THE EXPECTED VARIATION OF RANDOM BOUNDED INTEGER SEQUENCES OF FINITE LENGTH

RUDOLFO ANGELES, DON RAWLINGS,
LAWRENCE SZE, AND MARK TIEFENBRUCK

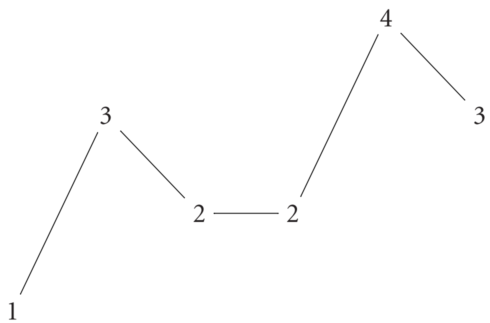
Received 24 November 2004 and in revised form 15 June 2005

From the enumerative generating function of an abstract adjacency statistic, we deduce the mean and variance of the variation on random permutations, rearrangements, compositions, and bounded integer sequences of finite length.

1. Introduction

When the finite sequence of integers $w = 1,3,2,2,4,3$ is sketched as below,

$w =$



(1.1)

its most compelling aspect is its vertical *variation*, that is, the sum of the vertical distances between its adjacent terms. Denoted by $\operatorname{var} w$, the vertical variation of the sequence in (1.1) is $\operatorname{var} w = 2+1+0+2+1 = 6$. Our purpose here is to compute the mean and variance of var on four classical sets of combinatorial sequences.

To formalize matters and place our problem in the context of other work, let $[m]^n$ denote the set of sequences $w = x_1x_2\cdots x_n$ of length n with each $x_i \in \{1,2,\ldots,m\}$. For a real-valued function f on $[m]^2$, the f*-adjacency number* of $w = x_1x_2\cdots x_n \in [m]^n$ is defined to be

$$\operatorname{adf} w = \sum_{k=1}^{n-1} f\left(x_k x_{k+1}\right). \tag{1.2}$$

International Journal of Mathematics and Mathematical Sciences 2005:14 (2005) 2277–2285
DOI: 10.1155/IJMMS.2005.2277

Table 1.1

Sequences	Expected value of var	Variance of var
S_n	$\dfrac{n^2-1}{3}$	$\dfrac{(n-2)(n+1)(4n-7)}{90}$
$[m]^n$	$\dfrac{(n-1)(m^2-1)}{3m}$	$\dfrac{(m^2-1)(6m^2n+6n-7m^2-2)}{90m^2}$
$R_n(\vec{i})$	$\dfrac{2}{n}\sum_{1\le x<y\le m}(y-x)i_xi_y$	See (3.10)
$C_n(m)$	$\dfrac{2(n-1)}{(m-1)^{\underline{n-1}}}\sum_{x=1}^{\lfloor m/2\rfloor}(m-2x)^{\underline{n-1}}$	See (3.18)

Some specializations of the f-adjacency number have been considered elsewhere. For instance, if $f(xy)$ is 1 when $x<y$ and 0 otherwise, then $\operatorname{adf} w$ is known as the *rise number* of w [1, 3, 4]. For the selection $f(xy)=|y-x|$, $\operatorname{adf} w = \operatorname{var} w$. In a sorting problem of computer science, Levcopoulos and Petersson [5] introduced the related notion of oscillation $(\operatorname{var} w - n + 1)$ as a measure of the presortedness of a sequence of n *distinct* numbers. In [6], compositions were enumerated by their *ascent variation*, the f-adjacency statistic induced by $f(xy)=y-x$ if $x<y$ and 0 otherwise. For the case $f(xy)=h(|y-x|)$ where h is a linear, convex, or concave increasing real-valued function, Chao and Liang [2] described the arrangements of n distinct integers for which adf achieves its extreme values.

Besides considering the distribution of var on the set $[m]^n$, we also consider it on the set of rearrangements $R_n(i_1,i_2,\dots,i_m)$ consisting of sequences of length $n=i_1+i_2+\cdots+i_m$ which contain l exactly i_l times, on the set of permutations $S_n=R_n(1,1,\dots,1)$ of $\{1,2,\dots,n\}$, and on the set of compositions of m into n parts $C_n(m)=\{x_1x_2\cdots x_n\in[m]^n : x_1+x_2+\cdots+x_n=m\}$. For $m,n\ge 2$, Table 1.1 displays the mean and variance of var on these four sets. The kth falling factorial of n is $n^{\underline{k}}=n(n-1)\cdots(n-k+1)$, $\vec{i}=(i_1,i_2,\dots,i_m)$, and, for r a real number, $\lfloor r\rfloor$ denotes the greatest integer less than or equal to r. The results in Table 1.1 are new. David and Barton [3, Chapter 10] present the distributions of several statistics (some f-adjacency numbers, some not) primarily on permutations. We also note that Tiefenbruck [7] derived a generating function for compositions with bounded parts by a close relative of var. We leave open questions concerning the asymptotic behavior of var.

2. Enumerative factorial moments for f-adjacencies

Before working specifically with var, we discuss the enumerative generating function for adf on sequences as developed by Fédou and Rawlings [4]. Let $[m]^*$ denote the set of sequences of 1, 2, …, m of finite length (including the empty sequence of length 0). For $w=x_1x_2\cdots x_n\in[m]^*$, we define $Z^w=z_{x_1}z_{x_2}\cdots z_{x_n}$. The enumerative generating function for adf over $[m]^*$ is then defined to be $G(p)=\sum_{w\in[m]^*}p^{\operatorname{adf} w}Z^w$.

By manipulating $G(p)$, we will obtain all of the information in Table 1.1 (and more). As a brief outline of our approach, note that the coefficient of $p^k z_1^{i_1}z_2^{i_2}\cdots z_m^{i_m}$ in $G(p)$ is

just the number of rearrangements w in $R_n(\vec{i})$ with adf $w = k$. Thus, by dividing the coefficient of $z_1^{i_1} z_2^{i_2} \cdots z_m^{i_m}$ in $G'(1)$ by the cardinality of $R_n(\vec{i})$, we will obtain the mean of adf. So, in general, we compute the dth *enumerative factorial moment* $G^{(d)}(1) = \sum_{w \in [m]^*} (\text{adf}\, w)^{\underline{d}} Z^w$.

From the work of Fédou and Rawlings [4], it follows that

$$G(p) = \frac{1}{D(p)}, \tag{2.1}$$

where

$$D(p) = 1 - \sum_{n \geq 1} \sum_{x_1 \cdots x_n \in [m]^n} Z^{x_1 \cdots x_n} \prod_{k=1}^{n-1} \left(p^{f(x_k x_{k+1})} - 1\right). \tag{2.2}$$

Examples are presented in [4, 6] for which D has a closed form. We do not know a closed form for D when adf = var (that is, when $f(x, y) = |y - x|$). Nevertheless, (2.1) is still useful in computing the mean and variance of var.

Although the formula for taking the d-fold derivative with respect to p of a function of the form in (2.1) is known, we provide a short derivation. To avoid the quotient and chain rules, rewrite (2.1) as $GD = 1$. Differentiating the latter d times, $d \geq 1$, and dividing by $d!$ gives

$$\sum_{j=0}^{d} \frac{G^{(d-j)}}{(d-j)!} \frac{D^{(j)}}{j!} = 0. \tag{2.3}$$

To solve for $G^{(d)}$, consider the system

$$\begin{aligned}
\frac{G^{(d)}}{d!}\frac{D^{(0)}}{0!} + \frac{G^{(d-1)}}{(d-1)!}\frac{D^{(1)}}{1!} + \frac{G^{(d-2)}}{(d-2)!}\frac{D^{(2)}}{2!} + \cdots + \frac{G^{(0)}}{0!}\frac{D^{(d)}}{d!} &= 0, \\
\frac{G^{(d-1)}}{(d-1)!}\frac{D^{(0)}}{0!} + \frac{G^{(d-2)}}{(d-2)!}\frac{D^{(1)}}{1!} + \cdots + \frac{G^{(0)}}{0!}\frac{D^{(d-1)}}{(d-1)!} &= 0, \\
&\vdots \\
\frac{G^{(1)}}{1!}\frac{D^{(0)}}{0!} + \frac{G^{(0)}}{0!}\frac{D^{(1)}}{1!} &= 0, \\
\frac{G^{(0)}}{0!}\frac{D^{(0)}}{0!} &= 1,
\end{aligned} \tag{2.4}$$

where the top d equations arise from repeated application of (2.3). Cramer's rule applied

to the above system yields

$$\frac{G^{(d)}}{d!}=\frac{(-1)^d}{D^{d+1}}\begin{vmatrix} \frac{D^{(1)}}{1!} & \frac{D^{(2)}}{2!} & \frac{D^{(3)}}{3!} & \cdots & & \frac{D^{(d)}}{d!} \\ \frac{D^{(0)}}{0!} & \frac{D^{(1)}}{1!} & \frac{D^{(2)}}{2!} & & & \frac{D^{(d-1)}}{(d-1)!} \\ 0 & \frac{D^{(0)}}{0!} & \frac{D^{(1)}}{1!} & & & \frac{D^{(d-2)}}{(d-2)!} \\ \vdots & & & & & \vdots \\ 0 & & \cdots & 0 & \frac{D^{(0)}}{0!} & \frac{D^{(1)}}{1!} \end{vmatrix} \tag{2.5}$$

which, when expanded, implies that

$$G^{(d)}=\sum_{\nu=1}^{d}\frac{(-1)^{\nu}}{D^{\nu+1}}\sum_{\substack{j_1+\cdots+j_\nu=d\\ j_k\geq 1}}\binom{d}{j_1\cdots j_\nu}D^{(j_1)}\cdots D^{(j_\nu)}. \tag{2.6}$$

To determine the enumerative factorial moment $G^{(d)}(1)$, we see from (2.2) that

$$D^{(j)}(1)=-\sum_{r=2}^{j+1}D_r^{(j)}, \tag{2.7}$$

where

$$D_r^{(j)}=\sum_{x_1\cdots x_r\in[m]^r}Z^{x_1\cdots x_r}\sum_{\substack{l_1+\cdots+l_{r-1}=j\\ l_k\geq 1}}\binom{j}{l_1\cdots l_{r-1}}\prod_{k=1}^{r-1}f\left(x_kx_{k+1}\right)^{\underline{l_k}}. \tag{2.8}$$

For instance,

$$\begin{gathered} D_2'=\sum_{xy\in[m]^2}f(xy)z_xz_y, \qquad D_2''=\sum_{xy\in[m]^2}f(xy)^{\underline{2}}z_xz_y, \\ D_3''=2\sum_{vxy\in[m]^3}f(vx)f(xy)z_vz_xz_y. \end{gathered} \tag{2.9}$$

Further setting $\vec{j}=(j_1,\dots,j_\nu)$, $s(\vec{j})=j_1+\cdots+j_\nu$,

$$\binom{d}{\vec{j}}=\binom{d}{j_1\cdots j_\nu}, \qquad \text{and} \qquad D_\mu^{(\vec{j})}=\sum_{\substack{r_1+\cdots+r_\nu=\mu\\ r_k\geq 2}}D_{r_1}^{(j_1)}\cdots D_{r_\nu}^{(j_\nu)}, \tag{2.10}$$

it follows from (2.6) and (2.7) that

$$G^{(d)}(1)=\sum_{\nu=1}^{d}\frac{1}{D^{\nu+1}(1)}\sum_{\substack{s(\vec{j})=d\\ j_k\geq 1}}\binom{d}{\vec{j}}\sum_{\mu=2\nu}^{d+\nu}D_\mu^{(\vec{j})}. \tag{2.11}$$

As $D(1) = 1 - (z_1 + \cdots + z_m)$, extracting the contributions made by all $w \in [m]^n$ from both sides of (2.11) gives the dth enumerative factorial moment of adf over $[m]^n$ as

$$\sum_{w\in[m]^n} (\operatorname{adf} w)^{\underline{d}} Z^w = \sum_{\nu=1}^{d} \sum_{\substack{s(\vec{j})=d \\ j_k \geq 1}} \binom{d}{\vec{j}} \sum_{\mu=2\nu}^{d+\nu} \binom{n+\nu-\mu}{\nu} \left(\sum_{i=1}^{m} z_i \right)^{n-\mu} D_\mu^{(\vec{j})} \tag{2.12}$$

valid for $d \geq 1$. When $d = 1, 2$, (2.9) and (2.12) imply that

$$\sum_{w\in[m]^n} \operatorname{adf} w\, Z^w = (n-1)\left(\sum_{i=1}^{m} z_i \right)^{n-2} \sum_{xy\in[m]^2} f(xy) z_x z_y \tag{2.13}$$

and that

$$\begin{aligned} \sum_{w\in[m]^n} (\operatorname{adf} w)^{\underline{2}} Z^w = {} & (n-1)\left(\sum_{i=1}^{m} z_i \right)^{n-2} \sum_{xy\in[m]^2} \left(f(xy)\right)^{\underline{2}} z_x z_y \\ & + 2(n-2)\left(\sum_{i=1}^{m} z_i \right)^{n-3} \sum_{vxy\in[m]^3} f(vx) f(xy) z_v z_x z_y \\ & + (n-2)(n-3)\left(\sum_{i=1}^{m} z_i \right)^{n-4} \left(\sum_{xy\in[m]^2} f(xy) z_x z_y \right)^2. \end{aligned} \tag{2.14}$$

3. Discussion of Table 1.1

The entries in Table 1.1 are consequences of (2.13) and (2.14) with $f(xy) = |y - x|$ and with appropriate substitutions for Z. For the mean and variance of var on the set of bounded sequences $[m]^n$, put $z_i = 1$ for $1 \leq i \leq m$. Noting that

$$\sum_{xy\in[m]^2} |y-x| = \sum_{1\leq x<y\leq m} 2(y-x) = 2\binom{m+1}{3}, \tag{3.1}$$

it follows from (2.13) that the mean of var on $[m]^n$ is

$$\frac{1}{m^n} \sum_{w\in[m]^n} \operatorname{var} w = \frac{2(n-1)m^{n-2}}{m^n} \binom{m+1}{3} = \frac{(n-1)(m^2-1)}{3m}. \tag{3.2}$$

As

$$\sum_{xy\in[m]^2} |y-x|^{\underline{2}} = 4\binom{m+1}{4} \tag{3.3}$$

and as

$$\sum_{vxy\in[m]^n}|x-v||y-x| = \sum_{1\le v<x<y\le m} 2(x-v)(y-x) + \sum_{1\le x<y\le v\le m} 4(v-x)(y-x) - \sum_{1\le x<y\le m} 2(y-x)^2 = \frac{7m^2-8}{10}\binom{m+1}{3}, \tag{3.4}$$

(2.14) implies that

$$\frac{1}{m^n}\sum_{w\in[m]^n}(\operatorname{var} w)^2 = \frac{4(n-1)}{m^2}\binom{m+1}{4} + \frac{(n-2)(7m^2-8)}{5m^3}\binom{m+1}{3} + \frac{4(n-2)(n-3)}{m^4}\binom{m+1}{3}^2. \tag{3.5}$$

Then, subbing the last result into

$$\frac{1}{m^n}\sum_{w\in[m]^n}(\operatorname{var} w)^2 + \frac{(n-1)(m^2-1)}{3m} - \left(\frac{(n-1)(m^2-1)}{3m}\right)^2 \tag{3.6}$$

and simplifying gives the variance of var as recorded in Table 1.1.

For $R_n(\vec{i})$, extracting the coefficient of $z_1^{i_1}z_2^{i_2},\cdots,z_m^{i_m}$ from (2.13) leads to

$$\sum_{w\in R_n(\vec{i})}\operatorname{var} w = 2(n-1)\sum_{1\le x<y\le m}(y-x)\binom{n-2}{i_1\cdots i_x-1\cdots i_y-1\cdots i_m}. \tag{3.7}$$

As the cardinality of $R_n(\vec{i})$ is

$$\binom{n}{i_1 i_2\cdots i_m} = \binom{n}{\vec{i}}, \tag{3.8}$$

it follows that the mean of var on $R_n(\vec{i})$ is

$$\binom{n}{\vec{i}}^{-1}\sum_{w\in R_n(\vec{i})}\operatorname{var} w = \frac{2}{n}\sum_{1\le x<y\le m}(y-x)i_x i_y. \tag{3.9}$$

Let $\vec{i}_{\backslash r} = (i_1,\dots,i_r-1,\dots,i_n)$. For example, $(3,2,1,4)_{\backslash 3\backslash 2\backslash 3} = (3,1,-1,4)$. The variance on $R_n(\vec{i})$ is then

$$\binom{n}{\vec{i}}^{-1}\sum_{w\in R_n(\vec{i})}\operatorname{var} w^2 + \frac{2}{n}\sum_{1\le x<y\le m}(y-x)i_x i_y - \left(\frac{2}{n}\sum_{1\le x<y\le m}(y-x)i_x i_y\right)^2, \tag{3.10}$$

where, upon extraction of the coefficient of $z_1^{i_1} z_2^{i_2}, \ldots, z_m^{i_m}$ from (2.14), we have

$$\begin{aligned}\sum_{w\in R_n(\vec{i})} (\operatorname{var} w)^{\underline{2}} = (n-1) \sum_{1\le x,\, y\le m} |y-x|^{\underline{2}} \binom{n-2}{\vec{i}_{\backslash x\backslash y}} \\ + 2(n-2) \sum_{1\le v,x,y\le m} |x-v|\,|y-x| \binom{n-3}{\vec{i}_{\backslash v\backslash x\backslash y}} \\ + (n-2)(n-3) \sum_{1\le u,v,x,y\le m} |v-u|\,|y-x| \binom{n-4}{\vec{i}_{\backslash u\backslash v\backslash x\backslash y}}.\end{aligned} \tag{3.11}$$

The permutation entries in Table 1.1 follow from (3.9) and (3.10). Selecting $m = n$ and $i_k = 1$ for $1 \le k \le n$ in (3.9) reveals the mean of var on S_n as

$$\frac{1}{n!} \sum_{w\in S_n} \operatorname{var} w = \frac{2}{n} \sum_{1\le x<y\le n} (y-x) = \frac{2}{n} \binom{n+1}{3} = \frac{n^2-1}{3}. \tag{3.12}$$

From (3.11), with $m = n$ and $i_k = 1$ for $1 \le k \le n$,

$$\begin{aligned}\sum_{w\in S_n} (\operatorname{var} w)^{\underline{2}} = (n-1)! \sum_{1\le x,\, y\le n} |y-x|^{\underline{2}} \\ + 2(n-2)! \sum_{1\le v,x,y\le n} |x-v|\,|y-x| \\ + (n-2)! \sum_{\substack{1\le u,v,x,y\le m \\ \{u,v\}\cap\{x,y\}=\varnothing}} |v-u|\,|y-x| \\ = \left(\frac{4}{15}\right)(n-2)!\left(10n^2+14n-27\right)\binom{n+1}{4}.\end{aligned} \tag{3.13}$$

So the variance of var on S_n is

$$\frac{1}{n!} \sum_{w\in S_n} \operatorname{var} w^{\underline{2}} + \frac{n^2-1}{3} - \left(\frac{n^2-1}{3}\right)^2 = \frac{(n-2)(n+1)(4n-7)}{90}. \tag{3.14}$$

For $w = x_1 \cdots x_n \in [m]^n$, let $\|w\| = x_1 + \cdots + x_n$. For the composition results in Table 1.1, set $z_k = q^k$ for $1 \le k \le m$. Then (2.13) implies that

$$\sum_{w\in[m]^n} \operatorname{var} w\, q^{\|w\|} = (n-1)q^{n-2}\left(\frac{1-q^m}{1-q}\right)^{n-2} \sum_{1\le x,\, y\le m} |y-x| q^{x+y} \tag{3.15}$$

and (2.14) leads to

$$\begin{aligned}\sum_{w\in[m]^n}(\operatorname{var}w)^{\underline{2}}q^{\|w\|} &= (n-1)q^{n-2}\left(\frac{1-q^m}{1-q}\right)^{n-2}\sum_{1\le x,\,y\le m}|y-x|^{\underline{2}}q^{x+y}\\ &\quad+2(n-2)q^{n-3}\left(\frac{1-q^m}{1-q}\right)^{n-3}\sum_{1\le v,x,y\le m}|x-v|\,|y-x|q^{v+x+y}\\ &\quad+(n-2)(n-3)q^{n-4}\left(\frac{1-q^m}{1-q}\right)^{n-4}\sum_{1\le u,v,x,y\le m}|v-u|\,|y-x|q^{u+v+x+y}.\end{aligned}\tag{3.16}$$

Extracting the coefficient of q^m from (3.15) to obtain

$$\begin{aligned}\sum_{w\in C_n(m)}\operatorname{var}w &= 2(n-1)\sum_{1\le x<y\le m}(y-x)\binom{m-1-x-y}{n-3}\\ &= 2(n-1)\sum_{1\le x\le\lfloor m/2\rfloor}\binom{m-2x}{n-1}\end{aligned}\tag{3.17}$$

and then dividing by the cardinality $\binom{m-1}{n-1}$ of $C_n(m)$ gives the mean of var as stated in Table 1.1. The variance is

$$\begin{aligned}\binom{m-1}{n-1}^{-1}\sum_{w\in C_n(m)}\operatorname{var}w^{\underline{2}}+\frac{2(n-1)}{(m-1)^{\underline{n-1}}}\sum_{1\le x\le\lfloor m/2\rfloor}(m-2x)^{\underline{n-1}}\\ -\left(\frac{2(n-1)}{(m-1)^{\underline{n-1}}}\sum_{1\le x\le\lfloor m/2\rfloor}(m-2x)^{\underline{n-1}}\right)^2,\end{aligned}\tag{3.18}$$

where, pulling the coefficient of q^m from (3.16), we have

$$\begin{aligned}\sum_{w\in C_n(m)}(\operatorname{var}w)^{\underline{2}} &= (n-1)\sum_{1\le x,\,y\le m}|y-x|^{\underline{2}}\binom{m-1-x-y}{n-3}\\ &\quad+2(n-2)\sum_{1\le v,x,y\le m}|x-v|\,|y-x|\binom{m-1-v-x-y}{n-4}\\ &\quad+(n-2)(n-3)\sum_{1\le u,v,x,y\le m}|v-u|\,|y-x|\binom{m-1-u-v-x-y}{n-5}.\end{aligned}\tag{3.19}$$

The sums in (3.19) are marginally simplified. For instance,

$$\sum_{1\le x,\,y\le m}|y-x|^{\underline{2}}\binom{m-1-x-y}{n-3}=4\sum_{1\le x\le\lfloor m/2\rfloor}\binom{m-2x}{n}.\tag{3.20}$$

As a part of the second sum on the right-hand side of (3.19), we note that

$$\begin{aligned}&\sum_{1\le v<x<y\le m}(x-v)(y-x)\binom{m-1-v-x-y}{n-4}\\&\quad=\sum_{2\le x\le\lfloor (m+1)/2\rfloor}\left(\binom{m-3x+1}{n}-\binom{m-2x+1}{n}+x\binom{m-2x}{n-1}\right).\end{aligned}\tag{3.21}$$

The four-fold sums arising in the last sum in (3.19) reduce to double sums.

Acknowledgment

This work is based on work supported by the National Science Foundation under Grant no. 0097392.

References

[1] L. Carlitz, *Enumeration of compositions by rises, falls and levels*, Math. Nachr. **77** (1977), 361–371.

[2] C.-C. Chao and W.-Q. Liang, *Arranging n distinct numbers on a line or a circle to reach extreme total variations*, European J. Combin. **13** (1992), no. 5, 325–334.

[3] F. N. David and D. E. Barton, *Combinatorial Chance*, Hafner, New York, 1962.

[4] J.-M. Fédou and D. P. Rawlings, *Adjacencies in words*, Adv. in Appl. Math. **16** (1995), no. 2, 206–218.

[5] C. Levcopoulos and O. Petersson, *Heapsort—adapted for presorted files*, Algorithms and Data Structures (Ottawa, ON, 1989), Lecture Notes in Comput. Sci., vol. 382, Springer, Berlin, 1989, pp. 499–509.

[6] D. P. Rawlings, *Restricted words by adjacencies*, Discrete Math. **220** (2000), no. 1-3, 183–200.

[7] M. Tiefenbruck, *Enumerating compositions with bounded parts by variation*, REU report, California Polytechnic State University, California, 2003.

Rudolfo Angeles: Department of Mathematics, College of Science and Mathematics, California, Polytechnic State University, San Luis Obispo, CA 93407, USA

Current address: Department of Statistics, Stanford University, Palo Alto, CA 94305-4065, USA

E-mail address: rangeles@stanford.edu

Don Rawlings: Department of Mathematics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407, USA

E-mail address: drawling@calpoly.edu

Lawrence Sze: Department of Mathematics, College of Science and Mathematics, California, Polytechnic State University, San Luis Obispo, CA 93407, USA

E-mail address: lsze@calpoly.edu

Mark Tiefenbruck: Department of Mathematics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407, USA

Current address: 8989 Jasmine Lane Street, Cottage Grove, MN 55016-3436, USA

E-mail address: mdash@alumni.northwestern.edu