Computing a Quantity of Interest from Observational Data

Ronald DeVore, Simon Foucart, Guergana Petrova, and Przemyslaw Wojtaszczyk *

June 13, 2017

Abstract

Scientific problems often feature observational data received in the form $w_1 = l_1(f), \ldots, w_m = l_m(f)$ of known linear functionals applied to an unknown function f from some Banach space $\mathcal{X}$, and it is required to either approximate f (the full approximation problem) or to estimate a quantity of interest $Q(f)$. In typical examples, the quantities of interest can be the maximum/minimum of f or some averaged quantity such as the integral of f, while the observational data consists of point evaluations. To obtain meaningful results about such problems, it is necessary to possess additional information about f, usually as an assumption that f belongs to a certain model class $\mathcal{K}$ contained in $\mathcal{X}$. This is precisely the framework of optimal recovery, which produced substantial investigations when the model class is a ball in a smoothness space, e.g. when it is a unit ball in Lipschitz, Sobolev, or Besov spaces. This paper is concerned with other model classes described by approximation processes, as studied in [14]. Its main contributions are: (i) designing implementable optimal or near-optimal algorithms for the estimation of quantities of interest, (ii) constructing linear optimal or near-optimal algorithms for the full approximation of an unknown function using its point evaluations.

While the existence of linear optimal algorithms for the approximation of linear functionals $Q(f)$ is a classical result established by Smolyak, a numerically friendly procedure that performs this approximation is not generally available. In this paper, we show that in classical recovery settings such linear optimal algorithms can be produced by constrained minimization methods. We illustrate these techniques on several examples involving the computation of integrals using point evaluation data.

In addition, we show that linearization of optimal algorithms can be achieved for the full approximation problem in the important situation where the l_j are point evaluations and $\mathcal{X}$ is a space of continuous functions equipped with the uniform norm. It is also revealed how quasi-interpolation theory enables the construction of linear near-optimal algorithms for the recovery of the underlying function.

AMS Subject Classification: 41A65, 46N40, 65Y20, 41A46, 41A05, 46A22, 90C05, 65D32, 46E22.

Key Words: optimal recovery, data fitting, Chebyshev centers, ℓ_1-minimization, quadrature formulas, reproducing kernel Hilbert spaces.

*This research was supported by the ONR Contracts N00014-15-1-2181, N00014-16-1-2706 the NSF Grant DMS 1521067, DARPA through Oak Ridge National Laboratory; by the NSF Grant DMS 1622134; and by National Science Centre, Poland grant UMO-2016/21/B/ST1/00241.

1 Introduction

In many scientific settings, data is given about an unknown function f, and the typical tasks consist in using this data to either recover f (full approximation of f) or to answer some questions about it (recovery of a quantity of interest for f), e.g., to evaluate its integral over a domain, its maximum on the domain, etc. Such problems fall into the realm of optimal recovery and form a major topic in information-based complexity, which provides a general theory describing optimal and near-optimal algorithms. This theory requires the specification of (i) the form of the data; (ii) a measure for the quality of performance; and (iii) what information about f is available, in addition to the data.

Most existing results on optimal recovery, as is the case here as well, treat the setting where the data is given by linear functionals and where the quality of performance is measured by a norm $\|\cdot\|_{\mathcal{X}}$ on a Banach space $\mathcal{X}$. The departing point in the present paper lies in (iii). It is common in optimal recovery to assume as an additional information about f that it lies in a compact subset $\mathcal{K}$ of $\mathcal{X}$, typically described by a smoothness condition such as membership to f in the unit ball of Lipschitz, Sobolev, or Besov spaces. We assume instead that the additional information about f quantifies how well it can be approximated by a fixed linear or nonlinear space. In this case, we show that one can explicitly construct optimal or near-optimal algorithms for both the full approximation problem and the quantity of interest problem and numerically implement them in classical scenarios. Our results build on those of [14], but emphasize more the numerical feasibility and computational performance of the recovery algorithms.

To give an overview of this paper, we consider first the task of approximating the function f itself, i.e., the full approximation problem, where f lives in an (infinite-dimensional) normed space $\mathcal{X}$. Our setting for (i) is that the data come in the form of a measurement map $M : \mathcal{X} \to \mathbb{R}^m$, where

$$M(f) := (l_1(f), \ldots, l_m(f)) = (w_1, \ldots, w_m) = w \in \mathbb{R}^m, \tag{1.1}$$

and where the l_j's are linear functionals on $\mathcal{X}$. Without loss of generality, we can assume that these measurement functionals are linearly independent. By an algorithm for approximating f from the data, we simply mean a (typically nonlinear) map $A : \mathbb{R}^m \to \mathcal{X}$ taking the data $w = M(f)$ and creating a substitute $A(w)$ for f. For (ii), we fix a norm $\|\cdot\|_{\mathcal{X}}$ on $\mathcal{X}$ and we characterize the performance of the algorithm by the error

$$E(f, M, A)_{\mathcal{X}} := \|f - A(M(f))\|_{\mathcal{X}}. \tag{1.2}$$

Concerning (iii), we note that without further information about f (other than the fact that $f \in \mathcal{X}$), the zero algorithm A_0, that is $A_0(w) = 0$ for every $w \in \mathbb{R}^m$, will be at least as good as any algorithm A, in the sense that there is at least one $g \in \mathcal{X} \setminus \{0\}$ such that $E(g, M, A)_{\mathcal{X}} \geq \|g\|_{\mathcal{X}} = E(g, M, A_0)_{\mathcal{X}}$. Indeed, for any nonzero h in the null space of M, we have $E(h, M, A)_{\mathcal{X}} + E(-h, M, A)_{\mathcal{X}} \geq 2\|h\|_{\mathcal{X}}$. Thus, to obtain meaningful results, we need additional information about f, as mentioned in (iii). This takes the form of a postulate that the functions f belong to a certain model class $\mathcal{K} \subset \mathcal{X}$. It is often assumed (e.g. in information-based complexity) that $\mathcal{K}$ is a compact set, but we will mostly deal with noncompact sets in this paper. Given this additional information, we consider the global performance of the proposed algorithm A,

$$E(\mathcal{K}, M, A)_{\mathcal{X}} := \sup_{f \in \mathcal{K}} E(f, M, A)_{\mathcal{X}}, \tag{1.3}$$

which is the maximal error over the model class $\mathcal{K}$, and the pointwise performance as defined by

$$E(\mathcal{K}_w, M, A)_{\mathcal{X}} := \sup_{f \in \mathcal{K}_w} E(f, M, A)_{\mathcal{X}}, \tag{1.4}$$

where $\mathcal{K}_w := \{f \in \mathcal{K} : \ M(f) = w\}$ is the set of all functions f in the given model class whose measurement vector is w.

We are interested in optimal or near-optimal recovery algorithms for approximation from data, namely, in the algorithms that attain or come close to attaining

$$E^*(\mathcal{K}, M)_{\mathcal{X}} := \inf_A E(\mathcal{K}, M, A)_{\mathcal{X}}, \tag{1.5}$$

where the infimum is taken over all maps $A : \mathbb{R}^m \to \mathcal{X}$. This is the central problem of optimal recovery. There is a simple theoretical description of best algorithms for recovery (see, for example, [7, 20, 21, 26]) using the so-called Chebyshev ball. Namely, given the measurement vector $w \in \mathbb{R}^m$, we let $B(g_w, R_w)$ be a smallest ball in $\mathcal{X}$ that contains $\mathcal{K}_w$. Then an optimal algorithm $A^* : \mathbb{R}^m \to \mathcal{X}$ is the mapping which sends w to g_w and whose performance is given by

$$E(\mathcal{K}_w, M, A^*)_{\mathcal{X}} = R_w. \tag{1.6}$$

This is the pointwise optimal error. The algorithm A^* also yields the global optimal error, namely

$$E^*(\mathcal{K}, M)_{\mathcal{X}} = E(\mathcal{K}, M, A^*)_{\mathcal{X}} = \sup_{w \in \mathbb{R}^m} R_w. \tag{1.7}$$

While the above provides a nice theoretical description of optimal algorithms, it is not readily applicable since it may be a formidable task to find the Chebyshev ball of $\mathcal{K}_w$. Typical results in optimal recovery consider special sets $\mathcal{K}$ (often unit balls of smoothness spaces) with the error metric chosen as an L_p norm or the uniform norm. The present paper focuses on different model classes, involving approximability criteria by a set V of functions used as approximants. Usually, V is a finite-dimensional linear space of polynomials, splines, or wavelets, or a nonlinear set encountered in sparse approximation. More precisely, we introduce the following model classes.

Approximation Set: *Given an $\varepsilon > 0$ and a set $V \subset \mathcal{X}$, we define the approximation set as*

$$\mathcal{K}(\varepsilon, V)_{\mathcal{X}} := \{f \in \mathcal{X} \ : \ \operatorname{dist}(f, V)_{\mathcal{X}} \le \varepsilon\}. \tag{1.8}$$

A finer model class is obtained by taking the intersection of such approximation sets.

Approximation Class: *Given a nonincreasing sequence of numbers $\varepsilon_0 \ge \cdots \ge \varepsilon_n \ge \varepsilon_{n+1} \ge \cdots \ge 0$ and a nested sequence of sets $V_0 \subset \cdots \subset V_n \subset V_{n+1} \subset \cdots \subset \mathcal{X}$, we define the approximation class as*

$$\mathcal{K}((\varepsilon_n), (V_n))_{\mathcal{X}} := \bigcap_{n \ge 0} \mathcal{K}(\varepsilon_n, V_n)_{\mathcal{X}}. \tag{1.9}$$

Our motivation for studying these two cases of model classes is threefold. Firstly, classical smoothness spaces can be described as approximation classes by using the existing theory of approximation, precisely, direct and inverse theorems. For instance, standard theorems about approximation by trigonometric polynomial, spline, and wavelet spaces V_n characterize approximation classes with $\varepsilon_n \asymp n^{-r}$ as Besov spaces (see [13]). Similar characterizations hold for nonlinear approximation as well (see [12]). Secondly, in some cases of interest to us, such as parametric PDEs where $\mathcal{K}$ is the solution manifold to the PDE as the parameter varies, our information about $\mathcal{K}$

tells us how well $\mathcal{K}$ can be approximated by certain finite-dimensional spaces V. For more on this, we refer the reader to the discussion in [5]. Lastly, any numerical algorithm for optimal recovery will be based on some form of approximation and its performance will be determined by the approximation sets or approximation classes for this form of approximation. A benefit of studying model classes consisting of approximation sets is that we can very precisely characterize optimal performance and describe optimal algorithms (see [5, 14]).

The majority of this paper studies the performance of algorithms on approximation sets. In the previous works [5, 14], we have constructed near-optimal algorithms for the approximation sets $\mathcal{K} = \mathcal{K}(\varepsilon, V)_{\mathcal{X}}$ for general Banach spaces $\mathcal{X}$ when V is a linear space of dimension $n \leq m$, and we have shown that the performance of such algorithms can be determined a priori. To recall the main points, we introduce, for any two sets $A, B \subset \mathcal{X}$, the quantity

$$\mu(A, B)_{\mathcal{X}} := \sup_{f \in A, g \in B - B} \frac{\|f\|_{\mathcal{X}}}{\|f - g\|_{\mathcal{X}}}, \tag{1.10}$$

which can be interpreted as the reciprocal of the angle between A and $B - B$. Notice that $\mu(A, B)_{\mathcal{X}} \geq 1$. The quantities $\mu(A, B)_{\mathcal{X}}$ and $\mu(B, A)_{\mathcal{X}}$ are generally different, but in case A and B are linear spaces, we have

$$\mu(A, B)_{\mathcal{X}} \leq 1 + \mu(B, A)_{\mathcal{X}}. \tag{1.11}$$

Now, let $\mathcal{N} := \{f \in \mathcal{X} : \ M(f) = 0\}$ be the null space of the measurement map M. Then, the performance estimates in [14] show that

$$E(\mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M, A)_{\mathcal{X}} \leq 2\, \mu(\mathcal{N}, V)_{\mathcal{X}}\, \varepsilon \tag{1.12}$$

holds for an appropriately chosen algorithm A, and it is proved that such an estimate cannot be improved in the sense that

$$E^*(\mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M)_{\mathcal{X}} \geq \ \mu(\mathcal{N}, V)_{\mathcal{X}}\, \varepsilon. \tag{1.13}$$

One can then view $\mu(\mathcal{N}, V)_{\mathcal{X}}$ as assessing the quality of the data when coupled with the assumption that the class of functions is well approximated by V.

Another interesting point proved in [14] (see Theorem 5.1 there) is the construction of near-optimal algorithms that do not require knowledge of the value of ε defining the model class $\mathcal{K} = \mathcal{K}(\varepsilon, V)_{\mathcal{X}}$. Such algorithms are universal in the sense that they can be applied to the general setting of data fitting. Thus, since an arbitrary $f \in \mathcal{X}$ always belongs to the approximation set $\mathcal{K}(\varepsilon, V)_{\mathcal{X}}$ with $\varepsilon := \operatorname{dist}(f, V)_{\mathcal{X}}$, the approximation $A(M(f))$ to f uses only the data $M(f)$ and the knowledge of the space V to guarantee the performance bound

$$\|f - A(M(f))\|_{\mathcal{X}} \leq C\, \mu(\mathcal{N}, V)_{\mathcal{X}}\, \operatorname{dist}(f, V)_{\mathcal{X}}, \qquad f \in \mathcal{X}. \tag{1.14}$$

The quantity $\mu(\mathcal{N}, V)_{\mathcal{X}}$ can be explicitly evaluated in several concrete situations discussed in [14] and [6]. Note that there are other results in this direction (see e.g. [25, 26, 2, 3]) addressing the full approximation problem in specific settings, such as $\mathcal{X}$ being the space of continuous functions, V a space of polynomials, and data consisting of values at equispaced points, or $\mathcal{X}$ being an L_2 space, V a space of polynomials, and data consisting of Fourier coefficients.

The present paper discusses two main topics: the full recovery of unknown functions f and the recovery of quantities of interest $Q(f)$, both using only the measurement $M(f) \in \mathbb{R}^m$. We

are particularly concerned with the construction of numerically friendly optimal or near-optimal algorithms. Note that, for the task of computing a quantity of interest $Q(f)$ instead of fully approximating f, one expects a smaller error than for the full approximation problem. We will clarify this gain.

We begin in §2 by discussing some results on optimal performance that are valid for the recovery of f as well as for the estimation of quantities of interest $Q(f)$. These simple and somewhat folklore extensions of the full approximation theory show that, for approximation sets and for the estimation of quantities of interest, the optimal performance can be explicitly determined.

In §3, we consider quantities of interest q that are linear functionals. It is a well-known fact, first discovered by Smolyak (see e.g. [4]), that the problem of approximating q admits optimal algorithms that are linear maps from $\mathbb{R}^m$ to $\mathbb{R}$. To our knowledge, the proofs of this fact are functional analytic and it is not apparent how to convert them into numerical algorithms. In §3.1, we give another proof of the existence of a linear optimal algorithm for the approximation of the linear functional q (see Theorem 3.1), and later in the paper we use the proof ingredients to actually implement the linear algorithm. In §3.2, we discuss the case of compressed sensing and the gain in approximating directly $q(\mathbf{x})$, rather than first finding an approximant $\tilde{\mathbf{x}}$ to the vector $\mathbf{x}$ and then calculating $q(\tilde{\mathbf{x}})$. In §3.3, we apply our theory to the concrete example of finding optimal quadrature formulas based on values at equidistant points in order to estimate the integral (our quantity of interest) of functions in $C[-1,1]$.

Next, the paper examines the full approximation of functions $f \in C(D)$ based on point values. After recalling some known results in §4.1, we establish in §4.2 the existence of a linear optimal algorithm for the recovery of f (see Theorem 4.2). Despite being a theoretically interesting fact, the theorem does not directly provide a numerical recipe for constructing such a linear optimal algorithm. As a remedy, we exploit in §4.3 quasi-interpolation theory to propose computable linear near-optimal maps for the recovery of f. Indeed, the fact that the space V admits explicit quasi-interpolants in many classical settings, such as polynomial and spline approximation, enables us to computationally construct linear near-optimal algorithms for the full approximation problem in $C(D)$. Finally, in §4.4, we apply our theory and discuss the full recovery of $f \in C[-1,1]$ based on values at equidistant points.

In closing, §5 puts our algorithms to the test of several examples involving construction of optimal quadrature formulas and computation of function approximants in $C[-1,1]$. We also provide details about our practical implementations.

2 Optimal performance for full recovery and quantities of interest

In this section, algorithmic considerations are left aside. The goal is merely to highlight what is theoretically achievable. We present easy extensions of the results from [5, 14] in two directions, when V may not be a linear space and when a full approximation of f may not be necessary but only a quantity of interest for f needs to be evaluated. In order to handle the cases of full approximation and estimation of functionals at once, we let Q denote a mapping from $\mathcal{X}$ into another normed space $\mathcal{Y}$. So in the case of full approximation, we simply take $\mathcal{Y} = \mathcal{X}$ and $Q = I$ (the identity map) and in the case of estimating a functional l applied to f (e.g. the integral of f over D), we take $\mathcal{Y} = \mathbb{R}$ and $Q = l$.

As before, to obtain meaningful results, we require the additional information that f belongs to some model class $\mathcal{K}$. The global performance of an algorithm $A : \mathbb{R}^m \to \mathcal{Y}$ that approximates

$Q(f)$ from the data $M(f)$ is characterized by the worst case error over $\mathcal{K}$, i.e.,

$$E(Q, \mathcal{K}, M, A) := \sup_{f \in \mathcal{K}} \|Q(f) - A(M(f))\|_{\mathcal{Y}}. \tag{2.1}$$

The global optimal performance is then given by

$$E^*(Q, \mathcal{K}, M) := \inf_A E(Q, \mathcal{K}, M, A), \tag{2.2}$$

where the infimum is taken over all maps $A : \mathbb{R}^m \to \mathcal{Y}$. We also consider the pointwise optimal performance given, for any $w \in \mathbb{R}^m$, by

$$E^*(Q, \mathcal{K}_w, M) := \inf_A \sup_{f \in \mathcal{K}_w} \|Q(f) - A(w)\|_{\mathcal{Y}}. \tag{2.3}$$

For any subset V of $\mathcal{X}$, we introduce the set

$$\mathcal{V} := V - V = \{f - g : f, g \in V\}, \tag{2.4}$$

and define the quantity

$$\mu(\mathcal{N}, V, Q)_{\mathcal{X}} := \sup_{\eta \in \mathcal{N}} \frac{\|Q(\eta)\|_{\mathcal{Y}}}{\operatorname{dist}(\eta, \mathcal{V})_{\mathcal{X}}}, \tag{2.5}$$

where $\mathcal{N}$ denotes again the null space of the measurement map M. Note that (2.5) agrees with (1.10) when $Q = I$. The crucial role played by $\mu(\mathcal{N}, V, Q)_{\mathcal{X}}$ is revealed in the following theorem, which follows from well-known results in optimal recovery [26, 20] adapted to our case of approximation sets.

Theorem 2.1. *Let $Q : \mathcal{X} \to \mathcal{Y}$ be any map between normed spaces $\mathcal{X}$ and $\mathcal{Y}$.*
(i) *For any set $\mathcal{K} \subset \mathcal{X}$ and any $w \in \mathbb{R}^m$,*

$$E^*(Q, \mathcal{K}_w, M) = \operatorname{rad}(Q(\mathcal{K}_w))_{\mathcal{Y}}, \tag{2.6}$$

where $\operatorname{rad}(S)_{\mathcal{Y}} := \inf\{r \ : \ S \subset B(y, r)$ for some $y \in \mathcal{Y}\}$ denotes the Chebyshev radius of a set $S \subset \mathcal{Y}$. If the infimum defining $\operatorname{rad}(Q(\mathcal{K}_w))_{\mathcal{Y}}$ is attained for a ball $B(y, r)$, then the algorithm assigning to $w \in \mathbb{R}^m$ the center y of this ball is an optimal algorithm.
(ii) *For any $\varepsilon > 0$ and any set $V \subset \mathcal{X}$,*

$$\begin{aligned} E^*(Q, \mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M) &= \sup_w \operatorname{rad}(Q(\mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}))_{\mathcal{Y}} \le \sup_w \operatorname{diam}(Q(\mathcal{K}_w(\varepsilon, V)))_{\mathcal{Y}} \\ &\le 2\, \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \varepsilon. \end{aligned} \tag{2.7}$$

(iii) *If $Q : \mathcal{X} \to \mathcal{Y}$ is a linear map, for any $\varepsilon > 0$ and any set $V \subset \mathcal{X}$ satisfying the property $\alpha V \subset V$ for all $\alpha > 0$, one has*

$$\sup_w \operatorname{diam}(Q(\mathcal{K}_w(\varepsilon, V)))_{\mathcal{Y}} = 2\, \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \varepsilon, \tag{2.8}$$

and

$$\mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \varepsilon \le E^*(Q, \mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M) \le 2\, \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \varepsilon. \tag{2.9}$$

Proof: (i) This well-known fact follows from the pointwise optimal performance (2.3), that is

$$E^*(Q, \mathcal{K}_w, M) = \inf_{y \in \mathcal{Y}} \sup_{g \in Q(\mathcal{K}_w)} \|g - y\|_{\mathcal{Y}} = \mathrm{rad}(Q(\mathcal{K}_w))_{\mathcal{Y}}. \tag{2.10}$$

(ii) For $w \in \mathbb{R}^m$, let us consider $f, g \in \mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}$. Since $f - g \in \mathcal{N}$, it follows from the definition (2.5) of $\mu := \mu(\mathcal{N}, V, Q)_{\mathcal{X}}$ that

$$\|Q(f-g)\|_{\mathcal{Y}} \le \mu \, \mathrm{dist}(f-g, \mathcal{V})_{\mathcal{X}} \le \mu \, [\mathrm{dist}(f, V)_{\mathcal{X}} + \mathrm{dist}(g, V)_{\mathcal{X}}] \le 2\,\mu\,\varepsilon. \tag{2.11}$$

Therefore, we have $\mathrm{diam}(Q(\mathcal{K}_w(\varepsilon, V)))_{\mathcal{Y}} \le 2\,\mu\,\varepsilon$ for all $w \in \mathbb{R}^m$, which is the rightmost inequality in (2.7). Since $E^*(Q, \mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M) = \sup_w E^*(Q, \mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}, M)$, the rest of (2.7) follows from (i) and the standard comparison between the Chebyshev radius and diameter of a set.

(iii) Because of (ii), we can clearly assume that $\mu > 0$. Let us fix $\mu' \in (0, \mu)$ and then choose $\eta \in \mathcal{N}$ such that $\|Q(\eta)\|_{\mathcal{Y}} > \mu' \, \mathrm{dist}(\eta, \mathcal{V})_{\mathcal{X}}$. Since, by the linearity of Q, we also have $\|Q(\alpha\eta)\|_{\mathcal{Y}} > \mu' \, \mathrm{dist}(\alpha\eta, \mathcal{V})_{\mathcal{X}}$ for any $\alpha > 0$, we may assume that $\mathrm{dist}(\eta, \mathcal{V})_{\mathcal{X}} = 2\varepsilon$. This means that there exist $v_1, v_2 \in V$ such that $h := (\eta - v_1 + v_2)/2$ satisfies $\|h\|_{\mathcal{X}} = \varepsilon$ (if existence is not guaranteed, we can argue using limiting arguments). Since $\eta = 2h + v_1 - v_2 \in \mathcal{N}$, we can define $w \in \mathbb{R}^m$ by

$$w := M(h + v_1) = M(-h + v_2). \tag{2.12}$$

Note that

$$\mathrm{dist}(h + v_1, V)_{\mathcal{X}} \le \|h\|_{\mathcal{X}} = \varepsilon, \qquad \mathrm{dist}(-h + v_2, V)_{\mathcal{X}} \le \| - h\|_{\mathcal{X}} = \varepsilon, \tag{2.13}$$

and therefore $h + v_1$ and $-h + v_2$ both belong to $\mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}$ for our choice of w. We derive that

$$\mathrm{diam}(Q(\mathcal{K}_w(\varepsilon, V)))_{\mathcal{Y}} \ge \|Q(h+v_1) - Q(-h+v_2)\|_{\mathcal{Y}} = \|Q(2h+v_1-v_2)\|_{\mathcal{Y}} = \|Q(\eta)\|_{\mathcal{Y}} > 2\,\mu'\,\varepsilon. \tag{2.14}$$

Since this is true for all $\mu' \in (0, \mu)$, we have that

$$\mathrm{diam}(Q(\mathcal{K}_w(\varepsilon, V)))_{\mathcal{Y}} \ge 2\,\mu\,\varepsilon, \tag{2.15}$$

and the desired equality (2.8) follows by taking (ii) into consideration. To prove (2.9), we only need to show the leftmost inequality. To do so, with w as defined above, we observe that

$$\begin{aligned} E^*(Q, \mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M) &\ge E^*(Q, \mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}, M) = \mathrm{rad}(Q(\mathcal{K}_w(\varepsilon, V)))_{\mathcal{Y}} \ge \frac{1}{2}\,\mathrm{diam}(Q(\mathcal{K}_w(\varepsilon, V)))_{\mathcal{Y}} \\ &= \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\,\varepsilon, \end{aligned} \tag{2.16}$$

where we used (2.6) and the fact that the diameter of a set is at most twice its Chebyshev radius. □

2.1 Theoretical optimal and near-optimal algorithms for full approximation

Theorem 2.1 reveals the optimal performance achievable for the full recovery of f or the evaluation of a quantity of interest for f from the measurement $M(f)$, but it does not provide an optimal or near-optimal algorithm to do so. In this subsection, we review what the existing theory (mainly the results from [14]) says about near-optimal algorithms in the case $Q = I$. This existing theory will be extended in several directions later on.

It is known (and established again in Theorem 2.1) that the mapping A^* sending w to the center of a Chebyshev ball of $\mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}$ is a pointwise optimal algorithm. However, A^* is only of theoretical interest, since it is usually impractical to find a Chebyshev center.[1]

On the other hand, any map A sending $w = M(f)$ to an arbitrary point in $\mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}$ is a pointwise near-optimal algorithm as it provides the performance bound

$$\|f - A(w)\|_{\mathcal{X}} \le 2\,\mathrm{rad}(\mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}), \qquad f \in \mathcal{K}(\varepsilon, V)_{\mathcal{X}}. \tag{2.17}$$

One way of producing such a point in $\mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}$ is to solve the minimization problem

$$A(w) := \operatorname*{argmin}_{f \in \mathcal{X}} \mathrm{dist}(f, V)_{\mathcal{X}} \qquad \text{subject to } M(f) = w, \tag{2.18}$$

since $A(w)$ will always belong $\mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}$ provided this set is nonempty. But again, this algorithm may only be of theoretical interest because implementing it numerically could be impractical, especially if V is not a convex set. However, the algorithm (2.18) has the attractive feature that it does not require an a priori knowledge of ε to be executed and performs according to the estimate

$$\|f - A(M(f))\|_{\mathcal{X}} \le 2\,\mu(\mathcal{N}, V)_{\mathcal{X}}\ \mathrm{dist}(f, V)_{\mathcal{X}}, \qquad f \in \mathcal{X}, \tag{2.19}$$

as a result of (2.7) applied with $\varepsilon = \mathrm{dist}(f, V)_{\mathcal{X}}$.

2.2 Theoretical optimal and near-optimal algorithms for quantities of interest

Instead of the full approximation of f, we consider in this subsection the approximation of a quantity of interest $Q(f)$ from the given measurement $M(f)$, where $Q : \mathcal{X} \to \mathbb{R}$ is a (not necessarily linear) functional. The implications of Theorem 2.1 in this case are again somewhat folklore and are included only for completeness of our presentation. Under the assumption that $\mathcal{N} \cap V = \{0\}$ (which we always make, since otherwise $\mu(\mathcal{N}, V)_{\mathcal{X}} = +\infty$), the set $\mathcal{K}_w(\varepsilon, V)$ is bounded for any $w \in \mathbb{R}^m$ since it has a finite diameter (see Theorem 2.1 (ii) in the case of full approximation). If its image under Q is also bounded[2], then the Chebyshev ball of $Q(\mathcal{K}_w(\varepsilon, V))$ is the interval $[Q_-(w), Q_+(w)]$, where

$$Q_-(w) := \inf_{f \in \mathcal{K}_w(\varepsilon, V)} Q(f) \qquad \text{and} \qquad Q_+(w) := \sup_{f \in \mathcal{K}_w(\varepsilon, V)} Q(f), \tag{2.20}$$

with center

$$A_Q^*(w) := \frac{Q_-(w) + Q_+(w)}{2}. \tag{2.21}$$

Clearly the theoretical algorithm A_Q^* is pointwise optimal with performance given by the Chebyshev radius

$$\delta(w) := E^*(Q, \mathcal{K}_w(\varepsilon, V)_{\mathcal{X}}, M) = \frac{Q_+(w) - Q_-(w)}{2}, \tag{2.22}$$

and the global optimal performance is

$$E^*(Q, \mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M) = \sup_{w \in \mathbb{R}^m} \frac{Q_+(w) - Q_-(w)}{2}. \tag{2.23}$$

[1] It is worth mentioning that the classical notion of Chebyshev center of a set considered in this paper is different from the more computable notion considered in [8], which corresponds to the center of the largest ball contained in the given set.

[2] There are various conditions on Q ensuring that $Q(\mathcal{K}_w(\epsilon, V))$ is bounded, e.g. Q being a Lipschitz map.

From here, we can improve the estimate (ii) of Theorem 2.1 for the Chebyshev radius when Q is a sublinear functional (e.g. $Q(f) = \max f$), as the following lemma indicates.

Lemma 2.2. *Let $Q : \mathcal{X} \to \mathbb{R}$ be a functional and let V be an arbitrary subset of $\mathcal{X}$.*

(i) *If Q is a sublinear functional, namely $Q(\alpha f) = \alpha Q(f)$, for $\alpha > 0$ and $Q(f+g) \le Q(f)+Q(g)$, then, for any $w \in \mathbb{R}^m$ for which $Q(\mathcal{K}_w(\varepsilon, V))$ is bounded, the Chebyshev radius $\delta(w)$ satisfies the estimate*

$$\delta(w) = E^*(Q, \mathcal{K}_w(\varepsilon, V), M) \le \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \varepsilon. \tag{2.24}$$

(ii) *If Q is a linear functional and $\mathcal{N}$ is the null space of the measurement map $M = (l_1, \ldots, l_m)$, then*

$$\mu(\mathcal{N}, V, Q)_{\mathcal{X}} \le \operatorname{dist}(Q, L)_{\mathcal{X}^*} \mu(\mathcal{N}, V)_{\mathcal{X}}, \tag{2.25}$$

where $L := \operatorname{span}\{l_1, \ldots, l_m\} \subset \mathcal{X}^*$.

(iii) *If Q is a linear functional and $V \subset \mathcal{X}$ satisfies the property $\alpha V \subset V$ for all $\alpha > 0$, then*

$$E^*(Q, \mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M) = \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \varepsilon \le \operatorname{dist}(Q, L)_{\mathcal{X}^*} \mu(\mathcal{N}, V)_{\mathcal{X}}\, \varepsilon. \tag{2.26}$$

Proof: If $f_+, f_- \in \mathcal{K}_w(\varepsilon, V)$ are functions for which $Q(f_\pm)$ equal $Q_\pm(w)$ (or are arbitrarily close to $Q_\pm(w)$ in case the extrema are not attained), then (2.22) and the fact that $f_+ - f_- \in \mathcal{N}$ give

$$\begin{aligned} 2\delta(w) &= Q(f_+) - Q(f_-) \le Q(f_+ - f_-) \le \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \operatorname{dist}(f_+ - f_-, \mathcal{V})_{\mathcal{X}} \\ &\le \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, [\operatorname{dist}(f_+, V)_{\mathcal{X}} + \operatorname{dist}(f_-, V)_{\mathcal{X}}] \le 2\, \mu(\mathcal{N}, V, Q)_{\mathcal{X}}\, \varepsilon, \end{aligned} \tag{2.27}$$

and the proof of (i) is completed. As for (ii), with l^* denoting a best approximation to Q from L relative to the norm in $\mathcal{X}^*$, estimate (2.25) follows from

$$\begin{aligned} \mu(\mathcal{N}, V, Q)_{\mathcal{X}} &= \sup_{\eta \in \mathcal{N}} \frac{|Q(\eta)|}{\operatorname{dist}(f, \mathcal{V})_{\mathcal{X}}} = \sup_{\eta \in \mathcal{N}} \frac{|(Q - l^*)(\eta)|}{\operatorname{dist}(f, \mathcal{V})_{\mathcal{X}}} \le \sup_{\eta \in \mathcal{N}} \frac{\|Q - l^*\|_{\mathcal{X}^*} \|\eta\|_{\mathcal{X}}}{\operatorname{dist}(f, \mathcal{V})_{\mathcal{X}}} \\ &= \operatorname{dist}(Q, L)_{\mathcal{X}^*} \mu(\mathcal{N}, V)_{\mathcal{X}}. \end{aligned} \tag{2.28}$$

A combination of the left inequality in (2.9), of (2.24), and of (2.25) yields (iii). □

From a numerical point of view, we have to determine if the optimal algorithm A_Q^* or some near-optimal algorithm A_Q' can be computed without an exhaustive search over all elements from $\mathcal{K}_w(\varepsilon, V)$. This can of course be done in the following simple way: use a construction for the full approximation problem to produce $f' = A(M(f)) \in \mathcal{K}_w(\varepsilon, V)$ and apply Q to form $Q(f')$, namely consider the algorithm $A_Q'(w) := Q(A(w))$. Clearly A_Q' is a pointwise near-optimal algorithm since $Q(f') \in [Q_-(w), Q_+(w)]$, and hence

$$|Q(f) - Q(f')| \le 2\delta(w). \tag{2.29}$$

The pressing question is now whether there exist other near-optimal algorithms for quantities of interest that can be computed at a fraction of the cost of solving the full approximation problem. We will show in §3.1 that the answer to this question is affirmative. We emphasize that the optimal performance is also improved by approximating the quantity of interest directly rather than performing the full approximation and then applying Q.

3 Linear real-valued quantities of interest

In this section, we consider quantities of interest $Q : \mathcal{X} \to \mathbb{R}$ that are linear functionals. We will denote them by q. It is a well-known fact that there is a linear map $A^* : \mathbb{R}^m \to \mathbb{R}$ which is an optimal algorithm for the computation of $q(f)$ from the measurements $w = M(f)$ whenever $f \in \mathcal{K}$, with $\mathcal{K}$ being a general symmetric convex set. The existence of such an A^* was first established by Smolyak and then discussed in several papers (see e.g. [4]). All proofs we have seen are based on supporting functionals for convex sets and it is not clear to us whether such an approach can be implemented numerically (with the possible exception of special cases where the linear optimal algorithm is unique).

In §3.1, we give another construction of a linear optimal algorithm in the case where $\mathcal{K}$ is an approximation set. The advantage of this new construction is that it can be implemented numerically. Indeed, later in this paper, we employ this proof to numerically construct such linear maps in a variety of classical settings. In §3.2, we give a striking example with a nonlinear space V, showing the benefit of approximating $q(f)$ directly rather than first approximating f in full and applying q next. In §3.3, we close the section with an application of the theory to the case of equispaced point evaluations.

3.1 A linear optimal algorithm when V is a linear space

We assume here that V is a linear subspace of $\mathcal{X}$. We prove that, in this favorable but not uncommon situation, one can find an optimal (not just near-optimal) linear algorithm A_q^* for the approximation of q. The map A_q^* is precomputed offline once and for all by solving a minimization problem and can be used for all observational data $w = M(f) \in \mathbb{R}^m$. This was also observed in [4] for a more general situation (see also [23] for the complex setting), but the proof given there is not constructive. Here, we propose an explicit construction of A_q^* that will be implemented in several examples (see §5.1 and §5.2) and will also be the basis for our treatment of the full approximation problem in §4.

As before, we denote by $L := \operatorname{span}\{l_1, \dots, l_m\}$, where the l_j's are the measurement functionals in $M = (l_1, \dots, l_m)$. It follows from our standing assumption $V \cap \mathcal{N} = \{0\}$ (recall that otherwise $\mu(\mathcal{N}, V)_{\mathcal{X}} = \infty$) that the set $L_q := \{l \in L : l(v) = q(v) \text{ for all } v \in V\}$ is nonempty. Thus, we may consider a best approximation to q from the convex set L_q, i.e.,

$$l^* \in \operatorname*{argmin}_{l \in L_q} \|q - l\|_{\mathcal{X}^*}. \tag{3.1}$$

Because $l^* \in L$, we can write

$$l^* = \sum_{j=1}^m a_j^* l_j \tag{3.2}$$

for some $a^* \in \mathbb{R}^m$. Such a vector a^* is not unique in general, as we shall see in §5.1.

Theorem 3.1. *Let $\mathcal{X}$ be a Banach space and $V \subset \mathcal{X}$ be a finite-dimensional subspace in $\mathcal{X}$ of dimension $n \le m$. Let $q \in \mathcal{X}^*$ be a linear functional on $\mathcal{X}$. If $a^* \in \mathbb{R}^m$ satisfies* (3.1)-(3.2)*, then the mapping $A_q^* : \mathbb{R}^m \to \mathbb{R}$ defined by*

$$A_q^*(w) = \sum_{j=1}^m a_j^* w_j, \qquad w \in \mathbb{R}^m, \tag{3.3}$$

is a linear optimal algorithm for the approximation of $q(f)$ from the data $w = M(f)$, namely

$$|q(f) - A_q^*(M(f))| \le \mu(\mathcal{N}, V, q)_{\mathcal{X}} \operatorname{dist}(f, V)_{\mathcal{X}} \qquad \text{for all } f \in \mathcal{X}. \tag{3.4}$$

Proof: Consider the linear functional λ defined on $\mathcal{N}+V$ by $\lambda(\eta) = q(\eta)$ for all $\eta \in \mathcal{N}$ and $\lambda(v) = 0$ for all $v \in V$. Any extension $\tilde{\lambda}$ of λ defined on the whole $\mathcal{X}$ is precisely of the form $\tilde{\lambda} = q - l$ for some $l \in L_q$. Since a Hahn–Banach extension of λ is an extension with minimal norm, we see that $q - l^*$ is a Hahn–Banach extension of $\lambda = (q - l^*)_{\mathcal{N}+V}$. As such, we have

$$\begin{aligned} \|q - l^*\|_{\mathcal{X}^*} &= \|q - l^*\|_{(\mathcal{N}+V)^*} = \sup_{\|\eta+v\|_{\mathcal{X}} \le 1} |(q - l^*)(\eta + v)| = \sup_{\|\eta+v\|_{\mathcal{X}} \le 1} |(q - l^*)(\eta)| \\ &= \sup_{\|\eta+v\|_{\mathcal{X}} \le 1} |q(\eta)| = \sup_{\operatorname{dist}(\eta, V)_{\mathcal{X}} \le 1} |q(\eta)| = \mu(\mathcal{N}, V, q)_{\mathcal{X}}, \end{aligned} \tag{3.5}$$

where the third equality uses the fact that $q - l^*$ vanishes on V, the fourth equality uses the fact that l^* vanishes on $\mathcal{N}$, and the last equality is the definition of $\mu(\mathcal{N}, V, q)_{\mathcal{X}}$. Then, for any $f \in \mathcal{X}$, picking $v \in V$ such that $\|f - v\|_{\mathcal{X}} = \operatorname{dist}(f, V)_{\mathcal{X}}$, we obtain

$$\begin{aligned} |q(f) - A_q^*(M(f))| &= |q(f) - l^*(f)| = |q(f - v) - l^*(f - v)| = |(q - l^*)(f - v)| \\ &\le \|q - l^*\|_{\mathcal{X}^*} \|f - v\|_{\mathcal{X}} = \mu(\mathcal{N}, V, q)_{\mathcal{X}} \operatorname{dist}(f, V)_{\mathcal{X}}. \end{aligned} \tag{3.6}$$

This proves (3.4), and therefore $E(q, \mathcal{K}(\varepsilon, V)_{\mathcal{X}}, M, A_q^*) \le \mu(\mathcal{N}, V, q)\, \varepsilon$. Together with Lemma 2.2, part (iii), this establishes the optimality of the algorithm A_q^*. □

Remark 3.2. *The algorithm A_q^* possesses the attractive feature that it does not require an a priori knowledge of ε to be executed.*

3.2 Nonlinear space V: the case of compressive sensing

We now place ourselves in the well-studied setting of compressive sensing to emphasize another advantage of directly computing linear quantities of interest $q(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^N$, from the measurements $M(\mathbf{x}) \in \mathbb{R}^m$. More precisely, we show that the approximation of $q(\mathbf{x})$ with a desired accuracy may be feasible whereas the full approximation to $\mathbf{x}$ with the same accuracy is impossible.

Here, we deal with the space $\mathcal{X} = \ell_p^N$, $1 \le p \le \infty$, and the set $V = \Sigma_s$ of s-sparse vectors in $\mathbb{R}^N$. Note that V is not a linear space and that $\mathcal{V} = \Sigma_{2s}$. The measurement map M can be viewed as an $m \times N$ matrix acting on vectors $\mathbf{x} \in \mathbb{R}^N$. One premise of compressive sensing is the user's ability to select good matrices M for the purpose of recovering exactly the elements of Σ_s. Our view is different here, as the measurements are given to us and we do not have the luxury of choosing them. Nonetheless, (2.18) provides an algorithm A with the performance bound

$$\|\mathbf{x} - A(M(\mathbf{x}))\|_{\ell_p^N} \le 2\, \mu(\mathcal{N}, \Sigma_s)_{\ell_p^N} \sigma_s(\mathbf{x})_{\ell_p^N}, \tag{3.7}$$

where $\sigma_s(\mathbf{x})_{\ell_p^N}$ is the error of best s-term approximation, i.e.,

$$\sigma_s(\mathbf{x})_{\ell_p^N} := \inf_{\mathbf{z} \in \Sigma_s} \|\mathbf{x} - \mathbf{z}\|_{\ell_p^N}. \tag{3.8}$$

Two major achievements of compressive sensing were the realizations that there exist many measurement matrices M yielding $\mu(\mathcal{N}, \Sigma_s)_{\ell_1^N} \le 2$ provided $m \gtrsim s \ln(eN/s)$ (this result is known in relation with the so-called null space property, see [9], [17, Chapter 11]) and that the ℓ_1-minimization

$$A(w) := \operatorname*{argmin}_{M\mathbf{z} = w} \|\mathbf{z}\|_{\ell_1^N} \tag{3.9}$$

constitutes a near-optimal algorithm for these favorable matrices. We note that, even though the performance estimate (3.7) holds for any measurement matrix M, it is unclear that constructing a near-optimal algorithm with an acceptable cost (like (3.9)) can be done for arbitrary M. We also note that, when $p > 1$, a number of measurements m much larger than $s \ln(eN/s)$ is needed to make $\mu(\mathcal{N}, \Sigma_s)_{\ell_p^N}$ bounded by an absolute constant. For example, when $p = 2$, one needs $m \asymp N$ even with $s = 1$.

The situation becomes more interesting for the approximation of a quantity of interest q that is a linear functional, namely

$$q(\mathbf{x}) = \langle \xi, \mathbf{x} \rangle, \tag{3.10}$$

where $\xi \in \mathbb{R}^N$ is a fixed vector assumed without loss of generality to satisfy $\|\xi\|_{\ell_2^N} = 1$. Let us restrict our attention to the most detrimental case, namely $\mathcal{X} = \ell_2^N$. Since the performance of the optimal algorithm is governed by $\mu(\mathcal{N}, \Sigma_s, q)_{\ell_2^N}$, see Lemma 2.2, part (iii), we inquire about the range of m making it possible to bound the quantity

$$\mu(\mathcal{N}, \Sigma_s, q)_{\ell_2^N} = \sup_{\eta \in \mathcal{N}} \frac{\langle \xi, \eta \rangle}{\sigma_{2s}(\eta)_{\ell_2^N}} \tag{3.11}$$

by an absolute constant. This clearly depends on ξ. At one extreme, if ξ is a row of M, then $\mu(\mathcal{N}, \Sigma_s, q)_{\ell_2^N} = 0$. At the other extreme, if ξ is an extremal vector in the definition of $\mu(\mathcal{N}, \Sigma_s)_{\ell_2^N}$, i.e., $\xi \in \mathcal{N}$, $\|\xi\|_{\ell_2^N} = 1$, and $\sigma_{2s}(\xi)_{\ell_2^N} = 1/\mu(\mathcal{N}, \Sigma_s)_{\ell_2^N}$, then

$$\mu(\mathcal{N}, \Sigma_s, q)_{\ell_2^N} \geq \frac{\langle \xi, \xi \rangle}{\sigma_{2s}(\xi)_{\ell_2^N}} = \mu(\mathcal{N}, \Sigma_s)_{\ell_2^N}. \tag{3.12}$$

In other words, the approximation of this quantity of interest is as bad as the full approximation, so that $m \asymp N$ measurements are needed even with $s = 1$.

In a more positive direction, we isolate the following simple result.

Lemma 3.3. *If $\xi \in \mathbb{R}^N$ is such that $\|\xi\|_{\ell_2^N} = 1$ and $|\xi_j| \leq C/\sqrt{N}$, $j = 1, \ldots, N$, then the linear functional $q(\cdot) = \langle \xi, \cdot \rangle$ satisfies the inequality*

$$\mu(\mathcal{N}, \Sigma_s, q)_{\ell_2^N} \leq C \mu(\mathcal{N}, \Sigma_s)_{\ell_1^N}. \tag{3.13}$$

Proof: For any $\eta \in \mathcal{N}$, we have

$$|q(\eta)| = |\langle \xi, \eta \rangle| \leq \frac{C}{\sqrt{N}} \|\eta\|_{\ell_1^N} \leq \frac{C}{\sqrt{N}} \mu(\mathcal{N}, \Sigma_s)_{\ell_1^N} \sigma_{2s}(\eta)_{\ell_1^N} \leq C \mu(\mathcal{N}, \Sigma_s)_{\ell_1^N} \sigma_{2s}(\eta)_{\ell_2^N}. \tag{3.14}$$

Dividing throughout by $\sigma_{2s}(\eta)_{\ell_2^N}$ and taking the supremum over η yields (3.13). □

This simple result has an interesting consequence for $m \times N$ matrices M with $m \asymp s \log(eN/s)$ and $\mu(\mathcal{N}, \Sigma_s)_{\ell_1^N}$ bounded by an absolute constant. Indeed, for such matrices and for ξ as in the lemma, approximating the quantity of interest $q(\mathbf{x}) = \langle \xi, \mathbf{x} \rangle$ can be performed to accuracy $\sigma_s(\mathbf{x})_{\ell_2^N}$ for all $\mathbf{x} \in \ell_2^N$, whereas the approximation of $\mathbf{x}$ itself to this accuracy is not possible because $m \asymp N$ does not hold.

3.3 Integral approximation from equispaced point evaluations

In this subsection, we discuss a concrete example where our abstract theory applies. We work with $\mathcal{X} = C[-1,1]$, the linear space $V = \mathcal{P}_{n-1}$ of algebraic polynomials of degree at most $n-1$, and a measurement map M given by point evaluations at the equispaced points $-1 = \tau_1 < \cdots < \tau_m = 1$ defined by

$$\tau_j = -1 + \frac{2(j-1)}{m-1}, \qquad j = 1, \ldots, m. \tag{3.15}$$

The quantity of interest is $q(f) = \int_{-1}^{1} f(x)\,dx$.

Let us take a step back, though, and not work with equispaced points and polynomial spaces right away. Thus, we are looking for quadrature formulas where the points $x_1, \ldots, x_m$ are specified. We have seen that the optimal performance is dictated by the quantity

$$\mu^{\mathrm{int}}(x_1, \ldots, x_m; V) := \sup_{\eta(x_1)=\cdots=\eta(x_m)=0} \frac{\int_{-1}^{1} \eta}{\mathrm{dist}(\eta, V)_{C[-1,1]}}. \tag{3.16}$$

We want to understand the dependence of this quantity on m and n, in particular for the equispaced points. We also seek to establish that one can do better (in terms of error and not only in terms of computational cost) when directly computing this integral rather than fully approximating f and taking the integral afterwards. We start by pointing out an alternative expression which is easier to handle than (3.16).

Lemma 3.4. *For any $x_1, \ldots, x_m \in [-1,1]$ and any n-dimensional subspace V of $C[-1,1]$,*

$$\mu^{\mathrm{int}}(x_1, \ldots, x_m; V) = 2 + \sup_{v \in V} \frac{\int_{-1}^{1} v}{\max\limits_{1 \le j \le m} |v(x_j)|}. \tag{3.17}$$

Proof: Let us denote

$$\sigma := \sup_{v \in V} \frac{\int_{-1}^{1} v}{\max\limits_{1 \le j \le m} |v(x_j)|}. \tag{3.18}$$

Given a function $\eta \in C[-1,1]$ with $\eta(x_1) = \cdots = \eta(x_m) = 0$, we consider an element $v \in V$ such that $\|\eta - v\|_{C[-1,1]} = \mathrm{dist}(\eta, V)_{C[-1,1]}$. From $\max_{1 \le j \le m} |v(x_j)| = \max_{1 \le j \le m} |(\eta - v)(x_j)| \le \|\eta - v\|_{C[-1,1]}$, we derive that

$$\frac{\int_{-1}^{1} \eta}{\mathrm{dist}(\eta, V)_{C[-1,1]}} = \frac{\int_{-1}^{1} (\eta - v + v)}{\|\eta - v\|_{C[-1,1]}} \le \int_{-1}^{1} \frac{\eta - v}{\|\eta - v\|_{C[-1,1]}} + \frac{\int_{-1}^{1} v}{\max_{1 \le j \le m} |v(x_j)|} \le 2 + \sigma. \tag{3.19}$$

Taking the supremum over η yields $\mu^{\mathrm{int}}(x_1, \ldots, x_m; V) \le 2 + \sigma$. To prove the reverse inequality, let us consider $v \in V$ such that $\max_{1 \le j \le m} |v(x_j)| = 1$ and $\int_{-1}^{1} v = \sigma$. For an arbitrarily small $\delta > 0$,

choose $h_\delta \in C[-1,1]$ with $\|h_\delta\|_{C[-1,1]} = 1$, $h_\delta(x_j) = -v(x_j)$ for all $j = 1, \dots, m$, and $h_\delta(x) = 1$ for all x in a set of measure $2 - \delta$. Setting $\eta := h_\delta + v \in C[-1,1]$, we have $\eta(x_1) = \cdots = \eta(x_m) = 0$ and $\operatorname{dist}(\eta, V)_{C[-1,1]} \le \|\eta - v\|_{C[-1,1]} = \|h_\delta\|_{C[-1,1]} = 1$. Therefore,

$$\mu^{\rm int}(x_1, \dots, x_m; V) \ge \frac{\int_{-1}^{1} \eta}{\operatorname{dist}(\eta, V)_{C[-1,1]}} \ge \int_{-1}^{1} (h_\delta + v) = \int_{-1}^{1} h_\delta + \int_{-1}^{1} v \ge 2 - 2\delta + \sigma. \tag{3.20}$$

Since $\delta > 0$ was arbitrary, we arrive at $\mu^{\rm int}(x_1, \dots, x_m; V) \ge 2 + \sigma$, which concludes the proof. □

Let us now return to the case of the space $\mathcal{P}_{n-1}$ and of equidistant points τ_j, $j = 1, \dots, m$. We use the notation

$$\mu^{\rm int}_{m,n} := \mu^{\rm int}(\tau_1, \dots, \tau_m; \mathcal{P}_{n-1}), \tag{3.21}$$

while $\tilde\mu_{m,n}$ stands for a quantity closely related to $\mu(\mathcal{N}, \mathcal{P}_{n-1})_{C[-1,1]}$ (see (4.73)-(4.74) below), namely

$$\tilde\mu_{m,n} := \max_{p \in \mathcal{P}_{n-1}} \frac{\|p\|_{C[-1,1]}}{\max\limits_{1 \le j \le m} |p(\tau_j)|}. \tag{3.22}$$

The following lemma relating $\mu^{\rm int}_{m,n}$ and $\tilde\mu_{m,n}$ holds.

Lemma 3.5. *For all integers $m \ge n \ge 1$, one has*

$$2 + \frac{8}{(n+1)^2} \tilde\mu_{m,\lfloor (n-1)/2 \rfloor} \le \mu^{\rm int}_{m,n} \le 4 + \frac{n-1}{m-1} \tilde\mu_{m,n}. \tag{3.23}$$

Proof: According to Lemma 3.4, it is enough to show that

$$\frac{8}{(n+1)^2} \tilde\mu_{m,\lfloor (n-1)/2 \rfloor} \le \sigma \le 2 + \frac{n-1}{m-1} \tilde\mu_{m,n}, \qquad \text{where} \quad \sigma := \max_{p \in \mathcal{P}_{n-1}} \frac{\int_{-1}^{1} p}{\max\limits_{1 \le j \le m} |p(\tau_j)|}. \tag{3.24}$$

To prove the upper estimate, let us consider $p \in \mathcal{P}_{n-1}$ with $\max_{1 \le j \le m} |p(\tau_j)| = 1$. Note that $\|p\|_{C[-1,1]} \le \tilde\mu := \tilde\mu_{m,n}$. For an arbitrary $\gamma > 1$, let

$$J := \{ j = 1, \dots, m-1 : \max_{[\tau_j, \tau_{j+1}]} p > \gamma \}. \tag{3.25}$$

Since there are at least two zeros of $p - \gamma$ in (τ_j, τ_{j+1}) for each $j \in J$, and since $p - \gamma$ has at most $n - 1$ zeros in total, we derive that $|J| \le (n-1)/2$. It follows that

$$\int_{-1}^{1} p = \sum_{j \notin J} \int_{\tau_j}^{\tau_{j+1}} p + \sum_{j \in J} \int_{\tau_j}^{\tau_{j+1}} p \le \sum_{j \notin J} \int_{\tau_j}^{\tau_{j+1}} \gamma + \sum_{j \in J} \int_{\tau_j}^{\tau_{j+1}} \tilde\mu \le 2\gamma + \frac{n-1}{2} \frac{2}{m-1} \tilde\mu = 2\gamma + \frac{n-1}{m-1} \tilde\mu. \tag{3.26}$$

We obtain the required upper bound by letting $\gamma \to 1$ and then by taking the maximum over p.

For the lower estimate, consider an extremal polynomial $q \in \mathcal{P}_{\lfloor (n-1)/2 \rfloor}$ in the definition of $\tilde{\mu}_{m,\lfloor (n-1)/2 \rfloor}$, normalized so that $\max_{1 \le j \le m} |q(\tau_j)| = 1$ and $\|q\|_{C[-1,1]} = \tilde{\mu}_{m,\lfloor (n-1)/2 \rfloor}$. Note that the polynomial $p := q^2 \in \mathcal{P}_{n-1}$ satisfies $\max_{1 \le j \le m} |p(\tau_j)| = 1$. Therefore, we have

$$\sigma \ge \int_{-1}^{1} p = \int_{-1}^{1} q^2 \ge \frac{2}{(\lfloor (n-1)/2 \rfloor + 1)^2} \, \|q\|^2_{C[-1,1]} \ge \frac{8}{(n+1)^2} \, \tilde{\mu}^2_{m,\lfloor (n-1)/2 \rfloor}, \tag{3.27}$$

where we have used the sharp comparison between the $C[-1,1]$ and $L_2[-1,1]$ norms on $\mathcal{P}_d$ with $d = \lfloor (n-1)/2 \rfloor$, that is

$$\|r\|_{C[-1,1]} \le \frac{(d+1)^2}{2} \|r\|_{L_2[-1,1]}, \qquad r \in \mathcal{P}_d. \tag{3.28}$$

The proof is now complete. □

The latter lemma provides an error estimate for the quadrature formula generated by our theory, as formalized in the following theorem.

Theorem 3.6. *Let*

$$\int_{-1}^{1} f \approx A^*(f(\tau_1), \dots, f(\tau_m)) \tag{3.29}$$

be a quadrature formula produced by Theorem 3.1 with $\mathcal{X} = C[-1,1]$,$V = \mathcal{P}_{n-1}$, and equispaced points $-1 = \tau_1 < \cdots < \tau_m = 1$. There are constants $c, C > 0$ such that, if $n \le c\sqrt{m \ln m}$, then

$$\left| \int_{-1}^{1} f - A^*(f(\tau_1), \dots, f(\tau_m)) \right| \le C \, \mathrm{dist}(f, \mathcal{P}_{n-1})_{C[-1,1]} \qquad \textit{for all } f \in C[-1,1]. \tag{3.30}$$

Proof: Theoren 3.1 ensures that, for all $f \in C[-1,1]$,

$$\left| \int_{-1}^{1} f - A^*(f(\tau_1), \dots, f(\tau_m)) \right| \le \mu^{\mathrm{int}}_{m,n} \, \mathrm{dist}(f, \mathcal{P}_{n-1})_{C[-1,1]}. \tag{3.31}$$

From Lemma 3.5 and the fact that $\tilde{\mu}_{m,n} \le \beta^{\frac{n^2}{m}}$ for some $\beta > 1$ (see [10]), when $n \le c\sqrt{m \ln m}$ with $c^2 = 1/(3 \ln \beta)$, we have

$$\mu^{\mathrm{int}}_{m,n} \le 4 + \frac{n}{m} \beta^{\frac{n^2}{m}} \le 4 + \frac{c\sqrt{\ln m}}{\sqrt{m}} e^{(\ln \beta) c^2 (\ln m)} = 4 + \frac{c\sqrt{\ln(m)}}{\sqrt{m}} m^{1/3} \le C, \tag{3.32}$$

which implies that (3.30) holds. □

We are now ready to deliver the main message of this section: given data at m equispaced points, the range of n enabling the approximation of the integral of continuous functions f with accuracy $\mathrm{dist}(f, \mathcal{P}_{n-1})_{C[-1,1]}$ is larger than the range enabling the approximation of the functions themselves with the same accuracy. Indeed, we know on the one hand (see Theorem 2.1, part (iii)

with $Q = I$) that no matter the procedure $A : \mathbb{R}^m \to C[-1,1]$, there is an $f \in C[-1,1]$ with $\operatorname{dist}(f, \mathcal{P}_{n-1})_{C[-1,1]} \le \varepsilon$ such that

$$\|f - A(f(\tau_1), \dots, f(\tau_m))\|_{C[-1,1]} \ge \mu_{m,n}\, \varepsilon, \tag{3.33}$$

where

$$\mu_{m,n} := \mu(\mathcal{N}, \mathcal{P}_{n-1})_{C[-1,1]} = \sup_{\eta(\tau_1)=\cdots=\eta(\tau_m)=0} \frac{\|\eta\|_{C[-1,1]}}{\operatorname{dist}(\eta, \mathcal{P}_{n-1})_{C[-1,1]}}. \tag{3.34}$$

Thus, if targeted estimates of the type $\|f - A(f(\tau_1), \dots, f(\tau_m))\|_{C[-1,1]} \le C \operatorname{dist}(f, \mathcal{P}_{n-1})_{C[-1,1]}$ are to be valid for all $f \in C[-1,1]$, we must have $\mu_{m,n} \le C$, which imposes $n \le c\sqrt{m}$, as will become apparent later in §4.4. On the other hand, Theorem 3.6 shows that (3.30) holds with n as large as $c\sqrt{m \ln m}$. In passing, we also point out that A^* constitutes a stable algorithm in the sense that $\|A^*\|_{\ell_\infty^m \to \mathbb{R}} \le C$.

4 Algorithms for full approximation in $C(D)$

We return in this section to the task of providing a full approximation to f in the setting $\mathcal{X} = C(D)$ from the measurements $w = M(f) = (f(x_1), \dots, f(x_m))$, where $x_j \in D$, $j = 1, \dots, m$. We also assume that V is a linear space of dimension $n \le m$.

We start with a discussion (see §4.1) about what is known on numerically executable near-optimal algorithms for the full recovery of f. We emphasize that, up to this point, the near-optimal algorithms are generally nonlinear and involve the solution of a convex optimization problem, similar to ℓ_1-minimization in terms of complexity (see [14] or **R2** in §4.1). Next, we further study the full approximation problem and assume for some results that V contains constant functions. If this is not the case, one can append the function $v \equiv 1$ to V and end up with an $(n+1)$-dimensional space $\bar{V} \supset V$. Note however that $\mu(\mathcal{N}, \bar{V})_{C(D)}$ will generally be larger than $\mu(\mathcal{N}, V)_{C(D)}$.

In §4.2, we first show that there always exist optimal algorithms which are linear. The proof of this fact is not constructive. In §4.3, we turn next to the question of practical constructions of linear algorithms which are near-optimal. Our constructions are based on quasi-projectors and can be carried out numerically in classical settings, for example, when the approximation space V is a space of trigonometric or algebraic polynomials. In §4.4, we close the section with an application of the theory to the case of equispaced point evaluations.

4.1 Known results

The results listed below were obtained in [14], except for (4.1), which is proved in the appendix.

R1: *Let $\mathcal{N}$ be the null space of the measurement map M. Then, we have*

$$\mu(\mathcal{N}, V)_{C(D)} = 1 + \mu(V, \mathcal{N})_{C(D)}, \tag{4.1}$$

and

$$\mu(V, \mathcal{N})_{C(D)} = \max_{v \in V} \frac{\|v\|_{C(D)}}{\max\limits_{1 \le j \le m} |v(x_j)|}. \tag{4.2}$$

R2: *Given the data $w = M(f) \in \mathbb{R}^m$, we consider the function $v := v(w) \in V$ defined by*

$$v := \operatorname*{argmin}_{u \in V} \max_{1 \le j \le m} |u(x_j) - w_j|, \tag{4.3}$$

and make a correction[3] *to interpolate the data by forming*

$$A(w) := v(w) + \sum_{i=1}^{m} (w_i - v(w)(x_i))\psi_i, \tag{4.4}$$

where $\psi_1, \ldots, \psi_m \in C(D)$ *are any functions with disjoint supports, norm one, and satisfying* $\psi_i(x_j) = \delta_{i,j}$. *Then the mapping* $A : \mathbb{R}^m \to C(D)$ *provides a pointwise near-optimal algorithm, i.e., for each* $w \in \mathbb{R}^m$,

$$\sup_{f \in \mathcal{K}_w} \|f - A(w)\|_{C(D)} \le 2 \operatorname{rad}(\mathcal{K}_w)_{C(D)}. \tag{4.5}$$

R3: *For each* $f \in C(D)$,

$$\|f - A(M(f))\|_{C(D)} \le 4\, \mu(\mathcal{N}, V)_{C(D)} \operatorname{dist}(f, V)_{C(D)}. \tag{4.6}$$

Estimates of this type cannot be improved, except for reduction of the constant 4.

Although the algorithm just described is pointwise and global near-optimal on approximation sets, its deficiency lies in its numerical implementation. Indeed, given the observational data $w = M(f)$, not only is the minimization problem (4.3) quite costly, but solving it has to be done whenever new data $w \in \mathbb{R}^m$ comes in. In contrast, we will show in §4.2 that there is a linear algorithm of the form

$$A(w) = \sum_{j=1}^{m} w_j \phi_j \tag{4.7}$$

for some suitably chosen functions $\phi_1, \ldots, \phi_m \in C(D)$ which is optimal. Furthermore, in many classical settings, i.e., for many choices for V, we will present an implementable numerical recipe to find a linear algorithm of the above form which is near-optimal. Such linear algorithms have the clear advantage of avoiding to solve (4.3) for each new set of measurements.

4.2 Linear optimal algorithms

We establish here the existence of a linear optimal algorithm $A^* : \mathbb{R}^m \to C(D)$ for the recovery of functions $f \in C(D)$ from the point values $w = M(f) = (f(x_1), \ldots, f(x_m))$ at m distinct points $x_1, \ldots, x_m \in D$. For any $x \in D$, we write δ_x for the linear functional defined by $\delta_x(f) := f(x)$. As before, we denote by $L \subset C(D)^*$ the subspace of all linear functionals spanned by the l_j's, where now $l_j = \delta_{x_j}$, $j = 1, \ldots, m$, that is

$$L := \left\{ \lambda_a = \sum_{j=1}^{m} a_j \delta_{x_j},\ a \in \mathbb{R}^m \right\}. \tag{4.8}$$

Continuing to work under the assumption that V is a linear subspace of $C(D)$ of dimension $n \le m$, we consider the subspace L_{δ_x} of L given by

$$L_{\delta_x} := \{\lambda \in L : \lambda(v) = \delta_x(v) \text{ for all } v \in V\}. \tag{4.9}$$

[3]The correction (4.4) may be omitted, since a pointwise near-optimal algorithm is already provided by $w \mapsto v(w)$, but it makes the algorithm A data-consistent.

We apply Theorem 3.1 to the linear functional $q = \delta_x$. In the case $x \notin \{x_1, \ldots, x_m\}$, since

$$\|\delta_x - \lambda_a\|_{\mathcal{X}^*} = 1 + \sum_{j=1}^{m} |a_j|, \tag{4.10}$$

a solution $a^* = a^*(x) \in \mathbb{R}^m$ to the minimization problem

$$\underset{a \in \mathbb{R}^m}{\text{minimize}} \sum_{j=1}^{m} |a_j| \qquad \text{subject to} \quad \sum_{j=1}^{m} a_j v(x_j) = v(x) \ \text{ for all } v \in V \tag{4.11}$$

provides a best approximation λ_{a^*} to δ_x from L_{δ_x} and we observe that (see (3.5))

$$\mu(\mathcal{N}, V, \delta_x)_{C(D)} = \|\delta_x - \lambda_{a^*}\|_{\mathcal{X}^*} = 1 + \sum_{j=1}^{m} |a_j^*(x)|. \tag{4.12}$$

On the other hand, when $x = x_j$ for some $j = 1, \ldots, m$, we have $\delta_x = \delta_{x_j}$, and therefore the best approximation to δ_x from the set L_{δ_x} is δ_{x_j}. We thus set $a^*(x_j) = e_j$, the j-th unit vector from the standard basis for $\mathbb{R}^m$. We now define the algorithm $A^* : \mathbb{R}^m \to C(D)$ by

$$A^*(w)(x) := A^*_{\delta_x}(w) = \sum_{j=1}^{m} w_j a_j^*(x). \tag{4.13}$$

Clearly, the linear mapping A^* is data-consistent, in the sense that

$$A^*(w)(x_j) = w_j \qquad \text{for all } w \in \mathbb{R}^m \text{ and all } j = 1, \ldots, m. \tag{4.14}$$

Moreover, it is a pointwise optimal algorithm for the recovery of $f(x)$, $x \in D$, since (3.4) reads

$$|f(x) - A^*(M(f))(x)| = |f(x) - A^*_{\delta_x}(M(f))| \le \mu(\mathcal{N}, V, \delta_x)_{C(D)} \operatorname{dist}(f, V)_{C(D)}. \tag{4.15}$$

As already observed by Bakhvalov (see [4]), it follows that, for any $x \in D$,

$$|f(x) - A^*(M(f))(x)| \le \mu(\mathcal{N}, V)_{C(D)} \operatorname{dist}(f, V)_{C(D)}, \tag{4.16}$$

so taking the supremum over x should provide a globally optimal algorithm, except that so far A^* is only known to map into $L_\infty(D)$ rather than in $\mathcal{X} = C(D)$. But let us recall that $a^*(x)$ may not be uniquely defined when $x \notin \{x_1, \ldots, x_m\}$, i.e., the set

$$\Lambda_x := \{a \in \mathbb{R}^m : \lambda_a \text{ is a best approximation to } \delta_x \text{ from } L_{\delta_x}\} \tag{4.17}$$

may contain more than one element. We are going to show that we can select an element in this set in such a way that $x \in D \mapsto a^*(x) \in \Lambda_x$ is continuous, so that the algorithm A^* introduced in (4.13) does indeed map into $C(D)$. The proof uses the following result.

Lemma 4.1. *Let $\theta_1, \ldots, \theta_N$ be N distinct points in $\mathbb{R}^n$ with convex hull $\mathcal{C} := \operatorname{conv}\{\theta_1, \ldots, \theta_N\}$. Then there exist continuous functions $\psi_i : \mathcal{C} \to \mathbb{R}$, $\psi_i \ge 0$, $i = 1, \ldots, N$, with*

$$\sum_{i=1}^{N} \psi_i(\theta) = 1, \qquad \sum_{i=1}^{N} \psi_i(\theta)\theta_i = \theta, \qquad \theta \in \mathcal{C}. \tag{4.18}$$

Moreover, these functions can be chosen to satisfy

$$\psi_i(\theta_j) = \delta_{i,j}, \qquad i, j = 1, \ldots, N. \tag{4.19}$$

This lemma, except for (4.19), can be found in [18]. We supply a proof of the complete statement in the appendix. We are now in the position to state and prove the main result of this subsection.

Theorem 4.2. *Let V be an n-dimensional subspace of $C(D)$ that contains the constant functions. Then there is a linear pointwise optimal algorithm $A^* : \mathbb{R}^m \to C(D)$ for the recovery of any $f \in C(D)$ from the measurement $M(f) = (f(x_1), \dots, f(x_m)) \in \mathbb{R}^m$, where x_j are m distinct points in D. More precisely, for each $x \in D$, one has*

$$|f(x) - A^*(M(f))(x)| \le \mu(\mathcal{N}, V, \delta_x)_{C(D)} \mathrm{dist}(f, V)_{C(D)} \qquad \text{for all } f \in C(D). \tag{4.20}$$

The algorithm A^* is also globally optimal for the full approximation problem, in the sense that

$$\|f - A^*(M(f))\|_{C(D)} \le \mu(\mathcal{N}, V)_{C(D)} \, \mathrm{dist}(f, V)_{C(D)} \qquad \text{for all } f \in C(D). \tag{4.21}$$

Proof: We shall rely on Lemma 4.1 to make a continuous selection $x \mapsto (a_1^*(x), \dots, a_m^*(x))$ of the functions appearing in the definition (4.13) of the algorithm A^*. Let $(v_1, \dots, v_n)$ denote a basis for V with $v_1 \equiv 1$. For each $x \in D$, we consider the vector $\theta(x) \in \mathbb{R}^n$ given by

$$\theta(x) := \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_n(x) \end{pmatrix} = \begin{pmatrix} 1 \\ v_2(x) \\ \vdots \\ v_n(x) \end{pmatrix}, \tag{4.22}$$

and we define $\theta_1, \dots, \theta_m, \theta_{m+1}, \dots, \theta_{2m} \in \mathbb{R}^n$ by

$$\theta_j := \theta(x_j), \qquad \theta_{m+j} = -\theta_j, \qquad j = 1, \dots, m. \tag{4.23}$$

Note that each θ_{m+i}, $i = 1, \dots, m$, is distinct from any θ_j, $j = 1, \dots, m$, because their first coordinates are different. It is easy to verify that the expression

$$\|\theta\| = \min \left\{ \sum_{j=1}^m |a_j| : \sum_{j=1}^m a_j \theta_j = \theta \right\}, \qquad \theta \in \mathbb{R}^n, \tag{4.24}$$

defines a norm on $\mathbb{R}^n$ (called the atomic norm for the set of atoms $\{\theta_1, \dots, \theta_m, \theta_{m+1}, \dots, \theta_{2m}\}$) whose unit ball is

$$\mathcal{C} := \mathrm{conv}\{\theta_1, \dots, \theta_m, \theta_{m+1}, \dots, \theta_{2m}\}. \tag{4.25}$$

We notice that, for every $x \in D$,

$$\|\theta(x)\| \ge 1, \tag{4.26}$$

since $v_1 \equiv 1$ implies $\sum_{j=1}^m |a_j| \ge \sum_{j=1}^m a_j = 1$ whenever $\sum_{j=1}^m a_j \theta_j = \theta(x)$. Moreover, we clearly have $\|\theta_j\| \le 1$. Therefore,

$$\|\theta_j\| = 1, \qquad j = 1, \dots, m. \tag{4.27}$$

We separate two cases.

Case 1: All of the $\theta_1, \dots, \theta_m$ are distinct. Since $\theta(x)/\|\theta(x)\| \in \mathcal{C}$, we use Lemma 4.1 to write

$$\frac{\theta(x)}{\|\theta(x)\|} = \sum_{j=1}^{2m} \psi_j \left(\frac{\theta(x)}{\|\theta(x)\|} \right) \theta_j, \tag{4.28}$$

so that

$$\theta(x) = \sum_{j=1}^{m} a_j^*(x)\theta_j, \qquad \text{with} \quad a_j^*(x) := \|\theta(x)\| \left(\psi_j \left(\frac{\theta(x)}{\|\theta(x)\|} \right) - \psi_{m+j} \left(\frac{\theta(x)}{\|\theta(x)\|} \right) \right). \tag{4.29}$$

Since the vector $\theta(x)$ varies continuously with $x \in D$ and the ψ_j's are continuous, the functions $a_1^*, \ldots, a_m^*$ thus defined are also continuous on D.

Next, we prove that the vector $a^*(x) = (a_1^*(x), \ldots, a_m^*(x))$ belongs to Λ_x for each $x \in D$. When $x = x_j$ for some $j = 1, \ldots, m$, this follows from (4.29), (4.27), and (4.19), since

$$a_i^*(x_j) = \psi_i(\theta_j) - \psi_{m+i}(\theta_j) = \delta_{i,j} - 0 = \delta_{i,j}, \tag{4.30}$$

hence $a^*(x_j) = e_j \in \Lambda_{x_j}$. In the case $x \notin \{x_1, \ldots, x_m\}$, we have to check that $a^*(x)$ minimizes $\sum_{j=1}^m |a_j|$ subject to the constraint $\sum_{j=1}^m a_j v(x_j) = v(x)$ for all $v \in V$. The latter constraint can be expressed as $\sum_{j=1}^m a_j \theta_j(x) = \theta(x)$, and thus we just need to check that $\sum_{j=1}^m |a_j^*(x)| \le \|\theta(x)\|$, which is easily obtained from

$$\begin{aligned} \sum_{j=1}^{m} |a_j^*(x)| &\le \sum_{j=1}^{m} \|\theta(x)\| \left(\psi_j \left(\frac{\theta(x)}{\|\theta(x)\|} \right) + \psi_{m+j} \left(\frac{\theta(x)}{\|\theta(x)\|} \right) \right) = \|\theta(x)\| \sum_{j=1}^{2m} \psi_j \left(\frac{\theta(x)}{\|\theta(x)\|} \right) \\ &= \|\theta(x)\|. \end{aligned} \tag{4.31}$$

Case 2: Some of the $\theta_1, \ldots, \theta_m$ coincide. Let there be exactly k distinct θ_j, and by reindexing if necessary, we can assume that $\theta_1, \ldots, \theta_k$ are distinct. We consider the continuous functions $a_1^*, \ldots, a_k^*$ associated with $\theta_1, \ldots, \theta_k$ as defined in Case 1. For each $j = 1, \ldots, k$, we denote by

$$I(j) := \{i : \ \theta_i = \theta_j\}, \tag{4.32}$$

and we consider nonegative continuous functions $g_i^{(j)}$, $i \in I(j)$, defined on D and satisfying

$$g_i^{(j)}(x_{i'}) = \delta_{i,i'}, \quad i, i' \in I(j), \qquad \text{and} \qquad \sum_{i \in I(j)} g_i^{(j)} \equiv 1. \tag{4.33}$$

For each $i = 1, \ldots, m$, we denote by $j(i)$ the index $j = 1, \ldots, k$ such that $\theta_j = \theta_i$, and we introduce

$$\tilde{a}_i(x) := g_i^{(j(i))}(x) a_{j(i)}^*(x), \qquad x \in D. \tag{4.34}$$

Clearly, $\tilde{a}_1, \ldots, \tilde{a}_m$ are continuous, so we only need to check that $\tilde{a}(x) = (\tilde{a}_1(x), \ldots, \tilde{a}_m(x)) \in \Lambda_x$. Suppose first that $x = x_{i'}$ for some $i' = 1, \ldots, m$. Notice that (4.29) yields $a_j^*(x_{i'}) = a_j^*(x_{j(i')})$ for all $j = 1, \ldots, m$. It follows that, for each $i = 1, \ldots, m$,

$$\tilde{a}_i(x_{i'}) = g_i^{(j(i))}(x_{i'}) a_{j(i)}^*(x_{j(i')}) = \begin{cases} g_i^{(j(i))}(x_{i'}) \times 0 = 0, & \text{if } j(i) \neq j(i'), \\ g_i^{(j)}(x_{i'}) \times 1 = \delta_{i,i'}, & \text{if } j(i) = j(i') =: j. \end{cases} \tag{4.35}$$

This shows that $\tilde{a}_i(x_{i'}) = \delta_{i,i'}$, i.e., that $\tilde{a}(x_{i'}) = e_{i'} \in \Lambda_{x_{i'}}$. Turning to the case $x \notin \{x_1, \ldots, x_m\}$, we start by verifying that $\sum_{i=1}^m \tilde{a}_i(x)\theta_i = \theta(x)$, which is seen from

$$\sum_{i=1}^{m} \tilde{a}_i(x)\theta_i = \sum_{j=1}^{k} \sum_{i \in I(j)} \tilde{a}_i(x)\theta_j = \sum_{j=1}^{k} \sum_{i \in I(j)} g_i^{(j)}(x) a_j^*(x)\theta_j = \sum_{j=1}^{k} a_j^*(x)\theta_j = \theta(x). \tag{4.36}$$

Finally, we have to prove that

$$\sum_{i=1}^{m} |\tilde{a}_i(x)| \le \sum_{i=1}^{m} |a_i| \qquad \text{whenever} \qquad \sum_{i=1}^{m} a_i\theta_i = \theta(x). \tag{4.37}$$

Since the latter can be written as

$$\theta(x) = \sum_{i=1}^{m} a_i\theta_i = \sum_{j=1}^{k} \Big(\sum_{i\in I(j)} a_i \Big)\theta_j, \tag{4.38}$$

the minimal property of $(a_1^*, \ldots, a_k^*)$ implies that

$$\sum_{j=1}^{k} |a_j^*(x)| \le \sum_{j=1}^{k} \Big| \sum_{i\in I(j)} a_i \Big| \le \sum_{j=1}^{k} \sum_{i\in I(j)} |a_i| = \sum_{i=1}^{m} |a_i|, \tag{4.39}$$

while we also have

$$\sum_{i=1}^{m} |\tilde{a}_i(x)| = \sum_{j=1}^{k} \sum_{i\in I(j)} |\tilde{a}_i(x)| = \sum_{j=1}^{k} \sum_{i\in I(j)} g_i^{(j)}(x)|a_j^*(x)| = \sum_{j=1}^{k} |a_j^*(x)|. \tag{4.40}$$

Combining (4.40) and (4.39) yields (4.37). The proof is now complete. □

One can drop the assumption that the space V contains the constant functions from Theorem 4.2 provided we do not insist on pointwise optimality. This is formally stated in the following result.

Theorem 4.3. *Let V be an arbitrary n-dimensional subspace of $C(D)$. Then there is a linear mapping $A^* : \mathbb{R}^m \to C(D)$ which is globally optimal for the recovery of any $f \in C(D)$ from the measurements $M(f) = (f(x_1), \ldots, f(x_m))$, in the sense that*

$$\|f - A^*(M(f))\|_{C(D)} \le \mu(\mathcal{N}, V)_{C(D)} \operatorname{dist}(f, V)_{C(D)} \qquad \text{for all } f \in C(D). \tag{4.41}$$

Proof: We use the same notation as in the proof of Theorem 4.2. Let us also denote by k the number of extreme points of $\mathcal{C}$ among $\theta_1, \ldots, \theta_m$, which we may assume to be $\theta_1, \ldots, \theta_k$. Since automatically $\|\theta_1\| = \cdots = \|\theta_k\| = 1$, we do not need the condition $1 \in V$ to construct an optimal algorithm $A^* : w \in \mathbb{R}^k \mapsto \sum_{j=1}^{k} w_j a_j^* \in C(D)$ based on the linear functionals $\delta_{x_1}, \ldots, \delta_{x_k}$. We keep the same notation for the algorithm defined on $\mathbb{R}^m$ by $A^*(w) = \sum_{j=1}^{k} w_j a_j^*$, $w \in \mathbb{R}^m$. We have

$$\|f - A^*(M(f))\|_{C(D)} \le \mu(\mathcal{N}', V)_{C(D)} \operatorname{dist}(f, V)_{C(D)}, \qquad f \in C(D), \tag{4.42}$$

where $\mathcal{N}' := \{\eta \in C(D) : \eta(x_1) = \cdots = \eta(x_k) = 0\}$. To finish the proof, we have to show that $\mu(\mathcal{N}', V)_{C(D)} = \mu(\mathcal{N}, V)_{C(D)}$, which by (4.2) is equivalent to

$$\max_{v\in V} \frac{\|v\|_{C(D)}}{\max\limits_{1\le j\le k} |v(x_j)|} = \max_{v\in V} \frac{\|v\|_{C(D)}}{\max\limits_{1\le i\le m} |v(x_i)|}. \tag{4.43}$$

The latter will follow from $\max_{1\le i\le m} |v(x_i)| = \max_{1\le j\le k} |v(x_j)|$ for any $v \in V$, and in turn from $|v(x_i)| \le \max_{1\le j\le k} |v(x_j)|$ for any $v \in V$ and any $i = 1, \ldots, m$. To see this, we notice that θ_i, as a convex combination of $\theta_1, \ldots, \theta_k$, can be written as

$$\theta_i = \sum_{h=1}^{k} \lambda_{i,h}\theta_h, \qquad \text{where } \lambda_{i,h} \ge 0 \quad \text{and} \quad \sum_{h=1}^{k} \lambda_{i,h} = 1. \tag{4.44}$$

This translates into the identity $v(x_i) = \sum_{h=1}^k \lambda_{i,h} v(x_h)$ valid for all $v \in V$, so we have

$$|v(x_i)| \le \sum_{h=1}^k \lambda_{i,h} |v(x_h)| \le \sum_{h=1}^k \lambda_{i,h} \max_{1 \le j \le k} |v(x_j)| = \max_{1 \le j \le k} |v(x_j)|, \tag{4.45}$$

as desired. This completes the proof. □

Remark 4.4. *Theorem* 4.3 *and Theorem* 2.1*, part* (iii) *with $Q = I$, imply that, for a linear subspace V of $C(D)$ and measurements of the type $M(f) = (f(x_1), \dots, f(x_m))$, one has*

$$E^*(\mathcal{K}(\varepsilon, V), M)_{C(D)} = \mu(\mathcal{N}, V)_{C(D)}\, \varepsilon, \qquad \mathcal{K}(\varepsilon, V) = \{f \in C(D) : \ \mathrm{dist}(f, V)_{C(D)} \le \varepsilon\}. \tag{4.46}$$

This statement could also be derived from results in [11], except that the results in [11] say that one can find linear algorithms that are arbritrarily close to optimal, whereas we additionally assert that the infimum defining $E^*(\mathcal{K}(\varepsilon, V), M)_{C(D)}$ is actually a minimum in the case of point evaluations.

Remark 4.5. *The results of [11] formalized an implicit understanding that the performance of linear algorithms for linear problems is closely connected to extensions of linear operators, a topic which has attracted some attention in geometry of Banach spaces (see [29] for a survey). Indeed, the main task can be interpreted as extending the inverse $(M_{|V})^{-1}$ of the restriction of M to V. The memoir [19] certainly influenced our view and in fact Theorem 4.3 could be deduced from general results contained in [19]. However, such a direct approach does not yield the pointwise optimality obtained in Theorem 4.2. Moreover, not relying on abstract results was more suited to a numerical treatment, especially in case of model classes given as approximation sets.*

4.3 Linear near-optimal algorithms

Theorem 4.2 shows that there always exist linear optimal algorithms for the recovery of f from the measurements $M(f) = (f(x_1), \dots, f(x_m))$ in the case $\mathcal{X} = C(D)$. However, it does not give a constructive procedure to create such an algorithm. In this section, we show how to create linear near-optimal algorithms using quasi-interpolants, which we introduce next.

4.3.1 Quasi-interpolants

Given a Banach space $\mathcal{X}$ and a subspace $V \subset \mathcal{X}$, we say that a bounded linear operator $P : \mathcal{X} \to \mathcal{X}$ is a quasi-projector for V if P has a finite-dimensional range and

$$P(v) = v \qquad \text{for all } v \in V. \tag{4.47}$$

When $\mathcal{X} = C(D)$ and the quasi-projectors are built on point evaluations, they shall be referred to as quasi-interpolants. We highlight below the fact that quasi-interpolants for V with a norm arbitrary close to 1 always exist in the case under consideration.

Lemma 4.6. *Let V be an n-dimensional subspace of $C(D)$, let $\xi_1, \dots, \xi_N$ be points in D, and let γ be a positive number. Then*
(i) the following conditions are equivalent:
(a) with $\mathcal{N}_\xi$ denoting the null space of the measurement map $M_\xi : f \mapsto (f(\xi_1), \dots, f(\xi_N))$,

$$\mu(V, \mathcal{N}_\xi)_{C(D)} = \max_{v \in V} \frac{\|v\|_{C(D)}}{\max\limits_{1 \le j \le N} |v(\xi_j)|} \le 1 + \gamma; \tag{4.48}$$

(b) there are functions $u_1, \dots, u_N \in C(D)$ such that the operator $P : C(D) \to C(D)$ defined as

$$P(f) := \sum_{i=1}^{N} f(\xi_i) u_i, \qquad f \in C(D), \tag{4.49}$$

is a quasi-interpolant for V and

$$\|P\|_{C(D)\to C(D)} \le 1 + \gamma; \tag{4.50}$$

(ii) there exist points $\xi_1, \dots, \xi_N \in D$, with $N \le (3 + 2/\gamma)^n$, such that the above conditions hold.

Proof: (i) The equality in (4.48) is just (4.2). If (a) holds, we call upon Theorem 4.3 applied to the space V and the points $\xi_1, \dots, \xi_N$ to produce an optimal algorithm $A^* : w \in \mathbb{R}^N \mapsto \sum_{i=1}^N w_i a_i^*$. Then the operator $P := A^* \circ M_\xi : C(D) \to C(D)$ has the form (4.49) with $u_i = a_i^*$ and satisfies

$$\|f - P(f)\|_{C(D)} \le \mu(\mathcal{N}_\xi, V)_{C(D)} \operatorname{dist}(f, V) \qquad \text{for all } f \in C(D). \tag{4.51}$$

The latter shows that P reproduces V, i.e., (4.47) holds. Therefore, P is a quasi-interpolant for V. Moreover, in view of (4.12) and (4.1),

$$\begin{aligned} \|P\|_{C(D)\to C(D)} &= \sup_{x\in D} \sum_{i=1}^{N} |a_i^*(x)| = \sup_{x\in D} \big(\mu(\mathcal{N}_\xi, V, \delta_x)_{C(D)} - 1\big) \\ &= \mu(\mathcal{N}_\xi, V)_{C(D)} - 1 = \mu(V, \mathcal{N}_\xi)_{C(D)} \le 1 + \gamma, \end{aligned} \tag{4.52}$$

which establishes (4.50), so (b) holds. Conversely, if (b) holds, for $v \in V$ and $x \in D$,

$$\begin{aligned} |v(x)| &= |P(v)(x)| \le \sum_{i=1}^{N} |v(\xi_i)||u_i(x)| \le \max_{1\le i\le N} |v(\xi_i)| \sum_{i=1}^{N} |u_i(x)| \\ &\le (1+\gamma) \max_{1\le i\le N} |v(\xi_i)|, \end{aligned} \tag{4.53}$$

and taking supremum over $x \in D$ and $v \in V$ yields (4.48), so (a) holds.
(ii) We let $\delta := \gamma/(1+\gamma)$. It is known (see e.g [22, Lemma 2.6]) that there exist $v_1, \dots, v_N \in V$, with $N \le (1 + 2/\delta)^n = (3 + 2/\gamma)^n$, such that $\|v_1\|_{C(D)} = \cdots = \|v_N\|_{C(D)} = 1$ and

$$\min_{1\le j\le N} \|v - v_j\|_{C(D)} \le \delta \qquad \text{for all } v \in V \text{ with } \|v\|_{C(D)} = 1. \tag{4.54}$$

For each $j = 1, \dots, N$, we select $\xi_j \in D$ such that $|v_j(\xi_j)| = 1$. Then, for any $v \in V$ with $\|v\|_{C(D)} = 1$, we can choose j such that $\|v - v_j\|_{C(D)} \le \delta$, and thereby obtain

$$|v(\xi_j)| \ge |v_j(\xi_j)| - |(v - v_j)(\xi_j)| \ge 1 - \delta = \frac{1}{1+\gamma}. \tag{4.55}$$

This proves that, for all $v \in V$,

$$\max_{1\le j\le N} |v(\xi_j)|/\|v\|_{C(D)} \ge 1/(1+\gamma). \tag{4.56}$$

and the inequality in (4.48) follows. □

For classical approximation spaces V, explicit quasi-interpolants can be constructed using much fewer points $\xi_1, \ldots, \xi_N$ than estimated in Lemma 4.6. We mention the following examples for spaces of polynomials.

Trigonometric polynomials: Consider the space $V = \mathcal{T}_{n-1}$ of trigonometric polynomials of degree at most $n-1$ defined on $D = [-\pi, \pi]$. This is a linear space of dimension $2n-1$. It is known that if $\xi_1, \ldots, \xi_N$ consist of $N = 4n-2$ equally spaced points, then there exist trigonometric polynomials $u_1, \ldots, u_N$ of degree at most N for which Lemma 4.6 (ii) holds with $\|P\|_{C(D)\to C(D)} \le 2$.

Algebraic polynomials: Consider the space $V = \mathcal{P}_{n-1}$ of algebraic polynomials of degree at most $n-1$ defined on $D = [-1, 1]$. This is a linear space of dimension n. From the above result for trigonometric polynomials, one can deduce that if $\xi_1, \ldots, \xi_N$ consist of the $N = 2n-1$ zeros of the Chebyshev polynomial T_N of degree N, then there exist algebraic polynomials $u_1, \ldots, u_N$ for which Lemma 4.6 (ii) holds with $\|P\|_{C(D)\to C(D)} \le 2$.

4.3.2 Construction of linear near-optimal algorithms from quasi-interpolants

We now show how linear near-optimal algorithms can be constructed in practice, still in the setting where $\mathcal{X} = C(D)$, V is a linear subspace of $\mathcal{X}$ of dimension $n \le m$, and with measurements about $f \in C(D)$ given as $M(f) = (f(x_1), \ldots, f(x_m))$ for some $x_1, \ldots, x_m \in D$. We start with a quasi-interpolant of the form (4.49), i.e.,

$$P(f) = \sum_{i=1}^{N} f(\xi_i) u_i, \qquad f \in C(D), \tag{4.57}$$

which has norm

$$\|P\|_{C(D)\to C(D)} = \left\| \sum_{i=1}^{N} |u_i| \right\|_{C(D)} \le 1 + \gamma. \tag{4.58}$$

For each $i = 1, \ldots, N$, Theorem 3.1 can be applied to the linear functional $q = \delta_{\xi_i}$, hence creating a vector $a^{(i)} = (a_1^{(i)}, \ldots, a_m^{(i)}) \in \mathbb{R}^m$ satisfying

$$\sum_{j=1}^{m} a_j^{(i)} v(x_j) = v(\xi_i) \qquad \text{for all } v \in V, \tag{4.59}$$

as well as

$$\sum_{j=1}^{N} |a_j^{(i)}| \le \mu(V, \mathcal{N})_{C(D)}. \tag{4.60}$$

The latter is obvious if $\xi_i = x_k$ for some $k = 1, \ldots, m$, since then $a^{(i)} = e_k$ and $\sum_{j=1}^{N} |a_j^{(i)}| = 1$, while if $\xi_i \notin \{x_1, \ldots, x_m\}$, we have

$$\sum_{j=1}^{m} |a_j^{(i)}(\xi_i)| = \mu(\mathcal{N}, V, \delta_{\xi_i})_{C(D)} - 1 \le \mu(\mathcal{N}, V)_{C(D)} - 1 = \mu(V, \mathcal{N})_{C(D)}. \tag{4.61}$$

We now introduce the map $A' : \mathbb{R}^m \to C(D)$ defined by

$$A'(w) := \sum_{i=1}^{N} \left(\sum_{j=1}^{m} a_j^{(i)} w_j \right) u_i = \sum_{j=1}^{m} w_j \phi_j, \qquad \phi_j := \sum_{i=1}^{N} a_j^{(i)} u_i. \tag{4.62}$$

Before proving that the linear algorithm A' is near optimal, we observe that, if the quasi-interpolant P is explicitly known, then A' is numerically constructed by solving N constrained ℓ_1-minimization problem (one for each for each $i = 1, \ldots, N$ to produce the vector $(a_1^{(i)}, \ldots, a_m^{(i)})$), so the the number of minimization problems is of order n when V is a space of trigonometric or algebraic polynomials of degree at most $n-1$.

Theorem 4.7. *Let V be an n-dimensional subspace of $C(D)$ and γ be an arbitrary positive number. Then the linear algorithm A' defined in (4.62) is globally near-optimal for the recovery of any $f \in C(D)$ from the measurements $M(f) = (f(x_1), \ldots, f(x_m))$, in the sense that*

$$\|f - A'(M(f))\|_{C(D)} \le (1+\gamma)\, \mu(\mathcal{N}, V)_{C(D)}\ \mathrm{dist}(f, V)_{C(D)} \qquad \text{for all } f \in C(D). \tag{4.63}$$

Proof: From (4.59) and the fact that P is a quasi-interpolant for V, we first observe that, for all $v \in V$,

$$A'(M(v)) = \sum_{i=1}^{N} \left(\sum_{j=1}^{m} a_j^{(i)} v(x_j) \right) u_i = \sum_{i=1}^{N} v(\xi_i) u_i = P(v) = v. \tag{4.64}$$

Next, from (4.60) and (4.58), we also observe that, for all $w \in \mathbb{R}^m$ and all $x \in D$,

$$\begin{aligned} |A'(w)(x)| &\le \sum_{i=1}^{N} \sum_{j=1}^{m} |a_j^{(i)}||w_j||u_i(x)| \le \mu(V,\mathcal{N})_{C(D)} \|w\|_{\ell_\infty^m} \sum_{i=1}^{N} |u_i(x)| \\ &\le (1+\gamma)\mu(V,\mathcal{N})_{C(D)} \|w\|_{\ell_m^\infty}. \end{aligned} \tag{4.65}$$

In other words, we have

$$\|A'\|_{\ell_\infty^m \to C(D)} \le (1+\gamma)\mu(V,\mathcal{N})_{C(D)}. \tag{4.66}$$

Now, for any $f \in C(D)$, choosing $v \in V$ such that $\|f - v\|_{C(D)} = \mathrm{dist}(f, V)_{C(D)}$, we derive that

$$\begin{aligned} \|f - A'(M(f))\|_{C(D)} &= \|f - v - A'(M(f-v))\|_{C(D)} \\ &\le \|f - v\|_{C(D)} + \|A'\|_{\ell_\infty^m \to C(D)} \|M(f-v)\|_{\ell_\infty^m} \\ &\le \big(1 + (1+\gamma)\mu(V,\mathcal{N})_{C(D)}\big)\, \|f - v\|_{C(D)} \\ &\le (1+\gamma)\, \mu(\mathcal{N}, V)_{C(D)}\ \mathrm{dist}(f, V)_{C(D)}. \end{aligned} \tag{4.67}$$

This justifies (4.63) and completes the proof. □

Remark 4.8. *The construction of the linear near-optimal operator $A' : \mathbb{R}^m \to C(D)$ requires explicit knowledge of the points $\xi_1, \ldots, \xi_N$ and the functions $u_1, \ldots, u_N$ in the expression (4.57) of a quasi-interpolant for V. We wish to point out that we could do with only the knowledge of $U := \mathrm{span}\{u_1, \ldots, u_N\}$, at least on a theoretical level. This is based on three simple observations (which incidentally give a fresh insight on the connection of extensions of operators with quasi-interpolants). First, we notice that, for an operator P from $C(D)$ to $C(D)$ and a constant C, the property*

$$\|f - P(f)\|_{C(D)} \le C\, \mu(\mathcal{N}, V)_{C(D)}\ \mathrm{dist}(f, V)_{C(D)} \qquad \text{for all } f \in C(D) \tag{4.68}$$

is readily equivalent to

$$P(v) = v \quad \text{for all } v \in V \qquad \text{and} \qquad \|I - P\|_{C(D) \to C(D)} \le C\, \mu(\mathcal{N}, V)_{C(D)}. \tag{4.69}$$

Second, we remark that the composition $A \circ M$ of the measurement map $M : C(D) \to \mathbb{R}^m$ and a linear recovery algorithm $A : \mathbb{R}^m \to C(D)$ is an operator from $C(D)$ to $C(D)$ of the form

$$P(f) = \sum_{j=1}^{m} f(x_j)\phi_j \qquad \text{for some } \phi_1, \dots, \phi_m \in C(D). \tag{4.70}$$

Third, we point out the easy-to-verify fact that such an operator satisfies

$$\|I - P\|_{C(D)\to C(D)} = 1 + \left\| \sum_{j=1}^{m} |\phi_j| \right\|_{C(D)} = 1 + \|P\|_{C(D)\to C(D)}. \tag{4.71}$$

Therefore, when U is the range of a quasi-interpolant for V with bounded norm, a linear near-optimal algorithm $A : \mathbb{R}^m \to C(D)$ with range equal to U can be obtained as a solution of

$$\underset{\phi_1,\dots,\phi_m \in U}{\text{minimize}} \left\| \sum_{j=1}^{m} |\phi_j| \right\|_{C(D)} \qquad \text{subject to} \quad \sum_{j=1}^{m} v(x_j)\phi_j = v \quad \text{for all } v \in V. \tag{4.72}$$

This problem may be hard to solve in practice, but we will see in §5.3 that it is possible to solve an ersatz problem and still produce a linear near-optimal algorithm.

4.4 Full approximation from equispaced point evaluations

This final subsection is dedicated to a concrete example where our abstract theory applies. We suppose here that $\mathcal{X} = C[-1, 1]$, that V is the linear space $\mathcal{P}_{n-1}$ of algebraic polynomials of degree at most $n - 1$, and that the measurement map M is given by point evaluations at the equispaced points $-1 = \tau_1 < \cdots < \tau_m = 1$. The problem of approximating $f \in C[-1, 1]$ using the particular measurements $M(f)$ is well studied, so our goal here is solely to indicate how our theory sheds new light on existing results.

Let us recall that the performance of an optimal recovery algorithm for the approximation set $\mathcal{K}(\varepsilon, \mathcal{P}_{n-1})_{C[-1,1]}$ is governed by the quantity

$$\mu_{m,n} := \mu(\mathcal{N}, \mathcal{P}_{n-1})_{C[-1,1]} = \sup_{\eta(\tau_1)=\cdots=\eta(\tau_m)=0} \frac{\|\eta\|_{C[-1,1]}}{\operatorname{dist}(\eta, \mathcal{P}_{n-1})_{C[-1,1]}}, \tag{4.73}$$

and that we have the relation (see **R1**)

$$\mu_{m,n} = 1 + \tilde{\mu}_{m,n}, \qquad \text{where} \quad \tilde{\mu}_{m,n} := \max_{p \in \mathcal{P}_{n-1}} \frac{\|p\|_{C[-1,1]}}{\max\limits_{1 \le j \le m} |p(\tau_j)|}. \tag{4.74}$$

We work in the setting where the number m of measurements is given to us and we are free to choose the dimension of the space V, that is the integer $n \le m$. Perhaps counterintuitively, it may not be the best idea to choose $n = m$, because then $\tilde{\mu}_{m,m}$ reduces to the Lebesgue constant for polynomial interpolation at m equispaced points, which is known to behave like $2^m/(em \ln(m))$ (see [27, 25]). For arbitrary $n \le m$, it was proved in [10] that there are absolute constants $\beta \ge \alpha > 1$ such that

$$\alpha^{\frac{n^2}{m}} \le \tilde{\mu}_{m,n} \le \beta^{\frac{n^2}{m}}, \qquad 0 \le n \le m. \tag{4.75}$$

Thus, it may be a better idea to chose n of the order of $\sqrt{m}$, because $\tilde{\mu}_{m,n}$ would then be bounded by an absolute constant.

Let us make this heuristic precise in the context of approximation classes. Previous studies primarily considered model classes which are finite balls in the spaces $\mathcal{B}_\rho$, $\rho > 1$, of functions that are analytic and bounded on the interior of the Bernstein ellipse $\mathcal{E}_\rho$, with norm

$$\|f\|_{\mathcal{B}_\rho} := \sup_{z \in \mathcal{B}_\rho} |f(z)|. \tag{4.76}$$

This Bernstein ellipse has focii ± 1 and its major and minor axes have lengths summing to ρ. It is well known that the functions $f \in \mathcal{B}_\rho$ are characterized by their error of polynomial approximation. Namely, we have

$$\operatorname{dist}(f, \mathcal{P}_n)_{C[-1,1]} \leq \|f\|_{\mathcal{B}_\rho} \rho^{-n} \qquad \text{for all } n \geq 0, \tag{4.77}$$

and conversly any function satisfying $\operatorname{dist}(f, \mathcal{P}_n)_{C[-1,1]} \leq C\rho^{-n}$ for all $n \geq 0$ is in the class $\mathcal{B}_\rho$ and $\|f\|_{B_\rho} \leq C$. It follows that the unit ball $U(\mathcal{B}_\rho)$ is equivalent to the approximation class $\mathcal{K}((\varepsilon_n), (\mathcal{P}_{n-1}))_{C[-1,1]}$ where $\varepsilon_n = \rho^{-n}$, $n \geq 1$. In particular, $f \in \mathcal{B}_\rho$ implies that f is in the approximation set $\mathcal{K}(\varepsilon_{n-1}, \mathcal{P}_{n-1})$ with $\varepsilon_{n-1} = \|f\|_{\mathcal{B}_\rho} \rho^{-n+1}$ for each $n \geq 1$.

An important and extensively studied question is whether one can stably recover $f \in \mathcal{B}_\rho$ from equispaced point values. It is shown in [24] that it is in fact impossible to find a stable recovery procedure while maintaining an approximation accuracy decaying exponentially with m. More precisely, for $\gamma > 1$, if $A_m : \mathbb{R}^m \to C[-1,1]$ are mappings yielding the estimates

$$\|f - A_m(f(\tau_1), \ldots, f(\tau_m))\|_{C[-1,1]} \leq C\gamma^{-m}, \qquad m \geq 0, \tag{4.78}$$

for all $f \in U(\mathcal{B}_\rho)$, then necessarily the condition numbers of A_m must grow exponentially as κ^m for some $\kappa > 1$. This statement remains valid if γ^{-m} and κ^m are replaced by γ^{-m^τ} and $\kappa^{m^{2\tau-1}}$ for any $\tau > 1/2$.

The results of the present paper, applied to the case $f \in \mathcal{B}_\rho$, bring forward the following observations.

Theorem 4.9. *Let $\rho > 1$ be fixed and $-1 = \tau_1 < \cdots < \tau_m = 1$ be the m equispaced points on $[-1,1]$.*
(i) *There are linear mappings $A_m : \mathbb{R}^m \to C[-1,1]$ such that, for all $f \in \mathcal{B}_\rho$,*

$$\|f - A_m(f(\tau_1), \ldots, f(\tau_m))\|_{C[-1,1]} \leq C\|f\|_{\mathcal{B}_\rho}\gamma^{-m}, \qquad m \geq m_*, \tag{4.79}$$

where $C > 0$ is an absolute constant, and where $\gamma > 1$ and m_* depend on ρ.
(ii) *There are stable linear mappings $A_m : \mathbb{R}^m \to C[-1,1]$ such that, for all $f \in \mathcal{B}_\rho$,*

$$\|f - A_m(f(\tau_1), \ldots, f(\tau_m))\|_{C[-1,1]} \leq C\|f\|_{\mathcal{B}_\rho}\rho^{-\sqrt{m}}, \qquad m \geq 0, \tag{4.80}$$

where $C > 0$ is an absolute constant.

Proof: According to Theorem 4.7, for all integers $m \geq n \geq 1$, we can construct a linear operator $A'_{m,n} : \mathbb{R}^m \to C[-1,1]$ such that, for all $f \in C[-1,1]$,

$$\|f - A'_{m,n}(f(\tau_1), \ldots, f(\tau_m))\|_{C[-1,1]} \leq C\tilde{\mu}_{m,n} \operatorname{dist}(f, \mathcal{P}_{n-1})_{C[-1,1]}. \tag{4.81}$$

We use (4.75) and (4.77) to deduce that, for all $f \in \mathcal{B}_\rho$,

$$\|f - A'_{m,n}(f(\tau_1), \ldots, f(\tau_m))\|_{C[-1,1]} \leq C\|f\|_{\mathcal{B}_\rho}\beta^{\frac{n^2}{m}}\rho^{-n}. \tag{4.82}$$

To obtain (i), we denote by $\tilde{\beta} := \max\{\beta, \rho\} > 1$ and $A_m := A'_{m,n}$, where n is chosen such that

$$n := \left\lceil \frac{\ln(\rho)}{2\ln(\tilde{\beta})} m \right\rceil \leq m \qquad \text{and therefore satisfies} \quad \frac{n}{m} \in \left[\frac{\ln(\rho)}{2\ln(\tilde{\beta})}, \frac{\ln(\rho)}{2\ln(\tilde{\beta})} + \frac{1}{m} \right). \tag{4.83}$$

Since the function $t \mapsto \ln(\tilde{\beta})t^2 - \ln(\rho)t$ increases for $t \geq \ln(\rho)/(2\ln(\tilde{\beta}))$, we derive that

$$\begin{aligned} \beta^{\frac{n^2}{m}} \rho^{-n} &\leq \left(\tilde{\beta}^{\frac{n^2}{m^2}} \rho^{-\frac{n}{m}} \right)^m = \left(e^{\left(\ln(\tilde{\beta}) \left(\frac{n}{m} \right)^2 - \ln(\rho) \left(\frac{n}{m} \right) \right)} \right)^m \\ &\leq \left(e^{\left(\ln(\tilde{\beta}) \left(\frac{\ln(\rho)}{2\ln(\tilde{\beta})} + \frac{1}{m} \right)^2 - \ln(\rho) \left(\frac{\ln(\rho)}{2\ln(\tilde{\beta})} + \frac{1}{m} \right) \right)} \right)^m = \left(e^{\left(-\frac{\ln(\rho)^2}{4\ln(\tilde{\beta})} + \frac{\ln(\tilde{\beta})}{m^2} \right)} \right)^m \\ &\leq \gamma^{-m}, \end{aligned} \tag{4.84}$$

where $\gamma := e^{\left(\ln(\rho)^2 / (20 \ln(\tilde{\beta})) \right)}$ and $m \geq m_* := \lceil \sqrt{5} \ln(\tilde{\beta}) / \ln(\rho) \rceil$.

To obtain (ii), we choose $A_m = A'_{m,n}$, where $n = \lceil \sqrt{m} \rceil \leq 2\sqrt{m}$, so that $\beta^{\frac{n^2}{m}} \rho^{-n} \leq \beta^4 \rho^{-\sqrt{m}}$. The stability of A_m is a consequence of (4.66), since $\|A_m\|_{\ell_\infty^m \to C[-1,1]} \leq C\tilde{\mu}_{m,n} \leq C\beta^4$. □

Let us comment on the significance of the two results from Theorem 4.9. In a way, (i) could have been obtained even for the interpolation operator $I_m : \mathbb{R}^m \to \mathcal{P}_{m-1}$ relative to the m equispaced points $\tau_1, \ldots, \tau_m$, provided ρ is sufficiently large. Indeed, for all $f \in U(\mathcal{B}_\rho)$,

$$\begin{aligned} \|f - I_m(f(\tau_1), \ldots, f(\tau_m))\|_{C[-1,1]} &\leq (1 + \|I_m\|_{C[-1,1] \to C[-1,1]}) \operatorname{dist}(f, \mathcal{P}_{n-1})_{C[-1,1]} \\ &\leq 2^m \rho^{-m}, \end{aligned} \tag{4.85}$$

so $\|f - I_m(f(\tau_1), \ldots, f(\tau_m))\|_{C[-1,1]}$ can be made smaller than γ^{-m} for some $\gamma > 1$, but only if $\rho > 2$. In Theorem 4.9, part (i), we do not have any restriction on $\rho > 1$ and our algorithm is still a linear operator. However, we need a prior knowledge of ρ to make our choice of n in $A_m = A'_{m,n}$. Of course, our A_m's, just like the I_m's, cannot be stable. The approximation procedure proposed in (ii), on the other hand, is stable and still guarantees good approximation performance. Note also that it does not require a prior knowledge of ρ. Approximation procedures with these features are also byproducts of results previously obtained in [14], as a consequence of **R3** in §4.1. Our new contribution is the linearity of the approximation procedures, thanks to the fact that $\mathcal{P}_{n-1}$ admits quasi-interpolants. We believe this result to be novel.

5 Implementation of the linear algorithms

In this last section, we test the linear optimal and near-optimal algorithms developed in this paper and discuss their implementation on three examples. Our MATLAB codes rely on CVX [1], a package for specifying and solving convex programs, and Chebfun [15], a package for computing with functions, and can be downloaded from the second author's webpage.

In all three cases, the measurements $M(f) = (f(x_1), \ldots, f(x_m))$ are point evaluations. For the first two examples, we estimate the quantity of interest $q(f) = \int_D f$ (i.e., we build a quadrature formula) relative to $\mathcal{X} = C(D)$ (see §5.1) or to a reproducing kernel Hilbert space $\mathcal{X}$ of functions defined on D (see §5.2). We let $V = \mathcal{P}_{n-1}$ be the space of algebraic polynomials of degree at most $n-1$, and perform the computations with three bases for V: the monomial basis $v_i(x) = x^{i-1}$,

the Chebyshev basis $v_i(x) = T_{i-1}(x)$, and the Legendre basis with $v_i(x) = P_{i-1}(x)$, $i = 1, \ldots, n$. Our experiments indicate that the result is sensitive to the choice of basis and that choosing the monomial basis is not recommended. For the third example, we execute the full approximation relative to $C(D)$ (see §5.3). We emphasize again the advantage that all of the proposed algorithms perform the approximations by simply computing an inner product between incoming data $(w_1, \ldots, w_m) = (f(x_1), \ldots, f(x_m))$ and a vector (of numbers or of functions) computed once and for all during an offline stage.

5.1 Optimal quadrature relative to $C(D)$

Given m distinct points $x_1, \ldots, x_m \in D$, the task of optimal approximation of the quantity of interest $q(f) = \int_D f$, $f \in C(D)$, using the information $M(f) = (f(x_1), \ldots, f(x_m))$ (see Theorem 3.1), is in fact the task of finding a vector $a^* = (a_1^*, \ldots, a_m^*)$ which solves the problem

$$\underset{a \in \mathbb{R}^m}{\text{minimize}} \sum_{j=1}^{m} |a_j| \qquad \text{subject to } \sum_{j=1}^{m} a_j v(x_j) = \int_D v \quad \text{for all } v \in V. \tag{5.1}$$

Indeed, for $q(f) = \int_D f$, we have

$$\left\| q - \sum_{j=1}^{m} a_j \delta_{x_j} \right\|_{C(D)^*} = \text{meas}(D) + \sum_{j=1}^{m} |a_j|, \tag{5.2}$$

because we can take $f \in C(D)$ with $\|f\|_{C(D)} = 1$, $f(x_i) = -\operatorname{sign} a_j$, $j = 1, \ldots, m$, and $\int_D f$ as close to $\text{meas}(D)$ as we wish.

Once a^* is constructed, the quadrature becomes

$$\int_D f \approx \sum_{j=1}^{m} a_j^* f(x_j), \tag{5.3}$$

where the latter formula is exact for all elements of the space V. We solve (5.1) as a linear program by introducing slack variables $s_1, \ldots, s_m$ and rewriting it as

$$\underset{a, s \in \mathbb{R}^m}{\text{minimize}} \sum_{j=1}^{m} s_j \qquad \text{subject to } Ba = c \quad \text{and} \quad -a \le s \le a, \tag{5.4}$$

where

$$B = (b_{i,j}), \quad b_{i,j} = v_i(x_j),\ i = 1, \ldots, n,\ j = 1, \ldots, m, \quad \text{and } c_i = \int_D v_i,\ i = 1, \ldots, n, \tag{5.5}$$

where $(v_1, \ldots, v_n)$ is a fixed basis for the space V.

We point out that a minimizer of (5.1) is not necessarily unique. We have witnessed some dependence on the choice of the basis $(v_1, \ldots, v_n)$ and on the solver selected to perform the minimization. In particular, we have observed that not all minimization algorithms produce an n-sparse solutions,

i.e., a solution with at most n-nonzero entries, even though it is known that there exists an n-sparse minimizer of (5.1) (see e.g. [16, Section 1.4]). Note that the existence of an n-sparse weight vector a^* in a quadrature formula (5.3), exact for all elements in the n-dimensional space V, seems to have been noticed earlier only for $V = \mathcal{P}_{n-1}$ and for the equispaced points $\tau_1, \ldots, \tau_m$ of $D = [-1, 1]$ under the additional assumption that there is a quadrature formula based on $\tau_1, \ldots, \tau_m$ which is exact on $\mathcal{P}_{n-1}$ and involves nonnegative weights (see [28]). In that paper, necessary conditions and sufficient conditions for the existence of such formulas were derived. Essentially, [28] showed that, if $m_*(n)$ is the minimal value of m for which the equispaced points $\tau_1, \ldots, \tau_m$ of $[-1, 1]$ support a nonnegative quadrature formula exact on $\mathcal{P}_{n-1}$, then

$$\left(\frac{n+1}{2\pi}\right)^2 + 1 \leq m_*(n) \leq n^2. \tag{5.6}$$

Let us observe that the existence of nonnegative quadrature formula based on $\tau_1, \ldots, \tau_m$ and exact on $\mathcal{P}_{n-1}$ is equivalent to the fact that the value of the minimum in (5.1) equals 2. This follows from the fact that a vector $a \in \mathbb{R}^m$ meeting the constraint $\sum_{j=1}^m a_i v(\tau_j) = \int_{-1}^{1} v$ for all $v \in \mathcal{P}_{n-1}$ satisfies

$$\sum_{j=1}^m |a_j| \geq \sum_{j=1}^m a_j = \int_{-1}^{1} 1 = 2, \tag{5.7}$$

with equality if and only if a is a nonnegative vector. We have computed solutions of (5.1) for $n = 1, \ldots, 160$ and $m = n, \ldots, 10,000$ (and made them available alongside our MATLAB reproducible file), so we can determine the exact value of $m_*(n)$ for $n = 1, \ldots, 160$. The result is displayed in Figure 5.1 and compared to the lower and upper bounds given in (5.6).

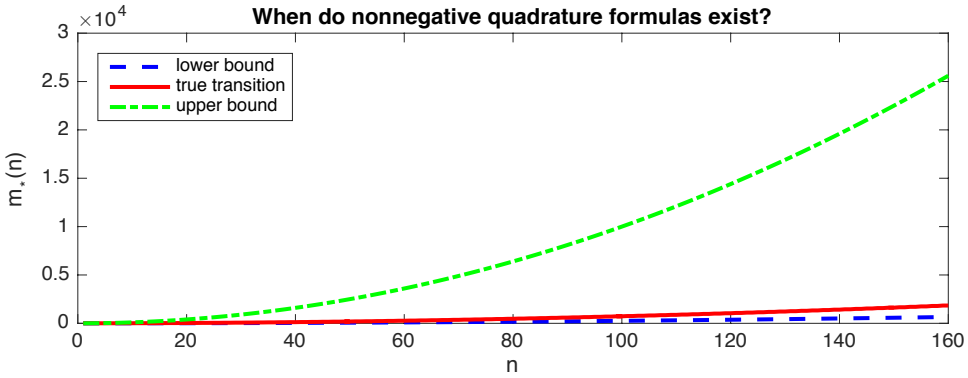


Figure 5.1: Smallest integer m such that the m equispaced points of $[-1, 1]$ support a nonnegative quadrature formula which is exact on the space of algebraic polynomials of degree $\leq n - 1$.

Remark 5.1. *If we select a solution $a^* \in \mathbb{R}^m$ of (5.4) which is n-sparse, then the data at x_j, $j \notin \operatorname{supp}(a^*)$, are not involved in the optimal quadrature. More than the sparsity, it is in fact known that we can select a solution $a^* \in \mathbb{R}^m$ of (5.4) for which the columns of B indexed by $\operatorname{supp}(a^*)$ are linearly independent. Therefore, for any $w \in \mathbb{R}^m$, we can find $v^* \in V$ such that $v^*(x_j) = w_j$ for all $j \in \operatorname{supp}(a^*)$, and thus*

$$A^*(w) = \sum_{j=1}^m a_j^* w_j = \sum_{j=1}^m a_j^* v^*(x_j) = \int_{-1}^1 v^*, \tag{5.8}$$

which shows that the optimal integration algorithm for the set $\mathcal{K}_w(\varepsilon, V)_{C[-1,1]}$, $\varepsilon > 0$, can be written as $\int_{-1}^1 v^*$ for some $v^* \in \mathcal{K}_w(\varepsilon, V)_{C[-1,1]}$.

Remark 5.2. *Another important feature of our quadrature is its stability. More precisely, in the presence of noise $\nu = (\nu_1, \dots, \nu_m)$ in the collected measurement $M(f) + \nu$, one wants to build a quadrature $\int_D f \approx \sum_{j=1}^m a_j f(x_j)$ for which the quantity*

$$\Delta(a_1, \dots, a_m) := \sup_{\nu \in \mathbb{R}^m} \left| \sum_{j=1}^m a_j (f(x_j) + \nu_j) - \sum_{j=1}^m a_j f(x_j) \right| \Big/ \|\nu\|_{\ell_\infty^m} \tag{5.9}$$

is minimal. Since

$$\Delta(a_1, \dots, a_m) = \sup_{\|\nu\|_{\ell_\infty^m} \le 1} \left| \sum_{j=1}^m a_j \nu_j \right| = \sum_{j=1}^m |a_j|, \tag{5.10}$$

the quadrature formula (5.3) yields a minimal Δ, and therefore is the most stable.

5.2 Optimal quadrature relative to a reproducing kernel Hilbert space

In this subsection, we again apply our theory to recover the quantity of interest $q(f) = \int_D f$ using point evaluations at given distinct points $x_1, \dots, x_m \in D$, hence building a quadrature formula

$$\int_D f \approx \sum_{j=1}^m a_j f(x_j), \tag{5.11}$$

The difference is that the underlying space $\mathcal{X}$ is no longer assumed to be $C(D)$, but rather a reproducing kernel Hilbert space, i.e., a Hilbert space of functions defined on the domain D for which point evaluations are continuous linear functionals. Let $\langle \cdot, \cdot \rangle$ denote the inner product on $\mathcal{X}$ and let us recall that there is a kernel $K(\cdot, \cdot)$ defined on $D \times D$ such that

$$f(x) = \langle f, K(\cdot, x) \rangle \qquad \text{for all } x \in D. \tag{5.12}$$

In particular, $f(x_j) = \langle f, K(\cdot, x_j) \rangle$ and the quantity of interest can be expressed as

$$q(f) = \int_D \langle f, K(\cdot, x) \rangle dx = \langle f, h \rangle, \qquad \text{where} \quad h := \int_D K(\cdot, x) dx. \tag{5.13}$$

In view of the identity

$$\begin{aligned}\left\| q - \sum_{j=1}^{m} a_j \delta_{x_j} \right\|_{\mathcal{X}^*} &= \sup_{\|f\|_{\mathcal{X}}=1} \left| \int_D f - \sum_{j=1}^{m} a_j f(x_j) \right| = \sup_{\|f\|_{\mathcal{X}}=1} \left| \langle f, h - \sum_{j=1}^{m} a_j K(\cdot, x_j) \rangle \right| \\ &= \left\| h - \sum_{j=1}^{m} a_j K(\cdot, x_j) \right\|_{\mathcal{X}}, \end{aligned} \tag{5.14}$$

an optimal quadrature using the measurements $M(f) = (f(x_1), \ldots, f(x_m))$ (see Theorem 3.1) is

$$\int_D f \approx \sum_{j=1}^{m} a_j^* f(x_j), \tag{5.15}$$

where the weights $a^* = (a_1^*, \ldots, a_m^*)$ are a solution of

$$\underset{a \in \mathbb{R}^m}{\text{minimize}} \left\| h - \sum_{j=1}^{m} a_j K(\cdot, x_j) \right\| \quad \text{subject to } \sum_{j=1}^{m} a_j \langle v, K(\cdot, x_j) \rangle = \langle v, h \rangle \quad \text{for all } v \in V. \tag{5.16}$$

This convex optimization problem could be solved using general purpose software, but we prefer to present a more explicit procedure. We define the positive definite matrix $P = (p_{i,j}) \in \mathbb{R}^{m \times m}$ by

$$p_{i,j} := K(x_i, x_j), \qquad i, j = 1, \ldots, m, \tag{5.17}$$

and the vector $y \in \mathbb{R}^m$ by

$$y = P^{-1/2} (h(x_1), \ldots, h(x_m))^\top. \tag{5.18}$$

We next express the square of the objective function as

$$\begin{aligned}\left\| h - \sum_{j=1}^{m} a_j K(\cdot, x_j) \right\|^2 &= \|h\|^2 - 2 \sum_{j=1}^{m} a_j \langle h, K(\cdot, x_j) \rangle + \sum_{j,\ell=1}^{m} a_j a_\ell \langle K(\cdot, x_j), K(\cdot, x_\ell) \rangle \\ &= \|h\|^2 - 2 \sum_{j=1}^{m} a_j h(x_j) + \sum_{j,\ell=1}^{m} a_j a_\ell K(x_j, x_\ell) \\ &= \|h\|^2 - 2 \langle a, P^{1/2} y \rangle + \langle P a, a \rangle \\ &= \|h\|^2 - 2 \langle P^{1/2} a, y \rangle + \langle P^{1/2} a, P^{1/2} a \rangle \\ &= \|h\|^2 - \|y\|_{\ell_2^m}^2 + \|y - P^{1/2} a\|_{\ell_2^m}^2. \end{aligned} \tag{5.19}$$

Then the minimization problem (5.16) can be replaced by

$$\underset{a \in \mathbb{R}^m}{\text{minimize}} \, \|y - P^{1/2} a\|_{\ell_2^m} \qquad \text{subject to } Ba = c, \tag{5.20}$$

where the matrix $B = (b_{i,j}) \in \mathbb{R}^{n \times m}$ and the vector $c \in \mathbb{R}^n$ are as in (5.5). This equality-constrained least squares problem can be recast as the standard least squares problem

$$\underset{z \in \mathbb{R}^{m-n}}{\text{minimize}} \, \|y - P^{1/2} \bar{a} - P^{1/2} Q_2 z\|_{\ell_2^m}, \tag{5.21}$$

where $\bar{a} \in \mathbb{R}^m$ is a fixed vector satisfying $B\bar{a} = c$ and $Q_2 \in \mathbb{R}^{m\times(m-n)}$ is the matrix coming from the QR factorization of the transpose of B written as

$$B^\top = \begin{bmatrix} Q_1 \mid Q_2 \end{bmatrix} \begin{bmatrix} R \\ \hline 0 \end{bmatrix}, \qquad Q_1 \in \mathbb{R}^{m\times n},\ Q_2 \in \mathbb{R}^{m\times(m-n)}. \tag{5.22}$$

The solution a^* of (5.16) is then given by $a^* = \bar{a} + Q_2 z^*$, where z^* is the solution of (5.21).

The method just described can be applied to any reproducing kernel Hilbert space $\mathcal{X}$ given via its kernel. An example considered in the reproducible file for such a space is the Sobolev space anchored at 0, which consists of all absolutely continuous functions defined on $[0,1]$ with square integrable first derivative, equipped with the inner product

$$\langle f, g\rangle = f(0)g(0) + \int_0^1 f'(t)g'(t)dt, \qquad f, g \in \mathcal{X}. \tag{5.23}$$

It is easily seen that it admits the kernel $K(t,x) = 2 + \min\{t,x\}$, $t, x \in [0,1]$, and that the function h is given by $h(t) = 7/2 + t - t^2/2$, $t \in [0,1]$.

5.3 Near-optimal recovery of functions in $C[-1,1]$

In §4.3.2, we have uncovered a practical construction of a linear near-optimal algorithm for the full approximation of functions $f \in \mathcal{K}(\varepsilon, V)_{C(D)}$ in case of a linear space V and of measurements $M(f) = (f(x_1), \ldots, f(x_m))$, see Theorem 4.7. The construction required an explicit knowledge of a bounded quasi-interpolant for V and, in terms of computational cost, the algorithm involves N (with $N \asymp n$ if $V = \mathcal{P}_{n-1}$ or $V = \mathcal{T}_{n-1}$) ℓ_1-minimization problems with m variables and n equality constraints, each of which is recast as a linear program with $2m$ variables, n equality constraints, and $2m$ inequality constraints.

Based on Remark 4.8, we consider here a slightly different procedure requiring only the explicit knowledge of the range U of a bounded quasi-interpolant P and of an approximation of the norm on U, i.e., the knowledge of points $\zeta_1, \ldots, \zeta_K$ such that

$$\mu(U, \mathcal{N}_\zeta)_{C(D)} = \max_{u\in U} \frac{\|u\|_{C(D)}}{\max\limits_{1\le k\le K} |u(\zeta_k)|} \le c \tag{5.24}$$

for some absolute constant $c \ge 1$. This procedure involves just one linear program with $\asymp mn$ variables, $\asymp n^2$ equality constraints, and $\asymp mn$ inequality constraints, see (5.27) below. Since a solution of (4.72) might not be computable in practice, we solve instead the problem

$$\underset{\varphi_1,\ldots,\varphi_m\in U}{\text{minimize}} \max_{1\le k\le K} \sum_{j=1}^m |\varphi_j(\zeta_k)| \qquad \text{subject to } \sum_{j=1}^m v(x_j)\varphi_j = v \quad \text{for all } v \in V. \tag{5.25}$$

This will still provide a linear near-optimal algorithm, by virtue of

$$\begin{aligned}\left\|\sum_{j=1}^m |\varphi_j|\right\|_{C(D)} &= \max_{\theta\in\{\pm1\}^m}\left\|\sum_{j=1}^m \theta_j\varphi_j\right\|_{C(D)} \le c\max_{\theta\in\{\pm1\}^m}\max_{1\le k\le K}\left|\sum_{j=1}^m \theta_j\varphi_j(\zeta_k)\right| \\ &= c\max_{1\le k\le K}\sum_{j=1}^m |\varphi_j(\zeta_k)| \le c\max_{1\le k\le K}\sum_{j=1}^m |\phi_j(\zeta_k)| \\ &\le c\left\|\sum_{j=1}^m |\phi_j|\right\|_{C(D)}, \end{aligned} \tag{5.26}$$

where $\phi_1,\ldots.\phi_m$ are functions from Remark 4.8.

Now, given a basis $(u_1,\ldots,u_N)$ for U such that $(u_1,\ldots,u_n)$ is a basis for V, we consider the matrix $A\in\mathbb{R}^{m\times N}$ containing the coefficients of the φ_i with respect to the basis $(u_1,\ldots,u_N)$, i.e., $\varphi_i=\sum_{\ell=1}^N a_{i,\ell}u_\ell$. The optimization problem (5.25) can be recast as a linear program by introducing slack variables $d_{j,k}$ and d such that $|\varphi_j(\zeta_k)|\le d_{j,k}$ for all $j=1,\ldots,m$ and $k=1,\ldots,K$ and that $\sum_{j=1}^m d_{j,k}\le d$ for all $k=1,\ldots,K$. Precisely, with $B\in\mathbb{R}^{n\times m}$ still representing the matrix with entries $b_{i,j}=u_i(x_j)$ and with $Z\in\mathbb{R}^{N\times K}$ representing the matrix with entries $z_{\ell,k}=u_\ell(\zeta_k)$, the minimization problem (5.25) is equivalent to the linear program

$$\begin{aligned}\underset{A\in\mathbb{R}^{m\times N},D\in\mathbb{R}^{m\times K},d\in\mathbb{R}}{\text{minimize}}\quad d \qquad \text{subject to}\quad & BA=\left[I_{n\times n}\mid 0_{n\times(N-n)}\right], \\ & -d_{j,k}\le (AZ)_{j,k}\le d_{j,k},\quad j=1,\ldots,m,\ k=1,\ldots,K, \\ & \sum_{j=1}^m d_{j,k}\le d,\quad k=1,\ldots,K. \end{aligned} \tag{5.27}$$

For illustration, Figure 5.2 displays the plots of the functions $\varphi_1,\ldots,\varphi_m$ created by solving (5.27) for equispaced points with $m=12$, $n=5$, $N=25$, and $K=50$. It is compared with the plots of the functions $a_1^*,\ldots,a_m^*$ resulting from Theorem 4.2 (in fact, of their discretizations on a fine grid).

6 Appendix

Finally, we provide full justifications for two unproven results we have relied on, namely (4.1) and Lemma 4.1. We start by restating (4.1).

Lemma 6.1. *Let V be an n-dimensional subspace of $C(D)$ and $x_1,\ldots,x_m\in D$ be m distinct points in D. If $\mathcal{N}:=\{\eta\in C(D):\eta(x_1)=\cdots=\eta(x_m)=0\}$, then*

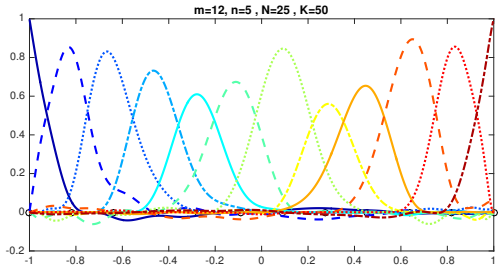
$$\mu(\mathcal{N},V)_{C(D)}=1+\mu(V,\mathcal{N})_{C(D)}. \tag{6.1}$$

Proof: In the view of (1.11), it is enough to establish that

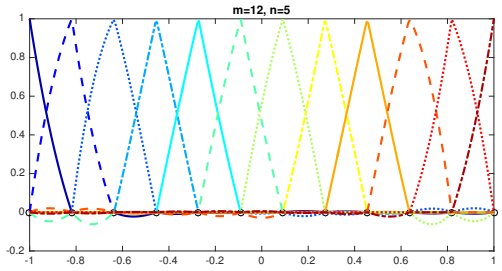
$$\mu(\mathcal{N},V)_{C(D)}\ge 1+\mu(V,\mathcal{N})_{C(D)}. \tag{6.2}$$

Let us denote by

$$\mu:=\mu(V,\mathcal{N})_{C(D)}=\max_{v\in V}\frac{\|v\|_{C(D)}}{\max\limits_{1\le j\le m}|v(x_j)|}, \tag{6.3}$$



(a) Functions φ_j



(b) Functions a_j^*

Figure 5.2: In case $V = \mathcal{P}_{n-1}$, (a) functions $\varphi_1, \ldots, \varphi_m$ obtained after solving (5.27) and (b) discretizations of functions $a_1^*, \ldots, a_n^*$ obtained by solving (4.11) for each point x from a grid of $[-1, 1]$.

and pick $v \in V$ with $\max_{1 \le j \le m} |v(x_j)| = 1$ and $\|v\|_{C(D)} = \mu$. If $\mu > 1$, choose $x^* \in D$ such that $|v(x^*)| = \mu$ and therefore $x^* \notin \{x_1, \ldots, x_m\}$. If $\mu = 1$, choose $x^* \in D \setminus \{x_1, \ldots, x_m\}$ such that $|v(x^*)| \ge \mu - \delta$ for an arbitrarily small $\delta > 0$. We introduce a function $h \in C(D)$ satisfying

$$h(x_j) = v(x_j),\ j = 1, \ldots, m, \qquad h(x^*) = -\mathrm{sgn}(v(x^*)), \qquad \|h\|_{C(D)} = 1. \tag{6.4}$$

Clearly, the function $\eta := v - h$ belongs to $\mathcal{N}$, and we have

$$\mu(\mathcal{N}, V)_{C(D)} \ge \frac{\|\eta\|_{C(D)}}{\|\eta - v\|_{C(D)}} = \frac{\|v - h\|_{C(D)}}{\|h\|_{C(D)}} \ge |v(x^*) - h(x^*)| \ge \mu - \delta + 1. \tag{6.5}$$

Since $\delta > 0$ was arbitrary, this proves (6.2). □

Next, we prove Lemma 4.1 stated in a slightly different version below.

Lemma 6.2. *Let $\theta_1, \ldots, \theta_N$ be N distinct points in $\mathbb{R}^n$ with convex hull $\mathcal{C} := \mathrm{conv}\{\theta_1, \ldots, \theta_N\}$. Then, there exist functions $\psi_j^{(N)} : \mathcal{C} \to \mathbb{R}$, $j = 1, \ldots, N$, such that*

(i) *$\psi_1^{(N)}, \ldots, \psi_N^{(N)}$ are continuous on $\mathcal{C}$;*

(ii) *for any linear function $\lambda : \mathbb{R}^n \to \mathbb{R}$ (in particular for $\lambda(\theta) = 1$ and $\lambda(\theta) = \theta$),*

$$\sum_{i=1}^{N} \psi_i^{(N)}(\theta) \lambda(\theta_i) = \lambda(\theta) \qquad \text{whenever } \theta \in \mathcal{C}; \tag{6.6}$$

(iii) *for all $i = 1, \ldots, N$, $\psi_i^{(N)}(\theta) \ge 0$ whenever $\theta \in \mathcal{C}$;*

(iv) *for all $i, j = 1, \ldots, N$, $\psi_i^{(N)}(\theta_j) = \delta_{i,j}$.*

Proof: We proceed by induction on $N \geq 1$. The result is clear for $N = 1$ and $N = 2$. Let us assume that it holds up to $N - 1$ for some integer $N \geq 3$ and that we are given N distinct points $\theta_1, \ldots, \theta_N \in \mathbb{R}^n$. We separate two cases.

Case 1: Each θ_j is an extreme point of $\mathcal{C} := \mathrm{conv}\{\theta_1, \ldots, \theta_N\}$. In this case, we invoke the result of Kalman [18] and consider the functions $\psi_1^{(N)}, \ldots, \psi_N^{(N)}$ from [18] satisfying (i)–(iii). Condition (iv) then occurs as a consequence of (ii)-(iii). Indeed, given $j = 1, \ldots, N$, one can find a linear function $\lambda : \mathbb{R}^n \to \mathbb{R}$ such that $\lambda(\theta_j) = 0$ and $\lambda(\theta_i) > 0$ for all $i \neq j$. Therefore,

$$\sum_{i=1}^{N} \psi_i^{(N)}(\theta_j)\lambda(\theta_i) = \lambda(\theta_j) = 0 \tag{6.7}$$

implies that $\psi_i^{(N)}(\theta_j) = 0$ for all $i \neq j$, and then $\psi_j^{(N)}(\theta_j) = 1$ follows from $\sum_{i=1}^{N} \psi_i^{(N)}(\theta_j) = 1$.

Case 2: One of the θ_j's belongs to the convex hull of the other θ_i's, say $\theta_N \in \mathrm{conv}\{\theta_1, \ldots, \theta_{N-1}\}$. Let $\psi_1^{(N-1)}, \ldots, \psi_{N-1}^{(N-1)}$ be the functions defined on $\mathcal{C} = \mathrm{conv}\{\theta_1, \ldots, \theta_{N-1}\} = \mathrm{conv}\{\theta_1, \ldots, \theta_N\}$ which are obtained from the induction hypothesis applied to the $N-1$ distinct points $\theta_1, \ldots, \theta_{N-1}$. Next, we introduce the set Ω, which has at least two elements, and the function τ, which is continuous on $\mathcal{C}$, given by

$$\Omega := \{j = 1, \ldots, N-1 : \psi_j^{(N-1)}(\theta_N) > 0\}, \qquad \tau(\theta) := \min_{j \in \Omega} \frac{\psi_j^{(N-1)}(\theta)}{\psi_j^{(N-1)}(\theta_N)}, \quad \theta \in \mathcal{C}. \tag{6.8}$$

Finally, we define functions $\psi_1^{(N)}, \ldots, \psi_N^{(N)}$ by

$$\psi_i^{(N)}(\theta) := \psi_i^{(N-1)}(\theta) - \psi_i^{(N-1)}(\theta_N)\tau(\theta), \quad i = 1, \ldots, N-1, \qquad \psi_N^{(N)}(\theta) := \tau(\theta). \tag{6.9}$$

These are continuous functions of $\theta \in \mathcal{C}$, so (i) is satisfied. To verify (ii), given a linear function $\lambda : \mathbb{R}^n \to \mathbb{R}$, we observe that

$$\begin{aligned} \sum_{i=1}^{N} \psi_i^{(N)}(\theta)\lambda(\theta_i) &= \sum_{i=1}^{N-1} \psi_i^{(N-1)}(\theta)\lambda(\theta_i) - \tau(\theta) \sum_{i=1}^{N-1} \psi_i^{(N-1)}(\theta_N)\lambda(\theta_i) + \tau(\theta)\lambda(\theta_N) \\ &= \lambda(\theta) - \tau(\theta)\lambda(\theta_N) + \tau(\theta)\lambda(\theta_N) = \lambda(\theta). \end{aligned} \tag{6.10}$$

As for (iii), given $\theta \in \mathcal{C}$, the fact that $\psi_N^{(N)}(\theta) \geq 0$ is clear from the definition of τ, and for $i = 1, \ldots, N-1$, the fact that $\psi_i^{(N)}(\theta) \geq 0$ is equivalent to $\psi_i^{(N-1)}(\theta_N)\tau(\theta) \leq \psi_i^{(N-1)}(\theta)$, which is obvious if $i \notin \Omega$ and follows from the definition of τ if $i \in \Omega$. Finally, to prove (iv), it is enough to verify that $\psi_i^{(N)}(\theta_i) = 1$ for all $i = 1, \ldots, N$, which clearly holds for $i = N$, and for $i = 1, \ldots, N-1$, it is the identity $\psi_i^{(N-1)}(\theta_N)\tau(\theta_i) = 0$, valid both when $i \notin \Omega$ and when $i \in \Omega$, that implies $\psi_i^{(N)}(\theta_i) = \psi_i^{(N-1)}(\theta_i) = 1$. We have now shown that the induction hypothesis holds for N, and this concludes the inductive proof. □

References

[1] CVX Research, Inc. CVX: MATLAB software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx` (2014).

[2] B. Adcock and A. C. Hansen, *Stable reconstructions in Hilbert spaces and the resolution of the Gibbs phenomenon*. Appl. Comput. Harm. Anal., **32**, 357–388 (2012).

[3] B. Adcock, A. C. Hansen, and C. Poon, *Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem*. SIAM J. Math. Anal., **45**, 3132–3167 (2013).

[4] N. S. Bakhvalov, *On the optimality of linear methods for operator approximation in convex classes of functions*. USSR Computational Mathematics and Mathematical Physics **11**, 244–249 (1971).

[5] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis methods*. SIAM Journal of Mathematical Analysis **43**, 1457–1472 (2011).

[6] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in reduced modeling*. SIAM/ASA Journal on Uncertainty Quantification **5**, 1–29 (2017).

[7] B. Bojanov, *Optimal recovery of functions and integrals*. First European Congress of Mathematics, Vol. I (Paris, 1992), 371–390, Progr. Math., **119**, Birkhäuser, Basel (1994).

[8] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press (2004).

[9] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k-term approximation*. J. A. M. S, **22**, 211–231 (2009).

[10] D. Coppersmith and T. Rivlin, *The growth of polynomials bounded at equally spaced points*. SIAM J. Math. Anal. **23**, 970–983 (1992).

[11] J. Creutzig and P. Wojtaszczyk, *Linear vs. nonlinear algorithms for linear problems*. Journal of Complexity **20**, 807–820 (2004).

[12] R. DeVore, *Nonlinear approximation*. Acta Numerica, **7**, 51-150 (1998).

[13] R. DeVore and G.G. Lorentz, Constructive Approximation. Springer Grundlehren, vol. 303 (1993).

[14] R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in Banach spaces*. P. Calcolo, To appear.

[15] T. A. Driscoll, N. Hale, and L. N. Trefethen, editors. Chebfun Guide. Pafnuty Publications, Oxford (2014).

[16] M. Elad, Sparse and Redundant Representations. Springer (2010).

[17] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing. Birkhäuser (2013).

[18] J. A. Kalman, *Continuity and convexity of projections and barycentric coordinates in convex polyhedra*. Pacific Math. J., **11**, 1017–1022 (1961).

[19] J. Lindenstrauss, *Extension property for compact operators*. Mem. Amer. Math. Soc., **48** (1964).

[20] C. Micchelli and T. Rivlin, *Lectures on optimal recovery*. Numerical analysis, (Lancaster, 1984), 21–93, Lecture Notes in Math., **1129**, Springer, Berlin (1985).

[21] C. Micchelli, T. Rivlin, and S. Winograd, *The optimal recovery of smooth functions*. Numerische Mathematik **26**, 191-200 (1976).

[22] V. Milman and G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*. Lecture Notes in Mathematics, vol. 1200, Springer-Verlag Berlin Heidelberg, 1986.

[23] K. Yu. Osipenko, *Best approximation of analytic functions from information about their values at a finite number of points*. Mathematical notes of the Academy of Sciences of the USSR, **19**(1), 17–23 (1976).

[24] R. Platte, L. Trefethen, and A. Kuijlaars, *Impossibility of fast stable approximation of analytic functions from equispaced samples*. SIAM Review, **53**, 308–318 (2011).

[25] A. Schönhage, *Fehlerfortpflantzung bei Interpolation*. Numer. Math., **3**, 62–71 (1961).

[26] J. Traub and H. Wozniakowski, A General Theory of Optimal Algorithms. Academic Press, New York (1980).

[27] A. H. Turetskii, *The bounding of polynomials prescribed at equally distributed points*. Proc. Pedag. Inst. Vitebsk **3**, 117–127 (1940). [Russian]

[28] M. W. Wilson, *Necessary and sufficient conditions for equidistant quadrature formula*. SIAM J. Numer. Anal., **7**(1), 134–141 (1970).

[29] M. Zippin, *Extension of bounded linear operators*. Handbook of the Geometry of Banach Spaces vol. 2, 1703–1741, North-Holland (2003).

[30] A. Zygmund, Trigonometric Series. Cambridge University Press (2002).

Ronald DeVore
Department of Mathematics, Texas A&M University, College Station, TX77840, USA
rdevore@math.tamu.edu

Simon Foucart
Department of Mathematics, Texas A&M University, College Station, TX77840, USA
foucart@tamu.edu

Guergana Petrova
Department of Mathematics, Texas A&M University, College Station, TX77840, USA
gpetrova@math.tamu.edu.

Przemyslaw Wojtaszczyk
Interdisciplinary Center for Mathematical and Computational Modelling,
University of Warsaw, 02-630 Warsaw, ul. Tyniecka 15/17, Poland
wojtaszczyk@icm.edu.pl