# Tractability results for classes of ridge functions

Sebastian Mayer
(joint work with Tino Ullrich, Jan Vybíral)

Workshop "Uniform Distribution Theory and Applications",
Oberwolfach, Germany, 2013.

Approximating functions which depend on a large number of variables can be a hard to solve problem. In many settings, it is known to suffer from the curse of dimensionality even for very smooth functions. So, if smoothness is not enough, what additional inner structure of functions could help to overcome the curse? In this regard, we investigate *ridge functions* $f(x) = g(a \cdot x)$ in high dimensions $d$ in our current work [3]. A ridge function is characterized by two components: a $d$-dimensional vector $a$, called the *ridge direction*, and a univariate function $g$, which is called the *profile*. Ridge functions provide a simple, coordinate-independent model to describe inherently one-dimensional structures hidden in a high-dimensional ambient space.

**The setting.** The problem we study is that of approximating ridge functions in the uniform norm in the worst-case by means of deterministic, adaptive algorithms. As information, the algorithms have only a limited amount of function values available. The function classes $\mathcal{F}$, where the ridge functions may be taken from, are introduced below. For the remainder of this text, let us call this setup the *sampling problem*. The key quantity, for which we find bounds from above and below, is the *n-th (adaptive) sampling number*

$$\begin{aligned}&g_{n,d}^{\mathrm{ada}}(\mathcal{F}, L_\infty)\\ &:= \inf\Big\{\sup_{f\in\mathcal{F}} \|f - S(f)\|_\infty :\ S \text{ adaptive algorithm using } n \text{ function values}\Big\}.\end{aligned}$$

So, let us finally introduce the ridge function classes. As domain $\Omega$, we fix the closed Euclidean unit ball $\Omega = \bar{B}_2^d$. Then, we impose two constraints: firstly, we restrict the norm of the feasible ridge directions, namely we require $\|a\|_p \leq 1$ for some $0 < p \leq 2$; secondly, we define the feasible profiles to be those which have a certain order of *Lipschitz*-smoothness $\alpha > 0$. We denote the so obtained classes by $\mathcal{R}_d^{\alpha,p}$. In a sense, the case $p = 2$ can be understood as all directions being possible. Letting $p$ get smaller and smaller, then imposes a more and more strict sparsity constraint on the feasible directions.

**Bounds on sampling numbers.** What could an algorithm that exploits the ridge structure look like? At least for $\alpha > 1$, there is an intuitive idea: find a point where the first derivative of the profile $g'$ is sufficiently large, do first-order Taylor around that point to recover the ridge direction $a$, and finally sample along $a$ to approximate $g$. Unfortunately, this cannot work for ridge functions on the Euclidean unit ball as given by the class $\mathcal{R}_d^{\alpha,p}$. Let us just mention the crucial point here: whatever algorithm we take, sampling the whole relevant domain of the profile is only then possible if the algorithm samples exactly along the ridge direction. Otherwise, there is always a range the algorithm cannot reach; there is no guarantee to find a point where the first derivative $g'$ is sufficiently large. It turns out that for functions from $\mathcal{R}_d^{\alpha,p}$, particularly in case $p = 2$, we cannot really do better than for general multi-variate Lipschitz functions. In terms of algorithms, this means that the best one can try is to spread the sampling points, in a certain sense, as uniformly as possible over the domain.

Let us formulate in more detail what bounds we get by the above reasoning. Both the lower and the upper bound on $g_{n,d}^{\mathrm{ada}}(\mathcal{R}_d^{\alpha,p}, L_\infty)$ are determined to a large extent by *(dyadic) entropy numbers* $e_k$. For the lower bound, it is the entropy numbers of the $p$-sphere in $\ell_2^d$; for the upper bound, it is the entropy numbers of the domain $\bar{B}_2^d$ in $\ell_{p'}^d$, where $p'$ is the dual index of $p$ given by $1/\max\{1,p\} + 1/p' = 1$. In formulas, if we let $k$ be the smallest integer such that $n \leq 2^{k-1}$, then we have bounds

$$e_k(\mathbb{S}_p^{d-1}, \ell_2^d)^{2\alpha} \lesssim g_{n,d}^{\mathrm{ada}}(\mathcal{R}_d^{\alpha,p}, L_\infty) \lesssim g_{n,d}^{\mathrm{lin}}(\mathcal{R}_d^{\alpha,p}, L_\infty) \lesssim e_{k-\Delta}(\bar{B}_2^d, \ell_{p'}^d)^{\alpha}, \qquad (1)$$

where $\Delta$ is a positive integer depending logarithmically on the dimension $d$ and the smoothness parameter $\alpha$. The behaviour of these entropy numbers is completely understood; for details, see [3, Subsection 2.3] and references therein, and [4]. If we just plug in these existing results on entropy numbers, then we obtain that sampling of ridge functions from $\mathcal{R}_d^{\alpha,2}$ is essentially as difficult as sampling of general multi-variate Lipschitz functions. For $p < 2$, the situation gets better, but still the bounds are far worse than those one would get for univariate Lipschitz functions. (Actually, we prove a slightly better lower bound than that shown in (1); this better bound is less easy to grasp, however.)

**Tractability results.** Translating the bounds on $g_{n,d}^{\mathrm{ada}}(\mathcal{R}_d^{\alpha,p}, L_\infty)$ into bounds on the *information complexity*

$$n(\varepsilon, d) := \inf\{n \in \mathbb{N} : g_{n,d}^{\mathrm{ada}}(\mathcal{F}, L_\infty) \leq \varepsilon\},$$

we see a surprisingly diverse picture in terms of degrees of tractability. Namely, we see almost the entire spectrum of degrees of tractability as introduced in the recent monographs by Novak and Woźniakowski. To be precise, the sampling problem

(a) suffers from the curse of dimensionality if $p = 2$ and $\alpha < \infty$,

(b) never suffers from the curse of dimensionality if $p < 2$,

(c) is intractable if $p < 2$ and $\alpha \leq \frac{1}{1/p-1/2}$,

(d) is weakly tractable if $p < 2$ and $\alpha > \frac{1}{1/\max\{1,p\}-1/2}$,

(e) is quasi-polynomially tractable if $\alpha = \infty$.

Let us emphasize the case (a). Not only do we understand here that the ridge structure alone does not help to overcome the curse of dimensionality. It even tells us that neither adaptivity nor non-linearity of algorithms leads to improvements. In the remaining cases, it is less obvious what to learn from the results. Especially, for $p \leq 1$, the choice of the domain $\Omega = \bar{B}_2^d$ becomes somewhat artificial. The natural choice then would be the cube $\Omega = [-1, 1]^d$.

Actually, in a recent work [1] it has been shown that, when $\Omega = [0, 1]^d$, $\alpha > 1$, and $p \leq 1$, then the sampling problem is *polynomially tractable*, provided that the feasible ridge directions are component-wise larger than zero. The results are based on an adaptive algorithm which is basically constructed like the intuitive step-wise scheme we have sketched above. This time, it works efficiently because on $\Omega = [0, 1]^d$ it is possible to reach the whole relevant range of the profile's domain without knowing the ridge direction a priori; one just has to sample along the direction $(1, \ldots, 1)$.

We have already argued that a comparable algorithm cannot work whenever the domain is the Euclidean unit ball. But, of course, there is one remedy: make it explicit knowledge about the problem that the profiles' first derivatives are uniformly bounded away from zero in some point, say the origin. In other words, for $\alpha > 1$ and some $0 < \kappa < 1$, we add the constraint $|g'(0)| \geq \kappa$ for all feasible profiles $g$ to the class $\mathcal{R}_d^{\alpha,p}$. In this we follow [2], where the approach has been worked out in broader generality. Now, the sampling problem becomes polynomially tractable, no matter what precise values the parameters $\alpha > 1$ and $p$ take.

# References


[1] A. Cohen, I. Daubechies, R. A. DeVore, G. Kerkyacharian, and D. Picard, *Capturing ridge functions in high dimensions from point queries*, Constr. Approx. **35**(2) (2012), 225–243.

[2] M. Fornasier, K. Schnass and J. Vybí ral, *Learning functions of few arbitrary linear parameters in high dimensions*, Found. Comput. Math. **12**(2) (2012), 229–262.

[3] S. Mayer, T. Ullrich, and J. Vybíral, *Entropy and sampling numbers of classes of ridge functions*, preprint.

[4] A. Hinrichs, and S. Mayer, *Entropy numbers of spheres in Banach and quasi-Banach spaces*, work in progress.