Optimal design for linear forms of the parameters in a Chebyshev regression

Michel Broniatowski, Giorgio Celant

Pub. Inst. Stat. Univ. Paris
60, fasc.3, 2016, 15-54

Optimal design for linear forms of the parameters in a Chebyshev regression

Michel Broniatowski, Giorgio Celant
LSTA, Université Pierre et Marie Curie, Paris, France
Dipartmento di Scienze Statistiche, Università degli Studi di Padova, Italy

January 24, 2017

Abstract

This paper considers the optimal design for the estimation of a linear form of the parameters of a regression model, when the regressor belong to a finite dimensional Haar linear space . A major interest is paid to the geometric approach, following Elfving's theorem. The classical Hoel-Levine design follows as a special case of a result due to Karlin and Studden.

1 Introduction and notation

This paper of pedagogical nature presents the derivation of designs in the context of the Chebyshev regression; specifically the focus is set on the designs leading to optimal estimation of a linear form of the coefficients in a Chebyshev regression model. For references on general notions on optimal designs for extrapolation and interpolation, we refer to [Celant and Broniatowski, 2016] ; for Chebyshev systems, we refer to the paper [Broniatowski and Celant, 2015]; the notation is inherited from those used in this paper, without further definition.

Consider a Chebyshev regression model defined by

$$Y_i(x_j) := (X(x_j))' \theta + \varepsilon_{i,j},\ i = 1, ..., n_j,\ j = 1, ..., r \tag{1}$$

where

$$\theta' := (\theta_0, ..., \theta_{g-1}) \in \mathbb{R}^g$$

$$E(\varepsilon_{i,j}) = 0, \text{ for all } (i,j)$$

$$cov(\varepsilon_{i,j}, \varepsilon_{i',j'}) = \begin{cases} \sigma^2 > 0 \text{ for } (i,j) = (i',\ j') \\ 0 \qquad\quad \text{for } (i,j) \neq (i',\ j') \end{cases}$$

and

$$X(x)' := (\varphi_0(x), ..., \varphi_{g-1}(x)), x \in [-1, 1]$$

where $\{\varphi_0, ..., \varphi_{g-1}\}$ is a Chebyshev system of functions defined on the observable domain $[-1, 1]$.

We make use of the usual notation for the observations. Denote the nodes taking into account their multiplicity, namely for each of them the number of replications of the experiment to be performed. On the node x_j let n_j be the number of replications. The discrete measure characterizing a design is therefore described as

$$\overbrace{x_0, ..., x_0}\ (n_0 \text{ times}), ..., \overbrace{x_j, ..., x_j}\ (n_j \text{ times}), ..., \overbrace{x_{g-1}, ..., x_{g-1}}(n_{g-1} \text{ times})$$

with

$$n_0 + ... + n_{g-1} = n.$$

This design can also be written by

$$t_1, ..., t_n$$

where $t_1, ..., t_{n_0}$ describe the n_0 equal values of x_0, and so on.

Taking into account the t_i's in place of the x_j's, the data set can be written through

$$Y = T\,\theta + \varepsilon \tag{2}$$

where

$$Y := \begin{pmatrix} y_1 \\ . \\ y_n \end{pmatrix}, T := \begin{pmatrix} \varphi_0(t_1) & ... & \varphi_{g-1}(t_1) \\ .. & ... & ... \\ \varphi_0(t_i) & ... & \varphi_{g-1}(t_i) \\ ... & .. & ... \\ \varphi_0(t_n) & ... & \varphi_{g-1}(t_n) \end{pmatrix},$$

$$\theta := \begin{pmatrix} \theta_0 \\ . \\ \theta_{g-1} \end{pmatrix}, \varepsilon := \begin{pmatrix} \varepsilon_1 \\ . \\ \varepsilon_n \end{pmatrix},$$

$$E(Y) = T\,\theta,\ var(\varepsilon) = \sigma^2 I_n,$$

and I_n is the identity matrix of order n. The vector of parameters θ belongs to $\mathbb{R}^g$ and is unknown; σ^2 is assumed non null.

A generic design ξ is defined through

$$\xi(x_j) = \frac{n_j}{n},\ j = 0, ..., g-1,\ \sum_{j=0}^{g-1} n_j = n,\ x_j \in [-1, 1].$$

A design therefore is defined as an element of $\mathcal{M}_d([-1, 1])$, the class of all probability measures with finite support on $[-1, 1]$. The generalization with repect to interpolation or extrapolation optimal designs lies in the estimation of a generic linear form of θ; in [Broniatowski and Celant, 2015] the estimation pertained to

$$X(x)'\ \theta = < X(x), \theta >$$

which amounts to estimate an interpolated value whenever $|x| \leq 1$, and an extrapolated one when $|x| > 1$.

Hence we consider a form given by $< \mathbf{c}, \theta >$ where the known vector $\mathbf{c}$ in $\mathbb{R}^g$ defines the linear form. The aim of the present paper lies in the quest for the optimal design for the estimation of this form, minimizing its variance among all unbiased linear estimators. This yields

$$\widehat{< \mathbf{c}, \theta >} := < \mathbf{c}, \widehat{\theta} >$$

where $\widehat{\theta}$ is the least square estimator of θ.

Example 1 *Let*

$$y(x) = \theta_0 \varphi_0(x) + ... + \theta_{g-1} \varphi_{g-1}(x) + \varepsilon(x),$$

where $\{\varphi_0, ..., \varphi_{g-1}\}$ *is a Chebyshev system on* $\mathbb{R}$, *and* $y(x)$ *is observable for* $x \in [-1, 1]$. *Let* $|u| \notin [-1, 1]$. *With*

$$\theta = (\theta_0, ..., \theta_g)'$$

and

Example 2

$$\mathbf{c} := (\varphi_0, ..., \varphi_{g-1})'$$

the inner product

$$< \mathbf{c}', \theta > = \sum_{j=0}^{g} \theta_j \varphi_j(u)$$

is a linear form which evaluates the extrapolated value at point u.

Let

$$Y = (Y(t_1), ..., Y(t_n))'$$

denote the observed sample. An equvalent writing for (2) is

$$Y = \begin{pmatrix} X(t_1) \\ . \\ X(t_n) \end{pmatrix} \theta + \varepsilon$$

denotes the model (1) where ε is the vector of errors,

$$\varepsilon' := (\varepsilon_1, ..., \varepsilon_n)$$

and the least square estimator of θ is

$$\widehat{\theta} := \left(\begin{pmatrix} X(t_1) \\ . \\ X(t_n) \end{pmatrix}' \begin{pmatrix} X(t_1) \\ . \\ X(t_n) \end{pmatrix} \right)^{-} \begin{pmatrix} X(t_1) \\ . \\ X(t_n) \end{pmatrix}' Y.$$

The measurements may be written grouping the replications of the measurements on the r nodes of the design, and denoting $n_1, .., n_r$ the number of replicates distinct nodes $(x_1, .., x_r)$. The *moment matrix* is defined by

$$M(\xi) := \sum_{j=1}^{r} \frac{n_j}{n} X(x_j) X(x_j)' .$$

Here ξ is the design, defined by

$$\xi(x_j) = \frac{n_j}{n}$$

for $1 \leq j \leq r$.

The contents of the paper is as follows. We first study the properties of the moment matrices associated to designs. Then we introduce the notion of estimable linear forms; those are forms which admit an unbiased linear estimator. The geometry of the class of moment matrices is the frame in which the estimable forms are best handled. Then we introduce the Elfving set of vectors which provides a link between estimable linear forms of the parameters and the corresponding optimal designs. Finally a Theorem by Studden and Karlin provides a complete characterization of the optimal designs; at this point some use is made of the theory of best uniform approximation of functions. It also leads to an effective way to obtain this design. We will consider optimal designs following Kiefer, Wolfowitz and Studden, mostly using the approach of Karlin and Studden [Karlin and Studden, 1966a], which makes use of a Theorem due to Elfving [Elfving, 1952] , [Pukelsheim, 2006] and [Studden, 2005]. The approach by Kiefer and Wolfowitz [Kiefer and Wolfowitz, 1964] makes use of some arguments from game theory instead.

This paper is based on [Elfving, 1954], [Hoel, 1966], [Karlin and Studden, 1966 [Pukelsheim, 2006], [Studden, 1971], [Studden, 2005], [Karlin and Studden, 1966a

2 Matrix of moments

To any $\xi \in \mathcal{M}_d([-1,1])$ we may associate the moment matrix, defining

$$M(\xi) := \sum_{i=0}^{r-1} \xi_i \, X(x_i) X'(x_i)$$

where $\xi_0, .., \xi_{g-1}$ denote the values of the measure ξ on its support $x_0, ..., x_{r-1}$.

Since $X(x) X'(x)$ is symmetric, so is $M(\xi)$ per ξ in $\mathcal{M}_d([-1,1])$. The class of all symmetric matrices of order g is denoted $\mathcal{S}(g)$ and it holds

$$\mathbf{M} \subset \mathcal{S}(g) .$$

Some special subsets in $\mathcal{S}(g)$ deserve interest for the sequel.

$$\mathcal{S}_{\geq 0}(g) := \{A \in \mathcal{S}(g) \text{ such that } x'Ax \geq 0 \text{ for any } x \in \mathbb{R}^g\},$$

$$\mathcal{S}^{+}(g) := \{A \in \mathcal{S}(g) \text{ such that } x'Ax > 0 \text{ for any } x \in \mathbb{R}^{g} \setminus \{0\}\}.$$

The set $\mathcal{S}_{\geq 0}(g)$ is the class of all symmetric positive semidefinite matrices of order g whereas the second class $\mathcal{S}^{+}(g)$ is the class of all symmetric definite positive matrices of order g.

We will also make use of $\mathcal{S}^{-}(g)$, $\mathcal{S}_{\leq 0}(g)$ with clear interpretation.

The classes $\mathcal{S}_{\geq 0}(g)$ and $\mathcal{S}^{+}(g)$ induce a partial ordering in $\mathcal{S}(g)$, the so-called *Loewner ordering*, defining

$$A \succsim B \iff A - B \in \mathcal{S}_{\geq 0}(g),$$

$$A \succ B \iff A - B \in \mathcal{S}^{+}(g).$$

It is customary to write $A - B \succsim 0$ in the first case and $A - B \succ 0$ in the second one. Loewner ordering enjoys a number of properties.

Let $\alpha \in \mathbb{R}^{+}$, and A, B, A_n belong to $\mathcal{S}_{\geq 0}(g)$, assuming that $\lim_{n\to\infty} A_n$ exists. Then

$$\alpha A \succsim 0,\ A + B \succsim 0,\ \lim_{n\to\infty} A_n \succsim 0;$$

see [Pukelsheim, 2006].

The mapping

$$tr : (\mathcal{S}(g))^{2} \to \mathbb{R},\ (A, B) \mapsto\ < A, B > := tr(A'B) = tr(AB)$$

defines an inner product on $\mathcal{S}(g)$ from which derive the norm and the distance

$$\|.\|_{tr} : \mathcal{S}(g) \to \mathbb{R}_{\geq 0},\ A \mapsto \|A\|_{tr} := \sqrt{tr(A^{2})},$$

$$dist(A, B) := \|A - B\| = \sqrt{tr\left((A - B)^{2}\right)}.$$

The closed ball with radius 1 and center 0 is defined by

$$\overline{S}_{tr}(0, 1) := \{A \in \mathcal{S}(g) : \|A\|_{tr} \leq 1\}. \tag{3}$$

The geometric structure of $\mathcal{S}_{\geq 0}(g)$ is described through the following Theorem.

Theorem 3 *(Pukelsheim p 29) $\mathcal{S}_{\geq 0}(g)$ is a convex closed cone. It is pointed. Furthermore $\mathcal{S}^{+}(g)$ is the relative interior of $\mathcal{S}_{\geq 0}(g)$ in $\mathcal{S}(g)$, which is denoted $int(\mathcal{S}_{\geq 0}(g))$.*

As a subset of $\mathcal{S}(g)$, the set $\mathbf{M}(\xi)$ enjoys specific geometric properties, as seen now.

Associated with the Chebyshev system of regressors $X(x)$ as x belongs to $[-1, 1]$ we define

$$\mathcal{H} := \{X(x) \in \mathbb{R}^{g} : x \in [-1, 1]\}. \tag{4}$$

We also define the *regression range* as the linear space generated by $\mathcal{H}$.

Theorem 4 $\mathbf{M} \subset \mathcal{S}(g)$ *is a compact and convex subset of* $\mathcal{S}_{\geq 0}(g)$.

Proof. See [Pukelsheim, 2006] p29. ■

3 Estimable forms

We introduce a linear form $\mathbf{c}$ in $\mathbb{R}^g$ which is identified with the vector $(c_0, .., c_{g-1})$ through

$$\theta \mapsto \sum_{j=0}^{g-1} c_j \theta_j :=< \mathbf{c}, \theta >, \tag{5}$$

assuming that $\mathbf{c} \neq \mathbf{0}$.

In order to emphasize the vector $\mathbf{c}$ of the coefficients of the parametric function $\sum_{j=0}^{g-1} c_j \theta_j$, we call a $\mathbf{c}$ - form the linear form $\theta \rightarrow \sum_{j=0}^{g-1} c_j \theta_j :=< \mathbf{c}, \theta >$.

Definition 5 *Let* $\mathbf{c} \in \mathbb{R}^g$ *and identify* $\mathbf{c}$ *with the linear form (5) where* $\mathbf{c} \in \mathbb{R}^g$; *we say that the linear form* $\mathbf{c}$ *is estimable if and only if there exists some unbiased linear estimator of* $< \mathbf{c}, \theta >$ *for* θ *defined in (1).*

Example 6 *Consider* $\mathbf{c} := (1, 0, .., 0)$ *; therefore* $< \mathbf{c}, \theta > = \theta_0$*; the linear form* $\mathbf{c}$ *is estimable if we can define an unbiased linear estimator of* θ_0.

From this definition $\theta \rightarrow < \mathbf{c}, \theta >$ is estimable if and only if there exists $\mathbf{u}' := (u_1, ..., u_n) \in \mathbb{R}^n$, such that $\widehat{< \mathbf{c}, \theta >} =< \mathbf{u}', Y >$ and

$$E(\mathbf{u}'Y) :=< \mathbf{c}, \theta >, \text{ for all } \theta \in \mathbb{R}^g,$$

which amounts to

$$\mathbf{u}'T\ \theta = \mathbf{c}'\theta, \text{ for all } \theta \in \mathbb{R}^g$$

which entails

$$\mathbf{u}'T\ = \mathbf{c}'.$$

Finally we see that a linear form is estimable if and only if

$$\mathbf{c} \in \operatorname{Im} T'. \tag{6}$$

In the next paragraph we obtain the explicit expression for the Gauss Markov estimator of a linear form.

Example 7 *A localization problem.*

$$Y = \theta + \varepsilon,\ \theta \in \mathbb{R},\ cov(\varepsilon, \varepsilon') = \begin{cases} \sigma^2 & \text{if } \varepsilon = \varepsilon' \\ 0 & \text{if } \varepsilon \neq \varepsilon' \end{cases}.$$

The least square estimator of θ *when a sample of* n *i.i.d. observations is at hand is given by*

$$\widehat{\theta} := \mathbf{u}'Y$$

where

$$Y := (Y_1, ..., Y_n)' \quad \text{and} \quad \mathbf{u}' := \left(\frac{1}{n}, ..., \frac{1}{n}\right)$$

Example 8 *A standard regression model*

$$Y = X\theta + \varepsilon,\ \varepsilon \text{ i.i.d.},\ \theta \in \mathbb{R}^k.$$

Example 9 *The least square estimator of* θ *under i.i.d. sampling is*

$$\widehat{\theta} := \mathbf{u}'Y$$

where

$$Y := (Y_1, ..., Y_n)' \quad \textit{and} \quad \mathbf{u}' := (\mathbf{X}'\mathbf{X})^{-1} X'$$

4 Matrix of moments and Gauss-Markov estimators of a linear form

4.1 Matrices of moments and estimable linear forms

We discuss the link which connects the matrix of moments and the estimable linear forms. The variance of the Gauss-Markov estimators of an estimable linear form will be derived.

Making use of the notation in Section 1, the matrixhe moment matrix takes the form

$$\begin{aligned} M(\xi) \quad &: \quad = \frac{1}{n} T'T = \sum_{x_j \in \operatorname{supp}(\xi)} \xi(x_j) \left(X(x_j)\right)' X(x_j) \\ &= \int_{[-1,1]} \left(X(x)\right)' X(x)\, \xi(dx). \end{aligned}$$

We first put forward an important property of estimable $\mathbf{c}-$ forms (see (6)), which is of algebraic nature.

Proposition 10 *A* $\mathbf{c}-$ *form is estimable if and only if*

$$\mathbf{c} \in \operatorname{Im} T'$$

if and only if

$$\mathbf{c} \in \operatorname{Im} M(\xi). \tag{7}$$

Proof. We first state that

$$\operatorname{Im} T' = \operatorname{Im}(T'T).$$

Given a matrix A in $\mathcal{M}_{p\times k}$ (the class of all matrices with p rows and k columns), and identifying such A with the corresponding linear mapping, it holds

$$Ker(A') = (\operatorname{Im} A)^{\perp}. \tag{8}$$

Now

$$Ker\,(T\,'T) = Ker\,(T)\,. \tag{9}$$

Letting in (8) $A = T\,'T$ we obtain

$$Ker\,(T\,'T) = (\mathrm{Im}\,(T\,'T))^{\perp}\,. \tag{10}$$

Now, letting in (8) $A = T$ we get

$$Ker\,(T\,') = (\mathrm{Im}\,T)^{\perp}\,. \tag{11}$$

Those two latest results , together with (9) yield

$$(\mathrm{Im}\,(T\,'T))^{\perp} = (\mathrm{Im}\,T)^{\perp}\,. \tag{12}$$

Taking the orthogonals on both sides in (12) it follows that

$$\mathrm{Im}\,(T\,'T) = \mathrm{Im}\,T.$$

■

It is common use to say that the **c** - form is estimable with respect to the measure ξ. In the sequel we will assume that the **c** - form is estimable w.r.t. ξ.

Solving the system of linear normal equations pertaining to (2)

$$M\,(\xi)\;\theta = \frac{1}{n}T\,'Y$$

we obtain the least square estimator, say $\widehat{\theta}$, of θ.

Since the matrix $M\,(\xi)$ is invertible (due to the fact that the family $\{\varphi_0, ..., \varphi_{g-1}\}$ is a linearly independent family of functions), it holds

$$\widehat{\theta} := (M\,(\xi))^{-1}\,T\,'Y.$$

A way to estimate the linear form

$$<\mathbf{c},\theta>=c_0\theta_0 + ... + c_{g-1}\theta_{g-1}$$

consists in the plug in of $\widehat{\theta}$ in place of the θ_j's, $j = 0, ..., g-1$, which yields the least square estimator

$$\widehat{<\mathbf{c},\theta>} = c_0\widehat{\theta}_0 + ... + c_{g-1}\widehat{\theta}_{g-1}. \tag{13}$$

Lemma 11 *The above estimator (13) is optimal within all linear unbiased ones, and it is the only estimator enjoying this property.*

Proof. Indeed let $\mathbf{u}'Y$ denote some other linear and unbiased estimator of the **c** - form. Then

$$\mathbf{u}'Y = \mathbf{c}'Y + \mathbf{d}'Y$$

for some $\mathbf{d}$ in $\mathbb{R}^n$. Using unbiasedness

$$E(\mathbf{u}'Y) = \mathbf{c}'\theta.$$

Therefore

$$E(\mathbf{u}'Y) = E(\mathbf{c}'Y + \mathbf{d}'Y) = \mathbf{c}'\theta + \mathbf{d}'E(Y) = \mathbf{c}'\theta + \mathbf{d}'T\,\theta = \mathbf{c}'\,\theta,$$

i.e.

$$\mathbf{d}'T\,\theta = 0,$$

i.e.

$$\mathbf{d}'T = 0.$$

Let us now evaluate the variance of this estimator. It holds

$$\begin{aligned} var(\mathbf{u}'Y) &= var(\mathbf{c}'Y + \mathbf{d}'Y) \\ &= var(\mathbf{c}'Y) + var(\mathbf{d}'Y) + 2cov(\mathbf{c}'Y, \mathbf{d}'Y). \end{aligned}$$

Since

$$\begin{aligned} cov(\mathbf{c}'Y, \mathbf{d}'Y) &= E\left\{(\mathbf{c}'Y - \mathbf{c}'\theta)'(\mathbf{d}'Y - \mathbf{c}'\theta)\right\} \\ &= \sigma^2\mathbf{c}'(M(\xi))^{-1}T'\mathbf{d} = 0, \end{aligned}$$

it follows that

$$var(\mathbf{u}'Y) = var(\mathbf{c}'Y) + \sigma^2\mathbf{d}'\mathbf{d}.$$

This variance reaches its minimal value if and only if

$$\mathbf{d} = \mathbf{0},$$

proving uniqueness. It follows that the Gauss - Markov estimator of the $\mathbf{c}$ - form is

$$\widehat{<\mathbf{c},\theta>} = c_0\widehat{\theta}_0 + ... + c_{g-1}\widehat{\theta}_{g-1}.$$

Furthermore

$$\begin{aligned} var\left(\widehat{<\mathbf{c},\theta>}\right) &= var\left(\mathbf{c}'\widehat{\theta}\right) \\ &= \sigma^2\mathbf{c}'\,(M(\xi))^{-1}\,\mathbf{c}. \end{aligned}$$

■

Once defined the optimal estimator of a $\mathbf{c}$ - form, we intend to characterize the optimal measure pertaining to this estimator.

The variance of the optimal estimator $\widehat{<\mathbf{c},\theta>}$ depends on the matrix $M(\xi)$ induced by the design. Now $M(\xi)$ is a symmetric positive definite matrix of order g. The optimal design defined through the minimization of the variance of $\widehat{<\mathbf{c},\theta>}$ will result from a study of a partial ordering of the symmetric matrices.

Some other form for the variance of the estimator of a linear form can be obtained also when the moment matrix is singular. Indeed the following important result holds. Denote $var(Y)$ the covariance matrix of the vector Y which is a symmetric positive semi definite matrix of order n.

Definition 12 *For a design ξ , let $\mathrm{F}(\xi)$ be the set of all unbiased linear estimators $< \mathbf{u}, Y >$ of the linear form $< \mathbf{c}, \theta >$, where the measure ξ is fixed.*

We will identify any element in $F(\xi)$ with the vector $\mathbf{u}$ in $\mathbb{R}^g$ such that this estimator writes $\mathbf{u}'Y$.

It holds

Proposition 13 *(Karlin and Studden) Let $\xi \in \mathcal{M}_d([-1,1])$. Assume that $\mathbf{supp}(\xi) := \{x_0, ..., x_{g-1}\}$ and denote*

$$\xi(x_i) := \frac{n_i}{n},$$

with $n_i > 0$ for all i. Assuming that $\mathrm{F}(\xi)$ is not voids, and denoting $< \gamma, Y >$ the Gauss-Markov estimator of $< \mathbf{c}, \theta >$, it holds

$$var(< \gamma, Y >) := \min_{\mathrm{F}(\xi)} (var(< \mathbf{u}, Y >)) = \frac{\sigma^2}{n} \sum_{i=1}^{s} \frac{(< \mathbf{v}_i, \mathbf{c} >)^2}{\lambda_i}$$

where $\mathbf{v}_i$ and λ_i are respectively the eigenvectors with norm 1 and eigenvalues of the matrix $M(\xi)$.

Proof. We assume without loss of generality that $\sigma^2 = 1$.

Step 1. We first prove that for any element $\mathbf{u}'Y$ in $F(\xi)$ it holds

$$var(\mathbf{u}'Y) \geq \sup_{0 \neq \mathbf{d} \in (Ker M(\xi))^{\perp}} \frac{1}{n} \frac{(< \mathbf{c}, \mathbf{d} >)^2}{< \mathbf{d}, M(\xi)\mathbf{d} >}.$$

Consider the inner product $< \mathbf{c}, \mathbf{d} >$ with $0 \neq \mathbf{d} \in (Ker M(\xi))^{\perp}$. Since the linear form $< \mathbf{c}, \theta >$ is estimable it holds $\mathbf{c} \in \mathrm{Im}(M(\xi))$ and therefore $\mathbf{c} \in \mathrm{Im} T'$; see Proposition 10.

Hence there exists some vector $\mathbf{u}$ such that $\mathbf{c} = T'\mathbf{u}$. Write henceforth

$$< \mathbf{c}, \mathbf{d} > = < T'\mathbf{u}, \mathbf{d} > = < \mathbf{u}, T\mathbf{d} > .$$

Applying Cauchy - Schwartz Inequality it holds

$$< \mathbf{u}, T\mathbf{d} > \leq \sqrt{< \mathbf{u}, \mathbf{u} >} \sqrt{< T\mathbf{d}, T\mathbf{d} >}.$$

Therefore

$$< \mathbf{c}, \mathbf{d} > \leq \sqrt{< \mathbf{u}, \mathbf{u} >} \sqrt{< T, T\mathbf{d} >}.$$

Now

$$var(\mathbf{u}'Y) = \mathbf{u}' var(Y)\mathbf{u} = \frac{\sigma^2}{n}\mathbf{u}'\mathbf{u} = \frac{1}{n} < \mathbf{u}, \mathbf{u} > .$$

Hence

$$\begin{aligned} (< \mathbf{c}, \mathbf{d} >)^2 &\leq < \mathbf{u}, \mathbf{u} > < T\mathbf{d}, T\mathbf{d} > \\ &= n var(\mathbf{u}'Y) < T\mathbf{d}, T\mathbf{d} > \\ &= n\, var(\mathbf{u}'Y) \ < \mathbf{d}, T'T\mathbf{d} > \\ &= n\, var(\mathbf{u}'Y) < \mathbf{d}, M(\xi)\mathbf{d} > . \end{aligned}$$

It follows that

$$var\left(\mathbf{u}'Y\right) \geq \frac{1}{n} \frac{\left(<\mathbf{c},\mathbf{d}>\right)^2}{<\mathbf{d}, M\left(\xi\right)\mathbf{d}>}.$$

Going to the supremum in both sides of this inequality we obtain

$$var\left(\mathbf{u}'Y\right) \geq \sup_{0 \neq \mathbf{d} \in (Ker M(\xi))^{\perp}} \frac{1}{n} \frac{\left(<\mathbf{c},\mathbf{d}>\right)^2}{<\mathbf{d}, M\left(\xi\right)\mathbf{d}>}.$$

Step 2. We now prove that equality holds for some element in $F\left(\xi\right)$. Namely we prove that there exists $\boldsymbol{\gamma} \in \mathbb{R}^n$ such that for fixed ξ ,

$$var\left(<\boldsymbol{\gamma}, Y>\right) := \min_{\mathbf{u} \in F(\xi)} \left(var\left(<\mathbf{u}, Y>\right)\right).$$

Clearly by definition $<\boldsymbol{\gamma}, Y>$ will then be the Gauss - Markov estimator of $<\mathbf{c}, \theta>$. Note that a basis of the linear space generated by the column vectors of $M\left(\xi\right)$ is given by $\{\mathbf{v}_i, i = 1, ..., s\}$ where $s := \dim \operatorname{Im} M\left(\xi\right)$. We assume the vectors $\mathbf{v}_i$'s to have norm 1. When $M(\xi)$ is of full rank then $s = g$. The condition for estimableness $\mathbf{c} \in \operatorname{Im} M\left(\xi\right)$ due to (7) may then be written as

$$\mathbf{c} = \sum_{i=1}^{s} <\mathbf{v}_i, \mathbf{c}> \mathbf{v}_i.$$

Therefore

$$\begin{aligned}
\left(<\mathbf{c},\mathbf{d}>\right)^2 &= \left(<\sum_{i=1}^{s} <\mathbf{v}_i, \mathbf{c}> \mathbf{v}_i, \mathbf{d}>\right)^2 \\
&= \left(\sum_{i=1}^{s} <\mathbf{v}_i, \mathbf{c}><\mathbf{v}_i, \mathbf{d}>\right)^2 \\
&= \left(\sum_{i=1}^{s} \frac{<\mathbf{v}_i, \mathbf{c}>}{\sqrt{\lambda_i}} <\mathbf{v}_i, \mathbf{d}> \sqrt{\lambda_i}\right)^2
\end{aligned}$$

Apply Cauchy - Schwartz Inequality to each of the components of the vectors

$$\left(\frac{<\mathbf{v}_1, \mathbf{c}>}{\sqrt{\lambda_1}}, ..., \frac{<\mathbf{v}_s, \mathbf{c}>}{\sqrt{\lambda_s}}\right)$$

and

$$\left(<\mathbf{v}_1, \mathbf{d}> \sqrt{\lambda_1}, ..., <\mathbf{v}_1, \underline{\mathbf{d}}> \sqrt{\lambda_s}\right).$$

We get

$$\left(<\mathbf{c},\mathbf{d}>\right)^2 \leq \left(\sum_{i=1}^{s} \frac{<\mathbf{v}_i, \mathbf{c}>^2}{\lambda_i}\right) \left(\sum_{i=1}^{s} <\mathbf{v}_i, \mathbf{d}>^2 \lambda_i\right).$$

From the spectral Theorem (see e.g. [Lang, 1989]) we get that

$$M(\xi) = \sum_{i=1}^{s} \lambda_i < \mathbf{v}_i, \mathbf{v}_i >;$$

hence

$$\begin{aligned} (< \mathbf{c}, \mathbf{d} >)^2 &\leq \left(\sum_{i=1}^{s} \frac{< \mathbf{v}_i, \mathbf{c} >^2}{\lambda_i} \right) \left(\sum_{i=1}^{s} < \mathbf{v}_i, \mathbf{d} >^2 \lambda_i \right) \\ &= \left(\sum_{i=1}^{s} \frac{< \mathbf{v}_i, \mathbf{c} >^2}{\lambda_i} \right) < \mathbf{d}, M(\xi)\, \mathbf{d} > . \end{aligned} \tag{14}$$

In this last display equality holds between the first and the second members in two cases (see [Karlin and Studden, 1966a] p788). Either when there exists some constant h such that

$$< \mathbf{v}_i, \mathbf{c} >^2 = h \lambda_i < \mathbf{v}_i, \mathbf{d} >,$$

or when $\mathbf{d}$ is proportional to

$$\mathbf{d}^* := \sum_{i=1}^{s} \frac{< \mathbf{v}_i, \mathbf{c} >^2}{\lambda_i} \mathbf{v}_i.$$

We only consider this latest case. See [Karlin and Studden, 1966a] for a complete treatment. Recall that $M(\xi) = \frac{1}{n} T'T$. Taking $\boldsymbol{\gamma} := \frac{1}{n} T' \mathbf{d}^*$ in

$$var(\mathbf{u}'Y) \geq \sup_{\mathbf{0} \neq \mathbf{d} \in (Ker M(\xi))^{\perp}} \frac{1}{n} \frac{(< \mathbf{c}, \mathbf{d} >)^2}{< \mathbf{d}, M(\xi)\, \mathbf{d} >} \tag{15}$$

we get equality in (14). In order to conclude the proof it is necessary to prove that the vector

$$\boldsymbol{\gamma} := \frac{1}{n} T' \mathbf{d}^*$$

belongs to the set $F(\xi)$. Now

$$\begin{aligned} T' \boldsymbol{\gamma} \quad &: \quad = \frac{1}{n} T' T \mathbf{d}^* = M(\xi)\, \mathbf{d}^* \\ &= \sum_{i=1}^{s} \lambda_i < \mathbf{v}_i, \mathbf{c} > \mathbf{v}_i = \mathbf{c} \end{aligned}$$

which closes the proof. ■

4.2 An alternative form of the lower bound of the variance of the estimator of the c form

It can easily be checked that whenever the inverse matrix of $M(\xi)$ is defined then the mapping

$$\mathbf{d} \rightarrow \frac{<\mathbf{c}, \mathbf{d}>^2}{<\mathbf{d}, M(\xi)\,\mathbf{d}>}$$

with $\mathbf{d} \neq \mathbf{0}$ assumes its maximal value for

$$\mathbf{d} = \mathbf{d}_0 := M^{-1}(\xi)\,\mathbf{c}.$$

Therefore

$$\sup_{\mathbf{d} \neq \mathbf{0}} \frac{<\mathbf{c}, \mathbf{d}>^2}{<\mathbf{d}, M(\xi)\,\mathbf{d}>} = \frac{<\mathbf{c}, \mathbf{d}_0>^2}{<\mathbf{d}_0, M(\xi)\,\mathbf{d}_0>}.$$

Indeed in the above Proposition 13 we proved that the variance of the Gauss-Markov estimator $\widehat{<\mathbf{c}, \theta>}$ of the form $<\mathbf{c}, \theta>$ is

$$var\,\widehat{<\mathbf{c}, \theta>} = \frac{1}{n} \sup_{\mathbf{d} \neq \mathbf{0}} \frac{<\mathbf{c}, \mathbf{d}>^2}{<\mathbf{d}, M(\xi)\,\mathbf{d}>}.$$

Assuming that $M(\xi)$ is invertible; direct calculation leads

$$var\,\widehat{<\mathbf{c}, \theta>} = \frac{1}{n}\mathbf{c}' M(\xi)\,\mathbf{c}. \qquad (16)$$

Setting $\sigma^2 = 1$ (since the optimal design does not depend on σ^2) it follows that the Gauss-Markov estimator is given by

$$\widehat{<\mathbf{c}, \theta>} = \mathbf{c}'\left(X'(x)\;X(x)\right)^{-1} X'(x)\,Y.$$

Hence

$$\begin{aligned} var\,\widehat{<\mathbf{c}, \theta>} &= var\left(\mathbf{c}'\left(X'(x)\;X(x)\right)^{-1} X'(x)\,Y\right) \\ &= \frac{\mathbf{c}'\left(X'(x)\;X(x)\right)^{-1} X'(x)\;X(x)\left(X'(x)\;X(x)\right)^{-1}\mathbf{c}}{n} \\ &= \frac{1}{n}\mathbf{c}' M^{-1}(\xi)\,\mathbf{c}. \end{aligned}$$

Now

$$\frac{<\mathbf{c}, \mathbf{d}_0>^2}{<\mathbf{d}_0, M(\xi)\,\mathbf{d}_0>} = \mathbf{c}' M^{-1}(\xi)\;\mathbf{c}.$$

Indeed

$$\begin{aligned} \frac{<\mathbf{c}, \mathbf{d}_0>^2}{<\mathbf{d}_0, M(\xi)\,\mathbf{d}_0>} &= \frac{<\mathbf{c}, M^{-1}(\xi)\,\mathbf{c}>^2}{<M^{-1}(\xi)\,\mathbf{c}, M(\xi)\,M^{-1}(\xi)\,\mathbf{c}>} \\ &= \mathbf{c}' M^{-1}(\xi)\,\mathbf{c}. \end{aligned}$$

In the above Proposition 13 the measure ξ is fixed in $\mathcal{M}_d([-1,1])$. Now since

$$M(\xi) := \frac{1}{n}T'T = \int_{[-1,1]} (X(x))' X(x)\, d\xi(x),$$

let ξ vary in $\mathcal{M}_d([-1,1])$; define the optimal design ξ which minimizes the variance as follows

$$\begin{aligned} var_{\xi^*}(<\gamma, Y>) &:= \min_{\xi\in\mathcal{M}_d([-1,1])} var(<\gamma, Y>) \\ &= \min_{\xi\in\mathcal{M}_d([-1,1])} \min_{F(\xi)} (var(<\mathbf{u}, Y>)) \\ &= \min_{\xi\in\mathcal{M}_d([-1,1])} \frac{1}{n} <\mathbf{c}, (M(\xi))^{-}\mathbf{c}> \\ &= \min_{\xi\in\mathcal{M}_d([-1,1])} \sup_{0\neq\mathbf{d}\in(KerM(\xi))^{\perp}} \frac{1}{n}\frac{(<\mathbf{c},\mathbf{d}>)^2}{<\mathbf{d}, M(\xi)\,\mathbf{d}>} \\ &= \min_{\xi\in\mathcal{M}_d([-1,1])} \sup_{0\neq\mathbf{d}\in(KerM(\xi))^{\perp}} \frac{(\mathbf{d}'\mathbf{c})^2}{\int_{[-1,1]} (\mathbf{d}'X(x))^2\, d\xi(x)}. \end{aligned}$$

Since $\mathbf{d}$ in $(KerM(\xi))^{\perp}$ can be chosen up to an arbitrary multiplicative constant (see formula (15)) we may assume that $\mathbf{d}'\mathbf{c} = 1$. Minimizing upon ξ choosing a measure whose support with g points consists in the points where the mapping $x \to \mathbf{d}'X(x)$ assumes its maximal values it holds

$$\begin{aligned} var_{\xi^*}(<\gamma, Y>) &= \min_{\xi\in\mathcal{M}_d([-1,1])} \sup_{0\neq\mathbf{d}\in(KerM(\xi))^{\perp}} \frac{1}{\int_{[-1,1]} (\mathbf{d}'X(x))^2\, d\xi(x)} \\ &= \min_{\xi\in\mathcal{M}_d([-1,1])} \frac{1}{\min_{0\neq\mathbf{d}\in(KerM(\xi))^{\perp}} (\max_x \mathbf{d}'X(x))^2}. \end{aligned}$$

Definition 14 *Let ξ belong to $\mathcal{M}_d([-1,1])$. If $<\mathbf{c},\theta>$ is estimable we denote*

$$d(\xi,\mathbf{c}) := \sup_{0\neq\mathbf{d}\in(KerM(\xi))^{\perp}} \frac{(<\mathbf{c},\mathbf{d}>)^2}{<\mathbf{d}, M(\xi)\,\mathbf{d}>}.$$

By the above discussion , denoting var_ξ the variance under a design ξ,

$$var_\xi \widehat{<\mathbf{c},\theta>} = \frac{1}{n} d(\xi,\mathbf{c}).$$

We now consider the relation between a linear form and the measure for which it is estimable.

We start with some preliminary result. Let A and B be two square non negative semi definite matrices with same dimension k. Then

$$Ker(A+B) = Ker(A) \cap Ker(B) \tag{17}$$

From (17), taking orthogonals and using (8) we get

$$\operatorname{Im}(A+B) = \operatorname{Im} A + \operatorname{Im} B.$$

We now state

Proposition 15 *(Pukelsheim p. 42) It holds* $\operatorname{Im}(M(\xi)) = span\{X(x) : x \in \mathbf{supp}(\xi)\}$

Proof. Since

$$M(\xi) = \sum_{j \le r} \xi(X(x_j)) X(x_j) (X(x_j))'$$

We have, by (??)

$$\begin{aligned} \operatorname{Im}(M(\xi)) &= \operatorname{Im}\left(\sum_{j \le r} \xi(X(x_j)) X(x_j) (X(x_j))' \right) \\ &= \sum_{j \le r} \operatorname{Im}\left(\xi(X(x_j)) X(x_j) (X(x_j))' \right). \end{aligned}$$

.Now the image of the matrix $\xi(X(x_j)) X(x_j) (X(x_j))'$ is the linear space spanned by its column vectors, i.e. $span(X(x_j))$.Therefore

$$\begin{aligned} \operatorname{Im}(M(\xi)) &= \sum_{j \le r} span(X(x_j)) \\ &= \left\{ \sum_{j \le r} k_j X(x_j) : k_j \in \mathbb{R} \text{ per } j = 0, ..., r \right\} \\ &= span\{X(x) : x \in \mathbf{supp}(\xi)\}. \end{aligned}$$

■

By the discussion in Section 4 $\mathbf{c}'\theta$ is estimable if and only if $\mathbf{c}$ belongs to $\operatorname{Im} M(\xi)$; therefore the above Proposition entails the following alternative definition to the corresponding Definition in 3.

Definition 16 *A* $\mathbf{c}-$ *form is estimable with respect to a measure* $\xi \in \mathcal{M}_d([-1,1])$ *if and only if* $\mathbf{c} \in span\{X(x) : x \in \mathbf{supp}(\xi)\}$.

In $\mathcal{M}_d([-1,1])$ consider the following optimal Problem

$$\begin{cases} Min_{\xi \in \mathcal{M}_d([-1,1])} d(\xi, \mathbf{c}) \\ \mathbf{c} \in span\{X(x) : x \in \mathbf{supp}(\xi)\} \end{cases} \tag{18}$$

The solution $\xi^* \in \mathcal{M}_d([-1,1])$ to Problem 18 is the discrete measure which minimizes the variance of the Gauss Markov estimator of $< \mathbf{c}', \theta >$.

5 Geometric interpretation of estimableness. Elfving set

5.1 Estimable forms and a convex subset of the regression range; the Elfving set

We already saw that estimableness is related to a precise geometric relation. This paragraph introduces the geometric context; we follow the presentation by [Pukelsheim, 2006].

The condition for estimableness of a linear form $< \mathbf{c}, \theta >$ is given by

$$\mathbf{c} \in \operatorname{Im} M(\xi).$$

This property may be extended independently from the measure ξ, for any element in $\mathcal{S}_{\geq 0}(g)$. We thus consider a generic matrix $A \in \mathcal{S}_{\geq 0}(g)$ such that $\mathbf{c} \in \operatorname{Im} A$, and also we consider all matrices $A \in \mathcal{S}_{\geq 0}(g)$ for which $\mathbf{c} \in \operatorname{Im} A$.

Definition 17 *The set*

$$\mathcal{A}(\mathbf{c}) := \{A \in \mathcal{S}_{\geq 0}(g) \text{ such that } \mathbf{c} \in \operatorname{Im} A\}$$

is called the feasibility cone.

The estimableness condition writes therefore

$$M(\xi) \in \mathcal{A}(\mathbf{c}) \cap \mathbf{M}$$

where

$$\mathbf{M} := \{M(\xi) : \xi \in \mathcal{M}_d([-1,1])\}.$$

That $M(\xi)$ belongs to $\mathbf{M}$ means that $M(\xi)$ is a moment matrix.

In $\mathcal{M}_d([-1,1])$ define the following optimization problem

$$\begin{cases} Min_{\xi \in \mathcal{M}_d([-1,1])} \mathbf{c}' M(\xi) \mathbf{c} \\ M(\xi) \in \mathcal{A}(\mathbf{c}) \cap \mathbf{M} \end{cases} \tag{19}$$

Problem 19 is analogous to Problem 18 in $\mathcal{M}_d([-1,1])$; see (16).

Proposition 18 *The feasibility cone $\mathcal{A}(\mathbf{c})$ for $< \mathbf{c}, \theta >$ is a convex subcone of $\mathcal{S}_{\geq 0}(g)$ which contains $\mathcal{S}^+(g)$.*

Proof. If $\delta > 0$ and $A \in \mathcal{A}(\mathbf{c})$ then since $\operatorname{Im} A = \operatorname{Im}(\delta A)$ it holds $\delta A \in \mathcal{A}(\mathbf{c})$, for any positive δ. Hence $\mathcal{A}(\mathbf{c})$ is a cone. By definition $\mathcal{A}(\mathbf{c}) \subset \mathcal{S}_{\geq 0}(g)$ and therefore $\mathcal{A}(\mathbf{c})$ is a subcone in $\mathcal{S}_{\geq 0}(g)$.

We prove that $\mathcal{A}(\mathbf{c})$ is convex. Let $\delta \in (0,1)$ and $A, B \in \mathcal{A}(\mathbf{c})$. Since A and B both belong to $\mathcal{S}_{\geq 0}(g)$ it holds

$$\operatorname{Im}(A+B) = \operatorname{Im} A + \operatorname{Im} B$$

(see [Pukelsheim, 2006] p 37). Hence for any A and B in $\mathcal{A}(\mathbf{c})$ it holds $(\delta A + (1-\delta) B) \in \mathcal{A}(\mathbf{c})$. ■

Since estimableness pertains to the expectation of an estimator and not to its variance we now characterize it using a generic unbiased linear estimator of a linear form.

Given

$$E(Y(x)) = E\left(\sum_{j=0}^{g-1} \theta_j \varphi_j(x) + \varepsilon_j\right) = X'(x)\,\theta,$$

an easy way in order to estimate a linear form $\mathbf{c}'\theta$ consists in a weighted mean. At this point it will be convenient to use the notation $Y_i(x_j)$ defined in (1) in order to define the estimator. For any node x_j define

$$\overline{Y}(x_j) := \frac{1}{n_j} \sum_{i=1}^{n_j} Y_i(x_j)$$

the mean of the observations on node x_j.

Denote

$$\widehat{\mathbf{c}'\theta} := \sum_{j=0}^{g-1} u_j \overline{Y}(x_j) \tag{20}$$

where the u_j 's are coefficients which should be determined in such a way that $\widehat{\mathbf{c}'\theta}$ is unbiased.

Hence the relation between the form $\mathbf{c}$ and the coefficients u_j's is dictated by

$$\mathbf{c}'\theta = E\left(\widehat{\mathbf{c}'\theta}\right) = E\left(\sum_{j=0}^{g-1} u_j \overline{Y}(x_j)\right)$$
$$= \sum_{j=0}^{g-1} u_j X'(x_j)\,\theta$$

which holds if and only if

$$\mathbf{c}' = \sum_{j=0}^{g-1} u_j X'(x_j).$$

Observe that there exists at least one index j such that $u_j \neq 0$. Indeed otherwise no data enters in the definition of the estimator. It follows that

$$\sum_{j=0}^{g-1} |u_j| \neq 0.$$

Henceforth dividing

$$\mathbf{c}' = \sum_{j=0}^{g-1} u_j X'(x_j)$$

by $\sum_{j=0}^{g-1} |u_j|$ we get

$$\frac{\mathbf{c}'}{\sum_{j=0}^{g-1} |u_j|} = \frac{\sum_{j=0}^{g-1} u_j X'(x_j)}{\sum_{j=0}^{g-1} |u_j|},$$

which, setting

$$\epsilon_j = sign(u_j) = \pm 1$$

becomes

$$\frac{\mathbf{c}'}{\sum_{j=0}^{g-1} |u_j|} = \sum_{j=0}^{g-1} \frac{\epsilon_j |u_j|}{\sum_{j=0}^{g-1} |u_j|} X'(x_j)$$

and therefore

$$\frac{\mathbf{c}}{\sum_{j=0}^{g-1} |u_j|} = \sum_{j=0}^{g-1} \frac{\epsilon_j |u_j|}{\sum_{j=0}^{g-1} |u_j|} X\ (x_j).$$

For $j = 0, ..., g-1$, the numbers

$$\eta_j := \frac{|u_j|}{\sum_{j=0}^{g-1} |u_j|}$$

define a discrete probability measure with support $\{x_0, ..., x_{g-1}\}$ included in $[-1, 1]$.

The condition for estimableness may thus be stated as follows.

Proposition 19 *Let $(u_0, .., u_{g-1})$ be defined by (20) and $(x_0, .., x_{g-1})$ be a set of nodes. The linear form $\mathbf{c}'\theta$ is estimable if and only if $\mathbf{c}/\sum_{j=0}^{g-1} |u_j|$ is a convex linear combination of the vectors $\epsilon_j X\ (x_j)$, $j = 0, ..., g-1$.*

Since ϵ_j is a sign function we conclude that $\mathbf{c}'\theta$ is estimable if and only if $\mathbf{c}/\sum_{j=0}^{g-1} |u_j|$ belongs to the convex hull generated by the set $\{\epsilon_j X\ (x_j), j = 0, ..., g-1\}$.

Call

$$\widetilde{\mathcal{R}}_+ := \{X\ (x_j), j = 0, ..., g-1\},$$
$$\widetilde{\mathcal{R}}_- := \{-X\ (x_j), j = 0, ..., g-1\},$$
$$\widetilde{\mathcal{R}} := \text{convex-hull}\left(\widetilde{\mathcal{R}}_+ \cup \widetilde{\mathcal{R}}_-\right)$$

Since $\pm X\ (x_j), j = 0, ..., g-1$ is a finite set of vectors, the set $\widetilde{\mathcal{R}}$ is a polytope.

An equivalent context for Proposition 19 is to introduce the sets $\mathcal{R}_+, \mathcal{R}_-$ and $\mathcal{R}$; indeed by Carathéodory Theorem 20 the convex hull of the points

$\{\epsilon_j X\ (x_j), j = 0, ..., g-1\}$ is generated by a finite collection of points. Introduce henceforth

$$\begin{aligned} \mathcal{R}_+ &:= \{X\ (x), x \in [-1,1]\}, \\ \mathcal{R}_- &:= \{-X\ (x), x \in [-1,1]\}, \\ \mathcal{R} &:= \text{convex-hull}\,(\mathcal{R}_+ \cup \mathcal{R}_-). \end{aligned}$$

Following Proposition 19 it holds

$$\mathbf{c}'\theta \text{ is estimable if and only if } \mathbf{c}/\sum_{j=0}^{g-1} |u_j| \in \mathcal{R}.$$

The set $\mathcal{R}$ is the *Elfving set.*

5.2 Geometry of the Elfving set

The Elfving set is symmetric and convex, by its very definition.

The points in $\mathcal{R}$ may be seen as expected values for probability measures η's. Indeed the random variable which assumes values $(\epsilon\,(X\ (x_j)))\,X\ (x_j)$, $j = 0, ..., g-1$ with probability η_j has expectation $\sum_{j=0}^{g-1} \eta_j\,(\epsilon\,(X\ (x_j))\,X\ (x_j))$; the reciprocal statement clearly holds. Thus to any point $\mathbf{z}$ in $\mathcal{R}$ we may associate a design η.

The Elfving set is contained in the regression range, namely in the linear space $span\,\{X(x) : x \in [-1,1]\}$. Indeed the convex combinations of $\{X(x) : x \in [-1,1]\}$ belong to this space.

In $span\,\{X(x) : x \in [-1,1]\}$ define the norm (*gauge or Elfving norm*)

$$\rho : span\,\{X(x) : x \in [-1,1]\} \to \mathbb{R}^+,\ \mathbf{z} \mapsto \rho(\mathbf{z}) := \inf\{\delta \geq 0 : \mathbf{z} \in \delta\mathcal{R}\}.$$

This norm is useful in order to locate any point $\mathbf{z} \in span\,\{X(x) : x \in [-1,1]\}$ with respect to $\mathcal{R}$.

For example if $\rho(\mathbf{z}) = 0$ then $\mathbf{z} \in 0\mathcal{R} = \mathbf{0}$. Therefore $\mathbf{z} = \mathbf{0}$ (which belongs to $\mathcal{R}$). If $\rho(\mathbf{z}) = 1 + \varepsilon$, with $\varepsilon > 0$ then $\mathbf{z} \notin \mathcal{R}$. If $\rho(\mathbf{z}) = 1/t$ then $\mathbf{z} \in (1/t)\,\mathcal{R}$ and $z \notin (1/t + \varepsilon)\,\mathcal{R}$. The larger t , the closer $\mathbf{z}$ to the null vector. Reciprocally small values of t make $\mathbf{z}$ close to the boundary $Fr\,(\mathcal{R})$ of $\mathcal{R}$. As t decreases the point moves away from $\mathcal{R}$.

Clearly $\mathbf{z} \in \mathcal{R}$ if and only if $\rho(\mathbf{z}) \leq 1$.

It follows that the Elfving set coincides with the closed sphere with radius 1 and center $\mathbf{z} = \mathbf{0}$ in $span\,\{X(x) : x \in [-1,1]\}$.

It holds

$$\mathcal{R} = \{\mathbf{z} \in span\,\{X(x) : x \in [-1,1]\} : \rho(\mathbf{z}) \leq 1\},$$

which yields that $\mathcal{R}$ is a compact .

We now characterize the boundary points of $\mathcal{R}$, namely

$$Fr(\mathcal{R}) := \{\mathbf{z} \in span\{X(x) : x \in [-1,1]\} : \rho(\mathbf{z}) = 1\}.$$

We fist recall a basic theorem pertaining to convex sets in $\mathbb{R}^{g-1}$, $g \geq 2$.

Theorem 20 *(Carathéodory) Let A be a non void subset in $\mathbb{R}^{g-1}$. Then any convex combination of elements in A can be written as a convex combination of at most g points.*

Proof. See e.g. [van Tiel, 1984] p 41. ■

Proposition 21 *Let $\mathbf{z} \in span\{X(x) : x \in [-1,1]\}$. Then there exists a discrete design ξ such that*

$$\frac{\mathbf{z}}{\rho(\mathbf{z})} = \int_{[-1,1]} (\epsilon(X(x)))\, X(x)\, \xi(dx).$$

Proof. The proof is in [Pukelsheim, 2006] p 48. ■

We use the fact that in any boundary point of a convex set there exists a tangent hyperplane to the convex set. This hyperplane divides $span\{X(x) : x \in [-1,1]\}$ into two subsets; the first one "below" contains $\mathcal{R}$ and the second one "above" does not contain any point in $\mathcal{R}$. This fact allows for the determination of the boundary of $\mathcal{R}$.

For $\mathbf{c}$ in $\mathcal{R}$ it holds

$$\frac{\mathbf{c}}{\rho(\mathbf{c})} \in Fr(\mathcal{R}).$$

Proposition 22 *There exists a vector $\mathbf{h}$ in $\mathbb{R}^g$ such that for any $\mathbf{z}$ in $\mathcal{R}$ it holds*

$$\mathbf{z}'\mathbf{h}\mathbf{h}'\mathbf{z} \leq 1.$$

Proof. The tangent hyperplane to $\mathcal{R}$ at point $\mathbf{c}/\rho(\mathbf{c})$ is defined as follows. A vector $\widetilde{\mathbf{h}}$ in $\mathbb{R}^g$ defines a linear form which in turn determines the tangent hyperplane to $\mathcal{R}$ at point $\mathbf{c}$ if for any $\mathbf{z} \in \mathcal{R}$,

$$\mathbf{z}'\,\widetilde{\mathbf{h}} \leq \frac{\mathbf{c}}{\rho(\mathbf{c})}\,\widetilde{\mathbf{h}}.$$

This relation states that all points in $\mathcal{R}$ lay "below" the hyperplane. Since $\mathcal{R}$ is a symmetric set, when $\mathbf{z}$ satisfies $\mathbf{z}'\,\widetilde{\mathbf{h}} \leq \frac{\mathbf{c}\,\widetilde{\mathbf{h}}}{\rho(\mathbf{c})}$ then the same holds for $-\mathbf{z}$ which also belongs to $\mathcal{R}$.

Hence

$$\begin{cases} \mathbf{z}'\,\widetilde{\mathbf{h}} \leq \frac{\mathbf{c}'\,\widetilde{\mathbf{h}}}{\rho(\mathbf{c})} \\ -\mathbf{z}'\,\widetilde{\mathbf{h}} \leq \frac{\mathbf{c}'\,\widetilde{\mathbf{h}}}{\rho(\mathbf{c})} \end{cases}.$$

It follows that for any $\mathbf{z} \in \mathcal{R}$, we have

$$\left|\mathbf{z}'\,\widetilde{\mathbf{h}}\right| \leq \frac{\mathbf{c}'\,\widetilde{\mathbf{h}}}{\rho(\mathbf{c})}.$$

The real number

$$\gamma := \frac{\mathbf{c}'\,\widetilde{\mathbf{h}}}{\rho(\mathbf{c})}$$

is therefore non negative. Furthermore it does not equal 0. Otherwise $\mathcal{R}$ has a void interior. Hence $\gamma > 0$. Define therefore the vector

$$\mathbf{h} := \frac{\widetilde{\mathbf{h}}}{\gamma}$$

from which

$$\mathbf{z}'\mathbf{h} \leq 1.$$

Also, since

$$\frac{\mathbf{c}\,\mathbf{h}}{\rho(\mathbf{c})} = 1$$

we get

$$\mathbf{z}'\,\mathbf{h}\,\mathbf{h}'\mathbf{z} \leq 1,$$

for any $\mathbf{z} \in \mathcal{R}$. ■

5.2.1 Some further development on the representation of the Elfving set.

So far we considered the inner representation of set Elfving set $\mathcal{R}$, through convex combination of elements of the kind $\pm X(x)$ with x in $[-1, 1]$. We now consider the fact that the Elfving is defined through an inner product; namely it holds, making use of Proposition 22,

$$\mathcal{R} = \left\{X(x) \in \mathbb{R}^g : (X(x))'\, N\ (X(x)) \leq 1\right\}$$

where

$$N = \mathbf{h}\,\mathbf{h}'$$

and $\mathbf{h}$ is the vector which defines the tangent hypeplane to $\mathcal{R}$ at point $\mathbf{c}/\rho(\mathbf{c})$

.

The matrix N clearly is symmetric and positive semi definite.
Define the inner product on $\mathbb{R}^g$ by

$$< \mathbf{z}', \mathbf{w} >_N := \mathbf{z}' N\ \mathbf{w}\ .$$

Thus, to such vector $\mathbf{h}$ we associate a quadratic form $< .,. >_N$ which is defined by the matrix N.

Geometrically, $\mathcal{R}$ is the closed unit ball with center $\mathbf{0}$ defined by

$$\mathcal{R} := \{\mathbf{z} \in span\{X(x) : x \in [-1, 1]\} : \rho(\mathbf{z}) \leq 1\}.$$

Consider the regression range

$$\mathcal{H} := \{X(x) : x \in [-1, 1]\} \subset \mathbb{R}^g.$$

Assume first that x may assume only a finite number of values in $[-1,1]$. Then $\mathcal{H}$ is finite dimensional, and $\mathcal{R}$ is a polytope with finite number of edges and vertices. When x may assume an infinite number of values in $[-1,1]$ then R coincides with a geometric set depending on the range of the matrix N. When N has full rank, then $\mathcal{R}$ is an hyperellipsoid (say ellipsoid),. When N is degenerate, then $\mathcal{R}$ is an hypercylinder (say a cylinder).

Citing [Pukelsheim, 2006] p 44 "Nevertheless Elfving sets and scalar product balls are linked to each other in an intrinsic way. A scalar product ball given by a positive define matrix N is an ellipsoid, because of the full rank of N; if we drop the full rank assumption, the ellipsoid may degenerate to a cylinder". For a nonnegative define matrix $N \in S_{\geq 0}(g)$, Pulkesheim calls the set of vectors $\{z \in \mathbb{R}^g : z'Nz \leq 1\}$ the cylinder induced by N. It includes the nullspace of N.

Elfving sets allow many shapes other than cylinders. However, we may approximate a given Elfving set $\mathcal{R}$ from the outside, by considering all cylinders that include $\mathcal{R}$. Since cylinders are symmetric and convex, inclusion of R equivalent to inclusion of the regression range.

Identifying a cylinder with the matrix inducing it, we define the set $\mathcal{N}$ of all a cylinders that includes $\mathcal{R}$ or $\mathcal{H}$ by

$$\begin{aligned} \mathcal{N} &: = \{N \in \mathcal{S}_{\geq 0}(g) : z'Nz \leq 1 \text{ for all } z \in \mathcal{R}\} \\ &= \{N \in \mathcal{S}_{\geq 0}(g) : x'Nx \leq 1 \text{ for all } x \in \mathcal{H}\}. \end{aligned} \tag{21}$$

Elements in $\mathcal{N}$ cast light on the connection between such approximation of $\mathcal{R}$ and moment matrices, which in turn are in relation with the design.

5.3 The relation between cylinders and the variance of the estimator of the c− form

The relation between the outer approximation of the Elfving set $\mathcal{R}$ through cylinders and the variance of the estimator of the $\mathbf{c}$ form can be captured noting that the optimal design is obtained solving

$$\begin{cases} min_{\xi} \ \mathbf{c}' \left(M(\xi)\right)^{-} \mathbf{c} \\ M(\xi) \in \mathcal{A}(\mathbf{c}) \cap \mathbf{M} \end{cases}. \tag{22}$$

where the constraint

$$M(\xi) \in A(\mathbf{c}) \cap \mathbf{M}$$

amounts to the estimableness of the linear form $\mathbf{c}'\theta$; see section 5 .

Optimality may now be characterized in terms of the moment matrix.

Definition 23 *A matrix moment $M(\xi)$ is called optimal in $\mathbf{M}$ for the form $\mathbf{c}'\theta$ when it belongs to $A(\mathbf{c})$ and when $\mathbf{c}' \left(M(\xi)\right)^{-} \mathbf{c}$ is minimal in $\mathbf{M}$.*

We observe that when M belongs to $A(\mathbf{c})$ then $\mathbf{c}' \left(M(\xi)\right)^{-} \mathbf{c}$ does not depend on the very inverse $\left(M(\xi)\right)^{-}$ used in order to evaluate $\mathbf{c}' \left(M(\xi)\right)^{-} \mathbf{c}$ (see [Pukelsheim, 2006] Chapter 2).

In (22) the variance $\mathbf{c}'\ (M(\xi))^{-}\ \mathbf{c}$ is written in terms of cylinders. We may write this geometrically, and the variance is minimal whenever, for any cylinder defined by a matrix $N \in \mathcal{N}$ (see Definition 21) it holds

$$\mathbf{c}'\ (M(\xi))^{-}\ \mathbf{c} \leq \mathbf{c}'\ N\ \mathbf{c}$$

or, in terms of the Loewner order

$$(M(\xi))^{-} \preceq N.$$

5.4 Lower bound for the variance

Let

$$\mathcal{N} := \left\{ \begin{array}{c} N \in S_{\geq 0}(g) \text{ such that } (X(x))' NX(x) \leq 1 \\ \text{for all } X(x) \text{ and any } x \in [-1,1] \end{array} \right\}$$

and assume that $M(\xi) \in A(\mathbf{c}) \cap \mathbf{M}$.

In this Section we quote a basic result, which states that for any matrix N in $\mathcal{N}$ it holds

$$\mathbf{c}'\ (M(\xi))^{-}\ \mathbf{c} \geq \mathbf{c}' N \mathbf{c}.$$

This is to say that any cylinder which contains the regression range is induced by a quadratic form which is a lower bound of the variance of $\widehat{\mathbf{c}'\theta}$.

The minimization of the variance of $\widehat{\mathbf{c}'\theta}$ amounts to the attainment of this lower bound, which is to say to the problem of the definition of the cylinder which realizes the infimum.

Recall that the variance of the Gauss Markov estimator of the $\mathbf{c}-$ form with design ξ is $\mathbf{c}'\ (M(\xi))^{-}\ \mathbf{c}$, where $(M(\xi))^{-}$ is the generalized inverse of $M(\xi)$, hence an element in $S_{\geq 0}(g)$. The next result compares this variance with homologue terms when $(M(\xi))^{-}$ is substituted by a generic element in $\mathcal{N}$, providing a lower bound for the variance upon all designs.

Theorem 24 *(Pukelsheim) Assume that* $M(\xi) \in A(\mathbf{c}) \cap \mathbf{M}(\mathcal{M}_d[-1,1])$. *Then for any* N *in* $\mathcal{N}$

$$Var\left(\widehat{<\mathbf{c},\theta>}\right) = \mathbf{c}'\ (M(\xi))^{-}\ \mathbf{c} \geq \mathbf{c}' N \mathbf{c},$$

Proof. We present a proof of this important result, more accessible than as presented in [Pukelsheim, 2006] pp 20, 21 and 46.

Step 1 We prove that

$$tr(M(\xi)\, N) \leq 1. \tag{23}$$

Integrate with respect to ξ in both sides of the inequality $(X(x))' NX(x) \leq 1$. Then

$$\int_{[-1,1]} (X(x))'\, N\ X(x)\ d\xi(x) \leq \int_{[-1,1]} 1\ d\xi(x) = 1.$$

Since $M(\xi) = X(x)(X(x))'$ and $tr(AB) = tr(BA)$, denoting $A := X(x)$ and $B := (X(x))'N$, we obtain

$$\begin{aligned}(X(x))'N\,X(x) &= tr\left((X(x))'N\,X(x)\right) \\ &= tr\left(X(x)(X(x))'\,N\right) = tr(M(\xi)N).\end{aligned} \tag{24}$$

Therefore (23) holds.

Step 2 We now prove that

$$tr(M(\xi)N) \geq \left(\underline{\mathbf{c}}'(M(\xi))^{-}\underline{\mathbf{c}}\right)^{-1}\underline{\mathbf{c}}'N\underline{\mathbf{c}}, \tag{25}$$

This follows from the fact that $M(\xi) \in A(\mathbf{c})$. Indeed if $M(\xi) \in \mathcal{A}(\mathbf{c})$ then by the Gauss Markov Theorem it can be proved that (see [Pukelsheim, 2006] pp 21 and 22)

$$M(\xi) \succeq \mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'.$$

Multiplying both sides by N,

$$M(\xi)N \succeq \mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N;$$

using the fact that when $A \succeq B$ then $tr(A) \geq tr(B)$), we obtain

$$\begin{aligned}tr(M(\xi)N) &\geq tr\left(\mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N\right) \\ &= tr\left(\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N\,\mathbf{c}\right) \\ &= \left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N\mathbf{c}.\end{aligned}$$

Step 3 We now prove the claim. From (25)

$$\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)\,tr(M(\xi)N) \geq \mathbf{c}'N\mathbf{c}.$$

Now by (23)

$$tr(M(\xi)N) \leq 1$$

and therefore, multiplying by $\mathbf{c}'(M(\xi))^{-}\mathbf{c}$ we obtain

$$tr(M(\xi)N)\,\mathbf{c}'(M(\xi))^{-}\mathbf{c} \leq \mathbf{c}'(M(\xi))^{-}\mathbf{c} \tag{26}$$

and finally

$$Var\left(\widehat{<\mathbf{c},\theta>}\right) = \mathbf{c}'(M(\xi))^{-}\mathbf{c} \geq \mathbf{c}'N\mathbf{c}.$$

We will see that this lower bound can be achieved, which yields a criterion for the optimality of the design ξ.

5.5 The lower bound can be achieved

We now prove that conditions on $N \in \mathcal{N}$ exist which imply

$$\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c} = \mathbf{c}'N\mathbf{c}.$$

We push forward three preliminary Lemmas. For fixed $\mathbf{c} \in \mathcal{R}$ denote $\widetilde{\mathbf{h}}$ the vector of the coefficients of the tangent hyperplane to $\mathcal{R}$ at point $\underline{\mathbf{c}}/\rho\left(\underline{\mathbf{c}}\right)$ and

$$\mathbf{h} := \frac{\widetilde{\mathbf{h}}}{\gamma}, \gamma = \left\langle \frac{\mathbf{c}}{\rho\left(\mathbf{c}\right)}, \widetilde{\mathbf{h}} \right\rangle.$$

Accordingly define the symmetric semidefinite positive matrix

$$N := \mathbf{h}\mathbf{h}'.$$

Lemma 25 *Let $M\left(\xi\right) \in \mathcal{A}\left(\mathbf{c}\right)$. Then*

$$M\left(\xi\right)\left(M\left(\xi\right)\right)^{-}\mathbf{c} = \mathbf{c} \tag{27}$$

and

$$\frac{T}{\sqrt{n}} = M\left(\xi\right)\left(M\left(\xi\right)\right)^{-}\frac{T}{\sqrt{n}}. \tag{28}$$

Furthermore

$$\text{If } G = \left(\left(M\left(\xi\right)\right)^{-}\right)^{-} \text{ then } \left(\frac{T}{\sqrt{n}}\right)' G\mathbf{c} = \left(\frac{T}{\sqrt{n}}\right)'\left(M\left(\xi\right)\right)^{-}\mathbf{c}. \tag{29}$$

Proof. See [Pukelsheim, 2006] p 41. ■

Lemma 26 *(Pukelsheim) Assume that*

$$M\left(\xi\right) \in \mathcal{A}\left(\mathbf{c}\right) \cap \mathbf{M}\left(\mathcal{M}_d\left[-1,1\right]\right).$$

Then $tr\left(M\left(\xi\right)N\right) = 1$ if and only if $\left(X\left(x\right)\right)' N\, X\left(x\right) = 1$ for all $x \in \boldsymbol{supp}(\xi)$.

Proof. See [Pukelsheim, 2006] p 45. ■

Lemma 27 *(Pukelsheim) Assume that*

$$M\left(\xi\right) \in \mathcal{A}\left(\mathbf{c}\right) \cap \mathbf{M}\left(\mathcal{M}_d\left[-1,1\right]\right).$$

Let N belong to $\mathcal{N}$. Then the two following assertions

(i)

$$tr\left(M\left(\xi\right)N\right) = \left(\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N\mathbf{c} = 1$$

and

(ii)

$$\begin{cases} \left(X\left(x\right)\right)' NX\left(x\right) = 1 \text{ for any } x \in \mathbf{supp}\left(\xi\right) \\ M\left(\xi\right)N = \mathbf{c}\left(\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N \end{cases}$$

are equivalent.

Proof. See [Pukelsheim, 2006] p 45. ■

The following Theorem indicates conditions for

$$\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}=\mathbf{c}'N\mathbf{c}$$

to hold.

Theorem 28 *(Pukelsheim) Assume that* $M\left(\xi\right)\in\mathcal{A}\left(\mathbf{c}\right)\cap\mathbf{M}\left(\mathcal{M}_{d}\left[-1,1\right]\right)$*. Then the two following assertions*

(i)

$$\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}=\mathbf{c}'N\mathbf{c}$$

and

(ii)

$$\begin{cases}\left(X\left(x\right)\right)'NX\left(x\right)=1 \text{ for any } x\in\mathbf{supp}\left(\xi\right)\\ M\left(\xi\right)N=\mathbf{c}\left(\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N\end{cases}$$

are equivalent.

Proof. We have by Lemma 27

$$trM\left(\xi\right)N=\left(\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N\mathbf{c}$$

and

$$trM\left(\xi\right)N=1.$$

Hence

$$\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}=\mathbf{c}'N\mathbf{c}. \tag{30}$$

■

We evaluate $\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}$. Since $\mathbf{c}/\rho\left(\mathbf{c}\right)=M\left(\xi\right)\mathbf{h}$ using (36)

$$\begin{aligned}\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}&=\left(\rho\left(\mathbf{c}\right)\right)^{2}\left(\frac{\mathbf{c}'}{\rho\left(\mathbf{c}\right)}\left(M\left(\xi\right)\right)^{-}\frac{\mathbf{c}}{\rho\left(\mathbf{c}\right)}\right)\\&=\left(\rho\left(\mathbf{c}\right)\right)^{2}\left(\left(M\left(\xi\right)\mathbf{h}\right)'\left(M\left(\xi\right)\right)^{-}\left(M\left(\xi\right)\mathbf{h}\right)\right)\\&=\left(\rho\left(\mathbf{c}\right)\right)^{2}\ \mathbf{h}'\left(M\left(\xi\right)\right)'\left(M\left(\xi\right)\right)^{-}M\left(\xi\right)\mathbf{h}.\end{aligned}$$

6 Elfving Theorem

Theorem 29 *(Elfving) Denoting*

$$\mathcal{H}:=\left\{X\left(x\right):x\in\left[-1,1\right]\right\}\subseteq\mathbb{R}^{g}$$

the regression range (which under the current hypotheses is a compact set),assume that the vector of coefficients $\mathbf{c}\in\mathbb{R}^{g}$ *lies in* $span\left(\mathcal{H}\right)$ *and has Elfving norm* $\rho\left(\mathbf{c}\right)>0$.

Then a design $\xi \in \mathcal{M}_d([-1,1])$ *is optimal for* $< \mathbf{c}', \theta >$ *in* $\mathcal{M}_d([-1,1])$ *if and only if there exists a function* ϵ *on* $\{X(x) : x \in [-1,1]\}$ *which takes values* ± 1 *on the support of* ξ *and such that such that*

$$\frac{\mathbf{c}}{\rho(\mathbf{c})} = \int_{[-1,1]} (\epsilon(X(x)))(X(x))\,\xi(dx). \tag{31}$$

Furthermore the optimal variance is $(\rho(\mathbf{c}))^2$.

Proof. (from Pukelsheim, pag. 51)

Direct part. Assume that there exists a function ϵ on $\{X(x) : x \in [-1,1]\}$ which on the support of ξ takes values ± 1 such that (31) holds ; we prove that there exists an optimal design for $< \mathbf{c}', \theta >$ in $\mathcal{M}_d([-1,1])$, and that the optimal variance is $(\rho(\mathbf{c}))^2$. We thus prove that $< \mathbf{c}', \theta >$ is estimable and that its Gauss Markov estimator has minimum variance for the measure ξ.

· We prove that $< \mathbf{c}', \theta >$ is estimable. By hypothesis,

$$\frac{\mathbf{c}}{\rho(\mathbf{c})} = \sum_{x \in \mathbf{supp}(\xi)} \xi(x)(\epsilon(X(x)))(X(x)),\ \epsilon(X(x)) = \pm 1. \tag{32}$$

Since $\rho(\mathbf{c}/\rho(\mathbf{c})) = 1$ it follows that $\mathbf{c}/\rho(\mathbf{c}) \in Fr(\mathcal{R})$. Hence there exists a tangent hyperplane which touches $\mathcal{R}$ in $\mathbf{c}/\rho(\mathbf{c})$. Let $\mathbf{h}$ be the vector of the coefficients of this hyperplane as defined in Proposition 22. Then

$$\epsilon(X(x))(X(x))'\mathbf{h} \leq 1 \text{ for any } X(x) \in span\{X(x) : x \in [-1,1]\}. \tag{33}$$

For this vector $\mathbf{h}$, the tangency condition on point $\mathbf{c}/\rho(\mathbf{c})$ provides

$$\frac{\mathbf{c}'}{\rho(\mathbf{c})}\mathbf{h} = 1.$$

Substituting (25) in this latest expression we obtain

$$1 = \frac{\mathbf{c}'}{\rho(\mathbf{c})}\mathbf{h} = \sum_{x \in \mathbf{supp}(\xi)} \xi(x)(\epsilon(X(x)))(X(x))'\mathbf{h}.$$

From (26), we get

$$\xi(x)(\epsilon(X(x)))(X(x))'\mathbf{h} \leq \xi(x)$$

and therefore

$$1 = \frac{\mathbf{c}'}{\rho(\mathbf{c})}\mathbf{h} = \sum_{x \in \mathbf{supp}(\xi)} \xi(x)(\epsilon(X(x)))(X(x))'\mathbf{h} \leq \sum_{x \in \mathbf{supp}(\xi)} \xi(x) = 1.$$

We deduce that

$$\sum_{x \in \mathbf{supp}(\xi)} \xi(x)(\epsilon(X(x)))(X(x))'\mathbf{h} = 1. \tag{34}$$

Assume that

$$(\epsilon(X(x)))(X(x))'\mathbf{h} \neq 1.$$

for some x. Multiply then on both sides by $\xi(x)$ and sum upon all points in the support of ξ. Then we get

$$\sum_{x \in \mathbf{supp}(\xi)} \xi(x)(\epsilon(X(x)))(X(x))'\mathbf{h} \neq 1,$$

a contradiction. Hence $\xi(x)(\epsilon(X(x)))(X(x))'\mathbf{h} = 1$ for all x in the support of ξ.

From (34) we get

$$(X(x))'\mathbf{h} = \frac{1}{\epsilon(X(x))} = \epsilon(X(x)) \tag{35}$$

and therefore, substituting $\epsilon(X(x))$ by $(X(x))'\mathbf{h}$ in (32) and noting that $(X(x))(X(x))' = M(\xi)$, we obtain

$$\begin{aligned}\frac{\mathbf{c}}{\rho(\mathbf{c})} &= \sum_{x \in \mathbf{supp}(\xi)} \xi(x)(\epsilon(X(x)))(X(x)) \\ &= \sum_{x \in \mathbf{supp}(\xi)} \xi(x)(X(x))(X(x))'\mathbf{h} \\ &= \sum_{x \in \mathbf{supp}(\xi)} \xi(x) M(\xi)\mathbf{h} = M(\xi)\mathbf{h}.\end{aligned}$$

This proves that

$$\frac{\mathbf{c}}{\rho(\mathbf{c})} = M(\xi)\mathbf{h} \tag{36}$$

and therefore

$$\frac{\mathbf{c}}{\rho(\mathbf{c})} \in \mathrm{Im} M(\xi),$$

which yields

$$M(\xi) \in \mathcal{A}\left(\frac{\mathbf{c}}{\rho(\mathbf{c})}\right).$$

Hence $< \mathbf{c}', \theta >$ is estimable.

We now prove that the matrix $M(\xi)^-$ is minimal in the Loewner order, which means that ξ provides a minimal variance Gauss Markov estimator of $< \mathbf{c}', \theta >$.

In the frontier point $\mathbf{c}'/\rho(\mathbf{c})$

$$\mathbf{h}' M(\xi)\mathbf{h} = \frac{\mathbf{c}'}{\rho(\mathbf{c})}\mathbf{h} = 1.$$

Hence

$$\mathbf{h}' M(\xi)\mathbf{h} = 1.$$

By Theorem 28, ξ is optimal if and only if

$$\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}=\mathbf{c}N\mathbf{c}'.$$

We evaluate $\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}$. Since $\mathbf{c}/\rho\left(\mathbf{c}\right)=M\left(\xi\right)\mathbf{h}$ using (36)

$$\begin{aligned}\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c} &= \left(\rho\left(\mathbf{c}\right)\right)^{2}\left(\frac{\mathbf{c}'}{\rho\left(\mathbf{c}\right)}\left(M\left(\xi\right)\right)^{-}\frac{\mathbf{c}}{\rho\left(\mathbf{c}\right)}\right)\\ &= \left(\rho\left(\mathbf{c}\right)\right)^{2}\left(\left(M\left(\xi\right)\mathbf{h}\right)'\left(M\left(\xi\right)\right)^{-}\left(M\left(\xi\right)\mathbf{h}\right)\right)\\ &= \left(\rho\left(\mathbf{c}\right)\right)^{2}\ \mathbf{h}'\left(M\left(\xi\right)\right)'\left(M\left(\xi\right)\right)^{-}M\left(\xi\right)\mathbf{h}.\end{aligned}$$

Now $M\left(\xi\right)$ is symmetric and $\left(M\left(\xi\right)\right)\left(M\left(\xi\right)\right)^{-}M\left(\xi\right)=M\left(\xi\right)$; hence since $\mathbf{h}'M\left(\xi\right)\mathbf{h}=1$,

$$\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}=\left(\rho\left(\mathbf{c}\right)\right)^{2}\ \mathbf{h}'M\left(\xi\right)\mathbf{h}=\left(\rho\left(\mathbf{c}\right)\right)^{2}.$$

This proves that if ξ is optimal then the variance of the estimator of $<\mathbf{c},\theta>$ equals $\left(\rho\left(\mathbf{c}\right)\right)^{2}$.

In order to prove optimality recall that

$$\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}=\mathbf{c}N\mathbf{c}'\text{ iff }\begin{cases}\left(X\left(x\right)\right)'NX\left(x\right)=1\text{ for any }x\in\mathbf{supp}\left(\xi\right)\\ M\left(\xi\right)N=\mathbf{c}\left(\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N,\end{cases}$$

by Theorem 28. Since $\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}=\left(\rho\left(\mathbf{c}\right)\right)^{2}$,

$$\frac{\mathbf{c}}{\rho\left(\mathbf{c}\right)}=M\left(\xi\right)\mathbf{h},\ \mathbf{h}\mathbf{h}'=N\text{ and }\mathbf{h}'M\left(\xi\right)\mathbf{h}=1,$$

we have

$$\begin{aligned}\mathbf{c}\left(\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N &= \mathbf{c}\left(\left(\rho\left(\mathbf{c}\right)\right)^{2}\right)^{-1}\mathbf{c}'N\\ &= \frac{1}{\left(\rho\left(\mathbf{c}\right)\right)^{2}}\left(\left(\rho\left(\mathbf{c}\right)\frac{\mathbf{c}}{\rho\left(\mathbf{c}\right)}\right)\left(\frac{\mathbf{c}'}{\rho\left(\mathbf{c}\right)}\rho\left(\mathbf{c}\right)\right)N\right)\\ &= \frac{\mathbf{c}}{\rho\left(\mathbf{c}\right)}\frac{\mathbf{c}'}{\rho\left(\mathbf{c}\right)}N=M\left(\xi\right)\mathbf{h}\left(\mathbf{h}'M\left(\xi\right)\mathbf{h}\right)\mathbf{h}'\\ &= M\left(\xi\right)\mathbf{h}\mathbf{h}'=M\left(\xi\right)N.\end{aligned}$$

Therefore

$$M\left(\xi\right)N=\mathbf{c}\left(\mathbf{c}'\left(M\left(\xi\right)\right)^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N.$$

In order to conclude about optimality it is enough to prove that

$$\left(X\left(x\right)\right)'NX\left(x\right)=1\ \text{ for any }\ x\in\mathbf{supp}\left(\xi\right).$$

By (35)

$$\begin{aligned}(X(x))' NX(x) &= \left((X(x))' \mathbf{h}\right)\left(\mathbf{h}' X(x)\right) \\ &= \epsilon(X(x))\,\epsilon(X(x)) \\ &= 1.\end{aligned}$$

We have proved optimality and also that

$$(\rho(\mathbf{c}))^2 = \min_{\xi} var\left(\widehat{<\mathbf{c}', \theta>}\right).$$

Reciprocal. Let us prove now that if ξ is optimal then

$$\frac{\mathbf{c}}{\rho(\mathbf{c})} = \int_{[-1,1]} (\epsilon(X(x)))\,(X(x))\,\xi(dx).$$

If ξ is the optimal measure to estimate $<\mathbf{c}', \theta>$, then $M(\xi) \in \mathcal{A}(\mathbf{c})$ and $\mathbf{c}'(M(\xi))^{-}\mathbf{c} = (\rho(\mathbf{c}))^2$. Furthermore since ξ is optimal it holds

$$\begin{cases} (X(x))' NX(x) = 1 \text{ for any } x \in \mathbf{supp}(\xi) \\ M(\xi) N = \mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N \end{cases}.$$

Now since $X(x)' NX(x) = 1$ for any $x \in \mathbf{supp}(\xi)$ it holds, using Theorem 28,

$$\begin{aligned}(X(x))' NX(x) &= (X(x))' \mathbf{h}\mathbf{h}' X(x) \\ &= \left((X(x))' \mathbf{h}\right)^2 \quad = 1\end{aligned} \tag{37}$$

for any $x \in \mathbf{supp}(\xi)$.

From $M(\xi) N = \mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N$ we get, multiplying on the right side in (37) by $\frac{\mathbf{h}}{\mathbf{h}'\mathbf{h}}$,

$$M(\xi) N\frac{\mathbf{h}}{\mathbf{h}'\mathbf{h}} = \mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'N\frac{\mathbf{h}}{\mathbf{h}'\mathbf{h}}.$$

Simplifying we have

$$M(\xi)\mathbf{h}\mathbf{h}'\frac{\mathbf{h}}{\mathbf{h}'\mathbf{h}} = \mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'\mathbf{h}\mathbf{h}'\frac{\mathbf{h}}{\mathbf{h}'\mathbf{h}}$$

and

$$M(\xi)\mathbf{h} = \mathbf{c}\left(\mathbf{c}'(M(\xi))^{-}\mathbf{c}\right)^{-1}\mathbf{c}'\mathbf{h}.$$

By the optimality of ξ, it holds $\mathbf{c}'(M(\xi))^{-}\mathbf{c} = (\rho(\mathbf{c}))^2$ and using $\mathbf{c}'\mathbf{h} = \rho(\mathbf{c})$, we get

$$M(\xi)\mathbf{h} = \mathbf{c}\frac{1}{(\rho(\mathbf{c}))^2}\mathbf{c}'\mathbf{h} = \frac{\mathbf{c}}{\rho(\mathbf{c})}.$$

Denote now $\epsilon(X(x)) := (X(x))' \mathbf{h}$. From (37), $1 = \left((X(x))' \mathbf{h}\right)^2$ for any $x \in \mathbf{supp}(\xi)$ it follows that $\sqrt{\epsilon(X(x))} = 1$. Hence $\epsilon(X(x)) = \pm 1$ for $x \in \mathbf{supp}(\xi)$.

Write now

$$\sum_{x \in \mathbf{supp}(\xi)} \xi(x) X(x) \epsilon(X(x)) =_{(a)} \sum_{x \in \mathbf{supp}(\xi)} \xi(x) X(x) (X(x))' \mathbf{h}$$

$$=_{(b)} \sum_{x \in \mathbf{supp}(\xi)} \xi(x) M(\xi) \mathbf{h}$$

$$= M(\xi) \mathbf{h} =_{(c)} \frac{\mathbf{c}}{\rho(\mathbf{c})}.$$

The equality (a) in the above display is obtained substituting $\epsilon(X(x))$ by $(X(x))' \mathbf{h}$. Equality (b) follows from $M(\xi) = X(x)(X(x))'$ and (c) from the fact that $M(\xi) \mathbf{h} = \frac{\mathbf{c}}{\rho(\mathbf{c})}$.

Therefore

$$\frac{\mathbf{c}}{\rho(\mathbf{c})} = \sum_{x \in \mathbf{supp}(\xi)} \xi(x) X(x) \epsilon(X(x)) \text{ with } \epsilon(X(x)) = \pm 1 \text{ for } x \in \mathbf{supp}(\xi).$$

■

Elfving Theorem assesses that the vectors in $\mathcal{R}$ to which an optimal measure is associated are necessarily frontier points of the Elfving set. Indeed clearly $\rho(\mathbf{c}/\rho(\mathbf{c})) = 1$.

In the next section and in the last one, we discuss the results by Kiefer, Wolfowitz and Studden; see [Studden, 1971] .

These authors have characterized optimal designs whose support consists in Chebyshev points. Our starting point is the optimal design which has been described above, through the Elfving Theorem 29.

7 Extension of Hoel - Levine result: Optimal design for a linear c− form

From [Dzyadyk and Shevchuk, 2008] we know by Borel Chebyshev Theorem that any continuous function f defined on a compact set in $\mathbb{R}$ has a uniquely defined best uniform approximation in the class of polynomials with prescribed degree. More generally given a finite class $\{\varphi_0, ..., \varphi_{g-1}\}$ of functions a necessary and sufficient condition for f in $C^{(0)}([-1,1])$ to admit a best uniform approximation $\varphi \in span\{\varphi_0, ..., \varphi_{g-1}\}$ is that $\{\{\varphi_0, ..., \varphi_{g-1}\}\}$ be a Chebyshev system in $C^{(0)}([-1,1])$; this is Haar Theorem. Finally the Borel Chebyshev equioscillation Theorem assesses that the resulting error of approximation by polynomials with fixed degree less or equal $g-1$ takes on its common maximal absolute values on $g+1$ points in $[-1,1]$ with alternating signs.

This important characterization of approximating schemes may be generalized through the following Lemma.

Lemma 30 *(Karlin and Studden) Let* $\{\varphi_0, ..., \varphi_{g-1}\}$ *be a Chebyshev system in* $[-1,1]$ *. Then there exists a unique element* $u(x) := \sum_{j=0}^{g-1} a_j^* \varphi_j(x)$ *in* $V :=$ *span* $\{\varphi_0, ..., \varphi_{g-1}\}$*, which enjoys the following properties:*

1)

$$|u(x)| \leq 1, \text{ for all } x \in [-1,1], \tag{38}$$

2) there exist g *points in* $[-1,1], \widetilde{x}_0, ..., \widetilde{x}_{g-1}$ *such that*

$$-1 \leq \widetilde{x}_0 < ... < \widetilde{x}_{g-1} \leq 1 \text{ and } u(\widetilde{x}_j) = (-1)^{g-1-j}, j = 0, ..., g-1. \tag{39}$$

Proof. see [Karlin and Studden, 1966a] p.792. ■

We now assume that $\mathbf{c} \neq \mathbf{0}$.

We state and prove the following Theorem, which extends Hoel - Levine result. The proof is due to Karlin and Studden.

Let

$$\mathcal{C} := \left\{ \begin{array}{c} \mathbf{c} := (c_0, ..., c_{g-1})' \in \mathbb{R}^g \text{ such that} \\ \det \begin{pmatrix} \varphi_0(x_0) & . & \varphi_0(x_{g-1}) & c_0 \\ \varphi_1(x_0) & . & \varphi_1(x_{g-1}) & c_1 \\ . & . & . & . \\ \varphi_{g-1}(x_0) & . & \varphi_{g-1}(x_{g-1}) & c_{g-1} \end{pmatrix} \neq 0 \end{array} \right\}. \tag{40}$$

This class $\mathcal{C}$ coincides with the estimable $\mathbf{c}-$ forms for designs supported by at least $g+1$ points in $[-1,1]$. See [Kiefer and Wolfowitz, 1965], p 1638.

For any $\mathbf{c}$, consider the projections π_i, on the axes $i = 0, ..., g-1$,

$$\pi_i : \mathcal{C} \to \mathbb{R}, \mathbf{c} \mapsto \pi_i(\mathbf{c}) := c_i$$

and let

$$z = \varphi_i^{\leftarrow}(\pi_i(\mathbf{c})) = \varphi_i^{\leftarrow}(c_i) \tag{41}$$

and therefore $\varphi_i(z) = \varphi_i(\varphi_i^{\leftarrow}(\pi_i(\mathbf{c}))) = c_i$.

Finally denote

$$d(\mathbf{c}, \xi) := \sup_{\mathbf{d}} \frac{< X(x), \mathbf{d} >^2}{< \mathbf{d}, M(\xi)\, \mathbf{d} >}$$

which is the variance of $\widehat{\langle \mathbf{c}, \theta \rangle}$ and

$$\mathbf{B} := \{x \in \mathbb{R} : u^2(x) = 1\}.$$

Theorem 31 *(Optimality) Let* $\{\varphi_0, ..., \varphi_{g-1}\}$ *be a Chebyshev system in* $C^{(0)}([-1,1])$*. Assume that the vector* $\mathbf{c}$ *belongs to the set* $\mathcal{C}$ *defined in (33) and let* $\xi \in \mathcal{M}_d([-1,1])$*. Then with* z *defined in (41)*

1- there exists a unique function

$$x \to u(x) := \sum_{j=0}^{g-1} a_j^* \varphi_j(x) \in span\{\varphi_0, ..., \varphi_{g-1}\}$$

such that

$$d(\mathbf{c},\xi) \geq u^2(z), \text{ for all } \xi \in \mathcal{M}([-1,1]) \ .$$

2-Let $\widetilde{x}_0 < .... < \widetilde{x}_{g-1}$, $j = 0, ..., g-1$, *be* g *points in* $\mathbf{B}$ *such that*

$$u(\widetilde{x}_j) = (-1)^{g-1-j}, j = 0, ..., g-1.$$

Define $l_{\widetilde{x}_j} : j = 0, ..., g-1$ *the Lagrange polynomials with degree* $g-1$ *defined on the nodes* $\widetilde{x}_j$. *Then*

$$d(\mathbf{c},\xi) = u^2(z)$$

if and only if

$$\xi := \xi^*$$

where ξ^* *is the measure with support* $\widetilde{x}_j$, $j = 0, ..., g-1$, *and*

$$\xi^*(\widetilde{x}_j) := \frac{\left|l_{\widetilde{x}_j}(z)\right|}{\sum_{j=0}^{g-1}\left|l_{\widetilde{x}_j}(z)\right|}.$$

3- If there exists $\widetilde{\mathbf{a}} := (\widetilde{a}_0, ..., \widetilde{a}_{g-1})' \in \mathbb{R}^g$ *such that the function* $x \to U(x) := \sum_{j=0}^{g-1} \widetilde{a}_j \varphi_j(x)$ *coincides with the constant function* $1_{[-1,1]} : [-1,1] \to \mathbb{R}, x \to 1_{[-1,1]}(x) = 1$, *then*

$$card\mathbf{B} = g \text{ and } \widetilde{x}_0 = -1, \widetilde{x}_{g-1} = 1.$$

Furthermore

$$d(\mathbf{c},\xi) = u^2(z) \text{ if and only if } \xi = \xi^*.$$

Remark 32 *Statement 1 means that the variance of the estimator of the* $\mathbf{c}$*-form* $< \mathbf{c}, \theta >$ *is bounded by below whatever* ξ. *Statement 2 means that for any vector* $\mathbf{c}$ *there exists an optimal measure* ξ *which provides optimality for the estimate of the* $\mathbf{c}$ *- form* $< \mathbf{c}, \theta >$. *Statement 3 assesses uniqueness.*

Proof. · Statement 2) is proved as follows. There exist g points in $\mathbf{B} \subset [-1,1]$, $\widetilde{x}_0, ..., \widetilde{x}_{g-1}$ such that : $-1 \leq \widetilde{x}_0 < ... < \widetilde{x}_{g-1} \leq 1$ and $u(\widetilde{x}_j) = (-1)^{g-1-j}$, $j = 0, ..., g-1$. Since $\{\varphi_0, ..., \varphi_{g-1}\}$ is a Chebyshev system in $[-1,1]$, the functions $\varphi_0, ..., \varphi_{g-1}$ are linearly independent. The function $u(x) := \sum_{j=0}^{g-1} a_j^* \varphi_j(x)$ is defined in a unique way when known at points $\widetilde{x}_0$, ..., $\widetilde{x}_{g-1}$. Further there exists a unique polynomial P_{g-1}, with degree $g-1$ which assumes the same values as u on $\widetilde{x}_0$, ..., $\widetilde{x}_{g-1}$. Therefore u equals its interpolation polynomial with degree $g-1$. Hence the system $u(\widetilde{x}_j) = P_{g-1}(\widetilde{x}_j)$, $j = 0, ..., g-1$, has a unique solution in the unknown numbers a_j^*'s. Considering the basis which consists in the elementary Lagrange polynomials $\{l_{\widetilde{x}_j} : j = 0, ..., g-1\}$ we may write P_{g-1} , and henceforth u as follows: $u(x) = \sum_{j=0}^{g-1} l_{\widetilde{x}_j}(x) u(\widetilde{x}_j)$. Consider $x = z$; it holds $u(z) = \sum_{j=0}^{g-1} l_{\widetilde{x}_j}(z) u(\widetilde{x}_j)$. Consider now each of the functions

φ_j in $\{\varphi_0, ..., \varphi_{g-1}\}$, for $j = 0, ..., g-1$. We write φ_j in the basis defined by the elementary Lagrange polynomials $l_{\widetilde{x}_j}$ through

$$\varphi_j(z) = \sum_{j=0}^{g-1} l_{\widetilde{x}_j}(z)\, \varphi_j(\widetilde{x}_j),\ j = 0, ..., g-1.$$

Since

$$l_{\widetilde{x}_j}(z) = (-1)^{g-1-j} \left| l_{\widetilde{x}_j}(z) \right|,$$

denoting $\epsilon_j := (-1)^{g-1-j}$, we get

$$\varphi_j(z) = \sum_{j=0}^{g-1} \epsilon_j \left| l_{\widetilde{x}_j}(z) \right| \varphi_j(\widetilde{x}_j),\ j = 0, ..., g-1.$$

Recall that $X(x) := \left(\varphi_0, ..., \varphi_{g-1}\right)'$; the g equalities above write as

$$X(z) = \sum_{j=0}^{g-1} \epsilon_j \left| l_{\widetilde{x}_j}(z) \right| X(\widetilde{x}_j).$$

Denoting

$$\xi := \frac{\left| l_{\widetilde{x}_j}(z) \right|}{\sum_{j=0}^{g-1} \left| l_{\widetilde{x}_j}(z) \right|},\ j = 0, ..., g-1$$

we have

$$\begin{aligned} \frac{X(z)}{\sum_{j=0}^{g-1} \left| l_{\widetilde{x}_j}(z) \right|} &= \sum_{j=0}^{g-1} \epsilon_j \frac{\left| l_{\widetilde{x}_j}(z) \right|}{\sum_{j=0}^{g-1} \left| l_{\widetilde{x}_j}(z) \right|} X(\widetilde{x}_j) \\ &= \sum_{j=0}^{g-1} \epsilon_j \xi_j^* X(\widetilde{x}_j). \end{aligned}$$

Denote

$$\beta := \frac{1}{\sum_{j=0}^{g-1} \left| l_{\widetilde{x}_j}(z) \right|}.$$

We then have

$$\beta X(z) = \sum_{j=0}^{g-1} \epsilon_j \xi_j^* X(\widetilde{x}_j).$$

By Elfving Theorem it follows that if we prove that $\beta X(z) \in Fr(\mathcal{R})$ then $\xi^*$is optimal.

We now prove that $\beta X(z) \in Fr(\mathcal{R})$. This follows from the fact that there exists a tangent hyperplane $\mathcal{R}$ in $\beta X(z)$, i.e.

$$\begin{cases} < (\mathbf{a}^*)', \beta \mathbf{c} >= 1 \\ < \mathbf{y}, \mathbf{a}^* > \leq 1, \text{ for any } \mathbf{y} \in \mathcal{R}. \end{cases}$$

where the vector $\mathbf{a}^*$ defines the hyperplane.

We have

$$\begin{aligned} u(z) &= \sum_{j=0}^{g-1} a_j^* \varphi_j(z) = <(\mathbf{a}^*)', X(z)> \\ &= \sum_{j=0}^{g-1} l_{\widetilde{x}_j}(z)\, u(\widetilde{x}_j). \end{aligned}$$

Therefore

$$\beta u(z) = <(\mathbf{a}^*)', \beta \mathbf{c}>.$$

By definition u alternates sign at points $\widetilde{x}_j$'s. Hence

$$u^2(\widetilde{x}_j) = 1$$

and

$$u(\widetilde{x}_j)\, l_{\widetilde{x}_j}(z) = \left| l_{\widetilde{x}_j}(z) \right|.$$

Now

$$u^2(z) = \left(\sum_{j=0}^{g-1} a_j^* \varphi_j(z) \right)^2 = \left(\sum_{j=0}^{g-1} l_{\widetilde{x}_j}(z)\, u(\widetilde{x}_j) \right)^2 = \left(\sum_{j=0}^{g-1} \left| l_{\widetilde{x}_j}(z) \right| \right)^2 = \frac{1}{\beta^2}$$

i.e.

$$\beta^2 = \frac{1}{u^2(z)}.$$

Clearly $u(z) > 0$. Indeed $u(z) = \sum_{j=0}^{g-1} \left| l_{\widetilde{x}_j}(z) \right|$. Hence $\beta = \frac{1}{u(z)}$ and therefore

$$<(\mathbf{a}^*)', \beta \mathbf{c}> = \beta u(z) = \frac{1}{u(z)} u(z) = 1.$$

By the Property (38) it holds

$$<(\mathbf{a}^*)', X(x)> = \sum_{j=0}^{g-1} a_j^* \varphi_j(x) \leq 1, \text{ for all } x \in [-1, 1].$$

We also have

$$<(\mathbf{a}^*)', -X(x)> = -\sum_{j=0}^{g-1} a_j^* \varphi_j(x) \leq 1, \text{ for any } x \in [-1, 1].$$

Therefore

$$<\mathbf{y}, \mathbf{a}^*> \leq 1, \text{ for all } \mathbf{y} \in \mathcal{R}.$$

Hence the hyperplane defined by the vector $\mathbf{a}^*$ is tangent to $\mathcal{R}$ in $\beta \underline{\mathbf{c}}$.

This proves claim 2).

· We prove 1), i.e. that $d(\mathbf{c},\xi) \geq u^2(z)$.

By Elfving's Theorem , β^{-2} is the minimum value of the variance. Hence

$$\beta^{-2} = \min_{\xi} d(\mathbf{c},\xi).$$

We have just seen that $u^2(z) = 1/\beta^2$. This proves the claim.

· We prove 3).

It holds

$$|u(x)| \leq 1 \quad \text{for } x \in [-1,1] \quad .$$

Also

$$-1 \leq u(x) \leq 1 \quad \text{for } x \in [-1,1]$$

i.e.

$$0 \leq u(x) + 1 \quad \text{for } x \in [-1,1] \quad .$$

It also holds

$$-2 \leq u(x) - 1 \leq 0 \quad \text{for } x \in [-1,1]$$

Hence

$$1 - u(x) \geq 0 \quad \text{for } x \in [-1,1] \quad .$$

Now $1-u(x)$ and $1+u(x)$ are non negative functions for $x \in [-1,1]$. Therefore $|1-u(x)| = 1-u(x)$ and $|1+u(x)| = 1+u(x)$.

From

$$u(\widetilde{x}_{g-1-j}) = (-1)^j \quad j = 0,...,g-1,$$

it follows that

$$1 - u(x) = 0 \text{ for } u(x) = 1$$

and

$$1 + u(x) = 0 \text{ for } u(x) = -1.$$

Therefore, considering the zero's in $(-1,1)$ with multiplicity 2, we have that $1-u(x)$ and $1+u(x)$ have g zero's in $[-1,1]$.

Assume now that there exists a vector of coefficients $\widetilde{\mathbf{a}} := (\widetilde{a}_0,...,\widetilde{a}_{g-1})'$ in $\mathbb{R}^g$, for which the function $U(x) := \sum_{j=0}^{g-1} \widetilde{a}_j \varphi_j(x)$ coincides with the constant function $\mathbf{1}_{[-1,1]} : [-1,1] \to \mathbb{R}$, $x \to \mathbf{1}_{[-1,1]}(x) = 1$. Then the functions

$$\sum_{j=0}^{g-1} \widetilde{a}_j \varphi_j(x) - u(x), \quad \sum_{j=0}^{g-1} \widetilde{a}_j \varphi_j(x) + u(x)$$

have g zero's in $[-1,1]$. These are then the points $\pm 1, \widetilde{x}_{g-1-j}$, $j = 0,...,g-3$.

Since there exists a unique linear combination of the Chebyshev system which assumes value 0 on the points $\pm 1, \widetilde{x}_{g-1-j}$, $j = 0,...,g-3$ it follows that

$$card\ (\mathbf{B}) = g.$$

We now prove that

$$d(\mathbf{c},\xi) = u^2(z) \quad \text{if and only if} \quad \xi = \xi^*,$$

where ξ^* is the measure defined at point 3) of this Theorem.

Assume $d(\mathbf{c},\xi) = u^2(z)$ and consider a generic finitely supported probability measure ξ whose support contains strictly the g points in $\mathbf{B}$. We prove that the variance associated to ξ is not optimal.

Since $c_i = \varphi_i(z)$,

$$\begin{aligned} d(\mathbf{c},\xi) &:= \sup_{\mathbf{d}} \frac{<\mathbf{c},\mathbf{d}>^2}{<\mathbf{d},M(\xi)\,\mathbf{d}>} \\ &= \sup_{\mathbf{d}} \frac{\left(\sum_{i=0}^{g-1} d_i\varphi_i(z)\right)^2}{\int_{[-1,1]}\left(\sum_{i=0}^{g-1} d_i\varphi_i(x)\right)^2 d\xi(x)} \\ &\geq \frac{u^2(x)}{\int_{[-1,1]} u^2(x)\,d\xi(x)}. \end{aligned}$$

But

$$u^2(x) < 1$$

and therefore

$$\frac{1}{u^2(x)} > 1.$$

It follows that

$$d(\mathbf{c},\xi) > u^2(x).$$

The inequality is strict since the support of ξ contains points not in $\mathbf{B}$. Using Karlin - Studden Theorem 30, and existence of U, $d(\mathbf{c},\xi) = u^2(x)$ if and only if ξ has support $\mathbf{B}$. Hence $d(\mathbf{c},\xi) > u^2(x)$. Therefore ξ is not optimal. Hence the optimal measure has support $\mathbf{B}$.

We now identify the support of the optimal measure, and then its masses at those points. The measure is optimal if and only if it has all its mass at points in $\mathbf{B}$, $\pm 1, \widetilde{x}_{g-1-j}$, $j = 0, ..., g-3$. By Elfving's Theorem it holds

$$\begin{aligned} \sum_{j=0}^{g-1} \epsilon_j p_j X(\widetilde{x}_j) &= \beta X(z), \\ \beta &= \frac{1}{\sum_{j=0}^{g-1} \left|l_{\widetilde{x}_j}(z)\right|}. \end{aligned}$$

This can be seen as a system of g equations with g unknown variables $\epsilon_j\ p_j$, $j = 0, ..., g-1$. It has a unique solution whenever

$$\det \begin{pmatrix} \varphi_0(\widetilde{x}_0) & \cdot & \varphi_{g-1}(\widetilde{x}_0) \\ \cdot & \cdot & \cdot \\ \varphi_0(\widetilde{x}_{g-1}) & \cdot & \varphi_{g-1}(\widetilde{x}_{g-1}) \end{pmatrix} \neq 0$$

which holds by independence of the elements in the Chebyshev system.

The solution writes

$$p_j := \frac{\left|l_{\widetilde{x}_j}(z)\right|}{\sum_{j=0}^{g-1}\left|l_{\widetilde{x}_j}(z)\right|}, \epsilon_j := (-1)^{g-1-j}, j = 0, ..., g-1.$$

Hence ξ is optimal if and only if $\xi = \xi^*$. We have proved claim 3). ■

Remark 33 *Considering* $\varphi_j(x) = x^j$, $j = 0, ..., g-1$, $z \in \mathbb{R}\backslash[-1,1]$, *the optimal measure is given by*

$$\xi^* := \frac{\left|l_{\widetilde{x}_j}(z)\right|}{\sum_{j=0}^{g-1}\left|l_{\widetilde{x}_j}(z)\right|}$$

with

$$\widetilde{x}_j := -\cos\left(\frac{j\pi}{g-1}\right)$$

for $j = 0, ..., g-1$. *Furthermore* $d(z,\xi) \geq T_{g-1}^2(z)$ *where* T_{g-1} *is the Chebyshev polynomial of first kind with degree* $g-1$. *Hence*

$$T_{g-1}(x) := \cos((g-1)\arccos x).$$

The polynomial T_{g-1} *has the same role as the function* $u(x) = \sum_{j=0}^{g-1} a_j^* \varphi_j(x)$. *Hence the result by Hoel - Levine appears as a special case of the above Theorem.*

Remark 34 *Considering* $\varphi_j(x) = x^j$, $j = 0, ..., g-1$, $z \in \mathbb{R} \setminus [-1,1]$, *the optimal measure is given by*

$$\xi^* := \frac{\left|l_{\widetilde{x}_j}(z)\right|}{\sum_{j=0}^{g-1}\left|l_{\widetilde{x}_j}(z)\right|}$$

with

$$\widetilde{x}_j := -\cos\left(\frac{j\pi}{g-1}\right)$$

for $j = 0, ..., g-1$. *Furthermore* $d(z,\xi) \geq T_{g-1}^2(z)$ *where* T_{g-1} *is the Chebyshev polynomial of first kind with degree* $g-1$. *Hence*

$$T_{g-1}(x) := \cos((g-1)\arccos x).$$

The polynomial T_{g-1} *has the same role as the function* $u(x) = \sum_{j=0}^{g-1} a_j^* \varphi_j(x)$. *Hence the result by Hoel - Levine appears as a special case of the above Theorem.*

References

[Broniatowski and Celant, 2015] M Broniatowski and G Celant. Optimal extrapolation design for the chebyshev regression. *Annales de l'ISUP*, 59(3):3–22, 2015.

[Celant and Broniatowski, 2016] Giorgio Celant and Michel Broniatowski. *Interpolation and extrapolation optimal designs. 1. Polynomial regression and approximation theory.* Mathematics and Statistics Series. ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[Dzyadyk and Shevchuk, 2008] V. K. Dzyadyk and I. A. Shevchuk. *Theory of uniform approximation of functions by polynomials.* Walter de Gruyter GmbH & Co. KG, Berlin, 2008. Translated from the Russian by Dmitry V. Malyshev, Peter V. Malyshev and Vladimir V. Gorunovich.

[Elfving, 1952] G. Elfving. Optimum allocation in linear regression theory. *Ann. Math. Statistics*, 23:255–262, 1952.

[Elfving, 1954] G. Elfving. Geometric allocation theory. *Skand. Aktuarietidskr.*, 37:170–190 (1955), 1954.

[Hoel, 1966] P. Hoel. A simple solution for optimal Chebyshev regression extrapolation. *Ann. Math. Statist.*, 37:720–725, 1966.

[Karlin and Studden, 1966a] S. Karlin and W. J. Studden. Optimal experimental designs. *Ann. Math. Statist.*, 37:783–815, 1966.

[Karlin and Studden, 1966b] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics.* Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[Kiefer and Wolfowitz, 1964] J. Kiefer and J. Wolfowitz. Optimum extrapolation and interpolation designs. I, II. *Ann. Inst. Statist. Math. 16 (1964), 79–108; ibid.*, 16:295–303, 1964.

[Kiefer and Wolfowitz, 1965] J. Kiefer and J. Wolfowitz. On a theorem of Hoel and Levine on extrapolation designs. *Ann. Math. Statist.*, 36:1627–1655, 1965.

[Lang, 1989] S. Lang. *Linear algebra.* Undergraduate Texts in Mathematics. Springer-Verlag, New York, third edition, 1989.

[Pukelsheim, 2006] F. Pukelsheim. *Optimal design of experiments*, volume 50 of *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. Reprint of the 1993 original.

[Studden, 1971] W. J. Studden. Optimal designs and spline regression. In *Optimizing methods in statistics (Proc. Sympos., Ohio State Univ., Columbus, Ohio, 1971)*, pages 63–76. Academic Press, New York, 1971.

[Studden, 2005] W. J. Studden. Elfving's theorem revisited. *J. Statist. Plann. Inference*, 130(1-2):85–94, 2005.

[van Tiel, 1984] J. van Tiel. *Convex analysis.* John Wiley & Sons, Inc., New York, 1984. An introductory text.