An equioscillation theorem for multivariate Chebyshev approximation

Alexandre Goldsztejn

[a] *CNRS, Ecole Centrale de Nantes, Nantes, France*

Abstract

The equioscillation theorem interleaves the Haar condition, the existence and uniqueness and strong uniqueness of the optimal Chebyshev approximation and its characterization by the equioscillation condition in a way that cannot extend to multivariate approximation: Rice [*Transaction of the AMS*, 1963] says "A form of alternation is still present for functions of several variables. However, there is apparently no simple method of distinguishing between the alternation of a best approximation and the alternation of other approximating functions. This is due to the fact that there is no natural ordering of the critical points." In addition, in the context of multivariate approximation the Haar condition is typically not satisfied and strong uniqueness may hold or not. The present paper proposes an multivariate equioscillation theorem, which includes such a simple alternation condition on error extrema, existence and a sufficient condition for strong uniqueness. To this end, the relationship between the properties interleaved in the univariate equioscillation theorem is clarified: first, a weak Haar condition is proposed, which simply implies existence. Second, the equioscillation condition is shown to be equivalent to the optimality condition of convex optimization, hence characterizing optimality independently from uniqueness. It is reformulated as the synchronized oscillations between the error extrema and the components of a related Haar matrix kernel vector, in a way that applies to multivariate approximation. Third, an additional requirement on the involved Haar matrix and its kernel vector, called strong equioscillation, is proved to be sufficient for the strong uniqueness of the solution. These three disconnected conditions give rise to a multivariate equioscillation theorem, where existence, characterization by equioscillation and strong uniqueness are separated, without involving the too restrictive Haar condition. Remarkably, relying on optimality condition of convex optimization gives rise to a quite simple proof. Instances of multivariate problems with strongly unique, non-strong but unique and non-unique solutions are presented to illustrate the scope of the theorem.

Keywords: Chebyshev approximation problem, multivariate approximation, equioscillation theorem, convex optimization

---

1. The multivariate equioscillation theorem

Given $f : X \to \mathbb{R}$ continuous, we consider the Chebyshev approximation problem

$$\min_{a\in\mathbb{R}^n} \max_{x\in X} \Big| \sum_{i=1}^{n} a_i\, \phi_i(x) - f(x) \Big|, \tag{1}$$

where basis function $\phi_i : X \to \mathbb{R}$ are continuous. The index set X is just assumed to be a compact topological space, so that continuity is well defined and entails the existence of the maximum (typically, we will just consider boxes in $\mathbb{R}^n$). The basis function vector $\phi : X \to \mathbb{R}^n$ is defined by $\phi(x) = (\phi_1(x), \dots, \phi_n(x))$. We define $e(a, x) = a^T \phi(x) - f(x)$ and

$$m(a) = \max_{x\in X} |e(a, x)|, \tag{2}$$

so that the Chebyshev approximation problem is $\min_{a\in\mathbb{R}^n} m(a)$.

Introducing non-square Haar matrices will be convenient not only to express the classical Haar condition. Given $x_1, \dots, x_k \in X$, we define the Haar matrix

$$H(x_1, \dots, x_k) = \begin{pmatrix} \phi(x_1) & \phi(x_2) & \cdots & \phi(x_k) \end{pmatrix} \in \mathbb{R}^{n\times k}. \tag{3}$$

When $x_1 < \cdots < x_k$ and $\phi(x) = (1, x, \dots, x^k)$, a Haar matrix is a Vandermonde matrix. The *Haar condition* holds for X and ϕ if and only if for all $x_1, \cdots, x_n \in X$ the square Haar matrix $H(x_1, \dots, x_n)$ is nonsingular. When $X = [\underline{x}, \overline{x}] \subseteq \mathbb{R}$ is a compact interval, the *equioscillation condition* is satisfied for $a \in \mathbb{R}^n$ if and only if there exist $x_1 < \cdots < x_{n+1} \in X$ such that $|e(a, x_i)| = m(a)$ for all $i \in \{1, \dots, n+1\}$ and $e(a, x_i)\, e(a, x_{i+1}) < 0$ hold for all $i \in \{1, \dots, n\}$. Values of $x \in X$ that satisfy $|e(a, x_i)| = m(a)$ will be called active indices at $a \in \mathbb{R}^n$. With this settings, the equioscillation theorem characterizes the solutions of the Chebyshev univariate approximation problem: if the Haar condition holds then the Chebyshev approximation problem has one strongly unique[1] solution, which is equivalently characterized by the equioscillation condition. The Haar condition in central in this statement: in Cheney's standard textbook [1], the Haar condition is assumed in the statement of the Alternation Theorem on page 75, as well in the Unicity Theorem on page 80 and in the Strong Unicity Theorem on page 80. As said previously, both the Haar condition and the equioscillation condition don't extend to multivariate approximation.

The Haar condition cannot be assumed anymore in the context of multivariate approximation: it entails that X is either a circle or a compact interval [2] (up to homeomorphism), hence it cannot hold for multivariate approximation with other index sets, e.g., boxes. We still need a similar but weaker condition, and define the *weak Haar condition* as follows: it holds for X and ϕ if and only

[1]Strong uniqueness means that not only $m(a) - m(a^*) > 0$ for $a \neq a^*$, but that $m(a) - m(a^*) \geq c\, \|a - a^*\|_{X,\phi}$ for some $c > 0$ and the norm $\|a - a^*\|_{X,\phi} = \max_{x\in X} \|(a - a^*)^T \phi(x)\|$. The definition is in fact independent of the norm since all norms are equivalent in $\mathbb{R}^n$.

if <u>there exist</u> $x_1, \cdots, x_n \in X$ such that the square Haar matrix $H(x_1, \ldots, x_n)$ is nonsingular. The weak Haar condition is likely to hold in most situations, e.g., $X = [-1,1] \times [-1,1]$, $\phi(x) = (1, x_1, x_2)$ where $H\big((0,0),(1,0),(1,1)\big)$ is nonsingular. The weak Haar condition has several consequences, among which $\text{card}(X) \geq n$, the existence of a minimizer of the Chebyshev approximation problem (cf. Section 3.1), the fact that $\|a\|_{X,\phi} = \max_{x \in X} |a^T \phi(x)|$ is a norm on $\mathbb{R}^n$ (the triangular inequality and absolute homogeneity are trivial, the positive definiteness follows directly from the weak Haar condition), with the consequence that strong uniqueness implies uniqueness.

The equioscillation condition requires a total order on indices and therefore does not make any sense for multivariate approximation. We reformulate it in a way that can apply more generally. To this end, we rely on the following lemma, which generalizes the well-known fact that Vandermonde matrices with one more column than rows have kernel vectors with oscillating signs, and its subsequent observation.

Lemma 1. *If $\phi : [\underline{x}, \overline{x}] \to \mathbb{R}^n$ is continuous and satisfies the Haar condition on $[\underline{x}, \overline{x}]$ then all Haar matrices $H(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n \times (n+1)}$, with $x_1 < \cdots < x_{n+1} \in X$, have a one dimensional kernel* $\text{span}\{u\}$ *with $0 \neq u \in \mathbb{R}^{n+1}$ having nonzero components with alternating signs, i.e., $u_i\, u_{i+1} < 0$ for all $i \in \{1, \ldots, n\}$.*

The proof of the lemma is given in Appendix A. The connection between the Haar matrix kernel vector component oscillations $u_i\, u_{i+1} < 0$, the equioscillation $e(a, x_i)\, e(a, x_{i+1}) < 0$ and the optimality of the Chebyshev approximation problem can be seen by introducing the matrix

$$G(x_1, \ldots, x_d) \;=\; \big(\epsilon(a, x_1)\phi(x_1) \;\cdots\; \epsilon(a, x_d)\phi(x_d)\big), \tag{4}$$

where $\epsilon(a, x) = \text{sign}(e(a, x))$, which will be used only when $e(a, x) \neq 0$ hence $\text{sign}(e(a, x)) \in \{-1, 1\}$. The capitale G is for "G"radient: the columns of $G(x_1, \ldots, x_d)$ are gradients of $|e(a, x_i)| = \epsilon(a, x)\, e(a, x)$, provided that $|e(a, x_i)| > 0$. When x_i are active indices for $a^* \in \mathbb{R}^n$, we will see that the columns of $G(x_1, \ldots, x_d)$ are actually subgradients of $m(a^*)$. The optimality condition of convex optimization states that a^* is optimal if and only if 0 is in the convex hull of the columns of $G(x_1, \ldots, x_d)$, or equivalently to $G(x_1, \ldots, x_d)$ having a kernel vector $\lambda \neq 0$ whose components are of the same sign (see Lemma 6 below). Finally, observe that by definition we have

$$G(x_1, \ldots, x_d)\, \lambda \;=\; H(x_1, \ldots, x_d)\, u\,, \text{ with } u_i = \epsilon(a^*, x_i)\, \lambda_i. \tag{5}$$

From this we see that all components λ_i have the same signs, i.e., a^* is optimal, if and only if all $u_i\, e(a^*, x_i)$ have the same sign. This condition generalizes the univariate equioscillation condition when the Haar condition is assumed: Lemma 1 proves that $u_i\, u_{i+1} < 0$, so if in addition $e(a, x_i)\, e(a, x_{i+1}) < 0$ then all $u_i\, e(a^*, x_i)$ have the same sign. In fact, requiring $u_i\, e(a^*, x_i) \geq 0$ for some kernel vector of $H(x_1, \ldots, x_d)$ is more convenient and equivalent since the opposite of a kernel vector is also in the kernel. These observations will be formalized in the next sections. Under this form, the equioscillation condition can be used for multivariate Chebyshev approximation problems. Uniqueness is not granted anymore in the context of multivariate Chebyshev approximation, so a technical sufficient condition for uniqueness is added below for the definition of the strong equioscillation condition.

Definition 2 (Multivariate equioscillation condition)**.** *The* (multivariate) equioscillation condition *is satisfied at $a \in \mathbb{R}^n$ if and only if there exist $x_1, \cdots, x_K \in X$, $K \geq 1$, such that $|e(a, x_i)| = m(a)$, and a kernel vector $0 \neq u \in \mathbb{R}^K$ of $H(x_1, \ldots, x_K)$ such that $e(a, x_i)\, u_i \geq 0$ hold for all $i \in \{1, \ldots, K\}$.*

The strong (multivariate) equioscillation condition *holds if and only if furthermore $H(x_1, \ldots, x_K)$ is full rank, hence $K \geq n + 1$, and u has no zero components.*

First note that by Lemma 1 the Haar condition and the (univariate) equioscillation condition together imply the strong multivariate equioscillation with $K = n + 1$. We draw several direct consequences of this definition: the active indices $x_1, \cdots, x_K \in X$ need not to be ordered anymore, so the definition can be applied in the context of multivariate approximation. In fact, the definition is not affected by any permutation the active indices $x_1, \cdots, x_K \in X$: indeed, the columns of the Haar matrix, together with the components of its kernel vectors, permute simultaneously with permutations of the active indices, so the sign synchronization $e(a, x_i)\, u_i \geq 0$ is not affected by permutation of the active indices. Another observation is that $K < n + 1$ implies that $H(x_1, \ldots, x_K)$ not full rank: indeed in this case $H(x_1, \ldots, x_K)$ has more rows than columns and has a non trivial kernel. In particular, if $K = 1$ then $H(x_1)$ has to be a zero matrix (see Example 11 for such a non trivial situation with infinitely many minimizers).

These conditions lead to the following generalization of the equioscillation theorem to multivariate Chebyshev approximation problems, where X is not restricted anymore to be any interval or circle.

Theorem 3 (Multivariate equioscillation theorem)**.** *Let $\phi : X \to \mathbb{R}^n$ be continuous. If the weak Haar condition holds for ϕ and X then there exists an optimal solution to the Chebyshev approximation problem. Furthermore, $a \in \mathbb{R}^n$ is a minimizer of this problem if and only it satisfies the multivariate equioscillation condition. Finally, if the strong multivariate equioscillation condition holds then the minimizer is strongly unique.*

The proof is given in Section 3. The usual equioscillation theorem for univariate approximation is a simple consequence of Theorem 3: on the one hand by Lemma 1 the equioscillation implies the multivariate equioscillation, on the other hand the Haar condition implies both the weak Haar condition and, together with Lemma 1, the strong multivariate equioscillation.

The rest of the paper is organized as follows: some basic definitions and properties of convex functions are presented in Section 2. Section 3 presents the proof of the multivariate equioscillation theorem. Finally, several instances of multivariate problems are presented in Section 5 to illustrate the scope of the theorem.

2. Standard convex properties of $m(a)$

The function $m : \mathbb{R}^n \to \mathbb{R}$ in (2) is called a pointwise supremum, and it's properties are well known in the context of nonsmooth convex optimization. The reader is referred to [3] for an introduction to convex optimization. First not that the function $|e(a,x)|$ is convex, and the pointwise supremum of convex functions is convex, hence $m(a)$ is convex. Convex functions defined in $\mathbb{R}^n$ being continuous, so is $m(a)$. The subdifferential of $m(a)$ is of central importance here. It is denoted by $\partial m(a)$ and is made of all subgradients: $u \in \mathbb{R}^n$ is a subgradient at $\bar{a}$ if it gives rise to a affine under-estimator, i.e., $m(a) \geq m(\bar{a}) + u^T(a - \bar{a})$. The optimality condition for unconstrained convex optimization then reads a^* is a minimizer of $m(a)$ is and only if $0 \in \partial m(a^*)$.

When $|e(a,x)|$ is defined in $\mathbb{R}^n \times X$ with values in $\mathbb{R}$, and X is compact, the pointwise supremum is actually an unconstrained maximum[2] and its subdifferential enjoys a simple explicit expression [3, Lemma 3.1.14]: $\partial m(a) = \mathrm{conv}\{\partial_a |e(a,x)| : x \in \mathrm{act}(a)\}$ with $\mathrm{act}(a) = \{x \in X : |e(a,x)| = m(a)\}$ is the set of active indices. Let us illustrate this formula on a simple finite pointwise maxima example, which presents a (strongly) unique minimizer.

Example 4. *Let $U \in \mathbb{R}^{2\times 3}$ with columns $u_1 = (-2,1)^T$, $u_2 = (1,1)^T$ and $u_3 = (1,-3)^T$. Define the pointwise maximum $m(a) = \max\{u_1^T a, u_2^T a, u_3^t a\} = \|U^T a\|_\infty$, which is a piecewise linear function. Figure 1 shows the level sets of $m(a)$. From darker to lighter level sets correspond to areas where $u_1^T a$, $u_2^T a$ or $u_3^t a$ is active. Red line are the place where to linear functions are equal, and the three linear functions are equal at the origin. The subdifferential are represented in blue: when one linear constraint is active, the pointwise maximum is differentiable and the subdifferential contains only the gradient. On red lines, where two linear constraints are active, the subdifferential is the convex hull of the two corresponding gradients, hence a segment. Finally, at the origin where the three functions are actives, the subdifferential is the convex hull of the three gradients, hence a triangle. From this analysis, we see that the origin is the only point where the subdifferential contains zero, hence the only minimizer of $m(a)$.*

In the typical situation where $\min_{a\in\mathbb{R}^n} m(a) > 0$, the subdifferential of $m(a)$ has a simple expression: in this case, we have that $m(a) > 0$ and $|e(a,x)|$ differentiable for all $a \in \mathbb{R}^n$ and $x \in \mathrm{act}(a)$. Therefore, by [3, Lemma 3.1.7] the subdifferential of $|e(a,x)|$ contains only its gradient, i.e., $\partial e(a,x) = \{\epsilon(a,x)\,\phi(x)\}$ with $\epsilon(a,x) = \mathrm{sign}\big(e(a,x)\big)$ whose value is inside $\{-1,1\}$ for $x \in \mathrm{act}(a)$. Finally we have

$$\partial m(a) = \mathrm{conv}\{\epsilon(a,x)\,\phi(x) : x \in \mathrm{act}(a)\}. \tag{6}$$

[2] In the convex analysis litterature, the wording pointwise supremum applied to infinite possibly compact index sets, and the working pointwise maximum is restricted to finite index sets.

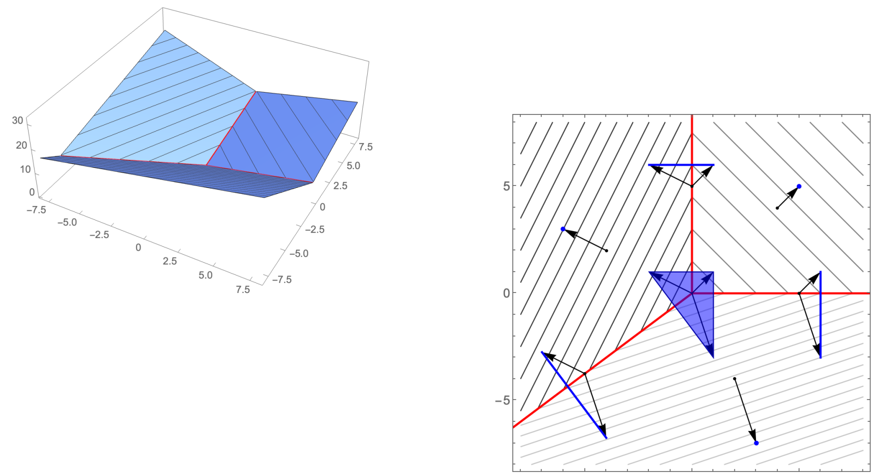

Figure 1: Pointwise maximum of Example 4, with the subdifferential evaluated at several points.

Also note that $\epsilon(a,x)\,\phi(x)$ for $x \in \operatorname{act}(\hat{a})$ are subgradients at $\hat{a}$, therefore they give rise to affine lower bounds

$$m(a) \geq m(\hat{a}) + (\epsilon(a,x)\,\phi(x))^T(a - \hat{a}). \tag{7}$$

This section is ended with the following two simple lemmas. The first provides a sufficient condition for the uniqueness of the minimizer of some pointwise maximum of linear functions. It applies directly to the pointwise maximum function of Example 4. In the sequel, the maximal component of the vector a u is denoted by $\max u$.

Lemma 5. *Let $U = \begin{pmatrix} u_1\, u_2\, \cdots\, u_d \end{pmatrix} \in \mathbb{R}^{n\times d}$, $d \geq n+1$, be full rank and suppose that $\lambda \in \ker(U)$ with $\lambda_i > 0$ for all $i \in \{1,\ldots,d\} = I$. Then the piecewise linear function $l(a) = \max\bigl(U^T a\bigr) = \max\{u_1^T a, \ldots, u_d^T a\}$ has the origin as unique minimizer.*

Proof. By way of contradiction, suppose $a \neq 0$ and $l(a) \leq 0$, i.e., $a^T u_i \leq 0$ for all $i \in I$ and in matrix form $U^T a \leq 0$. Now $U^T a \neq 0$ (because U^T is full rank and $a \neq 0$) therefore exists $i^* \in I$ such that $a^T u_{i^*} < 0$. Finally, $a^T u_{i^*} = -\sum_{i \in I\setminus\{i^*\}} \frac{\lambda_i}{\lambda_{i^*}} a^T u_i \geq 0$, since $\frac{\lambda_i}{\lambda_{i^*}} > 0$ and $a^T u_i \leq 0$, a contradiction. $\square$

Finally, the following lemma provides a simple matrix formulation of the property that the null vector belongs to a the convex hull of some given vectors.

Lemma 6. *Let* $U = \begin{pmatrix} u_1\, u_2\, \cdots\, u_m \end{pmatrix} \in \mathbb{R}^{n\times m}$. *Then* $0 \in \operatorname{conv}\{u_1, \ldots, u_m\}$ *if and only if* $0 \neq \lambda \in \ker(U)$ *with* $\lambda_i \geq 0$.

Proof. The only if part is a direct application of the convex hull definition. Now suppose that $0 \neq \lambda \in \ker(U)$ with $\lambda_i \geq 0$ and define $\mu = \frac{1}{\|\lambda\|_1}\lambda$. Then $\mu \in \ker(U)$, i.e., $\sum \mu_i u_i = 0$, $\mu_i \geq 0$ and $\sum \mu_i = \frac{\|\lambda\|_1}{\|\lambda\|_1} = 1$ hence $0 \in \operatorname{conv}\{u_1, \ldots, u_m\}$. □

3. Proof of the multivariate equioscillation theorem

First, if $\min_{a\in\mathbb{R}^n} m(a) = 0$ then by the weak Haar condition there is a unique solution ($a^T\phi(x_i) = f(x_i)$ with n linearly independent vectors $\phi(x_i)$ uniquely defines a), which is trivially identified with a vacuous multivariate equioscillation condition. We now suppose that $\min_{a\in\mathbb{R}^n} m(a) > 0$, hence the subdifferential formula (6) holds true.

3.1. Existence of a minimizer

By the weak Haar condition, there exists n distinct points $x_1, \ldots, x_n \in X$ such that $H(x_1, \ldots, x_n)$ is nonsingular. We build lower bounds for $m(a)$. First $m(a) \geq \max_{x\in X} \left(|a^T\phi(x)| - |f(x)|\right) \geq (\max_{x\in X}\ |a^T\phi(x)|) - M$ where M is an upper bound of the continuous function f inside the compact X. Finally performing the maximum over $\{x_1, \ldots, x_n\} \subseteq X$ we obtain the lower bound $(\max_{x\in\{x_1,\ldots,x_n\}}\ |a^T\phi(x)|) - M$. The pointwise maximum in this last lower bound is $\|H(x_1, \ldots, x_n)^T a\|_\infty$, which is coercive since $H(x_1, \ldots, x_n)$ is nonsingular (indeed $\sqrt{n}\,\|H^T a\|_\infty \geq \|H^T a\|_2 \geq \lambda_{\min}(HH^T)\,\|a\|_2$ while $\lambda_{\min}(HH^T) > 0$ because H is nonsingular). Continuous coercive functions defined in $\mathbb{R}^n$ have a minimizer.

3.2. Necessity of the multivariate equioscillation condition at the minimizer

Consider a minimizer $a^* \in \mathbb{R}^n$, hence $0 \in \partial m(a^*) = \operatorname{conv} A$, where is A given in (6). By Carathéodory theorem, 0 is the convex hull of $d \leq n+1$ vectors of A, i.e., by Lemma 6 we have

$$G(x_1, \ldots, x_d)\lambda = 0 \text{ with } \lambda \neq 0\,,\ \lambda_i \geq 0 \text{ and } x_1, \ldots, x_d \in \operatorname{act}(a^*). \qquad (8)$$

Therefore from (5) we obtain $H(x_1, \ldots, x_d)u = 0$ with $u_i = \epsilon(a^*, x_i)\lambda_i$. Finally $u_i\,\epsilon(a^*, x_i) = \epsilon(a^*, x_i)^2\,\lambda_i \geq 0$ proving that the multivariate equioscillation condition holds.

3.3. Sufficiency of the multivariate equioscillation condition for a minimizer

Now, we prove that the multivariate equioscillation condition at a^* implies $0 \in \partial m(a^*)$. This condition at a^* means there exists $x_1, \ldots, x_d \in \operatorname{act}(a^*)$ and a kernel vector $u \neq 0$ of $H(x_1, \ldots, x_d)$ such that $\epsilon(a^*, x_i)\,u_i \geq 0$. Again from (5) we have $G(x_1, \ldots, x_d)\lambda = 0$ with $u_i = \epsilon(a^*, x_i)\,\lambda_i$, so that $\lambda \neq 0$ and by the multivariate equioscillation condition $0 \leq \epsilon(a^*, x_i)\,u_i = \epsilon(a^*, x_i)^2\,\lambda_i = \lambda_i$. Finally, Lemma 6 proves that 0 is in the convex hull of the columns of $G(x_1, \ldots, x_d)$ which are subgradients of m at a^*. Therefore $0 \in \partial m(a^*)$ and a^* is a minimizer.

3.4. Strong equioscillation implies strong uniqueness

Consider a minimizer $a^* \in \mathbb{R}^n$, so $0 \in \partial m(a^*)$. We restart at (8) but with the additional assumption that the strong multivariate equioscillation condition holds: $H(x_1, \dots, x_d)$ is full rank, hence so is $G(x_1, \dots, x_d)$ and $d \geq n+1$, and $u_i > 0$, so $\lambda_i > 0$, for all $i \in \{1, \dots, d\}$. Since the columns of $G := G(x_1, \dots, x_d)$ are subgradients at a^* we have a piecewise affine lower bound $m(a) \geq m(a^*) + l(a - a^*)$ with $l(a-a^*) = \max\big(G^T(a-a^*)\big)$. The maximum is homogenous for positive scalar hence we have the lower bound $l(a-a^*) = \|a-a^*\|_1\, l\big(\frac{(a-a^*)}{\|a-a^*\|_1}\big) \geq \alpha \|a-a^*\|_1$ with $\alpha = \min_{\|a-a^*\|_1 = 1} l(a-a^*)$. By Lemma 5 the piecewise linear function l has 0 as unique minimizer, therefore $\alpha > 0$. Finally, $m(a) \geq m(a^*) + \alpha\|a-a^*\|_1$, which proves the strong uniqueness since $\|a\|_{X,\phi} = \max_{x \in X} |a^T\phi(x)|$ is a norm by the weak Haar condition, and all norms are equivalent in $\mathbb{R}^n$.

4. Related work

Rice [4] gave a necessary and sufficient condition for the optimality of multivariate Chebyshev approximation in termes of isolation of active indices[3]: the active indices $\operatorname{act}(a^*)$ are isolable if there is a function $a^T\phi(x)$ that has the same sign as $e(a^*, x)$ on $\operatorname{act}(a^*)$, i.e.,

$$\forall x \in \operatorname{act}(a^*)\ ,\ \epsilon(a^*, x)\ ((a^*)^T\phi(x)) > 0. \tag{9}$$

Informally, since $\epsilon(a^*, x)\phi(x)$ are the subgradients at a^*, Rice's isolation condition means that the vector a^* has a positive scalar product with all subgradients, therefore $-a^*$ is a descent direction and a^* cannot be a minimizer. When $\operatorname{act}(a^*) = \{x_1, \dots, x_d\}$, Equation (9) can be written $G(x_1 \dots, x_d)^T a^* > 0$, which is related to the multivariate equioscillation condition by Gordan's alternative theorem: the latter proves that there is no a^* such that $G(x_1 \dots, x_d)^T a^* > 0$ is and only if there exists $\lambda \neq 0$, $\lambda_i \geq 0$, such that $G(x_1 \dots, x_d)\lambda = 0$. That is, there is no descent direction is (informally) equivalent to the multivariate equioscillation condition.

The non-uniqueness necessary condition given in [4, Theorem 2 page 448] is related to the failure of the strong equioscillation: we can assume that $u_i \neq 0$ in the definition of multivariate equioscillation, otherwise the active indices x_i with $u_i = 0$ can be removed. If a^* is an optimal solution that is not unique then $H(x_1, \dots, x_K)$ is not full rank, therefore neither is $H(x_1, \dots, x_K)^T$ and there exists $0 \neq v \in \mathbb{R}^n$ such that $H(x_1, \dots, x_K)^T v = 0$. This later means $v^T\phi(x_i) = 0$ therefore the active indices x_i belong to $\{x \in X : v^T\phi(x) = 0\}$, called an isolating curve by Rice.

The optimality condition for convex optimization is used and reformulated in [6], giving rise to a condition for the optimality of multivariate Chebyshev approximation in terms of the nonempty intersection of two convex hulls of

[3] A similar argument is used in[5, Equation (7.6) page 74].

vectors related to gradients. The authors explain in [7] that this condition is a reformulation of Rice's isolation condition.

5. Some multivariate problem instances

The first example shows an instance with a strongly unique solution.

Example 7. *Let $X = [0,1]^n \subseteq \mathbb{R}^n$, $f(x) = x_1^2 + \cdots + x_n^2$ and $\phi(x) = (1, x_1, \ldots, x_n)$, so that we approximate f by affine functions $a^T\phi(x)$. The weak Haar condition is satisfied since $H\big(0, e_1, \ldots, e_n\big)$, where e_i is the i^{th} basis vector, is nonsingular, so there is an optimal solution. Let us test the approximation $a^* = (-\frac{n}{8}, 1, \ldots, 1)$ using the multivariate equioscillation theorem. The error absolute value $|e(a^*, x)|$ has $2^n + 1$ maximizers[4], which are all corners of the box X and its midpoint, whose error are $-\frac{n}{8}$ for the corners and $\frac{n}{8}$ for the midpoint. The Haar matrix at these active indices is as follows: its first 2^n columns correspond to corners, while the last column correspond to the midpoint. For example for $n = 3$ we have*

$$H\big((0,0,0), (0,0,1), (0,1,0), \ldots, (1,1,1), (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2})\big) \tag{10}$$

equals to

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \frac{1}{2} \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & \frac{1}{2} \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & \frac{1}{2} \end{pmatrix}. \tag{11}$$

We can observe and generalize to arbitrary n that each line contains exactly $\frac{2^n}{2}$ zeros and as many ones, and $\frac{1}{2}$ at the end. As a consequence, the vector $(-1, \ldots, -1, 2^n)$ is a kernel vector, which satisfies the multivariate equioscillation condition. Therefore the multivariate equioscillation theorem proves that $-\frac{n}{8} + x_1 + \cdots + x_2$ is one optimal solution of this Chebyshev approximation problem. Furthermore the Haar matrix is full rank and the kernel vector has only nonzero components, hence the strong multivariate equioscillation condition holds and this solution is strongly unique.

The second example shows an example where the strong multivariate equioscillation condition does not hold, but where further numerical investigations let us conjecture there is a unique solution. This is related to the singular Chebyshev approximation problems that were investigated in [8, 9, 10, 11].

Example 8. *Let $X = [0,1] \times [0,1]$, $f(x) = x_1^2 + 2x_2^2 - \frac{1}{2}x_1 x_2$ and $\phi(x) = (1, x_1, x_2)$, so that again we approximate f by affine functions $a^T\phi(x) = a_1 + a_2x_1 + a_3x_2$. Again the weak Haar condition holds and implies the existence of a*

[4] The error function $e(a^*, x) = (a^*)^T\phi(x) - f(x)$ is a strictly concave quadratic function, whose global maximizer is the midpoint of the box. Since it is strictly concave, all other local maximizers are strictly lower than the global maximizer, and strictly larger than the global minimizers. These latter can only be at the corners of the box.

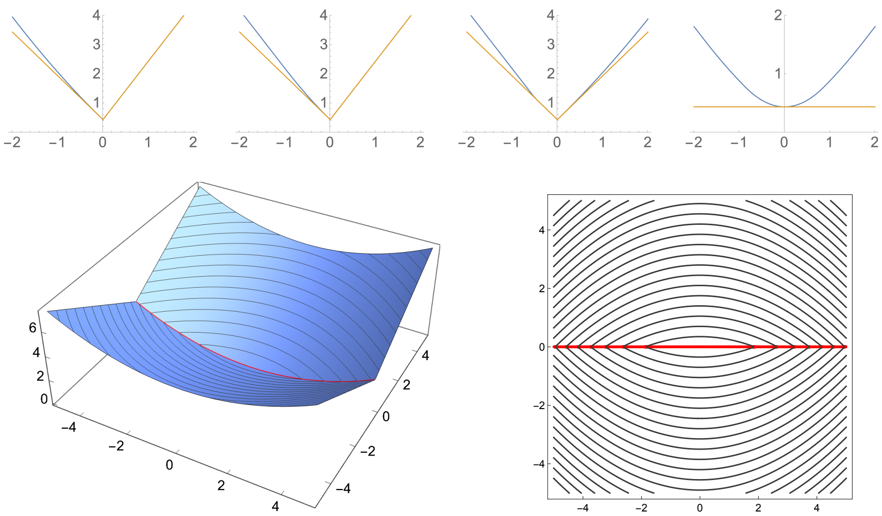

Figure 2: The four upper graphics show four directional restrictions of $m(a)$ in blue and its piecewise linear lower bound $l(a)$ in orange, showing a unique non-strongly unique minimizer. The two lower graphics show a similar situation with a finite pointwise maximum.

minimizer. Let us test the approximation $a^* = (-\frac{3}{16}, \frac{3}{4}, \frac{7}{4})$ *using the multivariate equioscillation theorem. The error absolute value* $|e(a^*, x)|$ *has three maximizers* $(0, 1)$, $(1, 0)$ *and* $(\frac{1}{2}, \frac{1}{2})$, *whose error values are respectively* $(-\frac{7}{16}, -\frac{7}{16}, \frac{7}{16})$. *The Haar matrix for these maximizers is*

$$H\big((0,1),(1,0),(\tfrac{1}{2},\tfrac{1}{2})\big) = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} \end{pmatrix}. \tag{12}$$

We can extract the kernel vector $u = (-1, -1, 2)$, *which satisfies the multivariate equioscillation condition. Therefore* $-\frac{3}{16} + \frac{3}{4}x_1 + \frac{7}{4}x_2$ *is one optimal solution of the Chebyshev approximation problem. The Haar matrix is not full rank, so the multivariate equioscillation theorem does not allow proving the uniqueness. Let us investigate in more details this situation. The subgradient matrix is*

$$G\big((0,1),(1,0),(\tfrac{1}{2},\tfrac{1}{2})\big) = \begin{pmatrix} -1 & -1 & 1 \\ 0 & -1 & \frac{1}{2} \\ -1 & 0 & \frac{1}{2} \end{pmatrix}, \tag{13}$$

whose kernel vector $\lambda = (1, 1, 2)$ *is positive as expected. The three columns of* G *are three subgradient, which give rise to a piecewise linear lower bound* $l(a) = \max\big(G^T(a - a^*)\big)$ *as defined in Subsection 3.4. But in the present case,* $l(a)$ *does not allow to prove uniqueness because it is constant in the direction* $v = (-1, 1, 1)$, *which is a kernel vector of* G^T. *The upper row of Figure 2 shows four directional restrictions of both* $m(a)$ *and its lower bound* $l(a)$. *The first three directions are columns of* G, *hence steepest ascent directions of the lower bound,*

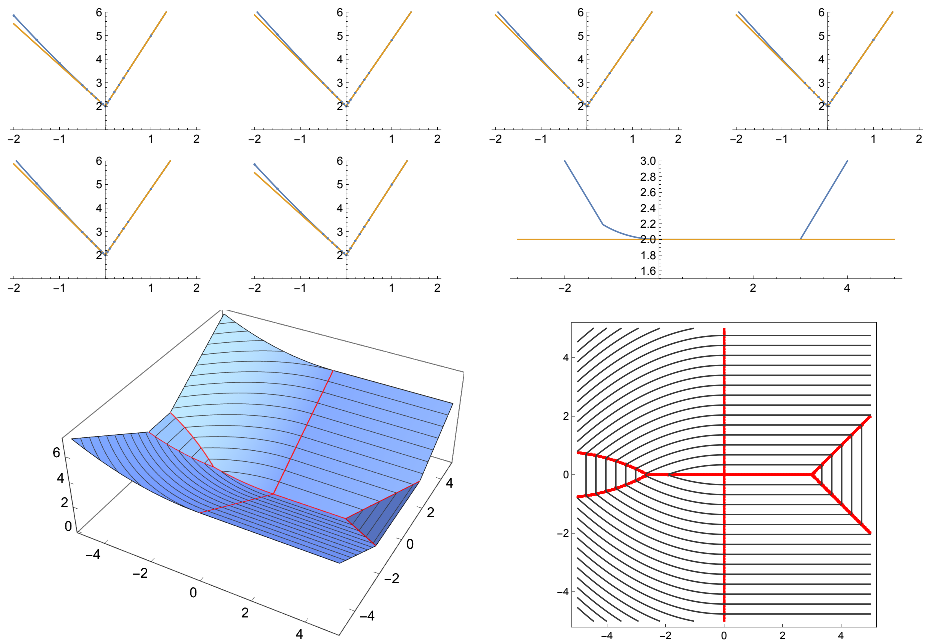

Figure 3: The seven upper graphics show seven directional restrictions of $m(a)$ in blue and its piecewise linear lower bound $l(a)$ in orange, showing infinitely many minimizers. The two lower graphics show a similar situation with a finite pointwise maximum.

and we observe expected strong univariate minimizers. The fourth direction is the kernel vector v of G^T, so as expected $l(a)$ is flat in this direction, and we can see that a second order curvature makes the minimizer unique, but not strongly unique. The lower two graphics of Figure 2 displays a finite pointwise maximum showing the same behavior: the unique minimizer is strongly unique in one direction, and not strongly unique in the other direction.

The following example is taken from [7] and shows a case with infinitely many optimal solutions to the Chebyshev approximation problem.

Example 9. *Let $X = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$, $f(x) = x_1^6 + x_2^6 + 3x_1^4x_2^2 + 3x_1^2x_2^4 + 6x_1x_2^2 - 2x_1^3$ and $\phi(x) = (1, x_1, x_2, x_1^2, x_2^2, x_1x_2)$. As shown in [7], for $a^* = (1, 0, 0, 0, 0, 0)$ the error extrema are $(-1, 0)$, $(\frac{1}{2}, -\frac{\sqrt{3}}{2})$, $\frac{1}{2}, \frac{\sqrt{3}}{2}$, $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$, $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$ and $(1, 0)$, with errors $(2, 2, 2, -2, -2, -2)$. The Haar matrix is*

$$H((-1,0),\ldots,(1,0)) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 1 \\ 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & 0 \\ 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 1 \\ 0 & \frac{3}{4} & \frac{3}{4} & \frac{3}{4} & \frac{3}{4} & 0 \\ 0 & -\frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{4} & -\frac{\sqrt{3}}{4} & 0 \end{pmatrix}, \tag{14}$$

with kernel vector $(1,1,1,-1,-1,-1)$*, which satisfies the multivariate equioscillation condition. As previously, the Haar matrix is not full rank so the strong multivariate equioscillation condition does not hold. In Figure 3, we again display the directional restrictions of the piecewise linear lower bound in the direction of each column of* G*, where we see expected strong univariate minimizers, and in the direction of the kernel vector* $v = (-1,0,0,1,1,0)$ *of* G^T*, which this time shows a continuum of minimizers. The kernel vector* v *is the direction where* $m(a)$ *is constant, therefore, as one expects,* $(a^* + tv)^T\phi(x) = 1 - t + tx_1^2 + tx_2^2$ *for* $t \in [1,3]$ *corresponds to the description of the minimizers given in [7]. The lower two graphics of Figure 3 displays a finite pointwise maximum showing the same behavior.*

The Haar matrix usually contains a row of 1, which corresponds to a constant basis function. In this case, no kernel vector can have components with constant sign, therefore the optimal solution must have some positive and negative error extrema, which corresponds to our intuition. However, the following simple example shows that when the constant basis function is not included some optimal solution may have error extrema with constant sign.

Example 10. *Let* $X = [-1,1] \times [-1,1]$*,* $f(x) = x_1^2 + x_2^2$ *and* $\phi(x) = (x_1, x_2)$*, so that we approximate* f *by linear functions* $a^T\phi(x) = a_1x_1 + a_2x_2$*. The weak Haar condition is satisfied since* $H((1,0),(0,1))$ *is nonsingular. Let us test the approximation* $a^* = (0,0)$ *using the multivariate equioscillation theorem. The error has four maximizers at the corners of the box with all the positive error* 2*, and no other extrema. The Haar matrix is*

$$H\big((-1,-1),(-1,1),(1,-1),(1,1)\big) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{pmatrix} \tag{15}$$

with kernel vector $(1,1,1,1)$*. The multivariate equioscillation condition is satisfied, therefore* 0 *is one best linear approximation of* $f(x)$ *in* X*. The Haar matrix is full rank and the kernel vector has no negative components, therefore the strong equioscillation condition holds and the solution is strongly unique.*

Finally, the following example shows an untypical case where there is only one active index and the multivariate equioscillation theorem succeeds. It is presented to show that $K = 1$ actually makes sense in the definition of the multivariate equioscillation condition.

Example 11. *Let* $X = [-1,1]\times[-1,1]$*,* $f(x) = 2 - x_1^2 - x_2^2$ *and* $\phi(x) = (x_1, x_2)$*, so that we approximate again* f *by linear functions* $a^T\phi(x) = a_1x_1 + a_2x_2$*. Let us test the approximation* $a^* = (0,0)$ *using the multivariate equioscillation theorem. The error is* 0 *on the four corners of the box, and has a unique maximizer at* $(0,0)$ *with error* 2*. The Haar matrix is*

$$H\big((0,0)\big) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{16}$$

with kernel vector $\mathbb{1}$*. The multivariate equioscillation condition is satisfied, therefore* 0 *is one best linear approximation of* $f(x)$ *in* X*. In fact, this seemingly strange situation is quite normal from the point of view of convex analysis: the column of the Haar matrix is a subgradient, and having a null subgradient entails being a minimizer.*

Appendix A. Proof of Lemma 1

It is well known if $\phi : X \to \mathbb{R}^n$ is continuous and satisfies the Haar condition on X then the determinant of all square Haar matrices $H(x_1, \ldots, x_n)$, with $x_1 < \cdots < x_n \in X$, have the same sign, see [1, proof of lemma page 74]. Now, the Haar matrix $H(x_1, \ldots, x_{n+1})$ is full rank (because Haar condition entails $\det H(x_1, \ldots, x_n) \neq 0$) so its kernel is dimension 1. Let us consider one kernel element $0 \neq u \in \mathbb{R}^{n+1}$, i.e., $\sum_{i=1}^{n+1} u_i \, \phi(x_i) = 0$. For an arbitrary $i \in \{1, \ldots, n+1\}$, we have $u_i \neq 0$ otherwise $\sum_{j \neq i} u_i \, \phi(x_j) = 0$ and $H_i := H(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1})$ would be singular contradicting Haar contidition. For $i \in \{1, \ldots, n\}$ we have $\phi(x_i) = -\sum_{j \neq i} \frac{u_j}{u_i} \, \phi(x_j)$ and replacing the column $\phi(x_i)$ in H_{i+1} by this sum shows that $\det H_{i+1}$ is equal to

$$\left| \, \phi(x_1) \ \cdots \ \phi(x_{i-1}) \ \big(-\sum_{j \neq i} \tfrac{u_j}{u_i} \, \phi(x_j)\big) \ \phi(x_{i+2}) \ \cdots \ \phi(x_{n+1}) \, \right| = -\tfrac{u_{i+1}}{u_i} \det H_i, \tag{A.1}$$

the last equality obtained using the multilinearity and alternativity of the determinant. The determinants $\det H_i$ and $\det H_{i+1}$ having the same sign, we conclude that u_i and u_{i+1} have opposite signs.

Appendix B. La Vallée Poussin

Theorem 12 (Multivariate La Vallée Poussin)**.** *Let $\phi : X \to \mathbb{R}^n$ be continuous and $p(x) = a^T \phi(x)$ be the approximation corresponding to $a \in \mathbb{R}^n$. Consider $x_1, \dots, x_K \in X$ such that the Haar matrix $H(x_1, \dots, x_K)$ has a kernel vector $0 \neq u \in \mathbb{R}^K$ satisfying the oscillating condition $e(a, x_i)\, u_i \geq 0$ for all $i \in \{1, \dots, K\}$. Then*

$$\min_{1 \leq k \leq K} |e(a, x_k)| \;\leq\; \min_{a \in \mathbb{R}^n} \max_{x \in X} |e(a, x)|. \tag{B.1}$$

Example 13. *Let us consider again Example 8: let $X = [0,1] \times [0,1]$, $f(x) = x_1^2 + 2x_2^2 - \frac{1}{2} x_1\, x_2$ and $\phi(x) = (1, x_1, x_2)$, whose optimal approximation is $a^* = (-\frac{3}{16}, \frac{3}{4}, \frac{7}{4}) = (-0.1875, 0.75, 1.75)$ with maximal error $\frac{7}{16} = 0.4375$. We now compute an approximate approximation by using* 100 *random samples in the domain and obtain $\hat{a} = (-0.21608, 0.68264, 1.69007)$. The maximal error is* 0.53344*, obtained at $x = (1, 0)$, which is larger than the optimal error as expected. The error of the finite linear relaxation is* 0.342148*, which is a lower bound of the optimal error as expected. We now apply the multivariate La Vallée Poussin theorem. We start by finding good samples points by locally maximizing the error starting from the actives samples at the optimal solution of the discretized finite LP, and find $\{(0,1), (1,0), (0.4613671, 0.48018839)\}$. Since samples must be aligned so that the Haar matrix has a one dimensional kernel, we apply the multivariate La Vallée Poussin theorem to the samples $\{(0,1), (1,0), (\frac{1}{2}, \frac{1}{2})\}$. The Haar matrix is*

$$H\big((0,1), (1,0), (\tfrac{1}{2}, \tfrac{1}{2})\big) = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} \end{pmatrix} \tag{B.2}$$

whose kernel vector is $u = (-1, -1, 2)$. The errors corresponding to the samples are $(-0.52601, -0.53344, 0.345275)$, hence the oscillation condition holds true. As a consequence, by the multivariate La Vallée Poussin theorem the smallest error in absolute value 0.345275 *is a lower bound of the error, which improves the lower bound* 0.342148 *obtained by the finite LP relaxation.*

References

[1] E. W. Cheney, Introduction to Approximation Theory, AMS Chelsea Publishing, 1982.

[2] J. C. Mairhuber, On Haar's theorem concerning Chebychev approximation problems having unique solutions, Proc. Amer. Math. Soc. 7 (1956) 609–615.

[3] Y. Nesterov, Lectures on Convex Optimization, Springer Cham, 2018.

[4] J. R. Rice, Tchebycheff approximation in several variables, Transaction of the AMS (1963).

[5] M. J. D. Powell, Approximation Theory and Methods, Cambridge University Press, 1981.

[6] N. Sukhorukova, J. Ugon, D. Yost, Chebyshev Multivariate Polynomial Approximation: Alternance Interpretation, Springer International Publishing, 2018, pp. 177–182.

[7] V. Roshchina, N. Sukhorukova, J. Ugon, Uniqueness of solutions in multivariate chebyshev approximation problems, Optimization Letters (2023).

[8] M. Osborne, G. Watson, A note on singular minimax approximation problems, Journal of Mathematical Analysis and Applications 25 (1969) 692–700.

[9] G. A. Watson, A multiple exchange algorithm for multivariate chebyshev approximation, SIAM Journal on Numerical Analysis 12 (1975) 46–52.

[10] R. Reemtsen, Modifications of the first remez algorithm, SIAM Journal on Numerical Analysis 27 (1990) 507–518.

[11] R. Reemtsen, Discretization methods for the solution of semi-infinite programming problems, Journal of Optimization Theory and Applications 71 (1991) 85–103.