International Journal of Mathematical Education in Science and Technology, Vol. 37, No. 1, 2006, 81–125

Taylor & Francis
Taylor & Francis Group

Classroom notes

Who should be the winner: a *post hoc* analysis of the 2000 US election

D. W. WU*
Department of Mathematics, Pacific Lutheran University, Tacoma, WA 98447, USA

(*Received 28 January 2004*)

The 2000 US presidential election between Al Gore and George W. Bush has been the most intriguing and controversial in American history. Using the Florida ballot data, Wu showed that the 2000 election result could have been reversed had the 'butterfly ballot effect' been eliminated. Through a combinatorial approach, Harger concluded that Gore should have won the election based on certain assumptions. In this paper, the author analyses the voting results of the 2000 election by applying the multiple regression technique to a set of national voters' data and discusses the prediction error under three different voting systems. The study indicates that party affiliation was the dominant factor in the 2000 election and that a proportional voting system was more predictable than popular and electoral systems.

1. Introduction

On the morning of 12 November 2000, Americans woke up without knowing who would be the nation's next president. In the following months, there were a great deal of controversial discussions and arguments on the election across the nation, especially during the Florida endless recounting process. With George W. Bush's meager lead, an adjustment of a decidedly small proportion of the votes could well reverse the outcome of the election. One of the most controversial accusations was the use of the 'butterfly ballot' in Florida. Using the Florida ballot data, Wu [1] showed that the 2000 election result could have been reversed had the butterfly ballot effect been eliminated. Through a combinatorial approach, Harger [2] concluded that Gore should have won the election based on certain assumptions over the likelihood of winning for each party in each state.

As the 2004 presidential election date approaches, we take another look at the 2000 national voter's demographic and social economic state-by-state data [3] by applying the multiple regression technique. Table 1 provides a data dictionary for this study.

*Email: wu@danewu.com

International Journal of Mathematical Education in Science and Technology
ISSN 0020–739X print/ISSN 1464–5211 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00207390500138199

Table 1. Definitions of variables.

Variable	Name	Definition
x_1	Female%	Percentage of the female voters in the state.
x_2	Black%	Percentage of the black voters in the state.
x_3	Young%	Percentage of the young voters (age < 45) in the state.
x_4	Jew%	Percentage of the Jewish voters in the state.
x_5	In(income)	Natural logarithm of the average family annual income.
x_6	Dem%	Percentage of the Democrat voters in the state.
x_7	Bush-vote	Actual number of votes Bush received from the state.
x_8	Gore-vote	Actual number of votes Gore received from the state.
y	B/G Ratio	The ratio of Bush-vote and Gore-vote.

Based on the US electoral voting system, for a given state (except Maine and Nebraska), if $y > 1(< 1)$, then Bush wins (loses) all the electoral votes for that state. If $y = 1$, then Bush would tie with Gore (recounting is necessary!) We aim to use variables $x_1, x_2, \ldots, x_6$ to 'predict' y. And then based on the fitted y value, we will calculate the 'predicted' votes for Bush in three different voting systems.

2. Regression modelling

Table 2 lists the Pearson's correlation coefficients between the predictor variable and the B/G Ratio as well as the corresponding p-values for testing the zero correlation. It is clear that party affiliation was the dominant factor in the 2000 election. Moreover, female, black, Jewish, low-income, and Democrat voters tended to be in favour of Gore.

Table 3 summarizes the results of forward regression. R^2 (coefficient of determination), MSE (mean square error), C_p (estimate of $TMSE/\sigma^2$), PRESS (prediction sum of squares), and simplicity are common criteria for regression model selection. These following four models are chosen for study in light of these criteria.

Model 1 $y = 3.39354 - 4.25112x_6.$

Model 2 $y = 3.23 - 4.02x_4 - 3.74x_6.$

Model 3 $y = 19.3 - 12.0x_1 + 0.439x_2 + 2.17x_3 - 2.32x_4 - 1.06x_5 - 3.39x_6.$

Model 4 $y = 5.56 - 12.7x_6 + 80.878.6x_4^2 - 22.3x_4x_6 + 8.65x_6^2.$

About 72% of the variation in the B/G ratio can be explained by Dem% alone. This simple model correctly predicts 37 states for Bush in terms of winning or losing the electoral votes for each state. The R^2 value is increased to 85.81% when the full linear model is applied. However, the number of states that are correctly 'predicted' is only increased by one. Comparing with the three linear models, the quadratic model (Model 4) yields the biggest R^2 (=88.80%), the least MSE (=0.18), the lowest PRESS (=1.8). And most importantly, Model 4 is unbiased since $C_p = 5.0$ [2]. This model is able to correctly 'predict' 40 states.

Table 2. Pearson's correlation coefficients and p-values.

Variable	Female%	Black%	Young%	Jew%	In(income)	Dem%
r	−0.478	−0.290	0.142	−0.562	−0.283	−0.851
p-value	<0.001	0.039	0.320	<0.001	0.041	<0.001

Table 3. Linear subset models and their criteria values.

Number of predictors	Variables in the model	R^2(%)	MSE	C_p	PRESS
1	x_6	72.36	0.27	37.2	4.07
2	x_6, x_4	77.70	0.24	23.2	3.43
3	x_6, x_4, x_5	79.41	0.24	20.0	3.31
4	x_6, x_4, x_5, x_1	82.38	0.22	13.1	3.12
5	x_6, x_4, x_5, x_1, x_3	84.41	0.21	7.7	2.77
6	$x_6, x_4, x_5, x_1, x_3, x_2$	85.81	0.20	6.8	2.87

Table 4. Model comparisons.

Model	Number of states correctly predicted	Predicted Bush's votes by popular voting system (Actual votes = 50456062)	Predicted Bush's votes by electoral voting system (Actual votes = 271)	Predicted Bush's votes by proportional voting system (Actual votes = 269.00)
1	37	56826177	399	280.64
2	38	53038044	339	270.98
3	38	52832184	384	272.56
4	40	51299683	353	270.47

Using our data set [4] (also see Appendix) and the 2000 electoral voters' data [5], we are able to calculate the total projected votes for Bush in three different voting systems (see table 4).

Please note that the actual vote is used in our calculation for Oklahoma due to the missing Dem% for the state.

3. Who should be the winner?

With a total of 50996582 votes (compared to 50456062 for Bush), Gore would have won the 2000 election were the popular voting system used. Or were the proportional voting system applied, Gore would also have exactly tied with Bush by claiming 269 (out of 538) electoral votes if the final vote were rounded to integer. The reality was that Bush won the election by claiming 271 electoral votes in the US 2000 electoral voting system in which 48 states and DC awarded all their electoral votes to the winner, regardless of the margin of victory. Only two states, Maine and Nebraska, allocated electoral votes proportionately. As a special note, Bush would still have

Table 5. Prediction error comparisons.

Model	Error by popular voting system (%)	Error by electoral voting system (%)	Error by proportional voting system (%)
1	12.63	47.23	4.32
2	5.12	25.09	0.74
3	4.71	41.70	1.32
4	1.67	30.26	0.55

received 271 votes if Maine and Nebraska did not allocate their electoral votes proportionately.

It is interesting that all of our regression models above 'predict' that Bush would win in three different voting systems. However, a regression model works very differently in 'predicting' Bush's votes in different voting systems. The proportional voting system seems to be most predictable (see table 5). If either a popular or proportional voting system were enforced, a minor mistake in the Florida recounting would not have altered the 2000 election result. Over the years, 'Do not blame me, I just voted with the majority' has become American's new mantra. Perhaps it is time for all Americans to seriously rethink the presidential voting system before it gets blamed again.

Appendix. The 2000 US presidential election source data.

State	Female%	Black%	Young%	Jew%	In(income)	Dem%	Bush-vote	Gore-vote	Electoral vote
Alabama	0.530453	0.242277	0.520396	0.003139	10.4972	0.525641	941173	692611	9
Alaska	0.472093	0.047887	0.574766	0.016854	10.8495	0.413793	167398	79004	3
Arizona	0.511313	0.040425	0.377261	0.045656	10.5219	0.473684	781652	685341	8
Arkansas	0.526943	0.144806	0.500519	0.001404	10.3010	0.626866	472940	422768	6
California	0.505186	0.085946	0.571440	0.073177	10.6861	0.551282	4567429	5861203	54
Colorado	0.509945	0.044489	0.527551	0.040123	10.7861	0.492958	883748	738227	8
Connecticut	0.523009	0.090984	0.510204	0.047783	10.8356	0.571429	561104	816659	8
Delaware	0.521515	0.701754	0.541237	0.045016	10.7329	0.575342	137288	180068	3
DC	0.540146	0.582278	0.527981	0.069333	10.5632	0.906977	18073	171923	3
Florida	0.522806	0.139070	0.466831	0.112000	10.4878	0.512821	2912790	2912253	25
Georgia	0.521806	0.274166	0.570677	0.022442	10.5824	0.518987	1419720	1116230	13
Hawaii	0.505495	0.078261	0.512061	0.017588	10.7004	0.704225	137845	205286	4
Idaho	0.506508	0.007786	0.532030	0.000000	10.4887	0.352113	336937	138637	4
Illinois	0.520427	0.144426	0.534068	0.039141	10.7449	0.588235	2019421	2589026	22
Indiana	0.521808	0.080373	0.528327	0.006836	10.6196	0.461538	1245836	901980	12
Iowa	0.520554	0.021077	0.495155	0.003571	10.6271	0.514706	634373	638517	7
Kansas	0.515381	0.058001	0.524962	0.011021	10.5315	0.434211	622332	399276	6
Kentucky	0.523889	0.069767	0.520722	0.004946	10.4312	0.541176	872520	638923	8
Louisiana	0.529339	0.298657	0.493190	0.005712	10.3950	0.585366	927871	792344	9
Maine	0.519628	0.007315	0.505165	0.017897	10.5696	0.534483	286616	319951	4
Maryland	0.522293	0.281908	0.546381	0.091656	10.8649	0.641026	813724	1143888	10
Mass	0.526537	0.059042	0.524642	0.068124	10.6963	0.585859	878502	1616487	12
Michigan	0.521745	0.135808	0.528269	0.022829	10.7416	0.536232	1953139	2170418	18
Minnesota	0.513955	0.030895	0.533559	0.014742	10.7630	0.537313	1109659	1168266	10
Mississippi	0.531998	0.333169	0.537372	0.000554	10.3902	0.487805	572844	404614	7
Missouri	0.524970	0.105094	0.517166	0.020988	10.6326	0.506494	1189924	1111138	11
Montana	0.508982	0.006329	0.474551	0.002924	10.3496	0.430769	240178	137126	3
Nebraska	0.519449	0.291895	0.515789	0.006951	10.5658	0.378378	433850	231776	5

(*continued*)

Appendix. Continued.

State	Female%	Black%	Young%	Jew%	In(income)	Dem%	Bush-vote	Gore-vote	Electoral vote
Nevada	0.492806	0.080769	0.522302	0.054264	10.6378	0.493333	301575	279978	4
NewHamp	0.513721	0.010022	0.549342	0.022676	10.7400	0.475410	273559	266348	4
NewJersey	0.524023	0.146275	0.514414	0.091772	10.8184	0.571429	1284173	1788850	15
NewMexico	0.516231	0.032485	0.523357	0.010090	10.3882	0.573333	286417	286783	5
NewYork	0.528359	0.178246	0.519739	0.142144	10.5981	0.616438	2403374	4107697	33
N. Carolina	0.523629	0.207537	0.527169	0.005540	10.5278	0.518987	1631163	1257692	14
N. Dakota	0.507338	0.008753	0.507338	0.002058	10.4005	0.506667	174852	95284	3
Ohio	0.526029	0.107611	0.518321	0.023705	10.5870	0.488372	2350363	2183628	21
Oklahoma	0.520158	0.079502	0.505138	0.002853	10.4018	*	744337	474276	8
Oregon	0.511858	0.021092	0.499011	0.021645	10.6143	0.541667	713577	720342	7
Pennsylvania	0.528621	0.091223	0.489132	0.044612	10.5452	0.500000	2281127	2485967	23
RhodeIsland	0.528552	0.049113	0.517241	0.020779	10.6675	0.716667	130555	249508	4
S. Carolina	0.527041	0.277362	0.529214	0.004171	10.5068	0.458333	786892	566037	8
S. Dakota	0.514760	0.009843	0.510129	0.000000	10.5061	0.402439	190700	118804	3
Tennessee	0.526765	0.152242	0.520843	0.006028	10.4908	0.513158	1061949	981720	11
Texas	0.513839	0.125646	0.558653	0.010366	10.5708	0.454545	3799639	2433746	32
Utah	0.510580	0.011339	0.608191	0.002909	10.7384	0.280000	515096	203053	5
Vermont	0.515217	0.008791	0.519565	0.025105	10.6366	0.516129	119775	149022	3
Virginia	0.519476	0.199128	0.550722	0.025227	10.7309	0.486111	1437490	1217290	13
Washington	0.508471	0.038072	0.537881	0.021079	10.7285	0.560606	1108864	1247652	11
W.Virginia	0.526836	0.031983	0.472830	0.002695	10.2899	0.582278	336473	295497	5
Wisconsin	0.515522	0.050117	0.521120	0.009195	10.7326	0.536232	1237279	1242987	11
Wyoming	0.501393	0.011461	0.495822	0.000000	10.5293	0.328947	147947	60481	3

References

[1] Harger, R.T., 2003, A combinatorial look at the election of a U.S. President, *International Journal of Mathematical Education in Science and Technology*, **34**, 459–463.
[2] Mendenhall, W. and Sincich, T., 1996, *A Second Course in Statistics: Regression Analysis*, 5th edn (Upper Saddle River, NJ: Prentice-Hall).
[3] Wu, D.W., 2002, Regression analyses on the butterfly ballot effect: a statistical perspective of the US 2000 election, *International Journal of Mathematical Education in Science and Technology*, **33**, 309–317.
[4] http://www.danewu.com/2000ElectionData.doc
[5] http://www.fec.gov/pages/elecvote.htm

Parallel curves: getting there and getting back

A. F. AGNEW and J. H. MATHEWS*

Mathematics Department, California State University Fullerton, Fullerton, CA, 92834, USA

(*Received 28 May 2004*)

This note takes up the issue of parallel curves while illustrating the utility of *Mathematica* in computations. This work complements results presented earlier. The presented treatment, considering the more general case of parametric curves, provides an analysis of the appearance of cusp singularities, and emphasizes the utility of symbolic algebra systems.

1. Introduction

Given a curve $F(x, y) = 0$ we want to construct a 'parallel curve' that is a specified distance δ from the given curve. We wish to retain some of the natural properties from the case of parallel lines; for example, we would like our parallel curves not to intersect, and we would like to maintain a constant distance between the curves. At first glance we might think that vertical translation $y = G(x) + \delta$ is the solution. If we consider the parabola $x = H(y) = y^2$, and it's $\delta = \frac{1}{2}$ vertical translate $x = H(y - \frac{1}{2}) = (y - \frac{1}{2})^2$, we learn quickly that these curves are not 'parallel' since they intersect. Is horizontal translation appropriate? Using $x = H(y) = y^2$ and its $\delta = \frac{1}{2}$ horizontal translate $x = \frac{1}{2} + H(y) = \frac{1}{2} + y^2$, we again find that the curves are not 'parallel' since a constant distance is not maintained between the curves. These two cases are illustrated in figures 1 and 2.

2. Background: the tangent and normal vector fields

The appropriate construction of a 'parallel curve' will involve the unit tangent and unit normal vector fields of a curve. Curves will be described in parametric form

*Corresponding author. Email: mathews@fullerton.edu

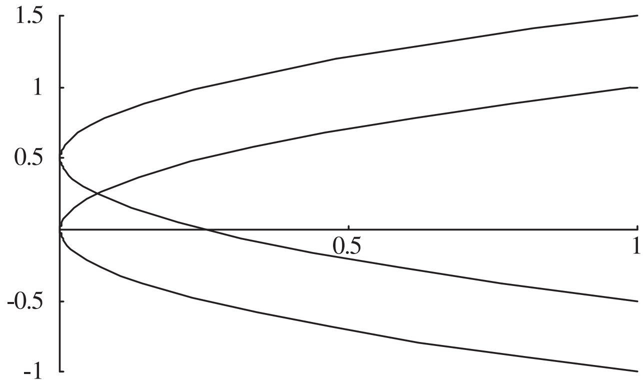

Figure 1. The curves $x = y^2$ and $x = (y - \frac{1}{2})^2$.

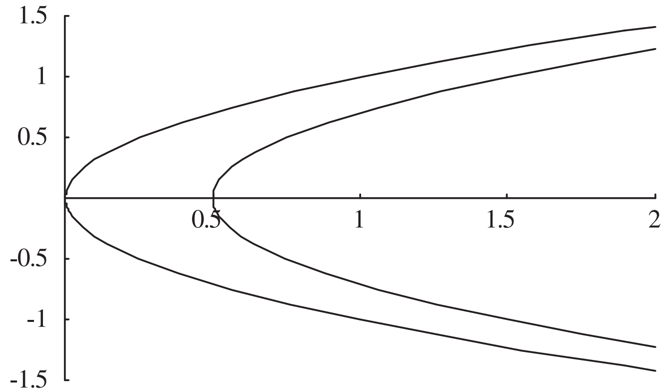

Figure 2. The curves $x = y^2$ and $x = \frac{1}{2} + y^2$.

with twice differentiable coordinate functions $x = f[t]$ and $y = g[t]$, so that the vector representation $\vec{\mathbf{r}}[t]$ will be written as

$$\vec{\mathbf{r}}[t] = \{f[t], g[t]\} \quad \text{for} \quad a < t < b.$$

The tangent vector field of the curve $\vec{\mathbf{r}}[t]$ in (1) is obtained by differentiating each component of $\vec{\mathbf{r}}[t]$ with respect to the parameter t,

$$\vec{\mathbf{t}}[t] = \{f'[t], g'[t]\}.$$

The unit tangent vector field $\vec{\mathbf{u}}[t]$ is then given by

$$\vec{\mathbf{u}}[t] = \frac{\vec{\mathbf{t}}[t]}{\left\|\vec{\mathbf{t}}[t]\right\|} = \frac{1}{\sqrt{(f'[t])^2 + (g'[t])^2}}\{f'[t], g'[t]\}.$$

A vector field perpendicular to $\vec{\mathbf{u}}[t]$ is a normal vector field, and one choice is $\vec{\mathbf{n}}[t] = \{g'[t], -f'[t]\}$. (the other choice is the opposite $-\vec{\mathbf{n}}[t]$, used, for example, in [1]).

The unit normal vector field, which we denote by $\vec{\mathbf{w}}[t]$, is given by

$$\vec{\mathbf{w}}[t] = \frac{\vec{\mathbf{n}}[t]}{\|\vec{\mathbf{n}}[t]\|} = \frac{1}{\sqrt{(f'[t])^2 + (g'[t])^2}}\{g'[t], -f'[t]\}.$$

3. The formula for a parallel curve

Let $\vec{\mathbf{r}}[t_0] = \{f[t_0], g[t_0]\}$ be a fixed but arbitrary point on a given curve $\vec{\mathbf{r}}[t] = \{f[t], g[t]\}$. The point in $\mathbb{R}^2$ that lies a distance δ from $\vec{\mathbf{r}}[t_0]$ in the normal direction $\vec{\mathbf{w}}[t_0]$ is $\vec{\mathbf{r}}[t_0] + \delta\vec{\mathbf{w}}[t_0]$. Repeating this procedure for each point of $\vec{\mathbf{r}}[t]$ results in a curve which we define to be the curve parallel to $\vec{\mathbf{r}}[t]$ at distance δ:

$$\vec{\mathbf{R}}[\delta, t] = \vec{\mathbf{r}}[t] + \delta\vec{\mathbf{w}}[t]$$

is the proposed parallel curve, and its coordinate form is

$$\vec{\mathbf{R}}[\delta, t] = \left\{f[t] + \delta\frac{g'[t]}{\sqrt{f'[t]^2 + g'[t]^2}}, g[t] - \delta\frac{f'[t]}{\sqrt{f'[t]^2 + g'[t]^2}}\right\}.$$

Note that this method of constructing a parallel curve, which we call 'normal translation', captures the notion of maintaining a fixed distance between the curves and apparently avoids creating intersections. We will investigate $\vec{\mathbf{R}}[\delta, t]$ and see that it is easy to explore examples with calculators or computers using the parametric graphing capability. However there are caveats to look out for in some situations. For purposes of illustration, the software *Mathematica* will be used in the following constructions, although the software Maple is equally capable for these tasks. Graphics calculators with symbolic capabilities can also be used. The general parallel curve $\vec{\mathbf{R}}[\delta, t]$ in coordinate form is constructed in *Mathematica* with the following commands.

$$\vec{\mathbf{r}}[t\_] = \{f[t], g[t]\}; \quad \text{Print}\left[\text{“}\vec{\mathbf{r}}[t] = \text{”}, \vec{\mathbf{r}}[t]\right];$$

$$\vec{\mathbf{t}}[t\_] = \vec{\mathbf{r}}'[t]; \quad \text{Print}\left[\text{“}\vec{\mathbf{t}}[t] = \text{”}, \vec{\mathbf{t}}[t]\right];$$

$$\vec{\mathbf{u}}[t\_] = \frac{1}{\sqrt{(f'[t])^2 + (g'[t])^2}}\vec{\mathbf{r}}'[t]; \quad \text{Print}\left[\text{“}\vec{\mathbf{u}}[t] = \text{”}, \vec{\mathbf{u}}[t]\right];$$

$$\vec{\mathbf{k}}[t\_] = \text{Simplify}\left[\vec{\mathbf{u}}'[t]\right]; \quad \text{Print}\left[\text{“}\vec{\mathbf{k}}[t] = \text{”}, \vec{\mathbf{k}}[t]\right];$$

$$\vec{\mathbf{n}}[t\_] = \{g'[t], -f'[t]\}; \quad \text{Print}\left[\text{“}\vec{\mathbf{n}}[t] = \text{”}, \vec{\mathbf{n}}[t]\right];$$

$$\vec{\mathbf{w}}[t\_] = \frac{1}{\sqrt{(f'[t])^2 + (g'[t])^2}}\{g'[t], -f'[t]\}; \quad \text{Print}\left[\text{“}\vec{\mathbf{w}}[t] = \text{”}, \vec{\mathbf{w}}[t]\right];$$

$$\vec{\mathbf{R}}[\delta\_, t\_] = \text{Simplify}\left[\vec{\mathbf{r}}[t] + \delta\vec{\mathbf{w}}[t]\right]; \quad \text{Print}\left[\text{“}\vec{\mathbf{R}}[\delta, t] = \text{”}, \vec{\mathbf{R}}[\delta, t]\right];$$

4. Parallel curves: examples and limitations

In our explorations of the parabola $\vec{\mathbf{r}}[t] = \{t^2, t\}$, we will make use of the quantities $\vec{\mathbf{t}}[t]$, $\vec{\mathbf{n}}[t]$ and $\vec{\mathbf{w}}[t]$ arising in the construction of the parallel curve. Hence, we compute:

$$\vec{\mathbf{t}}[t] = \vec{\mathbf{r}}'[t] = \{2t, 1\}$$

$$\vec{\mathbf{n}}[t] = \{1, -2t\}$$

$$\vec{\mathbf{w}}[t] = \frac{\vec{\mathbf{n}}[t]}{\|\vec{\mathbf{n}}[t]\|} = \left\{\frac{1}{\sqrt{1+4t^2}}, \quad \frac{-2t}{\sqrt{1+4t^2}}\right\}$$

The parallel curve that is a distance δ from the parabola $\vec{\mathbf{r}}[t] = \{t^2, t\}$ is

$$\vec{\mathbf{R}}[\delta, t] = \vec{\mathbf{r}}[t] + \delta\vec{\mathbf{w}}[t] = \left\{t^2 + \delta\frac{1}{\sqrt{1+4t^2}}, \quad t - \delta\frac{2t}{\sqrt{1+4t^2}}\right\}$$

The parallel curve $\vec{\mathbf{R}}[\delta, t]$ for the parabola $\vec{\mathbf{r}}[t] = \{t^2, t\}$ is constructed in *Mathematica* with the following commands:

$$f[t\_] = t^2; \quad g[t\_] = t;$$

$$\vec{\mathbf{r}}[t\_] = \{f[t], g[t]\}; \quad \text{Print}[\text{“}\vec{\mathbf{r}}[t] = \text{”}, \vec{\mathbf{r}}[t]];$$

$$\vec{\mathbf{r}}[t\_] = \vec{\mathbf{r}}'[t]; \quad \text{Print}\left[\text{“}\vec{\mathbf{t}}[t] = \text{”}, \vec{\mathbf{t}}[t]\right];$$

$$\vec{\mathbf{u}}[t\_] = \frac{1}{\sqrt{(f'[t]^2) + (g'[t])^2}}\,\vec{\mathbf{r}}'[t]; \quad \text{Print}[\text{“}\vec{\mathbf{u}}[t] = \text{”}, \vec{\mathbf{u}}[t]];$$

$$\vec{\mathbf{k}}[t\_] = \text{Simplify}[\vec{\mathbf{u}}'[t]]; \quad \text{Print}\left[\text{“}\vec{\mathbf{k}}[t] = \text{”}, \vec{\mathbf{k}}[t]\right];$$

$$\vec{\mathbf{n}}[t\_] = \{g'[t], -f'[t]\}; \quad \text{Print}[\text{“}\vec{\mathbf{n}}[t] = \text{”}, \vec{\mathbf{n}}[t]];$$

$$\vec{\mathbf{w}}[t\_] = \frac{1}{\sqrt{(f'[t]^2) + (g'[t])^2}}\,\{g'[t], -f'[t]\}; \quad \text{Print}[\text{“}\vec{\mathbf{w}}[t] = \text{”}, \vec{\mathbf{w}}[t]];$$

$$\vec{\mathbf{R}}[\delta\_, t\_] = \text{Simplify}[\vec{\mathbf{r}}[t] + \delta\vec{\mathbf{w}}[t]]; \quad \text{Print}\left[\text{“}\vec{\mathbf{R}}[\delta, t] = \text{”}, \vec{\mathbf{R}}[\delta, t]\right];$$

The graphs of parallel curves which lie a distance $\delta = 0, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}$, and $\frac{5}{10}$ from the given parabola are easy to construct with the parametric graphing software and are shown in figure 3.

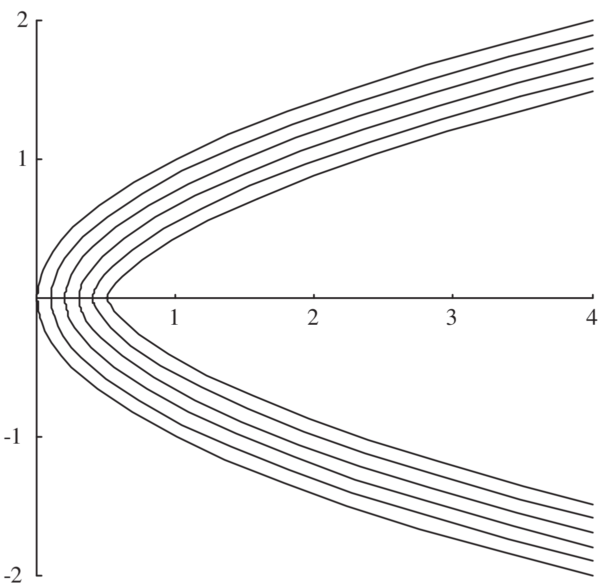

Figure 3. Parallel curves $\vec{\mathbf{R}}[\delta, t] = \{t^2 + \delta(1/\sqrt{1+4t^2}), t - \delta(2t/\sqrt{1+4t^2})\}$ for $\delta = 0, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}$, and $\frac{5}{10}$.

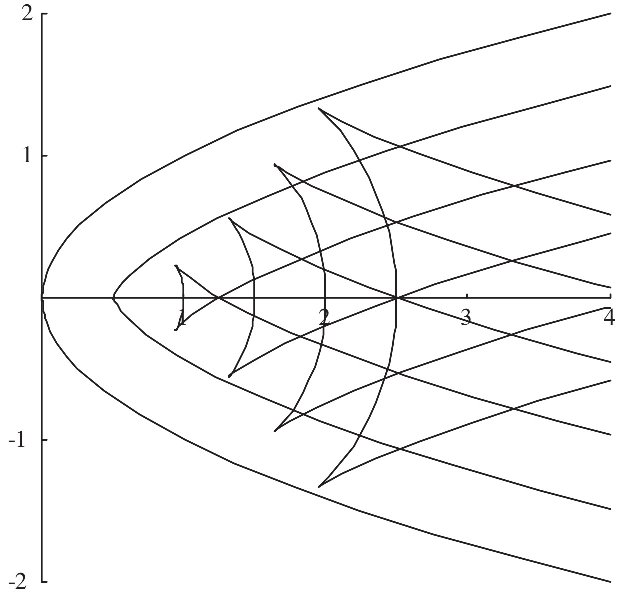

Figure 4. The curves $\vec{\mathbf{R}}[\delta, t] = \{t^2 + \delta(1/\sqrt{1+4t^2}),\ t - \delta(2t/\sqrt{1+4t^2})\}$ when $\delta = \frac{1}{2}, 1, \frac{3}{2}, 2,$ and $\frac{5}{2}$.

Why did we stop with $\delta = \frac{1}{2}$? Something strange will occur when $\delta > \frac{1}{2}$, and we must see it to believe it. The graphs of 'alleged' parallel curves which lie a distance $\delta = \frac{1}{2}$, 1, $\frac{3}{2}$, 2, and $\frac{5}{2}$ from the parabola are shown in figure 4.

Looking at figure 4, we observe that a portion of each curve $\vec{\mathbf{R}}[\delta, t]$ (for $\delta = 1$, $\frac{3}{2}$, 2, and $\frac{5}{2}$) is not parallel to the given parabola: although we are maintaining a fixed distance, the shape of these curves is distorted, and intersections among neighboring parallel curves have appeared. Clearly, we must address this issue in order to understand to what extent the normal translation technique yields true parallel curves. We now enter more fully into this interesting phenomenon of curve degeneration, that is, normally translated curves not having the same 'shape' as the original curve.

5. The parallel curve caveat: degeneration of parallel curves

In order for the construction 'to be a parallel curve', we will find later that the distance δ must satisfy $\delta \neq -(1/\kappa[t])$ for all t, where $\kappa[t]$ is the curvature at the point $\vec{\mathbf{r}}[t]$. This is a necessary condition, and if it is not heeded, then a portion of the curve $\vec{\mathbf{R}}[\delta, t]$ will not be parallel, but reflected. For the parabola $\vec{\mathbf{r}}[t] = \{f[t], g[t]\} = \{t^2, t\}$, the curvature is

$$\kappa[t] = \frac{f'[t]g''[t] - g'[t]f''[t]}{\left((f'[t])^2 + (g'[t])^2\right)^{3/2}} = \frac{-2}{(1+4t^2)^{3/2}}$$

The negative reciprocal of the curvature is

$$\frac{-1}{\kappa[t]} = \frac{1}{2}(1+4t^2)^{3/2}.$$

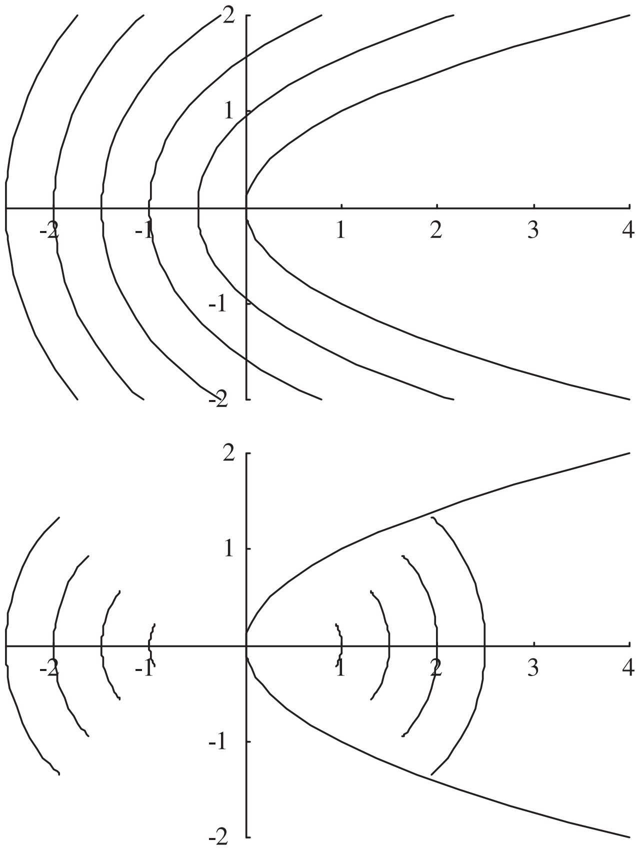

Figure 5. (a) Parallel curves $\vec{\mathbf{R}}[-\delta,\ t]$ when $\delta = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}$. (b) Portions of the curves $\vec{\mathbf{R}}[-\delta,\ t]$ and $\vec{\mathbf{R}}[\delta,\ t] = \vec{\mathbf{R}}[-\delta,\ t]$ when $\delta = 0, 1, \frac{3}{2}, 2, \frac{5}{2}$.

Thus, we see that the range of $-(1/\kappa[t])$ is $(\frac{1}{2}, \infty)$, and that the maximum value for δ must be $\frac{1}{2}$.

Consider the curves $\vec{\mathbf{R}}[\delta, t]$, constructed with negative values δ in our example of the parabola. (This corresponds to using $-\vec{\mathbf{w}}[t]$ for our unit normal vector field, not 'negative distances'.) These curves are indeed parallel and lie to the left of the given parabola, as shown in figure 5(a). If we look at a small portion of these parabolas and reflect them through the origin, their reflection is the non-parallel portion of $\vec{\mathbf{R}}[\delta, t]$ seen in figure 4. This situation is shown in figure 5.

6. The parallel curve caveat: the general case

In general, assume only that the given curve $\vec{\mathbf{r}}[t]$ is regular (i.e. $\|\vec{\mathbf{t}}[t]\| \neq 0$) and has at least two continuous derivatives, so that the curvature is well-defined and continuous. The appearance of degenerate curves is then concurrent with the appearance of singular points on the curves. Singular points occur for t values that are simultaneous solutions to $f'[t] = g'[t] = 0$, and so are points of irregularity. Thus, a regular curve may not necessarily map (via normal translation) to a regular curve.

Moreover, it follows that the tangent and normal vector fields vanish at singular points, and the unit tangent and normal vector fields are undefined. Let us look at how mapping via normal translation may introduce singularities and, moreover, determine the role of the radius of curvature $\rho[t]$. Assume (as above) that we are given $\vec{\mathbf{r}}[t]$, and we compute the curve normally translated at distance δ. Using *Mathematica*, we compute the derivative of the first component of $\vec{\mathbf{R}}[\delta, t]$:

$$D\Big[\vec{\mathbf{R}}[\delta, t][[1]], t\Big] \quad //\text{Factor} \quad //\text{Apart}$$

$$f'[t] + \frac{\delta f'[t]\,(-g'[t]f''[t] + f'[t]g''[t])}{(f'[t]^2 + g'[t]^2)^{3/2}}$$

Factoring out $f'[t]$ yields:

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

If we recognize the expression for the curvature in the first summand, we obtain (denoting the curvature of the *original* curve $\vec{\mathbf{r}}[t]$ by κ):

$$(\delta\kappa + 1) f'[t].$$

Similarly, for the second component we have $(\delta\kappa + 1)g'[t]$. It is worth noting here that the functional characterization of parallelism given in [2] may be derived as a special case of this equation. To see this, make the substitutions $t \to x, g \to u$, and $\vec{\mathbf{R}}[\delta,\ t][[2]] \to g$.

In general, when only one component of the derivative of a curve vanishes for a particular value of t, we merely have a local extremum. This is in contrast to points of irregularity, where *both* derivative components vanish, and we have a singularity. The derivative $\vec{\mathbf{r}}'[t]$ of the original curve may have a zero in the first or second component, $f'[t]$ or $g'[t]$, independently. However, the derivative of the translated curve $\vec{\mathbf{R}}'[\delta, t]$ may pick up additional zeros when the characteristic factor $(\delta\kappa + 1)$ vanishes, in which case *both* factors of $\vec{\mathbf{R}}'[\delta, t]$ must vanish, and $\vec{\mathbf{R}}[\delta, t]$ will have a singularity. Thus, by parallel translating, we may pick up new singularities, but not additional extrema. This is a crucial observation and captures the degenerating behavior fully. We note that having the same extrema in the non-degenerate translated curve indicates that the basic shape should be the same as the original curve. We ask the question: *When would it be* impossible *for the characteristic factor* $(\delta\kappa + 1)$ *to vanish?* The answer is when $\delta\kappa \neq -1$ for all t. In other words, δ must not assume a value $-(1/\kappa)$ for any t. For an alternative approach to parallel curves using 'inkages' including the issue of cusps, see [3] and the references therein.

7. Getting there and getting back: inverting the parallel curve mapping

It will now be shown that if we start with a normally translated parallel curve $\vec{\mathbf{R}}[t]$, then we can get back to the original curve $\vec{\mathbf{r}}[t]$ by using the parallel curve construction technique. We thus show that the 'parallel curve mapping' is invertible. Since we moved a distance δ from $\vec{\mathbf{r}}[t]$ to get to $\vec{\mathbf{R}}[t]$, we must move a distance $-\delta$ from $\vec{\mathbf{R}}[t]$ to move back to $\vec{\mathbf{r}}_1[t]$. This result may, perhaps, be expected. Our point, however, is that computations are quite messy and therefore provide a very nice illustration of the utility of a symbolic algebra system when exploring this material. Recall formula (7) which is the coordinate form for $\vec{\mathbf{R}}[t]$; it is

$$\vec{\mathbf{R}}[t] = \left\{ f[t] + \delta \frac{g'[t]}{\sqrt{f'[t]^2 + g'[t]^2}},\ g[t] - \delta \frac{f'[t]}{\sqrt{f'[t]^2 + g'[t]^2}} \right\} \quad \text{for } a < t < b.$$

A similar construction is used to form the parallel curve $\vec{\mathbf{r}}_1[t]$ which lies a distance $-\delta$ from $\vec{\mathbf{R}}[t]$, it is

$$\vec{\mathbf{r}}_1\,[t] = \left\{ F[t] + \delta \frac{G'[t]}{\sqrt{F'[t]^2 + G'[t]^2}},\ G[t] - \delta \frac{F'[t]}{\sqrt{F'[t]^2 + G'[t]^2}} \right\} \quad \text{for } a < t < b.$$

The formulas may look the same, but they really aren't. Notice the capital letters **F** and **G** are used and the opposite ± signs are used in the δ terms. Also they will be quite unmanageable for hand computations.

We will use *Mathematica* to assist with the simplifications for showing that $\vec{\mathbf{r}}_1[t] = \vec{\mathbf{r}}[t]$, and start by entering the general formulas for $\vec{\mathbf{r}}[t]$ and $\vec{\mathbf{R}}[t]$.

$$\text{Clear}\,[d, f, g, n, R, t, w];$$

$$\vec{\mathbf{r}}[t\_] = \{f[t], g[t]\}; \quad \text{Print}\left[\text{“}\vec{\mathbf{r}}[t] = \text{”}, \vec{\mathbf{r}}[t]\right];$$

$$\vec{\mathbf{w}}[t\_] = \frac{1}{\sqrt{(f'[t])^2 + (g'[t])^2}} \{g'[t],\ -f'[t]\};$$

$$\vec{\mathbf{R}}[t\_] := \vec{\mathbf{r}}[t] + \delta\vec{\mathbf{w}}[t]; \quad \text{Print}\left[\text{“}\vec{\mathbf{R}}[t] = \text{”}, \vec{\mathbf{R}}[t]\right];$$

Next, the coordinate functions $f_1[t]$ and $g_1[t]$ must be entered into *Mathematica*.

$$f_1[t\_] = F[t] - \delta \frac{1}{\sqrt{(F'[t])^2 + (G'[t])^2}} G'[t];$$

$$g_1[t\_] = G[t] - \delta \frac{1}{\sqrt{(F'[t])^2 + (G'[t])^2}} F'[t];$$

Here we have used a semicolon ';' at the end of the lines, to prevent the result from being displayed, because they are very unwieldy. Instead, we choose to look at 'pieces' of this formula and simplify them individually before we display the final results. To start, we type the commands $F'[t]$ and $G'[t]$, and let *Mathematica* do the computation. Then the commands Simplify $[F'[t]]$ and Simplify $[G'[t]]$ are used to simplify $F'[t]$ and $G'[t]$, respectively.

$$F[t\_] = \vec{\mathbf{R}}[t]_{[[1]]}; \quad \text{Print}\,[\text{“}F'[t] = \text{”}, F'[t]]; \quad \text{Print}\,[\text{“}F'[t] = \text{”}, \text{Simplify}\,[F'[t]]];$$

$$G[t\_] = \vec{\mathbf{R}}[t]_{[[2]]}; \quad \text{Print}\,[\text{“}G'[t] = \text{”}, G'[t]]; \quad \text{Print}\,[\text{“}G'[t] = \text{”}, \text{Simplify}\,[G'[t]]];$$

The *Mathematica* output is

$$F'[t] = f'[t] + \frac{\delta g''[t]}{\sqrt{f'[t]^2 + g'[t]^2}} - \frac{\delta g'[t](2f'[t]f''[t] + 2g'[t]g''[t])}{2(f'[t]^2 + g'[t]^2)^{3/2}}$$

$$F'[t] = \frac{f'[t]\left((f'[t]^2 + g'[t]^2)^{3/2} - \delta g'[t]f''[t] + \delta f'[t]g''[t]\right)}{(f'[t]^2 + g'[t]^2)^{3/2}}$$

and

$$G'[t] = g'[t] - \frac{\delta f''[t]}{\sqrt{f'[t]^2 + g'[t]^2}} - \frac{\delta f'[t](2f'[t]f''[t] + 2g'[t]g''[t])}{2(f'[t]^2 + g'[t]^2)^{3/2}}$$

$$G'[t] = \frac{g'[t]\left((f'[t]^2 + g'[t]^2)^{3/2} - \delta g'[t]f''[t] + \delta f'[t]g''[t]\right)}{(f'[t]^2 + g'[t]^2)^{3/2}}$$

We use the result from Simplify $[(F'[t])^2 + (G'[t])^2]$ in the computations that will follow.

$$\text{Simplify}\quad \left[(F'[t])^2 + (G'[t])^2\right] = \frac{\left((f'[t]^2 + g'[t]^2)^{3/2} - \delta g'[t]f''[t] + \delta f'[t]g''[t]\right)^2}{(f'[t]^2 + g'[t]^2)^2}$$

The *Mathematica* output is **True**

The term $1/\sqrt{(F'[t])^2 + (G'[t])^2}$ appears in $f_1[t]$ and $g_1[t]$, and the above result is used to simplify this quotient as follows

$$\frac{1}{\sqrt{(F'[t])^2 + (G'[t])^2}} = \sqrt{\frac{(f'[t]^2 + g'[t]^2)^2}{\left(\left(f'[t]^2 + g'[t]^2\right)^{3/2} - \delta g'[t]f''[t] + \delta f'[t]g''[t]\right)^2}}$$

$$\frac{1}{\sqrt{(F'[t])^2 + (G'[t])^2}} = \frac{f'[t]^2 + g'[t]^2}{(f'[t]^2 + g'[t]^2)^{3/2} - \delta g'[t]f''[t] + \delta f'[t]g''[t]}$$

Use these simplified quantities and $G'[t]$ to construct

$$f_1[t] = F[t] - \delta \frac{1}{\sqrt{F'[t]^2 + G'[t]^2}} G'[t], \quad \text{then get :}$$

$$\begin{aligned} f_1[t] = f[t] &+ \frac{\delta g'[t]}{\sqrt{f'[t]^2 + g'[t]^2}} \\ &- \delta \frac{f'[t]^2 + g'[t]^2}{\left(f'[t]^2 + g'[t]^2\right)^{3/2} - \delta g'[t]f''[t] + \delta f'[t]g''[t]} \\ &\times \frac{g'[t]\left(\left(f'[t]^2 + g'[t]^2\right)^{3/2} - \delta g'[t]f''[t] + \delta f'[t]g''[t]\right)}{\left(f'[t]^2 + g'[t]^2\right)^{3/2}} \end{aligned}$$

If the right side is entered into *Mathematica* it will be automatically simplified. (Note carefully that when a long line is entered into *Mathematica* the '–' sign must be at the end of the previous line that is to be continued.)

$$f[t]+\delta\frac{g'[t]}{\sqrt{f'[t]^2+g'[t]^2}}-\frac{\delta\left(f'[t]^2+g'[t]^2\right)}{\left(\left(f'[t]^2+g'[t]^2\right)^{3/2}-\delta g'[t]f''[t]+\delta f'[t]g''[t]\right)}$$
$$\times\frac{\left(g'[t]\left(\left(f'[t]^2+g'[t]^2\right)^{3/2}-\delta g'[t]f''[t]+\delta f'[t]g''[t]\right)\right)}{\left(f'[t]^2+g'[t]^2\right)^{3/2}}$$

The *Mathematica* output is $f[t]$

Hence, we have proven that $f_1[t]==f[t]$. In a similar fashion (note the symmetry), we may prove that $g_1[t]==g[t]$.

Thus, we have arrived at the original curve! The above computations have proven the following:

$$\vec{\mathbf{r}}_1[t]=\{f_1[t],g_1[t]\}=\{f[t],g[t]\}=\vec{\mathbf{r}}[t]$$

Therefore, the parallel $\vec{\mathbf{r}}_1[t]$ curve to the parallel curve $\vec{\mathbf{R}}[t]$ is the original curve $\vec{\mathbf{r}}[t]$. Q.E.D.

8. Appendix: Mathematica code

Mathematica code for the parallel curves $\vec{\mathbf{R}}[d,t]$ for $d=\{\frac{1}{10},\frac{2}{10},\frac{3}{10},\frac{4}{10},\frac{5}{10}\}$ in figure 3.

$$\text{Clear}[d];\quad \vec{\mathbf{p}}\,[d\_]=\vec{\mathbf{R}}\,[d,t];\quad d=\left\{\frac{1}{10},\frac{2}{10},\frac{3}{10},\frac{4}{10},\frac{5}{10}\right\};$$

$$gr1=\text{ParametricPlot}\Big[\text{Evaluate}\left[\text{Join}\left[\{\vec{\mathbf{r}}[t]\},\text{Transpose}\left[\vec{\mathbf{p}}[d]\right]\right]\right],$$
$$\{t,-2,2\},\text{PlotRange}\to\{\{0,4\},\{-2,2\}\}\Big];$$

$$\text{Print}\left[\text{“}\vec{\mathbf{r}}[t]=\text{”},\vec{\mathbf{r}}[t]\right];$$

$$\text{Print}[\text{“For }d=\text{”},d,\ \text{Parallel curve are:}\ ];$$

$$\text{Print}\Big[\text{“}\vec{\mathbf{R}}[d,t]=\text{”},\text{Join}\left[\{\vec{\mathbf{r}}[t]\},\text{Transpose}\left[\vec{\mathbf{p}}[d]\right]\right]\Big];$$

Mathematica code for the proposed parallel curves $\vec{\mathbf{R}}[d,t]$ in figure 4 is the same as for figure 3, with exception of the values $d=\{\frac{1}{2},1,\frac{3}{2},2,\frac{5}{2}\}$ for figure 4.
Mathematica code for formulas (8) and (9) for $\kappa[t]$, and $\rho[t]$; the curvature and radius of curvature, respectively.

$$\kappa[t\_]=\frac{f'[t]g''[t]-g'[t]f''[t]}{((f'[t])^2+g'[t])^2)^{3/2}};\quad \text{Print}[\text{“}\kappa[t]=\text{”},x[t]];$$

$$\rho[t\_]=\frac{1}{\kappa[t]};\quad \text{Print}[\text{“}\rho[t]=\text{”},\rho[t]];$$

$$\text{eqn}=\rho'[t]==0;\quad \text{Print}[\text{“Solve }\rho'[t]=\text{”},\text{eqn}];$$

$$\text{solset}=\text{Solve}[\text{eqn},t];\quad \text{Print}[\text{solset}];$$

$$\text{Print}[\text{“The extrema of }\rho\,[t]=\text{”},\rho[t]];\quad \text{Print}[\text{“}\rho[0]=\text{”},\rho[0]];$$

Mathematica code for the parallel curves $\vec{\mathbf{R}}[-d,t]$ for $d=\{\frac{1}{2},1,\frac{3}{2},2,\frac{5}{2}\}$ shown in figure 5(a).

$$\text{Clear}[d];\quad \vec{\mathbf{p}}[d\_]=\vec{\mathbf{R}}[d,t];\quad d=\left\{\frac{-1}{2},-1,\frac{-3}{2},-2,\frac{-5}{2}\right\};$$

$$gr1=\text{ParametricPlot}\left[\text{Evaluate}\left[\text{Join}\left[\left\{\vec{\mathbf{r}}[t]\right\},\text{Transpose}\left[\vec{\mathbf{p}}[d]\right]\right]\right],\right.$$
$$\left.\{t,-2,2\},\text{PlotRange}\to\{\{-2.5,4\},\{-2,2\}\}\right];$$
$$\text{Print}\left[\text{``}\vec{\mathbf{r}}[t]=\text{''},\vec{\mathbf{r}}[t]\right];$$
$$\text{Print}[\text{``For } d=\text{''},d,\text{parallel curve are: }];$$
$$\text{Print}\left[\text{``}\vec{\mathbf{R}}[d,t]=\text{''},\text{Join}\left[\left\{\vec{\mathbf{r}}[t]\right\},\text{Transpose}\left[\vec{\mathbf{p}}[d]\right]\right]\right];$$

Mathematica code for the non-parallel portion of the curves $\vec{\mathbf{R}}[-d,t]$ and $\vec{\mathbf{R}}[d,t]=-\vec{\mathbf{R}}[-d,t]$ for $d=\{1,\frac{3}{2},2,\frac{5}{2}\}$ and the given parabola, shown in figure 5(b).

$$\text{Clear}[c,d];$$
$$c[d\_]:=\text{If}\left[-1+2^{2/3}d^{2/3}<0,2,\sqrt{-\frac{1}{4}+\frac{d^{2/3}}{22^{1/3}}}\right];$$
$$s[t\_,h\_]:=\text{If}\ [\text{And}[c[h]<2,\ \text{Abs}[t]<c[h]],-1,1];$$
$$m=5;\quad \text{graphs}=\text{Table}[0,\{i,1,m+1\}];$$
$$\text{graphs}_{[[m+1]]}=\text{ParametricPlot}[\text{Evaluate}\,[\{\vec{\mathbf{r}}[t]\}],\{t,-2,2\},$$
$$\text{PlotRange}\to\{\{-2.5,4\},\{-2,2\}\}];$$
$$\text{For}\left[i=1,i\leqslant m,i++,h=\frac{i}{2};\right.$$
$$gr1=\text{ParametricPlot}\left[\text{Evaluate}\left[\left\{s[t,h]\vec{\mathbf{R}}[h,t]\right\}\right],\right.$$
$$\left.\{t,-c[h]+0.0001,c[h]-0.0001\}\right];$$
$$gr2=\text{ParametricPlot}\left[\text{Evaluate}\left[\left\{\vec{\mathbf{R}}[h,t]\right\}\right],\{t,-2.0001,-c[h]\}\right];.$$
$$gr3=\text{ParametricPlot}\left[\text{Evaluate}\left[\left\{\vec{\mathbf{R}}[h,t]\right\}\right],\{t,c[h],2.0001\}\right];$$
$$gr4=\text{ParametricPlot}\left[\text{Evaluate}\left[\left\{-s[t,h]\vec{\mathbf{R}}[h,t]\right\}\right],\right.$$
$$\left.\{t,-c[h]+0.0001,c[h]-0.0001\}\right];$$
$$\text{graphs}_{[[i]]}=\text{Show}[gr1,gr4,\text{PlotRange}\to\{\{-2.5,4\},\{-2,2\}\}];\ ];$$
$$\text{Show[graphs]};$$

References

[1] Stein, F.M., 1980, The curve parallel to a parabola is not a parabola: parallel curves, *Two-year Coll. Math. J.*, **11**, 239–246.

[2] Kroopnick, A., 1982, A note on parallel curves, *Two-year Coll. Math. J.*, **13**, 59–61.

[3] Yates, R.C., 1938, *Am. Math. Monthly*, **45**, 607–608.

A note on reverse derivations

M. SAMMAN*
Department of Mathematical Sciences, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

(*Received 5 July 2004*)

In this note, the notion of reverse derivation is studied. It is shown that in the class of semiprime rings, this notion coincides with the usual derivation when it maps a semiprime ring into its centre. However, we provide some examples to show that it is not the case in general.

Throughout, R denotes a ring with centre $Z(R)$(and not necessarily with 1). We write $[x, y]$ for $xy - yx$. Recall that a ring R is called prime if $aRb = 0$ implies $a = 0$ or $b = 0$; and it is called semiprime if $aRa = 0$ implies $a = 0$. A prime ring is obviously semiprime. An additive mapping d from R into itself is called a derivation if $d(xy) = d(x)y + xd(y)$, for all $x, y \in R$. We may recall a similar rule known as product rule for derivatives in the calculus texts, i.e. $d(fg) = d(f)g + fd(g)$. A considerable amount of work has been done on derivations and related maps on rings during the last decades.

Brešar and Vukman [1] have introduced the notion of a reverse derivation as an additive mapping d from a ring R into itself satisfying $d(xy) = d(y)x + yd(x)$, for all $x, y \in R$. Obviously, if R is commutative, then both derivation and reverse derivation are the same. In the present note, we explore more about reverse derivations. We will show that for a semiprime ring R, any reverse derivation becomes a derivation mapping R into its centre. For motivation and a close view on reverse derivations, we provide the following examples.

Example 1: Consider the ring

$$R = \left\{ \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} : \ a, b \in S \right\}$$

where S is a ring such that $S^2 \neq 0$. Define $d\colon \ R \longrightarrow R$ by

$$d\left(\begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}$$

Then it is easy to check that d is both a derivation and a reverse derivation.

Example 2: Consider the ring R as in Example 1. Define $d\colon \ R \longrightarrow R$ by

$$d\left(\begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}$$

*Email: msamman@kfupm.edu.sa

It is easy to see that d is a derivation. Now, let $x, y \in R$ such that x and y are both non-zero (this is possible because $S^2 \neq 0$ by hypothesis). Then, simple calculations show that $d(xy) \neq d(y)x + yd(x)$. So d is a derivation but is not a reverse derivation. Our final example will show that not every reverse derivation is a derivation.

Example 3: Consider the ring

$$R = \left\{ \begin{bmatrix} 0 & a & b & c \\ 0 & 0 & 0 & b \\ 0 & 0 & 0 & -a \\ 0 & 0 & 0 & 0 \end{bmatrix}; a, b, c \in \mathbb{R} \right\}$$

where $\mathbb{R}$ denotes the set of all real numbers. Define $d\colon R \longrightarrow R$ by

$$d\left(\begin{bmatrix} 0 & a & b & c \\ 0 & 0 & 0 & b \\ 0 & 0 & 0 & -a \\ 0 & 0 & 0 & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & 0 & 0 & -c \\ 0 & 0 & 0 & b \\ 0 & 0 & 0 & -a \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Let x, y be any elements of R, where

$$x = \begin{bmatrix} 0 & a & b & c \\ 0 & 0 & 0 & b \\ 0 & 0 & 0 & -a \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad y = \begin{bmatrix} 0 & e & f & g \\ 0 & 0 & 0 & f \\ 0 & 0 & 0 & -e \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Applying d, we can easily obtain

$$d(xy) = \begin{bmatrix} 0 & 0 & 0 & be - af \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = d(y)x + yd(x)$$

On the other hand, if we take the entries of the above matrices x and y as: $a = c = 1,\ b = 0$ and $f = g = 1,\ e = 0$, then

$$d(xy) = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = d(x)y + xd(y)$$

Hence d is a reverse derivation but not a derivation.

In order to prove our result we state the following lemma due to I. N. Herstein.

Lemma [2, Lemma 1.1.8]**:** Let R be a semiprime ring and let $a \in R$. Suppose that $a[x, y] = 0$, for all $x, y \in R$. Then $a \in Z(R)$.

The following result shows that for a semiprime ring R the reverse derivation is the same as the usual derivation when it maps R into its centre.

Proposition: A mapping d on a semiprime ring R is a reverse derivation if and only if it is a derivation that maps R into its centre.

Proof: Let R be a semiprime ring and $d\colon R \longrightarrow R$ a mapping on R. If d is a derivation which maps R into its centre then it is easy to see that d is a reverse derivation since, in this case, we have $d(xy) = d(x)y + xd(y) = yd(x) + d(y)x = d(y)x + yd(x)$. So let us suppose that d is a reverse derivation on R. Then

$$d(xy^2) = d(y^2)x + y^2d(x) = (d(y)y + yd(y))x + y^2d(x)$$

that is,

$$d(xy^2) = d(y)yx + yd(y)x + y^2d(x), \quad \text{for all } x, y \in R \tag{1}$$

Also,

$$d((xy)y) = d(y)xy + yd(xy) = d(y)xy + y(d(y)x + yd(x))$$

that is,

$$d(xy^2) = d(y)xy + yd(y)x + y^2d(x) \tag{2}$$

From (1) and (2), we get

$$d(y)yx = d(y)xy$$

$$d(y)[y, x] = 0, \quad \text{for all } x, y \in R \tag{3}$$

Replacing x by zx in (3), we get

$$d(y)[y, zx] = 0, \quad \text{for all } x, y, z \in R$$

So, we have

$$0 = d(y)[y, zx] = d(y)z[y, x] + d(y)[y, z]x$$

and using (3), the above equation gives

$$d(y)z[y, x] = 0, \quad \text{for all } x, y, z \in R \tag{4}$$

On the other hand, with the use of (3), we expand $d(u + y)[u + y, x] = 0$, to get

$$d(u)[y, x] + d(y)[u, x] = 0, \quad \text{for all } x, y, u \in R,$$

$$d(y)[u, x] = -d(u)[y, x] = d(u)[x, y] \tag{5}$$

Replacing z by $[u, x]zd(u)$ in (4) and using (5), we get

$$0 = d(y)[u, x]zd(u)[y, x] = -d(u)[y, x]zd(u)[y, x]$$

That is,

$$d(u)[y, x]zd(u)[y, x] = 0 \tag{6}$$

Since R is semiprime, by (6), we get

$$d(u)[y, x] = 0, \quad \text{for all } x, y, u \in R$$

By the lemma, this implies that $d(u) \in Z(R)$, for all $u \in R$. Hence $d(xy) = d(y)x + yd(x) = xd(y) + d(x)y$. This shows that d is a derivation on R which maps R into its centre.

Remark: In view of the above proposition one can easily notice that the ring R in Example 3 is not semiprime.

Acknowledgments

The author gratefully acknowledges the support provided by King Fahd University of Petroleum and Minerals during this research. The author also thanks the referee for useful suggestions for the improvement of the manuscript.

References

[1] Brešar, M. and Vukman J., 1989, On some additive mappings in rings with involution, *Aequations Math.*, **38**, 178–185.
[2] Herstein, I.N., 1976, *Rings with Involution* (Chicago: The University of Chicago Press).

Encouraging good mathematical writing

J. O'SHEA*
Department of Mathematics, NUI-Maynooth, Ireland

(*Received 19 July 2004*)

This paper is a report on an attempt to teach students in their first and second year of university how to write mathematics. The problems faced by these students are outlined and the system devised to emphasize the importance of communicating mathematics is explained.

1. Introduction

There have been many texts written on the topic of the importance of writing mathematics to the learning of mathematics, see for example [1] and [2]. Indeed one of the aims of the Irish school mathematics syllabus [3] is that 'Students should be able to communicate mathematics, both verbally and in written form'. However, students seem to be expected to learn how to use the language of mathematics by a process of osmosis: that is, they must pick it up from textbooks or notes. They are usually never formally taught or examined on the correct use of this language. Here we report on a very simple technique which has been used to alert students to the problems they have with mathematical exposition and to encourage them to solve these problems.

2. The problem

When correcting a piece of work written by a mathematics student, one often feels that the student understands the material but cannot communicate their knowledge

*Email: aoshea@maths.nuim.ie

effectively. By way of illustration, consider the following example from a first year calculus course:

Example: Let $f(x) = x^2 + 9$. Find the rate of change of f at $x = 3$.

The student may understand that they need to differentiate the function and then evaluate the derivative at the required point but often they will have problems writing this down. They may write something like:

$$f(x) = x^2 + 9 = 2x = 6$$

Of course this is nonsense, but it is clear that the student has performed the calculations correctly. It can be difficult to explain to a student why they have not received full marks.

Apart from problems with equals signs, students also have trouble with other mathematical symbols, using brackets correctly, omitting variables (for example: $f(x) = \sin$) and writing in sentences. They have difficulty explaining what they are doing and why they are doing it. A grader often finds themselves reading between the lines. Sometimes strange practices become fashionable: for example, my colleagues and I have noticed over the last two years that some students omit the line in fractions. So instead of writing $x = \frac{1}{2}$ they will write $\begin{smallmatrix}1\\2\end{smallmatrix}$. This is particularly disasterous for students who like to omit the brackets in binomial coefficients since now they have no way of distinguishing between $\binom{5}{2}$ and $\frac{5}{2}$.

Bad exposition is not only annoying to the reader but it can also lead to mistakes and misunderstandings. In the calculus example above, the students may look back on their work and see that $f(x) = 6$ and use this as the definition of the function from that point on. Faced with these kinds of problems, the mathematics department at NUI-Maynooth came up with a system designed to improve the mathematical writing of our students.

3. The system

Most of our first and second year courses are taught to large groups of students, with class sizes ranging from 100 to 200. These classes are then broken into small tutorial groups of 15–20 students which meet once a week. Our department has always had a tradition of taking up homework from students weekly. This homework is graded by the tutor.

However, in 2001 we realized that our homework system was running into some difficulties. One of the main problems was that our best tutors were often reluctant to take on lots of tutorials because of the amount of time they had to spend correcting assignments. (Since most lecturers would assign five or six homework questions each week, the tutor may have had to grade about 100 problems for each tutorial hour.)

It was decided to replace the traditional homework assignment with a tutorial quiz. These quizzes would consist of about four multiple choice questions and would be administered during the last ten minutes of each tutorial. Of course, these quizzes are very easy to correct and also have the benefit of focusing students' attention during tutorials. This had the effect of making tutorials a more active learning experience.

One of the functions of the old homework system was to get students used to writing mathematics, so we decided that we would still require students to hand in one question per week. This was called the exposition question. The purpose of the question was to ensure that students could not only solve mathematical problems but could also communicate their solution in a clear and correct manner. The marking scheme for the exposition question was designed to emphasize this dual role. The exposition question is marked out of ten, with five marks going for exposition and five marks for accuracy. The five exposition marks are awarded as follows: five marks if the answer is perfect, two marks if there is a minor error, no marks otherwise. The students' errors in the presentation of their argument are pointed out to them in a clear and unambiguous manner and in a way that is separate from the correctness of the argument itself. The marking scheme is purposely harsh; once a student has lost marks for making a particular mistake they usually do not make the same mistake again. By allotting half the marks for the hand-in question to exposition, the department is sending out a message that mathematical writing is important. Students are now aware that communicating their solution is just as important as using the right method to attack the problem.

In order to implement this grading system, it is necessary to train the tutors at the beginning of the academic year. It is vital that they are given clear guidelines so that they are comfortable with the system and that the department is seen to be acting consistently. However, the simplicity of the marking scheme allows it to be easily explained both to tutors and students.

Exposition is also taken into account when correcting final exams. For a question carrying 20 marks, two of these marks will be assigned to exposition. The student will receive two marks if their answer has no errors, one mark if there is a minor error and no marks otherwise.

4. Conclusions

It is quite difficult to measure the effect of our system since we did not count the number of exposition errors in the past. However, the feedback from tutors and lecturers has been positive. The improvement in mathematical writing is especially evident when correcting exams; it is now rare to see students making the classical blunders of yesteryear. The streamlining of grading has also led to experienced tutors being willing to take on more tutorial hours. In our opinion, the system has been easy to implement due to its simplicity and has had significant benefits. A similar system could work in any mathematics course and could even be used in schools and state examinations. Once educators are seen to value mathematical exposition, it will be easier to convince students of its importance.

References

[1] Meier, J. and Rishel, T., 1998, *Writing in the Teaching and Learning of Mathematics* (Mathematical Association of America).
[2] Sterret, A. (Ed.), 1990, *Using Writing to Teach Mathematics* (Mathematical Association of America).
[3] Irish Junior Certificate Mathematics Syllabus, http://www.education.ie/servlet/blobservlet/jc_maths_sy.pdf.

Using simple harmonic motion to help in the search for tautochrone curves

T. J. OSLER*† and T. R. CHANDRUPATLA‡
†Mathematics Department,
‡Mechanical Engineering Department, Rowan University,
Glassboro, New Jersey 08028, USA

(*Received 28 July 2004*)

The analysis of tautochrone problems involves the solution of integral equations. The paper shows how a reasonable assumption, based on experience with simple harmonic motion, allows one to greatly simplify such problems. Proposed solutions involve only mathematics available to students from first year calculus.

1. Introduction

Figure 1 shows a frictionless wire. A bead of unit mass starts at point (X, Y) from rest and descends to the bottom of the wire at O under the influence of a potential $V(y)$ describing a force in the downward direction. We assume that $V(0)=0$. We wish to shape the wire so that the bead completes the motion in a fixed time, regardless of the location of the starting point (X, Y) on the wire. We will call T the time for this descent from (X, Y) to the origin O. This is the tautochrone problem. In the classical tautochrone problem, we have the gravitational potential $V(y)=gy$ and the curve turns out to be an inverted cycloid. We will let s be the length of the curve measured from the origin O to the bead at the descending point P. Also, at the starting point (X, Y), let $s=s_0$. We begin with the statement of the conservation of energy

$$\frac{v^2}{2} + V(y) = V(Y) \tag{1}$$

where $v = -(ds/dt)$ is the velocity at the point P. In (1) the left side is the sum of the kinetic energy and potential energy at the point P, and the right side is the energy at the starting point (X, Y). From (1) we get at once

$$\frac{ds}{dt} = -\sqrt{2}\sqrt{V(Y) - V(y)} \tag{2}$$

and integrating we have

$$\int_0^T dt = T = \frac{1}{\sqrt{2}}\int_0^Y \frac{1}{\sqrt{V(Y) - V(y)}}\frac{ds}{dy}\, dy. \tag{3}$$

*Corresponding author. Email: Osler@rowan.edu

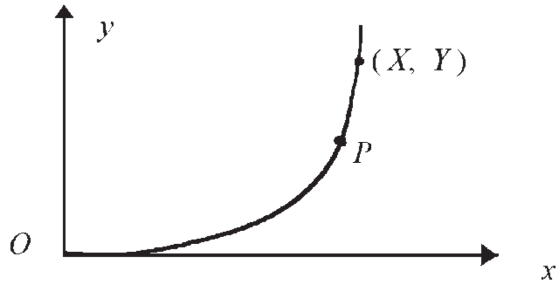

Figure 1. Bead at P on a frictionless wire.

Equation (3) is an integral equation, and is the starting point for most discussions of the tautochrone problem. The use of fractional derivatives to solve (3) dates back to Abel in the early nineteenth century, and a recent extension of this analysis is found in Flores and Osler [1]. Abel's work on this problem initiated the formal study of integral equations. If you already know the solution, you can 'cheat' and substitute the curve into the right side of (3) and show that the time T is independent of the starting ordinate Y. This is commonly done with textbook presentations with the familiar potential $V(y) = gy$, where the curve is known to be the inverted cycloid.

In this paper we avoid the need to solve the integral equation (3) by assuming a property suggested by our familiarity with simple harmonic motion. Simple harmonic motion exhibits a tautochrone-like feature since the period is independent of the amplitude. Specifically we will use the following.

Assumption of a harmonic property of tautochrone curves: If the bead on the curve shown in figure 1 moves under the given potential $V(y)$ in such a way that the arc length is given by

$$s = s_0 \cos kt, \tag{4}$$

(k is a constant), then the curve clearly has the tautochrone property and $k = \pi/2T$. We also have the differential equation

$$\frac{d^2s}{dt^2} + k^2 s = 0. \tag{5}$$

Remark: The above harmonic property is obviously true as stated since if (4) is true, then the time for the bead to go from $s = s_0$ to $s = 0$ is the constant $T = \pi/2k$. However we have no way of knowing, at this point in our analysis, that any tautochrone curves exist with this property. It is however a reasonable assumption, based on physical experience, to help us get started in our search for such curves. While it can be shown that all tautochrone curves have this harmonic property (see [2]), we will not address this problem here.

In the next section we will show how the assumption of this harmonic property yields a solution of the tautochrone problem that is short and elementary. We will need only mathematics from first year calculus. In addition, we will obtain equations for the tautochrone curve in the most physically desirable form. That is, we will find x and y in terms of the time parameter t. We have not seen this simple solution in the literature.

2. Solution assuming the harmonic property

We now solve our tautochrone problem by seeking curves with the harmonic properties (4) and (5). Starting with (2) we differentiate and get

$$\frac{d^2s}{dt^2} = \frac{1}{\sqrt{2}\sqrt{V(Y) - V(y)}}\frac{dV(y)}{ds}\frac{ds}{dt}$$

Replacing ds/dt on the right side by (2) we get

$$\frac{d^2s}{dt^2} = -\frac{dV(y)}{ds} \tag{6}$$

Comparing (6) with the harmonic property (5) we get

$$\frac{dV(y)}{ds} = k^2 s$$

Integrating and recalling that $V(0) = 0$ we get

$$V(y) = \frac{k^2}{2}s^2 \tag{7}$$

Using the harmonic relation (4) we have

$$V(y) = \frac{k^2}{2}s_0^2 \cos^2 kt = V(Y)\cos^2 kt \tag{8}$$

Since the function $V(y)$ is known, we can now solve for y in terms of time t to get

$$y = V^{-1}\big(V(Y)\cos^2 kt\big) \tag{9}$$

This is our first desired parametric equation for the tautochrone curve. We must now solve for x as a function of time.

From (4) we get $ds/dt = -ks_0 \sin kt = -\sqrt{2V(Y)}\,\sin kt$. If we call $G(y) = V^{-1}(y)$, then from (9) we get $dy/dt = -2kV(Y)G'(V(Y)\,\cos^2 kt)\cos kt\,\sin kt$. Since

$$\left(\frac{dx}{dt}\right)^2 = \left(\frac{ds}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2$$

we now have

$$\frac{dx}{dt} = -\sqrt{2V(Y)}\sqrt{1 - 2k^2 V(Y)\,G'(V(Y)\cos^2 kt)^2\cos^2 kt}\,\sin kt$$

We now integrate and get

$$x = X - \sqrt{2V(Y)}\int_0^t \sqrt{1 - 2k^2 V(Y)G'(V(Y)\cos^2 kt)^2 \cos^2 kt}\,\sin kt\,dt \tag{10}$$

For convenience we rewrite (10) as

$$x = X - \sqrt{2V(Y)}I(k;\,t) \tag{11}$$

where I is the integral in (10). Equations (9) and (10) are parametric equations for our tautochrone curve in terms of the parameter t.

Remark: If the expression under the radical in (10) becomes negative, then our solution fails, and no tautochrone curves exist with the harmonic property.

We know everything in (9) and (10) except the constant $k = \pi/2T$. When $t = T = \pi/2k$, we have $x = 0$, so (11) gives us

$$\sqrt{2V(Y)}\, I\left(k; \frac{\pi}{2k}\right) = X \tag{12}$$

Equation (12) allows us to solve for the constant k. The examples in the next section will make this clear.

3. Examples

Given the potential function $V(y)$, and the starting point (X, Y), equations (9) to (12) allow us to determine the tautochrone curve as a pair of parametric equations in terms of the parameter t. We will solve two examples.

Example 1: Let the potential be given by $V(y) = cy^2$, where c is a constant. Thus the inverse function $V^{-1}(y) = G(y) = \sqrt{y/c}$. From (9) we get

$$y = Y \cos kt. \tag{13}$$

After a little manipulation (10) becomes

$$x = X - \sqrt{2c}\, Y \int_0^t \sqrt{\frac{2c - k^2}{2c}} \sin kt\, dt$$

which integrates to

$$x = X - Y \frac{\sqrt{2c - k^2}}{k} (1 - \cos kt) \tag{14}$$

From (12), (or equivalently let $t = T = \pi/2k$ and $x = 0$ in (14)) we get

$$X = Y \frac{\sqrt{2c - k^2}}{k}$$

Thus

$$k = \frac{\sqrt{2c}\, Y}{\sqrt{X^2 + Y^2}}$$

Now (13) and (14) become

$$x = X \cos\left(\frac{\sqrt{2c}\, Y}{\sqrt{X^2 + Y^2}} t\right) \quad \text{and} \quad y = Y \cos\left(\frac{\sqrt{2c}\, Y}{\sqrt{X^2 + Y^2}} t\right)$$

and the tautochrone curve is the straight line from the origin to the point (X, Y).

Example 2: We now solve the classical tautochrone problem by using the potential $V(y) = g\, y$. We see that $G(y) = V^{-1}(y) = y/g$, so we have from (9)

$$y = \frac{k^2 s_0^2}{2g} \cos^2 kt = Y \cos^2 kt \tag{15}$$

From (10) we get

$$x = X - \sqrt{2gY} \int_0^t \sqrt{1 - \frac{2Yk^2}{g} \cos^2 kt}\, \sin kt\, dt$$

Performing the integration we have

$$x = X + \frac{g}{2k^2}\left[\sin^{-1}\left(k\sqrt{\frac{2Y}{g}}\cos kt\right) + k\sqrt{\frac{2Y}{g}}\cos kt\sqrt{1-\frac{2Yk^2}{g}\cos^2 kt}\right]$$
$$-\frac{g}{2k^2}\left[\sin^{-1}\left(k\sqrt{\frac{2Y}{g}}\right) + k\sqrt{\frac{2Y}{g}}\sqrt{1-\frac{2Yk^2}{g}}\right].$$

When $t = T$, we have $x = 0$ and we get

$$X = \frac{g}{2k^2}\left[\sin^{-1}\left(k\sqrt{\frac{2Y}{g}}\right) + k\sqrt{\frac{2Y}{g}}\sqrt{1-\frac{2Yk^2}{g}}\right] \tag{16}$$

Therefore

$$x = \frac{g}{2k^2}\left[\sin^{-1}\left(k\sqrt{\frac{2Y}{g}}\cos kt\right) + k\sqrt{\frac{2Y}{g}}\cos kt\sqrt{1-\frac{2Yk^2}{g}\cos^2 kt}\right] \tag{17}$$

Since we know X and Y, we can solve the transcendental equation (16) for k. (Notice that if $2Yk^2 > g$, then no solution of (16) exists.) Now (15) and (17) give us parametric equations for y and x in terms of t, with all constants determined.

If we let

$$\sin(\theta/2) = k\sqrt{\frac{2Y}{g}}\cos kt$$

then (17) becomes

$$x = \frac{g}{4k^2}(\theta + \sin\theta) \tag{18}$$

and (15) becomes

$$y = \frac{g}{2k^2}\sin^2\left(\frac{\theta}{2}\right)$$

which we rewrite as

$$y = \frac{g}{4k^2}(1 - \cos\theta) \tag{19}$$

Equations (18) and (19) are the equations of an inverted cycloid where the generating circle has radius

$$R = \frac{g}{4k^2} \tag{20}$$

and

$$\theta = 2\sin^{-1}\left(k\sqrt{\frac{2Y}{g}}\cos kt\right) \tag{21}$$

is the angle the generating circle has rotated as a function of time.

Remark: We note that our cycloid fails for certain initial points. Since the highest point on the inverted cycloid occurs at $(\pi R, 2R)$, we see that all starting points (X, Y)

must lie on or below the line $y = (2/\pi)x$, and above the x-axis. Thus we have the inequality

$$0 < Y \leqslant \frac{2}{\pi} X. \tag{22}$$

Finally we can summarize our solution. We start with a potential $V(y) = gy$ and a starting point at (X, Y) that satisfies the restriction (22). Next we solve (16) for k, and now (15) and (17) describe the solution as parametric equations for x and y in terms of t.

References

[1] Flores, E.V. and Osler, T.J., 1999, The tautochrone under arbitrary potentials using fractional derivatives, *American Journal of Physics*, **67**, 718–722.
[2] Osler, T.J. and Flores, E.V., 2001, The rotating tautochrone, *Journal of Applied Mechanics*, **68**, 353–356.

Computing logarithms digit-by-digit

MAYER GOLDBERG*
Department of Computer Science, Ben Gurion University,
P.O. Box 653, Beer Sheva 84105, Israel

(*Received 19 September 2004*)

In this work, we present an algorithm for computing logarithms of positive real numbers, that bears structural resemblance to the elementary school algorithm of long division. Using this algorithm, we can compute successive digits of a logarithm using a 4-operation pocket calculator. The algorithm makes no use of Taylor series or calculus, but rather exploits properties of the radix-d representation of a logarithm in base d. As such, the algorithm is accessible to anyone familiar with the elementary properties of exponents and logarithms.

1. Radixes and bases

The term *base* has two different meanings, both of which are used in this work, and it is important to distinguish amongst them. The *base of a number system* has to do with the power series representation of a number: If N is written as $a_0 a_1 \cdots a_k$ in base d, then $N = \sum_{j=0}^{k} a_j d^{k-j}$. The *base of a logarithm* has to do with representing a number as the power of another: If $\log_d x = y$, then $d^y = x$. These two distinct meanings of the term

*Email: gmayer@cs.bgu.ac.il

base are related in this work, so to prevent any ambiguity, we use the term *radix-d* to speak of the base-d representation of a number. Throughout the remainder of this work the term *base* will refer to the base of a logarithm.

2. The algorithm

Let M be a positive real number. Let the radix-d representation of the logarithm in base d of M be written as $a_0 \cdot a_1 a_2 \cdots$, then:

$$\log_d M = a_0 + \frac{a_1}{d} + \frac{a_2}{d^2} + \frac{a_3}{d^3} \cdots \quad (1)$$

Therefore

$$\begin{aligned} M &= d^{a_0 + (a_1/d) + (a_2/d^2) + (a_3/d^3) \cdots} \\ &= d^{a_0} \cdot d^{(a_1/d) + (a_2/d^2) + (a_3/d^3) \cdots} \end{aligned}$$

The central observation is that when we work in the radix-d, the first digit in the base-d expansion, or a_0 in equation (1), is always readily available. For example, the first digit of $\log_{10} 345$ is 2, because $10^2 \leqslant 345 < 10^3$. Similarly, the first digit of $\log_{10} 2468$ is 3, because $10^3 \leqslant 2468 < 10^4$.

Having extracted a_0, we now divide both sides by d^{a_0}, giving

$$\frac{M}{d^{a_0}} = d^{(a_1/d) + (a_2/d^2) + (a_3/d^3) \cdots}$$

Raising both sides to the d-th power, we get

$$\left(\frac{M}{d^{a_0}}\right)^d = d^{a_1 + (a_2/d) + (a_3/d^2 7) \cdots} \quad (2)$$

We see from equation (2) that $a_1 \cdot a_2 a_3 \cdots$ is the radix-d representation of $\log_d (M/d^{a_0})^d$, the first digit of which is a_1. Note our previous observation that the first digit of the radix-d representation of the logarithm in base d of some number is readily available, so the process of extracting the logarithm digit-by-digit can proceed.

We presented the algorithm for any choice d of both radix and base of logarithm, in order to show how these two notions are related in the algorithm. For all practical purposes, however, it would seem useful to consider the case where $d = 10$, i.e., computing logarithms in base 10 in decimal notation, or perhaps $d = 2$, where floating-point numbers would be represented in binary notation. The base of the logarithm is not as significant a choice, because it is a simple matter to convert logarithms from one base to another, by multiplying by such constants as $\ln 10$, $\log_2 10$, $\log_2 e$, etc. When computing logarithms on a pocket calculator, it is practical to use rational approximations for such constants. For example, $\ln 10$ can be approximated within $1.58 \cdot 10^{-9}$ by the fraction 5377/12381.[1]

[1] A great way to arrive at such rational approximations is using regular continued fractions [1].

2.1. *A worked-out example*

In this section, we present a worked-out example of computing $\log_{10} 1234.56$ in decimal (radix-10 notation). The various stages of the algorithm are tabulated below. In computing the first digit of the logarithm of the value in the second column, at each step of the algorithm, we need to count the number of digits to the left of the decimal point, minus one. This figure corresponds to the number of digits in the underlined portion of the value.

Expression		Value	Next digit
M	$=$	$1\underline{234}.56$	$a_0 = 3$
$M_1 = \left(\frac{M}{10^{a_0(=3)}}\right)^{10}$	$=$	82.2247369382767	$a_1 = 0$
$M_2 = \left(\frac{M_1}{10^{a_1(=0)}}\right)^{10}$	$=$	$1\underline{416511689}.4063$	$a_2 = 9$
$M_3 = \left(\frac{M_2}{10^{a_2(=9)}}\right)^{10}$	$=$	$3\underline{2}.523825911294$	$a_3 = 1$
$M_4 = \left(\frac{M_3}{10^{a_3(=1)}}\right)^{10}$	$=$	$1\underline{32439}.11735423$	$a_4 = 5$

Hence $\log_{10} 1234.56 = 3.0915\cdots$.

3. Using a pocket calculator

Adapting the algorithm for use on a pocket calculator requires that we address three issues:

- The error involved in the computation
- Raising numbers to the 10-th power on a simple, 4-operation calculator
- Overflow

3.1. *Accuracy and error*

In theory, had we been able to maintain all the digits obtained from raising numbers to the d-th power, we could have computed logarithms in base d with no loss of accuracy, one digit at a time, to any number of digits. In practice, though, calculators and computers will maintain only so many significant digits, and will round off the rest. As we iterate over the digit-extraction process, the roundoff error will propagate towards the more significant digits. After some iterations, the accumulated error will affect the number of digits to the left of the decimal point, and from that iteration onwards we will be "extracting" incorrect digits. To see how the error builds up, suppose we are computing $\log_d M$ in radix-d, for some integer $d > 1$, and some positive real number M. The error will increase as we raise numbers to the d-th power.

Consider $f(x) = x^d$. For small ϵ, we have

$$\begin{aligned} f(x+\epsilon) &\approx f(x) + \epsilon f'(x) \\ &= x^d + \epsilon\, dx^{d-1} \end{aligned}$$

Hence for a small error, we have

$$\begin{aligned} (x+\epsilon)^d &\approx x^d + \epsilon\, dx^{d-1} \\ &= x^d\left(1 + \frac{\epsilon d}{x}\right) \\ &\leqslant x^d(1 + \epsilon d) \end{aligned}$$

The upper bound on the error, the quantity ϵd, propogates the error one digit to the left.

In practice, we can use this algorithm to compute 7–8 correct digits on an 8-digit pocket calculator.

3.2. *Raising numbers to the 10-th power*

Raising a number to the 10-th power on a simple 4-operation pocket calculator can generally be done without re-entering the number, and without using the memory functions. Most pocket calculators support a feature known as *constant operations*, where given two arguments x, y, and one of the supported binary operations $\otimes \in \{+, -, \times, \div\}$, we can compute the nested, right-associated application

$$\underbrace{x \otimes (x \otimes \cdots (x \otimes y)}_{n \text{ times}} \cdots)$$

The key sequence that computes the above operation on most calculators is given by

$$\boxed{\text{key in } x} \otimes \otimes \boxed{\text{key in } y} \underbrace{= \cdots =}_{n \text{ times}}$$

On many calculators, it is even unnecessary to press $\otimes$ twice. When y is not given, the value for x, which appears on the display, is used.

Consequently, we raise a number x to the 10-th power by 9 successive multiplications of x by the *constant* x:

$$\boxed{\text{key in } x} \times \times \underbrace{= \cdots =}_{9 \text{ times}}$$

3.3. *Overflow*

Computing the 10-th power of a number that is less than 10 requires at least 10 calculator digits to represent, possibly with roundoff errors, but without an overflow error. Most 4-operation pocket calculators carry out calculations up to 8 digits, and hence overflow errors will occur. In this section, we discuss how to resume calculations after such an error.

Most calculators do not clear their display upon overflow errors. Rather, they display the correct digits, and shift the decimal point to the left by as many digits as the display can handle. They would then turn on the error annunciator (usually denoted by a small, capital **E**), and ignore all keyboard input except for the *clear error* key (usually marked $\boxed{\text{CE}}$). Pressing the *clear error* key removes the error condition, and enables further calculations, including constant operations.

For example, $1234567^2 = 1524155677489$. The result has 13 digits (to the left of the decimal point), so the calculation will cause an overflow error on an 8-digit pocket calculator, and result in an error annunciator turned on, and the display showing $\boxed{\text{E } 15241.556}$. Note that an 8-digit calculator displays the rounded answer with $13 - 8 = 5$ digits to the left of the decimal point.

In raising to the 10-th power numbers that are less than 10, there are two kinds of overflow situations:

$$\boxed{\text{E } d_0 \cdot d_1 d_2 d_3 d_4 d_5 d_6 d_7}$$

$$\boxed{\text{E } d_0 d_1 \cdot d_2 d_3 d_4 d_5 d_6 d_7}$$

representing 9- and 10-digit numbers, the logarithms of which are 8, 9 respectively. When such an overflow condition arises, we note either 8 or 0 in the expansion of the logarithm, reset the decimal point to $\boxed{\text{E } d_0 \cdot d_1 d_2 d_3 d_4 d_5 d_6 d_7}$, and continue with the algorithm.

4. Related work

John P. Killingbeck, in his book *The Creative Use of Calculators* [2, section 4.9], uses probability theory to arrive at an algorithm that amounts to a special case of the algorithm presented herein, where $d = 2$. The output of the Killingbeck's algortihm is a sequence of binary digits that are then converted to decimal and multiplied by $\log 2$ or $\ln 2$ in order to convert the logarithm to a more commonly-used base. The peculiar choice of d in Killingbeck's algortihm may have something to do with the probabilistic argument with which he arrives at his algorithm, and is otherwise unmotivated. The choice of $d = 2$ does, however, seem reasonable for working with representations of floating-point numbers on digital computers.

5. Conclusion

Even though logarithms are taught and used in high school, students are generally unable to compute logarithms in all but the simplest cases, e.g., when the logarithm is a rational number. Only after they reach college, and study calculus to Taylor series, are they able to compute the logarithm of any real number in any base. Luckily, this pedagogical wrinkle is easily ironed out, since, as this algorithm shows, logarithms can be computed one digit at a time.

We note that this algorithm is essentially the same as the well-known elementary school algorithm for long-division, where each operation in the long-division algorithm is replaced by a higher operation:

Long division	Finding a logarithm
Finding the first digit of the integer quotient	Finding the first digit of the logarithm
Reducing the dividend by the largest integer multiple of the divisor	Dividing the argument by the largest power of d that is smaller than the argument
Multiplying the difference by d, and iterating	Raising the quotient to the power d, and iterating

While the algorithm presented herein is not very efficient, it does offer several pedagogical and computational advantages:

- It assumed no calculus, and rather relies on the most elementary properties of powers and logarithms.
- The algorithm is easy to follow, especially since it is structurally similar to the elementary school algorithm for long division.
- The algorithm lends itself to rapid calculations on a pocket calculator.

Acknowledgements

The author is grateful to BRICS[2] for having hosted him and for providing a stimulating environment. The author is particularly grateful to Saurabh Agarwal, Dani Berend, Olivier Danvy, Ted Eisenberg, Kim Skak Larsen, and Kiril Morozov for their enthusiastic reception of the algorithm, and for their encouragement.

References

[1] Khinchin, A.Ya. and Aleksandr IAkovlevich, 1997, *Continued Fractions* (Moneola, New York: Dover Publications).
[2] John P. Killingbeck, 1981, *The Creative Use of Calculators* (Harmondsworth, UK: Penguin Books).

[2] Basic Research in Computer Science (www.brics.dk), funded by the Danish National Research Foundation.

Coin tossing with a new stopping rule: some interesting results

K. S. BHANU†, M. N. DESHPANDE† and C. P. CHOLKAR*‡
†Institute of Science, Nagpur, India
‡Dharampeth Science College, Nagpur, India

(*Received 29 March 2004*)

In this note, we discuss a coin tossing experiment with a new stopping rule. The two random variables involved in the experiment have some interesting properties.

1. Introduction

Coin tossing is a fascinating subject to study. Many ways of coin tossing are available in the literature. In this note we consider a coin tossing experiment with a new stopping rule and study the associated random variables. The study leads to many interesting results.

2. Experiment

Let n be a predetermined positive integer. An unbiased coin is tossed sequentially. For each outcome we score two points for head and one point for tail. We go on tossing till the total score reaches n or crosses n (i.e. the total score jumps from $n-1$ to $n+1$)

This experiment involves two random variables, H and T, where $H(T)$ denotes number of heads (tails) obtained before stopping.

The following notations are used in the sequel:

(i) $p(x, n)$ – probability that there are x heads and total score is n
(ii) $E_n(\,)$, $V_n(\,)$ – mean and variance for predetermined n.
(iii) $T(n)$ – probability that the total score is n when we stop.
(iv) Σ – summation over x

Here we study the properties of the random variables H and T. The technique of difference equation plays an important role in all our derivations.

2.1. *Distribution of H*

We denote by $p(x, n)$, the probability of getting x heads and a total score of n. This means that the number of tails must be $n-2x$ and total number of tosses must be $n-x$.

Thus $p(x, n)$, the probability of getting x heads, in $(n-x)$ tosses is given by,

$$\binom{n-x}{x}\left(\frac{1}{2}\right)^{n-x} \tag{1}$$

where $x = 0, 1, 2, \ldots [n/2]$ and $[x]$ denotes integral part of x.

---

*Corresponding author. Email: dpratap_ngp@sancharnet.in

Before deriving the distribution of H we prove some Lemmas.

Lemma 1:

$$T(n) = \Sigma p(x, n) = \frac{[2 + (-1/2)^n]}{3}$$

Proof: From I we can easily see that

$$p(x-1, n-1) + p(x, n) = 2p(x, n+1) \tag{2}$$

and summing over x we get

$$T(n-1) + T(n) = 2T(n+1)$$

Initial condition are $T(1) = 1/2$ and $T(2) = 3/4$.
The solution of the above difference equation gives the result.

Lemma 2:

$$\Sigma xp(x, n) = \frac{[(6n-2) + (6n+2)(-1/2)^n]}{27}$$

Proof: Let $A(n) = \Sigma x\, p(x, n)$. Multiplying 2 by x and summing over x we get

$$A(n-1) + T(n-1) + A(n) = 2A(n+1)$$

Initial conditions are $A(1) = 0$ and $A(2) = 1/2$

The solution of the above difference equation proves the lemma.

Lemma 3:

$$\Sigma x(x-1)p(x, n) = \frac{[(6n^2 - 18n + 8) + (12n^2 - 12n - 8)(-1/2)^n]}{81}$$

Proof is similar and hence omitted.

We note that the probability of stopping with x heads is

$$p(x, n) + \left(\frac{1}{2}\right)p(x-1, n-1) = \binom{n-x}{x}\left(\frac{1}{2}\right)^{n-x} + \binom{n-x}{x-1}\left(\frac{1}{2}\right)^{n-x+1}$$

The table below shows the probability distribution of H for a few n.

	Number of heads				
n	0	1	2	3	4
2	1/4	3/4	–	–	–
3	1/8	5/8	2/8	–	–
4	1/16	7/16	8/16	–	–
5	1/32	9/32	18/32	4/32	–
6	1/64	11/64	32/64	20/64	–
7	1/128	13/128	50/128	56/128	8/128
8	1/256	15/256	72/256	120/256	48/256

Since probability of $(H=x)$ is $p(x, n)+(1/2)\,p(x-1, n-1)$, we obtain moments of H using the above lemmas. These are given below.

$$E_n(H) = A(n) + \frac{A(n-1)}{2} + \frac{T(n-1)}{2}$$

$$= \frac{[(3n+1) - (-1/2)^n]}{9}$$

$$E_n(H^2) = \frac{[(3n^2+4n+3) - (4n+3)(-1/2)^n]}{27}$$

$$V_n(H) = \left\{\frac{6n+8}{81} + \frac{1}{81}\left(\frac{-1}{2}\right)^n\right\}\left\{1 - \left(\frac{-1}{2}\right)^n\right\}$$

For large n we note that

$$E_n(H) \cong \frac{3n+1}{9} \quad \text{and} \quad V_n(H) \cong \frac{6n+8}{81}$$

2.2. *Distribution of T*

We note that we stop either after reaching a total score of n or getting head after total score reaches $n-1$. If x heads are obtained and total score is n, then the number of tails will be $n-2x$ and it will be $(n-2x-1)$ if n is skipped and the total score is $n+1$. Thus the probability distribution of the number of tails, T is given by

$$P[T = n-2x] = p(x,n) \quad \text{and}$$

$$P[T = n-2x-1] = \frac{p(x, n-1)}{2}, \quad x = 0, 1, 2, \ldots, \left[\frac{(n+1)}{2}\right]$$

It may be noted that possible values of T are $0, 1, \ldots, n$.

The following table shows the probability distribution of T for a few n

	Number of tails							
n	0	1	2	3	4	5	6	7
2	2/4	1/4	1/4	–	–	–	–	–
3	2/8	4/8	1/8	1/8	–	–	–	–
4	4/16	4/16	6/16	1/16	1/16	–	–	–
5	4/32	12/32	6/32	8/32	1/32	1/32	–	–
6	8/64	12/64	24/64	8/64	10/64	1/64	1/64	–
7	8/128	32/128	24/128	40/128	10/128	12/128	1/128	1/128

The nature of the distribution of T is quite interesting. It has many local maxima. Its shape is zigzag. In practice, it is rare to find such a distribution.

We next evaluate the moments of the random variable T.

We note that

$$E_n(T) = \sum (n-2x)p(x,n) + \left(\frac{1}{2}\right)\left[\sum (n-2x-1)p(x,n-1)\right]$$
$$= nT(n) - 2A(n) + [(n-1)T(n-1)/2] - A(n-1)$$

Using the above lemmas we get

$$E_n(T) = \frac{[(3n+1) - (-1/2)^n]}{9}$$

Remark: Here it is interesting to note that $E_n(H) = E_n(T)$ for every n. It may be noted that possible values of H are $0, 1, \ldots, [(n+1)/2]$ and that of T are $0, 1, 2 \ldots, n$.

It is interesting to note that though supports are different, means of H & T are same.

After extensive algebraic work we get the following.

$$E_n(T^2) = \frac{9n^2 + 30n + 3}{81} + \frac{6n-3}{81}\left(-\frac{1}{2}\right)^n$$

$$V_n(T) = \frac{24n+2}{81} + \frac{12n-1}{81}\left(\frac{-1}{2}\right)^n - \frac{1}{81}\left(\frac{-1}{2}\right)^{2n}$$

For large n variance of T is $(24n+2)/81$

2.3. *Covariance between H and T*

From the above discussion we easily see that

$$E_n(HT) = \Sigma x(n-2x)p(x,n) + \left(\frac{1}{2}\right)[\Sigma (x+1)(n-2x-1)p(x,n-1)]$$

We can simplify RHS using above lemmas and obtain.

$$E_n(HT) = \frac{[3n+1-(-1/2)^n](n-1)}{27}$$

and covariance between H and T is

$$-\frac{12n+4}{81} + \frac{(3n+5)}{81}\left(\frac{-1}{2}\right)^n - \frac{1}{81}\left(\frac{-1}{2}\right)^{2n}$$

3. Concluding remarks

(1) It is seen that for large n, $V_n(H) \cong (6n+8)/81$ and $V_n(T) \cong (24n+2)/81$ and $\mathrm{Cov}_n(H,T)$ is $-(12n+4)/81$. Thus for large n correlation coefficient shall be -1 which is as expected.

(2) Total of our scores when we stop is either n or $n+1$ and probability that score equals n is $T(n)$. Thus expected score when we stop is

$$nT(n) + (n+1)(1-T(n)) = n + \frac{[1-(-1/2)^n]}{3}$$

Expected score is also equal to

$$2E_n(H) + E_n(T) = n + \frac{[1 - (-1/2)^n]}{3}.$$

(3) Distribution of H is always unimodal but that of T is zigzag.
(4) The following relation is noteworthy:

$$\frac{E_n(HT)}{E_n(H)} = \frac{(n-1)}{3}$$

On a class of logarithmic integrals

HONGWEI CHEN* and PARVIZ KHALILI
Department of Mathematics, Christopher Newport University,
Newport News, VA 23606, USA

(*Received 1 September 2004*)

In this note we give closed forms for a class of logarithmic integrals in terms of Bernoulli polynomials. This provides a method for unifying a large class of definite integrals.

1. Introduction

In this note we present a closed form evaluation of the following logarithmic integrals

$$I_n(\theta) = \int_0^1 \frac{\ln(1 - 2x\cos\theta + x^2)}{x} (-\ln x)^n \, dx, \tag{1}$$

where n is a nonnegative integer. This enables us to evaluate a large number of definite integrals. Some of these are well known, by which we mean that they can be computed by a computer algebra system or can be found in a table of integral. We will use Maple as sources for the former and Gradshteyn and Ryzhik [1] for the latter. In this note we have chosen to examine a selection of integrals that can be expressed in terms of Bernoulli polynomials. For example, we show the following integral representation of Bernoulli polynomial of even degree

$$B_{2(n+1)}(\theta) = (-1)^{n+1} \frac{(2n+1)(2n+2)}{(2\pi)^{2(n+1)}} \int_0^1 \frac{\ln\left(1 - 2x\cos 2\pi\theta + x^2\right)}{x} (\ln x)^{2n} \, dx, \tag{2}$$

*Corresponding author. Email: hchen@pcs.cnu.edu

In particular, for $n=0$

$$B_2(\theta) = -\frac{2}{(2\pi)^2}\int_0^1 \frac{\ln\bigl(1-2x\cos 2\pi\theta + x^2\bigr)}{x}\,dx.$$

This yields the main result of [2]:

$$\int_0^1 \frac{\ln\bigl(1-2x\cos\theta + x^2\bigr)}{x}\,dx = \frac{\pi^2}{6} - \frac{1}{2}(\theta-\pi)^2,$$

because $B_2(\theta) = \theta^2 - \theta + 1/6$.

As a consequence of our expression for (1), we also obtain

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^{2n+1}} = \frac{(-1)^{n+1}(2\pi)^{2n+1}}{2(2n+1)(2n)!}\,B_{2n+1}(1/4).$$

Since no closed form is known for $\zeta(n)$ for n odd, we consider this to be one of our most interesting results.

The rest of the note is organized as follows. In section 2 we evaluate (1). Examples involving specific θ are presented in section 3. In section 4 we give additional examples based on differentiation with respect to θ.

2. Derivation of the closed form

We begin with the Poisson kernel

$$\frac{1}{2} + \sum_{k=1}^{\infty} x^k \cos k\theta = \frac{1}{2}\cdot\frac{1-x^2}{1-2x\cos\theta + x^2}, \quad 0 \leqslant x < 1, \quad 0 \leqslant \theta \leqslant 2\pi.$$

Subtracting 1/2 from the left-hand side and then dividing by x, we have

$$\sum_{k=1}^{\infty} x^{k-1}\cos k\theta = \frac{\cos\theta - x}{1-2x\cos\theta + x^2}.$$

Since the series is uniformly convergent for $0 \leqslant x < 1$, integrating both sides from 0 to x, we get

$$2\sum_{k=1}^{\infty}\frac{1}{k}x^k\cos k\theta = -\ln(1-2x\cos\theta + x^2), \quad 0 \leqslant x \leqslant 1, \quad 0 < \theta < 2\pi. \tag{3}$$

This leads to

$$I_n(\theta) = -2\int_0^1 \sum_{k=1}^{\infty}\frac{1}{k}x^{k-1}\cos k\theta\,(-\ln x)^n\,dx.$$

Noting that for $0 < x < 1$,

$$|x^{k-1}(-\ln x)^n| \leqslant n^n/e^n(k-1)^n \quad \text{for } k \geqslant 2,$$

the series in the integrand is uniformly convergent. Therefore, integrating term by term once more gives

$$I_n(\theta) = -2\sum_{k=1}^{\infty}\frac{\cos k\theta}{k}\int_0^1 x^{k-1}(-\ln x)^n\,dx.$$

In view of the fact

$$\int_0^1 x^{k-1}(-\ln x)^n\,dx = \frac{\Gamma(n+1)}{k^{n+1}} = \frac{n!}{k^{n+1}},$$

we have

$$I_n(\theta) = -2\,(n!)\sum_{k=1}^{\infty}\frac{\cos k\theta}{k^{n+2}}. \tag{4}$$

We now distinguish between two cases:

Case 1: n is odd. Noting that

$$I_n(0) = -2\,(n!)\sum_{k=1}^{\infty}\frac{1}{k^{n+2}} = -2\,(n!)\,\zeta(n+2),$$

where $\zeta(n)$ indicates the Riemann zeta function, as we know that so far no one has found the exact value for $\zeta(n)$ for n odd, it seems that there is no closed form for I_n when n is odd.

Case 2: n is even. Using the relation [1, p. 46]

$$\sum_{k=1}^{\infty}\frac{\cos 2\pi k\theta}{k^{2n}} = \frac{(-1)^{n+1}(2\pi)^{2k}}{2(2n)!}\,B_{2n}(\theta)$$

in (4), where $B_n(\theta)$ is the nth Bernoulli polynomial, we have

$$I_{2n}(\theta) = \frac{(-1)^{n+1}(2\pi)^{2(n+1)}}{(2n+2)(2n+1)}\,B_{2(n+1)}(\theta/2\pi).$$

Thus, the desired result (2) follows directly from here. Moreover, the computations of I_n for n even are now in terms of the explicitly known Bernoulli polynomials. For example, since

$$B_2(\theta) = \theta^2 - \theta + \frac{1}{6};$$

$$B_4(\theta) = \theta^4 - 2\theta^3 + \theta^2 - \frac{1}{30};$$

$$B_6(\theta) = \theta^6 - 3\theta^5 + \frac{5}{2}\theta^4 - \frac{1}{2}\theta^2 + \frac{1}{42},$$

the first few $I_n(\theta)$ are found to be

$$I_0(\theta) = -2\pi^2 B_2(\theta/2\pi) = -\frac{1}{3}\pi^2 + \pi\theta - \frac{1}{2}\theta^2;$$

$$I_2(\theta) = \frac{4}{3}\pi^4 B_4(\theta/2\pi) = \frac{1}{12}\theta^4 - \frac{1}{3}\pi\theta^3 + \frac{1}{3}\pi^2\theta^2 - \frac{2}{45}\pi^4;$$

$$I_4(\theta) = -\frac{32}{15}\pi^6 B_6(\theta/2\pi) = -\frac{1}{30}\theta^6 + \frac{1}{5}\pi\theta^5 - \frac{1}{3}\pi^2\theta^4 + \frac{4}{15}\pi^4 x^2 - \frac{16}{315}\pi^6.$$

Remark: In [2], the result of I_0 was derived by the functional equation

$$f(\theta/2) + f(\pi - \theta/2) = \frac{1}{2} f(\theta).$$

The interested reader can verify that

$$I_n(\theta/2) + I_n(\pi - \theta/2) = \frac{1}{2^{n+1}} I_n(\theta).$$

The solutions of this equation provide another approach to evaluate (1).

3. Some special cases

We now compute examples of $I_n(\theta)$ for specific choices of values of n and θ. We tried to compute each example using Maple. In those cases where an answer was thereby obtained, we indicate the number of CPU time taken. In several cases Maple fails to given an answer.

Example 3.1: For $\theta = 0$ in I_2,

$$\int_0^1 \frac{\ln(1-x)(\ln x)^2}{x}\, dx = -\frac{\pi^4}{45}.$$

Maple computes this in 1.8 seconds.

Example 3.2: For $\theta = \pi$ in I_2,

$$\int_0^1 \frac{\ln(1+x)(\ln x)^2}{x}\, dx = \frac{7\pi^4}{360}.$$

Maple computes this in 2.77 seconds.

Example 3.3: For $\theta = 0$ in I_4,

$$\int_0^1 \frac{\ln(1-x)(\ln x)^4}{x}\, dx = -\frac{8\pi^6}{315}.$$

Maple computes this in 4.99 seconds.

Example 3.4: For $\theta = \pi$ in I_4,

$$\int_0^1 \frac{\ln(1+x)(\ln x)^4}{x}\, dx = \frac{31\pi^6}{1260}.$$

Maple computes this in 7.21 seconds.

In general, we have

$$\int_0^1 \frac{\ln(1+x)(-\ln x)^{n-1}}{x}\,dx = (n-1)!\left(1-\frac{1}{2^n}\right)\zeta(n+1);$$

$$\int_0^1 \frac{\ln(1-x)(-\ln x)^{n-1}}{x}\,dx = -(n-1)!\,\zeta(n+1).$$

Taking linear combinations of those results gives

$$\int_0^1 \frac{\ln(1-x^2)(-\ln x)^{n-1}}{x}\,dx = -\frac{1}{2^n}\zeta(n+1);$$

$$\int_0^1 \ln\left(\frac{1+x}{1-x}\right)(-\ln x)^{n-1}\frac{dx}{x} = 2((n-1)!)\left(1-\frac{1}{2^{n+1}}\right)\zeta(n+1).$$

Example 3.5: For $\theta = \pi/2$ in I_2,

$$\int_0^1 \frac{\ln(1+x^2)(\ln x)^2}{x}\,dx = \frac{7\pi^4}{2880}.$$

This could not be evaluated by Maple.

Example 3.6: For $\theta = \pi/4$ and $\theta = 3\pi/4$ in I_2, respectively

$$\int_0^1 \frac{\ln(1-\sqrt{2}x+x^2)(\ln x)^2}{x}\,dx = -\frac{1313\pi^4}{46080};$$

$$\int_0^1 \frac{\ln(1+\sqrt{2}x+x^2)(\ln x)^2}{x}\,dx = \frac{1327\pi^4}{46080}.$$

Linear combination yields

$$\int_0^1 \frac{\ln(1+x^4)(\ln x)^2}{x}\,dx = \frac{7\pi^4}{23040};$$

$$\int_0^1 \ln\left(\frac{1-\sqrt{2}x+x^2}{1+\sqrt{2}x+x^2}\right)(\ln x)^2\frac{dx}{x} = -\frac{11\pi^4}{192}.$$

Those could not be done by Maple either.

4. Differentiation of $I_n(\theta)$

It is easy to see that $I_n(\theta)$ is continuous differentiable from (4). Therefore, differentiating (4) term by term gives

$$I_n'(\theta) = \int_0^1 \frac{2\sin\theta}{1-2x\cos\theta+x^2}(-\ln x)^n\,dx = 2\,n!\sum_{k=1}^{\infty}\frac{\sin k\theta}{k^{n+1}}. \tag{5}$$

Using the relation [1, p. 46]

$$\sum_{k=1}^{\infty}\frac{\sin 2\pi k\theta}{k^{2n+1}} = \frac{(-1)^{n+1}(2\pi)^{2n+1}}{2(2n+1)!}B_{2n+1}(\theta)$$

we have

$$I'_{2n}(\theta) = \frac{(-1)^{n+1}(2\pi)^{2n+1}}{2n+1} B_{2n+1}(\theta/2\pi). \tag{6}$$

This deduces a integral representation for Bernoulli polynomials of odd degree

$$B_{2n+1}(\theta) = (-1)^{n+1}\frac{2(2n+1)}{(2\pi)^{2n+1}}\int_0^1 \frac{\sin(2\pi\theta)}{1-2x\cos 2\pi\theta + x^2}(\ln x)^{2n}\,dx.$$

Example 4.1: For $\theta = \pi/2$ in $I'_1(\theta)$,

$$\int_0^1 \frac{-\ln x\,dx}{1+x^2} = \sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^2} = \mathbf{G},$$

where $\mathbf{G} = 0.915965\cdots$ is the Catalan's constant, which Maple evaluated in 1.02 seconds.

Example 4.2: For $\theta = \pi/2$ in $I'_2(\theta)$,

$$\int_0^1 \frac{(\ln x)^2}{1+x^2}\,dx = \frac{\pi^3}{16}$$

(see [1, p. 572]). This could not be evaluated by Maple. On the other hand, setting $\theta = \pi/2$ in (5) gives

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^3} = 1 - \frac{1}{3^3} + \frac{1}{5^3} - \cdots = \frac{\pi^3}{32}.$$

Example 4.3: For $\theta = \pi/2$ in I'_4,

$$\int_0^1 \frac{(\ln x)^4}{1+x^2}\,dx = \frac{5\pi^5}{64}$$

(see [1, p. 574]). This could not be evaluated by Maple. This is equivalent to

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^5} = 1 - \frac{1}{3^5} + \frac{1}{5^5} - \cdots = \frac{5\pi^5}{1536}.$$

In general, we have

$$\int_0^1 \frac{(-\ln x)^n}{1+x^2}\,dx = n!\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^{n+1}}.$$

In particular,

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^{2n+1}} = \frac{(-1)^{n+1}(2\pi)^{2n+1}}{2(2n+1)(2n)!} B_{2n+1}(1/4).$$

Astonishingly, there is a closed form for the alternating odd term only $\zeta(n)$ for all $n \geqslant 1$.

Example 4.4: For $\theta = \pi/3$ in $I'_2(\theta)$,

$$\int_0^1 \frac{(\ln x)^2}{1-x+x^2}\,dx = \frac{10\pi^3}{81\sqrt{3}}$$

(see [1, p. 571]). This could not be done by Maple either.

Example 4.5: Let $\theta \to 0$. Then (5) yields

$$\int_0^1 \frac{(-\ln x)^n}{(1-x)^2}\,dx = \frac{1}{2}\lim_{\theta\to 0}\frac{I'_n(\theta)}{\sin\theta}.$$

In view of the fact that $B_3(\theta) = \theta^3 - 3/2\theta^2 + 1/2\theta$, we deduce the well-known Euler integral from (6) ([1, p. 572])

$$\int_0^1 \frac{(\ln x)^2}{(1-x)^2}\,dx = \frac{\pi^2}{3},$$

which Maple evaluated in 0.16 seconds.

Remark: Further differentiation of (1) will produce more examples of integrals that can be evaluated in closed form. For example,

$$\int_0^1 \frac{x(\ln x)^{2n}}{(1+x^2)^2}\,dx = \frac{(2\pi)^{2n}}{4}\,B_{2n}(1/4).$$

5. Conclusions

In terms of (1), we have demonstrated how to evaluate many classical integrals by specifying the θ or by differentiation with respect to it. Furthermore, we have been able to evaluate a large number of other integrals which cannot be found in [1] and cannot be evaluated by Maple. Within this framework, the interested reader is encouraged to discover additional integral formulas.

References

[1] Gradshteyn, I.S. and Ryzhik, I.M., 1994, *Table of Integrals, Series, and Products*, 5th edn (New York: Academic Press).
[2] Haruki, H. and Haruki, S., 1983, Euler's integrals. *Am. Math. Monthly*, **7**, 465–466.