An integer programming algorithm for constructing maximin distance designs from good lattice point sets

Alan R. Vazquez and Hongquan Xu

University of Arkansas and University of California, Los Angeles

Supplementary Sections

S1 Proofs

This section contains the proofs of selected theoretical results in the main text. We omit the proofs of Lemma 1 and Theorems 1 and 2 because they are straightforward or discussed in the main text. We therefore only provide the proofs of Lemmas 2 and 3, and Theorem 3 and 4. To this end, we need the auxiliary lemmas below.

Lemma S1. *Let N be even and $y = 0, 1, \ldots, N-1$. If $y' = N/2 + y \pmod{N}$, then $W(y) + W(y') = N - 1$.*

Proof. We note that, if $0 \leq y < N/2$, then $N/2 \leq y' < N$. Moreover, if $N/2 \leq y < N$, then $0 \leq y' < N/2$. In both cases, $W(y) + W(y') = N - 1$. $\square$

Lemma S2. *Let N be odd and $y = 0, 1, \ldots, N-1$. If $y' = (N-1)/2 - y \pmod N$, then $W(y) + W(y') = N - 1$.*

Proof. We note that, if $0 \leq y < N/2$, then $0 \leq y' < N/2$. Moreover, if $N/2 \leq y < N$, then $N/2 \leq y' < N$. In both cases, $W(y) + W(y') = N - 1$. □

PROOF OF LEMMA 2. For the GLP set $\mathbf{X} = (x_{i,j})$, we consider $\mathbf{X}_b = \mathbf{X} + b \pmod N$ with elements $x_{i,j}^{(b)} = x_{i,j} + b \pmod N$. We have that $\mathbf{X}_{N/2+b} = \mathbf{X} + b + N/2 \pmod N = \mathbf{X}_b + N/2 \pmod N$, for $b = 0, 1, \ldots, N/2 - 1$. So, the elements of $\mathbf{X}_{N/2+b}$ are $x_{i,j}^{(N/2+b)} = N/2 + x_{i,j}^{(b)} \pmod N$ where $i = 1, \ldots, N$, and $j = 1, \ldots, n$. Since N is even, Lemma S1 implies that

$$W\left(x_{i,j}^{(b)}\right) + W\left(x_{i,j}^{(N/2+b)}\right) = W\left(x_{i,j}^{(b)}\right) + W\left(N/2 + x_{i,j}^{(b)}\right) = N - 1. \quad \text{(S1.1)}$$

Let $\mathbf{Z}_b = W(\mathbf{X}_b)$ and $\mathbf{Z}_{N/2+b} = W(\mathbf{X}_{N/2+b})$ whose j-th columns are denoted as $\mathbf{z}_b^{(j)}$ and $\mathbf{z}_{N/2+b}^{(j)}$, respectively. From Equation (S1.1), we therefore have that $\mathbf{z}_b^{(j)} + \mathbf{z}_{N/2+b}^{(j)} = (N-1)\mathbf{1}_N$ for $b = 0, 1, \ldots, N/2 - 1$. □

PROOF OF LEMMA 3. We first prove Lemma 3(ii) and then Lemma 3(i).

PART (ii). We divide the proof into three steps. Step 1 characterizes the columns of a GLP set with an odd number of runs. Step 2 characterizes the columns in the matrix resulting from applying specific linear permutations and the Williams' transformation to this GLP set. Step 3 proves the main result.

Step 1. For N odd and $n = \phi(N)$, we consider the set $H = \{h_1, \ldots, h_n\}$ of positive integers that are smaller than and coprime to N, where $h_1 < h_2 < \cdots < h_n$. The elements of H are related to each other since $h_{n+1-k} = N - h_k$ for $k = 1, \ldots, n$. This is because, if h_k is coprime to N, then $N - h_k$ is also coprime to N. The GLP set $\mathbf{X}$ has the elements $x_{i,j} = ih_j \pmod N$ where $i = 1, \ldots, N$, and $j = 1, \ldots, n$. We therefore have that $x_{i,n+1-j} = N - x_{i,j}$; see also Lemma 2 of Yuan et al. (2017).

Step 2. Let $b' = (N-1)/2 - b \pmod N$. We will prove that

$$W\left(x_{i,j} + b \pmod N\right) + W\left(x_{i,n+1-j} + b' \pmod N\right) = N - 1, \qquad \text{(S1.2)}$$

which is equivalent to showing that

$$W\left(x_{i,j} + b \pmod N\right) + W\left(N - x_{i,j} + b' \pmod N\right) = N - 1. \qquad \text{(S1.3)}$$

First, we denote $y_{i,j} = x_{i,j} + b \pmod N$. Next, we have that

$$\begin{aligned} N - x_{i,j} + b' \pmod N &= N - x_{i,j} + \frac{(N-1)}{2} - b \pmod N \\ &= \frac{3N-1}{2} - y_{i,j} \pmod N = \frac{N-1}{2} - y_{i,j} \pmod N. \end{aligned}$$

Since N is odd, Lemma S2 implies that

$$W\left(y_{i,j}\right) + W\left(\frac{N-1}{2} - y_{i,j}\right) = N - 1,$$

for $j = 1, \ldots, n$, and $i = 1, \ldots, N$. Therefore, Equation (S1.3) holds and so does Equation (S1.2).

Step 3. Let $\mathbf{Z}_b = W(\mathbf{X}_b)$ and $\mathbf{Z}_{b'} = W(\mathbf{X}_{b'})$ whose j-th columns are denoted as $\mathbf{z}_b^{(j)}$ and $\mathbf{z}_{b'}^{(j)}$, respectively. From Equation (S1.2), we finally have that $\mathbf{z}_b^{(j)} + \mathbf{z}_{b'}^{(n+1-j)} = (N-1)\mathbf{1}_N$ with $b = 0, \ldots, N-1$, and $b' = (N-1)/2 - b \pmod N$.

PART (i). This is just a special case of Lemma 3(ii). Recall that $b \in \mathbb{Z}_N$. When N is odd, $\mathbb{Z}_N$ can be divided into two disjoint subsets: $G_1 = \mathbb{Z}_{(N+1)/2}$ and $G_2 = \mathbb{Z}_N \setminus \mathbb{Z}_{(N+1)/2}$. If $(N-1)/2$ is even, G_1 and G_2 have an odd and even cardinality, respectively. The mid-element of G_1 is $b^\star = (N-1)/4$. This implies that $b = b' = b^\star$ is a solution to $b' = (N-1)/2 - b \pmod N$. Consequently, for the linear permutation given by $b^\star$, the entries in $\mathbf{Z}_{b^\star} = W(\mathbf{X}_{b^\star})$ satisfy

$$W\left(x_{i,j} + b^\star \pmod N\right) + W\left(x_{i,n+1-j} + b^\star \pmod N\right) = N-1. \tag{S1.4}$$

Therefore, $\mathbf{z}_{b^\star}^{(j)} + \mathbf{z}_{b^\star}^{(n+1-j)} = (N-1)\mathbf{1}_N$ for $j = 1, \ldots, n$.

If $(N-1)/2$ is odd, then G_1 and G_2 have an even and odd cardinality, respectively. The mid-element of G_2 is $b^\star = (3N-1)/4$ (assuming $N > 3$ as we do here). This means that $b = b' = b^\star$ is a solution to $b' = (N-1)/2 - b \pmod N$ and so, the entries in $\mathbf{Z}_{b^\star}$ satisfy Equation (S1.4) too. As a result, $\mathbf{z}_{b^\star}^{(j)} + \mathbf{z}_{b^\star}^{(n+1-j)} = (N-1)\mathbf{1}_N$ for this $b^\star$. $\square$

PROOF OF THEOREM 3. For the GLP set $\mathbf{X} = (x_{i,j})$, we consider $\mathbf{Z}_b =$

$W(\mathbf{X}_b)$ with $\mathbf{X}_b = \mathbf{X} + b \pmod N$ and $b = 0, 1, \ldots, N-1$. We denote the L_1-distance between the u-th and v-th row of $\mathbf{Z}_b$ as $d_{u,v}(\mathbf{Z}_b)$. Since N is an odd prime, Theorem 1 of Wang et al. (2018) states that

$$d_{u,v}(\mathbf{Z}_b) = \begin{cases} (N^2-1)/3 + f(b) & \text{for } u = N \text{ or } v = N; \\ (N^2-1)/3 - 2f(b) & \text{for } u = N - v; \\ (N^2-1)/3 & \text{otherwise,} \end{cases} \tag{S1.5}$$

where $f(b) = (W(b) - (N-1)/2)^2 - (N^2-1)/12$ and $u \neq v$. In other words, if $\mathbf{A}^{(b)}$ is the distance matrix of $\mathbf{Z}_b$, the elements in $\mathbf{A}^{(b)}\mathbf{1}_{N-1}$ can take three different values.

We now consider the initial candidate set $\mathbf{C}^* = [\mathbf{Z}_0, \mathbf{Z}_1, \cdots, \mathbf{Z}_{N-1}]$ whose distance matrix is $\mathbf{B} = [\mathbf{A}^{(0)}, \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N-1)}]$. The distances between two rows in $\mathbf{C}^*$ are contained within the vector $\mathbf{B}\mathbf{1}_s = \mathbf{A}^{(0)}\mathbf{1}_{N-1} + \mathbf{A}^{(1)}\mathbf{1}_{N-1} + \cdots + \mathbf{A}^{(N-1)}\mathbf{1}_{N-1}$, where $s = N(N-1)$. From Equation (S1.5), we have that $d_{u,v}(\mathbf{C}^*) = N(N^2-1)/3$ for $u = 1, \ldots, N-1$, and $v = u+1, \ldots, N$. This is because

$$\sum_{b=0}^{N-1} f(b) = \sum_{b=0}^{N-1} [(b - (N-1)/2)^2 - (N^2-1)/12] = 0,$$

since $W(b)$ is a permutation of the elements in $\mathbb{Z}_N$.

Lemma 3 implies that $\mathbf{C}^* = [\mathbf{C}, (N-1)\mathbf{J} - \mathbf{C}]$ where $\mathbf{C}$ is the $N \times s/2$ reduced candidate set in Equation (4.2) or (4.3) in the main text, and $\mathbf{J}$

is an $N \times s/2$ matrix of ones. From Lemma 1, we therefore have that $d_{u,v}(\mathbf{C}) = N(N^2-1)/6$ for $u = 1, \ldots, N-1$, and $v = u+1, \ldots, N$. In other words, $\mathbf{C}$ is an L_1-equidistant design and so it is a maximin distance design. $\square$

PROOF OF THEOREM 4. To save space, we sketch the three main steps only. Step 1 characterizes the rows of a GLP set with an even number of runs. Step 2 characterizes the rows in the matrix resulting from applying linear permutations and the Williams' transformation to this GLP set. Step 3 shows the pairs of rows in this matrix that have the same distance.

Step 1. For N even and $n = \phi(N)$, we consider the set $H = \{h_1, \ldots, h_n\}$ of positive integers that are smaller than and coprime to N, where $h_1 < h_2 < \cdots < h_n$. The elements of H must necessarily be odd. Moreover, $h_{n+1-k} = N - h_k$ for $k = 1, \ldots, n$. The elements of the GLP set $\mathbf{X}$ are $x_{i,j} = ih_j \pmod N$ where $i = 1, \ldots, N$, and $j = 1, \ldots, n$. For $i = 1, \ldots, N/2$, we have that

$$x_{\frac{N}{2}+i,j} = \frac{N}{2} + x_{i,j} \pmod N,$$

since h_j is odd. As a result, $\mathbf{X}$ can be expressed as

$$\begin{pmatrix} \mathbf{C} \\ \mathbf{C} + \frac{N}{2} \pmod N \end{pmatrix},$$

where $\mathbf{C}$ is an $N/2\times n$ matrix with elements $c_{i,j} = x_{i,j}$, where $i = 1, \ldots, N/2$, and $j = 1, \ldots, n$.

Step 2. We consider $\mathbf{Z}_b = W(\mathbf{X}_b)$ with $\mathbf{X}_b = \mathbf{X} + b \pmod N$, $b = 0, \ldots, N/2-1$. Let $\mathbf{Y}_b = \mathbf{C} + b \pmod N$ with elements $y_{i,j}^{(b)}$. We have that

$$\mathbf{Z}_b = \begin{pmatrix} W(\mathbf{Y}_b) \\ \\ W(\mathbf{Y}_b + \frac{N}{2} \pmod N) \end{pmatrix}, \text{ for } b = 1, \ldots, N/2-1.$$

Since N is even, Lemma S1 implies that

$$W\left(y_{i,j}^{(b)}\right) + W\left(\frac{N}{2} + y_{i,j}^{(b)}\right) = N - 1,$$

for $i = 1, \ldots, N/2$, and $j = 1, \ldots, n$. Therefore,

$$W\left(\mathbf{Y}_b + \frac{N}{2} \pmod N\right) = (N-1)\mathbf{J} - W(\mathbf{Y}_b),$$

where $\mathbf{J}$ is the $N \times n$ matrix of ones.

Step 3. We now characterize the rows in the distance matrix of $\mathbf{Z}_b$, which we denote as $\mathbf{A}^{(b)}$. To this end, let $\mathbf{R}_b = W(\mathbf{Y}_b)$ with rows denoted by $\mathbf{r}_{u,b}$, $u = 1, \ldots, N/2$. So,

$$\mathbf{Z}_b = \begin{pmatrix} \mathbf{R}_b \\ \\ (N-1)\mathbf{J} - \mathbf{R}_b \end{pmatrix}. \tag{S1.6}$$

Each row in $\mathbf{A}^{(b)}$ equals the absolute difference between two different rows in $\mathbf{Z}_b$. From Equation (S1.6), it is easy to see that $|\mathbf{r}_{u,b} - \mathbf{r}_{v,b}| = |(N-1)\mathbf{1}_N - \mathbf{r}_{u,b} - ((N-1)\mathbf{1}_N - \mathbf{r}_{v,b})|$, where $(N-1)\mathbf{1}_N - \mathbf{r}_{u,b}$ is the u-th row of

$(N-1)\mathbf{J}-\mathbf{R}_b$, $u=1,\ldots,N/2-1$, and $v=u+1,\ldots,N/2$. Therefore, $\mathbf{A}^{(b)}$ has $\frac{1}{2}\left(\frac{N}{2}(\frac{N}{2}-1)\right)$ pairs of repeat rows. For the same values of u and v, we have that $|\mathbf{r}_{u,b}-((N-1)\mathbf{1}_N-\mathbf{r}_{v,b})| = |((N-1)\mathbf{1}_N-\mathbf{r}_{u,b})-\mathbf{r}_{v,b}|$, which involve distances between a row in $\mathbf{R}_b$ and a row in $(N-1)\mathbf{J}-\mathbf{R}_b$. So, there are also other $\frac{1}{2}\left(\frac{N}{2}(\frac{N}{2}-1)\right)$ pairs of repeat rows $\mathbf{A}^{(b)}$. This amounts to a total of $\frac{N}{2}(\frac{N}{2}-1)$ pairs of repeat rows in the distance matrix of $\mathbf{Z}_b$. Finally, we consider the distance matrix of the candidate set $\mathbf{C}$ in Equation (4.1) in the main text, which is $\mathbf{B}=[\mathbf{A}^{(0)},\mathbf{A}^{(1)},\cdots,\mathbf{A}^{(N/2-1)}]$. The main result then follows since the pairs of repeat rows remain in $\mathbf{B}$. $\square$

S2 Additional numerical comparisons

S2.1 Candidate sets generated at random

To obtain N-run LHDs using the IP problem formulation in Section 4.3, we may consider a candidate set comprising permutations of the elements in $\mathbb{Z}_N$ selected at random. Here, we explore the LHDs obtained from this set and compare them with those obtained from the candidate set in Section 4.2. To this end, we consider the first six design problems in Table 2. These design problems involve a candidate set with N runs and $N\phi(N)/2$ columns, with N ranging from seven to 14. Recall that the goal of these problems is to construct good N-run LHDs with $\phi(N)$ factors in terms of the L_1-distance.

For each value of N ranging from seven to 14, we generated 100 candidate sets each of which contains a total of $N\phi(N)/2$ randomly-chosen permutations of the elements in $\mathbb{Z}_N$. For each candidate set, we solved the IP problem formulation and reported the L_1-distance of the best LHD with $\phi(N)$ factors. Table S1 shows the distribution of the L_1-distances of the 100 LHDs generated in this way for each design problem. More specifically, the table shows the minimum and maximum of the distribution, as well as its 25th, 50th and 75th percentiles.

For N equal to 7, 8, 9, 10, 11, 12, 13 and 14, Table S1 shows that largest L_1-distances obtained from randomly-generated candidate sets are 14, 9, 16, 10, 35, 11, 49 and 21, respectively. These L_1-distances are smaller than those obtained from our candidate set in Section 4.2, which are shown in Table 2 in the main text. Therefore, our candidate set embeds more attractive LHDs in terms of the maximin distance criterion than comparable candidate sets generated at random.

S2.2 Randomly-chosen subsets of the candidate set

An alternative strategy to obtain LHDs from the candidate set in Section 4.2 is to draw a large random sample of subset of columns from this set,

Table S1: Distribution of L_1-distances of N-run k-factor LHDs constructed using randomly-generated candidate sets, with $k = \phi(N)$.

			Percentiles			
N	k	Min.	25	50	75	Max.
7	6	12	13	13	14	14
8	4	7	8	8	8	9
9	6	15	15	16	16	16
10	4	9	9	9	9	10
11	10	33	34	34	34	35
12	4	9	10	10	10	11
13	12	47	47	48	48	49
14	6	20	20	20	20	21

and select the best subset in terms of the L_1-distance. We compare this strategy with the use of the IP problem formulation for obtaining LHDs with N runs and $N-1$ factors, where $N = 7$, 11, 13 and 17. For these cases, the candidate set has 21, 55, 78 and 136 columns, respectively. For seven runs, the total number of 6-factor LHDs that can be generated is $21!/[(6!)(15!)] = 54,264$. For larger run sizes, this number exceeds 2.92×10^{10}.

For each value of N, Table S2 shows the distribution of the L_1-distances of 5,000 different LHDs with $N-1$ factors, obtained from randomly-chosen subsets of columns from the candidate set. The table shows the minimum

Table S2: Distribution of L_1-distances of N-run k-factor LHDs constructed using randomly-chosen subsets of columns from the candidate set, with $k = N - 1$.

			Percentiles			
N	k	Min.	25	50	75	Max.
7	6	7	10	10	11	14
11	10	14	24	26	27	35
13	12	22	33	36	38	46
17	16	38	58	62	65	78

and maximum of the distribution, as well as its 25th, 50th and 75th percentiles. For seven, 11, 13 and 17 runs, the largest L_1-distances are 14, 35, 46 and 78, respectively.

Table 2 in the main text shows the L_1-distance of LHDs constructed by the IP algorithm for the design problems in Table S2. The L_1-distance of the 7-, 11-, 13- and 17-run LHDs obtained from this algorithm equal 16, 39, 54 and 94, respectively; all of which are larger than the L_1-distances in Table S2. In fact, the Gurobi solver certified that these are the optimal L_1-distances for the candidate sets. The good computational performance of the solver for these problems is due to the state-of-the-art optimization methods that are implemented in it; see Section 4.1 in the main text.

We conclude that the IP algorithm with the Gurobi solver is the best way to search for the optimal subsets of columns of our candidate set.

References

Wang, L., Q. Xiao, and H. Xu (2018). Optimal maximin L_1-distance Latin hypercube designs based on good lattice point designs. *The Annals of Statistics 46*, 3741–3766.

Yuan, R., D. K. J. Lin, and M.-Q. Liu (2017). Nearly column-orthogonal designs based on leave-one-out good lattice point sets. *Journal of Statistical Planning and Inference 185*, 29–40.