Index assignment for two-channel quantization

József Balogh, János A. Csirik

Abstract

This paper concerns the design of a multiple description scalar quantization system for two identical channels for an unbounded discrete information source. This translates to the combinatorial problem of finding an arrangement of the integers into the infinite plane square grid so that each row and each column contains exactly N numbers, such that the difference between any two numbers in the same row (or column) is at most d, with d to be minimized for a given N. The best previous lower and upper bounds on the lowest d were $N^2/3+O(N)$ and $N^2/2+O(N)$. We give new lower and upper bounds, both of the form $3N^2/8+O(N)$. We also consider minimizing the maximal variance in any row or column and show that it must be at least $N^4/60+O(N^3)$, and that it does not have to be more than $3N^4/160+O(N^3)$.

I. Introduction

A *diversity system* provides several different channels for transmitting information from the source to the user. Thus, if a channel breaks down, an alternate path is available between the source and the user. Consider a diversity system with two channels. If the same message is sent over each channel and if both channels work, one of them is unused. We consider sending a different message over each channel. If only one channel works, the single message received over it is sufficient to achieve a minimum fidelity. On the other hand, should both channels work, the information received from both messages can be used to achieve a higher fidelity than would each message alone. The problem of coding, i.e., of designing pairs of messages to do so, is known as the multiple descriptions problem [1] and is a generalization of the problem of source coding subject to a fidelity criterion [2].

A multiple description scalar quantizer (MDSQ) is a scalar quantizer designed for operation in such a diversity system. The encoder of an MDSQ sends messages over each channel of the diversity system subject to a rate constraint. The decoder reconstructs the source sample based on the messages received from the channels that are currently working. The objective is to design a decoder-encoder pair that minimizes the distortion when both channels work, subject to constraints on the distortion when only one channel works. Thus, in the event that exactly one of the channels is broken, a minimum fidelity is guaranteed.

Applications of multiple description source codes arise in speech and video coding over packet-switched networks, where packet losses can result in a degradation in signal quality and there are significant delay constraints. For details and more examples see [3], [4]. The index assignment problem for MDSQ also has connections with the theory of graph bandwidth, where the bandwidth of a graph is defined as the maximal difference occurring between the endpoints of edges in an optimal labelling of the vertices using different integers. For more details, see [5].

JB is at Ohio State University, Columbus, Ohio, USA. JAC is at D. E. Shaw & Co., New York, New York, USA. This work was carried out while both authors were at AT&T Labs–Research, Florham Park, New Jersey, USA.

The design of an MDSQ system involves two steps: quantization of the source, and the index assignment problem of distributing the quantized signal over multiple channels. In this paper, we will only consider the index assignment problem, that is, an MDSQ design for a discrete information source. We investigate the case of two identical channels, where the rate of both channels is equal and the *maximum distortion* of either channel by itself is to be minimized. We also require that the source symbol be reconstructed perfectly if both channels work.

Let N and t be positive integers, with $N \geq 2$ and $t \geq N$. We consider the case of a discrete source that emits symbols that are integers in the interval $[1, tN]$, and two channels that can each transmit an integer in the range $[1, t]$. Thus, the parameter N can be interpreted as determining the amount ($\log_2 N$) by which the rate of each channel is less than that of the source if the source symbols are emitted according to a uniform probability distribution.

A solution to the index assignment problem then corresponds to an arrangement of the integers 1, 2, ..., tN in the places of a t-by-t square array. We consider discret positions in the array where an integer can be put to be referred to as *places*, each specified by two integer coordinates. For example, for $N = 3$ and $t = 7$, a possible arrangement is:

Row 7					19	16	21
Row 6		13			11		17
Row 5		9		18		8	
Row 4			5	12	14		
Row 3	3			10			15
Row 2	4		6			7	
Row 1	1	2	20				
	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7

Figure I.1

The interpretation is that the row number and the column number of a source symbol serve as the messages to be sent over the two channels. For example, in the arrangement above, the source symbol 13 would be sent as the message pair $(6, 2)$, since the number 13 is located in row 6 and column 2.[1] (Note that the requirement that the source symbol can be reconstructed exactly given the messages from both channels corresponds to the requirement that no place of the square array contains more than one source symbol.) If the second message is lost then the receiver will only know that the source symbol was one of the entries in row 6, so that it is one of $\{11, 13, 17\}$.

This paper investigates two different measures of maximum distortion.

The first measure to be investigated (the *spread*) corresponds to minimizing the maximum absolute difference between the original source symbol and its reconstruction, where the maximum is taken over all possible original source symbols and with either channel failing. This notion is independent of the probability distribution on the source symbols, and can be interpreted as a uniform upper bound on the expected error over all possible probability distributions on the source symbols and channel failure modes (where we discard the cases where both channels fail before taking the expectation).

[1]Here, and everywhere else in this article, we use the convention that column numbers increase to the right and row numbers increase toward the top of the page.

Assume now that one channel fails and only one message is received, and that it restricts the set of possible source symbols to a set $\mathcal{L}$. Letting $M = \max(\mathcal{L})$ and $m = \min(\mathcal{L})$, it is clear that the reconstruction $(M + m)/2$ minimizes the possible maximum of the absolute error, and that this maximum is $(M - m)/2$ (which will be attained if the original source symbol was m or M). Now let the *spread* of any row or column be the difference between the largest and the smallest numbers contained in it. Since the maximum error corresponding to any message is half the spread of the corresponding row or column, it follows that the arrangements that minimize the maximum distortion in this measure are those where the maximum spread of any row or column is as small as possible. For any arrangement, we shall use d to denote the maximum of the spreads of the rows and columns.

The second measure of distortion to be investigated (the *variance*) corresponds to minimizing the expected value of the square of the difference between the reconstruction and the original source symbol, under the assumption that the source symbols are emitted according to a uniform probability distribution, and taking the minimum over the set of all possible single messages that can be received (this set has $2t$ elements, corresponding to t possible messages on the first channel, and t on the second one).

Assume again that one channel fails and only one message is received, and that it restricts the set of possible source symbols to $\mathcal{L}$. In the second measure of distorsion, it is clear that the best reconstruction of the source symbol is the average of the elements of $\mathcal{L}$, and that the expected value of the square of the error is the variance of the elements of $\mathcal{L}$. Thus, the arrangements that minimize the maximum distortion in this measure are those where the maximum variance of any row or column is as small as possible. For any arrangement, we shall use V to denote the maximum of the variances of the rows and columns.

This paper concentrates on determining the optimal arrangements when N is fixed and t is much larger than N. To avoid the special cases of source symbols near the edge of the allowable range, we will consider instead the case where t is infinite, that is, we consider arrangements of all the integers in the plane square grid so that each row or column contains exactly N integers, where we seek to minimize the maximum spread (or variance) in any row or column. This will be referred to as the *infinite case*.[2]

In the construction of good arrangements for the infinite case, it will turn out that it is helpful to start by considering arrangements at the opposite extreme, where $t = N$. This corresponds to entirely filling up an N-by-N square by the first N^2 positive integers while minimizing the maximum spread (or variance), thus it will be referred to as the *square case*.

The question of the minimum spread in the square case has been fully resolved by Tanya Y. Berger-Wolf and Edward M. Reingold in [6]. They proved that the smallest possible spread is $N(N + 1)/2 - 1$ by explicitly constructing the corresponding arrangement and proving that no arrangement with lower maximum spread exists.

The question of the minimum spread in the infinite case was considered in [4] and in [6], where it was proved that the optimal arrangement has a maximum spread d of at least $N^2/3 + O(N)$ and at most $N^2/2 + O(N)$ (these are the best known results not contained in this paper, as far as the authors know). In Section III we show that, for even N, the

[2]In the infinite case, the spreads (and variances) of all the rows and columns do not necessarily have a maximum, so d (and V) must be defined as a supremum instead. However, in all arrangements of interest these suprema will be finite so this technical point will not play a role in the rest of this paper.

best maximum spread d is exactly $\lceil 3N^2/8 - 1/2 \rceil$. For odd N, we supply lower and upper bound that are both of the form $3N^2/8 + O(N)$. Both the upper bound (construction using large square blocks) and the lower bound (proof using a particular combination of local and global methods) of this paper contain new ideas that will extend to the case of non-identical channels (with different rates), and to more than two channels.

The investigation of the minimum variance in the square case will be carried out in Section IV, where we prove that it is at least $(1/24 + 4.2 \cdot 10^{-7})N^4 + O(N^3)$ and for all $N > 3$ there are assignments with maximum variance at most $N^4/20$. The authors do not know of any previous bounds in this case.

Finally, the investigation of the minimum variance in the infinite case will be carried out in Section V. We obtain that the minimum variance must be always at least $N^4/60 + O(N^3)$, and we have construction for all N where it is $3N^4/160 + O(N^3) = N^4/(53\frac{1}{3}) + O(N^3)$. The best previous upper bound known to the authors is $N^4/48 + O(N^3)$ in [7].

II. Notation

Given an arrangement of integers into places of a square grid, and for any set of integers $\mathcal{L}$ inside it, let $R(\mathcal{L})$ denote the set of *row neighbors* of $\mathcal{L}$, i.e., the set of integers which are in the same row as some element of $\mathcal{L}$. Similarly, define $C(\mathcal{L})$ to be the set of *column neighbors* of $\mathcal{L}$. For any integer x in the arrangement, define $R(x)$ as meaning $R(\{x\})$, and define $C(x)$ as meaning $C(\{x\})$. $R(x)$ and $C(x)$ are thus the row and column which contain x.

Given any finite set $\mathcal{L}$, let $|\mathcal{L}|$ denote its cardinality, $\min(\mathcal{L})$ and $\max(\mathcal{L})$ denote the smallest and the largest of its elements, respectively.

We let the *variance* of a list $(X_1, X_2, \ldots, X_n)$ of real numbers be denoted by

$$\mathrm{Var}(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_i (X_i - \overline{X})^2,$$

where $\overline{X} = \frac{1}{n} \sum_i X_i$ is the *mean.*

A list as above can be understood as a random variable with uniform probability distribution. For any random variable X, let $\mathbb{E}(X)$ and $\mathrm{Var}(X)$ denote its mean and variance.

III. Minimizing the spread in the infinite case

A. Outline of the argument

For the upper bound, we shall exhibit a construction in Section III-B. The crucial idea is that once we decided on the *footprint* of an arrangement (the places in the infinite grid where we are to put integers), it is relatively straightforward to permute the integers within the selected positions so as to achieve the smallest possible maximal spread. The footprint we use in Sec. III-B turns out to be composed of adjacent squares of $\lceil N/2 \rceil^2$ places, each containing $N^2/4$ integers if N is even (then the square is full) or with two adjacent squares containing $N(N+1)/2$ integers if N is odd. We shall denote any of these squares by $\mathcal{S}_i$. They are arranged along the diagonal which corresponds to increasing line and column numbers, plus other squares placed immediately below each of the diagonal ones. The numbering of the $\mathcal{S}_i$s increases from top to bottom and right to left, alternately. We may assume that the first square considered is $\mathcal{S}_0$ and that $\mathcal{S}_1$ is put below it, so $\mathcal{S}_{i+1}$ is below $\mathcal{S}_i$ if i is even, at its left if i is odd. No integer is put outside the numbered squares. Any arrangement considered in Sec. III-B is assumed to have this footprint. In the case of even N, the upper bound equals the lower bound so this is a best possible choice for a footprint. The description of the construction for odd N and further details are given in Section III-B.

The lower bound is investigated in Section III-C. To give an idea of the argument, consider first the lower bound $d \geq (N^2-1)/3$ due to Diggavi, Orlitsky and Vaishampayan ([7]). Pick an arbitrary row R of the arrangement, and let m and M be the smallest and largest elements of R. Then clearly $M-m \leq d$. Furthermore, every element of $C(R)$ is at most $m+2d$, since any element of R is at most $m+d$, and any element of $C(R)$ is at most d larger than the element of R in its column. Similarly, all elements of $C(R)$ are at least $M-2d$. As $C(R)$ contains N^2 distinct integers, all are in the interval $[M-2d, m+2d]$, which interval has length at most $3d$. It follows that $d \geq (N^2-1)/3$. The proof of the lower bound given in Section III-C extends this argument by considering not only the numbers in $C(R)$, but also in appropriately defined larger sets of numbers. The main result is stated as Theorem III.17.

B. *Upper bound on the optimal spread*

Theorem III.1: There exists an arrangement of the integers within an unbounded plane square grid, with each row and column containing exactly N numbers, $N \geq 2$, such that the difference between any two numbers in the same row (or column) is at most $\lceil 3N^2/8 - 1/2 \rceil$ (if N is even), or at most $3N^2/8 + N/4 + O(1)$ (if N is odd).

Proof: We will first describe the construction for even N. Examples will be given for $N = 8$. Start by filling in the numbers 1 through $N^2/4$ in an $N/2$-by-$N/2$ square as follows. Start by placing the numbers 1, 2, …, $N^2/8 - N/4$ below the northwest–southeast diagonal by filling the available spaces in the last row, then the second-last row, and so on, progressing from left to right in each row. Then place the numbers from $N^2/4$ down to $N^2/8 + N/4 + 1$ above the diagonal, by filling the available spaces in the last column, then the second-last column, and so on, progressing from top to bottom in each column. At this point, our square looks like this:

.	11	13	16
6	.	12	15
4	5	.	14
1	2	3	.

Figure III.1

Of the remaining numbers ($N^2/8 - N/4 + 1$ through $N^2/8 + N/4$), put the number $\lfloor N^2/8 + 1/2 \rfloor$ into the upper left corner, distribute the rest arbitrarily. We obtain

8	11	13	16
6	7	12	15
4	5	10	14
1	2	3	9

Figure III.2

Let $P_{i,j}$ denote the (i,j) entry of the $N/2$-by-$N/2$ matrix just constructed. Let $Q_{i,j}$ denote the number (if any) that is placed into the (i,j)th position of the infinite plane square grid in the arrangement we are currently constructing. If there is no number in the (i,j)th position, we shall say that "$Q_{i,j}$ is empty". Define $Q_{i,j}$ as follows. Let

$$\begin{aligned} i &= \frac{N}{2}a + i', \qquad \text{with } i' \in \{1, 2, \ldots, \tfrac{N}{2}\}, \\ j &= \frac{N}{2}b + j', \qquad \text{with } j' \in \{1, 2, \ldots, \tfrac{N}{2}\}, \end{aligned}$$

where a and b are integers. Then set

$$Q_{i,j} = \begin{cases} P_{i',j'} + aN^2/2 & \text{if } a = b, \\ P_{j',i'} + (a + \frac{1}{2})N^2/2 & \text{if } a + 1 = b, \\ \text{empty} & \text{otherwise.} \end{cases}$$

For $N = 8$, we obtain

						40	43	45	48	.	.
						38	39	44	47	.	.
						36	37	42	46	.	.
						33	34	35	41	.	.
		8	11	13	16	25	30	31	32		
		6	7	12	15	19	26	28	29		
		4	5	10	14	18	21	23	27		
		1	2	3	9	17	20	22	24		
.	.	-7	-2	-1	0						
.	.	-13	-6	-4	-3						
.	.	-14	-11	-9	-5						
.	.	-15	-12	-10	-8						

Figure III.3

It is easy to check that the maximal spread is $\lceil 3N^2/8 - 1/2 \rceil$, where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. Note that no smaller maximum spread can be expected with this layout, since three adjacent $N/2$-by-$N/2$ squares must contain $3N^2/4$ different integers within a range of $2d$.

For odd N, we can achieve $d = 3N^2/8 + N/4 + O(1)$. Start with the above construction for $N+1$ (which is even). Letting $i = \frac{N+1}{2}a + i'$ and $j = \frac{N+1}{2}b + j'$, with a, b integers and $i', j' \in \{1, 2, \ldots, \frac{N+1}{2}\}$, remove $Q_{i,j}$, letting the (i,j)th position empty, if

- *either* $a = b$, $i' + j' = \frac{N+3}{2}$, and $i' \leq \frac{N+1}{4}$;
- *or* $a + 1 = b$, $i' + j' = \frac{N+3}{2}$, and $i' > \frac{N+1}{4}$.

For example, for $N = 7$, we obtain

						43	45	48	. .
					38		44	47	. .
					36	37	42	46	. .
					33	34	35	41	. .
		11	13	16	25	30	31	32	
	6		12	15	19	26	28	29	
	4	5	10	14	18	21		27	
	1	2	3	9	17	20	22		
. .	-7	-2	-1	0					
. .	-13	-6	-4	-3					
. .	-14	-11		-5					
. .	-15	-12	-10						

Figure III.4

The remaining numbers will clearly form a set which has a bijective order-preserving map to the set of all integers. (In fact the set of all such maps is countable and is parametrized by an element left invariant.) To complete the construction, apply such a map to the remaining integers in the arrangement. (In the example presented, we used the map that keeps the number 10 invariant.)

						39	41	44	. .
					36		40	43	. .
					34	35	38	42	. .
					31	32	33	37	. .
		11	13	16	23	28	29	30	
	8		12	15	19	24	26	27	
	6	7	10	14	18	21		25	
	3	4	5	9	17	20	22		
. .	-5	0	1	2					
. .	-9	-4	-2	-1					
. .	-10	-7		-3					
. .	-11	-8	-6						

Figure III.5

The maximal spread in this case is easily seen to be $3N^2/8 + N/4 + O(1)$. ■

C. Lower bound on the spread

We now consider an arbitrary arrangement with a footprint where each row and column contains exactly N integers. We shall show that for this arrangement the corresponding maximal distance d satisfies $d \geq 3N^2/8 - 1/2$. We shall use the word 'line' to mean either a row, either a column in the arrangement, and will denote a line $\mathcal{L}$. Any line in the arrangement contains exactly N integers.

Definition III.2: Let $\mathcal{L}_0$ be an arbitrary row in the arrangement. Let m_1 be the smallest element of $\mathcal{N}_0 = C(\mathcal{L}_0)$. Let $\mathcal{L}_1$ be the column containing m_1. Let m_2 be the smallest element of $\mathcal{N}_1 = R(\mathcal{L}_1)$. Let $\mathcal{L}_2$ be the row containing m_2. Let m_3 be the smallest element of $\mathcal{N}_2 = C(\mathcal{L}_2)$.

Continue defining m_i, $\mathcal{L}_i$ and $\mathcal{N}_i$ for all i in the same manner. Specifically, if i is a positive odd integer and m_i has already been defined, let $\mathcal{L}_i$ be the column containing m_i, let $\mathcal{N}_i = R(\mathcal{L}_i)$. If i is a positive even integer and m_i has already been defined, let $\mathcal{L}_i$ be the row containing m_i, let $\mathcal{N}_i = C(\mathcal{L}_i)$. For any positive integer i, if $\mathcal{N}_i$ has already been defined, let m_{i+1} be the smallest element of $\mathcal{N}_i$.

Unwinding Definition III.2, we get columns $\mathcal{L}_1$, $\mathcal{L}_3$, $\mathcal{L}_5$, $\ldots$, and rows $\mathcal{L}_2$, $\mathcal{L}_4$, $\mathcal{L}_6$, $\ldots$. For all positive integers i, we have $m_i \in \mathcal{L}_i \subset \mathcal{N}_i$ and $\mathcal{L}_{i+1} \subset \mathcal{N}_i$. Also note that $\mathcal{N}_i$ has exactly N^2 elements.

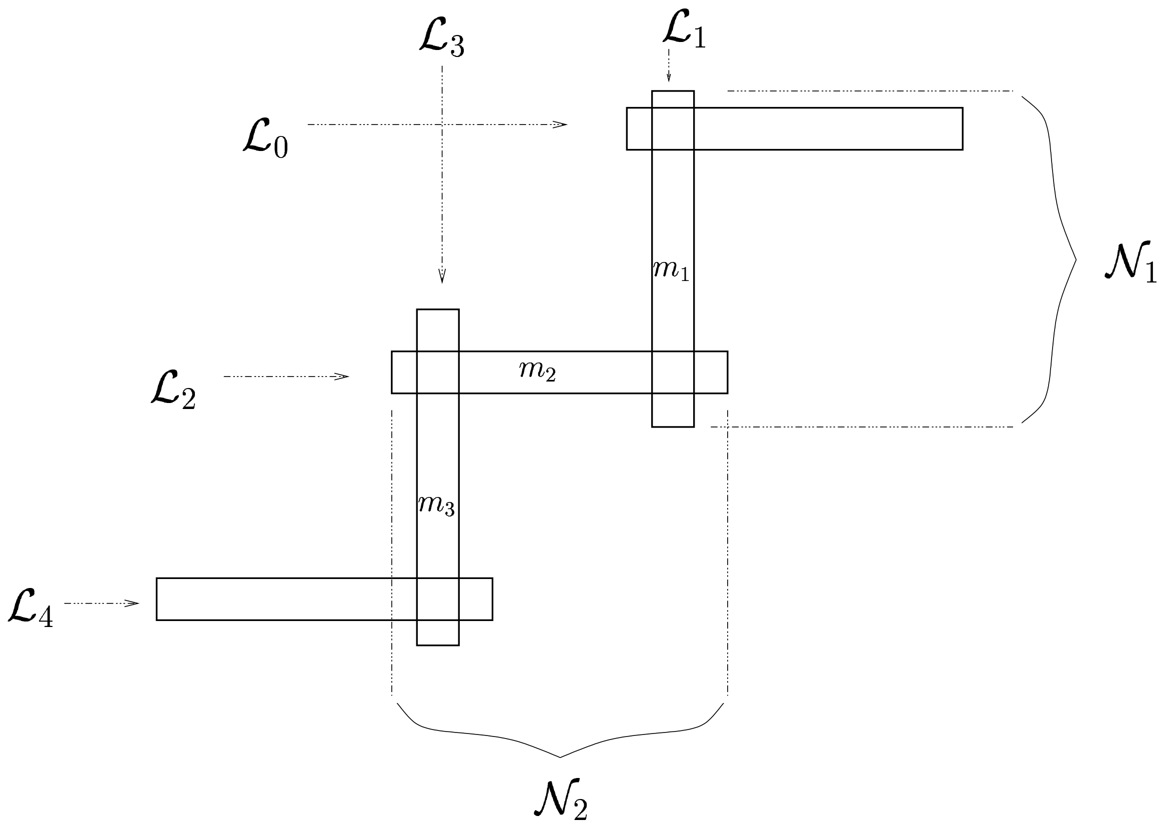

Figure III.6 The locations of $\mathcal{L}_i$s and m_is compared with $\mathcal{N}_1$ and $\mathcal{N}_2$.

Lemma III.3: For all $x \in \mathcal{N}_i$, we have

$$m_{i+1} \le x \le m_i + 2d.$$

Proof: The lower bound is valid by the definition of m_{i+1}.
Assume first that i is odd. Let x_0 be the unique element in the intersection of the column $\mathcal{L}_i$ and the row containing x. By the definition of d, we have $x - x_0 \le d$ and $x_0 - m_i \le d$. The sum of these two inequalities yields the upper bound of the lemma. The proof for even i works analogously (the rows and columns trade their roles). ■

Lemma III.4: For any positive integer i,

$$m_i - m_{i+1} \le d.$$

Proof: Let x_0 be the unique element of $\mathcal{L}_i \cap \mathcal{L}_{i+1}$. Then $x_0 - m_{i+1} \le d$ by the definition of d, and $m_i \le x_0$ by the definition of m_i. Thus, the lemma follows. ■

Corollary III.5: For any positive integers i and k,

$$m_i - m_{i+k+1} \leq kd.$$

Proof: Apply Lemma III.4 to i, $i+1$, $\ldots$, $i+k$, and combine the results. ■

Lemma III.6: For any positive integer i,

$$N^2 - 2d - 1 \leq m_i - m_{i+1}. \tag{1}$$

Proof: By Lemma III.3,

$$\max(\mathcal{N}_i) \leq m_i + 2d.$$

By the definition of m_{i+1},

$$m_{i+1} = \min(\mathcal{N}_i).$$

Since $\mathcal{N}_i$ has N^2 elements, we also have

$$\min(\mathcal{N}_i) + N^2 - 1 \leq \max(\mathcal{N}_i).$$

Summing the three displayed inequalities and rearranging yields the claimed result. ■

Lemma III.7: For any positive integer i,

$$|\mathcal{N}_i \cap \mathcal{N}_{i+3}| \leq 2d + 1 - (m_{i+1} - m_{i+3}).$$

Proof: If $\mathcal{N}_i \cap \mathcal{N}_{i+3}$ is empty then the inequality holds since the right hand side of the inequality is always at least 1 by Lemma III.4. Let us now assume that $\mathcal{N}_i \cap \mathcal{N}_{i+3}$ is not empty.

Since any element of $\mathcal{N}_i \cap \mathcal{N}_{i+3}$ is also in $\mathcal{N}_i$, by Lemma III.3 we have

$$m_{i+1} \leq \min(\mathcal{N}_i \cap \mathcal{N}_{i+3}).$$

Since any element of $\mathcal{N}_i \cap \mathcal{N}_{i+3}$ is also in $\mathcal{N}_{i+3}$, by Lemma III.3 we also have

$$\max(\mathcal{N}_i \cap \mathcal{N}_{i+3}) \leq m_{i+3} + 2d.$$

Finally, it is clearly also true that

$$\min(\mathcal{N}_i \cap \mathcal{N}_{i+3}) + |\mathcal{N}_i \cap \mathcal{N}_{i+3}| - 1 \leq \max(\mathcal{N}_i \cap \mathcal{N}_{i+3}).$$

Combining the three displayed inequalities proves this lemma. ■

We will prove the lower bound by contradiction. The following hypothesis embodies the opposite of the desired conclusion and the subsequent lemmas will be established assuming that this hypothesis holds. These lemmas will in turn be used to prove by contradiction the main theorem of this section.

Hypothesis III.8: Assume that

$$d < 3N^2/8 - 1/2.$$

The following Corollary demonstrates that the construction in Definition III.2 will never halt if this hypothesis holds, in the sense that the successive $\mathcal{N}_i$ always contain numbers that have not occurred in any $\mathcal{N}_i$ before.

Corollary III.9: If Hypothesis III.8 holds, then given a pair of distinct non-negative integers i and j, $m_i \neq m_j$, $\mathcal{L}_i \neq \mathcal{L}_j$ and $\mathcal{N}_i \neq \mathcal{N}_j$.

Proof: Assume Hypothesis III.8. It implies that $N^2 - 2d - 1$ is positive. Thus, $m_i \neq m_j$ follows by Lemma III.6. This implies that $\mathcal{L}_i \neq \mathcal{L}_j$ and $\mathcal{N}_i \neq \mathcal{N}_j$. ■

Lemma III.10: If Hypothesis III.8 holds, then for any positive integer i, and any integer $k \geq 2$,

$$\mathcal{N}_i \cap \mathcal{N}_{i+2k} = \emptyset.$$

Proof: For a proof by contradiction, assume that there is a row (respectively column) $\mathcal{Z}$ that contains elements of both $\mathcal{L}_i$ and $\mathcal{L}_{i+2k}$, if i is odd (respectively even). Let x_0 be the unique element of $\mathcal{L}_{i+2k} \cap \mathcal{Z}$, and let x_1 be the unique element of $\mathcal{L}_i \cap \mathcal{Z}$. Then $x_0 - m_{i+2k} \leq d$ and $x_1 - x_0 \leq d$. However, since x_1 is in $\mathcal{N}_{i-1}$, we also have $x_1 \geq m_i$. Combining these inequalities, we obtain that

$$m_i - m_{i+2k} \leq 2d.$$

Summing (1) for $i, \ldots, i+2k-1$, we obtain

$$2k(N^2 - 2d - 1) \leq m_i - m_{i+2k}. \tag{2}$$

Combining the two displayed inequalities above, we obtain

$$2kN^2 - 2k \leq (4k+2)d,$$

which contradicts Hypothesis III.8, thus proving the lemma. ■

Lemma III.11: If Hypothesis III.8 holds, then for any positive integer i, and any integer $k \geq 2$,

$$\mathcal{N}_i \cap \mathcal{N}_{i+2k+1} = \emptyset.$$

Proof: For a proof by contradiction, assume that $\mathcal{N}_i \cap \mathcal{N}_{i+2k+1}$ is not empty. Then the method of the proof of Lemma III.7 goes through to show that

$$1 \leq |\mathcal{N}_i \cap \mathcal{N}_{i+2k+1}| \leq 2d + 1 - (m_{i+1} - m_{i+2k+1}).$$

However, the repeated application of (1), similarly to (2), yields

$$2k(N^2 - 2d - 1) \leq m_{i+1} - m_{i+2k+1}.$$

This implies

$$2k(N^2 - 2d - 1) \leq 2d,$$

which contradicts Hypothesis III.8, thus proves the lemma. ■

Definition III.12: For any integer $i \geq 2$, let

$$\mathcal{H}_i = \mathcal{N}_{i-1} \cap \mathcal{N}_{i+1}.$$

For any integer $i \geq 3$, let

$$\mathcal{G}_i = \mathcal{N}_i - \mathcal{N}_i \cap \mathcal{N}_{i-2} - \mathcal{N}_i \cap \mathcal{N}_{i+2} = \mathcal{N}_i - \mathcal{H}_{i-1} - \mathcal{H}_{i+1}.$$

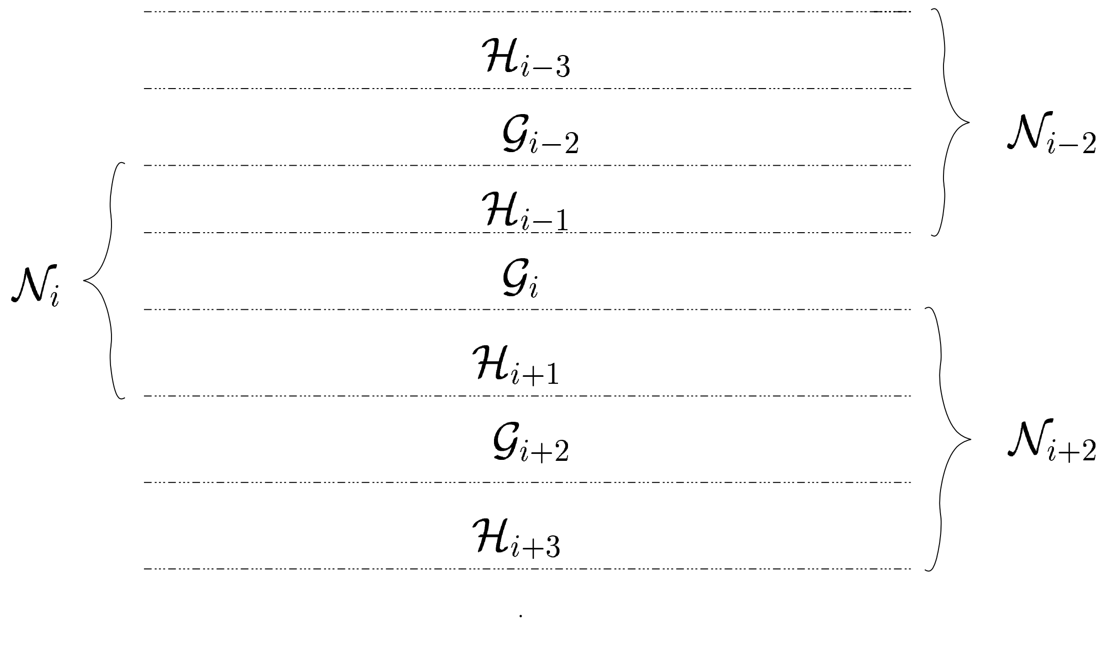

Figure III.7 The location of $\mathcal{H}_i$s and $\mathcal{G}_i$s compared with $\mathcal{N}_i$.

Lemma III.13: If Hypothesis III.8 holds, then for any integer $i \geq 2$,

$$|\mathcal{H}_i| = |\mathcal{N}_{i+1} \cap \mathcal{N}_{i-1}| \leq N^2/2.$$

Proof: Note that $\mathcal{H}_i$ is not empty since $\mathcal{L}_i \subset \mathcal{H}_i$. Since any element of $\mathcal{H}_i$ is also in $\mathcal{N}_{i-1}$, by Lemma III.3 we have

$$m_i \leq \min(\mathcal{H}_i).$$

Since any element of $\mathcal{H}_i$ is also in $\mathcal{N}_{i+1}$, by Lemma III.3 we also have

$$\max(\mathcal{H}_i) \leq m_{i+1} + 2d.$$

Finally, it is clearly also true that

$$\min(\mathcal{H}_i) + |\mathcal{H}_i| - 1 \leq \max(\mathcal{H}_i).$$

Combining the three displayed inequalities and using Lemma III.6, we obtain

$$|\mathcal{H}_i| \leq 4d + 2 - N^2,$$

which implies the lemma, since by Hypothesis III.8 the right hand side of the last inequality is at most $N^2/2$. ∎

Corollary III.14: If Hypothesis III.8 holds, then for any integer $i \geq 3$,

$$|\mathcal{G}_i \cap \mathcal{H}_i| \leq |\mathcal{G}_i|/2.$$

Proof: Assume first that i is odd. Every column intersecting $\mathcal{H}_i$ is entirely in $\mathcal{H}_i$. Therefore, Lemma III.13 implies that $\mathcal{H}_i$ has at most $N/2$ columns. Therefore, each row of $\mathcal{G}_i$ can contain no more than $N/2$ elements of $\mathcal{H}_i$. Similarly, each row intersecting $\mathcal{G}_i$ is entirely in $\mathcal{G}_i$, which implies that the size of $\mathcal{G}_i \cap \mathcal{H}_i$ is at most half the size of $\mathcal{G}_i$, as asserted. For i even, the proof is analogous (with rows and columns trading their roles). ■

Lemma III.15: If Hypothesis III.8 holds, then for any integer $i \geq 4$,

$$N^2 \leq 8d + 4 + m_{i+3} + m_{i+2} - m_{i-1} - m_{i-2} + |\mathcal{G}_{i+1}|/2 + |\mathcal{G}_i| + |\mathcal{G}_{i-1}|/2. \tag{3}$$

Proof: We will prove the lemma by giving an upper bound on $|\mathcal{N}_i| = N^2$.

By Lemmas III.10 and III.11, the set $\mathcal{N}_i$ is entirely contained in the union of the sets $\mathcal{N}_{i-3}$, $\mathcal{N}_{i-1}$, $\mathcal{N}_{i+1}$ and $\mathcal{N}_{i+3}$, which union is in turn equal to the (disjoint) union of the sets $\mathcal{N}_{i-3}$, $\mathcal{G}_{i-1}$, $\mathcal{H}_i$, $\mathcal{G}_{i+1}$ and $\mathcal{N}_{i+3}$. From Lemma III.7, we immediately obtain

$$\begin{aligned} |\mathcal{N}_i \cap \mathcal{N}_{i-3}| &\leq 2d + 1 - (m_{i-2} - m_i), \\ |\mathcal{N}_i \cap \mathcal{N}_{i+3}| &\leq 2d + 1 - (m_{i+1} - m_{i+3}). \end{aligned}$$

In order to estimate the remaining part of $\mathcal{N}_i$, namely its intersection with $\mathcal{G}_{i-1} \cup \mathcal{H}_i \cup \mathcal{G}_{i+1}$, we need to decompose $\mathcal{N}_i$ itself as the (disjoint) union $\mathcal{H}_{i-1} \cup \mathcal{G}_i \cup \mathcal{H}_{i+1}$. Doing so, we broke the remaining part of $\mathcal{N}_i$ into nine fragments. Three of them can be disposed of in the following way:

$$|\mathcal{G}_i \cap (\mathcal{G}_{i-1} \cup \mathcal{H}_i \cup \mathcal{G}_{i+1})| \leq |\mathcal{G}_i|.$$

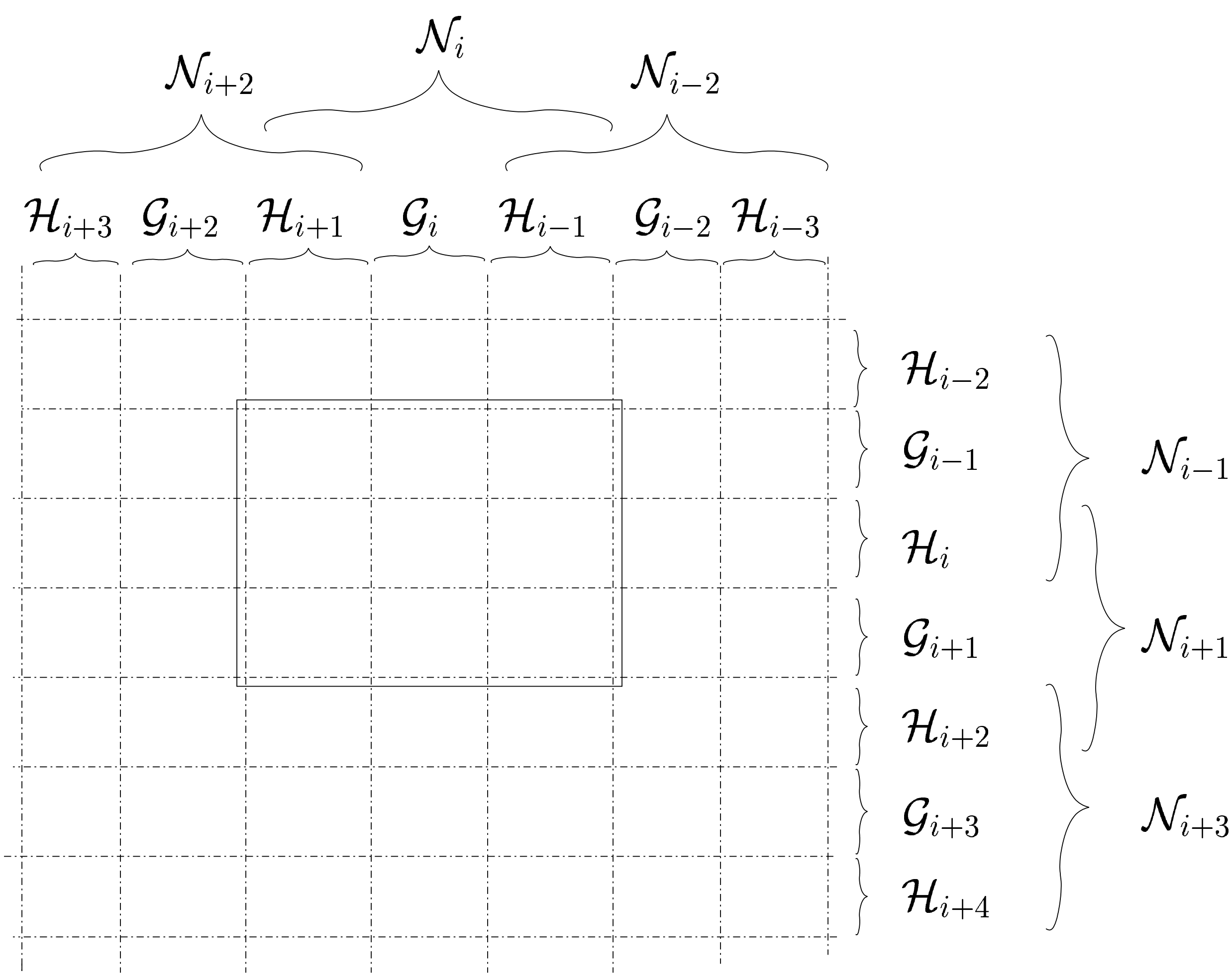

Figure III.8 The nine fragments of $\mathcal{N}_i$.

Four other fragments can be estimated from above using Lemma III.7 as follows:

$$\begin{aligned} |\mathcal{H}_{i-1} \cap (\mathcal{H}_i \cup \mathcal{G}_{i+1})| \le |\mathcal{N}_{i-2} \cap \mathcal{N}_{i+1}| &\le 2d + 1 - (m_{i-1} - m_{i+1}), \\ |\mathcal{H}_{i+1} \cap (\mathcal{H}_i \cup \mathcal{G}_{i-1})| \le |\mathcal{N}_{i+2} \cap \mathcal{N}_{i-1}| &\le 2d + 1 - (m_i - m_{i+2}). \end{aligned}$$

Finally, we can invoke Corollary III.14 twice to obtain

$$\begin{aligned} |\mathcal{G}_{i-1} \cap \mathcal{H}_{i-1}| &\le |\mathcal{G}_{i-1}|/2, \\ |\mathcal{G}_{i+1} \cap \mathcal{H}_{i+1}| &\le |\mathcal{G}_{i+1}|/2. \end{aligned}$$

Summing all the displayed equations in this proof yields the assertion of the lemma. ■

Lemma III.16: If Hypothesis III.8 holds, then for any integers $t \ge 3$ and $k \ge 2$, we have

$$|\mathcal{G}_t| + |\mathcal{G}_{t+4}| + \cdots + |\mathcal{G}_{t+4k-4}| \le 2d + 1 + m_t - m_{t+4k-3} - (k-1)N^2. \tag{4}$$

Proof: Let $\mathcal{T}$ denote the (disjoint) union of the sets $\mathcal{G}_t$, $\mathcal{N}_{t+2}$, $\mathcal{G}_{t+4}$, $\mathcal{N}_{t+6}$, $\ldots$, $\mathcal{N}_{t+4k-6}$ and $\mathcal{G}_{t+4k-4}$. By Lemma III.3 (and since the m_i form a decreasing sequence),

$$\begin{aligned} \max(\mathcal{T}) &\le m_t + 2d, \\ m_{t+4k-3} &\le \min(\mathcal{T}). \end{aligned}$$

We also have

$$\min(\mathcal{T}) + |\mathcal{T}| - 1 \le \max(\mathcal{T}).$$

Since each $\mathcal{N}_i$ has size N^2, we also know that

$$|\mathcal{T}| = |\mathcal{G}_t| + |\mathcal{G}_{t+4}| + \cdots + |\mathcal{G}_{t+4k-4}| + (k-1)N^2.$$

Combining the displayed (in)equalities, we obtain the claimed result. ■

Theorem III.17: Assume that the integers have been arranged within a plane square grid, with each row and column containing exactly N numbers, and that the supremum of the row and column spreads is d. Then

$$3N^2/8 - 1/2 \le d.$$

Proof: We prove the theorem by contradiction. Assume that $d < 3N^2/8 - 1/2$ (this is exactly Hypothesis III.8.)

Let k be any positive integer with $k \ge 2$.

First of all, sum (3) with i taking each of the values 4, 5, 6, $\ldots$, $4k+3$ to obtain

$$4kN^2 \le 4k(8d+4) + 2(|\mathcal{G}_4| + |\mathcal{G}_5| + \cdots + |\mathcal{G}_{4k+3}|) + \mu_1^+ - \mu_1^- + \gamma. \tag{5}$$

Here μ_1^+ and μ_1^- involve values of various m_i, specifically

$$\begin{aligned} \mu_1^+ &= m_{4k+6} + 2m_{4k+5} + 2m_{4k+4} + 2m_{4k+3} + m_{4k+2}, \\ \mu_1^- &= m_6 + 2m_5 + 2m_4 + 2m_3 + m_2, \end{aligned}$$

whereas γ involves various values of $|\mathcal{G}_i|$,

$$\gamma = |\mathcal{G}_3|/2 - |\mathcal{G}_4|/2 - |\mathcal{G}_{4k+3}|/2 + |\mathcal{G}_{4k+4}|/2.$$

The sum of $|\mathcal{G}_i|$ in (5) can be estimated from above by summing (4) for each of $t = 7$, $t = 6$, $t = 5$ and $t = 4$, and multiplying with 2, to obtain

$$2(|\mathcal{G}_4| + |\mathcal{G}_5| + \cdots + |\mathcal{G}_{4k+3}|) \leq 16d + 8 - 8(k-1)N^2 + \mu_2^+ - \mu_2^-, \tag{6}$$

where

$$\begin{aligned} \mu_2^+ &= 2m_7 + 2m_6 + 2m_5 + 2m_4, \\ \mu_2^- &= 2m_{4k+4} + 2m_{4k+3} + 2m_{4k+2} + 2m_{4k+1}. \end{aligned}$$

Now combining (5) and (6), we obtain

$$(12k - 8)N^2 \leq d(32k + 16) + (16k + 8) + \gamma + \mu_1^+ + \mu_2^+ - \mu_1^- - \mu_2^-,$$

which rearranges to

$$3N^2 \leq (8d + 4) + \frac{1}{k}\left[2N^2 + 4d + 2 + \gamma/4 + (\mu_1^+ - \mu_2^-)/4 + (\mu_2^+ - \mu_1^-)/4\right]. \tag{7}$$

Note that

$$\mu_1^+ - \mu_2^- = (m_{4k+6} - m_{4k+2}) + 2(m_{4k+5} - m_{4k+1}).$$

Hence, by Corollary III.5,

$$0 \geq \mu_1^+ - \mu_2^- \geq -4d - 2(4d) = -12d.$$

Similary, we can show that $|\mu_2^+ - \mu_1^-| \leq 12d$. It is also clear that the absolute value of γ is no bigger than N^2. Since none of these bounds depends on k, we may let k tend to (positive) infinity in (7) to obtain

$$3N^2 \leq 8d + 4,$$

which implies

$$3N^2/8 - 1/2 \leq d.$$

This contradicts Hypothesis III.8, thereby completing the proof (by contradiction) of Theorem III.17.

In summary, we have shown that

$$3N^2/8 - 1/2 \leq d.$$

■

Remark III.18: For odd N, the methods of Theorem III.17 extend to show that

$$3N^2/8 + N/8 - 1/2 \leq d.$$

The proof of Theorem III.17 can be modified to prove Remark III.18 as follows. Changing the assertion in Hypothesis III.8 to $d < 3N^2/8 + N/8 - 1/2$, instead of the assertion of Lemma III.13 we can prove

$$|\mathcal{H}_i| = |\mathcal{N}_{i+1} \cap \mathcal{N}_{i-1}| < N(N+1)/2.$$

However, since $|\mathcal{H}_i|$ must be divisible by N, this in fact implies

$$|\mathcal{H}_i| \leq N(N-1)/2. \tag{8}$$

Using this, we can strengthen the claim of Lemma III.14 to

$$|\mathcal{G}_i \cap \mathcal{H}_i| \leq \frac{N-1}{2N}|\mathcal{G}_i|.$$

Recall that by Lemmas III.10 and III.11, the set $\mathcal{N}_i$ of N^2 elements is the disjoint union of $\mathcal{H}_{i-1}$, $\mathcal{G}_i$ and $\mathcal{H}_{i+1}$. Therefore (8) also implies that

$$|\mathcal{G}_i| \geq N,$$

which, with Lemma III.16, implies

$$N^2/4 + N/4 \leq \lim_{k\to\infty} \frac{m_{t+4k} - m_t}{k}. \tag{9}$$

Then the proof can proceed as before, though using (6) with a factor of $2-1/N$ instead of 2, and finally utilizing (9) to conclude that

$$3N^2/8 + N/8 - 1/4 \leq d,$$

which contradicts the modified Hypothesis III.8, thus proving Remark III.18.

IV. Minimizing the variance in the square case

A. *Outline of argument*

We shall begin with showing in Section IV-B that $(N^4-1)/24$ is a lower bound on the row and column variances in an N-by-N square arrangement. The proof of this statement is based on an algebraic inequality relating the average row or column variance to the variance of the set of all numbers in the square.

In Section IV-C, we give an upper bound on the row and column variances by constructing appropriate square arrangements. First we exhibit an explicit arrangement where the maximal row or column variance is $N^4/16 + O(N^3)$. Then we prove a lemma that allows us to combine two arrangements with small N into a single one with larger N, such that the row and column variances of the bigger arrangement are well controlled by the row and column variances of the smaller arrangements. We then use this combination lemma, along with arrangements for $N \leq 10$ constructed by hand, to construct square arrangements for any N with row and column variances not exceeding $N^4/20 + O(N^3)$.

In Section IV-D, we revisit the issue of a lower bound on the row and column variances in square arrangements by showing that the algebraic inequality used in Section IV-B cannot be sharp in this application. Upon refining the inequality, this non-sharpness turns out to be related to the fact that the sum of two independent random variables of (nearly) equal variance cannot be a (nearly) uniformly distributed random variable. We prove a theorem in this direction and then use it to slightly improve the lower bound to $N^4(1/24+\eta) - 1/24$, where η is close to 4.2×10^{-7} (this is stated as Theorem IV.11).

B. A lower bound on the variance in a square

The following general theorem yields the crucial ingredient for the lower bound presented in Theorem IV.2.

Theorem IV.1: Let $X_{i,j}$ $(1 \leq i, j \leq N)$ be arbitrary real numbers. Then

$$\mathrm{Var}(X_{1,1}, X_{1,2}, \ldots, X_{N,N}) \leq \frac{1}{N}\sum_i \mathrm{Var}(X_{i,1}, X_{i,2}, \ldots, X_{i,N}) + \frac{1}{N}\sum_j \mathrm{Var}(X_{1,j}, X_{2,j}, \ldots, X_{N,j}). \tag{10}$$

Proof: Let H denote the N-by-N matrix whose diagonal elements are equal to $N-1$, and whose other elements are -1. This matrix H has eigenvalues 0 (of multiplicity one) and N (of multiplicity $N-1$), and is therefore positive semi-definite.

Let G denote the Kronecker (tensor) square of the matrix H. Then G has eigenvalues 0 (of multiplicity $2N-1$) and N^2 (of multiplicity $(N-1)^2$), and is also positive semi-definite.

Consider the quadratic form Q_G defined by the matrix G, which is a form in N^2 variables. Since the matrix G is positive semi-definite, we have

$$Q_G(X_{1,1}, X_{1,2}, \ldots, X_{N,N}) \geq 0. \tag{11}$$

The coefficients of the various monomials in the expansion of this inequality are easily seen to be as follows: The coefficients of the terms of the form $X_{i,j}^2$ (where i and j need not be distinct) are $(N-1)^2$, the coefficients of the terms of the form $X_{i,j}X_{l,k}$ (with $i \neq l$ and $j \neq k$) are 2, and the coefficients of the remaining terms (either $X_{i,j}X_{i,k}$ or $X_{i,j}X_{l,j}$) are all $2(1-N)$. There are no other terms.

By expanding the definitions of the variances in (10), we see that (10) is in fact the same as N^{-4} times (11). This completes the proof. ■

Theorem IV.1 can be generalized to higher dimensions. This will be dealt with in a forthcoming paper [8].

Theorem IV.2: Assume that the integers 1 through N^2 have been arranged in an N-by-N matrix P, and that the largest variance of any row or column is V. Then

$$(N^4-1)/24 \leq V.$$

Proof: An easy calculation shows that the variance of the integers from 1 through N^2 is

$$\mathrm{Var}(1, 2, \ldots, N^2) = (N^4-1)/12. \tag{12}$$

Now we can apply Theorem IV.1 to the matrix P:

$$(N^4-1)/12 \leq \frac{1}{N}\sum_i \mathrm{Var}(P_{i,1}, P_{i,2}, \ldots, P_{i,N}) + \frac{1}{N}\sum_j \mathrm{Var}(P_{1,j}, P_{2,j}, \ldots, P_{N,j}). \tag{13}$$

Since every row or column variance on the right hand side is at most V, the statement of the theorem follows immediately. ■

C. Upper bound on the optimal variance in a square

Theorem IV.3: For any N, there exists an arrangement of the integers 1 through N^2 where the variance of the numbers in a single row or column is at most $N^4/16 + O(N^3)$.

Proof: First, let N be any even integer and define the N-by-N matrix P (containing the integers 1 through N^2) by

$$P_{i,j} = \begin{cases} (i-1)N/2 + j & \text{if } j \leq N/2, \\ (i-1)N/2 + j + (N^2 - N)/2 & \text{otherwise.} \end{cases} \quad (14)$$

It is easy to check that the variance of each row of P is $N^4/16 + N^2/48 - 1/12$, and the variance of each column is $N^4/48 - N^2/48$.

This construction can also be used to give a construction with maximal variance $N^4/16 + O(N^3)$ for odd N. For any odd N,

- *Step A.* take the construction just given for $N + 1$;
- *Step B.* omit a row and a column to get an N-by-N square; and
- *Step C.* apply the unique order-preserving bijection between the remaining numbers and the set $\{1, 2, \ldots, N^2\}$.

In Step A, each row and column variance is bounded above by $(N+1)^4/16+O((N+1)^3) = N^4/16 + O(N^3)$, as described for the case of even N above. To estimate the change of variance in Step B we need the following observation.:

Fact IV.4: If a set of n numbers has variance v then any subset of $n-1$ numbers in it has variance no more than $nv/(n-1)$.

In Step B, using Fact IV.4, the variance of each row or column increases by no more than a factor of $1 + 1/N$, and is therefore still bounded above by $N^4/16 + O(N^3)$ (of course the constant implicit in the O factor increases). Finally, Step C cannot increase variances either since no pairwise distances are increased between the numbers (as stated in Lemma V.2(iii) in Section V-C), and therefore we can conclude that in the new arrangement, the maximal variance obtained in any row or column will not exceed $N^4/16+O(N^3)$. ■

By checking all the possible arrangements, we can verify that the construction given above is optimal for $N = 2$. However, for any specific N over 2, it is easy to find an arrangement that has a lower maximal variance than the general construction given above. For example, for N between 3 and 10, we have found arrangements[3] with the following maximal variances:

N	3	4	5	6	7	8	9	10
V	4.7	12.2	29.6	59.8	109.7	184.5	299.9	455.2
$10^2V/N^4$	5.80	4.77	4.74	4.61	4.57	4.50	4.57	4.55

Figure IV.1

Known good arrangements can be combined to form larger arrangements with relatively low maximal row and column variances. The following lemma (which is stated with enough generality to be usable in Section V-B for infinite arrangements) explains how.

Lemma IV.5: Assume that an arrangement A (resp. B) has exactly N (resp. N') integers in every row and column, and that the maximal variance within a row or column is at most V (resp. V'). The arrangement A must be an N-by-N square, whereas the arrangement B can be either an N'-by-N' square or an arrangement in the whole plane square grid. Then there is an arrangement C with exactly NN' integers in every row and column, and the maximal variance within a row or column is at most $V + N^4V'$. The

[3]These arrangements are available at `www.csirik.net/square-variances.html`.

arrangement C is square (of size NN'-by-NN') or infinite, depending on whether B is square or infinite.

Proof: Define $C_{i,j}$ as follows. For

$$\begin{aligned} i &= Na + i', \qquad \text{with } i' \in \{1, 2, \dots, N\}, \\ j &= Nb + j', \qquad \text{with } j' \in \{1, 2, \dots, N\}, \end{aligned}$$

where a and b are integers, set

$$C_{i,j} = \begin{cases} A_{i',j'} + N^2(B_{a+1,b+1} - 1) & \text{if } B_{a+1,b+1} \text{ is not an empty place} \\ \text{empty place} & \text{if } B_{a+1,b+1} \text{ is an empty place.} \end{cases}$$

It is then clear that

$$\begin{aligned} \mathrm{Var}(C_{i,\bullet}) &= \mathrm{Var}(A_{i',\bullet}) + N^4 \mathrm{Var}(B_{a,\bullet}) \le V + N^4 V', \\ \mathrm{Var}(C_{\bullet,j}) &= \mathrm{Var}(A_{\bullet,j'}) + N^4 \mathrm{Var}(B_{\bullet,b}) \le V + N^4 V', \end{aligned}$$

which proves the lemma. ■

For example, if we apply Lemma IV.5 with A being the optimal arrangement for $N = 2$ (with $V = 1$), and B being the optimal arrangement for $N' = 3$ (with $V' = 14/3$), namely

$$A: \begin{matrix} 3 & 4 \\ 1 & 2 \end{matrix}, \qquad B: \begin{matrix} 6 & 7 & 9 \\ 3 & 5 & 8 \\ 1 & 2 & 4 \end{matrix},$$

we obtain

$$\begin{matrix} 23 & 24 & 27 & 28 & 35 & 36 \\ 21 & 22 & 25 & 26 & 33 & 34 \\ 11 & 12 & 19 & 20 & 31 & 32 \\ 9 & 10 & 17 & 18 & 29 & 30 \\ 3 & 4 & 7 & 8 & 15 & 16 \\ 1 & 2 & 5 & 6 & 13 & 14 \end{matrix},$$

with a maximal row or column variance of $227/3 \approx 75.7$.

Theorem IV.6: For any positive integer N, let $V(N)$ be the smallest possible maximal variance in a row or column when the numbers 1 through N^2 are arranged in an N-by-N square. Let $c(N)$ be defined as

$$c(N) = V(N)/N^4.$$

Then

(a) $c(2N) \le c(N) + 1/(16N^4)$,
(b) $c(N) \le c(N+1)(1 + 1/N)^5$.

Proof: Part (a) follows from applying Lemma IV.5 to the optimal construction for A of size 2-by-2 and the optimal construction for B of size N-by-N.

Part (b) follows from the argument given in the proof of Theorem IV.3: take the optimal construction for an $(N+1)$-by-$(N+1)$ square, delete a row and a column, and adjust the remaining numbers (bijectively) to lie in the interval $[1, N^2]$. The inequality follows from Fact IV.4. ■

It is clear that the results of Theorem IV.6 can be used to improve the upper bound on all $c(N)$ if sufficiently good constructions are available for small N. For example, given arrangements for all integers from 1 through N_0, we can use Theorem IV.6(a) to upper bound good arrangements for all even integers in the range $[N_0+1, 2N_0]$, and then use Theorem IV.6(b) to upper bound good arrangements for all odd integers in the same range. This procedure can be iterated and will clearly give a universal upper bound on $c(N)$, which would depend on how good the initial set of arrangements were. Using this method and the constructions for $N \leq 10$ given above, we can get

$$c(N) < 1/20 \qquad \text{(for } N \geq 4\text{)}.$$

For example,

$$\begin{aligned} c(16) &\leq c(8) + 1/(16 \cdot 8^4), \\ c(32) &\leq c(16) + 1/(16 \cdot 16^4) \leq c(8) + 1/(16 \cdot 8^4) + 1/(16 \cdot 16^4), \end{aligned}$$

and so on, from which it follows that

$$c(N) \leq 1/22 \qquad \text{(for } N = 2^k \geq 8\text{)},$$

where k is of course meant to be an integer.

D. A better lower bound on the variance in a square

Since the lower bound obtained with this construction does not agree with that given in Theorem IV.2, it is natural to ask whether the theorem could be improved. The matrix P given in the proof of Theorem IV.3 illustrates that it is possible for the result of Theorem IV.2 not to be sharp, even though the result of Theorem IV.1 is sharp. The problem is that Theorem IV.2 allows us to give a lower bound on the *average* of the row and column variances of any square matrix P containing the numbers 1 through N^2, but for any particular matrix, either the inequality of Theorem IV.1 is not sharp, or the average and the maximal row and column variances are not equal. In this section, we will develop these ideas to slightly improve the lower bound of Theorem IV.2. Larger improvements should be possible by more tightly controlling the various estimates given below. First, we need the following more precise version of Theorem IV.1.

Theorem IV.7: Let $X_{i,j}$ $(1 \leq i, j \leq N)$ denote the elements of an N-by-N matrix. Assume that $\sum_{i,j} X_{i,j} = 0$. Then

$$\mathrm{Var}(X_{1,1}, \ldots, X_{N,N}) + \frac{1}{N^2} \sum_{i,j} H_{i,j}^2 = \frac{1}{N} \sum_i \mathrm{Var}(X_{i,1}, \ldots, X_{i,N}) + \frac{1}{N} \sum_j \mathrm{Var}(X_{1,j}, \ldots, X_{N,j}), \tag{15}$$

where

$$H_{i,j} = X_{i,j} - \frac{1}{N} \sum_k X_{i,k} - \frac{1}{N} \sum_k X_{k,j},$$

for all $1 \leq i, j \leq N$.

Proof: Adding $(\sum_{i,j} X_{i,j})^2/N^4$ to the right-hand side of (15), we get a polynomial identity that is easy to verify. However, by assumption we have $\sum_{i,j} X_{i,j} = 0$, which yields the statement of the theorem.[4] ■

Let us now return to the matrix P as in Theorem IV.2. We will scale P to simplify the calculations. Define an N-by-N matrix U by $U_{i,j} = (P_{i,j} - (N^2+1)/2)/N^2$. Thus $\sum_{i,j} U_{i,j} = 0$ and the elements of U all lie within $[-1/2, 1/2]$. Clearly, the largest row or column variance α in U will be equal to V/N^4 and all other interesting properties of P will be represented in U, too. We can write

$$U = F + G + H,$$

where each element of F is the average of the elements in the corresponding row of U, each element of G is the average of the elements in the corresponding column of U, and H is defined as $U - F - G$.

Let us now consider the matrices U, F, G and H as random variables on the probability space $T_R \times T_C$ (where T_R and T_C are N-element sets representing rows and columns, respectively), where U is endowed with the uniform distribution, and F, G and H are defined as before.

Regardless of how the $N \times N$ matrix P containing all integers from 1 to N^2 is obtained, namely either using a construction rule or drawing it at random, we shall consider its entries as randomly distributed with uniform distribution. Similarly, the entries of the scaled matrix U associated with P will be considered as a uniformly distributed random variable taking its values in the interval $[-1/2, 1/2]$. Moreover, although the entries of U are in finite number so this random variable is discrete, we shall approximate it by a continuous random variable uniformly distributed over the interval $[-1/2, 1/2]$, an approximation which becomes more and more accurate as N tends to infinity. We shall also denote this random variable U, the context unambiguously indicating if U denotes a scaled matrix or the associated continuous random variable. Then, Theorem IV.7 implies

$$\mathrm{Var}(U) + \mathbb{E}(H^2) \leq 2\alpha \tag{16}$$

where $\mathrm{Var}(U)$ denotes the variance of the random variable associated with matrix U, and $\mathbb{E}(H^2)$ similarly denotes the average of the squares of all the entries of H. Remember that α denotes the maximal row or column variance that occurs in the matrix U.

Let $\overline{U_{\bullet,j}}$ denote the average of all elements of row j in the matrix U. Now our next aim is to give an upper bound on $\mathrm{Var}(F)$ (remember that in any row of F every number is the same, namely the average of the entries in the corresponding row of U). Then, for any $1 \leq k \leq N$, the inequality between the arithmetic and the quadratic means implies that

$$\left(\frac{1}{N}\sum_j U_{k,j} - \frac{1}{N}\sum_j \overline{U_{\bullet,j}}\right)^2 = \left(\frac{1}{N}\sum_j (U_{k,j} - \overline{U_{\bullet,j}})\right)^2 \leq \frac{1}{N}\left(\sum_j \left(U_{k,j} - \overline{U_{\bullet,j}}\right)^2\right).$$

Summing both sides over k and dividing by N, we obtain

$$\mathrm{Var}(F) \leq \frac{1}{N}\sum_j \mathrm{Var}(U_{1,j}, U_{2,j}, \ldots, U_{N,j}) \leq \alpha, \tag{17}$$

[4]We discovered the identity (15) by calculating the contribution of the component of X that lies outside of the 0-eigenspace of the matrix G (in Theorem IV.2) to the value of the quadratic form Q_G defined by G.

and similarly

$$\mathrm{Var}(G) \leq \alpha. \tag{18}$$

Note that F and G are independent random matrices, since the entries of F (of G) are constant within rows (within columns).

We can now see clearly why Theorem IV.2 cannot be sharp. If that bound were sharp, then (16) would imply that $H = 0$ and thus

$$U = F + G.$$

Then (17) and (18) together with $\mathrm{Var}(U) = \mathrm{Var}(F) + \mathrm{Var}(G)$ would imply that

$$\mathrm{Var}(F) = \mathrm{Var}(G) = \frac{1}{2}\mathrm{Var}(U).$$

The non-sharpness of Theorem IV.2 in this point of view is related to (but does not yet immediately follow from) the following general fact.

Fact IV.8: A random variable that is uniformly distributed on some interval of the real line cannot be the sum of two independent random variables of same variance.

This fact can be proved using the following lemma.

Lemma IV.9: Let F and G be independent real-valued random variables with $\mathbb{E}(F) = \mathbb{E}(G) = 0$, and let $n_2 = \mathbb{E}((F+G)^2)$ and $n_4 = \mathbb{E}((F+G)^4)$. Then $a = \mathbb{E}(F^2)$ satisfies

$$4a^2 + n_4 \geq n_2(n_2 + 4a).$$

Proof: We have (using $\mathbb{E}(F) = 0$ to eliminate the cross term)

$$n_2 = \mathbb{E}((F+G)^2) = \mathbb{E}(F^2) + \mathbb{E}(G^2),$$

and hence

$$\mathbb{E}(G^2) = n_2 - a.$$

Clearly

$$\mathbb{E}(F^4) \geq (\mathbb{E}(F^2))^2 = a^2$$

and

$$\mathbb{E}(G^4) \geq (\mathbb{E}(G^2))^2 = (n_2 - a)^2.$$

This implies that (again using $\mathbb{E}(F) = 0$ and $\mathbb{E}(G) = 0$ to eliminate some terms)

$$n_4 = \mathbb{E}((F+G)^4) = \mathbb{E}(F^4) + 6(\mathbb{E}(F^2))(\mathbb{E}(G^2)) + \mathbb{E}(G^4) \geq a^2 + 6a(n_2 - a) + (n_2 - a)^2,$$

which rearranges to yield the statement of the lemma. ■

Proof: (Proof of Fact IV.8.) Without loss of generality, we can assume that the uniform distribution has mean 0, and that it is written as a sum of independent random variables F and G where both F and G have mean 0. Applying Lemma IV.9 with $a = \mathbb{E}(F^2) = \mathbb{E}(G^2) = n_2/2$, we obtain

$$n_4/n_2^2 \geq 2.$$

However, for a uniform distribution we have $n_4/n_2^2 = 9/5$, which shows that a decomposition of the proposed type is not possible. ■

The general question of how a uniformly distributed random variable can be broken up as a sum of two independent random variables has been studied extensively. For details, and a general characterization of those pairs of independent random variables that sum to a uniform distribution, the reader is referred to [9, Section 1.4], [10], [11] and the references contained therein. ¿From the general characterization we can infer that if F and G are independent random variables with $\mathrm{Var}(F) \geq \mathrm{Var}(G)$ and $F+G$ is uniformly distributed, then $\mathrm{Var}(F)/\mathrm{Var}(G) \geq 3$, and this is achieved only when F is uniformly distributed over the two discrete values $-1/4$ and $1/4$, and G is uniformly distributed over the shorter interval $[-1/4, 1/4]$.[5]

The following theorem is a more general version of Lemma IV.9 that allows us to cope with the fact that in the current application U is not exactly uniformly distributed, and that the variances of F and G are not exactly equal. It will allow us to improve the lower bound given in Theorem IV.2.

Theorem IV.10: Let μ, λ be arbitrary positive real numbers. Let U, H, F and G be random variables of mean 0 with

$$U - H = F + G,$$

such that F and G are independent, $\mathbb{E}(U^2) = m_2$, $\mathbb{E}(U^4) = m_4$, and assume that almost surely $|U| < \mu$ and $|H| < \lambda$.

If $0 < 2m_2^2 - m_4$ then there is an $\varepsilon = \varepsilon(m_2, m_4, \mu, \lambda) > 0$ such that the following inequalities cannot be satisfied simultaneously:

$$\mathbb{E}(F^2) = \frac{1}{2}m_2 + a, \qquad \text{with } |a| < \varepsilon, \tag{19}$$
$$\mathbb{E}(H^2) \leq 2\varepsilon. \tag{20}$$

(The proof of the theorem provides clues on how to determine such an $\varepsilon(m_2, m_4, \mu, \lambda)$.)

Proof: We will apply Lemma IV.9 to the random variables F and G. Accordingly, let us define

$$\begin{aligned} n_2 &= \mathbb{E}((F+G)^2) = \mathbb{E}((U-H)^2), \\ n_4 &= \mathbb{E}((F+G)^4) = \mathbb{E}((U-H)^4). \end{aligned}$$

Note that $\mathbb{E}|H| \leq \sqrt{\mathbb{E}(H^2)} \leq \sqrt{2\varepsilon}$, so we can bound the difference between n_2 and m_2 as follows:

$$|n_2 - m_2| = |\mathbb{E}((U-H)^2) - \mathbb{E}(U^2)| \leq 2|\mathbb{E}(UH)| + \mathbb{E}(H^2) \leq 2\mu\sqrt{2\varepsilon} + 2\varepsilon. \tag{21}$$

Similarly, we can obtain

$$|n_4 - m_4| \leq 4|\mathbb{E}(U^3H)| + 6\mathbb{E}(U^2H^2) + 4|\mathbb{E}(UH^3)| + \mathbb{E}(H^4) \leq 4\mu^3\sqrt{2\varepsilon} + 6\mu^2 2\varepsilon + 4\mu\lambda 2\varepsilon + \mu^2 2\varepsilon. \tag{22}$$

The crucial point with these inequalities is that there are continuous functions f_1 and f_2,[6] with $f_1(0) = f_2(0) = 0$, such that

$$n_2 = m_2 + b, \qquad \text{with } |b| < f_1(\varepsilon), \tag{23}$$
$$n_4 = m_4 + c, \qquad \text{with } |c| < f_2(\varepsilon). \tag{24}$$

[5]The best result in this direction implied by Lemma IV.9 is $\mathrm{Var}(F)/\mathrm{Var}(G) \geq (3+\sqrt{5})/2 = 2.62$.

[6]The functions f_1 and f_2 also depend on μ and λ, but our notation does not express this dependence for simplicity.

Now applying Lemma IV.9, we obtain

$$4(\mathbb{E}(F^2))^2 + n_4 \geq n_2(n_2 + 4\mathbb{E}(F^2)).$$

Using (19), (23), and (24), this transforms to

$$4a^2 + c - 4m_2 b - b^2 - 4ab \geq 2m_2^2 - m_4. \tag{25}$$

We can give an upper bound on the left hand side using (23), (21), (24), (22), $|a| < \varepsilon$, $-b^2 \leq 0$, $|4ab| < 4\varepsilon f_1(\varepsilon)$ and $m_2 < \mu^2$, to obtain

$$4\varepsilon^2 + f_2(\varepsilon) + 4\mu^2|f_1(\varepsilon)| + 4\varepsilon f_1(\varepsilon) \geq 2m_2^2 - m_4. \tag{26}$$

We assumed in the statement of the theorem that the right hand side here is positive. On the other hand, the left hand side is a continuous function of ε, and is 0 when $\varepsilon = 0$, so we get a contradiction if ε is chosen small enough. This proves the theorem. ■

Theorem IV.11: Assume that the integers 1 through N^2 have been arranged in an N-by-N matrix P, and the largest variance of numbers in any row or column is V. Then

$$N^4(1/24 + \eta) - 1/24 \leq V,$$

where η is close to $4.2 \cdot 10^{-7}$.

Proof: In terms of the scaled matrix U, we need to show that

$$1/24 + \eta - 1/(24N^4) \leq \alpha.$$

Since all elements of U lie in $[-1/2, 1/2]$, we have

$$|U| \leq 1/2 = \mu.$$

By construction, all elements of F and G must lie in $[-1/2, 1/2]$ too, so we have

$$|H| \leq 3/2 = \lambda.$$

A quick calculation shows that

$$2m_2^2 - m_4 = \frac{1}{720} + \frac{1}{72N^2} - \frac{11}{720N^4} > \frac{1}{720},$$

for all $N \geq 2$.

We now use (25) in the proof of Theorem IV.10 to obtain a lower bound on ε. We can substitute the values $\lambda = 3/2$ and $\mu = 1/2$ and use the upper bounds on a, b and c obtained in the proof of Theorem IV.10 (i.e., $|a| < \varepsilon$), (21) and (22) to obtain

$$\frac{3\sqrt{2}}{2}\sqrt{\varepsilon} + \frac{23}{2}\varepsilon + 4\sqrt{2}\varepsilon\sqrt{\varepsilon} + 12\varepsilon^2 > \frac{1}{720}.$$

This implies that ε is larger than the threshold η beyond which the above inequality obtains, with η close to 4.2×10^{-7}, thereby proving the theorem. ■

V. Minimizing the variance in the infinite case

A. Outline of the argument

Section V-B gives an upper bound on the row and column variances in the infinite case by construction. The footprint of this construction is the same as the one in Section III-B, and we use the combination lemma and the square arrangements of Section IV-C to fill the footprint.

Finally, in Section V-C we prove a lower bound on the row and column variances in the infinite case by a more involved version of the argument of Section IV-B combined with some ideas from Section III-C.

In summary, the main result of this section is the following theorem:

Theorem V.1: Assume that the integers have been arranged within a plane square grid, with each row and column containing exactly N numbers, and the variance of numbers in any row or column is at most V. Then

$$N^4/60 - 1/60 \leq V.$$

For any N, an arrangement exists where the variance of any row or column is at most $3N^4/160 = N^4/53\frac{1}{3}$.

B. Upper bound on the optimal variance

Proof: (Proof of Theorem V.1, upper bound.) We shall use Lemma IV.5 to construct assignments with small maximal variances V. Suppose that N is even (odd N can be handled as in Section IV-C). Let A be an $(N/2)$-by-$(N/2)$ matrix containing the integers 1 through $N^2/4$, with small maximal variance $c_A \cdot (N/2)^4$. Let B be the optimal infinite assignment of the integers to the infinite square grid where each row and column contains exactly two numbers, namely which assigns $2n$ to the position (n, n) and $2n + 1$ to the position $(n+1, n)$. The variance in any row or column of B is of course $1/4$. Let us apply Lemma IV.5 to A and B. The result is an assignment containing N numbers in each row and column, with a maximal row or column variance of

$$\left(\frac{1}{64} + \frac{c_A}{16}\right) N^4.$$

Let us now consider what this gives to us. In Theorem IV.3, we gave an explicit construction of an N-by-N square of maximal variance $N^4/16 + O(N^3)$ for any N. Choosing it for A we get an upper bound of $(5/256)N^4 + O(N^3) = (1/51.2)N^4 + O(N^3)$ in the infinite case. In Section IV-C we also showed that by 'blowing up' examples for small N, we can get squares with maximal variance no more than $N^4/20$. This gives us for the infinite case a construction with an upper bound of $(3/160)N^4 = N^4/53\frac{1}{3}$. ■

Remark. If N is a power of two, we can construct squares with maximal variances $N^4/22$ (see Section IV-C). In the infinite case, this yields constructions with maximal variances of $N^4/54.15$. However, with this method we cannot get close to the lower bound because, by Theorem IV.11, the maximal variance for an N-by-N square is always at least $N^4/24$, but even a square of maximal variance $N^4/24$ would give an infinite construction with $(7/384)N^4 = N^4/54.857$. It is far from clear if the method described in this section gives the optimal construction. On the other hand, we believe that the lower bound in Theorem V.1 is not sharp, either.

C. Lower bound on the variance

Let us fix N. Let us consider an assignment to the infinite grid, where each column and row contains N numbers, and the maximum of the variances is finite and as small as possible. Let d denote the maximal spread of this assignment and V the maximum of the variances of rows and columns. Trivially

$$\frac{d^2}{2N} \leq V.$$

If $V > N^4/2$ then the lower bound of Theorem V.1 is satisfied, and there is nothing to prove. Otherwise, we can deduce that

$$d < N^{5/2}. \tag{27}$$

We shall need the following lemma.

Lemma V.2: Let $x_1, \ldots, x_N, x, y, m$ be real numbers and N a positive integer. (In the following, the notation $\sum_{i<j}$ will be used as an abridgement of $\sum_{j=1}^{N} \sum_{i<j} \cdot$.) Then

(i)

$$\mathrm{Var}(x_1, \ldots, x_N) = \frac{N-1}{N^2} \sum x_i^2 - \frac{2}{N^2} \sum_{i<j} x_i \cdot x_j.$$

(ii)

$$\sum_{i<j} (x_i - x_j)^2 = (N-1) \sum x_i^2 - 2 \sum_{i<j} x_i \cdot x_j.$$

(iii)

$$\mathrm{Var}(x_1, \ldots, x_N) = \frac{1}{N^2} \sum_{i<j} (x_i - x_j)^2.$$

(iv)

$$(x-y)^2 = 2(m-x)^2 + 2(m-y)^2 - 4\left(m - \frac{x+y}{2}\right)^2.$$

(v)

$$\sum_{i=1}^{N^2} (x-i)^2 \geq \frac{N^6 - N^2}{12}.$$

(vi) For real numbers $A_1, \ldots, A_N, B_1, \ldots, B_N$,

$$\sum_{i,j} (A_i - B_j)^2 = \sum_{i<j} (A_i - A_j)^2 + \sum_{i<j} (B_i - B_j)^2 + (A_1 + \cdots + A_N - B_1 - \cdots - B_N)^2.$$

(vii) Fix a row R of the assignment. Then

$$\sum_{x \in R, y \in C(R)} (x-y)^2 \geq \frac{N^7 - N^3}{12} + N^3 \mathrm{Var}(R).$$

Proof: (i) follows from the very definition of variance, (ii), (iv) and (vi) are easy identities, (iii) follows from (i) and (ii), (v) follows from $\mathrm{Var}(1,\ldots,N^2) = (N^4-1)/12$. To prove (vii), we shall use (v) and notice that $C(R)$ contains N^2 distinct numbers. Let

$$m = \frac{1}{N^2}\sum_{y\in C(R)} y.$$

Then we have

$$\sum_{x\in R, y\in C(R)} (x-y)^2 = \sum_{x\in R}\sum_{y\in C(R)} (x-y)^2 = \sum_{x\in R}\sum_{y\in C(R)} ((x-m)+(m-y))^2 =$$

$$= N^2\sum_{x\in R}(x-m)^2 + N\sum_{y\in C(R)}(m-y)^2 \geq N^3\mathrm{Var}(R) + \frac{N^7-N^3}{12}.$$

■

Now we are ready to start to prove the main result of this section.

Proof: (Proof of Theorem V.1, lower bound.) Fix a row R. Recall that $C(z)$ denotes the column which contains the number z. Let

$$\mathrm{Var}_x(R) = \frac{1}{N^2}\sum_{x_i\in R}(x-x_i)^2. \tag{28}$$

By Lemma V.2 (iii),

$$\sum_{x\in R}\mathrm{Var}_x(R) = 2\mathrm{Var}(R). \tag{29}$$

Consider the sum

$$Z(R) = \sum_{x\in R, y\in C(R)} (x-y)^2. \tag{30}$$

By Lemma V.2 (vii) we have

$$Z(R) \geq \frac{N^7-N^3}{12} + N^3\mathrm{Var}(R). \tag{31}$$

For the right hand side of (30), we apply Lemma V.2 (iv), with $m = R\cap C(y)$ (recall that $C(R)$ denotes the numbers in the same column as an element of R):

$$Z(R) = \sum_{x\in R, y\in C(R), z=R\cap C(y)} \left(2(x-z)^2 + 2(y-z)^2 - 4\left(z-\frac{x+y}{2}\right)^2\right). \tag{32}$$

For the first term of the summation from (32), we apply Lemma V.2 (iii):

$$2N\sum_{x,z\in R}(x-z)^2 = 4N^3\cdot\mathrm{Var}(R).$$

For the second term, we obtain from (28)

$$2N\sum_{z\in R, y\in C(z)}(y-z)^2 = 2N^3\sum_{z\in R}\mathrm{Var}_z(C(z)).$$

To handle the third term, we use Lemma V.2 (vi) with $A_i = z - x_i$ and $B_i = -z + y_j$:

$$\sum_{x\in R}\sum_{z\in R}\sum_{y\in C(z)} 4\left(z - \frac{x+y}{2}\right)^2 = \sum_{x\in R}\sum_{z\in R}\sum_{y\in C(z)} (2z - x - y)^2$$

$$\geq \sum_{z\in R}\left(\sum_{x_i<x_j\in R}(x_i - x_j)^2 + \sum_{y_i<y_j\in C(z)}(y_i - y_j)^2\right) = N^3\mathrm{Var}(R) + N^2\sum_{z\in R}\mathrm{Var}(C(z)).$$

Combining the last three displayed equalities, we get that

$$\frac{N^7 - N^3}{12} + N^3\mathrm{Var}(R) \leq Z(R) \leq 3N^3\mathrm{Var}(R) + 2N^3\sum_{z\in R}\mathrm{Var}_z(C(z)) - N^2\sum_{z\in R}\mathrm{Var}(C(z)).$$

We can conclude that

$$\frac{N^4 - 1}{12} \leq 2\mathrm{Var}(R) + 2\sum_{z\in R}\mathrm{Var}_z(C(z)) - \frac{1}{N}\sum_{z\in R}\mathrm{Var}(C(z)). \tag{33}$$

Similarly, if C is any column, we obtain

$$\frac{N^4 - 1}{12} \leq 2\mathrm{Var}(C) + 2\sum_{z\in R}\mathrm{Var}_z(R(z)) - \frac{1}{N}\sum_{z\in R}\mathrm{Var}(R(z)), \tag{34}$$

where $R(z)$ of course stands for the row containing the number z.

We shall define *distance* of numbers (non-empty places) in an assignment. Consider a graph on the set of integers as vertices, where two integers are connected if they are in the same row or column in the arrangement. Then the distance of two numbers in the arrangement is defined to be the distance of the corresponding vertices in this graph.

Fix a non-empty place x of the assignment. Denote $B(k,x)$ the set of numbers which are located within distance k from x, and let $S(k,x) = B(k,x)\setminus B(k-1,x)$. By definition of $B(k,x)$, we have

$$\max\{y|y\in B(k,x)\} \leq x + kd$$

and

$$\min\{y|y\in B(k,x)\} \geq x - kd,$$

hence

$$|B(k,x)| \leq 2kd + 1.$$

Since

$$\cup_{i=0}^{k} S(i,x) = B(k,x),$$

and the sets $S(i,x)$ are disjoint from each other, there exist arbitrarily large integers k such that

$$|S(k,x)| \leq 2d. \tag{35}$$

It is possible that there is a k such that $B(k,x) = B(k+1,x)$. In this case for all $K > k$, we have $B(k,x) = B(K,x)$. We shall handle this case at the end of the proof. First suppose that the sets $B(k,x)$ are increasing.

Let k be a large integer with

$$0 < |S(k,x)| \leq 2d. \tag{36}$$

Consider the complete set of lines $\mathcal{L}$ containing an element of $B(k-1,x)$. Let us call the cardinality of this set t. Note that for "large" k, t is "large" also. For each of these rows (resp. columns), add up the statement of equation (33) (resp. equation (34)) for that row (resp. column).

Let us compare the left hand side and the right hand side of the obtained inequality. On the left hand side clearly we have $t(N^4-1)/12$. The right hand side looks rather complex. Let us consider the terms on the right hand side. First fix a row L which is entirely in $B(k-1,x)$, and count how many times $Var(L)$ occurs.

- It occurs with multiplicity 2 when (33) is added for L (the first term of the right hand side of (33));
- It occurs with multiplicity 4 when (34) is added for all the columns intersecting L (the second term of the right hand side of (34));
- It occurs -1 times when (34) is added up for the columns intersecting L (the third terms on the right hand side of (34)).

This gives that for rows L entirely contained in $B(k-1,x)$, $\mathrm{Var}(L)$ occurs 5 times. A similar argument applies to columns as well.

The number of rows and columns which are not entirely contained in $B(k-1,x)$, but intersecting it, is at most $4d$, because each such row or column must contain an element of $S(k,x)$, and any element of $S(k,x)$ can only be contained in one row and one column. These rows and columns contain at most $4dN$ numbers and each can be in one additional row or column. Hence we obtain that the number of rows and columns intersecting the ones intersecting (but not contained in) $B(k-1,x)$ is at most $4dN$. The variance of each of these rows and columns can occur on the right hand side at most 5 times. The variance of each row or column is at most V, hence we obtain:

$$t\frac{N^4-1}{12} \leq 5tV + (4d+4dN)V. \tag{37}$$

We can choose t to be arbitrarily large, but N is fixed, d and V are bounded as functions of N. Thus, dividing (37) by t gives the result $(N^4-1)/60 \leq V$.

Finally, consider the case when $B(k,x) = B(k+1,x)$. Similarly to the other case we add up (33) and (34) for all rows and columns intersecting $B(k-1,x)$. In this case, all these rows and columns are in fact contained in $B(k-1,x)$. On the left hand side of the sum we obtain $t(N^4-1)/12$; on the right hand side, for all the t rows and columns intersecting $B(k-1,x)$, the variance shows up exactly 5 times for the same reasons as before, and there are no partial sums. (Note that here we did not need to use the fact that t is large.)

This concludes the proof of the theorem. ∎

Acknowledgments

The authors thank Vinay Vaishampayan, Suhas Diggavi, Jim Reeds, Gábor Pete, Jeff Lagarias, David Applegate and Henry Landau for helpful discussions. The authors especially thank two anonymous referees and in particular to Gérard Battail (Associate Editor) for numerous helpful suggestions which have contributed greatly to the improvements in exposition in the final version.

References

[1] A.A. El Gamal and T.M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[2] C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 4, pp. 142–163, Mar. 1959.

[3] J.G. Proakis, *Digital Communications*, Macgraw–Hill, New York, 1983.

[4] V.A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 93, pp. 821–834, 1993.

[5] L.H. Harper T.Y. Berger-Wolf and M.A. Harris, "Bandwidth of clique products of cliques," preprint, 2002.

[6] T.Y. Berger-Wolf and E.M. Reingold, "Index assignment for multichannel communication under failure," *IEEE Trans. Inform. Theory*, vol. 48(10), pp. 2656–2668, Oct 2002.

[7] V.A. Vaishampayan, "Personal communication to the authors about previous research carried out by S. Diggavi, A. Orlitsky and V.A. Vaishampayan," 2001.

[8] J. Balogh and C. Smyth, "On the variance of Shannon product of graphs," *preprint*

[9] I. Ruzsa and G. Székely, *Algebraic Probability Theory*, Wiley, New York, 1988.

[10] T. Lewis, "The factorization of the rectangular distribution," *J. Appl. Probability*, vol. 4, pp. 529–542, 1967.

[11] A. Tortrat, "Sur un théorème de Lewis et la décomposition en facteurs premiers de la loi rectangulaire," *J. Appl. Probability*, vol. 6, pp. 177–185, 1969.