



HAL
open science

Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials

Francis Bach, Alessandro Rudi

► **To cite this version:**

Francis Bach, Alessandro Rudi. Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials. 2022. hal-03843458v1

HAL Id: hal-03843458

<https://hal.science/hal-03843458v1>

Preprint submitted on 8 Nov 2022 (v1), last revised 16 Apr 2023 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials

Francis Bach and Alessandro Rudi
Inria, Ecole Normale Supérieure
PSL Research University
francis.bach@inria.fr, alessandro.rudi@inria.fr

November 8, 2022

Abstract

We consider the unconstrained optimization of multivariate trigonometric polynomials by the sum-of-squares hierarchy of lower bounds. We first show a convergence rate of $O(1/s^2)$ for the relaxation with degree s without any assumption on the trigonometric polynomial to minimize. Second, when the polynomial has a finite number of global minimizers with invertible Hessians at these minimizers, we show an exponential convergence rate with explicit constants. Our results also apply to the minimization of regular multivariate polynomials on the hypercube.

1 Introduction

Sum-of-squares hierarchies provide an elegant framework for global optimization for a variety of hard optimization problems. Starting from continuous polynomial optimization and combinatorial optimization problems [1, 2], they now apply to many other infinite-dimensional optimization problems such as optimal transport or optimal control (see a thorough review in [3, 4]).

Within optimization, they are most often cast as the minimization of multivariate polynomials over sets defined by essentially arbitrary polynomial constraints. They work by solving a sequence of semi-definite programming problems of increasing sizes, often referred to as a sum-of-squares (SOS) “hierarchy” of optimization problems.

The convergence rate of the minimal values of these problems towards the optimal value is empirically much faster than can actually be shown. Current theoretical results can be summarized as follows:

- In dimension one, there is no need for hierarchies, as the most direct formulations are tight [5].
- In higher dimensions, the hierarchies are always converging, due to powerful representation results of strictly positive polynomials [6, 7]. However, finite convergence can only be shown when strict second-order local optimality conditions are satisfied, but without a bound on the level at which the finite convergence is achieved [8]. In terms of asymptotic convergence rates, they are quite slow, at best $O(1/s^2)$ in the simplest situations for the relaxation with polynomials of degree s [9, 10].

In this paper, we focus on one of the simplest formulations of minimizing polynomials on $[-1, 1]^d$, which, as we show below through the use of Chebyshev polynomials, can be formulated as minimizing specific instances of trigonometric polynomials on $[0, 1]^d$, which will be our main focus, since for unconstrained optimization of trigonometric polynomials, most results simplify.

We make the following contributions:

- We provide in Section 3 an $O(1/s^2)$ convergence result for the level of the hierarchy corresponding to trigonometric polynomials of degree s , *without any assumptions*, that extends the work of [10] for polynomials on $[-1, 1]^d$, with a similar proof technique (taken from [9]), but with simpler arguments and explicit constants.
- When we add local optimality conditions similar to [8], then we prove in Section 4 an exponential convergence rate with explicit (but more complex) constants. The proof technique is taken from [11, 12] who showed convergence rates that were faster than any polynomial in s , but without explicit constants.

Our proof techniques deviate from previous work on polynomial hierarchies by a strong focus on smoothness properties of the optimization problems rather than its algebraic properties. More precisely, this allows us (1) to use square roots and matrix square roots (which will typically lead to non-polynomial functions when taken on polynomials) together with their differentiability properties, and (2) to consider all infinitely differentiable functions with a certain control of all derivatives, which trigonometric polynomials are only a sub-class of.

2 Problem set-up

Periodic functions and trigonometric polynomials. We consider 1-periodic continuous functions f on \mathbb{R}^d , which we restrict to $f : [0, 1]^d \rightarrow \mathbb{R}$, with summable Fourier series, that is, for which the “F-norm”:

$$\|f\|_F = \sum_{\omega \in \mathbb{Z}^d} |\hat{f}(\omega)|$$

is finite, where $\hat{f}(\omega) = \int_{[0,1]^d} f(x) e^{-2i\pi\omega^\top x} dx$ is the Fourier series of f . We can then represent such functions as sums of exponential $f(x) = \sum_{\omega \in \mathbb{Z}^d} \hat{f}(\omega) e^{2i\pi\omega^\top x}$, where the series is uniformly convergent.

We consider real-valued functions f , that is, such that $\hat{f}(-\omega) = \hat{f}^*(\omega)$ for all $\omega \in \mathbb{Z}^d$. This implies we can write $f(x)$ as real linear combinations of $\cos 2\pi\omega^\top x$ and $\sin 2\pi\omega^\top x$, and thus as a linear combination of monomials in $\cos 2\pi x_1, \dots, \cos 2\pi x_d, \sin 2\pi x_1, \dots, \sin 2\pi x_d$. This includes, but is not limited to, trigonometric polynomials of degree $2r$, which corresponds to functions with vanishing Fourier series coefficients $\hat{f}(\omega)$ for $\|\omega\|_\infty > 2r$, that is,

$$f(x) = \sum_{\|\omega\|_\infty \leq 2r} \hat{f}(\omega) e^{2i\pi\omega^\top x}.$$

Hierarchies of SOS optimization problems. We consider the maximization of c such that $f - c$ is a sum of squares of trigonometric polynomials of degree s . We denote by $c^*(f, s)$ the optimal value. The principle behind SOS hierarchies is that when f is a trigonometric polynomial, this optimization problem can be solved as a finite-dimensional semi-definite programming (SDP) problem (see, e.g. [13]), and thus be solved with a variety of algorithms (see, e.g., [14]).

If f is a trigonometric polynomial of degree $2r$ with $r \leq s$, then the value is finite, and we always have $c_*(f, s) \leq \inf_{x \in [0,1]^d} f(x)$. Our main goal is to provide a bound:

$$0 \leq \inf_{x \in [0,1]^d} f(x) - c_*(f, s) \leq \varepsilon(f, s), \quad (1)$$

depending on simple properties of f , and that tends to zero when s tends to $+\infty$ with an explicit dependence in s .

Beyond polynomials. When f is not a trigonometric polynomial (of sufficiently low degree), then the SDP is not feasible (and the value thus equal to $-\infty$), but as shown in [12], by using $c - \|f - c - g\|_F$ as an objective function (with g an SOS trigonometric polynomials of degree $2s$), we always get feasible problems with values less than the minimal value of f . They can then be solved with appropriate sampling schemes (see [12] for details).

2.1 Relationship with polynomial hierarchies on $[-1, 1]^d$

In this section, we show how results on trigonometric polynomials on $[0, 1]^d$ lead to results on regular polynomials on $[-1, 1]^d$.

Given a real polynomial P on \mathbb{R}^d of degree $2r$, we define the function $f : [0, 1]^d \rightarrow \mathbb{R}$ as

$$f(y) = P(\cos 2\pi y_1, \dots, \cos 2\pi y_d),$$

which is a trigonometric polynomial on $[0, 1]^d$.

If the function f is a sum of squares of trigonometric polynomials, it is the sum of terms of the form $[Q(\cos 2\pi y_1, \dots, \cos 2\pi y_d, \sin 2\pi y_1, \dots, \sin 2\pi y_d)]^2$, where Q is a regular multivariate polynomial.

We can then use the unique decomposition of multivariate trigonometric polynomials as¹

$$Q(\cos 2\pi y_1, \dots, \cos 2\pi y_d, \sin 2\pi y_1, \dots, \sin 2\pi y_d) = \sum_{J \subset \{1, \dots, d\}} Q_J(\cos 2\pi y_1, \dots, \cos 2\pi y_d) \prod_{j \in J} \sin 2\pi y_j,$$

where Q_J is a multivariate polynomial. Then, when taking the square, we get the following terms for any $J, J' \subset \{1, \dots, d\}$:

$$Q_J(\cos 2\pi y_1, \dots, \cos 2\pi y_d) Q_{J'}(\cos 2\pi y_1, \dots, \cos 2\pi y_d) \prod_{j \in J} \sin 2\pi y_j \prod_{j' \in J'} \sin 2\pi y_{j'}.$$

¹This is a simple consequence of the definitions of Chebyshev polynomials of the first and second kinds (see, e.g., [15]), that show that for $\omega \geq 1$, $\cos 2\pi\omega z$ is a polynomial in $\cos 2\pi z$, while $\sin 2\pi(\omega + 1)z$ is the product of $\sin 2\pi z$ and a polynomial in $\cos 2\pi z$.

When $J = J'$, and writing $x_1 = \cos 2\pi y_1, \dots, x_d = \cos 2\pi y_d$ for $x \in [-1, 1]^d$, we get the term

$$Q_J(x_1, \dots, x_d)^2 \prod_{j \in J} (1 - x_j^2), \quad (2)$$

while for $J \neq J'$, the sum of all terms coming from all squares must vanish because the original trigonometric polynomial f has no sine terms.

Thus, using Chebyshev polynomials, we get exactly the Schmüdgen representation [7] of polynomials on $[-1, 1]^d$, as sum of terms of the form in Eq. (2) for all subsets $J \subset \{1, \dots, d\}$. Therefore, existence of an SOS decomposition for f leads to the existence of the corresponding Schmüdgen representation for P on $[-1, 1]^d$. Thus our results also provide convergence rates for this hierarchy, and we thus actually extend results from [10].

Note that our explicit results need to express a polynomial in the basis of Chebyshev polynomials, and then we consider the ℓ_1 -norm of the associated coefficients.

2.2 Review of existing results

In this section, we briefly review results about SOS hierarchies for the particular case of unconstrained optimization of trigonometric polynomials:

- If $d = 1$, and f is a trigonometric polynomial of degree $2r$, it is well-known that $\varepsilon(f, s) = 0$ as soon as $s \geq r$, as all non-negative trigonometric polynomials are sums-of-squares [16, 17].
- When $d > 1$, any *strictly positive* trigonometric polynomial is a sum-of-squares [18, 13], but there exist non-negative polynomials which are not SOS [19]. Thus SOS hierarchies have to converge, but cannot be always finitely convergent.
- When the set of zeroes of the non-negative function f is finite and with invertible Hessians at these points, the hierarchy is finitely convergent, but with no a priori bound on the required degree [8].

The goal of this paper is to provide upper-bounds of $\varepsilon(f, s)$ in Eq. (1) for $d > 1$, first without assumptions with a rate $O(1/s^2)$ (Section 3), and then with stronger assumptions regarding the Hessian at optimum and explicit exponential rates (Section 4).

3 $O(1/s^2)$ convergence without assumptions for polynomials

We now show that the hierarchy of degree s leads to a convergence rate in $O(1/s^2)$ with explicit simple constants and few assumptions.

Theorem 1 *For any trigonometric polynomial f of degree less than $2r$, we have, for any $s \geq 3r$:*

$$\varepsilon(f, s) \leq \|f - f_*\|_{\mathbb{F}} \cdot \left[\left(1 - \frac{6r^2}{s^2} \right)^{-d} - 1 \right] \sim_{s \rightarrow +\infty} \|f - f_*\|_{\mathbb{F}} \cdot \frac{6r^2 d}{s^2}.$$

Proof We here follow the proof technique of [9, 10] based on integral operators, by adapting it to trigonometric polynomials of degree $2r$, which are easier to deal with than spherical harmonics or regular polynomials through the use of Fourier series. We consider the following integral operator on 1-periodic functions on $[0, 1]^d$ to \mathbb{R} , defined as

$$Th(x) = \int_{[0,1]^d} |q(x-y)|^2 h(y) dy, \quad (3)$$

for a well-chosen 1-periodic function q which is a trigonometric polynomial of degree s . By design, if h is a non-negative function, then Th is a sum of squares of polynomials of degree less than s . We will find h such that $Th = f - f_* + b$ for a constant $b \geq 0$, for f_* the minimal value of f , which will prove the result, since then $f = f_* - b + Th$, and $f_* - b$ is smaller than the value of the SOS relaxation $c_*(f, s)$, leading to $f_* - c_*(f, s) \leq b$.

In the Fourier domain, since convolutions lead to pointwise multiplication and vice-versa, we have for all $\omega \in \mathbb{Z}^d$:

$$\widehat{Th}(\omega) = \hat{q} * \hat{q}(\omega) \cdot \hat{h}(\omega),$$

and thus the candidate h is defined by its Fourier series, which is equal to zero for $\|\omega\|_\infty \leq 2r$, and to

$$\frac{\hat{f}(\omega) + (b - f_*)1_{\omega=0}}{\hat{q} * \hat{q}(\omega)}$$

otherwise. We then have

$$\|f - f_* + b - h\|_\infty = \left\| \sum_{\omega \in \mathbb{Z}^d} \hat{f}(\omega) \left(1 - \frac{1}{\hat{q} * \hat{q}(\omega)}\right) \right\|_\infty.$$

If we impose that $\hat{q} * \hat{q}(\omega) = 1$, then we get: $\|f - f_* + b - h\|_\infty = \left\| \sum_{\omega \neq 0} \hat{f}(\omega) \left(1 - \frac{1}{\hat{q} * \hat{q}(\omega)}\right) \right\|_\infty$.

Using that $\|f - f_*\|_\infty \leq \|f - f_*\|_F = \sum_{\omega \neq 0} |\hat{f}(\omega)|$, we get:

$$\|f - f_* + b - h\|_\infty \leq \|f - f_*\|_F \cdot \max_{\|\omega\|_\infty \leq 2r} \left| \frac{1}{\hat{q} * \hat{q}(\omega)} - 1 \right|.$$

The goal is now to find a good function $q : [0, 1]^d \rightarrow \mathbb{R}$ with Fourier support within the ball of radius s , so that $\hat{q} * \hat{q}(\omega)$ is close to 1 for $\|\omega\|_\infty \leq 2r$, and simply check when $\|f - f_* + b - h\|_\infty \leq b$.

A simple candidate is $\hat{q}(\omega) = \frac{1}{(2s+1)^{d/2}} 1_{\|\omega\|_\infty \leq s}$; we can then compute the convolution and obtain that $\hat{q} * \hat{q}(\omega) = \prod_{i=1}^d \left(1 - \frac{|\omega_i|}{2s+1}\right)_+ \geq \left(1 - \frac{2r}{2s+1}\right)^d$, leading to $b = \|f - f_*\|_F \cdot \left[\left(1 - \frac{2r}{2s+1}\right)^{-d} - 1\right]$. When s goes to infinity, we have the equivalent $b \sim \|f - f_*\|_F \cdot \frac{rd}{s} = O(1/s)$, which thus converges to zero, but at a slow rate.

A better candidate leads to a rate in $O(1/s^2)$ (like in [9, 10]), and is

$$\hat{q}(\omega) = a \prod_{i=1}^d \left(1 - \frac{|\omega_i|}{s}\right)_+,$$

with a a normalizing constant. A tedious computation including sums of powers of consecutive integers leads to, for any $\|\omega\|_\infty \leq s$ (note that $\hat{q} * \hat{q}(\omega)$ is only equal to zero for $\|\omega\|_\infty > 2s$),

$$\hat{q} * \hat{q}(\omega) = a^2 \prod_{i=1}^d \left[\frac{2s}{3} + \frac{1}{s} - \frac{\omega^2}{s} + \frac{|\omega|}{2s^2}(\omega^2 - 1) \right].$$

Thus we need $a^2 = \frac{1}{(\frac{2s}{3} + \frac{1}{s})^d}$ to get $\hat{q} * \hat{q}(0) = 1$ and thus

$$\hat{q} * \hat{q}(\omega) \geq \prod_{i=1}^d \left(1 - \frac{1}{\frac{2s}{3} + \frac{1}{s}} \frac{\omega_i^2}{s} \right)_+ \geq \prod_{i=1}^d \left(1 - \frac{3\omega_i^2}{2s^2} \right)_+,$$

which is greater than $(1 - \frac{6r^2}{s^2})_+^d$, when in addition $\|\omega\|_\infty \leq 2r$. This leads to, for $s \geq 3r \geq \sqrt{6}r$,

$$b \leq \|f - f_*\|_F \cdot \left[\left(1 - \frac{6r^2}{s^2} \right)^{-d} - 1 \right] \sim \|f - f_*\|_F \cdot \frac{6r^2 d}{s^2}.$$

■

We can make the following observations:

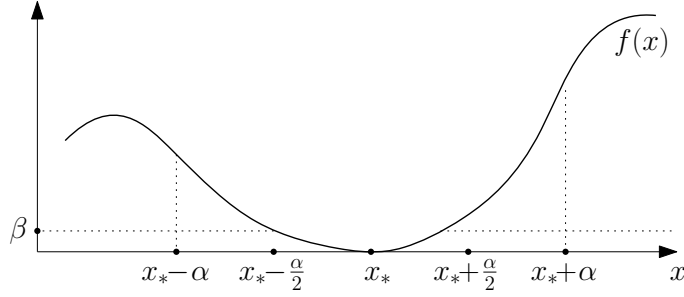
- The proposed bound follows a series of earlier bounds with a similar behavior in $O(1/s^2)$ for the convergence rate of Lasserre's SOS hierarchies, and uses the same proof technique based on integral operators [9, 10, 20]. The most closely related is the one of [10], which considers regular polynomials on $[-1, 1]^d$ with Schmüdgen's representation, but with a different choice for the function q in Eq. (3). As shown in Section 2.1, our bound applies to this case as well through a change of variable; it differs in the choice of normalization of coefficients (for us, ℓ_1 -norm of the expansion in Chebyshev polynomials), but leads to explicit constants on problem dimensions that scale better.
- Note that we could extend this result to other types of regularity beyond finite support and bounded F-norm, with the asymptotic bound $\|f - f_*\|_F \cdot \frac{6r^2 d}{s^2} + \sum_{\|\omega\|_\infty > 2r} |\hat{f}(\omega)|$, and by optimizing over $r \leq s$.
- We believe the proof technique based on integral operators cannot lead to a better rate than $O(1/s^2)$, with the following informal argument. In order to obtain a faster rate in the simplest one-dimensional case, the function $r : [0, 1] \rightarrow \mathbb{R}$ defined as $r(x) = |q(x)|^2$, should be so that its Fourier series $\hat{r}(\omega)$ is of the form $f(\omega/s)$ for a function $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $f''(0) = 0$ and with support in $[-2, 2]$. Thus, when s gets large, $r(x) = \sum_{|\omega| \leq 2s} f(\omega/s) e^{2i\pi\omega x}$ should be proportional to the Fourier transform of f . Thus the Fourier transform of f should be non-negative with $f''(0) \propto \int_{\mathbb{R}} x^2 \hat{f}(x)^2 dx = 0$, which is impossible.
- A natural open question is the optimality of the “assumption-free” bound in $O(1/s^2)$ (regardless of the proof technique). We show in the next section that adding extra assumptions lead to significantly better rates.

4 Exponential convergence with local optimality conditions

We consider the simplest situation where the minimum of f is attained at a unique point x_* on the torus, and we assume that the Hessian $f''(x_*)$ is invertible. This implies that there exist “conditioning” constants $\alpha \in [0, 1/2)$, $\beta > 0$, and $\lambda > 0$ such that:

$$\|x - x_*\|_\infty \leq \alpha \Rightarrow f''(x) \succcurlyeq \lambda I \quad \text{and} \quad \|x - x_*\|_\infty \geq \frac{\alpha}{2} \Rightarrow f(x) - f(x_*) \geq \beta,$$

that is, (a) in the ℓ_∞ -ball of radius α around x_* , the Hessian of f has strictly positive eigenvalues greater than λ (which we can take to be $\frac{1}{2}\lambda_{\min}(f''(x_*))$), and hence f is strictly convex, and (b) away from a slightly smaller ball, $f - f(x_*)$ is strictly positive and greater than $\beta > 0$. See illustration below.



The proof technique is based on the one introduced in Lemma 1 and Theorem 2 of [11] (for the non-periodic case and without explicit constants) and extends directly to situations where the global minimum is attained at finitely many points with the same local Hessian condition (see also [21] for cases where minimizers are whole manifolds).

Note that in that regime, the hierarchy is known to be finitely convergent [8], but without bounds on the required degree s . The following theorem gives an explicit bound on the convergence rate for any infinitely differentiable function with a specific growth condition for derivatives:²

Theorem 2 *Assume that $f : [0, 1]^d \rightarrow \mathbb{R}$ is infinitely differentiable and such that $|\nabla^m f(x)[\delta, \dots, \delta]| \leq \|f - f_*\|_{\mathbb{F}} (4\pi r)^m \|\delta\|_1^m$ for all $m \geq 0$, $x \in [0, 1]^d$ and $\delta \in \mathbb{R}^d$. Assume there exist $x_* \in [0, 1]^d$, as well as, $\alpha \in [0, 1/2)$, $\beta > 0$, and $\lambda > 0$ such that:*

$$\|x - x_*\|_\infty \leq \alpha \Rightarrow f''(x) \succcurlyeq \lambda I \quad \text{and} \quad \|x - x_*\|_\infty \geq \frac{\alpha}{2} \Rightarrow f(x) - f(x_*) \geq \beta.$$

Then, we have:

$$\varepsilon(r, f) \leq \Delta_1 \exp\left(-\left(\frac{s}{\Delta_2}\right)^{1+\xi}\right),$$

for any $\xi \in (0, 1/2]$, with

$$\Delta_1 = (\beta + \lambda d^3)(20B^2 d^5)^{d+1} \quad \text{and} \quad \Delta_2 = d \max\left\{\frac{275}{\alpha \xi}, \frac{8\pi r \|f - f_*\|_{\mathbb{F}}}{\beta}, \frac{6}{\lambda} \|f - f_*\|_{\mathbb{F}} (4\pi r)^3\right\}. \quad (4)$$

Before describing the proof, we can make a few simple observations:

²We denote by $\nabla^m f(x)$ the symmetric m -th order tensor of m -th order derivatives.

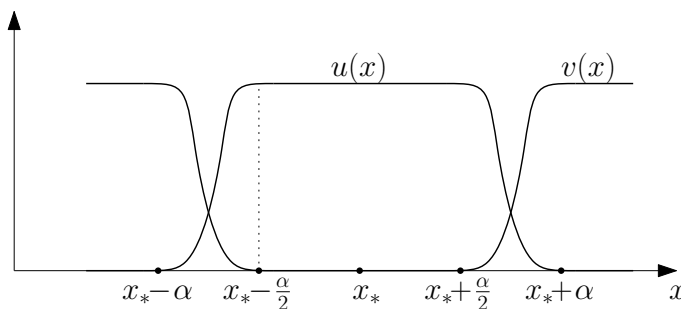
- Trigonometric polynomials of degree $2r$ satisfy the required growth condition.
- The result extends a prior result [12], that was showing convergence rates faster than any power of s , but without explicit constants, which are needed to obtain the exponential rate.
- We could easily consider weaker growth conditions for the m -th order derivatives.
- We could optimize over $\xi \in [0, 1/2)$ to get a better dependence in s .

4.1 Proof technique

We consider two infinitely differentiable 1-periodic functions $u, v : \mathbb{R}^d \rightarrow [0, 1]$ such that

$$\|x - x_*\|_\infty \leq \frac{\alpha}{2} \Rightarrow u(x) = 1 \quad \text{and} \quad \|x - x_*\|_\infty \geq \alpha \Rightarrow u(x) = 0,$$

and for all $x \in \mathbb{R}^d$, $u(x)^2 + v(x)^2 = 1$. See illustration below.



These are usually referred to as partitions of the unity and will be built in Appendix A.2 using standard tools. Following [11], we can then decompose f as, using Taylor's formula with integral remainder:

$$\begin{aligned} f(x) - f(x_*) &= v(x)^2[f(x) - f(x_*)] + u(x)^2[f(x) - f(x_*)] \\ &= \left[v(x)\sqrt{f(x) - f(x_*)} \right]^2 + \left[u(x)^2 \int_0^1 (1-t)(x - x_*)^\top f''(x_* + t(x - x_*))(x - x_*) dt \right] \\ &= \left[v(x)\sqrt{f(x) - f(x_*)} \right]^2 + \left[u(x)^2 (x - x_*)^\top R(x)(x - x_*) \right] \\ &= \left[v(x)\sqrt{f(x) - f(x_*)} \right]^2 + \left[u(x)^2 \sum_{i=1}^d (x - x_*)^\top R(x)^{1/2} u_i u_i^\top R(x)^{1/2} (x - x_*) \right] \\ &= \left[v(x)\sqrt{f(x) - f(x_*)} \right]^2 + \sum_{i=1}^d \left[u(x)(x - x_*)^\top R(x)^{1/2} u_i \right]^2, \end{aligned}$$

with $R(x) = \int_0^1 (1-t)f''(x_* + t(x - x_*))dt \succcurlyeq \frac{\lambda}{2}$ if $\|x - x_*\|_\infty \leq \alpha$, and $(u_1, \dots, u_d) \in \mathbb{R}^{d \times d}$ any orthonormal basis of \mathbb{R}^d .

We thus get an explicit SOS decomposition with $d + 1$ functions as

$$f(x) - f(x_*) = \sum_{i=1}^{d+1} g_i(x)^2,$$

with

$$\begin{aligned} g_i(x) &= u(x)(x - x_*)^\top R(x)^{1/2} u_i \text{ for } i \in \{1, \dots, d\}, \\ g_{d+1}(x) &= v(x) \sqrt{f(x) - f(x_*)}, \end{aligned}$$

which are infinitely differentiable functions (just taking the square root of $f - f_*$ without taking care of the region around the minimizer like we do above would not lead to a differentiable function).

We consider the truncations \bar{g}_i obtained by keeping in g_i only frequencies such that $\|\omega\|_\infty \leq s$, leading to, using lemmas from [12] about the F-norm (see also [22, Section I.6]):

$$\begin{aligned} \left\| \sum_{i=1}^{d+1} g_i^2 - \sum_{i=1}^{d+1} \bar{g}_i^2 \right\|_{\mathbb{F}} &\leq \sum_{i=1}^{d+1} (\|g_i\|_{\mathbb{F}} + \|\bar{g}_i\|_{\mathbb{F}}) \|g_i - \bar{g}_i\|_{\mathbb{F}} \\ &\leq 2 \sum_{i=1}^{d+1} \|g_i\|_{\mathbb{F}} \sum_{\|\omega\| > s} |\hat{g}_i(\omega)|. \end{aligned}$$

Denoting $\|f\|_{\mathbb{F},s} = \sum_{\|\omega\|_\infty > s} |\hat{f}(\omega)|$, we need to find bounds on $\|g_i\|_{\mathbb{F}}$ and $\|g_i\|_{\mathbb{F},s}$, for $i \in \{1, \dots, d+1\}$.

Since these functions are C^∞ (i.e., infinitely differentiable), the decay in s is faster than any power, as already noted in [12]. In the present paper, we provide explicit constants that allow to obtain an exponential convergence rate.

4.2 Precise bound

Using the lemmas from Appendix A, as described in Appendix B, we get explicit bounds on all derivatives:

$$\begin{aligned} \nabla^m g_{d+1}[\delta, \dots, \delta] &\leq 3\beta^{1/2} \max \left\{ \frac{275}{\alpha\eta}, \frac{8\pi r \|f - f_*\|_{\mathbb{F}}}{\beta} \right\}^m \|\delta\|_1^m m! \cdot m^{\eta m} \\ \nabla^m g_i[\delta, \dots, \delta] &\leq d\sqrt{\lambda} \max \left\{ \frac{275}{\alpha\eta}, \frac{6}{\lambda} \|f - f_*\|_{\mathbb{F}} (4\pi r)^3 \right\}^m \|\delta\|_1^m m! \cdot m^{\eta m}, \end{aligned}$$

for $i \in \{1, \dots, d\}$. Thus, with $B = \max \left\{ \frac{275}{\alpha\eta}, \frac{8\pi r \|f - f_*\|_{\mathbb{F}}}{\beta}, \frac{6}{\lambda} \|f - f_*\|_{\mathbb{F}} (4\pi r)^3 \right\} \geq 275$, we get from Lemma 1, for all $k \geq d+1$:

$$\begin{aligned} \|g_{d+1}\|_{\mathbb{F}} &\leq 3\beta^{1/2} \left(2 + \frac{Bd(d+1)}{2\pi} \right)^{d+1} (d+1)^{\eta(d+1)} \cdot 2(2e)^{d-2} \\ \|g_{d+1}\|_{\mathbb{F},s} &\leq 3\beta^{1/2} \left(2 + \frac{dBk}{2\pi} \right)^k k^{\eta k} \cdot 2(2e)^{d-2} s^{d-k} \\ \|g_i\|_{\mathbb{F}} &\leq d\sqrt{\lambda} \left(2 + \frac{Bd(d+1)}{2\pi} \right)^{d+1} (d+1)^{\eta(d+1)} \cdot 2(2e)^{d-2} \\ \|g_i\|_{\mathbb{F},s} &\leq d\sqrt{\lambda} \left(2 + \frac{dBk}{2\pi} \right)^k k^{\eta k} \cdot 2(2e)^{d-2} s^{d-k}, \end{aligned}$$

and thus a bound

$$\begin{aligned} c &\leq (36\beta + 4\lambda d^3) \left(2 + \frac{Bd(d+1)}{2\pi} \right)^{d+1} \left(2 + \frac{dBk}{2\pi} \right)^k (2e)^{2d-4} s^{d-k} k^{\eta k} \\ &\leq (\beta + \lambda d^3) (5Bd^2)^{d+1} \left(\frac{dBk}{6} \right)^k s^{d-k} k^{\eta k}. \end{aligned}$$

The main term is of the form $\left(\frac{k^{1+\eta}dB}{6s}\right)^k$. We then select $k = \left(\frac{6s}{edB}\right)^{1/(1+\eta)}$, leading to the term

$$\exp\left(-\left(\frac{6s}{edB}\right)^{1/(1+\eta)}\right) \leq \exp\left(-\left(\frac{2s}{dB}\right)^{1/(1+\eta)}\right).$$

Overall, we get, using the identity $u^d e^{-\alpha u} \leq \left(\frac{d}{e\alpha}\right)^d$:

$$\begin{aligned} c &\leq (\beta + \lambda d^3)(5Bd^2)^{d+1} s^d \exp\left(-\left(\frac{2s}{dB}\right)^{1/(1+\eta)}\right) \\ &\leq (\beta + \lambda d^3)(5Bd^2)^{d+1} \left(d(1+\eta)(dB)^{1/(1+\eta)}\right)^{d(1+\eta)} \exp\left(-\left(\frac{s}{dB}\right)^{1/(1+\eta)}\right) \\ &\leq (\beta + \lambda d^3)(5Bd^2)^{d+1} \left(2d(dB)^{1/2}\right)^{2d} \exp\left(-\left(\frac{s}{dB}\right)^{1/(1+\eta)}\right) \\ &\leq (\beta + \lambda d^3)(20B^2d^5)^{d+1} \exp\left(-\left(\frac{s}{dB}\right)^{1/(1+\eta)}\right). \end{aligned}$$

We then consider $\xi = 1 - \frac{1}{1+\eta}$ to obtain the constants in Eq. (4).

5 Discussion

Our convergence results could be extended in a number of interesting ways:

- While convergence rates in $O(1/s^2)$ already exist for the Boolean hypercube [20], it would be interesting to obtain improved rates with some form of local condition.
- Our proof technique relies on Fourier series and can be extended to all cases where such tools can be used, such as on the Euclidean hypersphere [9] and beyond [23].
- Almost all the techniques that we used to derive explicit constants can be extended easily to the more general kernel case [11] (noting that the function q that we used is a specific instance of a translation-invariant periodic kernel), as well as the case where minimizers are manifolds [21].
- It would be interesting to extend our second result to provide an explicit bound on the degree for finite convergence.
- We only focused on the unconstrained global optimization problem, but adding constraints and extending to more general problems (e.g., optimal control and optimal transport) is natural.

Acknowledgements

We thank Monique Laurent for helpful discussions related to this work. We acknowledge support from the French government under the management of the Agence Nationale de la Recherche as part of the ‘‘Investissements d’avenir’’ program, reference ANR-19-P3IA0001 (PRAIRIE 3IA Institute). This work was also supported by the European Research Council (grants SEQUOIA 724063 and REAL 947908).

A Generic lemmas about derivatives

In this appendix, we state and prove a series of lemmas about derivatives, Fourier decays, square roots, and partitions of unity.

A.1 From derivatives to Fourier decay

Lemma 1 (From derivatives to Fourier decay) *Assume that $g : [0, 1]^d \rightarrow \mathbb{R}$ is C^∞ and such that for all $\delta \in \mathbb{R}^d$ and $m \geq 0$,*

$$\nabla^m g(x)[\delta, \dots, \delta] \leq C \cdot B^m \cdot \|\delta\|_1^m \cdot m! \cdot \kappa(m),$$

with κ non-decreasing. Then, for $k \geq d + 1$,

$$\begin{aligned} \|g\|_{\mathbb{F}} &\leq C \left(2 + \frac{dBk}{2\pi}\right)^k \kappa(k) \cdot 2(2e)^{d-2} \\ \|g\|_{\mathbb{F},s} &\leq C \left(2 + \frac{dBk}{2\pi}\right)^k \kappa(k) 2(2e)^{d-2} (s+1)^{d-k}. \end{aligned}$$

Proof We will consider bounds on a function g of the form

$$|\hat{g}(\omega)| \leq D(k) \frac{1}{(2 + \|\omega\|_1)^k}, \quad (5)$$

for a constant $D(k)$ to be determined, since it implies, for $k \geq d + 1$:

$$\sum_{\|\omega\|_\infty \geq s} |\hat{g}(\omega)| \leq D(k) \sum_{\omega \in \mathbb{Z}^d} \frac{1}{(2 + \|\omega\|_1)^k} = D(k) \sum_{t=s}^{\infty} \frac{1}{(2+t)^k} \binom{d+t-1}{d-1},$$

by counting the number of $\omega \in \mathbb{Z}^d$ such that $\|\omega\|_1 = t$. This then leads to the desired results (in particular by taking $s = 0$).

We first start by a simple upper bound on $\binom{d+t-1}{d-1}$, as (using the identity $n^n \leq n!e^{n-1}$ applied to $n = d-1$):

$$\begin{aligned} \binom{d+t-1}{d-1} &= \frac{1}{(d-1)!} (t+1) \cdots (t+d-1) \leq \frac{(t+d-1)^{d-1}}{(d-1)!} \\ &\leq 2^{d-2} \frac{t^{d-1} + (d-1)^{d-1}}{(d-1)!} \leq \frac{2^{d-2}}{(d-1)!} t^{d-1} + (2e)^{d-2}. \end{aligned}$$

using the bound $n! \geq \frac{n^n}{e^{n-1}}$ for any integer n . This leads to:

$$\begin{aligned} \sum_{\omega \in \mathbb{Z}^d} |\hat{g}(\omega)| &\leq D(k) \sum_{t=0}^{\infty} \frac{1}{(2+t)^k} \left(\frac{2^{d-2}}{(d-1)!} t^{d-1} + (2e)^{d-2} \right) \\ &\leq D(k) \left[\frac{2^{d-2}}{(d-1)!} \frac{1}{k-d} + (2e)^{d-2} \frac{1}{k-1} \right] \leq D(k) \left[\frac{2^{d-2}}{(d-1)!} + \frac{(2e)^{d-2}}{d} \right] \leq 2(2e)^{d-2} D(k). \\ \sum_{\|\omega\|_\infty \geq s} |\hat{g}(\omega)| &\leq D(k) \sum_{t=s}^{\infty} \frac{1}{(2+t)^k} \left(\frac{2^{d-2}}{(d-1)!} t^{d-1} + (2e)^{d-2} \right) \\ &\leq D(k) \left[\frac{2^{d-2}}{(d-1)!} \frac{1}{(s+1)^{k-d}} + \frac{(2e)^{d-2}}{d} \frac{1}{(s+1)^{k-1}} \right] \leq 2(2e)^{d-2} (s+1)^{d-k} D(k). \end{aligned}$$

To obtain Eq. (5), we need to be able to bound $|\hat{g}(\omega)| |\omega_j|^{\alpha_1} \dots |\omega_d|^{\alpha_d}$ for any $\alpha_1 + \dots + \alpha_d = k$, which we obtain from bounds on $\nabla^k g(x)[\delta, \dots, \delta]$ for all δ and k , where $\nabla^k g(x)$ is the k -th order tensor of partial derivatives. Indeed, if $|\nabla^k g(x)[\delta, \dots, \delta]| \leq E(k) \|\delta\|_1^k$ for all $\delta \in \mathbb{R}^d$, then we have from Lemma 3:

$$|\nabla^k g(x)[\delta_1, \dots, \delta_k]| \leq \frac{1}{k!} E(k) \left(\sum_{i=1}^k \|\delta_i\|_1 \right)^k,$$

which we will use with δ_j being canonical basis vectors and for which we have: $\sum_{i=1}^k \|\delta_i\|_1 = \|\alpha\|_1 = k$, which overall leads to, for any $\alpha \in \mathbb{N}^d$ such that $\|\alpha\|_1 = k$:

$$|\partial_\alpha g(x)| \leq |\nabla^k g(x)[\delta_1, \dots, \delta_k]| \leq \frac{1}{k!} E(k) k^k,$$

with $E(k) = C \cdot B^k \cdot k! \cdot \kappa(k)$ (by assumption).

Then, by expanding $(2 + \|\omega\|_1)^k$ with the multinomial formula, and using $\hat{g}(\omega) \prod_{i=1}^d |2\pi\omega_i|^{\alpha_i} \leq \sup_{x \in [0,1]^d} |\partial_\alpha g(x)|$, we get:

$$\begin{aligned} |\hat{g}(\omega)| \sum_{\|\alpha\|_1=k} \frac{k!}{\alpha_0! \alpha_1! \dots \alpha_d!} 2^{\alpha_0} |\omega_j|^{\alpha_1} \dots |\omega_d|^{\alpha_d} &\leq \sum_{\|\alpha\|_1=k} \frac{k!}{\alpha_0! \alpha_1! \dots \alpha_d!} 2^{\alpha_0} C \left(\frac{B}{2\pi} \right)^{k-\alpha_0} k^{k-\alpha_0} \kappa(k) \\ &\leq C \left(2 + \frac{dBk}{2\pi} \right)^k \kappa(k). \end{aligned}$$

This leads to $D(k) \leq C \left(2 + \frac{dBk}{2\pi} \right)^k \kappa(k)$, and thus the desired result. ■

Lemma 2 (Derivatives of products) *Assume that $g_1, g_2 : [0, 1]^d \rightarrow \mathbb{R}$ is C^∞ and such that for all $\delta \in \mathbb{R}^d$, $\nabla^m g_1(x)[\delta, \dots, \delta] \leq C_1 \cdot B_1^m \cdot \|\delta\|_1^m \cdot m! \cdot \kappa_1(m)$, and $\nabla^m g_2(x)[\delta, \dots, \delta] \leq C_2 \cdot B_2^m \cdot \|\delta\|_1^m \cdot m! \cdot \kappa_2(m)$, Then*

$$\nabla^m (g_1 g_2)(x)[\delta, \dots, \delta] \leq C_1 C_2 \kappa_1(m) \kappa_2(m) \|\delta\|_1^m (m+1)! \max\{B_1, B_2\}^m.$$

Proof Using Leibniz formula applied to $\varphi_1(t) = g_1(x + t\delta)$, $\varphi_2(t) = g_2(x + t\delta)$, we have:

$$\begin{aligned} (\varphi_1 \varphi_2)^{(m)}(0) &= \sum_{i=0}^m \binom{m}{i} \varphi_1^{(i)}(0) \varphi_2^{(m-i)}(0) \\ &\leq C_1 C_2 \|\delta\|_1^m \sum_{i=0}^m \binom{m}{i} B_1^i B_2^{m-i} i! (m-i)! \kappa_1(i) \kappa_2(m-i) \\ &\leq C_1 C_2 \kappa_1(m) \kappa_2(m) \|\delta\|_1^m m! \sum_{i=0}^m B_1^i B_2^{m-i} \leq C_1 C_2 \kappa_1(m) \kappa_2(m) \|\delta\|_1^m (m+1)! \max\{B_1, B_2\}^m. \end{aligned}$$
■

Lemma 3 (Polarization) *Let $u : E^m \rightarrow \mathbb{R}$ be a symmetric m -multi-linear form on some normed vector space E . Then for all $z_1, \dots, z_m \in E$, we have:*

$$|u[z_1, \dots, z_m]| \leq \frac{1}{m!} E(m) \left(\sum_{i=1}^m \|z_i\|_1 \right)^m \cdot \sup_{\|z\|_1 \leq 1} u(z, \dots, z).$$

Proof We use the polarization identity for the m -multilinear form $u : E^m \rightarrow E$ and its diagonal $\tilde{u} : z \mapsto u(z, \dots, z)$, see [24, Eq. (A.4)],

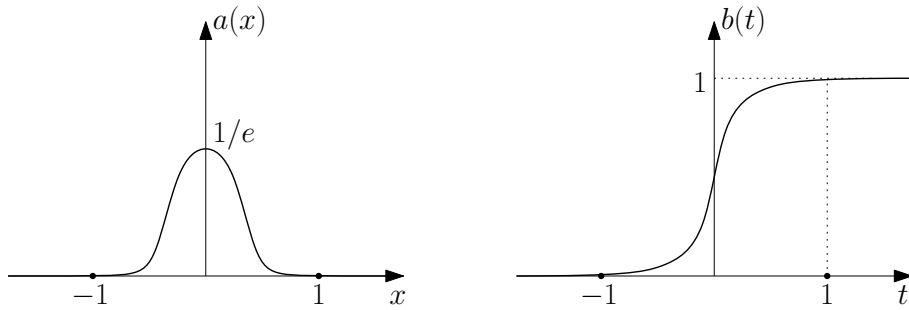
$$u(z_1, \dots, z_m) = \frac{1}{2^m m!} \sum_{\varepsilon \in \{0,1\}^m} (-1)^{\|\varepsilon\|_1} \tilde{u} \left(\sum_{i=1}^m (-1)^{\varepsilon_i} z_i \right),$$

which leads to

$$|u(z_1, \dots, z_m)| \leq \frac{1}{2^m m!} \sum_{\varepsilon \in \{0,1\}^m} \left(\sum_{i=1}^m \|z_i\|_1 \right)^m \sup_{\|z\|_1 \leq 1} |\tilde{u}(z)| = \frac{1}{m!} \left(\sum_{i=1}^m \|z_i\|_1 \right)^m \sup_{\|z\|_1 \leq 1} |\tilde{u}(z)|. \quad \blacksquare$$

A.2 Partitions of unity

Following [25, Section 3.1], we consider for $\eta \in (0, 1]$, the function $a : \mathbb{R} \rightarrow \mathbb{R}$ defined as $a(x) = \exp(-(1 - x^2)^{-1/\eta})$ on $[-1, 1]$, and zero otherwise. We then consider the function $b : \mathbb{R} \rightarrow \mathbb{R}$, defined as $b(t) = \frac{\int_{-\infty}^t a(x) dx}{\int_{-\infty}^{+\infty} a(x) dx}$, which is non-decreasing, equal to zero for $t \leq -1$, and equal to 1 if $t \geq 1$. These two functions are infinitely differentiable on \mathbb{R} . See illustrations below.



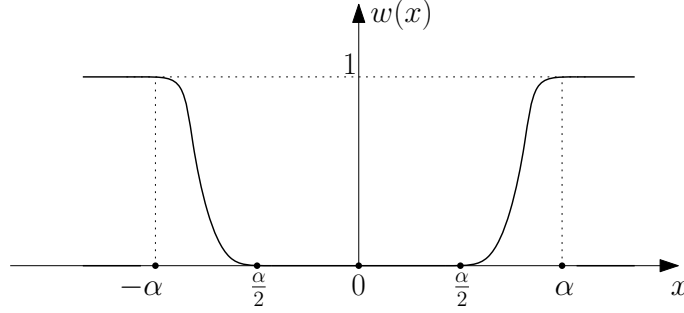
We have, from [25, Section 3.1], $|a^{(m)}(x)| \leq \left(\frac{16}{\eta}\right) m^{(1+\eta)m}$, for any $m \geq 0$, and any $x \in [-1, 1]$. Moreover, we have

$$\begin{aligned} \int_{-\infty}^{+\infty} a(x) dx &\geq 2 \int_0^{\sqrt{\eta/2}} \exp(-(1 - x^2)^{-1/\eta}) dx \\ &\geq \sqrt{2\eta} \exp(-(1 - \eta/2)^{-1/\eta}) = \sqrt{2\eta} \exp\left(-\exp\left(-\frac{1}{\eta} \log\left(1 - \frac{\eta}{2}\right)\right)\right). \end{aligned}$$

Using $\log(1-x) \geq -(2 \log 2)x$ for $x \in [0, 1/2]$, we get the lower bound³

$$\int_{-\infty}^{+\infty} a(x) dx \geq \sqrt{2\eta} \exp(-\exp(\log 2)) = \sqrt{2}e^{-2}\sqrt{\eta} \geq \sqrt{\eta}/8.$$

We consider the function w defined on $[0, 1]$ as $w(x) = b[\frac{4}{\alpha}(|x| - \frac{3\alpha}{4})]$, and extended by 1-periodicity to \mathbb{R} . It is of the form plotted below



Moreover we have:

$$\forall x \in [0, 1], |b^{(m+1)}(x)| \leq 8\sqrt{1/\eta} \left(\frac{64}{\alpha\eta}\right)^m m^{(1+\eta)m},$$

which leads to for $m > 0$

$$\forall x \in [0, 1], |b^{(m)}(x)| \leq 8\sqrt{1/\eta} \frac{\alpha\eta}{64} \left(\frac{64}{\alpha\eta}\right)^m m^{(1+\eta)m} \leq c^m m^{(1+\eta)m},$$

with $c = \frac{64}{\alpha\eta}$, an equality which is also valid for $m = 0$ (where we only know that $|b(x)| \leq 1$).

We then consider the functions

$$u(x) = \sin \left[\frac{\pi}{2} \prod_{i=1}^d (1 - w(x_i - (x_*)_i)) \right] \quad (6)$$

$$v(x) = \cos \left[\frac{\pi}{2} \prod_{i=1}^d (1 - w(x_i - (x_*)_i)) \right]. \quad (7)$$

These functions satisfy exactly the constraints from Section 4.1, that is, $u(x)^2 + v(x)^2 = 1$ for all x , and, as soon as $\|x - x_*\|_\infty \leq \alpha/2$, $u(x) = \sin(\pi/2) = 1$, as well as, when $\|x - x_*\|_\infty \geq \alpha$, $u(x) = 0$. The next lemma provides bounds on their derivatives.

Lemma 4 *For the functions u defined in Eq. (6) and Eq. (7), we have for any $\delta \in \mathbb{R}^d$ and $m > 0$:*

$$|\nabla^m u[\delta, \dots, \delta]| \leq \left(\frac{275}{\alpha\eta} \|\delta\|_1\right)^m m! \cdot m^{\eta m}, \quad (8)$$

with the same bound for v in Eq. (7).

³Note that the bound from from [25] is incorrectly independent of η .

Proof We consider the function $g(t) = u(x + \delta t) = \sin \left[\frac{\pi}{2} f(t) \right]$. We can expand the derivatives of the product function f using Leibniz formula to get for all t :

$$\begin{aligned} |f^{(m)}(t)| &\leq \frac{\pi}{2} \left| \sum_{\alpha_1 + \dots + \alpha_d = m} \binom{m}{\alpha_1, \dots, \alpha_d} \prod_{i=1}^d c^{\alpha_i} \alpha_i^{(1+\eta)\alpha_i} \delta_i^{\alpha_i} \right| \\ &\leq \frac{\pi}{2} \sum_{\alpha_1 + \dots + \alpha_d = m} \binom{m}{\alpha_1, \dots, \alpha_d} \prod_{i=1}^d [cm^{1+\eta} |\delta_i|]^{\alpha_i} = \frac{\pi}{2} [cm^{1+\eta} \|\delta\|_1]^m \leq \frac{\pi}{2} \frac{1}{e} [cem^\eta \|\delta\|_1]^m m! , \end{aligned}$$

using $m^m \leq m!e^{m-1}$.

We have, using Faà di Bruno's formula (see, e.g., [26]) for the sine function,

$$\begin{aligned} |g^{(m)}(x)| &\leq \sum_{k=1}^m B_{m,k} \left(\frac{\pi}{2e} [c1^\eta \|\delta\|_1]^1 1!, \dots, \frac{\pi}{2e} [c(m-k+1)^\eta \|\delta\|_1]^{m-k+1} (m-k+1)! \right) \\ &\leq \sum_{k=1}^m B_{m,k} \left(\frac{\pi}{2e} [cm^\eta \|\delta\|_1]^1 1!, \dots, \frac{\pi}{2e} [cm^\eta \|\delta\|_1]^{m-k+1} (m-k+1)! \right) \end{aligned}$$

using the fact that Bell polynomials have non-negative coefficients (and are thus non-decreasing functions). Thus, using that $B_{m,k}(\alpha\beta z_1, \dots, \alpha\beta^{m-k+1} z_{m-k+1}) = \alpha^k \beta^m B_{m,k}(z_1, \dots, z_{m-k+1})$, we get:

$$\begin{aligned} |g^{(m)}(x)| &\leq \sum_{k=1}^m \left(\frac{\pi}{2e} \right)^k [cm^\eta \|\delta\|_1]^m B_{m,k}(1!, \dots, (m-k+1)!) \\ &= [cm^\eta \|\delta\|_1]^m \sum_{k=1}^m \left(\frac{\pi}{2e} \right)^k \frac{(m-1)!}{(k-1)!} \binom{m}{k} \text{ using an explicit formula for Bell polynomials,} \\ &\leq [cm^\eta \|\delta\|_1]^m \sum_{k=1}^m \left(\frac{\pi}{2e} \right)^k m! \binom{m}{k} \\ &= [cm^\eta \|\delta\|_1]^m m! (1 + \pi/2e)^m \leq \left[\frac{64e(1 + \pi/(2e))}{\alpha\eta} m^\eta \|\delta\|_1 \right]^m m! \leq \left[\frac{275}{\alpha\eta} m^\eta \|\delta\|_1 \right]^m m! . \end{aligned}$$

■

A.3 Square root of a lower-bounded function

Since our SOS decomposition relies on a square root for the function g_{d+1} , we need the following lemma.

Lemma 5 *We consider a C^∞ function g defined on a neighborhood of zero (on the real line) such that $g(0) \geq c > 0$ and such that for all $m \in \mathbb{N}$,*

$$|g^{(m)}(0)| \leq C \cdot B^m$$

with $C \geq c$. For $h(x) = \sqrt{g(x)}$, we have:

$$|h^{(k)}(0)| \leq 3c^{1/2} \left(2B \frac{C}{c} \right)^k k! .$$

Proof We use the Faà di Bruno's formula (see, e.g., [26]) to get, with the k -th derivative of \sqrt{y} being $y^{\frac{1}{2}-k}(-1)^{k-1}\frac{1}{2^{k-1}}\frac{(2k)!}{(k)!2^{2k}} = y^{\frac{1}{2}-k}b_k = y^{\frac{1}{2}-k}k!C_{k-1}2^{1-2k}$ for $k > 0$, where $C_k = \frac{1}{k+1}\binom{2k}{k}$ is the Catalan number. Using the classical bound $C_n = \frac{1}{n+1}\frac{(2n)!}{(n!)^2} \leq 2 \cdot 4^n$, we get $|b_k| \leq k!$.

Faà di Bruno's formula leads to, with the Bell polynomials $B_{k,i}$, and Stirling numbers of the second kind $s(k, i)$:

$$\begin{aligned}
h^{(k)}(0) &= \sum_{i=1}^k g(0)^{\frac{1}{2}-i} b_i B_{k,i}(g'(0), \dots, g^{(k-i+1)}(0)) \\
|h^{(k)}(0)| &\leq \sum_{i=1}^k c^{\frac{1}{2}-i} i! B_{k,i}(CB, CB^2, \dots, CB^{k-i+1}) \\
&= \sum_{i=1}^k B^k C^i c^{\frac{1}{2}-i} i! B_{k,i}(1, 1, \dots, 1) \\
&= B^k \sqrt{c} \sum_{i=1}^k \left(\frac{C}{c}\right)^i i! |s(k, i)| \text{ using properties of Bell polynomials,} \\
&\leq B^k c^{1/2} \left(\frac{C}{c}\right)^k \sum_{i=0}^k i! |s(k, i)| \text{ which is the ordered Bell number } A_k, \\
&\leq 3c^{1/2} \left(2B\frac{C}{c}\right)^k k!
\end{aligned}$$

using the bound $A_k \frac{x^k}{k!} \leq \frac{1}{2-e^x}$, taken at $x = 1/2$. ■

When applied to $g(t) = f(x + t\delta) - f_*$, where f satisfied the assumptions of Theorem 2, we can take

$$C = \|f - f_*\|_{\mathbb{F}} \quad \text{and} \quad B = 4\pi r \|\delta\|_1,$$

and when $f - f_* \geq \beta$, obtain the bound for $h(t) = \sqrt{g(t)}$:

$$h^{(k)}(0) \leq 3\beta^{1/2} \left(\frac{8\pi r \|f - f_*\|_{\mathbb{F}}}{\beta}\right)^k k! . \quad (9)$$

A.4 Matrix square root of a lower-bounded function

Since our SOS decomposition relies on a matrix square roots for the functions g_1, \dots, g_d , we need the following lemma.

Lemma 6 *We consider a C^∞ function $G : \mathbb{R} \rightarrow \mathbb{R}^{d \times d}$ with values in positive semidefinite matrices and defined on a neighborhood of zero (on the real line) such that $G(0) \succ cI$, with $c > 0$, and such that for all $m \in \mathbb{N}$,*

$$\|G^{(m)}(0)\|_{\text{op}} \leq C \cdot B^m,$$

with $C \geq c$. For $h(x) = \text{tr}[MG(x)^{1/2}]$, with M a symmetric matrix with unit spectral norm, we have:

$$|h^{(k)}(0)| \leq 3c^{1/2} \left(2B\frac{C}{c}\right)^k k! .$$

Proof We use results from [27] and Lemma 7 below, with the operator norm on the set of symmetric matrices and the symmetric square root, where [27, Theorem 1.1] exactly shows that we can take $\alpha(k) = \frac{k!C_{k-1}}{2^{k-1}}c^{1/2-k}$, which is exactly the bound on k -th derivative of the square root which we used in the lemma above. Thus, the exact same derivations can be applied. ■

Lemma 7 *We consider functions $f : \mathbb{R}^a \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}^a$, and $\varphi = f \circ g : \mathbb{R} \rightarrow \mathbb{R}$ that are infinitely differentiable. For a certain norm $\|\cdot\|$ on \mathbb{R}^a , we assume that*

$$|\nabla^k f(g(0))[\delta_1, \dots, \delta_k]| \leq \alpha(k)\|\delta_1\| \cdots \|\delta_k\|,$$

for some $\alpha(k) > 0$. Then for any $n \geq 1$,

$$|\varphi^{(n)}(0)| \leq \sum_{k=1}^n \alpha(k)B_{n,k}(\|g^{(1)}(0)\|, \dots, \|g^{(n-k+1)}(0)\|).$$

Proof We follow the proof of Faà di Bruno's formula that considers a Taylor expansion of g around zero as, for any $m > 0$:

$$g(t) - g(0) = \sum_{k=1}^m \frac{t^k}{k!} g^{(k)}(0),$$

and of f around $g(0)$, as

$$f(g(0) + \delta) - f(g(0)) = \sum_{k=1}^m \frac{1}{k!} \nabla^k f(g(0))[\delta, \dots, \delta].$$

Thus $f(g(t))$ can be expanded as a polynomial in t , with coefficients composed of factors of the form $c\nabla^k f(g(0))[g^{\alpha_1}(0), \dots, g^{\alpha_k}(0)]$, with a *non-negative* coefficient c . Each of them can then be bounded by the term $c\alpha(k)\|g^{\alpha_1}(0)\| \cdots \|g^{\alpha_k}(0)\|$, which is then equivalent to the formula obtained by applying the univariate Faà di Bruno's formula, with a function with derivatives $\alpha(k)$, and the other one with derivatives $\|g^k(0)\|$. We then use the usual formulation with Bell polynomials. ■

We can now apply it to bound derivatives of $g : x \mapsto (x - x_*)^\top R(x)^{1/2}u$. We consider $\varphi(t) = g(x + t\delta)$, and we have, using Leibniz formula:

$$\varphi^{(m)}(0) = (x - x_*)^\top \frac{\partial^m}{\partial t^m} R(x + t\delta)^{1/2}u + m\delta^\top \frac{\partial^{m-1}}{\partial t^m} R(x + t\delta)^{1/2}u.$$

We have

$$h(t) = R(x + t\delta) = \int_0^1 (1 - u)f''(x_* + u(x + t\delta - x_*))du,$$

with derivatives which can be computed as:

$$v^\top h^{(m)}(0)v = \int_0^1 (1 - u)\nabla^{m+2} f(x_* + u(x + t\delta - x_*))[\delta, \dots, \delta, u, u]u^m du.$$

In operator norm, it is less than the supremum over $\|v\|_2 = 1$ of:

$$\|f - f_*\|_{\mathbb{F}}(4\pi r)^{m+2}\|\delta\|_1^m\|v\|_1^2 \int_0^1 (1-u)u^m du \leq \|f - f_*\|_{\mathbb{F}}(4\pi r)^{m+2}\|\delta\|_1^m \frac{d}{m^2}.$$

This leads to constants $C = \|f - f_*\|_{\mathbb{F}}(4\pi r)^2 d$ and $B = 4\pi r\|\delta\|_1$ for the function h , and thus, to a function g with all derivatives of order m less than (using Lemma 6):

$$\begin{aligned} & \sqrt{d} \cdot 3\sqrt{\lambda/2} \left(\frac{4}{\lambda}\|f - f_*\|_{\mathbb{F}}(4\pi r)^3 d\right)^m m! + m \cdot 3\sqrt{\lambda/2} \left(\frac{4}{\lambda}\|f - f_*\|_{\mathbb{F}}(4\pi r)^3 d\right)^{m-1} m! \\ & \leq d\sqrt{\lambda} \left(\frac{6}{\lambda}\|f - f_*\|_{\mathbb{F}}(4\pi r)^3 d\right)^m m! . \end{aligned} \tag{10}$$

B Getting bounds on $\|g_i\|_{\mathbb{F}}$ and $\|g_i\|_{\mathbb{F},s}$

To get such bound, we first realize that all of these functions are products of two functions, so we use Lemma 2 with the estimates in Eq. (10) and Eq. (8) for all $i \in \{1, \dots, d\}$, and the estimates in Eq. (9) and Eq. (8) for $i = d + 1$. We can then use Lemma 1 to obtain the bound in Section 4.2.

Note that for g_{d+1} , we need to consider two cases: one where v is uniformly zero, and thus g_{d+1} is zero as well, and one where v is strictly positive, where $f - f_*$ is lower-bounded by β , and we can apply bounds on derivatives of products.

References

- [1] Jean-Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001. (cited on page 1)
- [2] Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003. (cited on page 1)
- [3] Jean-Bernard Lasserre. *Moments, Positive Polynomials and their Applications*. World Scientific, 2010. (cited on page 1)
- [4] Didier Henrion, Milan Korda, and Jean Bernard Lasserre. *The Moment-SOS Hierarchy: Lectures In Probability, Statistics, Computational Geometry, Control And Nonlinear PDEs*. World Scientific, 2020. (cited on page 1)
- [5] Yurii Nesterov. Squared functional systems and optimization problems. In *High Performance Optimization*, pages 405–440. Springer, 2000. (cited on page 1)
- [6] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993. (cited on page 1)
- [7] Konrad Schmüdgen. *The Moment Problem*. Springer, 2017. (cited on pages 1 and 4)
- [8] Jiawang Nie. Optimality conditions and finite convergence of Lasserre’s hierarchy. *Mathematical Programming*, 146(1):97–121, 2014. (cited on pages 1, 2, 4, and 7)

- [9] Kun Fang and Hamza Fawzi. The sum-of-squares hierarchy on the sphere and applications in quantum information theory. *Mathematical Programming*, 190(1):331–360, 2021. (cited on pages 1, 2, 5, 6, and 10)
- [10] Monique Laurent and Lucas Slot. An effective version of Schmüdgen’s Positivstellensatz for the hypercube. *Optimization Letters*, pages 1–16, 2022. (cited on pages 1, 2, 4, 5, and 6)
- [11] Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach. Finding global minima via kernel approximations. Technical Report 2012.11978, arXiv, 2020. (cited on pages 2, 7, 8, and 10)
- [12] Blake Woodworth, Francis Bach, and Alessandro Rudi. Non-convex optimization with certificates and fast rates through kernel sums of squares. In *Proceedings of the Conference on Learning Theory*, 2022. (cited on pages 2, 3, 8, and 9)
- [13] Alexandre Megretski. Positivity of trigonometric polynomials. In *International Conference on Decision and Control*, volume 4, pages 3814–3817, 2003. (cited on pages 3 and 4)
- [14] Christoph Helmberg, Franz Rendl, Robert J. Vanderbei, and Henry Wolkowicz. An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, 6(2):342–361, 1996. (cited on page 3)
- [15] Bogdan Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*, volume 103. Springer, 2007. (cited on page 3)
- [16] Leopold Fejér. Über trigonometrische Polynome. *Journal für die reine und angewandte Mathematik*, (146):55–82, 1916. (cited on page 4)
- [17] Friedrich Riesz. Über ein Problem des Herrn Carathéodory. *Journal für die reine und angewandte Mathematik*, (146):83–87, 1916. (cited on page 4)
- [18] Mihai Putinar. Sur la complexification du problème des moments. *Comptes Rendus de l’Académie des sciences. Série 1, Mathématique*, 314(10):743–745, 1992. (cited on page 4)
- [19] Aaron Naftalovich and M. Schreiber. Trigonometric polynomials and sums of squares. In *Number Theory*, pages 225–238. Springer, 1985. (cited on page 4)
- [20] Lucas Slot and Monique Laurent. Sum-of-squares hierarchies for binary polynomial optimization. *Mathematical Programming*, pages 1–40, 2022. (cited on pages 6 and 10)
- [21] Ulysse Marteau-Ferey, Francis Bach, and Alessandro Rudi. Second order conditions to decompose smooth functions as sums of squares. Technical Report 2202.13729, arXiv, 2020. (cited on pages 7 and 10)
- [22] Yitzhak Katznelson. *An Introduction to Harmonic Analysis*. Cambridge University Press, 2004. (cited on page 9)
- [23] Walter Rudin. *Fourier Analysis on Groups*. Courier Dover Publications, 2017. (cited on page 10)
- [24] Erik G. F. Thomas. A polarization identity for multilinear maps. *Indagationes Mathematicae*, 25(3):468–474, 2014. (cited on page 13)
- [25] Arie Israel. The eigenvalue distribution of time-frequency localization operators. Technical Report 1502.04404, arXiv, 2015. (cited on pages 13 and 14)

- [26] Charalambos A. Charalambides. *Enumerative Combinatorics*. Chapman and Hall, 2002. (cited on pages 15 and 16)
- [27] Pierre Del Moral and Angele Niclas. A Taylor expansion of the square root matrix function. *Journal of Mathematical Analysis and Applications*, 465(1):259–266, 2018. (cited on page 17)