Homogeneous algorithms and solvable problems on cones

D. Krieg, P. Kritzer

RICAM-Report 2023-30

Homogeneous algorithms and solvable problems on cones

David Krieg and Peter Kritzer

November 27, 2023

Abstract

We consider linear problems in the worst case setting. That is, given a linear operator and a pool of admissible linear measurements, we want to approximate the values of the operator uniformly on a convex and balanced set by means of algorithms that use at most n such measurements. It is known that, in general, linear algorithms do not yield an optimal approximation. However, as we show in this paper, an optimal approximation can always be obtained with a homogeneous algorithm. This is of interest to us for two reasons. First, the homogeneity allows us to extend any error bound on the unit ball to the full input space. Second, homogeneous algorithms are better suited to tackle problems on cones, a scenario that is far less understood than the classical situation of balls. We illustrate our results by several examples.

1 Introduction

We consider problems given by a solution operator $S\colon \mathcal{F} \to \mathcal{G}$ between normed spaces $\mathcal{F}$ and $\mathcal{G}$ over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, a class $\Lambda \subseteq \mathbb{K}^{\mathcal{F}}$ of admissible measurements, and a class $F \subset \mathcal{F}$ of inputs. We want to approximate the solution $S(f)$ for some unknown $f \in F$ based on the outcome of a finite number n of measurements $L_1(f), \ldots, L_n(f)$. The measurements $L_1, \ldots, L_n$ shall be contained in the class Λ and may be chosen adaptively, i.e., the choice of $L_i \in \Lambda$ may depend on the already computed $L_1(f), \ldots, L_{i-1}(f)$.[1] This results in an *information mapping* of the form

$$N\colon \mathcal{F} \to \mathbb{K}^n, \quad N(f) = (L_1(f), \ldots, L_n(f)). \tag{1}$$

We consider algorithms of the form

$$A_n\colon \mathcal{F} \to \mathcal{G}, \quad A_n(f) = \varphi(N(f)), \tag{2}$$

where $\varphi\colon \mathbb{K}^n \to \mathcal{G}$ is an arbitrary mapping and often called the *recovery map*. If the L_i are not chosen adaptively, i.e., $L_1, \ldots, L_n \in \Lambda$ are the same for all $f \in \mathcal{F}$, then the algorithm A_n is called non-adaptive. The error of an algorithm A_n is measured in the norm of $\mathcal{G}$ and in the worst case over the given input class $F \subset \mathcal{F}$, that is,

$$\mathrm{err}(A_n) = \mathrm{err}(A_n, S, F) := \sup_{f \in F} \|S(f) - A_n(f)\|_{\mathcal{G}}. \tag{3}$$

A problem of this form is called a *linear problem* if the following conditions hold.

(1) The input class F is non-empty, convex (i.e., $\lambda f + (1-\lambda)g \in \mathcal{C}$ for all $f, g \in \mathcal{C}$ and $\lambda \in (0,1)$), and balanced (i.e., $\lambda f \in \mathcal{C}$ for all $f \in \mathcal{C}$ and $\lambda \in \mathbb{K}$ with $|\lambda| \leq 1$),

---

[1] In principle, also the number $n = n(f)$ of measurements could be chosen adaptively; however, in the setting of nth minimal worst-case errors considered in the first three sections, such algorithms can be identified with algorithms using a fixed n (which can be chosen as the maximum of all $n(f)$). In Section 4, the adaptive choice of n will be of importance.

(2) the solution operator $S\colon \mathcal{F} \to \mathcal{G}$ is linear,

(3) the class Λ of admissible measurements contains only linear functionals.

Typical examples of linear problems are *approximation problems*, where we have $S(f) = f$, and *integration problems*, where $\mathcal{F}$ is some class of integrable functions and $S(f)$ is the integral of f. Examples for the class Λ of admissible measurements are the class $\Lambda^{\mathrm{all}} := \mathcal{F}'$ of all continuous linear functionals (*linear information*) or the class Λ^{std} of all function evaluations (*standard information*) if $\mathcal{F}$ is a function space. The input class F often equals the unit ball of the space $\mathcal{F}$ but the model also allows for so-called approximation sets of the form

$$F = \bigl\{ f \in \mathcal{F} \mid \mathrm{dist}_{\mathcal{F}}(f, \mathcal{V}) \leq \delta \bigr\},$$

where $\mathcal{V}$ is a (typically finite-dimensional) subspace of $\mathcal{F}$ and $\delta > 0$. We refer to [31, Section 4.2] and the references therein for further details on the model of linear problems.

The class of all algorithms of the form (2), which we denote by $\mathcal{A}_n$, is quite huge and contains very complicated and impractical mappings, so it is natural to ask whether an (almost) optimal error bound can already be achieved with simpler algorithms, obeying a special structure. One class of particularly simple algorithms are *linear algorithms*. It is a classical result in Information-Based Complexity that linear algorithms are optimal for linear problems in the case that $\mathcal{G} = \mathbb{K}$, that is, if the solution operator S is a linear functional. Then we have

$$\inf_{A_n \in \mathcal{A}_n^{\mathrm{lin}}} \mathrm{err}(A_n) = \inf_{A_n \in \mathcal{A}_n} \mathrm{err}(A_n), \tag{4}$$

where $\mathcal{A}_n^{\mathrm{lin}}$ denotes the class of all linear algorithms of the form (2). This result goes back to Smolyak [37] and Bakhvalov [1], see also [8, 10, 34] for similar results in the presence of noise. Other instances where linear algorithms are optimal for linear problems are when $\mathcal{F}$ is a pre-Hilbert space or when $\mathcal{G}$ is a space of bounded functions with the sup-norm, see Mathé [28] and Creutzig and Wojtaszczyk [4]. On the other hand, linear algorithms are *not* optimal for all linear problems. An example where (4) does not hold goes back to Kashin, Garnaev, and Gluskin [15, 22] and is important in the area of compressive sensing. It is given by the approximation problem

$$S\colon \ell_1^m \to \ell_2^m, \quad S(f) = f,$$

on the unit ball F of ℓ_1^m with $\Lambda = \Lambda^{\mathrm{all}}$. Here, non-linear algorithms are *much* better than linear algorithms if the dimension m is large compared to the number n of measurements. Indeed, it can be shown that, in this case, the left-hand side of (4) equals $\sqrt{(m-n)/m}$ for all $n < m$, while the right-hand side is of order $\sqrt{\log(m/n)/n}$. See, e.g., [33, VI Theorem 2.7] and [11, Theorem 1.1]. Furthermore, there are even linear problems for which every linear algorithm has an infinite error, but for which a nonlinear algorithm with finite error exists, see [40].

In this paper, we consider the larger class of *homogeneous algorithms*. These are algorithms with the property that $A_n(\lambda f) = \lambda A_n(f)$ for all $f \in \mathcal{F}$ and all $\lambda \in \mathbb{K}$. If the same property holds for all real $\lambda \geq 0$, we call an algorithm *positively homogeneous*. When linear algorithms are bad, one may hope that at least the homogeneity is reconcilable with a small error, and in fact, this is known to be the case for various examples (see Remark 7). Here, we show that homogeneous algorithms are optimal for *all* linear problems up to a factor of at most two.

Theorem 1. *Let $(S\colon \mathcal{F} \to \mathcal{G}, F, \Lambda)$ be a linear problem and let $\mathcal{G}$ be complete. For $n \in \mathbb{N}$, we let $\mathcal{A}_n^*$ denote the class of all homogeneous and non-adaptive algorithms of the form (2), and, as above, let $\mathcal{A}_n$ denote the more general class of all algorithms of the form (2). Then*

$$\inf_{A_n \in \mathcal{A}_n^*} \mathrm{err}(A_n) \leq 2 \inf_{A_n \in \mathcal{A}_n} \mathrm{err}(A_n). \tag{5}$$

2

Remark 1. The infimum on the right-hand side of (5) is referred to as the *nth minimal error* (with respect to all admissible algorithms, where the class of admissible algorithms depends on the precise problem specifications).

Remark 2. In fact, we are going to show that homogeneous recovery maps are optimal for every fixed non-adaptive information mapping N, see Proposition 1. Theorem 1 is then implied by the known optimality of non-adaptive information mappings, see [14, 38]; we also refer to [30] for a discussion of the power of adaption.

Remark 3. Assuming additional structure, there is a result by Bartle and Graves [2] which implies the existence of a positively homogeneous and continuous approximate spline mapping for every continuous and linear information mapping N, see [9, Theorem 3.3]. The notion of approximate splines is explained in Remark 5. Using this result, it is easy to obtain a (less general) version of Theorem 1. Here, we take a more general approach, which is also easier since we disregard continuity.

There are two reasons why we are interested in results of this type. First, in the prominent case that F is the unit ball of $\mathcal{F}$, it is easily seen that any (positively) homogeneous algorithm A_n satisfies

$$\|S(f) - A_n(f)\|_{\mathcal{G}} \leq \mathrm{err}(A_n) \cdot \|f\|_{\mathcal{F}} \quad \text{for all } f \in \mathcal{F}.$$

In this sense, a homogeneous algorithm with small error is not only good on the unit ball of $\mathcal{F}$ but instead on the full space. This is usually not the case for non-homogeneous algorithms. In particular, as stated in Corollary 1, the nth minimal error and the complexity of a linear problem do not change (up to a factor of two) if we switch from the error criterion $\mathrm{err}(A_n)$ in (3) to the error criterion

$$\widetilde{\mathrm{err}}(A_n) := \sup_{f \in \mathcal{F} \setminus \{0\}} \frac{\|S(f) - A_n(f)\|_{\mathcal{G}}}{\|f\|_{\mathcal{F}}}.$$

Second, homogeneous algorithms are better suited for problems that are defined on cones as considered, e.g., in [6, 25]. By a *cone* we generally understand any subset $\mathcal{C}$ of a $\mathbb{K}$-vector space $\mathcal{F}$ which satisfies that $\lambda f \in \mathcal{C}$ for all $f \in \mathcal{C}$ and $\lambda > 0$. Problems on cones are usually not solvable with algorithms that use a fixed number of measurements, see Proposition 2. In other words, they are not *uniformly solvable*. However, they are often solvable with algorithms that use an adaptive number of measurements, see [6, 25]. In other words, they are *weakly solvable*. In Theorem 2, we provide a statement on the solvability of such problems on a certain family of cones, and this insight is based on the optimality of homogeneous algorithms for linear problems.

Theorem 2. *Let $(S, B_{\mathcal{F}}, \Lambda)$ and $(T, B_{\mathcal{F}}, \Lambda)$ be uniformly solvable linear problems, where $B_{\mathcal{F}}$ is the unit ball of $\mathcal{F}$, and let $t > 0$. Then the problem $(S, \mathcal{C}_t, \Lambda)$ is weakly solvable, where*

$$\mathcal{C}_t := \{f \in \mathcal{F} \colon \|f\|_{\mathcal{F}} \leq t \, \|Tf\|_{\mathcal{H}}\}.$$

More precisely, we have the cost bound (15).

We refer to Section 4 for the precise definitions of the solvability notions and the explicit cost bound (15). To illustrate Theorem 2 with an example, which is discussed among other examples in Section 4, consider the approximation problem $S \colon W_2^1([0,1]) \to L_2(0,1)$ with standard information on the input set

$$\mathcal{C}_t := \left\{f \in W_2^1([0,1]) \colon \|f'\|_2 \leq t \, \|f\|_2\right\}.$$

Here, there is no algorithm $A_n \in \mathcal{A}_n$ that uses a fixed number n of function values and has a finite error $\mathrm{err}(A_n) < \infty$. But a prescribed error $\varepsilon > 0$ can still be guaranteed with a varying

3

finite number of function evaluations. The algorithm, which does not require knowledge of $\|f\|_2$, needs at most $\mathcal{O}(t\varepsilon^{-1}\|f\|_2)$ function values to find an ε-approximation of any $f \in \mathcal{C}_t$.

Remark 4. We mentioned above that linear algorithms are, in general, not optimal for linear problems, but they are optimal if the target space $\mathcal{G}$ is the space $B(X)$ of bounded functions on a compact Hausdorff space X (see [4]). It is shown in [32] that for any normed space $\mathcal{G}$ there exists a compact Hausdorff space X and a subspace $\mathcal{G}'$ of $B(X)$ such that $\mathcal{G}$ and $\mathcal{G}'$ are isometrically isomorphic. Thus, any linear problem $S\colon \mathcal{F} \to \mathcal{G}$ may be interpreted as a linear problem $S'\colon \mathcal{F} \to \mathcal{G}'$, which implies that the nth minimal error can be achieved with a linear algorithm $A'_n\colon \mathcal{F} \to B(X)$. Thus, in a certain sense, linear algorithms are always optimal for linear problems if we sufficiently blow up the target space. However, since A'_n maps to $B(X)$ and not to the subspace $\mathcal{G}'$, there is in general no meaningful interpretation of the approximation $A'_n(f)$ in terms of the original space of solutions $\mathcal{G}$.

Notation. In the following, depending on the situation, we might sometimes write Sf instead of $S(f)$, Nf instead of $N(f)$, and similarly for other mappings, when it eases readability. Furthermore, we will use the notion of (positively) homogeneous mappings also for mappings between subsets of $\mathbb{K}$-vector spaces.

2 Homogeneous algorithms and linear problems

The following proposition is at the heart of our findings. It is a statement on homogeneous solution operators $S\colon \mathcal{F} \to \mathcal{G}$ and homogeneous information mappings $N\colon \mathcal{F} \to \mathbb{K}^n$. In short, it implies that there always exists a *homogeneous* recovery map $\varphi\colon \mathbb{K}^n \to \mathcal{G}$ such that the corresponding algorithm $A_n = \varphi \circ N$ is optimal up to a factor of at most 2.

For a precise statement, we introduce the diameter of an (information) mapping $N\colon \mathcal{F} \to \mathbb{K}^n$,

$$\mathrm{diam}(N) = \mathrm{diam}(N, S, F) := \sup\left\{\|Sf - Sg\|_{\mathcal{G}}\colon f, g \in F, Nf = Ng\right\},$$

which measures the maximal uncertainty in the solution under the a priori knowledge $f \in F$ if the information about f is given by N. The diameter of information relates to the minimal error that can be achieved with this information by

$$\inf_{\varphi\colon \mathbb{K}^n \to \mathcal{G}} \mathrm{err}(\varphi \circ N) \leq \mathrm{diam}(N) \leq 2 \inf_{\varphi\colon \mathbb{K}^n \to \mathcal{G}} \mathrm{err}(\varphi \circ N). \tag{6}$$

The infimum appearing in (6) is commonly referred to as the radius of information and denoted by $\mathrm{rad}(N)$. We refer to [31, Chapter 4] for a proof of (6) and for background on these concepts.

Proposition 1. *Let F be a balanced, convex and non-empty subset of a $\mathbb{K}$-vector space $\mathcal{F}$, let $\mathcal{G}$ be a complete normed space, and let $S\colon \mathcal{F} \to \mathcal{G}$ be positively homogeneous. Then, for every positively homogeneous mapping $N\colon \mathcal{F} \to \mathbb{K}^n$ and any $\delta > 0$, there is a positively homogeneous mapping $\varphi\colon \mathbb{K}^n \to \mathcal{G}$ such that*

$$\mathrm{err}(\varphi \circ N) \leq (1 + \delta)\,\mathrm{diam}(N). \tag{7}$$

In particular,

$$\mathrm{err}(\varphi \circ N) \leq (2 + 2\delta) \inf_{\psi\colon \mathbb{K}^n \to \mathcal{G}} \mathrm{err}(\psi \circ N). \tag{8}$$

Moreover, if S and N are homogeneous, we can also choose φ to be homogeneous.

Remark 5. The main step in the proof will be to define a (positively) homogeneous approximate spline mapping. Here, a mapping $s\colon N(\mathcal{F}) \to \mathcal{F}$ is called an *approximate spline*, if it satisfies

4

- $N(s(y)) = y$ for all $y \in N(\mathcal{F})$ and

- $\|s(Nf)\|_F \le (1 + \delta)\|f\|_F$ for all $f \in \mathcal{F}$,

where $\delta > 0$ is a given parameter and $\| \cdot \|_F$ is the Minkowski semi-norm induced by F, see (9). It is called an *exact spline* in the case $\delta = 0$. Exact splines are studied in [38, Chapter 4], for example. In certain situations, see [9, Theorem 3.4], there exists a unique exact spline, and then the spline is necessarily (positively) homogeneous. But in general, an exact spline does not exist. Here, we show that there always exists a (positively) homogeneous approximate spline. The desired recovery map is then given by $\varphi = S \circ s$.

The existence of a positively homogeneous approximate spline was already noted in [9, Theorem 3.3], a result that is attributed to Bartle and Graves [2]. This result makes extra assumptions on $\mathcal{F}$, N and F (e.g., that N is linear and continuous and that $\| \cdot \|_F$ is a norm), but in return the approximate spline is also continuous.

Proof. We first observe that the positively homogeneous case implies the homogeneous case. Indeed, by the first part of the statement, we can find a positively homogeneous mapping $\widetilde{\varphi}$ with the desired error bound (7). We let M be the set of vectors in $\mathbb{K}^n$ which are either zero or whose first non-zero coordinate is real and positive. Then each $y \in \mathbb{K}^n \setminus \{0\}$ can be uniquely expressed as $y = \lambda(y) \cdot y_+$ with some $y_+ \in M$ and some $\lambda(y) \in \mathbb{K}$ with $|\lambda(y)| = 1$. We replace the mapping $\widetilde{\varphi}$ by the mapping $\varphi \colon \mathbb{K}^n \to \mathcal{G}$ with $\varphi(y) := \lambda(y) \cdot \widetilde{\varphi}(y_+)$. It is easily verified that the new mapping φ is homogeneous and that the error bound (7) is preserved under this replacement whenever S and N are homogeneous.

Moreover, Equation (8) follows from Equation (7) via (6). Thus our task is to prove the first statement on the existence of a positively homogeneous mapping φ that satisfies (7).

For this, we first note that $\mathrm{span}(F) = \{\lambda f \colon f \in F, \lambda \ge 0\}$. Without loss of generality, we assume that $\mathcal{F} = \mathrm{span}(F)$ since the quantities $\mathrm{err}(\varphi \circ N)$ and $\mathrm{diam}(N)$ depend on F but not $\mathcal{F}$. Note furthermore that the result is trivially correct if $\mathrm{diam}(N) = \infty$. Hence, we may assume for the rest of the proof that $\mathrm{diam}(N) < \infty$.

We define a semi-norm on $\mathcal{F}$ by the Minkowski functional

$$\|f\|_F := \inf \{r > 0 \colon f/r \in F\}. \tag{9}$$

Then

$$\{f \in \mathcal{F} \colon \|f\|_F < 1\} \subset F \subset \{f \in \mathcal{F} \colon \|f\|_F \le 1\}. \tag{10}$$

Moreover, we define

$$K \colon N(\mathcal{F}) \to [0, \infty), \quad K(y) := \inf_{f \in N^{-1}(y)} \|f\|_F.$$

This mapping is positively homogeneous; we have $K(\lambda y) = \lambda K(y)$ for $y \in \mathbb{K}^n$ and $\lambda \ge 0$.

We now define a mapping $\varphi \colon N(\mathcal{F}) \to \mathcal{G}$. For this, we split $\mathcal{F}$ into the two disjoint cones

$$\mathcal{F}_0 := \{f \in \mathcal{F} \colon K(Nf) = 0\} \quad \text{and} \quad \mathcal{F}_+ := \{f \in \mathcal{F} \colon K(Nf) > 0\}$$

and thereby also split $N(\mathcal{F})$ into the two disjoint cones

$$N(\mathcal{F}_0) = \{y \in N(\mathcal{F}) \colon K(y) = 0\} \quad \text{and} \quad N(\mathcal{F}_+) = \{y \in N(\mathcal{F}) \colon K(y) > 0\}.$$

We will define two positively homogeneous mappings $\varphi_0 \colon N(\mathcal{F}_0) \to \mathcal{G}$ and $\varphi_+ \colon N(\mathcal{F}_+) \to \mathcal{G}$ such that

$$\|Sf - \varphi_0(Nf)\|_{\mathcal{G}} \le \mathrm{diam}(N) \quad \text{for all } f \in F \cap \mathcal{F}_0 \tag{11}$$

5

and

$$\|Sf - \varphi_+(Nf)\|_{\mathcal{G}} \le (1+\delta)\operatorname{diam}(N) \quad \text{for all } f \in F \cap \mathcal{F}_+. \tag{12}$$

Then the mapping

$$\varphi \colon \mathbb{K}^n \to \mathcal{G}, \quad \varphi(y) = \begin{cases} \varphi_0(y) & \text{if } y \in N(\mathcal{F}_0), \\ \varphi_+(y) & \text{if } y \in N(\mathcal{F}_+), \\ 0 & \text{if } y \notin N(\mathcal{F}), \end{cases}$$

is positively homogeneous and satisfies (7). Thus, it only remains to show (11) and (12) to complete the proof.

We start with the definition of φ_+ and the proof of (12). Let $y \in N(\mathcal{F}_+)$. If $K(y) = 1$, we choose $s(y)$ as an arbitrary element from the pre-image $N^{-1}(y)$ that satisfies

$$\|s(y)\|_F < (1+\delta)K(y). \tag{13}$$

For arbitrary $y \in N(\mathcal{F}_+)$, we have $K(y/K(y)) = 1$ and put $s(y) = K(y)s(y/K(y))$. Then the mapping $s \colon N(\mathcal{F}_+) \to \mathcal{F}$ is positively homogeneous, which also implies that (13) holds for all $y \in N(\mathcal{F}_+)$. Moreover, for all $y \in N(\mathcal{F}_+)$, the element $s(y) \in \mathcal{F}$ satisfies $N(s(y)) = y$ (i.e., it is an approximate spline). We put $\varphi_+(y) = S(s(y))$.

The mapping $\varphi_+ \colon N(\mathcal{F}_+) \to \mathcal{G}$ is positively homogeneous as a composition of positively homogeneous mappings. To bound the error of $\varphi_+ \circ N$, we note that for all $f \in \mathcal{F}_+$ with norm $\|\cdot\|_F$ at most $1/(1+\delta)$, which implies $f \in F$, we have $K(Nf) \le 1/(1+\delta)$ and thus by (13) that $s(Nf)$ is contained in F. Moreover, f and $s(Nf)$ yield the same information. Therefore,

$$\|Sf - \varphi_+(Nf)\|_{\mathcal{G}} \le \operatorname{diam}(N) \quad \text{for all } f \in \mathcal{F}_+ \text{ with } \|f\|_F \le 1/(1+\delta).$$

The bound (12) is now obtained by recalling (10) and scaling due to the positive homogeneity of S and $\varphi_+ \circ N$.

We continue with the definition of φ_0 and the proof of (11). To this end, let $y \in N(\mathcal{F}_0)$ and hence $K(y) = 0$. Then there is a sequence $(f_k) \subset N^{-1}(y)$ such that $\|f_k\|_F \to 0$. This implies that (Sf_k) is a Cauchy sequence in $\mathcal{G}$. Indeed, assume to the contrary that it is no Cauchy sequence. Then there is some $c > 0$ such that for all n_0 there are $k, m > n_0$ with $\|Sf_k - Sf_m\|_{\mathcal{G}} \ge c$. Given any $R > 0$, we can choose n_0 large enough such that $\|f_k\|_F$ and $\|f_m\|_F$ are smaller than c/R. Then $h_k = Rf_k/c$ and $h_m = Rf_m/c$ are in F and satisfy $Nh_k = Nh_m$ and $\|Sh_k - Sh_m\|_{\mathcal{G}} \ge R$. This yields $\operatorname{diam}(N) = \infty$, which would be a contradiction.
As $\mathcal{G}$ is complete, the sequence (Sf_k) has a limit g and we define $\varphi_0(y) := g$. This definition is independent of the choice of the sequence (f_k). Indeed, let (f_k^*) be another sequence in $N^{-1}(y)$ with $\|f_k^*\|_F \to 0$. Assume to the contrary that (Sf_k^*) has a different limit than (Sf_k), i.e., there is some $c > 0$ with $\|Sf_k - Sf_k^*\|_{\mathcal{G}} \ge c$ for infinitely many k. For any $R > 0$, choosing k large enough, we thus have $h_k := Rf_k/c \in F$ and $h_k^* := Rf_k^*/c \in F$ with $N(h_k) = N(h_k^*)$ and $\|Sf_k - Sf_k^*\|_{\mathcal{G}} \ge R$, again leading to the contradiction $\operatorname{diam}(N) = \infty$.

It is not hard to check that the mapping φ_0 is indeed positively homogeneous: We have $\varphi_0(0) = 0$ (since the value of $\varphi_0(0)$ is independent of the concrete choice of the sequence $(f_k) \subset N^{-1}(0)$ from above) and for $\lambda > 0$ and $y \in N(\mathcal{F}_0) \setminus \{0\}$, we know that $\varphi_0(\lambda y)$ may be expressed as the limit of a sequence $(S(\lambda f_k))$, where $(f_k) \subset N^{-1}(y)$ such that $\|f_k\|_F \to 0$, and hence $Sf_k \to \varphi_0(y)$; thus $\varphi_0(\lambda y) = \lambda \varphi_0(y)$. Moreover, for every $y \in N(\mathcal{F}_0)$, we have that $\varphi_0(y)$ is contained in the closure $\overline{S(N^{-1}(y) \cap F)}$. Thus, for all $f \in N^{-1}(y) \cap F$, we have

$$\|Sf - \varphi_0(Nf)\|_{\mathcal{G}} \le \operatorname{diam}(\overline{S(N^{-1}(y) \cap F)}) = \operatorname{diam}(S(N^{-1}(y) \cap F)) \le \operatorname{diam}(N).$$

Here, the diameter of a set in $\mathcal{G}$ is defined in the usual way. This gives (11). $\qquad\square$

We now turn to the setting of *linear problems*. Here, a classical result from [14, 38] states that the minimal diameter of information can be obtained with non-adaptive information mappings, see, e.g., [31, Chapter 4.2]. We note that this result is usually stated in the case $\mathbb{K} = \mathbb{R}$, but the proof is exactly the same in the case $\mathbb{K} = \mathbb{C}$. Since the proof is short and elegant anyway, let us repeat it.

Lemma 1 (see [14, 38]). *Let $(S\colon \mathcal{F} \to \mathcal{G}, F, \Lambda)$ be a linear problem and $n \in \mathbb{N}$. For any $\delta > 0$, there exists a non-adaptive information mapping $N\colon \mathcal{F} \to \mathbb{K}^n$ such that*

$$\operatorname{diam}(N) \leq (2 + \delta) \inf_{A_n \in \mathcal{A}_n} \operatorname{err}(A_n).$$

Proof. Let $A_n \in \mathcal{A}_n$ be an arbitrary algorithm and N be the corresponding (possibly adaptive) information mapping. Let $L_1, \ldots, L_n \in \Lambda$ be the measurement maps that the algorithm chooses for the input $f = 0$ and consider the non-adaptive information mapping $N^{\mathrm{non}} := (L_1, \ldots, L_n)$. Let further $f, g \in F$ with $N^{\mathrm{non}}(f) = N^{\mathrm{non}}(g)$. Since F is convex and balanced, the function $h = \frac{f-g}{2}$ is contained in F, and since N^{non} is linear, we have $N^{\mathrm{non}}(h) = 0$. But this means that the adaptive information mapping N recursively chooses the same measurement maps as for the zero input, and thus also $N(h) = 0$. The same statements hold for the function $-h$, and thus A_n cannot distinguish between h and $-h$; we have $A_n(h) = A_n(-h)$. This gives

$$\operatorname{err}(A_n) \geq \frac{1}{2} \|S(h) - S(-h)\|_{\mathcal{G}} = \frac{1}{2} \|S(f) - S(g)\|_{\mathcal{G}}.$$

Taking the supremum over all such f and g, we obtain $2 \operatorname{err}(A_n) \geq \operatorname{diam}(N^{\mathrm{non}})$. $\square$

For a linear problem, any non-adaptive information mapping is also homogeneous. Hence, Proposition 1 results in Theorem 1, the statement of which is repeated here for completeness.

Theorem 1. *Let $(S\colon \mathcal{F} \to \mathcal{G}, F, \Lambda)$ be a linear problem and let $\mathcal{G}$ be complete. For $n \in \mathbb{N}$, we let $\mathcal{A}_n^*$ denote the class of all homogeneous and non-adaptive algorithms of the form (2), and, as above, let $\mathcal{A}_n$ denote the more general class of all algorithms of the form (2). Then*

$$\inf_{A_n \in \mathcal{A}_n^*} \operatorname{err}(A_n) \leq 2 \inf_{A_n \in \mathcal{A}_n} \operatorname{err}(A_n).$$

Proof of Theorem 1. Let $\delta > 0$. By Lemma 1 there exists a non-adaptive (and thus homogeneous) information mapping $N\colon \mathcal{F} \to \mathbb{K}^n$ such that

$$\operatorname{diam}(N) \leq (2 + \delta) \inf_{A_n \in \mathcal{A}_n} \operatorname{err}(A_n).$$

By Proposition 1, there exists a homogeneous mapping $\varphi\colon \mathbb{K}^n \to \mathcal{G}$ such that the homogeneous algorithm $A_n^* := \varphi \circ N$ satisfies

$$\operatorname{err}(A_n^*) \leq (1 + \delta) \operatorname{diam}(N) \leq (1 + \delta)(2 + \delta) \inf_{A_n \in \mathcal{A}_n} \operatorname{err}(A_n),$$

and the proof is finished. $\square$

Remark 6. By Theorem 1, homogeneous recovery maps are essentially (i.e., up to an absolute multiplicative constant) optimal for linear problems while linear recovery maps are not. To compensate for the non-linearity, one may ask whether homogeneous recovery maps with a finite-dimensional image are optimal. But this is not the case. There exists a linear solution operator $S\colon \mathcal{F} \to \mathcal{G}$ such that for any homogeneous information mapping N and every homogeneous recovery map φ with finite-dimensional image, the algorithm $\varphi \circ N$ has infinite error, while $\varphi \circ N$ has finite error for other choices of φ, see [40, Remark 2.1].

Remark 7. For many linear problems, as for example in numerical integration, one naturally uses algorithms that are not only homogeneous, but even linear. So one may ask to what extent homogeneous algorithms are relevant at all from a practical point of view. Let us thus mention a few specific examples of homogeneous and non-linear algorithms that are used for linear problems in the literature.

- ℓ^1-minimization or basis pursuit is proven to be vastly superior to any linear method in the area of compressive sensing, see, e.g., the book [12].

- The median of means is a popular estimator for the computation of integrals. When used in a randomized setting, the median leads to a probability amplification, see, e.g., [25, Proposition 2.2]. In the deterministic sense, it can be used to obtain certain universality properties, see [16, 17].

- The paper [19] considers a randomized setting and gives an example of a linear problem where adaptive algorithms achieve a better rate of convergence than non-adaptive algorithms. The algorithms that achieve the optimal rate are homogeneous, but not linear.

Let us now assume that $\mathcal{F}$ is a normed space and that the input class F is the unit ball of $\mathcal{F}$. Then the nth minimal worst case error of the problem $(S\colon \mathcal{F} \to \mathcal{G}, F, \Lambda)$ is defined as

$$\mathrm{err}(n) := \mathrm{err}(n, S, \Lambda) := \inf_{A_n \in \mathcal{A}_n} \mathrm{err}(A_n, S, F)$$

with the worst case error $\mathrm{err}(A_n, S, F)$ as defined in (3), i.e.,

$$\mathrm{err}(A_n) = \mathrm{err}(A_n, S, F) = \sup_{\|f\|_{\mathcal{F}} \leq 1} \|S(f) - A_n(f)\|_{\mathcal{G}}.$$

It has recently been shown in [39, Lemma 1.3] for a broad class of approximation problems that the nth minimal worst case error over the unit ball coincides up to a factor of at most eight[2] with the nth minimal relative error on the full input space, i.e., with

$$\widetilde{\mathrm{err}}(n) := \widetilde{\mathrm{err}}(n, S, \Lambda) := \inf_{A_n \in \mathcal{A}_n} \widetilde{\mathrm{err}}(A_n, S),$$

where

$$\widetilde{\mathrm{err}}(A_n) := \widetilde{\mathrm{err}}(A_n, S) := \sup_{f \in \mathcal{F} \backslash \{0\}} \frac{\|S(f) - A_n(f)\|_{\mathcal{G}}}{\|f\|_{\mathcal{F}}}.$$

As a corollary to Theorem 1, we obtain that this result is true for all linear problems, where the factor eight may be replaced by the factor two.

Corollary 1. *Consider a linear problem $(S\colon \mathcal{F} \to \mathcal{G}, F, \Lambda)$, where F is the unit ball of $\mathcal{F}$ and $\mathcal{G}$ is complete. For any $n \in \mathbb{N}$, we have*

$$\mathrm{err}(n) \leq \widetilde{\mathrm{err}}(n) \leq 2\,\mathrm{err}(n).$$

Proof. The second inequality follows from Theorem 1 and the fact that $\mathrm{err}(A_n) = \widetilde{\mathrm{err}}(A_n)$ whenever A_n is homogeneous. Namely,

$$\widetilde{\mathrm{err}}(n) \leq \inf_{A_n \text{ homogeneous}} \widetilde{\mathrm{err}}(A_n) = \inf_{A_n \text{ homogeneous}} \mathrm{err}(A_n) \leq 2\,\mathrm{err}(n).$$

The first inequality uses that any algorithm A_n with $A_n(0) = 0$ satisfies

$$\mathrm{err}(A_n) = \sup_{f \in \mathcal{F} \backslash \{0\}\colon \|f\|_{\mathcal{F}} \leq 1} \|S(f) - A_n(f)\|_{\mathcal{G}} \leq \sup_{f \in \mathcal{F} \backslash \{0\}\colon \|f\|_{\mathcal{F}} \leq 1} \frac{\|S(f) - A_n(f)\|_{\mathcal{G}}}{\|f\|_{\mathcal{F}}} \leq \widetilde{\mathrm{err}}(A_n)$$

---

[2]The formula in [39, Lemma 1.3] shows the factor four but only compares with the minimal error of non-adaptive algorithms, which is why an additional factor of two comes into play.

and hence
$$\mathrm{err}(n) \leq \inf_{A_n \colon A_n(0)=0} \mathrm{err}(A_n) \leq \inf_{A_n \colon A_n(0)=0} \widetilde{\mathrm{err}}(A_n).$$

It remains to show that
$$\inf_{A_n \colon A_n(0)=0} \widetilde{\mathrm{err}}(A_n) \leq \widetilde{\mathrm{err}}(n). \tag{14}$$

To this end, suppose that we have an algorithm A_n such that $A_n = \varphi \circ N$ with $\varphi(0) \neq 0$. We then introduce a mapping $\widetilde{\varphi}$ such that $\widetilde{\varphi}(0) = 0$ and $\widetilde{\varphi}(y) = \varphi(y)$ for all $y \neq 0$ and set $A_n^* := \widetilde{\varphi} \circ N$. We claim that $\widetilde{\mathrm{err}}(A_n^*) \leq \widetilde{\mathrm{err}}(A_n)$, which proves (14).

Indeed, let us consider the set $N^{-1}(0)$ of those $f \in \mathcal{F}$ for which $N(f) = 0$, and for which then $A_n(f) \neq 0$, but $A_n^*(f) = 0$. If all such f satisfy that $S(f) = 0$, then A_n^* is at least as good as A_n and thus $\widetilde{\mathrm{err}}(A_n^*) \leq \widetilde{\mathrm{err}}(A_n)$. If, on the other hand, there exist $f \in N^{-1}(0)$ such that $S(f) \neq 0$, then we argue that $\widetilde{\mathrm{err}}(A_n) = \infty$ and thus $\widetilde{\mathrm{err}}(A_n^*) \leq \widetilde{\mathrm{err}}(A_n)$ again holds.

Since S is linear, we have $f \neq 0$. On the other hand, we have $N(cf) = 0$ for any positive constant c. Then, however, by the linearity of S,
$$\widetilde{\mathrm{err}}(A_n) \geq \frac{\|S(cf) - \varphi(0)\|_{\mathcal{G}}}{\|cf\|_{\mathcal{F}}} = \frac{\|cS(f) - \varphi(0)\|_{\mathcal{G}}}{c\,\|f\|_{\mathcal{F}}}.$$

Letting c tend to zero, we obtain the asserted identity $\widetilde{\mathrm{err}}(A_n) = \infty$. $\qquad\square$

3 Homogeneous algorithms in other settings

Recall from the introduction that our problems are described by a solution operator S mapping from a vector space $\mathcal{F}$ to a normed space $\mathcal{G}$, an input set $F \subset \mathcal{F}$, and a class $\Lambda \subset \mathbb{K}^{\mathcal{F}}$ of admissible measurements. The problem is called linear if

(1) the worst case error is considered on a non-empty, convex, and balanced subset F of $\mathcal{F}$,

(2) the solution operator $S \colon \mathcal{F} \to \mathcal{G}$ is linear,

(3) the class Λ of admissible measurements contains only linear functionals.

In Section 2, we have seen that (positively) homogeneous algorithms are optimal for linear problems. It is natural to ask whether this extends to a more general class of (positively) homogeneous problems. Suppose, for example, that we replace one or several of the above conditions by the weaker conditions,

(1') the worst case is considered on a class $F \subset \mathcal{F}$ satisfying $\lambda f \in F$ for $f \in F$ and $0 \leq \lambda \leq 1$,

(2') the solution operator $S \colon \mathcal{F} \to \mathcal{G}$ is positively homogeneous,

(3') the class Λ of admissible measurements contains only positively homogeneous functionals.

Are positively homogeneous algorithms still optimal for such positively homogeneous problems? We do not know the answer to this question.

Open Problem 1. Are positively homogeneous algorithms optimal for positively homogeneous problems? That is, if we consider a class of problems where one or several of the conditions (1)–(3) are replaced by their respective weaker variants (1')–(3'), do there exist constants $C, c \in \mathbb{N}$ such that, for all such problems and all $n \in \mathbb{N}$, we have
$$\inf_{\substack{A \in \mathcal{A}_{cn} \\ A \text{ pos. hom.}}} \mathrm{err}(A) \leq C \inf_{A \in \mathcal{A}_n} \mathrm{err}(A) \quad?$$

Of course, a similar question can be asked for homogeneous problems instead of positively homogeneous problems. Examples of such positively homogeneous problems could be the computation of some norm or the maximum of a function $f \in F$.

If we only consider positively homogeneous problems where the input set F is convex and balanced, Proposition 1 implies that positively homogeneous algorithms are optimal whenever positively homogeneous information mappings are optimal (in the sense that the corresponding diameter of information is close to minimal). In particular, for such problems, positively homogeneous algorithms are optimal whenever non-adaptive algorithms are optimal.

This leads to the question whether non-adaptive algorithms are optimal for homogeneous problems. Unfortunately, this is not the case. The optimality of non-adaptive algorithms does not stay true if any of the conditions (1)–(3) is replaced by its respective weaker variant, as illustrated by the following examples.

Example 1. An example where the properties (1'), (2), and (3) are satisfied is given in [23], see also [31, Section 4.2.1]. Here, F is the class of all monotonically increasing, α-Hölder continuous functions on an interval with Hölder-constant at most one and $\alpha < 1$, S is the embedding into L_∞ and Λ is the class of all function evaluations. Then the error of non-adaptive algorithms is of order $n^{-\alpha}$ while the error of adaptive algorithms is of polynomial order n^{-1}, where n is the number of measurements. We also refer to [13] for related results.

Example 2. We give an example where the properties (1), (2'), and (3) are satisfied. We let F be the unit ball in the space $\mathcal{F}$ of all Lipschitz-continuous functions on $[0, 1]$ with the norm

$$\|f\|_{\mathcal{F}} := \max\left\{\|f\|_\infty, \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}\right\}.$$

We do bisection in order to approximate a point $z \in [0, 1/2]$ with $f(z) = \frac{1}{2}(f(0) + f(1/2))$. The bisection method converges to a solution $z(f)$ of the above equation, while the approximation $z_n(f)$ obtained after n bisection steps (using $n+1$ function values) satisfies $|z(f) - z_n(f)| \leq 2^{-n}$. We consider the homogeneous solution operator $S(f) = f(z(f) + 1/2)$ and the class Λ^{std} of all function evaluations. The adaptive algorithm $A_n(f) = f(z_n(f)+1/2)$ has an error of at most 2^{-n}.

On the other hand, any non-adaptive algorithm using n function values has an error of at least $1/(8n)$. Indeed, there is an interval $(a, a + 1/(2n)) \subset [0, 1/2]$ that does not contain any sampling point. The algorithm thus cannot distinguish the two functions f and g defined by the properties that $f(0) = g(0) = 0$, f is constant on $[0, a]$ and $[a + 1/(4n), 1/2]$, g is constant on $[0, a + 1/(4n)]$ and $[a + 1/(2n), 1/2]$, and both functions are linearly increasing with slope one on the rest of the interval $[0, 1]$. We illustrate this situation in Figure 2. However, since $S(f)$ and $S(g)$ differ by $1/(4n)$, the error of the algorithm is at least $1/(8n)$. Thus, the error of certain adaptive algorithms converges exponentially and the error of non-adaptive algorithms converges only linearly. This problem is clearly related to the zero-finding problem (see, e.g., [35] and the references therein for detailed information on the zero-finding problem).

Example 3. We also give an example where the properties (1), (2), and (3') are satisfied. We let $\mathcal{F}$ be defined as in the previous example and consider the M-fold Cartesian product $\mathcal{F}_M := \mathcal{F} \times \cdots \times \mathcal{F}$, with elements denoted by $f = (f_1, \ldots, f_M)$, equipped with the norm

$$\|f\|_{\mathcal{F}_M} := \sum_{i=1}^{M} \|f_i\|_{\mathcal{F}}.$$

Let F_M be the unit ball of $\mathcal{F}_M$ and let $\mathcal{G}_M$ be the same set of functions as $\mathcal{F}_M$, but with the norm

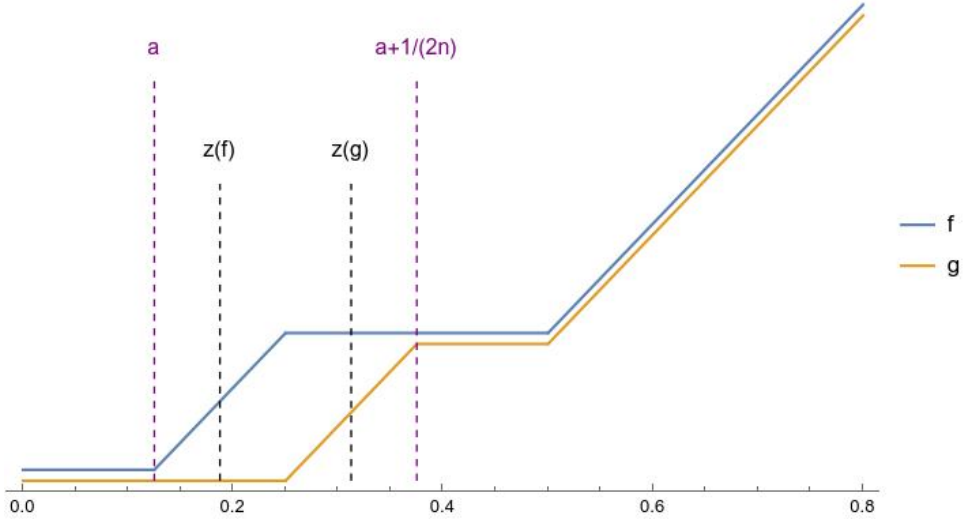$$\|f\|_{\mathcal{G}_M} := \max_{i \leq M} \|f_i\|_\infty.$$

Figure 1: The situation described in Example 2.

Furthermore, let S be the embedding of $\mathcal{F}_M$ into $\mathcal{G}_M$.

The class Λ shall consist of the homogeneous measurements $f_i(x)$ with $i \leq M$ and $x \in [0,1]$ and $\|f_i\|_{\mathcal{F}}$ with $i \leq M$. An adaptive algorithm using at most n measurements is given by first computing the M norms $\|f_i\|_{\mathcal{F}}$ and then approximating each f_i by a piecewise constant function that interpolates m_i function values of f_i at equi-spaced points, where $m_i = 0$ if $\|f_i\|_{\mathcal{F}} < 2/(n-M)$ (i.e., in this case we approximate f_i by the zero function), and $m_i = \lceil \|f_i\|_{\mathcal{F}}(n-M)/2 \rceil \leq \|f_i\|_{\mathcal{F}}(n-M)$ otherwise. This algorithm has an error of at most $2/(n-M)$.

On the other hand, for a non-adaptive algorithm taking n measurements, there exists some i where at most n/M samples of f_i are computed (recall the form of Λ). Then there is some $f^* \in F$ that equals zero at all those points but satisfies $\|f^*\|_\infty \geq M/(2n)$ (by choosing f^* to be linear with slope 1 left of the midpoints of the intervals between the interpolation nodes, and slope -1 right of them, respectively). The algorithm cannot distinguish $f \in \mathcal{F}_M$ defined by $f_j = 0$ for $j \neq i$ and $f_i = f^*$ from its negative since all measurements for f and $-f$ yield the same value (including the possible norm-evaluations). Thus, it has an error of at least $M/(2n)$. Taking, for instance, $M = n/2$, we obtain that all non-adaptive algorithms have an error of at least $1/4$, while a suitable adaptive algorithm can achieve an error of at most $4/n$.

Another interesting question is whether the optimality of homogeneous algorithms for linear problems remains true in the randomized setting. Recall from the introduction that our deterministic algorithms are described as the composition of an information mapping $N\colon \mathcal{F} \to \mathbb{K}^n$ and a recovery map $\varphi\colon \mathbb{K}^n \to \mathcal{G}$. In the randomized setting, N and φ can be random and the randomized error of the algorithm $A = \varphi \circ N$ is defined as

$$e^{\mathrm{ran}}(A) := \sup_{f \in F} \mathbb{E}\left[\|S(f) - A(f)\|_{\mathcal{G}}\right].$$

We refrain from a precise description of the setting and refer to [31, Section 4.3.3] and the references therein. We denote by $\mathcal{A}_n^{\mathrm{ran}}$ the class of all randomized algorithms with cardinality at most n. The algorithm A is called homogeneous if almost every realization of A is homogeneous.

Open Problem 2. Are homogeneous algorithms optimal for linear problems in the randomized setting? That is, do there exist constants $C, c \in \mathbb{N}$ such that, for all linear problems and all

11

$n \in \mathbb{N}$, we have

$$\inf_{\substack{A \in \mathcal{A}_{cn}^{\mathrm{ran}} \\ A \text{ homogeneous}}} e^{\mathrm{ran}}(A) \ \leq \ C \inf_{A \in \mathcal{A}_n^{\mathrm{ran}}} e^{\mathrm{ran}}(A) \ ?$$

Again, there is a relation to the question whether non-adaptive algorithms are optimal. But the answer to this question differs from the deterministic setting. It only turned out recently that there exist linear problems where adaptive randomized algorithms are significantly better than non-adaptive ones. The first such example was found by Heinrich in [18]. We refer to [19, 26] for further progress on this matter.

As a consequence, the optimality of homogeneous algorithms for linear problems is likely harder to prove in the randomized setting than it was in the deterministic setting (if possible at all). In the deterministic setting, we already knew that we could take N non-adaptive, and hence homogeneous, and only had to prove that we can also take φ homogeneous. In the randomized setting, it is already unclear (to us) whether an almost minimal error can be achieved with homogeneous N and arbitrary φ.

4   Solvable problems and problems on cones

Also in this section, our problems are given by a solution operator $S \colon \mathcal{F} \to \mathcal{G}$ between normed spaces $\mathcal{F}$ and $\mathcal{G}$ over $\mathbb{K}$, a class $\Lambda \subseteq \mathbb{K}^{\mathcal{F}}$ of admissible measurements, and a class $F \subset \mathcal{F}$ of inputs. In the previous sections, we studied linear and homogeneous problems in the worst case setting. Those problems were solvable in the following sense.

Definition 1. *We say that the problem* $(S \colon \mathcal{F} \to \mathcal{G}, F, \Lambda)$ *is* uniformly solvable *iff, for any* $\varepsilon > 0$, *there is some* $n \in \mathbb{N}$ *and an algorithm* A_n *of the form* (2) *such that* $\mathrm{err}(A_n) \leq \varepsilon$. *In this case, we define the* ε-complexity *of the problem by*

$$\mathrm{comp}(\varepsilon) \ := \ \mathrm{comp}(\varepsilon, S, F, \Lambda) \ := \ \min\{n \in \mathbb{N} \mid \exists A_n \colon \mathrm{err}(A_n) \leq \varepsilon\}.$$

We now turn to problems which are not uniformly solvable. Specifically, we are interested in the case that the input set $F \subset \mathcal{F}$ is a cone. Then, in general, the worst case error of any algorithm using homogeneous measurements with fixed n is infinite. More precisely, the following holds.

Proposition 2. *Let* $S \colon \mathcal{F} \to \mathcal{G}$ *be a homogeneous mapping between normed spaces, let* Λ *be a class of homogeneous functionals on* $\mathcal{F}$, *and let* $F \subset \mathcal{F}$ *be a cone. Then the problem* (S, F, Λ) *is uniformly solvable if and only if there exists some* $n \in \mathbb{N}$ *such that* $\mathrm{err}(n) = 0$.

Proof. We prove the stronger statement that any algorithm with finite error can be turned into an algorithm with arbitrarily small error and the same cost. Due to (6), this is equivalent to saying that every information mapping of the form (1) with finite diameter can be turned into an information mapping with the same cost and a diameter arbitrarily close to zero. To this end, assume that N is an information mapping of the form (1) such that $\mathrm{diam}(N) < \infty$. Given $r > 0$, we consider the new information mapping N_r, which chooses those measurement maps L_i that the mapping N would choose for rf instead of f. Note that the measurement results of rf are available due to $L(rf) = rL(f)$ for all $L \in \Lambda$ and $f \in F$. If $f, g \in F$ are such that $N_r(f) = N_r(g)$, then this implies $N(rf) = N(rg)$. Thus,

$$\|Sf - Sg\|_{\mathcal{G}} \ = \ \frac{1}{r}\|S(rf) - S(rg)\|_{\mathcal{G}} \ \leq \ \frac{\mathrm{diam}(N)}{r}$$

and hence the diameter of N_r is bounded by $\mathrm{diam}(N)/r$, which can be made arbitrarily small by choosing r sufficiently large. $\qquad \square$

In order to study problems that are not uniformly solvable, we consider a more general class of algorithms, where not only the measurements $L_i \in \Lambda$ themselves, but also the cardinalities of the algorithms, i.e., the number of those measurements, may be chosen adaptively. Roughly speaking this means that, after each measurement, we may decide based on the already computed measurements whether we compute another measurement and which one. Formally, those algorithms can be described as follows.

An *information mapping* is a function $N \colon F \to c_{00}$, where $c_{00} := \bigcup_{n \in \mathbb{N}} \mathbb{K}^n$, for which there are functions $L_j \colon F \times \mathbb{K}^{j-1} \to \mathbb{K}$ with $L_j(\cdot, y_1, \ldots, y_{j-1}) \in \Lambda$ for all $y_1, \ldots, y_{j-1} \in \mathbb{K}$ and a Boolean function $\Delta \colon c_{00} \to \{0, 1\}$ such that $N(f) = (y_1, \ldots, y_{n(f)})$ with $y_j = L_j(f, y_1, \ldots, y_{j-1})$ and $n(f) = \min\{n \in \mathbb{N} \mid \Delta(y_1, \ldots, y_n) = 0\}$. That is, if we have already obtained the data $y_1, \ldots, y_{j-1}$, we next perform the measurement $L_j(\cdot, y_1, \ldots, y_{j-1})$, unless $\Delta(y_1, \ldots, y_{j-1}) = 0$.

Now, an *algorithm A* for the problem (S, F, Λ) is defined as a pair (N, φ) given by an information mapping N as above and an arbitrary mapping $\varphi \colon c_{00} \to \mathcal{G}$, which is used to turn our finite information $N(f) \in c_{00}$ into an approximation of the solution $S(f) \in \mathcal{G}$. This results in a mapping $\varphi \circ N \colon F \to G$ used to approximate the solution operator $S \colon F \to \mathcal{G}$. In slight abuse of notation, we denote this mapping also by A, that is, we write $A = \varphi \circ N$. We denote the class of all such algorithms by $\mathcal{A}^+$. We write $\operatorname{cost}(A, f) := n(f)$ for the (information) *cost* of the algorithm at $f \in F$, and for $F_0 \subseteq F$, we set

$$\operatorname{cost}(A, F_0) := \sup\{\operatorname{cost}(A, f) \mid f \in F_0\}.$$

We are now ready to define a weaker notion of solvability.

Definition 2. *We say that an algorithm $A \in \mathcal{A}^+$ is ε-approximating for S on F, iff $\|Sf - Af\|_{\mathcal{G}} \leq \varepsilon$ for all $f \in F$. The problem $(S \colon \mathcal{F} \to \mathcal{G}, F, \Lambda)$ is called weakly solvable if, for any $\varepsilon > 0$, there exists an ε-approximating algorithm $A \in \mathcal{A}^+$.*

This notion of solvability is inspired by [25], where an analogous notion of solvability is considered in a randomized setting. Note that an ε-approximating algorithm has finite cost for each input $f \in F$, but it may have infinite cost on the full input set F.

As mentioned above, we are mainly interested in problems where the input set is a cone. Problems on cones have lately been considered in [25] and [6], see also [20, 21]. Here, we study the following general kind of cones.

Assumption 1. Let $S \colon \mathcal{F} \to \mathcal{G}$ be a linear operator between normed spaces $\mathcal{F}$ and $\mathcal{G}$ and let Λ be a class of linear functionals on $\mathcal{F}$. We consider the problem $(S, \mathcal{C}_t, \Lambda)$ on the cone

$$\mathcal{C}_t := \left\{ f \in \mathcal{F} \colon \|f\|_{\mathcal{F}} \leq t \, \|Tf\|_{\mathcal{H}} \right\},$$

where $T \colon \mathcal{F} \to \mathcal{H}$ is any linear mapping to another normed space $\mathcal{H}$ and $t > 0$ is a constant, which we call the *inflation factor*.

This setting covers a situation considered in [25], where $S(f)$ is the integral of a function $f \in L_2$ and $T \colon L_4 \to L_2$ is the identity. Then, essentially, $\mathcal{C}_t$ is the cone of all functions with bounded kurtosis. The setting also covers a situation considered in [6], where S is a diagonal operator mapping a Schauder basis of $\mathcal{F}$ onto a Schauder basis of $\mathcal{G}$ and $T \colon \mathcal{F} \to \mathcal{F}$ is a basis projection with finite rank. We will come back to these examples later.

The problem $(S, \mathcal{C}_t, \Lambda)$ as defined in Assumption 1 is included in the assumptions of Proposition 2. Therefore, it is usually not uniformly solvable. Here, we show that the problem is weakly solvable for a wide range of operators S and T. We recall that we write $B_{\mathcal{F}}$ to denote the unit ball in the space $\mathcal{F}$, and by $r B_{\mathcal{F}}$ the ball with radius r.

Theorem 2. *Let Assumption 1 hold. If $(S, B_{\mathcal{F}}, \Lambda)$ and $(T, B_{\mathcal{F}}, \Lambda)$ are uniformly solvable, then the problem $(S, \mathcal{C}_t, \Lambda)$ is weakly solvable for any $t > 0$. More precisely, for any $\varepsilon > 0$, there is an ε-approximating algorithm $A_\varepsilon \in \mathcal{A}^+$ such that, for all $f \in \mathcal{C}_t \setminus \{0\}$,*

$$\mathrm{cost}(A_\varepsilon, f) \leq \mathrm{comp}\left(\frac{1}{5t}, T, B_{\mathcal{F}}, \Lambda\right) + \mathrm{comp}\left(\frac{\varepsilon}{7t\,\|Tf\|_{\mathcal{H}}}, S, B_{\mathcal{F}}, \Lambda\right). \qquad (15)$$

To interpret the cost bound, let us assume that the operator T is bounded and that $\|T\|$ and t are absolute constants (and not very big). Corollary 1 implies that $\mathrm{comp}(\widetilde{\varepsilon}/R, S, B_{\mathcal{F}}, \Lambda) \leq \mathrm{comp}(\widetilde{\varepsilon}, S, 2RB_{\mathcal{F}}, \Lambda)$. By putting $c = 14t\|T\|$, we get for small enough $\varepsilon > 0$ (assuming that $\mathrm{comp}(\varepsilon, S, B_{\mathcal{F}}, \Lambda) \to \infty$ for $\varepsilon \to 0$) that

$$\mathrm{cost}(A_\varepsilon, \mathcal{C}_t \cap rB_{\mathcal{F}}) \leq 2\,\mathrm{comp}(\varepsilon, S, crB_{\mathcal{F}}, \Lambda),$$

i.e., the cost of the algorithm on a ball is proportional to the complexity of the problem S on a ball of comparable radius.

Proof. Let $\varepsilon > 0$ and $t > 0$. Due to the solvability of $(T, B_{\mathcal{F}}, \Lambda)$ and Theorem 1, there is a homogeneous algorithm $Q_m \colon \mathcal{F} \to \mathcal{H}$ that uses at most m measurements such that

$$\|Tf - Q_m f\|_{\mathcal{H}} \leq (2t)^{-1} \cdot \|f\|_{\mathcal{F}}, \quad \forall f \in \mathcal{F},$$

where $m = \mathrm{comp}(1/(5t), T, B_{\mathcal{F}}, \Lambda)$. This implies that, for all $f \in \mathcal{C}_t$, we have

$$\|f\|_{\mathcal{F}} \leq t\,\|Tf\|_{\mathcal{H}} \leq t\,\|Q_m f\|_{\mathcal{H}} + t\,\|Tf - Q_m f\|_{\mathcal{H}} \leq t\,\|Q_m f\|_{\mathcal{H}} + \frac{1}{2}\|f\|_{\mathcal{F}}$$

and thus

$$\|f\|_{\mathcal{F}} \leq 2t\,\|Q_m f\|_{\mathcal{H}}.$$

Our algorithm consists of two steps. First, given the unknown $f \in \mathcal{C}_t$, we compute $Q_m f$. If $Q_m f = 0$, then we know that $f = 0$ and thus obtain the exact solution $S(f) = 0$. In case that $Q_m f \neq 0$, we adaptively choose the cost $k := \mathrm{comp}(\varepsilon/((4 + 2\delta)t\|Q_m f\|_{\mathcal{H}}), S, B_{\mathcal{F}}, \Lambda)$ for the second step, where $\delta > 0$ is arbitrary. By the solvability of $(S, B_{\mathcal{F}}, \Lambda)$ and Theorem 1, there exists a homogeneous algorithm $A_k \colon \mathcal{F} \to \mathcal{G}$ that uses at most k measurements, whose worst case error on $B_{\mathcal{F}}$ is at most $\varepsilon/(2t\,\|Q_m f\|_{\mathcal{H}})$. In particular, for any $f \in \mathcal{C}_t$,

$$\|Sf - A_k f\|_{\mathcal{G}} \leq \frac{\varepsilon}{2t\,\|Q_m f\|_{\mathcal{H}}} \cdot \|f\|_{\mathcal{F}} \leq \varepsilon.$$

Thus, our two-step procedure gives us an ε-approximating algorithm that uses in total $m + k$ measurements from Λ. Noting that

$$\|Q_m f\|_{\mathcal{H}} \leq \|Tf - Q_m f\|_{\mathcal{H}} + \|Tf\|_{\mathcal{H}} \leq \frac{\|f\|_{\mathcal{F}}}{2t} + \|Tf\|_{\mathcal{H}} \leq \frac{3}{2}\|Tf\|_{\mathcal{H}},$$

we have

$$k \leq \mathrm{comp}\left(\frac{\varepsilon}{3(2 + \delta)t\,\|Tf\|_{\mathcal{H}}}, S, B_{\mathcal{F}}, \Lambda\right),$$

because $\mathrm{comp}(\cdot, S, B_{\mathcal{F}}, \Lambda)$ is decreasing in its first argument. The desired cost bound follows. $\square$

Remark 8. In the proof of Theorem 2, we used homogeneous algorithms Q_m and A_k which are optimal in the sense of the complexities of the problems of approximating T and S for certain error thresholds. Since such optimal algorithms are often not known in practice, we point out that the same approach works with arbitrary homogeneous algorithms of our choice. We can take any homogeneous algorithm Q_m for approximating T with arbitrary cost m and worst case error smaller than $1/(2t)$ and then any homogeneous algorithm A_k for S with arbitrary cost k and worst case error smaller than $\varepsilon/(2t\|Q_m f\|_{\mathcal{H}})$, and take $A_k f$ as an approximation for Sf (or zero, if $Q_m f = 0$). This gives us an ε-approximating algorithm on $\mathcal{C}_t$ with the total cost $m + k$.

We now discuss several examples.

4.1 Bounded kurtosis

Consider the the integration problem

$$S\colon \mathcal{F} \to \mathbb{R}, \quad S(f) = \int_0^1 f(x)\, \mathrm{d}x,$$

where $\mathcal{F}$ shall be the space of all continuous real-valued functions on $[0,1]$, equipped with the 4-norm. The input set is given by the cone

$$\mathcal{C}_t := \{ f \in \mathcal{F} \colon \|f\|_4 \le t \|f\|_2 \}$$

with some $t > 1$ and the information is given by the class Λ^{std} of function evaluations.

The paper [25] proves that such problems are weakly solvable in a randomized sense. In our deterministic setting, we quickly realize that the problems $S\colon \mathcal{F} \to \mathbb{R}$ and $T\colon \mathcal{F} \to L_2$ are not uniformly solvable on the unit ball of $\mathcal{F}$ and thus Theorem 2 does not apply. And indeed, the integration problem on $\mathcal{C}_t$ is not weakly solvable. In fact, there exists no ε-approximating algorithm for any $\varepsilon > 0$:

To see this, let $A \in \mathcal{A}^+$ be an arbitrary algorithm and let $\varepsilon > 0$ be arbitrary. We let $x_1, \ldots, x_n \in [0,1]$ be the measurement points that the algorithm uses for the input zero. Then, for any $\delta \in (0,1)$, one can easily find a continuous function $f\colon [0,1] \to [0, 4\varepsilon]$ such that f equals zero at all those points and the Lebesgue measure of the set $f^{-1}(4\varepsilon)$ is at least $1 - \delta$. We get

$$\|f\|_4 \le 4\varepsilon \quad \text{and} \quad \|f\|_2 \ge 4\varepsilon \cdot (1 - \delta)^{1/2}$$

and thus $f \in \mathcal{C}_t$ if δ is sufficiently small. Since $A(f) = A(0)$ and $S(f) - S(0) \ge 4\varepsilon \cdot (1 - \delta)$, the algorithm has an error of at least $2\varepsilon(1 - \delta)$ for one of the functions f or zero. Thus, the algorithm is not ε-approximating for the integration problem on $\mathcal{C}_t$.

In the next section, we present a simple example that follows a similar pattern, but which is weakly solvable.

4.2 Inverse Poincaré

To give another explicit example, let us consider the approximation problem

$$S\colon W_2^1([0,1]) \to L_2([0,1]), \quad S(f) = f,$$

where $W_2^1([0,1])$ is the univariate Sobolev space of all continuous functions $f\colon [0,1] \to \mathbb{R}$ that possess a weak derivative $f' \in L_2([0,1])$, equipped with the norm $\|f\|_{W_2^1}^2 = \|f\|_2^2 + \|f'\|_2^2$. The input set is given by the cone

$$\mathcal{C}_t := \left\{ f \in W_2^1([0,1]) \colon \|f'\|_2 \le t \|f\|_2 \right\}$$

of all functions from $\mathcal{F}$ satisfying a reverse Poincaré type inequality with constant $t > 0$ and the information is given by the class Λ^{std} of all function evaluations. This situation matches Assumption 1 if $\mathcal{F} = W_2^1$, $\mathcal{G} = \mathcal{H} = L_2$, and $T = S$, where the cone $\mathcal{C}_t$ from above is contained in the cone from (17) with inflation factor $\sqrt{1 + t^2}$.

It is a well-known fact that the problem S is uniformly solvable on balls in $\mathcal{F}$ and that there is a constant $c > 0$ such that

$$\mathrm{comp}(\varepsilon, S, B_{\mathcal{F}}, \Lambda^{\mathrm{std}}) \le c\varepsilon^{-1}$$

for all $\varepsilon > 0$, see, e.g., [31, Section 4.2.4]. Thus, Theorem 2 yields that the L_2-approximation problem is weakly solvable on the full cone $\mathcal{C}_t$ and gives rise to ε-approximating algorithms A_ε that satisfy

$$\text{cost}(A_\varepsilon, f) \leq 14\, ct \cdot \varepsilon^{-1} \|f\|_2$$

for all $f \in \mathcal{C}_t$ with $\|f\|_2 \geq \varepsilon$.

We conclude this example by noting that also the integration problem is weakly solvable on $\mathcal{C}_t$. The algorithm A_ε^* defined by $A_\varepsilon^*(f) := \int_0^1 A_\varepsilon(f)(x)\,\mathrm{d}x$ satisfies the same cost bound and is an ε-approximating algorithm for the integration problem on $\mathcal{C}_t$ since

$$\left| \int_0^1 f(x)\,\mathrm{d}x - \int_0^1 A_\varepsilon(f)(x)\,\mathrm{d}x \right| \leq \int_0^1 |f(x) - A_\varepsilon(f)(x)|\,\mathrm{d}x \leq \big\| f - A_\varepsilon(f) \big\|_2.$$

4.3 Diagonal operators between Banach spaces

We now consider a more general example analogous to [6].

Let $\mathcal{F}$ and $\mathcal{G}$ be two Banach spaces and let $S \colon \mathcal{F} \to \mathcal{G}$ be a continuous linear operator. We assume that there exist unconditional Schauder bases $\{u_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \mathbb{I}}$ of $\mathcal{F}$ and $\{v_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \mathbb{I}}$ of $\mathcal{G}$ such that

$$S(u_{\boldsymbol{k}}) = v_{\boldsymbol{k}}, \quad \forall \boldsymbol{k} \in \mathbb{I}. \tag{16}$$

By the spectral theorem, this situation occurs, for example, if $\mathcal{F}$ and $\mathcal{G}$ are separable Hilbert spaces and S is a compact and injective linear operator with a dense range. In this case, we can choose $\{u_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \mathbb{I}}$ as an orthonormal basis of eigenfunctions of S^*S. But also in the general case, all functions $f \in \mathcal{F}$ and $g \in \mathcal{G}$ can be written as a convergent series,

$$f = \sum_{\boldsymbol{k} \in \mathbb{I}} (f, u_{\boldsymbol{k}})_{\mathcal{F}} \cdot u_{\boldsymbol{k}}, \qquad g = \sum_{\boldsymbol{k} \in \mathbb{I}} (g, v_{\boldsymbol{k}})_{\mathcal{G}} \cdot v_{\boldsymbol{k}},$$

where the coefficients $(f, u_{\boldsymbol{k}})_{\mathcal{F}}$ and $(g, v_{\boldsymbol{k}})_{\mathcal{G}}$ are unique. We have for all $f \in \mathcal{F}$ that

$$S(f) = \sum_{\boldsymbol{k} \in \mathbb{I}} (f, u_{\boldsymbol{k}})_{\mathcal{F}} \cdot v_{\boldsymbol{k}}.$$

Let $\mathbb{I}_0 \subseteq \mathbb{I}$ be a finite index set. We denote the span of the functions $u_{\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{I}_0$, by $\mathcal{F}_0$ and the span of the functions $v_{\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{I}_0$, by $\mathcal{G}_0$. Moreover, we consider the basis projections $T \colon \mathcal{F} \to \mathcal{F}_0$ and $Q \colon \mathcal{G} \to \mathcal{G}_0$, given by

$$T(f) = \sum_{\boldsymbol{k} \in \mathbb{I}_0} (f, u_{\boldsymbol{k}})_{\mathcal{F}} \cdot u_{\boldsymbol{k}}, \qquad Q(g) = \sum_{\boldsymbol{k} \in \mathbb{I}_0} (g, v_{\boldsymbol{k}})_{\mathcal{G}} \cdot v_{\boldsymbol{k}},$$

which are bounded linear operators. Let further S_0 be the restriction of S to $\mathcal{F}_0$. Consider the cone $\mathcal{C}_t$ defined analogously to [6, Equation (7.18)], i.e.,

$$\mathcal{C}_t := \{ f \in \mathcal{F} \colon \|f\|_{\mathcal{F}} \leq t \|Tf\|_{\mathcal{F}} \}, \quad t \geq 1. \tag{17}$$

While this problem was dealt with in [6] for the case of arbitrary linear information $\Lambda = \Lambda^{\mathrm{all}}$, it remained an open question how to deal with it if one allows only standard information, which seems to be more relevant for practical cases. In our approach, any $\Lambda \subseteq \Lambda^{\mathrm{all}}$ is allowed, and we may consider $\Lambda = \Lambda^{\mathrm{std}}$ whenever $\mathcal{F}$ is a space of functions.

The problem matches our Assumption 1 if we let $\mathcal{H} = \mathcal{F}$. In particular, Theorem 2 applies. That is, the problem is weakly solvable on $\mathcal{C}_t$, whenever S and T are solvable on the unit ball of $\mathcal{F}$. We want to argue that the solvability of S implies the solvability of T and simplify the

cost bound from Theorem 2. To this end, let A_n be an algorithm for S with fixed cost n. We put $M_n = S_0^{-1} Q A_n$. This is an algorithm for the approximation of T that uses at most n measurements. We have

$$\|Tf - M_n f\|_{\mathcal{F}} \leq \|S_0^{-1}\| \cdot \|STf - QA_n f\|_{\mathcal{G}} \overset{ST=QS}{\leq} \|S_0^{-1}\| \cdot \|Q\| \cdot \|Sf - A_n f\|_{\mathcal{G}}.$$

This means that

$$\mathrm{comp}(\varepsilon, T, B_{\mathcal{F}}, \Lambda) \leq \mathrm{comp}\left(\frac{\varepsilon}{\|S_0^{-1}\| \cdot \|Q\|}, S, B_{\mathcal{F}}, \Lambda \right).$$

Summing up, we obtain the following corollary.

Corollary 2. *Consider the problem $(S, \mathcal{C}_t, \Lambda)$ as defined in this section. If S is uniformly solvable on the unit ball, then S is weakly solvable on $\mathcal{C}_t$. There exist ε-approximating algorithms A_ε using information from Λ such that, for all $f \in \mathcal{C}_t$, we have*

$$\mathrm{cost}(A_\varepsilon, f) \leq 2 \cdot \mathrm{comp}\left(\min\left\{ \varepsilon_0, \frac{c \cdot \varepsilon}{\|f\|_{\mathcal{F}}} \right\}, S, B_{\mathcal{F}}, \Lambda \right), \tag{18}$$

where

$$\varepsilon_0 := (5t \cdot \|S_0^{-1}\| \cdot \|Q\|)^{-1} \quad \text{and} \quad c := (7t\,\|T\|)^{-1}.$$

Let us see how this formula looks like in the Hilbert space setting mentioned after equation (16), where S is a compact operator between Hilbert spaces. Here, $\|T\| = \|Q\| = 1$ and $\|S_0^{-1}\|$ is the reciprocal of the smallest singular number $\sigma_{\min}$ of the mapping S_0. Thus, we have

$$\mathrm{cost}(A_\varepsilon, f) \leq 2\,\mathrm{comp}\left(\frac{\varepsilon}{7t\|f\|_{\mathcal{F}}}, S, B_{\mathcal{F}}, \Lambda \right)$$

for all $0 < \varepsilon < \sigma_{\min} \cdot \|f\|_{\mathcal{F}}$.

4.3.1 Simple cones based on a pilot sample

In the paper [6], a setting as in the beginning of Section 4.3 is considered, but with more particular requirements regarding the norms in $\mathcal{F}$ and $\mathcal{G}$. Indeed, let us additionally assume that

$$\|f\|_{\mathcal{F}} := \left\| \left(\frac{(f, u_{\boldsymbol{k}})_{\mathcal{F}}}{\lambda_{\boldsymbol{k}}} \right)_{\boldsymbol{k} \in \mathbb{I}} \right\|_\rho, \quad \text{for some } \rho \in [1, \infty],$$

$$\|g\|_{\mathcal{G}} := \left\| \left((g, v_{\boldsymbol{k}})_{\mathcal{G}} \right)_{\boldsymbol{k} \in \mathbb{I}} \right\|_\tau, \quad \text{for some } \tau \in [1, \rho],$$

where the $\lambda_{\boldsymbol{k}}$ are positive reals, which we assume to be ordered:

$$\lambda_{\boldsymbol{k}_1} \geq \lambda_{\boldsymbol{k}_2} \geq \lambda_{\boldsymbol{k}_3} \geq \cdots > 0.$$

In this context, we write $\left\| (\boldsymbol{x}_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathbb{I}} \right\|_\rho$ to denote the ℓ_ρ-norm of the sequence $(\boldsymbol{x}_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathbb{I}}$, and analogously for ℓ_τ. We again consider $S : \mathcal{F} \to \mathcal{G}$ such that $S(u_{\boldsymbol{k}}) = v_{\boldsymbol{k}}$ for all $\boldsymbol{k} \in \mathbb{I}$, and choose $\mathbb{I}_0$ as the set of integers $\{1, 2, \ldots, n_1\}$. Then the cone in (17) takes the special form

$$\mathcal{C}_t = \left\{ f \in \mathcal{F} \colon \|f\|_{\mathcal{F}} \leq t \left\| \left(\frac{(f, u_{\boldsymbol{k}_i})_{\mathcal{F}}}{\lambda_{\boldsymbol{k}_i}} \right)_{1 \leq i \leq n_1} \right\|_\rho \right\}.$$

17

This means that the cone $\mathcal{C}_t$ consists of functions whose norm can be bounded in terms of the pilot sample $(f, u_{\boldsymbol{k}_1})_{\mathcal{F}}, \ldots, (f, u_{\boldsymbol{k}_{n_1}})_{\mathcal{F}}$, and the mapping T is given by

$$T(f) = \sum_{i=1}^{n_1} (f, u_{\boldsymbol{k}_i})_{\mathcal{F}} \cdot v_{\boldsymbol{k}_i}.$$

The paper [6] then studies an adaptive algorithm A_n, defined by

$$A_n(f) = \sum_{i=1}^{n} (f, u_{\boldsymbol{k}_i})_{\mathcal{F}} \cdot v_{\boldsymbol{k}_i},$$

where $n = n(f, \varepsilon)$ is allowed to depend on $f \in \mathcal{F}$ and on an error threshold $\varepsilon > 0$, and in general n may be different from n_1. The definition of A_n requires that the information class Λ allows access to $(f, u_{\boldsymbol{k}_i})_{\mathcal{F}}$. Note that this is in general not fulfilled if we consider function spaces and the information class Λ^{std}, but for the sake of comparison let us consider the case that Λ contains the measurements $(f, u_{\boldsymbol{k}_i})_{\mathcal{F}}$.

Using the notation in Theorem 2 and its proof, we are allowed to approximate T by $Q_m = T$, where $m = n_1$, because we assume that we have access to the $(f, u_{\boldsymbol{k}_i})_{\mathcal{F}}$. Hence we obviously have $\mathrm{comp}(\varepsilon, T, B_{\mathcal{F}}, \Lambda) \leq n_1$ for any error threshold ε.

Then Theorem 2 or Corollary 2 implies the existence of an ε-approximating algorithm A_ε such that

$$\mathrm{cost}(A_\varepsilon, f) \leq \mathrm{comp}\left(\frac{\varepsilon}{7t\,\|Tf\|_{\mathcal{F}}}, S, B_{\mathcal{F}}, \Lambda\right) + n_1.$$

As outlined in [6], we have

$$\mathrm{comp}\left(\frac{\varepsilon}{7t\,\|Tf\|_{\mathcal{F}}}, S, B_{\mathcal{F}}, \Lambda\right) = \min\left\{n \in \mathbb{N}_0 \colon \|((\lambda_{\boldsymbol{k}})_{\boldsymbol{k} \geq n})\|_{\rho'} \leq \frac{\varepsilon}{7t\,\|Tf\|_{\mathcal{F}}}\right\}, \qquad (19)$$

where ρ' is such that $1/\rho + 1/\rho' = 1/\tau$. In order to determine the minimum in (19), one needs to find the minimal n such that

$$\|Tf\|_{\mathcal{F}} \cdot \|((\lambda_{\boldsymbol{k}})_{\boldsymbol{k} \geq n})\|_{\rho'} = \left\|\left(\frac{(f, u_{\boldsymbol{k}_i})_{\mathcal{F}}}{\lambda_{\boldsymbol{k}_i}}\right)_{1 \leq i \leq n_1}\right\|_{\rho} \cdot \|((\lambda_{\boldsymbol{k}})_{\boldsymbol{k} \geq n})\|_{\rho'} \leq \frac{\varepsilon}{7t}.$$

This corresponds to the analysis done in [6, Proof of Theorem 7.5] and leads (up to constants) to the same result. Indeed, it is shown in that paper that the cost is essentially optimal. Furthermore, using these findings, one can then analyze the problem from the viewpoint of Information-Based Complexity, and do a tractability analysis as discussed in [6]; differences in the constant are due to the fact that the results in [6] are tailored to this particular situation, whereas Theorem 2 and Corollary 2 are formulated in a much more general way.

A drawback of the approach just outlined is that we need information that is in general not contained in Λ^{std}, but we need direct access to the $(f, u_{\boldsymbol{k}_i})_{\mathcal{F}}$. If this information is not available, then one way is to approximate the $(f, u_{\boldsymbol{k}_i})_{\mathcal{F}}$, possibly by only using information from Λ^{std}, which may be more practical for applications. In this case, Q_m will in general no longer be equal to T as above, but a proper approximation of T. Theorem 2 and Corollary 2 enable us to analyze this situation. We present an example of such a situation in the following section.

4.3.2 Simple cones based on a pilot sample: L_2-approximation in weighted Korobov spaces using standard information

We now give a concrete example of the situation described in Section 4.3.1, but with an algorithm that only uses information from Λ^{std}.

Let $\mathcal{G} = L_2([0,1]^d)$ and let $\mathcal{F} \subset L_2([0,1]^d$ be the weighted Korobov space of one-periodic functions on $[0,1]^d$ with smoothness $\alpha > 1/2$, i.e.,

$$\mathcal{F} = \left\{ f \colon f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \widehat{f}(\boldsymbol{k}) e^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \quad \|f\|_{\mathcal{F}} := \left\| \left(\widehat{f}(\boldsymbol{k}) \sqrt{r_{2\alpha,\boldsymbol{\gamma}}(\boldsymbol{k})} \right)_{\boldsymbol{k}\in\mathbb{Z}^d} \right\|_2 < \infty \right\}.$$

Here, $\boldsymbol{\gamma} = (\gamma_j)_{j\geq 1}$ is a non-increasing sequence of reals in $(0,1]$, and

$$r_{2\alpha,\boldsymbol{\gamma}}(\boldsymbol{k}) = \prod_{j=1}^d r_{2\alpha,\gamma_j}(k_j), \quad \text{with} \quad r_{2\alpha,\gamma_j}(k_j) = \begin{cases} 1 & \text{if } k_j = 0, \\ \gamma_j^{-1} \, |k_j|^{2\alpha} & \text{otherwise,} \end{cases}$$

for $\boldsymbol{k} = (k_1, \ldots, k_d)$ (note that we always have $r_{2\alpha,\boldsymbol{\gamma}}(\boldsymbol{k}) \geq 1$). The space $\mathcal{F}$ is a function class commonly considered in the literature on quasi-Monte Carlo rules, see, e.g., [5].

Let $S(f) = f$, i.e., S is the embedding of $\mathcal{F}$ in $\mathcal{G}$. This means that in the notation of the previous section we have $\rho = \tau = 2$, $\mathbb{I} = \mathbb{Z}^d$, $u_{\boldsymbol{k}} = v_{\boldsymbol{k}} = e^{2\pi\mathrm{i}\boldsymbol{k}\cdot\circ}$, and

$$\lambda_{\boldsymbol{k}} = (r_{2\alpha,\boldsymbol{\gamma}}(\boldsymbol{k}))^{-1/2}, \quad \text{for } \boldsymbol{k} \in \mathbb{Z}^d.$$

For a constant $M > 1$, let

$$\mathbb{I}_0 = \mathcal{A}_{d,M} := \{\boldsymbol{k} \in \mathbb{Z}^d \colon r_{2\alpha,\boldsymbol{\gamma}}(\boldsymbol{k}) \leq M\},$$

so $\mathcal{A}_{d,M}$ contains the indices $\boldsymbol{k}$ corresponding to the "largest" Fourier coefficients of a given $f \in \mathcal{F}$, or putting it differently, $\mathcal{A}_{d,M}$ contains the collection of those $|\mathcal{A}_{d,M}|$ indices $\boldsymbol{k}$ for which the values of $\lambda_{\boldsymbol{k}} = \sqrt{r_{2\alpha,\boldsymbol{\gamma}}(\boldsymbol{k})}$ are the largest. In the notation of Section 4.3.1, $|\mathcal{A}_{d,M}| = n_1$.

Let the cone $\mathcal{C}_t$ be defined as in (17), which in this case yields

$$\mathcal{C}_t = \left\{ f \in \mathcal{F} \colon \|f\|_{\mathcal{F}} \leq t \left\| \left(\widehat{f}(\boldsymbol{k}) \sqrt{r_{2\alpha,\boldsymbol{\gamma}}(\boldsymbol{k})} \right)_{\boldsymbol{k}\in\mathcal{A}_{d,M}} \right\|_2 \right\}.$$

Let also the mappings T, Q, and S_0 be defined as in the beginning of Section 4.3.

As shown in [6], an—in the sense of the cost—optimal algorithm using information from Λ^{all} would be to approximate $f \in \mathcal{C}_t$ by the truncated Fourier series

$$A_n(f) = \sum_{\boldsymbol{k}\in\mathcal{A}_{d,M'}} \widehat{f}(\boldsymbol{k}) e^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}},$$

with $M' = M'(\varepsilon, f)$ chosen adaptively, and $\mathcal{A}_{d,M'}$ defined anaolously to $\mathcal{A}_{d,M}$ which exactly corresponds to the algorithm A_n in Section 4.3.1.

However, we would now like to take a different route, where we only use information from Λ^{std}. Recall that it is described in Remark 8 how we can find an ε-approximating algorithm on the cone $\mathcal{C}_t$ using only information from Λ^{std}. The algorithm consists of two successive parts:

- a homogeneous algorithm Q_m for approximating T whose worst case error on the unit ball is smaller than the constant $1/(2t)$,

19

- and a homogeneous algorithm A_k for approximating S whose worst case error on the unit ball is smaller than $\varepsilon/(2t\|Q_m f\|_{\mathcal{H}})$.

For the first part, we can consider an algorithm $Q_m = \widetilde{A}_m$ that uses an N-point rank-1 lattice rule to approximate Fourier coefficients of $f \in \mathcal{C}_t$, and then replaces the $\widehat{f}(\boldsymbol{k})$ by these approximations. Such algorithms were analyzed in, e.g., [27], see also [5] for an overview. The integration nodes used in the lattice rule are given by

$$\boldsymbol{x}_\ell = \left\{ \frac{\ell \boldsymbol{g}}{N} \right\}, \quad 0 \le \ell \le N - 1,$$

where $\{y\} = y - \lfloor y \rfloor$ denotes the fractional part of a real y, and where $\boldsymbol{g}$ is suitably chosen *generating vector* of the lattice rule. Further details on the choice of such $\boldsymbol{g}$ can be found again in [27] and [5]. Then, $\widetilde{A}_m$ is defined by

$$(\widetilde{A}_m(f))(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \mathcal{A}_{d,M}} \left(\frac{1}{N} \sum_{\ell=0}^{N-1} f\left(\left\{ \frac{\ell \boldsymbol{g}}{N} \right\} \right) e^{-2\pi \mathrm{i} \ell \boldsymbol{k} \cdot \boldsymbol{g}/N} \right) e^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \tag{20}$$

for $f \in \mathcal{C}_t$ and $\boldsymbol{x} \in [0,1]^d$. Note that $\widetilde{A}_m$ is linear, and in particular homogeneous. Moreover, $\widetilde{A}_m$ uses $m = N$ function evaluations as information measurements.

For the algorithm A_k in the second part, one can again use an algorithm of the form (20) with index set $\mathcal{A}_{d,M'}$ instead of $\mathcal{A}_{d,M}$ and lattice size k instead of N, where M' and k are adapted to the outcome of $\widetilde{A}_m(f)$. Alternatively, we can use a least-squares estimator of the form

$$A_k(f) = \operatorname*{argmin}_{g \in \mathcal{T}(\mathcal{A}_{d,M'})} \sum_{i=1}^{k} |f(x_i) - g(x_i)|^2.$$

Here, $\mathcal{T}(I)$ denotes the space of all trigonometric polynomials with frequencies from $I \subset \mathbb{Z}^d$. This way, the information cost k will be much lower than with algorithms of the form (20), see [7, 24, 29] for upper bounds using least-squares and [3, 36] for lower bounds using lattices. The points x_i of the least-squares algorithm can simply be chosen as realizations of i.i.d. uniformly distributed random variables on $[0,1]^d$, see [24]. The drawback of this construction is that it only works with high probability and there is always a (very) small probability that the worst-case error of the resulting algorithm A_k will be larger than the desired threshold $\varepsilon/(2t\|Q_m f\|_{\mathcal{H}})$.

No matter how we decide, we end up with an ε-approximating algorithm using information from Λ^{std} such that (18) holds, with the same notation as in Corollary 2. We stress that this is an advancement in comparison to the results in [6], where there were no tools available to analyze algorithms and complexity for information from Λ^{std}.

4.3.3 Multi-layer cones

As mentioned in [6, Section 7.4], it may be advantageous to consider a definition of cones that modifies the definition (17), in order to keep better track of the decay of the quantities $(f, u_{\boldsymbol{k}})_{\mathcal{F}}$. This would help in limiting the cost of adaptive algorithms in situations where the $(f, u_{\boldsymbol{k}})_{\mathcal{F}}$ decay fast. Indeed, consider a partition of $\mathbb{I}$ into a countable number of pairwise disjoint subsets $(\mathbb{I}_j)_{j \ge 0}$, i.e., $\mathbb{I} = \bigcup_{j=0}^{\infty} \mathbb{I}_j$. For $j \ge 1$, define

$$T_j \colon \mathcal{F} \to \mathcal{F}, \quad T_j(f) := \sum_{\boldsymbol{k} \in \mathbb{I}_j} (f, u_{\boldsymbol{k}})_{\mathcal{F}} \cdot u_{\boldsymbol{k}}.$$

That is, the T_j are the basis projections corresponding to the finite index sets $\mathbb{I}_j$.

Then, as in [6, Section 7.4], one can modify (17) to

$$\widetilde{\mathcal{C}} := \left\{ f \in \mathcal{F} \colon \|T_{j+\ell}f\|_{\mathcal{F}} \leq tv^{\ell}\|T_j f\|_{\mathcal{F}}, \ \forall j, \ell \in \mathbb{N} \right\}, \tag{21}$$

where $v < 1 < t$. Opposed to the cones considered in (17), which were defined via a single reference layer, let us call $\widetilde{\mathcal{C}}$ a multi-layer cone.

One quickly realizes (using the triangle inequality) that such multi-layer cones are actually contained in a single-layer cone as in (17) with T being the basis projection for coefficients $\boldsymbol{k} \in \mathbb{I}_0 \cup \mathbb{I}_1$ and suitably chosen inflation factor. As such, solvability statements and upper bounds carry over from Corollary 2.

We leave it for future research to improve upon the bounds resulting from this inclusion. Noting that the information class Λ^{all} has been dealt with in [6], this problem is especially interesting for the case of information from the class Λ^{std}.

References

[1] N. S. Bakhvalov. On the optimality of linear methods for operator approximation in convex classes of functions (in Russian), Zh. Vychisl. Mat. Mat. Fiz. 11, 244–249, 1971.

[2] R. Bartle and L. Graves. Mappings between function spaces. Trans. Am. Math. Soc. 72, 400–413, 1952.

[3] G. Byrenheid, L. Kämmerer, T. Ullrich, T. Volkmer. Tight error bounds for rank-1 lattice sampling in spaces of hybrid mixed smoothness. Numer. Math., 136, 993–1034, 2017.

[4] J. Creutzig and P. Wojtaszczyk. Linear vs. nonlinear algorithms for linear problems, J. Complexity 20, 807–820, 2004.

[5] J. Dick, P. Kritzer, F. Pillichshammer. *Lattice Rules*. Springer, Cham, 2022.

[6] Y. Ding, F.J. Hickernell, P. Kritzer, S. Mak. Adaptive approximation for multivariate linear problems with inputs lying in a cone. In: F.J. Hickernell, P. Kritzer (eds.) *Multivariate Algorithms and Information-Based Complexity*, 109–145, DeGruyter, Berlin/Boston, 2020.

[7] M. Dolbeault, D. Krieg, and M. Ullrich. A sharp upper bound for sampling numbers in L_2. Appl. Comput. Harmon. Anal. 63, 113–134, 2023.

[8] D. L. Donoho. Statistical estimation and optimal recovery. The Annals of Statistics, 22(1), 238–270, 1994.

[9] R. DeVore, G. Petrova, P. Wojtaszczyk. Data assimilation and sampling in Banach spaces. Calcolo, 54:963–1007, 2017.

[10] S. Foucart, G. Paouris. Near-optimal estimation of linear functionals with log-concave observation errors. arXiv:2301.07228, 2023.

[11] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich. The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$, J. Complexity 26, 629–640, 2010.

[12] S. Foucart, H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, New York, 2013.

[13] P. Gaillard, S. Gerchinovitz, E. de Montbrun. Adaptive approximation of monotone functions. arXiv:2309.07530, 2023.

[14] S. Gal and C. A. Micchelli. Optimal sequential and non-sequential procedures for evaluating a functional. Appl. Anal. 10, 105–120, 1980.

[15] A. Yu. Garnaev and E. D. Gluskin. The widths of a Euclidean ball. Soviet Math. Dokl. 30 (1984), 200–204.

[16] T. Goda, P. L'Ecuyer. Construction-free median quasi-Monte Carlo Rules for function spaces with unspecified smoothness and general weights. SIAM J. Sci. Comp. 44, A2765—A2788, 2022.

[17] T. Goda, K. Suzuki, M. Matsumoto. A universal median quasi-Monte Carlo integration. arXiv:2209.13186, 2023.

[18] S. Heinrich. Randomized Complexity of Parametric Integration and the Role of Adaption I. Finite Dimensional Case arXiv:2306.13471, 2023.

[19] S. Heinrich. Randomized Complexity of Parametric Integration and the Role of Adaption II. Sobolev Spaces. arXiv:2306.13499, 2023.

[20] F.J. Hickernell, L.A. Jiménez Rugama. Reliable adaptive cubature using digital sequences. In: R. Cools, D. Nuyens (eds.), *Monte Carlo and quasi-Monte Carlo Methods 2014*, 367–383, Springer, Cham, 2016.

[21] F.J. Hickernell, L.A. Jiménez Rugama, D. Li. Adaptive quasi-Monte Carlo methods for cubature. In: J. Dick, F.Y. Kuo, H. Woźniakowski (eds.), *Contemporary Computational Mathematics—a celebration of the 80th birthday of Ian Sloan*, 597–619, Springer, Cham, 2018.

[22] B. S. Kashin. Diameters of some finite-dimensional sets and classes of smooth functions. Math. USSR, Izv., 11:317–333, 1977.

[23] N. P. Korneichuk. Optimization of active algorithms for recovery of monotonic functions from Hölder's class. J. Complexity 10, 265–269, 1994.

[24] D. Krieg and M. Ullrich. Function values are enough for L_2-approximation. Foundations of Computational Mathematics, 21(4), 1141-1151, 2021.

[25] R. J. Kunsch, E. Novak, D. Rudolf. Solvable integration problems and optimal sample size selection. J. Complexity 53, 40–67, 2019.

[26] R. J. Kunsch, E. Novak, M. Wnuk. Randomized approximation of summable sequences – adaptive and non-adaptive. arXiv:2308.01705, 2023.

[27] F.Y. Kuo, I.H. Sloan, H. Woźniakowski. Lattice rules for multivariate approximation in the worst case setting. In: H. Niederreiter, D. Talay (eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2004*, 289–330. Springer, Berlin, 2006.

[28] P. Mathé. *s*-Numbers in Information-Based Complexity. J. Complexity 6, 41–66, 1990.

[29] N. Nagel, M. Schäfer, and T. Ullrich. A new upper bound for sampling numbers. Found. Comput. Math. 22(2), 445–468, 2022.

[30] E. Novak. On the power of adaption. J. Complexity 12, 199–237, 1996.

[31] E. Novak, H. Woźniakowski. *Tractability of Multivariate Problems. Volume I: Linear Information.* EMS, Zürich, 2008.

[32] E.W. Packel. Linear problems (with extended range) have linear optimal algorithms. Aeq. Math. 31, 18–25, 1986.

[33] A. Pinkus. *n-Widths in Approximation Theory.* Springer, Berlin, 1985.

[34] L. Plaskota. *Noisy Information and Computational Complexity.* Cambridge University Press, Cambridge, UK, 1996.

[35] K.A. Sikorski. *Optimal Solution of Nonlinear Euqations.* Oxford University Press, Oxford, UK, 2001.

[36] S. A. Smolyak. Interpolation and quadrature formulas for the classes W_s^α and E_s^α. Dokl. Akad. Nauk SSSR, 131(5), 1028–1031, 1960.

[37] S. A. Smolyak. On optimal restoration of functions and functionals of them (in Russian), Candidate Dissertation, Moscow State University, 1965.

[38] J. F. Traub, H. Woźniakowski. *A General Theory of Optimal Algorithms.* Academic Press, NewYork, 1980.

[39] F. Voigtlaender. L^p sampling numbers for the Fourier-analytic Barron space. Submitted, 2022, https://arxiv.org/abs/2208.07605.

[40] A.G. Werschulz, H. Woźniakowski. Are linear algorithms always good for linear problems? Aeq. Math. 31, 202–211, 1986.

Authors' addresses:

David Krieg
Department of Analysis
Johannes Kepler University Linz
Altenbergerstr. 69, 4040 Linz, Austria
david.krieg@jku.at

Peter Kritzer
Johann Radon Institute for Computational and Applied Mathematics (RICAM)
Austrian Academy of Sciences
Altenbergerstr. 69, 4040 Linz, Austria
peter.kritzer@oeaw.ac.at