically sampled data, they are likely to produce artifacts with visually striking large-scale structure, which could be misleading to a reader of the image and lead to a misdiagnosis.

Image Enhancement

Either during or after interpolation, we have the option of enhancing the image to sharpen blurred regions, enhance edges, or otherwise bring out details. However, we found that the tricubic method performed well enough that most of these methods were either inappropriate or unnecessary. The human eye is extremely adept at interpolation of obscured detail, and tricubic interpolation tends to capitalize on this by producing blurry but suggestive output in regions of uncertainty. The enhancement algorithms that we examined revealed no details that the eye could not already interpolate. Given the dangers in introducing artificial detail in medical imaging, we decided to leave our tricubic method unenhanced.

Anti-Aliasing

For sharp or undersampled data, it can be beneficial to blur high-contrast edges slightly. However, this is counterproductive in our case; the detail in our images is significant, but the edges of objects are not generally aliased.

Sharpening

Traditional sharpening algorithms work by moving a pixel's value away from the average of its neighbors, perhaps weighted so that the sharpening will be localized to the edges of objects.

A major problem with this sort of sharpening is that it can produce jagged edges and exaggerate the effects of noise in data (**Figure 6**). Since tricubic interpolation can produce blurry output in heavily interpolated regions, such sharpening could be of use. However, we found that it made little visible difference in cross sections of real data and revealed no significant new detail.
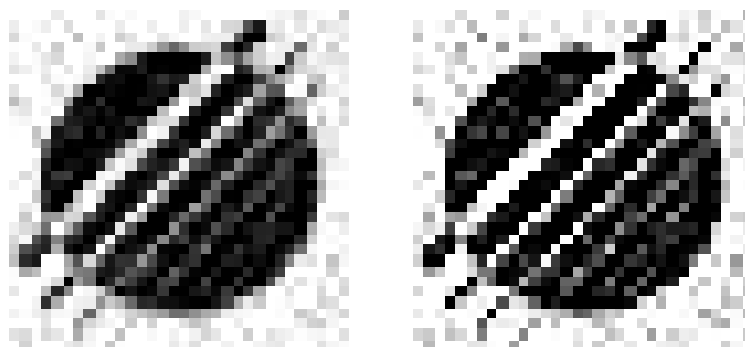
Figure 6. Oversharpening increases noise and aliases edges.

Edge Fitting

Many algorithms work to enhance jagged or blurry edges by finding the edges in an image (the areas where pixels differ from the average of their neighbors), fitting curves to them, and enhancing them, either by anti-aliasing or selective sharpening. Such algorithms, however, are usually used as an aesthetic enhancement and not in situations that call for scientific accuracy; although their output can be very pleasing to the eye, they are guilty in the extreme of introducing artificial detail. Thus, although they might produce visually pleasing results in our problem, they should be used very judiciously if at all.

Image Staining

Even with very clear interpolations and a graphical aid to help visualize the orientation of various cross sections, it can be difficult for a human to mentally compose cross sections into a solid and to identify the same object in different cuts. To assist in this visualization process, it would be nice to allow the user to "stain" portions of the data with a color. A physician could mark a reference point or an anomalous object and be sure of its location in different cross sections. Such a feature should work so as to make a colored area visible even in cross sections that do not exactly intersect the region that the user originally marked. Thus, a marked region should be "feathered" or blurred outward slightly from the plane in which the user places it.

Implementing Our Algorithm

Specifying the Plane of Cross Section

We allow the user to specify the plane for a cross section by either:

- selecting three noncollinear points in the plane of the cross section, a good method for selecting an initial arbitrary cut based on image features; or

- selecting an arbitrary point to be included and specifying two angles, known as *Euler angles*. The first angle is the angle between the xy-plane and the plane of cross section; the second is the angle between the x-axis and the intersection of the cross-sectional plane with the xy-plane. This is a good method for continuously traversing the image or for fine-tuning the orientation of a particular cut.

To calculate the cross section, we need to transform the input data into a triplet $(\vec{p}, \hat{x}, \hat{y})$, where $\vec{p}$ is an arbitrary point on the plane and $(\hat{x}, \hat{y})$ forms an orthonormal basis for the plane.

Three-Point Representation

To obtain $(\vec{p}, \hat{x}, \hat{y})$ from the three-point representation $(\vec{p_1}, \vec{p_2}, \vec{p_3})$, we take the cross product of the two vectors $\vec{p_2} - \vec{p_1}$ and $\vec{p_3} - \vec{p_1}$ to produce the normal vector $\vec{n}$, which is perpendicular to the plane. Then we solve the system of equations

$$\vec{n} \cdot \hat{x} = \vec{n} \cdot \hat{y} = 0, \qquad \hat{x} \cdot \hat{x} = \hat{y} \cdot \hat{y} = 1, \qquad \hat{x} \cdot \hat{y} = 0, \qquad \hat{x}_z = 0$$

for $\hat{x}$ and $\hat{y}$. The first two equations ensure that the basis vectors are in the plane; the next two ensure they are of unit magnitude; the fifth makes them perpendicular. These five equations in six unknowns do not specify a unique basis, so we need one more constraint. We choose $\hat{x}_z = 0$ as that last constraint because it simplifies the resulting formulas greatly. Finally, we let $\vec{p} = \vec{p_1}$.

Point Plus Euler Angles

For input of the form $(\vec{p}, \phi, \theta)$, we think of the plane as a rotation of the xy-plane, with the origin set to $\vec{p}$. We first rotate the plane by ϕ around the x-axis, and then rotate it by θ around the z-axis. The resulting transformations are given by

$$\hat{x} = R_\theta R_\phi \hat{i}, \qquad \hat{y} = R_\theta R_\phi \hat{j},$$

where $\hat{i}$, $\hat{j}$ are the standard basis for the xy plane and

$$R_\phi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{pmatrix}, \qquad R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The vectors that we obtain from these methods are not always the best ones for our purposes. We would like the display orientation to correspond to the user's concept of the volume; that is, up should remain up and left should remain left whenever possible. Therefore, we try to align the basis vectors as

closely as possible with the xy-plane's basis. To do this, we rotate the basis vectors in the plane so as to maximize $\hat{x} \cdot \vec{i}$, thus bringing the $\hat{x}$ vector as close as possible to the true x-axis. We then reverse the direction of $\hat{y}$ (effectively flipping the image over) if that reversal increases the value of $\hat{y} \cdot \vec{j}$.

Once we have our adjusted basis, we calculate where, if anywhere, the plane intersects each of the 12 edges of the data volume. These edge intersection points define the boundary of the cross section of the volume. We define the data volume to be the parallelepiped with corners at $(0,0,0)$ and $(x_{\max}, y_{\max}, z_{\max})$. Each edge has two fixed coordinates; thus, we can compute each edge intersection by solving two equations in two unknowns. For example, to compute the point of intersection with the edge running along the z-axis, we solve the equation system

$$p_x + c_1 \hat{x}_x + c_2 \hat{y}_x = 0, \qquad p_y + c_1 \hat{x}_y + c_2 \hat{y}_y = 0$$

(where p_x is the x-component of $\vec{p}$, $\hat{x}_x$ is the x-component of $\hat{x}$, and so on) for c_1 and c_2. This system will always have a unique solution unless the plane is parallel to the z-axis. If that happens, either the plane does not intersect the z-axis, or else the z-axis lies in the plane; in the latter case, we take the two endpoints of the edge at 0 and $z_{\max}$ to be the intersection points.

If we have unique values for c_1 and c_2, we then solve for the z-coordinate of the intersection: $z_{\text{intercept}} = p_z + c_1 \hat{x}_z + c_2 \hat{y}_z$. If $z_{\text{intercept}} \in [0, z_{\max}]$, then the plane does indeed intersect this edge of the data volume. Otherwise, it intersects the line defined by the edge at a point outside the volume.

When we have all the edge intersection points, we can define a rectangle bounding the cross section in terms of the basis vectors: Just take the maximum and minimum values of c_1 and c_2 over all the points. Finally, we calculate the upper left corner of the bounding rectangle and proceed to the interpolation.

Performing the Interpolation

At this phase in the computation, we scale the individual components of the $\vec{p}$, $\hat{x}$, and $\hat{y}$ to account for the possibility of voxels with different sizes in different dimensions, which could result from MRI slices taken far apart. In other words, if the actual size of a voxel is $(a\ b\ c)$, we scale from the basis

$$(1\ 0\ 0) \qquad (0\ 1\ 0) \qquad (0\ 0\ 1)$$

(which reflects geometric reality) to the basis

$$(a\ 0\ 0) \qquad (0\ b\ 0) \qquad (0\ 0\ c)$$

(which is appropriate for our data array). This method presumes that the size of a voxel in each dimension is constant and thus that MRI slices are spaced evenly.

The cross-section sample points are now described by $\vec{p} + a\hat{x} + b\hat{y}$, where a and b are integers. From this, we can easily construct a double loop to traverse the cross section.

At each of these cross-section sample points, we perform a tricubic interpolation. Since doing so involves two neighbor points in each direction, we define the value of a sample point outside the data array to be a uniform dark gray, so that we can interpolate for points near the edge.

The assumption that the data slices are equally spaced allows a number of simplifications in the Lagrange polynomials and thus in the code to perform the interpolation. Allowing for uneven voxel size within a dimension (as might result from an uneven series of slices) would require substantial extension of this portion of the program.

Performing Image Staining

[EDITOR'S NOTE: We omit the authors' description of implementation of the staining feature.]

Testing the Algorithm

Correctness Testing

As a simple test that the algorithm was working properly, we used it to take sections at various angles through two different geometric objects (see **Figure 7**):

- a cube, filled with smaller cubes alternating black and white in a checkerboard pattern; and

- a torus, filled according to a variable grayscale gradient, with three perpendicular cylinders of different diameters, filled with white, intersecting at the center of the torus.

We chose these objects because their correct cross sections at any given angle are readily identifiable. The algorithm did indeed take correct cross sections of these objects at a variety of angles.

Real-World Testing

To provide test data reflecting conditions actually encountered in diagnosis, we downloaded four series of axial (xy-plane) MRI slices from the Whole Brain Atlas [Johnson and Becker 1997], a database of information on brain anatomy and pathology. We converted these slices into four three-dimensional arrays of test data using Adobe Photoshop.
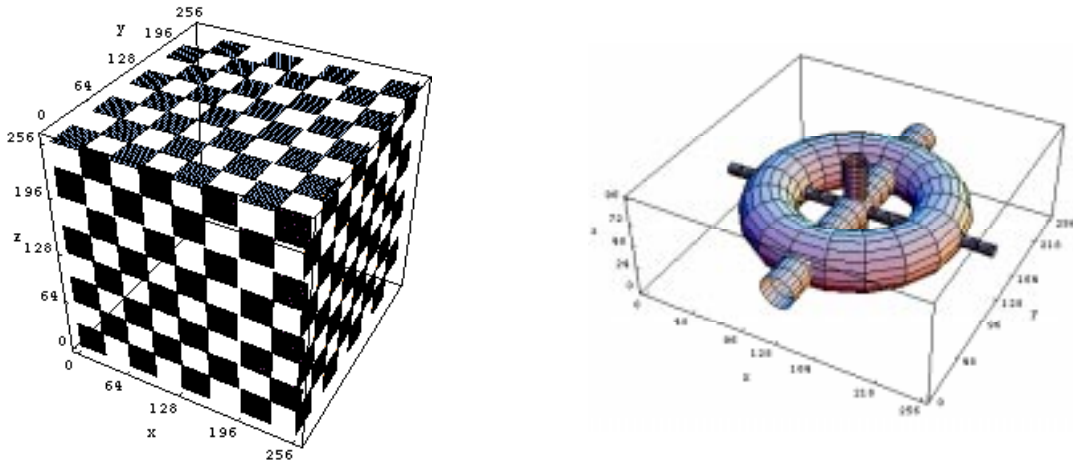
Figure 7. Geometric test objects.

- Data Set 1 was from a normal, healthy brain;

- Data Set 2 was from a brain containing a type of tumor known as a glioblastoma;

- Data Set 3 was from a brain affected by cerebral hemorrhaging; and

- Data Set 4 was from the brain of a woman with advanced Alzheimer's disease.

Examples of the resulting cross-sectional images are shown in **Figure 8**. We found that the algorithm worked better in perpendicular planes than in oblique planes, as expected. However, almost all the images were quite sharp and clear, preserving object boundaries and shapes excellently. Note in particular the series of oblique cross sections of Data Set 2, showing the shape and boundaries of the glioblastoma with great clarity.

When our image data contained artifacts, these too were preserved; they occurred most notably in the images from Data Set 3, which also produced by far the lowest-quality images. The primary reason for this is that it contained only 24 slices, as against 54 for Data Set 1, 56 for Data Set 2, and 45 for Data Set 4. Thus, the pixels in this data set were much more "stretched" in the z-direction, forcing the algorithm to do more vertical interpolation in taking cross sections.

Finally, we found that staining worked well in highlighting important image features in different cross sections. [EDITOR'S NOTE: We must omit the authors' two-color figures, which strikingly highlight the hemorrhage in Data Set 3.]
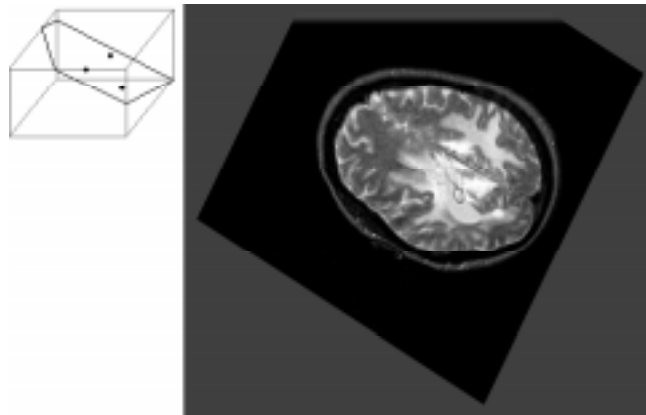
Problems in Our Algorithm

Our interpolation produces a slight increase in blurriness or fuzziness of edges characteristic of most image interpolation methods. Different cuts of the
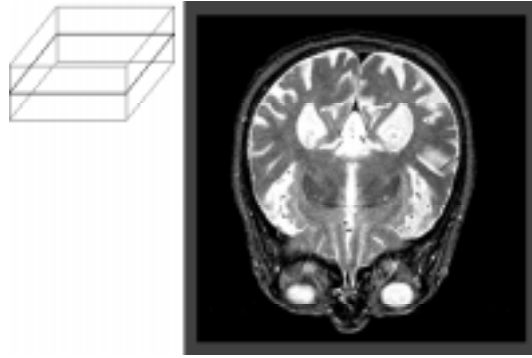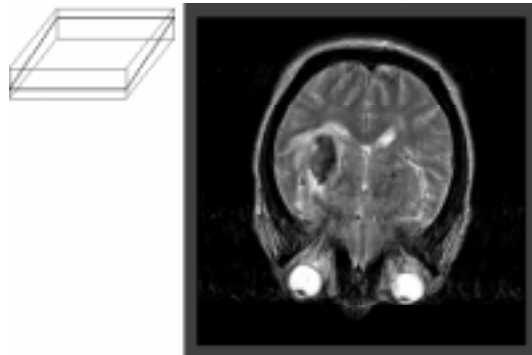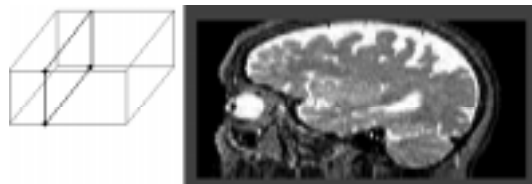
Slice from Data Set 1, the brain of a healthy elderly woman.

Slice from Data Set 2, the brain of a man with glioblastoma. The large bright mass at lower right is the tumor.

Slice from Data Set 3, the brain of a man with acute cerebral hemorrhage (the dark mass on the right side of the brain). Note the image artifacts; this is our lowest-quality data set.

Slice from Data Set 4, the brain of a woman in the advanced stages of Alzheimer's disease. Note the enlarged lateral ventricles (bright structures in the center of the brain) and unusually large, bright convolutions at the top of the image.

Figure 8. Sample slices from the four data sets.

data can produce widely varying results; compare (see **Figure 9**):

A. a cut along a horizontal plane, in which the points in the cut correspond to actual data points and we have maximal clarity (**Figure 9a**);

B. a cut along a horizontal plane halfway between two actual slices, in which points on the slice are blurred between the neighboring layers (**Figure 9b**);

C. a vertical cut, in which the image is blurry in the z dimension, where the voxels are very tall (**Figure 9c**); and

D. a maximally oblique cut, in we are cutting across the long diagonal of a voxel (**Figure 9d**; note the jaggedness along the edge of the skull).

We do have some control over this blurring. Examples A and B are essentially the same image, but the former is slightly clearer; a smarter algorithm could take this into account. However, these are special cases; in general, we cannot avoid moving through areas that require a high degree of interpolation.

We can illustrate this problem with the extreme case of a three-dimensional checkerboard of $1 \times 1 \times 1$-voxel black and white squares. Tricubic interpolation gives an even 50% gray at points midway between data points; thus, a cross section of this object shows high-contrast checkering in areas with a low degree of interpolation, and gray in areas that are heavily interpolated. Note, for example, the effect of shifting a horizontal plane up half a pixel just as we did in examples A and B above (**Figure 10a**). Because neighboring pixels differ so much, the blurring effect is much more pronounced.

A nearly horizontal but slightly oblique plane passes closer and farther from data points, producing an interference pattern (**Figure 10b**). The gray areas in this image correspond to points where sharp edges would blur slightly in real sample data. We could move these interference patterns around by translating the plane of cross section and rotating our basis vectors. In certain circumstances, this can actually decrease the overall blurriness, as in examples A and B. However, we cannot eliminate the interference pattern without compromising the integrity of the interpolation.

It might seem that we could perturb the sample plane slightly, bending it toward data points (i.e., avoiding the gray areas). However, as the perturbation increases, this algorithm degrades to nearest-neighbor, which distorts proportions locally and essentially defeats the purpose of interpolation. We experimented with restricting this perturbation to a direction normal to the plane; however, we found that unadorned bicubic interpolation worked best.

It is important to realize that the checkerboard is a very poor model of actual data—it is nothing but very high-contrast noise, which is not at all typical of MRI data. Its usefulness lies in illustrating the fundamental problem of discrete sampling: We simply cannot avoid approximating the values for a significant portion of an arbitrary cut.

The blurring that tricubic interpolation produces, however, does not mangle detail beyond the one-voxel level. Even a $2 \times 2 \times 2$-voxel checkerboard shows its
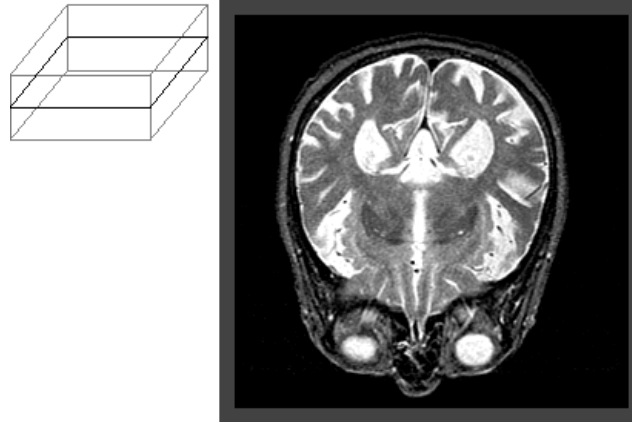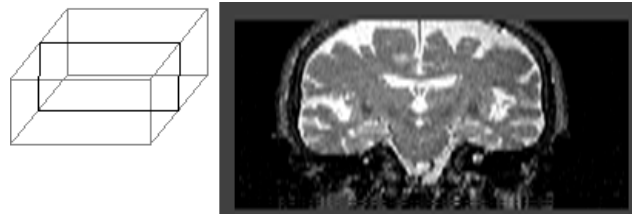
a. A cut along a horizontal plane; points in the cut correspond to actual data points and we have maximal clarity.
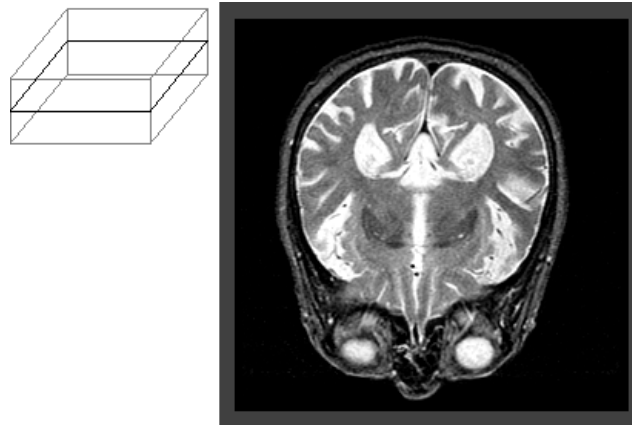
b. A cut along a horizontal plane halfway between two actual slices; points are blurred between the neighboring layers.

c. A vertical cut; the image is blurry in the z dimension, where the voxels are very tall.
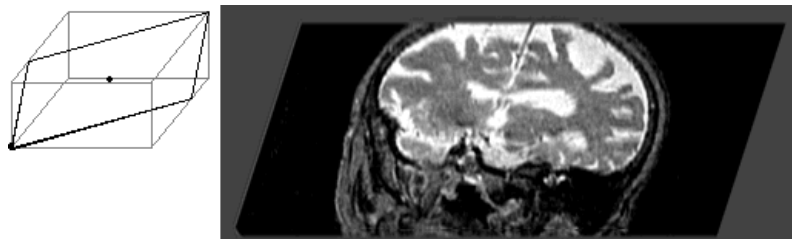
d. A maximally oblique cut, across the long diagonal of a voxel; note the jaggedness along the edge of the skull.

Figure 9. Slices along various planes through Data Set 1.

a. A cut along a data plane (center) and a cut halfway between data planes (right).

b. A cut along a nearly horizontal but slightly oblique plane produces interference bands.

c. A cut along a grossly oblique plane produces two families of interference bands.

d. A cut along an oblique plane through a checkerboard with $2 \times 2 \times 2$ voxels.
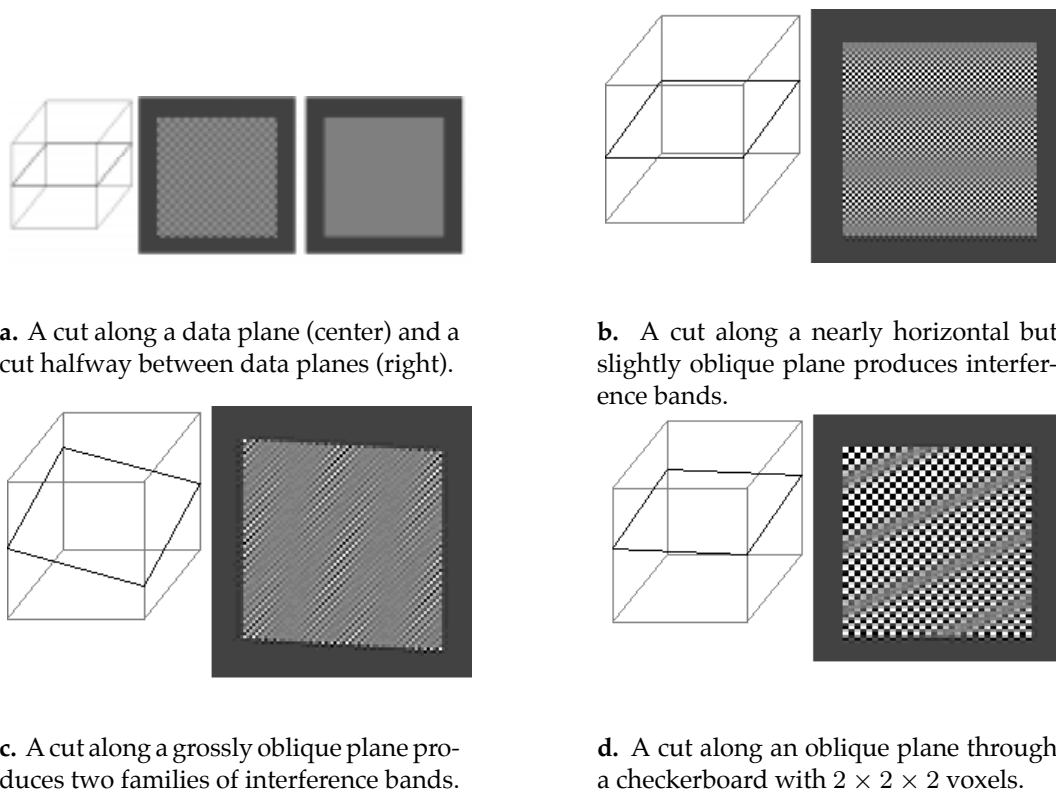
Figure 10. Various cross sections of 3-D checkerboards.

checkered pattern more clearly than the $1 \times 1 \times 1$ at oblique angles (**Figures 10cd**), and real data are even better behaved.

One other problematic situation is machine calibration. Suppose that the user has scanned a solid cube to align the machine and now wishes to know the exact angle at which that cube is oriented in the data array. The user could use our algorithm to align a cut with the top of the cube. However, the precision of the image degrades when the angle between the cross section and the cube is very small, especially if the cube is only slightly offset from the axes of the data array. The pixellated top of the cube produces a mild interference pattern, and the user would have to re-scan once or twice to align the scanner past sample resolution. In this case, an edge-fitting enhancement algorithm would be entirely appropriate.

Conclusion

Our algorithm's strengths working with real MRI data are:

- robust and intuitive specification of the cross-section plane,

- preservation of large-scale image features,

- preservation of fine detail present in the source image,

- preservation of image proportions given knowledge of the voxel dimensions,

- preservation of features of diagnostic interest,

- conceptually useful coloring of three-dimensional features in the image, and

- real-time performance sufficient to allow interactive exploration of the data.

There is, of course, a good deal of room for improvement. Our current implementation is not as fast as it should be, nor as easy to use. There are circumstances where it would be useful to offer several interpolation options (e.g., nearest-neighbor for cross sections of line drawings), or to perform edge-fitting and sharpening on the interpolated image (e.g., for the calibration situation described above). It would also be nice to extend the algorithm to deal with unequally spaced slices; this would require a more general (and significantly slower) implementation of cubic interpolation. However, our algorithm serves its principal purpose very well, giving good results on a variety of real data.

Acknowledgment

References

Ballinger, Ray. 1997. Gainesville VAMC MRI Teaching File. `http://www.xray.ufl.edu/~rball/teach/mriteach.html` .

——— . 1998. The MRI Tutor. `http://128.227.164.224/mritutor/index.html` .

Hornak, Joseph. 1997. The Basics of MRI. `http://www.cis.rit.edu/htbooks/mri/` .

Johnson, Keith, and Alex Becker. 1997. The Whole Brain Atlas. `http://www.med.harvard.edu/AANLIB/home.html` .

Mahan, Steven L. 1996. Resolution enhancement. `http://aurora.phys.utk.edu/~mahan/enhancement.html` .

Makivic, Miloje. 1996. Bicubic interpolation. `http://www.npac.syr.edu/projects/nasa/MILOJE/final/node36.html` .

Mnuk, Michal. 1997. Lagrange interpolation (in German). `http://www.risc.uni-linz.ac.at/people/mmnuk/FHS/MTD/MAT2/Skriptum/K5node3.html`.

Rodriguez, Paul. 1995. MRI Indications for the Referring Physician. `http://www.gcnet.com/maven/aurora/mri/toc.html` .

MRI Slice Picturing

Ni Jiang
Chen Jun
Li Ling
Tsinghua University
Beijing, China

Advisor: Ye Jun

Summary

We set up two coordinate systems, one in the object space and the other on the computer screen. We introduce six parameters to describe the slice plane, and we formulate the coordinate mapping from the screen to the object space.

We designed six alternative algorithms that use the given data to estimate the density at any location in space and produce a slice of a three-dimensional array. Some of the algorithms exploit global information and some are self-adaptive; all but one have advantages in certain circumstances.

We extended a well-known two-dimensional model of a human head model to build a three-dimensional model of a head, consisting of 10 ellipsoids of different size, orientation, and density. We produced the data sets by sampling in the object space (the head model) at evenly spaced intervals; the dimension of the data set is $128 \times 128 \times 128$.

We devised several test slices to test our model and algorithms. Some test slices have a complex shape, some are critical in their position, and some are really disasters to most algorithms. We also tried different sampling intervals to verify our ideas about the model.

Based on subjective and objective comparisons, we summarize the strengths and weaknesses of the algorithms. For common use, we suggest the gradient algorithm and our GNP-integrated algorithm. In most cases, both can render well slices with both sharp and smooth edges.

Facts about MRI

MRI has several features relevant to the problem:

关注数学模型
获取更多资讯

- **High precision.** The scanning precision of MRI is about 1–3 mm. That is, MRI can easily distinguish nuances at the size of 1–3 mm. Commonly used MRI slices are no larger than 25 cm × 25 cm [Gao 1996].

- **High contrast.** One of the advantages of MRI is the high contrast of its images, which makes the boundaries of the organs sharp enough for diagnosis [Frommhold and Otto 1985, Gao 1996].

- **Long performance time.** The performance time of MRI is several minutes. For example, a typical scanning of a two-dimensional image ($128 \times 128 \times 256$) with pulse repeat time $T_R = 1.5$ s needs about 6 min [Gao 1996]. The time required is still one of the main drawbacks of MRI. Thus, we cannot expect that the given data set will be thorough enough to produce a good slice picture (that may require too much time). Our algorithms should not be too complex or time consuming.

- **Reconstruction Algorithms.** Two commonly used methods to reconstruct the three-dimensional information from the raw data produced by MRI are Projection Reconstruction (PR) and Fourier Transformation. They require that the data be evenly sampled through space, so we assume that.

Assumptions, Coordinates, and Notations

Assumptions

From the request of the problem and some facts of MRI, we take the following assumptions:

- The dimension of the examined object is 256 mm × 256 mm × 256 mm, if not smaller; this is big enough in most cases. If a larger object is scanned, we can divide up the data into several cubes.

- The desired precision of pictured slices is 1 mm. We will picture the slices produced by our algorithms on the computer screen, using one pixel to present an area of 1 mm × 1 mm.

- The given data set is a three-dimensional array $A(i, j, k)$ sampled in the whole object space with evenly spaced intervals along the coordinate axes. Such intervals are about 2–4 mm and are big enough for MRI to scan in not too long a time. Later we discuss the case of the data not being evenly spaced.

- $A(i, j, k)$ takes an integer value from 0 through 255, indicating the water density, from high density to low density. On our screen, 0 is represented by black and 255 is represented by white.

- The examined object consists of several different components. We assume that the density does not change much within one component.

- The object is the body of some animal or of a human being. Since the organs or tissues in such body are likely to be tender, we can image that the boundaries are smooth and sharp in most cases. Exceptions will happen only between some kind of bones, such as the backbone (they are sharp but not smooth) or some sick tissues.

- The unknown density of a location is affected by all the given data. However, the distance between points plays an important role in this problem. Locations far away (for instance, 50 mm) from the unknown point are assumed to have little or no effect.

Coordinate Systems

We set up two coordinate systems, one in the data (or object) space and one on the computer screen, as presented in **Figures 1–2**. The units (pixel in the screen image) in these two systems are both 1 mm, for the sake of convenience. Since the object is of 256 mm × 256 mm × 256 mm, the data space is just $0 \le x, y, z < 256$.
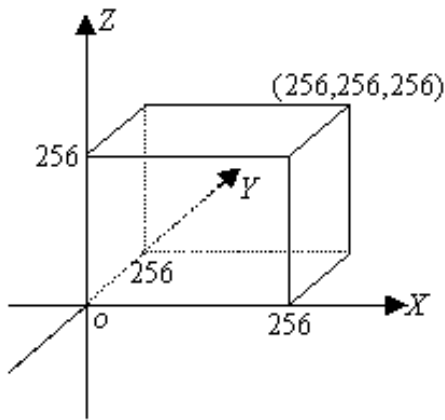
Figure 1. The data space coordinate system.

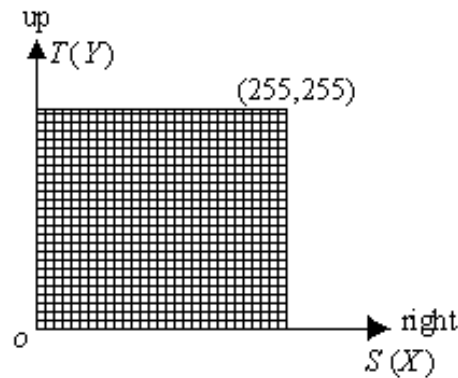**Figure 2.** The screen image coordinate system. The origin O is the left bottom corner of the screen image.

Notation

"Density" indicates the water concentration in a small region of the scanned object at some location. The phrase "unknown point" or "unknown location" means the point (or the location) where the density of the object is not given as known data, hence needs to be calculated.

The symbols commonly used in this paper are:

$A(i, j, k)$	The given three-dimensional data indicating the density of a location. In some contexts, A also represents the location.
s_X, s_Y, s_Z	The three sampling intervals along the Cartesian axes. Thus, $A(i, j, k)$ is the density of location $(i \cdot s_X, j \cdot s_Y, k \cdot s_Z)$.
$\alpha, \beta, \gamma, x_0, y_0, z_0$	The six parameters to define a slice plane.
$D(x, y, z)$	The density of the object at location (x, y, z).

Analysis of the Problem

For a plane slicing the object, we want to know the density of the object throughout the plane. If we can convert the coordinates of the points in the slice plane to the real 3-D coordinates in the object, and calculate the corresponding densities, the problem is solved. The first step is simple, with some knowledge of the space geometry. But how about the second step?

Can the Unknown Density Be Known?

From the famous Nyquist sampling theorem, we know that to reconstruct the whole density information of the scanned object *exactly*, the sampling intervals must satisfy the inequality

$$\max(s_X, s_Y, s_Z) \leq \frac{1}{2f_m}, \tag{1}$$

where f_m is the upper limit of the spatial frequency of the density. In our problem, **(1)** would need to be satisfied if a slice is required to be pictured exactly; but we don't need to do that.

On the one hand, the inequality could never be satisfied, since the f_m in an object is always very large—infinity, in reality. No sampling intervals can satisfy such an inequality! On the other hand, to picture the slice we do not need to know *exactly* what the unknown is. Since the grayscale is from 0 to 255, an error of less than 1 grayscale unit is acceptable. In fact, a blur to some extent is always allowable and unavoidable. In this sense, we can know the unknown density.

How to Know the Unknown Density?

Since we cannot know the unknown density exactly, we must estimate it. We can choose from:

- **Simplicity and Complexity.**   Our goal is to find an effective but simple algorithm to produce any slice of the object. We also believe that the real object is too complex to describe or estimate by only one kind of algorithm. So our motto is "If it works, it's good enough," and we tried to find several different algorithms to deal with the different aspects of the real object.

- **Local and Global Information.** Global information is alluring but very difficult to use. As human beings, we can easily locate a vessel or a bone in an MRI image and outline it, using our global impression (thus we can do reconstruction). But it is difficult for computers to know a shape rather than a number. Current algorithms can outline an image, but the information used by computers is local (e.g., the difference between adjacent pixels) but not global. So we base our main idea on local information but remain alert to global information. Our experiments show that appropriately using even a little global information brings great benefit.

- **Static and Self-adaptive Algorithm.** There are several advantages to a static algorithm: it is fast and simple (in most cases), it is often designed by aiming at some aspect of the real application and may be effective in that aspect, and it is easily controlled and safe. Similar to global information, a self-adaptive algorithm is powerful but difficult to control.

Description of Our Model

Our model consists of three parts:

- the given data,

- the description of the slice plane, and

- the algorithm to estimate the density of the object at any location, whether or not this location is included in the given data.

The description of the slice plane is used to convert the coordinates of a point on the screen to coordinates in space, while the density-estimating algorithm obtains the density of the point from the known data. Thus, the slice can be easily displayed on the screen.

We propose six different algorithms to estimate the density, discussed in detail in the next section. The given data are already described in the problem and the subsection on **Assumptions**, so here we treat the mapping from the screen to space.

Slice Plane

In order to define a slice plane at any orientation and any location in space, we perform four steps to transform from the XY-plane to any other plane in the space:

1. Put a plane, say P, with its own coordinate system ST (the same as the screen coordinate system), onto the XY-plane and make the two coordinates exactly the same origin and orientation.

2. Rotate P around its normal line (i.e., the z-axis) by angle α to make the orientation of ST differ from XY.

3. Rotate the normal line of P around the origin to a prescribed orientation. In **Figure 3**, this orientation is defined by angles β and γ.
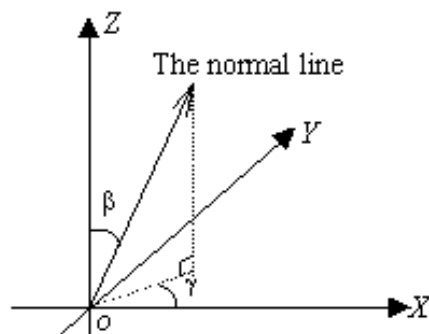

Figure 3. Rotate the normal line to a prescribed orientation.

4. Perform a translation to the plane P, moving the origin of P to some predefined point in the space, say (x_0, y_0, z_0).

Thus, using six parameters $(\alpha, \beta, \gamma, x_0, y_0, z_0)$, we can define a plane anywhere, with a coordinate system the same as the screen system.

Mapping from the Screen to the Space

Since the coordinate systems of the screen and of the slice plane are the same, we can change the screen coordinate to the slice plane and then use the the transformation in the last subsection to convert the slice plane to space coordinates.

Suppose a pixel in the screen is at position (s, t) and the corresponding point in the space is at (x, y, z). From the transformation, we get the mapping equation from the screen to the space and thereby solve the first step of the problem:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s \\ t \\ 0 \end{pmatrix}$$
$$+ \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}.$$

Density-Estimating Algorithms

Consider a pixel on the screen and the corresponding point U at location (x, y, z) in the object space. The task of the density-estimating algorithm is to estimate the density at U, which we denote by $D(x, y, z)$.

We tried five basic types of density-estimating algorithms. Based on experimental results, we designed an all-around effective method, which we call *GNP-integrated*.

Trilinear Interpolation

In general, linear interpolation can produce satisfactory results. In three-dimensional space, we use trilinear interpolation, which interpolates from the eight neighbors in three directions. That is,

$$
\begin{aligned}
D(x, y, z) = {} & A(i, j, k) \cdot (1 - u) \cdot (1 - v) \cdot (1 - w) \\
& + A(i + 1, j, k) \cdot u \cdot (1 - v) \cdot (1 - w) + A(i, j + 1, k) \cdot (1 - u) \cdot v \cdot (1 - w) \\
& + A(i, j, k + 1) \cdot (1 - u) \cdot (1 - v) \cdot w + A(i + 1, j + 1, k) \cdot u \cdot v \cdot (1 - w) \\
& + A(i + 1, j, k + 1) \cdot u \cdot (1 - v) \cdot w + A(i, j + 1, k + 1) \cdot (1 - u) \cdot v \cdot w \\
& \qquad\qquad\qquad\qquad\qquad + A(i + 1, j + 1, k + 1) \cdot u \cdot v \cdot w, \quad \textbf{(2)}
\end{aligned}
$$

with

$$
i = \left\lfloor \frac{x}{s_X} \right\rfloor, \; j = \left\lfloor \frac{y}{s_Y} \right\rfloor, \; k = \left\lfloor \frac{z}{s_Z} \right\rfloor,
$$
$$
u = \frac{x}{s_X} - i, \; v = \frac{y}{s_Y} - j, \; w = \frac{z}{s_Z} - k,
$$

where $\lfloor x \rfloor$ is the largest integer no larger than x.

This method uses the density values of eight neighbors to resolve the density at U. Discontinuity at boundaries can be mitigated by this approach in nearly all cases. However, this method tends to blur some sharp edges, because of its intrinsic low-pass filtering attribute.

Nearest-Neighbor

With the preservation of edge sharpness in mind, we tried the nearest-neighbor method, which assigns to U the density of its nearest neighbor in space.

This method is fairly simple and the computational load is very low. The effect produced by the method is quite unstable, although sometimes it really gives good results. Nevertheless, it partly preserves edge sharpness, and its power can be amplified if properly combined with other methods.

Median

The idea comes from the median filtering that is famous in signal and image processing. Median filtering can preserve the sharp edge of the signal from great damage while smoothing the signal. In our algorithm, we assign the median of the density of U's eight neighbors to U. The result of this algorithm, as expected, gives sharp edges but has obvious feather-out, which results in an unrealistic contour.

Power-Control

Since we believe that the distance between points is very important, we can conceive that each point within a reasonable distance from U has a "power" to control the density of U, forcing the density of U to be similar to its own, and that such power decreases with distance. The overall result should be the average of the densities of those points, taking their power into account.

We define the power of $A(i, j, k)$ over distance d as:

$$p = \frac{1}{1 + e^{5(d/d_0 - 1)}},$$

where $d = \sqrt{(x - i \cdot s_X)^2 + (y - j \cdot s_Y)^2 + (z - k \cdot s_Z)^2}$ and d_0 is a distance threshold (when $d = d_0$, the power is $\frac{1}{2}$). Then the density of U is estimated as

$$D(x, y, z) = \frac{\displaystyle\sum_{d_\xi \leq 2d_0} p_\xi \cdot A(i_\xi, j_\xi, k_\xi)}{\displaystyle\sum_{d_\xi \leq 2d_0} p_\xi}, \tag{3}$$

with the summation over all the known points within a distance $2d_0$. We use $d_0 = 1$ mm when the sampling interval is 2 mm.

Though **(3)** has some similarity to trilinear interpolation **(2)**, the nonlinearity in the definition of power makes the edge produced by this algorithm smoother—but also more blurred.

Here we could adopt another type of power, called the *optimal interpolation function*:

$$p = \text{sinc}(\pi d) = \frac{\sin(\pi d)}{\pi d}, \text{ where } d = \sqrt{\left(\frac{x}{s_X} - i\right)^2 + \left(\frac{y}{s_Y} - j\right)^2 + \left(\frac{z}{s_Z} - k\right)^2}.$$

This kind of power is famous, for it is an ideal low-pass filter in the frequency field, and it is used to reconstruct the original signal (from frequency information) in the sampling theorem. But since f_m is very large in this problem, we cannot expect that such a power will do a good job.

Gradient

The methods above are all based on the effect of one point on another. If the effect of a point-pair to one unknown point is considered, we can introduce the gradient method.
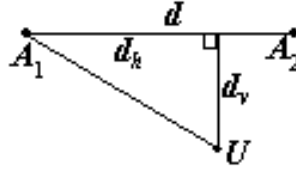

Figure 4. Gradient in a point-pair.

Figure 4 shows two given data values $A_1(i_1, j_1, k_1)$, $A_2(i_2, j_2, k_2)$ and the unknown point U. The distance between A_1 and A_2 is d, the projection of $\overrightarrow{A_1U}$ on $\overrightarrow{A_1A_2}$ is d_h (which is negative when the angle between $\overrightarrow{A_1U}$ and $\overrightarrow{A_1A_2}$ is an obtuse angle), and d_v is the distance from U to $\overrightarrow{A_1A_2}$. The density of point U, if only estimated by the gradient from A_1 to A_2, is

$$D(x, y, z) = A_1 + \frac{d_h}{d}(A_2 - A_1).$$

However, when other data-pairs in the neighborhood of point U are considered, the density D is a weighted average of all the effects, and the weight (similar to the "power") is defined as:

$$p = \begin{cases} e^{-d_v}, & \text{when } d_h \geq 0; \\ \frac{1}{4}e^{-d_v}, & \text{when } d_h < 0. \end{cases}$$

This algorithm exploits not only the density information around the unknown point U but also the tendency of the density in a local volume. This makes it self-adaptive to some extent. Further, we can add some global information to the algorithm. For example, in our implementation, when A_1 and A_2 are close enough ($|A_1 - A_2| < 20$), we multiply the weight p by 3; in such a case, A_1 and A_2 are deemed to be in the same component, which makes the probability that U is also in the same component very large. Similarly, when $|A_1 - A_2| > 80$, we multiply the weight p by 0.7, since A_1 and A_2 may be in different components.

GNP-Integrated

From experimental results (see the next section), we found out that the gradient and power-control methods are good at making smooth but slightly blurred edges, while the nearest-neighbor method always gives a high-contrast image with rough edges. In an attempt to combine their advantages, we integrated these three methods into one algorithm that we call *GNP-integrated*. It

can be described in brief as the combination of the gradient, nearest-neighbor, and power-control methods in the proportions $3 : 2 : 1$.

Test of the Algorithms

Data Sets: The Head Model

Suitable data sets are necessary for testing and demonstrating the algorithms, as well as for comparing different algorithms. Real MRI data would be best. However, besides the inconvenience involved in getting such data, another annoying problem is that we would have great difficulty comparing the pictured slices and the actual slices, since we actually can't have the latter!

Motivated by the widely accepted two-dimensional Sheep–Logan (S-L) head model [Gao 1996], where ten ellipses, different in location, shape, orientation, and intensity, constitute an object representing a head section, we designed a 3-D head model made up of ten ellipsoids different in location, shape, orientation, and density. The empty space inside the head model is filled with an ambient color that differs from that of the background outside the model.

We adopt ellipsoids for our data model because of their simplicity and because the combination of varied ellipsoids can imitate many real objects, such as a brain or a stomach. We designed three types of ellipsoids with different density distributions:

- Type 1: Uniform density.

- Type 2 : The density changes linearly from the center to the surface.

- Type 3: The same as type 2 but with additive random noise of a specific standard deviation. (In our experiments, we don't analyze this type, since noise filtering is beyond our concern in this paper.)

With the sampling intervals specified, data sets can be produced easily by determining in which ellipsoid a sample point lies. Such data sets are large; for example, when the intervals are all 2 mm, the data set is $128 \times 128 \times 128 = 2$ MB.

In addition, we can also compute the actual slice with our head model. The computation process is similar to the data set producing process.

Experiments and Results

An important issue is how to test the output of different algorithms. The two main objectives of this problem, maintaining the sharpness of the edges and the smoothness of the contours, are difficult to measure numerically, so we made comparisons by visual inspection (subjective as it is).

At the same time, although the RMS (root mean square) error cannot comprehensively and rationally reflect the quality of a pictured slice, it is still helpful to the assessment of a pictured slice. So we take the minimization of the RMS error as our third goal. (This makes sense only for our simulated data, since the actual slice is unknowable in the real world.)

We also must take the computational loads into consideration, since the data set is comparatively large.

We did a number of comparisons, from which we present some representative slices. Typical slices are presented in **Figures 5–8**. We examine each in detail and then draw some conclusions. (In all cases, $s_X = s_Y = s_Z = 2$.)

In **Figure 5**, the slice is in the middle of the scanned object and parallel to the XZ-plane. In this case, the slice traverses all the ten ellipsoids, so the overall performance of each method is easy to evaluate. In **Figure 6**, the slice is oblique and all algorithms work well, except for the power-control using the sinc function. In **Figure 7**, the slice in **Figure 6** is translated by just a tiny distance, but the performances of some algorithms fall dramatically. In **Figure 8**, the slice plane is at an odd angle and in a critical position, which gives our algorithms a chance to show their performance in an awful situation.

Assessment of the Algorithms

Except for the power-control method (with the sinc function), which is clearly unsuited to this problem, each method has advantages and disadvantages.
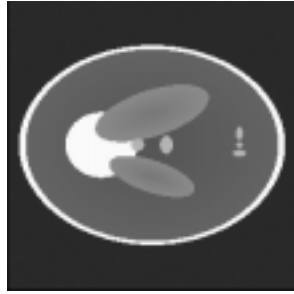
Trilinear

The trilinear method works well in common cases. It usually has a small RMS error and takes a short time. But it has a tendency to blur the picture, and it disappointed us in the awful situation (see **Figure 8b**). When the contrast of different components in the scanned object is low, this method is not recommended.
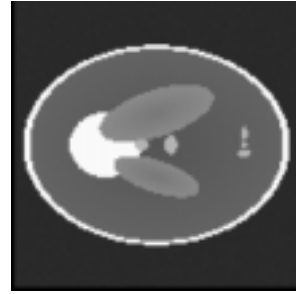
Nearest-Neighbor and Median

Both the nearest-neighbor method and the median method preserve a sharp edge and take the least time. But they lose the smoothness of the contours and cannot discriminate small objects (see **Figures 6cd**). A small translation of the slice plane also causes them to produce many more zigzag contours (see **Figure 7cd**). Consequently, they have big RMS errors. In spite of these weaknesses, when the data set is very large and the time element is more important, or when the zigzag has little bad effect on the result, these methods are desirable.

a. Actual slice.

b. Trilinear interpolation (14.5).

c. Nearest-neighbor (20.0).
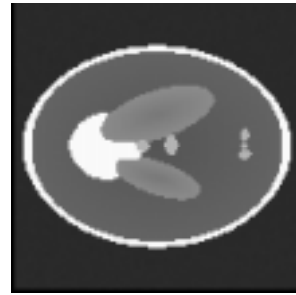
d. Median (23.9).
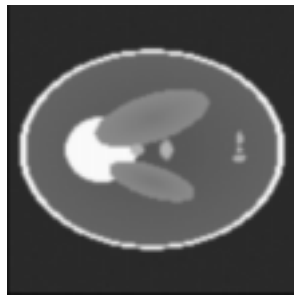
e. Power-control (15.3).
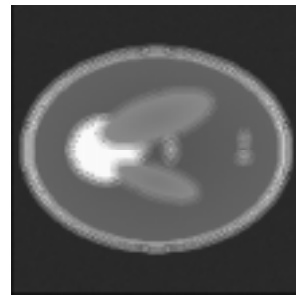
f. Power-control (sinc) (29.3).
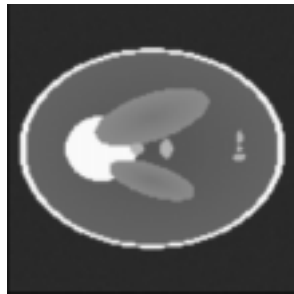
g. Gradient (12.9).
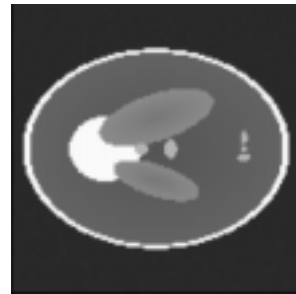
h. GNP-integrated (14.4).

Figure 5. A slice in the middle of the object and parallel to the XZ-plane, with parameters $x_0 = 0$, $y_0 = 128$, $z_0 = 0$, $\alpha = 0$, $\beta = 90°$, $\gamma = 90°$. The number following each algorithm name is the RMS error.
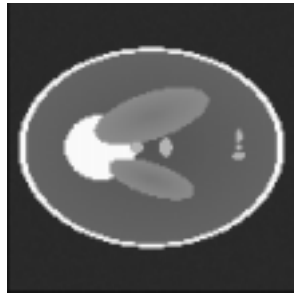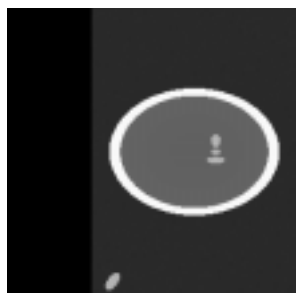
a. Actual slice.
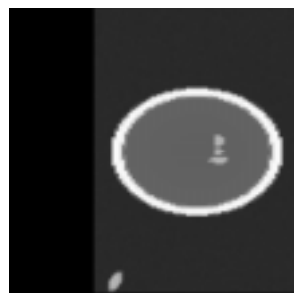
b. Trilinear interpolation (12.3).

c. Nearest-neighbor (15.5).

d. Median (18.7).

e. Power-control (13.7).
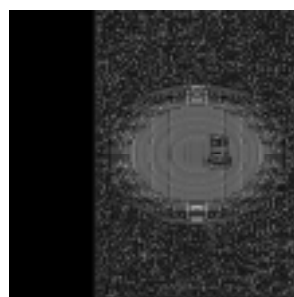
f. Power-control (sinc) (55.8).
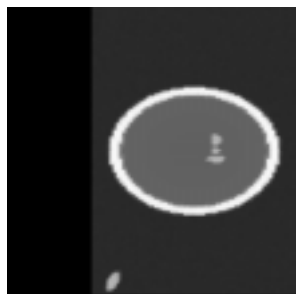
g. Gradient (11.3).
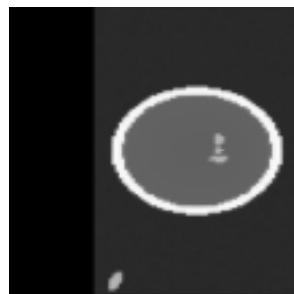
h. GNP-integrated (12.0).

Figure 6. An oblique slice, with parameters $x_0 = 0$, $y_0 = 128$, $z_0 = 0$, $\alpha = 0$, $\beta = 45°$, $\gamma = 90°$. The number following each algorithm name is the RMS error. The black area on the left of each slice is outside the scanned object.
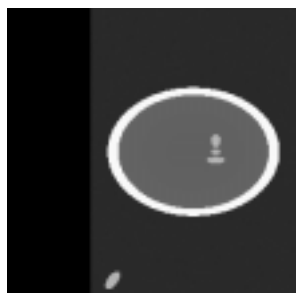
a. Actual slice.

b. Trilinear interpolation (13.0).
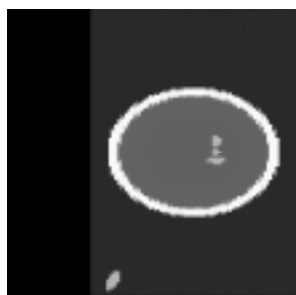
c. Nearest-neighbor (18.8).

d. Median (20.6).

e. Power-control (13.8).
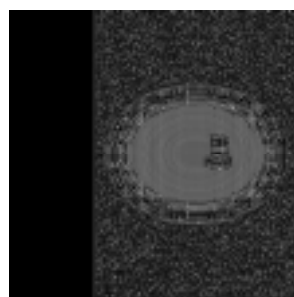
f. Power-control (sinc) (55.0).
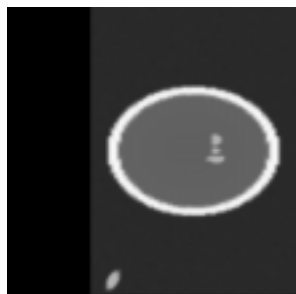
g. Gradient (12.3).
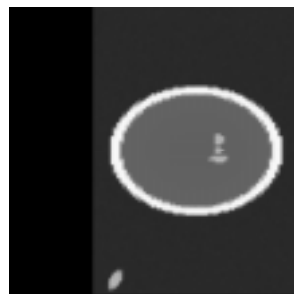
h. GNP-integrated (13.9).

Figure 7. The oblique slice of **Figure 6** translated a tiny distance (one unit in the y-direction), with parameters $x_0 = 0$, $y_0 = 129$, $z_0 = 0$, $\alpha = 0$, $\beta = 45°$, $\gamma = 90°$. The number following each algorithm name is the RMS error.
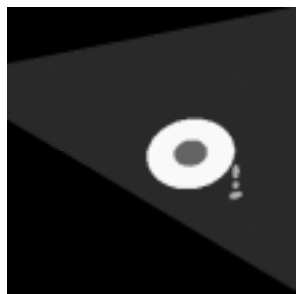
a. Actual slice.
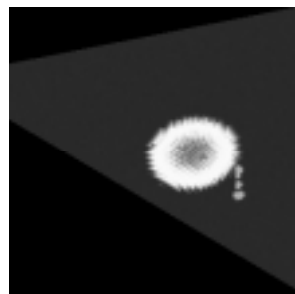
b. Trilinear interpolation (12.3).
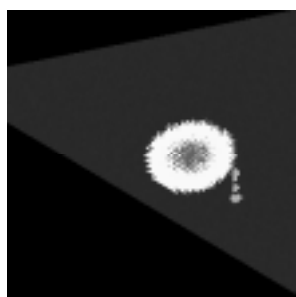
c. Nearest-neighbor (17.5).
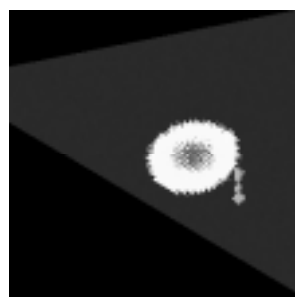
d. Median (18.0).
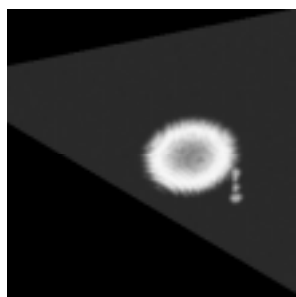
e. Power-control (13.8).
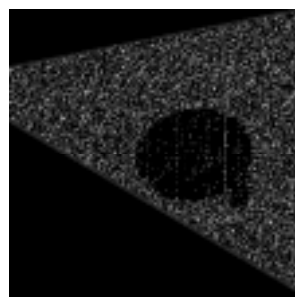
f. Power-control (sinc) (68.0).
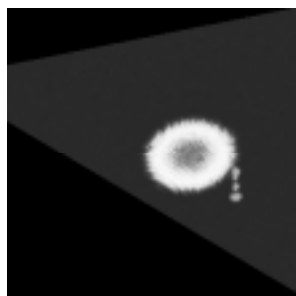
g. Gradient (12.2).
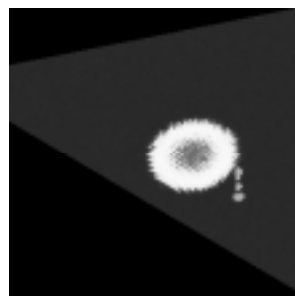
h. GNP-integrated (13.4).

Figure 8. An oblique slice at an odd angle and in a critical position, with parameters $x_0 = 0$, $y_0 = 126$, $z_0 = 0$, $\alpha = 0$, $\beta = 70°$, $\gamma = 60°$. The number following each algorithm name is the RMS error.

Gradient

The gradient method has the amazing advantage that it has the minimal RMS error in all cases. More important, the slice as pictured by this method has rather satisfactory smoothness *and* sharpness, as is obvious in any of the figures.

GNP-Integrated

The GNP-integrated method has RMS error a little larger than that of the trilinear method; however, it excels over any other algorithm when the sharpness and smoothness are taken into account.

Conclusion

The gradient and GNP-integrated algorithms are the most competent if the runtime (10–14 s on a Pentium 166 for our implementation) is not a serious consideration (the other algorithms take 2–3 s). They are especially powerful in awful situations when some critical oblique slice is desired.

What Happens When the Interval Is Too Large?

To verify our discussion on the sampling intervals (see the section **Can the Unknown Density Be Known?**), we also tested the result when $s_X = s_Y = s_Z = 4$. As expected, the quality of the produced slices deteriorated. For example, some connected thin boundaries in **Figure 9** are broken because of insufficiency of the data.

Figure 9. The slice of **Figure 5a**, with sampling interval 4 mm, as rendered by the gradient method; compare with **Figures 5a** and **5g**.

If a data set is not sampled at evenly spaced intervals, or if the data are too scattered, the user should first use simple interpolation to construct a data set with evenly spaced sampling intervals.

Strengths and Weaknesses

- We present several good algorithms that can be selected by the user to fit different situations.

- We present a clear assessment of different algorithms, based on experimentation on simulated data for a head.

- We implemented all of the algorithms in a Windows 95 user-oriented computer simulation with easy input, suitable for repeated experimental research.

- We tried to find objective measurements of sharpness and smoothness but time did not permit.

References

Frommhold, H., and R.Ch. Otto. 1985. *New Methods of Medicine Imaging and Their Application* (German). 1988. Chinese translation by Wang Zhen and Gu Ying. Beijing, China: Medician Technology Press.

Gao, S.K. 1996. Imaging System in Medicine. Beijing, China: Dept. of Electrical Engineering, Tsinghua University.

Judge's Commentary:
The Outstanding Scanner Papers

William P. Fox
Dept. of Mathemats
Francis Marion University
Florence, SC 29501
`wfox@fmarion.edu`

Each of the participating schools is to be commended for its fine effort. The judges did not witness a wide range of mathematical modeling by the participants to obtain their solutions. Most teams recognized this problem as an image processing problem.

According to the problem statement, the current family of MRIs slice a three-dimensional scanned image, vertically or horizontally. One component of the problem required teams to obtain an oblique slice. Teams used one of three basic methods to obtain their oblique slice:

- The method seen most often was to create a plane, $Ax + By + Cz = D$, and then rotate it using a standard matrix transformation.

- Selecting two points in three-space and defining a plane between those points.

- Selecting one point and two angles to define their plane.

Teams realized that a critical element was mapping the coordinates of their oblique plane through their three-dimensional data set in order to obtain a gray scale scheme (0–255) for the elements in the plane. The three-dimensional data set was defined by three integers, while the points in the oblique plane were real numbers. Methods had to be developed to interpolate the gray scale values for all the points in the oblique plane.

Methods chosen by the teams included:

- a nearest neighbor algorithm, using eight or more points;

- a weighted point algorithm;

- splines (linear through cubic); and

- Lagrangian polynomials.

Teams usually tried more than one method to obtain the grayscale values. Comparisons of methodologies were generally sketchy and lacked analysis. Teams that critically compared and analyzed their methodology and results, and reached valid conclusions, impressed the judges.

The problem statement required teams to design and test an algorithm to produce sections of three-dimensional arrays made by planes in any orientation in space, preserving as closely as possible the original grayscale values. The use of grayscale was a distinguishable characteristic. Some teams used color to enhance their presentation; this was acceptable provided that the grayscale was not totally removed. Several teams suggested color because their grayscale resolution could not detect certain diagnostic elements; this was viewed as a fatal flaw, since the problem statement required the use of grayscale.

The judges felt that to distinguish teams better, they would focus more closely on the MUST and SHOULD requirements of the problem statement.

- The team's algorithm MUST produce a picture of the slice of the three-dimensional array by a plane in space. This became one critical element for teams to move beyond the Successful Participant category. Judges wanted to see a picture, not a matrix portrayal. Pictures were closely scrutinized to see if they appeared to be oblique slices.

- The teams SHOULD:

  - Design data sets to test and demonstrate their algorithm.
  - Produce data sets that reflect conditions likely to be of diagnostic value.
  - Characterize data sets that limit the effectiveness of their algorithm.

Thus, judges looked for a good description of the data sets chosen and a description of the elements of diagnostic value. Verbal descriptions stating teams were looking for tumors or anomalies in body parts were acceptable. Teams also created spheres inside cubes as their representative data set. Provided teams put something of diagnostic value inside their lager 3-D elements, their data sets were still acceptable.

The characterization of data sets that limited the effectiveness of the team's algorithm was the most avoided "SHOULD" requirement. A verbal description of any data set content that limited effectiveness was acceptable by the judges in order to separate the top quality papers.

Another important element not uniformly accomplished was some kind of error analysis. Very few teams even checked for accuracy their integer values in the plane against the corresponding integer point in their three-space data set. The judges praised those teams that accomplished that. Almost every team referred to their pictures—"outputs"—to explain or attempt to show accuracy. Teams, as their only basis for analysis, used the "blurry versus sharp" edges.

Style and clarity of presentation was viewed as another critical element. Teams' organization and ability to explain their methodologies separated participants. Good organization and a solid layout helped distinguish teams.

The judges who evaluated this problem were impressed by the quality and completeness of the solutions presented. We were amazed at how much work was accomplished during that weekend.

About the Author

William P. Fox is Chairman of the Dept. of Mathematics and Professor of Mathematics at Francis Marion University. He received his M.S. in operations research from the Naval Postgraduate School in 1982 and earned his Ph.D. from Clemson University in 1990. He has served as a judge and as the associate contest director of the MCM. Bill will be the contest director for the new High School Mathematical Contest in Modeling under a grant through COMAP.

Proposer's Commentary:
The Outstanding Scanner Papers

Yves Nievergelt
Department of Mathematics, MS 32
Eastern Washington University
526 5th Street
Cheney, WA 99004–2431
`yves.nievergelt@mail.ewu.edu`

Once again, the problem came from the laboratory of Dr. Mark F. Dubach, who is studying the effects of intracerebral drug injections on monkeys with brain diseases at the University of Washington's Regional Primate Research Center in Seattle, WA.

This year, the striking novelty about the winning solutions of this purely mathematical modeling problem is the student teams' mastery of several electronic tools, which they used very adeptly along with mathematics.

- The first such electronic tool is the World Wide Web, which teams used in various degrees to find general medical information about Magnetic Resonance Imaging, real and simulated three-dimensional data sets for the human brain, and mathematical algorithms for two-dimensional interpolation. Two teams, however, found all the information they needed in printed form, and then adroitly generated their own test data.

- The second electronic tool lies in computer graphics, which all teams employed efficiently to communicate their results. For this problem, as one team noted, there does not seem to exist any numerical estimate of performance—such as a root-mean square or any other norm—that can substitute for the final visual medical diagnostic, and hence graphics may remain the best way to compare algorithms to reality.

- The third electronic tool consists of computer programming, which the teams utilized for the change of coordinates, in effect an isometric parametrization of a plane in space, and for three-dimensional interpolation.

- The fourth electronic tool, used appropriately by all teams, is the preparation of a final document containing prose, mathematical formulae, and graphics.

All these tools helped, of course, with the essential part of the problem, namely, mathematics. Within mathematics, the teams demonstrated a good command of concepts and details. As a first example, one crucial place for concepts is at the start, where all teams realized that the practical problem could be cast as a mathematical problem in three-dimensional interpolation. As a second example, one place where detail became important is in the generalization from one- or two-dimensional to three-dimensional interpolation. While one team (Tsinghua University) already knew the result, other teams (Eastern Oregon University, Harvey Mudd College) offered excellent explanations and proofs of their mathematical generalizations.

Finally, all teams demonstrated an efficient use of their time in balancing time devoted to searches and time devoted to in-house production for such items as data and algorithms. Such a balancing act between finding and reinventing the wheel can be critical in practice to deliver a working computer program in time. For example, none of the teams appears to have used a three-dimensional interpolation computer program from the World Wide Web, perhaps because it is not obvious where to get one. Indeed, a search of Netlib at `http://netlib2.cs.utk.edu` for "three-dimensional interpolation" shows such one- and two-dimensional routines as `toms/474` (bicubic interpolation) but does not reveal any specifically three-dimensional routines. Nevertheless, such routines exist, but finding them and using them may demand more time than available. For instance, there is a multidimensional (with an unlimited number of dimensions) interpolation routine using nonuniform rational B-splines (NURBS) at `http://dtnet33-199.dt.navy.mil/dtnurbs/about.htm` .

About the Author

Yves Nievergelt graduated in mathematics from the École Polytechnique Fédérale de Lausanne (Switzerland) in 1976, with concentrations in functional and numerical analysis of PDEs. He obtained a Ph.D. from the University of Washington in 1984, with a dissertation in several complex variables under the guidance of James R. King. He now teaches complex and numerical analysis at Eastern Washington University.

Prof. Nievergelt is an associate editor of *The UMAP Journal*. He is the author of many UMAP Modules, a bibliography of case studies of applications of lower-division mathematics (*The UMAP Journal* 6 (2) (1985): 37–56), and *Mathematics in Business Administration* (Irwin, 1989).

Alternatives to the Grade Point Average for Ranking Students

Jeffrey A. Mermin
W. Garrett Mitchener
John A. Thacker
Duke University
Durham, NC 27708–0320
`wgm2@acpub.duke.edu`

Advisor: Greg Lawler

Introduction

The customary ranking of students by grade point average (GPA) encourages students to take easy courses, thereby contributing to grade inflation. Furthermore, many ties occur, especially when most grades are high. We consider several alternatives to the *plain GPA ranking* that attempt to eliminate these problems while ranking students sensibly. Each is based on computing a revised GPA, called an *ability score*, for each student. We evaluate these alternative methods within the context of the fictitious ABC College, where grades are inflated to the extreme that the average grade is A−.

- The *standardized GPA* replaces each grade by the number of standard deviations above or below the course mean. Students are then ordered by the average of their revised grades.

- The *iterated adjusted GPA* compares the average grade given in a course to the average GPA of students taking it, thereby estimating how difficult the course is. It repeatedly adjusts the grades until average grade equals the average GPA and uses the corrected GPA to determine rank.

- The *least-squares method* assumes that the difference between two students' grades in a course is equal to the difference between their ability scores. It then sets up a large matrix of linear equations, with an optional handicap for courses taken outside a student's major, and solves for the ability scores with a least-squares algorithm.

关注数学模型

获取更多资讯

An acceptable ranking method must reward students for scoring well, while taking into account the relative difficulties of their courses. It must clearly distinguish the top 10% of students. Preferably, the method should make allowances for the fact that students often earn lower grades in courses outside their majors and should not discourage them from taking such courses.

We used a small simulated student body to explore how the different methods work and to test the effects of changing a single grade. The least-squares method gave the most intuitive and stable results, followed by the iterated adjusted, the standardized, and finally the plain GPA. Under the least-squares and iterated adjusted methods, when a certain student's grade was changed in one course, that student and other students in that course changed position but most of the other students moved very little.

We used a larger simulated student body, generated by a computer program, to compare the iterated adjusted and standardized algorithms. They agree on most of students in the top decile, around 89% if plus and minus grades are included. They did not agree well with the plain GPA ranking, due to massive ties in the latter.

All four methods are more reliable when plus and minus grades are included, since a great deal of information is lost if only letter grades are given.

We recommend the least-squares method, since it is not very sensitive to small changes in grades and yields intuitive results. It can also be adapted to encourage well-roundedness of students, if the college chooses.

However, if there are more than about 6,000 students, the least-squares method can be prohibitively difficult to compute. In that case, we recommend the iterated adjusted GPA, which is easier to calculate and is the best of the remaining methods.

We recommend against the standardized GPA, because it does not properly correct for course difficulty, makes assumptions that are inappropriate for small or specialized courses, and produces counterintuitive results. We also recommend against the plain GPA, because it assumes that all courses are graded on the same scale and results in too many ties when grades are inflated.

To avoid confusion, we use the following terminology: A *class* is a group of students who all graduate at the same time, for example, the class of 1999. A *course* is a group of students being instructed by a professor, who assigns a grade to each student.

Assumptions and Hypotheses

- It is possible to assign a single number, or "ability score" (this will be the revised GPA), to each student, which indicates the student's relative scholastic ability and, in particular, the student's worthiness for the scholarship. In other words, we can rank students.

- The rank should be transitive; that is, if X is ranked higher than Y, and Y is ranked higher than Z, then X should be ranked higher than Z. We can

therefore completely order students by rank.

- The performances of an individual student in all courses are positively correlated, since:

    - There is a degree of general aptitude corresponding to the ability score that every student possesses.

    - All instructors, while their grade averages may differ, rank students within their courses according to similar criteria.

- While there may be a difference between grades in courses that reflects the student's aptitude for the particular subjects, this has only a small effect, because:

    - Students select electives in a manner highly influenced by their skill at the subjects available, that is, students tend to select courses at which they are most talented.

    - All students should major in an area of expertise, so that they are most talented at courses within or closely related to their majors.

    - The college may require courses that reflect its emphasis; even if the required courses could be considered "unfair" because they are weighted towards one subject (e.g., writing), that is the college's choice and highly ranked students must do well in such required courses.

- Not all courses have the same difficulty. That is, it is easier to earn a high grade in some courses than in others.

- The correspondence of grades to grade points is as follows: A=4.0, B=3.0, C=2.0, D=1.0, F=0.0. A plus following a grade raises the grade point by one-third, while a minus lowers it by the same amount (i.e., A$- \approx 3.7$, while C$+ \approx 2.3$).

- Students take a fixed courseload for each semester for eight semesters.

- The average grade given at ABC College is A$-$. Thus we assume that the average GPA of students is at least 3.5, the smallest number that rounds to an A$-$.

- In general, X should be ranked ahead of Y (we write X > Y) if:

    - X has better grades than Y, and

    - X takes a more challenging courseload than Y, and

    - X has a more well-rounded courseload (we recognize that this point is debatable).

Analysis of Problem and Possible Models

The Problem with Plain GPA Ranking

The traditional method of ranking students, commonly known as the grade-point average, or GPA, consists of taking the mean of the grade points that a student earns in each course and then comparing these values to determine the student's class rank.

The immediate problem with the plain GPA ranking is that it does not sufficiently distinguish between students. When the average grade is an A−, all above-average students within any class receive the same grade, A. Thus, with only four to six classes per semester, fully one sixth of the student body can be expected to earn a 4.0 or higher GPA.[1] This makes it all but impossible to distinguish between the first and second deciles with anything resembling reliability. Furthermore, any high-ranking student earning a below-average grade, for any reason, is brutally punished, dropping to the bottom of the second decile, if not farther. This is a result of the extremely high average grade; if the average grade were lower, there would be a margin for error for top students.

Unfortunately, the plain GPA exacerbates its own problems by encouraging the grade inflation that makes it so useless. Since the plain GPA does not correct for course difficulty, students may seek out courses in which it is easy to get a good grade. Faced with the prospect of declining enrollment and poor student evaluations, instructors who grade strictly may feel pressure to relax their grading standards. Instructors who grade easily may be rewarded with high enrollment and excellent evaluations, potentially leading to promotion. The entire process may create a strong push towards grade inflation, since the plain GPA punishes both the student taking a difficult course and the instructor teaching it. Any system intended to replace the plain GPA should address this problem, so that grade inflation will be arrested and hopefully reversed.

Another potential concern is that the plain GPA encourages specialization by students. Since students tend to perform better in courses related to their majors, the GPA rewards students who take as few courses outside their "comfort zone" as possible and punish students who attempt to expand their horizons. We note, however, that individual colleges may or may not regard this as a problem; the relative values of specialization and well-roundedness are open to debate.

Three Possible Solutions

Several potential alternatives to GPA ranking directly compare grades within each course. Under such a system, the following considerations come into play:

---

[1]Repeated trials of the process described later yield this result.

- It is not possible to compare students just to others in their own class. Students often take courses in which all other students belong to another class.

- We have to compute rankings separately each semester, because the pool of students changes due to graduation and matriculation.

- It is not possible to take into account independent studies, because there is nobody to compare to.

- It is not possible to take into account pass/fail courses, because they do not assign relative grades.

We recognize three potential solutions to this problem. The following sections describe them in more detail.

- For the *standardized GPA* each student is given a revised GPA based on the student's grade's position in the distribution of grades for each course.

- The *iterated adjusted GPA* attempts to correct for the varying difficulties of courses. In theory, every grade given to a student should be approximately equal to the student's GPA, so that the average grade given in a course should be about equal to the average GPA of students in that course. This scheme repeatedly adjusts all the grade points in each course until the average grade in every course equals the average GPA of the enrolled students.

- The *least squares* method assumes that, other things being equal, the difference between two students' grades will be equal to the difference in their ability scores. It attempts to find these ability scores by solving the system of equations generated by each course (for example, if student X gets an A but student Y gets a B, then $X - Y = 4.0 - 3.0 = 1.0$). Since in any nontrivial population this system has no solution, methods of least-squares approximation are used to approximate these values. The students are then ranked according to ability score.

Standardized GPA

How It Works

The standardized GPA is perhaps the simplest method and one most in keeping with the dean's suggestion. In each course, we determine how many standard deviations above or below the mean each student's grade is. This *standard score* becomes the student's "grade" for the class, the student's standard scores are averaged for a *standardized GPA*, and students are ranked by standardized GPA. This is a quantified version of the dean's suggestion to rank each student as average, below average, or above average in each class, and then combine the information for a ranking.

Strengths

- The standardized GPA is not much more difficult to calculate than the plain GPA measurement.

- Each course can be considered independently. Instead of waiting for all results to come in, the registrar can calculate the standardized scores for each course as grades come in, possibly saving time in sending grades out.

- The standard deviations do correct for differing course averages, for example, getting a B+ when the course average is a C+ looks better than getting an A− when the course average is an A. At the same time, this method continues to rank students in the order in which they scored in each course. Student X is thus always ranked above student Y if X and Y take similar courses and X has better grades.

Weaknesses

The standardized GPA suffers from many of the same problems as the plain GPA.

- It does not reward students who have a more well-rounded courseload. Instead, students are punished severely if they perform at less than the course average; for example, a student who takes a course outside his or her major is likely to score worse than students majoring in the course's subject.

- The plain GPA makes no distinction between easy and difficult courses and thus encourages easy courses. The standardized GPA attempts to correct this but ends up claiming that a low average grade is equivalent to a difficult course. This is not always true and has some interesting quirks:

  - Higher-level courses may be populated only by students who excel both in the subject of the course and in general, so only high grades are given. But if all grades are high, this method treats the course as easy!

  - This method boosts one student's grade if the other students in the course have lower scores.

  - Additionally, ability scores may be significantly raised by adding poor students to the course.

- The standardized GPA does *not* assume that instructors assign grades based on a normal curve or to fit any other prespecified distribution. Not all instructors grade on the normal curve or even on any curve. Some courses may require grades to fit some other distribution in order to be fair, for example, if all the students are extraordinarily talented.

- The method does not compensate for the skill of the students when deciding the difficulty of a course. A good student who takes courses with other good

students will look worse than a slightly less able student who takes courses among significantly less able students. The difficulty of a course should be measured not only by the grades of its students but also by the aptitudes of those students.

Consequences

Grading based on deviation from the mean fosters cutthroat competition among students, since any student's ability score may be significantly raised by lowering the ability scores of other students.

Iterated Adjusted GPA

How it Works

Rather than directly comparing students, this method compares courses. Suppose that a course is unusually difficult. Then students should receive lower grades in that course relative to their others, so the average grade in that course should be lower than the average GPA of all students enrolled in it. We should therefore be able to correct for courses that are unusually difficult by adding a small amount to the point value of every grade given in that course. Likewise, we can correct for easy courses by subtracting a small amount. Of course, once we have corrected everyone's grades, their new GPAs will be different, and most likely some courses will need further correction. The iterated adjusted GPA method makes ten corrections to all grades, then sorts students in order of corrected GPA. (Our numerical experiments show that ten iterations are sufficient to bring the difference between the average GPA and the average grade down to zero, to three decimal places.)

Strengths

- This algorithm is fairly quick to compute, taking only a couple of minutes for 1,000 students, 200 courses, and 6 courses per student.

- The computation is straightforward to explain and easily understood by non-experts.

Weaknesses

- All grades from all courses must be known in order to run the computation.

- The corrected grades cannot be computed independently by students.

- There is no guarantee that the corrected GPAs will be comparable across semesters; to compute overall class rank at graduation, it will be necessary to average ranks across semesters, rather than average corrected GPAs.

Consequences

This method systematically corrects for instructor bias in giving grades, thus eliminating the tendency of students to select easy courses, and therefore makes progress toward reversing grade inflation. The total correction made for each course may be used as an indicator of the course's grade bias.

This algorithm tends to "punish" students in courses where grades are unusually high. If students score high in a course relative to their other grades, it could be because the course was easy or because the students put forth extra effort. If the course was easy, then the punishment is due; if the difference was due to extra effort, then such effort is not typical of the students in question and the punishment is arguably due.

Although the correction can be applied to very small classes and independent studies, strange things are likely to happen. If a student in an independent study gets a grade above his GPA, he is punished by the correction, and if he gets a lower grade, he is rewarded—which is clearly undesirable. Using the sample data set presented later in **Table 1**, we experimented with independent studies and determined that they had minimal impact on the rank order. However, to avoid the possibility of such strange results, independent studies should be ignored in the computation.

The Least-Squares Algorithm

How It Works

The least-squares method assumes that the difference between two students' abilities will be reflected in the difference between their grades. Hence, if X and Y take the same course, and get grades A and B, then we have a difference $X - Y = 4.0 - 3.0 = 1.0$. We further assume that students majoring in natural science fields perform better in natural science courses than in humanities courses, and vice versa, and that the difference is of approximately the same for all students; we call it H_H. Hence, if, in the example above, students X and Y are taking a mathematics course, but X is majoring in physics and Y is majoring in literature, we have $X - (Y + H_H) = 1.0$.

A course with N students generates $N(N + 1)/2$ such linear equations; the abilities of each student are the solution to the set of all such equations from every course offered during the semester. In practice, these equations never have a solution. Hence, methods of least-squares approximation must be employed. The system is converted into the matrix equation $Ax = b$, where A is the matrix of the coefficients of the left-hand side of each equation, x is the

vector of the abilities of each student and the constant H_H, and b is the right-hand side of each equation. Multiplication by the transpose of A yields the equation $A^T A x = A^T b$. This matrix equation has a one-dimensional solution set, with nullspace equal to scalar multiples of $(1 \ 1 \ 1 \ \ldots \ 1 \ 0)^T$, where the 1s correspond to the student's abilities and the 0 to the constant H_H. Thus, one student's ability score may be assigned arbitrarily, and the rest will then be well determined. This arbitrary assignment will in no way affect the ordering of any two students' ability scores, or the magnitude of the difference between two students. After these scores are determined, the difference between a 2.0 and the median score is added to every student's score, so that the scores will be easily interpretable in terms of the plain GPA. These scores can be averaged over all eight semesters to produce a ranking at graduation.

Strengths

- Least squares corrects for the difficulty of every student's courseload.

- Least squares can reward students for carrying a well-rounded courseload. This second strength is extremely flexible, and deserves further enumeration.

  - If a school wishes not to account for well-roundedness, the factor H_H may be omitted, with no consequence except that the ability scores will no longer consider the balance or specialization in each student's courseload.

  - If a school wishes to emphasize several areas of specialization rather than just two, it could do so by replacing H_H with constants representing the difficulty of the transitions between each pair.

  - A school wanting to assure that certain emphasized courses (e.g., a freshman writing course) not unduly benefit students majoring in some departments could categorize such courses as belonging to every area of specialization, or to none.

  - Similarly, if a school wishes to dictate that certain de-emphasized courses (e.g., physical education) not reward students with a well-roundedness correction, it may also dictate that they be categorized in every area of specialization or in none.

  - Other corrections may be made for students with special circumstances; for example, if a student double-majors in two different areas of specialization, each well-roundedness correction might be replaced by the average of the two corrections from each of the student's major areas.

Weaknesses

The most glaring weakness of this method is that it involves huge amounts of computation and may severely tax computing resources at larger universities.

For a student body of 6,000, with 120 courses of size 20 and each student taking 4 courses, we have $1{,}200 \times 21(21+1)/2 \approx 250{,}000$ pairs of grades. This results in a sparse A with 250,000 rows, 6,000 columns, and only 4 nonzero entries in each column (for the 4 courses that the student took). Then $A^T A$ has 36,000,000 entries; at 4 bytes per entry, keeping it in memory requires 144 MB, barely within range of current medium-size computers. Computing $A^T A$ takes on the order of $250{,}000 \times 6{,}000^2 = 9 \times 10^{12}$ multiplications, computing $A^T b$ takes only about 1.5×10^9 multiplications, and solving $A^T A x = A^T b$ takes about $6{,}000^3 = 2.2 \times 10^{11}$ operations. Thus, the time to solve the system is about 10^{13} operations, which would take $50{,}000$ sec ≈ 14 hr on a 200 MHz computer.

The memory needed increases with the square of the number of students and quickly becomes infeasible with this approach and current technology.

Consequences

An immediate consequence of changing to this ranking will be that, so long as the average grade remains an A−, all ability scores will be tightly packed into a range between about 1.0 and about 3.0; no student will appear to carry an A average. This will likely result in instructors widening their grading scales, in order to reward their best students, thus reducing grade inflation to something more reasonable.

A Small Test Population

We postulate a minicollege, with 18 students (A–R), that offers only the following courses: Math, Physics, Computer Science, Physical Education, Health, English, French, History, Philosophy, Psychology, and Music History.

Math, Physics, and English are generally believed to be prohibitively difficult courses, while Physical Education, Health, and Music History are generally considered to be very easy. Students' transcripts are listed in **Table 1**. Just looking at these transcripts, without analyzing them numerically, we find that we should have the following, which any valid ranking system *must satisfy* (recall that X > Y means that X should be ranked above Y):

- A > B; C > D; and E > F, and so on, because A, C, etc., carry better grades than B, D, etc., in courseloads of similar difficulty.

- O, D > J because O and D have slightly better grades than J in more difficult courseloads.

- E > D because E has better grades in a more difficult courseload.

  We also recognize the following relationships as *desirable*:

- O > Q, R and P > R, because O and P have almost as good grades and much more difficult schedules.

Table 1.
Transcripts of the test population.
A star indicate the student's major. "CPS" means Computer Science and "PhysEd" means
Physical Education.

Student	Courses
A	PhysEd 4.3, Health 4.0, *History 3.0, Math 2.3
B	PhysEd 4.3, Health 3.3, *Psychology 2.0, CPS 2.0
C	Math 4.0, *Physics 4.3, CPS 4.0, Philosophy 3.7
D	*Math 4.0, Physics 3.7, CPS 4.0, French 3.0
E	*Math 4.3, Physics 4.0, English 3.3, History 3.7
F	Physics 3.7, *CPS 4.0, French 3.7, History 3.0
G	Math 4.0, *CPS 4.3, Health 4.0, English 3.7
H	CPS 3.0, *Physics 4.0, PhysEd 4.0, Psychology 3.0
I	English 4.0, French 4.3, CPS 3.7, *Philosophy 4.3
J	English 3.7, *French 4.0, Music History 4.0, Math 2.7
K	*English 4.3, Philosophy 4.0, Psychology 4.0, Music History 4.3
L	English 3.7, *History 4.0, Psychology 4.0, Music History 4.0
M	Music History 4.3, Psychology 4.3, *French 4.3, PhysEd 4.0
N	*Music History 4.0, Psychology 4.0, French 4.0, Health 4.0
O	Physics 4.0, English 3.3, *Math 4.0, Philosophy 4.0
P	Physics 3.0, *English 3.7, Math 3.3, Philosophy 4.0
Q	PhysEd 4.0, Health 4.3, Music History 4.3, *Psychology 4.3
R	PhysEd 4.0, Health 4.0, Music History 4.0, *CPS 4.0

- M > Q and N > R, because M and Q have similar grades but M has a more difficult schedule, and similarly for N and R.

- K > M, N, Q, R because K has similar grades in a much more difficult schedule.

- C, G, and K should be ranked near each other because they have similar grades in similar schedules.

- P > J because P has similar grades against a significantly more difficult schedule and has higher grades in the two classes that they share.

If we postulate that the well-roundedness of a student's schedule should affect rank, we also find the following relationships:

- E > C, D because E has almost as good grades in a more difficult, much more well-rounded schedule.

- I > K, M because I has similar grades against a more well-rounded schedule.

The rankings of this sample population are given in the **Table 2**. A comparison of the different methods relative to the criteria that we have set out is in **Table 3**. Least squares does best, followed by iterated adjusted, standardized, and plain.

Table 2.

Rankings of the sample population under the various methods.

Rank	With +/−								Without +/−							
	Plain		Standardized		Iterated		LS		Plain		Standardized		Iterated		LS	
1	Q	4.25	K	0.84	K	4.22	E	2.32	R	4.00	G	0.53	L	4.12	G	2.25
2	M	4.25	I	0.81	I	4.17	I	2.26	Q	4.00	L	0.49	G	4.07	I	2.19
3	K	4.17	Q	0.60	M	4.09	G	2.24	C	4.00	C	0.39	I	4.06	E	2.17
4	I	4.08	M	0.52	C	4.08	C	2.24	N	4.00	I	0.36	C	4.05	C	2.17
5	R	4.00	G	0.39	G	4.07	O	2.18	M	4.00	N	0.34	K	4.03	F	2.14
6	N	4.00	C	0.22	L	4.06	K	2.14	G	4.00	K	0.27	N	3.96	O	2.04
7	C	4.00	E	0.21	E	4.05	M	2.05	K	4.00	M	0.24	E	3.92	L	2.04
8	G	4.00	L	0.16	Q	4.02	Q	2.03	I	4.00	Q	0.24	M	3.89	D	2.03
9	L	3.92	N	−0.01	O	3.96	R	2.01	L	4.00	R	0.23	J	3.87	R	2.02
10	E	3.83	O	−0.03	N	3.90	F	2.01	O	3.75	F	0.11	D	3.84	K	1.98
11	O	3.83	R	−0.20	R	3.76	P	1.94	J	3.75	J	0.07	F	3.84	J	1.95
12	D	3.67	D	−0.26	D	3.74	D	1.93	F	3.75	E	0.07	O	3.83	N	1.90
13	J	3.58	A	−0.27	F	3.69	L	1.92	E	3.75	D	−0.12	Q	3.81	M	1.88
14	F	3.58	F	−0.28	J	3.66	N	1.87	D	3.75	O	−0.15	R	3.80	P	1.87
15	H	3.50	H	−0.45	P	3.62	J	1.74	H	3.50	H	−0.30	P	3.58	Q	1.85
16	P	3.50	J	−0.49	H	3.41	H	1.60	P	3.50	P	−0.60	H	3.39	H	1.58
17	A	3.42	P	−0.59	A	3.36	A	1.44	A	3.25	A	−0.61	A	3.18	A	1.26
18	B	2.92	B	−1.16	B	2.76	B	0.89	B	2.75	B	−1.56	B	2.59	B	1.54

Table 3.

Number of criteria satisfied by each method on the minicollege data set, for +/− grades.

	Plain	Standardized	Iterated	Least Squares
Required (20)	all	all	all	all
Desirable (13)	5	6	8	9
Well-roundedness (4)	1	2	2	all

Test Population Redux (No +/− Grades)

We now take the test population and drop all pluses and minuses from the grades. Again, we determine some basic *required* relationships that any valid ranking system must satisfy:

- A > B; C > D; and G > H since A, C, and G have better grades in similar courses.

We also recognize the following relationships as *desirable*:

- O > P because O has slightly better grades in the same courseload.

- E > F because E has the same grades in a more difficult courseload.

- O > Q, R because O has almost equivalent grades in a much more difficult courseload.

- C > I, G because C has the same grades in a more difficult courseload.

- I > K, L because I has the same grades in a more difficult courseload.

- K, L > M, N because K and L have the same grades in a more difficult courseload.

- M, N > Q, R because M and N have the same grades in a more difficult courseload.

If we postulate that the well-roundedness of a student's schedule should affect rank, we also find that C, E, G, and I should be ranked near each other because

- E has slightly worse grades in a more difficult, better-rounded courseload; and

- C has the same grades as G and I in a slightly more difficult, slightly less well-rounded courseload.

The rankings of this sample population are given in the right-hand half of **Table 2**. **Table 4** gives a comparison of the methods.

Table 4.

Number of criteria satisfied by each method on the minicollege data set (no +/− grades).

	Plain	Standardized	Iterated	Least Squares
Required (3)	all	all	all	all
Desirable (12)	1	6	9	9
Well-roundedness (6)	3	1	3	4

Stability

How Well Do the Models Agree?

We have four ways of ordering students: plain GPA, standardized GPA, iterated adjusted GPA, and least squares. Since all four are more or less reasonable, they should agree fairly well with each other. One way to test agreement is to plot each student's rank under one method with his rank under the others. If the plot is scattered randomly, then the rankings do not agree about anything. If the plot is a straight line, then the rankings agree completely.

To get an idea for how each model works, we created by means of a computer simulation a population of 1,000 students and 200 courses, with 6 courses per student. The details of the simulator are explained in the **Appendix**. We implemented all of the algorithms except least squares, which was too difficult for the available time. A single run of the simulation is analyzed here, but these results are typical of other runs.

With Plus and Minus Grades

See **Figures 1–3** for graphs of the agreement, using simulated students and courses, and allowing plus and minus grades. The comparisons to plain GPA rankings are rather scattered, especially toward the lower left corner, where

the highest rankings are. The plain GPA rankings do not appear to agree particularly well with either the iterated adjusted or the standardized rankings. There are lots of scattered points, which is due mostly to the facts that there are lots of ties in plain GPA rankings (especially near the top of the class) and that tied students are ordered more or less at random. Very few ties are present in any of the other methods.
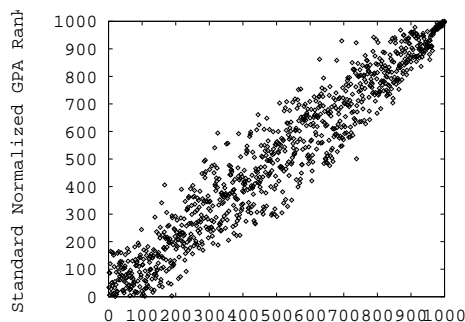
Figure 1. Plain GPA rankings vs. standardized GPA rankings, using simulated students.
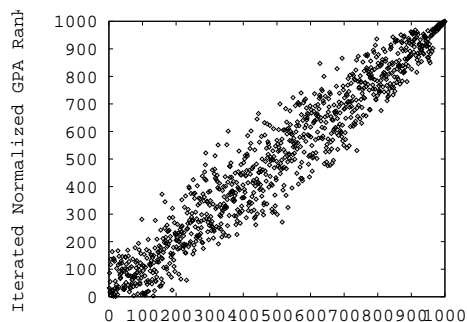
Figure 2. Plain GPA rankings vs. iterated adjusted GPA rankings, using simulated students.
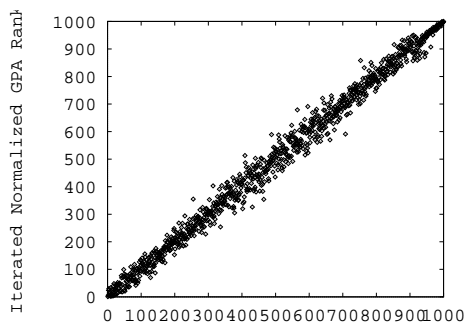
Figure 3. Standardized GPA rankings vs. iterated adjusted GPA rankings, using simulated students.
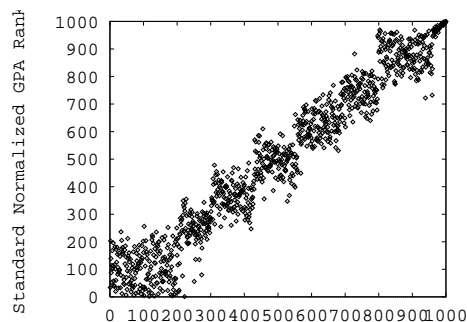
Figure 4. Plain GPA rankings vs. standardized GPA rankings, using simulated students, with no plus or minus grades.

The iterated adjusted and standardized rankings are in better agreement, with fewer outlying points. These two methods agree on 89 of the 100 students in the top decile.

Without Plus and Minus Grades

See **Figures 4–6** for graphs of the agreement, using simulated students and courses, and disallowing plus and minus grades.

A great deal of information is lost without the use of plus and minus grades. In particular, there are many more ties in the plain GPA-based ranking, which show up as large squares of scattered points. The large square at the bottom left shows the massive tie among people with 4.0 averages. Again, the plain GPA is not in good agreement with the nontraditional methods due to these
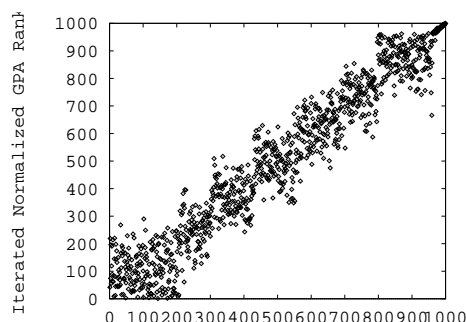
Figure 5. Plain GPA rankings vs. iterated adjusted GPA rankings, using simulated students, with no plus or minus grades.
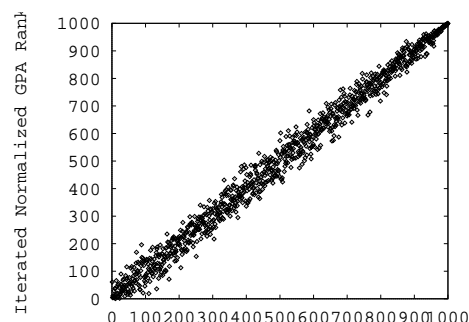
Figure 6. Standardized GPA rankings vs. iterated adjusted GPA rankings, using simulated students, with no plus or minus grades.

ties. Both new models agree with each other on 79 of the 100 students in the top decile. Apparently, the loss of information is responsible for the greater lack of agreement.

How Much Does Changing One Grade Affect the Outcome?

If one grade of one student is changed, the student's rank can be expected to change as well. For plain GPA rankings, changing one student's grades can only move that student from one place to another. In the nontraditional rankings, each student's rank is determined relative to the other students, and one changed grade might trigger a chain of rank changes.

To test sensitivity, the sample population was modified slightly: Student Q's grade of A+ in Music History was changed to a C−, a very drastic change. The change was tested including plus and minus grades and using only whole letter grades. (When only letter grades are considered, the change is to a C.)

- Using the GPA ranking and plus and minus grades, Q dropped from 1st to 14th; with only whole letter grades, Q dropped from 2nd to 16th. In both cases, there were no changes in the order of other students except to make room for Q.

- For the standardized GPA with plus and minus grades, Q dropped from 3rd to 12th. L, N, and J improved several places, apparently because they also took Music History and benefitted from the drop in mean grade. R improved one spot, apparently for the same reason. K dropped by one. Without plus and minus grades, Q dropped 8 places, and J, L, and N improved one rank each. Student I dropped three places, perhaps because of how N and K benefitted from Music History.

- The iterated adjusted GPA including plus and minus grades was rather stable. Q dropped 9 places, and J and L improved a couple of places each,

benefitting from the apparent increase in the difficulty of Music History. G dropped two places, possibly because he scored lower in Health. When only whole letter grades are used, Q dropped from 13th to 16th. J and K improved a couple of places, benefitting from the increased difficulty of Music History, while G dropped again, three places this time. O and F switched places for no obvious reason.

- Using least squares and plus and minus grades, Q dropped 9 places. Other members of the Music History course J and K improved a bit, and L improved a lot. With letter grades only, Q dropped from 15th to 16th, and J, K, and L improved. For no obvious reason, E and C switched places. O dropped by two because of improvements by K and L.

Thus, it would seem that plain GPA ranking is the most stable, since at most one person changes rank and the rest move up or down at most one rank to compensate. The next most stable seems to be least squares, followed by iterated adjusted, and finally standardized. In each scheme, the coursemates of the person whose grade changed are most likely to change rank. There were a few chain-reaction reorderings, which are harder to explain. Also, having plus and minus grades appears to improve stability in general.

How Does Course Size Affect the Outcome?

Another simulation was run with 1,000 students, 500 courses, and 6 courses per student. Courses came out smaller, and the correlation between the standardized ranking and the iterated adjusted ranking was weaker. This is probably due to the fact that standard deviations computed on smaller data sets tend to be less reliable, as are average grades and average GPAs.

Strengths and Weaknesses of Each Model and Recommendations

If the college wishes to promote well-roundedness over specialization (we would suggest this), and has a fairly small population (fewer than about 6,000 students), we recommend the least-squares method. Otherwise, we recommend the iterated adjusted GPA method.

We feel that the least-squares method is superior to the other two because:

- It does not punish students for attempting to expand their horizons.

- It produces results more consistent with intuitive observation than do the iterated or standardized GPA.

- It is more flexible than either the iterated or standardized GPA.

- It is clear and easily understood.

  The iterated adjusted GPA method has a few definite advantages as well:

- It is significantly faster than the least-squares method.

- If the well-roundedness of students is not a consideration, it produces results that are roughly as consistent with intuitive observation as the least-squares method.

  We feel that the standardized GPA method is decidedly inferior, and should not be recommended, because:

- It makes no attempt to correct for schedule difficulty or well-roundedness.

- It assumes that all courses have the same range of ability among their students.

- It produces results that are no more consistent with intuitive observation than those produced by the plain GPA.

Further Recommendations

Transition from GPA Ranking

The three methods given here all rank an entire student body for one semester of courses. Thus, to rank students just within a single class, we must either average their ability scores (revised GPAs) or their ranks within their class over each semester. The new system could be phased in at any time if grades for enough preceding years are kept on record. The new ranking algorithm could be applied to students who have graduated to determining rankings for the next class. However, we recommend careful testing on several past years of data as well as current grades. The administration should be prepared for a great deal of student and faculty opposition because it is a new, untested system. The standardized and iterated adjusted schemes are likely to encounter opposition because they directly alter the point values of grades during computation. The least-squares method simply reinterprets them and is less likely to make instructors feel that their authority has been violated.

Transfer Students

ABC College will have to come up with its own policy concerning the ranking of transfer students. One option is to translate transferred grades to an equivalent grade in a particular course at ABC. That allows the ranking algorithm to run on the maximum amount of information. However, someone will have to compare all other colleges to ABC very carefully to create the official

translation policy. Another possibility is to ignore transferred grades when computing the rankings. That avoids the problem of estimating how grades at other schools compare to ABC's, but at the expense of throwing out a lot of information.

Importance of Plus and Minus Grades

It seems that plus and minus grades are extremely helpful in determining class rank, especially since grades are so heavily inflated. Without them, ABC has to rank its students primarily on the basis of just two grades, A and B, and a considerable fraction of the students have exactly the same grades. With pluses and minuses, there are six different grades, A+, A, A−, B+, B, and B−, which come into play, thus differentiating students more precisely. All four ranking systems appear to work better when plus and minus grades are used. ABC should encourage its instructors to use them with care.

Appendix: Details of the Simulation

Simulating Courses

We want to take the following things into consideration when creating courses:

- Students tend to pick more courses in areas they are comfortable in. In particular, they are required to select courses in their majors.

- Courses vary in subject matter. Some require a lot of math and scientific experience, while others focus more on human nature, history, and literature.

- Courses vary in difficulty. Here, we are not considering the difficulty of the material, but rather how difficult it is to get a good grade in the course. Students generally prefer courses where they expect to get better grades.

- Students are able to estimate their grade in a course fairly accurately.

Each simulated course c therefore has three attributes. The first two are fractions, c_s and c_h, which represent how much the course emphasizes the sciences and the humanities, respectively. Since these are fractions of the total effort required for a course, we have $c_s + c_h = 1$. In the simulation, c_s is determined by generating uniformly distributed random numbers between 0 and 1, and $c_h = 1 - c_s$.

The third attribute c_e is the "easiness" of the course, that is, how easy it is to get a good grade. This number represents the tendency of the instructor to give higher or lower grades. In the simulation, c_e is determined by taking a uniformly distributed random number between -0.5 and 0.5, indicating that

instructors may skew their grades by up to half a letter grade up or down. We use a uniform distribution rather than a normal distribution so as to make the courses vary in difficulty over the entirety of a small range.

Simulating Students

We want to take the following things into consideration when creating simulated students:

- Students have varying strengths and weaknesses. In particular, some students have different ability levels in the sciences and humanities. Students prefer courses within their comfort zones.

- Students prefer getting higher grades.

Each simulated student S has two attributes, S_s and S_h. Both of these are numbers representing grades that indicate the student's abilities in the sciences and humanities, respectively. Both range from 0 to $g_{\max}$, which is either 4.0 or 4.3 depending on the grading scale.

Given a course c and a student S, the grade for that student in that course is given by

$$g = \min\left(S_s c_s + S_h c_h + c_e, g_{\max}\right). \tag{1}$$

In the simulation, S_s and S_h are determined by taking random numbers from a normal distribution with mean 3.5 and standard deviation 1.0, with a maximum of $g_{\max}$.

Generating a Simulated Population

The simulated population is created by first generating a number of courses and a number of students. A courseload is selected for each student S by repeating the following: First, a course c is selected at random. If the student is weak in science ($S_s < 2.5$) and the course is heavy in science ($c_s > 0.75$), then the course is rejected. Similarly, if the student is weak in humanities and the course is heavy in humanities, the course is rejected. If the student estimates his or her overall grade at less than 2.5, the course is rejected. This process of selection and rejection is repeated until a course is not rejected, but at most ten times, and then the last course is taken no matter what. The selected course is then added to the student's schedule and the grade computed as stated in **(1)**, rounded to the nearest possible grade.

The rejection process allows for the students' preferences in selecting courses, and the fact that at most ten courses can be rejected allows for distribution requirements.

Analysis of the Simulated Data

The simulation program was used to create 1,000 students and 200 courses, where the courseload was six. Thus, there were around $1{,}000 \times 6/200 \approx 30$ people in each course, which is reasonable. Two runs were made, one with only whole grades, and one with $+$ and $-$ grades allowed.

We can determine a lower bound for the average GPA at ABC College. Suppose we have N students, each of whom takes M courses. Denote by g_{ij} the grade of student i in that student's jth course. Then the average grade for that entire class is given by

$$\frac{\sum_{i=1}^{N} \sum_{j=1}^{M} g_{ij}}{NM},$$

and the average GPA is given by

$$\frac{\sum_{i=1}^{N} \frac{\sum_{j=1}^{M} g_{ij}}{M}}{N}.$$

The two are equal, so if the average grade at ABC College is A$-$, then the average GPA should be no more than 3.5. Any GPA less than 3.5 would be rounded to a B$+$ or less, and those greater than 3.5 would be rounded to A$-$ or better. In the both data sets, the median GPA was 3.5, which agrees with the information given about ABC College.

Strengths and Weaknesses of the Simulation

The computation runs very quickly—in a few minutes—even though it was written in a high-level interpreted language (Python). It is very flexible and can be adjusted to reflect different grade distributions, as may be found in different colleges. It takes into account variation in student interest and in course material.

However, most of the courses turn out roughly the same size. Many colleges have a high proportion of small, seminar-style courses, and there are almost always some very large lectures. The simulation ranks the whole school together and does not distinguish among the classes. There are only two majors in the simulation, sciences and humanities; and while there are forces within the simulation that push students into taking more courses in their preferred area of knowledge, there are no guarantees that the resulting schedules accurately reflect major requirements. There are also no prerequisites enforced, and thus no courses that are predominantly populated by freshmen and seniors. This also means that the simulator cannot realistically create courses for more than one year.

A Case for Stricter Grading

Aaron F. Archer
Andrew D. Hutchings
Brian Johnson
Harvey Mudd College
Claremont, California 91711
{aarcher,ahutchin,bjohnson}@hmc.edu

Advisor: Michael Moody

Abstract

We develop a ranking method that corrects the student's grades to take into account the harshness or leniency of the instructor's grading tendencies.

We simulate grade assignment to a student based on the student's inherent ability to perform well, the student's specific aptitude for the course, the difficulty of the course, and the harshness or leniency of the instructor's grading.

We assume that we have access to a instructor's previous grading history, so that we can judge how harsh or lenient a grader each instructor is. After making this determination, we adjust each grade given by that instructor to systematically correct for that instructor's bias.

After correction, the student body has an aggregate GPA of approximately 2.7, corresponding to an uninflated grade of B−. The corrected GPA values do a considerably better job of accurately ranking the students by ability, especially for students in the bottom eight deciles.

Assumptions

1. We wish to evaluate students purely based on their ability to perform well in courses.

2. Each student has a quality attribute that is not directly measurable but influences the ability to do well in courses. The ideal ranking of students is by highest quality attribute.

3. The instructor has an accurate perception of each student's performance in a course.

4. The cause of high grades is lenient grading practices by the average instructor at ABC College.

5. A more lenient instructor tends to grade all students higher, not just students of a certain ability level.

6. Scholarship selection is completed in the first half of a student's undergraduate career, to allow her to enjoy the scholarships while she is still in school.

7. Because the students are early in their careers at ABC College, they are still taking primarily general education courses, rather than courses in their major. Therefore, we assume that they select courses randomly, and thereby we model the breadth of course selection across disciplines.

8. Since the students know that their grades are going to be adjusted to filter out the harshness of their instructors' grading, they do not gravitate toward courses taught by lenient instructors.

9. Each student has a varying aptitude for each course. Presumably, a student has more aptitude for courses in her major. But since general course requirements tend to be broad and these are the courses we are examining, we assume that a student's aptitude for a course is random.

10. Each course has an inherent difficulty. In an easy course it is difficult to differentiate the high ability students from the rest, whereas tougher material produces a greater spread of performances.

11. Instructors know when a course is difficult. Presumably all students (even the top ones) will attain a lesser mastery of more difficult material, but the instructor will take this into account when assigning grades.

12. The college is on a semester system and each student takes four courses per semester.

13. A student's performance in a course is not influenced by which other students are taking the course. Neither is the student's grade, since we assume that instructors do not grade the students in a given course on a curve but rather on some absolute standard of performance.

14. An instructor's harshness in grading does not depend on the course and remains constant over a period of several years. Data on instructors' grading histories are available.

15. All instructors rate a student's performance the same, but they have different standards for what grade that performance should earn.

Practical Considerations

The concept of a single quality attribute that describes each student is not one that plays well politically and in the media. Not many people would advocate that a student's overall ability to do well in courses can be accurately characterized by a single real number. Therefore, our adjusted measure of student ability should be some sort of adjusted GPA, which will be easier for a general audience to accept and understand. This does not present a problem from the modeling point of view, as long as we know how quality rankings correspond to GPA values, and vice versa.

Ultimately, as we construct our model, we will run into a fundamental grading problem. The average grade at ABC College is an A−, which corresponds to a 3.67 GPA. Grade point averages that are this high result in very uninteresting grade distributions. The majority of the grades must be A+, A, or A−. In other words, if we look at the transcript of any above-average student at ABC, we will probably see a page full of A+, A, and A− grades. In this kind of environment, it will be extremely difficult to pick out the top few students, because the top half of the school is separated by only about 0.6 grade points. In contrast, the bottom half of the school is spread over the remaining 3.67 grade points, so it will be much easier to rank them by ability.

One radical solution to this dilemma is to require additional feedback on student performance from the instructors. We outline one possible system here, before we move on to a less radical approach. In addition to giving grades on the usual A to F scale, we could require an instructor to give each student a ranking between 1 and 10. At least one student in each course must receive a 1, and at least one student must receive a 10. This forces a spread in the instructor's rankings, so that even an easy-grading instructor (all A+ grades) must rank the better-performing students above the less able students. Next, the instructor is allowed to give a context to the scale. If the instructor has taught that course before, she would be asked to rate the current course in terms of previous ones. We ask the instructor to identify, on some absolute scale of ability, which interval corresponds to the 1 to 10 relative scale for the course. For example, if the instructor felt that her best student was about as competent as a student at the 90th percentile, then she would identify the right endpoint of the scale with the 90th percentile of absolute student ability. If she felt that her worst student was the poorest student to attend the college over an entire five-year span, then she would identify the left end of the relative scale with that point on the absolute scale. This two-stage evaluation system forces the students to be differentiated by performance puts the measures of performance into an absolute (rather than instructor-dependent) context.

What Characterizes a Good Evaluation Method?

As we attempt to rank the students at ABC College, we assume that the students have underlying quality scores that are reflected in their grades. We try to approximate the ranking induced by the hidden quality values. It may be inappropriate (for political reasons) to refer to our rankings as "estimated student qualities," so we instead calculate an adjusted GPA.

As we calculate adjusted GPA values, we keep in mind several goals:

- We wish to allocate correctly the available scholarships to the top 10% of the student body. To test whether or not we succeed, we must compare the ranking induced by our adjusted GPA values with the actual ranking of the students by intrinsic quality. Our first measure of the accuracy of our adjusted GPAs will just be the number of scholarships that we correctly allocated to deserving students.

- If the top-ranked student somehow fails to receive a scholarship, this is considerably more unjust than if a student who just barely deserves a scholarship misses out. Thus, we compute a second measure of accuracy by summing the severity of the mistakes made in awarding scholarships.

- It is important for all of the student rankings to be accurate, not just the top 10%, because they are used for much more than just scholarship determination. For instance, class rank is often cited in graduate school and job applications. Therefore, we consider a third measure of accuracy that gives a total error measure for our entire set of adjusted GPA rankings, rather than for just the top decile.

Modeling College Composition and Grade Assignment

According to Assumption 13, we do not need to consider the other students in a course when we determine a student's performance in the course and the grade the student receives; in other words, the composition of students in the course does not significantly affect the students' ability to learn, and none of the instructors grades on a curve. Thus, we model a student's grade as a function of

- her inherent quality,

- her aptitude for the specific course,

- the difficulty of the course, and

- the harshness of the instructor grading the course.

We treat each of these quantities as real-valued random variables and generate their values by computer.

We let q_i denote the inherent quality of student i. We will consider q_i to be distributed normally with mean 0 and standard deviation σ_q. This is reasonable, since we know that the normal distribution gives a good approximation for many characteristics of a large population.

We let $c_{i,j}$ represent the random course aptitude adjustment for student i when she takes course j. Again, it makes sense to let $c_{i,j}$ be normally distributed about 0, and we denote the standard deviation of this aptitude adjustment by σ_c. We let the net aptitude of student i in course j be $q_i + c_{i,j}$, which is normally distributed with mean 0 and standard deviation $\sqrt{\sigma_q^2 + \sigma_c^2}$. We choose our unit of measure so that $\sigma_q^2 + \sigma_c^2 = 1$. Furthermore, we estimate that a student's intrinsic quality influences her success at least five times as much as her aptitude adjustment for the particular course she is taking. Hence, we choose $\sigma_c < 0.2$.

Next, we consider how the difficulty of a particular course affects the grades that the instructor gives. We assume (see Assumption 10) that a difficult course spreads out the distribution of grades given; this means that poor students tend to do worse in difficult courses, but also that excellent students will do better, since they are being given an opportunity to excel. Conversely, in an easy course, the grades tend to bunch closer together, since the poor students are being given an opportunity to excel and the best students' performances are limited by the ease of the subject matter. Let d_j denote the difficulty of course j; then this interpretation leads us to consider a performance rating $N_{i,j}$ of student i in course j given by

$$N_{i,j} = (q_i + c_{i,j})d_j,$$

where d_j is a positive number, equal to 1 for a course of average difficulty, greater than 1 for a difficult course, and less than 1 for an easy course. Note that we are assuming the performance of a student in a course is random only in that the student's inherent ability is modified by a random aptitude adjustment factor. Once this factor is applied, the student's performance is determined, given the difficulty of the course.

Finally, we must take into account the grading philosophy of the instructor. Notice that a difficult course does not shift the performance distribution to the left, because the performance is measured relative to the instructor's expectation. We assume that the instructor is aware of the difficulty of the course and compensates accordingly in grading. This brings up a delicate distinction. An instructor's harshness does not reflect her expectation level but only her tendencies in grading. That is, we assume that the instructor's harshness does not pertain to her assessment of a student's performance but rather to what grade she thinks that performance deserves.

Let h_k denote the harshness of instructor k; then we should let the student's grade depend on $N_{i,j} - h_k$, since the harshness causes a systematic bias in all

of the grades that the instructor gives. We let a harshness of 0 correspond to an average instructor at an institution without grade inflation. At Duke University and presumably at other institutions, 2.7 was the average GPA prior to the grade inflation that began to appear in the 1970s [Gose 1997]. Therefore, letting G denote the grading function that maps real numbers to discrete letter grades, we should center $G(0)$ on a grade of B$-$. Furthermore, among instructors who grade on a curve, an interval of one letter grade is often equated with one sample standard deviation in the course scores. So, we let one standard deviation for a course of average difficulty correspond to a whole letter grade in our model. Thus, the instructors in our model are grading on a virtual curve; that is, they grade on an absolute standard that simulates grading on a curve in a hypothetical course in which the full distribution of students is enrolled.

This analysis leads to a grading function

$$G: \quad \to \{0, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$$

defined by

$$G(x) = \begin{cases} 0, & \text{if } x \le -\frac{11}{6}; \\ 3, & \text{if } -\frac{11}{6} < x \le -\frac{9}{6}; \\ 4, & \text{if } -\frac{9}{6} < x \le -\frac{7}{6}; \\ \vdots \\ 12, & \text{if } \frac{7}{6} < x \le \frac{9}{6}; \\ 13, & \text{if } x > \frac{9}{6}, \end{cases}$$

where the values $0, \dots, 13$ represent the letter grades F, D, D+, C$-$, C, C+, B$-$, B, B+, A$-$, A, A+. To convert the numeric value to grade points, we divide by 3; thus, an A$-$ average means a GPA of 3.67. If student i takes course j taught by instructor k, she will receive a grade of

$$G(N_{i,j} - h_k) = G\big((q_i + c_{i,j})d_j - h_k\big).$$

For convenience, we define $l(g)$ and $r(g)$ to be the left-hand and right-hand endpoints of the interval on which $G = g$. For instance, $r(13) = \infty$ and $l(12) = \frac{7}{6}$.

A simple calculation reveals that in a course where $d = 1$ and $h = 0$, the expected grade is 2.63 (see **Table 2** and **Figure 1**). We intend harshness 0 to represent a reasonable level of strictness in grading. It centers the grades at B$-$, which is the exact middle of all passing grades, and yields a GPA in line with the "reasonable" historical number of 2.7 at Duke University.

We can visualize the grading method by graphing a $\text{normal}(0, 1)$ density function, which represents N (in the case where difficulty is 1), with the x-axis partitioned into intervals representing grades according to the grading function G (see **Figure 2**). A difficult course spreads out the distribution, resulting in more Fs (because the poor students cannot keep up) and more As (because the top students have an opportunity to shine). A positive (negative) harshness effectively shifts the grade intervals to the right (left).
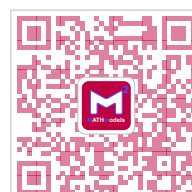
Table 1.

Symbol table.

$c_{i,j}$	course aptitude adjustment for student i taking course j
d	estimate of course difficulty
d_j	difficulty of course j
G	grading function, from performance rating to letter grade
$\overline{g}$	average grade given by instructor, from historical data
g_{adj}	adjusted grade that an instructor of harshness zero would give
h_k	harshness of instructor k
I	interval in which student's performance value is estimated to lie
$l(g), r(g)$	endpoints of performance rating interval corresponding to letter grade g
$N_{i,j}$	performance rating of student i in course j
N_{est}	estimate of student performance value
$N_{i,\mathrm{est}}$	estimate of student i's performance value
Φ	standard normal cumulative distribution function
q_i	inherent quality of student i
$q_{i,\mathrm{est}}$	estimate of inherent quality of student i
σ_q	SD of inherent quality of student i
σ_i	SD of course aptitude adjustment of student i taking course j
$\sigma(d, h)$	SD of grades given by a professor of harshness h in a course of difficulty d
t_0	left endpoint for performance rating interval corresponding to grade of A+
t_1	right endpoint for performance rating interval corresponding to grade of F

Table 2.

Expected grade on a 4-point scale as a function of course difficulty d and instructor harshness h.

|						d					
h	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
-3.0	4.33	4.33	4.32	4.31	4.30	4.29	4.27	4.25	4.23	4.20	4.18
-2.8	4.33	4.32	4.31	4.30	4.28	4.27	4.24	4.22	4.19	4.16	4.13
-2.6	4.32	4.31	4.30	4.28	4.26	4.24	4.21	4.18	4.15	4.12	4.08
-2.4	4.31	4.30	4.28	4.25	4.22	4.19	4.16	4.13	4.10	4.06	4.03
-2.2	4.29	4.27	4.24	4.21	4.18	4.14	4.11	4.07	4.03	4.00	3.96
-2.0	4.26	4.22	4.19	4.15	4.11	4.08	4.04	4.00	3.96	3.92	3.88
-1.8	4.19	4.15	4.11	4.07	4.03	3.99	3.95	3.91	3.87	3.83	3.79
-1.6	4.10	4.06	4.02	3.98	3.94	3.90	3.86	3.82	3.78	3.73	3.69
-1.4	3.97	3.94	3.90	3.86	3.82	3.78	3.74	3.70	3.66	3.62	3.58
-1.2	3.82	3.79	3.76	3.72	3.69	3.65	3.62	3.58	3.54	3.50	3.46
-1.0	3.64	3.62	3.60	3.57	3.54	3.51	3.48	3.44	3.41	3.37	3.33
-0.8	3.45	3.44	3.43	3.41	3.38	3.35	3.32	3.29	3.26	3.23	3.19
-0.6	3.26	3.25	3.24	3.23	3.21	3.19	3.16	3.13	3.10	3.07	3.04
-0.4	3.06	3.06	3.05	3.04	3.03	3.01	2.99	2.96	2.94	2.91	2.88
-0.2	2.86	2.86	2.86	2.85	2.84	2.82	2.80	2.78	2.76	2.74	2.72
0.0	2.66	2.66	2.66	2.65	2.64	2.63	2.61	2.60	2.58	2.57	2.55
0.2	2.46	2.46	2.46	2.45	2.44	2.43	2.42	2.41	2.40	2.39	2.38
0.4	2.26	2.26	2.25	2.24	2.23	2.23	2.22	2.21	2.21	2.20	2.20
0.6	2.06	2.05	2.04	2.03	2.03	2.02	2.02	2.02	2.02	2.02	2.02
0.8	1.85	1.84	1.83	1.82	1.81	1.81	1.82	1.82	1.83	1.84	1.85
1.0	1.63	1.62	1.61	1.60	1.60	1.61	1.62	1.63	1.64	1.66	1.67
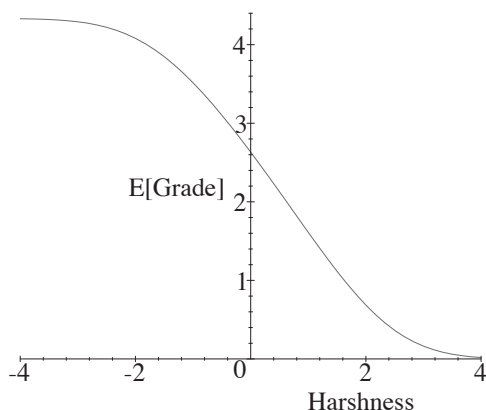
Figure 1. Expected grade on a 4-point scale as a function of instructor harshness h, given that course difficulty $d = 1$.
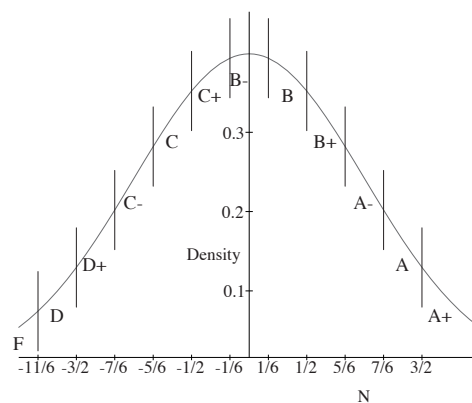

Figure 2. The probability density of the performance variable N. The vertical bars represent the grade ranges for an instructor of zero harshness. For $h > 0$, the ranges shift to the right by h, making it harder to earn a high grade.

Since we have no data on the students, courses, instructors, or grades at ABC College, we generated a random set of students, courses, and instructors. Since we don't know the exact composition of the college, we generated various scenarios.

- We assigned each instructor to teach five courses per year, which is a typical teaching load at many colleges.

- We computed a quality variable q for each student by sampling randomly from a $\mathrm{normal}(0, \sigma_q)$ distribution.

- We assigned each student to eight courses per year (also a standard load for many colleges), for either one or two years, using a uniform probability of selecting each course.

- We generated course aptitude adjustments c for each course that a student enrolled in by sampling from a $\mathrm{normal}(0, \sigma_c)$ distribution.

- We assigned difficulties to each course by sampling from a symmetric beta distribution centered at 1. A typical choice would be $\mathrm{beta}(3, 3)$ on $[0.7, 1.3]$ (see **Figure 3**).

- We assigned harshnesses to instructors by sampling from an asymmetric beta distribution skewed and translated towards leniency (to represent the tendency to inflate grades at the college). A typical choice would be $\mathrm{beta}(2, 3)$ on $[-2, 0]$ (see **Figure 4**). One can use **Table 2** to guide the choice of distribution according to the average GPA that we desire.

We did not restrict ourselves to considering the case where the average GPA at the college is 3.67. In early 1997, Duke University considered revising
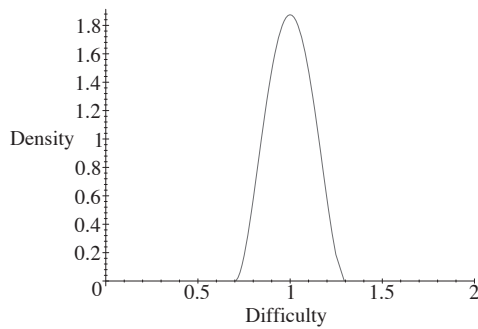
Figure 3. Typical course difficulty distribution: $\mathrm{beta}(3,3)$ on $[0.7, 1.3]$.
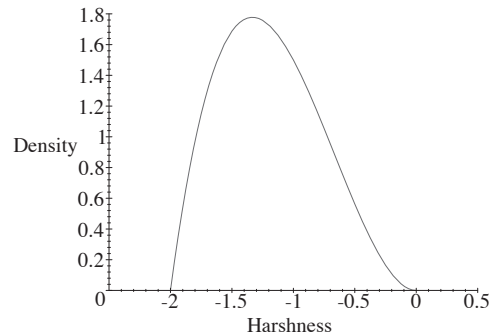

Figure 4. Typical instructor harshness distribution: $\mathrm{beta}(2,3)$ on $[-2, 0]$.

its calculation of GPAs to take into account the difficulty of the course in which a grade was received, the quality of the other students in the course, and the historical grading tendencies of the instructor. The reason for this move was alarm that the average GPA at the University had risen from 2.7 in 1969 to 3.3 by fall 1996 [Gose 1997]. If an average GPA of 3.3 is considered evidence of rampant grade inflation, then 3.3 is a more likely estimate of the average GPA at ABC College than 3.67 is.

Some word on our rationale for choosing distributions is in order here. The normal distribution is a standard choice for representing abilities in a population. Our model of how course difficulty affects student performance and grades loses validity for d outside the range of approximately $[0.5, 2]$, and of course a negative difficulty makes no sense at all. Thus, the normal distribution is not an appropriate choice. We chose a beta distribution because it takes values on a finite interval. Similarly, any harshness value outside the range $[-3, 2]$ is patently ridiculous (see **Figure 1**). In fact, a harshness value of -2 is fairly ridiculous; but to obtain a school-wide average GPA of 3.67, we have to allow that some instructors are that lenient. In any event, it behooves us to choose a distribution over a finite interval. We also desire an asymmetric distribution, to represent the tendency at the college toward lenient grading. Thus, a $\mathrm{beta}(a, b)$ distribution with $a < b$ is appealing for our purposes.

Our model indicates that the ABC College administrators' concerns about being unable to distinguish among the top students are justified. Indeed, when we generate an incoming freshman class of 500 students and make the instructors lenient enough to yield a 3.64 average GPA, 60 of these students *still have a straight A+ average after two years!* In this environment it is clearly necessary to search for a better evaluation method than simple GPA.

The Modified GPA algorithm

Our algorithm for establishing a class rank involves a number of distinct stages. We first attempt to gain additional information from the instructor's historical grade awards and use this information to refine our knowledge of the courses and instructors that the students are taking currently. Using an estimate of the instructor's harshness based on the grades that he has given historically, we correct the grades awarded in a particular course by estimating the mean value by which any leniency or harshness changed a student's letter grade. Incorporating this correction factor allows us to provide an adjusted GPA measure that represents more fully the actual performances and, hence, the quality of the students.

We assume that the instructors have been teaching at the college for at least two years and that we have access to the grades that they assigned during those years. We simulate these data just as we generated the data for the students, as described in section **Modeling College Composition and Grade Assignment**. New students are generated randomly as in the original student data. Given the actual harshness of the instructors and the difficulties of the courses taught, numerical performances and grades for all of each instructor's courses are generated as above.

From these historical data, we compute the average grade $\overline{g}$ granted by each instructor. One can calculate the expected grade granted in a course as a function of difficulty d and harshness h (see **Table 2** and **Figure 1**). For a given value of d, this function decreases monotonically with h, so we can calculate the inverse function. Assuming that $d = 1$ we estimate the instructor's harshness h_{est} by evaluating this inverse function at $\overline{g}$.

Notice that we never even try to estimate the difficulty or take the actual grade distribution into account. Despite this crude method of estimating harshness, we achieve surprisingly good results. Using courses of 40 students each, the harshness that we estimate is usually within about 0.05 of the actual harshness, though it is not too uncommon to err by as much as 0.12. The error tends to decrease the closer the actual harshness is to zero.

We can now adjust the grades of the students in each of an instructor's courses based on our harshness estimate h_{est} for that instructor. For simplicity we assume $d = 1$ for the course. The fact that a student receives a grade g in a course with a instructor of harshness h means that the student's performance value N lies in the interval

$$(l(g) + h,\ r(g) + h] \ .$$

Thus we estimate that N lies in the interval

$$I = (l(g) + h_{\text{est}}, r(g) + h_{\text{est}}] .$$

We estimate N to be the expected value of the distribution of N given that N lies in I. For grades other than A+ and F, this interval has width $\frac{1}{3}$. Assuming

$d = 1$, the a priori distribution of N is standard normal. Given that it lies in I, the probability density function is just the indicator function for I times the standard normal density times a constant factor. Since the density function for the standard normal distribution is almost linear over any interval of width $\frac{1}{3}$, we estimate N as

$$N_{est} = h_{\text{est}} + \frac{l(g) + r(g)}{2}.$$

If the grade is A+ or F, we can calculate the expected value analytically, with the following results:

$$E[N|N > t_0] = \frac{e^{-t_0^2/2}}{\sqrt{2\pi}(1 - \Phi(t_0))} \tag{1}$$

$$E[N|N \leq t_1] = -\frac{e^{-t_1^2/2}}{\sqrt{2\pi}\Phi(t_1)} \tag{2}$$

where Φ is the standard normal cumulative distribution function and the relevant t values are $t_0 = l(A+)$ and $t_1 = r(F)$. So when the student's grade is A+ or F, we set N_{est} to **(1)** or to **(2)**, respectively.

Now that we have a value for N_{est}, we assign the student an adjusted grade for the course. The adjusted grade g_{adj} is the grade that an instructor of zero harshness would have given, except that we assign a real number grade instead of an integer. Specifically,

$$g_{\text{adj}} = 3N_{\text{est}} + 8.$$

To avoid a discontinuity at $N_{\text{est}} = r(F)$, we treated an F as a grade of 2 for this purpose.

Since the grades given by a lenient grader exhibit a smaller spread and hence do not differentiate the students as well as those given by a strict grader, we grant them less import when calculating a student's adjusted GPA. Specifically, the student's GPA is a sum of the grades received weighted by $\sigma(1, h_{\text{est}})$, where h_{est} is the estimated harshness of the instructor who assigned the grade and $\sigma(d, h)$ is the standard deviation of grades given by a instructor of harshness h in a course of difficulty d (see **Figure 5**).

If we wished to refine this method, we might use the spread of grades in a course to estimate its difficulty both when examining the instructors' grading histories and when adjusting grades at the end. However, even without this enhancement, the basic method that we have just outlined performs well, as we demonstrate in **Results of the Model**. One has to be very clever to estimate the difficulty of a course in a way that is numerically stable. The most obvious method is to note that for student i, we have $N_i = q_i d$, then use $N_{i,\text{est}}$ and some estimate $q_{i,\text{est}}$ of q_i that we pull from some other source to estimate

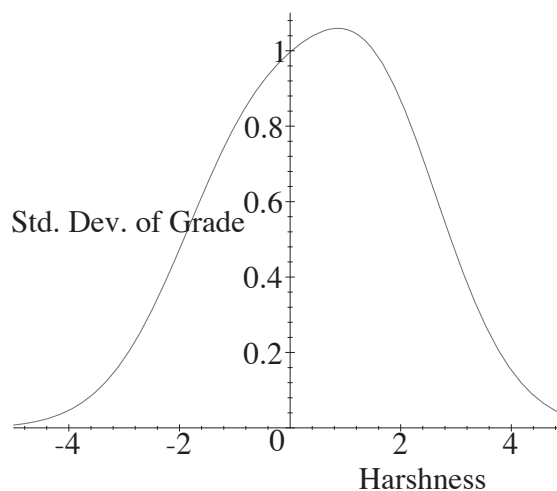$$d \approx \frac{N_{i,est}}{q_{i,est}}.$$

Figure 5. Standard deviation of grade distribution as a function of h, assuming difficulty $d = 1$.

The difficulty is that q_i is likely to be near zero, so the error $|q_i - q_{i,\text{est}}|$ is magnified when we divide.

Results of the Model

We generated a number of scenarios to elucidate the features both of our simulated data and of the modified GPA algorithm. We ran these simulations on a test student population of 500 students, each of whom took 16 courses, with average course size 40.

We use a number of these scenarios to demonstrate the results of our simulations. Plots of actual quality ranks vs. rank by GPAs or adjusted GPAs demonstrate the effectiveness of each ranking method over all tiers of students. For a perfect ranking of students, this plot would lie along the line $y = x$.

We define three error metrics to aid in the comparison between the ranking generated by our revised GPA method and the raw GPA ranking.

- We define a *misassigned scholarship candidate* to be a student who either received a scholarship but was not in the top 10% in quality, or who was in the top 10% of quality but did not receive a scholarship. A simple count of the number of misassigned scholarship candidates measures the method's effectiveness at identifying the highest caliber of student. We refer to this quantity as the MS (Missed Scholarship) metric for a given estimated ranking.

- For each student who is ranked incorrectly, the rank errs by some number of places. Summing these rank errors over all students gives us a measure of how our ranking compares to the actual quality ranking across the entire spectrum of students. We refer to this as the SE (Scaled Error) metric.

- Finally, to determine the injustice with which scholarships are assigned, we sum over all misassigned scholarship candidates the distance between their quality ranking and the scholarship cutoff rank. We refer to this measure of error as the SI (Scholarship Injustice) metric.

The first scenario has difficulty scaled to be between 0.7 and 1.3, while the harshness distribution is relatively lenient, with values ranging between -2.1 and -0.1. The variation due to course material is set to be 0. This yields, as one might expect, a student population with rampant grade inflation. Overall GPA is 3.64, with 60 students receiving perfect A+ averages. **Figure 6** plots GPA rank against actual quality rank, where we observe significant discrepancies between the estimated and actual rankings. At this level of grade inflation, the top tiers of students are almost entirely indistinguishable by GPA. Attempting to correct for harshness by using the corrected GPA does not significantly improve the results at the high ranks. It does, however, improve the SE number from 9,124 to 5,352, representing a superior evaluation of the middle and lower tiers of students (see **Figure 7**).

We now alter the parameters in our model to fit what we feel is a far more realistic situation. Harshness is set to vary between -1.569 and $.431$, yielding a scenario that has an average GPA of 3.34. Now only 38 students have perfect A+ averages. The effect of ranking based on the adjusted GPA is definitely apparent, in **Figures 8–9**. The discrepancies are clearly less for middle-ranked and low-ranked students. The SE measure improves from 6,916 using simple GPAs to 4,842 using adjusted GPAs.

Note the loss of ranking accuracy at high quality levels. The two methods of ranking perform nearly identically, with raw GPA giving an MS of 6 while the adjusted GPA rank gives an MS of 7.

Table 3 summarizes results for a sampling of the scenarios, all of which use the usual range of difficulty from 0.7 to 1.3.

As our simulations show, the modified GPA ranking fares well against the raw GPA method, with SE numbers significantly lower in each trial. This means that the students in the lower deciles are ranked much more accurately in each case. This suggests a certain robustness and indicates that judgments based on the modified GPA rank will be more fair.

As in all other trials performed, both the raw and adjusted GPA ranking methods performed poorly at the high end of the ability curve, according to all three measures.

Strengths and Weaknesses

One weakness of our model is that it does not allow for a completely analytic solution to the scholarship selection problem. Computer simulation is the only means we have to test and evaluate our methods against the simple-minded raw GPA ranking method.
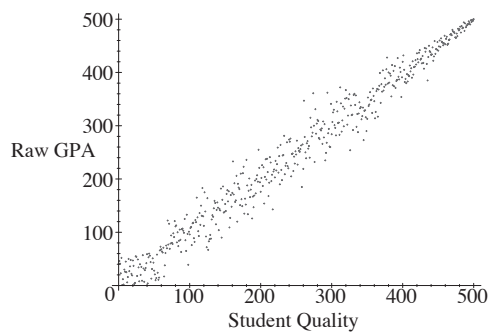
Figure 6. Wild grade inflation resulting in an average GPA of 3.64. The raw GPA estimate makes significant mistakes in the entire range but is especially inaccurate in the top two deciles.
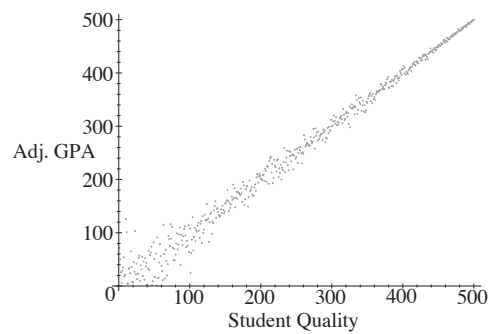

Figure 7. The same scenario as **Figure 6** with rank determined by GPAs modified for instructor harshness.
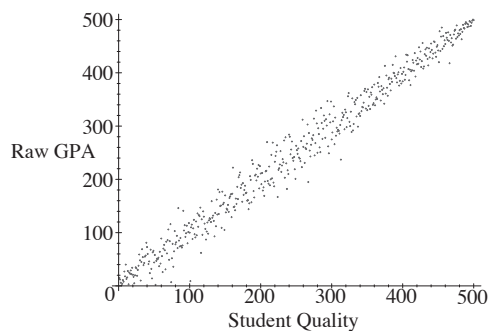
Figure 8. A more reasonable scenario. The raw GPA rank maintains some level of inaccuracy throughout the spectrum of student ability.
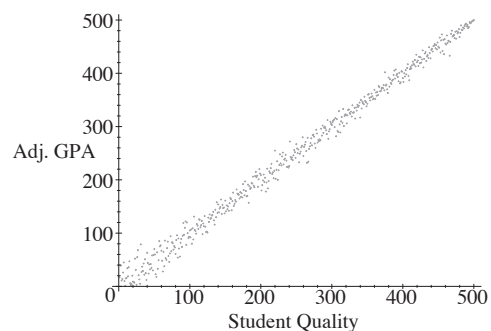
Figure 9. Same scenario as **Figure 8** but with ranks determined by the modified GPAs.

Table 3.

The relevant information for several simulations. Note how the modified GPA ranking produces smaller SE numbers in all cases, representing greater overall accuracy.

Trial	Harsh low	Harsh high	Raw GPA	Raw MS	Adj. MS	Raw SE	Adj. SE	Raw SI	Adj. SI
1	−2.1	−0.1	3.64	9	16	9124	5352	139	118
2	−1.55	0.35	3.22	8	10	7178	3122	247	596
3	−1.569	0.331	3.34	6	7	6916	4842	215	197
4	−1.9	0.1	3.48	8	10	8896	5424	101	226
5	−1.49	0.51	3.19	8	5	7454	4410	153	81

Potentially the greatest weakness in our model and techniques is the lack of a good ranking of the top two deciles. Whether or not a robust method exists is, we believe, debatable. Nothing we have seen indicates that the information required to form a confident ability ranking is even contained in the GPA information we have. It is likely that complete rank-ordering cannot be achieved given the information our model provides. We do not know this to be true, but it is certainly consistent with the results we have witnessed.

Another weakness is that our model cannot take into account the effect of curved grading systems and the possibility of student grades being altered by the performances of the fellow students in the course. Similarly, other interactions between the entities in our model, such as the formation of study groups, can affect the performance (as distinguished from grade) of a student in a course in a way that is dependent upon the other members of the course. Our model also includes parameters, namely, the course difficulties, that are difficult to estimate accurately, and thus remain a complete unknown throughout our attempts to rank based on ability.

In spite of these features, our model has a number of compelling features. By changing just a few parameters, one can generate an entirely new scenario that has a plausible distribution of grades and GPAs. Furthermore, it takes into account the three functional parts of any educational experience: the students, the instructors, and the courses. Arguably, no model could be complete without accounting for variations of each of the three parts.

Despite all of our problems in classifying the scholarship winners, the adjusted GPA method we use is almost uncannily good at identifying the lower deciles, which in a real context is important to the students and the school.

From a practical standpoint, our model and methods are fairly simple to implement. The number and size of the calculations performed is linear in the size of the student body and could be executed with modest computer resources at even a large institution.

To sum up, it would behoove ABC College to use our ranking system, since it more accurately identifies the bottom eight deciles of student ability. However, if the administration seeks to accurately rank the top tier of students, it must realize that a bloated aggregate GPA from excessively lenient grading can quickly lead to a situation where no amount of calculations and statistics can recover the desired information about the intrinsic quality of the students.

References

Gose, Ben. 1997. Duke rejects a controversial plan to revise the calculation of grade-point averages. *Chronicle of Higher Education* 43 (21 March 1997): A53.

Grade Inflation: A Systematic Approach to Fair Achievement Indexing

Amanda M. Richardson
Jeff P. Fay
Matthew Galati
Stetson University
421 N. Woodland Blvd.
Deland, FL 32720
`vgalati@bellatlantic.net, jfay@steton.edu`

Advisor: Erich Friedman

Background

Constantly rising grade-point averages at universities across the nation have made it increasingly difficult to distinguish between "excellent" and "average" students. For example, *The Chronicle of Higher Education* found that the mean grade-point average (GPA) at Duke University was 3.3 in 1997, up from the 1969 mean of 2.7. It also found that Duke is not alone in this trend.

Average grades have consistently increased while the system of measurement has remained unchanged. Receiving an A in a course does not necessarily denote exceptional performance, since the percentage of students receiving As has increased dramatically over the last few decades. In 1995, the *Yale Daily News* reported that As and Bs constituted 80% of grades at Yale. According to the *New York Times* (4 June 1994), nearly 90% of grades at Stanford were As or Bs, and an estimated 43% of grades at Harvard and 40% at Princeton were As. This situation has led many universities to seek new methods for ranking student performance.

Some say that expectations and grading difficulty have dropped from the faculty point of view, while others argue that the quality of student has been on the rise. Whatever the cause, a major problem arises when scholarship foundations or graduate schools try to distinguish exactly who deserves to be

in the top 10%, etc. For this reason, an alteration of the current quantitative ranking system is necessary.

One approach is a system of *quality points* based on comparative performance within each course. In this system, a student's quality points for a given course can be calculated based on performance relative to students' overall performance in the course. To obtain this objective, overall course performance may be measured by the mean or by the median of all the grades in the course. Then the student's individual performance can be measured in terms of standard deviations from the mean/median.

Many schools are looking for a feasible system to reach this goal. For example, in 1994 the faculty at Dartmouth voted to include on a student's transcript the course size and median course grade next to each grade, plus a summary telling in how many courses the student surpassed the median, met the median, or performed lower than the median (*Boston Globe*, 4 June 1994). This year, Duke University's faculty considered (but did not implement) using an "achievement index" (AI) to rank students. The factors considered in this index would include the course's difficulty level, the grades received by all the other students taking the course, and the grades those students received in other courses.

Most arguments against this type of indexed system appear qualitative rather than quantitative. For example, in an article in *The Chronicle* at Duke, Prof. Robert Erickson explains that "the reason [the AI system] won't work is that the faculty will not agree to have their grading system tinkered with." Many faculty fear that such a system may put their students at a disadvantage until such a system became more widely used, and hence students may seek to attend other schools. No one wants to be the guinea pig.

A quantitative question that arises when determining the best index is, when determining "average" performance in a course, should the index use the course's mean grade or its median grade? Some argue for the median, since it is more robust, while others argue for the mean, since it is a better estimator when the distribution is close to a normal distribution.

Our model attempts to find a solution to the problem of ranking.

Assumptions

- The sample data that we made up, with grades for 68 students, effectively represents the entire population of 2,000 students at ABC College.

- Past performance has no effect on performance in any given course.

- Results from one semester can be extrapolated to a ranking system cumulative over semesters.

- Course size and difficulty level do not change the model's effectiveness.

- The system is implemented at the administrative level, after standard grades are reported by professors.

- In comparing a plus-minus grading system (including A+, A, A−, B+, etc.) to one with only straight letter grades (A, B, C, etc.), the latter's A would encompass the former's A+, A, and A−.

- The model may compare students at large universities with those at small colleges without loss of generality.

Motivation for the Model

The goal is to find a way to order, or compare, students with only slightly varying grades; so an index rating students' performance relative to other students is germane. To arrive at this end, we must first determine how students in a particular course performed overall. Assuming that we have determined which estimator (mean or median) to use, we measure a student's standing by how many standard deviations from the given estimator the student's grade lies. Thus, if the student's grade is 2 standard deviations above the estimator, the student receives 2 quality points; if the student's grade equals the mean/median, the student receives 0 points; and if the student's grade falls below the estimator, say by 1.2 standard deviations, negative points are issued (−1.2).

The Model

We now determine the best index to use in implementing such a system. The question of using mean or median as our standard of comparison is of utmost importance. Perhaps instead of choosing one or the other for all courses, we should determine which is more effective for a particular course. The mean is the preferable estimator for data resembling a normal distribution (**Figure 1**) but is more sensitive to outlying data than the median when the distribution is strongly skewed. Thus, the skewness of the distribution of grades in a given course can help to determine which estimator to use for grade comparison in that particular course.

Through trial and error, we decided that if a course's distribution has skewness greater than 0.2, we would use the median as the estimator. If the skewness is smaller than that, the distribution is sufficiently close to normal that we can use the mean as the estimator.

In our indexing system, after determining the most appropriate estimator for comparison, we define a student's relative performance in terms of standard deviations. The quality points awarded for a given course are then weighted by the courses' credit hours. Finally, a student's overall index is computed by summing total points and dividing by total credit hours.
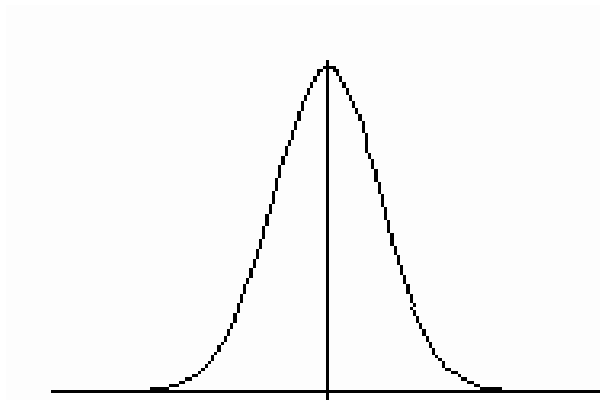
Figure 1. A normal distribution.

Let G_i be the student grade for course i, $G_i \in [0, \text{MaxGrade}]$, where course i has C_i credit hours. Then

$$\text{GPA} = \frac{\sum G_i C_i}{\sum C_i},$$

where the sums are taken from $i = 1$ to the number n of courses taken by the student.

Our procedure is as follows. Let $A_i = a_0, a_1, a_2, \dots, a_{n_i}$ be all of the n grades for course i. Let μ be the mean of A_i, χ its median, and σ its standard deviation. The skewness S is defined as the third moment about the mean divided by σ^3:

$$S = \frac{\sum (a_i - \mu)^3}{\sigma^3},$$

where the sum is over all n_i students in the course.

If $|S| < 0.2$, we use μ as the estimator; if $|S| \geq 0.2$, we use χ as the estimator. We define

$$\text{Index} = \begin{cases} \dfrac{\sum \dfrac{G_i - \mu}{\sigma} \cdot C_i}{\sum C_i}, & \text{if } |S| < 0.2; \\[4ex] \dfrac{\sum \dfrac{G_i - \chi}{\sigma} \cdot C_i}{\sum C_i}, & \text{if } |S| \geq 0.2. \end{cases}$$

Analysis

First, we offer some justification of our assumptions about course size and difficulty level.

Since the model seeks a comparative ranking system, the number of students in a course is not the issue; rather, each student's performance relative to each other student in the course is the important factor. Hence, course size is not directly involved in the model.

The difficulty level of a course, although not explicitly dealt with in determining a student's index rating, is accounted for indirectly in our model, in the overall course performance. For example, one would suspect that a relatively easy course would tend to have a large percentage of high grades, while grades in a more challenging course might be more evenly distributed or even tend toward lower grades. While the standard grading system would rate students purely on the grade they received, and thus not take difficulty level into account, the comparative nature of our indexing system causes inherent dependence on this factor.

Sensitivity

In the analysis of our data, we have seen a number of cases where the system used had a great effect on the ranking of particular students. We offer three students as examples of what can occur in our minicollege of 68 students. [EDITOR'S NOTE: We do not reproduce here the authors' complete table of grades, indexes, and ranks under the various systems.]

Student 603 is ranked 15th by GPA with plus/minus grades (**Figure 2**), with grades of B, B, A, A. This student's rank would be the same for GPA with straight letter grades. However, using our index ranking system, Student 603's rank drops dramatically. With plus/minus grades, the student's rank falls three deciles from the 30th percentile to the 60th percentile; that is, the student's rank drops from 15th to 43rd. With straight letter grades, the drop is even more drastic: The student falls to 5th and plummets to the 70th percentile.
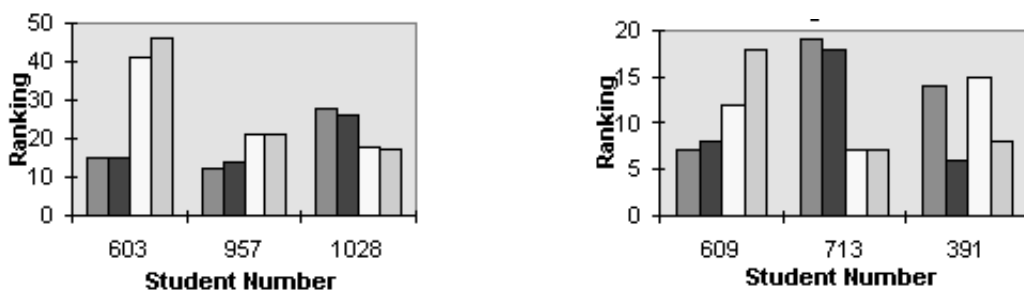
Figure 2. Rankings of some students under various systems. The leftmost bar is for GPA with plus/minus grades; the bar second from the left is for GPA with straight letter grades; the bar second from the right is for our index system with plus/minus grades; and the rightmost bar is for our index system with straight letter grades.

To understand better how such an event could occur, we take a closer look at the courses that Student 603 took. In the first two courses, the student received a B. However, the value of that B is what is under consideration in our model. The grades in the first course are 16 As, 13 Bs, and 1 C. Due to the large skewness coefficient, our formula uses the median, which is an A. So the student's grade is actually "below average." The second course has an

average of an A−; here too, a B is a relatively low grade. In both of the other two courses, the mean/median grade is an A, and Student 603's As in these courses represent average performance in them. Our ranking system takes the average of the deviations from the estimator to find an index for the student. Specifically, the student has deviated from the estimator by a factor of 0.6387.

Student 957 is ranked 12th by GPA with plus/minus grades, with grades A, A, A−, B+, A, A, A, B+, A, B, A. In the first six courses, the lowest overall grade out of all of the students is a C, and the mean/median for these courses is A, A, A, B+, A, and A. It is obvious from these statistics that the student's ranking has been overestimated. Under our system, Student 957 is ranked 23rd and drops from the 20th percentile to the 40th percentile. Although this is a dramatic drop, it is not as drastic as for Student 603. Looking more carefully at the other courses, we focus on the tenth course. Here, the mean/median is C+, while Student 957 received a B.

Similarly, Student 1028 is ranked 28th by GPA. However, when his grades are compared with those in his respective courses, his ranking drops to 22nd. Once again, the strong level of grades throughout his courses lowers the "value" of the grades that he received.

What about students who fall out of the top 10% under the new rating system? The new system is supposed to determine better which students are worthy of scholarship and advancement, so this distinction is key.

Student 609 is ranked 7th in GPA with plus/minus grades and 8th with straight letter grades, with grades A, A, B+, A. However, due to the relatively high average grades in these courses, this student suffers a loss in ranking under our system, to 12th with plus/minus grades. This drop causes Student 609 to fall out of the first decile.

Student 713 benefits from our system. This student's grades are B, A−, A, B, and B, but the mean/median for each course is below the student's respective grades. Due to this, the student's ranking rises from 19th to 8th. As in the case for Student 609, the new system of ranking has a large effect on the awarding of scholarship.

We need to ensure that our model is not too susceptible to a single grade change in a single course. How does such a change affect the overall ranking of this person and the overall rankings of other students? Since students' rankings are dependent on other students' grades, a single change could affect other students' rankings across the board.

The most extreme scenario is when the course size is small and the grades are skewed. For example, let's take the case of a course with two students. Suppose that student X and student Y both receive As, so the mean/median is an A. If student X's grade changes to an F, the mean/median becomes a C. This change increases student Y's index, since student Y is no longer "average" in this course but above average. This change in student Y's index could potentially alter the rank of a few students. This scenario is the most extreme case but would also be extremely rare.

With a much larger course, a similar grade change would have minimal

effect. Suppose that a course has 20 students with a distribution highly skewed towards the high end. The grades could be as follows:

A+, A+, A, A, A, A, A, A−, A−, A−, A−, A−, B+, B, B, B, B, C+, C, F.

We use the median (A−) instead of the mean. Suppose that a student who received an A+ should have received an F. Once the correction is made, the grades would still be skewed and we would still use the median, which would still be approximately A−. Thus, the only person who would be affected would be the one student whose grade was corrected.

In certain circumstances, a grade change could potentially change the median of a skewed course. Since the grades in the course are skewed, the median will not change by much and thus the change will not constitute a major problem. A few students may move ranks, and in some cases, deciles.

If the distribution is more normal, then we use the mean; and as long as the course size is relatively large, the effect of a grade change is minimal.

Strengths

Implementing our system would involve just the introduction of a computer program that would use data already in the current system in the registrar's office.

Since our index system is implemented at the administrative level, professors would not have to alter their methods of grading at all. The index system is merely a new method of interpreting the grades currently issued by professors.

Another strength of the index model relates to the issue of grade inflation itself. One problem with grade inflation is that it may not be universal. In other words, certain departments or colleges may be more or less affected by grade inflation. Thus, for employers or graduate institutions seeking the best candidates from a wide variety of undergraduate institutions, our model using the index system takes away the problem of comparing universities with varying levels of grade inflation.

Weaknesses

One possible flaw with our index model is the lack of consideration of the quality of student in a given course. For example, consider courses X and Y in which all students earn a letter grade of an A. Our model gives each student the same quality points. Perhaps consideration should be given to the quality of students taking a given course. If all the students in course X also received As in their other courses, while students in class Y had a wide range of grades in their other courses, then performing at the "average" level in course X is

theoretically more difficult than in course Y, so awarding equal points does not effectively differentiate as we would desire.

Another uncertainty may arise with courses with multiple sections. It may happen that higher-level students all take a certain section of the course, thus making the comparisons of course performance invalid. For this reason, larger universities especially may need to group all sections of a given course before computing the mean/median and calculating the comparative index to rate each student's performance in that course.

Another weakness is our trial-and-error choice of 0.2 as the value of skewness that determines which formula to use for the index. Our data are limited to fairly small course sizes, and different values may be work better for larger course sizes. In rare cases when only one student is enrolled in a course (say, for a senior project or independent study), the student's grade would determine the mean/median and thus would always equal the estimator, resulting in the issuance of zero quality points. Thus, a student can never be rewarded for doing well or hurt by doing poorly in such a situation.

Future Models

More research into the effects of different skewness factors is necessary before implementation of such an index system. Also, including a method of evaluating the quality of students in a given course would further help the comparative ranking idea essential to this model. Looking at each student's grades outside a given course may help determine the quality of student. Thus, if a course is full of straight-A students and a particular student performs "above average" in that course according to our index system, that student should be awarded higher points than a student performing "above average" in a course full of students who have lower grades outside that course.

Also, further investigation into the choice of mean and median may yield more effective determination of which is the better estimator for a given course or perhaps for a particular university as a whole. Consideration may also be given to the different effects an index system has in large universities compared to smaller colleges and private universities.

Judge's Commentary:
The Outstanding Grade Inflation Papers

Daniel Zwillinger
Waltham, MA 02453
`zwillinger@alum.mit.edu`

Grade point average (GPA) is the most widely used summary of undergraduate student performance. Unfortunately, combining student grades using simple averaging to obtain a GPA score results in systematic biases against students enrolled in more rigorous curricula and/or taking more courses. Here is an example [Larkey and Caulkins 1992] of four students (call them I–IV) in which Student I always obtains the best grade in every course that she takes and Student IV always obtains the worse grade in every course that he takes, yet Student I has a lower GPA than Student IV does:

	Student I	Student II	Student III	Student IV	Course GPA
Course 1	B+			B−	3.00
Course 2	C+		C		2.15
Course 3			A	B+	3.65
Course 4	C−	D			1.35
Course 5		A		A−	3.85
Course 6	B+			B	3.15
Course 7		B+	B		3.15
Course 8	B+	B	B−	C+	2.83
Course 9		B	B−		2.85
Student GPA	2.78	2.86	2.88	3.0	

The MCM problem was to determine a "better" ranking than one using pure GPAs; this problem has no simple "solution." Johnson [1997] refers to many studies of this topic and suggests a technique that was considered—but not accepted—by the faculty at Duke University.

Each participating team is to be commended for its efforts in tackling this problem. As in any open-ended mathematical modeling problem, there is not only great latitude for innovative solution techniques but also the risk of finding

关注数学模型

获取更多资讯

no results valuable in supporting one's thesis. Solutions submitted contained a wide variety of approaches, including graph theory and fuzzy logic.

Unfortunately, several teams were confused as to the exact problem that the dean wanted solved. Assigning students to deciles, by itself, was not the problem; for example, deciles could be assigned from any list of student names by choosing the first 10% of the students to be in the first decile, etc. The dean wanted meaningful deciles, based on students' relative course performance. Simply re-scaling GPAs so that the average became lower (and the top 10% became more spread out) would not change the inherent problem.

The problem statement suggested that relative rankings of students within courses should be used to evaluate student performance. With this assumption, possible approaches include:

- using relative ranking *with* grade information

  (A useful additional assumption might be that faculty would give grades based on an absolute concept of what constitutes mastery of a course.)

- using relative ranking *without* grade information

In the latter approach (chosen by most teams), an instructor who assigns As to all students in a course provides exactly the same information as an instructor who assigns all Cs to the same students in another course.

Specific items that the judges looked for in the papers included:

- Reference to ranking problems in other fields that use relative performance results, such as chess and golf.

- A detailed worked-out example, illustrating the method(s) proposed, even if there were only 4 students in the example.

- Computational results (when appropriate) and proper consideration of large datasets. Teams that used only a small sample in their computational analysis (say 20 students) did not appreciate many of the difficulties with implementing a grade adjustment technique.

- Mention (if not use) of the fact that even though the GPA may not be as "good" a discriminator as the various solutions obtained by the teams, it seems reasonable that there be some correlation between the two.

- A response indicating understanding of the question about the changing of an individual student's grade. Such a grade change could affect that student's ranking, but if it affected many other students' ranks then the model is probably unstable.

- A clear, concise, complete, and meaningful list of assumptions. Needed assumptions included:

- The average grade was A−. (This must be assumed, as it was stated in the problem statement—amazingly, several teams assumed other starting averages!)
- in an {A+, A, A−, ...} system, not *all* grades were A−. (Otherwise, there is no hope for distinguishing student performance.)

Many teams confused assumptions with the results that they were trying to obtain. Teams also made assumptions that were not used in their solution, were naive, or needed further justification. For example,

- Many teams assumed a continuous distribution of grades. As an approximation of a discrete distribution, this is fine. However, several teams allowed grades higher than A+, and other teams neglected to convert to a discrete one when actually simulating grades.
- Several teams assumed that teachers routinely turn in a percentage score or course ranking with each letter grade. This, of course, would be very useful information but is not realistic.
- Low grades in a course do not necessarily imply that a course is difficult. A course could be scheduled only for students who at "at risk." Likewise, a listing of faculty grading does not necessarily allow "tough" graders to be identified: an instructor may teach only "at risk" students.

- The most straightforward approaches to solving this problem were:

  - Use of information about how a specific student in a course compared to the statistics of a course. For example, "Student 1's grade was 1.2 standard deviations above the mean, Student 2's grade was equal to the mean, ..." The numbers {1.2, 0, ...} can be used to construct a ranking.
  - Use of information about how a specific student in a course compared to other specific students. For example, "in Course 1, Student 1 was better than Student 2, Student 1 was better than Student 3, ..." This information can be used to construct a ranking.

The judges rewarded mention of these techniques, even if other techniques were pursued.

- Other features of an outstanding paper included:

  - Clear presentation throughout
  - Concise abstract with major results stated specifically
  - A section devoted to answering the specific questions raised in the problem statement or stating why answers could not be given.
  - Some mention of whether the data available (i.e., sophomores being ranked with only two years worth of data) would lead to statistically valid conclusions.

None of the papers had all of the components mentioned above, but the outstanding papers had many of these features. Specific pluses of the outstanding papers included:

- Duke team

    – Their summary was exemplary. By reading the summary, you could tell what they were proposing and why, what the issues they saw were, and what the models they produced were.

    – Their use of least squares to solve an overdetermined set of equations was innovative.

    – Their figures of raw and "adjusted" GPAs clearly and visually showed the correlation between the two and also the amount of "error" caused by exclusive use of GPAs.

- Harvey Mudd team

    – Their sections on "Practical Considerations" and "What Characterizes a Good Evaluation Method" demonstrated a clear understanding of the problem.

    – Their figures of "Raw GPA" versus "Student Quality" clearly and visually showed the correlation between the two and also the amount of "error" caused by exclusive use of GPAs.

- Stetson team

    – Their use of the median, as well the mean, in comparing a specific student to the statistics of a course was innovative. (Use of the median reduces the effects of outliers.)

    – They interpreted the results of rank adjustment for specific individuals in their sample.

    – An awareness of the problem as indicated by literature references in their "Background Information" section.

Reference

Johnson, Valen E. 1997. An alternative to traditional GPA for evaluating student performance. *Statistical Science* 12 (4): 251–278.

Larkey, P., and J. Caulkins. 1982. Incentives to fail. Working Paper 92–51. Pittsburgh, PA: Heinz School of Public Policy and Management, Carnegie Mellon University.

About the Author

Daniel Zwillinger attended MIT and Caltech, where he obtained a Ph.D. in applied mathematics. He taught at Rensselaer Polytechnic Institute, worked in industry (Sandia Labs, Jet Propulsion Lab, Exxon, IDA, Mitre, BBN), and has been managing a consulting group for the last five years. He has worked in many areas of applied mathematics (signal processing, image processing, communications, and statistics) and is the author of several reference books.

Practitioner's Commentary: The Outstanding Grade Inflation Papers

Valen E. Johnson
Institute of Statistics and Decision Sciences
Duke University
Box 90251
Durham, NC   27708-0251
`valen@isds.duke.edu`

Introduction

I would like to begin my comments by congratulating all three teams on their innovative solutions to what is both a difficult and important societal problem. The depth of thought given to this problem in the short time each team had to generate a solution is very impressive, and all teams present solutions that are surprisingly close to proposals that have appeared in the educational research literature. In fact, the three proposed solutions span the range of previously proposed alternatives to GPAs in terms of model complexity, ranging from relatively simple to highly complex. In the discussion that follows, I focus on the problems associated with each proposal rather than their strengths. I chose this course not because the proposals are weak but instead because they are of sufficient quality to merit serious criticism.

As all three teams note, GPA plays an important role in our educational system. In the particular scenario presented, an adjusted GPA is needed to more equitably allocate scholarships to students at ABC College. More generally, however, GPA is arguably the single most important summary of a student's academic performance while in college. It plays a critical role in determining the success of a student in the job market and is influential in determining whether or not a student is admitted to professional or graduate school.

A more subtle influence of GPA is the impact that GPA has on student course selection. Because GPA is perceived to play such a critical role in a student's career, it is now common for students to select courses based on their expectations of how courses will be graded. In a recent survey of Duke University

undergraduates, 69% of participating students indicated that expected grading policy had some influence in their decision to enroll in courses taken to satisfy a distributional requirement! This fact suggests that fewer "hard" courses are taken by undergraduates as a result of differential grading policies and probably causes a net decrease in the number of science and mathematics courses taken. To a large extent, it also explains the spiraling assignment of grades that has taken place over the last decade. Students gravitate towards courses that are graded leniently, and professors soften grading standards to ensure adequate course enrollments and favorable course evaluations.

Changing the way GPA is computed can solve all of these problems, but implementing such a change is a difficult proposition. Any change to the current GPA system will be opposed by faculty and students who do not benefit from the change, and every meaningful modification of the GPA will produce a sizable proportion of each. For this reason, it is crucial that any modification to the current GPA system be both logically consistent and fair. Additionally, any change to the GPA must be understandable by nonstatisticians, at least at a rudimentary level. Simplicity is a benefit. Unfortunately, the fact that essentially all American universities report traditional GPA attests to the fact that no alternative is available that is

- simple,

- fair, and

- logically consistent.

Compromises in one or more of these three criteria are therefore inevitable. In evaluating the proposed solutions to ABC College's grade inflation problem, I will attempt to consider all three of these criteria and indicate the extent to which I feel each is compromised.

Stetson University

The proposal by the team from Stetson University represents the simplest solution to the problem. Their proposal essentially is to standardize grades in each class using either the median or mean grade, depending on the value of the coefficient of skewness.[1]

From a statistical standpoint, the use of the median in the standardization offers robustness to outliers, or extreme observations. For grade data, extreme observations are usually Ds, and Fs. One or two Fs in a class can significantly affect the mean grade assigned but does not affect the median grade any more than one or two "below average" grades would.

---

[1]The formula provided for the standardization appears to have a minor error. As presented, each student's grades would always be standardized by either the mean grade or median grade. Presumably, it was the intent of this group to use either the mean or median grade on a course-by-course basis.

Unfortunately, the use of the median grade in the standardization process introduces several potential difficulties.

- First, for highly discrete data (i.e., data taking on only two or three values) the median value can be very uninformative. As an illustration of how bad the median can be, consider the problem of estimating the probability of success for a new cancer treatment. If cures are coded as 1s and deaths as 0s, the median estimate for the probability of a cure in a clinical trial conducted with an odd number of patients must be either 0 or 1! A similar difficulty arises when analyzing student grades when only two or three unique grades are assigned. When grades are inflated, the median grade is likely to be either an A or A−, but this says little about the relative proportions of As and A−s that were awarded, and less still about the proportion of Bs, Cs, Ds and Fs.

- Next, if skewness and outliers are considered problematic, then should not a robust estimate of the variance also be used? The Stetson team uses the sample variance to standardize the grades, but this estimate of the spread of a distribution is more sensitive to outliers than is the sample mean when estimating location. As an alternative to the sample variance, I would recommend that the interquartile range (IQR) be used as a measure of distributional spread when the median is used as a measure of centrality. Like the median, the IQR is robust to outliers; for normally distributed data, it is nominally equal to 1.35 standard deviations.

A compromise between the median (which is robust against outliers) and the mean (which is often statistically efficient) would be a *trimmed mean*. To compute a trimmed mean, a fixed proportion of the extreme values of the data is ignored. Thus, to compute the 10% trimmed mean, the lowest and highest 5% of grades are thrown out and the mean of the remaining data is computed. Trimmed means have the advantage of offering some robustness against outliers while at the same time maintaining good statistical efficiency. Use of the trimmed mean in the Stetson team's standardization procedure might also eliminate the problem of deciding when to switch between mean and median, which in itself introduces some potentially large jumps in the standardized values for small changes in skewness.

Although the Stetson team's proposal wins in terms of simplicity, it is somewhat weaker in terms of fairness and consistency.

- In my opinion, fairness is compromised because the standardization procedure does not account for the quality of students within a class, as the team members themselves comment. At my institution (Duke University), there are many courses known to be populated by top students, and if implemented, this proposal would encourage students to opt out of these courses. Accounting for the quality of students in a course is an important facet of GPA adjustment, and I would hesitate to recommend any method that didn't account for this aspect of classroom grading. Implementing such a method

would encourage students to register for lower-level classes with less talented students.

- Consistency is also a problem. To see why, consider two students who take identical courses through their senior years, and receive A+ in all of their courses. In the last semester of their senior year, the second student develops an interest in art history and takes an introductory course in that subject (in addition to the other courses that both he and the first student take). Both students again receive A+s in all of their courses, but unfortunately everyone in the art history course also receives an A+. Which student graduates on top?

  According to the Stetson team's adjustment method, the first student beats out the second student for valedictorian, even though the second student tied the first student in all of the courses they took together and got an A+ in the one course he took above their normal course load. Why? The standardized grade for an A+ in the art history course is 0, which when averaged into the other A+s the student received would lower his adjusted GPA. It is interesting to note that the same problem exists if the art history course is replaced with an independent study course, though in that case it is not clear how the estimate of the standard deviation would be determined.

Duke University[2]

The team from Duke University discusses three proposals for adjusting GPA, the first of which is a nonrobust version of the standardization scheme proposed by the Stetson team. Their other proposals are based on regression-type adjustments to traditional GPA. In their iterated GPA adjustment, the grades from each class are adjusted for the difference between the mean grade assigned in a class and the mean adjusted GPA of students in that class. This difference is then used to compute new adjusted GPAs, which led to new adjustments to the class grades. The team's least-squares estimate of the adjusted GPA is based on the assumption that students tend to receive higher grades in classes taken in their academic majors. As I understand this proposal, they estimate the adjustments for each combination of major and course, conditionally on observed grades.

Both adjustment schemes are quite similar to an adjusted GPA proposed in a more formal framework by Caulkin et al. [1996], though the "least-squares" proposal is also similar to the pairwise course/department differences estimated by Strenta and Elliot [1987], Goldman and Widawski [1976], and Goldman et al.

---

[2]EDITOR'S NOTE: When he prepared his comments, Prof. Johnson was not aware that a paper from a Duke University team was among the papers that he reviewed. He is not affiliated with the Duke University Mathematics Dept. nor with any members of the Duke team. The copies of the Outstanding papers that he read were the same as those read by the judges; they contained code numbers but no identification of the papers' authors or their institutions.

[1974]. For comparison, the models proposed by Caulkin et al. [1996] have the general form

$$\text{Grade}_{ij} = \text{True GPA}_i + \text{Course effect}_j + \epsilon_{ij}.$$

Under this model, the grade received by student i in course j is assumed to be an additive function of their "true GPA" plus a course effect for the jth course, plus a (normally distributed) random error. Like the Duke team, Caulkin et al. [1996] also propose an iterative procedure for estimating all student's "true GPA" along with all course effects.

From a technical standpoint, I regard this modeling approach as a significant improvement on simple standardization schemes. This approach implicitly accounts for both the grading policies of individual instructors and the quality of students within each class. Algorithmically, such models are comparatively simple to estimate and can be computed in less than two minutes on a PC for datasets containing 12,000 students and 17,000 classes.

The primary drawback of these regression-type models is the assumption that grades are intervally scaled. In other words, it usually does not make sense to assume that the difference between an A and A− is the same as the difference between a C and C−, or that a D added to a B is equal to an A. Typical grading scales assign more probability to As and Bs than to Cs and Ds, and by not taking the ordinal nature of grade data[3] into account, substantial statistical efficiency is lost. These models also suffer from the paradox presented above for standardized GPAs but to a lesser extent. The art history course would lower the second student's "true GPA," but an independent study course would leave it unaffected.

Substantively, I have several minor objections to the modeling assumptions made by the Duke team. Perhaps most important, they premise their model on the assumption that "it is possible to assign a single number, or 'ability score,' to each student, which indicates their relative scholastic ability, and in particular, their worthiness of the scholarship." A similar assumption is made by the team from Harvey Mudd College. In fact, this assumption is not necessary or appropriate. Each student's true GPA can instead be interpreted as the ability score for the student *in courses that that student chose to take.* Some courses are required by the university to satisfy distributional requirements, but most will be courses that students choose to take in their areas of interest and competence. The model proposed by Caulkin et al. [1996] and the variation of it proposed by the Duke team can be applied without difficulty to colleges in which, say, humanities students are completely separated from engineering students in the sense that they have no common classes. In such cases, "true GPA" corresponds to the ability of students in the classes that they took. The grades assigned to engineering students in engineering classes should not be used to estimate the abilities of engineering students in humanities classes.

I also feel that it is important to distinguish difficult courses from courses that are graded stringently. They are not (always) the same, so it does not follow

---

[3]Ordinal data are data that are recorded in ordered categories.

that students should be penalized for taking courses that are graded leniently. In fact, I would argue that any student who receives the highest grade in all classes that she takes should be awarded a high adjusted GPA.

As a final comment on this proposal, I think it is dangerous to devise a ranking algorithm that rewards students for the type of curriculum chosen. For some students and some majors, a "well-rounded" courseload is desirable. For others, a more concentrated curriculum is apropos. A student who has satisfied the relevant distributional requirements for their university should be free to choose whatever courses they wish—without penalty. Indeed, I lament the fact that American mathematics undergraduates are often ill-prepared for graduate studies in statistics programs because they have taken so few mathematics courses.

Harvey Mudd College

The proposal by the team from Harvey Mudd College is also very interesting. Though there are several technical problems in their model specification, the paradigm proposed by this team is surprisingly close to a statistical model called the *Graded Response Model* (GRM). GRMs are normally introduced in advanced graduate-level statistics courses, and I was impressed by the extent to which this team exposed the underlying assumptions of these models.

As suggested by the Harvey Mudd College team, the basic assumption of a GRM is that instructors choose thresholds on an underlying achievement scale and assign grades to students based on the grade intervals into which their classroom achievement is observed to fall. Letting z_{ij} denote the classroom performance of student i in class j, and $\gamma_F^j, \gamma_D^j, \ldots, \gamma_B^j$ the upper cutoffs for each grade on the ability scale, the GRM assumes that student i receives a grade of, say, C in class j if

$$\gamma_D < z_{ij} < \gamma_C.$$

A further assumption of the model is that the mean ability of student i, say z_i, for the courses that student i chooses to take, is related to his performance in class j according to

$$z_{ij} = z_i + \epsilon_{ij}.$$

Here, ϵ_{ij} denotes a random deviation. In the GRM, grade cutoff vectors, mean student achievements, and the distribution of the error terms are estimated jointly. Details of this model are discussed in further in Johnson [1997].

In terms of the GRM, the Harvey Mudd team's q_i is roughly equivalent to z_i, while c_{ij} is comparable to ϵ_{ij}, and d_j plays a role similar to the variance of the distribution of the error term ϵ_{ij} if this distribution is assumed to be the same for all students in a given class. The primary difference between the proposed model and the GRM is that the grade-cutoff vectors γ^j are fixed in the former and estimated in the latter. Although the team takes a reasonable approach toward fixing these cutoffs, doing so leads to several inconsistencies in the

resulting model. Partially to overcome these difficulties, the team introduces a harshness term h_k. This harshness term models a uniform shift of the grade-cutoff values for all classes taught by professor k. The team estimates values of h_k from the mean grades assigned by each professor.

The most important technical defect in this model is caused by the assumption that the grade-cutoffs for each class can be obtained by a contraction and shift of the baseline cutoffs. When harshness is 0, it follows that large increases in d_j result in more extreme grades (that is, more As and Fs) and fewer middle grades. Shifts in harshness can change the As to Bs or Cs, but there is still a gap between the high and low marks. This is not typical of the grading patterns observed in actual grade data. To accommodate the distribution of grades actually observed, it is usually necessary to adjust the relative widths of the intervals associated with each grade on a class-by-class basis.

By estimating the grade-cutoffs separately for each class, several of the more controversial assumptions made by this team regarding the properties of undergraduate grades can be eliminated. For example, if grade-cutoffs are estimated individually for each class, it is not necessary to assume that one letter grade corresponds to one standard deviation in student achievement, or that professors do not grade on curves or compare performances of students within classes, or that professors uniformly adjust for course difficulty.

Other questionable model assumptions include the statements that

- students select courses randomly,

- students do not gravitate to courses that are graded leniently, and

- professors have accurate perceptions of student achievements.

None of these assumptions is required for the GRM; by liberalizing the interpretation of this model's parameters, they could be eliminated here as well.

Conclusion

Of the three models proposed, only the model proposed by the team from Harvey Mudd College seems to handle the two-student paradox mentioned above. Their model also attempts to combine information about the grading patterns of instructors across classes, which is an aspect of model fitting not normally included even in GRMs. The primary substantive disadvantage of this team's proposal is its complexity. It is clearly the most difficult model to explain.

In summary, all three teams proposed models that would improve the rankings of students within most undergraduate institutions. Importantly, each proposal would also reduce the incentives introduced by traditional GPA for students to enroll in "easy" classes, and would therefore improve the academic environment at colleges where they were applied. Of course, the greatest weakness of each proposal is that it is only a proposal! I encourage each team truly to

make their models an *application* of mathematics by lobbying for the adoption of an adjusted GPA at their institution.

References

Caulkin, J., P. Larkey, and J. Wei, J. 1996. Adjusting GPA to reflect course difficulty. Working paper, Heinz School of Public Policy and Management, Carnegie Mellon University.

Goldman, R., D. Schmidt, B. Hewitt, and R. Fisher, R. 1974. Grading practices in different major fields. *American Education Research Journal* 11: 343–357.

Goldman, R., and Widawski, M. 1976. A within-subjects technique for comparing college grading standards: implications in the validity of the evaluation of college achievement. *Educational and Psychological Measurement* 36: 381–390.

Johnson, Valen E. 1997. An alternative to traditional GPA for evaluating student performance. *Statistical Science* 12 (4): 251–278.

Strenta, A., and R. Elliott. 1987. Differential grading standards revisited. *Journal of Educational Measurement* 24: 281–291.

About the Author

Valen E. Johnson is Associate Professor of Statistics and Decision Sciences at Duke University. His research interests include statistical image analysis, ordinal data modeling, and Markov Chain Monte Carlo simulation methods.