UNDERSTANDING STRUCTURES THROUGH VARIATION: INSIGHTS INTO GENOME EVOLUTION AND POPULATION HISTORY

by

JONATHAN A. SHORTT

B.S., Brigham Young University, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Human Medical Genetics and Genomics Program

2018

This thesis for the Doctor of Philosophy degree by

Jonathan A. Shortt

has been approved for the

Human Medical Genetics and Genomics Program

by

Tamim Shaikh, Chair

David D. Pollock, Advisor

Elizabeth J. Carlton

Stephanie Santorico

James Sikela

Date: May 18, 2018

Shortt, Jonathan A. (Ph.D., Human Medical Genetics and Genomics Program)

Understanding structures through variation: insights into genome evolution and population history

Thesis directed by Professor David D. Pollock

ABSTRACT

As everyone knows, a row of dominoes, all standing on edge, will undergo a chain reaction when the first domino is tipped. This paper addresses the question of minimizing the total travel time of the chain reaction in two situations: when the dominoes have uniform spacing and when the dominoes may have arbitrary spacing. The solution of these problems requires a combination of symbolic calculations, in order to analyze the travel time in terms of elliptic integrals, and numerical calculations, in order to solve high-dimensional minimization problems.

The form and content of this abstract are approved. I recommend its publication.

Approved: David D. Pollock

DEDICATION

This file is dedicated to Gary, my pet chimpanzee, who watched at my side as I proved theorems and wrote my thesis.

ACKNOWLEDGMENT

This thesis would not have been possible without the generous support of Great Aunt Penelope.

TABLE OF CONTENTS

TABLES

Table

FIGURES

Figure

1. Introduction

The geometry of two dominoes, separated by a distance d, is shown in Figure 1.1. The angular displacement of a single domino is denoted θ, measured from the upright position, and it is governed by the familiar pendulum equation

$$\theta''(t) - k^2 \sin\theta(t) = 0, \tag{1.1}$$

where $k^2 = \frac{3g}{2h}$, g is the acceleration due to gravity, and h is the height of the domino. The factor of $\frac{3}{2}$ in k accounts for the moment of inertia of a thin rod, and the minus sign in this equation is due to the fact that θ is measured from the upright position. The critical angle at which one domino strikes its neighbor is given by $\theta^* = \sin^{-1}(\frac{d}{h})$. The striking point is a distance $y = \sqrt{h^2 - d^2}$ from the ground.[9]

FIGURE GOES HERE

Figure 1.1: *Two successive dominoes in the chain are separated by a distance d. The critical angle at which the first domino strikes its neighbor is given by $\theta^* = \sin^{-1}(\frac{d}{h})$.*

Multiplying (1.1) by θ' allows us to write

$$\frac{1}{2}\frac{d}{dt}(\theta'(t))^2 - k^2 \sin\theta(t)\theta'(t) = 0. \tag{1.2}$$

Integrating (1.2), letting $\theta'(0) = \omega_1$, and noting that $\theta(0) = 0$, leads to

$$(\theta'(t))^2 - \omega_1^2 = 2k^2(1 - \cos\theta(t)).$$

After some rearrangement, we find that the angular velocity of the domino is given by[3]

$$\theta'(\theta) = \sqrt{\frac{3g(1-\cos\theta) + h\omega_1^2}{h}}. \tag{1.3}$$

The angular velocity equation can also be derived from energy conservation arguments.

Two important calculations result from the angular velocity equation. First, we assume that only the horizontal component of the angular velocity at impact is imparted to the next domino. Using $\cos\theta^* = \frac{y}{h}$, this means that the initial velocity of the next domino as it begins its fall is

$$\omega_2 = \theta'(\theta^*)\cos\theta^* = \frac{y}{h^2}\sqrt{3g(h-y) + h^2\omega_1^2}.$$

Local mode analysis can be extended easily to two or more dimensions. In two dimensions, the Fourier modes have the form

$$e_{jk}^{(m)} = A(m)e^{i(\theta_1+\theta_2)}, \tag{1.4}$$

where $-\pi < \theta_1, \theta_2 \le \pi$ are the wavenumbers in the x- and y-directions, respectively. Substituting this representation into the error updating step generally leads to an expression for the change in the amplitudes of the form

$$A(m+1) = G(\theta_1, \theta_2)A(m).$$

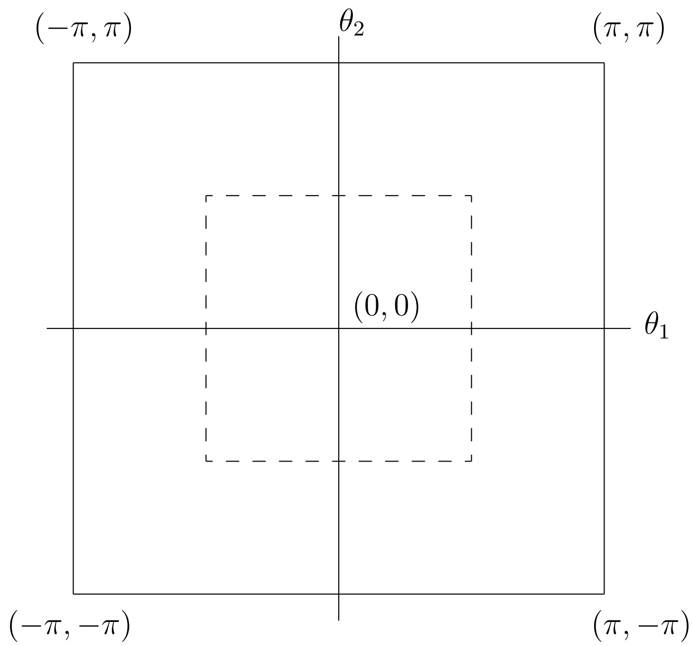
The amplification factor now depends of two wavenumbers. The smoothing factor is the maximum magnitude of the amplification factor over the oscillatory modes. As we see in Fig. 1.2, the oscillatory modes correspond to $\pi/2 \le |\theta_i| \le \pi$ for either $i = 1$ or $i = 2$; that is,

$$\mu = \max_{\pi/2\le|\theta_i|\le\pi} |G(\theta_1, \theta_2)|.$$

Now let's change the subject.

Theorem 1.1 *Let f be continuously differentiable on $[a, b]$. Then there exists $\xi \in (a, b)$ such that $f'(\xi)(b - a) = f(b) - f(a)$.*

Proof: Intuitively clear. ∎



.

Figure 1.2: *The oscillatory modes in two dimensions correspond to the wave numbers $\pi/2 \leq |\theta_i| < \pi$ for either $i = 1$ or $i = 2$; this is the region outside of the dashed box.*

Corollary 1.2 *Let f be continuously differentiable on $[a, b]$ with $f(a) = f(b) = 0$. Then there exists $\xi \in (a, b)$ such that $f'(\xi) = 0$.*

Proof: Even more obvious. ■

Table 2.1: *Statistics on coarsening for AMG applied to frozen asphalt.*

Level	Number of Rows	Number of nonzeros	Density (% full)	Average entries per row
0	4192	28832	0.002	6.9
1	2237	29617	0.006	13.2
2	867	14953	0.020	17.2
3	369	7345	0.054	19.9
4	152	3144	0.136	20.7
5	69	1129	0.237	16.4
6	30	322	0.358	10.7
7	20	156	0.390	7.8
8	18	125	0.383	6.9
9	3	9	1.000	3

2. Main Results

To understand how *AMG* performs on this problem, we first examine the coarsening statistics in Table 2.1. One interesting phenomenon is that the relative density of the operators (percentage of non-zeros) increases as the grids become coarser. This increasing density can be seen in the last column, where the average number of entries per row increases through several coarse levels. Fortunately, the operator complexity is not adversely affected: a short calculation shows that the operator complexity is 2.97, while the grid complexity is 1.89. It is also interesting to note that the first coarse-grid operator has more actual nonzero coefficients than does the fine-grid operator. This is relatively common for *AMG* on unstructured grids.

The convergence of the method is summarized in Table 2.2. Using the standard 2-norm, we see that after 11 V-cycles, the residual norm reached 10^{-7} and the process attained an asymptotic convergence factor of 0.31 per V-cycle. Clearly, *AMG* does

Table 2.2: *Convergence of AMG V-cycles applied to popcorn in terms of the 2-norm (Euclidean norm) of the residual.*

V-cycle	$\|\|\mathbf{r}\|\|_2$	Ratio of $\|\|\mathbf{r}\|\|_2$
0	1.00e+00	–
1	4.84e-02	0.05
2	7.59e-03	0.16
3	1.72e-03	0.23
4	4.60e-04	0.27
5	1.32e-04	0.29
6	3.95e-05	0.30
7	1.19e-05	0.30
8	3.66e-06	0.31
9	1.12e-06	0.31
10	3.43e-07	0.31
11	1.05e-07	0.31
12	3.22e-08	0.31

not converge as rapidly for this problem as for the previous model problem, where we saw a convergence factor of about 0.1. Nevertheless, in light of the unstructured grid and the strong, non-aligned jump discontinuities in the coefficients, the convergence properties are quite acceptable.

APPENDIX A. Theorems

Four score and seven years ago our fathers brought forth, upon this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived, and so dedicated, can long endure. We are met here on a great battlefield of that war. We have come to dedicate a portion of it as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But in a larger sense we can not dedicate – we can not consecrate – we can not hallow this ground. The brave men, living and dead, who struggled, here, have consecrated it far above our poor power to add or detract. The world will little note, nor long remember, what we say here, but can never forget what they did here. It is for us, the living, rather to be dedicated here to the unfinished work which they have, thus far, so nobly carried on. It is rather for us to be here dedicated to the great task remaining before us – that from these honored dead we take increased devotion to that cause for which they here gave the last full measure of devotion – that we here highly resolve that these dead shall not have died in vain; that this nation shall have a new birth of freedom; and that this government of the people, by the people, for the people, shall not perish from the earth.

APPENDIX B. Contradictions

When in the Course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the Laws of Nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the pursuit of Happiness. That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed. That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness.

REFERENCES

[1] R.E. Bank, *PLTMG: A Software Package for Solving Elliptic Partial Differential Equations – Users' Guide 7.0*, vol. 15, Frontiers in Applied Mathematics, SIAM Books, Philadelphia, 1994.

[2] J.H. Bramble, *Multigrid Methods*, vol. 294, Pitman Research Notes in Mathematical Sciences, Longman Scientific and Technical, Essex, England, 1993.

[3] A. Brandt, *Multi-level adapative solutions to boundary value problems*, Mathematics of Computation, 31, 1977, pp. 333-390.

[4] A. Brandt, *Multigrid techniques: 1984 guide with applications to fluid dynamics*, GMD-studien Nr. 85, Gesellschaft für Mathematik und Datenverarbeitung, St. Agustin, 1984.

[5] A. Brandt, *Algebraic multigrid theory: the symmetric case*, Appl. Math. Comput., 19, 1986, pp. 23-56.

[6] A. Brandt, S.F. McCormick, J. Ruge, *Algebraic multigrid (AMG) for sparse matrix equations*, in Sparsity and its Applications, D.J. Evans, ed., Cambridge University Press, 1984.

[7] J.W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[8] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM Books, Philadelphia, 1996.

[9] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1996

[10] W. Hackbusch, *Multigrid Methods and Applications*, vol. 4, Computational Mathematics, Springer-Verlag, Berlin, 1985.

[11] W. Hackbusch and U. Trottenberg, *Multigrid Methods*, vol. 960, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1982.

[12] D. Jespersen, *Multigrid Methods for Partial Differential Equations*, in Studies in Numerical Analysis, vol. 24, Studies of Mathematics, MAA, Washington D.C., 1984.

[13] S.F. McCormick, ed., *Multigrid Methods*, vol. 3, Frontiers in Applied Mathematics, SIAM Books, Philadelphia, 1987.