Elfving’s Theorem Revisited

W.J. Studden

Department of Statistics, Purdue University, 1399 Mathematical Science Building, West Lafayette, IN 47907-1399, USA

Received 11 April 2003; received in revised form 24 May 2003; accepted 30 May 2003
Available online 29 July 2004

Dedicated to Professor Herman Chernoff on his 80th Birthday

Abstract

The importance of Elfving’s geometrical result is illustrated by showing how his original proof can be used to prove a number of Elfving-type results without the use of equivalence theorems. It is indicated that some of the elementary equivalence theorems follow easily from the Elfving results. It is also shown that Elfving Theorem is equivalent to a special case of a general theorem in approximation theory.

MSC: primary 60J15; secondary 33C45

Keywords: Elfving set

1. Introduction

The Elfving Theorem referred to in the title is an important, simple, elegant geometrical result which allows one to choose an experimental design which will minimize the variance of unbiased estimates of a single linear combination $c'\theta$ in an ordinary linear regression model $Y(z) = f'(z)\theta + \text{error}$ where $f'(z)\theta = \sum_{i=1}^{p} f_i(z)\theta_i$. Uncorrelated observation are chosen at $z_1, \ldots, z_n$ from some available set $Z$, and the objective is to determine the optimal choice of the $z_i$. This result is from Elfving (1952) and will be described more fully below.

Professor Chernoff had many research interests, one of which was the design of experiments. During the years 1958–62, I was a graduate student at Stanford and took at least one...
course from Herman which was on asymptotic theory. Part on my TA/RA duties while at Stanford was to help prepare the graphs for the paper by Herman titled, Optimal Accelerated Life Designs for Estimation, which was published in Technometrics in 1962. (I actually did very little work, if any, on this project since the assistant doing the computations and graphics was very reliable.) In this paper the lifetime of some object depended on some variable, say $z$, in a way that larger $z$ values produced shorter lifetimes. In investigating this relationship large $z$ values are used to accelerate or shorten the test times due to cost considerations. The objective was to determine which $z$ values should be used to run the tests. The paper uses a simple geometrical result of Elfving which Herman explained to me. This paper contains the first good application of Elfving’s result. I shortly became more involved with this result as Professors Karlin and Kiefer were also working on design problems. Herman wrote an article in 1999 in Statistical Science describing Elfving’s impact on the design of experiments and describes the result with some examples. The same volume contains papers by Nordström and Fellman also concerned with Elfving’s work.

The purpose of the present paper is to reiterate the importance of Elfving’s geometrical result. It will be shown that some of the results in design theory that rely on ‘equivalence theorems’ can be derived using Elfving’s original argument. The basic result is described in Section 2. In Section 3 it will be indicated that Elfving result is equivalent to some very fundamental results in approximation theory. A simple extension of Elfving’s result is described in Section 4 and Dette’s Elfving-type result is proven in Section 5 following Elfving’s original argument. The relationship between Elfving’s Theorem and some simple equivalence theorems is discussed briefly in Section 6.

There are a number of other results, that are not discussed here, where the Elfving set $E$, (see (2.8) below) plays a dominant role. For example, the inball radius of $E$ is related to the E-optimality or Elfving’s min–max criterion. This is discussed in Pukelsheim and Studden (1993), Dette and Studden (1993) and Heiligers et al. (1995). It should also be mentioned that Dette (1993) has included extensions, of the main result discussed here, for which the result in Section 4 is a special case.

2. Elfving’s Theorem

We are interested in estimating a single linear form $c'\theta$ in the model where observations are given by

$$Y(z_i) = f'(z_i)\theta + \varepsilon_i = \sum_{i=1}^{p} f_i(z_i)\theta_i + \varepsilon_i \quad i = 1, \ldots, n. \quad (2.1)$$

The observations are uncorrelated and have constant variance which will be assumed to be one. A judicious choice of the $z_i$ from some set $Z$ to minimize the variance of the least squares estimate of $c'\theta$ is the main objective at this point. An approximate design is written as

$$\xi = \left( z_1, \ldots, z_r \right), \quad (2.2)$$
so that if \( n \) observations are allowed \( np_i \) are chosen at \( z_i \). All of the results to follow are concerned only with approximate designs. The normalized information matrix of the design is

\[
M(\xi) = \sum_{i=1}^{r} p_i f(z_i) f'(z_i)
\]  
(2.3)

and the variance of the LSE of \( c'\theta \) is equal to \( c'M^{-1}(\xi)c/n \), where \( M^{-1}(\xi) \) is any generalized inverse of \( M(\xi) \). We assume that any design used will be such that \( c'\theta \) is estimable. Rather than perturbating \( \xi \) and arriving at an ‘equivalence theorem’ giving a condition on \( \xi \) for it to be optimal Elfving proceeds with a more elementary approach as follows.

Let \( \bar{Y}_i \) denote the average of the observations taken at \( z_i \) and let \( \sum_{i=1}^{r} l_i \bar{Y}_i = l'\bar{Y} \) be a linear estimate of \( c'\theta \). The unbiasedness of the estimate result in

\[
c = \sum_{i=1}^{r} f(z_i) l_i
\]  
(2.4)

and the variance is

\[
\frac{1}{n} \sum_{i=1}^{r} \frac{l_i^2}{p_i}.
\]  
(2.5)

From Schwartz inequality

\[
\sum_{i=1}^{r} \frac{l_i^2}{p_i} \geq \left( \sum_{i=1}^{r} |l_i| \right)^2
\]  
(2.6)

and from (2.4)

\[
c \sum_{i=1}^{r} |l_i| = \frac{\sum_{i=1}^{r} |l_i| \varepsilon_i f(z_i)}{\sum_{i=1}^{r} |l_i|} = \sum_{i=1}^{r} p_i \varepsilon_i f(z_i),
\]  
(2.7)

where \( \varepsilon_i = \pm 1 \) and \( p_i = |l_i|/\sum_{j=1}^{r} |l_j| \). The right-hand side of (2.7) is contained in the Elfving set

\[
\partial \mathcal{E} = \text{convex-hull}[ \pm f(z) | z \in Z].
\]  
(2.8)

If \( \rho \) is such that \( \rho c \) is on the boundary of \( \partial \mathcal{E} \), then

\[
\sum_{i=1}^{r} \frac{l_i^2}{p_i} \geq \left( \sum_{i=1}^{r} |l_i| \right)^2 \geq \frac{1}{\rho^2}.
\]  
(2.9)

Now equality can be achieved in both places if and only if we choose \( z_i, p_i \) and \( \varepsilon_i \) in (2.7) to produce \( \rho c \) and then take \( l_i = \varepsilon_i p_i / \rho \).

**Elfving’s Theorem.** The discrete design \( \{z_i, p_i\} \) is optimal for estimating \( c'\theta \) if and only if there exist \( \varepsilon_i = \pm 1 \) and \( \rho \) such that

\[
\rho c = \sum_{i=1}^{r} p_i \varepsilon_i f(z_i) \varepsilon_i \partial \mathcal{E}.
\]  
(2.10)
Moreover, $p^{-2} = c'M^- (\xi)c$ is the minimum variance.

The reader can try out the theory by considering a simple linear regression $EY_i = \theta_0 + \theta_1 z_i$ with observations $z_i \in [0, 1]$ and finding the optimal design of estimating $\theta_0 + \theta_1 z_0$ for $z_0 > 0$, in which case we are extrapolating to $z_0$.

3. Approximation theory

It is relatively well known that the minimization of $c'M^- (\xi)c$ with respect to the design $\xi$ is equivalent to a problem in the theory of approximation of functions. Thus, $c'M^- (\xi)c = \max_d \frac{(d'c)^2}{d'M(\xi)d}$ = \max_d \frac{(d'c)^2}{\int (d'(f(z)))^2 d\xi} \tag{3.1}$

and

$\min_\xi c'M^- (\xi)c = \frac{1}{\min_d \max_z (d'(f(z)))^2}, \tag{3.2}$

where the min is now subject to $d'c = 1$. The interchange of min and max should be given some justification.

In case $c' = (0, \ldots, 0, 1)$ we are approximating the last function $f_p$ by a linear combination of the rest. Considering this case is actually no loss of generality since for the more general $c'\theta$ one can reparameterize by using $\delta = K\theta$ for some nonsingular $k \times k$ matrix $K$ with last row equal to $c'$. The value of $\rho$ in Eq. (2.10) is the value of the best approximation, using a sup norm, of $f_p$ by a linear combination of the other functions.

On p. 178 of Singer (1970), Theorem 1.3 (2), we find the following results. Let $E = C(Z)$ denote the linear space of continuous functions on some compact space $Z$ with sup-norm, $G = \{g_1, \ldots, g_{p-1}\}$ be a $(p - 1)$ dimension subspace and $h$ be some function not in $G$, then, the function $g_0 \in G$ is the best approximation to $h$, if and only if there exists $r$ points, $z_1, \ldots, z_r$ ($1 \leq r \leq p$) and $r$ numbers $\mu_1, \ldots, \mu_r \neq 0$ with $\sum |\mu_i| = 1$ such that

$$\sum_j \mu_j g_i(z_j) = 0, \quad i = 1, \ldots, p - 1,$$

$$\sum_j \mu_j (h(z_j) - g_0(z_j)) = |h - g_0|. \tag{3.3}$$

The resemblance to Elfving’s Theorem is clear. One half of the proof is not difficult. If conditions (3.3) hold then for any $g \in G$

$$|h - g_0| = \sum \mu_i (h(z_i) - g(z_i)) \leq \sum |\mu_i| |h - g| = |h - g|.$$

The other direction is more interesting and one way to derive it is by finding the $\rho$ so that $\rho(0, \ldots, 0, 1)$ is a boundary point of the Elfving set with the functions $g_1, \ldots, g_{p-1}, h$. 
In Singer, of course, no mention is made of Elfving and it is not used to prove the result. Singer considers his Theorem 3.1 as a special case of a more general result, which also appears in most books on functional analysis. This result says that if \( E \) is a normed linear space and \( G \) is a linear subspace then \( g_0 \) is a best approximation to some \( h \) in \( E \), if and only if there exists a linear functional in the dual \( E^* \) such that

\[
|l| = 1, \quad l(g) = 0 \quad \text{for all } g \in G \quad \text{and} \quad l(h) = |h - g_0|.
\]

This result is geometrically evident, but in full generality is a consequence of the Hahn–Banach Theorem. To obtain Theorem 3.1 of Singer one has to identify the dual \( E^* \) for the space of continuous functions \( C(Z) \) and show that the linear functional needed here is given by

\[
l(e) = \sum \mu_i e(z_i),
\]

that is, \( l \) is a linear combination of the point functionals \( l_i(e) = e(z_i) \), with \( \sum |\mu_i| = 1 \). The point functionals are extremal functionals amongst those of norm one, and the condition \( \sum_i |\mu_i| = 1 \) is required to give \( |l| = 1 \).

4. A simple extension

In Studden (1971) a number of strange contortions were used to extend the simple Elfving result to the situation where one is interested in minimizing the sum of variances for unbiased estimates of \( c'_i\theta, \ i = 1, \ldots, s \). This becomes somewhat trivial if Elfving’s original idea is followed. In fact, with some slight change of notation the outline in Section 2 can be essentially copied.

If \( l'_i\bar{Y} \) is an unbiased estimate of \( c'_i\theta \), then, as before \( c_i = \sum_j f(z_j)l_{ij} \) or

\[
C = \sum_j f(z_j)l'_j,
\]

(4.1)

where \( C \) is the matrix with columns \( c_1, \ldots, c_s \) and \( l'_j = (l_{1j}, \ldots, l_{sj}) \). Then the variance of \( l'_i\bar{Y} \) is \( \sum_j l_{ij}^2/p_j \) and the sum is

\[
\sum_i \left( \sum_j \frac{l_{ij}^2}{p_j} \right) = \sum_j \frac{|l_j|^2}{p_j}, \quad \left( |l_j|^2 = \sum_i l_{ij}^2 \right)
\]

\[
\geq \left( \sum_j |l_j| \right)^2.
\]
We can now proceed as before and rewrite (2.7) in the form

$$\frac{C}{\sum_i |l_i|} = \frac{\sum_i |l_i| f(z_i)\varepsilon'_i}{\sum_i |l_i|} = \sum_i p_i f(z_i)\varepsilon'_i,$$

where now $\varepsilon'_i = l'_i/|l_i|$ is an $s$-vector of length one ($\varepsilon'_i\varepsilon_i = 1$) and $p_j = |l_j|/\sum_k |l_k|$.

It is now apparent that we should define an extended Elfving set of $p \times s$ matrices

$$\mathcal{E}_s = \text{convex-hull}\{f(z)\varepsilon'|z \in Z, \varepsilon \in \mathbb{R}^s, \varepsilon'\varepsilon = 1\}.$$ 

It then follows that $\sum_j |l_j| \geq \rho$ where $\rho$ is such that $\rho C$ is on the boundary of $\mathcal{E}_s$. Thus the sum of variances is minimized when

$$\rho = \sum_i p_i f(z_i)\varepsilon'_i \varepsilon \hat{\varepsilon}\mathcal{E}_s;$$

the optimal design and estimate are determined as before, and $\rho^{-2}$ gives the minimum sum of variances.

5. Dette’s Theorem

Dette (1993) extended Elfving’s result in a more interesting manner by considering $s$ different models with means

$$f_j'(z)\theta_j = \sum_k f_{jk}(z)\theta_{jk}$$

having the same design space $Z$. (Note here that $f_j(z)$ and $\theta_j$ are vectors of length $k_j$, i.e. $f_j'(z) = (f_{j1}(z), \ldots, f_{jk_j}(z))$ and $\theta_j' = (\theta_{j1}, \ldots, \theta_{jk_j})$.)

For the $j$th model $c'_j\theta_j$ is to be estimated. The criterion here is changed to

$$V(\xi) = \sum_j \beta_j \ln(\text{Var}(c'_j\hat{\theta}_j(\xi))), \quad \beta_j > 0, \quad \sum_j \beta_j = 1, \quad (5.1)$$

which allows for investigating various models and robustness questions and also produces an Elfving-type result for D-optimality. The D-optimality comes about by starting with one model, say, $g'(z)\hat{\theta}$, and then letting $\beta_j = 1/m$,

$$f_1' = (g_1),$$
$$f_2' = (g_1, g_2),$$
$$\vdots$$
$$f_m' = (g_1, g_2, \ldots, g_m)$$

and $c'_i = (0, \ldots, 0, 1)$. If $M_i(\xi) = \int f_i(z)f_i'(z)\,d\xi(z), \quad i = 1, \ldots, m$ then the optimally criterion easily reduces to the maximization of the determinant of the information matrix of $M_m(\xi)$ since $c_i M_i^{-\dagger}(\xi)c_i = |M_i|/|M_i|$. 

As before, the design \( \{z_j, p_j\} \) must be such that \( c_i' \theta_i \) is estimable with respect to the \( i \)th model so that there are \( l_{ij} \) with

\[
c_i = \sum_j f_i(z_j)l_{ij} \quad \text{and} \quad c_i' M_i^{-1}(\zeta)c_i = \sum_j l_{ij}^2 / p_j = \rho_i^{-2}. \tag{5.2}
\]

Note that \( \sum_i \beta_i \ln(\rho_i^{-2}) \) is minimized over \( p_j \) by choosing

\[
p_j = \sqrt{\sum_i \beta_i l_{ij}^2 \rho_i^2} \tag{5.3}
\]

(first show that \( p_j \) is proportional to the right side and then show the sum over the left side is one using Schwartz inequality). The \( p_j \) are not defined explicitly by the above since \( \rho_i^2 \) depends on the \( p_j \), however, proceeding as before we write

\[
\rho_i c_i = \sum_j p_j f_i(z_j) \varepsilon_{ij} \quad \text{where} \quad \varepsilon_{ij} = \frac{l_{ij} \rho_i}{\sqrt{\sum_i \beta_i l_{ij}^2 \rho_i^2}}. \tag{5.4}
\]

Note that \( \sum_i \beta_i \varepsilon_{ij}^2 = 1 \), in which case the corresponding Elfving set is

\[
\mathcal{E}(\beta) = \text{convex-hull} \left\{ (\varepsilon_1 f_1'(z), \ldots, \varepsilon_m f_m'(z) | z \in Z \text{ and } \sum_i \beta_i \varepsilon_{ij}^2 = 1 \right\}. \tag{5.5}
\]

The Elfving set is now written as one long vector since the models with \( f_i(z) \) are allowed to be of different dimensions. Hyperplanes bounding this Elfving set on one side are denoted by \( (a_1', a_2', \ldots, a_m', 1) \) so that

\[
\sum_i a_i' f_i(z) \varepsilon_i \leq 1 \quad \text{for all} \quad z, \quad \text{whenever} \quad \sum_i \beta_i \varepsilon_{ij}^2 = 1. \tag{5.6}
\]

Dette’s Theorem. The design \( \zeta = \{z_i, p_i\} \) is \( V(\zeta) \) optimal if and only if there exists \( \gamma_i, \varepsilon_{ij} \) with \( \sum_i \beta_i \varepsilon_{ij}^2 = 1 \) such that

\[
\gamma_i c_i = \sum_j p_j f_i(z_j) \varepsilon_{ij}, \quad i = 1, \ldots, m, \quad (\gamma_1 c_1', \ldots, \gamma_m c_m') \in \mathcal{E}(\beta) \tag{5.7}
\]

and for the hyperplane \( (a_1', a_2', \ldots, a_m', 1) \) supporting \( \mathcal{E}(\beta) \) at \( (\gamma_1 c_1', \ldots, \gamma_m c_m') \)

\[
\gamma_i a_i' c_i = \beta_i. \tag{5.8}
\]

(It is shown further that \( \gamma_i^{-2} = c_i' M_i(\zeta)c_i. \))

Proof. Dette uses the equivalence theorem conditions in (6.6) below. Our purpose here is to derive the result directly following Elfving and produce the equivalence theorem as a consequence. Suppose that the conditions in (5.7) and (5.8) hold. If we consider any other
competing design then following (5.2) and (5.4) we have that \((\rho_1 c_1', \ldots, \rho_m c_m')\) is in \(\mathcal{E}(\beta)\) implying that
\[
\sum_i \rho_i a_i' c_i \leq 1.
\]
This implies from (5.8) that
\[
\sum_i \rho_i \frac{\beta_i}{\gamma_i} \leq 1,
\]
and hence
\[
\prod_i \left( \frac{\rho_i}{\gamma_i} \right)^{\beta_i} \leq \sum_i \rho_i \frac{\beta_i}{\gamma_i} \leq 1
\]
implying that
\[
\prod \left( \frac{1}{\gamma_i^{\beta_i}} \right) \leq \prod \left( \frac{1}{\rho_i^2} \right)^{\beta_i}.
\]
Thus the conditions in the theorem imply optimality.

Suppose now that \(\bar{\zeta} = \{z_j, p_j\}\) is optimal. Then there are optimal estimates \(l_i \bar{Y}\) with variances \(\gamma_i^{-2} = \sum j l_{ij}^2/p_j\). We will show that (5.7) and (5.8) now hold with these \(\gamma_i\). To begin it is easy to show that if \(\bar{\zeta} = \{z_j, p_j\}\) is optimal then \((\gamma_1 c_1', \ldots, \gamma_m c_m')\) must be a boundary point of \(\mathcal{E}(\beta)\). Then following Dette (1993, p. 758) one has the corresponding hyperplane \((a_1', \ldots, a_m', 1)\) for which
\[
a_i = \beta_i \gamma_i M_i^{-} (\bar{\zeta}) c_i
\]
for some generalized inverses \(M_i^{-} (\bar{\zeta})\). In this case
\[
\gamma_i a_i' c_i = \beta_i \gamma_i^2 c_i' M_i^{-} (\bar{\zeta}) c_i = \beta_i,
\]
since we have chosen \(\gamma_i^{-2} = c_i' M_i^{-} (\bar{\zeta}) c_i\). \(\square\)

6. Equivalence Theorems

An excellent detailed analysis of equivalence theorems for optimal designs is given in Pukelsheim (1993). Here we indicate that in simple situations the equivalence theorem will follow once one has an Elfving-type theorem. Comparing the equivalence theorem for D-optimality from Dette’s result with the usual equivalence theorem for this case produces an interesting (well known) identity for quadratic forms.

One of the simplest equivalence theorems is for estimating the single linear form \(c' \theta\) described in Section 2. This states that the design \(\bar{\zeta}\) is optimal for estimating \(c' \theta\) if and only if there exists some generalized inverse \(M^{-} (\bar{\zeta})\) such that
\[
\frac{(c' M^{-} (\bar{\zeta}) f(z))^2}{c' M^{-} (\bar{\zeta}) c} \leq 1 \quad \text{for all } z \in Z.
\]

(6.1)
This is related to Elfving’s Theorem by the fact that the quantity
\[ a' = \frac{c'M^{-}(\xi)}{\sqrt{c'M^{-}(\xi)c}} \]
describes the support plane \((a,1)\) to the Elfving set \(\mathcal{E}\) at the boundary point \(\rho c\). Thus if
\[ \rho c = \sum p_i \varepsilon_i f(z_i) \varepsilon \hat{\mathcal{E}}, \tag{6.2} \]
then there exists a support plane \((a,1)\) such that
\[ a' \varepsilon f(z) \leq 1 \text{ for all } \varepsilon = \pm 1 \text{ and } z \in \mathcal{Z} \text{ and } a'(\rho c) = 1 \tag{6.3} \]
It is easily seen that the first part of (6.3) holds if and only if
\[ (a' f(z))^2 \leq 1 \text{ for all } z \in \mathcal{Z}. \tag{6.4} \]
If (6.3) holds then necessarily
\[ a' \varepsilon_i f(z_i) = 1 \text{ or } \varepsilon_i = a' f(z_i). \]
If these \(\varepsilon_i\) are inserted into (6.2) the
\[ \rho c = M(\xi) a. \tag{6.5} \]
In this case there exists some inverse \(M^{-}(\xi)\) such that \(a = \rho M^{-}(\xi)c\) and inserting this into (6.4) produces the equivalence condition (6.1).

The main problem in finding the optimal design is to find boundary points \(\varepsilon_i f(z_i)\) of \(\mathcal{E}\) so that \(c\) is in the convex cone generated by these points.

We turn now to the equivalence theorem corresponding to Dette’s result. This states that in the setting of Section 5 the design \(\rho c\) is \(V(\xi)\) optimal if and only if there exist inverses \(M_i^{-}(\xi)\) such that
\[ \sum \beta_i \left( c_i' M_i^{-}(\xi) f_i(z) \right)^2 \leq 1 \text{ for all } z \in \mathcal{Z}. \tag{6.6} \]
This can be obtained in a manner similar to the above using the support plan \((a'_1, \ldots, a'_m, 1)\) and the fact that (5.4.1) is equivalent to
\[ \sum_i (a'_i f_i(z))^2 \leq 1, \text{ whenever } \sum_i \beta_i \varepsilon_i^2 = 1. \tag{6.7} \]
A rather interesting result then emerges for the case of D-optimality described at the beginning of Section 5. Recall that for this situation \(c_i\) is the vector \((0, \ldots, 0, 1)\) of length \(i\) and \(f_i = (g_1, \ldots, g_i)\). If one has only one model with regression functions \(g = (g_1, \ldots, g_m)\) then the design \(\xi\) is D-optimal if and only if maximizes the determinant of \(M_g(\xi)\). The well-known equivalence theorem in this case, says that \(\xi\) is D-optimal if and only if
\[ g'(z) M_g^{-1}(\xi) g(z) \leq m \text{ for all } z \in \mathcal{Z}. \tag{6.8} \]
The question arises as to whether (6.8) is equivalent to (6.6) with the $c_i$ and $f_i$ as described above and $\beta_i = 1/m$. Thus we have

$$\sum_i \frac{(c_i' M_i^{-1}(\xi) f_i(z))^2}{c_i' M_i^{-1}(\xi)c_i} = g'(z)M_g^{-1}(\xi)g(z).$$

That these are actually the same (as they must be) follows from taking an LU decomposition of $M_g(\xi) = LU$ with lower and upper triangular matrices $L$ and $U$. The details are left to the reader.

7. For further reading

The following references could also be of interest to the reader. Chernoff (1962, 1999); Fellman (1999); Nordström (1999).

References