A numerical algorithm for zero counting. IV: An adaptive speedup

Josué Tonelli-Cueto*
Inria Paris & IMJ-PRG
Sorbonne Université
Paris, FRANCE
josue.tonelli.cueto@bizkaia.eu

Abstract

In this paper, we provide an adaptive version of the algorithm for counting real roots of real polynomial systems of Cucker, Krick, Malajovich and Wschebor. We show that, unlike the original algorithm, the adaptive version runs in finite expected time, while preserving the good properties of the original: numerically stable, highly parallelizable and good probabilistic run-time. Moreover, our probabilistic complexity analysis will be robust not limiting itself to KSS random polynomial systems.

1 Introduction

The trilogy of papers *A numerical algorithm for zero counting* [26, 27, 28] by Cucker, Krick, Malajovich and Wschebor is the first milestone of the so-called grid method [21, 22]. Their algorithm, from now on `CKMW`, is numerically stable, highly parallelizable and has a good run-time with high probability. The latter still holds under very general probabilistic assumptions [36, 37]. So, as with many trilogies in the last decade, we must answer the question: why does this classic trilogy need a sequel?

In one sentence: `CKMW` has infinite expected run-time (even for KSS polynomial systems). The shadow of this fact affects subsequent descendants of `CKMW` for computing the homology of real smooth projective varieties [29], basic semialgebraic sets [13] and general semialgebraic sets [15, 16] (cf. [76]). In this way, we cannot hope for finite expected run-times for the latter problems if we don't have such finite expected run-time for the most simple problem: counting real zeros.

In this sequel paper, we present an adaptive version of `CKMW`, which we call `aCKMW`, whose run-time has finite expectation, and that preserving all the nice properties of the original algorithm: numerical stability, highly parallelizability, and good probabilistic run-time. Our result is an step towards the holy grail of real numerical algebraic geometry: a numerical algorithm for computing the homology groups of semialgebraic sets in expected single exponential time. Unfortunately, our results are not enough to solve this problem.

Without more hesitation, let's state our main result in an informal way. To see the full technical version, see Theorem 2.21 in the next section where we explain in full technical detail all the claims of the theorem.

*This work and the author were supported by a postdoctoral fellowship of the 2020 "Interaction" program of the Fondation Sciences Mathématiques de Paris. Partially supported by ANR JCJC GALOP (ANR-17-CE40-0009), the PGMO grant ALMA, and the PHC GRAPE.

Theorem A. *There is a numerically stable algorithm* `aCKMW` *that given a real homogeneous polynomial system*

$$\begin{cases} f_1(X_0, X_1, \ldots, X_n) = & 0 \\ \qquad\qquad \vdots & \vdots \\ f_n(X_0, X_1, \ldots, X_n) = & 0 \end{cases}$$

with f_i of degree d_i computes the number of projective real roots f. Moreover, `aCKMW` *has the following properties:*

(i) It can be modified to produce approximations à la Smale *of all projective real roots, i.e., approximations such that the projective Newton's method converge quadratically at them.*

(ii) For a wide class of random polynomial systems, the expected run-time is

$$2^{\mathcal{O}(n \log n)} \mathbf{D}^n N + 2^{\mathcal{O}(n \log n)} (N + n\mathcal{D})\mathcal{D}^2),$$

where $\mathbf{D} := \max_{i=1}^n d_i$ is the maximum degree, $\mathcal{D} := \prod_{i=1}^n d_i$ the Bézout bound, and $N := \sum_{i=1}^n \binom{n+d_i}{n}$ is the number of coefficients of the system.

(iii) The algorithm can be parallelized so that it has

$$\mathcal{O}(n \log(n\mathbf{D}))$$

expected run-time with $2^{\mathcal{O}(n \log n)}(\mathbf{D}^n + \mathcal{D}^2)N$ expected number of processors, under the same probabilistic assumptions as in (ii).

(iv) The algorithm can be run in floating-point arithmetic (with adaptive precision) in a way to guarantee that the output is correct and preserving the above probabilistic complexity bounds up to a logarithmic factor.

As it happened at the time when the original `CKMW` appeared, `aCKMW` is not the first algorithm on the table for counting (and computing) real zeros of real polynomial systems. So why should we care about this new algorithm? Because this algorithm is the first numerically stable algorithm for this problem that has a proven finite expected run-time which is polynomial in the degree and quasi-linearly exponential in the number of variables.

Relation to other algorithms

To put our algorithm and its complexity analysis into context, we will review briefly the literature on algorithms for finding real zeros of real polynomial systems. Our purpose here is to showcase the differences and similarities of our algorithms to other algorithms.

In the symbolic realm, there are several efficient methods for counting zeros (cf. [3, Ch. 12]): critical points method [43, 18], rational univariate representations [64, 41, 9], Gröbner bases [46]... Some of these methods are implemented and work reasonably well in practice (`RootFinding[Isolate]` in `Maple` [64], `FGb` [38], `RAGlib` [68]). We can see that the complexity estimates of our algorithm are similar to those of the symbolic methods, which are of the form $\mathbf{D}^{\mathcal{O}(n)}$ (using the notations above). However, as it happened with the original algorithm, the adaptive version that we give is guaranteed to be numerically stable (at the price of not being able to handle ill-posed systems), unlike its symbolic alternatives.

In the numerical realm, homotopy continuation is the queen. To solve the considered problem, the main path would be to compute all the complex roots and then counts the real ones. To compute all the complex zeros, there are many ways: total degree homotopy [63] (cf. [12, 18.4.1]), real homotopies [55], repeatedly computing one random root until roots are obtained [5, §10.2], monodromy [33]... The majority of these methods are implemented in many ways (`PHCpack` [78], `Bertini` [4], `NAG4M2` [54] `Hom4PS-3` [19], `juliahomotopy.jl` [11]) and perform very well in practice. Moreover, one can even certify the final count of real zeros efficiently in practice using floating-point interval arithmetic [67, 53, 10] (cf. [59, Ch. 5]), or using the not so efficient rational arithmetic [47].

However, despite all the progress in numerical complexity theory on solving complex polynomial systems through homotopy continuation [12, III], the full resolution of the original Smale's 17th problem [51] and theoretical achievements beyond its scope [52, 14], the numerical complexity of finding the real zeros (or finding just one complex zero) of real polynomial systems through homotopy continuation (or any other numerical method) remains a widely open problem[1]. Hence, by means of our algorithm and its complexity analysis, we show for the first time that the problem of computing the real zeros of a real polynomial system can be solved numerically in finite expected time. In the future, we should aim at improving our complexity-theoretical knowledge of homotopy continuation, so that homotopy continuation becomes the complexity-theoretical queen also in the realm of real numerical algebraic geometry.

In the so-called symbolic-numerical realm, the class of relevant algorithms is those known as subdivision methods. In the univariate setting, subdivision methods are near optimal to compute the real roots of real univariate polynomials [60, 69] (cf. [50]). In the multivariate case, there are not near optimal algorithms and one must consider a larger zoo of subdivision methods. The simplest ones are based on some interval version of Newton's method [59, Ch. 5] or some test based on Miranda's theorem [80], which work not only for polynomials, but any C^1-function with simple zeros. A more sophisticated class exploits the Bernstein basis to exclude roots using linear programming [70], normal cones [35] or reduction-to-one-dimension techniques [58], and for including roots they exploit Miranda's theorem [39, 40]. We note that subdivision methods based in reduction techniques can be extended to more general functions [77] and to polynomials in the monomial basis [57]. In practice, these algorithms work reasonably well (see, in addition to the mentioned references, `GlobSol` [48], `INTLAB` [66] and `IntervalRootFinding.jl` [7]). However, the complexity of these algorithms is not well understood [81, p. 32]. In the cases where a complexity analysis is available, the complexity estimates seem to depend on quantities that are either unknown or hard to interpret (e.g. $N_{\mathcal{D}}(\mathbf{f})$ in [58, Theorem 5.7], the cost of the oracles in [57, Proposition 5.2.] or the λs in [80, §7]).

Anyone familiar with these subdivision methods described in the previous paragraph will notice that `aCKMW` is very similar to a subdivision method. This similarity is not surprising as the underlying idea of the grid method (i.e., approximate the space with a cloud of points to capture an object in it) is 'dual' to that of subdivision methods (i.e., subdivide the space into smaller regions). In particular, our algorithm is very similar to those subdivision methods of [59, Ch. 5], but it uses a variation of Smale's α-theory [74] (cf. [31]) instead of interval versions of Newton's method. In other words, our algorithm does not use any fancy methods beyond Newton's method in it. We note that our complexity analysis of `aCKMW` is better than the ones existing for subdivision methods in two ways: 1) the

[1] Up to the knowledge of this author and without considering papers related to the grid method, [8] and [61] are the only existing works considering the numerical complexity of solving real polynomial systems.

condition-based estimate is easy to interpret, and 2) the probabilistic estimate allows us to understand how the algorithm works in practice. We think that this kind of complexity analysis might help in understanding theoretically and not only in practice the subdivision methods.

Background, techniques and achievements

The core of the paper is to make the algorithm `CKMW` adaptive and analyze the complexity by taking this into account. However, the ideas of this paper didn't generate *ex nihilo*. Here, we describe briefly what was known and what this paper brings to the table in terms of techniques of complexity analysis.

The idea that an adaptive grid method might provide acceleration was already in the air when `CKMW` was proposed in [26]. However, there were two main obstacles: 1) how to find an efficient procedure to make the grid method adaptive, and 2) how to exploit the adaptive nature to the algorithm to obtain a better complexity estimate.

The existence of complexity analyses that could take advantage of the latter became clear while analyzing the complexity of the Plantinga-Vegter algorithm in [23] (cf. [25]). In this work, the continuous amortization technique of [17] made it clear that, for the adaptive grid method, the condition-based bound of the run-time should be in terms of an expression of the form

$$\mathbb{E}_{\mathfrak{x}\in\mathbb{S}^n}\kappa(f,\mathfrak{x})^n, \tag{1.1}$$

where $\kappa(f,x)$ is a local condition number measuring how near is f of being ill-posed around x. This 'averaged-over-the-sphere' expression contrasts with the 'supremum-over-the-sphere' expression,

$$\sup_{x\in\mathbb{S}^n}\kappa(f,x)^n, \tag{1.2}$$

used to bound the run-time of all previous algorithms based on the (non-adaptive) grid method. The change from (1.2) to (1.1) has important probabilistic consequences. While (1.2) does not have finite expectation[2] when we randomize f, (1.1) does indeed have finite expectation.

The above change in the bound can be seen as a real analogue of the change from non-adaptive homotopy to adaptive homotopy done by Shub [71] in complex numerical algebraic geometry. In the complex setting, this change meant passing from infinite variance to finite variance for the run-time of homotopy continuation [6]. As we have seen above, in the real setting, the change is more dramatic as we pass from infinite expected run-time to finite run-time. However, as it happened with [71], the above change does not come with an adaptive grid method.

The first construction of an algorithm using an adaptive grid method was done by Han [44, 45][3], but his algorithm has two major flaws: 1) it is not constructive, and 2) the bound of its run-time involves

$$\mathbb{E}_{\mathfrak{x}\in\mathbb{S}^n}\kappa(f,\mathfrak{x})^{2n}, \tag{1.3}$$

which does not have a finite expectation when we randomize f. Again, unlike in the complex setting, the real setting demands extra care. One can make a constructive grid

[2]Note that in previous works, one goes around the fact that (1.2) does not have finite expectation using variation of the notion of weak complexity [2]. In this way, one says that (1.2) is sufficiently small with high probability.

[3]We warn the reader about the numerous mistakes of these references. This means that any statement can be false beyond trivial corrections.

method, as done by the author in [76, Ch. $4^{\S 3}$] to show that one can estimate (1.2) in finite expected time[4]. Building on this work, Eckhardt [34] proposes a fully constructive version of Han's adaptive algorithm for computing the homology of basic semialgebraic sets[5]. Unfortunately, Eckhardt's algorithm's run-time's bound still involves (1.3), which does not have finite expectation (with respect a random f).

Hence, after all these developments, one can see that obtaining an adaptive grid method that works in finite expected time is non-trivial. The main reasons for this are the following:

(1) The procedure to select an approximating cloud of points from the adaptive grid seems to require a quadratic condition in $\kappa(f,x)$. The latter is enough to force (1.3) into the complexity estimates.

(2) The post-processing step of the approximating cloud of points has to be done in a way that it avoids pairwise comparisons among all points. If pairwise comparisons, or even worse, k-subset comparisons as it required for current homology computing algorithm, are required, the naive complexity estimates will be at quadratic in (1.1). A priori, we don't expect squares of (1.1) to behave a lot more differently than (1.3), and, in particular, to have finite expectations.

In general, the post-processing step is the hardest part of subdivision methods. For example, although one might think that the analysis here is similar to the one done for the Plantinga-Vegter algorithm [23] (cf. [25]), we note that there we didn't deal with the selection or processing step. Dealing with them in a way that the complexity estimates don't blow up can be a challenging problem. In particular, for the Plantinga-Vegter algorithm this is so due to the need of very exact sign evaluations [81, p. 32].

In this paper, we solve these two issues for zero counting. For (1), we recover the use of Smale's β in Smale's α-criterion without using bounds of the form

$$\beta(f,x) \leq \mu(f,x)\frac{\|f(x)\|}{\|f\|_W}, \tag{1.4}$$

which were the ones responsible for the quadratic condition in (1). The reason this change works is because we can guarantee that $\beta(f,x)$ is small, and so that Smale's α-criterion is satisfied, if x is near enough a root. For (2), we use strongly that the topology of a zero dimensional set is discrete. Now, doing this alone is not enough for the bounds in Theorem A. For this, we need to incorporate several tricks:

- *Change of norm*: Instead of using the Weyl norm $\| \ \|_W$, we use the real L_∞-norm $\| \ \|_\infty$ following the ideas in [24]. Doing this, allows us to have N instead of N^{n+1} is the estimate of the expected run-time.

- *Row normalization of the systems*: We normalize our polynomial system, equation by equation. In other words, instead of normalizing f as $f/\|f\|_\infty$, we normalize it as $(f_i/\|f_i\|)_i$. This substitutes a $\mathbf{D}^{2n}$ factor is the estimate of the run-time by $\mathcal{D}^2$ the significantly smaller.

- *Lipschitz-constant-decreasing normalization of condition*: Instead of dealing with local condition numbers $\kappa(f,x)$ that are $\mathbf{D}$-Lipschitz in x (2nd Lipschitz property), we

[4]This publication is the first one containing results related to that part of the PhD thesis of the author.

[5]Let us note that the work of Eckhardt was highly non-trivial, as it had to provide completely new lower bounds for the local reach of a basic semialgebraic sets in [44, 45] are completely wrong.

introduce additional normalization by the diagonal matrix of degrees Δ to force this functions to be 1-Lipschitz in x. As the operation above, this allows us to transform several $\mathbf{D}^n$ in the estimates by $\mathcal{D}$.

We note that the factor $\mathbf{D}^n$ comes from the need to compute $\| \ \|$, but this computation can be avoided at the price of turning `aCKMW` into a Montecarlo algorithm.

In the future, one should expect to extend the adaptive grid method to the computation of homology of algebraic and semialgebraic sets. The ideas exposed here generalize easily to solve (1) for this problem (at least in the algebraic setting). The main challenge remains in the solution of (2) when we cannot rely on the zero set being a discrete set. An alternative line of work is to consider the original application of the grid method into feasibility problems [30, 20], whose complexity continues open [12, P.18].

Structure of the paper

In the next section, we explain in full detail the content of our algorithm, the probabilistic assumptions and all the complexity estimates. Then in Section 3, we provide all the results for the variation of Smale's α-theory that we will be using; in Section 4, we analyze the condition-based and probabilistic complexity of the algorithm; and in section 6, we discuss briefly the finite precision of the algorithm.

2 Main Ingredients and Overview

Let $n, q \in \mathbb{N}$, $\boldsymbol{d} := (d_1, \ldots, d_q) \in \mathbb{N}^q$,

$$\mathcal{H}_{n,d_i} := \{g \in \mathbb{R}[X_0, \ldots, X_n] \mid g \text{ is homogeneous of degree } d_i\},$$

the set of real homogeneous polynomials of degree d_i in $X_0, \ldots, X_n$, and

$$\mathcal{H}_{n,\boldsymbol{d}}[q] := \prod_{i=1}^{q} \mathcal{H}_{n,d_i} = \{f \in \mathbb{R}[X_0, \ldots, X_n]^q \mid \text{for all } i,\ f_i \text{ is homogeneous of degree } d_i\},$$

the set of polynomial q-tuples f such that f_i is an homogeneous polynomial of degree d_i in $X_0, \ldots, X_n$. We also introduce the following matrix

$$\Delta = \operatorname{diag}(\boldsymbol{d}) = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_q \end{pmatrix} \in \mathbb{R}^{q \times q}. \tag{2.1}$$

Associated to the above object, we consider the following constans:

- $\mathbf{D} := \|\boldsymbol{d}\|_\infty = \max_{i=1}^q d_i$, the *maximum degree*.
- $\mathcal{D} := \prod_{i=1}^q d_i \leq \mathbf{D}^q$, the *Bézout bound*.
- $N := \sum_{i=1}^q \binom{d_i+n}{n} \leq q(1+\mathbf{D})^n$, the *input size*.

Given $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$, we will consider the following notations:

- $\partial^k f \in \mathbb{R}[X_0, \ldots, X_n]^{q \times (n+1)^{\otimes k}}$ is the tensor whose entries are given by the polynomials $\frac{\partial^k}{\partial X_{j_1} \cdot \partial X_{j_k}} f_i$. Note that given $v_1, \ldots, v_k \in \mathbb{R}^{n+1}$, $\partial^k f(v_1, \ldots, v_k)$ is a q-tuple of homogeneous polynomials (such that the ith component is of degree $\max\{0, d_i - k\}$).

- $\partial_x^k f$ is the value of $\partial^k f$ at $x \in \mathbb{R}^{n+1}$.
- $\mathrm{D}_x f : \mathrm{T}_x\mathbb{S}^n \to \mathbb{R}^q$ is the *tangent map* of f (as a map on the sphere) at $x \in \mathbb{S}^n$. By abuse of notation, we will write $\mathrm{D}_x f = \partial_x f(\mathbb{I} - xx^*)$ for $x \in \mathbb{S}^n$.

In what follows, we will describe the algorithm and the complexity results of this paper. Before that, we will introduce the main ingredients.

2.1 Main Ingredients

The main ingredients for our algorithm are the following ones: the L_∞-norm, the condition number, the δ-theory, and an adaptive grid. We will mainly do the exposition working on the sphere $\mathbb{S}^n$, as it is conceptually simpler. However, we note that all the relevant results can be adapted to the projective space, as the majority of the definitions are representative-invariant.

2.1.1 L_∞-norm and exclusion lemma

In previous works in real numerical algebraic geometry, the usual norm to use was the Weyl norm. The *Weyl norm* of $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ is given by

$$\|f\|_W := \sqrt{\sum_i \sum_{|\alpha|=d_i} \binom{d_i}{\alpha}^{-1} f_{i,\alpha}^2}$$

where $f_i = \sum_{|\alpha|=d_i} f_{i,\alpha} X^\alpha$. The disadvantage of this norm is the appearance of factors of the form $N^{\mathcal{O}(n)}$ in the complexity estimates of the algorithms. This can be avoided as shown in [24] by using other norms, such as the L_∞-norm. The *L_∞-norm* of $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ is given by

$$\|f\|_\infty := \max_{x\in\mathbb{S}^n} \|f(x)\|_\infty = \max_{x\in\mathbb{S}^n} \max_i |f_i(x)|.$$

We note that this norm is invariant not only under the action of the orthogonal group $O(n+1)$ on $\mathcal{H}_{n,\boldsymbol{d}}[q]$, but also under the action of $O(n+1)^d$ on $\mathcal{H}_{n,\boldsymbol{d}}[q]$. The following result by Kellogg [49, Theorem IV] (which we cite in the form given in [24, Corollary 2.20]) will be the most useful result for us.

Theorem 2.1 (Kellogg's Theorem). *Let $f \in \mathcal{H}_{n,d}$ be an homogeneous polynomial of degree d. Then for all $k \in \mathbb{N}$, $x \in \mathbb{S}^n$ and $v_1, \ldots, v_k \in \mathbb{R}^{n+1}$,*

$$\left| \frac{1}{k!} \partial_x^k f(v_1, \ldots, v_k) \right| \leq \binom{d}{k} \|f\|_\infty \|v_1\|_2 \cdots \|v_k\|_2.$$

In particular, $f : \mathbb{S}^n \to \mathbb{R}$ is $d\|f\|_\infty$-Lipschitz (with respect both the geodesic and Euclidean metrics in $\mathbb{S}^n$). □

The above theorem can be applied component by component to $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ in order to obtain an analogue of the exclusion lemma [26, Lemma 3.1]. Given $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$, the *row-normalization* of f is the following polynomial tuple

$$\hat{f} := (f_i / \|f_i\|_\infty)_i \in \mathcal{H}_{n,\boldsymbol{d}}[q]. \tag{2.2}$$

Note that the row-normalization is better than the naive normalization $f/\|f\|_\infty$, because it makes all polynomials of the same magnitude.

Proposition 2.2 (Exclusion lemma). *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$. Then $\Delta^{-1}\hat{f} : \mathbb{S}^n \to [0,1]$ is a well-defined 1-Lipschitz map (with respect both the geodesic and Euclidean metrics on $\mathbb{S}^n$). In particular, for $x \in \mathbb{S}^n$ such that $f(x) \neq 0$,*

$$B_{\mathbb{S}}\left(x, \left\|\Delta^{-1}\hat{f}(x)\right\|_\infty\right)$$

does not contain any zero of f. □

The above result will allow us to exclude points that are very far away from the zero set of f.

Remark 2.3. We note that [26] used originally the norm $\max_i \|f_i\|_W$ because it was friendly to the ∞-norm. Our choice for the L_∞-norm has this advantage, with the addition of a better probabilistic behaviour.

2.1.2 (Re-scaled) condition number

Given $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$, the usually used condition numbers in numerical algebraic geometry are

$$\mu(f,x) := \|f\|_W \left\|\mathrm{D}_x f^\dagger \Delta^{\frac{1}{2}}\right\|_{2,2} \quad \text{and} \quad \kappa(f,x) := \left(\|f(x)\|_2^2/\|f\|_W^2 + \mu(f,x)^{-2}\right)^{-\frac{1}{2}},$$

where $A^\dagger := A^*(AA^*)^{-1}$ is the pseudoinverse, $\mu(f,x) = \infty$ if $\mathrm{D}_x f$ is not surjective (by convention), and $\| \ \|_{2,2}$ is the induced norm with the Euclidean norm $\| \ \|_2$ in both the domain and codomain. The μ-condition number controls the conditioning of f around a root x, the κ-condition number controls the conditioning of f around any point x. Because of this, the μ-condition number tends to appear in the complex setting and the κ-condition number in the real one.

In [24, §3], these norms are adapted to the L_∞-norm by changing $\|f\|_W$ by $\sqrt{q}\|f\|_\infty$ and $\Delta^{\frac{1}{2}}$ by Δ. However, the scaling kept still favors the geometric approach of [12, Ch. 14] instead of a more complexity-focused approach. To correct this, we introduce the following variant of the μ-condition.

Definition 2.4. Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$. The *ν-condition number* of f at x is the quantity given by

$$\nu(f,x) := \|f\|_\infty \left\|\mathrm{D}_x^{-1} f^\dagger \Delta^2\right\|_{\infty,2},$$

where $\| \ \|_{\infty,2}$ is the operator norm with the ∞-norm in the domain and the Euclidean norm in the codomain, if $\mathrm{D}_x f$ is surjective, and ∞, otherwise.

Remark 2.5. Note that for the row-normalization of f, we have $\nu(\hat{f},x) = \left\|\mathrm{D}_x^{-1}\hat{f}^\dagger\Delta^2\right\|_{\infty,2}$.

The importance of the ν-condition lies in its use for controlling Newton's method through the means of δ-theory. We discuss this in the following point. The theorem below shows the so-called Lipschitz properties of ν (term introduced in [76]). We note that our choice of scaling is so that the 2nd Lipschitz property has constant one. We postpone its proof until Section 3.

Theorem 2.6. *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$. Then the following holds:*

- ***1st Lipschitz property***. *For all* $g, \tilde{g} \in \mathcal{H}_{n,\boldsymbol{d}}[q]$,

$$\left| \frac{\|\tilde{g}\|_\infty}{\nu(\tilde{g}, x)} - \frac{\|g\|_\infty}{\nu(\tilde{g}, x)} \right| \leq \|\Delta^{-1}(\tilde{g} - g)\|_\infty.$$

  In particular, $\nu(f, x) \geq 1$.

- ***2nd Lipschitz property***. *The map*

$$\begin{aligned} \mathbb{S}^n &\to [0, 1] \\ x &\mapsto \frac{1}{\nu(f, x)} \end{aligned}$$

  is 1*-Lipschitz.*

The condition number that will play a role in controlling the complexity around each point will be the following one.

Definition 2.7. Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$. The $\mathtt{C}$*-condition number* of f at x is the quantity given by

$$\mathtt{C}(f, x) := \min \left\{ \frac{\|f\|_\infty}{\|\Delta^{-1} f(x)\|_\infty}, n\nu(f, x) \right\}.$$

Note that $\mathtt{C}(f, x)$ becomes ∞ if and only if x is a singular zero of f. This shows that $\mathtt{C}(f, x)$ controls the conditioning of f around x.

2.1.3 δ-Theory

Recall that the *Newton's spherical operator* for $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ is the partial map given by

$$\begin{aligned} \mathrm{N}_f : \mathbb{S}^n &\dashrightarrow \mathbb{S}^n \\ x &\mapsto \frac{x - \mathrm{D}_x f^\dagger f(x)}{\|x - \mathrm{D}_x f^\dagger f(x)\|}. \end{aligned}$$

One can easily see that N_f is defined on the domain

$$\operatorname{dom} \mathrm{N}_f := \{x \in \mathbb{S}^n \mid \mathrm{D}_x f \text{ is surjective}\},$$

and the fixed points of N_f are precisely the non-singular zeros of f. Moreover, note that N_f is equivalent to projecting onto $\mathbb{S}^n$ a Newton step for $f_{|\mathrm{T}_x\mathbb{S}^n}$.

Smale's α-theory [74] provides point-wise bounds to determine if there is a zero of an analytic function near given point and if convergence of Newton's method happens. This is done in term of three parameters for $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$:

- α*-estimate*:

$$\alpha(f, x) := \beta(f, x)\gamma(f, x).$$

- β*-estimate*:

$$\beta(f, x) := \|\mathrm{D}_x f^\dagger f(x)\|_2.$$

- γ-*estimate*:

$$\gamma(f,x) := \max\left\{1, \sup_{k\geq 2}\left\|\frac{1}{k!}\mathrm{D}_x f^\dagger \partial_x^k f\right\|_{2,2}^{\frac{1}{k-1}}\right\}$$

  where $\|\ \|_{2,2}$ is the operator norm of the multilinear map

$$(v_1,\ldots,v_k) \mapsto \frac{1}{k!}\mathrm{D}_x f^\dagger \partial_x^k f(v_1,\ldots,v_k)$$

  with respect the Euclidean norms.

We note that the β-estimate measures the length of the Newton step, in the sense that

$$\mathrm{dist}_{\mathbb{S}}(x, \mathrm{N}_f(x)) = \arctan \beta(f,x) \leq \beta(f,x). \tag{2.3}$$

We note that for small values of $\beta(f,x)$, the case in which we are interested, $\arctan\beta(f,x)$ and $\beta(f,x)$ are the same for any practical effects. More precisely,

$$|\beta(f,x) - \arctan\beta(f,x)| \leq \frac{1}{3}\beta(f,x)^3. \tag{2.4}$$

Let us state Smale's α-theorem in a simple way (with explicit, although not optimal constants) for the spherical case.

Theorem 2.8 (α-theorem). *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$. If $\alpha(f,x) \leq 1/20$, then:*

1. *$\{\mathrm{N}_f^k(x)\}_{k\in\mathbb{N}}$ is a well-defined convergent sequence.*
2. *The limit point of $\{\mathrm{N}_f^k(x)\}_{k\in\mathbb{N}}$, $N_f^\infty(x)$, is a non-singular zero of f.*
3. *For all $k \geq 0$, $\mathrm{dist}_{\mathbb{S}}\left(N_f^k(x), N_f^\infty(x)\right) \leq \frac{3}{2}\left(\frac{1}{2}\right)^{2^k-1}\beta(f,x)$. In particular,*

$$\mathrm{dist}_{\mathbb{S}}\left(x, N_f^\infty(x)\right) \leq \frac{3}{2}\beta(f,x) \leq \frac{3}{40}\frac{1}{\gamma(f,x)}.$$

□

Following previous work, in particular [26], we would use the estimate

$$\beta(f,x) \leq \overline{\beta}(f,x) := \nu(f,x)\frac{\|\Delta^{-1}f(x)\|_\infty}{\|f\|_\infty} \tag{2.5}$$

for Smale's β, and the estimate

$$\gamma(f,x) \leq \overline{\gamma}(f,x) := \frac{1}{2}(\mathbf{D}-1)\nu(f,x), \tag{2.6}$$

the latter a variant of the Higher Derivative Estimate [12, Theorem 16.1] (for the proof in this setting, see [24, §3]), for Smale's γ. These two estimates together are not enough to obtain finite expected run-time. The reason for this is that $\overline{\beta}(f,x)\overline{\gamma}(f,x)$ can only be guaranteed to be sufficiently small if

$$(\mathbf{D}-1)\nu(f,x)^2\,\mathrm{dist}_{\mathbb{S}}(x, \mathcal{Z}_{\mathbb{S}}(f)) < \frac{1}{10},$$

thus forcing (1.3) into our estimates.

A priori, the above situation might seem unavoidable given that Smale's α-criterion is a sufficient condition. However, there is no need to estimate Smale's β using $\overline{\beta}$, we can just compute β without affecting the overall complexity. When we do this, the following converse (whose proof we left for Section 3) shows that if x is sufficiently near the zero set, the α-test must hold.

Proposition 2.9 (Converse of Smale's α-theorem). *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$. If $\gamma(f,x)\operatorname{dist}_{\mathbb{S}}(x,\mathcal{Z}_{\mathbb{S}}(f)) < 1$, then*

$$\alpha(f,x) \leq \frac{\gamma(f,x)\operatorname{dist}_{\mathbb{S}}(x,\mathcal{Z}_{\mathbb{S}}(f))}{1-\gamma(f,x)\operatorname{dist}_{\mathbb{S}}(x,\mathcal{Z}_{\mathbb{S}}(f))}.$$

In particular, if $(\mathbf{D}-1)\nu(f,x)^2\operatorname{dist}_{\mathbb{S}}(x,\mathcal{Z}_{\mathbb{S}}(f)) < 1/11$, then $\alpha(f,x) < \beta(f,x)\overline{\gamma}(f,x) < 1/20$.

However, the Higher Derivative Estimate would force several powers of $\mathbf{D}^n$ into the complexity. To avoid this, and obtain the best possible bound, we develop the δ-theory which works exclusively in terms of the ν-condition without involving Smale's γ at all. The letter 'δ', in δ-theory, is in honor of Dedieu, in whose exposition [31] we base our adaptation.

We will prove in Section 3 the following two adaptation of the two results above. Note that the constants in the results are not optimal.

Theorem 2.10 (δ-theorem). *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$ such that $\nu(f,x) < \infty$. If*

$$\delta(f,x) := \nu(f,x)\beta(f,x) \leq \frac{1}{4},$$

then:

1. *$\{\mathrm{N}_f^k(x)\}_{k\in\mathbb{N}}$ is a well-defined convergent sequence.*
2. *The limit point of $\{\mathrm{N}_f^k(x)\}_{k\in\mathbb{N}}$, $N_f^\infty(x)$, is a non-singular zero of f.*
3. *For all $k \geq 0$,*

$$\operatorname{dist}_{\mathbb{S}}\left(N_f^k(x), N_f^\infty(x)\right) \leq \frac{3}{2}\left(\frac{3}{4}\right)^k\left(\frac{1}{3}\right)^{2^k-1}\beta(f,x) \leq \frac{3}{8}\left(\frac{3}{4}\right)^k\left(\frac{1}{3}\right)^{2^k-1}\frac{1}{\nu(f,x)}.$$

In particular,

$$\operatorname{dist}_{\mathbb{S}}\left(x, N_f^\infty(x)\right) \leq \frac{3}{2}\beta(f,x) \leq \frac{3}{8}\frac{1}{\nu(f,x)}.$$

Theorem 2.11 (Converse of the δ-theorem). *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$, $x \in \mathbb{S}^n$ such that $\nu(f,x) < \infty$ and $c > 0$. If there is a regular zero $\zeta \in \mathbb{S}^n$ of f such that*

$$\operatorname{dist}_{\mathbb{S}}(x,\zeta) < \frac{c}{\nu(f,x)},$$

then

$$\delta(f,x) \leq c\left(1+\frac{c}{2}\right).$$

Corollary 2.12. *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$ such that $\nu(f,x) < \infty$. If*

$$\overline{B}_{\mathbb{S}}\left(x, \frac{1}{5}\frac{1}{\nu(f,x)}\right) \cap \mathcal{Z}_{\mathbb{S}}(f) \neq \varnothing,$$

then $\delta(f,x) \leq \frac{1}{4}$. □

The following technical proposition is needed to guarantee that, in the zero dimensional case, the above balls not only contain one zero, but that they don't contain more than one zero. Its proof is at the end of Section 3.

Proposition 2.13. *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ and $x \in \mathbb{S}^n$ such that $\nu(f,x) < \infty$. Then*

$$\overline{B}_{\mathbb{S}}\left(x, \frac{1}{3}\frac{1}{\nu(f,x)}\right)$$

contains at most one zero of f.

2.1.4 An adaptive grid

The way in which our grid will be constructed is by considering a grid of points in the boundary of the cube, which we will refine locally in the boundary of the cube itself. We only project the points to the sphere when an evaluation is needed, which allows to store the points in the cube with exact floating-point representation.

Consider the n-cube $[-1,1]$ and together with it the following set of $2(n+1)$-maps $\text{Ю}_{k,\sigma} : [-1,1]^n \to \mathbb{S}^n$, where $(k,\sigma) \in \{0,\dots,n\} \times \{+1,-1\}$, and given by

$$\text{Ю}_{k,\sigma}(x) := \frac{1}{\sqrt{1+\|x\|_2^2}} \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ \sigma \\ x_i \\ \vdots \\ x_n. \end{pmatrix} \tag{2.7}$$

Note that all the above maps together is the same as considering the boundary fo the $(n+1)$-cube, $\partial[-1,1]^{n+1}$, together with the central projection, $x \mapsto x/\|x\|$, onto the sphere $\mathbb{S}^n$. In this way, we have that the $\text{Ю}_{k,\sigma}[-1,1]^n$ cover the whole $\mathbb{S}^n$.

In the original [26], one would consider a uniform grid in $[-1,1]^n$ and then project it onto $\mathbb{S}^n$ to obtain the desired grid on the sphere. Making these concrete, with a small variation, one would consider the grid

$$\mathcal{U}_\ell := [-1,1]^n \cap 2^{-\ell}\,(1+2\mathbb{Z})^n\,. \tag{2.8}$$

The main property of this grid is that every point $x \in \mathbb{S}^n$, there is some $(p,k,\sigma) \in \mathcal{U}_\ell \times \{0,\dots,n\} \times \{-1,1\}$ such that $\text{dist}_{\mathbb{S}}(x, \text{Ю}_{k,\sigma}(p)) < \sqrt{n}2^{-\ell}$.

In [26], the refinement of the grid was global. In other words, one would just pass from $\mathcal{U}_\ell^{\mathbb{S}^n}$ to $\mathcal{U}_{\ell+1}^{\mathbb{S}^n}$ whenever the verification of the conditions failed at one point (even if they failed at just one point). To make this process adaptive, we do the above process locally, i.e., we only refine at points were things went wrong.

In our construction, we will not only require that we cover the sphere. We will also require to have an efficient cover of the cubes involved, like in subdivision methods such as the Plantinga-Vegter algorithm [62]. Based on this, we introduce the following definition.

Definition 2.14. An *adaptive cubical grid* is a subset

$$\mathcal{G} \subset [-1,1]^n \times \mathbb{N} \times \{0,\dots,n\} \times \{-1,1\}$$

such that for all $(k,\sigma) \in \{0,\dots,n\} \times \{-1,1\}$,

$$\mathcal{B}_{k,\sigma} := \left\{ \overline{B}_\infty \left(x, 2^{-\ell} \right) \mid (x,\ell,k,\sigma) \in \mathcal{G} \right\},$$

where $\overline{B}(y,r) := \{x \in \mathbb{R}^n \mid \|y-x\|_\infty \leq r\}$, is a subdivision of $[-1,1]^n$. The latter means that $\bigcup \mathcal{B}_{k,\sigma} = [-1,1]^n$ and that for any $B, \tilde{B} \in \mathcal{B}_{k,\sigma}$, $B \cap \tilde{B}$ has volume zero.

For a point (x,ℓ,k,σ) in such an adaptive cubical grid, we consider the following *refinement operator*:

$$\mathtt{R}(x,\ell,k,\sigma) := \left\{ \left(x + 2^{-\ell-1} v, \ell+1, k, \sigma \right) \mid v \in \{-1,1\}^n \right\}. \tag{2.9}$$

Geometrically, $\mathtt{R}(x,\ell,k,\sigma)$ produces a subdivision of $\overline{B}_\infty \left(x, 2^{-\ell}\right)$ into 2^n equal cubes. Also note that

$$\mathcal{U}_{\ell+1} \times \{\ell+1\} \times \{0,\dots,n\} \times \{-1,1\} = \bigcup_{(x,\ell,k,\sigma) \in \mathcal{U}_\ell \times \{\ell\} \times \{0,\dots,n\} \times \{-1,1\}} \mathtt{R}(x,\ell,k,\sigma),$$

which shows why $\mathtt{R}$ is a local refinement of the global refinement described above.

The following proposition is the geometric reason of why our adaptive grid covers the full space.

Proposition 2.15. *(i) For every $\ell \in \mathbb{N}$, $\mathcal{U}_\ell \times \{\ell\} \times \{0,\dots,n\} \times \{-1,1\}$ is an adaptive cubical grid.*

(ii) For every adaptive cubical grid $\mathcal{G}$ and every $(x,\ell,k,\sigma) \in \mathcal{G}$,

$$(\mathcal{G} \setminus \{(x,\ell,k,\sigma)\}) \cup \mathtt{R}(x,\ell,k,\sigma)$$

is an adaptive cubical grid.

(iii) If $\mathcal{G}$ is an adaptive cubical grid, then

$$\left\{ \overline{B}_{\mathbb{S}} \left(\mathrm{IO}_{k,\sigma}(x), \sqrt{n} 2^{-\ell} \right) \mid (x,\ell,k,\sigma) \in \mathcal{G} \right\},$$

is a cover of $\mathbb{S}^n$.

2.2 The algorithm

We now have all ingredients to present our adaptive version of `aCKMW`. The latter is given in pseudocode in page 14. For comparison, we have also given the original `CKMW` in pseudocode in page 15, written in the best way for comparison. Let us now give the definition of approximations *à la Smale*.

Definition 2.16. Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ and $\zeta \in \mathcal{Z}_{\mathbb{S}}(f)$. An *approximation à la Smale* of ζ is some $(x,r) \in \mathbb{S}^n \times [0,1]$ such that not only $\{\mathrm{N}_f^k(z)\}$ converges to ζ, where N_f is Newton's spherical operator, but such that for all $k \in \mathbb{N}$,

$$\mathrm{dist}_{\mathbb{S}}(\mathrm{N}_f^k(z), \zeta) \leq \left(\frac{1}{2} \right)^{2^k - 1} r.$$

Algorithm 2.1: aCKMW

Input : $f \in \mathcal{H}_{n,\boldsymbol{d}}[n]$
Precondition : All the zeros of f are simple

/* Computation of the $\|f_i\|_\infty$ */
forall $i \in \{1,\dots,n\}$ **do**
  $\ell \leftarrow 1 + \lceil \log(n) + \log d_i \rceil$
  $Q_i \leftarrow \left(1 - \frac{1}{8n}\right) \max\{|f_i(\mathrm{IO}_{k,+1}(x))| \mid (x,k) \in \mathcal{U}_\ell \times \{0,\dots,n\}\}$
$\boldsymbol{Q} \leftarrow \mathrm{diag}(Q_1,\dots,Q_n)$
/* Initialization of the grid */
$\ell \leftarrow \lceil \frac{1}{2}\log(n) \rceil$
$\mathcal{G} \leftarrow \mathcal{U}_\ell \times \{\ell\} \times \{0,\dots,n\} \times \{+1\}$ // $\mathcal{U}_\ell$ def. in (2.8)
/* Exclusion-Inclusion-Refinement Cycle */
$\tilde{Z} \leftarrow \varnothing$
repeat
  Take $(x,\ell,k,+1) \in \mathcal{G}$
  $\xi \leftarrow \mathrm{IO}_{k,+1}(x)$ // $\mathrm{IO}_{k,+1}$ def. in (2.7)
  /* Exclusion test */
  if *for some* i, $2^{\ell-1}|f_i(\xi)| \geq \sqrt{n} d_i Q_i$ **then**
    $\mathcal{G} \leftarrow \mathcal{G} \setminus \{(x,\ell,k,+1)\}$
  /* Inclusion test */
  else if $6\left(1-\frac{1}{8n}\right)^{-1} n^{3/2} \left\|\mathrm{D}_\xi f^\dagger \Delta^2 \boldsymbol{Q}\right\|_{\infty,\infty} \leq 2^\ell$ **then**
    $\delta \leftarrow 5\sqrt{n} \left\|\mathrm{D}_\xi f^\dagger \Delta^2 \boldsymbol{Q}\right\|_{\infty,\infty} \left\|\mathrm{D}_x f^\dagger f(x)\right\|_2$ // Δ def. in (2.1)
    if $\delta < 1$ **then**
      $\tilde{Z} \leftarrow \tilde{Z} \cup \left\{\left(\xi, 1.5\left\|\mathrm{D}_x f^\dagger f(x)\right\|_2\right)\right\}$
    $\mathcal{G} \leftarrow \mathcal{G} \setminus \{(x,\ell,k,+1)\}$
  /* Refinement */
  else
    $\mathcal{G} \leftarrow (\mathcal{G} \setminus \{(x,\ell,k,+1)\}) \cup \mathtt{R}(x,\ell,k,+1)$ // R def. in (2.9)
until $\mathcal{G} = \varnothing$
/* Elimination of redundant approximations */
$Z \leftarrow \varnothing$
forall $(x,r) \in \tilde{Z}$ **do**
  $\tilde{Z} \leftarrow \tilde{Z} \setminus \{(x,r)\}$
  $Z \leftarrow Z \cup \{(x,r)\}$
  forall $(y,s) \in \tilde{Z}$ **do**
    if $\mathrm{dist}_{\mathbb{S}}(x,y) < r+s$ ***or*** $\mathrm{dist}_{\mathbb{S}}(x,-y) < r+s$ **then**
      $\tilde{Z} \leftarrow \tilde{Z} \setminus \{(y,s)\}$
/* Return of the approximation of the real zero set */
return Z

Output : Finite set $Z \subseteq \mathbb{S}^n \times [0,1]$
Postcondition: $\#Z = \#\mathcal{Z}_{\mathbb{P}}(f)$
For all $(z,r) \in Z$, (z,r) approximates some $\zeta \in \mathcal{Z}_{\mathbb{S}}(f)$ *à la Smale*

Algorithm 2.2: CKMW

Input : $f \in \mathcal{H}_{n,\boldsymbol{d}}[n]$
Precondition : All the zeros of f are simple

/* Computation of $\|f\|_W$ */
$Q \leftarrow \|f\|_W$
/* Initialization of the grid */
$\ell \leftarrow \lceil \frac{1}{2}\log(n) \rceil$
/* Exclusion-Inclusion-Refinement Cycle */
repeat
  $b \leftarrow$ `true`, $\mathcal{G} \leftarrow \mathcal{U}_\ell \times \{0,\dots,n\}$ // $\mathcal{U}_\ell$ def. in (2.8)
  $\tilde{Z} \leftarrow \varnothing$
  repeat
    Take $(x,k) \in \mathcal{G}$
    $\xi \leftarrow \mathrm{IO}_{k,+1}(x)$ // $\mathrm{IO}_{k,+1}$ def. in (2.7)
    /* Exclusion test */
    if $2^\ell \|f(\xi)\|_2 \geq \sqrt{n\mathbf{D}}\, Q$ **then**
      $\mathcal{G} \leftarrow \mathcal{G} \leftarrow \setminus\{(x,k)\}$
    /* Inclusion test */
    else if $2^{\ell+2}\mathbf{D}^{3/2}Q^2\|\mathrm{D}_\xi f^\dagger \Delta^{1/2}\|_{2,2}^2 \leq \sqrt{n}$ **then**
      $\overline{\alpha} \leftarrow 10Q\mathbf{D}^{3/2}\|\mathrm{D}_\xi f^\dagger \Delta^{1/2}\|_{2,2}^2 \|f(\xi)\|_2$ // Δ def. in (2.1)
      if $\overline{\alpha} < 1$ **then**
        $\tilde{Z} \leftarrow \tilde{Z} \cup \{(\xi, 2\|\mathrm{D}_\xi f^\dagger \Delta^{1/2}\|_{2,2}\|f(\xi)\|_2)\}$
      $\mathcal{G} \leftarrow \mathcal{G} \setminus \{(x,k)\}$
    /* Refinement */
    else
      $\ell \leftarrow \ell + 1$, $b \leftarrow$ `false`, $\mathcal{G} = \varnothing$
  until $\mathcal{G} = \varnothing$
until $b =$ `true`
/* Elimination of redundant approximations */
$Z \leftarrow \varnothing$
forall $(x,r) \in \tilde{Z}$ **do**
  $\tilde{Z} \leftarrow \tilde{Z} \setminus \{(x,r)\}$
  $Z \leftarrow Z \cup \{(x,r)\}$
  forall $(y,s) \in \tilde{Z}$ **do**
    if $\mathrm{dist}_{\mathbb{S}}(x,y) < r+s$ ***or*** $\mathrm{dist}_{\mathbb{S}}(x,-y) < r+s$ **then**
      $\tilde{Z} \leftarrow \tilde{Z} \setminus \{(y,s)\}$
/* Return of the approximation of the real zero set */
return Z

Output : Finite set $Z \subseteq \mathbb{S}^n \times [0,1]$
Postcondition: $\#Z = \#\mathcal{Z}_{\mathbb{P}}(f)$
For all $(z,r) \in Z$, (z,r) approximates some $\zeta \in \mathcal{Z}_{\mathbb{S}}(f)$ *à la Smale*

Note that the importance of approximations *à la Smale* is that they can be refined to any needed degree of precision fast. This is because, we can guarantee that the number of binary digits duplicates at each Newton iteration (quadratic convergence). Note that the δ-theorem (Theorem 2.10, provides a stronger convergence condition, but the given one is the usual one.

We let the technical details of `aCKMW` to Section 4 where we will show that the algorithm is correct. Now, we give the intuition of the algorithm, so that one can see how all the ingredients come together.

The algorithm is divided three main parts:

1. Computation of the norms
2. Cycle of Exclusion-Inclusion-Refinement
3. Elimination of redundant approximations

The computation of the norms is the most expensive part of the algorithm. While in `CKMW`, one only has to compute the Weyl norm; in `aCKMW`, we are required to compute the L_∞-norms. This amounts to maximize a polynomial in the sphere, and it is the responsible for the $\mathbf{D}^n$ in the complexity estimates. As shown in [24, §3], the L_∞-norms are needed for obtaining the best complexity bounds for the grid method, despite the difficulty to compute them.

In the cycle of Exclusion-Inclusion-Refinement, we do an exclusion test, an inclusion test, if the latter fails; and we refine, if both tests fails. The main idea is to make the grid finer and finer until we can certify at each point of the grid that either there are no roots around (exclusion) or that there is a root around (inclusion). For the exclusion test, we use the exclusion lemma (Proposition 2.2); for the inclusion test, we use the δ-theorem (Theorem 2.10); and for the refinement we use the refinement operator `R` in (2.9).

In the above cycle, we can see several important difference between `CKMW` and `aCKMW`:

- *Row-Normalization*: We observe that while the normalization in `CKMW` is of the form $f/\|f\|_W$, the normalization in `aCKMW` is the row-normalization defined in (2.2).
- *Use of Smale's β*: Although `aCKMW` uses δ-theory instead of α-theory, this is not the most important difference with `CKMW`. Note that δ-theory is nothing more than α-theory adapted to homogeneous polynomials on the sphere. The main difference is that `aCKMW` uses Smale's β directly, without estimating it with expression of the form (1.4) and (2.5).

  This difference between how β is used by both algorithm is the reason behind the different form of the inequalities in line 13 of `aCKMW` and line 11 of `CKMW`. These conditions allow us to certify two things: 1) there is at most one root in $\overline{B}_{\mathbb{S}}(\xi, \sqrt{n}2^{-\ell})$. 2) If there is such a root in $\overline{B}_{\mathbb{S}}(\xi, \sqrt{n}2^{-\ell})$ the inclusion test will be successful. Without these two guarantees we risk undercounting the number of real roots.
- *Local refinement*: In `CKMW`, if any of the two tests fail, even if it just fails at a single point, we refine the full grid. The latter means the lost of any information obtained through the execution of the algorithm. In `aCKMW`, we avoid this lost of information, through means of local refinements. Of course, this is the adaptive character of the algorithm.

In the elimination of redundant of redundant approximations, `aCKMW` and `CKMW` are very similar. The theoretical justification for the procedure in `aCKMW` is Proposition 2.13, which guarantees that each approximation has only one zero in the ball considered around. To avoid testing all possible pairs, we test one approximation against all other approximations, removing the ones that approximate the same root. Once we do this for one root, we don't have to test the rest of the approximations against the removed roots.

Remark 2.17. We observe that instead of computing the $\|\mathrm{D}_\xi f^\dagger \Delta^2 Q\|_{\infty,2}$ appearing in the definition of $\nu(Q^{-1}f, \xi)$, we are computing the estimate $\|\mathrm{D}_\xi f^\dagger \Delta^2 Q\|_{\infty,\infty}$. The reason for this is that while computing the first operator norm is computationally expensive, computing $\|\mathrm{D}_\xi f^\dagger \Delta^2 Q\|_{\infty,\infty}$ is very cheap.

Remark 2.18. Note that we are not considering points (x, ℓ, k, σ) only with $\sigma = +1$. This is because we are working in $\mathbb{P}^n$ and symmetry allows us to disregard half of the points to check.

Beyond the zero-dimensional case

If we are interested in the zero dimensional case, we can adapt turn `aCKMW` into `Sampling` effortlessly. Unfortunately, there is no way of redundant approximations. Combining the postcondition in `Sampling` with the results in [34, §2.3], where an adaptive version of the Niyogi-Smale-Weinberger theorem is proven following on work of [45, 44], we can guarantee topological correctness. By this, we mean that we can guarantee that the inclusion $\mathcal{Z}_{\mathbb{S}}(f) \subset \bigcup\{\overline{B}_{\mathbb{P}}(z, r) \mid (z, r) \in Z \text{ or } (-z, r) \in Z\}$ is an homotopy equivalence. Unfortunately, we are unaware of any way of exploiting this to obtain an algorithm that computes the homology or Betti numbers of smooth projective algebraic sets in finite expected time. We hope to explore this possibility, the main motivation of this paper, in future work.

The condition-based complexity analysis of `Sampling` is almost identical to that of `aCKMW` and we omit it. The probabilistic complexity analysis is harder than that of `aCKMW` but it can be done using similar techniques to the ones we use in this paper. However, we omit such proofs, as we plan to include probabilistic complexity analyses under more general hypotheses in future work.

2.3 Complexity results

In Section 4, we will perform a condition-based complexity of our algorithms in terms of

$$\mathbb{E}_{\mathfrak{r}\in\mathbb{S}^n} \mathtt{C}(f, \mathfrak{r})^n \log^l \mathtt{C}(f, \mathfrak{r}),$$

where `C` was given in Definition 2.7. In these estimates l will vary depending on whether we only count arithmetic operations (the main case that we will consider in Section 4) or if we count bit operations (up to some degree) of the finite precision version (discussed in Section 5).

Now, condition-based complexity bounds explain why and how a numerical algorithm is faster depending on the input. Unfortunately, these input-dependent estimates do not give a good idea about the complexity of an algorithm. A long tradition in numerical complexity, going back to Goldstine and von Neumann [42] and popularized by Demmel [32] and Smale [73, 72], is to randomize the input and consider the probabilistic behaviour of the algorithm. In this way, when we talk about expected run-time, we do so with respect a distribution or family of distributions of the input space.

Algorithm 2.3: `Sampling`

Input : $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$
Precondition : $\mathcal{Z}_{\mathbb{P}}(f)$ is smooth

/* Computation of the $\|f_i\|_\infty$ */
forall $i \in \{1, \dots, q\}$ **do**
  $\ell \leftarrow 1 + \lceil \log(n) + \log d_i \rceil$
  $\boldsymbol{Q}_i \leftarrow 2 \max\{|f_i(\mathrm{IO}_{k,+1}(x))| \mid (x,k) \in \mathcal{U}_\ell \times \{0, \dots, n\}\}$
$\boldsymbol{Q} \leftarrow \mathrm{diag}(\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_q)$
/* Initialization of the grid */
$\ell \leftarrow \lceil \frac{1}{2} \log(n) \rceil$
$\mathcal{G} \leftarrow \mathcal{U}_\ell \times \{\ell\} \times \{0, \dots, n\} \times \{+1\}$ // $\mathcal{U}_\ell$ def. in (2.8)
/* Exclusion-Inclusion-Refinement Cycle */
$Z \leftarrow \varnothing$
repeat
  Take $(x, \ell, k, +1) \in \mathcal{G}$
  $\xi \leftarrow \mathrm{IO}_{k,+1}(x)$ // $\mathrm{IO}_{k,+1}$ def. in (2.7)
  /* Exclusion test */
  if *for some* i, $2^{\ell-1}|f_i(\xi)| \geq \sqrt{n} d_i Q_i$ **then**
    $\mathcal{G} \leftarrow \mathcal{G} \setminus \{(x, \ell, k, +1)\}$
  /* Inclusion test */
  else if $6\left(1 - \frac{1}{8n}\right)^{-1} n^{3/2} \left\|\mathrm{D}_\xi f^\dagger \Delta^2 \boldsymbol{Q}\right\|_{\infty,\infty} \leq 2^\ell$ **then**
    $\delta \leftarrow 5\sqrt{n} \left\|\mathrm{D}_\xi f^\dagger \Delta^2 \boldsymbol{Q}\right\|_{\infty,\infty} \left\|\mathrm{D}_x f^\dagger f(x)\right\|_2$ // Δ def. in (2.1)
    if $\delta < 1$ **then**
      $Z \leftarrow Z \cup \left\{\left(\xi, \sqrt{n} 2^{-\ell}\right)\right\}$
    $\mathcal{G} \leftarrow \mathcal{G} \setminus \{(x, \ell, k, +1)\}$
  /* Refinement */
  else
    $\mathcal{G} \leftarrow (\mathcal{G} \setminus \{(x, \ell, k, +1)\}) \cup \mathtt{R}(x, \ell, k, +1)$ // R def. in (2.9)
until $\mathcal{G} = \varnothing$
/* Return of the redundant approximation of $\mathcal{Z}_{\mathbb{P}(f)}$ */
return Z

Output : Finite set $Z \subseteq \mathbb{S}^n \times [0,1]$
Postcondition: $\mathcal{Z}_{\mathbb{S}}(f) \subset \bigcup\{\overline{B}_{\mathbb{P}}(z,r) \mid (z,r) \in Z \text{ or } (-z,r) \in Z\}$
For all $(z,r) \in Z$, $6\nu(\hat{f}, z)r < 1$ and $\mathrm{dist}_{\mathbb{S}}(z, \mathcal{Z}_{\mathbb{S}}(f)) < \frac{3}{10\nu(\hat{f},z)}$

As it usual for numerical algorithms, these kind of condition-based complexity estimates are not illustrative of the average behaviour of the algorithm. Following , we randomize the input of the algorithm to be able to talk about the expected run-time, which should give an idea of how the algorithm could work in practice. Moreover, we will consider also an smoothed analysis of our algorithm.

2.3.1 Probabilistic model

Let us recall some probabilistic notions that will be needed.

(i) A *centered* random variable is a random variable $\mathfrak{x}$ such that $\mathbb{E}\mathfrak{x} = 0$.

(ii) A *subgaussian* random variable is a random variable $\mathfrak{x} \in \mathbb{R}$ for which there is a constant $K > 0$ such that for all $p \geq 1$,

$$(\mathbb{E}|\mathfrak{x}|^p)^{\frac{1}{p}} \leq K\sqrt{p}.$$

The smallest such constant K is called the *ψ_2-norm* of $\mathfrak{x}$ and it is denoted by $\|\mathfrak{x}\|_{\psi_2}$.

(iii) A random variable $\mathfrak{x} \in \mathbb{R}$ has the *anti-concentration property with constant* ρ if for all $u \in \mathbb{R}$ and $\varepsilon > 0$,

$$\mathbb{P}(|\mathfrak{x} - u| < \varepsilon) \leq 2\rho\varepsilon.$$

Note that this is equivalent to $\mathfrak{x}$ having a density (with respect to the Lebesgue measure) bounded by ρ.

The following definition introduce the main class of random polynomials that we will consider in our probabilistic considerations. This class was originally introduced in [23] for single polynomials. Note that 'dobro' (добро) comes from Russian and it means 'good'.

Definition 2.19. A *dobro random polynomial system* $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ with parameters $\boldsymbol{K} \in \mathbb{R}^n_>$ and $\boldsymbol{\rho} \in \mathbb{R}^n_>$ is a tuple of random polynomials

$$\left(\sum_{|\alpha|=d_1} \binom{d_1}{\alpha}^{\frac{1}{2}} \mathfrak{c}_{1,\alpha} X^\alpha, \ldots, \sum_{|\alpha|=d_n} \binom{d_n}{\alpha}^{\frac{1}{2}} \mathfrak{c}_{n,\alpha} X^\alpha \right)$$

such that the $\mathfrak{c}_{i,\alpha} \in \mathbb{R}$ are independent centered subgaussian random variables with ψ_2-norm at most K_i and anti-concentration property with constant ρ_i.

The class of dobro random polynomials contains the two following important classes of random polynomial systems:

(KSS) A *KSS random polynomial system*[6] is a dobro random polynomial system $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ such that the $\mathfrak{c}_{i,\alpha}$ are i.d.d. normal random variables of mean zero and variance one. For a KSS random polynomial system, we have for each i, $K_i \leq \frac{1}{2}$ and $\rho_i = \frac{1}{\sqrt{2\pi}}$.

(W) A *Weyl random polynomial system* is a dobro random polynomial system $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ such that the $\mathfrak{c}_{i,\alpha}$ are i.d.d. random variables uniformly distributed in $[-1, 1]$. For a Weyl random polynomial system, we have for each i, $K_i = \rho_i = \frac{1}{2}$.

[6]Note that we use KSS for polynomial systems where not all polynomials have the same degree!

We note that our probabilistic estimates for dobro polynomial systems will depend on

$$\text{д}(\mathfrak{f}) := \left(\prod_{i=1}^{n} K_i(\mathfrak{f})\rho_i(\mathfrak{f})\right)^{\frac{1}{n}}$$

which we will call *dobro constant* of $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[n]$. We observe that this quantity is invariant not only under multiplication of scalars of $\mathfrak{f}$, but under multiplication by (possibly different) escalars of the polynomials in $\mathfrak{f}$. Note that $\text{д}(\mathfrak{f}) < 1/5$, if $\mathfrak{f}$ is KSS; and $\text{д}(\mathfrak{f}) = 1/4$, if $\mathfrak{f}$ is Weyl.

Given $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ dobro, then

$$\text{д}(\mathfrak{f}) \geq \frac{1}{6}. \tag{2.10}$$

The above inequality gives an idea of how small the dobro constant can be in the estimates. To prove the above inequality, one has just to consider $\mathbb{P}_{\mathfrak{x}}(|\mathfrak{x}| \leq Kt) \geq 1 - \mathbb{P}_{\mathfrak{x}}(|\mathfrak{x}| > Kt)$ for a centered, subgaussian random variable $\mathfrak{x}$ with the anti-concentration property.

The above random systems correspond to an average probabilistic model. We now define its smoothed version so that we can do the smoothed analysis of Spielman and Teng [75].

Definition 2.20. Let $\sigma > 0$. A *σ-smoothed dobro random polynomial system* with parameters $\boldsymbol{K} \in \mathbb{R}^n_>$ and $\boldsymbol{\rho} \in \mathbb{R}^n_>$ is a random polynomial system of the form

$$\mathfrak{q}_\sigma = (f_1 + \sigma\|f_1\|_\infty \mathfrak{f}_1, \ldots, f_n + \sigma\|f_n\|_\infty \mathfrak{f}_n)$$

where $f \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ is fixed and $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ is a dobro random polynomial system with parameters K and ρ. The *dobro constant* of $\mathfrak{q}_\sigma$ is given by $\text{д}(\mathfrak{q}_\sigma) = \text{д}(\mathfrak{f})$.

The idea of the smoothed analysis is that we have a fixed input, in this case f, that is perturbed by a some random perturbation, in this case $\mathfrak{f}$, of a certain magnitude controlled by a paremeter σ. Observe that as σ goes to 0, $\mathfrak{q}_\sigma$ just becomes the fixed f and that as σ goes to ∞, $\mathfrak{q}_\sigma$ becomes the random $\mathfrak{f}$. In this way, smoothed analysis bridges between the worst case estimates and the average ones. In practice, this probabilistic model is very realistic as one usually deals with fixed input subjected to some random errors.

2.3.2 Main result

We now state the technical version of Theorem A in the introduction.

Theorem 2.21. *Algorithm* `aCKMW` *is correct. When the input is a random polynomial system $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[n]$, it satisfies the following:*

(i) If $\mathfrak{f}$ is dobro, then the expected run-time is

$$\mathcal{O}\left(2^{n\log n+3n}\mathbf{D}^n N + \text{д}(\mathfrak{f})^2 2^{4n\log n+16n}\mathcal{D}^2(N+n\mathcal{D})\right).$$

In particular, if $\mathfrak{f}$ is KSS or Weyl, the expected run-time is

$$\mathcal{O}\left(2^{n\log n+3n}\mathbf{D}^n N + 2^{4n\log n+12n}\mathcal{D}^2(N+n\mathcal{D})\right).$$

(ii) If $\mathfrak{f}$ *is* σ*-smoothed dobro, then the expected run-time is*

$$\mathcal{O}\left(2^{n\log n+3n}\mathbf{D}^n N + \text{д}(\mathfrak{f})^2 2^{4n\log n+16n}\mathcal{D}^2(N+n\mathcal{D})\left(1+\frac{1}{\sigma}\right)^{2n}\right).$$

In particular, if $\mathfrak{f}$ *is* σ*-smoothed KSS/Weyl, the expected run-time is*

$$\mathcal{O}\left(2^{n\log n+3n}\mathbf{D}^n N + 2^{4n\log n+12n}\mathcal{D}^2(N+n\mathcal{D})\left(1+\frac{1}{\sigma}\right)^{2n}\right).$$

Moreover, `aCKMW` *can be parallelized. The parallel version, on random input* $\mathfrak{f}\in\mathcal{H}_{n,\boldsymbol{d}}[n]$*, satisfies the following:*

(iii) If $\mathfrak{f}$ *is dobro, then the expected run-time is*

$$\mathcal{O}(n(\log(n\mathbf{D})+\log(6\text{д}(\mathfrak{f})))$$

with expected number of processor $2^{\mathcal{O}(n\log n)}(\mathbf{D}^n+\text{д}(\mathfrak{f})^{2n}\mathcal{D}^2)N$*. A similar estimate holds in the smoothed case.*

Finally, `aCKMW` *can be executed in floating-point arithmetic with almost identical complexity estimates to those of (i) and (ii).*

3 Sampling points: δ-theory

We prove the δ-theorem in a similar way to which Smale's α-theorem is proven in [31, Ch. 4]. For this, first we will prove Propositions 3.1 (which will follow from proving Theorem 2.6) and 3.2, which give the general variational properties of the ν-condition and the β-estimate, and the Proposition 3.5, which gives the variational properties along a Newton step. We use them to prove the δ-theorem. We finish with the proofs of Proposition 2.9

Proposition 3.1 (Variation of ν). *Let* $f\in\mathcal{H}_{n,\boldsymbol{d}}[q]$ *and* $x,y\in\mathbb{S}^n$*. If* $\nu(f,x)\operatorname{dist}_{\mathbb{S}}(x,y)<1$*, then*

$$\nu(f,y)\leq\frac{1}{1-\nu(f,x)\operatorname{dist}_{\mathbb{S}}(x,y)}\nu(f,x).$$

Proposition 3.2 (Variation of β). *Let* $f\in\mathcal{H}_{n,\boldsymbol{d}}[q]$ *and* $x,y\in\mathbb{S}^n$*. If* $\nu(f,x)\operatorname{dist}_{\mathbb{S}}(x,y)<1$*, then*

$$\beta(f,y)\leq\frac{1}{1-\nu(f,x)\operatorname{dist}_{\mathbb{S}}(x,y)}\,\beta(f,x)+\frac{1+\nu(f,x)\operatorname{dist}_{\mathbb{S}}(x,y)}{1-\nu(f,x)\operatorname{dist}_{\mathbb{S}}(x,y)}\,\operatorname{dist}_{\mathbb{S}}(x,y)$$

Proof of Theorem 2.6. **1st Lipschitz property**. Note the following analog of the max-min theorem:

$$\frac{\|f\|_\infty}{\nu(f,x)}=\left\|(\Delta^{-2}\mathrm{D}_x f)^\dagger\right\|_{\infty,2}^{-1}=\sup_{\substack{V\leq\mathbb{R}^{n+1}\\ \dim V=q}}\inf_{\substack{v\in V\\ v\neq 0}}\frac{\|\Delta^{-2}\mathrm{D}_x f v\|_\infty}{\|v\|_2}. \tag{3.1}$$

If $\mathrm{D}_x f$ is not surjective, then both sizes are zero and the equality holds. If $\mathrm{D}_x f$ is surjective, then

$$\left\|(\Delta^{-2}\mathrm{D}_x f)^\dagger\right\|_{\infty,2}^{-1} = \inf_{w\neq 0} \frac{\|w\|_\infty}{\|(\Delta^{-2}\mathrm{D}_x f)^\dagger w\|_2}$$
$$= \inf_{\substack{v\in \ker \mathrm{D}_x f^\perp \\ v\neq 0}} \frac{\|\Delta^{-2}\mathrm{D}_x f v\|_\infty}{\|(\Delta^{-2}\mathrm{D}_x f)^\dagger \Delta^{-2}\mathrm{D}_x f v\|_2} = \inf_{\substack{v\in (\ker \mathrm{D}_x f)^\perp \\ v\neq 0}} \frac{\|\Delta^{-2}\mathrm{D}_x f v\|_\infty}{\|v\|_2},$$

giving that the left-hand side is bounded by the right-hand side. For the other inequality, take a subspace V of $\mathbb{R}^{n+1}$ of dimension q such that $V\cap \ker \mathrm{D}_x f = 0$, so that the orthogonal projection $P : V \to (\ker \mathrm{D}_x f)^\perp$ is injective. Otherwise, the infimum on the right hand side would be zero. Then

$$\inf_{\substack{v\in V \\ v\neq 0}} \frac{\|\Delta^{-2}\mathrm{D}_x f v\|_\infty}{\|v\|_2} \leq \inf_{\substack{v\in V \\ v\neq 0}} \frac{\|\Delta^{-2}\mathrm{D}_x f v\|_\infty}{\|Pv\|_2} = \inf_{\substack{v\in V \\ v\neq 0}} \frac{\|\Delta^{-2}\mathrm{D}_x f (Pv)\|_\infty}{\|Pv\|_2}$$
$$= \inf_{\substack{v\in (\ker \mathrm{D}_x f)^\perp \\ v\neq 0}} \frac{\|\Delta^{-2}\mathrm{D}_x f v\|_\infty}{\|v\|_2} = \left\|(\Delta^{-2}\mathrm{D}_x f)^\dagger\right\|_{\infty,2},$$

where we use that projecting in the orthogonal complement of the kernel does not alter the image. Thus the wanted equality follows.

Using the above max-min theorem and that taking maximums and minimums preserves Lipschitz properties, we have that

$$\left|\frac{\|\tilde{g}\|_\infty}{\nu(\tilde{g},x)} - \frac{\|g\|_\infty}{\nu(g,x)}\right| \leq \|\Delta^{-2}\mathrm{D}_x(\tilde{g}-g)\|_{2,\infty}.$$

Now, by Kellogg's theorem (Theorem 2.1),

$$\|\Delta^{-2}\mathrm{D}_x h\|_{2,\infty} \leq \max_i \sup_{v\neq 0} \frac{1}{d_i^2}\frac{|\partial_x h_i v|}{\|v\|_2} \leq \max_i \frac{\|h_i\|_\infty}{d_i} = \|\Delta^{-1} h\|_\infty, \tag{3.2}$$

so the 1st Lipschitz inequality follows. For the other inequality, take in the 1st Lipschitz inequality, $g = f$ and $\tilde{g} = 0$, so that

$$\nu(f,x) \geq \frac{\|f\|_\infty}{\|\Delta^{-1} f\|_\infty} \geq 1.$$

2nd Lipschitz property. Let $y, \tilde{y} \in \mathbb{S}^n$ and $u \in O(n+1)$ be the planar rotation sending x to y. Then

$$\left|\frac{1}{\nu(f,\tilde{y})} - \frac{1}{\nu(f,y)}\right| = \frac{1}{\|f\|_\infty}\left|\frac{\|f^u\|_\infty}{\nu(f^u,y)} - \frac{\|f\|_\infty}{\nu(f,y)}\right|$$

where $f^u := f(uX)$, by the chain rule and the invariance of the L_∞-norm. Thus, by the 1st Lipschitz property,

$$\left|\frac{1}{\nu(f,\tilde{y})} - \frac{1}{\nu(f,y)}\right| \leq \frac{\|\Delta^{-1}(f^u - f)\|_\infty}{\|f\|_\infty}.$$

Now, $\|\Delta^{-1}(f^u - f)\|_\infty = \max_{z\in\mathbb{S}^n} \|\Delta^{-1} f(uz) - \Delta^{-1} f(z)\|_\infty$. By Kellogg's theorem (Theorem 2.1), $\Delta^{-1} f$ is $\|f\|_\infty$-Lipschitz, since each $d_i^{-1} f_i$ is $\|f_i\|_\infty$-Lipschitz. Thus

$$\|\Delta^{-1} f(uz) - \Delta^{-1} f(z)\|_\infty \leq \|f\|_\infty \operatorname{dist}_{\mathbb{S}}(z, uz) \leq \|f\|_\infty \operatorname{dist}_{\mathbb{S}}(y, \tilde{y})$$

where the last inequality follows because u is the planar rotation taking y to $\tilde{y}$. □

Proof of Proposition 3.1. By the 2nd Lipschitz property of the ν-condition (Theorem 2.6), $x \mapsto 1/\nu(f,x)$ is 1-Lipschitz (with respect the geodesic distance on $\mathbb{S}^n$). The proposition is just a rewriting of this condition. □

Lemma 3.3. *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x, y \in \mathbb{S}^n$. Then*

$$\|\mathrm{D}_y f^\dagger \mathrm{D}_x f\|_{2,2} \leq \frac{1}{1 - \nu(f,x)\,\mathrm{dist}_{\mathbb{S}}(x,y)}.$$

Lemma 3.4. *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x, y \in \mathbb{S}^n$ be such that $y \neq -x$. Then*

$$\|\mathrm{D}_x f^\dagger f(y)\|_2 \leq \left\|\mathrm{D}_x f^\dagger f(x) + \mathrm{dist}_{\mathbb{S}}(x,y)\mathrm{D}_x f^\dagger \mathrm{D}_x f \upsilon_{x,y}\right\|_2 + \frac{1}{2}\,\nu(f,x)\,\mathrm{dist}_{\mathbb{S}}(x,y)^2$$

where $\upsilon_{x,y} := (y - \langle y, x\rangle x)/\sin \mathrm{dist}_{\mathbb{S}}(x,y)$ is the tangent unit vector at x of the geodesic joining x to y.

Proof of Proposition 3.2. Note that

$$\beta(f,y) = \|\mathrm{D}_y f^\dagger \mathrm{D}_x f \mathrm{D}_x f^\dagger f(y)\|_2 \leq \|\mathrm{D}_y f^\dagger \mathrm{D}_x f\|_{2,2} \|\mathrm{D}_x f^\dagger f(y)\|_2,$$

since $\mathrm{D}_x f \mathrm{D}_x f^\dagger = \mathbb{I}$. The propositions follows now from Lemmas 3.3 and 3.4, $1/2 \leq 1$, the triangle inequality and that $\mathrm{D}_x f^\dagger \mathrm{D}_x f$ is an orthogonal projection. □

Proof of Lemma 3.3. Note that for $f, g \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $z \in \mathbb{S}^n$,

$$\begin{aligned}
\|\mathrm{D}_z g^\dagger \mathrm{D}_z f\|_{2,2} &\leq \|\mathrm{D}_z g^\dagger \mathrm{D}_z g\|_{2,2} + \|\mathrm{D}_z g^\dagger \mathrm{D}_z (f-g)\|_{2,2} \\
&\leq 1 + \|\mathrm{D}_z g^\dagger \mathrm{D}_z (f-g)\|_{2,2} && \left(\mathrm{D}_z g^\dagger \mathrm{D}_z g \text{ is an orthogonal projection}\right) \\
&\leq 1 + \nu(g,z)\frac{\|\Delta^{-2}\mathrm{D}_z(f-g)\|_{2,\infty}}{\|g\|_\infty} && \text{(Definition of } \nu\text{-condition)} \\
&\leq 1 + \nu(g,z)\frac{\|\Delta^{-1}(f-g)\|_\infty}{\|g\|_\infty}. && \text{(Kellogg's theorem, applied as in (3.2))}
\end{aligned}$$

Now, let $u \in O(n+1)$ be the planar rotation taking x to y. Then

$$\|\mathrm{D}_y f^\dagger \mathrm{D}_x f\| = \|\mathrm{D}_x (f^u)^\dagger \mathrm{D}_x f\|.$$

Hence, arguing as in the proof of the 2nd Lipschitz property, but using the inequality above, we get

$$\|\mathrm{D}_y f^\dagger \mathrm{D}_x f\| \leq 1 + \nu(f,y)\,\mathrm{dist}_{\mathbb{S}}(x,y).$$

Finally, Proposition 3.1 finishes the proof. □

Proof of Lemma 3.4. For any smooth path $\vartheta : [0,1] \to \mathbb{R}^{n+1}$, we have by Taylor's theorem that

$$\|\vartheta(1)\|_2 \leq \|\vartheta(0) + \vartheta'(0)\|_2 + \frac{1}{2}\max_{s\in[0,1]}\|\vartheta''(s)\|_2.$$

Now, let $[0,1] \ni t \mapsto x_t \in \mathbb{S}^n$ a constant speed geodesic joining x and y and $\vartheta : [0,1] \to \mathbb{R}^{n+1}$ the smooth path given by $\vartheta(t) = \mathrm{D}_x f^\dagger f(x_t)$. By the chain rule,

$$\vartheta'(0) = \mathrm{dist}_{\mathbb{S}}(x,y)\mathrm{D}_x f^\dagger \mathrm{D}_x f \upsilon_{x,y} = \mathrm{dist}_{\mathbb{S}}(x,y)\upsilon_{x,y},$$

since $\partial_x f v_{x,y} = \mathrm{D}_x f v_{x,y}$ since $x \in \ker \mathrm{D}_x f$ and $v_{x,y}$ is orthogonal to x; and

$$\vartheta''(s) = \mathrm{D}_x f^\dagger \left(\partial^2_{x_s} f\left(\dot{x_s}, \dot{x_s}\right) - \Delta f(x_s) \operatorname{dist}_{\mathbb{S}}(x,y)^2 \right),$$

since $\ddot{x_s} = -\operatorname{dist}_{\mathbb{S}}(x,y)^2 x_s$ and $\mathrm{D}_x f(x_t) = \Delta f(x_t)$ by Euler's formula for homogeneous polynomials.

Now,

$$\|\vartheta''(s)\|_2 \leq \nu(f,x) \frac{1}{\|f\|_\infty} \left\| \Delta^{-2} \partial^2_{x_s} f\left(\dot{x_s}, \dot{x_s}\right) - \Delta^{-1} f(x_s) \operatorname{dist}_{\mathbb{S}}(x,y)^2 \right\|_\infty .$$

Now, by the triangle inequality, the above is bounded by

$$\max_i \left(|d_i^{-2} \partial^2_{x_s} f\left(\dot{x_s}, \dot{x_s}\right)| + |d_i^{-1} f(x_s)| \operatorname{dist}_{\mathbb{S}}(x,y)^2 \right).$$

Applying Kellogg's inequality (Theorem 2.1, we get that

$$|d_i^{-2} \partial^2_{x_s} f\left(\dot{x_s}, \dot{x_s}\right)| \leq \left(1 - \frac{1}{d_i}\right) \|f\|_\infty \operatorname{dist}_{\mathbb{S}}(x,y)^2 \quad \text{and} \quad |d_i^{-1} f(x_s)| \leq \frac{1}{d_i} \|f\|_\infty,$$

since $\|\dot{x}_s\|_2 = \operatorname{dist}_{\mathbb{S}}(x,y)$. Hence

$$\frac{1}{2} \max_{s \in [0,1]} \|\vartheta''(s)\|_2 \leq \frac{1}{2}$$

and the result follows. □

Proposition 3.5 (Variation along a Newton step). *Let $f \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ and $x \in \mathbb{S}^n$ be such that $\delta(f,x) < 1$. Then $\mathrm{N}_f^2(x)$ is defined,*

$$\beta(f, \mathrm{N}_f(x)) \leq \frac{1}{2} \frac{1 + \delta(f,x)}{1 - \delta(f,x)} \delta(f,x) \beta(f,x)$$

and

$$\delta(f, \mathrm{N}_f(x)) \leq \frac{1}{2} \frac{1 + \delta(f,x)}{(1 - \delta(f,x))^2} \delta(f,x)^2.$$

Proof. We note that $\mathrm{N}_f^2(x)$ is well-defined, since $\mathrm{D}_{\mathrm{N}_f(x)} f$ is surjective as $\nu(f, \mathrm{N}(f,x)) < \infty$ by Proposition 3.1 and $\nu(f,x) \operatorname{dist}_{\mathbb{S}}(x, \mathrm{N}_f(x)) \leq \delta(f,x) < 1$.

For the bound on $\beta(f, \mathrm{N}_f(x))$, we argue as in Proposition 3.2 using Lemmas 3.3 and 3.4. The main trick is that when we apply Lemma 3.4, we get that

$$\begin{aligned} \left\| \mathrm{D}_x f^\dagger f(x) - \operatorname{dist}_{\mathbb{S}}(x, \mathrm{N}_f(x)) \mathrm{D}_x f \mathrm{D}_x f^\dagger v_{x, \mathrm{N}_f(x)} \right\| &= \beta(f,x) |\beta(f,x) - \arctan \beta(f,x)| \\ &\leq \frac{1}{3} \beta(f,x)^3 \end{aligned}$$

due to $v_{x,\mathrm{N}_f(x)} = -\mathrm{D}_x f^\dagger f(x) / \beta(f,x)$, (2.3) and (2.4). Hence

$$\beta(f, \mathrm{N}_f(x)) \leq \frac{\frac{1}{2} + \frac{1}{3}\beta(f,x)}{1 - \delta(f,x)} \delta(f,x) \beta(f,x).$$

The desired bound for $\beta(f, \mathrm{N}_f(x))$ follows after noting that

$$\frac{1}{2} + \frac{1}{3}\beta(f,x) \leq \frac{1}{2} + \frac{1}{3}\delta(f,x) \leq \frac{1}{2}(1 + \delta(f,x)),$$

since $\nu(f,x) \geq 1$ by the 1st Lipschitz property (Theorem 2.6). For the inequality regarding $\delta(f, \mathrm{N}_f(x))$, we combined the inequality for β obtained with the one in Proposition 3.1. □

Proof of the δ-Theorem (Theorem 2.10). We will show by induction that the Newton sequence is well-defined and that

$$\delta(f, \mathrm{N}_f^k(x)) \leq \left(\frac{1}{3}\right)^{2^k-1} \delta(f,x)$$

and

$$\beta(f, \mathrm{N}_f^k(x)) \leq \left(\frac{3}{4}\right)^k \left(\frac{1}{3}\right)^{2^k-1} \beta(f,x)$$

using Proposition 3.5.

Clearly both claims are true for $k = 0$, so the base case is true. We now show the induction step. By Proposition 3.5 and the induction hypothesis, $\mathrm{N}_f^{k+1}(x) = \mathrm{N}_f(\mathrm{N}_f^k(x))$ is well-defined,

$$\begin{aligned}\beta(f, \mathrm{N}_f^{k+1}(x)) &\leq \frac{1}{2}\frac{1+\delta(f,\mathrm{N}_f^k(x))}{1-\delta(f,\mathrm{N}_f^k(x))}\delta(f,\mathrm{N}_f^k(x))\beta(f,\mathrm{N}_f^k(x)) \\ &\leq \frac{1}{2}\frac{1+\delta(f,x)}{1-\delta(f,x)}\delta(f,x)\left(\frac{3}{4}\right)^k\left(\frac{1}{3}\right)^{2^{k+1}-2}\beta(f,x)\end{aligned}$$

and

$$\begin{aligned}\delta(f, \mathrm{N}_f^{k+1}(x)) &\leq \frac{1}{2}\frac{1+\delta(f,\mathrm{N}_f^k(x))}{(1-\delta(f,\mathrm{N}_f^k(x)))^2}\,\delta(f,\mathrm{N}_f^k(x))^2 \\ &\leq \frac{1}{2}\frac{1+\delta(f,x)}{(1-\delta(f,x))^2}\delta(f,x)\left(\frac{1}{3}\right)^{2^{k+1}-2}\delta(f,x),\end{aligned}$$

where we have used that both

$$t \mapsto \frac{1}{2}\frac{1+t}{1-t} \text{ and } t \mapsto \frac{1}{2}\frac{1+t}{(1-t)^2}$$

are monotonous on t for $t \in [0,1]$ and that $\beta(f, \mathrm{N}_f^k(x)) \leq \beta(f,x)$ and $\delta(f, \mathrm{N}_f(x)) \leq \delta(f,x)$. Now, by assumption, $\delta(f,x) \leq 1/4$, so

$$\frac{1}{2}\frac{1+\delta(f,x)}{(1-\delta(f,x))}\delta(f,x) < \frac{1}{4} \quad \text{and} \quad \frac{1}{2}\frac{1+\delta(f,x)}{(1-\delta(f,x))^2}\delta(f,x) < \frac{1}{3}. \tag{3.3}$$

Hence our inductive claims hold. This shows 1 and 2.

For 3, we note that for all $k >= 0$,

$$\mathrm{dist}_{\mathbb{S}}\left(\mathrm{N}_f^k(x), \mathrm{N}_f^{k+1}(x)\right) \leq \left(\frac{3}{4}\right)^k\left(\frac{1}{3}\right)^{2^k-1}\beta(f,x) \leq \frac{1}{4}\left(\frac{3}{4}\right)^k\left(\frac{1}{3}\right)^{2^k-1}\frac{1}{\nu(f,x)}.$$

Adding the sum, the desired claim follows. □

Remark 3.6. In the above proof, we can see that to improve the constants in the δ-theorem, we can just play we the values of $\delta(f,x)$ and the bounds obtained in (3.3). For getting quadratic convergence, we can use $\delta(f,x) < 1/3$; and for getting convergence, $\delta(f,x) \leq \frac{5-\sqrt{17}}{2} \approx 0.4385\ldots$

We now prove the converses to the α-theorem and to the δ-theorem.

Proof of Proposition 2.9. Let $\zeta \in \mathcal{Z}_{\mathbb{S}}$ be the nearest zero to x. Then, by Taylor's theorem,

$$-\mathrm{D}_x f^\dagger f(x) = \mathrm{D}_x f^\dagger f(\zeta) - \mathrm{D}_x f^\dagger f(x) = \sum_{k=1}^{\infty} \frac{1}{k!} \mathrm{D}_x f^\dagger \partial_x^k (\zeta - x, \dots, \zeta - x).$$

Now, taking norms,

$$\beta(f,x) = \|\mathrm{D}_x f^\dagger \mathrm{D}_x f(\zeta - x)\|_2 + \sum_{k=2}^{\infty} \left\| \frac{1}{k!} \mathrm{D}_x f^\dagger \partial_x^k (\zeta - x, \dots, \zeta - x) \right\|_2 .$$

where the first summand side is bounded by $\|\zeta - x\|_2$, since $\mathrm{D}_x f^\dagger \mathrm{D}_x f$ is an orthogonal projection, and the kth summand in the series is bounded by $\gamma(f,x)^{k-1}\|\zeta - x\|_2^k$. Hence, we have that,

$$\beta(f,x) \leq \mathrm{dist}_{\mathbb{S}}(x, \mathcal{Z}_{\mathbb{S}}(f)) \left(\sum_{k=0}^{\infty} \left(\gamma(f,x)\, \mathrm{dist}_{\mathbb{S}}(x, \mathcal{Z}_{\mathbb{S}}(f))\right)^k \right) = \frac{\mathrm{dist}_{\mathbb{S}}(x, \mathcal{Z}_{\mathbb{S}}(f))}{1 - \gamma(f,x)\, \mathrm{dist}_{\mathbb{S}}(x, \mathcal{Z}_{\mathbb{S}}(f))}$$

The rest follows from the Higher Derivative Estimate (2.6). □

Proof of Theorem 2.11. We argue as in Lemma 3.4, taking the same map ϑ, but with $y \in \mathcal{Z}_{\mathbb{S}}(f)$ the nearest point to x. Using Taylor's theorem for a smooth map $\vartheta : [0,1] \to \mathbb{S}^n$, we have that

$$\|\vartheta(1) - \vartheta(0)\|_2 \leq \|\vartheta'(0)\|_2 + \frac{1}{2} \max_{s \in [0,1]} \|\vartheta''(s)\|_2.$$

Taking ϑ as in the proof of Lemma 3.4, , and arguing analogously, we obtain that

$$\beta(f,x) \leq \mathrm{dist}_{\mathbb{S}}(x, \mathcal{Z}_{\mathbb{S}}(f)) + \frac{1}{2}\nu(f,x)\, \mathrm{dist}_{\mathbb{S}}(x, \mathcal{Z}_{\mathbb{S}}(f))^2.$$

Multiplying by $\nu(f,x)$ gives the desired bound. □

We finish the section with the proof of Proposition 2.13.

Proof of Proposition 2.13. The proof follows a similar idea to the proof in of [1, p. 161], but with he technical difficulties associated to working in the sphere. Without loss of generality, we assume that $\|f\|_\infty = 1$.

Let $\zeta_0, \zeta_1 \in \overline{B}_{\mathbb{S}}(x, 1/(3\nu(f,x)))$ be distinct zeros of f. Let $[0,1] \ni t \mapsto \zeta_t$ be the constant speed geodesic joining them. For each i, we have that $t \mapsto f_i(\zeta_t)$ takes the value 0 at the extremes of $[0,1]$. Therefore, by Rolle's theorem, for each i, there is some t_i so that

$$\mathrm{D}_{\zeta_{t_i}} f_i(\dot{\zeta}_{t_i}) = 0.$$

Let $y = \zeta_{1/2}$ be the point equidistant to ζ_0 and ζ_1 in the geodesic joining them, and $v = \dot{\zeta}_{1/2}$ the tangent vector of the considered constant speed geodesic. Let $u_1, \dots, u_n$ be the planar rotations that send y respectively to $\zeta_{t_1}, \dots, \zeta_{t_n}$. These planar rotations can be obtained by rotating the plane containing ζ_0, ζ_1 and the geodesic joining them. Note that, because of this, we must have that

$$u_i v = \dot{\zeta}_{t_i},$$

since these rotations must preserve tangent vectors along the geodesic joining ζ_0 and ζ_i.

Let $g = (f_i^{u_i})$. By assumption and orthogonal invariance, $\|g\|_\infty = \|f\|_\infty = 1$. By construction, $\mathrm{D}_y g : \mathrm{T}_y\mathbb{S}^n \to \mathbb{R}^n$ is not surjective, since $v \in \ker \mathrm{D}_y g$. Note that here it is essential that g is an n-tuple and not a q-tuple. Therefore

$$\nu(g, y) = \infty.$$

Now, we show that the above equality cannot happen. Therefore we cannot have two roots inside the considered ball. By the Lipschitz properties in Theorem 2.6,

$$\begin{aligned}
\frac{1}{\nu(g,y)} &\geq \frac{1}{\nu(g,x)} - \mathrm{dist}_{\mathbb{S}}(x,y) && \text{(2nd Lipschitz prop. of } \nu) \\
&\geq \frac{1}{\nu(f,x)} - \left\|\Delta^{-1}(g-f)\right\|_\infty - \mathrm{dist}_{\mathbb{S}}(x,y) && \text{(1st Lipschitz prop. of } \nu) \\
&\geq \frac{1}{\nu(f,x)} - \max_i \left\|d_i^{-1}\left(f_i^{u_i} - f_i\right)\right\|_\infty - \mathrm{dist}_{\mathbb{S}}(x,y) \\
&\geq \frac{1}{\nu(f,x)} - \max_i \max_{z\in\mathbb{S}^n} \mathrm{dist}(z, u_i z) - \mathrm{dist}_{\mathbb{S}}(x,y) && \text{(Kellogg's Theorem (2.1)).}
\end{aligned}$$

As u_i is the planar rotation taking $y = \zeta_{1/2}$ to ζ_{t_i} along the geodesic where they lie, we have that

$$\mathrm{dist}(z, u_i z) \leq \mathrm{dist}_{\mathbb{S}}(\zeta_{1/2}, \zeta_{t_i}) \leq \frac{1}{2}\,\mathrm{dist}_{\mathbb{S}}(\zeta_0, \zeta_1).$$

Hence

$$\frac{1}{\nu(g,y)} \geq \frac{1}{2\nu(f,x)},$$

since $y, \zeta_0, \zeta_1 \in \overline{B}_{\mathbb{S}}(x, 1/(3\nu(f,x)))$. But this means that

$$\nu(g,y) \leq 2\nu(f,x) < \infty,$$

obtaining the desired contradiction. □

4 Complexity Analysis

We now give the complexity analysis, which is the core of this paper. First, we give several condition-based complexity analyses; then we use them to produce probabilistic complexity analyses.

4.1 Condition-based complexity analysis

The following theorem gives the condition-based estimate from where all our probabilistic estimates will be deduced.

Theorem 4.1. *Algorithm* `aCKMW` *is correct. When the input is a polynomial system* $f \in \mathcal{H}_{n,\boldsymbol{d}}[n]$*, its run-time is bounded by*

$$\mathcal{O}\left[2^{n\log n+3n}\mathbf{D}^n N + \mathsf{vol}_n(\mathbb{S}^n)\left(\frac{3n}{2}\right)^n (N + n\mathcal{D})\left(\mathbb{E}_{\mathfrak{r}\in\mathbb{S}^n}\left(\frac{\mathtt{C}\left(\hat{f},\mathfrak{r}\right)^n}{\|\mathfrak{r}\|_\infty^{n+1}}\right)\right)\right].$$

Remark 4.2. We note that

$$\mathsf{vol}_n(\mathbb{S}^n)\mathbb{E}_{\mathfrak{x}\in\mathbb{S}^n}\frac{1}{\|\mathfrak{x}\|_\infty^{n+1}} = 2^n(n+1), \tag{4.1}$$

due to the change of variables formula. Because of this, we don't estimate further the elements in the statement above.

Recall that in both algorithms we are just bounding the number of arithmetic operations. We divide the analysis between the three main part of `aCKMW`: 1) computation of the norms, 2) the Exclusion-Inclusion-Refinement Cycle, and 3) the elimination of redundant approximations. The proof of Theorem 4.1 follows from combining Propositions 4.3, 4.5 and 4.6.

4.1.1 Computation of the norms (Lines 1-3)

To compute the L_∞-norms, the main result is [24, Proposition 4.2].

Proposition 4.3. *Let $f \in \mathcal{H}_{n,d}$ be an homogeneous polynomial of degree d and $\mathcal{G} \subset \mathbb{S}^n$ such that for every $x \in \mathbb{S}^n$, $\mathrm{dist}_{\mathbb{S}}(x, \mathcal{G}) \leq \eta$. If $d\eta \leq 1$, then*

$$\max_{x\in\mathcal{G}} \|f(x)\|_\infty \leq \|f\|_\infty \leq \frac{1}{1-\frac{\mathbf{D}^2}{2}\eta^2}\max_{x\in\mathcal{G}}\|f(x)\|_\infty.$$

□

Using the above proposition, one proves the following result in a similar way to [24, Proposition 4.4].

Proposition 4.4. *In lines 7-3 of `aCKMW`, the computed Q_i satisfy that*

$$\left(1-\frac{1}{8n}\right)Q_i \leq \|f_i\|_\infty \leq Q_i.$$

Moreover, this computation requires

$$\mathcal{O}\left(2^{n\log n+3n}\mathbf{D}^n N\right)$$

arithmetic operations.

Proof. We note that $|\mathcal{U}_\ell| = 2^{\ell n}$. In this way, for each f_i, we do $\mathcal{O}\left(n2^{n\lceil 1+\log n+\log d_i\rceil}\right)$ evaluations, taking each one of them $\mathcal{O}\left(\binom{n+d_i}{n}\right)$ arithmetic operations (see [12, Lemma 16.31]). The property of the Q_i is justified by Proposition 4.3, since, by construction, for all i and all $x \in \mathbb{S}^n$, $\mathrm{dist}_{\mathbb{S}}(x, \text{Ю}_{k,+1}(\mathcal{U}_\ell)) \leq 1/d_i$. □

4.1.2 Exclusion-Inclusion-Refinement Cycle (Lines 7-20)

The next proposition is reminiscent of the continuous amortization of Burr, Krahmer and Yap [17].

Proposition 4.5. *Lines 7-20 of `aCKMW` satisfy the following:*

(a) *If for all i, $\|f_i\|_\infty \leq Q_i$, then, at the end of the execution, we have that for all $(z,r) \in \tilde{Z}$, $\overline{B}_{\mathbb{S}}(z,r)$ contains exactly one root of f, of which (z,r) is an approximation à la Smale. And reciprocally, for each $\zeta \in \mathcal{Z}_{\mathbb{S}}(f)$, there is some $(z,r) \in Z$ such that $\zeta \in \overline{B}_{\mathbb{S}}(z,r) \cup \overline{B}_{\mathbb{S}}(-z,r)$.*

(b) *The total number of points that were processed in the cycle is bounded by*

$$\left(\frac{3n}{2}\right)^n \int_{x\in\mathbb{S}^n} \frac{\mathtt{C}\left(\hat{f},x\right)^n}{\|x\|_\infty^{n+1}}\,\mathrm{d}x.$$

(c) *The total number of arithmetic operations is bounded by*

$$\mathcal{O}\left(\left(\frac{3n}{2}\right)^n N \int_{x\in\mathbb{S}^n} \frac{\mathtt{C}\left(\hat{f},x\right)^n}{\|x\|_\infty^{n+1}}\,\mathrm{d}x\right).$$

Proof of Proposition 2.15. (i) This is by construction.

(ii) The refinement operator substitutes a cube by 2^n copies of cube with half the width. Therefore it preserves the cubical grid property.

(iii) This follows from the fact that the $\mathrm{IO}_{k,\sigma}$ are 1-Lipschitz with the Euclidean distance in the domain and the geodesic distance in the codomain. Note that the $\sqrt{n}$ comes from the relation between the ∞-norm and the Euclidean norm. □

Proof of Proposition 4.5. (a) Notice that

$$\left(1-\frac{1}{8n}\right)\|\Delta^{-1}\hat{f}(\xi)\|_\infty \leq \|\mathcal{Q}^{-1}\Delta^{-1}f(\xi)\|_\infty \leq \|\Delta^{-1}\hat{f}(\xi)\|_\infty \tag{4.2}$$

and that

$$\frac{1}{\sqrt{n}}\nu(\hat{f},x) \leq \|\mathrm{D}_\xi f^\dagger \Delta^2 \mathcal{Q}\|_{\infty,\infty} \leq \left(1-\frac{1}{8n}\right)^{-1}\nu(\hat{f},x) \tag{4.3}$$

since for all i, $\left(1-\frac{1}{8n}\right) \leq \|f_i\|_\infty/\mathcal{Q}_i \leq 1$.

If the inequality in line 11 holds, then

$$\|\mathcal{Q}^{-1}\Delta^{-1}f(\xi)\|_\infty \geq \sqrt{n}2^{1-\ell},$$

and so, by (4.2),

$$\|\Delta^{-1}\hat{f}(\xi)\|_\infty \geq \sqrt{n}2^{1-\ell}.$$

Hence the exclusion lemma (Proposition 2.2) guarantees that there are not any roots inside $\overline{B}_{\mathbb{S}}(\xi,\sqrt{n}2^{1-\ell})$, so the exclusion is justified.

If the inequality in line 13 is satisfied, then, by (4.3),

$$6\left(1-\frac{1}{8n}\right)^{-1} n\nu(\hat{f},\xi) \leq 2^\ell$$

and so

$$\overline{B}_{\mathbb{S}}(\xi,\sqrt{n}2^{-\ell}) \subseteq \overline{B}_{\mathbb{S}}\left(\xi,\left(1-\frac{1}{8n}\right)\frac{1}{6\sqrt{n}\nu(\hat{f},\xi)}\right).$$

By the converse of the δ-theorem (Proposition 2.11), if there is a root ζ of f in $\overline{B}_{\mathbb{S}}(\xi, \sqrt{n}2^{-\ell})$, then

$$5\left(1-\frac{1}{8n}\right)^{-1}\sqrt{n}\delta(\hat{f},x)<1,$$

and so, by (4.3),

$$5\sqrt{n}\|\mathrm{D}_\xi f^\dagger \Delta^2 \mathcal{Q}\|_{\infty,\infty}\|\mathrm{D}_\xi f^\dagger f(\xi)\|_2 \leq 1.$$

Thus we conclude that the inclusion test would be passed in this case. Note that by construction of the adaptive grid (Proposition 2.15), the above situation happens to every possible projective root.

Now, if the inclusion test is passed, this means that, by (4.3), the hypothesis of the δ-theorem (Theorem 2.10) holds, and so there is a root $\zeta \in \mathcal{Z}_{\mathbb{S}}(f)$ such that $(\xi, 1.5\|\mathrm{D}_\xi f^\dagger f(\xi)\|_2)$ is an approximation *à la Smale*of ζ, and such that

$$\mathrm{dist}_{\mathbb{S}}(\xi,\zeta) \leq 1.5\|\mathrm{D}_\xi f^\dagger f(\xi)\|_2 \leq \frac{3}{10}\frac{1}{\nu(\hat{f},x)} < \frac{1}{3\nu(\hat{f},x)}.$$

Hence ζ is the unique root of f in $\overline{B}_{\mathbb{S}}\left(\xi, \frac{1}{3\nu(\hat{f},\xi)}\right)$, by Proposition 2.13. Note that this means that if there was a root in and such that . This justifies the inclusion of the point. Moreover, note that if there was a root in $\overline{B}_{\mathbb{S}}(\xi,\sqrt{n}2^{-\ell}) \subset \overline{B}_{\mathbb{S}}\left(\xi, \frac{1}{3\nu(\hat{f},\xi)}\right)$, it has to be ζ. Thus (a) follows, as this happens to every root.

(b) Let $(x,\ell,k,+1) \in \mathcal{G}$ be a point that either passes either the test in line 11 or the test in line 13, and $(y,\ell-1,k,+1)$ the parent of $(x,\ell,k,+1)$, i.e., $(x,\ell,k,+1) \in \mathtt{R}(y,\ell-1,k,+1)$. We define $\hat{x} = \text{Ю}_{k,+1}(x)$ and $\hat{y} = \text{Ю}_{k,+1}(y)$. Note that $(y,\ell-1,k,+1)$ does not satisfy the conditions in lines line 11 and 13. Therefore, by applying (4.2) and (4.3), we have that

$$\|\Delta^{-1}\hat{f}(\hat{y})\|_\infty \leq \left(1-\frac{1}{8n}\right)^{-1}\sqrt{n}2^{-\ell} \quad \text{and} \quad \nu(\hat{f},\hat{y})^{-1} \leq 3n\left(1-\frac{1}{8n}\right)^{-2}\sqrt{n}2^{-\ell}.$$

Therefore for all $\hat{z} \in \text{Ю}_{k,+1}(\overline{B}_\infty(\hat{x},2^{-\ell}))$,

$$\left(1-\frac{1}{8n}\right)^{-1}\sqrt{n}2^{-\ell} \geq \|\Delta^{-1}\hat{f}(\hat{y})\|_\infty \geq \|\Delta^{-1}\hat{f}(\hat{z})\|_\infty - 2\sqrt{n}2^{-\ell},$$

by Proposition 2.2, and

$$3n\left(1-\frac{1}{8n}\right)^{-2}\sqrt{n}2^{-\ell} \geq \nu(\hat{f},\hat{y})^{-1} \geq \nu(\hat{f},\hat{z})^{-1} - 2\sqrt{n}2^{-\ell},$$

by the 2nd Lipschitz property of ν (Theorem 2.6). Hence for all $z \in \overline{B}_\infty(x,2^{-\ell})$,

$$2^{\ell-1} \leq \min\left\{\frac{\sqrt{n}}{\left\|\Delta^{-1}\hat{f}\left(\text{Ю}_{k,+1}(z)\right)\right\|_\infty}, \left(n+\frac{1}{2}\right)\sqrt{n}\,\nu\left(\hat{f},\text{Ю}_{k,+1}(z)\right)\right\},$$

and so for all $z \in \overline{B}_\infty(x,2^{-\ell})$,

$$2^{\ell-1} \leq 3\sqrt{n}\,\mathtt{C}\left(\hat{f},\text{Ю}_{k,+1}(z)\right). \tag{4.4}$$

Now,

$$1 \leq \left(\frac{3n}{2}\right)^n \int_{z\in\overline{B}_\infty(x,2^{-\ell})} \mathtt{C}\left(\hat{f}, \text{Ю}_{k,+1}(z)\right)^n \,\mathrm{d}z.$$

Therefore the final set of points, those that pass either the test in line 11 or the one in line 13, is bounded by

$$\left(\frac{3n}{2}\right)^n \sum_{k=0}^{n+1} \int_{z\in\overline{B}_\infty(0,1)} \mathtt{C}\left(\hat{f}, \text{Ю}_{k,+1}(z)\right)^n \,\mathrm{d}z.$$

Now, for all k, $\partial_z \text{Ю}_{k,+1}$ has singular values

$$\frac{1}{\sqrt{1+\|z\|^2}}, \dots, \frac{1}{\sqrt{1+\|z\|^2}}, \frac{1}{1+\|z\|^2}$$

and so

$$|\det \mathrm{D}_{\hat{z}} \text{Ю}^{-1}_{k,+1}| = \frac{1}{\left(1+\left\|\text{Ю}^{-1}(\hat{z})\right\|^2\right)^{\frac{n+1}{2}}} = \frac{1}{\|z\|_\infty^{n+1}}.$$

And so, doing a change of variables, we get the bound

$$\frac{1}{2}\left(\frac{3n}{2}\right)^n \int_{\hat{z}\in\mathbb{S}^n} \frac{\mathtt{C}\left(\hat{f},\hat{z}\right)^n}{\|\hat{z}\|_\infty^{n+1}} \,\mathrm{d}\hat{z}.$$

This concludes the proof of (b) since the number of total points considered in the cycle is at most the double of the final number of points.

For the number of arithmetic operations, we apply [12, Proposition 16.32]. □

4.1.3 Elimination of redundant approximations (Lines 21-27)

The following proposition finishes the proof of Theorem 4.1.

Proposition 4.6. *Lines 21-27 of* `aCKMW` *produces a set satisfying the postcondition of the algorithm. The computation in them takes at most* $\mathcal{O}\left(n\mathcal{D}\left|\tilde{\mathcal{Z}}\right|\right)$ *arithmetic operations. In particular, it requires*

$$\mathcal{O}\left(\left(\frac{3n}{2}\right)^n (n\mathcal{D}) \int_{x\in\mathbb{S}^n} \frac{\mathtt{C}\left(\hat{f},x\right)^n}{\|x\|_\infty^{n+1}} \,\mathrm{d}x\right)$$

arithemtic operations.

Proof. Correctness follows from Proposition 4.5(i), as if two $\overline{B}_{\mathbb{S}}(z,r)$ intersect, they must approximate the same root.

We compare each elements of $\tilde{Z}$ with the remainder elements of $\tilde{Z}$ that haven't been removed. Now, there are at most $\mathcal{D}$ points that we will select. So we will do at most $\mathcal{D}|\tilde{Z}|$ comparison. Note that each comparison requires $\mathcal{O}(n)$ operations, so the desired bound follows. The last part is Proposition 4.5(ii). □

4.2 Probabilistic complexity analysis

We prove now the probabilistic statements of Theorem 2.21. We will focus in proving points (i) and (ii) as the rest are similar.

4.2.1 Probabilistic tools

The following proposition is a version of [65, Theorem 1.1] with the explicit constants of [56].

Proposition 4.7 (Anti-concentration bound). *Let $\mathfrak{x} \in \mathbb{R}^N$ be a random vector such that its components $\mathfrak{x}_i$ are random variables with anti-concentration property with constant ρ. Then, for every orthogonal projection $P : \mathbb{R}^N \to \mathbb{R}^N$ and measurable $U \subseteq \mathbb{R}^k$,*

$$\mathbb{P}_{\mathfrak{x}}\left(A\mathfrak{x} \in U\right) \leq \mathsf{vol}(U)(\sqrt{2}\rho)^k.$$

□

The following technical lemmas will be useful for some of our estimates. The first one is just Stirling's approximation and the second one a simple application of the change of coordinates.

Lemma 4.8. *For all $n \geq 1$,*

$$\frac{1}{\sqrt{2\pi n}}\left(\frac{2\pi e}{n}\right)^{\frac{n}{2}} \leq \omega_n \leq \frac{1}{\sqrt{\pi n}}\left(\frac{2\pi e}{n}\right)^{\frac{n}{2}}$$

where ω_n is the volume of the n-dimensional ball $B(0,1) \subset \mathbb{R}^n$. □

Lemma 4.9. *[12, Lemma 2.31] Let $r \in [0, 1/2]$ and $x \in \mathbb{S}^n$. Then*

$$\omega_n\ (0.95r)^n \leq \omega_n \sin^n r \leq \mathsf{vol}_n(B_{\mathbb{S}}(x,r)) \leq \omega_n\, r^n \tag{4.5}$$

where ω_n is the volume of the n-dimensional ball $B(0,1) \subset \mathbb{R}^n$. □

As the following integral will appear once and once again, we put its computation in a lemma.

Lemma 4.10. *Let $\alpha \geq 1$ and $\beta > 1$. Then*

$$\int_1^\infty \frac{\ln^\alpha t}{t^\beta}\,\mathrm{d}t = \frac{\Gamma(\alpha+1)}{(\beta-1)^{\alpha+1}} \leq \mathcal{O}\left(\left(\frac{\alpha}{e(\beta-1)}\right)^{\alpha+1}\right).$$

□

4.2.2 Tail bound for $\mathtt{C}$: Average case

We prove now the tail bound for $\mathtt{C}(\mathfrak{f}, x)$ for $\mathfrak{f}$ dobro and $x \in \mathbb{S}^n$.

Theorem 4.11. *Let $x \in \mathbb{S}^n$ and $\mathfrak{f} \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ dobro. Then for all $t \geq 2n$,*

$$\mathbb{P}_{\mathfrak{f}}\left(\mathtt{C}(\hat{\mathfrak{f}}, x) \geq t\right) \leq 2^{n\log(n)+13.5n}\mathcal{D}^2_{\text{Д}}(\mathfrak{f})^2\frac{\ln^{2n}(t)}{t^{n+1}}$$

Corollary 4.12. *Let $x \in \mathbb{S}^n$ and $\mathfrak{q}_\sigma \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ dobro. Then*

$$\mathbb{E}_{\mathfrak{q}_\sigma}\mathtt{C}(\hat{\mathfrak{f}}, x)^n = \mathcal{O}\left(2^{3n\log(n)+13.5n}\mathcal{D}^2_{\text{Д}}(\mathfrak{f})^2\right).$$

The proof rely on a lemma for controlling the L_∞-norm of a dobro random polynomial, which is given in [24, Proposition 4.30]. We give the optimized version for just a single polynomial (using [24, Proposition 4.23]).

Lemma 4.13. *Let $\mathfrak{f} \in \mathcal{H}_{n,d}$ be a dobro random polynomial with parameters K and ρ. Then, for all $t > 0$,*

$$\mathbb{P}\left(\|\mathfrak{f}\|_\infty \geq t\right) \leq 2\exp\left(-\frac{t^2}{128K^2 n \ln(ed)}\right).$$

□

Proof of Theorem 4.11. If $\mathtt{C}(\hat{f}, x) \geq t$, then for all i,

$$|f_i(x)| \leq d_i \|f_i\|_\infty t^{-1},$$

and there is $v \in \mathbb{S}(\mathrm{T}_x\mathbb{S}^n)$ such that for all i,

$$|\mathrm{D}_x f_i(v)| \leq n d_i^2 \|f_i\|_\infty t^{-1}.$$

To see the last one we use the max-min principle (3.1).

Now, by Kellogg's theorem (Theorem 2.1), the second condition implies that

$$\{w \in \mathbb{S}(\mathrm{T}_x\mathbb{S}^n) \mid \forall i,\ |\mathrm{D}_x f_i(v)| \leq 2n d_i^2 \|f_i\|_\infty t^{-1}\} \supset \overline{B}_{\mathbb{S}}(v, nt^{-1}).$$

Now, by Lemmas 4.8 and 4.9, the latter means that

$$\mathbb{P}_{\mathfrak{v}\in\mathbb{S}(\mathrm{T}_x\mathbb{S}^n)}\left(\forall i,\ |\mathrm{D}_x f_i(\mathfrak{v})| \leq 2n d_i^2 \|f_i\|_\infty t^{-1}\right) \geq 0.25\sqrt{n}(0.95nt^{-1})^{n-1},$$

when $t \geq 2n$.

We consider the following events for the random $\mathfrak{f}$:

- A_i: $|\mathfrak{f}_i(x)| \leq d_i \|\mathfrak{f}_i\|_\infty t^{-1}$.
- B_i: $|\mathrm{D}_x \mathfrak{f}_i(\mathfrak{v})| \leq 2n d_i^2 \|\mathfrak{f}_i\|_\infty t^{-1}$.

By the above and the implication bound, we have that

$$\begin{aligned}
&\mathbb{P}_{\mathfrak{f}}(\mathtt{C}(\hat{\mathfrak{f}}, x) \geq t) \\
&\quad \leq \mathbb{P}_{\mathfrak{f}}\left(\cap_i A_i \ \&\ \mathbb{P}_{\mathfrak{v}\in\mathbb{S}(\mathrm{T}_x\mathbb{S}^n)}\left(\cap_i B_i\right) \geq 0.25\sqrt{n}(0.95nt^{-1})^{n-1}\right) \\
&\quad \leq \mathbb{P}_{\mathfrak{f}}\left(\mathbb{P}_{\mathfrak{v}\in\mathbb{S}(\mathrm{T}_x\mathbb{S}^n)}\left(\cap_i (A_i \cap B_i)\right) \geq 0.25\sqrt{n}(0.95nt^{-1})^{n-1}\right) && (A_i \text{ independent of } \mathfrak{v}) \\
&\quad \leq \frac{4}{\sqrt{n}}\left(\frac{1.1t}{n}\right)^{n-1} \mathbb{E}_{\mathfrak{f}}\mathbb{P}_{\mathfrak{v}\in\mathbb{S}(\mathrm{T}_x\mathbb{S}^n)}\left(\cap_i (A_i \cap B_i)\right) && (\text{Markov's inequality}) \\
&\quad \leq \frac{4}{\sqrt{n}}\left(\frac{1.1t}{n}\right)^{n-1} \mathbb{E}_{\mathfrak{v}\in\mathbb{S}(\mathrm{T}_x\mathbb{S}^n)}\mathbb{P}_{\mathfrak{f}}\left(\cap_i (A_i \cap B_i)\right) && (\text{Tonelli's theorem}) \\
&\quad \leq \frac{4}{\sqrt{n}}\left(\frac{1.1t}{n}\right)^{n-1} \mathbb{E}_{\mathfrak{v}\in\mathbb{S}(\mathrm{T}_x\mathbb{S}^n)} \prod_i \mathbb{P}_{\mathfrak{f}}\left(A_i \cap B_i\right) && (A_i \cap B_i \text{ independent})
\end{aligned}$$

Now, we only need to bound each $\mathbb{P}_{\mathfrak{f}}(A_i \cap B_i)$ when $\mathfrak{v}$ is fixed to a constnat v. By a simple probabilistic argument, we have that

$$\begin{aligned}
&\mathbb{P}_{\mathfrak{f}}(A_i \cap B_i) \\
&\quad = \mathbb{P}_{\mathfrak{f}}(|\mathfrak{f}_i(x)| \le d_i \|\mathfrak{f}_i\|_\infty t^{-1},\, |\mathrm{D}_x \mathfrak{f}_i(v)| \le 2nd_i^2 \|\mathfrak{f}_i\|_\infty t^{-1}) \\
&\quad \le \mathbb{P}_{\mathfrak{f}}(\|\mathfrak{f}_i\|_\infty \ge Q_i) + \mathbb{P}_{\mathfrak{f}}(\|\mathfrak{f}_i\|_\infty \le Q_i,\, |\mathfrak{f}_i(x)| \le d_i \|\mathfrak{f}_i\|_\infty t^{-1},\, |\mathrm{D}_x \mathfrak{f}_i(v)| \le 2nd_i^2 \|\mathfrak{f}_i\|_\infty t^{-1}) \\
&\quad \le \mathbb{P}_{\mathfrak{f}}(\|\mathfrak{f}_i\|_\infty \ge Q_i) + \mathbb{P}_{\mathfrak{f}}(\|\mathfrak{f}_i\|_\infty \le Q_i,\, |\mathfrak{f}_i(x)| \le d_i Q_i t^{-1},\, |\mathrm{D}_x \mathfrak{f}_i(v)| \le 2nd_i^2 Q_i t^{-1}) \\
&\quad \le \mathbb{P}_{\mathfrak{f}}(\|\mathfrak{f}_i\|_\infty \ge Q_i) + \mathbb{P}_{\mathfrak{f}}(|\mathfrak{f}_i(x)| \le d_i Q_i t^{-1},\, |\mathrm{D}_x \mathfrak{f}_i(v)| \le 2nd_i^2 Q_i t^{-1})
\end{aligned}$$

For the first summand, we apply Lemma 4.13. For the second one, we apply Proposition 4.7, since, in the orthogonal coordinates of the Weyl basis,

$$f_i \mapsto \begin{pmatrix} f_i(x) \\ d_i^{-1/2} \mathrm{D}_x f_i(v) \end{pmatrix}$$

is an orthogonal projection (by [12, §16.3] for example). In this way, we get

$$\mathbb{P}_{\mathfrak{f}}(A_i \cap B_i) \le 2 \exp\left(-\frac{Q_i^2}{128K^2 n \ln(ed_i)} \right) + 8nd_i^{\frac{3}{2}} \rho_i^2 Q_i^2 t^{-2}.$$

By substituting $Q_i^2 = 256K^2 n \ln(ed_i) \ln t$, we get

$$\mathbb{P}_{\mathfrak{f}}(A_i \cap B_i) \le 2t^{-2} + 2^{11} n^2 d_i^{\frac{3}{2}} \ln(ed_i) K_i^2 \rho_i^2 \ln^2(t) t^{-2} \le 2^{12} n^2 d_i^{\frac{3}{2}} \ln(ed_i) K_i^2 \rho_i^2 \ln^2(t) t^{-2}.$$

Putting everything together and some simple computations give the desired bound. □

Proof of Corollary 4.12. We use the fact that

$$\mathbb{E}_{\mathfrak{f}} \mathtt{C}(\hat{f}, x)^n = \int_1^\infty \mathbb{P}_{\mathfrak{f}}(\mathtt{C}(\hat{f}, x) \ge t^{1/n})\, \mathrm{d}t$$

together with Theorem 4.11 and Lemma 4.10. □

The results here prove (i) in Theorem 2.21, when combined with Theorem 4.1 and (4.1).

4.2.3 Tail bound for C: Smoothed case

The smoothed case is very similar to the average case and the result almost identical.

Theorem 4.14. *Let $x \in \mathbb{S}^n$ and $\mathfrak{q}_\sigma \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ σ-smoothed dobro. Then for all $t \ge 2n$,*

$$\mathbb{P}_{\mathfrak{q}_\sigma}\left(\mathtt{C}(\hat{\mathfrak{f}}, x) \ge t \right) \le 2^{n \log(n) + 13.5n} \mathcal{D}^2_{\text{Д}}(\mathfrak{f})^2 \frac{\ln^{2n}(t)}{t^{n+1}} \left(1 + \frac{1}{\sigma} \right)^{2n}.$$

Corollary 4.15. *Let $x \in \mathbb{S}^n$ and $\mathfrak{q}_\sigma \in \mathcal{H}_{n,\boldsymbol{d}}[n]$ σ-smoothed dobro. Them*

$$\mathbb{E}_{\mathfrak{q}_\sigma} \mathtt{C}(\hat{\mathfrak{f}}, x)^n = \mathcal{O}\left(2^{3n \log(n) + 13.5n} \mathcal{D}^2_{\text{Д}}(\mathfrak{f})^2 \left(1 + \frac{1}{\sigma} \right)^{2n} \right).$$

We only need to substitute Lemma 4.13 in the proof of Theorem 4.11 by the following lemma to prove Theorem 4.14.

Lemma 4.16. *Let $\mathfrak{q}_\sigma \in \mathcal{H}_{n,\boldsymbol{d}}[q]$ be σ-smoothed dobro. Then for all i,*

$$\mathbb{P}\left(\|\mathfrak{q}_{\sigma,i}\|_\infty \geq t\right) \leq 2\exp\left(-\frac{t^2}{128K_i^2 n\ln(ed_i)}\right).$$

Proof. By the triangular inequality,

$$\mathbb{P}(\|\mathfrak{q}_{\sigma,i}\|_\infty \geq t\|f\|) \leq \mathbb{P}(\|\mathfrak{f}_i\|_\infty \geq (t-1)/\sigma).$$

Here, Lemma 4.13 finishes the proof. □

Sketch of proof of Theorem 4.14. Note that the probabilistic assumptions only enter at the end of the proof of Theorem 4.11. There we use Lemma 4.16 instead of Lemma 4.13. □

Proof of Corollary 4.15. As the proof of Corollary 4.12, but with Theorem 4.14 instead of Theorem 4.11. □

The results here prove (ii) in Theorem 2.21, when combined with Theorem 4.1 and (4.1).

5 Finite precision and Parallelization

We discuss briefly here how to get the results for finite precision and parallelization.

5.1 Finite precision

To see the computation in finite precision, one should just follow the results from the original [26]. The main difference is that we have to guarantee that the precision grows sufficiently in terms of ℓ, which measure the fineness of the grid. This is similar to what was done in [25] with the Plantinga-Vegter algorithm.

We note, however, that one does not need to be as precise as in [25] with the finite precision analysis. If one wants certification is enough to use interval arithmetic with an increasing of the precision compatible with bounds obtained.

Regarding the complexity, when we take into account the finite precision, the condition based complexity will depend on expression of the form

$$\mathbb{E}_{\mathfrak{x}\in\mathbb{S}^n}\left(\frac{\mathtt{C}\left(\hat{f},\mathfrak{x}\right)^n}{\|\mathfrak{x}\|_\infty^{n+1}}\log^l \mathtt{C}\left(\hat{f},\mathfrak{x}\right)\right).$$

These expression can be easily controlled using the probabilistic bounds given.

5.2 Parallelization

Parallelization of the grid methods is easy. We only have to perform the cycles through the points in the grid in parallel. We refer to [76, $4^{\S 3}$-1] to more details of how the parallelization of the grid method can be done in parallel.

The only point to be careful is the elimination of redundant approximations. In it, we must substitute the existing method, by a tournament method. In this method, we would star with pairs of approximations. In each round, we compare all survivor of one group

against all survivors of the other group. This can be parallelized, as in each match we will perform at most $\mathcal{D}^2$ comparissons. In this way, the parallel run-time of this method is

$$\mathcal{O}(\log |\tilde{Z}\mathcal{D}|)$$

with $\mathcal{O}(\mathcal{D}^2|\tilde{\mathcal{Z}}|)$ operations at most.

Acknowledgments.

I am grateful to Evgenia Lagoda for moral support and Gato Suchen for useful suggestions for this paper.

References

[1] O. Aberth. *Introduction to precise numerical methods*. Elsevier/Academic Press, Amsterdam, second edition, 2007.

[2] D. Amelunxen and M. Lotz. Average-case complexity without the black swans. *J. Complexity*, 41:82–101, 2017.

[3] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[4] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler. Bertini: Software for numerical algebraic geometry. Available at bertini.nd.edu with permanent doi: dx.doi.org/10.7274/R0H41PB5.

[5] C. Beltrán and L. M. Pardo. Fast linear homotopy to find approximate zeros of polynomial systems. *Found. Comput. Math.*, 11(1):95–129, 2011.

[6] Carlos Beltrán and Michael Shub. A note on the finite variance of the averaging function for polynomial system solving. *Found. Comput. Math.*, 10(1):115–125, 2010.

[7] L. Benet and D. P. Sanders. Intervalrootfinding.jl, 2014. `https://juliaintervals.github.io/IntervalRootFinding.jl/`.

[8] C. E. Borges and L. M. Pardo. On the probability distribution of data at points in real complete intersection varieties. *J. Complexity*, 24(4):492–523, 2008.

[9] C. Brand and M. Sagraloff. On the complexity of solving zero-dimensional polynomial systems via projection. In *Proceedings of the 2016 ACM International Symposium on Symbolic and Algebraic Computation*, pages 151–158. ACM, New York, 2016.

[10] P. Breiding, K. Rose, and S. Timme. Certifying zeros of polynomial systems using interval arithmetic, November 2020. arXiv:2011.05000.

[11] P. Breiding and S. Timme. Homotopycontinuation.jl: A package for homotopy continuation in julia. In J. H. Davenport, M. Kauers, G. Labahn, and J. Urban, editors, *Mathematical Software – ICMS 2018*, pages 458–465, Cham, 2018. Springer International Publishing.

[12] P. Bürgisser and F. Cucker. *Condition: The geometry of numerical algorithms*, volume 349 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, 2013.

[13] P. Bürgisser, F. Cucker, and P. Lairez. Computing the homology of basic semialgebraic sets in weak exponential time. *J. ACM*, 66(1):5:1–5:30, December 2018.

[14] P. Bürgisser, F. Cucker, and P. Lairez. Rigid continuation paths II. Structured polynomial systems. arXiv:2010.10997, October 2020.

[15] P. Bürgisser, F. Cucker, and J. Tonelli-Cueto. Computing the Homology of Semialgebraic Sets. I: Lax Formulas. *Foundations of Computational Mathematics*, 20(1):71–118, 2020. On-line from May of 2019.

[16] P. Bürgisser, F. Cucker, and J. Tonelli-Cueto. Computing the Homology of Semialgebraic Sets. II: General formulas. *Foundations of Computational Mathematics*, 1 2021. On-line from January of 2021. arXiv: 1903.10710.

[17] M. A. Burr, F. Krahmer, and C. K. Yap. Continuous amortization: A non-probabilistic adaptive analysis technique. *Electronic Colloquium on Computational Complexity*, Report. No. 136, 2009.

[18] J. Canny, D. Yu. Grigor˜ ev, and N. N. Vorobjov, Jr. Finding connected components of a semialgebraic set in subexponential time. *Appl. Algebra Engrg. Comm. Comput.*, 2(4):217–238, 1992.

[19] T. Chen, T.-L. Lee, and T.-Y. Li. Hom4PS-3: a parallel numerical solver for systems of polynomial equations based on polyhedral homotopy continuation methods. In *Mathematical software—ICMS 2014*, volume 8592 of *Lecture Notes in Comput. Sci.*, pages 183–190. Springer, Heidelberg, 2014.

[20] F. Cucker. Approximate zeros and condition numbers. *J. Complexity*, 15(2):214–226, 1999.

[21] F. Cucker. Grid methods in computational real algebraic (and semialgebraic) geometry. *Chin. Ann. Math. Ser. B*, 39(2):373–396, 2018.

[22] F. Cucker. Recent advances in the computation of the homology of semialgebraic sets. In F. Manea, Barnaby Martin, D. Paulusma, and G. Primiero, editors, *Computing with Foresight and Industry*, pages 1–12, Cham, 2019. Springer International Publishing.

[23] F. Cucker, A. A. Ergür, and J. Tonelli-Cueto. Plantinga-Vegter Algorithm Takes Average Polynomial Time. In *Proceedings of the 2019 on International Symposium on Symbolic and Algebraic Computation*, ISSAC '19, pages 114–121. ACM, New York, 2019. (arXiv:1901.09234v2).

[24] F. Cucker, A. A. Ergür, and J. Tonelli-Cueto. Functional norms, condition numbers and numerical algorithms in algebraic geometry. Manuscript. To appear in arXiv by the end of February, 2020.

[25] F. Cucker, A. A. Ergür, and J. Tonelli-Cueto. On the Complexity of the Plantinga-Vegter Algorithm, April 2020. arXiv:2004.06879.

[26] F. Cucker, T. Krick, G. Malajovich, and M. Wschebor. A numerical algorithm for zero counting. I. Complexity and accuracy. *J. Complexity*, 24(5-6):582–605, 2008.

[27] F. Cucker, T. Krick, G. Malajovich, and M. Wschebor. A numerical algorithm for zero counting. II. Distance to ill-posedness and smoothed analysis. *J. Fixed Point Theory Appl.*, 6(2):285–294, 2009.

[28] F. Cucker, T. Krick, G. Malajovich, and M. Wschebor. A numerical algorithm for zero counting. III: Randomization and condition. *Adv. in Appl. Math.*, 48(1):215–248, 2012.

[29] F. Cucker, T. Krick, and M. Shub. Computing the homology of real projective sets. *Found. Comput. Math.*, 18:929–970, 2018.

[30] F. Cucker and S. Smale. Complexity estimates depending on condition and round-off error. *J. ACM*, 46(1):113–184, 1999.

[31] J.-P. Dedieu. *Points fixes, zéros et la méthode de Newton*, volume 54 of *Mathématiques & Applications*. Springer, Berlin, 2006.

[32] J. W. Demmel. The probability that a numerical analysis problem is difficult. *Math. Comp.*, 50:449–480, 1988.

[33] T. Duff, C. Hill, A. Jensen, K. Lee, A. Leykin, and J. Sommars. Solving polynomial systems via homotopy continuation and monodromy. *IMA J. Numer. Anal.*, 39(3):1421–1446, 2019.

[34] A. Eckhardt. An Adaptive Algorithm for Computing the Homology of Semialgebraic Sets. Master's thesis, Technsiche Universität Berlin, 2020.

[35] G. Elber and M.-S. Kim. Geometric constraint solver using multivariate rational spline functions. In *Proceedings of the Sixth ACM Symposium on Solid Modeling and Applications*, SMA '01, page 1–10, New York, NY, USA, 2001. Association for Computing Machinery.

[36] A. A. Ergür, G. Paouris, and J. M. Rojas. Probabilistic condition number estimates for real polynomial systems I: A broader family of distributions. *Found. Comput. Math.*, 19(1):131–157, 2019.

[37] A. A. Ergür, G. Paouris, and J. Maurice Rojas. Probabilistic Condition Number Estimates for Real Polynomial Systems II: Structure and Smoothed Analysis, September 2018. (arXiv:1809.03626).

[38] J.-C. Faugère. FGb: A Library for Computing Gröbner Bases. In K. Fukuda, J. Hoeven, M. Joswig, and N. Takayama, editors, *Mathematical Software - ICMS 2010*, volume 6327 of *Lecture Notes in Computer Science*, pages 84–87, Berlin, Heidelberg, September 2010. Springer Berlin / Heidelberg.

[39] J. Garloff and A. P. Smith. Investigation of a subdivision based algorithm for solving systems of polynomial equations. In *Proceedings of the Third World Congress of Nonlinear Analysts, Part 1 (Catania, 2000)*, volume 47, pages 167–178, 2001.

[40] J. Garloff and A. P. Smith. Solution of systems of polynomial equations by using Bernstein expansion. In *Symbolic algebraic methods and verification methods (Dagstuhl, 1999)*, pages 87–97. Springer, Vienna, 2001.

[41] M. Giusti, G. Lecerf, and B. Salvy. A Gröbner free alternative for polynomial system solving. *J. Complexity*, 17(1):154–211, 2001.

[42] H. H. Goldstine and J. von Neumann. Numerical inverting of matrices of high order. II. *Proc. Amer. Math. Soc.*, 2:188–202, 1951.

[43] D. Yu. Grigoriev and N. N. Vorobjov. Solving systems of polynomial inequalities in subexponential time. *J. Symbolic Comput.*, 5(1-2):37–64, 1988.

[44] J. Han. An Adaptive Grid Algorithm for Computing the Homology Group of Semialgebraic Set. Master's thesis, Université Paris Sud, 2018.

[45] J. Han. An Adaptive Grid Algorithm for Computing the Homology Group of Semialgebraic Set, March 2019. (arXiv:1903.02388).

[46] A. Hashemi and D. Lazard. Sharper complexity bounds for zero-dimensional Gröbner bases and polynomial system solving. *Internat. J. Algebra Comput.*, 21(5):703–713, 2011.

[47] J. D. Hauenstein and F. Sottile. Algorithm 921: alphaCertified: certifying solutions to polynomial systems. *ACM Trans. Math. Software*, 38(4):Art. 28, 20, 2012.

[48] R. B. Kearfott. Globsol: History, composition, and advice on use. In C. Bliek, C. Jermann, and A. Neumaier, editors, *Global Optimization and Constraint Satisfaction*, pages 17–31, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[49] O. D. Kellogg. On bounded polynomials in several variables. *Math. Z.*, 27(1):55–64, 1928.

[50] A. Kobel, F. Rouillier, and M. Sagraloff. Computing real roots of real polynomials ... and now for real! In *Proceedings of the 2016 ACM International Symposium on Symbolic and Algebraic Computation*, pages 303–310. ACM, New York, 2016.

[51] P. Lairez. A deterministic algorithm to compute approximate roots of polynomial systems in polynomial average time. *Found. Comput. Math.*, 17(5):1265–1292, 2017.

[52] P. Lairez. Rigid continuation paths I. Quasilinear average complexity for solving polynomial systems. *J. Amer. Math. Soc.*, 33(2):487–526, 2020.

[53] K. Lee. Certifying approximate solutions to polynomial systems on Macaulay2. *ACM Commun. Comput. Algebra*, 53(2):45–48, 2019.

[54] A. Leykin. Numerical algebraic geometry. *Journal of Software for Algebra and Geometry*, 3:5–10, June 2011.

[55] T. Y. Li and Xiao Shen Wang. Solving real polynomial systems with real homotopies. *Math. Comp.*, 60(202):669–680, 1993.

[56] G. Livshyts, G. Paouris, and P. Pivovarov. On sharp bounds for marginal densities of product measures. *Israel Journal of Mathematics*, 216(2):877–889, 2016.

[57] A. Mantzaflaris, B. Mourrain, and E. Tsigaridas. On continued fraction expansion of real roots of polynomial systems, complexity and condition numbers. *Theoret. Comput. Sci.*, 412(22):2312–2330, 2011.

[58] B. Mourrain and J. P. Pavone. Subdivision methods for solving polynomial equations. *J. Symbolic Comput.*, 44(3):292–306, 2009.

[59] A. Neumaier. *Interval methods for systems of equations*, volume 37 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1990.

[60] V. Y. Pan. Univariate polynomials: nearly optimal algorithms for numerical factorization and root-finding. *J. Symbolic Comput.*, 33(5):701–733, 2002. Computer algebra (London, ON, 2001).

[61] G. Paouris, K. Phillipson, and J. M. Rojas. A faster solution to Smale's 17th problem I: real binomial systems. In *ISSAC'19—Proceedings of the 2019 ACM International Symposium on Symbolic and Algebraic Computation*, pages [323–330]. ACM, New York, 2019.

[62] Simon Plantinga and Gert Vegter. Isotopic approximation of implicit curves and surfaces. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, SGP '04, pages 245–254, New York, NY, USA, 2004. ACM.

[63] J. Renegar. On the efficiency of Newton's method in approximating all zeros of a system of complex polynomials. *Math. Oper. Res.*, 12(1):121–148, 1987.

[64] F. Rouillier. Solving zero-dimensional systems through the rational univariate representation. *Appl. Algebra Engrg. Comm. Comput.*, 9(5):433–461, 1999.

[65] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Not. IMRN*, 19:9594–9617, 2015.

[66] S. M. Rump. INTLAB - INTerval LABoratory. In Tibor Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. `http://www.ti3.tuhh.de/rump/`.

[67] S. M. Rump. Verification methods: rigorous results using floating-point arithmetic. *Acta Numer.*, 19:287–449, 2010.

[68] M. Safey El Din. Raglib: A library for real solving polynomial systems of equations and inequalities. `https://www-polsys.lip6.fr/~safey/RAGLib/`.

[69] M. Sagraloff and K. Mehlhorn. Computing real roots of real polynomials. *J. Symbolic Comput.*, 73:46–86, 2016.

[70] E. C. Sherbrooke and N. M. Patrikalakis. Computation of the solutions of nonlinear polynomial systems. *Comput. Aided Geom. Design*, 10(5):379–405, 1993.

[71] M. Shub. Complexity of Bezout's theorem. VI. Geodesics in the condition (number) metric. *Found. Comput. Math.*, 9(2):171–178, 2009.

[72] S. Smale. Complexity theory and numerical analysis. In A. Iserles, editor, *Acta Numerica*, pages 523–551. Cambridge University Press, 1997.

[73] Steve Smale. On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc. (N.S.)*, 13(2):87–121, 1985.

[74] Steve Smale. Newton's method estimates from data at one point. In *The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985)*, pages 185–196. Springer, New York, 1986.

[75] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians*, volume I, pages 597–606, 2002.

[76] J. Tonelli-Cueto. *Condition and Homology in Semialgebraic Geometry.* Doctoral thesis, Technische Universität Berlin, DepositOnce Repository, December 2019. http://dx.doi.org/10.14279/depositonce-9453.

[77] P. Van Hentenryck, D. McAllester, and D. Kapur. Solving polynomial systems using a branch and prune approach. *SIAM J. Numer. Anal.*, 34(2):797–827, 1997.

[78] J. Verschelde. Algorithm 795: Phcpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Trans. Math. Softw.*, 25(2):251–276, June 1999.

[79] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.

[80] J. Xu and C. Yap. Effective subdivision algorithm for isolating zeros of real systems of equations, with complexity analysis. In *ISSAC'19—Proceedings of the 2019 ACM International Symposium on Symbolic and Algebraic Computation*, pages 355–362. ACM, New York, 2019.

[81] C. Yap. Towards soft exact computation (*invited talk*). In *Computer algebra in scientific computing*, volume 11661 of *Lecture Notes in Comput. Sci.*, pages 12–36. Springer, Cham, 2019.

A Proofs of cited results of [24]

We include for completeness the proof from results cited from [24], so that they are available easily to the reader.

Proof of Theorem 2.1. By [49, Theorem IV], we have that

$$|\partial_x f(v)| \leq d\|f\|_\infty \|v\|.$$

The result follows applying induction. □

Proof of Proposition 4.3. Let x_* be the maximum of $|f|$ on $\mathbb{S}^n$, $x \in \mathcal{G}$ such that

$$\operatorname{dist}_{\mathbb{S}}(x_*, x) \leq \eta$$

and $[0,1] \ni t \mapsto x_t$ the geodesic on $\mathbb{S}^n$ going from x_* to x with constant speed. Then, for the function $t \mapsto M(t) := f(x_t)$, we have that $|M(1)| \leq |M(0)| + |M'(0)| + \max_{s\in[0,1]} \frac{M''(s)}{2}$ by Taylor's theorem. Furthermore, $|M(0)| = |f(x_*)| = \|f\|_\infty$, $|M(1)| = |f(x)|$ and $M'(0) = 0$. The latter is due to the fact that x_* is an extremal point of f and so of M. Now,

$$M''(t) = \partial^2_{x_t} f(\dot{x}_t, \dot{x}_t) - df(x_t) \operatorname{dist}_{\mathbb{S}}(x_*, x)^2,$$

since $\ddot{x}_t = -\operatorname{dist}_{\mathbb{S}}(x_*, x)^2 x_t$, as x_t is a geodesic on $\mathbb{S}^n$ of constant speed $\operatorname{dist}_{\mathbb{S}}(x_*, x)$, and $\partial_{x_t} f(x_t) = df(x_t)$ by Euler's formula for homogeneous polynomials. Then, by Kellogg's Theorem (Theorem 2.1),

$$\max_{s\in[0,1]} \frac{|M''(s)|}{2} \leq \binom{d}{2}\|f\|_\infty + \frac{d}{2}\|f\|_\infty = \frac{d^2}{2}\|f\|_\infty.$$

Thus $\|f\|_\infty \leq |f(x)| + \frac{d^2}{2}\|f\|_\infty \eta^2$, and the desired inequality follows. □

For proving Lemma 4.13, the following two propositions of [24] are needed. Note that these are just explicit version of results in [79].

Proposition A.1 (Subgaussian tail bounds). *Let $\mathfrak{x} \in \mathbb{R}$ be a random variable.*

1. *If $\mathfrak{x}$ is subgaussian with ψ_2-norm at most K, then for all $t > 0$, $\mathbb{P}(|\mathfrak{x}| \geq t) \leq 2e^{-\frac{t^2}{2K^2}}$.*
2. *If there are $C > 1$ and $K > 0$ such that for all $t > 0$, $\mathbb{P}(|\mathfrak{x}| \geq t) \leq Ce^{-\frac{t^2}{K^2}}$, then $\mathfrak{x}$ is subgaussian with ψ_2-norm at most $K\left(1 + \sqrt{2\ln C}\right)$.*

Proposition A.2 (Hoeffding inequality). *Let $\mathfrak{x} \in \mathbb{R}^N$ be a random vector such that its components $\mathfrak{x}_i$ are centered subgaussian random varianbles with ψ_2-norm at most K and $a \in \mathbb{S}^{N-1}$. Then, $a^*\mathfrak{x}$ is a subgaussian random variable with ψ_2-norm at most $\frac{5}{4}K$. In particular, for all $t \geq 0$,*

$$\mathbb{P}_{\mathfrak{x}}\left(|a^*\mathfrak{x}| \geq t\right) \leq 2e^{-\frac{8t^2}{25K^2}}.$$

Proof of Proposition 4.13. Fix $\eta \in [0, 1/d]$. By the proof of Proposition 4.3, we have that $\|\mathfrak{f}\|_\infty > t$ implies $\mathsf{vol}\left\{x \in \mathbb{S}^n \mid \|\mathfrak{f}(x)\|_\infty \geq \left(1 - \frac{d^2}{2}\eta^2\right)t\right\} \geq \mathsf{vol}B_{\mathbb{S}}(x_*, \eta)$, where $x_* \in \mathbb{S}^n$ maximizes $\|f(x)\|_\infty$. Therefore

$$\mathbb{P}\left(\|\mathfrak{f}\|_\infty \geq t\right) \leq \mathbb{P}_{\mathfrak{f}}\left(\mathbb{P}_{\mathfrak{x}\in\mathbb{S}^n}\left(\|\mathfrak{f}(\mathfrak{x})\|_\infty \geq \left(1 - \frac{d^2}{2}\eta^2\right)t\right) \geq \mathsf{vol}B_{\mathbb{S}}(x_*, \eta)/\mathsf{vol}\mathbb{S}^n\right).$$

Now, by Stirling's approximation [12, Eq. (2.14)] and [12, Lemma 2.31] (plus some estimations of $\int_0^\eta \sin^{n-1}\theta\,\mathrm{d}\theta$), we have that

$$\mathsf{vol}B_{\mathbb{S}}(x_*, \eta)/\mathsf{vol}\mathbb{S}^n \geq 3\sqrt{n+1}\left(\frac{2}{5}\eta\right)^n.$$

In this way,

$$\begin{aligned}
&\mathbb{P}\left(\|\mathfrak{f}\|_\infty \geq t\right)\\
&\quad\leq \mathbb{P}_{\mathfrak{f}}\left(\mathbb{P}_{\mathfrak{x}\in\mathbb{S}^n}\left(\|\mathfrak{f}(\mathfrak{x})\|_\infty \geq \left(1 - \frac{d^2}{2}\eta^2\right)t\right) \geq 3\sqrt{n+1}\left(\frac{2}{5}\eta\right)^n\right)\\
&\quad\leq \frac{1}{3\sqrt{n+1}}\left(\frac{5}{2\eta}\right)^n \mathbb{E}_{\mathfrak{f}}\mathbb{P}_{\mathfrak{x}\in\mathbb{S}^n}\left(\|\mathfrak{f}(\mathfrak{x})\|_\infty \geq \left(1 - \frac{d^2}{2}\eta^2\right)t\right) \qquad \text{(Markov's inequality)}\\
&\quad\leq \frac{1}{3\sqrt{n+1}}\left(\frac{5}{2\eta}\right)^n \mathbb{E}_{\mathfrak{x}\in\mathbb{S}^n}\mathbb{P}_{\mathfrak{f}}\left(\|\mathfrak{f}(\mathfrak{x})\|_\infty \geq \left(1 - \frac{d^2}{2}\eta^2\right)t\right) \qquad \text{(Tonelli's theorem)}\\
&\quad\leq \frac{1}{3\sqrt{n+1}}\left(\frac{5}{2\eta}\right)^n \max_{x\in\mathbb{S}^n}\mathbb{P}_{\mathfrak{f}}\left(\|\mathfrak{f}(x)\|_\infty \geq \left(1 - \frac{d^2}{2}\eta^2\right)t\right)
\end{aligned}$$

Since $\mathfrak{f}$ is dobro, for all i and $x \in \mathbb{S}^n$, $\mathfrak{f}_i(x)$ is a subgaussian random variable with ψ_2-norm at most $\frac{5}{4}K$. Note that we are using that in the coordinates of an orthogonal monomial basis for the Weyl norm, the following holds: 1) a dobro random polynomial looks like random vector whose components are independent and subgaussian of ψ_2-norm at most K, and 2) evaluation at a point of the sphere becomes inner product with a vector of norm one.

Hence

$$\mathbb{P}\left(\|\mathfrak{f}\|_\infty \geq t\right) \leq \frac{q}{3\sqrt{n+1}} \left(\frac{5}{2\eta}\right)^n \exp\left(-\left(1-\frac{d^2}{2}\eta^2\right)^2 \frac{8t^2}{25K^2}\right).$$

The claim follows taking $\eta = 5/(6d)$. For the other inequalities on the moments use Proposition A.1. □

Proof of Proposition A.1. This is just [79, Proposition 2.5.2] with a twist. For the first part, we only have to follow the constants in the proof. For the second one, note that

$$\mathbb{E}|\mathfrak{x}|^p = K^p \left(2\ln C\right)^{\frac{p}{2}} + \int_0^\infty u^{p-1} e^{-\frac{u^2}{2K}}\, \mathrm{d}u,$$

which follows from

$$\mathbb{P}(|\mathfrak{x}| > u) \leq \begin{cases} 1 & \text{if } u \leq K\sqrt{2\ln C} \\ e^{-\frac{u^2}{2K^2}} & \text{if } u \geq K\sqrt{2\ln C}, \end{cases}$$

dividing the integration domain into $[0, K\sqrt{2\ln C}]$ and $[K\sqrt{2\ln C}, \infty]$, and applying some straightforward calculations and bounds.

Now, applying the change of variables $t = \frac{u^2}{2K}$ and Stirling's inequality, we obtain

$$\int_0^\infty u^{p-1} e^{-\frac{u^2}{2K}}\, \mathrm{d}u = K^p 2^{\frac{p}{2}-1} \Gamma\left(\frac{p}{2}\right) \leq 2K^p e^{-\frac{p}{2}} p^{\frac{p}{2}}.$$

Hence

$$\mathbb{E}|\mathfrak{x}|^p \leq K^p \left(\left(2\ln C\right)^{\frac{p}{2}} + 2p^{\frac{p}{2}}\right),$$

from where the second part follows. □

Proof of Proposition A.2. This is an application of [79, Proposition 2.6.1], where we only have to find the explicit constants hidden in the proofs of [79, (2.5) and (2.6)] —the constants are given as absolute constants in the statement, but one can find the precise constants in the proofs—. Note however that for us the ψ_2-norm is the K_1 in [79, Proposition 2.5.2], while in [79] it is the constant K_4 of that proposition.

The last claim is just applying Proposition A.1. □