Statistics and Applications {ISSN 2454-7395(online)}
Volume 19, No. 1, 2021 (New Series), pp 11-28

Unbalanced Two-symbol $E(s^2)$-Optimal Designs

Dursun A. Bulutoglu[1], Kashinath Chatterjee[2], Stelios D. Georgiou[3], Christos Koukouvinos[4], Kenneth J. Ryan[5] and Stella Stylianou[3]

[1]*Department of Mathematics and Statistics, Air Force Institute of Technology, WPAFB, OH 45433*

[2]*Department of Statistics, Visva Bharati University, Santiniketan, West Bengal, India*

[3]*Department of Mathematics, School of Science, RMIT University, Melbourne, VIC 3001, Australia*

[4]*Department of Mathematics, National Technical University of Athens, Zografou 15773, Athens, Greece*

[5]*Department of Statistics, West Virginia University, Morgantown, West Virginia 26506, USA*

Received: 27 April 2020; Revised: 15 June 2020; Accepted: 20 June 2020

Abstract

Two-symbol supersaturated designs (SSDs) are two-symbol arrays in which the number of rows is no larger than the number of columns. In this paper, a lower bound for the E(s^2) value of SSDs that are not necessarily balanced is derived. The sharpness of the newly derived lower bound is analyzed theoretically by using constructions of E(s^2)-optimal SSDs and computationally by using the NOA$_4$ algorithm in Ryan and Bulutoglu (2007). Applications of the newly derived E(s^2) lower bound to searching for D-optimal designs and equiangular lines are discussed.

Key words: Two-symbol unbalanced SSD; E(s^2)-optimal designs; D-optimal designs; Lower bound.

AMS Subject Classifications: 62M20, 62F15

1. Introduction

Two-symbol supersaturated designs (SSDs) are two symbol arrays with the following properties. The number of rows N does not exceed the number of columns m and no pair of columns are fully aliased, *i.e.*, there exists no pair of identical columns up to permuting the symbols within a column. A two-symbol array is *balanced* if each of the two symbols in a column appears the same number of times when the number of rows N is even or the absolute difference between the frequencies of the occurrences of the two symbols in each column is 1 when N is odd. A two symbol array that is not balanced is called *unbalanced*. Two-symbol SSDs are commonly coded with symbol set $\{0, 1\}$ or $\{\pm 1\}$ and are particularly useful in screening experimentation due to their row-size economy (Georgiou, 2014). It has long been assumed in the literature that SSDs should be balanced. However, unbalanced SSDs are of interest to practitioners who are willing to compromise on the balance property due to high costs. In particular, unbalanced SSDs are useful when restrictions embedded in the problem at hand makes it infeasible to use a balanced SSD. Such SSDs are also

Correponding Author: Dursun Bulutoglu
Email: dursun.bulutoglu@gmail.com

preferable in cases where certain symbols of some columns need to be examined but are expensive to set.

WLOG, assume that each column of the N-row array (SSD) with m columns each with symbols from $\{0, 1\}$ has 0 at most $\lfloor N/2 \rfloor$ times, where $\lfloor . \rfloor$ is the floor function. Define k_l to be the number of columns in which 0 appears l times for $l = 0, \ldots, \lfloor N/2 \rfloor$. Clearly, $\sum_l k_l = m$. Denote this class of arrays by $\mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$. Let $\mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ denote the corresponding class of arrays when -1 is used instead of 0. For even N, if we choose $k_l = 0$ for $0 \le l < \lfloor N/2 \rfloor$ and $k_{\lfloor N/2 \rfloor} = m$, we get a balanced array (SSD), and we get an unbalanced array if $k_l \neq 0$ for at least one l with $l < \lfloor N/2 \rfloor$. We call the vector $(k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor})$ the *balancedness structure* of each $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ or each $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$.

Example 1: Consider the 4-row SSD

$$\mathbf{D} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in \mathcal{D}(4, 2^4, (0, 1, 3)).$$

This SSD is unbalanced with $k_1 = 1$, $k_2 = 3$ and $\sum_l k_l = m = 4$.

For $m \ge N$, an N row, m column, two-symbol array $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]$ with entries from $\{\pm 1\}$ is a supersaturated design if it has no two columns $\mathbf{x}_i$ and $\mathbf{x}_j$ such that $i \neq j$ and $\mathbf{x}_i^{\top} \mathbf{x}_j \in \{-N, N\}$. The $\mathrm{E}(s^2)$ value of $\mathbf{X}$ is defined as

$$\mathrm{E}(s^2) = \frac{\sum_{i \neq j} s_{ij}^2}{m(m-1)},$$

where $s_{ij} = \mathbf{x}_i^{\top} \mathbf{x}_j$ for $1 \le i \neq j \le m$. A two-symbol SSD is mapped to a $\{\pm 1\}$ SSD by assigning $+1$ to one symbol and -1 to the other symbol in each column. We call a resulting $\{\pm 1\}$ SSD a corresponding $\{\pm 1\}$ SSD. Then each of the concepts defined for a $\{\pm 1\}$ SSD is defined for a two-symbol SSD via one of its corresponding $\{\pm 1\}$ SSDs. The $\mathrm{E}(s^2)$ value is used to compare two-symbol SSDs with the same number of rows and columns (Georgiou, 2014). An SSD with a smaller $\mathrm{E}(s^2)$ value is more desirable (Georgiou, 2014), and an SSD with the smallest possible $\mathrm{E}(s^2)$ value is called $\mathrm{E}(s^2)$-optimal. For a detailed review of the $\mathrm{E}(s^2)$ optimality criteria for two-symbol SSDs, the reader is referred to (Georgiou, 2014).

Ryan and Bulutoglu (2007) and Das *et. al.* (2008) gave the sharpest known lower bound for balanced SSDs with even N. Bulutoglu and Ryan (2008) and Suen and Das (2010) derived an improved $\mathrm{E}(s^2)$ lower bound for two-symbol SSDs with odd N. For unbalanced SSDs, the best known $\mathrm{E}(s^2)$ lower bounds are not applicable. We generalize the results in Bulutoglu and Ryan (2008) and Suen and Das (2010) to unbalanced SSDs. We derive a lower bound for the $\mathrm{E}(s^2)$ value of unbalanced two-symbol SSDs and present some families of $\mathrm{E}(s^2)$-optimal unbalanced SSDs. Part of our derivation is based on an adaptation of the derivation in Bulutoglu and Ryan (2008).

For an SSD $\mathbf{X} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ let $s_{\max} = \max_{i<j} |s_{ij}|$ and $f_{s_{\max}}$ be the frequency of $s_{\max}$ in $\{s_{ij}\}_{i<j}$. Then $\mathbf{X}$ is called *minimax-optimal* if no other SSD in $\mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ has a smaller $s_{\max}$ or smaller $f_{s_{\max}}$ at the smallest possible $s_{\max}$. For balanced SSDs, Ryan and Bulutoglu (2007) and Bulutoglu and Ryan (2008) used minimax optimality as a secondary

criterion for picking an SSD among E(s^2)-optimal SSDs. Finding a minimax-optimal and balanced SSD among E(s^2)-optimal and balanced SSDs is a very difficult problem as mentioned in Morales and Bulutoglu (2018). Some of the unbalanced infinite families of E(s^2)-optimal SSDs in Section 3 are also minimax-optimal.

There are no known theories or construction methods of SSDs for any choice of $k_l, 0 \leq l \leq \lfloor N/2 \rfloor$, because an achievable lower bound for E(s^2) is not known for the general case and it is not possible to prove E(s^2)-optimality without resorting to full enumeration. Therefore, there is a need to develop a general sharp lower bound that also covers unbalanced arrays. In this paper, we generalize the best known E(s^2) lower bound for balanced two-symbol SSDs to all two-symbol SSDs with a given balancedness structure. Additionally, we describe how our newly derived E(s^2) lower bound can be used to speed up search algorithms for finding two-symbol D-optimal designs in general.

An SSD can be thought as a frame, *i.e.*, a spanning set for its column space. Moreover, certain E(s^2)-optimal SSDs are tight frames (Morales and Bulutoglu, 2018). Another motivation for generalizing the best known E(s^2) lower bound for balanced two-symbol SSDs to all two-symbol SSDs with a given balancedness structure is that there is no balancedness requirement for frames. Furthermore, certain E(s^2)-optimal and minimax-optimal SSDs are equiangular tight frames and imply the existence of certain strongly regular graphs (Morales and Bulutoglu, 2018; Waldron, 2009).

This paper is organized as follows. In Section 2, we derive a previously unknown lower bound for the E(s^2) value of an unbalanced two-symbol SSD with symbols from $\{\pm 1\}$ given its column sums. After providing a naive E(s^2) lower bound, Section 3 theoretically analyzes the Section 2 bound in terms of its achievability and provides families of unbalanced E(s^2)-optimal SSDs achieving the Section 2 bound. Some of these SSDs are optimal with respect to the minimax criterion as well. Section 4 provides computational test results obtained by using the NOA$_4$ algorithm in Ryan and Bulutoglu, (2007) for the achievability of the Section 2 bound. Finally, in Section 5, we discuss two possible applications of our newly derived E(s^2) lower bound. In particular, in Section 5.1, we provide an application to searching for D-optimal designs. Moreover, in Section 5.2, for a given t such that $0 < t$, we discuss an application to finding upper bounds on the maximum number of columns for a two-symbol $\{\pm 1\}$ SSD with N rows whose each pairwise column angle is in $[\arccos(t/N), \arccos(-t/N)]$.

2. A General Lower Bound

In this section, we derive a previously unknown lower bound for the E(s^2) value of unbalanced SSDs. We first provide some definitions and lemmas that will be useful in proving the desired lower bound.

For a 2-symbol array let $c_{i_1 i_2}$ be the number of coincidences in the i_1'th and i_2'th rows for $1 \leq i_1 \neq i_2 \leq N$. The following lemma provides the number of coincidences in all the different pairs of rows in a given SSD.

Lemma 1: For $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ or $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$,

$$2 \sum_{i_1 \neq i_2} c_{i_1 i_2} = \sum_l (N - 2l)^2 k_l + mN(N - 2).$$

Proof: For given $k_l, 0 \le l \le \lfloor N/2 \rfloor$ and $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, we have

$$\mathbf{X}_0\mathbf{X}_0^\top = \begin{pmatrix} m & 2c_{12} - m & \cdots & 2c_{1N} - m \\ 2c_{12} - m & m & \cdots & 2c_{2N} - m \\ \vdots & \vdots & \vdots & \vdots \\ 2c_{1N} - m & 2c_{2N} - m & \cdots & m \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \mathbf{1}_N^\top \mathbf{X}_0\mathbf{X}_0^\top \mathbf{1}_N &= 2\sum_{i_1 \ne i_2} c_{i_1 i_2} - mN(N-2) \\ &= \sum_l (N-2l)^2 k_l. \end{aligned}$$

Let $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$. For $1 \le j \le m$, let n_α^j be the number of times $\alpha \in \{0, 1\}$ appears in the j'th column of $\mathbf{D}$. Also, for $1 \le j_1 \ne j_2 \le m$, $\alpha, \beta \in \{0, 1\}$, let $n_{\alpha\beta}^{j_1 j_2}$ be the number of times the symbol combination (α, β) appears as rows of the $N \times 2$ array obtained by concatenating j_1'th and j_2'th columns of $\mathbf{D}$. Then

$$s_{j_1 j_2}^2 = 4\sum_{\alpha,\beta} (n_{\alpha\beta}^{j_1 j_2})^2 - 2\sum_\alpha (n_\alpha^{j_1})^2 - 2\sum_\beta (n_\beta^{j_2})^2 + N^2. \tag{1}$$

Based on (1), we can express the $\mathrm{E}(s^2)$ value in a convenient form so as to obtain another lower bound.

Lemma 2: For $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$,

$$\mathrm{E}(s^2) = \frac{4\sum_{j_1 \ne j_2, \alpha, \beta} (n_{\alpha\beta}^{j_1 j_2})^2 - 4(m-1)\sum_l [l^2 + (N-l)^2] k_l + m(m-1)N^2}{m(m-1)}.$$

Proof: For $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, we have

$$\begin{aligned} m(m-1)\mathrm{E}(s^2) &= \sum_{j_1 \ne j_2} s_{j_1 j_2}^2 \\ &= \sum_{j_1 \ne j_2} \left[4\sum_{\alpha,\beta} (n_{\alpha\beta}^{j_1 j_2})^2 - 2\sum_\alpha (n_\alpha^{j_1})^2 - 2\sum_\beta (n_\beta^{j_2})^2 + N^2 \right] \\ &= 4\sum_{j_1 \ne j_2, \alpha, \beta} (n_{\alpha,\beta}^{j_1 j_2})^2 - 2(m-1)\left[\sum_{j_1,\alpha} (n_\alpha^{j_1})^2 + \sum_{j_2,\beta} (n_\beta^{j_2})^2 \right] + m(m-1)N^2 \\ &= 4\sum_{j_1 \ne j_2, \alpha, \beta} (n_{\alpha\beta}^{j_1 j_2})^2 - 4(m-1)\sum_l [l^2 + (N-l)^2] k_l + m(m-1)N^2. \end{aligned}$$

For a given $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, let l_{j_i} be the number of zeros in the j_i'th column of **D**. Then we have

$$\begin{aligned} n_{00}^{j_1 j_2} &= l_{j_2} - n_{10}^{j_1 j_2} \\ n_{01}^{j_1 j_2} &= l_{j_1} - l_{j_2} + n_{10}^{j_1 j_2} \\ n_{11}^{j_1 j_2} &= N - l_{j_1} - n_{10}^{j_1 j_2}. \end{aligned} \tag{2}$$

Let $\sum_{\alpha,\beta}(n_{\alpha\beta}^{j_1 j_2})^2 = f(l_{j_1}, l_{j_2}, n_{10}^{j_1 j_2})$. Then by equations (2),

$$f(l_{j_1}, l_{j_2}, n_{10}^{j_1 j_2}) = 4(n_{10}^{j_1 j_2})^2 + (4l_{j_1} - 4l_{j_2} - 2N)n_{10}^{j_1 j_2} + l_{j_2}^2 + (l_{j_1} - l_{j_2})^2 + (N - l_{j_1})^2.$$

For fixed l_{j_1}, l_{j_2}, $f(l_{j_1}, l_{j_2}, n_{10}^{j_1 j_2})$ is a convex function of $(n_{10}^{j_1 j_2})$. By differentiating $f(l_{j_1}, l_{j_2}, n_{10}^{j_1 j_2})$ with respect to $n_{10}^{j_1 j_2}$ we see that $f(l_{j_1}, l_{j_2}, n_{10}^{j_1 j_2})$ is minimized at

$$\widehat{n}_{10}^{j_1 j_2}(l_{j_1}, l_{j_2}) = \left[\frac{N + 2l_{j_2} - 2l_{j_1}}{4}\right],$$

where $[x]$ is the integer closest to x. Define

$$\widehat{n}_{10}(i, j) = \left[\frac{N + 2j - 2i}{4}\right].$$

Then we have

$$\begin{aligned} \sum_{j_1 \neq j_2, \alpha, \beta} (n_{\alpha,\beta}^{j_1 j_2})^2 &\geq \sum_i k_i(k_i - 1) f(i, i, \widehat{n}_{10}(i, i)) + \sum_{i \neq j} k_i k_j f(i, j, \widehat{n}_{10}(i, j)) \\ &= \theta_1^* = \sum_i k_i(k_i - 1) f\left(i, i, \left[\frac{N}{4}\right]\right) + \sum_{i \neq j} k_i k_j f\left(i, j, \left[\frac{N + 2j - 2i}{4}\right]\right). \end{aligned} \tag{3}$$

The following lemma, whose proof follows from Lemma 2 and inequality (3), provides a lower bound for the $\mathrm{E}(s^2)$ value of $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$.

Lemma 3: For $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$,

$$\mathrm{E}(s^2) \geq \mathrm{LB}_1,$$

where

$$\mathrm{LB}_1 = \frac{4\theta_1^* - 4(m-1)\sum_l [l^2 + (N-l)^2] k_l + m(m-1)N^2}{m(m-1)}.$$

By routine algebra we get the following result.

Lemma 4: For balanced and two-symbol arrays (SSDs)

$$\mathrm{LB}_1 = \mathrm{LB}_1(N, (0, \ldots, 0, m)) = \begin{cases} 1, & \text{if } N \text{ is odd}, \\ 0, & \text{if } N = 0 \pmod 4, \\ 4, & \text{if } N = 2 \pmod 4. \end{cases}$$

When $N = 2 \pmod 4$, the bound in Lemma 4 for $m = N + 1$ is achievable if a skew-symmetric Hadamard matrix of order $N+2$ exists (Morales *et al.* 2019). Such a Hadamard matrix is conjectured to exist for each N divisible by 4 (Koukouvinos and Stylianou, 2008). It is plain to verify the following remark.

Remark 1: For balanced SSDs, $m \geq N + 3$ or $N = 0 \pmod 4$, $\mathrm{LB}_1(N, (0, \ldots, 0, m))$ is strictly smaller than the $\mathrm{E}(s^2)$ lower bound in Das *et al.* (2008) and Ryan and Bulutoglu (2007). For balanced SSDs, $N = 2 \pmod 4$ and $m \leq N + 2$, $\mathrm{LB}_1(N, (0, \ldots, 0, m))$ is the same as the $\mathrm{E}(s^2)$ lower bound in Das *et al.* (2008) and Ryan and Bulutoglu (2007). For balanced SSDs and odd N, $\mathrm{LB}_1(N, (0, \ldots, 0, m))$ cannot be sharper than the $\mathrm{E}(s^2)$ lower bound in Bulutoglu and Ryan (2008). In particular, when $N = 3 \pmod 4$ and $m = N$, $\mathrm{LB}_1(N, (0, \ldots, 0, m))$ equals to the $\mathrm{E}(s^2)$ lower bound in Bulutoglu and Ryan (2008). A numerical check suggests that these are the only odd N cases for which equality is satisfied.

For $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, let $\Delta = \sum_l (N - 2l)^2 k_l$ and

$$F(p) := 8p^2 + 4N^2 - 8Np - 4N + 4\max\{|-mN + \Delta + qN(N-1)| - 4p^2 - 2N^2 + 4Np + 2N, 0\}.$$

Then

$$F(p) = \begin{cases} F_1(p) & \text{if } |-mN + \Delta + qN(N-1)| - 4p^2 - 2N^2 + 4Np + 2N \leq 0, \\ F_2(p) & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} F_1(p) &= 8p^2 + 4N^2 - 8Np - 4N \\ F_2(p) &= -8p^2 - 4N^2 + 8Np + 4N + 4|-mN + \Delta + qN(N-1)|. \end{aligned}$$

The following theorem provides another $\mathrm{E}(s^2)$ lower bound for $\mathbf{D}$.

Theorem 1: There is a unique q such that $-2N \leq (Nm - \Delta)/(N-1) - qN < 2N$ and $m + q \equiv 2 \ (mod 4)$. Let

$$p_-^* = \frac{N - \sqrt{2N - N^2 + |-mN + \Delta + qN(N-1)|}}{2},$$

$$p_+^* = \frac{N + \sqrt{2N - N^2 + |-mN + \Delta + qN(N-1)|}}{2},$$

$g(q) := (m+q)^2 N - q^2 N^2 - mN^2 - 2q\Delta$, and $K = \sum_i k_{2i+1}$. For odd N, let

$$\mathrm{LB}_2 = \begin{cases} \dfrac{16\left\lceil \frac{g(q) + F_1\left(\left\lfloor \frac{N}{2} \right\rfloor\right) - m(m-1)}{16} \right\rceil^+ + m(m-1)}{m(m-1)}, & \text{if } \left|\frac{Nm - \Delta}{N-1} - qN\right| < N, \\ \dfrac{16\left\lceil \frac{g(q) + \min\{F_1(\lfloor p_-^* \rfloor), F_2(\lceil p_-^* \rceil)\} - m(m-1)}{16} \right\rceil^+ + m(m-1)}{m(m-1)}, & \text{otherwise,} \end{cases}$$

where $\lceil . \rceil$ is the ceiling function. For $N \equiv 0 \pmod 4$, let

$$\mathrm{LB}_2 = \begin{cases} \dfrac{32\left\lceil \frac{g(q)+F_1\left(\lfloor \frac{N}{2} \rfloor\right)-8K(m-K)}{32} \right\rceil^+ + 8K(m-K)}{m(m-1)}, & \text{if } \left|\frac{Nm-\Delta}{N-1} - qN\right| < N, \\ \dfrac{32\left\lceil \frac{g(q)+\min\{F_1(\lfloor p_-^* \rfloor), F_2(\lceil p_-^* \rceil)\}-8K(m-K)}{32} \right\rceil^+ + 8K(m-K)}{m(m-1)}, & \text{otherwise.} \end{cases}$$

For $N \equiv 2 \pmod 4$, let

$$\mathrm{LB}_2 = \begin{cases} \dfrac{\theta\left\lceil \frac{g(q)+F_1\left(\lfloor \frac{N}{2} \rfloor\right)-4K(K-1)-4(m-K)(m-K-1)}{\theta} \right\rceil^+ + 4K(K-1)+4(m-K)(m-K-1)}{m(m-1)}, & \text{if } \left|\frac{Nm-\Delta}{N-1} - qN\right| < N, \\ \dfrac{\theta\left\lceil \frac{g(q)+\min\{F_1(\lfloor p_-^* \rfloor), F_2(\lceil p_-^* \rceil)\}-4K(K-1)-4(m-K)(m-K-1)}{\theta} \right\rceil^+ + 4K(K-1)+4(m-K)(m-K-1)}{m(m-1)}, & \text{otherwise,} \end{cases}$$

where

$$\theta = \begin{cases} 64 & \text{if } K = 0, \\ 32 & \text{otherwise,} \end{cases}$$

and $\lceil x \rceil^+ = \max\{0, \lceil x \rceil\}$. Then

$$\mathrm{E}(s^2) \geq \mathrm{LB}_2 \,.$$

Proof: The proof is an adaptation of the proof of Theorem 1 in Bulutoglu and Ryan (2008). For general $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \dots, k_{\lfloor N/2 \rfloor}))$, the $F(p)$ in Bulutoglu and Ryan (2008) becomes

$$F(p) := 8p^2 + 4N^2 - 8Np - 4N + 4\max\{|-mN + \Delta + qN(N-1)| - 4p^2 - 2N^2 + 4Np + 2N, 0\},$$

so that $F(p)$ is continuous. Moreover,

$$F'(p) = \begin{cases} 16p - 8N, & \text{if } |-mN + \Delta + qN(N-1)| - 4p^2 - 2N^2 + 4Np + 2N < 0, \\ -16p + 8N, & \text{if } |-mN + \Delta + qN(N-1)| - 4p^2 - 2N^2 + 4Np + 2N > 0. \end{cases}$$

Hence, $F(p)$ has all of its local minima at p such that

$$-4p^2 - 2N^2 + 4Np + 2N + |-mN + \Delta + qN(N-1)| = 0,$$

and these p's are p_-^* and p_+^*. By the reflection symmetry of $F_1(p)$ and $F_2(p)$ along the axis $y = F_1(p_-^*)$, both of these local minima are in fact global minima, and $F(p)$ for $p \in \{0, 1, \dots, (N+1)/2\}$ is minimized at $p = \lfloor p_-^* \rfloor$ or $p = \lceil p_-^* \rceil$. The result now follows from

$$s_{ij} \equiv \begin{cases} N \pmod 4, & \text{if the } i\text{'th and } j\text{'th columns of } \mathbf{D} \text{ have both even} \\ & \text{or odd number of } -1\text{'s,} \\ (N+2) \pmod 4, & \text{otherwise.} \end{cases}$$

Remark 2: LB_2 is cheaper to compute compared to the corresponding lower bound in Bulutoglu and Ryan (2008). This is because computing LB_2 requires computing only $F_1(\lfloor p_-^* \rfloor)$ and $F_2(\lceil p_-^* \rceil)$ instead of computing $F(p)$ for $p = 1, 2, \ldots, (N+1)/2$.

By Lemma 3, we obtain the following theorem.

Theorem 2: For $\mathbf{D} \in \mathcal{D}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$,

$$\mathrm{E}(s^2) \geq \mathrm{LB} = \mathrm{LB}(N, (k_0, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) = \max\{\mathrm{LB}_1, \mathrm{LB}_2\}. \tag{4}$$

For balanced SSDs with an odd number of rows, it is easy to show that LB_2 reduces to the $\mathrm{E}(s^2)$ lower bound provided in Bulutoglu and Ryan (2008). (We numerically verified this for $7 \leq N \leq 41$ and $N \leq m \leq 4N$.) For balanced SSDs with an even number of rows, LB is still a valid lower bound; however, it cannot be sharper than the $\mathrm{E}(s^2)$ lower bound of Ryan and Bulutoglu (2007) or of Das *et al.* (2008). This is because the assumed set of constraints for the hypothetical SSD in the derivation of LB_2 are satisfied by the hypothetical SSD in the derivation of the lower bound Ryan and Bulutoglu (2007) and that in Das *et al.* (2008). In fact, we observed in 700 of the 1,314 even N cases with $8 \leq N \leq 40$ and $N \leq m \leq 4N$ that the $\mathrm{E}(s^2)$ lower bound of Ryan and Bulutoglu (2007) or of Das *et al.* (2008) is sharper than the lower bound in Bulutoglu and Ryan (2008). Hence, for balanced and even N SSDs, LB should not be used. Numerically comparing LB_2 to the bound in Ryan and Bulutoglu (2007) and that in Das *et al.* (2008) provided us with a check for the correctness of LB_2. We implemented a similar check for the correctness of LB_1. We observed that LB_1 is most useful when m is close to N and is dominated by LB_2 as m increases.

3. Theoretical Achievability

In this section, we first derive a naive $\mathrm{E}(s^2)$ lower bound for an N row, m column, two-symbol SSD. This bound does not depend on the column sums of the SSD. Then we show that if this naive $\mathrm{E}(s^2)$ lower bound is achievable then LB_2 is also achievable.

Let $\mathbf{X}$ be an N row, m column SSD whose column symbols are from $\{\pm 1\}$. Then the off-diagonal entries of $\mathbf{XX}^\top$ are odd if and only if m is odd. When m is divisible by 4, it is possible for $\mathbf{X}$ to have mutually orthogonal rows. When $m = 2 \pmod 4$, $\mathbf{X}$ can have at most $\lfloor N/2 \rfloor \lceil N/2 \rceil$ pairs of orthogonal rows. From these facts, we immediately get the following naive $\mathrm{E}(s^2)$ lower bound

$$\mathrm{E}(s^2) \geq \begin{cases} \frac{N(N-1) - mN^2 + Nm^2}{m(m-1)}, & \text{if } m \text{ is odd}, \\ \frac{-mN^2 + Nm^2}{m(m-1)}, & \text{if } m = 0 \pmod 4), \\ \frac{4(\lfloor \frac{N}{2} \rfloor(\lfloor \frac{N}{2} \rfloor - 1) + \lceil \frac{N}{2} \rceil(\lceil \frac{N}{2} \rceil - 1)) - mN^2 + Nm^2}{m(m-1)}, & \text{if } m = 2 \pmod 4). \end{cases} \tag{5}$$

The lower bound $\mathrm{LB}_2(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ is based on a derivation where $\mathbf{XX}^\top$ has off-diagonals from the set $\{-4, -2, 0, 2\}$ or the set $\{-2, 0, 2, 4\}$ for even $\sum_i k_i$ and from the set $\{-3, -1, 1, 3\}$ for odd $\sum_i k_i$, where the off-diagonal entries sum to $\sum_{l=0}^n (N - 2l)^2 k_l - Nm$. For an SSD $\mathbf{X}$ achieving the naive $\mathrm{E}(s^2)$ lower bound (5), $\mathbf{XX}^\top$ has off-diagonals from the set $\{-2, 0, 2\}$ for even $\sum_i k_i$ and from the set $\{\pm 1\}$ for odd $\sum_i k_i$. Moreover, the entries of such an $\mathbf{XX}^\top$ sum to $\sum_{l=0}^n (N - 2l)^2 k_l$. Since $\{\pm 1\} \subseteq \{-3, -1, 1, 3\}$ and $\{-2, 0, 2\} \subseteq \{-2, 0, 2, 4\} \cap \{-4, -2, 0, 2\}$, we conclude that $\mathrm{LB}_2(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ is always at least as sharp as the naive $\mathrm{E}(s^2)$ lower bound (5). Hence, an SSD $\mathbf{X}$ achieving the naive $\mathrm{E}(s^2)$ lower bound (5) also achieves $\mathrm{LB}_2(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, i.e., if the naive $\mathrm{E}(s^2)$ lower bound (5) is achievable, then $\mathrm{LB}_2(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ is equal to the naive $\mathrm{E}(s^2)$ lower bound (5).

There are cases in which LB_2 is strictly larger than the naive $\mathrm{E}(s^2)$ lower bound (5). In particular, for balanced SSDs when N is odd LB_2 reduces to the $\mathrm{E}(s^2)$ lower bound derived in Bulutoglu and Ryan (2008), and there are SSDs ($\mathbf{X}$'s) achieving this lower bound such that the off-diagonal entries of $\mathbf{X}\mathbf{X}^\top$ are not all from the set $\{\pm 1\}$.

For each N, m combination, an $\mathrm{E}(s^2)$-optimal SSD achieving the naive bound (5) can be constructed by using Hadamard matrices. A $t \times t$ matrix $\mathbf{H}_t$ of ± 1's is called a Hadamard matrix if $\mathbf{H}_t^\top \mathbf{H}_t = t\mathbf{I}_t$, where $\mathbf{I}_t$ is the $t \times t$ identity matrix. It is well-known that t must be divisible by 4 for a $t \times t$ Hadamard matrix to exist. It is conjectured that $t \times t$ Hadamard matrix exists whenever t is divisible by 4. Let $\mathbf{1}_r$ be the column of all $1s$ of length r. It is easy to show that any Hadamard matrix can be put in the form

$$\mathbf{H}_t = \left(\begin{array}{cc|cc} 1 & 1 & \mathbf{1}^\top_{\frac{t}{2}-1} & \mathbf{1}^\top_{\frac{t}{2}-1} \\ 1 & -1 & \mathbf{1}^\top_{\frac{t}{2}-1} & -\mathbf{1}^\top_{\frac{t}{2}-1} \\ \hline \mathbf{1}_{\frac{t}{2}-1} & \mathbf{1}_{\frac{t}{2}-1} & \multicolumn{2}{c}{\multirow{2}{*}{\mathbf{A}}} \\ \mathbf{1}_{\frac{t}{2}-1} & -\mathbf{1}_{\frac{t}{2}-1} & & \end{array}\right) \tag{6}$$

by applying signed column and/or row permutations (by right and/or left multiplying with permutation matrices that are right or left multiplied by ± 1 diagonal matrices).

Let $\mathbf{X}$ be an N row, m column $\mathrm{E}(s^2)$-optimal SSD achieving the naive $\mathrm{E}(s^2)$ lower bound (5). In what follows, we describe how $\mathbf{X}$ can be constructed provided that the Hadamard conjecture is true. If m is divisible by 4, then $\mathbf{X}$ can be taken to be any N rows of a Hadamard matrix $\mathbf{H}_m$. If $m = 1 \pmod 4$ ($m = 3 \pmod 4$), then $\mathbf{X}$ can be constructed by first adding (deleting) any column with entries in $\{-1, 1\}$ to (from) a Hadamard matrix $\mathbf{H}_{m-1}$ ($\mathbf{H}_{m+1}$) followed by picking any N rows from the resulting matrix. If $m = 2 \pmod 4$, let $\mathbf{H}_{m+2}$ be a Hadamard matrix. Let $\mathbf{A}$ be obtained from $\mathbf{H}_{m+2}$ after putting $\mathbf{H}_{m+2}$ in form (6). Then any N rows of $\mathbf{A}$ can be taken to be $\mathbf{X}$.

Two arrays $\mathbf{D}_1$ and $\mathbf{D}_2$ are *equivalent* if $\Pi_1 \mathbf{D}_1 \Pi_2 = \mathbf{D}_2$ for some signed permutation matrices Π_1 and Π_2 (i.e., permutation matrices that are right or left multiplied by ± 1 diagonal matrices). If $\mathbf{X}_1$ and $\mathbf{X}_2$ are equivalent SSDs, then $\mathrm{SS}(\mathbf{X}_1\mathbf{X}_1^\top) = \mathrm{SS}(\mathbf{X}_2\mathbf{X}_2^\top)$, where $\mathrm{SS}(\mathbf{M})$ is the sum squares of the entries of a matrix $\mathbf{M}$. Hence, if $\mathbf{X}_1$ is an $\mathrm{E}(s^2)$-optimal SSD achieving the naive $\mathrm{E}(s^2)$ lower bound (5), then any other SSD equivalent to $\mathbf{X}_1$ is also $\mathrm{E}(s^2)$-optimal and achieves the naive $\mathrm{E}(s^2)$ lower bound (5).

Not every $N, m, k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}$ combination allows a $\mathbf{D} \in \mathcal{D}^\pm(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ that achieves the naive $\mathrm{E}(s^2)$ lower bound (5). If $\mathbf{D} \in \mathcal{D}^\pm(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ achieves the naive $\mathrm{E}(s^2)$ lower bound (5), then the fact that $\sum_{l=0}^{n} (N - 2l)^2 k_l$ must be equal to the sum of the entries of $\mathbf{D}\mathbf{D}^\top$ implies the constraint

$$\sum_{l=0}^{n}(N-2l)^2k_l = \begin{cases} Nm+2\alpha_1-2\beta_1, & \text{if } m \text{ is odd,} \\ Nm, & \text{if } m=0 \text{ (mod 4),} \\ Nm+4\alpha_2-4\beta_2, & \text{if } m=2 \text{ (mod 4)} \end{cases} \tag{7}$$

on $N, m, k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}$ for some non-negative integers $\alpha_1, \beta_1, \alpha_2, \beta_2$. The next theorem follows from the derivation of $\mathrm{LB}_2(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$.

Theorem 3: The naive $\mathrm{E}(s^2)$ lower bound (5) is equal to $\mathrm{LB}_2(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ if and only if the constraint in equation (7) is satisfied.

Circulant matrices can be used to construct unbalanced, $\mathrm{E}(s^2)$-optimal, minimax-optimal SSDs and D-optimal designs. A matrix is called *circulant* if each row vector is shifted one element to the right relative to the preceding row vector. A circulant matrix $\mathbf{A} = circ(\mathbf{a})$ is determined by its first row $\mathbf{a}$. Each row of $\mathbf{A}$ is a cyclic shift of the vector $\mathbf{a}$ to the right.

The dual Gram matrix and the Gram matrix of a matrix $\mathbf{A}$ are defined to be $\mathbf{A}\mathbf{A}^\top$ and $\mathbf{A}^\top\mathbf{A}$. If $\mathbf{x} = (x_0, x_1, \ldots, x_{t-1})$ is a vector of length t, the *periodic autocorrelation function* $P_{\mathbf{x}}(s)$ (abbreviated as PAF) is defined, reducing $i+s$ modulo t, as

$$P_{\mathbf{x}}(s) = \sum_i x_i x_{i+s} \quad \text{for } s = 0, 1, \ldots, t-1.$$

The (dual) Gram matrix of a circulant matrix is also circulant and can be calculated by using the periodic autocorrelation function of its first row.

Let t be odd and $\mathbf{A}, \mathbf{B}$ be $t \times t$ circulant matrices with entries in $\{\pm 1\}$. Let $\mathbf{a} = (a_0, a_1, \ldots, a_{t-1})$ and $\mathbf{b} = (b_0, b_1, \ldots, b_{t-1})$ be the first rows of $\mathbf{A}$ and $\mathbf{B}$. Also, let

$$P_{\mathbf{a}}(s) + P_{\mathbf{b}}(s) = \gamma_s \quad \text{for } s = 1, 2, \ldots, t-1, \tag{8}$$

where $|\gamma_s| = \gamma$ is a constant positive real number. Then

$$\mathbf{A}\mathbf{A}^\top + \mathbf{B}\mathbf{B}^\top = (2t-\gamma)\mathbf{I}_t + \gamma\mathbf{J}_t,$$

where $\mathbf{I}_t$ is the $t \times t$ identity matrix and $\mathbf{J}_t$ is a $t \times t$ matrix of ± 1's whose diagonal entries are all 1's. If $\gamma = 2$ and $\mathbf{J}_t$ is the $t \times t$ matrix of 1's, then the $2t \times 2t$ matrix

$$\mathbf{C}_2 = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B}^\top & \mathbf{A}^\top \end{pmatrix} \tag{9}$$

has the maximum determinant (see, Ehlich 1964) among all the $2t \times 2t$ $\{\pm 1\}$-matrices, i.e., D-optimal.

Theorem 4: Let $N = 2t$, t odd and $\mathbf{a}, \mathbf{b}$ two vectors of length t with entries from $\{\pm 1\}$ satisfying equation (8) for $\gamma_s \in \{-2, 2\}$. Let $a = \sum_i a_i$ and $b = \sum_i b_i$. Then an unbalanced, two-symbol, $\mathrm{E}(s^2)$-optimal and minimax-optimal SSD with N rows and $m = N$ columns achieving the lower bound $\mathrm{LB}(N, (0, \ldots, 0, k_{t-(a+b)/2} = t, 0, \ldots, 0, k_{t-(a-b)/2} = t, 0, \ldots, 0)$ can be constructed. If $\gamma_s = 2$ for $s = 1, 2, \ldots, t-1$, then the constructed design is also D-optimal.

Proof: Let $\mathbf{J}_t$ be a $t \times t$ matrix of ± 1's. Use circulant matrices $\mathbf{A}$ and $\mathbf{B}$ given in (9). Then $\mathbf{C}_2$ in (9) satisfies

$$\mathbf{C}_2\mathbf{C}_2^\top = \mathbf{C}_2^\top\mathbf{C}_2 = \begin{pmatrix} (2t-2)\mathbf{I}_t + 2\mathbf{J}_t & \mathbf{0}_{t\times t} \\ \mathbf{0}_{t\times t} & (2t-2)\mathbf{I}_t + 2\mathbf{J}_t \end{pmatrix}.$$

Thus, $\max_{i<j} |s_{ij}| = 2$, and $\mathbf{C}_2$ achieves the naive $\mathrm{E}(s^2)$ lower bound (5). Moreover, if $\mathbf{J}_t$ is the $t \times t$ matrix of 1's, then $\mathbf{C}_2$ is D-optimal.

Since $\mathbf{a} = (a_0, a_1, \ldots, a_{t-1})$ and $\mathbf{b} = (b_0, b_1, \ldots, b_{t-1})$ satisfy equation (8), t should be odd with $2(2t-1) = a^2 + b^2$. WLOG, we may assume that $a \geq b > 0$. Also, 1 and -1 appear $(t+a)/2$ and $(t+b)/2$ times in $\mathbf{a}$ and $(t-a)/2$ and $(t-b)/2$ times in $\mathbf{b}$. Each of the first t columns of the generated SSD by construction (9) has a column sum of $(a-b)$. So, the number of -1's in each of the first t columns is $l_2 = t - (a-b)/2$. Similarly, each of the last t columns has a column sum of $(a+b)$. So, the number of -1's in each of the last t columns is $l_1 = t - (a+b)/2$. Thus, $k_{l_2} = k_{t-(a-b)/2} = t$, $k_{l_1} = k_{t-(a+b)/2} = t$.

In the examples below, we denote 1 by + and -1 by $-$.

Example 2: Let

$$\begin{aligned} \mathbf{a} &= (+,+,+,+,+,+,+,+,-,+,+,-,-) \\ \mathbf{b} &= (+,+,+,-,-,-,+,-,+,+,-,+,-). \end{aligned}$$

Vectors $\mathbf{a}$ and $\mathbf{b}$ satisfy $P_{\mathbf{a}}(s) + P_{\mathbf{b}}(s) = 2$, $s = 1, 2, \ldots, t-1$. Moreover, $a = \sum_i a_i = 7$ and $b = \sum_i b_i = 1$. Let $\mathbf{A} = circ(\mathbf{a})$ and $\mathbf{B} = circ(\mathbf{b})$. Using construction (9) we obtain $\mathbf{D}$ such that $\mathbf{D} \in \mathcal{D}^{\pm}(26, 2^{26}, (k_9 = 13, k_{10} = 13))$ and $\mathrm{E}(s^2) = \mathrm{LB} = \mathrm{LB}_2 = 1.92$. Hence, $\mathbf{D}$ is an unbalanced, $\mathrm{E}(s^2)$-optimal, minimax-optimal SSD and a D-optimal design.

Example 3. Let

$$\begin{aligned} \mathbf{a} &= (+,+,+,+,+,-,+,-,-,+,+,+,-) \\ \mathbf{b} &= (+,+,+,+,+,-,+,-,-,+,+,+,-). \end{aligned}$$

Vectors $\mathbf{a}$ and $\mathbf{b}$ satisfy $P_{\mathbf{a}}(s)+P_{\mathbf{b}}(s) = 2$ for $s = 1, 2, \ldots, t-1$. Moreover, $a = \sum_i a_i = \sum_i b_i = 5$. Let $\mathbf{A} = circ(\mathbf{a})$ and $\mathbf{B} = circ(\mathbf{b})$. Using construction (9), we obtain $\mathbf{D}$ such that $\mathbf{D} \in \mathcal{D}^{\pm}(26, 2^{26}, (k_8 = 13, k_{13} = 13))$ and $\mathrm{E}(s^2) = \mathrm{LB} = \mathrm{LB}_2 = 1.92$. Hence, $\mathbf{D}$ is an unbalanced, $\mathrm{E}(s^2)$-optimal, minimax-optimal SSD and a D-optimal design.

4. Testing Achievability Computationally

We implemented a computational study to test the achievability of $\mathrm{LB}(N, (k_0, \ldots, k_{\lfloor N/2 \rfloor}))$. In searching for an SSD $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ achieving $\mathrm{LB}(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, we assumed that $k_0 = 1$. First we show that this assumption can be made without losing generality. To do this, we need the following two lemmas.

Lemma 5: Let $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$. Then $[\mathbf{1}_N\, \mathbf{D}] \in \mathcal{D}^{\pm}(N, 2^{(m+1)}, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, and $\mathbf{D}$ is $\mathrm{E}(s^2)$-optimal if and only if $[\mathbf{1}_N\, \mathbf{D}]$ is $\mathrm{E}(s^2)$-optimal.

Proof: Observe that

$$N^2 + \mathrm{SS}(\mathbf{D}^\top\mathbf{D}) + 2\sum_l k_l(N-2l)^2 = \mathrm{SS}([\mathbf{1}_N\, \mathbf{D}]^\top[\mathbf{1}_N\, \mathbf{D}]).$$

Hence, $\mathrm{SS}(\mathbf{D}^\top\mathbf{D})$ is minimized if and only if $\mathrm{SS}([\mathbf{1}_N\,\mathbf{D}]^\top[\mathbf{1}_N\,\mathbf{D}])$ is minimized. The result follows because for any SSD $\mathbf{X} \in \mathcal{D}^{\pm}(N', 2^{m'}, (k'_0, \ldots, k'_{\lfloor N'/2 \rfloor}))$ we have

$$\mathrm{E}(s^2) = \frac{(\mathbf{X}^T\mathbf{X}) - N'^2m'}{m'(m'-1)}.$$

Hence, $\mathbf{X}$ is $\mathrm{E}(s^2)$-optimal if and only if $\mathrm{SS}(\mathbf{X}^\top\mathbf{X})$ is minimized.

Lemma 6: Let $k_0 = 0$ and $\sum_l k_l = m$. Then

$$(m+1)m\,\mathrm{LB}(N, (1, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) = m(m-1)\,\mathrm{LB}(N, (0, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) + 2\sum_l k_l(N-2l)^2.$$

Proof: Each $\mathrm{LB}_i(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ for $i = 1, 2$ is derived based on a hypothetical $\mathbf{D}_i^* \in \mathcal{D}^{\pm}(N, 2^m, (0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$, where

$$\mathrm{LB}_i(N, (k_0, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) = \frac{\mathrm{SS}(\mathbf{D}_i^{*\top}\mathbf{D}_i^*) - N^2m}{m(m-1)}.$$

If $\mathbf{D}_i^* \in \mathcal{D}^{\pm}(N, 2^m, (0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ then $[\mathbf{1}_N\,\mathbf{D}_i^*] \in \mathcal{D}^{\pm}(N, 2^{(m+1)}, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$. Now,

$$\begin{aligned}
\mathrm{LB}_i(N, (0, k_1 \ldots, k_{\lfloor \frac{N}{2} \rfloor})) &= \frac{\mathrm{SS}(\mathbf{D}_i^{*\top}\mathbf{D}_i^*) - N^2m}{m(m-1)}, \\
\mathrm{LB}_i(N, (1, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) &= \frac{\mathrm{SS}([\mathbf{1}_N\,\mathbf{D}_i^*]^\top[\mathbf{1}_N\,\mathbf{D}_i^*]) - N^2(m+1)}{(m+1)m}.
\end{aligned}$$

Hence, by Lemma 5

$$(m+1)m\,\mathrm{LB}_i(N, (1, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) = m(m-1)\,\mathrm{LB}_i(N, (0, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) + 2\sum_l k_l(N-2l)^2.$$

Now, the result follows from

$$\mathrm{LB}(N, (k_0, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) = \max_{i\in\{1,2\}}\{\mathrm{LB}_i(N, (k_0, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor}))\}$$

and

$$(m+1)m \max_{i\in\{1,2\}} \mathrm{LB}_i(N, (1, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) =$$
$$m(m-1) \max_{i\in\{1,2\}} \mathrm{LB}_i(N, (0, k_1, \ldots, k_{\lfloor \frac{N}{2} \rfloor})) + 2\sum_l k_l(N-2l)^2.$$

The definition of an SSD $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ requires that it has no pair of *aliased* columns, i.e., columns $\mathbf{d}_i$ and $\mathbf{d}_j$ such that $\mathbf{d}_i^\top\mathbf{d}_j \in \{-N, N\}$. Hence, we must have $k_0 \leq 1$ for each SSD $\mathbf{D}$.

Theorem 5: An SSD $\mathbf{D}$ achieving $\mathrm{LB}(N, (0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ exists if and only if an SSD $[\mathbf{1}_N\,\mathbf{D}]$ achieving $\mathrm{LB}(N, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ exists.

Proof: The result follows immediately from Lemmas 5 and 6.

By Theorem 5, WLOG, we can restrict our search for SSDs achieving $\mathrm{LB}(N, (k_0, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ to SSDs with $k_0 = 1$. Accordingly, we wrote a C program for the NOA_p for $p = 2, 4, 8$ algorithms together with the derived $\mathrm{E}(s^2)$ lower bound $\mathrm{LB}(N, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ to search for SSDs in $\mathcal{D}^{\pm}(N, 2^m, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ achieving this bound.

There is a one-to-one correspondence between all SSDs achieving $\mathrm{LB}(N, (1, 0, \ldots, 0, m-1))$ and all balanced SSDs achieving $\mathrm{LB}(N, (0, \ldots, 0, m-1))$ obtained by deleting the all ones column. The NOA_p for $p = 2, 4, 8$ algorithms could find a balanced $\mathrm{E}(s^2)$-optimal SSD for $N = 14, 16$ in all cases except the $N = 14, m = 16$ case (Bulutoglu and Ryan, 2008). For the cases with $N \in \{14, 16\}$ and $15 \leq m \leq 70$ for which $\mathrm{LB}(N, (0, \ldots, 0, m-1))$ is (not) equal to $\mathrm{E}(s^2)$-lower bound in Bulutoglu and Ryan (2008), we were (not) able to find an SSD in $\mathcal{D}^{\pm}(N, 2^m, (1, 0, \ldots, 0, m))$ achieving $\mathrm{LB}(N, (1, 0, \ldots, 0, m))$ except for the $N = 14, m = 17$ case (this case with the all 1's column corresponds to the $N = 14, m = 16$ case in Ryan and Bulutoglu (2007). These observations confirm the correctness of our C program.

For each of the number of rows N and number of columns m combinations in Table 1 we randomly generated 100 vectors $(1, k_1, \ldots, k_{\lfloor N/2 \rfloor})$ such that $k_0 = 1$ and $\sum_l k_l = m$. Then for each of these 100 vectors $(1, k_1, \ldots, k_{\lfloor N/2 \rfloor})$, we used the NOA_4 exchange algorithm (Bulutoglu and Ryan, 2008; Ryan and Bulutoglu, 2007) to search for an SSD $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ achieving $\mathrm{LB}(N, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$. The complexity of running the NOA_p algorithm increases with p. However, for each random starting design, increased p increases the probability of converging to an SSD which has no aliased columns by definition. After experimenting with NOA_2, NOA_4, and NOA_8 we decided to use NOA_4 as a compromise between speed and avoidance of converging to a design with aliased columns. In Table 1, the column N reports N, the column m reports a range of m for which this experiment was conducted, and the column *numiter* reports the number of random starting designs that were used each time the NOA_4 algorithm was run. (We changed *numiter* only with N.) For each N, m combination such that m is within the reported range of m, the numbers of successes column of Table 1 reports the number of times out of 100 sampled vectors $(1, k_1, \ldots, k_{\lfloor N/2 \rfloor})$ an SSD achieving $\mathrm{LB}(N, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ was found. The number of successes in Table 1, which are in fact each a percent out of 100 sampled vectors $(1, k_1, \ldots, k_{\lfloor N/2 \rfloor})$, can only underestimate the true percentage of the sampled vectors $(1, k_1, \ldots, k_{\lfloor N/2 \rfloor})$ where our bound is achievable. Hence, for each vector $(1, k_1, \ldots, k_{\lfloor N/2 \rfloor})$, we needed a sufficient number of random starting designs to avoid significantly underestimated true percentages. We observed that for the balanced cases of $N = 10$ and $N = 14$, $numiter = 10^6$ and $numiter = 10^7$ were sufficient. (We were able to determine this, as for the balanced cases we know exactly when $\mathrm{LB}(N, (0, \ldots, 0, m))$ is achievable.) Hence, we set $numiter = 10^7$ for $11 \leq N \leq 14$ and $numiter = 10^6$ for $N \leq 10$. However, just because these values of *numiter* are sufficient in the balanced cases does not guarantee that they will be sufficient for the corresponding unbalanced cases.

Our Table 1 estimates do decrease with increased N. The output of our computational experiments also provided us with the iteration number at which an $\mathrm{E}(s^2)$-optimal SSD achieving $\mathrm{LB}(N, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$ was found. We used this information to perform a statistical analysis to determine the significance of underestimation. Our statistical analysis suggests that the true percentage of the sampled vectors $(1, k_1, \ldots, k_{\lfloor N/2 \rfloor})$ where our bound is achievable $\mathrm{LB}(N, (1, k_1, \ldots, k_{\lfloor N/2 \rfloor}))$

Table 1: Numbers of times an SSD achieving $\mathrm{LB}(N,(1,k_1,\ldots,k_{\lfloor N/2\rfloor}))$ **found for a randomly generated set of** 100 **vectors** $(1,k_1,\ldots,k_{\lfloor N/2\rfloor})$ **such that** $1+\sum_l k_l = m$

N	m	*numiter*	numbers of successes each out of 100
7	7-20	10^6	17 52 34 36 68 68 55 41 67 72 55 51 45 69
8	8-35	10^6	12 18 10 44 57 45 50 63 76 65 58 59 69 83 64 68 67 75 69 46 45 56 55 45 45 53 40 30
9	9-69	10^6	4 4 15 37 45 33 31 47 54 76 61 68 74 75 70 63 63 66 82 72 56 60 69 65 52 45 52 50 37 37 40 44 34 30 33 23 28 18 18 22 15 14 7 12 17 7 3 5 9 4 2 3 5 1 1 2 2 5 0 0 2
10	10-63	10^6	2 4 25 33 18 11 45 64 57 52 54 61 66 58 53 68 59 63 60 67 61 62 59 54 58 59 61 59 59 47 55 55 52 49 45 50 46 38 34 34 36 34 27 30 24 16 26 28 21 26 18 11 13 13
11	11-50	10^7	2 8 20 8 2 13 29 56 52 41 36 42 46 54 43 54 58 58 53 53 53 56 53 49 55 46 50 50 46 41 40 38 39 39 36 29 30 34 36 33
12	12-43	10^7	0 0 0 0 12 36 56 32 18 21 28 57 28 27 35 48 39 32 26 32 32 31 32 32 33 29 34 30 25 33 30 22
13	13-35	10^7	0 0 0 3 14 27 20 2 4 12 31 25 21 17 18 29 27 27 28 18 36 28 20
14	14-35	10^7	0 0 3 12 2 1 0 4 7 9 6 1 4 4 7 5 1 3 6 4 3 3

is underestimated significantly only in the $N \geq 13$ rows of Table 1.

5. Application

In this section we discuss two possible applications of our newly derived $\mathrm{E}(s^2)$ lower bound.

5.1. Application to searching for D-optimal designs

Let $\mathbf{D}$ be an $N\times m$ ($m \leq N$) matrix with entries from $\{\pm 1\}$ representing an N row, two-symbol and m column array. Let $\lambda_1,\lambda_2,\ldots,\lambda_m$ be all the non-zero eigenvalues of $\mathbf{D}^\top\mathbf{D}$, where $\mathrm{rank}(\mathbf{D}^\top\mathbf{D}) = m$. By the spectral decomposition theorem,

$$\begin{aligned}
\sum_i \lambda_i &= \mathrm{Tr}(\mathbf{D}^\top\mathbf{D}) = Nm,\\
\sum_i \lambda_i^2 &= \mathrm{Tr}((\mathbf{D}^\top\mathbf{D})^2) = \mathrm{Tr}((\mathbf{D}\mathbf{D}^\top)^2) = \mathrm{SS}(\mathbf{D}^\top\mathbf{D}) = \mathrm{SS}(\mathbf{D}\mathbf{D}^\top),\\
\mathrm{Det}(\mathbf{D}^\top\mathbf{D}) &= \prod_i \lambda_i.
\end{aligned}$$

If $\sum_i \lambda_i^2 = \theta$, then $\prod_i \lambda_i \leq d(\theta)$ for some $d(\theta) \geq 0$, where θ is some positive integer. This is equivalent to $-\log(\prod_i \lambda_i) \geq -\log(d(\theta))$ by the monotonicity of the $-\log(\cdot)$ function. To find such $d(\theta)$ that is as small as possible we consider the following smooth, non-convex nonlinear programming (NLP) problem

$$\begin{aligned} &\min \sum_i -\log(\lambda_i) \\ &\text{subject to: } \sum_i \lambda_i = Nm, \\ &\sum_i \lambda_i^2 = \theta, \quad \lambda_i \geq 0. \end{aligned} \tag{10}$$

NLP (10) was solved analytically in Cheng (1978). Cheng (1978) showed that the minimum is attained at a point which has constant coordinates $\lambda^* = N$ when $\theta = N^2 m$ or has two distinct coordinates $\lambda_1^* > \lambda_2^* > 0$ when $\theta > N^2 m$, where λ_1^* has multiplicity n, λ_2^* has multiplicity $m - n$, and

$$\begin{aligned} \lambda_1^* &= N + \sqrt{\frac{(m-n)(\theta - N^2 m)}{mn}}, \\ \lambda_2^* &= N - \sqrt{\frac{n(\theta - N^2 m)}{(m-n)m}}. \end{aligned}$$

Let

$$d(\theta, N, m, n) = \left(N + \sqrt{\frac{(m-n)(\theta - N^2 m)}{mn}}\right)^n \left(N - \sqrt{\frac{n(\theta - N^2 m)}{(m-n)m}}\right)^{m-n}.$$

Cheng (1978) also showed that $d(\theta, N, m, n)$ is a strictly decreasing function of n. For $\theta > N^2 m$, this result implies that

$$d(\theta) = d(\theta, N, m, 1) = \left(N + \sqrt{\frac{(m-1)(\theta - N^2 m)}{m}}\right)\left(N - \sqrt{\frac{(\theta - N^2 m)}{(m-1)m}}\right)^{m-1}$$

is a valid upper bound for $\text{Det}(\mathbf{D}^\top \mathbf{D}) = \prod_i \lambda_i$. Then for fixed N and m, by differentiating $\log(d(\theta))$, we see that $d(\theta)$ is a strictly decreasing function of θ for $\theta > 0$. Hence, we get the following theorem.

Theorem 6: Let $\theta \in \mathbb{Z}$ be such that $\theta > N^2 m$. Then

$$\text{SS}(\mathbf{D}^\top \mathbf{D}) = \text{SS}(\mathbf{D}\mathbf{D}^\top) = \sum_i \lambda_i^2 \geq \theta,$$

implies

$$\text{Det}(\mathbf{D}^\top \mathbf{D}) = \prod_i \lambda_i \leq d(\theta).$$

The following example shows how Theorem 6 can be used to derive upper bounds for the $\text{SS}(\mathbf{D}^*(\mathbf{D}^*)^\top)$, where $\mathbf{D}^*$ is a two-symbol, D-optimal design.

Example 4: For $N = 22$ and $m = 22$, the largest possible $\text{Det}(\mathbf{D}^\top \mathbf{D})$ of $20^{12} \times (6400000)^2$ for a two-symbol design with entries from $\{\pm 1\}$ is given by Chasiotis *et al.* (2018). Then by Theorem 6, for a D-optimal design $\mathbf{D}^*$ with 22 rows and 22 columns, we must have

$$\text{SS}((\mathbf{D}^*)^\top \mathbf{D}^*) = \text{SS}(\mathbf{D}^*(\mathbf{D}^*)^\top) \leq 11,920.$$

This proves that a D-optimal design in this case cannot be balanced. Since for a balanced 22 row and 22 column design $\mathbf{D}$, $\text{SS}(\mathbf{D}^\top\mathbf{D}) \geq 12,496$.

Next, we provide a connection between D-optimal designs and the $\text{E}(s^2)$ lower bound that we derived in Section 2. Let $\mathbf{D} = (d_{ij})$ be a sought after D-optimal design with N rows and m columns and $\mathbf{d}_1^\top$ be the first row of $\mathbf{D}$. Let

$$\hat{\mathbf{D}} = \mathbf{D}\text{diag}(\mathbf{d}_1),$$

where $\text{diag}(\mathbf{d}_1)$ is the $m \times m$ diagonal matrix whose (j, j)'th entry is the j'th entry of $\mathbf{d}_1$. Now, $\hat{\mathbf{D}}^\top$ can be viewed as a not necessarily balanced SSD. Let $\text{E}(s^2)$ be the $\text{E}(s^2)$ value of $\hat{\mathbf{D}}^\top$. Then the $\text{E}(s^2)$ values of $\hat{\mathbf{D}}^\top$ satisfies

$$\text{SS}(\hat{\mathbf{D}}^\top\hat{\mathbf{D}}) = \text{SS}(\hat{\mathbf{D}}\hat{\mathbf{D}}^\top) = \text{E}(s^2)N(N-1) + Nm^2.$$

Hence we get the following corollary to Theorem 6.

Corollary 1: Let $\theta \in \mathbb{Z}$ be such that $\theta > N^2m$. Then

$$\text{E}(s^2) \geq \frac{\theta - Nm^2}{N(N-1)}$$

implies

$$\text{Det}(\mathbf{D}^\top\mathbf{D}) = \text{Det}(\hat{\mathbf{D}}^\top\hat{\mathbf{D}}) \leq d(\theta).$$

Now for a given value of $d(\theta)$, when searching for a $\mathbf{D}$ with

$$\text{Det}(\mathbf{D}^\top\mathbf{D}) > d(\theta),$$

the search can be restricted to $\hat{\mathbf{D}}^\top$ with balancedness structure $(1, k_1, \ldots, k_{\lfloor m/2 \rfloor})$ such that $\hat{\mathbf{D}}^\top \in \mathcal{D}^{\pm}(m, 2^N, (1, k_1, \ldots, k_{\lfloor m/2 \rfloor}))$ and

$$\text{LB}(m, (1, k_1, \ldots, k_{\lfloor \frac{m}{2} \rfloor})) \leq \frac{\theta - Nm^2}{N(N-1)}. \tag{11}$$

The θ in (11) can be decreased by using the methods of Chasiotis *et al.* (2018). This restriction should decrease the search space significantly and improve algorithm performance for finding D-optimal designs in Brent *et al.* (2011).

Requiring an SSD to be minimax optimal has the benefit of reducing the search space and can be useful in the search for a D-optimal design. In fact, there is a 22 row, 22 column, D-optimal design that can be viewed as an unbalanced, $\text{E}(s^2)$-optimal, minimax-optimal SSD achieving the naive $\text{E}(s^2)$ lower bound (5), (see Chasiotis *et al.* 2018). However, finding even a balanced $\text{E}(s^2)$-optimal and minimax-optimal SSD in general is a very difficult problem (Morales and Bulutoglu, 2018).

5.2. Application to finding upper bounds on the maximum number of columns

The following is an important theoretical problem in the SSD literature (Cheng and Tang, 2001).

Problem 1: For a given $0 \le t \le N$ find the maximum number of columns $B(N, t, (0, \dots, 0, m))$ such that an SSD $\mathbf{D} \in \mathcal{D}^{\pm}(N, 2^m, (0, \dots, 0, m))$ with $s_{\max} \le t$ exists. The generalization of this problem to unbalanced SSDs is determining $B(N, t, (k_0, k_1, \dots, k_{\lfloor N/2 \rfloor}))$. Multiplying a subset of rows of an SSD $\mathbf{D}$ does not change the $s_{\max}$ of $\mathbf{D}$. Hence, by Lemma 1 in (Cheng and Tang, 2001), WLOG it suffices to find $B(N, t, (1, k'_1, \dots, k'_{\lfloor N/2 \rfloor}))$, where $k'_i = 0$ for $1 \le i < (N-t)/2$.

By using $\mathrm{E}(s^2)$ lower bounds on balanced SSDs, Cheng and Tang (2001) found an upper bound on $B(N, t, (0, \dots, 0, m))$. Our newly derived $\mathrm{E}(s^2)$ lower bounds for unbalanced SSDs can be used to generalize the upper bound on $B(N, t, (0, \dots, 0, m))$ to an upper bound on $B(N, t, (1, k'_1, \dots, k'_{\lfloor N/2 \rfloor}))$ with $k'_i = 0$ for $1 \le i < (N-t)/2$.

The following is an important theoretical problem in the frame theory literature (Szöllösi and Östergård, 2018).

Problem 2: For given $a \in (0, 1)$ and $d \in \mathbb{Z}^{\ge 0}$, find the maximum number of equiangular lines in $\mathbb{R}^d$ with pairwise angle $\arccos(a)$.

For an SSD $\mathbf{X} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \dots, k_{\lfloor N/2 \rfloor}))$, let $s_{\min} = \min_{i<j} |s_{ij}|$. Let $C(N, t, d)$ be the maximum m such that an SSD $\mathbf{X} \in \mathcal{D}^{\pm}(N, 2^m, (k_0, k_1, \dots, k_{\lfloor N/2 \rfloor}))$ with $t = s_{\max} = s_{\min}$ and $\mathrm{rank}(\mathbf{X}) = d \le N$ exists. Since multiplying a subset of rows of an SSD $\mathbf{D}$ does not change the $s_{\max}$ and $s_{\min}$ of $\mathbf{D}$, WLOG we can assume $\mathbf{X} \in \mathcal{D}^{\pm}(N, 2^m, (1, k'_1, \dots, k'_{\lfloor N/2 \rfloor}))$, where $k'_{(N-t)/2} = m - 1$ and $k'_i = 0$ for $i \notin \{0, (N-t)/2\}$. Hence, it is plain to see that $C(N, t, d) \le B(N, t, (1, k'_1, \dots, k'_{\lfloor N/2 \rfloor}))$.

The number $C(N, t, d)$ is a lower bound on the number of equiangular lines in $\mathbb{R}^d$ with pairwise angle $\arccos(t/N)$. Hence, determining $C(N, t, d)$ provides information on the solution of Problem 2 for $a = t/N$ and $B(N, t, (1, k'_1, \dots, k'_{\lfloor N/2 \rfloor}))$, where $k'_{(N-t)/2} = m - 1$ and $k'_i = 0$ for $i \notin \{0, (N-t)/2\}$ bounds $C(N, t, d)$. We propose determining $C(N, t, d)$ for many N, t, d combinations by using the upper bound $B(N, t, (1, k'_1, \dots, k'_{\lfloor N/2 \rfloor}))$ as needed as a future research project.

Acknowledgements

The research of Dr. Dursun A. Bulutoglu was supported by the Air Force Office of Scientific Research grant F4FGA08087J006.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

References

Brent, R. P., Orrick W. P., Osborn, J. and Zimmermann, P. (2011). Maximal determinants and saturated d -optimal designs of orders 19 and 37. *Arxiv*. *http://arxiv.org/pdf/ 1112.4160.pdf*

Bulutoglu, D. A. and Ryan, K. J. (2008). $E(s^2)$-optimal supersaturated designs with good minimax properties when *n* is odd. *Journal of Statistical Planning and Inference*, **138**, 1754-1762.

Chasiotis, V., Kounias, S. and Farmakis, N. (2018). The *d*-optimal saturated designs of order 22. *Discrete Mathematics*, **341**, 380–387.

Cheng, C. S. (1978). Optimality of certain asymmetrical experimental designs. *Annals of Statistics*, **6**, 1239–1261.

Cheng, C. S. and Tang, B. (2001). Upper bounds on the number of columns in supersaturated designs. *Biometrika*, **88**, 1169–1174.

Das, A., Dey, A., Chan, L-Y. and Chatterjee, K. (2008). $E(s^2)$- optimal supersaturated designs. *Journal of Statistical Planning and Inference*, **138**, 3749–3757.

Ehlich, H. (1964). Determinantenabschätzung für binäre matrizen. *Mathematische Zeitschrift*, **83**, 123–132.

Georgiou, S. D. (2014). Supersaturated designs: A review of their construction and analysis. *Journal of Statistical Planning and Inference*, **144**, 92–109.

Koukouvinos, C. and Stylianou, S. (2008). On skew-hadamard matrices. *Discrete Mathematics*, **308**, 2723–2731.

Morales, L. B. and Bulutoglu, D. A. (2018). On $E(s^2)$-optimal and minimax-optimal supersaturated designs with 20 rows and 76 columns. *Journal of Combinatorial Designs*, **26**, 344–355.

Morales, L. B., Bulutoglu, D. A. and Arasu, K. T. (2019). The maximum number of columns in supersaturated designs with $s_max = 2$. *Journal of Combinatorial Designs*, **27**, 448–472.

Ryan, K. J. and Bulutoglu, D. A. (2007). $E(s^2)$-optimal supersaturated designs with good minimax properties. *Journal of Statistical Planning and Inference*, **137**, 2250–2262.

Suen, C. S. and Das, A. (2010). $E(s^2)$-optimal supersaturated designs with odd number of runs. *Journal of Statistical Planning and Inference*, **140**, 1398–1409.

Szöllösi, F. and Östergård, P. R. J. (2018). Enumeration of seidel matrices. *European Journal of Combinatorics*, **69**, 169–184.

Waldron, S. (2009). On the construction of equiangular frames from graphs. *Linear Algebra and its Applications*, **431**, 2228–2242.