SIAM J. OPTIM.
Vol. 33, No. 3, pp. 2137–2159

EXPONENTIAL CONVERGENCE OF SUM-OF-SQUARES HIERARCHIES FOR TRIGONOMETRIC POLYNOMIALS*

FRANCIS BACH† AND ALESSANDRO RUDI†

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 minimizing regular multivariate polynomials on the hypercube.

Key words. polynomial optimization, sum of squares, semidefinite programming

MSC codes. 90C22, 11E25

DOI. 10.1137/22M1540818

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 [13, 22], they now apply to many other infinite-dimensional optimization problems such as optimal transport or optimal control (see a thorough review in [15, 10]).

Within optimization, they are often cast as the minimization of multivariate polynomials over sets defined by essentially arbitrary polynomial constraints. They work by solving a sequence of semidefinite 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 [20].
- In higher dimensions, under mild assumptions, the hierarchies are always converging, due to powerful representation results of strictly positive polynomials [24, 29]. 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 [21]. Similar finite convergence results may be obtained in other situations, such as convexity [14, 4].
- In terms of asymptotic convergence rates (in dimension greater than one), they are quite slow, at best $O(1/s^2)$ in the simplest situations for the relaxation with polynomials of degree s [7, 16, 31], already improving on more generic results with rates in $O(1/s^c)$ for an unspecified value of c [30, 1, 2].

*Received by the editors December 12, 2022; accepted for publication (in revised form) March 30, 2023; published electronically August 11, 2023.
https://doi.org/10.1137/22M1540818

Funding: This work was supported by 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).

†Inria, Éole Normale Supérieure, PSL Research University, Paris, France (francis.bach@inria.fr, alessandro.rudi@inria.fr).

Hierarchies for polynomial optimization come in two main types, using two different representations for nonnegative polynomials under polynomial constraints. The "Putinar representation" adds as many polynomials as the number of constraints [24], while the "Schmüdgen representation" adds an exponential number [29]. In this paper, we focus on one of the simplest formulations of minimizing polynomials on $[-1,1]^d$ with the Schmüdgen representation, 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 primary 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 [16] for polynomials on $[-1,1]^d$, using a similar proof technique (taken from [7]) but simpler arguments and explicit constants.
- When we add local optimality conditions similar to [21], we prove in section 4 an exponential convergence rate with explicit (but more complex) constants. The proof technique is taken from [26, 34], who showed convergence rates faster than any polynomial in s, but without explicit constants.

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

2. Problem setup.

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

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

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 complex exponentials $f(x) = \sum_{\omega\in\mathbb{Z}^d} \hat{f}(\omega) e^{2i\pi\omega^\top x}$, where the series is uniformly convergent. A key property of the F-norm is its relationship with the L_∞-norm, that is, $\|f\|_\infty \leqslant \|f\|_{\mathrm{F}}$.

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,

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

We denote by x_* any minimizer of f on $[0,1]^d$ and by f_* the minimal value (which does not depend on the chosen minimizer).

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 the optimal value by $c_*(f,s)$. The principle behind SOS hierarchies is that when f is a trigonometric polynomial, this optimization problem can be solved as a finite-dimensional semidefinite programming (SDP) problem that we describe in section 2.1, and thus can be solved with a variety of algorithms (see, e.g., [9]).

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

$$0 \leqslant \inf_{x\in[0,1]^d} f(x) - c_*(f,s) \leqslant \varepsilon(f,s), \tag{2.1}$$

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 [34], by using $c - \|f-c-g\|_{\mathrm{F}}$ 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 [34] for details).

2.1. Semidefinite programming formulations. In this section, we provide an explicit description of the semidefinite program for the SOS relaxation, as well as the associated spectral relaxation. For trigonometric polynomials, the optimization problems can be compactly written.

For an integer s, we consider the feature map $\varphi : [0,1]^d \to \mathbb{C}^{(2s+1)^d}$, indexed by $\omega \in \{-s,\dots,s\}^d$ with values

$$\varphi_\omega(x) = \frac{1}{(2s+1)^{d/2}} \exp(2i\pi\omega^\top x). \tag{2.2}$$

It satisfies $\|\varphi(x)\| = 1$ for all $x \in [0,1]^d$, where $\|\cdot\|$ denotes the standard Hermitian norm.

We can represent any trigonometric polynomial of degree $2s$ as a quadratic form in $\varphi(x)$, that is, we can write f (nonuniquely) as $f(x) = \varphi(x)^* F \varphi(x)$, where F is a Hermitian matrix of dimension $(2s+1)^d \times (2s+1)^d$. We denote by $\mathcal{V}_s$ the set of multivariate Hermitian Toeplitz matrices in dimension $(2s+1)^d \times (2s+1)^d$, that is, Hermitian matrices Σ such that $\Sigma_{\omega\omega'}$ depends only on $\omega - \omega' \in \mathbb{Z}^d$. It turns out that the span of all matrices $\varphi(x)\varphi(x)^*$ for $x \in [0,1]^d$ is exactly $\mathcal{V}_s$. We denote by $\mathcal{V}_s^\perp$ the orthogonal complement of $\mathcal{V}_s$ for the dot-product $(M,N) \mapsto \operatorname{tr}(M^*N)$.

Primal-dual formulations. The SOS relaxation is obtained by solving

$$\max_{c\in\mathbb{R},\ A\succcurlyeq 0} \ c \quad \text{such that} \ \ \forall x \in [0,1]^d,\ f(x) = c + \varphi(x)^* A \varphi(x).$$

It can be rewritten using $\mathcal{V}_s$ as

$$\begin{aligned} &\max_{c\in\mathbb{R},\ A\succcurlyeq 0} \ c \quad \text{such that} \quad \forall x \in [0,1]^d,\ \operatorname{tr}\left[\varphi(x)\varphi(x)^*(F - cI - A)\right] = 0 \\ &= \max_{c\in\mathbb{R},\ A\succcurlyeq 0,\ Y\in\mathcal{V}_s^\perp} \ c \quad \text{such that} \quad F - cI - A + Y = 0 \\ &= \max_{Y\in\mathcal{V}_s^\perp} \ \lambda_{\min}(F+Y), \end{aligned} \tag{2.3}$$

whose optimal value is $c_*(f,s)$. Its dual can be written as, using standard semidefinite duality,

$$
\begin{aligned}
\max_{Y\in\mathcal{V}_s^\perp} \lambda_{\min}(F+Y) &= \min_{\Sigma\succcurlyeq 0} \max_{Y\in\mathcal{V}_s^\perp} \operatorname{tr}[\Sigma(F+Y)] \ \text{ such that } \ \operatorname{tr}(\Sigma)=1 \\
&= \min_{\Sigma\succcurlyeq 0} \operatorname{tr}(\Sigma F) \ \text{ such that } \ \operatorname{tr}(\Sigma)=1,\ \Sigma\in\mathcal{V}_s,
\end{aligned} \tag{2.4}
$$

which corresponds to an outer approximation of the convex hull of all $\varphi(x)\varphi(x)^*$, $x\in[0,1]^d$, by the set of positive semidefinite matrices such that $\operatorname{tr}(\Sigma)=1$ and $\Sigma\in\mathcal{V}_s$.

Spectral relaxation. We can further relax the problem by equivalently setting $Y=0$ in (2.3), or removing the constraint $\Sigma\in\mathcal{V}_s$ in (2.4), and we simply obtain $\lambda_{\min}(F)$, which is the natural spectral relaxation of the minimization of $\varphi(x)^*F\varphi(x)$, by only considering that $\|\varphi(x)\|=1$. This relaxation is appealing computationally as it can be solved in quadratic time in the dimension of F as opposed to more than cubic for the SDP corresponding to the SOS problem, but it leads in general to slow rates (see Appendix B).

2.2. 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 $f:[0,1]^d\to\mathbb{R}$ 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 $\left[Q(\cos 2\pi y_1,\dots,\cos 2\pi y_d,\sin 2\pi y_1,\dots,\sin 2\pi y_d)\right]^2$, where Q is a regular multivariate polynomial.

We can then use the unique decomposition of multivariate trigonometric polynomials as[1]

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

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

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

When $J=J'$, 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

$$Q_J(x_1,\dots,x_d)^2\prod_{j\in J}(1-x_j^2), \tag{2.5}$$

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.

[1] This is a simple consequence of the definitions of Chebyshev polynomials of the first and second kinds (see, e.g., [6]), which show that for $\omega\geqslant 1$, $\cos 2\pi\omega z$ 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 $\cos 2\pi z$.

Thus, using Chebyshev polynomials, we get precisely the Schmüdgen representation [29] of polynomials on $[-1,1]^d$, as the sum of terms of the form in (2.5) for all subsets $J \subset \{1,\dots,d\}$. Therefore, the 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. We therefore actually extend results from [16], which themselves provide a quantitative rate in $O(1/s^2)$, improving on the rates of the form $O(1/s^c)$, for an unspecified value of c, obtained in the more general setup of all Schmüdgen representations by [30] (see [1] for a similar result for Putinar representations).

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.

Transfer of local optimality conditions. While Theorem 3.1 (section 3) will apply directly to regular polynomials with the construction above, Theorem 4.1 (section 4) will require the function f to have finitely many isolated second-order strict minimizers. We show below that local second-order strict optimality conditions for the minimization of a regular polynomial on $[-1,1]^d$ translates to second-order strict optimality conditions for the corresponding problem on trigonometric polynomials.

By symmetry, any $x \in (-1,1)^d$ is represented by 2^d potential y's such that $x_i = \cos 2\pi y_i$, for $i \in \{1,\dots,d\}$, and if the minimum of P on $[-1,1]^d$ is attained in x_* in the interior $(-1,1)^d$, represented by $y_* \in [0,1]^d$ (any of the 2^d possible ones), we have $\frac{\partial P}{\partial x_i}(x_*) = 0$ for all $i \in \{1,\dots,d\}$, and thus $\frac{\partial f}{\partial y_i}(y_*) = -2\pi \sin[2\pi (y_*)_i] \frac{\partial P}{\partial x_i}(x_*) = 0$, and

$$\begin{aligned}\frac{\partial^2 f}{\partial y_i \partial y_j}(y_*) &= -1_{i=j}(2\pi)^2 \cos[2\pi(y_*)_i]\frac{\partial P}{\partial x_i}(x_*)\\ &\qquad + (2\pi)^2 \sin[2\pi(y_*)_i]\sin[2\pi(y_*)_j]\frac{\partial^2 P}{\partial x_i \partial x_j}(x_*)\\ &= (2\pi)^2 \sin[2\pi(y_*)_i]\sin[2\pi(y_*)_j]\frac{\partial^2 P}{\partial x_i \partial x_j}(x_*).\end{aligned}$$

Since $x_* \in (-1,1)^d$, $\sin[2\pi(y_*)_i] \neq 0$ for all $i \in \{1,\dots,d\}$, and thus, if the Hessian of P at x_* is positive definite, so is the one f at y_*, and therefore we obtain 2^d strict second-order minimizers for the trigonometric polynomial if the original polynomial had such a minimizer in the interior of $[-1,1]^d$.

If the minimizer x_* is on the boundary, we obtain a similar result. Indeed, assume without loss of generality that $(x_*)_i = 1$ for $i \in \{1,\dots,r\}$ and $(x_*)_i \in (-1,1)$ for $i \in \{r+1,\dots,d\}$. We consider the following standard sufficient conditions for a strict local minimizer: $\frac{\partial P}{\partial x_i}(x_*) < 0$ for $i \in \{1,\dots,r\}$, $\frac{\partial P}{\partial x_i}(x_*) = 0$ for $i \in \{r+1,\dots,d\}$, and the square submatrix of the Hessian corresponding to indices in $\{r+1,\dots,d\}$ is positive definite. Then, using the partial derivative computations above, we have $\frac{\partial f}{\partial y_i}(y_*) = -2\pi\sin[2\pi(y_*)_i]\frac{\partial P}{\partial x_i}(x_*) = 0$ for all $i \in \{1,\dots,d\}$, since either $\frac{\partial P}{\partial x_i}(x_*) = 0$ or $\sin[2\pi(y_*)_i] = 0$. Moreover, the Hessian of f is block diagonal with one block composed of a diagonal matrix with elements $-(2\pi)^2\cos[2\pi(y_*)_i]\frac{\partial P}{\partial x_i}(x_*)$ (which are strictly positive for $i \in \{1,\dots,r\}$) and another block with elements $(2\pi)^2\sin[2\pi(y_*)_i]\sin[2\pi(y_*)_j]\frac{\partial^2 P}{\partial x_i \partial x_j}(x_*)$, which is a positive definite block by assumption. Thus the Hessian is positive definite, and we obtain a second-order strict minimizer.

2.3. 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 $s \geqslant r$, as all nonnegative trigonometric polynomials are sums of squares [8, 25].
- When $d = 2$, then for any trigonometric polynomial f, the relaxation is tight with s sufficiently large (but unknown a priori bound), that is, $\varepsilon(f, s)$ is equal to zero for s greater than some $s_0(f)$ (as a consequence of [28, Corollary 3.4]).
- When $d > 1$, any *strictly positive* trigonometric polynomial is a sum of squares [23, 18], but there exist nonnegative polynomials which are not sum of squares [19]. Thus SOS hierarchies have to converge but cannot always be finitely convergent.
- When the set of zeros of the nonnegative 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 [21].

The goal of this paper is to provide upper bounds of $\varepsilon(f, s)$ in (2.1) 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. Since no assumptions are made on polynomials except their degrees, this directly leads to an approximation result for moment matrices presented in section 3.1.[2]

THEOREM 3.1. *For any trigonometric polynomial f of degree less than $2r$, we have, for any $s \geqslant 3r$, and for $\bar{f} = \hat{f}(0)$ the mean value of f,*

$$\varepsilon(f, s) \leqslant \|f - \bar{f}\|_{\mathrm{F}} \cdot \left[\left(1 - \frac{6r^2}{s^2}\right)^{-d} - 1\right] \sim_{s \to +\infty} \|f - \bar{f}\|_{\mathrm{F}} \cdot \frac{6r^2 d}{s^2}.$$

Proof. We here follow the proof technique of [7, 16] 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

$$Th(x) = \int_{[0,1]^d} |q(x - y)|^2 h(y) dy, \tag{3.1}$$

for a well-chosen 1-periodic function q which is a trigonometric polynomial of degree s. The function $x \mapsto |q(x - y)|^2$ is an element of the finite-dimensional cone of SOS polynomials of degree s; thus, by design, if h is a nonnegative 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 $b \geqslant 0$, 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 $f_* - c_*(f, s) \leqslant b$.

In the Fourier domain, since convolutions lead to pointwise multiplication, and vice versa, we have for all $\omega \in \mathbb{Z}^d$, where $\hat{q} * \hat{q}(\omega)$ is shorthand for $(\hat{q} * \hat{q})(\omega)$,

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

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

[2] Sections 3 and 4 are independent, and thus can read in any order.

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

otherwise. If we impose that $\hat{q} * \hat{q}(0) = 1$, we then have

$$\begin{aligned} f - f_* + b - h &= \sum_{\omega \in \mathbb{Z}^d} \hat{f}(\omega)\Big(1 - \frac{1}{\hat{q} * \hat{q}(\omega)}\Big) \exp(2i\pi\omega^\top \cdot) \\ &= \sum_{\omega \neq 0} \hat{f}(\omega)\Big(1 - \frac{1}{\hat{q} * \hat{q}(\omega)}\Big) \exp(2i\pi\omega^\top \cdot). \end{aligned}$$

We then get $\|f - f_* + b - h\|_\infty = \big\| \sum_{\omega \neq 0} \hat{f}(\omega)\big(1 - \frac{1}{\hat{q}*\hat{q}(\omega)}\big) \exp(2i\pi\omega^\top \cdot)\big\|_\infty$.

Using that $\|\cdot\|_\infty \leqslant \|\cdot\|_{\mathrm{F}}$, we get

$$\|f - f_* + b - h\|_\infty \leqslant \sum_{\omega \neq 0} |\hat{f}(\omega)| \cdot \max_{\|\omega\|_\infty \leqslant 2r} \Big| \frac{1}{\hat{q} * \hat{q}(\omega)} - 1 \Big| \leqslant \|f - \bar{f}\|_{\mathrm{F}} \cdot \max_{\|\omega\|_\infty \leqslant 2r} \Big| \frac{1}{\hat{q} * \hat{q}(\omega)} - 1 \Big|.$$

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

A simple candidate is $\hat{q}(\omega) = \frac{1}{(2s+1)^{d/2}} 1_{\|\omega\|_\infty \leqslant s}$, based on a "box kernel"; we can then compute the convolution and obtain that $\hat{q} * \hat{q}(\omega) = \prod_{i=1}^d \big(1 - \frac{|\omega_i|}{2s+1}\big)_+ \geqslant \big(1 - \frac{2r}{2s+1}\big)^d$, leading to $b = \|f - \bar{f}\|_{\mathrm{F}} \cdot \big[\big(1 - \frac{2r}{2s+1}\big)^{-d} - 1\big]$. When s goes to infinity, we have the equivalent $b \sim \|f - \bar{f}\|_{\mathrm{F}} \cdot \frac{rd}{s} = O(1/s)$, which thus converges to zero, but at a slow rate.

A better candidate leads to a rate in $O(1/s^2)$ (as in [7, 16]) and is based on a "triangular kernel," such as

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

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

$$\hat{q} * \hat{q}(\omega) = a^2 \prod_{i=1}^d \Big[\frac{2s}{3} + \frac{1}{3s} - \frac{\omega_i^2}{s} + \frac{|\omega_i|}{2s^2}(\omega_i^2 - 1)\Big]. \tag{3.2}$$

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

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

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

$$b \leqslant \|f - \bar{f}\|_{\mathrm{F}} \cdot \Big[\Big(1 - \frac{6r^2}{s^2}\Big)^{-d} - 1\Big] \sim \|f - \bar{f}\|_{\mathrm{F}} \cdot \frac{6r^2 d}{s^2}.$$

Above, the asymptotic equivalent is taken with s tending to infinity, with r and d being fixed. ∎

We make the following observations:

- The proposed bound follows a series of earlier bounds with 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 [7, 16, 31, 32]. The most closely related is that of [16], which considers regular polynomials on $[-1,1]^d$ with Schmüdgen's representation, but with a different choice for the function q in (3.1). As shown in section 2.2, our bound also applies to this case through a change of variable; it differs in the choice of normalization of coefficients (for us, the ℓ_1-norm of the expansion in Chebyshev polynomials).
- Note that we could extend this result to other types of regularity beyond finite support and bounded F-norm, with the asymptotic bound $\|f-\bar{f}\|_{\mathrm{F}} \cdot \frac{6r^2 d}{s^2} + \sum_{\|\omega\|_\infty > 2r} |\hat{f}(\omega)|$, and by optimizing over $r \leqslant s$.
- 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. To obtain a faster rate in the simplest one-dimensional case, the function $r:[0,1]\to\mathbb{R}$ defined as $r(x) = |q(x)|^2$ should be such that its Fourier series $\hat{r}(\omega)$ is of the form $f(\omega/s)$ for a function $f:\mathbb{R}\to\mathbb{R}$ such that $f''(0)=0$ and with support in $[-2,2]$. Thus, when s gets large, $r(x) = \sum_{|\omega|\leqslant 2s} f(\omega/s) e^{2i\pi\omega x}$ should be proportional to the Fourier transform of f. Thus the Fourier transform of f should be nonnegative 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 leads to significantly better rates.
- As shown in Appendix B, it turns out that a simple spectral relaxation of the problem already achieves a rate in $\|f-\bar{f}\|_{\mathrm{F}} \cdot \frac{rd}{s}$ which is worse than the $O(1/s^2)$ rate that we show in this section, but not representative of the empirical differences between the two methods. Our following result will show an explicit benefit of the SOS relaxation by obtaining exponential convergence rates (with extra assumptions on f).

3.1. Approximation of moment matrices. We denote by $\mathcal{K}_s$ the closure of the convex hull of all Hermitian matrices $\varphi(x)\varphi(x)^* \in \mathbb{C}^{(2s+1)^d \times (2s+1)^d}$ for $x \in [0,1]^d$. It contains exactly all moment matrices; SOS relaxations can then be interpreted by relaxing it to the set $\widehat{\mathcal{K}}_s$ of "pseudo-moment matrices" Σ such that $\Sigma \in \mathcal{V}_s$, $\Sigma \succcurlyeq 0$, and $\mathrm{tr}(\Sigma) = 1$. For $r \leqslant s$, we denote by $\Pi_s^{(r)}$ the linear operator on $\mathcal{V}_s$ that sets of elements $H_{\omega\omega'}$ to zero as soon as $\|\omega\|_\infty > r$ or $\|\omega'\|_\infty > r$, and we multiply all other elements by $(2s+1)^d/(2r+1)^d$ (making it essentially an element of $\mathcal{V}_r$). A classical duality argument leads to the following corollary of Theorem 3.1. See the proof in Appendix C.

Corollary 3.2. *For any $s \geqslant 3r$, and any $\Sigma \in \widehat{\mathcal{K}}_s$, there exists $\Sigma' \in \mathcal{K}_s$ such that*

$$\|\Pi_s^{(r)}(\Sigma - \Sigma')\|_{\mathrm{Frob}} \leqslant \frac{\sqrt{2}}{(2r+1)^d}\Big[\Big(1 - \frac{6r^2}{s^2}\Big)^{-d} - 1\Big],$$

where $\|M\|_{\mathrm{Frob}}$ denotes the Frobenius norm of M.

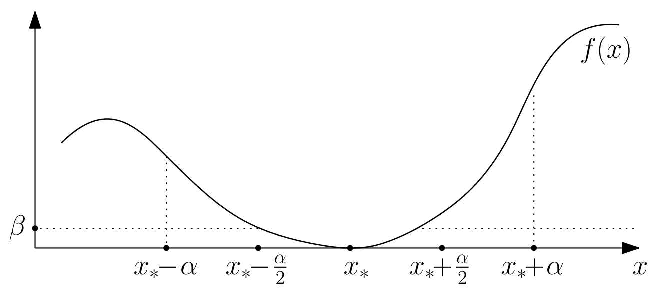
This corollary shows that matrices in $\mathcal{K}_r$ can be well approximated by projections of matrices in $\widehat{\mathcal{K}}_s$. Note that the factor $(2r+1)^{-d}$ is an outcome of our choice of normalization to unit traces.

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

$$(4.1) \qquad \|x - x_*\|_\infty \leqslant \alpha \Rightarrow f''(x) \succcurlyeq \lambda I \quad \text{and} \quad \|x - x_*\|_\infty \geqslant \frac{\alpha}{2} \Rightarrow f(x) - f(x_*) \geqslant \beta,$$

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 the illustration below in one dimension.



The proof technique is based on the one introduced in Lemma 1 and Theorem 2 of [26] (for the nonperiodic case and without explicit constants) and can be extended directly to situations where the global minimum is attained at finitely many points with the same local Hessian condition (see also [17] for cases where minimizers are whole manifolds).

Note that in that regime, the hierarchy is known to be finitely convergent [21], 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. We denote by $\nabla^m f(x)$ the symmetric mth order tensor of mth order derivatives, with element $\nabla^m f(x)_{j_1,\dots,j_m} = \frac{\partial^m f}{\partial x_{j_1}\cdots\partial x_{j_m}}(x)$, where $j_1,\dots,j_m \in \{1,\dots,d\}$. Throughout the proofs, we will use the notation $\nabla^m f(x)[\delta,\dots,\delta] \in \mathbb{R}$ to denote the contraction of the tensor along the m copies of δ, that is, $\nabla^m f(x)[\delta,\dots,\delta] = \sum_{j_1,\dots,j_m=1}^d \nabla^m f(x)_{j_1,\dots,j_m}\delta_{j_1}\cdots\delta_{j_m}$. We consider bounds on derivatives of the form

$$\|\nabla^m f\|_\infty = \max_{x\in[0,1]^d} \sup_{\|\delta\|_1\leqslant 1} |\nabla^m f(x)[\delta,\dots,\delta]|.$$

Up to a constant that depends on m, this is equivalent to imposing a bound on all partial derivatives (see Appendix E for a precise relationship). This can also be seen as a bound on all directional derivatives, that is, of all $|g^{(m)}(0)|$, for $g(t) = f(x + t\delta)$.

THEOREM 4.1. *Assume that $f : [0,1]^d \to \mathbb{R}$ is infinitely differentiable and such that $\|\nabla^m f\|_\infty \leqslant \|f - f_*\|_{\mathrm{F}}(4\pi r)^m$ for all $m \geqslant 0$. Assume there exist $x_* \in [0,1]^d$, as well as $\alpha \in [0,1/2), \beta > 0$, and $\lambda > 0$, such that* (4.1) *is satisfied. Then we have*

$$\varepsilon(s,f) \leqslant \triangle_1 \exp\Big(-\Big(\frac{s}{\triangle_2}\Big)^{1+\xi}\Big)$$

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

$$(4.2) \qquad \triangle_1 = (\beta + \lambda d^3)\big(32B^3d^6\big)^{d+1}, \ \triangle_2 = dB,$$

where $B = \max\big\{\frac{275}{\alpha\xi}, \frac{8\pi r\|f-f_*\|_{\mathrm{F}}}{\beta}, \frac{6}{\lambda}\|f - f_*\|_{\mathrm{F}}(4\pi r)^3\big\}$.

Before describing the proof, we can make a few simple observations:
- Trigonometric polynomials of degree $2r$ satisfy the required growth condition, because for $f : x \mapsto e^{2i\pi\omega^\top x}$ where $\omega \in \mathbb{Z}^d$ such that $\|\omega\|_\infty \leqslant 2r$, we have $\|\nabla^m f\|_\infty \leqslant (2\pi \cdot 2r)^m$.
- The result extends a prior result [34] that was showing convergence rates faster than any power of s, but without explicit constants, which are needed to obtain the exponential rate. When the conditioning constants λ, α, β tend to zero, the constant Δ_2 in (4.2) tends to infinity, and the rate is not informative. In this situation, we could add a regularizer and optimize its strength to obtain a rate.
- We could easily consider weaker growth conditions for the mth order derivatives (with slower convergence rates), such as $\|\nabla^m f\|_\infty = O(r^m m!)$.
- We could optimize over $\xi \in [0, 1/2)$ to get a better dependence in s.
- The result can be extended to functions with finitely many isolated second-order strict minimizers (following [26, Theorem 2]).

4.1. Proof technique. The main technical result is to show that the nonnegative function $f - f_*$ can be approximated by a trigonometric polynomial g which is a sum of squares of polynomials of degree at most s, with an error bound measured in the norm $\|\cdot\|_{\mathrm{F}}$ as $\|f - f_* - g\|_{\mathrm{F}} \leqslant \varepsilon'(s, f)$. Thanks to the following technical lemma, whose proof is in Appendix D, this leads to the desired result with $\varepsilon(s, f) = (2s+1)^d \varepsilon'(s, f)$.

LEMMA 4.2. *Assume f is a trigonometric polynomial of degree less than $2s$, with minimal value f_*. If there exists a trigonometric polynomial g which is a sum of squares of polynomials of degree at most s such that $\|f - f_* - g\|_{\mathrm{F}} \leqslant \varepsilon'(s, f)$, then the optimal value $c_*(f, s)$ of (2.3) and (2.4) satisfies*

$$0 \leqslant f_* - c_*(f, s) \leqslant (2s+1)^d \varepsilon'(s, f).$$

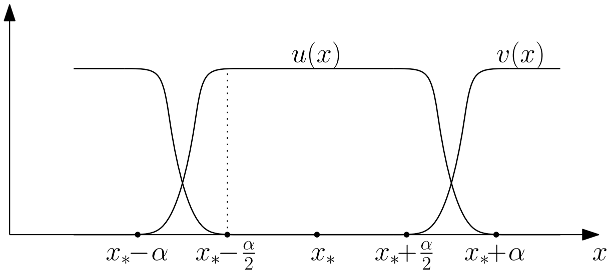
To obtain the desired approximant g, we follow the approach of [34] and build an exact representation of $f - f_*$ as the sum of squared infinitely differentiable functions. We then truncate the Fourier series of these functions to obtain the approximation.

To provide the exact SOS decomposition, following [26], we provide a decomposition around x_*, where the function $f - f_*$ has a zero, and away from x_*, where the function is strictly positive. This is then glued together with "partitions of unity" which we now present.

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

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

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



These are usually referred to as partitions of the unity and will be built in section 4.2 using standard tools. Following [26], we can then decompose f as, using Taylor's formula with integral remainder,

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

with $R(x)=\int_0^1(1-t)f''(x_*+t(x-x_*))dt\succcurlyeq\frac{\lambda}{2}$ if $\|x-x_*\|_\infty\leqslant\alpha$, and $(z_1,\dots,z_d)\in\mathbb{R}^{d\times d}$ any orthonormal basis of $\mathbb{R}^d$.

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

$$\forall x\in[0,1]^d,\ f(x)-f(x_*)=\sum_{i=1}^{d+1}g_i(x)^2,$$

with

$$g_i(x)=u(x)(x-x_*)^\top R(x)^{1/2}z_i \text{ for } i\in\{1,\dots,d\}, \tag{4.3}$$

$$g_{d+1}(x)=v(x)\sqrt{f(x)-f(x_*)}, \tag{4.4}$$

which are infinitely differentiable functions (just taking the square root of $f-f_*$ without taking care of the region around the minimizer as 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 $\|\omega\|_\infty\leqslant s$, leading to the following, using lemmas from [34] about the F-norm (see also [12, section I.6]):

$$\begin{aligned}
\Big\|f-f_*-\sum_{i=1}^{d+1}\bar g_i^2\Big\|_{\mathrm F}&=\Big\|\sum_{i=1}^{d+1}g_i^2-\sum_{i=1}^{d+1}\bar g_i^2\Big\|_{\mathrm F}\leqslant\sum_{i=1}^{d+1}(\|g_i\|_{\mathrm F}+\|\bar g_i\|_{\mathrm F})\|g_i-\bar g_i\|_{\mathrm F}\\
&\leqslant 2\sum_{i=1}^{d+1}\|g_i\|_{\mathrm F}\cdot\sum_{\|\omega\|>s}|\hat g_i(\omega)|=2\sum_{i=1}^{d+1}\|g_i\|_{\mathrm F}\cdot\|g_i\|_{\mathrm F,s},
\end{aligned} \tag{4.5}$$

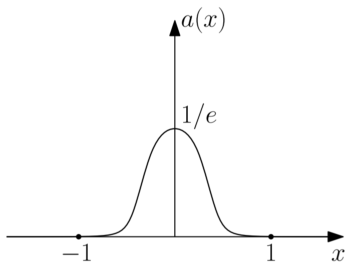
where we denote $\|f\|_{\mathrm F,s}=\sum_{\|\omega\|_\infty>s}|\hat f(\omega)|$. We thus need to find bounds on $\|g_i\|_{\mathrm F}$ and $\|g_i\|_{\mathrm F,s}$, for $i\in\{1,\dots,d+1\}$, and then multiply the bound in (4.5) above by the term $(2s+1)^d$ from Lemma 4.2.

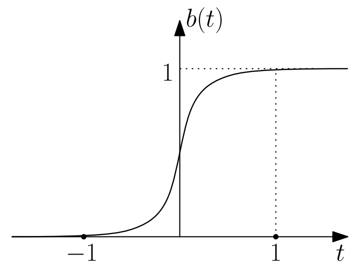
Since these functions are C^∞ (i.e., infinitely differentiable), the decay of their Fourier series is faster than any power, as already noted in [34]. In the present paper, we provide explicit constants that allow us to obtain an exponential convergence rate.

We will obtain bounds on Fourier series coefficients of the functions g_i defined in (4.3) and (4.4) by bounding their derivatives. Since they are defined as products, we need to bound the derivatives of each part: the partitions of unity u and v (section 4.2), the scalar square root $(f-f_*)^{1/2}$ (section 4.3), and the matrix square root $R^{1/2}$ (section 4.4). The bounds are then put together in section 4.5.

The key to obtaining bounds on order m derivatives is to track the dependence in m, with bounds of the form $c_1^m m!^{1+c_2}$ for constants c_1, c_2.

4.2. Partitions of unity. Following [11, section 3.1], we consider for $\eta \in (0,1]$ the function $a : \mathbb{R} \to \mathbb{R}$ defined as $a(x) = \exp(-(1-x^2)^{-1/\eta})$ on $[-1,1]$, and zero otherwise. We then consider the function $b : \mathbb{R} \to \mathbb{R}$, defined as $b(t) = \frac{\int_{-\infty}^{t} a(x)dx}{\int_{-\infty}^{+\infty} a(x)dx}$, which is nondecreasing, equal to zero for $t \leqslant 1$, and equal to 1 if $t \geqslant 1$. These two functions are infinitely differentiable on $\mathbb{R}$. See the illustrations below.





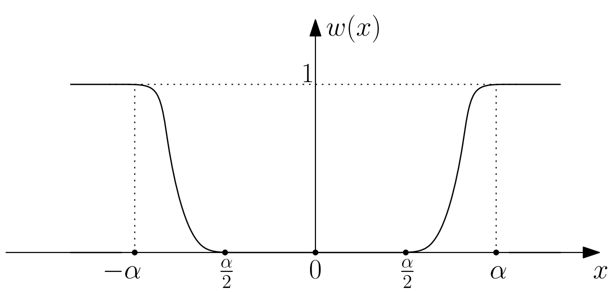
We have, from [11, section 3.1], $|a^{(m)}(x)| \leqslant \big(\frac{16}{\eta}\big)^m m^{(1+\eta)m}$ for any $m \geqslant 0$ and any $x \in [-1,1]$. Moreover, we have

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

Using $\log(1-x) \geqslant -(2\log 2)x$ for $x \in [0,1/2]$, we get the lower bound[3]

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

We consider the function w defined on $[-\frac{1}{2},\frac{1}{2}]$ as $w(x) = b\big[\frac{4}{\alpha}\big(|x| - \frac{3\alpha}{4}\big)\big]$, and extended by 1-periodicity to $\mathbb{R}$. It is of the form plotted below.



[3]Note that the bound from from [11] is incorrectly independent of η.

Moreover we have, through the explicit expression of w and the bounds on the derivatives of a and b,

$$\forall x \in \Big[-\frac{1}{2},\frac{1}{2}\Big],\ |w^{(m+1)}(x)| \leqslant 8\sqrt{1/\eta}\Big(\frac{64}{\alpha\eta}\Big)^m m^{(1+\eta)m},$$

which, for $m>0$, leads to

$$\forall x \in \left[-\frac{1}{2},\frac{1}{2}\right],\ |w^{(m)}(x)| \leqslant 8\sqrt{1/\eta}\frac{\alpha\eta}{64}\left(\frac{64}{\alpha\eta}\right)^m m^{(1+\eta)m} \leqslant c^m m^{(1+\eta)m},$$

with $c=\frac{64}{\alpha\eta}$, an equality which is also valid for $m=0$ (where $|w(x)|\leqslant 1$).

We then consider the functions

$$u(x) = \sin\Big[\frac{\pi}{2}\prod_{i=1}^d \big(1-w(x_i-(x_*)_i)\big)\Big] \tag{4.6}$$

$$v(x) = \cos\Big[\frac{\pi}{2}\prod_{i=1}^d \big(1-w(x_i-(x_*)_i)\big)\Big]. \tag{4.7}$$

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 $\|x-x_*\|_\infty \leqslant \alpha/2$, $u(x)=\sin(\pi/2)=1$, as well as when $\|x-x_*\|_\infty \geqslant \alpha$, $u(x)=0$. The next lemma provides bounds on their derivatives.

Lemma 4.3. *For the function u defined in* (4.6) *and* (4.7), *we have for all $m>0$*

$$\|\nabla^m u\|_\infty \leqslant \Big(\frac{275}{\alpha\eta}\Big)^m m!\cdot m^{\eta m}, \tag{4.8}$$

with the same bound for v as in (4.7).

Proof. We consider the function $g(t)=u(x+\delta t)=\sin\big[\frac{\pi}{2}f(t)\big]$, with $f(t)=\prod_{i=1}^d\big(1-w(x_i-(x_*)_i)\big)$. We can expand the derivatives of the product function f using the Leibniz formula to get, for all t,

$$\begin{aligned}
|f^{(m)}(t)| &\leqslant \frac{\pi}{2}\Bigg|\sum_{\alpha_1+\cdots+\alpha_d=m}\binom{m}{\alpha_1,\ldots,\alpha_d}\prod_{i=1}^d c^{\alpha_i}\alpha_i^{(1+\eta)\alpha_i}\delta_i^{\alpha_i}\Bigg|\\
&\leqslant \frac{\pi}{2}\sum_{\alpha_1+\cdots+\alpha_d=m}\binom{m}{\alpha_1,\ldots,\alpha_d}\prod_{i=1}^d\big[cm^{1+\eta}|\delta_i|\big]^{\alpha_i}\\
&= \frac{\pi}{2}\big[cm^{1+\eta}\|\delta\|_1\big]^m \leqslant \frac{\pi}{2e}\big[cem^{\eta}\|\delta\|_1\big]^m m! \quad \text{using } m^m \leqslant m!e^{m-1}.
\end{aligned}$$

Applying Faà di Bruno's formula (see, e.g., [3]) for the sine function, with the Bell polynomials $B_{m,k}$, we have

$$\begin{aligned}
|g^{(m)}(t)| &\leqslant \sum_{k=1}^m B_{m,k}\Big(\frac{\pi}{2e}\big[ce1^{\eta}\|\delta\|_1\big]^1 1!,\ldots,\frac{\pi}{2e}\big[ce(m-k+1)^{\eta}\|\delta\|_1\big]^{m-k+1}(m-k+1)!\Big)\\
&\leqslant \sum_{k=1}^m B_{m,k}\Big(\frac{\pi}{2e}\big[cem^{\eta}\|\delta\|_1\big]^1 1!,\ldots,\frac{\pi}{2e}\big[cem^{\eta}\|\delta\|_1\big]^{m-k+1}(m-k+1)!\Big),
\end{aligned}$$

using the fact that Bell polynomials have nonnegative coefficients (and are thus nondecreasing functions over the positive orthant). Thus, using that

$$B_{m,k}(\alpha\beta z_1,\ldots,\alpha\beta^{m-k+1}z_{m-k+1}) = \alpha^k\beta^m B_{m,k}(z_1,\ldots,z_{m-k+1}),$$

we get, using an explicit formula for Bell polynomials,[4]

$$
\begin{aligned}
|g^{(m)}(t)| &\leqslant \sum_{k=1}^{m} \Big(\frac{\pi}{2e}\Big)^{k} \big[cem^{\eta}\|\delta\|_1\big]^{m} B_{m,k}\big(1!,\dots,(m-k+1)!\big) \\
&= \big[cem^{\eta}\|\delta\|_1\big]^{n} \sum_{k=1}^{m} \Big(\frac{\pi}{2e}\Big)^{k} \frac{(m-1)!}{(k-1)!}\binom{m}{k}, \\
&\leqslant \big[cem^{\eta}\|\delta\|_1\big]^{n} \sum_{k=1}^{m} \Big(\frac{\pi}{2e}\Big)^{k} m!\binom{m}{k} = \big[cem^{\eta}\|\delta\|_1\big]^{m} m!(1+\pi/2e)^{m} \\
&\leqslant \Big[\frac{64e(1+\pi/(2e))}{\alpha\eta} m^{\eta}\|\delta\|_1\Big]^{m} m! \leqslant \Big[\frac{275}{\alpha\eta} m^{\eta}\|\delta\|_1\Big]^{m} m!\,,
\end{aligned}
$$

which leads to $|\nabla^m u[\delta,\dots,\delta]| \leqslant \big[\frac{275}{\alpha\eta} m^{\eta}\|\delta\|_1\big]^{m} m!$, and thus to the desired result. ∎

4.3. Scalar square root. Since our SOS decomposition relies on the square root of the function $f-f_*$ for the function g_{d+1} in (4.4), we need to bound square roots of functions which are strictly positive and bounded away from zero. By applying Lemma 4.4 below to the function $g: t \mapsto f(x+t\delta) - f(x_*)$, for an arbitrary $\delta \in \mathbb{R}^d$ such that $\|\delta\|_1 \leqslant 1$ and $\|x-x_*\|_\infty \geqslant \frac{\alpha}{2}$, with $c=\beta$, $C=\|f-f_*\|_{\mathrm{F}}$ and $D=4\pi r\|\delta\|_1$, we obtain that for $h: x \mapsto \sqrt{f(x)-f_*}$,

$$
\|\nabla^m h(x)\|_\infty = \max_{\|\delta\|_1\leqslant 1} |\nabla^m h(x)[\delta,\dots,\delta]| \leqslant 3\beta^{1/2}\Big(\frac{8\pi r\|f-f_*\|_{\mathrm{F}}}{\beta}\Big)^{k} k!\,. \tag{4.9}
$$

LEMMA 4.4. *We consider a C^∞ function g defined on a neighborhood of zero (on the real line) such that $g(0)\geqslant c>0$ and such that for all $m\in\mathbb{N}$, $|g^{(m)}(0)|\leqslant C\cdot D^m$, with $C\geqslant c$. For $b(t)=\sqrt{g(t)}$, we have $|b^{(k)}(0)|\leqslant 3c^{1/2}(2D\frac{C}{c})^k k!$.*

Proof. We will use Faà di Bruno's formula (see, e.g., [3]), with the kth derivative of $y\mapsto\sqrt{y}$ being $y^{\frac12-k}(-1)^{k-1}\frac{1}{2k-1}\frac{(2k)!}{(k)!2^{2k}} = y^{\frac12-k}b_k = y^{\frac12-k}k!C_{k-1}2^{1-2k}$ for $k>0$, where $C_k=\frac{1}{k+1}\binom{2k}{k}$ is the Catalan number. Using the classical bound $C_n=\frac{1}{n+1}\frac{(2n)!}{(n!)^2}\leqslant 2\cdot 4^n$, we get $|b_k|\leqslant k!$. Faà di Bruno's formula leads to the following, with the Bell polynomials $B_{k,i}$, and Stirling numbers of the second kind $s(k,i)$:

$$
b^{(k)}(0) = \sum_{i=1}^{k} g(0)^{\frac12-i} b_i B_{k,i}(g'(0),\dots,g^{(k-i+1)}(0)).
$$

This leads to

$$
\begin{aligned}
|b^{(k)}(0)| &\leqslant \sum_{i=1}^{k} c^{\frac12-i} i! B_{k,i}(CD, CD^2,\dots,CD^{k-i+1}) = \sum_{i=1}^{k} D^k C^i c^{\frac12-i} i! B_{k,i}(1,1,\dots,1) \\
&= D^k\sqrt{c}\sum_{i=1}^{k}\Big(\frac{C}{c}\Big)^{i} i!|s(k,i)| \text{ using properties of Bell polynomials,} \\
&\leqslant D^k c^{1/2}\Big(\frac{C}{c}\Big)^{k}\sum_{i=0}^{k} i!|s(k,i)| \text{ which is the ordered Bell number } A_k, \\
&\leqslant 3c^{1/2}\Big(2D\frac{C}{c}\Big)^{k} k!
\end{aligned}
$$

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

[4]See a summary of properties in https://en.wikipedia.org/wiki/Bell_polynomials.

4.4. Matrix square root. Since our SOS decomposition relies on matrix square roots for the functions $g_1, \dots, g_d$ in (4.3), we need the following lemma ($\|M\|_{\mathrm{op}}$ denotes the largest singular value of the matrix M), which can be seen as a matrix extension of Lemma 4.4.

LEMMA 4.5. *We consider a C^∞ function $G : \mathbb{R} \to \mathbb{R}^{d\times d}$ with values in positive semidefinite matrices and defined on a neighborhood of zero (on the real line) such that $G(0) \succcurlyeq cI$, with $c > 0$, and such that for all $m \in \mathbb{N}$, $\|G^{(m)}(0)\|_{\mathrm{op}} \leqslant C \cdot D^m$, with $C \geqslant c$. For $h(x) = \mathrm{tr}[MG(x)^{1/2}]$, with M a symmetric matrix such that $\|M\|_{\mathrm{op}} = 1$, we have $|h^{(k)}(0)| \leqslant 3c^{1/2}\big(2D\frac{C}{c}\big)^k k!$.*

Proof. We use results from [5] and Lemma 4.6 below, with the operator norm on the set of symmetric matrices and the symmetric square root, where [5, Theorem 1.1] exactly shows that we can take $\alpha(k) = \frac{k!C_{k-1}}{2^{2k-1}} c^{1/2-k}$, which is exactly the bound on the kth derivative of the square root which we used in Lemma 4.4 above. Thus, the exact same derivations can be applied. ∎

LEMMA 4.6. *We consider functions $f : \mathbb{R}^a \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}^a$, and $\varphi = f \circ g : \mathbb{R} \to \mathbb{R}$ that are infinitely differentiable. For a certain norm $\|\cdot\|$ on $\mathbb{R}^a$, we assume that*

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

for some $\alpha(k) > 0$. Then for any $n \geqslant 1$, for the Bell polynomials $B_{n,k}$, we have

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

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 *nonnegative* 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 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 $g : x \mapsto (x - x_*)^\top R(x)^{1/2} z$ for $z \in \mathbb{R}^d$. We consider $\varphi(t) = g(x + t\delta)$. We have, using the Leibniz formula,

$$\varphi^{(m)}(0) = (x - x_*)^\top \frac{\partial^m}{\partial t^m} R(x + t\delta)^{1/2} z + m\delta^\top \frac{\partial^{m-1}}{\partial t^{m-1}} R(x + t\delta)^{1/2} z. \tag{4.10}$$

We have, from expressions in section 4.1, with $h(t) = R(x + t\delta)$,

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

with derivatives which can be computed as, for any $v \in \mathbb{R}^d$,

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

Using assumptions from Theorem 4.1, in the operator norm, $h^{(m)}(0)$ is less than the supremum over $\|v\|_2 = 1$ of (using $\|v\|_1^2 \leqslant d\|v\|_2^2$ and integration)

$$\|f - f_*\|_{\mathrm{F}} (4\pi r)^{m+2} \|\delta\|_1^m \|v\|_1^2 \int_0^1 (1-s)s^m ds \leqslant \|f - f_*\|_{\mathrm{F}} (4\pi r)^{m+2} \|\delta\|_1^m \frac{d}{m^2}.$$

Moreover, we have $h(t) \succcurlyeq \int_0^1 (1-s)\lambda I ds = \frac{\lambda}{2} I$. This leads to constants $c = \frac{\lambda}{2}$, $C = \|f - f_*\|_{\mathrm{F}} (4\pi r)^2 d$, and $D = 4\pi r \|\delta\|_1$ for the function h, and thus to the function $g : x \mapsto (x - x_*)^\top R(x)^{1/2} z$, with all derivatives of order m less than (using Lemma 4.5 and (4.10))

$$\sqrt{d} \cdot 3\sqrt{\lambda/2} \Big(\frac{4}{\lambda} \|f - f_*\|_{\mathrm{F}} (4\pi r)^3 d \Big)^m m! + m \cdot 3\sqrt{\lambda/2} \Big(\frac{4}{\lambda} \|f - f_*\|_{\mathrm{F}} (4\pi r)^3 d \Big)^{m-1} (m-1)!,$$

which is less than

$$d\sqrt{\lambda} \Big(\frac{6}{\lambda} \|f - f_*\|_{\mathrm{F}} (4\pi r)^3 d \Big)^m m! \quad . \tag{4.11}$$

4.5. Precise bound. We start with bounds on all derivatives of functions g_i, $i = 1, \ldots, d+1$, defined in (4.3) and (4.4), and then translate them into bounds on their Fourier series coefficients and thus $\|g_i\|_{\mathrm{F}}$ and $\|g_i\|_{\mathrm{F,s}}$.

To get our bound, we first realize that all of these functions are products of two functions, and thus we can use Lemma 4.7 below, proved in Appendix E.2, that bounds derivatives of products.

LEMMA 4.7 (derivatives of products). *Assume that $h_1, h_2 : [0,1]^d \to \mathbb{R}$ is C^∞ and such for all $m \geqslant 0$, $\|\nabla^m h_1\|_\infty \leqslant C_1 \cdot B_1^m \cdot m! \cdot \kappa_1(m)$ and $\|\nabla^m h_2\|_\infty \leqslant C_2 \cdot B_2^m \cdot m! \cdot \kappa_2(m)$. Then $\|\nabla^m (h_1 h_2)\|_\infty \leqslant C_1 C_2 \kappa_1(m) \kappa_2(m)(m+1)! \max\{B_1, B_2\}^m$.*

With the estimates in (4.8) and (4.11), we get

$$\forall i \in \{1, \ldots, d\}, \ \|\nabla^m g_i\|_\infty \leqslant d\sqrt{\lambda} \max \Big\{ \frac{275}{\alpha\eta}, \frac{6}{\lambda} \|f - f_*\|_{\mathrm{F}} (4\pi r)^3 \Big\}^m m! \cdot m^{\eta m}. \tag{4.12}$$

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. We thus get explicit bounds on all derivatives, from (4.8) and (4.9):

$$\|\nabla^m g_{d+1}\|_\infty \leqslant 3\beta^{1/2} \max \Big\{ \frac{275}{\alpha\eta}, \frac{8\pi r \|f - f_*\|_{\mathrm{F}}}{\beta} \Big\}^m m! \cdot m^{\eta m}. \tag{4.13}$$

We can now use Lemma 4.8 below (see the proof in Appendix E.1) that relates the growth of derivatives to the (truncated) F-norm.

LEMMA 4.8 (from derivatives to Fourier decay). *Assume that $g : [0,1]^d \to \mathbb{R}$ is C^∞ and such that for all $m \geqslant 0$, $\|\nabla^m g\|_\infty \leqslant C \cdot B^m \cdot m! \cdot \kappa(m)$, with κ nondecreasing. Then, for $k \geqslant d+1$,*

$$\|g\|_{\mathrm{F}} \leqslant C \Big(2 + \frac{dBk}{2\pi} \Big)^k \kappa(k) \cdot 2(2e)^{d-2},$$
$$\|g\|_{\mathrm{F},s} \leqslant C \Big(2 + \frac{dBk}{2\pi} \Big)^k \kappa(k) 2(2e)^{d-2} (s+1)^{d-k}.$$

With $B = \max\{\frac{275}{\alpha\eta}, \frac{8\pi r\|f-f_*\|_{\mathrm{F}}}{\beta}, \frac{6}{\lambda}\|f - f_*\|_{\mathrm{F}}(4\pi r)^3\} \geqslant 275$, we get from Lemma 4.8 above, (4.12), and (4.13), for all $k \geqslant d+1$,

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

and thus a bound from (4.5),

$$\begin{aligned}
c &\leqslant 8(9\beta + \lambda d^3)\Big(2 + \frac{Bd(d+1)}{2\pi}\Big)^{d+1}\Big(2 + \frac{dBk}{2\pi}\Big)^{k}(2e)^{2d-4}s^{d-k}k^{\eta k}\\
&\leqslant (\beta + \lambda d^3)\big(6Bd^2\big)^{d+1}\Big(\frac{dBk}{6}\Big)^{k}s^{d-k}k^{\eta k}.
\end{aligned}$$

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

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

Overall, using the identity $e^{-z} \leqslant \big(\frac{c}{ez}\big)^c$, applied to $c = \frac{3d}{2}$ and $z = \frac{2}{5}\big(\frac{s}{dB}\big)^{1/(1+\eta)}$, and multiplying the bound in (4.5) above by the term $(2s+1)^d$ from Lemma 4.2, and using $\eta \in (0,1]$, we get

$$\begin{aligned}
\varepsilon(f,s) &\leqslant (2s+1)^d(\beta + \lambda d^3)\big(6Bd^2\big)^{d+1}s^d\exp\Big(-\Big(\frac{2s}{dB}\Big)^{1/(1+\eta)}\Big)\\
&\leqslant (\beta + \lambda d^3)\big(9Bd^2\big)^{d+1}s^{3d/2}\exp\Big(-\Big(\frac{2s}{dB}\Big)^{1/(1+\eta)}\Big)\\
&\leqslant (\beta + \lambda d^3)\big(9Bd^2\big)^{d+1}s^{2d}\exp\Big(-\Big(\frac{s}{dB}\Big)^{1/(1+\eta)}\Big)\exp\Big(-\frac{2}{5}\Big(\frac{s}{dB}\Big)^{1/(1+\eta)}\Big)\\
&\leqslant (\beta + \lambda d^3)\big(9Bd^2\big)^{d+1}s^{2d}\Big(\frac{15}{4e}(dB/s)^{1/(1+\eta)}\Big)^{2d}\exp\Big(-\Big(\frac{s}{dB}\Big)^{1/(1+\eta)}\Big)\\
&\leqslant (\beta + \lambda d^3)\big(9Bd^2\big)^{d+1}\Big(\frac{15}{4e}d^2B\Big)^{2d}\exp\Big(-\Big(\frac{s}{dB}\Big)^{1/(1+\eta)}\Big)\\
&\leqslant (\beta + \lambda d^3)\big(32B^3d^6\big)^{d+1}\exp\Big(-\Big(\frac{s}{dB}\Big)^{1/(1+\eta)}\Big).
\end{aligned}$$

We then consider $\xi = 1 - \frac{1}{1+\eta} \in (0,1/2]$ to obtain the constants in (4.2).

5. Discussion. Our convergence results could be extended in several ways:

- While explicit polynomial convergence rates already exist for the Boolean hypercube [32], it would be interesting to obtain improved rates with some form of local condition.
- Our proof technique relies on Fourier series and the characterization of various orders of differentiability using the corresponding orthonormal basis. It could thus be extended to all cases where such tools can be used, such as on the Euclidean hypersphere [7] and beyond [27].

- Almost all the techniques that we used to derive explicit constants can be extended easily to the more general kernel case [26] (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 [17].
- 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.

Appendix A. Computation of convolutions.

Given the function $b : \mathbb{Z} \to \mathbb{R}$ defined as $b(\omega) = (s - |\omega|)_+$, we need to compute the convolution $b{*}b(\omega)$ for $|\omega| \leqslant s$. Since b is even, so is $b{*}b$, and we can thus consider $\omega \in [0, s]$. We want to show that $b{*}b(\omega) = \frac{s(2s^2+1)}{3} - \frac{\omega}{2} - s\omega^2 + \frac{\omega^3}{2}$.

We can split the sum $b{*}b(\omega) = \sum_{i\in\mathbb{Z}} b(i)b(\omega - i)$ as follows:

$$\begin{aligned}
&\sum_{i=\omega-s}^{0}(s+i)(s-\omega+i) + \sum_{i=1}^{\omega}(s-i)(s-\omega+i) + \sum_{i=\omega+1}^{s}(s-i)(s-i+\omega) \\
&= \sum_{i=\omega-s}^{0}\big[i^2 + (2s-\omega)i + s(s-\omega)\big] + \sum_{i=1}^{\omega}\big[-i^2 + \omega i + s(s-\omega)\big] \\
&\quad + \sum_{i=\omega+1}^{s}\big[i^2 - (2s+\omega)i + s(s+\omega)\big].
\end{aligned}$$

Then, using $\sum_{i=1}^{t} i = \frac{t(t+1)}{2} = \frac{1}{2}(t^2+t)$ and $\sum_{i=1}^{t} i^2 = \frac{t(t+1)(2t+1)}{6} = \frac{1}{6}(2t^3+3t^2+t)$, we get

$$\begin{aligned}
&\frac{1}{6}(s-\omega)(s-\omega+1)(2s-2\omega+1) - \frac{1}{2}(s-\omega)(s-\omega+1)(2s-\omega) \\
&\quad + s(s-\omega)(s-\omega+1) - 2\frac{1}{6}\omega(\omega+1)(2\omega+1) + \frac{1}{2}\omega(\omega+1)(2s+2\omega) + s(s-\omega)\omega \\
&\quad + \frac{1}{6}s(s+1)(2s+1) - \frac{1}{2}s(s+1)(2s+\omega) + s(s+\omega)(s-\omega),
\end{aligned}$$

leading to

$$\begin{aligned}
&\frac{1}{6}(s-\omega)(s-\omega+1)(\omega+1-4s) + s(s-\omega)(2s+\omega+1) \\
&\quad + \frac{1}{6}(2s^3+3s^2+s) - \frac{1}{3}(2\omega^3+3\omega^2+\omega) + (\omega^2+\omega)(s+\omega) - \frac{1}{2}(s^2+s)(2s+\omega) \\
&= \Big[\frac{1}{6} - \frac{2}{3} + 1\Big]\omega^3 + \Big[\frac{1}{6}(1-4s-s-1-s) - s - 1 + s + 1\Big]\omega^2 \\
&\quad + \Big[\frac{1}{6}((s+1)(4s-1) - s(1-4s) + s(s+1)) + s^2 - s(2s+1) - \frac{1}{3} + s - \frac{1}{2}(s^2+s)\Big]\omega \\
&\quad + \Big[s(s+1)\frac{1}{6}(1-4s) + s^2(2s+1) + \frac{1}{6}(2s^3+3s^2+s) - s(s^2+s)\Big] \\
&= \frac{1}{2}\omega^3 - s\omega^2 - \frac{1}{2}\omega + \frac{s}{3} + \frac{2}{3}s^3.
\end{aligned}$$

To get (3.2), we then use $\hat{q} * \hat{q}(\omega) = a^2 \prod_{i=1}^{d} \frac{1}{s^2} b{*}b(|\omega_i|)$.

Appendix B. Performance of the spectral relaxation. Given a trigonometric polynomial f of degree $2r$, with $r \leqslant s$, we can represent it in quadratic form in $\varphi(x)$ as defined in (2.2),

$$f(x) = \varphi(x)^\top F \varphi(x) \text{ with } F_{\omega\omega'} = \hat{f}(\omega - \omega') \prod_{i=1}^{d} \Big(1 - \frac{|\omega_i - \omega_i'|}{2s+1}\Big)^{-1},$$

which is the unique Toeplitz representation F for f. We denote by $g : [0,1]^d \to \mathbb{R}$ the function with Fourier series $\hat{g}(\omega) = \hat{f}(\omega) \prod_{i=1}^{d} \big(1 - \frac{|\omega_i|}{2s+1}\big)^{-1}$.

For any $z \in \mathbb{C}^{(2s+1)^d}$ of unit norm, we have

$$\begin{aligned} z^* F z &= \sum_{\|\omega\|_\infty, \|\omega'\|_\infty \leqslant s} z_\omega z_{\omega'}^* \int_{[0,1]^d} g(x) \exp(-2i\pi(\omega - \omega')^\top x) dx \\ &= \int_{[0,1]^d} g(x) \Big| \sum_{\|\omega\|_\infty \leqslant s} z_\omega \exp(-2i\pi\omega^\top x) \Big|^2 dx \\ &\geqslant \inf_{x' \in [0,1]^d} g(x') \cdot \int_{[0,1]^d} \Big| \sum_{\|\omega\|_\infty \leqslant s} z_\omega \exp(-2i\pi\omega^\top x) \Big|^2 dx = \inf_{x' \in [0,1]^d} g(x'). \end{aligned}$$

Thus $\lambda_{\min}(F) \geqslant \inf_{x \in [0,1]^d} g(x)$. We have, moreover,

$$\begin{aligned} \|f - g\|_\infty &\leqslant \sum_{\omega \in \mathbb{Z}^d} |\hat{f}(\omega)| \cdot \Big| \prod_{i=1}^{d} \Big(1 - \frac{|\omega_i|}{2s+1}\Big)^{-1} - 1 \Big| \\ &\leqslant \|f - \bar{f}\|_{\mathrm{F}} \Big[\Big(1 - \frac{2r}{2s+1}\Big)^{-d} - 1 \Big] \sim_{s \to +\infty} \|f - \bar{f}\|_{\mathrm{F}} \cdot \frac{rd}{s}, \end{aligned}$$

which leads to

$$0 \geqslant \lambda_{\min}(F) - f_* \geqslant -\|f - \bar{f}\|_{\mathrm{F}} \Big[\Big(1 - \frac{2r}{2s+1}\Big)^{-d} - 1 \Big] \sim_{s \to +\infty} -\|f - \bar{f}\|_{\mathrm{F}} \cdot \frac{rd}{s}.$$

Appendix C. Proof of Corollary 3.2. For $\|\tau\|_\infty \leqslant 2s$, let $\Omega(\tau)$ denote the set of $(\omega, \omega') \in \mathbb{Z}^d \times \mathbb{Z}^d$ such that $\|\omega\|_\infty \leqslant r$, $\|\omega'\|_\infty \leqslant r$, and $\omega - \omega' = \tau$. We consider the norm Θ on the set of Hermitian matrices of dimension $(2r+1)^d$ defined as

$$\Theta(\Sigma) = (2r+1)^d \sum_{\|\tau\|_\infty \leqslant 2r} \Big(|\Omega(\tau)|^{-1} \sum_{(\omega,\omega') \in \Omega(\tau)} |\Sigma_{\omega\omega'}|^2 \Big)^{1/2}.$$

This norm is constructed so that, for a trigonometric polynomial f of degree less than $2r$ represented by a Hermitian matrix F, then $\|f\|_{\mathrm{F}} = \inf_{Y \in \mathcal{V}_r^\perp} \Theta^*(F+Y)$ (taking into account the normalizing factor defining φ). Thus, we have, applying Theorem 3.1,

$$\begin{aligned} \inf_{\Sigma' \in \mathcal{K}_s} \Theta(\Pi_s^{(r)}(\Sigma - \Sigma')) &= \inf_{\Sigma' \in \mathcal{K}_s} \sup_{\Theta^*(F) \leqslant 1} \operatorname{tr}[F \Pi_s^{(r)}(\Sigma - \Sigma')] \\ &= \sup_{\Theta^*(F) \leqslant 1} \operatorname{tr}[F \Pi_s^{(r)} \Sigma] + \inf_{\Sigma' \in \mathcal{K}_s} \operatorname{tr}[-F \Pi_s^{(r)} \Sigma'] \\ &\leqslant \sup_{\Theta^*(F) \leqslant 1} \operatorname{tr}[F \Pi_s^{(r)} \Sigma] + \inf_{\Sigma' \in \widehat{\mathcal{K}}_s} \operatorname{tr}[-F \Pi_s^{(r)} \Sigma'] + \Big[\Big(1 - \frac{6r^2}{s^2}\Big)^{-d} - 1 \Big] \\ &\leqslant \Big[\Big(1 - \frac{6r^2}{s^2}\Big)^{-d} - 1 \Big], \end{aligned}$$

by selecting $\Sigma' = \Sigma$ in the bound above. The bound using the Frobenius norm is obtained by computing a lower bound on Θ^* as done in Appendix D below (but applying to r instead of s).

Appendix D. Proof of Lemma 4.2.

Proof. Assuming $f_* = 0$ without loss of generality, let f be represented by the Hermitian matrix F, and g by the positive semidefinite Hermitian matrix G, that is, for all $x \in [0,1]^d$, $f(x) = \varphi(x)^* H\varphi(x)$ and $g(x) = \varphi(x)^* G\varphi(x)$. For $\|\tau\|_\infty \leqslant 2s$, if $\Omega(\tau)$ is the set of $(\omega,\omega') \in \mathbb{Z}^d \times \mathbb{Z}^d$ such that $\|\omega\|_\infty \leqslant s$, $\|\omega'\|_\infty \leqslant s$, and $\omega - \omega' = \tau$, then, using that the space $\mathcal{V}_s$ of Hermitian Toeplitz matrices H is characterized by equal values for $H_{\omega\omega'}$ for $(\omega,\omega') \in \Omega(\tau)$ for each τ,

$$\begin{aligned}
\|f-g\|_{\mathrm{F}} &= \sum_{\|\omega\|_\infty \leqslant 2s} |\hat{f}(\omega) - \hat{g}(\omega)| \\
&= \frac{1}{(2s+1)^d} \inf_{Y \in \mathcal{V}_s^\perp} \sum_{\|\tau\|_\infty \leqslant 2s} \Big(|\Omega(\tau)| \sum_{(\omega,\omega')\in\Omega(\tau)} |(F-G+Y)_{\omega\omega'}|^2 \Big)^{1/2} \\
&\geqslant \inf_{Y\in\mathcal{V}_s^\perp} \frac{\min_{\|\tau\|_\infty \leqslant 2s} |\Omega(\tau)|^{1/2}}{(2s+1)^d} \left(\sum_{\|\tau\|_\infty \leqslant 2s} \sum_{(\omega,\omega')\in\Omega(\tau)} |(F-G+Y)_{\omega\omega'}|^2 \right)^{1/2} \\
&= \frac{\sqrt{2}}{(2s+1)^d} \inf_{Y\in\mathcal{V}_s^\perp} \|F-G+Y\|_{\mathrm{Frob}} \geqslant \frac{\sqrt{2}}{(2s+1)^d} \inf_{Y\in\mathcal{V}_s^\perp} \|F-G+Y\|_{\mathrm{op}}.
\end{aligned}$$

We can then take the maximizer above $Y \in \mathcal{V}_s^\perp$, and we have

$$c_*(f,s) \geqslant \lambda_{\min}(F+Y) \geqslant \lambda_{\min}(G) - \|F-G+Y\|_{\mathrm{op}} \geqslant 0 - \frac{(2s+1)^d}{\sqrt{2}} \varepsilon'(s,f),$$

where $\|\cdot\|_{\mathrm{Frob}}$ denotes the Frobenius norm and $\|\cdot\|_{\mathrm{op}}$ the largest singular value. ∎

Appendix E. Proof of generic lemmas about derivatives.

In this appendix, we prove lemmas about derivatives and Fourier decays.

E.1. Proof of Lemma 4.8.

Proof. We will show a bound on the Fourier series of g of the form

$$|\hat{g}(\omega)| \leqslant D(k) \frac{1}{(2+\|\omega\|_1)^k}, \tag{E.1}$$

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

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

by counting the number of $\omega \in \mathbb{Z}^d$ such that $\|\omega\|_1 = t$. This 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 $n^n \leqslant n! e^{n-1}$ applied to $n = d-1$)

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

This leads to

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

Proof of (E.1). To obtain (E.1), we need to be able to bound the product $|\hat{g}(\omega)||\omega_j|^{\alpha_1}\cdots|\omega_d|^{\alpha_d}$ for any α such that $\alpha_1+\cdots+\alpha_d=k$. For this, we need uniform bounds on all partial derivatives, which we need to obtain from bounds on $\nabla^k g(x)[\delta,\ldots,\delta]$ for all δ and k. From the polarization Lemma E.1, we have

$$|\nabla^k g(x)[\delta_1,\ldots,\delta_k]|\leqslant\frac{1}{k!}\left(\sum_{i=1}^k\|\delta_i\|_1\right)^k\cdot\|\nabla^k g\|_\infty\leqslant\frac{1}{k!}\left(\sum_{i=1}^k\|\delta_i\|_1\right)^k C\cdot B^k\cdot k!\cdot\kappa(k),$$

by definition of $\|\nabla^k g\|_\infty$ and because of the assumptions of the lemma. For any α such that $\alpha_1+\cdots+\alpha_d=k$, the partial derivative $\partial_\alpha g(x)=\frac{\partial^k g}{\partial x_1^{\alpha_1}\cdots x_d^{\alpha_d}}(x)$ can be written as $\frac{\partial^k g}{\partial x_{j_1}\cdots\partial x_{j_k}}(x)$ for $j_1,\ldots,j_k\in\{1,\ldots,d\}$. Thus, applying the inequality above with δ_i the indicator vector of the set $\{j_i\}$ for each $i\in\{1,\ldots,k\}$, we get

$$|\partial_\alpha g(x)|\leqslant|\nabla^k g(x)[\delta_1,\ldots,\delta_k]|\leqslant C\cdot B^k\cdot k^k\cdot\kappa(k).$$

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

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

This leads to $D(k)\leqslant C\big(2+\frac{dBk}{2\pi}\big)^k\kappa(k)$, and thus the desired result. ∎

LEMMA E.1 (polarization). *Let $u:E^m\to\mathbb{R}$ be a symmetric m-multilinear form on some normed vector space E. Then for all $z_1,\ldots,z_m\in E$, we have*

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

Proof. We use the polarization identity for the m-multilinear form $u:E^m\to E$ and its diagonal $\tilde{u}:z\mapsto u(z,\ldots,z)$ (see [33, eq. (A.4)]),

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

which leads to

$$
\begin{aligned}
|u(z_1,\dots,z_m)| &\leqslant \frac{1}{2^m m!}\sum_{\varepsilon\in\{0,1\}^m}\left(\sum_{i=1}^m \|z_i\|_1\right)^m \sup_{\|z\|_1\leqslant 1}|\tilde u(z)| \\
&= \frac{1}{m!}\left(\sum_{i=1}^m \|z_i\|_1\right)^m \sup_{\|z\|_1\leqslant 1}|\tilde u(z)|,
\end{aligned}
$$

which is the desired result. ∎

E.2. Proof of Lemma 4.7.

Proof. Using the Leibniz formula applied to $\varphi_1(t)=h_1(x+t\delta)$, $\varphi_2(t)=h_2(x+t\delta)$, we have

$$
\begin{aligned}
(\varphi_1\varphi_2)^{(m)}(0) &= \sum_{i=0}^m \binom{m}{i}\varphi_1^{(i)}(0)\varphi_2^{(m-i)}(0) \\
&\leqslant C_1C_2\|\delta\|_1^m \sum_{i=0}^m \binom{m}{i} B_1^i B_2^{m-i} i!(m-i)!\kappa_1(i)\kappa_2(m-i) \\
&\leqslant C_1C_2\kappa_1(m)\kappa_2(m)\|\delta\|_1^m m! \sum_{i=0}^m B_1^i B_2^{m-i} \\
&\leqslant C_1C_2\kappa_1(m)\kappa_2(m)\|\delta\|_1^m (m+1)!\max\{B_1,B_2\}^m.
\end{aligned}
$$
∎

Acknowledgments. We thank Monique Laurent, Jean-Bernard Lasserre, and Milan Korda for helpful discussions about this work. The comments of the anonymous reviewers were greatly appreciated.

REFERENCES

[1] L. Baldi and B. Mourrain, *On the effective Putinar's Positivstellensatz and moment approximation*, Math. Program., 200 (2023), pp. 71–103.
[2] L. Baldi, B. Mourrain, and A. Parusinski, *On Lojasiewicz Inequalities and the Effective Putinar's Positivstellensatz*, Technical report, https://arxiv.org/abs/2212.09551, 2022.
[3] C. A. Charalambides, *Enumerative Combinatorics*, Chapman and Hall, 2002.
[4] E. De Klerk and M. Laurent, *On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems*, SIAM J. Optim., 21 (2011), pp. 824–832, https://doi.org/10.1137/100814147.
[5] P. Del Moral and A. Niclas, *A Taylor expansion of the square root matrix function*, J. Math. Anal. Appl., 465 (2018), pp. 259–266.
[6] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*, Signals Commun. Technol. 103, Springer, 2007.
[7] K. Fang and H. Fawzi, *The sum-of-squares hierarchy on the sphere and applications in quantum information theory*, Math. Program., 190 (2021), pp. 331–360.
[8] L. Fejér, *Über trigonometrische Polynome*, J. Reine Angew. Math., 146 (1916), pp. 55–82.
[9] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz, *An interior-point method for semidefinite programming*, SIAM J. Optim., 6 (1996), pp. 342–361, https://doi.org/10.1137/0806020.
[10] D. Henrion, M. Korda, and J.-B. Lasserre, *The Moment-SOS Hierarchy: Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs*, World Scientific, 2020.
[11] A. Israel, *The Eigenvalue Distribution of Time-Frequency Localization Operators*, Technical report, https://arxiv.org/abs/1502.04404, 2015.
[12] Y. Katznelson, *An Introduction to Harmonic Analysis*, Cambridge University Press, 2004.
[13] J.-B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM J. Optim., 11 (2001), pp. 796–817, https://doi.org/10.1137/S1052623400366802.

[14] J.-B. Lasserre, *Convexity in semialgebraic geometry and polynomial optimization*, SIAM J. Optim., 19 (2009), pp. 1995–2014, https://doi.org/10.1137/080728214.
[15] J.-B. Lasserre, *Moments, Positive Polynomials and Their Applications*, World Scientific, 2010.
[16] M. Laurent and L. Slot, *An effective version of Schmüdgen's Positivstellensatz for the hypercube*, Optim. Lett., 17 (2023), pp. 515–530.
[17] U. Marteau-Ferey, F. Bach, and A. Rudi, *Second Order Conditions to Decompose Smooth Functions as Sums of Squares*, Technical report, https://arxiv.org/abs/2202.13729, 2020.
[18] A. Megretski, *Positivity of trigonometric polynomials*, in 42nd IEEE International Conference on Decision and Control, Vol. 4, 2003, pp. 3814–3817.
[19] A. Naftalovich and M. Schreiber, *Trigonometric polynomials and sums of squares*, in Number Theory, Springer, 1985, pp. 225–238.
[20] Y. Nesterov, *Squared functional systems and optimization problems*, in High Performance Optimization, Springer, 2000, pp. 405–440.
[21] J. Nie, *Optimality conditions and finite convergence of Lasserre's hierarchy*, Math. Program., 146 (2014), pp. 97–121.
[22] P. A. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Math. Program., 96 (2003), pp. 293–320.
[23] M. Putinar, *Sur la complexification du problème des moments*, C. R. Acad. Sci. Paris Sér. 1 Math., 314 (1992), pp. 743–745.
[24] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana Univ. Math. J., 42 (1993), pp. 969–984.
[25] F. Riesz, *Uber ein Problem des Herrn Carathéodory*, J. Reine Angew. Math., 1916 (1916), pp. 83–87.
[26] A. Rudi, U. Marteau-Ferey, and F. Bach, *Finding Global Minima via Kernel Approximations*, Technical report, https://arxiv.org/abs/2012.11978, 2020.
[27] W. Rudin, *Fourier Analysis on Groups*, Courier Dover, 2017.
[28] C. Scheiderer, *Sums of squares on real algebraic surfaces*, Manuscripta Math., 119 (2006), pp. 395–410.
[29] K. Schmüdgen, *The Moment Problem*, Springer, 2017.
[30] M. Schweighofer, *On the complexity of Schmüdgen's Positivstellensatz*, J. Complex., 20 (2004), pp. 529–543.
[31] L. Slot, *Sum-of-squares hierarchies for polynomial optimization and the Christoffel–Darboux kernel*, SIAM J. Optim., 32 (2022), pp. 2612–2635, https://doi.org/10.1137/21M1458338.
[32] L. Slot and M. Laurent, *Sum-of-squares hierarchies for binary polynomial optimization*, Math. Program., 197 (2023), pp. 621–660.
[33] E. G. F. Thomas, *A polarization identity for multilinear maps*, Indagationes Math., 25 (2014), pp. 468–474.
[34] B. Woodworth, F. Bach, and A. Rudi, *Non-convex optimization with certificates and fast rates through kernel sums of squares*, in Proceedings of the Conference on Learning Theory, 2022.