



**HAL**  
open science

## Some overview on unbiased interpolation and extrapolation designs

Michel Broniatowski, Giorgio Celant

► **To cite this version:**

Michel Broniatowski, Giorgio Celant. Some overview on unbiased interpolation and extrapolation designs. *Annales de l'ISUP*, 2014, 58 (3), pp. 27-60. hal-00961611

**HAL Id: hal-00961611**

**<https://hal.science/hal-00961611v1>**

Submitted on 20 Mar 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.

# Some overview on unbiased interpolation and extrapolation designs

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

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

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

March 20, 2014

## Abstract

This paper considers the construction of optimal designs due to Hoel and Levine and Guest. It focuses on the relation between the theory of the uniform approximation of functions and the optimality of the designs. Some application to accelerated tests is also presented. The multivariate case is also handled in some special situations.

AMS Classification: 62K05, 62N04

Key words: Optimal design, Extrapolation, Interpolation, Lagrange polynomials, Legendre polynomials, Chebyshev points

## 1 Introduction

### 1.1 Definition of the model and of the estimators

We assume that we are given an interval where the explanatory variable  $x$  takes its value; for simplicity assume that  $x$  belongs to  $[-1, 1]$ . On this interval the response  $Y$  can be observed. An additive noise causes  $Y$  to be only partly related to the input  $x$ . This noise is assumed to be independent upon the value of  $x$ , which is commonly referred to as a homoscedastic hypothesis on the model. For a given input  $x$  the measure  $Y(x)$  can be written as

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

where  $f$  is some unknown function and the generic real valued random variable  $\varepsilon$  has some unknown distribution; however it is assumed that the two first moment of  $\varepsilon$  are finite. The function  $f$  might be defined on a larger interval than  $[-1, 1]$ . All possible measurements of  $f$  can only be achieved on  $[-1, 1]$ . It may occur that we are interested in some estimation of  $f(x)$  for some  $x$  where  $f$  is not measured; when  $x$  belongs to  $[-1, 1]$  this is an interpolation problem. At times we may be interested in some approximation of  $f(x)$  for  $x$  outside  $[-1, 1]$ ; this is an extrapolation problem.

We will discuss optimal designs in those contexts. Defining a design results in a two fold description. Firstly it is based on a set of measurements points, say  $x_0, \dots, x_{g-1}$  in  $[-1, 1]$ . Those are the nodes of the design. Next, for any node  $x_j$ , we define an integer  $n_j$  which is the number of replicates of the measurement performed under the condition  $x_j$ . We thus inherit of the  $n_j$  measurements  $Y_1(x_j), \dots, Y_{n_j}(x_j)$ . Those measurements are supposed to be independent. Note that we do not assume any probabilistic structure on the  $x_j$ 's which therefore will not be considered as sampled under any device. The  $x_j$ 's are determined by the experimenter and their choice will follow from a strictly deterministic procedure.

Obviously this simple model will allow for a simple estimate of  $f(x_j)$  for all  $j$ , assuming without loss of generality that the error  $\varepsilon$  has expectation 0.

The design is therefore defined by the family of the nodes (their number  $g$  is fixed by the experimenter), and the so called frequencies  $n_j$ 's,  $0 \leq j \leq g - 1$ .

Obviously the total number of experiments is limited, for reasons which have to do with the context of the study. Those reasons might be related to the cost of each individual experiment, or by other considerations. For example in phase 1 clinical trials it is usually assumed that only very few patients can be eligible for the trial. Call  $n$  this number of trials to be performed. The resulting constraint on the  $n_j$ 's is therefore

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

Let us now define formally this model.

For any  $i = 0, \dots, g - 1$ ,  $x_i$  is a node and  $y_k(x_i)$  is the  $k$ -th measurement of  $Y(x_i)$  when  $k$  runs in  $1, \dots, n_i$ . Therefore

$$\begin{cases} y_1(x_i) = f(x_i) + \varepsilon_{1,i} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_{n_i}(x_i) = f(x_i) + \varepsilon_{n_i,i} \end{cases}$$

where  $n_i > 1$  and  $n_i \in \mathbb{N}$ , together with  $n := \sum_{i=0}^{g-1} n_i$  where  $n$  is fixed. Obviously the r.v.'s  $\varepsilon_{j,i}$ ,  $1 \leq j \leq n_i$ ,  $i = 0, \dots, g-1$ , are not observed. They are i.i.d. copies of a generic r.v.  $\varepsilon$ . Furthermore  $E(\varepsilon) = 0$ ,  $var(\varepsilon) = \sigma^2$ .

We assume that  $f$  is a polynomial with known degree  $g-1$ . Therefore it is completely determined if known the values of  $f$  in  $g$  distinct points. Note that the knowledge of  $g$  is an important and strong assumption. Denote further

$$I := \{x_0 < \dots < x_{g-1}\} \subset [-1, 1]$$

the family of nodes.

The aim of this chapter is to discuss the operational choice of the design; we will thus propose some choices for the nodes and the so-called frequencies  $n_j/n$  which, all together, define the design. This will be achieved discussing some notion of optimality.

When no random effect is present, existence and uniqueness of the solution of the linear system with  $g$  equations and  $g$  variables  $\theta := (\theta_0, \dots, \theta_{g-1})$ ,

$$y(x_i) = \sum_{j=0}^{g-1} \theta_j x_i^j, (x_0, \dots, x_{g-1}) \in [-1, 1]^g, 0 \leq i \leq g-1$$

allow to identify the function  $f$  at any  $x$  in  $\mathbb{R}$ . Changing the canonical basis in  $P_{g-1}(X)$  into the family of the elementary Lagrange polynomials

$$l_i(x) := \prod_{j=0, j \neq i}^{g-1} \frac{x - x_j}{x_i - x_j} \quad (1)$$

yields

$$f(x) = \sum_{i=0}^{g-1} f(x_i) l_i(x).$$

In the present random setting,  $f(x_i)$  is unknown. This suggests to consider the estimator  $\widehat{f}$  of  $f$  defined by

$$\mathcal{L}_n(\widehat{f})(x) := \sum_{i=0}^{g-1} \widehat{f(x_i)} l_i(x),$$

where  $\widehat{f(x_i)}$  denotes some estimate of  $f$  on a generic node  $x_i$ .

Turn to the estimate of  $f(x_i)$  namely the simplest one, defined by

$$\widehat{f}(x_i) := \bar{Y}(x_i) := \frac{1}{n_i} \sum_{j=1}^{n_i} Y_j(x_i),$$

which solves

$$\widehat{f}(x_i) = \arg \min_{\mu \in \mathbb{R}} \sum_{j=1}^{n_i} (Y_j(x_i) - \mu)^2,$$

provides the optimal linear unbiased estimator of  $f(x_i)$ . It follows that  $\widehat{f}(x) := \mathcal{L}_n(\widehat{f})(x)$  is unbiased since for all  $x \in \mathbb{R}$ ,

$$E\left(\mathcal{L}_n(\widehat{f})(x)\right) := \sum_{i=0}^{g-1} E\left(\widehat{f}(x_i)\right) l_i(x) = \mathcal{L}_n(f)(x) = f(x).$$

Since  $\mathcal{L}_n(\widehat{f})(x)$  is linear with respect to the parameters  $f(x_i)$ ,  $i = 0, \dots, g-1$ , using Gauss Markov Theorem,  $\mathcal{L}_n(\widehat{f})(x)$  is optimal, i.e. has minimal variance. The variance of the estimator  $\mathcal{L}_n(\widehat{f})(x)$  is

$$\text{var}\left(\mathcal{L}_n(\widehat{f})(x)\right) = \text{var}\left(\sum_{i=0}^{g-1} \text{var}\left(\widehat{f}(x_i)\right) (l_i(x))^2\right) = \sigma^2 \sum_{i=0}^{g-1} \frac{(l_i(x))^2}{n_i}, \quad (2)$$

which depends explicitly on the frequency  $n_i$  of the observations of  $f$  on the nodes  $x_i$ 's.

We now proceed to the formal definition of a design. The set

$$\left\{ ((n_0, \dots, n_{g-1}), (x_0, \dots, x_{g-1})) \in \mathbb{N}^g \times [-1, 1]^{g-1} : n := \sum_{i=0}^{g-1} n_i, n \text{ fixed} \right\}$$

determines a discrete probability measure  $\xi$  with support  $I$ , a finite subset in  $[-1, 1]$ , by

$$\xi(x_i) := \frac{n_i}{n}, \quad i = 0, \dots, g-1.$$

Turning to (2) we observe that the accuracy of the design depends on the point  $x$  where the variance of  $\mathcal{L}_n(\widehat{f})(x)$  is calculated.

Since all estimators of the form  $\mathcal{L}_n(\widehat{f})(x)$  are unbiased, their accuracy depend only on their variance, which in turn depends both on  $x$  and on the measure  $\xi$ . Optimizing on  $\xi$  for a given  $x$  turns to an optimal choice for  $I$  and for the family of the  $n_i$ 's, under the constraint

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

Such designs  $\xi_x$  are called Hoel-Levine extrapolation designs when  $x$  lies outside  $[-1, 1]$ . When  $x$  belongs to  $[-1, 1]$  then clearly the optimal design for the criterion of the variance of the estimator of  $f(x)$  results in performing all the  $n$  measurements at point  $x$ . The non trivial case is when the optimality is defined through a control of the uniform variance of the estimator of  $f(x)$ , namely when  $\xi$  should minimize

$$\sup_{x \in [-1, 1]} \text{var}_\xi \mathcal{L}_n(\widehat{f})(x).$$

Those designs  $\xi$  are called interpolation designs, or Guest or Legendre designs.

The notation to be kept is as follows. The set of all probability measures on the interval  $[-1, 1]$  supported by  $g$  distinct points in the interval  $[-1, 1]$  is denoted  $\mathcal{M}_{[-1, 1]}^*$ , which therefore is the class of all designs.

The purpose of this paper is to present a unified view on this classical field which links the theory of the uniform approximation of functions and the statistical theory of experimental designs, unifying notation and concepts. Its content comes mostly from the seminal works by Hoel and Levine [8], Guest [6]. Some other important references in the context is Kiefer and Wolfowicz [11] and Studden [16].

## 1.2 Some facts from the theory of the approximation of functions

We briefly quote some basic facts from classical analysis, to be used extensively. For fixed  $n$  let  $\mathcal{P}_n$  denote the class of all polynomials with degree less or equal  $n$  defined on  $[-1, 1]$  and let  $P_n(x) := a_0 + \dots + a_n x^n$ . Some coordinates of the vector of coefficients  $(a_0, \dots, a_n)$  can take value 0.

For a continuous function  $f$  defined on  $[-1, 1]$  denote

$$e := f - P_n$$

which is a continuous function on  $[-1, 1]$ , as is  $|e|$ .

Applying Weierstrass Theorem it follows that  $|e|$  attains its maximal value in  $[-1, 1]$ , for at least one point  $x$ . We denote

$$\mu = \mu(a_0, a_1, \dots, a_n) := \max_{x \in [-1, 1]} |e(x)| \geq 0$$

which yields to define

$$\alpha := \inf_{(a_0, a_1, \dots, a_n)} \mu(a_0, a_1, \dots, a_n), \quad (3)$$

the minimum of the uniform error committed substituting  $f$  by a polynomial in  $\mathcal{P}_n$ . This lower bound exists.

**Definition 1** A polynomial  $P^*$  in  $\mathcal{P}_n$  with coefficients  $(a_0^*, \dots, a_n^*)$  such that

$$\sup_{x \in [-1, 1]} |f(x) - P^*(x)| = \alpha,$$

where  $\alpha$  is defined in (3) is called a best approximating polynomial of  $f$  with degree  $n$  in the uniform sense.

This best approximation thus satisfies

$$P^*(x) := \arg \inf_{P_n \in \mathcal{P}_n} \sup_{x \in [-1, 1]} |e(x)|.$$

The polynomial  $P^*$  may be of degree less than  $n$ . Define

$$e^*(x) := f(x) - P^*(x) \quad (4)$$

and

$$E(f) := \max_{x \in [-1, 1]} |e^*(x)|. \quad (5)$$

### 1.2.1 Existence of the best approximation

The following result answers the question of attainment for the least uniform error when approximating  $f$  by a polynomial in  $\mathcal{P}_n$ .

**Theorem 2** Let  $f$  be some continuous function defined on  $[-1, 1]$ . For any integer  $n$  there exists a unique  $P^*$  in  $\mathcal{P}_n$  such that  $\|f - P^*\|_\infty = \inf_{P_n \in \mathcal{P}_n} \|f - P_n\|_\infty$ .

### 1.2.2 Uniqueness of the best approximation

The proof of the uniqueness Theorem strongly relies on a Theorem by Chebyshev which explores the number of changes of the sign of the error, which we state now. Consider  $P^*(x)$  and  $E(f)$  as defined in (4) and (5).

**Theorem 3** (*Borel-Chebyshev*) *A polynomial  $P^*$  in  $\mathcal{P}_n$  is a best uniform approximation of a function  $f$  in  $\mathcal{C}^{(0)}[-1, 1]$  if and only if the function  $x \rightarrow e^*(x)$  equals  $E(f)$  with alternating signs for at least  $n + 2$  values of  $x$  in  $[-1, 1]$ .*

**Theorem 4** *Let  $f$  be a continuous function defined on  $[-1, 1]$ . The best uniform approximating polynomial is unique.*

For a complete treatment of the above arguments, see e.g. [5], [15] and [12].

## 2 Optimal extrapolation designs; Hoel Levine or Chebyshev designs

We consider the problem of approximating  $f(x)$  for any fixed  $x$  in the interval  $[c, -1]$  for some  $c < -1$ . More generally the optimal design which is obtained in this section is valid for any  $c$  such that  $|c| > 1$ .

As seen previously, since  $\mathcal{L}_n(\hat{f})(x)$  is an unbiased estimate of  $f(x)$  a natural criterion for optimality in the class of all unbiased linear estimates is the variance.

We therefore consider the problem

$$\xi_x^* := \arg \min_{\xi \in \mathcal{M}_{[-1,1]}^*} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right)$$

to which we will propose a suboptimal solution.

Denoting generically  $\mathbf{n}^* := (n_0^*, \dots, n_{g-1}^*)$  and  $(x^*) := (x_0^*, \dots, x_{g-1}^*)$ ,  $\mathbf{n} := (n_0, \dots, n_{g-1}) \in \mathbb{N}^g$  and  $(x) := (x_0, \dots, x_{g-1})$  this problem can thus be written as

$$\left\{ \begin{array}{l} (\mathbf{n}^*, (x^*)) = \arg \min_{(\mathbf{n}, (x))} \sum_{i=0}^{g-1} \frac{(l_i(x))^2}{n_i} \\ \mathbf{n} : \sum_{i=0}^{g-1} n_i = n, n \text{ fixed} \\ -1 \leq x_0 < \dots < x_{g-1} \leq 1, (x) \in [-1, 1]^g \end{array} \right. .$$



This is an integer programming problem (w.r.t.  $\mathbf{n}$ ) with an inequality constraint in  $\mathbb{R}$ . As commonly done we find a proxy to the solution, considering the  $n_i$ 's as real numbers. The resulting solution  $(n_0^*, \dots, n_{g-1}^*)$  will be substituted by the integer part of each of the  $n_i^*$ 's.

We therefore get to the following optimization problem in the  $2g$  real variables

$$\left\{ \begin{array}{l} ((w_0^*, \dots, w_{g-1}^*), (x_0^*, \dots, x_{g-1}^*)) = \arg \min_{((w_0, \dots, w_{g-1}), (x_0, \dots, x_{g-1}))} \sum_{i=0}^{g-1} \frac{(l_i(x))^2}{w_i} \\ (w_0, \dots, w_{g-1}) \in \mathbb{R}^g, w > 0, \sum_{i=0}^{g-1} w_i = n, n \text{ fixed} \\ -1 \leq x_0 < \dots < x_{g-1} \leq 1, (x_0, \dots, x_{g-1}) \in [-1, 1]^g. \end{array} \right. \quad (6)$$

Since the mapping

$$((w_0, \dots, w_{g-1}), (x_0, \dots, x_{g-1})) \rightarrow \sum_{i=0}^{g-1} \frac{(l_i(x))^2}{w_i}$$

is continuous, the optimization problem (6) can be solved in a two steps procedure.

The principle is as follows: fix the vector  $\underline{\mathbf{x}} := (x_0, \dots, x_{g-1})'$  and look for the minimum with respect to the vector  $\underline{\mathbf{w}} := (w_0, \dots, w_{g-1})'$ . Once obtained the optimizer  $\underline{\mathbf{w}}^*$  vary  $\underline{\mathbf{x}}$  and determine the resulting minimum value of the variance for fixed  $\underline{\mathbf{w}} = \underline{\mathbf{w}}^*$ .

Define therefore , with fixed  $\underline{\mathbf{x}}$

$$(2_a) \left\{ \begin{array}{l} \min_{(w_0, \dots, w_{g-1}) \in \mathbb{R}^g} \sum_{i=0}^{g-1} \frac{(l_j(x))^2}{w_j} \\ \sum_{i=0}^{g-1} w_i = n, n \text{ fixed} \end{array} \right. \quad (7)$$

Denote  $(w_0^*, \dots, w_{g-1}^*)$  the solution of this problem.

The optimal design  $\xi^*$  results then as the solution of the following problem, assuming in the present case that  $x < x_0$

$$(2_b) \left\{ \begin{array}{l} \min_{(x_0, \dots, x_{g-1}) \in \mathbb{R}^g} \sum_{i=0}^h \frac{(l_j(x))^2}{w_j^*} \\ x < -1 < x_0 < \dots < x_{g-1}, \{x_0, \dots, x_{g-1}\} \subset [-1, 1] \end{array} \right. \quad (8)$$

**Step 1.** We handle Problem (7).

**Proposition 5** *The solution of problem (7) exists and is unique. It is given by*

$$w_j^* := n \frac{|l_j(x)|}{\sum_{j=0}^{g-1} |l_j(x)|}, \quad j = 0, \dots, g-1.$$

**Proof.** Applying the Karush-Kuhn-Tucker Theorem (see e.g. [1]), we obtain

$$\begin{cases} \frac{\partial}{\partial w_j} \left( \sum_{i=0}^{g-1} \frac{(l_i(x))^2}{w_i} + \lambda \left( \sum_{i=0}^{g-1} w_i - n \right) \right) = 0 \\ \lambda \left( \sum_{i=0}^{g-1} w_i - n \right) = 0 \\ \lambda \geq 0 \end{cases},$$

$$\frac{\partial}{\partial w_j} \left( \sum_{i=0, i \neq j}^{g-1} \frac{(l_i(x))^2}{w_j} + \frac{(l_j(x))^2}{w_j} + \lambda \left( \sum_{i=0, i \neq j}^{g-1} w_i + w_j - n \right) \right) = 0,$$

$$\lambda = \left( \frac{l_j(x)}{w_j} \right)^2,$$

$$w_j = \frac{|l_j(x)|}{\sqrt{\lambda}}.$$

Since  $\sum_{i=0}^{g-1} w_i = n$ , we get

$$\left( \frac{l_j(x)}{w_j} \right)^2 = \lambda = \left( \frac{\sum_{j=0}^{g-1} |l_j(x)|}{n} \right)^2.$$

Finally we get the solution of the problem (2<sub>a</sub>), namely

$$w_j^* := n \frac{|l_j(x)|}{\sum_{j=0}^{g-1} |l_j(x)|}, \quad j = 0, \dots, g-1.$$

■

**Step 2.** Solving Problem (8) is more tedious and requires some technical arguments, which we develop now.

Substituting  $n_j$  in (2) by  $w_j^*$  the variance of  $\mathcal{L}_n(\hat{f})(x)$  becomes

$$\text{Var} \mathcal{L}_n(\hat{f})(x) = \sigma^2 \left( \sum_{i=0}^{g-1} |l_i(x)| \right)^2.$$

Hence minimizing this variance under the nodes turns into the following problem

$$\begin{cases} \min_{(x_0, \dots, x_{g-1}) \in \mathbb{R}^g} \sum_{i=0}^{g-1} |l_i(x)| \\ x < x_0 < \dots < x_{g-1}, \{x_0, \dots, x_{g-1}\} \subset [-1, 1] \end{cases} \quad (9)$$

Since  $x < x_0 < \dots < x_{g-1}$ , we have

$$|l_i(x)| = (-1)^i l_i(x). \quad (10)$$

We consider a suboptimal solution to Problem (9). Indeed no general solution presently exists to the determination of a system of nodes which minimizes the evaluation of the Lebesgue function  $t \rightarrow \sum_{i=0}^{g-1} |l_i(t)|$  at some fixed point  $t = x < -1$ .

Define

$$t \rightarrow T_{g-1}(t) := \sum_{i=0}^{g-1} (-1)^i l_i(t) \quad (11)$$

a polynomial with degree  $g - 1$  defined on  $\mathbb{R}$ . This polynomial does not depend any longer of  $x$  but only on the nodes. Up to the multiplicative constant  $\sigma$  it coincides with the standard deviation of  $\mathcal{L}_n(\widehat{f})(x)$  when evaluated at point  $x$ . We provide the optimal choice of the nodes under this representation.

Observe the properties of  $T_{g-1}$  on  $[-1, 1]$ . It holds

$$T_{g-1}(x_j) = (-1)^j$$

for all  $j$  between 0 and  $g - 1$ . Hence  $T_{g-1}$  takes its extreme values in  $g$  points; among those points, all interior ones in  $[-1, 1]$  are points where  $T_{g-1}$  changes curvature. Those are obtained as the roots of the first derivative of  $T_{g-1}$ ; indeed  $t \rightarrow T_{g-1}(t)$  is monotonous when  $t$  does not belong to  $[-1, 1]$ . Hence we obtain that  $g - 2$  points among the  $x_i$ 's are interior points in  $[-1, 1]$ . It follows that the two remaining ones are  $-1$  and  $1$ .

We now identify  $T_{g-1}$  and therefore the nodes.

**Lemma 6** *The polynomial  $T_{g-1}$  is the solution of the differential equation*

$$1 - T_{g-1}^2(x) = \frac{1}{(g-1)^2} (1-x^2) \left( \frac{dT_{g-1}(x)}{dx} \right)^2. \quad (12)$$

**Proof.** It should hold

$$\left\{ \begin{array}{l} T_{g-1}(x_j) = (-1)^j \\ \left( \frac{dT_{g-1}(x)}{dx} \right)_{x=x_j} = 0, \quad j = 1, \dots, g-2 \end{array} \right. .$$

Observe that

a) the degree of  $(1 - T_{g-1}^2(x))$  equals  $2(g-1)$  and  $1 - T_{g-1}^2(x_j) = 0$ , which implies that the set of roots of  $1 - T_{g-1}^2(x)$  are  $\{x_0, \dots, x_{g-1}\}$ ;

b) since  $T_{g-1}(x) = a_0 + a_1x + \dots + a_{g-1}x^{g-1}$ , it follows that  $\frac{dT_{g-1}(x)}{dx} = a_1 + a_2x + \dots + (g-1)a_{g-1}x^{g-2}$ , and

$$\left( \frac{dT_{g-1}(x)}{dx} \right)^2 = \beta_1 + \dots + \beta_{g-2}x^{2(g-2)-2}.$$

This implies that

$$\text{degree}(1 - x^2) \left( \frac{dT_{g-1}(x)}{dx} \right)^2 \leq 2(g-1).$$

Furthermore

$$(1 - x^2) \left( \frac{dT_{g-1}(x)}{dx} \right)^2 = 0$$

when  $x \in \{-1, 1\} = \{x_0, x_{g-1}\}$ , and

$$T'_{g-1}(x) = 0 \text{ when } x \in \{x_1, \dots, x_{g-2}\}.$$

From a) and b) we deduce that the polynomials  $1 - T_{g-1}^2(x)$  and  $(1 - x^2) \left( \frac{dT_{g-1}(x)}{dx} \right)^2$  share the same roots and the same degree, hence are equal up to a constant  $K$ , i.e.

$$1 - T_{g-1}^2(x) = K(1 - x^2) \left( \frac{dT_{g-1}(x)}{dx} \right)^2. \quad (13)$$

We determine  $K$ . Let

$$T_{g-1}(x) = \sum_{i=0}^{g-1} a_i x^i.$$

Then the coefficient of  $x^{2(g-1)}$  equals  $a_{g-1}^2$ . Also since

$$\frac{dT_{g-1}(x)}{dx} = \sum_{i=1}^{g-1} i a_i x^{i-1}$$

then the term with highest degree in  $\left(\frac{dT_{g-1}(x)}{dx}\right)^2$  is

$$(g-1)^2 a_{g-1}^2 x^{2(g-1)-2}.$$

By (13) it should hold that the coefficient of greatest degree of  $1 - T_{g-1}^2(x)$  equals the corresponding one of  $K(1-x^2)\left(\frac{dT_{g-1}(x)}{dx}\right)^2$ . This means

$$-a_{g-1}^2 x^{2(g-1)} = -Kx^2 (g-1)^2 a_{g-1}^2 x^{2(g-1)-2}$$

which yields

$$1 - T_{g-1}^2(x) = \frac{1}{(g-1)^2} (1-x^2) \left(\frac{dT_{g-1}(x)}{dx}\right)^2.$$

This is a differential equation with separable variables with solution the  $T_{g-1}(x)$  which we look for. ■

We first solve the differential equation (12).

**Lemma 7** *The solution of (12) under the boundary conditions  $T_{g-1}(-1) = T_{g-1}(1) = 1$  is*

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

**Proof.** Denote  $y = T_{g-1}(v)$ . For  $v = x_{g-1} = 1$ , it should hold  $y = T_{g-1}(1) = 1$ . Therefore at  $v = 1$ ,  $T_{g-1}$  has a maximum. This proves that for  $v < 1$ , with  $v$  close to 1,  $y = T_{g-1}(v)$  is increasing. Let  $v^*$  the point to the left of  $x_{g-1}$  such that from  $v^*$  to  $x_{g-1}$  the function  $y$  is always increasing, and is not increasing before  $v^*$ . Clearly  $v^*$  is a minimizer of  $T_{g-1}$  and  $y(v^*) = -1$ . Therefore on  $[v^*, 1]$  the first derivative  $y'$  is positive.

We can write, therefore

$$\sqrt{(y')^2} = |y'| = y'.$$

It follows that the equation  $1 - y^2 = \frac{1}{(g-1)^2} (1 - v^2)(y')^2$  may be written as

$$\frac{g-1}{\sqrt{1-v^2}} = \frac{y'}{\sqrt{1-y^2}}.$$

Take the primitive on all terms, namely

$$(g-1) \int \frac{dv}{\sqrt{1-v^2}} = \int \frac{y'}{\sqrt{1-y^2}} dv + c.$$

Some attention yields

$$(g-1) \arccos v = \arccos y(v) + c'.$$

Hence

$$\cos((g-1) \arccos v) = \cos(\arccos y + c').$$

Take  $v = 1$ ; then  $y = 1$  so that

$$\cos((g-1) \arccos 1) = \cos(\arccos 1 + c').$$

Now  $\arccos 1 = 2r\pi$ , with  $r = 0, 1, \dots$ . Hence  $\cos(h \arccos 1) = 1$ , writing  $rh = r'$ , with  $r' \in \mathbb{Z}$ . It follows that  $1 = \cos(\arccos 1 + c')$ . Write  $\arccos 1 + c' = \beta$ ; this implies  $1 = \cos \beta$ , i.e.  $\beta = 2r''\pi$ , with  $r'' \in \mathbb{N}$ .

Note that

$$\arccos 1 + c' = 2h'\pi$$

and therefore

$$c' = 2h'\pi - \arccos 1 = 2h'\pi - 2m\pi = 2\pi(h' - m) = 2h''\pi, \text{ with } h'' \in \mathbb{Z}.$$

For the constant  $c'$  we may therefore consider any multiple of  $\pi$ , including 0, i.e.  $\cos((g-1) \arccos v) = \cos(\arccos y + c')$  with  $c' = 0$ .

A solution of the initial differential equation is therefore given by

$$y = \cos((g-1) \arccos v), \text{ for } v \in [v^*, 1].$$

The polynomial  $T_{g-1}$  increases from  $v^*$  to 1, it should decrease at the left of  $v^*$ . Define  $v^{**}$  the point where it starts its decrease. The point  $v^{**}$  is therefore a maximum point with  $y(v^{**}) = 1$  and  $y$  decreases on  $[v^{**}, v^*]$ . Therefore  $y' < 0$  in  $[v^{**}, v^*]$  and

$$-\frac{y'}{\sqrt{1-y^2}} = \frac{g-1}{\sqrt{1-v^2}},$$

since  $\sqrt{(y')^2} = |y'| = -y'$ .

Therefore

$$\begin{aligned}\int -\frac{y'}{\sqrt{1-y^2}}dy &= \int \frac{g-1}{\sqrt{1-v^2}}dv + c, \arccos y \\ &= (g-1)\arccos v + c,\end{aligned}$$

which together with a similar argument as previously yields to adopt  $c = 0$ .

Since  $v^*$  coincides with  $x_{g-2}$ , and  $v^{**} = x_{g-3}$ , we iterate the above arguments for all nodes until  $x_0$ ; we conclude that  $y = \cos((g-1)\arccos v)$ .

Proceeding as above on any of the intervals between the nodes concludes the proof. ■

We now obtain the roots  $\tilde{x}_k$  of the derivative of  $x \rightarrow T_{g-1}(x)$ .

**Proposition 8** *It holds*

$$\tilde{x}_k = \cos\left(\frac{k\pi}{g-1}\right), \text{ for } k = 1, \dots, g-2. \quad (14)$$

**Proof.** We get the roots  $\tilde{x}_k$  through a first order differentiation. With  $\theta = (g-1)\arccos v$  it holds

$$y' = T'_{g-1}(v) = -\sin\theta \frac{d\theta}{dv} = (\sin(h\arccos v)) \frac{g-1}{\sqrt{1-v^2}}.$$

Note that

$$\frac{h}{\sqrt{1-v^2}} \sin(h\arccos v) = 0, \sin(h\arccos v) = 0$$

hence

$$h\arccos v = k\pi, \arccos v = \frac{k\pi}{g-1}, x_k = \cos\left(\frac{k\pi}{g-1}\right).$$

Then we get

$$\begin{aligned}y(\tilde{x}_k) &= \cos((g-1)\arccos x_k) \\ &= \cos\left((g-1)\frac{k\pi}{g-1}\right) = (-1)^k,\end{aligned}$$

which yields

$$\tilde{x}_k = \cos\left(\frac{k\pi}{g-1}\right), \text{ for } k = 1, \dots, g-2.$$

We rename the  $\tilde{x}_k$ 's in increasing order. ■

Note that taking  $k = 0$  and  $k = g - 1$  in (14) we recover  $T_{g-1}(-1) = T_{g-1}(1) = 1$  so that all  $g$  points  $\tilde{x}_k$ ,  $0 \leq k \leq g - 1$  are points of maximal or minimal value of  $T_{g-1}$  in  $[-1, 1]$ .

The optimal design is therefore given by

$$\xi^* := \left\{ \left( \left[ n \frac{|l_k(x)|}{\sum_{j=0}^{g-1} |l_j(x)|} \right]; \cos \left( \frac{k\pi}{g-1} \right) \right), \text{ for } k = 0, \dots, g-1 \right\}.$$

**Definition 9** *The nodes*

$$\tilde{x}_k = \cos \left( \frac{k\pi}{g-1} \right), \text{ for } k = 0, \dots, g-1.$$

are the Chebyshev nodes on  $[-1, 1]$ .

**Example 10** (Hoel - Levine) Consider  $\mathbb{R} \rightarrow \mathbb{R}, x \mapsto P_3(x) := \sum_{j=0}^3 \theta_j x^j$ ,  $\theta_j$ ,  $j = 0; 1; 2; 3$ . We intend to estimate  $P_3(2)$  using  $n = 52$  observations in  $[-1; 1]$ . The model writes as  $y_i(x_j) := P_3(x) + \varepsilon_i$  with  $\varepsilon_i \sim N(0; 1)$  i.i.d, for all  $i$  and  $j$ ,  $i = 1, \dots, n_j$ ,  $j = 0, \dots, 3$ ,  $\sum_{j=0}^3 n_j = 52$ . The optimal design  $\bar{\xi}^*$  is given by:

$$\left( x_j = \cos \left( \frac{j\pi}{h} \right), \left[ n \frac{|l_j(x)|}{\sum_{j=0}^{g-1} |l_j(x)|} \right] \right).$$

Therefore  $x_0 = -1, x_1 = -\frac{1}{2}, x_2 = \frac{1}{2}, x_3 = 1$  and  $\bar{\xi}^*(-1) = \frac{5}{52}, \bar{\xi}^*(-\frac{1}{2}) = \frac{12}{52}, \bar{\xi}^*(\frac{1}{2}) = \frac{20}{52}, \bar{\xi}^*(1) = \frac{15}{52}$ . Instead of this optimal design, consider the design with nodes  $\text{supp}(\xi) = \left\{ -1, \frac{-1}{3}, \frac{1}{3}, 1 \right\}$  and weights  $\xi(-1) = \xi(\frac{-1}{3}) = \xi(\frac{1}{3}) = \xi(1) = \frac{13}{52}$ , it holds  $\text{var}_\xi \left( \mathcal{L}_3(\widehat{P}_3)(2) \right) \sim 20 > \text{var}_{\xi^*} \left( \mathcal{L}_3(\widehat{P}_3)(2) \right) \sim 13$  where  $\text{var}_\xi$  and  $\text{var}_{\bar{\xi}^*}$  denote respectively the variance under the corresponding design.

## 2.1 Hoel Levine optimal design and the uniform approximation of functions

We now make some comment pertaining to the polynomial  $x \rightarrow T_{g-1}(x)$  in relation with the theory of the best uniform polynomial approximation of functions on  $[-1, 1]$ .



Consider the monomial  $x \rightarrow x^{g-1}$  on  $[-1, 1]$  and define its best uniform  $q_{g-2}$  approximation in the linear space of all polynomials with degree less or equal  $g - 2$ . By Theorem 3 there exist  $g$  equi-oscillation points  $z_0, \dots, z_{g-1}$  in  $[-1, 1]$  where

$$z_k^{g-1} - q_{g-2}(z_k) = (-1)^k \sup_{x \in [-1, 1]} |x^{g-1} - q_{g-2}(x)|.$$

Some analysis proves that the coefficient of  $x^{g-1}$  in  $T_{g-1}$  is  $C := 1/2^{g-2}$ .

Observe that the polynomial  $x \rightarrow \tilde{T}_{g-1}(x) := CT_{g-1}(x)$  shares the same oscillation properties and degree as  $x \rightarrow x^{g-1} - q_{g-2}(x)$  on  $[-1, 1]$ , but for its extreme values. Now by Theorem 3 those properties identify  $T_{g-1}$  with the error function  $x \rightarrow x^{g-1} - q_{g-2}(x)$  up to a multiplicative constant.

Consider now the null function  $x \rightarrow 0(x)$  defined on  $[-1, 1]$  and let  $p_{g-1}$  be a generic polynomial defined on  $[-1, 1]$  with degree  $g-1$  and whose coefficient of  $x^{g-1}$  is 1. Then clearly

$$\sup_{x \in [-1, 1]} \left| \tilde{T}_{g-1}(x) - 0(x) \right| \leq \sup_{x \in [-1, 1]} |p_{g-1}(x) - 0(x)|.$$

Note that in the present case when  $x \rightarrow 0(x)$  is to be approximated, uniqueness of the best uniform approximating function is defined up to a multiplicative constant; therefore we may say that  $T_{g-1}$  is, up to the multiplicative constant  $C$  the best uniform approximation of  $x \rightarrow 0(x)$  by polynomials with degree  $g - 1$  in  $[-1, 1]$ .

This fact is the entry point to more general optimal designs when estimating functions outside the context of polynomials; see [10].

### 3 Uniform interpolation optimal designs (Guest)

Consider the uniform variance of the estimator (with respect to  $x$ ). A natural strong criterion for optimality is defined through

$$\min_{\{n_j \in \mathbb{N}^*, j=0, \dots, g-1: \sum_{j=0}^{g-1} n_j = n\}} \max_{x \in [-1; 1]} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right). \quad (15)$$

In this section two goals will be reached. First we obtain the optimal design  $\xi^*$  solving (15). Then we will show that extrapolation designs are of a different nature with respect to interpolation ones, since, as seen below,

$$\text{var}_{\xi^*} \left( \mathcal{L}_n(\hat{f})(x) \right) \neq \min_{\xi} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right), \text{ for } x > 1$$

where the minimum upon  $\xi$  all designs depends on  $x$ . Here we consider an extrapolation design with  $x > 1$ .

Define the Legendre polynomials on  $[-1, 1]$ .

**Definition 11** *The Legendre polynomial of order  $g - 1$  on  $[-1, 1]$  is defined by*

$$\begin{aligned} P_{g-1}(x) &:= \frac{1}{2^{g-1}(g-1)!} \frac{d^{g-1}}{dx^{g-1}} (x^2 - 1)^{g-1} \\ &= 2^{-(g-1)} \sum_{j=0}^{\lfloor \frac{g-1}{2} \rfloor} (-1)^{g-1-j} \binom{g-1}{j} \binom{2(g-1-j)}{g-1} x^{g-1-2j}. \end{aligned}$$

**Remark 12** *The relation  $P_{g-1}(x) := \frac{1}{2^{g-1}(g-1)!} \frac{d^{g-1}}{dx^{g-1}} ((x^2 - 1)^{g-1})$  is known as Rodriguez formula.*

**Remark 13** *Clearly  $P_{g-1}$  has  $g - 1$  roots in  $(-1, 1)$ , as seen now. Indeed the polynomial  $(x^2 - 1)^{g-1}$  has degree  $2(g - 1)$ , and it has multiple roots at points  $\pm 1$ . By Rolle's Theorem its derivative admits a root inside  $(-1, 1)$ . This derivative assumes also the value 0 at  $\pm 1$ , since it has at least three roots in  $[-1, 1]$ . Apply once more Rolle's theorem to the second derivative, which takes value 0 at  $\pm 1$ , since it has at least four roots. Proceeding further, the  $(g - 1)$ -th derivative has  $g - 1$  roots in  $(-1, 1)$ . Up to a constant this derivative is the Legendre polynomial  $P_{g-1}$ .*

**Remark 14** *The value of  $P_{g-1}(x)$  at  $x = \pm 1$  can be obtained. Indeed it holds*

$$(x^2 - 1)^{g-1} = (x - 1)^{g-1} (x + 1)^{g-1}.$$

*By Leibnitz formula*

$$\begin{aligned} &\frac{d^{g-1} ((x - 1)^{g-1} (x + 1)^{g-1})}{dx^{g-1}} \\ &= \sum_{j=0}^{g-1} \binom{g-1}{j} \frac{d^j ((x - 1)^{g-1})}{dx^j} \frac{d^{g-1-j} (x + 1)^{g-1}}{dx^{g-1-j}}. \end{aligned}$$

For  $j = 0, \dots, g - 2$ , it holds

$$\left( \frac{d^j ((x - 1)^{g-1})}{dx^j} \right)_{x=1} = 0$$

and

$$\left( \frac{d^{g-1} (x - 1)^{g-1}}{dx^{g-1}} \right)_{x=1} = (g - 1)!.$$

Henceforth

$$\frac{d^{g-1} ((x - 1)^{g-1} (x + 1)^{g-1})}{dx^{g-1}} = (g - 1)! 2^{g-1}.$$

This yields

$$P_{g-1}(1) = 1 \text{ and } P_{g-1}(-1) = (-1)^{g-1}. \quad (16)$$

We need some facts about the Lagrange elementary polynomials; denoting

$$\pi(x) := \prod_{j=0}^{g-1} (x - x_j)$$

it holds

**Lemma 15** *It holds (i)*

$$\begin{aligned} \left( \frac{d\pi(x)}{dx} \right)_{x=x_j} &= 0 \text{ for } j = 1, \dots, g - 2, \\ \text{iff } \left( \frac{d^2\pi(x)}{dx^2} \right)_{x=x_j} &= 0 \text{ for } j = 1, \dots, g - 2. \end{aligned}$$

(ii)  $\pi(x) = \alpha(x^2 - 1)\phi_{g-2}(x)$ , with

$$\phi_{g-2}(x) = \frac{dP_{g-1}(x)}{dx}$$

where  $P_{g-1}$  is the Legendre polynomial of order  $g - 1$  on  $[-1, 1]$ .

Finally (iii)

$$\left( \frac{d}{dx} l_j(x) \right)_{x=x_j} = 0 \text{ iff } \left( \frac{dP_{g-1}(x)}{dx} \right)_{x=x_j} = 0.$$

**Proof.** Denoting

$$\pi(x) := \prod_{j=0}^{g-1} (x - x_j)$$

write

$$l_j(x) = \frac{\pi(x)}{(x - x_j) \left( \frac{d\pi(x)}{dx} \right)_{x_j}}.$$

We have

$$\pi(x) = (x - x_j) \left( \frac{d\pi(x)}{dx} \right)_{x_j} l_j(x),$$

$$\frac{d\pi(x)}{dx} = (x - x_j) \left( \frac{d\pi(x)}{dx} \right)_{x_j} \frac{dl_j(x)}{dx} + \left( \frac{d\pi(x)}{dx} \right)_{x_j} l_j(x)$$

and

$$\frac{d^2\pi(x)}{dx^2} = \left( \frac{d\pi(x)}{dx} \right)_{x_j} \left\{ (x - x_j) \frac{d^2l_j(x)}{dx^2} + 2 \frac{dl_j(x)}{dx} \right\}.$$

This last display proves (i).

In

$$\pi(x) := \prod_{j=0}^{g-1} (x - x_j)$$

the  $x_j$ 's,  $j = 0, \dots, g - 1$ , are the abscissas where the variance function is minimax. Indeed in (10) we have proved that the absolute value of the elementary Lagrange polynomial takes value 1, which is its maximal value, when evaluated on the nodes. Hence the variance

$$\text{var} \left( \mathcal{L}_n(\widehat{f})(x) \right) = \sum_{i=0}^{g-1} (l_i(x))^2 \frac{\sigma^2}{n_i}$$

takes its maximal values at points  $x_j$ 's.

Hence  $\{-1, 1\} \subset \{x_j, j = 0, \dots, g - 1\}$  and the remaining  $g - 2$   $x_j$ 's are points of maximal value of the variance inside  $(-1; 1)$ . Write the polynomial  $\pi(x)$  as

$$\pi(x) = \alpha (x^2 - 1) \phi_{g-2}(x),$$

where

$$\alpha \phi_{g-2}(x) := \prod_{j=0, j \neq -1; 1}^{g-1} (x - x_j).$$

The polynomial  $\phi_{g-2}$ , with degree  $g-2$  is determined through the conditions

$$\left( \frac{d^2\pi(x)}{dx^2} \right)_{x=x_j} = 0 \text{ for } j = 1, \dots, g-2.$$

Since

$$\frac{d\pi(x)}{dx} = 2\alpha x \phi_{g-2}(x) + \alpha(x^2 - 1) \frac{d(\phi_{g-2}(x))}{dx}$$

and

$$\frac{d^2\pi(x)}{dx^2} = 2\alpha \phi_{g-2}(x) + 4\alpha x \frac{d(\phi_{g-2}(x))}{dx} + \alpha(x^2 - 1) \frac{d^2(\phi_{g-2}(x))}{dx^2}$$

those conditions amount to the system

$$\begin{cases} 0 = 2\alpha \phi_{g-2}(x_1) + 4\alpha x_1 \left( \frac{d(\phi_{g-2}(x))}{dx} \right)_{x=x_1} + \alpha(x_1^2 - 1) \left( \frac{d^2(\phi_{g-2}(x))}{dx^2} \right)_{x=x_1} \\ \dots\dots\dots \\ 0 = 2\alpha \phi_{g-2}(x_j) + 4\alpha x_j \left( \frac{d(\phi_{g-2}(x))}{dx} \right)_{x=x_j} + \alpha(x_j^2 - 1) \left( \frac{d^2(\phi_{g-2}(x))}{dx^2} \right)_{x=x_j} \\ \dots\dots\dots \\ 0 = 2\alpha \phi_{g-2}(x_{g-2}) + 4\alpha x_{g-2} \left( \frac{d(\phi_{g-2}(x))}{dx} \right)_{x=x_{g-2}} + \alpha(x_{g-2}^2 - 1) \left( \frac{d^2(\phi_{g-2}(x))}{dx^2} \right)_{x=x_{g-2}} \end{cases}$$

Now the derivative of the Legendre polynomial  $P_{g-1}$  is precisely the solution of this system (see [6]). Hence

$$\phi_{g-2}(x) = \frac{dP_{g-1}(x)}{dx}.$$

This closes the proof of (ii).

We prove (iii). It holds

$$l_j(x) = \frac{\pi(x)}{(x - x_j) \left( \frac{d\pi(x)}{dx} \right)_{x_j}}.$$

By

$$\frac{d}{dx} l_j(x) = \frac{\alpha(x^2 - 1) \frac{d^2}{dx^2} P_{g-1}(x) + 2\alpha x \frac{d}{dx} P_{g-1}(x)}{(x - \tilde{x}_j) K} - \frac{\alpha(x^2 - 1) \frac{d}{dx} P_{g-1}(x)}{(x - \tilde{x}_j)^2 K}$$

for some constant  $K$ . When  $x = \tilde{x}_j$  then (iii) follows. ■

We now obtain the optimal design. It holds

**Proposition 16** *The nodes of the optimal design  $\xi^*$  are the  $g - 2$  solutions of the equation*

$$\frac{d}{dx}P_{g-1}(x) = 0$$

and  $-1, 1$ . *The optimal frequencies are defined by the relation*

$$n_j = \frac{g}{n}.$$

**Proof.** Keeping the notation  $l_j$  defined in (1), we have

$$\text{var} \left( \mathcal{L}_n(\widehat{f})(x) \right) = \sum_{j=0}^{g-1} l_j^2(x) \frac{\sigma^2}{n_j}.$$

Since  $\frac{\sigma^2}{n_j} > 0$ , any  $\frac{\sigma^2}{n_j}$  should be minimal in order to make the sum minimal. Hence  $(n_0^*, \dots, n_{g-1}^*)$  should solve

$$\begin{cases} \min_{(n_0, \dots, n_{g-1})} \left( \frac{1}{n_0} + \dots + \frac{1}{n_{g-1}} \right) \\ \sum_{j=0}^{g-1} n_j = n \end{cases} . \quad (17)$$

Hence

$$n_j^* = \frac{g}{n}.$$

The polynomial  $\sum_{j=0}^{g-1} l_{x_j}^2(x) \frac{\sigma^2}{n_j}$  has degree  $2g - 2$ , and indeed has  $g - 1$  roots of order 2. This function is a decreasing function of  $x$  on  $(-\infty, -1)$  and an increasing function of  $x$  on  $(1, +\infty)$ ; the points  $-1$  and  $1$  are therefore points of local maximal value of the variance. The variance has therefore  $g$  local extrema in  $[-1, 1]$ . Hence there exist  $g - 2$  local extrema for the variance inside  $(-1, 1)$ ; they lie between the roots of  $\text{var} \left( \mathcal{L}_n(\widehat{f})(x) \right)$ . These extrema are maxima, since the variance is a sum of squares and takes value 0  $g - 2$  times.

We suppose that we have those points at hand; call them  $\tilde{x}_j, j = 0, \dots, g - 1$ . On those points  $\tilde{x}_j$  the function  $\text{var} \left( \mathcal{L}_n(\widehat{f})(x) \right)$  takes the value

$$\sum_{i=0}^{g-1} l_i^2(\tilde{x}_j) \frac{\sigma^2}{n_i} = \frac{\sigma^2}{n_j}$$

with

$$l_i(x) := \prod_{j=0, j \neq i}^{g-1} \frac{x - \tilde{x}_j}{\tilde{x}_i - \tilde{x}_j}.$$

The function  $x \rightarrow l_j^2(x)$  takes its maximal value for  $x = \tilde{x}_j$ , with  $l_j^2(\tilde{x}_j) = 1$  independently on  $j$ . Therefore it holds

$$\max_{x \in [-1;1]} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right) = \max_{x \in [-1;1]} \sum_{j=0}^{g-1} l_j^2(x) \frac{\sigma^2}{n_j}.$$

The principle leading to the optimal design should now be made precise. The largest variance of  $\left( \mathcal{L}_n(\hat{f})(x) \right)$  should be attained on the points of measurements, in order to be able to control it. Consider two nodes  $\tilde{x}_i$  and  $\tilde{x}_k$ . Then

$$\text{var} \left( \mathcal{L}_n(\hat{f})(\tilde{x}_i) \right) = \text{var} \left( \mathcal{L}_n(\hat{f})(\tilde{x}_k) \right) = \frac{\sigma^2 g^2}{n}.$$

Hence

$$\max_{x \in [-1;1]} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right) = \frac{\sigma^2 g^2}{n}.$$

The nodes should hence be the points of maximal value of the variance, which equals  $\frac{\sigma^2 g^2}{n}$ .

The first derivative of  $\text{var} \left( \mathcal{L}_n(\hat{f})(x) \right)$  writes

$$\frac{d}{dx} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right) = \frac{2g}{n} \sum_{j=0}^{g-1} l_j(x) \frac{d}{dx} l_j(x).$$

It follows that finding the  $g - 2$  internal nodes  $\tilde{x}_j$ 's results in finding the solutions of the equation

$$\frac{d}{dx} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right)_{x=\tilde{x}_j} = 0 \tag{18}$$

which by the above argument turns out to solve

$$\left( \sum_{j=0}^{g-1} l_j(x) \frac{d}{dx} l_j(x) \right)_{x=\tilde{x}_j} = 0$$

which yields, since

$$l_i(\tilde{x}_j) = \delta_{i,j}$$

$$\left( \frac{d}{dx} l_j(x) \right)_{x=\tilde{x}_j} = 0 \text{ for all } j = 1, \dots, g-2.$$

This is a system of  $g-2$  equations in the  $g-2$  variables  $\tilde{x}_1, \dots, \tilde{x}_{g-2}$ . This system has precisely  $g-2$  solutions, solving (18).

Apply Lemma 15 (iii) to conclude. ■

We now characterize the performance of the optimal design  $\xi^*$  through an evaluation of the minimax variance (15)

$$\min_{\{n_j \in \mathbb{N}^*, j=0, \dots, g-1: \sum_{j=0}^{g-1} n_j = n\}} \max_{x \in [-1; 1]} \text{var} \left( \mathcal{L}_n(\hat{f})(x) \right).$$

**Lemma 17** *The Legendre polynomial  $P_{g-1}$  is a solution of the following differential equation (so-called Legendre equation)*

$$(1-x^2) \frac{d^2 f(x)}{dx^2} - 2x \frac{df(x)}{dx} + g(g-1) f(x) = 0$$

i.e

$$\frac{d}{dx} \left( (x^2-1) \frac{d}{dx} f(x) \right) = g(g-1) f(x) \quad (19)$$

**Proof.** For an analytic function  $f$  on  $D$ , by Cauchy formula, it holds

$$f^{(g-1)}(x) = \frac{(g-1)!}{2\pi i} \int_{\gamma} \frac{f(z)}{(z-x)^g} dz$$

where  $x$  is an interior point in  $D$  and  $\gamma$  is a regular circuit in  $D$  with  $x$  in its interior. The variable  $z$  runs on  $\gamma$  in the positive sense. Apply this formula to the analytic function

$$f(z) = (z^2-1)^{g-1}, g = 0, 1, 2, \dots$$

By Rodriguez formula we obtain the following relation, known as Schläfli formula

$$P_{g-1}(x) = \frac{1}{2\pi i} \int_{\gamma} \frac{(z^2-x)^{g-1}}{2^{g-1}(z-x)^g} dz.$$



Substituting now  $f$  by  $P_{g-1}$  in Legendre equation and applying the above formula, we get

$$\begin{aligned} & (1-x^2) \frac{d^2 P_{g-1}}{dx^2} - 2x \frac{dP_{g-1}}{dx} + g(g-1) P_{g-1} \\ &= \frac{g}{2^{g-1} 2\pi i} \int_{\gamma} \left( \frac{d}{dz} \left( \frac{(z^2-1)^g}{(z-x)^{g+1}} \right) \right) dz. \end{aligned}$$

Now

$$\frac{g}{2^{g-1} 2\pi i} \int_{\gamma} \left( \frac{d}{dz} \left( \frac{(z^2-1)^g}{(z-x)^{g+1}} \right) \right) dz = 0.$$

This can be written through

$$\frac{d}{dx} \left( (x^2-1) \frac{dP_{g-1}(x)}{dx} \right) = g(g-1) P_{g-1}(x).$$

Indeed

$$\frac{d}{dx} \left( (x^2-1) \frac{dP_{g-1}(x)}{dx} \right) = 2x \frac{dP_{g-1}(x)}{dx} + (x^2-1) \frac{d^2 P_{g-1}}{dx^2}$$

and therefore

$$\frac{d}{dx} \left( (x^2-1) \frac{dP_{g-1}(x)}{dx} \right) = g(g-1) P_{g-1}(x)$$

which is

$$2x \frac{dP_{g-1}(x)}{dx} + (x^2-1) \frac{d^2 P_{g-1}}{dx^2} - g(g-1) P_{g-1}(x) = 0$$

which proves the claim. ■

We evaluate the local variance of the design of Guest for any point  $x$  in  $[-1, 1]$ .

We now turn back to the points where the variance assumes its maximal values. It holds

$$\pi(x) = \alpha (x^2-1) \frac{dP_{g-1}(x)}{dx}$$

hence

$$\begin{aligned}\frac{d\pi(x)}{dx} &= \frac{d}{dx} \left( \alpha (x^2 - 1) \frac{dP_{g-1}(x)}{dx} \right) \\ &= \alpha g (g - 1) P_{g-1}(x)\end{aligned}$$

by Lemma 17.

Therefore

$$\frac{d^2\pi(x)}{dx^2} = \alpha g (g - 1) \frac{dP_{g-1}(x)}{dx}.$$

We evaluate the minimax variance, which we denote by  $var_{\xi^*}$ .

It holds

$$\begin{aligned}var_{\xi^*} \left( \mathcal{L}_n(\widehat{f})(x) \right) &= \sum_{j=0}^{g-1} l_j^2(x) \frac{\sigma^2}{n_j} \\ &= \sum_{j=0}^{g-1} \left( \frac{\pi(x)}{(x - x_j) \left( \frac{d\pi(x)}{dx} \right)_{x_j}} \right)^2 \frac{g\sigma^2}{n} \\ &= \left( (x^2 - 1) \frac{dP_{g-1}(x)}{dx} \right)^2 \frac{\sigma^2}{g(g-1)^2 n} \sum_{j=0}^{g-1} \left( \frac{1}{(x - x_j) P_{g-1}(x_j)} \right)^2.\end{aligned}$$

Making use of Lobatto formula and after some calculus (see [7]), we obtain

$$var_{\xi^*} \left( \mathcal{L}_n(\widehat{f})(x) \right) = \left( 1 + \frac{x^2 - 1}{g(g-1)} \left( \frac{d^2P_{g-1}(x)}{dx^2} \right)^2 \right) \frac{g\sigma^2}{n}.$$

As a consequence, using Guest minimax design, the maximal variance of the interpolation is obtained at the boundaries  $x = -1$  and  $x = 1$ . By symmetry the minimal variance of the interpolation holds when  $x = 0$ .

In the extrapolation zone, namely for large  $|x| > 1$

$$\frac{dP_{g-1}(x)}{dx} \sim (g-1) \frac{(2(g-1))!}{2^{g-1} ((g-1)!)^2} x^{g-2}.$$

In the extrapolation zone this yields to the approximation

$$\text{var}_{\xi^*} \left( \mathcal{L}_n(\widehat{f})(x) \right) \sim (g-1) \left( \frac{(2(g-1)!)}{2^{g-1} ((g-1)!)^2} \right)^2 x^{2(g-1)} \frac{\sigma^2}{g-1}.$$

Considering the points  $x_0 = -1, x_{g-1} = 1$  which are also points of maximum variance we see that the maximal variance cannot exceed  $\frac{g}{n} \sigma^2$ .

We have obtained the optimal minimax design in the interpolation range. We now prove that this design is not suitable for extrapolation.

## 4 The interplay between the Hoel-Levine and the Guest designs

Without loss of generality we may consider the case when  $c > 1$ ; by

$$\text{var}_{\xi} \left( \mathcal{L}_n(\widehat{f})(x) \right) = \sum_{j=0}^{g-1} l_j^2(x) \frac{\sigma^2}{n_j}$$

the variance of  $\mathcal{L}_n(\widehat{f})(x)$ , say  $\text{var}_{\xi} \left( \mathcal{L}_n(\widehat{f})(x) \right)$  is an increasing function of  $x$  for  $x > 1$  for any design  $\xi$  since the mapping  $x \rightarrow l_j^2(x)$  increases for  $x \geq 1$ . It follows that for any  $c > 1$  the Hoel Levine design  $\xi_c$  is the minimax optimal extrapolation design on  $(1, c)$  namely it solves

$$\min_{\xi \in \mathcal{M}_1^*} \max_{x \in (1, c]} \text{var}_{\xi} \left( \mathcal{L}_n(\widehat{f})(x) \right).$$

However there is no reason that  $\xi_c$  be minimax optimal on whole  $[-1, c]$  since it might not solve

$$\min_{\xi \in \mathcal{M}_1^*} \max_{x \in [-1, c]} \text{var}_{\xi} \left( \mathcal{L}_n(\widehat{f})(x) \right).$$

We consider the optimal minimax design on  $[-1, c]$  with  $c > 1$  and discuss its existence and properties.

On  $[-1, 1]$  the optimal minimax design is Guest's design. We will prove (see Proposition 21 hereunder) that this design is not minimax optimal on  $[-1, c]$  with  $c > 1$  for large  $c$ .

At the contrary we prove (Proposition 18) hereunder that the Hoel Levine design  $\xi_c$  is minimax optimal on  $[1, c]$ .

Finally we prove (Proposition 19) that there exists a unique  $c^* \gg 1$  such that  $\xi_{c^*}$  is minimax optimal on  $[-1, c^*]$ .

**Proposition 18** *The Hoel Levine optimal design  $\xi_c$  is minimax optimal on  $[1, c]$  for  $c > 1$  as proved in Section 2 (substitute  $c < -1$  by  $c > 1$ ).*

**Proof.** This is a consequence of the fact that  $x \rightarrow \text{var}_{\xi_c} \left( \mathcal{L}_n(\widehat{f})(x) \right)$  is an increasing function on  $[1, c]$ . ■

**Proposition 19** *There exists  $c_1 \gg 1$  such that the Hoel Levine design  $\xi_{c^*}$  is minimax optimal on  $[-1, c_1]$ , i.e. it solves*

$$\min_{\xi \in \mathcal{M}_1^*} \max_{x \in [-1, c_1]} \text{var}_{\xi} \left( \mathcal{L}_n(\widehat{f})(x) \right).$$

**Proof.** We have seen that for  $1 < x < c$ , the solution provided by Hoel and Levine is minimax optimal. We now consider the case when  $[1, c]$  is substituted by  $[-1, c]$  with  $c > 1$ .

In this case the minimax optimal solution still holds as the Hoel - Levine design if  $c$  "large enough" .

Indeed let  $\text{var}_{\eta} \left( \mathcal{L}_n(\widehat{f})(x) \right)$  be the variance under a design  $\eta$  whose support consists in the Chebyshev nodes in  $[-1, 1]$ . The design  $\eta$  at this point is not defined in a unique way, since the values of  $\eta(x_j)$  is not specified.

The function

$$x \rightarrow \text{var}_{\eta} \left( \mathcal{L}_n(\widehat{f})(x) \right)$$

is continuous on  $[-1, 1]$ . Denote

$$v_{\eta}^* := \max_{x \in [-1, 1]} \text{var}_{\eta} \left( \mathcal{L}_n(\widehat{f})(x) \right).$$

Assume that there exists some  $c > 1$  which does not depend on  $\eta$  such that

$$v_{\eta}^* < \text{var}_{\eta} \left( \mathcal{L}_n(\widehat{f})(c) \right). \quad (20)$$

In such case it holds

$$\min_{\eta \in \mathcal{M}_{[-1,1]}^*} v_\eta^* < \min_{\eta \in \mathcal{M}_{[-1,1]}^*} \text{var}_\eta \left( \mathcal{L}_n(\widehat{f})(c) \right).$$

The minimizing measure on the right hand side of the above display is precisely the extrapolation Hoel Levine design at  $c$  since the function  $x \rightarrow \text{var}_\eta \left( \mathcal{L}_n(\widehat{f})(x) \right)$  is increasing for  $x > 1$ .

It remains to prove that such  $c$  satisfying (20) exists.

For a given  $c$  let

$$\begin{aligned} R(c) &:= \frac{\max_{[-1;1]} \left( \text{var}_\eta \left( \mathcal{L}_n(\widehat{f})(x) \right) \right)}{\text{var}_\eta \left( \mathcal{L}_n(\widehat{f})(c) \right)} \\ &= \frac{\max_{[-1;1]} \left( \sum_{j=0}^{g-1} \frac{l_j^2(x)}{n_j} \right)}{\sum_{j=0}^{g-1} \frac{l_j^2(c)}{n_j(c)}}, \end{aligned}$$

with

$$n_j(c) := \frac{|l_j(c)|}{\sum_{i=0}^{g-1} |l_i(c)|}$$

where the  $n_j(c)$ ,  $0 \leq j \leq g-1$  are the optimal frequencies of the Hoel - Levine design evaluated in  $x = c$ .

We intend to prove that some  $c > 1$  exists for which  $R(c) < 1$ .

If this holds then

$$\begin{aligned} R(c) &= \frac{\left( \sum_{i=0}^{g-1} |l_i(c)| \right) \left( \max_{[-1;1]} \left( \sum_{j=0}^{g-1} \frac{l_j^2(x)}{|l_j(c)|} \right) \right)}{\left( \sum_{i=0}^{g-1} |l_i(c)| \right) \left( \sum_{j=0}^{g-1} |l_j^2(c)| \right)} \\ &= \frac{\left( \max_{[-1;1]} \left( \sum_{j=0}^{g-1} \frac{l_j^2(x)}{|l_j(c)|} \right) \right)}{\left( \sum_{j=0}^{g-1} |l_j(c)| \right)}. \end{aligned}$$

Any of the  $|l_j(c)|$  is an increasing function of  $c$  for  $c > 1$ ; therefore  $R(c)$  is a decreasing function of  $c$  for  $c > 1$ . Since each  $l_j(c) \rightarrow \infty$  as  $c \rightarrow \infty$ ,  $R(c)$  will approach 0 as  $c \rightarrow \infty$ .

Since  $l_j(c) \rightarrow 0$  as  $c \rightarrow 1$ , for all  $j$ ,  $R(c)$  will become infinite as  $c \rightarrow 1$ . But  $R(c)$  is a continuous function of  $c$  for  $c > 1$ ; consequently there will exist a unique value of  $c$ , denote by  $c_1$ , satisfying  $R(c_1) = 1$ . For  $c > c_1$ ,  $R(c) < 1$ ; this entails that  $c$  exists with (20).

The proof of Proposition 19 is completed. ■

**Remark 20** *The analytic derivation of  $c_1$  is presented in [13].*

It follows from the same type of arguments as that just used to reject the possibility of a Legendre (or Guest) design for  $c > 1$  that the Hoel - Levine design cannot be optimum for  $c < c_1$ . From continuity considerations one would expect the optimum design to gradually change from the Guest spacing and weighting to the Hoel - Levine spacing and weighting as  $c$  increases from 1 to  $c_1$ . This is still an open question.

**Proposition 21** *The Guest design  $\xi^*$  is not minimax optimal on  $[-1, c]$  for any  $c > 1$ , which is to say that it not an optimal extrapolating design.*

**Proof.** By Proposition 18 the Hoel Levine design on  $[1, c]$  is minimax optimal for large  $c > 1$ . By uniqueness of the optimal design, following from the optimization problem, we deduce that Guest design cannot coincide with this design. ■

## 5 Confidence bound for interpolation/extrapolation designs

Using a minimax optimal design we may produce a confidence bound for  $f(x)$  at any point  $x$  in  $[-1, 1]$  or for  $x$  far away from  $[-1, 1]$ . We thus consider two cases for the location of  $x$ . When  $x$  belongs to  $[-1, 1]$  then the optimal design is the Guest one. By Proposition 19 the Hoel Levine design is minimax on  $[-1, c_1]$  for large  $c_1$ . The minimax variance on  $[-1, c_1]$  is therefore the variance of  $\widehat{f(c_1)}$  since  $\text{var}(\widehat{f(x)})$  is an increasing function of the variable  $x$  for  $x > 1$ .

Write

$$f(x) = \sum_{j=0}^{g-1} l_j(x) f(x_j) = \mathbf{l}(x) \mathbf{f}(\tilde{\mathbf{x}})$$

where  $\tilde{\mathbf{x}} := (x_0, \dots, x_{g-1})'$  are the Chebychev nodes,  $\mathbf{l}(x) := (l_0(x), \dots, l_{g-1}(x))'$  and

$$\mathbf{f}(\tilde{\mathbf{x}}) := \begin{pmatrix} f(x_0) \\ \cdot \\ f(x_j) \\ \cdot \\ f(x_{g-1}) \end{pmatrix}.$$

Assume that  $y_i(x_j) := f(x_j) + \varepsilon_{i,j}$  with  $\varepsilon_{i,j} \sim N(0, 1)$  i.i.d, for all  $i$  and  $j$ ,  $i = 1, \dots, n_j$ ,  $j = 0, \dots, g-1$ ,  $\sum_{j=0}^{g-1} n_j = n$ , where the  $n$  observations are measured on  $[-1; 1]$ . The pointwise unbiased estimator of  $f(c)$ ,  $c > 1$ , is given by

$$\widehat{f(c)} := \mathbf{l}(c) \widehat{\mathbf{f}(\tilde{\mathbf{x}})}$$

where

$$\widehat{\mathbf{f}(\tilde{\mathbf{x}})} = \left( \frac{1}{n_0} \sum_{i=1}^{n_0} y_i(x_0), \dots, \frac{1}{n_{g-1}} \sum_{i=1}^{n_{g-1}} y_i(x_{g-1}) \right).$$

Since the distribution of the  $y_i(x_j)$ 's is  $N(f(x_j), 1)$  for all  $i = 1, \dots, n_j$  and every  $j = 0, \dots, g-1$ , the variance of the estimator  $\widehat{f(c_1)}$  is given by

$$\begin{aligned} \text{var}(\widehat{f(c_1)}) &= \text{var} \left( \sum_{j=0}^{g-1} l_j(c_1) \frac{\sum_{i=1}^{n_j} y_i(x_j)}{n_j} \right) \\ &= \sum_{j=0}^{g-1} \frac{(l_j(c_1))^2}{n_j}. \end{aligned}$$

where the  $n_j$ 's are the frequencies of the Hoel Levine design evaluated at point  $c_1$  (which is indeed the minimax optimal design on  $[-1, c_1]$  as argued above). The confidence set for  $f(c_1)$  is given by

$$C_n := (\mathbf{l}(c_1))' \widehat{\mathbf{f}(\tilde{\mathbf{x}})} \pm \sqrt{p_\alpha \sum_{j=0}^{g-1} \frac{(l_j(c_1))^2}{n_j}}.$$

where

$$\Pr(N(0, 1) > p_\alpha) = 1 - \alpha$$

and  $N(0, 1)$  is a random variable distributed with a standard normal law. It holds

$$\Pr(C_n \ni f(c_1)) \geq 1 - \alpha.$$

When the variance of the  $\varepsilon_i$ 's are unknown then it can be approximated by

$$s^2 := \frac{\sum_{j=0}^{g-1} (n_j - 1) s_j^2}{n - g - 2}$$

where

$$s_j^2 := \frac{\sum_{i=1}^{n_j} (y_i(x_j) - (\sum_{i=1}^{n_j} y_i(x_j)) / n_j)^2}{n - g - 2}.$$

The confidence area for  $f(x)$  becomes

$$C_n := (\mathbf{1}(x))' \widehat{\mathbf{f}}(\tilde{\mathbf{x}}) \pm \sqrt{q_{\alpha/2} \sum_{j=0}^{g-1} \frac{(l_j(c_1))^2}{n_j s_j}}$$

where

$$\Pr(|t_{g-2}| > q_{\alpha/2}) = 1 - \alpha$$

where  $t_{g-2}$  is a Student r.v. with  $g - 2$  degrees of freedom.

## 6 An application of the Hoel - Levine design, a multivariate case

### 6.1 Some examples

The above discussion may be applied for more general situations including the regression models. We refer to the location/scale models, which are of broad interest. Let

$$Z = \frac{Y(x) - \mu(x)}{\sigma},$$

$(\sigma, \mu) \in \mathbb{R}^+ \times \mathbf{F}$ , with  $(\sigma, \mu)$  unknown and  $\mathbf{F}$  a known class of functions. The scale parameter  $\sigma$  is constant w.r.t.  $x$  and  $Z$  is a r.v. which is absolutely continuous w.r.t. the Lebesgue measure. Its distribution is assumed to be known and does not depend on  $x$ .



Write

$$f(x) := \mu(x) + \sigma E(Z),$$

$$\varepsilon := \sigma Z - \sigma E(Z),$$

and therefore write the location/scale model as

$$Y(x) = f(x) + \varepsilon.$$

We consider some examples.

**Example 22** *The importance of the Weibull distribution in Reliability is well known. Denote  $T$  a Weibull r.v. with distribution function*

$$F(t) = 1 - \exp\left(-\left(\frac{t}{\mu(x)}\right)^\beta\right), t \geq 0.$$

It can be written into

$$\ln T = \ln \mu(x) + \frac{1}{\beta} \ln(-\ln(1 - F(T))),$$

and therefore

$$Y(x) = \ln \mu(x) + \sigma Z.$$

where we wrote

$$Y(x) := \ln T, \sigma := \frac{1}{\beta}, Z := \ln(-\ln(1 - F(T))).$$

The model is therefore

$$Z = \frac{Y(x) - \ln \mu(x)}{\sigma}.$$

Observe that

$$\Pr(Z > t) = e^{-e^t}, t > 0.$$

Thus  $Z$  is the Gumbel standard r.v.

Write the above model defining

$$\varepsilon := \sigma Z - \sigma E(Z),$$

so that

$$Y(x) = f(x) + \varepsilon,$$

where

$$f(x) := \ln \mu(x) + \sigma E(Z).$$

**Example 23** For a Gaussian r.v.  $X \sim N(\mu(x), \sigma^2)$ , it holds

$$Z = \frac{X - \mu(x)}{\sigma} \sim N(0, 1).$$

**Example 24** A regression model is clearly of the preceding type.

**Example 25** Assume that  $T$  is logistic, i.e.

$$F(t) = 1 - \left( 1 + \exp\left(\frac{t - f(x)}{\beta}\right) \right)^{-1}.$$

When  $\beta = 1$ , we may write

$$1 - F(t) = \frac{1}{1 + \exp(t - f(x))},$$

$$1 + \exp(t - f(x)) = \frac{1}{1 - F(t)}, \quad \exp(t - f(x)) = \frac{F(t)}{1 - F(t)}.$$

It is enough to state

$$Z := \ln \frac{F(t)}{1 - F(t)}, \quad \varepsilon := Z - E(Z), \quad Y := T$$

to get

$$Y = f(v) + E(Z) + \varepsilon.$$

## 6.2 Multivariate optimal designs; a special case

We extend the results of the above sections to a bivariate setting in a reliability context; extension to multivariate similar cases is straightforward.

We consider the extrapolation problem with two variables.

Let

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \mathbf{x} := (x, y) \mapsto f(x, y) := \sum_{i_1=0}^{g_1-1} \sum_{i_2=0}^{g_2-1} a_{i_1 i_2} x^{i_1} y^{i_2}, \quad a_{i_1 i_2} \in \mathbb{R},$$

be a polynomial in the two variables  $x, y$  with partial degrees  $g_i - 1$ ,  $i = 1, 2$  in the variables  $x, y$ . The polynomial  $f$  has  $M_1 := g_1 g_2$  unknown coefficients.

In order to determine these coefficients we observe  $f$  on a finite set  $\mathcal{E}$  in  $\mathbb{R}^2$ . The fact that  $\mathcal{E}$  consists in  $M_1$  distinct points in  $\mathbb{R}^2$  is not sufficient for the estimation of the coefficients; it is necessary that these points do not belong to an algebraic curve (or algebraic hypersurface in higher dimension). Indeed the identification for a polynomial with many variables usually does not have a unique solution. For example consider  $n$  points  $\{(x_i, y_i) : i = 0, \dots, n-1\} \subset \mathbb{R}^2$ , together with  $n$  known values of  $f$  on those points  $\{f(x_i, y_i) : i = 0, \dots, n-1\}$ ; then there may not exist a unique polynomial  $P(x, y)$ , such that

$$f(x_i, y_i) = P(x_i, y_i), \quad (x_i, y_i) \in \mathcal{E}.$$

Indeed it is enough to consider the case when the  $n$  distinct points are on a line in  $\mathbb{R}^2$ . In this case there exists an infinite number of planes  $z = ax + by + c$ , which contain the  $n$  points  $(x_i, y_i)$ .

We will therefore assume that the  $M_1$  points which define  $\mathcal{E} \subset \mathbb{R}^2$  do not belong to an algebraic curve. This implies existence and uniqueness for a polynomial which coincides with  $f$  on  $\mathcal{E}$ , with partial degree  $g_1 - 1$  with respect to  $x$  and  $g_2 - 1$  w.r.t.  $y$ . Denote  $P_{\mathcal{E}}(f)(\cdot)$  this polynomial.

It can be proved that  $(x, y) \rightarrow P_{\mathcal{E}}(f)(x, y)$  satisfies

$$f(x, y) = P_{\mathcal{E}}(f)(x, y) = \sum_{(x_i, y_i) \in \mathcal{E}} f(x_i, y_i) Q_i(\mathcal{E}, (x, y))$$

where the polynomials  $Q_i(\mathcal{E}, (x, y))$  do not depend on  $f$ . Indeed we may make  $Q_i(\mathcal{E}, \cdot)$  explicit; see e.g. [9] p 248-251.

Consider  $\mathcal{E}$  a finite subset of the compact set  $S := \times_{i=1}^2 [a_i, b_i]$ . Let the points  $(x_i, y_i)$  in  $\mathcal{E}$  be

$$(x_{i_j}, y_{i_j})$$

$i_j = 0, \dots, g_j - 1$  and  $j = 1, 2$ .

Define the elementary Lagrange polynomial in two variables by

$$l_{i_1 i_2}(x, y) := l_{i_1}(x) l_{i_2}(y) \tag{21}$$

where

$$l_{i_1}(x) := \frac{\prod_{\substack{h_1=0 \\ h_1 \neq i_1}}^{g_1-1} (x - x_{h_1})}{\prod_{\substack{h_1=0 \\ h_1 \neq i_1}}^{g_1-1} (x_{i_1} - x_{h_1})}, \quad l_{i_2}(z) := \frac{\prod_{\substack{h_2=0 \\ h_2 \neq i_2}}^{g_2-1} (z - z_{h_2})}{\prod_{\substack{h_2=0 \\ h_2 \neq i_2}}^{g_2-1} (z_{i_2} - z_{h_2})}$$

are the elementary Lagrange polynomials with respect to the coordinates  $x$  and  $y$ .

Clearly

$$l_{i_1 i_2}(x, y) = \begin{cases} 1 & \text{if } (x, y) = (x_{i_1}, y_{i_2}) \\ 0 & \text{otherwise} \end{cases}.$$

The set

$$\{l_{i_1 i_2}(x, y) : i_j = 0, \dots, g_j - 1, j = 1, 2\}$$

is a basis for the linear space of all polynomials with partial degree with respect to the coordinate  $x_j$  less or equal  $g_j - 1$ , for  $j = 1, 2$ .

The Gram matrix associated with this basis

$$(l_{i_1 i_2}(x, y))_{i_1=0, \dots, g_1-1, i_2=0, \dots, g_2-1}$$

is therefore invertible; by uniqueness we have

$$Q_i(\mathcal{E}, (x, y)) = l_{i_1 i_2 i}(x, y).$$

Therefore

$$P_{\mathcal{E}}(f)(x, y) := \mathcal{L}(P_{\mathcal{E}}(f))(x, y)$$

where we wrote

$$\mathcal{L}(P_{\mathcal{E}}(f))(x, y) = \sum_{i_1=0}^{g_1-1} \sum_{i_2=0}^{g_2-1} f(x_{i_1}, y_{i_2}) l_{i_1 i_2}(x, y). \quad (22)$$

The above formula (22) holds true since the nodes  $(x_{i_1}, y_{i_2})$  belong to a rectangle (see [9]).

The polynomial  $\mathcal{L}(P_{\mathcal{E}}(f))(x, y)$  is called the bivariate Lagrange polynomial. The points in  $\mathcal{E}$  are the nodes for the interpolation of  $f$ .

When  $f$  is not a polynomial but merely a function defined on  $S$ , which can be extended by continuity on an open set  $\mathcal{O}$  which contains  $S$ , then the Lagrange interpolation scheme may be used as an approximation scheme on  $\mathcal{O}$ ; see [3].

By uniqueness we adopt the notation

$$P_{\mathcal{E}}(f)(x, y) = \mathcal{L}(P_{\mathcal{E}}(f))(x, y) = f(x, y).$$

We now assume the following model

$$Z := \frac{Y(x, y) - f(x, y)}{\sigma}$$

where  $Z$  is a r.v. totally known in distribution with finite expectation  $E(Z)$  and finite variance  $\eta^2$ ; the scale parameter  $\sigma > 0$  is unknown and does not depend on  $x, y$ ; the coefficients of  $P(x, y)$ ,  $a_{i_1 i_2} \in \mathbb{R}$ , are also unknown. We will see that the optimal design does not depend on the constants  $\sigma^2$  nor  $\eta^2$ .

Denote

$$\varepsilon(x, y) := \sigma Z - \sigma E(Z)$$

whose variance equals  $\eta^2 \sigma^2$ .

It holds

$$Y(x, y) = f(x, y) + \sigma E(Z) + \varepsilon(x, y).$$

We assume further that  $f(x, y)$  can be observed only on a subset  $S$  in  $\mathbb{R}^2$ . In the setting of accelerated runs, this subset  $S$  is the stressed domain;

it will be assumed that it is a rectangle  $[a_1, b_1] \times [a_2, b_2]$  in  $\mathbb{R}^2$ , a choice which is achievable by the experimenter. This shape allows for an important simplification for the definition and the calculation of the optimal design.

The lexicographic order on  $\mathbb{R}^2$  is defined as follows; for  $(x, y)$  and  $(z, t)$  in  $\mathbb{R}^2 \times \mathbb{R}^2$ ,  $(x, y) \lesssim (z, t)$  iff  $x \leq z, y \leq t$ .

Denote  $\mathbf{a} := (a_1, a_2)$  be the point in  $\mathbb{R}^2$  which describes the threshold between the standard operational values of the environment and the stressed conditions. With respect to  $\mathbf{a}$  the stressed region is a rectangle  $[\mathbf{a}, \mathbf{b}]$  north-east with respect to  $\mathbf{a}$ , with south-west corner at  $\mathbf{a}$ , whereas the unstressed domain is the south west quadrant  $U$  with north east corner at  $\mathbf{a}$ . We denote  $\mathbf{u}$  a point in

$$U := \{(x, y) \in \mathbb{R}^2 : (x, y) \lesssim \mathbf{a}\}.$$

We intend to find an optimal design in order to estimate the value of the polynomial  $f$  at point  $\mathbf{u}$ , hence we look for  $\mathcal{E}$  and for the number of observations on any of the points in  $\mathcal{E}$ , in such a way to make the variance of the estimate of  $P(\mathbf{u})$  minimal.

Let  $(x_{i_1}, y_{i_2}) \in S$  be a node, i.e. a stress configuration. We consider now the set of trials under this configuration. Denoting  $\underline{i} := (i_1, i_2) \in \times_{j=1}^2 \{0, \dots, g_j - 1\}$ , we define  $n(\underline{i})$  be the total number of replications of the measurement  $Y$  at point  $(x_{i_1}, y_{i_2})$ . We denote  $\underline{\mathbf{Y}}(\underline{i})$  the vector of these measurements; We assume that the coordinates of  $\underline{\mathbf{Y}}(\underline{i})$  are ordered; this is the common procedure when looking at lifetimes of a number  $n(\underline{i})$  of identical systems operating in parallel during the trial. So  $\underline{\mathbf{Y}}(\underline{i}) := (Y_{(1)}(\underline{i}), \dots, Y_{(n(\underline{i}))}(\underline{i}))$  is an ordered sample obtained from an i.i.d. sample with size  $n(\underline{i})$ .

The system of equations which represents the observations is therefore

$$(1) \left\{ \begin{array}{l} y_{(1)}(x_{i_1}, y_{i_2}) = f(x_{i_1}, y_{i_2}) + \sigma E(Z) + \varepsilon_1(x_{i_1}, y_{i_2}) \\ \dots \\ y_{(k)}(x_{i_1}, y_{i_2}) = f(x_{i_1}, y_{i_2}) + \sigma E(Z) + \varepsilon_k(x_{i_1}, y_{i_2}) \\ \dots \\ y_{(n(\underline{i}))}(x_{i_1}, y_{i_2}) = f(x_{i_1}, y_{i_2}) + \sigma E(Z) + \varepsilon_{n(\underline{i})}(x_{i_1}, y_{i_2}) \end{array} \right. .$$

It holds

$$\sum_{i_1=0}^{g_1-1} \sum_{i_2=0}^{g_2-1} n(\underline{i}) = M_1$$

with  $M_1 \in \mathbb{N} \setminus \{0\}$  a fixed number.

We assume that the vectors of measurements  $\underline{\mathbf{Y}}(\underline{i})$  and  $\underline{\mathbf{Y}}(\underline{h})$  are independent for  $\underline{i} \neq \underline{h}$ .

The system (1) may be written as

$$\underline{\mathbf{Y}}(\underline{i}) = X\theta + \varepsilon(\underline{i})$$

where  $X$  is the matrix with  $n(\underline{i})$  rows and 2 columns

$$X(\underline{i}) := \begin{pmatrix} 1 & E(Z) \\ \cdot & \cdot \\ 1^{n(\underline{i})} & E(Z) \end{pmatrix}$$

and  $\theta$  is a column vector with two rows

$$\theta := \begin{pmatrix} f(x_{i_1}, y_{i_2}) \\ \sigma \end{pmatrix}.$$

Finally

$$\underline{\varepsilon}(\underline{i}) := \begin{pmatrix} \varepsilon_1(\underline{i}) \\ \cdot \\ \varepsilon_{n(\underline{i})}(\underline{i}) \end{pmatrix}, \underline{\mathbf{Y}}(\underline{i}) := \begin{pmatrix} y^{(1)}(x_{i_1}, y_{i_2}) \\ \cdot \\ y^{(n(\underline{i}))}(x_{i_1}, y_{i_2}) \end{pmatrix}.$$

Denote

$$\Omega^{-1}(\underline{i}) := (\text{cov}(y^{(a)}(\underline{i}), y^{(b)}(\underline{i})))_{a,b}^{-1}$$

which we assume to exist for all  $\underline{i}$ . In the above display,  $\Omega^{-1}(\underline{i})$  is a matrix of order  $n(\underline{i})$ ; the matrix  $\Omega(\underline{i})$  is not the identity matrix of order  $n(\underline{i})$  since the vector of observations is ordered according to its coordinates.

$$\Omega^{-1}(\underline{i}) := \begin{pmatrix} \omega_{1,1} & \cdot & \omega_{1,n(\underline{i})} \\ \cdot & \cdot & \cdot \\ \omega_{n(\underline{i}),1} & \cdot & \omega_{n(\underline{i}),n(\underline{i})} \end{pmatrix}.$$

The expected value of the measurement  $Y$  at point  $(x_{i_1}, y_{i_2})$  equals  $f(x_{i_1}, y_{i_2}) + \sigma E(Z)$ . Denote  $m_{i_1, i_2}$  its GLS estimator

$$m_{\underline{i}} := m_{i_1, i_2} = [(X'(\underline{i}) \Omega^{-1}(\underline{i}) X(\underline{i}))^{-1} X'(\underline{i}) \Omega^{-1}(\underline{i}) \underline{\mathbf{Y}}(\underline{i})].$$

This estimator is strongly consistent.  
We now define the estimator outside of the nodes.  
Observe that for  $\mathbf{u} := (x, y)$

$$\begin{aligned} \mathcal{L}(P_{\mathcal{E}}(f))(x, y) + \sigma E(Z) &:= f(x, y) + \sigma E(Z) \\ &= \sum_{(x_i, y_i) \in \mathcal{E}} (f(x_i, y_i) + \sigma E(Z)) l_{i_1 i_2}(x, y). \end{aligned}$$

Denote  $m(\mathbf{u})$  the resulting estimator of  $f(\mathbf{u}) + \sigma E(Z)$

$$\begin{aligned} m(\mathbf{u}) &:= \sum_{(x_i, y_i) \in \mathcal{E}} m_{\underline{i}} l_{i_1 i_2}(\mathbf{u}) \\ &= \sum_{(x_i, y_i) \in \mathcal{E}} m_{\underline{i}} l_{i_1}(x) l_{i_2}(y). \end{aligned}$$

This factorization relies on the fact that  $S$  is a rectangle. We now evaluate the variance of the unbiased estimator  $m(\mathbf{u})$ ; by independence of the measurements on the nodes

$$\begin{aligned} \text{Var}(m(\mathbf{u})) &= \sum_{\underline{i}=(i_1, i_2)}^2 (l_{i_1}(x) l_{i_2}(y))^2 \text{var}(m_{\underline{i}}) \\ &= \sum_{\underline{i}=(i_1, i_2)} (l_{i_1}(x) l_{i_2}(y))^2 G(\sigma^2 \eta^2, \Omega^{-1}(\underline{i}), X'(\underline{i})), \end{aligned}$$

where

$$G(\sigma^2 \eta^2, \Omega^{-1}(\underline{i}), X'(\underline{i})) := \sigma^2 \eta^2 (X'(\underline{i}) \Omega^{-1}(\underline{i}) X(\underline{i}))^{-1} X'(\underline{i}). \quad (23)$$

Note that  $\text{var}(\mathcal{L}(P_{\mathcal{E}}(f))(x, y)) \rightarrow 0$ , for  $n(\underline{i}) \rightarrow \infty$ , due to the fact that the generalized least-squares estimator is consistent under the present conditions.

The optimal design results as the solution to the following optimization problem,

$$\left\{ \begin{array}{l} \min \sum_{\underline{i}=(i_1, i_2)} (l_{i_1 i_2}(x, y))^2 G(\sigma^2 \eta^2, \Omega^{-1}(\underline{i}), X'(\underline{i})) \\ \mathbf{u} \in U \\ M_1 = \sum_{\underline{i}} n(\underline{i}). \end{array} \right.$$



where the minimization is held on all choices of the set of measurements (nodes)  $\mathcal{E}$  and all frequencies  $n(\underline{i})$ .

Although the problem generally has a numerical solution, in some practical cases it is possible to obtain an analytic solution.

We explore a special case.

Define

$$\Gamma := \sum_{m,u}^{n(\underline{i})} \omega_{m,u}.$$

Let  $E(Z) = 0$  and the distribution of  $Y$  be symmetric around  $E(Y(u))$ . In this case,  $G(\sigma^2\eta^2, \Omega^{-1}(\underline{i}), X'(\underline{i}))$  becomes:

$$G_1 := G(\sigma^2\eta^2, \Omega^{-1}(\underline{i}), X'(\underline{i})) = \frac{\sigma^2\eta^2}{\Gamma}$$

which depends on  $\underline{i}$  through  $\Gamma$ ; (See [14] for the proof).

In some cases  $G_1$  may be simplified as follows,

$$G_2 := g(\sigma^2\eta^2, \Omega^{-1}(\underline{i}), X'(\underline{i})) = \frac{\sigma^2\eta^2}{n(\underline{i})}. \quad (24)$$

Indeed a necessary and sufficient condition for  $G_2$  is

$$\begin{pmatrix} 1 \\ \cdot \\ 1 \end{pmatrix}' \Omega(\underline{i}) = \begin{pmatrix} 1 \\ \cdot \\ 1 \end{pmatrix}'.$$

(see [4]).

In many cases the function  $G$  takes on the form

$$\frac{\sigma^2\eta^2}{\alpha n(\underline{i}) + \beta} (1 + o(1))$$

where  $\alpha, \beta$  are constants depending on the (known) distribution of the random variable  $Z$ , extending (24); see [2].

The problem of determining the optimal design becomes

$$\begin{cases} \min \sum_{(i_1, i_2)} \frac{l_{(i_1, i_2)}(\mathbf{u})^2}{\alpha n(\underline{i}) + \beta} \\ M_1 = \sum_{\underline{i}} n(\underline{i}), n(\underline{i}) \in \mathbb{R}^+. \end{cases}$$

where the minimum holds on the choice of the nodes  $\mathcal{E}$  and on the frequencies. Fix  $(x_{i_1}, y_{i_2})$  and apply the Theorem of Karush-Kuhn-Tucker to

$$\left\{ \begin{array}{l} \min \sum_{(i_1, i_2)} \frac{(l_{(i_1, i_2)}(\mathbf{u}))^2}{\alpha n(\underline{i}) + \beta} \\ M_1 = \sum_{\underline{i}} n(\underline{i}), n(\underline{i}) \in \mathbb{R}^+ \end{array} \right. .$$

where the minimization is held on the frequencies  $n(\underline{i})$ . We obtain

$$[n^*(\underline{i})] = \frac{|l_{(i_1, i_2)}(\mathbf{u})| \left( \alpha M_1 + \beta \prod_{j=1}^2 g_j \right)}{\sum_{(i_1, i_2)=0} |l_{(i_1, i_2)}(\mathbf{u})|} - \beta.$$

Clearly,  $n^*(\underline{i})$  depends on the  $(x_{i_1}, y_{i_2})$ 's. We substitute  $n^*(\underline{i})$  in the variance formula, i.e. in

$$\sum_{i=0}^l (l_{(i_1, i_2)}(\mathbf{u}))^2 \frac{\sigma^2 \eta^2}{\alpha n(\underline{i}) + \beta},$$

with  $l := (g_1 - 1)(g_2 - 1)$  to obtain

$$\text{Var}(m(\mathbf{u})) = \frac{\sigma^2 \eta^2}{\alpha M_1 + \beta \left( \prod_{j=1}^2 g_j \right)} \left( \sum_{i=0}^l |l_{(i_1, i_2)}(\mathbf{u})| \right)^2.$$

The optimal design hence will not depend on the value of  $\sigma^2 \eta^2$ . Optimizing with respect to the  $(x_{i_1}, y_{i_2})$ 's under the constraint  $\mathbf{u} \in U$ , yields

$$\min_{\{(x_{i_1}, y_{i_2})_{i_1, i_2}\}} \sum_{i=0}^l |l_{(i_1, i_2)}(\mathbf{u})|.$$

This is the same as the following two problems with one variable

$$\min \sum_{i_j=0}^{g_j-1} |l_{(i_1, i_2)}(\mathbf{u})|, \quad j = 1, 2.$$

With  $\mathbf{a} := (a_1, a_2)$ , the minimization is held when  $j = 1$  on the abscissas  $x_j$ 's all larger than  $a_1$  and smaller than  $b_1$  and on the ordinates  $y_j$ 's, all larger than  $a_2$  and smaller than  $b_2$  when  $j = 2$ , since

$$\sum_{i=0}^g |l_{(i_1, i_2)}(\mathbf{u})| = \sum_{i_1=0}^{g_1-1} |l_{i_1}(u_1)| \sum_{i_2=0}^{g_2-1} |l_{i_2}(u_2)| \quad (25)$$

with  $\mathbf{u} := (u_1, u_2)$ .

Now minimizing the product in (25) results in two independent minimizations, one for each of the two factors, under the corresponding constraint on the respective terms  $u_1$  and  $u_2$ . It follows that the optimal design is the combination of two Hoel Levine marginal optimal designs.

Therefore the solution coincides with the previously obtained one, namely the Hoel Levine design of Section 2, i.e.

$$s_j^*(i_j) = \frac{a_j + b_j}{2} + \frac{b_j - a_j}{2} \cos\left(\frac{g_j - 1 - i_j}{g_j - 1} \pi\right), \quad i_j = 0, \dots, g_j - 1, j = 1, 2$$

with  $S := [a_1, b_1] \times [a_2, b_2]$ .

## References

- [1] Mokhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty, *Nonlinear programming*, third ed., Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006, Theory and algorithms. MR 2218478 (2006k:90001)
- [2] Giorgio Celant, *Extrapolation and optimal designs for accelerated runs*, Ann. I.S.U.P. **47** (2003), no. 3, 51–84. MR 2056619
- [3] Christian Coatméc, *Approximation et interpolation des fonctions différentiables de plusieurs variables*, Ann. Sci. École Norm. Sup. (3) **83** (1966), 271–341. MR 0232143 (38 #469)
- [4] F. Downton, *Least-squares estimates using ordered observations*, Ann. Math. Statistics **25** (1954), 303–316. MR 0061334 (15,810b)
- [5] 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. MR 2447076 (2009f:30001)

- [6] P. G. Guest, *The spacing of observations in polynomial regression*, Ann. Math. Statist. **29** (1958), 294–299. MR 0094883 (20 #1392)
- [7] F. B. Hildebrand, *Introduction to numerical analysis*, McGraw-Hill Book Company, Inc., New York-Toronto-London, 1956. MR 0075670 (17,788d)
- [8] P. G. Hoel and A. Levine, *Optimal spacing and weighting in polynomial prediction*, Ann. Math. Statist. **35** (1964), 1553–1560. MR 0168102 (29 #5367)
- [9] Lee W. Johnson and Ronald Dean Riess, *Numerical analysis*, second ed., Addison-Wesley Publishing Co., Reading, Mass., 1982. MR 668699 (83j:65001)
- [10] Samuel Karlin and William J. Studden, *Optimal experimental designs*, Ann. Math. Statist. **37** (1966), 783–815. MR 0196871 (33 #5055)
- [11] J. Kiefer and J. Wolfowitz, *Optimum extrapolation and interpolation designs. I, II*, Ann. Inst. Statist. Math. **16** (1964), 79–108; *ibid.* **16** (1964), 295–303. MR 0178549 (31 #2806)
- [12] A. N. Kolmogorov and S. V. Fomin, *Elementy teorii funktsii i funktsionalnogo analiza*, fifth ed., “Nauka”, Moscow, 1981, With a supplement “Banach algebras” by V. M. Tikhomirov. MR 630899 (83a:46001)
- [13] A. Levine, *A problem in minimax variance polynomial extrapolation*, Ann. Math. Statist. **37** (1966), 898–903. MR 0195215 (33 #3418)
- [14] E. H. Lloyd, *On the estimation of variance and covariance*, Proc. Roy. Soc. Edinburgh. Sect. A. **63** (1952), 280–289. MR 0048756 (14,64i)
- [15] Bryan P. Rynne and Martin A. Youngson, *Linear functional analysis*, second ed., Springer Undergraduate Mathematics Series, Springer-Verlag London Ltd., London, 2008. MR 2370216 (2008i:46001)
- [16] W. J. Studden, *Optimal designs on Tchebycheff points*, Ann. Math. Statist **39** (1968), 1435–1447. MR 0231497 (37 #7050)