



# Optimal extrapolation design for the Chebyshev regression

Michel Broniatowski, Giorgio Celant

## ► To cite this version:

Michel Broniatowski, Giorgio Celant. Optimal extrapolation design for the Chebyshev regression. Annales de l'ISUP, 2015, 59 (3), pp.3-22. hal-00984010

**HAL Id: hal-00984010**

**<https://hal.science/hal-00984010>**

Submitted on 26 Apr 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimal extrapolation design for the Chebyshev regression

Michel Broniatowski<sup>(1)</sup>, Giorgio Celant<sup>(2,\*)</sup>

<sup>(1)</sup> LSTA, Université Pierre et Marie Curie, Paris, France

<sup>(2)</sup> Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Italy

<sup>(\*)</sup> Corresponding author: [giorgio.celant@stat.unipd.it](mailto:giorgio.celant@stat.unipd.it)

## Abstract

This paper introduces optimal designs in the context of a regression model when the regression function is assumed to be generated by a Chebyshev system of functions. The criterion for optimality is the variance of a Gauss Markov estimator for an extrapolated value.

**Key words:** Chebyshev system; optimal design; extrapolation design; Borel-Chebyshev Theorem

## 1 Introduction

This paper deals with a natural extension of the Hoel Levine optimal extrapolation design, as described in [Hoel, 1966]. We recall that this classical design results as a consequence of the following fact.

A design is defined as a discrete probability measure  $\xi$  on a set of measurements points  $x_0, \dots, x_{g-1}$  which for notational convenience belong to the observable environmental set  $[-1, 1]$ , denoting  $n_i/n := \xi(x_i)$  the frequency of replications of the experiment to be performed at point  $x_i$ ,  $0 \leq i \leq g-1$ , where the  $n_i$ 's satisfy  $n_0 + \dots + n_{g-1} = n$ . The points  $x_i$ 's are the nodes of the design, and  $\xi(x_i)$  is the so-called frequency of the design at node  $x_i$ . Recall that the model writes

$$Y(x) = f(x) + \epsilon(x)$$

for  $x$  in  $[-1, 1]$ , the real valued function  $f$  is unknown but belongs to a specified class of functions, and the random variable  $\epsilon(x)$  is centered, with a finite variance, in the present context. Observations are performed under the design, with the constraint

$$n_1 + \dots + n_{g-1} = n$$

on the global budget of the design. Replications of the  $n_i$  measurements  $Y_j(x_i)$ ,  $1 \leq j \leq n_i$  are independent. Independence also holds from node to node, which is to assume that all measurements errors due to the r.v.'s  $\epsilon(x)$  are independent. The model is supposed to be homoscedasticity; hence the variance of  $\epsilon(x)$  may not depend on  $x$ .

For a given  $c$  not in  $[-1, 1]$  consider an estimate of  $f(c)$  with smallest variance among all unbiased estimators of  $f(c)$  which are linear functions of the observations  $Y_j(x_i)$ ,  $1 \leq j \leq n_i, 0 \leq i \leq g-1$ , hence under a given design  $\xi$ . An optimal design achieves the minimal variance among all such designs. This design is achieved by the Hoel Levine design when the function  $f$  is assumed to belong to the class of all polynomials defined on  $\mathbb{R}$  with degree less or equal  $g-1$ , hence to the span of the class of monomials  $\{1, x, \dots, x^{g-1}\}$ .

The main mathematical argument in order to obtain the Hoel Levine design lies in the solution of the following basic question: find a polynomial with equioscillations in  $g+1$  points in  $[-1, 1]$  which assumes maximal absolute values all equal to 1 at those points. Up to a multiplicative constant such a polynomial results as the best polynomial approximation of the null function on  $[-1, 1]$  by polynomials with degree  $g-1$ . Existence and uniqueness of this polynomial follows from the Borel-Chebyshev Theorem. We refer to [Dzyadyk and Shevchuk, 2008] for details and derivation of these results.

The aim is now to provide a larger context for similar questions, assuming that the function  $f$  may belong to some other functional class, still in a finitely generated set of functions.

**Definition 1** *The system of functions  $(\varphi_0, \dots, \varphi_{g-1})$  in  $\mathcal{C}(\mathbb{R})$  is a Chebyshev (or Haar) system on  $[-1, 1]$  when*

- 1)  $(\varphi_0, \dots, \varphi_{g-1})$  are linearly independent
- 2) Any equation

$$a_0\varphi_0(x) + \dots + a_{g-1}\varphi_{g-1}(x) = 0$$

with  $(a_0, \dots, a_{g-1}) \neq (0, \dots, 0)$  has at most  $g$  roots in  $[-1, 1]$ .

Denote

$$V := \text{span} \{ \varphi_0, \dots, \varphi_{g-1} \} \subset \mathcal{C}([-1, 1])$$

the linear space generated by the Chebyshev system  $(\varphi_0, \dots, \varphi_{g-1})$ .

Haar Theorem (see [Dzyadyk and Shevchuk, 2008]) states that the two following assertions are equivalent:

- a)  $\{ \varphi_0, \dots, \varphi_{g-1} \}$  is a Chebyshev system in  $\mathcal{C}([-1, 1])$

b) for any  $f$  in  $\mathcal{C}([-1, 1])$  there exists a unique best uniform approximation in  $V$ .

In the sequel we assume that the system  $\{\varphi_0, \dots, \varphi_{g-1}\}$  is a Chebyshev system in  $\mathcal{C}([-1, 1])$  and in  $\mathcal{C}([-1, c])$  with  $c > 1$ . This implies that no non null linear combination of the  $\varphi_i$ 's may have roots in  $(1, c]$ .

We also make use of the following result. The following properties are equivalent

**Proposition 2** 1)  $\{\varphi_0, \dots, \varphi_{g-1}\}$  is a Chebyshev system;

2) for any set of  $g$  points  $(x_0, \dots, x_{g-1})$  in  $[-1, 1]$  such that  $x_i \neq x_j$ , and for any  $(y_0, \dots, y_{g-1})$  in  $\mathbb{R}^g$ , there exists a unique function  $g$  in  $V$  such that  $g(x_k) = y_k$ ;

3) for any  $g$  points  $(x_0, \dots, x_{g-1})$  in  $[-1, 1]$  such that  $x_i \neq x_j$ , the determinant

$$\Gamma := \det G, \quad G := \begin{pmatrix} \varphi_0(x_0) & \varphi_0(x_j) & \varphi_0(x_{g-1}) \\ \vdots & \vdots & \vdots \\ \varphi_i(x_0) & \varphi_i(x_j) & \varphi_i(x_{g-1}) \\ \vdots & \vdots & \vdots \\ \varphi_{g-1}(x_0) & \varphi_{g-1}(x_j) & \varphi_{g-1}(x_{g-1}) \end{pmatrix}.$$

does not equal 0.

**Proof.** Assume 3) holds. With the set of  $g$  points  $(x_0, \dots, x_{g-1})$  in  $[-1, 1]$  such that  $x_i \neq x_j$ ,  $\Gamma = 0$  iff the matrix  $G$  is not invertible, which is to say that the system of equations defined through  $0 = \sum_{i=0}^{g-1} a_i g_i(x_j)$ ,  $j = 0, \dots, g-1$ , admits a solution  $\mathbf{a}^* := (a_0^*, \dots, a_{g-1}^*)$  different from  $(0, \dots, 0)$  in  $\mathbb{R}^g$ . Define  $g := \sum_{i=0}^{g-1} a_i^* g_i$ , an element in  $V$  which is not the function  $x \rightarrow 0$ . Since  $\sum_{i=0}^{g-1} a_i^* g_i(x) = 0$  for  $x$  in  $\{x_0, \dots, x_{g-1}\}$  it follows that  $g$  has  $g$  distinct roots in  $[-1, 1]$ . It follows that whenever  $\Gamma = 0$ ,  $\{\varphi_0, \dots, \varphi_{g-1}\}$  is not a Chebyshev system. It follows that 3) is equivalent to 1). Now 2) is equivalent to 3). Indeed when  $G$  is invertible then for any  $(y_0, \dots, y_{g-1})$  in  $\mathbb{R}^{g-1}$  the system  $\sum_{i=0}^{g-1} a_i g_i(x_j) = y_j$ ,  $j = 0, \dots, g-1$ , has a unique solution, which means that there is a unique  $g$  in  $V$  with  $g(x_j) = y_j$  for all  $j$ . ■

We therefore introduce the basic definition

**Definition 3** A regression model

$$Y(x) = f(x) + \epsilon(x)$$

is a Chebyshev regression model iff  $f$  belongs to  $V := \text{span}\{\varphi_0, \dots, \varphi_{g-1}\}$  where  $(\varphi_0, \dots, \varphi_{g-1})$  is Chebyshev system (or Haar system) of functions in  $\mathcal{C}([-1, 1])$ .

The following result stands as a generalization of the Borel Chebyshev Theorem and improves on the Haar Theorem

**Theorem 4** .(Generalization of Chebyshev-Borel Theorem) *Let*

$$\{\varphi_0, \dots, \varphi_{g-1}\}$$

*be a Chebyshev system on  $[-1, 1]$  , and  $g$  is any function in  $\mathcal{C}([-1, 1])$  . Then there exists a unique function  $h$  in  $V := \text{span}\{\varphi_0, \dots, \varphi_{g-1}\}$  defined on  $[-1, 1]$  , which achieves*

$$\sup_{x \in [-1, 1]} |g(x) - h(x)| = \inf_{f \in V} \sup_{x \in [-1, 1]} |g(x) - f(x)| .$$

*Furthermore  $h$  is the only function in  $V$  such that  $p := g - h$  attains its unique maximal values in at least  $g + 1$  points in  $[-1, 1]$ ; the sign of  $p$  on those points alternates.*

**Proof.** See [Achieser, 1992]. ■

**Remark 5** *The above function  $h$  plays a similar role as the function  $T_{g-1}$  (Chebyshev polynomial of the first kind) in the polynomial regression case; see [Broniatowski and Celant, 2014].*

The notation  $\mathcal{M}_d([-1, 1])$  designates the class of all discrete probability measures with support in  $[-1, 1]$  .

The aim of this paper is to present the contribution of Hoel [Hoel, 1966] to the construction of optimal designs for the extrapolated value of the regression function as treated by Kiefer and Wolfowitz [Kiefer and Wolfowitz, 1965]. The model and the Gauss Markov estimator are defined in the next Section. An orthogonalization procedure allows to express the extrapolated value as a parameter in an adequate regression model. Finally the support of the optimal design will be obtained through geometrical arguments; the number of replications of the experiments on the nodes will then be deduced.

## 2 The model and Gauss Markov estimator

We consider a Chebyshev system on  $[-1, 1]$

$$\{\varphi_0, \dots, \varphi_{g-1}\} .$$

For any  $x \in [-1, 1]$  we assume that we may observe a r.v.  $y(x)$ , such that, denoting  $\theta := (\theta_0, \dots, \theta_{g-1})'$

$$f(x) := E(Y(x)) = \sum_{j=0}^{g-1} \theta_j \varphi_j(x) = (X(x))' \theta. \quad (1)$$

We notice that the function

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto f(x)$$

is continuous on  $\mathbb{R}$ . Indeed since the system of the  $g$  equations in  $\theta$

$$\begin{cases} f(x_0) = \sum_{j=0}^{g-1} \theta_j \varphi_j(x_0) \\ ..... \\ f(x_{g-1}) = \sum_{j=0}^{g-1} \theta_j \varphi_j(x_{g-1}) \end{cases}$$

has a unique solution whenever  $(f(x_0), \dots, f(x_{g-1}))'$  is known, for any  $(x_0, \dots, x_{g-1})' \in [-1, 1]^g$  with  $-1 \leq x_0 < \dots < x_{g-1} \leq 1$ , the function  $f$  can be extended on  $\mathbb{R}$ ; this extension is continuous since so are the  $\varphi_i$ 's.

Recall that the measurements can be performed only on  $[-1, 1]$ , and not for  $|x| > 1$ .

## 2.1 Examples of Chebyshev systems

Here is a short list of classical chebyshev systems. We refer to the classical treatises of [Karlin and Studden, 1966] for a extensive study of those systems and their applications in analysis and in statistics.

- a)  $\{\varphi_0(x) = 1, \varphi_1(x) = x^3\}$  is a Chebyshev system on whole  $\mathbb{R}$ ,
- b)  $\{\varphi_0(x) = 1; \varphi_1(x) = \frac{1}{x^3}\}$  is a Chebyshev system on  $(0, +\infty)$ ,
- c)  $\{1, \cos x, \cos 2x, \dots, \cos nx\}$  is a Chebyshev system on  $[0, \pi)$ ,
- d)  $\{1, \sin x, \cos x, \sin 2x, \cos 2x, \dots, \sin nx, \cos nx\}$  is a Chebyshev system on  $\mathbb{R}/2\pi$
- e)  $\{\sin x, \sin 2x, \dots, \sin nx\}$  is a Chebyshev system on  $[0, \pi)$ ,
- f)  $\{\varphi_0(x) = x^2 - x, \varphi_1(x) = x^2 + x, \varphi_2(x) = x^2 + 1\}$  is a Chebyshev system on  $\mathbb{R}$ ,
- g)  $\{x^{a_0}, \dots, x^{a_n}, \text{ where } 0 = a_0 < \dots < a_n\}$  is a Chebyshev system on  $[0, +\infty)$ ,
- h)  $\{e^{a_0x}, \dots, e^{a_nx}, \text{ where } 0 = a_0 < \dots < a_n\}$  is a Chebyshev system on  $\mathbb{R}$ ,
- i)  $\{1, \sinh x, \cosh x, \dots, \sinh nx, \dots, \cosh nx\}$  is a Chebyshev system on  $\mathbb{R}$ .

j)  $\{(x+a_0)^{-1}, \dots, (x+a_n)^{-1}, \text{ where } 0 = a_0 < \dots < a_n\}$  is a Chebyshev system on  $[0, +\infty)$ ,

k)  $\{1, \log x, x, x \log x, x^2, x^2 \log x, \dots, x^n, x^n \log x\}$  is a Chebyshev system on  $(0, \infty)$ , ....

Finally note that being a Chebyshev system is a linear property; indeed if  $(\varphi_0, \dots, \varphi_{g-1})$  is a Chebyshev system then any other basis of  $\text{span}\{\varphi_0, \dots, \varphi_{g-1}\}$  is a Chebyshev system.

## 2.2 Description of the dataset coming from the experiment

Given the set of nodes  $-1 \leq x_0 < \dots < x_{g-1} \leq 1$ , the experiment is described through the following measurements

[illegible]

or through the more synthetic form, with

$$\begin{aligned} \underline{Y}(x_i) &:= (Y_1(x_i), \dots, Y_{n_j}(x_i))', \quad X(x_i) := (\varphi_0(x_i), \dots, \varphi_{g-1}(x_i))', \\ \underline{\varepsilon}(x_i) &:= (\varepsilon_1(x_i), \dots, \varepsilon_{n_i}(x_i))', \end{aligned}$$

$$\underline{Y}(x_i) = \begin{pmatrix} \varphi_0(x_i) & \dots & \varphi_{g-1}(x_i) \\ \dots & \dots & \dots \\ \varphi_0(x_i) & \dots & \varphi_{g-1}(x_i) \end{pmatrix} \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_{g-1} \end{pmatrix} + \underline{\varepsilon}(x_i), \quad i = 0, \dots, g-1.$$

Denote,

$$X_i := \begin{pmatrix} \varphi_0(x_i) & \dots & \varphi_{g-1}(x_i) \\ \dots & \dots & \dots \\ \varphi_0(x_i) & \dots & \varphi_{g-1}(x_i) \end{pmatrix}.$$

The matrix  $X_i$  has  $n_i$  lines and  $g$  columns. All lines of  $X$  equal  $(X(x_i))'$ .

Denote

$$\mathcal{H} := \text{Im } X(x) := \{X(x) \in \mathbb{R}^g : x \in [-1, 1]\}. \quad (2)$$

The set  $\mathcal{H}$  is called the regression range.

It may be at time convenient to attribute distinct indices to the same  $x_j$  when repeated  $n_j$  times.

The discrete measure defined through

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

with

$$n_0 + \dots + n_{g-1} = n$$

will hence be written as

$$t_1, \dots, t_n \quad (3)$$

with  $t_1 = t_2 = \dots = t_{n_0} = x_0, \dots, t_{n_0+\dots+n_{g-2}+1} = t_{n_0+\dots+n_{g-2}+2} = \dots = t_{n_0+\dots+n_{g-2}+n_{g-1}} = x_{g-1}$ ; hence  $t_1, \dots, t_n$  indicates the same point  $x_0$  repeated  $n_0$  times, etc.

The system which describes the  $n$  observations writes therefore as

$$Y = C \underline{\theta} + \underline{\varepsilon}$$

where

$$\mathbf{Y} := \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad C := \begin{pmatrix} \varphi_0(t_1) & \dots & \varphi_{g-1}(t_1) \\ \vdots & \dots & \vdots \\ \varphi_0(t_i) & \dots & \varphi_{g-1}(t_i) \\ \vdots & \dots & \vdots \\ \varphi_0(t_n) & \dots & \varphi_{g-1}(t_n) \end{pmatrix},$$

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

$$E(\mathbf{Y}) = C \underline{\theta}, \quad \text{var}(\underline{\varepsilon}) = \sigma^2 I_n,$$

and  $I_n$  is the Identity matrix of order  $n$ .

The Gauss Markov estimator of  $f(x) = E(y(x))$  is the solution of the linear system



$$X_i' X_i \theta = X_i' \underline{Y}(x_i), \quad i = 0, \dots, g-1.$$

It holds

$$X_i' X_i = \begin{pmatrix} \varphi_0(x_i) & \dots & \varphi_0(x_i) \\ \dots & \dots & \dots \\ \varphi_{g-1}(x_i) & \dots & \varphi_{g-1}(x_i) \end{pmatrix} \begin{pmatrix} \varphi_0(x_i) & \dots & \varphi_{g-1}(x_i) \\ \dots & \dots & \dots \\ \varphi_0(x_i) & \dots & \varphi_{g-1}(x_i) \end{pmatrix} = n_i M_i$$

where

$$M_i := \begin{pmatrix} (\varphi_0(x_i))^2 & \dots & \varphi_0(x_i) \varphi_k(x_i) & \dots & \varphi_0(x_i) \varphi_{g-1}(x_i) \\ \dots & \dots & \dots & \dots & \dots \\ \varphi_h(x_i) \varphi_0(x_i) & \dots & \varphi_h(x_i) \varphi_k(x_i) & \dots & \varphi_h(x_i) \varphi_{g-1}(x_i) \\ \dots & \dots & \dots & \dots & \dots \\ \varphi_{g-2}(x_i) \varphi_0(x_i) & \dots & \varphi_{g-2}(x_i) \varphi_k(x_i) & \dots & \varphi_{g-2}(x_i) \varphi_{g-1}(x_i) \\ \varphi_{g-1}(x_i) \varphi_0(x_i) & \dots & \varphi_{g-1}(x_i) \varphi_k(x_i) & \dots & (\varphi_{g-1}(x_i))^2 \end{pmatrix}.$$

We have

$$M_i = X(x_i) X'(x_i).$$

In

$$X(x_i) X'(x_i) \theta = X_i' \underline{Y}(x_i), \quad i = 0, \dots, g-1$$

sum both sides with respect to  $i$  to obtain

$$\sum_{i=0}^{g-1} X_i X_i' \theta = \sum_{i=0}^{g-1} X_i' \underline{Y}(x_i).$$

Therefore

$$n \left( \sum_{i=0}^{g-1} \frac{n_i}{n} M_i \right) \theta = \sum_{i=0}^{g-1} X_i' \underline{Y}(x_i).$$

Denote

$$\xi_i := \xi(x) := \begin{cases} \frac{n_i}{n} & \text{if } x = x_i \\ 0 & \text{if } x \notin \{x_0, \dots, x_{g-1}\} \end{cases}.$$

The matrix

$$M(\xi) := \sum_{i=0}^{g-1} \frac{n_i}{n} M_i = \sum_{i=0}^{g-1} \xi_i M_i \quad (4)$$

is the *moment matrix* of the measure  $\xi$ .

By definition

$$\text{supp}(\xi) = \{x_0, \dots, x_{g-1}\}.$$

Since

$$M_i = X(x_i) X'(x_i)$$

we may write

$$M(\xi) = \sum_{i=0}^{g-1} \xi_i M_i = \sum_{i=0}^{g-1} \xi_i X(x_i) X'(x_i) = \int_{[-1,1]} X(x) X'(x) d\xi(x).$$

Specific study of this matrix is needed for the estimation of linear forms of the coefficients  $\theta_i$ 's. This area has been developed by Elfving see e.g. [Pukelsheim, 1993]), out of the scope of the present paper.

### 3 An expression of the extrapolated value through an orthogonalization procedure

We will consider an alternative way, developed by Kiefer and Wolfowitz [Kiefer and Wolfowitz, 1965] as follows. It has the main advantage that up to a coefficient  $\gamma_{g-1}$  which depends on the values of  $f$  on the  $x_j$ 's, the estimate of  $f(c)$  is  $\varphi_{g-1}(c)$ . It follows that only the coefficient  $\gamma_{g-1}$  has to be estimated, a clear advantage. Recall that  $c$  does not belong to  $[-1, 1]$ .

It is more convenient, at this stage, to introduce the following notation. It will be assumed that  $n$  measurements of  $Y$  are performed, namely

$$Y(t_1), \dots, Y(t_n)$$

where the  $t_i$ 's belong to  $[-1, 1]$ . The points of measurement  $t_1, \dots, t_n$  might be distinct or not, as defined in (3). Obviously when defining the optimal design with nodes  $x_0, \dots, x_{g-1}$ , then  $n_j$  values of the  $t_i$ 's coincide on  $x_j$  for  $0 \leq j \leq g-1$ . In order to define the estimator, and not the design, it is however more convenient to differentiate between all the measurements  $Y(t_i)$ ,  $1 \leq i \leq n$ . This allows to inherit from the classical geometric least square formalism.

We consider the basis of  $V$  defined as follows:

Set for all  $j$  between 0 and  $g-2$

$$h_j(x) := \varphi_j(x) - \frac{\varphi_j(c)}{\varphi_{g-1}(c)} \varphi_{g-1}(x) \quad (5)$$

and

$$h_{g-1}(x) := \varphi_{g-1}(x).$$

Clearly  $(h_0, \dots, h_{g-1})$  generate  $V$ . Also  $(h_0, \dots, h_{g-1})$  is a Chebyshev system on  $[-1, c]$ .

Denote  $(\gamma_0, \dots, \gamma_{g-1})$  the coordinates of  $f$  on  $(h_0, \dots, h_{g-1})$ , namely

$$f(x) = \sum_{j=0}^{g-1} \gamma_j h_j(x).$$

We evaluate the coefficients  $\gamma_j$  with respect to the  $\theta'_k$ s defined in (1). It holds

$$\gamma_j := \theta_j \text{ for } j = 0, \dots, g-2$$

and

$$\gamma_{g-1} := \frac{\sum_{j=0}^{g-1} \theta_j \varphi_j(c)}{\varphi_{g-1}(c)}$$

assuming  $\varphi_{g-1}(c) \neq 0$ , and obviously we have

$$f(x) = \sum_{j=0}^{g-1} \gamma_j h_j(x) = \sum_{j=0}^{g-1} \theta_j \varphi_j(x).$$

In  $x = c$  we get

$$f(c) := \sum_{j=0}^{g-1} \gamma_j h_j(c) = \sum_{j=0}^{g-1} \theta_j \varphi_j(c).$$

By the definition of  $\gamma_{g-1}$  we have

$$\gamma_{g-1} := \frac{\sum_{j=0}^{g-1} \theta_j \varphi_j(c)}{\varphi_{g-1}(c)}$$

and therefore we have proved

**Lemma 6**

$$f(c) = \gamma_{g-1} \varphi_{g-1}(c). \quad (6)$$

## 4 The Gauss Markov estimator of the extrapolated value

It holds

$$f(c) = \sum_{i=1}^{g-1} \theta_i \varphi_i(c)$$

where the  $\theta_i$ 's are defined through  $g$  equations of the form

$$f(x_j) = \sum_{i=0}^{g-1} \theta_i \varphi_i(x_j)$$

with  $-1 \leq x_j \leq 1$  for all  $0 \leq j \leq g-1$ .

Replace  $f(x_j)$  by its estimate

$$\widehat{f(x_j)} := \frac{1}{n_j} \sum_{i=1}^{n_j} Y_i(x_j).$$

Under the present model,  $\widehat{f(x_j)}$  is an unbiased estimate of  $f(x_j)$ . Determine  $\widehat{\theta}_i$  though the system defined by

$$\widehat{f(x_j)} = \sum_{i=0}^{g-1} \hat{\theta}_i \varphi_i(x_j).$$

The resulting  $\hat{\theta}_i$ 's are unbiased and so is

$$\widehat{f(c)} = \sum_{i=0}^{g-1} \widehat{\theta}_i \varphi_i(c).$$

The natural optimality criterion associated to this procedure is the variance of the estimate  $\widehat{f(c)}$  which depends on the location of the nodes and on the weights  $n_j$ 's.

We now write the above Gauss Markov estimator of  $f(c)$  on the new basis  $(h_0, \dots, h_{g-1})$ . Substituting the function  $f$  by its expansion on the basis  $(h_0, \dots, h_{g-1})$  the model write as

[illegible]

because of (5),

$$\underline{Y}(\mathbf{t}) = T\theta + \underline{\varepsilon}$$

where  $\mathbf{t} := (t_1, \dots, t_n)'$

$$T := \begin{pmatrix} h_0(t_1) & \dots & h_{g-2}(t_1) & \varphi_{g-1}(t_1) \\ \vdots & \ddots & \vdots & \vdots \\ h_0(t_i) & \dots & h_{g-2}(t_i) & \varphi_{g-1}(t_i) \\ \vdots & \ddots & \vdots & \vdots \\ h_0(t_n) & \dots & h_{g-2}(t_n) & \varphi_{g-1}(t_n) \end{pmatrix}, \theta := \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_{g-2} \\ \gamma_{g-1} \end{pmatrix}, \underline{\varepsilon} := \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Recall that we intend to estimate  $\gamma_{g-1}$ . We make a further change of the basis of  $V$ . We introduce a vector  $G_{g-1}$ , which together with  $h_0, \dots, h_{g-2}$  will produce a basis  $(h_0, \dots, h_{g-2}, G_{g-1})$  for which the vector  $G_{g-1}$  is orthogonal to any of the  $h_j$ ,  $0 \leq j \leq g-2$ . The aim of this construction is to express  $f(c)$  as a linear combination of the components of  $G_{g-1}$ . Since  $G_{g-1}$  belongs to  $V = \text{span}(h_0, \dots, h_{g-1})$  we write

$$G_{g-1}(t_i) := h_{g-1}(t_i) - \sum_{j=0}^{g-2} \delta_j h_j(t_i)$$

for some vector  $\underline{\delta} := (\delta_0, \dots, \delta_{g-1})'$ .

We impose the following condition

$$\left\langle \begin{pmatrix} G_{g-1}(t_1) \\ \vdots \\ G_{g-1}(t_n) \end{pmatrix}, \begin{pmatrix} h_j(t_1) \\ \vdots \\ h_j(t_n) \end{pmatrix} \right\rangle = 0, \text{ for all } j = 0, \dots, g-2,$$

where the above symbol  $\langle, \rangle$  is the inner product in  $\mathbb{R}^n$ . The  $\delta_j$ 's in  $\mathbb{R}$  are to be chosen now.

The linear system

$$\sum_{i=1}^n G(t_i) h_j(t_i) = 0, \text{ for } j = 0, \dots, g-2$$

with  $g-1$  equations has  $g-1$  unknown variables  $\delta_j$ .

Once obtained the solution  $\delta_j^*$ ,  $j = 0, \dots, g-2$ , and since

$$h_{g-1}(t) = G_{g-1}(t) + \sum_{j=0}^{g-2} \delta_j h_j(t),$$

we may write  $f(t)$  for any  $t$

$$\begin{aligned} f(t) &= \sum_{j=0}^{g-1} \gamma_j h_j(t) = \gamma_0 h_0(t) + \dots + \gamma_{g-2} h_{g-2}(t) \\ &\quad + \gamma_{g-1} G_{g-1}(t) + \gamma_{g-1} \delta_0 h_0(t) + \dots + \gamma_{g-1} \delta_{g-2} h_{g-2}(t) \\ &= (\gamma_0 + \gamma_{g-1} \delta_0) h_0(t) + \dots + (\gamma_{g-2} + \gamma_{g-1} \delta_{g-2}) h_{g-2}(t) + \gamma_{g-1} G_{g-1}(t) \\ &= \alpha_0 h_0(t) + \dots + \alpha_{g-2} h_{g-2}(t) + \alpha_{g-1} G_{g-1}(t), \end{aligned}$$

where the  $\alpha'_j$ s are defined by

$$\alpha_j := \begin{cases} \gamma_j + \gamma_{g-1}\delta_j & \text{for } j = 0, \dots, g-2 \\ \gamma_{g-1} & \text{for } j = g-1 \end{cases}.$$

The point is that  $\gamma_{g-1}$  appears as the coefficient of  $G_{g-1}$ , namely the last term in the regression of  $f(t)$  on the regressors  $(h_0, \dots, h_{g-2}, G_{g-1})$ . Furthermore  $G_{g-1}$  is orthogonal to the other regressors. The system which describes the data is now written by

$$\underline{Y}(\mathbf{t}) = \tilde{T} \tilde{\theta} + \varepsilon$$

where

$$\tilde{T} := \begin{pmatrix} h_0(t_1) \dots h_{g-2}(t_1) G_{g-1}(t_1) \\ \vdots \\ h_0(t_i) \dots h_{g-2}(t_i) G_{g-1}(t_i) \\ \vdots \\ h_0(t_n) \dots h_{g-2}(t_n) G_{g-1}(t_n) \end{pmatrix}, \tilde{\theta} := \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{g-2} \\ \gamma_{g-1} \end{pmatrix}.$$

The minimum least square estimation of  $\gamma_{g-1}$  is obtained through the normal equations imposing

$$\left( \underline{Y}(\mathbf{t}) - \tilde{T} \hat{\tilde{\theta}} \right) \in V^\perp$$

where  $\hat{\tilde{\theta}}$  hence designates the least square estimator of the vector of coefficients  $\tilde{\theta}$ , and where  $V^\perp$  is the orthogonal linear space of  $V$ .

We have, denoting  $\widehat{\gamma_{g-1}}$  the least square estimator of  $\gamma_{g-1}$ , and noting that  $V = \text{span}\{h_0, \dots, h_{g-2}, G_{g-1}\}$

$$\left\langle \begin{pmatrix} Y(t_1) - \sum_{j=0}^{g-2} \alpha_j h_j(t_1) - \widehat{\gamma_{g-1}} G_{g-1}(t_1) \\ \vdots \\ Y(t_n) - \sum_{j=0}^{g-2} \alpha_j h_j(t_n) - \widehat{\gamma_{g-1}} G_{g-1}(t_n) \end{pmatrix}, \begin{pmatrix} h_j(t_1) \\ \vdots \\ h_j(t_n) \end{pmatrix} \right\rangle = 0,$$

for  $j = 0, \dots, g-2$  and

$$\left\langle \begin{pmatrix} Y(t_1) - \sum_{j=0}^{g-2} \alpha_j h_j(t_1) - \widehat{\gamma_{g-1}} G_{g-1}(t_1) \\ \vdots \\ Y(t_n) - \sum_{j=0}^{g-2} \alpha_j h_j(t_n) - \widehat{\gamma_{g-1}} G_{g-1}(t_n) \end{pmatrix}, \begin{pmatrix} G_{g-1}(t_1) \\ \vdots \\ G_{g-1}(t_n) \end{pmatrix} \right\rangle = 0$$

Hence

$$\sum_{i=1,\dots,n} \left( Y(t_i) - \sum_{j=0}^{g-2} \alpha_j h_j(t_i) - \widehat{\gamma_{g-1}} G_{g-1}(t_i) \right) G_{g-1}(t_i) = 0. \quad (7)$$

Inserting the orthogonality condition

$$\sum_{i=1}^n G(t_i) h_j(t_i) = 0, \text{ for } j = 0, \dots, g-2,$$

in (7) we have

$$\sum_{j=1}^n Y(t_j) G_{g-1}(t_j) - \widehat{\gamma_{g-1}} \sum_{j=1}^n G_{g-1}^2(t_j) = 0,$$

and

$$\widehat{\gamma_{g-1}} = \frac{\sum_{j=1}^n Y(t_j) G_{g-1}(t_j)}{\sum_{j=1}^n G_{g-1}^2(t_j)}.$$

Finally we obtain the explicit form of the estimator of  $f(c)$ . It holds

**Proposition 7** *The least square estimator (Gauss Markov) of the extrapolated value  $f(c)$  is*

$$\widehat{f(c)} = \varphi_{g-1}(c) \widehat{\gamma_{g-1}} = \varphi_{g-1}(c) \frac{\sum_{j=1}^n Y(t_j) G_{g-1}(t_j)}{\sum_{j=1}^n G_{g-1}^2(t_j)}.$$

## 5 The Optimal extrapolation design for the Chebyshev regression

### 5.1 The support of the optimal design

We determine the support of the optimal design for the extrapolation of  $f$  at point  $c$ .

Recall that a design  $\xi^*$  is optimal if and only if it produces a Gauss Markov estimator of  $f(c)$  with minimal variance among all such estimators built upon other designs.

We note that the variance of  $\widehat{f(c)}$  depends on the  $x_j$ 's since

$$\begin{aligned} \text{var} \left( \widehat{f(c)} \right) &= (\varphi_{g-1}(c))^2 \frac{\sum_{j=1}^n \text{var}(Y(t_j)) G_{g-1}^2(t_j)}{\left( \sum_{j=1}^n G_{g-1}^2(t_j) \right)^2} \\ &= \frac{(\sigma \varphi_{g-1}(c))^2}{\sum_{j=1}^n G_{g-1}^2(t_j)}. \end{aligned}$$

The design is defined through a discrete probability measure  $\xi \in \mathcal{M}_d$  ( $[-1, 1]$ ) with support  $(x_0, \dots, x_{g-1})$  with  $\xi(x_j) := n_j/n$  and  $n_j$  equals the number of the  $t'_i$ s which equal  $x_j$ , for  $0 \leq j \leq g-1$ .

We now determine the support of the optimal design denoted  $\xi^*$ .

$$\begin{aligned} \xi^* &:= \arg \min_{\xi \in \mathcal{M}_d([-1, 1])} \frac{1}{\sum_{j=0}^{g-1} n_j G_{g-1}^2(x_j)} = \arg \max_{\xi \in \mathcal{M}_d^*} \sum_{i=0}^{g-1} n_j G_{g-1}^2(x_j) \\ &= \arg \max_{\xi \in \mathcal{M}_d([-1, 1])} \sum_{i=0}^{g-1} n_j \left( h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j h_j(x_i) \right)^2. \end{aligned}$$

The solution can be obtained in a simple way through some analysis of the objective function. By convenience in order to use simple geometric arguments and to simplify the resulting expressions it is more convenient to write the derivation of the optimal design in terms of the  $t'_i$ s.

The function

$$\sum_{i=1}^n \left( h_{g-1}(t_i) - \sum_{j=0}^{g-2} \delta_j h_j(t_i) \right)^2 = \left\| \begin{pmatrix} h_{g-1}(t_1) - \sum_{j=0}^{g-2} \delta_j h_j(t_1) \\ \vdots \\ h_{g-1}(t_n) - \sum_{j=0}^{g-2} \delta_j h_j(t_n) \end{pmatrix} \right\|^2$$

is the distance from the orthogonal projection of the vector

$$\mathbf{h} := (h_{g-1}(t_1) \dots h_{g-1}(t_n))'$$

on the linear space  $V$  generated by the family  $\{h_0, \dots, h_{g-2}, G_{g-1}\}$ .

Therefore by the minimal projection property

$$\sum_{i=1}^n \left( h_{g-1}(t_i) - \sum_{j=0}^{g-2} \delta_j h_j(t_i) \right)^2 = \min_{\psi \in V} \text{dist}(\mathbf{h}, \psi).$$

Let  $\underline{\delta} := (\delta_0 \dots \delta_{g-2})'$ .



The optimal design is obtained through a two steps procedure. Fix the frequencies  $n_0, \dots, n_{g-1}$  with sum  $n$  and determine the discrete measure  $\xi$  on  $[-1, 1]$  which minimizes  $\text{var}_\xi \left( \widehat{f(c)} \right)$  among all  $\xi$ 's with support  $\mathbf{x} := (x_0, \dots, x_{g-1})$  and masses  $\xi(x_j) = n_j/n$ ,  $0 \leq j \leq g-1$ . The optimization is performed upon the  $x_j$ 's.

The optimal design solves therefore the problem

$$\begin{aligned} \xi^* &= \arg \max_{\xi \in \mathcal{M}_d([-1,1])} \min_{\psi \in V} \text{dist}(\mathbf{h}, \psi) \\ &= \arg \max_{\mathbf{x} \in [-1,1]^g} \min_{\underline{\delta} \in \mathbb{R}^{g-1}} \sum_{i=0}^{g-1} n_i \left( h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j h_j(x_i) \right)^2 \\ &= \arg \max_{\xi \in \mathcal{M}_d([-1,1])} \min_{\underline{\delta} \in \mathbb{R}^{g-1}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \xi(dx). \end{aligned}$$

The integrand  $\left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2$  is always non negative. Henceforth it is enough to minimize its square root w.r.t.  $\mathbf{x}$ . This optimization turns therefore to be independent of the  $n_j$ 's.

Denote  $\delta_j^*$ ,  $j = 0, \dots, g-2$ , the values which minimize  $\text{dist}(\mathbf{h}, \psi)$  w.r.t.  $\delta_j$ . The optimality condition writes

$$\begin{aligned} \max_{x \in [-1,1]} \left| h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j^* h_j(x) \right| &= \min_{\underline{\delta} \in \mathbb{R}^{g-1}} \max_{x \in [-1,1]} \left| h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right| \\ &= \min_{p \in W} \max_{x \in [-1,1]} |h_{g-1}(x) - p(x)| \end{aligned} \tag{8}$$

where

$$W := \text{span} \{h_0, \dots, h_{g-2}\}. \tag{9}$$

If we prove that  $\{h_0, \dots, h_{g-2}\}$  is a Chebyshev system on  $[-1, 1]$ , then clearly the support of the optimal measure  $\xi^*$  consists in the points of maximal value in  $[-1, 1]$  for the function

$$|h_{g-1}(x) - p^*(x)|$$

where  $p^*$  is the best uniform approximating polynomial of  $h_{g-1}$  in  $W$ . Indeed the support of  $\xi^*$  consists in the set of points where

$$\left| h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j^* h_j(x) \right|$$

in (8) attains its maximal value for  $p = p^*$  the best uniform approximation of  $h_{g-1}$  in  $W$ .

This is the major argument of the present derivation, which justifies all of the uniform approximation theory in this context.

**Definition 8** *The vector  $\underline{\delta}^*$  in  $\mathbb{R}^{g-1}$  is a Chebyshev vector iff it designates the vector of the coefficients of  $p^*$ , where  $p^*$  is the best uniform approximating polynomial of  $h_{g-1}$  in  $W$  defined in (9). It is defined through (8).*

Now writing

$$\underline{\delta}^* := (\delta_0^*, \dots, \delta_{g-2}^*)'$$

we define the set of all points  $\tilde{x}$  in  $[-1, 1]$  where the distance between  $h_{g-1}$  and its best approximation on the  $h_k$ ,  $0 \leq k \leq g-2$  is minimal. These points are precisely the support of the optimal design  $\xi^*$ . Formally we define

$$E := \min_{\underline{\delta} \in \mathbb{R}^{g-1}} \max_{x \in [-1, 1]} \left| h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right| \quad (10)$$

and

$$B(\underline{\delta}^*) := \left\{ \tilde{x} \in [-1, 1] : \left| h_{g-1}(\tilde{x}) - \sum_{j=0}^{g-2} \delta_j^* h_j(\tilde{x}) \right| = E \right\}. \quad (11)$$

It holds (see Proposition 10 below)

$$\xi^*(B(\underline{\delta}^*)) = 1.$$

We prove that  $\{h_0, \dots, h_{g-2}\}$  is a Chebyshev system on  $[-1, 1]$ .

**Proposition 9** (Hoel) *The functions  $h_0, \dots, h_{g-2}$  are a Chebyshev system on  $[-1, 1]$ .*

**Proof.** For any choice of  $\{x_0, \dots, x_g\}$  in  $[-1, 1]$ , with  $x_0 < \dots < x_{g-1}$ , since the family  $\{\varphi_0, \dots, \varphi_{g-1}\}$  is a Chebyshev system on  $[-1, 1]$ , we have, by Proposition 2, assuming a positive sign of the determinant, without loss of generality

$$0 < \det \begin{pmatrix} \varphi_0(x_0) & \varphi_0(x_1) & \cdot & \varphi_0(x_{g-2}) & \varphi_0(c) \\ \varphi_1(x_0) & \varphi_1(x_1) & \cdot & \varphi_1(x_{g-2}) & \varphi_1(c) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \varphi_{g-2}(x_0) & \varphi_{g-2}(x_1) & \cdot & \varphi_{g-2}(x_{g-2}) & \varphi_{g-2}(c) \\ \varphi_{g-1}(x_0) & \varphi_{g-1}(x_1) & \cdot & \varphi_{g-1}(x_{g-2}) & \varphi_{g-1}(c) \end{pmatrix}.$$

For  $j = 0, \dots, g-1$ , the operations

$$\varphi_j(x_i) \mapsto \varphi_j(x_i) - \varphi_j(c) \frac{\varphi_g(x_i)}{\varphi_g(c)}$$

do not change the value of the determinant.

Hence,

$$\begin{aligned} 0 &< \det \begin{pmatrix} \varphi_0(x_0) & \varphi_0(x_1) & \cdot & \varphi_0(x_{g-2}) & \varphi_0(c) \\ \varphi_1(x_0) & \varphi_1(x_1) & \cdot & \varphi_1(x_{g-2}) & \varphi_1(c) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \varphi_{g-2}(x_0) & \varphi_{g-2}(x_1) & \cdot & \varphi_{g-2}(x_{g-2}) & \varphi_{g-2}(c) \\ \varphi_{g-1}(x_0) & \varphi_{g-1}(x_1) & \cdot & \varphi_{g-1}(x_{g-2}) & \varphi_{g-1}(c) \end{pmatrix} \\ &= \det \begin{pmatrix} h_0(x_0) & h_0(x_1) & \cdot & h_0(x_{g-2}) & \varphi_0(c) \\ h_1(x_0) & h_1(x_1) & \cdot & h_1(x_{g-2}) & \varphi_1(c) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ h_{g-2}(x_0) & h_{g-2}(x_1) & \cdot & h_{g-2}(x_{g-2}) & \varphi_{g-2}(c) \\ 0 & 0 & \cdot & 0 & \varphi_{g-1}(c) \end{pmatrix}. \end{aligned}$$

By the Laplace Theorem pertaining to determinants, we get

$$\begin{aligned} 0 &< \det \begin{pmatrix} h_0(x_0) & h_0(x_1) & \cdot & h_0(x_{g-2}) & \varphi_0(c) \\ h_1(x_0) & h_1(x_1) & \cdot & h_1(x_{g-2}) & \varphi_1(c) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ h_{g-2}(x_0) & h_{g-2}(x_1) & \cdot & h_{g-2}(x_{g-2}) & \varphi_{g-2}(c) \\ 0 & 0 & \cdot & 0 & \varphi_{g-1}(c) \end{pmatrix} \\ &= \varphi_{g-1}(c) \det \begin{pmatrix} h_0(x_0) & h_0(x_1) & \cdot & h_0(x_{g-2}) \\ h_1(x_0) & h_1(x_1) & \cdot & h_1(x_{g-2}) \\ \cdot & \cdot & \cdot & \cdot \\ h_{g-2}(x_0) & h_{g-2}(x_1) & \cdot & h_{g-2}(x_{g-2}) \end{pmatrix} := \varphi_{g-1}(c) \times \Delta. \end{aligned}$$

Therefore the two real numbers  $\varphi_{g-1}(c)$  and  $\Delta$  have same sign.

Since  $\varphi_{g-1}(c) \neq 0$  we deduce that

$$\det \begin{pmatrix} h_0(x_0) & h_0(x_1) & \cdot & h_0(x_{g-2}) \\ h_1(x_0) & h_1(x_1) & \cdot & h_1(x_{g-2}) \\ \cdot & \cdot & \cdot & \cdot \\ h_{g-2}(x_0) & h_{g-2}(x_1) & \cdot & h_{g-2}(x_{g-2}) \end{pmatrix} \neq 0.$$

Hence the family  $\{\varphi_0, \dots, \varphi_{g-1}\}$  is a Chebyshev system in  $\mathcal{C}([-1, 1])$ .  
In the same way we can prove that it is a Chebyshev system in  $[-1, c]$ .

■

## 5.2 The frequencies of the optimal design

Once characterized the points  $x$  in  $\text{supp } \xi^*$ , we characterize the values of the  $\xi^*(x)$ 's. The following Proposition produces a sufficient condition in order that the measure  $\xi^*$  be optimal, which can be phrased as

$$\begin{aligned} & \min_{\delta \in \mathbb{R}^{g-2}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \xi^*(dx) \\ & \leq \min_{\delta \in \mathbb{R}^{g-2}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \xi(dx) \end{aligned}$$

for any  $\xi$  in  $\mathcal{M}_d([-1, 1])$ . Uniqueness might not hold.

**Proposition 10** (Kiefer-Wolfowitz) *Let  $B(\underline{\delta}^*)$  be defined as in (11). If  $\underline{\delta}^*$  is Chebyshev vector and  $\xi(B(\underline{\delta}^*)) = 1$  and if*

$$\int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j^* h_j(x) \right) h_i(x) \xi(dx) = 0, \quad \text{for } i = 0, \dots, g-2,$$

then  $\xi$  is optimal.

**Proof.** Let  $\xi \in \mathcal{M}_d([-1, 1])$  with  $\xi(B(\underline{\delta}^*)) = 1$ . The hypothesis

$$\int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j^* h_j(x) \right) h_i(x) \xi(dx) = 0,$$

for  $i = 0, \dots, g-2$ , indicates that

$$h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_i)$$

is orthogonal to the linear space  $W$  generated by  $\{h_0, \dots, h_{g-2}\}$ . Thus  $\sum_{j=0}^{g-2} \delta_j^* h_j$  is the orthogonal projection of  $h_{g-1}$  on  $W$ . The inner product is

$$\langle \mathbf{v}, \mathbf{w} \rangle := \int_{[-1,1]} \mathbf{v}(x) \mathbf{w}(x) \xi(dx).$$

By the minimal projection property

$$\begin{aligned}
A(\xi) &:= \min_{\underline{\delta} \in \mathbb{R}^{g-2}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \xi(dx) \\
&= \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j^* h_j(x) \right)^2 \xi(dx) \\
&= \sum_{\tilde{x} \in \text{supp} \xi} \left( h_{g-1}(\tilde{x}) - \sum_{j=0}^{g-2} \delta_j^* h_j(\tilde{x}) \right)^2 \xi(\tilde{x}) \\
&\geq E^2 \sum_{\tilde{x} \in \text{supp} \xi} \xi(\tilde{x}) \\
&= E^2 \geq \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j^* h_j(x) \right)^2 \nu(dx) \\
&\geq \min_{\underline{\delta} \in \mathbb{R}^{g-2}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \nu(dx), \\
&\geq \max_{v \in \mathcal{M}_d([-1,1])} \min_{\underline{\delta} \in \mathbb{R}^{g-2}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \nu(dx) \\
&= \min_{\underline{\delta} \in \mathbb{R}^{g-2}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \xi^*(dx) =: A(\xi^*)
\end{aligned}$$

The measure  $\nu$  which appears in lines 5 and followings in the above displays are arbitrary measures in  $\mathcal{M}_d([-1,1])$ .

Since by definition

$$\xi^* := \arg \max_{v \in \mathcal{M}_d([-1,1])} \min_{\underline{\delta} \in \mathbb{R}^{g-2}} \int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j h_j(x) \right)^2 \nu(dx) \tag{12}$$

i.e.

$$A(\xi^*) \geq A(\xi).$$

Hence  $A(\xi^*) = A(\xi)$ .

■

## 6 Identification of the optimal design

In this Section we provide an explicit solution for the optimal design and prove its uniqueness.

By the Borel-Chebyshev Theorem 4 there exist at least  $g$  points

$$x_0 < \dots < x_{g-1}$$

in  $[-1, 1]$  on which the best uniform approximation of  $h_{g-1}$ , namely  $\sum_{j=0}^{g-2} \delta_j^* h_j$ , satisfies the following conditions

$$h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_i) = (-1)^i E.$$

We now see that there are exactly  $g$  points on which the function  $\left| h_{g-1} - \sum_{j=0}^{g-2} \delta_j^* h_j \right|$  equals  $E$ .

Since  $\{h_0, \dots, h_{g-1}\}$  is a Chebyshev system the linear combination

$$\sum_{i=0}^{g-1} a_i h_i$$

cannot have more than  $g$  roots in  $[-1, 1]$ . Hence the function

$$\left| h_{g-1} - \sum_{j=0}^{g-2} \delta_j^* h_j \right|$$

which is the absolute value of a linear combination of the Chebyshev system  $\{h_0, \dots, h_{g-1}\}$  cannot have more than  $g$  roots. Therefore  $\left| h_{g-1} - \sum_{j=0}^{g-2} \delta_j^* h_j \right|$  cannot have more than  $g + 1$  maximal values.

As seen previously the support of the optimal measure  $\xi^*$  consists in the points of maximal value in  $[-1, 1]$  for the function

$$\left| h_{g-1} - \sum_{j=0}^{g-2} \delta_j^* h_j \right|.$$

Applying the Borel-Chebyshev Theorem we now determine the support of  $\xi^*$ .

Since  $E$  is known the support is the vector  $(x_1^*, \dots, x_g^*)$  which solves the linear system

$$h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_i) = (-1)^i E, \quad i = 1, \dots, g.$$

We apply the sufficient condition provided by Kiefer and Wolfowitz above, Proposition 10.

This condition states that the values  $\xi^*(x_i)$ ,  $i = 0, \dots, g - 1$ , satisfy the system

$$\begin{cases} \sum_{i=0}^{g-1} \left( h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_i) \right) h_r(x_i) \xi^*(x_i) = 0 \\ r = 0, \dots, g-2 \end{cases}.$$

In the  $x_i$ 's it holds

$$E = \left| h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_i) \right|$$

and

$$\int_{[-1,1]} \left( h_{g-1}(x) - \sum_{j=0}^{g-2} \delta_j^* h_j(x) \right) h_i(x) \xi^*(dx) = 0, \quad \text{for } i = 0, \dots, g-2.$$

Therefore

$$\begin{aligned} 0 &= \sum_{i=0}^{g-1} \left( h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_i) \right) h_r(x_i) \xi^*(x_i) \\ &= E \sum_{i=0}^{g-1} (-1)^i h_r(x_i) \xi^*(x_i), \text{ for } r = 0, \dots, g-2. \end{aligned}$$

The optimal extrapolation design  $\{(x_i, \xi^*(x_i)) : i = 0, \dots, g-1\}$  thus solves

$$\left\{ \begin{array}{l} h_{g-1}(x_0) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_0) = +E \\ \dots\dots\dots \\ h_{g-1}(x_i) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_i) = (-1)^i E \\ \dots\dots\dots \\ h_{g-1}(x_{g-1}) - \sum_{j=0}^{g-2} \delta_j^* h_j(x_{g-1}) = (-1)^{g-1} E \\ \sum_{i=0}^{g-1} (-1)^i h_0(x_i) \xi^*(x_i) = 0 \\ \dots\dots\dots \\ \sum_{i=0}^{g-1} (-1)^i h_r(x_i) \xi^*(x_i) = 0 \\ \dots\dots\dots \\ \sum_{i=0}^{g-1} (-1)^i h_{g-2}(x_i) \xi^*(x_i) = 0 \end{array} \right.$$

In practice we first evaluate  $\delta_j^*$  for  $0 \leq j \leq g-2$  through (8). Note that  $E$  is known by (10). The above system consists in  $2g-1$  equations in the  $2g$  unknown quantities  $\{(x_i, \xi^*(x_i)) : i = 0, \dots, g-1\}$ . Add the constraint

$$\xi^*(x_0) + \dots + \xi^*(x_{q-1}) = 1$$

to obtain a linear system with a unique solution.

The first  $g$  equations determine the nodes, by Borel Chebyshev Theorem. The last  $g - 1$  ones determine the values of the  $n'_j$ s by the Proposition of Kiefer and Wolfowitz 10. Hence there is a unique optimal design solving the minimal variance problem for the extrapolation.

## References

- [Achieser, 1992] N. I. Achieser. *Theory of approximation*. Dover Publications, Inc., New York, 1992. Translated from the Russian and with a preface by Charles J. Hyman, Reprint of the 1956 English translation.
- [Broniatowski and Celant, 2014] M. Broniatowski and G. Celant. Some overview on unbiased interpolation and extrapolation designs. *arXiv:1403.5113*, 2014.
- [Dzyadyk and Shevchuk, 2008] Vladislav K. Dzyadyk and Igor 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.
- [Hoel, 1966] Paul G. Hoel. A simple solution for optimal Chebyshev regression extrapolation. *Ann. Math. Statist.*, 37:720–725, 1966.
- [Karlin and Studden, 1966] Samuel Karlin and William 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, 1965] J. Kiefer and J. Wolfowitz. On a theorem of Hoel and Levine on extrapolation designs. *Ann. Math. Statist.*, 36:1627–1655, 1965.
- [Pukelsheim, 1993] Friedrich Pukelsheim. *Optimal design of experiments*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1993. A Wiley-Interscience Publication.