ELSEVIER

Discrete Mathematics 213 (2000) 123–139

DISCRETE MATHEMATICS

Optimal linear arrangement of a rectangular grid

Peter Fishburn[a], Prasad Tetali[b, *], Peter Winkler[c]

[a] AT&T Labs-Research, 180 Park Avenue, Florham Park, NJ 07932-0971, USA

[b] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA

[c] Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, USA

Received 21 April 1997; revised 1 September 1997; accepted 14 February 1998

Abstract

An *optimal linear arrangement* of a finite simple graph $G=(V,E)$ with vertex set V, edge set E, and $|V|=N$, is a map f from V onto $\{1,2,\ldots,N\}$ that minimizes $\sum_{\{u,v\}\in E}|f(u)-f(v)|$. We determine optimal linear arrangements for $m\times n$ rectangular grids where $V=\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}$ and $E=\{\{(i,j),(k,\ell)\}\colon |i-k|+|j-\ell|=1\}$. When $m\geqslant n\geqslant 5$, they are disjoint from bandwidth-minimizing arrangements for which f minimizes the maximum $|f(u)-f(v)|$ over E. The different solutions to the bandwidth and linear arrangement problems for rectangular grids is reminiscent of Harper's result (J. Soc. Ind. Appl. Math. 12 (1964) 131–135; J. Combin. Theory 1 (1966) 385–393) of different bandwidth and linear arrangement solutions for the hypercube graph with vertex set $\{0,1\}^n$ and edge set $\{\{(x_1,x_2,\ldots,x_n),\ (y_1,y_2,\ldots,y_n)\}\colon \sum_i |x_i-y_i|=1\}$.

1. Introduction

In this work, we report an optimal labeling (with integers $1,\ldots,mn$) of the vertices of an $m\times n$ grid graph which minimizes the sum of the *weights* of the edges, where the weight of an edge is the absolute difference in the labels of its incident vertices. This problem on a general graph is commonly referred to as the *linear arrangement problem*, and sometimes also as the *wire-length problem*. For an excellent overview of results on this problem and other *discrete isoperimetric problems*, the reader is referred to Bezrukov [1] and Chavez and Harper [2].

Since proving our results, we were informed by Sergej Bezrukov about the original results of Muradjan and Piliposjan on the linear arrangement problem for the rectangular grid. It turns out that this work (see [5–7]) appears only in Russian (with abstracts in Armenian), and due to space constraints with the volume where these results were published, the proofs are very sketchy; Muradjan apparently never published

* Corresponding author.
E-mail address: tetali@math.gatech.edu (P. Tetali)

0012-365X/00/$ - see front matter
PII: S0012-365X(99)00173-9

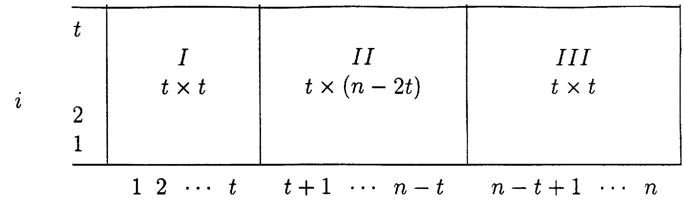


Fig. 1.

the complete version of his results. For this reason, we were encouraged by Bezrukov to publish our results with proofs in detail. A brief summary of Muradjan's results can also be found in [1]. Larry Harper informed us of the work of Mitchison and Durbin [4] (who were also seemingly unaware of Muradjan's work) on the exact solution of the linear arrangement problem on the $(n \times n)$ *square* grid.

Notation and terminology: Throughout, G_{mn} for $m \geqslant n \geqslant 2$ is the graph with vertex set $V_{mn} = \{1,2,\ldots,m\} \times \{1,2,\ldots,n\}$ and edge set

$$E_{mn} = \{\{(i,j),(k,\ell)\}\colon (i,j),(k,\ell) \in V_{mn} \text{ and } |i-k| + |j-\ell| = 1\}.$$

We refer to a map f from V_{mn} onto $\{1,2,\ldots,mn\}$ as an *assignment* and to the $m \times n$ matrix $[a_{ij}]$ with $a_{ij} = f(i,j)$ as an *assignment matrix*. Rows are numbered $1,\ldots,m$ from bottom to top; columns $1,\ldots,n$ from left to right.

An assignment f is *doubly monotonic* if $f(i+1,j) > f(i,j)$ and $f(i,j+1) > f(i,j)$ for all applicable (i,j) and is *complementary* if

$$f(i,j) + f(m+1-i, n+1-j) = mn+1$$

for all $(i,j) \in V_{mn}$.

A few special definitions are needed before we state our main theorem.

Given $t \in \{1,2,\ldots,\lfloor n/2 \rfloor\}$, we consider a 3-part partition of the first t rows of an assignment matrix (see Fig. 1).

The following definitions apply to a doubly monotonic f with assignment matrix $[a_{ij}]$:

(I) Section I is an *up staircase* if, for each $i \in \{1,\ldots,t\}$, $a_{i1}, a_{i2}, \ldots, a_{ii}$ are consecutive with $a_{ii} = i^2$.

(II) Section II is a *vertical slats* section if, for each $j \in \{t+1,\ldots,n-t\}$, $a_{1j}, a_{2j}, \ldots, a_{tj}$ are consecutive with $a_{1j} = (j-1)t + 1$.

(III) Section III is a *down staircase* if, for each $i \in \{1,\ldots,t\}$, $a_{i,n-i+1}, a_{i,n-i+2}, \ldots, a_{in}$ are consecutive with $a_{in} = nt - t(t+1)/2 + i(i+1)/2$, and, for each $j \in \{n-t+1,\ldots,n-1\}$, $a_{1j}, a_{2j}, \ldots, a_{n-j,j}$ are consecutive with $a_{1j} = a_{n-j+1,n} - (n-j+1)^2 + 1$.

The first t rows of an assignment matrix for G_{mn} or V_{mn} *have pattern* $R_t(m,n)$ if (I), (II) and (III) hold for the given t. $R_4(m,11)$ is given in Fig. 2.

In addition, row i of $[a_{ij}]$ is a *horizontal slat* if $a_{i1}, a_{i2}, \ldots, a_{in}$ are consecutive with $a_{i1} = n(i-1) + 1$ and $a_{ij} = ni$.

4	13	14	15	16	20	24	28	41	42	43	44
3	7	8	9	12	19	23	27	31	38	39	40
2	3	4	6	11	18	22	26	30	33	36	37
1	1	2	5	10	17	21	25	29	32	34	35
	1	2	3	4	5	6	7	8	9	10	11

Fig. 2.

Main Theorem. *Suppose $m \geqslant n \geqslant 2$. Denote by F the set of all assignments from V_{mn} onto $\{1,2,\dots,mn\}$, and for all $f \in F$ define $L(f)$ by*

$$L(f) = \sum_{\{u,v\}\in E_{mn}} |f(u) - f(v)|.$$

For each $t \in \{1,2,\dots,\lfloor n/2 \rfloor\}$ denote by $f^{(t)}$ the member of F that is doubly monotonic, complementary, has pattern $R_t(m,n)$ in the first t rows of its assignment matrix with horizontal slats in rows $t+1$ through $m-t$. Then

$$\min_{f\in F} L(f) = L(f^{(t^*)})$$
$$= m(n^2+n-1) - n - t^*[2(t^*)^2 - 6nt^* + 3n^2 + 3n - 2]/3,$$

where t^* denotes a value of $t \in \{1,2,\dots,\lfloor n/2 \rfloor\}$ that maximizes $2t^3 - 6nt^2 + (3n^2 + 3n - 2)t$.

When $m=n=9$, the unique maximizing t is $t^*=3$, and the optimal linear arrangement for G_{99} according to the preceding theorem is as follows. The top three rows are determined by complementarity and pattern $R_3(9,9)$ in the bottom three rows (see Fig. 3).

An optimal arrangement for the 20×20 grid is given below in Fig. 4.

Section 2 notes that we need only consider doubly monotonic assignments in minimizing $L(f)$. Section 3 proves the main theorem when $m = n$, and Section 4 extends the proof to $m > n$.

Remark. 1. The proof of the Main Theorem shows that every optimal assignment is essentially of the form given, but there are several sources of technical non-uniqueness. The construction in the lower left and right corners admits some small variation; there are the obvious symmetries of the rectangle and square; and for some values of n there are two maximizing values of t (e.g. $t^* = 1$ or 2 at $n = 4$).

2. As $n \to \infty$, $t^*/n \to 1 - 1/\sqrt{2} = 0.2928932\dots$.

3. $f^{(1)}$ is a bandwidth-minimizing assignment (see e.g. [3]). For $n \geqslant 5$, it follows from our work that the $m \times n$ bandwidth and linear arrangement problems have no common assignment.

60	61	63	66	69	72	77	80	81
58	59	62	65	68	71	76	78	79
55	56	57	64	67	70	73	74	75
46	47	48	49	50	51	52	53	54
37	38	39	40	41	42	43	44	45
28	29	30	31	32	33	34	35	36
7	8	9	12	15	18	25	26	27
3	4	6	11	14	17	20	23	24
1	2	5	10	13	16	19	21	22

Fig. 3.

4. Let f be the optimal assignment given by the Main Theorem for the $n \times n$ grid. We define a real function $g_n : [0,1]^2 \to [0,1]$ as follows:

$$g_n(x,y) = \frac{1}{n^2} f(\lceil xn \rfloor, \lceil yn \rfloor),$$

where $\lceil z \rfloor$ is the nearest integer to z. Then g_n tends in measure to a monotonic, measure-preserving function $g : [0,1]^2 \to [0,1]$ which minimizes the value

$$\int_0^1 g(x,1)\,\mathrm{d}x + \int_0^1 g(1,y)\,\mathrm{d}y - \int_0^1 g(x,0)\,\mathrm{d}x - \int_0^1 g(0,1)\,\mathrm{d}y.$$

The piecewise-quadratic layout of g is given in Fig. 5 below, with $t = 1 - 1/\sqrt{2}$.

2. Doubly monotonic assignments

Given $G_{mn} = (V_{mn}, E_{mn})$ for $m \geqslant n \geqslant 2$, let F denote the set of all maps f from V_{mn} onto $\{1,2,\ldots,mn\}$, and define L on F by

$$L(f) = \sum_{\{u,v\} \in E_{mn}} |f(u) - f(v)|.$$

We observe first that L is minimized by a doubly monotonic f.

Lemma 1. *$L(f)$ is minimized by a doubly monotonic $f \in F$.*

Proof. It is well known (see e.g. [8]) that if $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_K$ and $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_K$ then $\sum_{k=1}^{K} |a_k - b_{\sigma(k)}|$ is minimized over permutations σ on $\{1,2,\ldots,K\}$ by the identity

296	298	301	305	310	316	322	328	334	340	346	352	358	364	375	384	391	396	399	400
295	297	300	304	309	315	321	327	333	339	345	351	357	363	374	383	390	395	397	398
293	294	299	303	308	314	320	326	332	338	344	350	356	362	373	382	389	392	393	394
290	291	292	302	307	313	319	325	331	337	343	349	355	361	372	381	385	386	387	388
286	287	288	289	306	312	318	324	330	336	342	348	354	360	371	376	377	378	379	380
281	282	283	284	285	311	317	323	329	335	341	347	353	359	365	366	367	368	369	370
261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260
221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
161	162	163	164	164	165	167	168	169	170	171	172	173	174	175	176	177	178	179	180
141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
31	32	33	34	35	36	42	48	54	60	66	72	78	84	115	116	117	118	119	120
21	22	23	24	25	30	41	47	53	59	65	71	77	83	89	110	111	112	113	114
13	14	15	16	20	29	40	46	52	58	64	70	76	82	88	93	106	107	108	109
7	8	9	12	19	28	39	45	51	57	63	69	75	81	87	92	96	103	104	105
3	4	6	11	18	27	38	44	50	56	62	68	74	80	86	91	95	98	101	102
1	2	5	10	17	26	37	43	49	55	61	67	73	79	85	90	94	97	99	100

Fig. 4.

permutation. Moreover, it is obvious that if $a_1 < a_2 < \cdots < a_K$ then $\sum_{k<K} |a_{\sigma(k+1)} - a_{\sigma(k)}|$ is minimized by the identity permutation (or its inverse). Given any $f \in F$ with $f(i,j)=a_{ij}$, define f^* by first rearranging the a_{ij} in each row i from smallest to largest values and then rearranging each column j from smallest to largest values in the first rearrangement. It is easily checked that f^* is doubly monotonic and, by the preceding observations, that $L(f^*) \leqslant L(f)$. Consequently, $L(f)$ attains its minimum over F at some doubly monotonic f. $\square$

We consider only doubly monotonic assignments henceforth and let F_{mn} denote the set of doubly monotonic maps from V_{mn} onto $\{1,2,\ldots,mn\}$. Given $f \in F_{mn}$ with $a_{ij} = f(i,j)$, it follows that $L(f)$ is fully determined by the border values of the

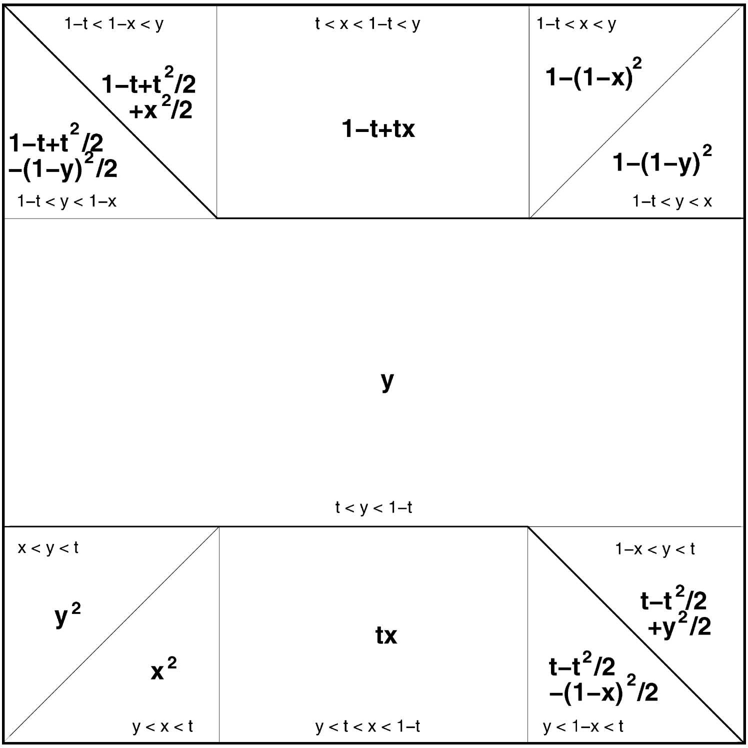


Fig. 5.

assignment:

$$L(f) = \sum_{i=1}^{m} (a_{in} - a_{i1}) + \sum_{j=1}^{n} (a_{mj} - a_{1j})$$
$$= \sum_{i=2}^{m-1} a_{in} + \sum_{j=2}^{n-1} a_{mj} - \sum_{i=2}^{m-1} a_{i1} - \sum_{j=2}^{n-1} a_{1j} + 2(mn-1),$$

where $2(mn-1) = a_{mn} - a_{11} + a_{mn} - a_{11}$. As a small convenience, we transpose $2(mn-1)$ and define $L^*(f)$ as $L(f) - 2(mn-1)$, so

$$L^*(f) = \sum_{i=2}^{m-1} a_{in} + \sum_{j=2}^{n-1} a_{mj} - \sum_{i=2}^{m-1} a_{i1} - \sum_{j=2}^{n-1} a_{1j}.$$

This puts our problem in an isoperimetric form. Subject to double monotonicity, we seek assignments with relatively small values of a_{in} and a_{mj} (last column and top row) and relatively large values of a_{i1} and a_{1j} (first column and bottom row). The corner values, $a_{11}(=1)$, a_{1n}, a_{m1}, and $a_{mn}(=mn)$, do not figure explicitly in $L^*(f)$.

13	14	15	16		11	12	15	16
9	10	11	12		9	10	13	14
5	6	7	8		3	4	7	8
1	2	3	4		1	2	5	6
	($t^* = 1$)					($t^* = 2$)		

Fig. 6.

3. Square grids

We assume throughout this section that $m = n$ and will prove the Main Theorem for this square-grid case. Extensions to $m > n$ are considered in the next section.

Theorem 1. *Given $m = n \geqslant 2$, define $f^{(t)}$ for $t \in \{1, 2, \ldots, \lfloor n/2 \rfloor\}$ as the member of F_{nn} that has pattern $R_t(n, n)$ in the first t rows, has horizontal slats in rows $t + 1$ through $n - t$, and is complementary. Let t^* be a value of t that maximizes*

$$2t^3 - 6nt^2 + (3n^2 + 3n - 2)t \text{ over } t \in \{1, 2, \ldots, \lfloor n/2 \rfloor\}.$$

Then $f^{(t^*)}$ minimizes $L(f)$ over $f \in F_{nn}$ with

$$L(f^{(t^*)}) = n(n^2 + n - 2) - t^*[2(t^*)^2 - 6nt^* + 3n^2 + 3n - 2]/3.$$

When $n \in \{2, 3, 4\}$, $t^* = 1$ maximizes $2t^3 - 6nt^2 + (3n^2 + 3n - 2)t$ over $t \in \{1, 2, \ldots, \lfloor n/2 \rfloor\}$, and in these cases one optimal assignment is composed entirely of horizontal slats with

$$L(f^{(1)}) = n(n^2 - 1).$$

Optimality for these cases is easily verified by inspection. When $n = 4$, we have a joint t-maximizer at $t^* = 2$. This gives the second of the following assignments; see Fig. 6.

For both, $L^* = 30$ and $L = 60$. The second has pattern $R_2(4, 4)$ in rows 1 and 2, and rows 3 and 4 follow from these by the complementarity equation $a_{ij} + a_{5-i,5-j} = 17$.

We assume henceforth in this section that $n \geqslant 5$. We then have $t^* \geqslant 2$ in Theorem 1. This is reflected in the following lemma. As before $a_{ij} = f(i, j)$.

Lemma 2. *Suppose $m = n \geqslant 5$. If $f \in F_{nn}$ minimizes $L(f)$ then $a_{22} = 4$.*

Proof. Assume without loss of generality that, $a_{12} = 2$. The lemma says that a minimizing f must have

$$\begin{matrix} 3 & 4 \\ 1 & 2 \end{matrix}$$

in the lower left corner of its assignment matrix. Suppose $f \in F_{nn}$ does not have this pattern. Then one of the following occurs in the lower left corner:

$$\begin{matrix} 1 & 2 & 3 & 4, \end{matrix} \qquad \begin{matrix} 3 & & \\ 1 & 2 & 4\text{'} \end{matrix} \qquad \begin{matrix} 4 & & \\ 1 & 2 & 3\text{'} \end{matrix} \qquad \begin{matrix} 4 & \\ 3 & \\ 1 & 2 \end{matrix}.$$

By interchanging 3 and 4 in the third array, or by interchanging 2 and 3 followed by matrix transposition in the fourth array, we obtain an f with the same $L^*(f)$ as the second array. Hence it suffices to consider only the first two patterns.

Suppose row 1 begins 1 2 3 4. Then double monotonicity implies the following for rows 1 and 2:

$$\begin{matrix} k+1 & p & & & & \\ 1 & 2 & 3 & 4 & \cdots & k \end{matrix}, \qquad 4\leqslant k\leqslant n,\ k+1<p.$$

We now move 3 and 4 into position (2,1) and (2,2), and put $k+1$ and p into the first row, shifting other things in row 1 leftward to preserve monotonicity. The new array is

$$\begin{matrix} 3 & 4 & & & & & \\ 1 & 2 & \cdots & k & k+1 & \cdots & p \end{matrix} \qquad (k \text{ not in row 1 if } k=4)$$

and it is easily seen to be doubly monotonic. The changes remove 4 from the border area with negative coefficients in $L^*(f)$, i.e., the left side and bottom row, but add at least $p>k+1>4$, so there is a net reduction in L^*. Hence an optimal arrangement cannot have 1 2 3 4 in row 1.

Suppose row 1 begins 1 2 4 with $a_{21}=3$. Let $a_{22}=p\geqslant 5$. Increase entries $4,5,\dots,p-1$ by 1 each, then enter 4 in position (2,2), so now $a_{22}=4$. The changes preserve double monotonicity, increase the lower border by at least 1, do not decrease the left border, and do not change the upper or right borders between the corners. Hence the changes cause a net decrease in L^*, so the array in the first sentence of this paragraph cannot be part of an optimal arrangement. □

We now extend Lemma 2, beginning with $\begin{smallmatrix}34\\12\end{smallmatrix}$ in an optimal arrangement. Double monotonicity requires $a_{13}=5$ or $a_{31}=5$, and because $m=n$ we assume, without loss of generality, that $a_{13}=5$ to obtain

$$\begin{matrix} 3 & 4 & p \\ 1 & 2 & 5 \end{matrix} \quad \text{with } p\geqslant 6.$$

If $p>6$, we increase each of entries $6,7,\dots,p-1$ by 1 and enter 6 in position (2,3). This preserves double monotonicity and increases the lower and left border sum by least 1. If $a_{n2}<p$, it will also increase the top border sum (between corners) by exactly 1, but in this case monotonicity implies that a_{31} through $a_{n-1,1}$ each increases by 1 along the left border, and the net result in any case will be a reduction of L^*. Hence, when $n\geqslant 5$, an optimal arrangement *must* have

$$\begin{matrix} 3 & 4 & 6 \\ 1 & 2 & 5 \end{matrix},$$

given our choices of 2 and then 5 in row 1.

Define $P_t(k)$ as the $t\times(t+k)$ matrix whose first $t\times t$ section is an up staircase and whose next k columns are vertical slats with $a_{1j}=(j-1)t+1$ for $j=t+1,\dots,t+k$. When $t\leqslant n/2$ and $k=n-2t$, we obtain $R_t(n,n)$ by adjoining a $t\times t$ down staircase to the right end of $P_t(n-2t)$.

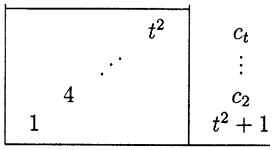


Fig. 7.

Lemma 3. *Suppose $n \geqslant r \geqslant 5$ and assume, without loss of generality that, if $f \in F_{nn}$ has a $t \times t$ up staircase in its lower left corner with $1 \leqslant t < r$ then $f(1,t+1) = t^2+1$. Then every $f \in F_{nn}$ that minimizes $L(f)$ includes one of the $P_t(r-2t)$ for $t = 2,3,\ldots,\lfloor r/2 \rfloor$.*

Proof. We prove the lemma by induction on r. Lemma 2 and the paragraph that follows its proof verify Lemma 3 for $r = 5 : P_2(1)$ is

$$\begin{matrix} 3 & 4 & 6 \\ 1 & 2 & 5 \end{matrix}.$$

Assume that the conclusion of Lemma 3 holds for an arbitrary $r \geqslant 5$. As we consider cases for $r+1$, we repeatedly use the argument noted above for $p > 6$. Its general form is: 'If $p > c$, increase each of entries $c, c+1, \ldots, p-1$ by 1 and enter c into the position previously occupied by p. This preserves double monotonicity and produces a net reduction in L^*.' We abbreviate the statement in quotes by: *replace p by c*. In each such instance the second sentence of the quote is easily seen to be true and we often omit details.

Our induction hypothesis assumes that an optimal f for $n \geqslant r$ includes $P_t(r-2t)$ for some $t \in \{2,\ldots,\lfloor r/2 \rfloor\}$. Suppose in fact that $n \geqslant r+1$, we are to show that f also includes $P_t(r+1-2t)$ for some $t \in \{2,\ldots,\lfloor (r+1)/2 \rfloor\}$. We divide the hypothesis' cases into $r-2t=0$, $r-2t=1$ and $r-2t \geqslant 2$. Given that an optimal f includes $P_t(r-2t)$, and $n \geqslant r+1$, we prove that

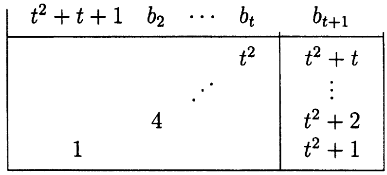
$$r-2t=0 \Rightarrow f \text{ includes } P_t(1),$$

$$r-2t=1 \Rightarrow f \text{ includes } P_{t+1}(0) \text{ or } P_t(2),$$

$$r-2t=k \geqslant 2 \Rightarrow f \text{ includes } P_t(k+1).$$

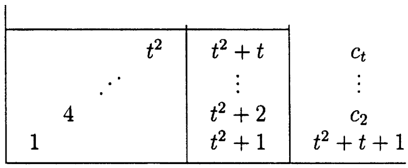
Suppose f includes the $t \times t$ up staircase $P_t(0)$ with $r = 2t$. By convention, $a_{1,t+1} = f(1,t+1) = t^2+1$. Let $c_2, c_3, \ldots, c_t$ denote the next $t-1$ entries in column $t+1$ of f's assignment matrix (see Fig. 7).

If $c_2 > t^2+2$, replace c_2 by t^2+2; if $c_3 > t^2+3$, replace c_3 by $t^2+3; \ldots;$ if $c_t > t^2+t$, replace c_t by t^2+t. Each replacement decreases L^*, so our assumption of optimality for f implies that its $t+1$st column begins $t^2+1, t^2+2, \ldots, t^2+t$, hence that f includes $P_t(1)$. To check for decreasing L^* in the final step, suppose $c_{t-1} = t^2+t-1$



$a_{t+1,1} = t^2 + t + 1$

Fig. 8.



$a_{1,t+2} = t^2 + t + 1$

Fig. 9.

and $c_t > t^2 + t$. Then, by double monotonicity, either $a_{t+1,1} = t^2 + t$ (left border) or $a_{1,t+2} = t^2 + t$ (bottom border), and this entry increases to $t^2 + t + 1$ in the replacement of c_t by $t^2 + t$. If $t > 2$ and $a_{ni} < c_t$, for $i \in \{2, 3, \ldots, t\}$, so that all of a_{n2} through a_{ni} increase by 1 in the replacement, then all of $a_{t+1,1}$ through $a_{n-1,1}$ in column 1 also increase by 1; there are $n - t - 1$ of the latter, and $i - 1 \leqslant t - 1$ positions in the top row, and because $n - t - 1 > t - 1$, i.e., $n > 2t = r$, there is a net reduction in L^* in the upper left part of the assignment matrix. If $t > 2$ and $a_{2n} < c_t$, so that a_{2n} increases by 1 in the replacement, all of $a_{1,t+2}$ through $a_{1,n-1}$ in row 1 also increase by 1; there are $n - t - 2$ of the latter, and $t - 2$ positions in column n strictly between rows 1 and t, so because $n - t - 2 > t - 2$ there is a net reduction in L^* in the lower right part of the matrix.

Suppose next that f includes $P_t(1)$ with $r = 2t + 1$. $P_t(1)$ is the $t \times t$ up staircase followed by a vertical slat of height t in column $t + 1$ with $a_{1,t+1} = t^2 + 1$. We consider two cases according to whether $a_{t+1,1}$ or $a_{1,t+2}$ is $t^2 + t + 1$ (see Figs. 8 and 9).

Suppose $a_{t+1,1} = t^2 + t + 1$. If $b_2 > t^2 + t + 2$, replace b_2 by $t^2 + t + 2$; if $b_3 > t^2 + t + 3$, replace b_3 by $t^2 + t + 3; \ldots$; if $b_{t+1} > t^2 + t + (t+1) = (t+1)^2$, replace b_{t+1} by $(t+1)^2$. This produces the $(t+1) \times (t+1)$ up staircase $P_{t+1}(0)$. Suppose $a_{1,t+2} = t^2 + t + 1$. If $c_2 > t^2 + t + 2$, replace c_2 by $t^2 + t + 2; \ldots$; if $c_t > t^2 + 2t$, replace c_t by $t^2 + 2t$. This produces $P_t(2)$. It follows that f includes either $P_{t+1}(0)$ or $P_t(2)$.

Finally, suppose f includes $P_t(k)$ with $k = r - 2t \geqslant 2$. Then, by double monotonicity, either $a_{1,t+k+1} = t^2 + kt + 1$ or $a_{t+1,1} = t^2 + kt + 1$. Suppose $a_{1,t+k+1} = t^2 + kt + 1$; see Fig. 10.

			t^2	t^2+t		t^2+kt	c_t
		$\cdot^{\cdot^{\cdot}}$		$\vdots$	$\cdots$	$\vdots$	$\vdots$
	4			t^2+2		$t^2+(k-1)t+2$	c_2
1				t^2+1		$t^2+(k-1)t+1$	t^2+kt+1
1	2	$\cdots$	t	$t+1$	$\cdots$	$t+k$	$t+k+1$

Fig. 10.

$t+1$	t^2+kt+1	b_2	$\cdots$	b_t	b_{t+1}	$\cdots$	b_{t+k}
t				t^2	t^2+t		t^2+kt
			$\cdot^{\cdot^{\cdot}}$		$\vdots$	$\cdots$	$\vdots$
2		4			t^2+2		$t^2+(k-1)t+2$
1	1				t^2+1		$t^2+(k-1)t+1$

Fig. 11.

If $c_2 > t^2+kt+2$, replace c_2 by $t^2+kt+2; \ldots$; if $c_t > t^2+(k+1)t$, replace c_t by $t^2+(k+1)t$. This produces $P_t(k+1)$. Suppose $a_{t+1,1} = t^2+kt+1$; see Fig. 11.

If $b_2 > t^2+kt+2$, replace b_2 by $t^2+kt+2; \ldots$; if $b_{t+k} > t^2+kt+(t+k)$, replace b_{t+k} by $t^2+kt+(t+k)$. Despite the fact that these replacements reduce L^*, the resulting $(t+1)\times(t+k)$ array is not a P array and cannot be part of an optimal f. To verify suboptimality, observe that the resulting array has precisely the same dimensions as $P_{t+1}(k-1)$ and contains the same entries, namely 1 through $(t+1)(t+k)$. However, its left and bottom border sum is less than the similar sum for $P_{t+1}(k-1)$, so its L^* contribution is greater than that of $P_{t+1}(k-1)$. Specifically, the difference of the left and bottom border sums of $P_{t+1}(k-1)$ and the preceding array is

$$\left\{(t^2+t+1)+(t^2+1)+\sum_{i=1}^{k-1}[(t+1)^2+(i-1)t+1)\right\}$$

$$-\left\{(t^2+kt+1)+\sum_{i=1}^{k}[t^2+(i-1)t+1]\right\} = k-1.$$

Hence, if optimal f includes $P_t(k)$ with $k = r-2t \geqslant 2$, then it also includes $P_t(k+1)$. $\square$

We now conclude the proof of Theorem 1 for $n\geqslant 5$. Assume that $f \in F_{nn}$ minimizes $L(f)$. Set $r=n$ in Lemma 3. Then f's assignment matrix includes one of the $P_t(n-2t)$ for $t=2,3,\ldots,\lfloor n/2 \rfloor$. In doing this we have assumed, without loss of generality, that for $k>0$ in $P_t(k)$, the k slats which follow the $t\times t$ up staircase extend rightward

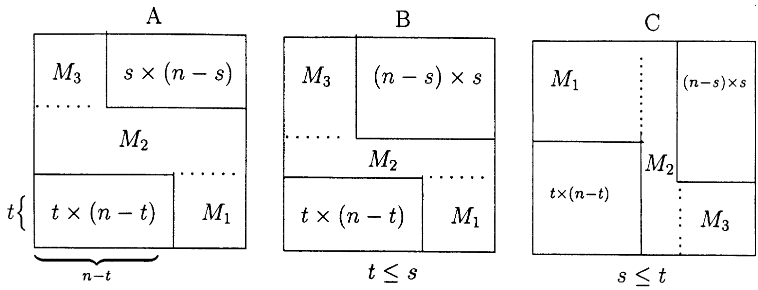


Fig. 12.

along the bottom rather than upward along the left side: they are height-t vertical slats rather than width-t horizontal slats.

A symmetric conclusion holds for the upper right section of the matrix. If we define b_{ij} by $b_{ij} = n^2 + 1 - a_{ij}$, so $b_{nn} = 1$, the b_{ij} increase monotonically away from the upper right corner and all signs in L^* are reversed when its a_{ij} are replaced by the b_{ij}. However, we can no longer assume for $k > 0$ in the upper right counterpart of $P_t(k)$ that the slat section adjacent to the upper right $t \times t$ section extends leftward along the top rather than downward along the right side. The sum of the top row and right column border values within the rectangular sections will be identical, but they fit differently with the $P_t(n-2t)$ lower-left rectangle of the preceding paragraph.

To evaluate the possibilities, assume that a doubly monotonic assignment matrix has the $t \times (n-t)$ submatrix $P_t(n-2t)$ in the lower left and either an $s \times (n-s)$ or an $(n-s) \times s$ submatrix counterpart of $P_s(n-2s)$ in the upper right, with $s, t \in \{2, 3, \ldots, \lfloor n/2 \rfloor\}$. Each submatrix has $n-2$ border entries excluding the corner entry. Fig. 12 outlines the possibilities.

In each of A, B and C, M_1 and M_3 are square submatrices. For example, M_1 is $t \times t$ in A and B, and $(n-t) \times (n-t)$ in C. Submatrix M_2 is $(n-s-t) \times n$ in A, $(s-t) \times n$ in B, and $n \times (t-s)$ in C. By construction, the smallest integers in $\{1, 2, \ldots, n^2\}$ are in the $t \times (n-t)$ submatrix, and the largest are in the $s \times (n-s)$ or $(n-s) \times s$ submatrix. Given those arrays, L^* is minimized in each case by down staircases in the lower right square submatrices (M_1 for A and B, M_3 for C), down staircase counterparts in the upper left square submatrices (M_3 for A and B, M_1 for C), and width-n horizontal slats (A, B) or height-n vertical slats (C) in M_2. We can do no better under double monotonicity, which requires $a_{jn} - a_{1,n-j+1} \geqslant j^2 - 1$ in the lower right corner, $a_{nj} - a_{n-j+1,1} \geqslant j^2 - 1$ in the upper left corner, and $a_{jn} - a_{j1} \geqslant n - 1$ and $a_{nj} - a_{1j} \geqslant n - 1$ for width-n and height-n rows and columns. To allow the down staircases and full-length slats in minimization of L^* while preserving double monotonicity, all integers assigned to M_1 are less than those assigned to M_2, which in turn are less than those assigned to M_3.

The contributions to L^* from the border entries of M_1, M_2 and M_3, excluding the extreme corners, are

$$L^*[A] = \sum_{i=2}^{t}(i^2-1) + \sum_{i=2}^{s}(i^2-1) + (n-1)(n-s-t),$$

$$L^*[B] = \sum_{i=2}^{t}(i^2-1) + \sum_{i=2}^{n-s}(i^2-1) + (n-1)(s-t),$$

$$L^*[C] = \sum_{i=2}^{s}(i^2-1) + \sum_{i=2}^{n-t}(i^2-1) + (n-1)(t-s).$$

It suffices to compare $L^*[A]$ and $L^*[B]$, assuming $t \leqslant s$. Clearly, $L^*[A]=L^*[B]$ if $s=n/2$. Otherwise $L^*[A] < L^*[B]$:

$$L^*[B] - L^*[A] = [2n^3 - (6s+3)n^2 + (6s^2+6s+1)n - s(4s^2+2)]/6$$

and calculus shows that the right-hand side is positive if $s < n/2$. Hence, for fixed s and t, L^* is minimized by A. We proceed with A. Computation gives

$$L(A \text{ with } s \text{ and } t) = n(n^2+n-2) - s[2s^2 - 6ns + 3n^2 + 3n - 2]/6$$
$$- t[2t^2 - 6nt + 3n^2 + 3n - 2]/6.$$

Consequently, $L(f)$ for $f \in F_{nn}$ is minimized when $s = t = k$, where k is a value in $\{2, 3, \ldots, \lfloor n/2 \rfloor\}$ that maximizes $2k^3 - 6nk^2 + (3n^2+3n-2)k$. We denoted such a k in Theorem 1 by t^*. By the construction for A, the corresponding f has pattern $R_{t^*}(n,n)$ in the first t^* rows, horizontal slats in rows t^*+1 through $n-t^*$, and can be presumed to be complementary by using the complementary counterpart of $R_{t^*}(n,n)$ in the top t^* rows.

4. Nonsquare grids

We assume throughout this section that $m > n$ and consider the Main Theorem for this case. The result mentioned earlier, that one optimal $f \in F_{nn}$ for $n \leqslant 4$ consists entirely of horizontal slats, applies also to $f \in F_{mn}$ for $n \leqslant 4$. This is obvious for $n=2$ and nearly so for $n=3$, but requires a little effort to verify it for $n=4$. We omit the $n=4$ proof, which can be patterned after the ensuing proof of Theorem 2, and assume henceforth that $n \geqslant 5$.

Theorem 2. *Given $m > n \geqslant 5, f \in F_{mn}$ minimizes $L(f)$ if it has pattern $R_{t^*}(m,n)$ in the first t^* rows, has horizontal slats in rows t^*+1 through $m-t^*$, and is complementary, where t^* is a value of t that maximizes $2t^3 - 6nt^2 + (3n^2+3n-2)t$ over $t \in \{2, 3, \ldots, \lfloor n/2 \rfloor\}$. The minimum of $L(f)$ for $f \in F_{mn}$ is*

$$m(n^2+n-1) - n - t^*[2(t^*)^2 - 6nt^* + 3n^2 + 3n - 2]/3.$$

We begin the proof with an extension of Lemmas 2 and 3. To account for the asymmetry of $m > n$, we refer to the $t \times (t+k)$ lower left matrix $P_t(k)$ as defined in the preceding section as a *horizontal* $P_t(k)$. A *vertical* $P_t(k)$ can be defined as the $(t+k) \times t$ transpose of a horizontal $P_t(k)$, or as the lower left $(t+k) \times t$ matrix which consists of a $t \times t$ up staircase directly beneath k width-t horizontal slats with $a_{j1} = (j-1)t + 1$ for $j = t+1, \dots, t+k$. These two versions of a vertical $P_t(k)$ are not identical (in the initial $t \times t$ part), but they have identical border sums.

For the upper right section of an $m \times n$ assignment matrix, we define a *horizontal* $\hat{P}_t(k)$ as the $t \times (t+k)$ matrix that is the complementary counterpart of a horizontal $P_t(k)$. That is, for $i = m-t+1, \dots, m$ and $j = n-t-k+1, \dots, n$, a_{ij} for $\hat{P}_t(k)$ equals $mn + 1 - a_{m+1-i, n+1-j}$, where the latter a comes from $P_t(k)$. Similarly, a *vertical* $\hat{P}_t(k)$ is the $(t+k) \times t$ complementary counterpart in the upper right section of a vertical $P_t(k)$.

Lemma 4. *Every $f \in F_{mn}$ that minimizes $L(f)$ includes a horizontal or vertical $P_t(n-2t)$ for some $t \in \{2, 3, \dots, \lfloor n/2 \rfloor\}$, and includes a horizontal or vertical $\hat{P}_t(n-2t)$ for some $t \in \{2, 3, \dots, \lfloor n/2 \rfloor\}$, subject to variations in $t \times t$ lower left and upper right matrices that do not change border sums.*

Proof. All operations in the proofs of Lemmas 2 and 3 remain valid (including 'replace p by c') when $m > n$ excepting those based on the symmetry created by $m = n$. Joint consideration of horizontal and vertical $P_t(k)$ accounts for the asymmetry introduced by $m > n$. The joint extension of Lemma 3 for $m > n \geqslant r \geqslant 5$ under the relaxation for $t \times t$ noted at the end of Lemma 4 yields the $P_t(n-2t)$ part of Lemma 4 when we set $r = n$, and the $\hat{P}_t(n-2t)$ part follows from the $P_t(n-2t)$ part by complementarity when b_{ij} is defined by $b_{ij} = mn + 1 - a_{ij}$. □

We complete the proof of Theorem 2 in a manner similar to that for Theorem 1 with $n \geqslant 5$ based on Fig. 12. Fig. 13 illustrates four situations of Lemma 4. We use $P_t(n-2t)$ in the lower left and $\hat{P}_s(n-2s)$ in the upper right.

M_1, M_2 and M_3 for A, B and C are assigned integers in the manner described for A and B in Fig. 12 with minimum border contributions to L^* as follows:

$$L^*[A] = \sum_{i=2}^{t} (i^2 - 1) + \sum_{i=2}^{s} (i^2 - 1) + (n-1)(m - s - t),$$

$$L^*[B] = \sum_{i=2}^{t} (i^2 - 1) + \sum_{i=2}^{n-s} (i^2 - 1) + (n-1)(m - n + s - t),$$

$$L^*[C] = \sum_{i=2}^{n-t} (i^2 - 1) + \sum_{i=2}^{n-s} (i^2 - 1) + (n-1)(m - 2n + s + t).$$

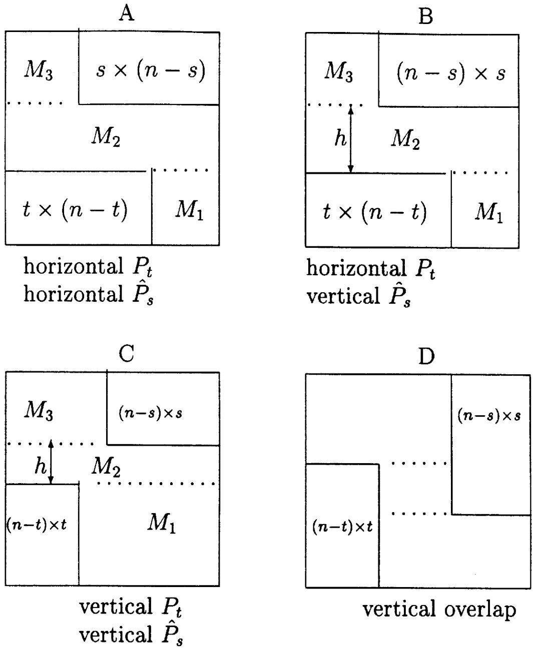


Fig. 13.

We presume in the B and C cases that there is no vertical overlap, i.e., that $h \geqslant 0$. This requires $m \geqslant n - s + t$ for B and $m \geqslant 2n - s - t$ for C. We return to overlap cases shortly.

Comparisons of the preceding L^* sums show in all cases that $L^*[A]$ has minimum value. For example,

$$3[L^*[C] - L^*[A]] = 2n^3 - 3(s+t+1)n^2 + (3s^2 + 3s + 3t^2 + 3t + 1)n$$
$$-s(2s^2+1) - t(2s^2+1)$$

and this is positive when $s, t \leqslant n/2$ unless $s = t = n/2$, in which case $L^*[C] = L^*[A]$. Hence A will minimize $L(f)$ for $f \in F_{mn}$ with best choices of s and t, which are identical to those for the square grid case as stated in Theorems 1 and 2. The additional height of $m - n$ simply means that this many new horizontal slats are inserted into the midsection of an optimal assignment of Theorem 1 to produce an optimal assignment for Theorem 2. The minimum of $L(f)$ at the conclusion of Theorem 2 is given by straightforward calculation.

Because $m > n$, vertical overlap cases like D in Fig. 14 require somewhat different treatment. We consider the situation shown by D in more detail: see Fig. 14.

Contributions to $L^*[D]$ from sections (a) and (c) must be as great as those given by down staircase patterns, and the contribution from (b) must be at least $h(m-1+m-n)$

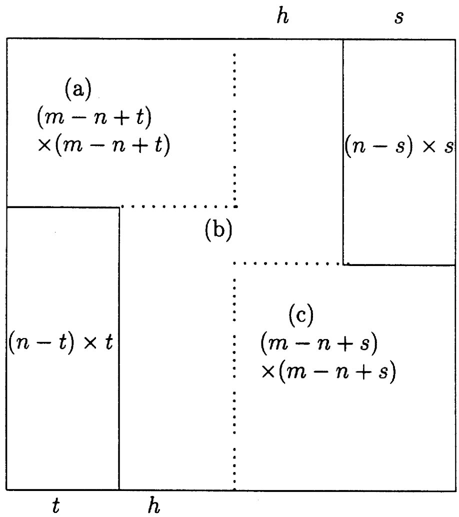


Fig. 14.

by double monotonicity. Therefore,

$$L^*[D]-L^*[A] \geqslant \sum_{i=2}^{m-n+t}(i^2-1)+\sum_{i=2}^{m-n+1}(i^2-1)+[2n-(m+s+t)](2m-n-1)$$

$$-\sum_{i=2}^{t}(i^2-1)-\sum_{i=2}^{s}(i^2-1)-(n-1)(m-s-t).$$

With $p=m-n$, we have

$$6[L^*[D]-L^*[A]] \geqslant 4p^3+p^2[3(2t+1)+3(2s+1)]$$
$$+p[(2t+1)^2+2t(t+1)+(2s+1)^2+2s(s+1)]$$
$$-12p(p+s+t).$$

Because $s,t\geqslant 2$, $3(2t+1)+3(2s+1)\geqslant 30$ and $(2t+1)^2+2t(t+1)>12t$, so the $-12p(p+s+t)$ term is more than offset by positive terms. Hence $L^*[A]<L^*[D]$. Other vertical overlaps, in which one of P_t and $\hat{P}_s$ is horizontal, have the same conclusion. We omit their details.

Acknowledgements

We thank Larry Harper for bringing our attention to the work of Mitchison and Durbin, and Sergej Bezrukov for several useful references, including the work of Muradjan. We also thank Steve Horton for a careful reading of an earlier draft of this manuscript.

References

[1] S. Bezrukov, Edge isoperimetric problems on graphs, preprint, 1996.
[2] J.D. Chavez, L.H. Harper, Discrete isoperimetric problems and pathmorphisms, preprint, 1996.
[3] J. Chvátalová, Optimal labeling of a product of two paths, Discrete. Math. 2 (1975) 249–253.
[4] G. Mitchison, R. Durbin, Optimal numberings of an $n \times n$ array, SIAM J. Algebra Discrete Math. 7 (1986) 571–582.
[5] D.O. Muradjan, T.E. Piliposjan, Minimal numberings of vertices of a rectangular lattice, Akad. Nauk. Armjan. SSR. Dokl. 70 (1980) 21–27 (see MR 81m:05121) (in Russian).
[6] D.O. Muradjan, Minimal numberings of a two-dimensional cylinder, Akad. Nauk. Armjan. SSR. Dokl. 75 (1982) 114–119 (see MR 84m:05041) (in Russian).
[7] D.O. Muradjan, Some estimates for the length of an arbitrary graph, Mat. Voprosy Kibernet. Vychisl. Tekhn. (1985) 79–86 (in Russian).
[8] Y. Rinott, Multivariate majorization and rearrangement inequalities with some applications to probability and statistics, Israel J. Math. 15 (1973) 60–77.