A Discrete Fourier Transform Approach Searching for Compatible Sequences and Optimal Designs

S.D. Georgiou
Department of Statistics and
Actuarial-Financial Mathematics,
University of the Aegean,
Karlovassi 83200, Samos, Greece
E-mail: stgeorgiou@aegean.gr

K. Drosou and C. Koukouvinos
Department of Mathematics
National Technical University of Athens
Zografou 15773, Athens, Greece
E-mail: drosou.kr@gmail.com, ckoukouv@math.ntua.gr

In this paper, we apply a new method of evaluating the Discrete Fourier Transform that requires significantly less computational effort. In this evaluation, the Discrete Fourier Transform is defined over the support of the sequence of interest. The method can be applied to search for sequences with zero periodic autocorrelation function. As an example we apply the procedure and we were able to classify weighing matrices $W(2n, 9)$ constructed from two sequences of length n and weight $(5, 4)$ for all $400 \le n \le 500$.

Povzetek: Predstavljena je nova metoda Fourierjeve transformacije.

1 Introduction

Sequences with zero or low autocorrelation function have been widely used in Statistics and in particular in the theory of optimal experimental designs. In many cases the experimenter wishes to develop and study an empirical linear regression model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad (1)$$

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

It is well known that the least square estimator for the coefficient vector $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ with covariance matrix $Cov(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X'X})^{-1}$. Orthogonality of the design matrix $\mathbf{X}$ is essential to create models with optimal variance. More details on linear regression analysis and optimal designs can be found in [2, 13]. Sequences with a zero

autocorrelation function can be used to generate orthogonal design matrices that achieve the optimal covariance matrix for the estimator of $\widehat{\boldsymbol{\beta}}$. Such sequences are also known as *compatible sequences*.

Two level and three level design matrices are commonly used for screening and weighing experiments. Recently, new methods for constructing three level screening designs, from weighing matrices, were proposed in the literature. For example, in [17] the authors used $W(n, n-1)$ to construct three-level screening designs while their results were generalized to the case of $W(n, k)$ in [6]. The designs constructed by the methods of the above papers, have their main effects orthogonal to each other, orthogonal to any quadratic effects and orthogonal to any two-factors interactions. For quantitative factors, the linear model (1) with such designs can be used for screening out the main effects in the presence of active second order terms, such as two-factor interactions or pure quadratic effects, in the true model.

Such orthogonal designs of experiments can be easily constructed from sequences with zero autocorrelation function. The construction of such sequences is important but the needed computational effort is sometimes enormous, making the search for such desirable designs infeasible. Some known algorithms for developing sequences with zero autocorrelation function and the related optimal ex-

perimental designs can be found in [1, 8] and the references therein. A method that uses the Discrete Fourier Transform (DFT) was recently developed in the literature (see [3]).

In this paper, we proposed a new evaluation of the DFT that reduces the computational effort. This evaluation can be used in searching for sequences with zero autocorrelation function. It is applicable to sequences with two, three or more levels and the complexity of the method does not depend on the length of the sequences but only on the number of their non zero elements (weight). So, when searching for weighing matrices of order n and weight k, constructed from a number of suitable circulant matrices (sequences), the complexity depends only on k. In this way, for a fixed weight k, one may search for large optimal weighing designs. Moreover, this method can test each of the required sequences independently and decide whether or not this sequence can be a candidate for a set of compatible sequences. As an example, we apply the suggested method to search for sequences with zero periodic autocorrelation function that can be used to classify a special type of weighing matrices.

2 Preliminary Results

A weighing matrix $W = W(n, k)$ is a square matrix with entries 0, ± 1 having k non-zero entries per row and column and having zero inner product of distinct rows. Hence W satisfies $WW^T = kI_n.$. The number k is called the *weight* of W. Weighing matrices have long been studied because of their use in weighing experiments as first studied by Hotelling [10] and later by Raghavarao [12] and others [1, 14].

Given a set of ℓ sequences,

$$A = \{A_j : A_j = (a_{j0}, a_{j1}, ..., a_{j(n-1)}), \ j = 1, ..., \ell\}, \quad (2)$$

of length n, the *periodic autocorrelation function* (abbreviated as PAF) $P_A(s)$ is defined, reducing $i + s$ modulo n, as

$$P_A(s) = \sum_{j=1}^{\ell} \sum_{i=0}^{n-1} a_{ji} a_{j,i+s}, \quad s = 0, 1, ..., n - 1. \quad (3)$$

The set A of the above sequences is called *compatible* if $\sum_{j=1}^{\ell} P_{A_j}(s) = a$, $s = 1, 2, ..., n - 1$. Moreover, if $a = 0$ then the sequences A are said to have *zero periodic autocorrelation function* (zero PAF).

Notation. We use the following notation throughout this paper.

1. We use $\overline{x}$ to denote $-x$.

2. We use $R = (r_{ij})$ to denote the $n \times n$ *back diagonal matrix* whose elements satisfy $r_{ij} = \begin{cases} 1, & when \ i + j = n + 1 \\ 0, & otherwise \end{cases}$ $i, j = 1, 2, ..., n$.

3. Let $A = (a_0, a_1, ..., a_{n-1})$ where $a_i \in \{0, \pm 1\}$. The *support of A* is the set $SP_A = \{i : a_i \neq 0\}$.

The *discrete Fourier transform (DFT)* of a sequence $B = (b_0, b_1, ..., b_{n-1})$ of length n is given by

$$DFT_B(k) = \mu_B(k) = \sum_{i=0}^{n-1} b_i \omega^{ik}, \quad k = 0, 1, ..., n - 1,$$

$$(4)$$

where ω is the primitive n-th root of unity, $e^{\frac{2\pi i}{n}}$. If we take the squared magnitude of each term in the DFT of B, the resulting sequence is called the *power spectral density (PSD)* of B and will be denoted by $|\mu_B(k)|^2$.

We make use of the following well-known theorem (see [16, chapter 10]).

Theorem 1. *Let B be a sequence of length n with elements from the set $\{0, \pm 1\}$. The PSD of this sequence is equal to the DFT of its periodic autocorrelation function:*

$$|\mu_B(k)|^2 = \sum_{j=0}^{n-1} P_B(j) \omega^{jk}. \quad (5)$$

3 The Support Based Discrete Fourier Transform and Power Spectral Density

The constant value of the PSD of compatible sequences can be easily calculated using the elements of the sequences (see [3]). The following Lemma illustrates an alternative method for calculating this value.

Lemma 1. *Let $A = \{A_j : A_j = (a_{j0}, a_{j1}, ..., a_{j(n-1)}), \ j = 1, ..., \ell\}$, be a set of ℓ sequences of length n, with zero periodic autocorrelation function. Then we have that*

$$\sum_{j=1}^{\ell} |\mu_{A_j}(k)|^2 = \sum_{j=1}^{\ell} P_{A_j}(0) = \sum_{i=0}^{n-1} \sum_{j=1}^{\ell} a_{ji}^2 = c.$$

Proof. Using equation (5) and the fact that the sequences have zero PAF we obtain

$$\sum_{j=1}^{\ell} |\mu_{A_j}(k)|^2 = \sum_{j=1}^{\ell} \sum_{s=0}^{n-1} P_{A_j}(s) \omega^{sk}$$

$$= \sum_{s=0}^{n-1} \sum_{j=1}^{\ell} P_{A_j}(s) \omega^{sk}$$

$$= \sum_{j=1}^{\ell} P_{A_j}(0) + \sum_{s=1}^{n-1} \left(\sum_{j=1}^{\ell} P_{A_j}(s) \right) \omega^{sk}$$

$$= \sum_{j=1}^{\ell} P_{A_j}(0) = \sum_{i=0}^{n-1} \sum_{j=1}^{\ell} a_{ji}^2 = c.$$

$$\square$$

Thus, when searching for sequences with zero autocorrelation, it is very simple to find the constant value of the PSD since this constant will be the sum of squares of the elements of the sequences.

Let $A = (a_0, a_1, \ldots, a_{n-1})$ with $a_i \in \{0, \pm 1\}$ and let SP_A be the support of A. The *discrete Fourier transform (DFT)* of the sequence A can be defined on SP_A by

$$DFT_A(k) = \mu_A(k) = \sum_{i \in SP_A} a_i \omega^{ik}, k = 0, 1 \ldots, n-1,$$
(6)

where ω is the primitive n-th root of unity, $e^{\frac{2\pi i}{n}}$.

Lemma 2. *Suppose that we have a set A of compatible sequences, as in (2), with $P_A(s) = a$ for $s = 1, 2, \ldots, n-1$. Then*

$$\sum_{j=1}^{\ell} |\mu_{A_j}(k)|^2 = c,$$
(7)

where $k = 0, 1, \ldots n-1$ and $c = \sum_{j=1}^{\ell} P_{A_j}(0) - a$.

Proof. The result is a straightforward generalization of a case proved in [7] for four sequences and thus the proof is omitted. □

Remark 1. Since $|\mu_X(k)|^2 \geq 0$ and $\sum_{j=1}^{\ell} |\mu_{A_j}(k)|^2 = c$ we have the following. A sequence X should satisfy $|\mu_X(k)|^2 \leq c$ for all $k = 0, 1, \ldots, n-1$ in order to be selected as a candidate for the compatible set A. When searching for sequences with elements from the set $\{-1, 0, 1\}$ the computational effort of the PSD does not depend on the length but only on the number of non-zero elements (weight) of the sequence.

Example 1. Suppose we wish to search for a weighing design of order $2n$ and weight w, constructed from two sequences A, B of length n each and $2n \geq w$. We need $2n$ summations and $2n$ multiplications to calculate the DFT (or the PSD) using the classic definition of DFT but only w summations and w multiplications are needed using the support based definition. If $w << n$ then the new definition is extremely fast (by comparison), while when $w = n$ it will be shown that the needed computational effort of the PSD is reduced in half. Note that the above calculation concerns only one candidate pair of compatible sequences from the total number of possible pairs in the search space. Since the search space in such problems is huge and exponentially increasing with n, it is clear that any reduction in the computational effort at each point of the search space results in a huge reduction of the total computational effort (in absolute terms).

Lemma 3. *Suppose we have a set A of $\ell = 2m$ sequences of length n, as in (2), with elements from the set $\{-1, 1\}$ and $P_A(s) = 0$ for $s = 1, 2, \ldots, n-1$. Then the set*

$$B = \{ \quad B_{2j-1} = \tfrac{1}{2}(A_{2j-1} + A_{2j}), \\ B_{2j} = \tfrac{1}{2}(A_{2j-1} - A_{2j}), \ j = 1, \ldots, m\}$$

is a set of ℓ sequences of length n with elements from the set $\{-1, 0, 1\}$ and $P_B(s) = 0$ for $s = 1, 2, \ldots, n-1$. The total weight of the new sequences, in B, is nm.

Remark 2. Using Lemma 3, we are able to get the benefits of the new definition of DFT even in the case of sequences with elements from the set $\{-1, 1\}$. Suppose that we are interested in searching for four sequences with elements from the set $\{-1, 1\}$, zero PAF and length n. By using the proposed definition of DFT we need $2n$ summations/multiplications for each evaluation of the DFT while the old definition of DFT requires $4n$ calculations. Generally, each calculation of the DFT will be reduced in half. If recursively ℓ nested evaluations of the DFT are used in the algorithm, then we have a reduction of calculation by a scale factor $1/2^{\ell}$.

4 An Illustrating Example

In this section we illustrate the use of the suggested procedure in searching for weighing matrices constructed from suitable circulant matrices (sequences). The computational advantages of this approach, as these were discussed in the previous section, are illustrated through numerical examples. We present in details the case of weighing matrices $W(2n, 9)$ that can be constructed from two circulants.

Weighing matrices $W(2n, 9)$ constructed from two sequences are of great interest but hard to find, since the necessary conditions for their existence are not sufficient (see [9]). It is well known that if there exist a $W(2n, k)$ constructed from two circulant matrices of order n, then $k = a^2 + b^2$, where a and b are the row (and column) sums of A and B respectively. The next theorem gives a known construction of weighing matrices by using two sequences with elements from the set $\{0, 1, -1\}$ and constant PSD (equivalently, by using two circulant matrices A_1, A_2 with elements from the set $\{0, 1, -1\}$ satisfying $A_1 A_1^T + A_2 A_2^T = kI$).

Theorem 2. (Geramita and Seberry (1979), Theorem 4.46) *If there exist two circulant matrices A_1, A_2 of order n with elements from the set $\{0, 1, -1\}$, satisfying*

$$\sum_{i=1}^{2} A_i A_i^T = kI,$$

then there exists a $W(2n, k)$.

The construction of the corresponding weighing matrix is achieved by using either of the following two arrays

$$D = \begin{pmatrix} A_1 & A_2 \\ -A_2^T & A_1^T \end{pmatrix} \quad , \quad D = \begin{pmatrix} A_1 & A_2 R \\ -A_2 R & A_1 \end{pmatrix}$$
(8)

The case of weighing matrices constructed from two circulants A and B having $(|SP_A|, |SP_B|) = (9, 0)$, which is actually the case where a circulant weighing matrix exists, was resolved in [15]. In [15] it was shown that a

circulant weighing matrix $W(n, 9)$ exist if and only if n is multiple of 13 or 24 (i.e. $13 \mid n$ or $24 \mid n$). This case implies that there exists one sequence with elements from the set $\{0, 1, -1\}$ having weight 9 and zero PAF. Thus, a circulant weighing matrix $W(n, 9)$ exists and a weighing matrix $W(2n, 9)$, constructed from two circulant matrices, exists (take one to be the matrix with all elements zero and the other to be the circulant weighing matrix $W(n, 9)$ as it is given in [15]). For more details on this case see [15]. If $(|SP_A|, |SP_B|) = (8, 1)$ and $P_A(s) + P_B(s) = 0$, $\forall\ s = 1, 2, \ldots, n-1$, and there exists a $W(2n, 9)$, constructed from two circulant matrices, such that $9 = a^2 + 1^2$ which is not possible since 8 is not a perfect square. So the case $(8, 1)$ is not permitted.

Two pairs of sequence are said to be equivalent if the one can be constructed from the other by some transformations. More specifically, we recall the following definition.

Definition 1. We say that two pairs of $(0, \pm1)$ sequences ((A,B) and (C,D)) of length n are *equivalent* iff one can be obtained from the other by applying some of the following transformations.

1. Multiply one or both sequences of the pair by -1.

2. Reverse one or both sequences of the pair.

3. Take circulant permutation of one or both sequences of the pair.

4. Multiply the elements of the support of both sequences by ℓ, $(\ell, n) = 1$.

We call the corresponding weighing matrices, constructed from the two circulant matrices whose first rows are the sequences (A, B) and (C, D), *equivalent*. More details on weighing matrices constructed from two circulant matrices can be found in [11].

In this section we classify the weighing matrices $W(2n, 9)$ for $400 \leq n \leq 500$ constructed from two sequences of weight 5 and 4 respectively. Results for $n < 100$ were presented in [4], for $n \leq 400$ in [5], and the results for $400 \leq n \leq 500$ are new and given in Table 4. One representative of each order was known, and were presented in [5]. Thus, in Table 4, we only present the number of inequivalent solutions for each order. All inequivalent sequences and the corresponding weighing designs are available on request.

In the next example we illustrate numerically the computational gain from the suggested approach in the case of $n = 500$.

Example 2. Following Example 1, we need just 9 calculations for each evaluation of the PSD in contrast to the 1000 calculations needed for the old definition. Note that we have about 500^4 sequences in the search space and thus the proposed algorithm require 2×10^{10} while the old definition needs 4×10^{12} simple calculations. So, the time needed is approximately 1 day by applying the old definition and just

n	400	403	405	406	407	408	410	413
N	16	5	5	8	5	8	5	1
n	414	415	416	418	420	424	425	427
N	2	1	10	12	27	3	6	1
n	429	430	432	434	435	437	440	441
N	10	5	7	7	6	6	20	3
n	442	444	445	448	450	451	455	456
N	8	2	1	15	10	6	7	9
n	459	460	462	464	465	468	469	470
N	3	11	17	6	5	7	1	4
n	472	473	475	476	480	481	483	484
N	3	6	7	13	20	5	5	8
n	485	488	490	492	493	494	495	496
N	1	3	11	3	4	9	10	5
n	497	500						
N	1	11						

Table 1: Number N of inequivalent solutions for the construction of $W(2n, 9)$, when $(|SP_A|, |SP_B|) = (5, 4)$ and $400 \leq n \leq 500$.

10 minutes with the new evaluation. As an extreme example consider a large search space (for n=100000) where the required search time using the old definition is more than a year (infeasible). The reduction of time (simple computations) will be by a scale factor $9/2n$ (i.e., 999.99% less) and thus the required time will be just a few hours.

5 Discussion

In this paper, we proposed a new evaluation of the DFT that reduces the computational effort. This evaluation can be used in searching for sequences with zero autocorrelation function. It is applicable to sequences with two, three or more levels and the complexity of the method does not depend on the length of the sequence but only on the number of non-zero elements (weight). So, when searching for weighing matrices of order n and weight w, constructed from a number of suitable circulant matrices (sequences), the complexity depends only on w. In this way, for fixed weight w, one may search for large optimal weighing designs. Moreover, this method can test each of the required sequences independently and decide whether or not this sequence can be a candidate for a set of compatible sequences. As an example, we applied the suggested method to search for sequences with zero periodic autocorrelation function that can be used to classify a special type of weighing matrices $W(2n, 9)$.

The reduction of the computational effort could lead to many new numerical results and help to resolve other open cases for weighing matrices. Moreover, the support based approach may be used for theoretical investigation of weighing matrices and other optimal designs. The same approach can be applied in many other cases where circulant matrices are used for the construction of optimal designs (see [2, 8, 12]).

6 Acknowledgments

References

[1] Craigen, R. and Kharaghani, H. *Orthogonal designs*, in: C.J. Colbourn and J.H. Dinitz (eds.), *H*andbook of Combinatorial Designs, 2 ed., CRC Press, Boca Raton, FL, 2007, pp. 273–280.

[2] V.V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York, 1972.

[3] R.J. Fletcher, M. Gysin and J. Seberry, Application of the discrete Fourier transform, to the search for generalised Legendre pairs and Hadamard matrices, *Australas. J. Combin.*, 23 (2001), 75–86.

[4] S. Georgiou and C. Koukouvinos, New infinite classes of weighing matrices, *Sankhya Ser. B*, 64 (2002), 26–36.

[5] S. Georgiou, Signed differences for weighing designs, *Sankhya Ser. B*, 72 (2010), 107–121.

[6] S.D. Georgiou, S. Stylianou, and M. Aggarwal, Efficient three-level screening designs using weighing matrices, (submitted).

[7] S. Georgiou, C. Koukouvinos and S. Stylianou, On good matrices, skew Hadamard matrices and optimal designs, *Computational Statistics and Data Analysis*, 41 (2002), 171–184.

[8] A.V. Geramita, and J. Seberry, *Orthogonal Designs: Quadratic Forms and Hadamard Matrices*, Marcel Dekker, New York, 1979.

[9] J. Horton and J. Seberry, When the necessary conditions are not sufficient: sequences with zero autocorrelation function, *Bulletin ICA*, 27 (1999), 51–61.

[10] H. Hotelling, Some improvements in weighing and other experimental techniques, *Ann. Math. Stat.*, 16 (1944), 294–300.

[11] C. Koukouvinos and J. Seberry, New weighing matrices and orthogonal designs constructed using two sequences with zero autocorrelation function—a review, *J. Statist. Plann. Inference*, 81 (1999), 153–182.

[12] D. Raghavarao, *Constructions and Combinatorial Problems in Design of Experiments*, New York, 1971.

[13] G.A.F. Seber and A.J. Lee, *Linear Regression Analysis*, Wiley, New York, 2003.

[14] J. Seberry and M. Yamada, Hadamard matrices, sequences and block designs, in *Contemporary Design Theory—a Collection of Surveys*, eds J.H. Dinitz and D.R. Stinson, Wiley, New York, 1992, pp. 431–560.

[15] Yoseph Strassler, *The Classification of Circulant Weighing Matrices of Weight 9*, PhD Thesis, Bar-Ilan University, Ramat-Gan, 1997.

[16] S.A. Tretter, *Introduction to Discrete-time Signal Processing*, Wiley, New York, 1976.

[17] L. Xiao, D.K.J. Lin, and F. Bai, Constructing Definitive Screening Designs Using Conference Matrices, *Journal of Quality Technology*, 44 (2012), 2–8.