SIAM J. OPTIM.
Vol. 34, No. 1, pp. 616–641

SECOND ORDER CONDITIONS TO DECOMPOSE SMOOTH FUNCTIONS AS SUMS OF SQUARES*

ULYSSE MARTEAU-FEREY†, FRANCIS BACH†, AND ALESSANDRO RUDI†

Abstract. We consider the problem of decomposing a regular nonnegative function as a sum of squares of functions which preserve some form of regularity. In the same way as decomposing nonnegative polynomials as sum of squares of polynomials allows one to derive methods in order to solve global optimization problems on polynomials, decomposing a regular function as a sum of squares allows one to derive methods to solve global optimization problems on more general functions. As the regularity of the functions in the sum of squares decomposition is a key indicator in analyzing the convergence and speed of convergence of optimization methods, it is important to have theoretical results guaranteeing such a regularity. In this work, we show second order sufficient conditions in order for a p times continuously differentiable nonnegative function to be a sum of squares of $p-2$ differentiable functions. The main hypothesis is that, locally, the function grows quadratically in directions which are orthogonal to its set of zeros. The novelty of this result, compared to previous works is that it allows sets of zeros which are continuous as opposed to discrete, and also applies to manifolds as opposed to open sets of $\mathbb{R}^d$. This has applications in problems where manifolds of minimizers or zeros typically appear, such as in optimal transport, and for minimizing functions defined on manifolds.

Key words. nonnegative functions, manifolds, sum of squares, global optimization, second order

MSC codes. 90C26, 58C05, 53A99, 90C56, 11E25

DOI. 10.1137/22M1480914

1. Introduction.

The relationship between nonnegative functions and functions decomposable as sums of squares is a fundamental question in both theoretical and applied mathematics. From a theoretical viewpoint, the decomposability of a nonnegative function in terms of sum of squares is the basis of important theoretical objects and properties: quadratic modules [18] in algebraic geometry, regularizing operators such as Laplacians or sub-Laplacians in (sub-)Riemannian geometry [6, 12], nonnegative symbols in pseudodifferential calculus [13, 34]. From an applied viewpoint, representing a nonnegative function in terms of sum of squares simplifies the analysis of probability representations and optimization problems [16, 19]. Restricting oneself to the case of nonnegative polynomials, this has been applied to global optimization and generalized methods of moments [10, 16]. In a more recent line of work, the decomposition of nonnegative p-times differentiable functions allowed us to derive simple and fast optimization algorithms in the context of global optimization [28], the Kantorovich problem in optimal transport [35], and some formulations of optimal control [5]. Moreover, it allowed us to obtain effective and concise representations for probability densities, with applications in probabilistic inference, sampling, and machine learning [20, 27].

*Received by the editors February 25, 2022; accepted for publication (in revised form) August 6, 2023; published electronically February 8, 2024.
https://doi.org/10.1137/22M1480914
Funding: This work was funded in part by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001(PRAIRIE 3IA Institute). We also acknowledge support from the European Research Council (grants SEQUOIA 724063 and REAL 947908), and support by grants from Région Ile-de-France.
†Inria/PSL Research University, Paris, France (ulysse.marteau@gmail.com, francis.bach@inria.fr, alessandro.rudi@gmail.com).

1.1. Motivation: Widening the scope of optimization methods based on analytical sum of squares decompositions. The main motivation of this work is to extend the class of functions under which optimization methods based on sum of squares decompositions [4, 28] have fast convergence rates. For simplicity, we will focus on the method introduced by [28], but the same reasoning can be applied to any method whose convergence properties depend on a regular sum of squares decomposition of the target function.

Regular sum of squares decompositions imply fast convergence of optimization methods. We will denote the method introduced by [28] as Algorithm 1 with AnalyticalSoSOpt. In a nutshell, given a function f of class C^p defined on a bounded d-dimensional box in $\mathbb{R}^d$, and n random points sampled from that box, AnalyticalSoSOpt returns $\widehat{x} = \text{AnalyticalSoSOpt}(f, n)$, an approximation of the minimizer of f. Denote with $f_\star$ the minimum of f.

The original work shows that AnalyticalSoSOpt(f,n) has a near optimal convergence rate in n (close to the optimal $n^{-p/d}$ rate, [23]), in polynomial time in n, under the assumption that f can be decomposed as a sum of squares of functions of class C^{p-2} for some $p \geq 2$.[1] We have simplified the result in the following proposition.

PROPOSITION 1.1 (see Theorem 3 from [28] and its implications for the proof of Theorem 6). *If $f - f_\star$ is a sum of squares of functions of class C^{p-2},* AnalyticalSoSOpt *satisfies*

$$f(\text{AnalyticalSoSOpt}(f,n)) - f_\star \leq C_{d,f}\, n^{-p/d+1/2+3/d}. \tag{1.1}$$

Understanding under which conditions a function f can be decomposed as its minimum plus a sum of squares of functions of class C^{p-2} is therefore crucial to understand for which functions f the algorithm AnalyticalSoSOpt converges at the fast rate given in (1.1) (by fast, we mean that the regularity p allows it to go faster than the $n^{-1/d}$ rate).

In this paper, we therefore provide interpretable sufficient conditions for nonnegative functions f to be decomposable as sum of squares of class C^{p-2}, which directly implies the fast convergence of AnalyticalSoSOpt when applied to these functions (plus their minimum) by virtue of Proposition 1.1. Our main theorem, summarized in Theorem 1.3, is a *nichtnegativstellensatz* for regular functions.

Preserving regularity in the analytical sum of squares decomposition. The main difficulty in analytical sum of squares decompositions is not to find a sum of squares decomposition in itself, but to find one where the different functions in the decomposition preserve some regularity properties of the original function. Indeed, if no constraints are added on the regularity of the decomposition, this is a trivial problem as writing $f = (\sqrt{f})^2$ would offer an immediate solution. However, regularity is not necessarily preserved when taking the square root: the map $(x,y) \mapsto x^2 + y^2$ is smooth and a sum of smooth squares, but its square root is not differentiable at $(0,0)$.

1.2. Precise setting and results. In this work, we will concentrate on decomposing nonnegative functions f of class C^p for $p \in \mathbb{N} \cup \{\infty\}$, $p \geq 2$, on open sets of $\mathbb{R}^d$ (or any d-dimensional manifolds M, but we keep to $\mathbb{R}^d$ in this introduction). We will

[1] We take $p \geq 2$ and use $p-2$ in the decomposition to be coherent with the results presented in the paper, in which f is of class C^p and the decomposition of class C^{p-2}.

show that under a certain condition on the set of zeros $\mathcal{Z}$ of f, it can be decomposed as[2]

$$f = \sum_{i \in I} f_i^2, \qquad f_i \in C^{p-2}, \tag{1.2}$$

where (f_i) is an at most countable family and has locally finite support. Two elements are important in (1.2): the locally finite aspect and the regularity of the functions f_i, i.e., $p-2$. This is a consequence of the fact that we will consider *second order sufficient conditions,* hence the loss of two derivatives.

1.2.1. Intuition and previous results. We start by proving that this decomposition holds locally in a neighborhood of any $x_0 \in \mathbb{R}^d$. It is then possible to invoke a result to "glue" the local decompositions together; we develop the tools to do so in subsection 3.2 (this is one of the key differences between results for polynomials and results for functions). For any fixed $x_0 \in \mathbb{R}^d$, if $f(x_0) > 0$, then $f_1 := \sqrt{f}$ is well defined and of class C^p around x_0, and so (1.2) holds locally around x_0 since $f = f_1^2$. The crux of the problem is to determine whether f can be decomposed as a sum of squares around a point in the set of zeros $\mathcal{Z}$ of f, i.e., the set of points x such that $f(x) = 0$. Since f is nonnegative, all such points are necessarily minimizers of f, hence the following *necessary second-order condition*:

$$\forall x_0 \in \mathcal{Z},\ \nabla f(x_0) = 0,\ \nabla^2 f(x_0) \succeq 0. \tag{1.3}$$

Around any $x_0 \in \mathcal{Z}$, f can be approximated by a parabola since the eigenvalues of $\nabla^2 f(x_0)$ are nonnegative: $f(x) = x^\top \nabla^2 f(x_0) x + o(\|x\|^2)$ using a Taylor expansion. Since any parabola can be written as the sum of at most d squares of linear functions (just write the eigendecomposition of $\nabla^2 f(x_0)$), we see that up to the $o(\|x\|^2)$ factor, we can indeed write f as a sum of at most d squares around x_0. The whole difficulty of the following results is to go beyond this $o(\|x\|^2)$ approximation and have an exact decomposition, using the Taylor expansion with integral remainder.

In the case where $\nabla^2 f(x_0) \succ 0$, i.e., the Hessian has strictly positive eigenvalues, this decomposition can be made exact. We will call this condition the *strict Hessian condition* (SHC) at x_0. This result exists in recent work: it is a particular case of Theorem 2 of [28], applied to the set $\mathcal{H} = C^{p-2}$. Precisely, it states the following.

THEOREM 1.2 (Theorem 2 of [28]). *Let f be a nonnegative function of class C^p for $p \geq 2$, $p \in \mathbb{N} \cup \{\infty\}$, and assume that the zeros $\mathcal{Z}$ of f satisfy the strict Hessian condition*

$$\forall x_0 \in \mathcal{Z},\ \nabla^2 f(x_0) \succ 0. \tag{1.4}$$

If f has a finite number $m = |\mathcal{Z}|$ of zeros, then f satisfies (1.2) *with $dm+1$ functions f_i.*

This situation is illustrated on the left hand side of Figure 1, where the Hessian is positive definite at all four zeros of f and hence satisfies the SHC: by Theorem 1.2, it can be decomposed as a sum of squares. It is not the case on the right-hand side, where there is a continuous subspace of zeros; in that case, f does not satisfy the SHC.

[2] If $p = \infty$, we take the convention $p - k = \infty$ for any fixed $k \in \mathbb{N}$

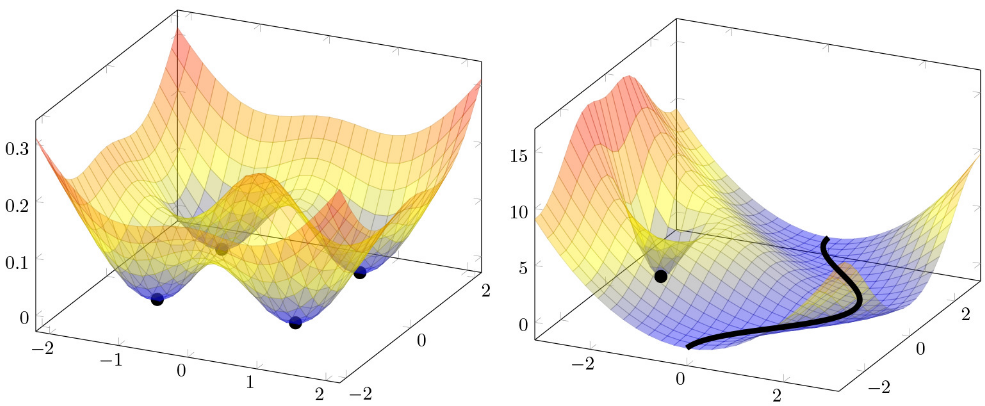

FIG. 1. *Plots of functions $z = f(x,y)$, where the zeros of f are highlighted in black. left: f satisfies the SHC, right: f satisfies the normal Hessian condition but not the SHC.*

1.2.2. Main contribution: A nichtnegativstellensatz. While the SHC condition (1.4) already offers a nice result in Theorem 1.2, there is a big gap with (1.3). Previous results in the literature show that satisfying (1.3) is not sufficient to be decomposed as a sum of squares of C^{p-2} functions as soon as the dimension d is greater than 3 (see Theorem 1.5 in the background section for more details). On the other hand, (1.4) is very restrictive and implies that the set $\mathcal{Z}$ of zeros is discrete. In some situations such as that of [35], however, the set of zeros has a natural structure, which can be a submanifold of $\mathbb{R}^d$. In this paper, we show that if the set $\mathcal{Z}$ of zeros is a submanifold of $\mathbb{R}^d$ such that the Hessian of f along this manifold is positive along all directions which are not tangent to $\mathcal{Z}$, then (1.2) still holds. This is the case for the function depicted in the right-hand side of Figure 1, and illustrates the difference between previous works and our contributions. More formally, we prove the following result.

THEOREM 1.3. *Let f be a nonnegative function of class C^p, where $p \in \mathbb{N} \cup \{\infty\}$, $p \geq 2$, and let $\mathcal{Z}$ denote the set of zeros of f. If $\mathcal{Z}$ is a submanifold of $\mathbb{R}^d$ of class C^1 such that*

$$\forall x_0 \in \mathcal{Z} \;\; \forall h \in \mathbb{R}^d \setminus T_{x_0}\mathcal{Z},\; h^\top \nabla^2 f(x_0) h > 0, \tag{1.5}$$

then f satisfies (1.2) *and $\mathcal{Z}$ is of class C^{p-1}. Here, $T_{x_0}\mathcal{Z}$ denotes the tangent space to $\mathcal{Z}$ at x_0, which is a vector subspace of $\mathbb{R}^d$.*

This theorem is proved as Theorem 2.9 in section 2, and the assumption (1.5) will be referred to as the normal Hessian condition (or NHC). The NHC assumption encompasses that of the SHC assumption of Theorem 1.2; in that case, the results presented in this paper make the result tighter by removing the assumption that $\mathcal{Z}$ is finite and by needing only $d+1$ squares to represent the function, and not $d|\mathcal{Z}|+1$ (see the full version of Theorem 2.9).

We want to emphasize that the NHC is different from the classical nonlinear optimization second order sufficient conditions. These conditions usually imply that the function grows at least quadratically in all directions pointing inside the domain of definition of the function f (which would include the set $T_x\mathcal{Z}$ in our setting; see [17]), and which imply that the set of minimizers is discrete. On the contrary, the goal of this work is to go beyond isolated minima.

The proof techniques used to prove this theorem differ from the proof of [28] and use tools from differential geometry and Morse theory. In particular, the proof extends naturally to functions defined on d-dimensional manifolds, which is the object of section 3 and Theorem 3.9. This opens the way to new problems, which are more naturally defined on standard manifolds like the d-dimensional sphere S^d or the d-dimensional torus $\mathbb{T}^d \approx (S^1)^d$.

1.2.3. Consequence: Convergence guarantees for analytic sum of squares algorithms. As explained in subsection 1.1, having such a nichtnegativstellensatz implies fast convergence of algorithms such as AnalyticalSoSOpt. In [28], convergence of AnalyticalSoSOpt was established for functions satisfying the SHC. Using Theorem 1.3, we actually show that satisfying the NHC is sufficient to guarantee the same speed of convergence. This therefore extends the scope of AnalyticalSoSOpt to a broader class of functions, which contains (i) functions defined on a manifold and not only on a box and (ii) functions with a continuous set of zeros, which was not the case in the original result by [28]. In subsection 2.3, we give examples of interesting functions which satisfy the NHC and verify that AnalyticalSoSOpt indeed returns a good approximation of the minimizer on toy examples in different dimensions.

1.3. Background.

1.3.1. Regular decompositions of functions. The problem of decomposing C^p functions as sums of squares has appeared in the context of symbolic calculus, in the proof of the Fefferman–Phong inequality, which is an important regularity result for partial differential operators (see [9] for the original article and [6] for the link with sum of squares decompositions, as well as [34]). In this context, the following result is proved (with $C^{k,1}_{loc}$ denoting the set of k times differentiable functions with locally Lispchitz kth derivative).

THEOREM 1.4 (Fefferman–Phong [9], Theorem 1.1 of [8]). *Let Ω be an open set of $\mathbb{R}^d$, $d \geq 1$, and $f \in C^{3,1}_{loc}(\Omega)$ be a nonnegative function. Then f can be written as a finite sum of squares of $C^{1,1}_{loc}(\Omega)$ functions.*

In the context of preserving regularity, a natural question which arises is whether increasing the regularity of f can increase the regularity of the functions in a sum of squares decomposition. In [7, 8], it is shown that the general answer (under no further assumptions) is negative. More precisely, if f is a function defined on a neighborhood of 0, a local decomposition of f around 0 of class $\mathcal{C}$ is a finite family $(f_i)_{i\in I}$ of functions of class $\mathcal{C}$ defined on an open neighborhood U of 0 such that $\sum_{i\in I} f_i^2 = f$ on U.

THEOREM 1.5 (Theorem 2.1 of [8]). *In all the following cases, there exists $f \in C^\infty$ defined on an open neighborhood of 0 in $\mathbb{R}^d$ such that the following hold:*
- *if $d \geq 4$, f has no local decomposition of class C^2;*
- *if $d = 3$, f has no local decomposition of class C^3.*

The case $d = 1$ is explored in [7]: it is shown in Theorem 1 that if f is of class C^{2m} for m finite, then f can be written as the sum of squares of two functions of class C^m. Moreover, this is shown to be tight: there exists a function $f \in C^{2m}$ with no local decomposition as a sum of squares of functions of class C^{m+k}, for $k \geq 1$. The case $d = 2$ has been explored less in the literature (some results exist when dealing with flat minima; see, for example, Theorem 2 of [7]). To summarize, these results show that without additional assumptions, as soon as the dimension is greater than 3, inheriting the C^p regularity properties of the function f in the sum of squares decomposition is not possible in a satisfactory way, and motivates the introduction of additional geometric assumptions such as the NHC.

1.3.2. Polynomials and analytic functions. Decomposing nonnegative polynomial, and later analytic, functions as sums of squares has been related to important problems in algebraic geometry during the 20th century. In 1927, on his way to the resolution of Hilbert's 17th problem, Artin [3] proved that any nonnegative polynomial is a sum of squares of rational functions (that is, formal fractions of polynomials $P(x)/Q(x)$). Similarly, it has been established that under certain conditions on its null set, a semidefinite (i.e., nonnegative) analytic function on $\mathbb{R}^n$ can be expressed as a potentially infinite sum of squares of meromorphic functions, which are, in a sense, the analytic analog of rational functions (see [1, 29] for an introduction to those results). Moreover, Hilbert had earlier proved that there exist nonnegative polynomials which cannot be written as sum of squares of polynomials in [11] (for more than 3 variables and with degree at least 6 for example), and it can also be established that certain semidefinite analytic functions cannot be expressed as a sum of squares of analytic functions (again, see [1]). In algebraic geometry, an important question is to understand under which sufficient conditions a polynomial P which is positive on an algebraic set, i.e., defined by polynomial inequalities of the form $Q_i \geq 0$ for polynomials Q_i, can be written in the form $P = P_0 + \sum_{i=1}^{N} P_i Q_i$, where the P_i are sum of squares. The theoretical literature regroups these results under the name "positivstellensatz" (or nichtnegativstellensatz if P is assumed to be nonnegative). The most often seen in the sum of squares optmization literature are the Stengle [33], Schmügden [31], and Putinar [26] positivstellensätzen. If these algebraic geometry considerations seem far from applications and from decomposing smooth functions as sums of squares (indeed, polynomials are much more rigid than smooth functions) at first glance, they are actually related in two ways. First, as smooth functions can be locally approximated by polynomials, results on polynomials give a good intuition of the difficulties one can encounter at the local level when decomposing a function as a sum of squares (see [7, 8]). Second, the certificates given by positivstellensatz on the decomposability of certain nonnegative polynomials can be algorithmically checked in some cases, using semidefinite programming. This has paved the way to so-called sum of squares hierarchies, and optimization of polynomial objective functions with polynomial constraints. These have been developed by Lasserre [16] (based on the Putinar postivstellensatz [26]) and Parrilo [24] (based on the Stengle and Schmügden positivstellensatz [31, 33]). Using these theoretical results, they can provide certificates of lower bounds for certain optimization problems (or upper bound in the dual "moment problem"; see [16]). Moreover, to have more interpretable results for these more applied settings, theses works have motivated more practical positivstellensätzen or nichtnegativstellensätzen, like that in [18], which provides a condition for writing a polynomial with a finite set of zeros as a sum of squares. This result has in turned been used in [22] in order to show the finite convergence of the Lasserre hierarchy under second order conditions which are analoguous to those of constrained nonlinear optimization [17] (these are called boundary Hessian conditions and become the SHC in the unconstrained case).

1.3.3. Analytical sum of squares algorithms and comparison with polynomial optimization. In the same spirit as the polynomial hierarchies but for regular functions, recent works [4, 19, 27, 28] have developed models and methods based on sum of squares to optimize regular functions (such as AnalyticalSoSOpt, defined in subsection 1.1). The computational properties of these methods such as (1.1) are based on the fact that C^pfunctions can be well approximated by functions of the form $\sum_i \alpha_i k(\cdot, x_i)$, where k is a so-called positive definite kernel [2] and can be adapted to the regularity.

Let us very briefly describe AnalyticalSoSOpt from [28]. Let f be a C^p function and Ω its domain (we can assume it is a box). As in the case of polynomial optimization (see [16]), the starting point of this method is to cast the optimization problem as a convex problem: finding the supremum of lower bounds

$$f_* = \min_{x\in\Omega} f(x) = \sup_{c\in\mathbb{R}} c \;\text{ s.t. }\; f(x) - c \geq 0,\; x\in\Omega \tag{1.6}$$

$$\geq f_s = \sup_{c\in\mathbb{R}} c \;\text{ s.t. }\; f - c = \sum_{i\in I} g_i^2,\; g_i \in C^{p-2}(\Omega),\; I \text{ finite.} \tag{1.7}$$

The constraint $f - c = \sum_{i\in I} g_i^2$ is further discretized and enforced on a discrete set of randomly sampled points $(x_1, \ldots, x_n)$, and the problem is regularized with a trace norm to avoid interpolation. The constraints that $g_i \in C^{p-2}(\Omega)$ is relaxed to g_i belonging to a reproducing kernel Hilbert space with kernel k containing these functions (such as a Sobolev space), in order to cast the problem as an n dimensional semidefinite program using the representer theorem from [19]. The approximation of the minimizer $\widehat{x} = \text{AnalyticalSoSOpt}(f, n)$ is obtained by using certain properties of the dual variable, described in [28].

Compared to polynomial optimization methods, these "analytical" methods can tackle functions beyond polynomials. This comes at the cost of losing the exactness and a posteriori guarantees given by polynomial optimization methods (trading it off for convergence rates in the number of evaluation points n). However, they are not designed to tackle more polynomials than pure polynomial methods. Examples of polynomials which are not sums of squares of polynomials are typically homogeneous of degree greater than 4, and hence cannot satisfy second order assumptions such as the NHC. Moreover, the set of zeros of nonnegative polynomials can have algebraic variety structures (such as the set $x^2 = y^2 = z^2$ for the Robinson polynomial), which is out of scope for the NHC as these varieties are not differentiable manifolds. This analytical sum of squares framework must therefore be seen as a parallel rather than competing line of work to polynomial optimization.

Organization of the work. In section 2, we formalize the different notions needed to state Theorem 1.3 in the case of nonnegative functions defined on open sets of $\mathbb{R}^d$. In particular, we start by presenting a local decomposition in Theorem 2.3, which will be the cornerstone of the work. In section 3, we extend Theorem 1.3 to the manifold setting, and detail the procedure in which we glue local decompositions into a global one, using traditional tools from differential geometry. In section 4, we formally prove Theorem 2.3. We finish by a discussion on the result presented in this paper, as well as possible extensions in section 5.

2. Decomposition as sums of squares given second order conditions (Euclidean case).

In this section, we present our results on decomposing a C^p function f as a sum of squares of C^{p-2} functions on open sets of $\mathbb{R}^d$. After some notations and definitions on submanifolds, we present the cornerstone technical result of this paper in subsection 2.1, Theorem 2.3, as well as a sketch of its proof. In subsection 2.2, we present Theorem 2.9, which shows that a function can be decomposed as a locally finite sum of squares of functions defined on Ω under the NHC. Finally, in subsection 2.3, we present some classes of functions satisfying the NHC and show that AnalyticalSoSOpt indeed optimizes them correctly.

Definitions and notations. In general, given two topological sets M and N as well as $x_0 \in M$ and $y_0 \in N$, we will say that $\phi : (x_0, M) \to (y_0, N)$ is a local map satisfying a property (P) if there exists an open neighborhood U of x_0 in M such that $\phi : U \to N$ is well defined, satisfies $\phi(x_0) = y_0$, and property (P). We will say that $\phi : U \subset \mathbb{R}^d \to \mathbb{R}^e$ defined on an open set U is of class C^k if it is k times differentiable, and its derivatives of order k are continuous. For any function $\phi : (x, \mathbb{R}^d) \to \mathbb{R}^e$ of class C^1, we denote with $d\phi(x)$ its differential at x. It is an element of $\mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e)$, the set of linear maps from $\mathbb{R}^d$ to $\mathbb{R}^e$. We will write $d\phi(x)\xi$ or $d\phi(x)[\xi]$, the evaluation of $d\phi(x)$ at ξ. The Jacobian of ϕ at x is the matrix $J_\phi(x) \in \mathbb{R}^{e\times d}$ which is the matrix of $d\phi(x)$ in the canonical bases. Writing the coordinates of ϕ: $\phi = (\phi^1, \dots, \phi^e)$, we have $[J_\phi]_{ij} = \frac{\partial \phi^i}{\partial x^j}(x)$.

For a comprehensive introduction to submanifolds, see Chapter 1 of [15], section 2.2 of [25] (in French), or [32]. Intuitively, a d_0 dimensional submanifold N is a subset of $\mathbb{R}^d$ which can be locally parametrized by $\mathbb{R}^{d_0}$. More formally, a map $\phi : U \to \mathbb{R}^d$ defined on an open neighborhood U of 0 in $\mathbb{R}^{d_0}$ is said to be a local parametrization of N around x_0 of class C^k for $k \geq 1$ if ϕ is of class C^k, and if there exists an open set $V \subset \mathbb{R}^d$ such that the following conditions are satisfied:

(i) $\phi(0) = x_0$, $\phi(U) = N \cap V$, and $\phi : U \to \phi(U)$ is a homeomorphism, i.e., it is bijective and has a continuous inverse;

(ii) its differential at 0 is injective (one to one), i.e., $d\phi(t_0) \in \mathrm{Hom}(\mathbb{R}^{d_0}, \mathbb{R}^d)$ is injective.

N is said to be a submanifold of $\mathbb{R}^d$ and of class C^k if there exists a local parametrization ϕ of class C^k around each point $x \in N$. On any connected component of N, the dimension is well defined (it must be the same at every point). Similarly, the subspace $T_x N := d\phi(x)\mathbb{R}^{d_0}$ is uniquely defined for all x and is called the tangent space to N at x (see Figures 2 and 3).

Example 2.1. All open sets of $\mathbb{R}^d$ are submanifolds of $\mathbb{R}^d$. The d-dimensional sphere S^d is a submanifold of $\mathbb{R}^{d+1}$. S^1 is represented in the left hand side (l.h.s.) of Figure 2 and S^2 in the l.h.s. of Figure 3. Given a submanifold N of $\mathbb{R}^d$, the intersection of N with any open set of $\mathbb{R}^d$ is a submanifold of $\mathbb{R}^d$.

2.1. Local decomposition as a sum of squares. In this section, f will always denote a nonnegative function defined on an open set of $\mathbb{R}^d$. We will assume that f is of class C^p for $p \in \mathbb{N} \cup \{\infty\}$, $p \geq 2$. We will also denote with $\mathcal{Z}$ the set of zeros of f. In this section, we will make local assumptions on the Hessian of f at points $x \in \mathcal{Z}$ such that the function f can be decomposed as a sum of squares locally around x.

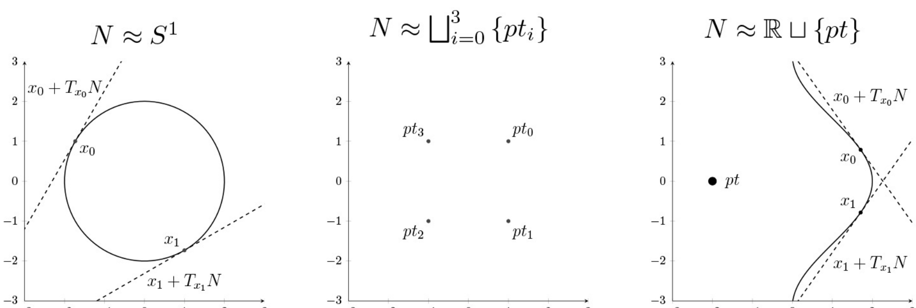

FIG. 2. *Examples of submanifolds of* $\mathbb{R}^2$*; points are denoted with pt.* ***Left:*** *connected submanifold of dimension* 1 *(a circle).* ***Center:*** *a submanifold of* 4 *connected components which are all points, i.e., of dimension* 0 *(their tangent space is not represented since it is reduced to* $\{0\}$*).* ***Right:*** *a submanifold of two connected components, one point pt of dimension* 0 *and one of dimension* 1*.*

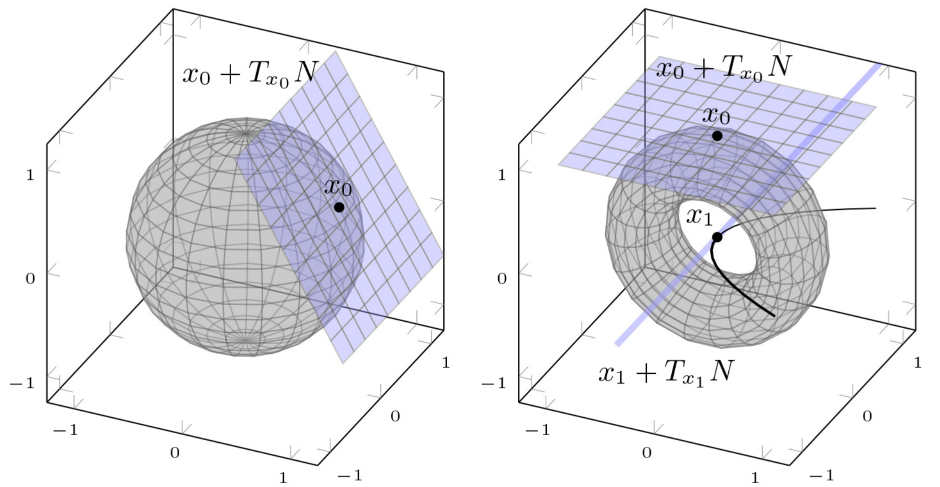

FIG. 3. *Two examples of submanifolds of $\mathbb{R}^3$. The blue affine spaces represent tangent spaces.* ***Left:*** *connected submanifold of dimension 2 (the sphere S^1).* ***Right:*** *a submanifold of two connected components, one of dimension 2 (homeomorphic to the torus $\mathbb{T}^2$ on which lies x_0), and one of dimension 1 on which lies x_1.*

We will denote with $d^2f(x)$ the second differential of f (which we will sometimes call abusively its Hessian), which is a symmetric bilinear form on $\mathbb{R}^d$. We denote with $d^2f(x)[\xi,\eta]$ its evaluation on vectors ξ,η. We denote with $\nabla^2 f(x) \in \mathbb{R}^{d\times d}$ the Hessian matrix of f at x, which is the matrix of $d^2f(x)$ in the canonical basis of $\mathbb{R}^d$, and we have $d^2f(x)[\xi,\eta] = \eta^\top \nabla^2 f(x)\xi$. For any vector subspace space $S \subset \mathbb{R}^d$, and any bilinear form H on $\mathbb{R}^d$, we denote with $H|_S$ the restriction of H to S, which is a bilinear form on F. We say that a bilinear form H is positive semidefinite if $H[\xi,\xi] \geq 0$ for any $\xi \in \mathbb{R}^d$, and is positive definite if $H[\xi,\xi] > 0$ for all $\xi \in \mathbb{R}^d \setminus \{0\}$. We use the same terminology for matrices.

We are now ready to state Theorem 2.3, which is the cornerstone of this work. For the rest of this section (subsection 2.1), let $x_0 \in \mathbb{R}^d$ and $f : (x_0, \mathbb{R}^d) \to \mathbb{R}$ be a nonnegative C^p function, for $p \in \mathbb{N} \cup \{\infty\}$, $p \geq 2$, such that $f(x_0) = 0$. We claim that if there is a submanifold of class C^1 and of dimension d_0 around x_0 of zeros of f, and if the Hessian of f at x_0 has rank $d - d_0$ (which we will call the normal Hessian condition), then it can be decomposed as a sum of squares as in (1.2).

DEFINITION 2.2 (normal Hessian condition). *Let $\mathcal{Z}$ denote the set of zeros of f. We say that f satisfies the NHC at $x_0 \in \mathcal{Z}$ with regularity $k \in \mathbb{N} \cup \{\infty\}$, $k \geq 1$, if $\mathcal{Z}$ is a submanifold of dimension of class C^k around x_0 (i.e., this holds on an open neighborhood of x_0), and such that $d^2f(x_0)|_{T_{x_0}\mathcal{Z}^\perp}$ is positive definite, that is*

$$\forall h \in T_{x_0}\mathcal{Z}^\perp \setminus \{0\},\ d^2f(x_0)[h,h] > 0. \tag{2.1}$$

The proof of Theorem 2.3 can be found in section 4. To illustrate the definition of the NHC, we refer to Figure 4 which represents the local behavior of functions f defined locally around a point $x_0 \in \mathbb{R}^2$ in the set of zeros and which satisfies the NHC for $d_0 = 1$.

THEOREM 2.3. *Assume f satisfies the NHC at x_0 (Definition 2.2) with regularity k, and let d_0 be the dimension of the submanifold $\mathcal{Z}$ in the neighborhood x_0. There exists an open neighborhood U of x_0 in $\mathbb{R}^d$ on which f is defined and such that $U \cap \mathcal{Z}$*

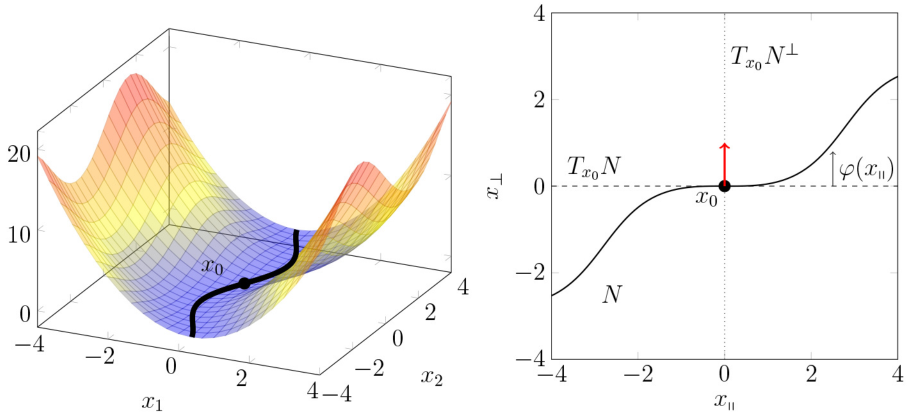

FIG. 4. *Local view of the function around a minimum lying on a* 1*-dimensional manifold. Left: function around the minimum* x_0*. Right: decomposition of* $\mathbb{R}^2$ *at* x_0 *between tangent space and normal tangent space* $T_{x_0}N + T_{x_0}N^\perp$*, and positive eigenvector of the Hessian in red. Reparametrization in the right coordinate system, and representation of the map* $\varphi(x)$ *given by the Morse lemma. (Note that color appears only in the online article.)*

is a submanifold of $\mathbb{R}^d$ *of dimension* d_0 *and of class* $C^{\max(k,p-1)}$*, and there exist functions* $f_i \in C^{p-2}(U)$*, where* $1 \le i \le d - d_0$ *such that*

$$\forall x \in U,\ f(x) = \sum_{i=1}^{d-d_0} f_i^2(x). \tag{2.2}$$

The proof of this theorem relies on the Morse lemma, initially stated in [21] in order to decompose a function at a critical point.

LEMMA 2.4 (Morse lemma [21, Lemma 2.2]). *Let* $f : (x_0, \mathbb{R}^d) \to \mathbb{R}$ *be of class* C^p*,* $p \in \mathbb{N} \cup \{\infty\}$*,* $p \ge 2$*, and let* x_0 *be a critical point of* f *(i.e.,* $df(x_0) = 0$*) such that* $d^2 f(x_0)$ *is nonsingular with index* s *(that is with* s *positive eigenvalues). Then there exists a* C^{p-2} *local coordinate system around* x_0*:* $z_1, \ldots, z_d : (x_0, \mathbb{R}^d) \to \mathbb{R}$ *such that* $f = f(x_0) + \sum_{i=1}^{s} z_i^2 - \sum_{i=s+1}^{d} z_i^2$ *in a neighborhood of* x_0*.*

Note that under the SHC, this lemma directly yields the desired local decomposition of f as a sum of squares, as the index $s = d$ since the Hessian is positive definite. In [12, Lemma C.6.1], which we restate and prove as Lemma B.3, this lemma is extended to include the case where the Hessian is nonsingular with respect to certain coordinates, which is necessary to include manifolds of zeros and work under the NHC (see also (2.3)).

Main steps of the proof of Theorem 2.3. The main steps of this proof are represented geometrically in Figure 4.

Step 1. We show that under the NHC at x_0, we have $T_{x_0}N = \ker(\nabla^2 f(x_0))$ and hence that $d^2 f(x_0)|_{T_{x_0}N^\perp}$ is positive definite.

Step 2. Reparametrizing f on a basis adapted to $T_{x_0}N^\perp \oplus T_{x_0}N$ as $f(x_\perp, x_{\shortparallel})$, we apply the Morse lemma (see Lemma B.3), which decomposes the function f into the form

$$f(x_\perp, x_{\shortparallel}) = f(\varphi(x_{\shortparallel}), x_{\shortparallel}) + \sum_{i=1}^{d-d_0} f_i^2(x_\perp, x_{\shortparallel}) \tag{2.3}$$

for a certain function φ of class C^{p-1} and f_i of class C^{p-2} in a certain open set around x_0 (for an easy visualization, see Figure 4).

Step 3. We characterize the manifold of zeros around x_0.

Step 4. We show that the first term of the result of the Morse lemma is equal to zero using the previous characterization, which shows (2.2). ∎

Example 2.5 (case where $d_0 = 0$). When $d_0 = 0$, the NHC at x_0 is simply the SHC (1.4), that is the condition that x_0 be a strict minimum. In that case, Theorem 2.3 simply states that there exists an open neighborhood U of x_0 such that $U \cap \mathcal{Z} = \{x_0\}$ and on which f can be decomposed as the sum of d squares.

Remark 2.6 (smoothing effect). Theorem 2.3 induces a smoothing effect; indeed, if we simply assume that there exists a d_0 dimensional manifold of class C^1 of zeros with which the NHC is satisfied, one sees that this manifold is actually of class C^{p-1} in a neighborhood of x_0.

2.2. Global decomposition as a sum of squares for functions on $\mathbb{R}^d$. In this section, we fix f to be a nonnegative C^p function defined on an open subset Ω of $\mathbb{R}^d$. Once again, we assume $p \in \mathbb{N} \cup \{\infty\}$, $p \geq 2$. The goal is to find a condition on f to be written as a sum of squares of functions defined on Ω. In this work, this condition is simply that the NHC holds at every $x_0 \in \mathcal{Z}$.

DEFINITION 2.7 (global NHC). *We say that f satisfies the global NHC (with regularity $k \geq 1$) if for all x_0 in its null set $\mathcal{Z}$, f satisfies the local NHC at x_0 (with regularity $k \geq 1$).*

This definition of the global NHC as a local condition holding everywhere can be reformulated in a more global and geometric way.

Remark 2.8 (reformulation of the NHC). f satisfies the global NHC (with regularity $k \geq 1$) if and only if its null set $\mathcal{Z}$ is a submanifold of class $C^{\max(k,p-1)}$, and if f is positive normally to $\mathcal{Z}$, *i.e.,*

$$\forall x \in \mathcal{Z} \;\; \forall h \in T_x\mathcal{Z}^\perp \setminus \{0\},\; d^2 f(x)[h,h] > 0. \tag{2.4}$$

Examples of functions satisfying the global NHC can be found in Figure 5. They have manifolds of zeros $\mathcal{Z}$ which are depicted in the same order in Figure 2. There can be more than one connected component in $\mathcal{Z}$ (see the second and third examples in Figure 5).

Under this geometric condition, we will show in Theorem 2.9 that f can be written as a sum of squares of C^{p-2} functions with locally finite support, defined below.

Locally finite support. Let X be a topological space (see [14] for full definitions). We say that a family (S_i) of subsets of X is locally finite if for every $x \in X$, there exists an open set U_x containing x which intersects a finite number of the S_i, i.e., $|\{i \in I \,:\, U_x \cap S_i \neq \emptyset\}| < \infty$. A family (f_i) of functions on a topological space X has locally finite support if the family of supports $(\mathrm{supp}(f_i))_{i\in I}$ is locally finite (recall that $\mathrm{supp}(f_i) = \overline{\{x \,:\, f(x) \neq 0\}}$). In particular, if (f_i) has locally finite support, the function $\sum_{i\in I} f_i^2$ is well defined and it is also of class C^q if the functions are of class C^q. Using this terminology, the global result can be stated as follows.

THEOREM 2.9. *If f satisfies the global NHC in* Definition 2.7*, there exists an at most countable family $(f_i)_{i\in I} \in (C^{p-2}(\Omega))^I$ with locally finite support such that*

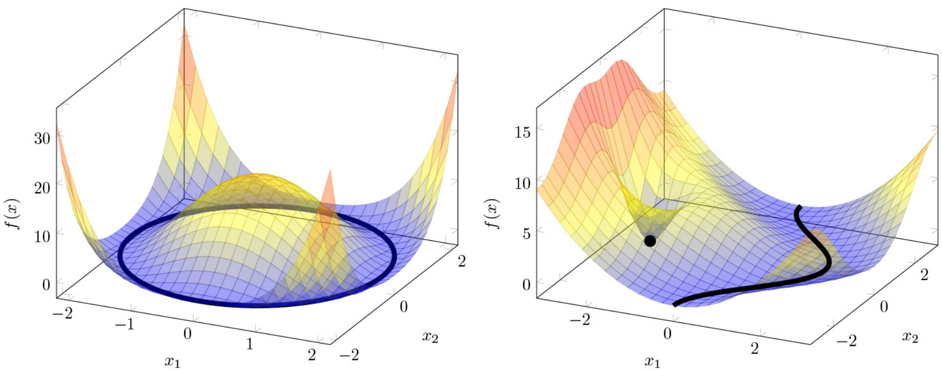

FIG. 5. *Examples of functions f which satisfy the global NHC, with submanifolds $\mathcal{Z}$ of zeros corresponding to the first and last submanifolds presented in Figure* 2.

$$\forall x \in \Omega,\ f(x) = \sum_{i\in I} f_i(x)^2. \tag{2.5}$$

Moreover,

- *if f satisfies the strict Hessian condition, $\mathcal{Z}$ is discrete and we can find such a decomposition such that $|I| \leq d+1$;*
- *if $\mathcal{Z}$ is compact, then $|I|$ can be taken to be finite.*

For the formal proof of this result, we refer to the next section, where this result will be proved more generally for functions defined on manifolds (see Theorem 3.9 and section 3).

Main steps of the proof. For subtleties pertaining to the SHC case, we refer to section 3. The gluing done in that section is slightly more elaborate.

Step 1. Since the local NHC holds at any point in $\mathcal{Z}$, using Theorem 2.3 shows that at any point x, there exists an open neighborhood U_x of x, an integer n_x, and functions $(f_{x,j})_{1\leq j\leq n_x}$ of class C^{p-2} on U_x such that $f = \sum_{j=1}^{n_x} f_{x,j}^2$ on U_x. The collection of sets U_x is then an open covering of $\mathcal{Z}$. Since $\mathbb{R}^d$ is Hausdorff and second-countable (see section 3 for precise definitions), only at most countable subsets of them are necessary to cover $\mathcal{Z}$ (even a finite number if $\mathcal{Z}$ is included in a compact, since it is then itself a compact). Denote with $(U_i)_{i\in I}$ this open covering, and replace x by i to denote the associated $f_{i,j}$ and n_i.

Step 2. Since $\mathcal{Z}$ is closed, as the set of zeros of a continuous function, the set $U_{>0} := \{x \in \Omega \,:\, f(x) > 0\}$ is open and the map $f_1 := \sqrt{f} : U_{>0} \to \mathbb{R}$ is of class C^p and satisfies $f_1^2 = f$. We can therefore add $U_{>0}$ to the collection (U_i) and still guarantee the following property: for all $i \in I$, there exists $n_i \in \mathbb{N}$ and $f_{i,j} \in C^{p-2}(U_i)$ such that $f = \sum_{j=1}^{n_i} f_{i,j}^2$. Moreover, (U_i) becomes an open covering of Ω; in particular, if U_i was a finite covering of $\mathcal{Z}$, it now becomes a finite covering of Ω.

Step 3. Using Lemma 3.6, we can take a partition of unity (χ_i) adapted to the open covering $\bigcup_i U_i$ such that $\sum_i \chi_i^2 = 1$ and which is locally finite. Define $\tilde{f}_{i,j} = f_{i,j}\chi_i$ which is now defined on the whole of Ω (indeed, it can be extended as zero to $\Omega \setminus U_i$ since the support ot χ_i is included in U_i). The $\tilde{f}_{i,j}$ satisfy $\sum_{i,j} \tilde{f}_{i,j}^2 = f$ on the whole of Ω, and is a finite family if the covering U_i is finite (if $\mathcal{Z}$ is assumed to be compact for example). ∎

2.3. Examples of functions satisfying the NHC, and applying optimization schemes.

2.3.1. Optimal transport. A natural example to motivate the results of this work is that of [35], in which the authors need to represent a nonnegative function f defined on a product $X \times Y$ of open sets of $\mathbb{R}^d$ as sum of squares to apply a method similar to AnalyticalSoSOpt to solve an optimization problem. The aim is to minimize the transport cost between two probability measures μ and ν, which can be cast as a linear problem with a nonnegativity constraint of the form $f(x,y) \geq 0$. The optimal nonnegative function f_* is of the form $f_*(x,y) = \frac{1}{2}\|x-y\|^2 - u_*(x) - v_*(y)$, where u_* and v_* are so-called Kantorovich potentials. A consequence of Brenier's theorem for quadratic optimal transport [30, Theorem 1.22], and proved in [35, Theorem 5], the two following facts hold. (i) The null set of f_* is the graph of a function $T : X \to Y$ called the transport map and (ii) orthogonally to this graph, f_* grows quadratically. Hence, implicitly, the authors show that the NHC holds for their problem, and hence that f can be decomposed as a sum of squares. Using tools similar to AnalyticalSoSOpt, this allows us to derive an algorithm to compute an approximation of the optimal transport cost with convergence rate of $n^{-m/d}$ and in a time which is polynomial in n, where n is the number of samples available from μ, ν and m is the regularity of the Kantorovich potentials.

2.3.2. An example class of functions satisfying the NHC. Consider functions $F : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$ of the form

$$F(x,y) = g(\|y - f(x)\|^2),\ f : \mathbb{R}_1^d \to \mathbb{R}^{d_2}, g : \mathbb{R}_+ \to \mathbb{R}, \tag{2.6}$$

where f, g are of class C^p, $p \geq 2$, and g reaches its unique minimum at 0 and satisfies $g'(0) > 0$. It is easy to check that such functions F satisfy the NHC and have a continuous set of minimizer: the graph of f. With complex g, these functions can also be quite elaborate; for example, they can have multiple local minimizers which are not global. Applying Theorem 2.9 for functions of type (2.6), and combining it with Proposition 1.1, AnalyticalSoSOpt will converge with almost optimal rates, depending on the smoothness p of the functions. Note that these functions do not satisfy the SHC: the novelty of this work is to show that such functions can be optimized using AnalyticalSoSOpt with good convergence rates.

We illustrate the good convergence of AnalyticalSoSOpt by applying it to functions F defined as above with $g = \tanh$ and $f_i(x_1, \ldots, x_{d_1}) = cos(w_i \cdot x)$, $1 \leq i \leq d_2$, where w_i satisfy $\sum_{j=1}^{d_1} |w_i| = 1$. The w_i have been generated randomly, and we test AnalyticalSoSOpt for different dimensions $(d_1, d_2) = (d, d)$ with d ranging from 1 to 5. In Table 1, we report the performance of AnalyticalSoSOpt for certain values of n, compared to the naive method of taking the discretization point from the n points where f is minimal. We compute $\epsilon = n^{1/d}(\min_{1 \leq i \leq n} f(x_i) - f(\text{AnalyticalSoSOpt}(f, n)))$ to measure how much better our method is compared to the naive random sampling (positive = better). We report the mean and standard deviation of ϵ on five different seeds.

3. Global decomposition as a sum of squares for functions on manifolds.

In this section, we present results analogous to those of section 2 but in the more general context of manifolds. After a brief recap of the terminology and definitions related to manifolds, in subsection 3.1, we will adapt the definitions of the local and global NHCs, as well as state the equivalent result to Theorem 2.3 in the context manifolds. In subsection 3.2, we will introduce the tools to glue local decompositions

TABLE 1
Performance of AnalyticalSoSOpt *w.r.t. the naive baseline, for different values of d and n.*

d	n	ϵ
1	30	0.16 ± 0.15
2	100	0.13 ± 0.14
3	300	0.15 ± 0.13
4	300	0.10 ± 0.12
5	400	0.14 ± 0.20

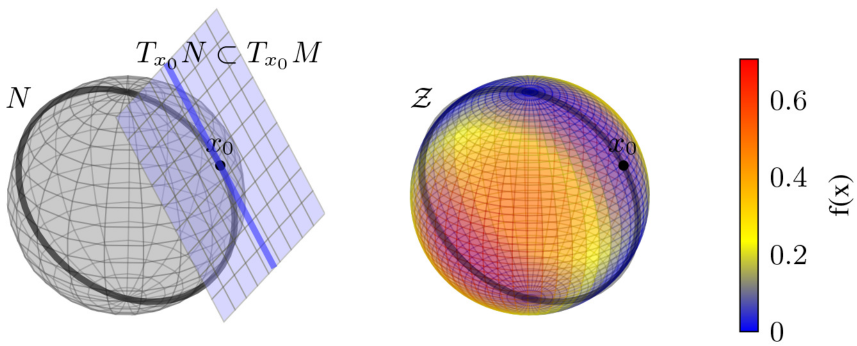

FIG. 6. ***Left:*** *Representation of the manifold $M = S^2$ as well as a submanifold N homeomorphic to a circle. The tangent spaces at a given point $x_0 \in N \subset M$ are represented as well. Right: Representation of a nonnegative function on the sphere as a color map; it satisfies the NHC, and its null space $\mathcal{Z}$ is represented in black. (Note that color appears only in the online article.)*

as sum of squares together. Finally, in subsection 3.3, we prove Theorem 3.9, the equivalent of Theorem 2.9 in the broader context of manifolds.

Additional definitions and notations for manifolds. In this section, we recall the basic terminology of manifolds. We assume that a reader of this section is familiar with this topic. The main idea behind the introduction of manifolds as opposed to submanifolds of $\mathbb{R}^d$ is to consider the intrinsic geometric object, and not its relation to the euclidean space it is embedded in (as such an embedding is not unique). An example of manifold as well as a representation of the tangent space is provided in the left of Figure 6. For introductions to manifolds, we refer to [15, 25, 32]. Informally, a manifold M of dimension d is a set which "looks like $\mathbb{R}^d$" locally. This means that at every point $x \in M$, we can find a chart ϕ which is a homeomorphism from an open neighborhood U of x onto an open set of $\mathbb{R}^d$. The family of charts $\mathcal{A} = (\phi_\rangle)$ is called an atlas. We further assume that M is a second-countable,[3] Hausdorff[4] space. It is of class C^k if all transition maps $\phi_i \circ \phi_j^{-1}$ are of class C^k. If M is a manifold of class at least C^1, we can define its tangent space T_xM at each point x. A map from M to $\mathbb{R}^p$ is of class C^q if its composition with any chart is of class C^q. If $q \geq 1$, we will denote with $df(x) \in \mathrm{Hom}(T_xM, \mathbb{R}^p)$ its differential at x.

Example 3.1. All sub-manifolds of $\mathbb{R}^d$ are manifolds. The notions of regularity, dimension, and tangent space coincide.

[3]A topological space is said to be second-countable if there exists a countable sequence of open sets U_n such that any open set U in the topology is a reunion of a part of the U_n.

[4]A topological space is Hausdorff if for any two points $x \neq x'$, there exist two open sets U, V such that $x \in U$ and $x' \in V$ and $U \cap V = \emptyset$

3.1. Assumptions in the manifold case. In this section, we formulate the local NHC in the case of manifolds, and rewrite Theorem 2.3 in this setting. We also extend the definitions of being positively normal and the global NHC (Definition 2.7).

Fix $p \in \mathbb{N} \cup \{\infty\}$, $p \geq 2$, and a manifold M of regularity at least C^p and of dimension $d \in \mathbb{N}$. To start with, let $f : \Omega \to \mathbb{R}$ be a nonnegative function defined on an open set of M and of class C^p. As before, define $\mathcal{Z}$ to be the set of zeros of f.

Contrary to the $\mathbb{R}^d$ case, the second differential of the function f cannot be identified as a symmetric bilinear form everywhere. However, it is the case at so-called critical points, i.e., points $x \in \Omega$ such that $df(x) = 0$. In particular, since all the zeros of a C^1 nonnegative function are critical points, this Hessian will be defined at all points in the set of zeros $\mathcal{Z}$ of f.

Lemma 3.2 (definition of the Hessian). *Let x be a critical point of f. Then there exists a unique symmetric bilinear form $H_f(x) : T_xM \times T_xM \to \mathbb{R}$ such that for any local chart $\phi : (M, x) \to (\mathbb{R}^d, 0)$ it holds that*

$$\forall \xi, \eta \in T_xM \times T_xM,\ H_f(x)[\xi, \eta] = d^2(f \circ \phi^{-1})(0)[d\phi(x)\xi, d\phi(x)\eta].$$

In order to prove this lemma, we simply define the bilinear form as such for a given chart ϕ around x, and then show that this definition does not depend on the chart ϕ using the fact that x is a critical point. This is completely proved in section 2 of [21]. In order to formulate the definition of the NHC in the setting of manifolds, we further need a definition of what a submanifold of M is. A subset $N \subset M$ is said to be a submanifold of M of class C^k if M is of class C^k and if, for any $x \in N$ and any local chart $\phi : U \to \mathbb{R}^d$ defined on a neighborhood of x, $\phi(U \cap N)$ is a submanifold of $\mathbb{R}^d$ of class C^k. In the literature, this is also called a proper submanifold (see on the left of Figure 6 for an example).

Definition 3.3 (NHC for a manifold). *Let $x \in \Omega$ be a point in the domain of f. We say that f satisfies the NHC at a point x in its null set $\mathcal{Z}$ (with regularity $k \geq 1$) if $\mathcal{Z}$ is a submanifold of M of class C^k around x, and such that*

$$\forall \xi \in T_xM \setminus T_xN,\ H_f(x)[\xi, \xi] > 0. \tag{3.1}$$

Using any local chart around the point x in the domain of f, one can apply Theorem 2.3 to obtain the following theorem as a corollary.

Theorem 3.4. *Assume f satisfies the NHC at x_0 with regularity k, and let d_0 be the dimension of the submanifold $\mathcal{Z}$ around x_0. There exists an open neighborhood U of x_0 in M on which f is defined and such that $U \cap \mathcal{Z}$ is a submanifold of M of dimension d_0 and of class $C^{\max(k,p-1)}$; and there exist functions $f_i \in C^{p-2}(U)$ where $1 \leq i \leq d - d_0$ such that*

$$\forall x \in U,\ f(x) = \sum_{i=1}^{d-d_0} f_i^2(x). \tag{3.2}$$

Exactly in the same way as for the definition of the NHC for manifolds, we can similarly extend the definition of a function satisfying the global NHC with regularity $k \geq 1$ in Definition 2.7. We will therefore say that $f : M \to \mathbb{R}$ which is nonnegative satisfies the global NHC with regularity $k \geq 1$ if it satisfies the local NHC with regularity $k \geq 1$ at every point in its set of zeros $\mathcal{Z}$ or, equivalently if it is positive normally to $\mathcal{Z}$ (see (2.4)) which is a submanifold of M of class $C^{\max(k,p-1)}$. On the right-hand side of Figure 6, we represent a function which satisfies the NHC on the sphere S^2 through a colormap, with a continuous set of zeros. The goal is to prove that such a function can be decomposed as a sum of squares on S^2.

3.2. Gluing local decompositions to form a global one. In this section, we present and develop the tools to glue local decompositions such as Theorem 3.4 into a global one, which will lead to Theorem 3.9.

The first result we need is a simple result to "extend" a function defined on an open set U of M to M by multiplying it by a function defined on M whose support lies in U (Lemma 3.5). The second one is a variant of the fundamental result of existence of partitions of unity on a manifold, adapted to our sum of squares setting (Lemma 3.6). Recall that the support of a function has been defined in subsection 2.2. The proof of these results can be found in subsection A.1.

Lemma 3.5 (extension lemma). *Let $q \in \mathbb{N}$, M be a manifold of class at least C^q. Let U be an open set of M, $g : U \to \mathbb{R}$ be a C^q function defined on U, and $\chi : M \to \mathbb{R}$ be a C^q function defined on the whole of M but with support included in U. Then the function $\chi g : U \to \mathbb{R}$ extended as 0 on $M \setminus U$, is of class C^q on the whole of M and has support included in* $\operatorname{supp}(\chi) \subset U$. *We still denote with χg its extension to M.*

Lemma 3.6 (gluing lemma). *Let $(U_i)_{i\in I}$ be an open covering of a manifold M of class C^k (i.e., $\bigcup_{i\in I} U_i = M$). There exists a family of functions $\chi_i : M \to [0,1]$ of class C^k with locally finite support, such that* $\operatorname{supp}(\chi_i) \subset U_i$ *for all $i \in I$ and satisfying:*

$$\sum_{i\in I} \chi_i^2 = 1.$$

We can now proceed from local to global in two steps. First, we use the gluing lemma to glue decompositions in a single connected component of the manifold of zeros (Lemma 3.7). We then glue these different decompositions into a single global one (Lemma 3.8).

Lemma 3.7. *Assume f satisfies the global NHC. Let N be a connected component of its manifold of zeros $\mathcal{Z}$. There exists an open neighborhood U of N as well as a locally finite, at most countable family $(f_j)_{j\in J}$ of functions of class C^{p-2} such that*

$$\forall x \in U,\ f(x) = \sum_{j\in J} f_j(x)^2. \tag{3.3}$$

Moreover, we can find J such that (a) $|J| = d$ if $N = \{x_0\}$ is a single point and (b) J is finite if N is compact.

Proof. The case where $N = \{x_0\}$ is simply Theorem 3.4 applied to x_0. In the other cases, note that for all $x \in N$, by Theorem 2.3 since the NHC is satisfied at x, there exists an open neighborhood U_x of x as well as functions $(f_{x,i})_{1\le i\le d}$ of class C^{p-2} such that $f = \sum_{i=1}^d f_{x,i}^2$ on U_x. Since $(U_x)_{x\in N}$ covers N, we can extract a covering $(U_{x_j})_{j\in J}$ of N such that (a) J is finite if N is compact and (b) J is at most countable otherwise, since N is second-countable and Hausdorff. Denote with $(U_j)_{j\in J}$ this open covering, and replace x by j to denote the associated $f_{j,i}$. Denote with U the open set $\bigcup_j U_j$.

Applying Lemma 3.6 to the manifold U, we can find a family of functions $(\chi_j)_{j\in J}$ with locally finite support, such that $\operatorname{supp}(\chi_j) \subset U_j$ and $\sum_j \chi_j^2 = 1$ on U. By the extension Lemma 3.5, we can therefore define the functions $\tilde{f}_{j,i} := \chi_j\ f_{j,i}$ for $i \in \{1,\dots,d\}$ and $j \in J$ which are defined on the whole of M. Since $\operatorname{supp}(\tilde{f}_{j,i}) \subset \operatorname{supp}(\chi_j)$ and since $1 \le i \le d$, the support of $(\tilde{f}_{j,i})$ is also locally finite. To conclude, we use the property that $\sum_j \chi_j^2 = 1$ on U as well as the fact that $\sum_i f_{j,i}^2 = f$ on $\mathrm{supp}(\chi_j) \subset U_j$ to show that $\sum_{i,j} \tilde{f}_{j,i}^2 = f$ on U. The number of functions $\tilde{f}_{j,i}$ is finite if N is compact since J is finite, and is at most countable else since J is at most countable. □

LEMMA 3.8. *Let $\mathcal{Z} = \sqcup_{i \in I} N_i$ be the manifold of zeros decomposed along its connected components. Assume that there exists an index set J, such that for all $i \in I$, there exists an open neighborhood U_i of N_i on which f can be decomposed as a sum of squares indexed by J:*

$$\forall i \in I,\ \exists (f_{i,j})_{j \in J} \in (C^{p-2}(U_i))^J,\ \forall x \in U_i,\ f(x) = \sum_{j \in J} f_{i,j}(x)^2, \tag{3.4}$$

and such that the families $(f_{i,j})_{j \in J}$ are all locally finite. Then there exists a locally finite family $(g_j)_{j \in J \cup \{\star\}}$ of C^{p-2} functions on M (we add an extra element $\star$ to J), such that

$$\forall x \in M,\ f(x) = \sum_{j \in J \cup \{\star\}} g_j(x)^2. \tag{3.5}$$

Proof. By Lemma A.4, there exist disjoint open sets $V_i \subset M$ such that $N_i \subset V_i$, since $\mathcal{Z}$ is a proper submanifold of M by Definition 2.7 (directly adapted to the manifold case). Hence, we can assume that the U_i are disjoint (considering instead $U_i \cap V_i$, the property still holds). Define $U_\star = \{f > 0\}$. Since the U_i's cover $\mathcal{Z}$, $U_\star \cup \bigcup_{i \in I} U_i$ covers M since f is nonnegative; take $\chi_\star, (\chi_i)_{i \in I}$ to be a gluing family adapted to that covering given by Lemma 3.6. For any $i, i' \in I$, we have $\chi_i \chi_{i'} = 0$ since χ_i is supported on U_i and the U_i are disjoint. Consider the function $g_j = \sum_{i \in I} \chi_i f_{i,j}$, which is well defined on M and C^{p-2} by Lemma 3.5. We have $g_j^2 = \sum_{i \in I} \chi_i^2 f_{i,j}^2$ since $\chi_i \chi_{i'} = 0$ when $i \neq i'$.

Assertion: the family $(g_j)_{j \in J}$ has locally finite support. Let $x \in \mathbb{R}^d$ and assume $g_j(x) \neq 0$. Then there exists $i \in I$ such that $\chi_i(x) > 0$, and hence there exists an open set U_x around x such that $U_x \subset U_i$. But in that case, $\chi_{i'}(x') = 0$ for all other i' and for all $x' \in U_x$ since the U_i are disjoint and $\chi_{i'}$ is supported on $U_{i'}$. Moreover, since $(f_{i,j})_{j \in J}$ is locally finite, there exists an open set V_x around x as well as a finite $J_0 \subset J$ such that $f_{i,j} = 0$ on V_x for all $j \in J \setminus J_0$. Hence, for any $j \in J \setminus J_0$ and any $x' \in U_x \cap V_x$, we have $f_{i,j}(x') = 0$ and $\chi_{i'}(x') = 0$ thus $g_j(x') = 0$. Thus, $U_x \cap V_x \subset M \setminus \mathrm{supp}(g_j)$ for all $j \notin J_0$; the family g_j is locally finite.

Conclusion. Define $g_\star = \chi_\star \sqrt{f}$, which is of class C^p since $\chi_\star$ is supported on $\{f > 0\}$. Since the addition of one function changes nothing to the locally finite property of a family of functions, the family $(g_j)_{j \in J} \cup g_\star$ is still locally finite. Using the fact that $g_j^2 = \sum_{i \in I} \chi_i^2 f_{i,j}^2$, that $\sum_{j \in I \cup \{\star\}} \chi_i^2 = 1$, and (3.4), it holds that

$$\begin{aligned} \sum_{j \in J \cup \{\star\}} g_j^2 &= \chi_\star^2 f + \sum_{j \in J} g_j^2 = \chi_\star^2 f + \sum_{j \in J} \sum_{i \in I} \chi_i^2 f_{i,j}^2 \\ &= \chi_\star^2 f + \sum_{i \in I} \sum_{j \in J} \chi_i^2 f_{i,j}^2 = \chi_\star^2 f + \sum_{i \in I} \chi_i^2 f = f. \end{aligned}$$ □

3.3. Main results.

We are now ready to state our main result on manifolds. On the right-hand side of Figure 6, we represent a case where this theorem applies for a nonnegative function defined on S^2.

THEOREM 3.9. *Let M be a manifold and $f : M \to \mathbb{R}$ be a nonnegative map of class C^p. Assume f satisfies the global NHC. Then there exists I which is at most* *countable and functions* $f_i \in C^{p-2}(M)$ *for* $i \in I$ *such that the family* (f_i) *has locally finite support and*

$$\forall x \in M,\ f(x) = \sum_{i\in I} f_i(x)^2. \tag{3.6}$$

Moreover,

- *if* f *satisfies the strict Hessian condition,* $\mathcal{Z}$ *is discrete and we can find such a decomposition such that* $|I| \leq d+1$*;*
- *if* $\mathcal{Z}$ *is compact, then* $|I|$ *can be taken to be finite.*

Proof. The proof of this theorem is a simple consequence of Lemmas 3.7 and 3.8. The global NHC Definition 2.7 shows that $\mathcal{Z}$ is a submanifold of M. Let N_i denote the connected components of $\mathcal{Z}$. By Lemma A.4, we can find disjoints open sets U_i such that $N_i \subset U_i$.

General case. Without any more assumptions, we know from Lemma 3.7 that on any connected component N_i, we can have a decomposition of the form $f = \sum_{j\in J} f_{i,j}^2$ with $f_{i,j} \in C^{p-2}$ a family with locally finite support on an open neighborhood V_i of N_i. Moreover, we know that J is at most countable. Adding zeros when necessary, and reindexing, we can assume that $J = \mathbb{N}$. Now applying Lemma 3.8, we prove the general case.

Compact case. If we assume that N is compact, since the U_i cover N, necessarily the number of connected components is finite (just extract a finite covering of N from the U_i). We know from Lemma 3.7 that on any connected component N_i, we can have a decomposition of the form $f = \sum_{j=1}^{n_i} f_{i,j}^2$ with $f_{i,j} \in C^{p-2}$ and $n_i \in \mathbb{N}$ on an open neighborhood V_i of N_i, since N_i is compact. Hence, up to adding $f_{i,j} = 0$, we can assume that $n_i = n = \max_i(n_i)$ since there are a finite number of connected components. Now applying Lemma 3.8 with $J = \{1,\ldots,n\}$, the result is proven in the compact case with $n+1$ functions.

SHC case. If we assume that the SHC holds, every connected component N_i is a singleton $\{x_i\}$; we know from Lemma 3.7 we can have a decomposition of the form $f = \sum_{j=1}^{d} f_{i,j}^2$ with $f_{i,j} \in C^{p-2}$ on an open neighborhood V_i of N_i, since N_i is compact. Now applying Lemma 3.8 with $J = \{1,\ldots,d\}$, the result is proven with $d+1$ functions. ∎

Remark 3.10. The difference between the number of functions in the SHC case is better than the one obtained in [28]. This is because of the two step procedure in the gluing: first in a connected component, and then between connected components. The long term goal is to be able to prove that we need only a finite number $N(d)$ of functions per connected component (in the compact case), and hence to have an explicit bound after gluing the connected components together, rather than just relying on a compact extraction argument, which is not as precise.

4. Proof of the local decomposition as a sum of squares. In this section, we formally prove the key result of the paper, Theorem 2.3.

Proof. Assume the slightly weaker condition that there exists a submanifold N of dimension d_0 and regularity k which contains x_0 and is included in $\mathcal{Z}$ (but not equal a priori; we will show that it is equal to $\mathcal{Z}$ around x_0), such that $d^2 f(x_0)|_{T_{x_0}N^\perp} \succ 0$. N can be locally parametrized around x_0 (see, for instance, Theorem 1.21 of [15]): there exists an open neighborhood $\widetilde{W}_0$ of 0 in $\mathbb{R}^{d_0}$, an open neighborhood U_{x_0} of x_0 in $\mathbb{R}^d$, and a C^k immersion $\phi : \widetilde{W}_0 \to U_{x_0}$ of class C^k such that ϕ is a homeomorphism from $\widetilde{W}_0$ onto $U_{x_0} \cap N$. Since the result is local, without loss of generality, we assume that $N = \mathrm{im}(\phi)$ for a C^k immersion $\phi : (0, \mathbb{R}^{d_0}) \to (x_0, \mathbb{R}^d)$. We will denote with $T_{x_0} := d\phi_0(\mathbb{R}^{d_0}) = T_{x_0}N$ the tangent space to N at x_0.

Before starting the proof, recall that for any $x \in \mathcal{Z}$, it holds that $df(x) = 0$ and $d^2 f(x) \succeq 0$ (or equivalently $\nabla^2 f(x) \succeq 0$). Moreover, if $A \in \mathbb{S}_+(\mathbb{R}^d)$ is a symmetric positive semidefinite matrix, if a vector $k \in \mathbb{R}^d$ satisfies $k^\top Ak = 0$, then $Ak = 0$ (this is a trivial consequence of the spectral theorem by decomposing k along an orthonormal basis of eigenvectors).

Step 1*: characterizing the null-space of the Hessian.* We will prove that under the assumptions of the theorem, (a) T_{x_0} is equal to the null-space $\ker(\nabla^2 f(x_0))$ of the Hessian of f at x_0 and (b) that for any supplementary S to T_{x_0}, the restricted Hessian $\nabla^2 f(x_0)|_S$ is positive definite.

To prove (a), assume that there exists an element in $k \in T_{x_0}$ such that $\nabla^2 f(x_0)k \neq 0$. Since $\nabla^2 f(x_0)$ is positive semidefinite, this implies that $k^\top \nabla^2 f(x_0)k > 0$. Let $h \in \mathbb{R}^{d_0}$ such that $d\phi_0 h = k$, and let $x_t = \phi(th)$ which is defined for t in an open neighborhood of 0. Using the Taylor expansion of f around x_0,

$$f(x) - f(x_0) - df(x_0)[x - x_0] = \tfrac{1}{2}(x - x_0)^\top \nabla^2 f(x_0)(x - x_0) + \epsilon(x - x_0)\|x - x_0\|^2,$$

where $\epsilon(x) \underset{\|x\|\to 0}{\to} 0$. Now applying this for x_t, since $f(x_t) = f(x_0) = 0$ and $df(x_0) = 0$, it holds that

$$0 = \frac{1}{2}(x_t - x_0)^\top \nabla^2 f(x_0)(x_t - x_0) + \epsilon(x_t - x_0)\|x_t - x_0\|^2.$$

Using the fact that ϕ is differentiable at 0 yields $x_t - x_0 = td\phi(0)[h] + o_{t\to 0}(t) = tk + o_{t\to 0}(t)$. Injecting this into the equation above yields

$$0 = t^2\, \tfrac{1}{2} k^\top \nabla^2 f(x_0)k + o_{t\to 0}(t^2).$$

Hence, necessarily, $k^\top \nabla^2 f(x_0)k = 0$, which is a contradiction. This proves that $T_{x_0} \subset \ker(\nabla^2 f(x_0))$, and in particular, $d_0 \leq \dim(\ker(\nabla^2 f(x_0)))$. Since the rank of $\nabla^2 f(x_0)$ is actually $d - d_0$, the rank theorem shows that $\dim(\ker(\nabla^2 f(x_0))) = d_0$ and hence $T_{x_0} = \ker(\nabla^2 f(x_0))$.

To prove (b), we just need to prove that the restriction to any supplementary to the null space of $\nabla^2 f(x_0)$ is positive definite. Using the small result at the beginning of the proof, any vector $k \in \mathbb{R}^d \setminus T_{x_0}$ satisfies $k^\top \nabla^2 f(x_0)k > 0$. In particular, this means that the restriction of $\nabla^2 f(x_0)$ to any supplementary subspace S of T_{x_0} is positive definite.

Step 2*: applying the Morse lemma.* Let $P = (P_1, P_2) \in O_d(\mathbb{R})$ be the matrix of an orthonormal basis adapted to the decomposition $\mathbb{R}^d = T_{x_0}^\perp \oplus T_{x_0}$. Note that $P_1 \in \mathbb{R}^{d\times(d-d_0)}$ and $P_2 \in \mathbb{R}^{d\times d_0}$ are also orthonormal matrices, and that since P_1 spans $T_{x_0}^\perp$, in particular $P_1^\top \nabla^2 f(x_0)P_1 \succ 0$.

Define $g : (x', y') \in \mathbb{R}^{d-d_0} \times \mathbb{R}^{d_0} \mapsto f(P_1x' + P_2y' + x_0) = f(\mathcal{A}(x', y'))$, where $\mathcal{A}(x', y') = P(x', y') + x_0$ is an isometry[5] ($\mathcal{A}^{-1}x = P^\top(x - x_0)$). We have $\nabla_{x'} g(0, 0) = P_1^\top \nabla f(x_0) = 0$ and $\nabla^2_{x'x'} g(0, 0) = P_1^\top \nabla^2 f(x_0)P_1 \succ 0$, which is nonsingular with index $s = d - d_0$. We can therefore apply the Morse lemma, Lemma B.3, to g: there exists two open neighborhoods of zero $V \subset \mathbb{R}^{d-d_0}, W \subset \mathbb{R}^{d_0}$ as well as $\varphi : W \to V$ of class C^{p-1} such that $\{(x', y') \in V \times W : \nabla_{x'} g(x', y') = 0\} = \{(x', y') \in V \times W : x' = \varphi(y')\}$ and $g : V \times W \to \mathbb{R}^{d-d_0}$ of class C^{p-2} such that

[5]An isometry is simply a map which preserves distances, and can be defined as an orthogonal transformation plus an affine shift.

$$\forall (x',y') \in V \times W,\ g(x',y') = g(\varphi(y'),y') + \sum_{i=1}^{d-d_0} g_i^2(x',y'). \tag{4.1}$$

We see that if we can show that $g(\varphi(y'),y')=0$ in a neighborhood of $(0,0)$, since we can go back to the original coordinate system through $\mathcal{A}^{-1}$, we will have shown the theorem.

Step 3: characterizing $\mathcal{Z}$ in a neighborhood of x_0. Denote with $G_\varphi = \{(\varphi(y),y) : y \in W\}$ the graph of φwhich is a submanifold of class C^{p-1} of $\mathbb{R}^{d-d_0} \times \mathbb{R}^d$ (see theorem 1.21, point (iv) of [15]). Since $\mathcal{A}$ is an isometry, the set $\mathcal{A}(G_\varphi)$ is also a submanifold of class C^{p-1} of $\mathbb{R}^d$.

Let $\widetilde{W} = \phi^{-1}(\mathcal{A}(V\times W))$; it is an open neighborhood of 0. Note that $\phi(\widetilde{W}) \subset \mathcal{Z} \cap \mathcal{A}(V\times W)$ by assumption, and since for any $x\in\mathcal{Z}$, we have $\nabla f(x)=0$, it holds in particular that for any $x \in \mathcal{Z}\cap\mathcal{A}(V\times W)$, we have $\nabla_{x'} g(\mathcal{A}^{-1}(x)) = P_1^\top \nabla f(x) = 0$. Hence, by the result of the Morse lemma, it holds that $\mathcal{A}^{-1}(\phi(\widetilde{W})) \subset \mathcal{A}^{-1}(\mathcal{Z}) \cap (V \times W) \subset G_\varphi$.

Define $\psi : (x',y') \in V\times W \mapsto (x'-\varphi(y'),y')$ which is a C^{p-1} diffeomorphism onto its image with inverse $(t,u)\mapsto(t+\varphi(u),u)$. Note that ψ maps G_φ onto $\{0_{\mathbb{R}^{d-d_0}}\}\times W$. If π_2 denotes the canonical projection $\pi_2 : \mathbb{R}^{d-d_0}\times\mathbb{R}^{d_0} \to \mathbb{R}^{d_0}$, we see that $\pi_2\circ\psi$ maps G_φ injectively onto $W\subset\mathbb{R}^{d_0}$.

Take $\Phi = \pi_2\circ\psi\circ\mathcal{A}^{-1}\circ\phi : \widetilde{W} \to \mathbb{R}^{d_0}$, which is well defined by the definition of $\widetilde{W}$, and C^1 by composition. It is an immersion at 0. Indeed (i) ϕ maps 0 onto x_0 and is an immersion at 0 by assumption, hence $d\phi_0$ is injective; (ii) $\psi\circ\mathcal{A}^{-1}$ is a C^{p-1} diffeomorphism from $\mathcal{A}(V\times W)$ (containing x_0) to its image, and hence its differential is invertible at x_0, and thus by composition, the differential $d(\psi\circ\mathcal{A}^{-1}\circ\phi)(0)$ is injective; (iii) since $\mathcal{A}^{-1}(\phi(\widetilde{W}))\subset G_\varphi$ by a previous statement, and since $\psi(G_\varphi)\subset\{0\}\times W$ also by a previous statement, it holds that the differential $d(\psi\circ\mathcal{A}^{-1}\circ\phi)(0)\mathbb{R}^{d_0}\subset\{0\}\times\mathbb{R}^{d_0}$ and hence applying π_2 does not change the injectivity of the differential; hence Φ is an immersion at 0. But since $d\Phi_0$ is a linear map from $\mathbb{R}^{d_0}$ to $\mathbb{R}^{d_0}$, $d\Phi_0$ being injective is equivalent to $d\Phi_0$ being invertible. Hence, by the local inversion theorem, Theorem B.1, there exists an open neighborhood of 0 $\widetilde{W}'\subset\widetilde{W}$ and an open neighborhood of 0 $W'\subset W$ such that Φ is a C^1 diffeomorphism from $\widetilde{W}'$ to W'.

Define $U = (\pi_2\circ\psi\circ\mathcal{A}^{-1})^{-1}(W') = \mathcal{A}(\psi^{-1}(\mathbb{R}^{d-d_0}\times W'))$, which is an open neighborhood of x_0. Since Φ is a diffeomorphism from $\widetilde{W}'$ to W', we have $\phi(\widetilde{W}')\subset U$. Moreover, since ψ is defined on $V\times W$, we have $U\subset\mathcal{A}(V\times W)$. Finally, let $u\in U\cap\mathcal{A}(G_\varphi)$. Since $u\in U$, there exists $\tilde{w}'\in\widetilde{W}'$ such that $\pi_2\circ\psi\circ\mathcal{A}^{-1}(\phi(\tilde{w}')) = \pi_2\circ\psi\circ\mathcal{A}^{-1}(u)$. Moreover, since $\pi_2\circ\psi$ is injective on G_φ, and since both $\mathcal{A}^{-1}(\phi(\tilde{w}'))$ and $\mathcal{A}^{-1}(u)$ belong to G_φ (the first using the previous point since $\widetilde{W}'\subset\widetilde{W}$ and the second by assumption), we have $\mathcal{A}^{-1}(\phi(\tilde{w}')) = \mathcal{A}^{-1}(u)$ and hence $u=\phi(\tilde{w}')$ since $\mathcal{A}$ is one to one. This shows that $U\cap\mathcal{A}(G_\varphi)\subset\phi(\widetilde{W}')$.

Moreover, a previous point shows that $\mathcal{A}^{-1}(\phi(\widetilde{W}))\subset\mathcal{A}^{-1}(\mathcal{Z})\cap(V\times W)\subset G_\varphi$. Now since $\mathcal{A}$ is one to one and since $\widetilde{W}'\subset\widetilde{W}$ we have $\phi(\widetilde{W})\subset\mathcal{Z}\cap(\mathcal{A}(V\times W))\subset\mathcal{A}(G_\varphi)$. Since $\phi(\widetilde{W}')\subset U$, we therefore have $\phi(\widetilde{W}')\subset\mathcal{Z}\cap U\subset\mathcal{A}(G_\varphi)\cap U$. Combining this with the previous result, we finally have

$$\phi(\widetilde{W}')\subset\mathcal{Z}\cap U\subset\mathcal{A}(G_\varphi)\cap U\subset\phi(\widetilde{W}') \implies \phi(\widetilde{W}')=\mathcal{Z}\cap U=\mathcal{A}(G_\varphi)\cap U. \tag{4.2}$$

Step 4: conclusion. (4.2) shows that $\phi(\widetilde{W}')=\mathcal{Z}\cap U=\mathcal{A}(G_\varphi)\cap U$.

On the one hand, this shows that $U\cap\mathcal{Z}$ is the intersection between an open set U and a submanifold $\mathcal{A}(G_\varphi)$ of $\mathbb{R}^d$ of class C^{p-1} (since it is the composition of the graph of φ which is C^{p-1}, which is a C^{p-1} manifold by [15], by an isometry which is in particular a diffeomorphism). Moreover, since ϕ is a C^k immersion which is a homeomorphism on its image, $\phi(\widetilde{W}')$ is a submanifold of class C^k. Thus, $U \cap \mathcal{Z}$ is a submanifold of $\mathbb{R}^d$ of class $C^{\max(k,p-1)}$.

On the other, since $\mathcal{A}^{-1}(U) \subset V \times W$, (4.1) becomes

$$\forall u \in U,\ g(\mathcal{A}^{-1}(u)) = g(\varphi(y'), y') + \sum_{i=1}^{d-d_0} g_i^2(\mathcal{A}^{-1}(u)),\ (x', y') = \mathcal{A}^{-1}(u). \tag{4.3}$$

Let $u \in U$ and write $(x', y') = \mathcal{A}^{-1}(u)$. First, $\mathcal{A}(\varphi(y'), y') \in \mathcal{A}(G_\varphi)$. Moreover, since $\mathcal{A}^{-1}u \in \psi^{-1}(\mathbb{R}^{d-d_0} \times W')$ by the definition of U, this shows that $y' \in W'$ and hence $(\varphi(y'), y') = \psi^{-1}(0, y') \in \psi^{-1}(\mathbb{R}^{d-d_0} \times W')$. This in turn shows that $\mathcal{A}(\varphi(y'), y') \in U$. Hence, $\mathcal{A}(\varphi(y'), y') \in \mathcal{A}(G_\varphi) \cap U = \mathcal{Z} \cap U$ and thus $g((\varphi(y'), y')) = f(\mathcal{A}(\varphi(y'), y')) = 0$. Finally, using this in (4.3), recalling that $g = f \circ \mathcal{A}$, and defining $f_i : u \in U \mapsto g_i(\mathcal{A}^{-1}u)$, we have

$$\forall u \in U,\ f(u) = \sum_{i=1}^{d-d_0} f_i^2(u). \tag{4.4}$$

We see that f_i is of class C^{p-2} since g_i was of class C^{p-2} and $\mathcal{A}^{-1}$ is an isometry; this concludes the proof of the theorem. ∎

5. Discussion and possible extensions. In this work, we have provided second order sufficient conditions in order for a nonnegative C^p function to be written as a sum of squares of C^{p-2} functions. We hope this will help provide a theoretical basis to algorithms which use functional sum of squares methods such as [27, 28, 35], which rely on the smoothness of such decompositions. The main avenue of future research we would like to explore is, as in the polynomial case, to handle functions f which are nonnegative on a constrained set defined by inequalities $f_i \geq 0$. This would imply both a theoretical component, similar to this one, and an algorithmic one, similar to [28].

Appendix A. Around partitions of unity and gluing functions. In this section, we detail a few topological properties of manifolds, in order to (a) decompose a manifold or a submanifold in connected components and (b) use partitions of unity as a tool to glue functions together. These specific properties are needed for subsection 3.2. For basics on topological spaces (what is a topology, the notion of continuity, of homeomorphism), we refer to Chapter 1 of [14]. The main references for manifolds can be found in [15, 25, 32]. Recall from subsection 3.3 the definition of a manifold M equipped with its atlas $\mathcal{A}$ of class C^k, and of a chart on M. A chart ϕ is said to be of class $C^{k'}$ for $k' \leq k$ if it is compatible with the atlas up to k' smoothness, i.e., if the transitions maps $\phi \circ \phi_i^{-1}$ and $\phi_i \circ \phi^{-1}$ are all $C^{k'}$. A priori, the atlas of a manifold of class C^k is not unique in the sense that more than one atlas generates the same structure. To make it so, and to be able to say *the atlas* of M of class C^k, we consider the *maximal atlas* on M, i.e., the collection of all charts of class C^k on M.

A.1. Paracompactness and partitions of unity. The main point of asking a (differential) manifold to be second-countable and Hausdorff, (and not just to be locally homeomorphic to $\mathbb{R}^d$), is for the manifold to be paracompact, and and hence to be equipped with partitions of unity. In this section, we introduce the main definitions and results on this topic.

Recall that a family of subsets (U_α) of a space X is said to be a covering of X if $\bigcup_\alpha U_\alpha = X$. It is said to be locally finite if for any $x \in X$, there exists an open neighborhood U of x which intersects only a finite number of the U_α. A family (V_β) is said to be a refinement of (U_α) if for all β, there exists an α such that $V_\beta \subset U_\alpha$.

A topological space X is said to be *paracompact* if for any open covering (U_α) of X, there exists an open refinement (V_β) of (U_α) such that (V_β) is locally finite, and is an open covering of X. The following lemma is proved in the first part of proposition 2.3 of [25] or can be found in Theorem 2.13 of [32].

LEMMA A.1. *A manifold is paracompact.*

In [32], a manifold is defined to be a metric space locally like $\mathbb{R}^d$. In proposition 2.2 of [25], it is shown that being metric and second countable is equivalent to the countable Hausdorff condition (under the condition of being locally homeomorphic to $\mathbb{R}^d$). Spivak's condition in [32] is, however, a bit more general; in fact, it allows a manifold M to be a union of a possible noncountable connected component (as theorem 2 of [32] shows that any connected component of a metric space locally homeomorphic to $\mathbb{R}^d$ is actually second-countable).

Paracompactness is an important property as it yields the existence of partitions of unity. The following lemma is standard (a proof can be found in [25, Proposition 2.3]). The result is of course also true for $k = 0$, but is more technical to prove.

LEMMA A.2 (standard gluing lemma, [25]). *Let $(U_i)_{i\in I}$ be an open covering of a manifold M of class C^k (i.e., $\bigcup_{i\in I} U_i = M$). There exists a family of functions $\chi_i : M \to [0,1]$ of class C^k such that* $\mathrm{supp}(\chi_i) \subset U_i$ *for all $i \in I$ and with locally finite support satisfying*

$$\sum_{i\in I} \chi_i = 1.$$

We now prove the two technical results need in subsection 3.2.

Proof of Lemma 3.5. The proof of this lemma is immediate. Indeed, by multiplication, we already know that χg is well defined and C^q on U. Moreover, for any point x in $V = M \setminus \mathrm{supp}(\chi)$, which is an open set, $(\chi g)(x) = 0$ (by definition if $x \in M \setminus U$ and since $\chi(x) = 0$ if $x \in U$) and hence is C^q on V. Since $V \cup U = M$ as $\mathrm{supp}(\chi) \subset U$, the property holds. Moreover, since $\chi g = 0$ on V, $\mathrm{supp}(\chi g) \subset \mathrm{supp}(\chi) \subset U$. ☐

Proof of Lemma 3.6. The proof of this result is a consequence of Lemma A.2. Indeed, this result shows that there exists a family of function $\widetilde{\chi}_i : M \to [0,1]$ of class C^k such that (i) for all $i \in I$, $\mathrm{supp}(\widetilde{\chi}_i) \subset U_i$, (ii) the support of $(\widetilde{\chi}_i)$ is locally finite, and (iii) $\sum_i \widetilde{\chi}_i = 1$.

Define $\phi = \sum_i \widetilde{\chi}_i^2$. Since $\sum_i \widetilde{\chi}_i = 1$, and $\widetilde{\chi} \geq 0$, necessarily, $\phi > 0$. Hence $\sqrt{\phi}$ is of class C^k, and hence $\chi_i := \widetilde{\chi}_i/\sqrt{\phi}$ is of class C^k, and satisfies all the desired properties. ☐

A.2. Connected components. Connectedness is a key topological notion for manifolds, and allows us to decompose a manifold into separate blocks. Recall that two points x, x' of a topological set X are connected if there exists no two open sets U, V such that $X = U \cup V$, $x \in U$, and $x' \in V$. Since being connected is an equivalence relation, we can partition X into classes with respect to that relation, which are called "connected components." Connected components are both open and closed.[6] On a

[6]For more details on connected components, see [14]

connected component of a manifold, the dimension d of the charts $\phi : U \to \mathbb{R}^d$ is the same, and is called the dimension of that connected component (for more details, see any of the references on manifolds). As a manifold M is assumed to be second-countable, it has at most a countable number of connected components. Recall that a submanifold is defined in the main text as follows (such a definition can be found in section 2.4.2 of [25]).

DEFINITION A.3. *Let M be a manifold of class $C^{k'}$, $k \leq k'$. N is a submanifold of M of class C^k if for any $x \in N$, and any chart $\phi : U \to \mathbb{R}^d$ defined around x and of class C^k, $\phi(U \cap N)$ is a submanifold (in the sense of $\mathbb{R}^d$; see* [15]*) of $\mathbb{R}^d$ around $\phi(x)$. It is equivalent to asking the existence of one such chart per point x.*

Let N be a submanifold of class C^k of a manifold M of class $C^{k'}$. Then it is naturally a manifold of class C^k in its own right. Indeed, consider that (i) N is equipped with the topology of M, i.e., V is open in N if and only if $V = U \cap N$ for some open set of M, and (ii) the atlas of N is (the completion of) the set of restrictions of charts $\phi|_{U\cap N}$, where $\phi : U \to \mathbb{R}^d$ is a C^k chart on M such that $\phi(U \cap N) \subset \mathbb{R}^{d'} \times \{0\}$, where d' is the dimension of N at $x \in U$ (we identify $\mathbb{R}^{d'} \times \{0\} \approx \mathbb{R}^{d'}$). The second-countable Hausdorff condition directly follows from that of M. Moreover, the C^k compatibility of the charts is evident. From now on, when considering a submanifold $N \subset M$ as a manifold, it will be with this structure. The reason for the introduction of all these concepts is to obtain the following lemma, which while it seems natural, we have not found as such in the literature.

LEMMA A.4. *Let N be a submanifold of a manifold M. Let $(N_i)_{i\in I}$ be the connected components of N. There exists a collection of disjoint open sets $(U_i)_{i\in I}$ of M such that each $N_i \subset U_i$.*

Proof. This proof relies mainly on paracompactness.

Step 1. For all $x \in N$, there exists U_x, an open set in M such that $\overline{U_x} \cap N$ is included in the unique connected component of x in N. Indeed, by Definition A.3, there exists a chart $\phi : U \to N$, where U is an open neighborhood of x. But since $\phi(U \cap N)$ is a submanifold of $\mathbb{R}^d$ of class C^k around $\phi(x)$, by Theorem 2.5 of [25], there exists a C^k diffeomorphism $\psi : (\phi(x), V) \to (0, W)$, where V is such that $\psi(\phi(U \cap N) \cap V) = W \cap (\mathbb{R}^{d'} \times \{0\})$ for some d'. Taking $\widetilde{\phi} = \psi \circ \phi$ on $\widetilde{U} = \phi^{-1}(V) \cap U$, we have a chart of class C^k around x such that $\widetilde{\phi} : \widetilde{U} \to W \subset \mathbb{R}^d$ such that $\widetilde{\phi}(\widetilde{U} \cap N) = W \cap (\mathbb{R}^{d'} \times \{0\})$. Now let r be a radius such that the closed ball $\overline{B}(0, r) \subset W$. Set $U_x = \widetilde{\phi}^{-1}(B(0, r))$, which is an open neighborhood of x included in $\widetilde{U}$. Note that $\overline{U_x} \subset \widetilde{\phi}^{-1}(\overline{B}(0, r)) \subset \widetilde{U}$ since $\widetilde{\phi}$ is continuous. Since $\overline{U}_x \cap N = \widetilde{\phi}^{-1}(\overline{B}(0, r) \cap (\mathbb{R}^{d'} \times \{0\}))$ which is connected, we have that $\overline{U}_x \cap N$ is connected and hence is included in the unique connected component of x.

Step 2. Consider the collection of open sets U_x. By paracompactness of $U := \bigcup_{x\in N} U_x$ (it is a manifold), we can find an open cover (U_α) of U which is locally finite, and which still satisfies the condition that for all α, $\overline{U}_\alpha \cap N$ is included in at most one connected component of N. Let $(N_i)_{i\in I}$ denote the connected components of N. For $i \in I$, let $(V_{i,\alpha})_{\alpha\in A_i}$ denote the collection of open sets U_α such that $\overline{U}_\alpha \cap N \subset N_i$ and $\overline{U}_\alpha \cap N \neq \emptyset$. These collections satisfy (a) the $(V_{i,\alpha})_{\alpha\in A_i}$ cover N_i; (b) the collection $(V_{i,\alpha})_{i\in I,\, \alpha\in A_i}$ has locally finite support; and (c) $\overline{V_{i,\alpha}} \cap N_i \subset N_i$ for all $i \in I$, $\alpha \in A_i$.

Step 3. For all $i \in I$, define $F_i = \bigcup_{j\in I\setminus\{i\},\beta\in A_j} \overline{V}_{j,\beta}$ and for all $\alpha \in A_i$, consider that the set $W_{i,\alpha} = V_{i,\alpha} \setminus F_i$. $W_{i,\alpha}$ is open, and $W_{i,\alpha} \cap N = V_{i,\alpha} \cap N$. Indeed, let $x \in W_{i,\alpha}$. Since the $(V_{j,\beta})$ are locally finite, there exists $V_x \subset V_{i,\alpha}$ such that V_x

intersects a finite number of the $V_{j,\beta}$ and hence of the $\overline{V}_{j,\beta}$. Hence, $V_x \setminus F_i$ is still open. Hence $W_{i,\alpha}$ is open. The second condition comes from the fact that $\overline{V}_{i,\alpha} \cap N \subset N_i$, and that the connected components are disjoint Finally, taking $W_i = \bigcup_{\alpha \in A_i} W_{i,\alpha}$, the W_i satisfy all the desired properties (they are disjoint thanks to the previous point and cover N_i since the $V_{i,\alpha}$ covered N_i). ∎

Appendix B. Morse lemma. In order for this article to be self-contained, we restate the following classical lemmas from differential geometry and topology. Recall that a C^k-diffeomorphism is a map $\phi : U \subset \mathbb{R}^d \to V \subset \mathbb{R}^{d'}$ which is of class C^k and whose inverse is of class C^k (in that case, necessarily, $d = d'$). The following results are classical.

THEOREM B.1 (Theorem 1.13 of [15]). *Let $f : (x_0, \mathbb{R}^d) \to \mathbb{R}^d$ be a function of class C^k ($k \geq 1$) defined around x_0 and such that $df(x_0)$ is invertible. Then there exists a neighborhood U of x_0 such that $f(U)$ is open and $f : U \to f(U)$ is a C^k diffeomorphism.*

THEOREM B.2 (Theorem 1.18 of [15]). *Let $f : (x_0, \mathbb{R}^{d_1}) \to (y_0, \mathbb{R}^{d_2})$ be a function of class C^k ($k \geq 1$) defined around x_0 s.t. $df(x_0)$ is surjective and $f(x_0) = y_0$. Then there exists an open neighborhood U of x_0 in $\mathbb{R}^{d_1}$, V of y_0 in $\mathbb{R}^{d_2}$, as well as a function $g : V \to U$ of class C^k such that $g(y_0) = x_0$ and $f \circ g = Id_{\mathbb{R}^{d_2}}$.*

We restate and reprove Lemma C.6.1 from [13], which is a generalization of the so-called Morse lemma (see Lemma 2.2 of [21]), and which is the basis of Morse theory. We will consider a function of two variables $f(x,y)$ defined on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. We will denote with $\nabla_x f(x,y)$ its gradient with respect to the first variable taken at point (x,y); it is an element of $\mathbb{R}^{d_1}$. Similarly, we will use the notation $\nabla^2_{xx} f(x,y) \in \mathbb{R}^{d_1 \times d_1}$ to denote the Hessian matrix taken with respect to the first coordinate at point (x,y). It is symmetric.

LEMMA B.3 (Lemma C.6.1 from [13]). *Let $d_1, d_2 \in \mathbb{N}, p \in \mathbb{N} \cup \{\infty\}$ with $p \geq 2$. Let $f : (x,y) \in U_0 \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \mapsto f(x,y) \in \mathbb{R}$ be a C^p function defined on a neighborhood U_0 of $(0,0)$. Assume that $\nabla_x f(0,0) = 0$ and that $\nabla^2_{xx} f(0,0)$ is nonsingular with index s (that is with s positive eigenvalues).*

There exists an open convex neighborhood V of 0 in $\mathbb{R}^{d_1}$ and an open convex neighborhood W of 0 in $\mathbb{R}^{d_2}$ such that $V \times W \subset U_0$, a map $\varphi \in C^{p-1}(W, V)$, and a map $z \in C^{p-2}(V \times W, \mathbb{R}^{d_1})$ such that for any $(x,y) \in V \times W$ $\nabla_x f(x,y) = 0$ if and only if $x = \varphi(y)$, and

$$\forall (x,y) \in V \times W,\ f(x,y) = f(\varphi(y), y) + \sum_{i=1}^{s} z_i(x,y)^2 - \sum_{i=s+1}^{d_1} z_i(x,y)^2. \tag{B.1}$$

To simplify the proof, we first show an intermediate result which gives φ.

LEMMA B.4. *Under the assumptions of Lemma* B.3*, there exist two open convex neighborhoods of zero, $V_0 \subset \mathbb{R}^{d_1}$, $W_0 \subset \mathbb{R}^{d_2}$, and $\varphi : W_0 \to V_0$ of class C^{p-1} such that (a) $V_0 \times W_0 \subset U_0$ and (b) $\forall (x,y) \in V_0 \times W_0$, $\nabla_x f(x,y) = 0 \Leftrightarrow x = \varphi(y)$.*

Proof. Consider the map $\psi : (x,y) \in U_0 \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \mapsto (\nabla_x f(x,y), y)$. Its Jacobian at $(0,0)$ is of the form $\left(\begin{smallmatrix} H & \star \\ 0 & \boldsymbol{I}_{d_2} \end{smallmatrix}\right)$. Since H is nonsingular, this matrix is nonsingular. Applying the local inversion lemma Theorem B.1, there exists an open neighborhood $U_1 \subset U_0$ such that ψ is a C^{p-1} diffeomorphism from U_1 to $\psi(U_1)$.

Let $\widetilde{V}_0 \subset \mathbb{R}^{d_1}$, $\widetilde{W}_0 \subset \mathbb{R}^{d_2}$ be open convex neighborhoods of 0 such that $\widetilde{V}_0 \times \widetilde{W}_0 \subset U_1 \cap \psi(U_1)$. Define $\varphi : w \in \widetilde{W}_0 \mapsto \pi_1(\psi^{-1}(0,w)) \in \mathbb{R}^{d_1}$. Defining $V_0 = \tilde{V}_0$ and $W_0 \subset \mathbb{R}^{d_2}$

to be an open convex neighborhood of 0 included in $\varphi^{-1}(V_0)\cap\widetilde{W}_0$, we have $\varphi(W_0)\subset V_0$ and $V_0\times W_0\subset U_1\subset U_0$.

Moreover, for any $(x,y)\in V_0\times W_0\subset U_1\cap\psi(U_1)$, $\nabla_1 f(x,y)=0$ if and only if $\psi(x,y)=(0,y)\in\psi(U_1)$, if and only if $(x,y)=\psi^{-1}(0,y)=(\varphi(y),y)$, if and only if $x=\varphi(y)$. ∎

We can now prove our main result.

Proof of Lemma B.3. Fix V_0,W_0 satisfying the properties of Lemma B.4. Let $(x,y)\in V_0\times W_0$. For $t\in[0,1]$, define $x_t=\varphi(y)+t(x-\varphi(y))$. By the convexity of V_0, $(x_t,y)\in V_0\times W_0\subset U_1\subset U_0$ for all $t\in[0,1]$. Thus, the map $g: t\in[0,1]\mapsto f(x_t,y)$ is well defined, and we can apply the Taylor formula $g(1)=g(0)+g'(0)+\int_0^1(1-t)g''(t)dt$ and the fact that $g'(0)=\nabla_x f(\varphi(y),y)\cdot(x-\varphi(y))=0$ to obtain

$$f(x,y)=f(\varphi(y),y)+(x-\varphi(y))^\top\left(\int_0^1(1-t)\nabla^2_{xx}f(x_t,y)dt\right)(x-\varphi(y)).$$

Defining $B: V_0\times W_0\to S(\mathbb{R}^{d_1})$, such that $B(x,y):=2\int_0^1(1-t)\nabla^2_{xx}f(x_t,y)dt$, the previous equation can simply be written $f(x,y)=f(\varphi(y),y)+\frac{1}{2}(x-\varphi(y))^\top B(x,y)(x-\varphi(y))$. Note that $B\in C^{p-2}(V_1\times W_1,S(\mathbb{R}^{d_1}))$ and $B(0,0)=H$. Now define $G: R\in\mathbb{R}^{d_1\times d_1}\mapsto R^\top HR\in S(\mathbb{R}^{d_1})$ which is C^∞ and whose differential in $\boldsymbol{I}_{\mathbb{R}^{d_1}}$ is surjective (see [13]). Theorem B.2 shows there exists a neighborhood $\mathcal{O}$ of H in $S(\mathbb{R}^{d_1})$ and a C^∞ function $F:\mathcal{O}\to\mathbb{R}^{d_1\times d_1}$ such that $(G\circ F)(B)=B$ for all $B\in\mathcal{O}$. Let $V\subset\mathbb{R}^{d_1},\widetilde{W}\subset\mathbb{R}^{d_2}$ be two open convex neighborhoods of 0 such that $V\times\widetilde{W}\subset B^{-1}(\mathcal{O})$. Let W be an open convex neighborhood of 0 such that $W\subset\widetilde{W}\cap\varphi^{-1}(V)$ and define $\widetilde{z}(x,y)=(F\circ B)(x,y)(x-\varphi(y))$. The function $\widetilde{z}$ satisfies, by the definition of F, $f(x,y)=f(\varphi(y),y)+\frac{1}{2}\widetilde{z}(x,y)^\top H\widetilde{z}(x,y)$. Taking an eigendecomposition of H in the form $H=\sum_{i=1}^s|\lambda_i|u_iu_i^\top-\sum_{i=s+1}^{d_1}|\lambda_i|u_iu_i^\top$, where s is the index of H, and setting $z_i(x,y)=\sqrt{|\lambda_i|/2}u_i^\top\widetilde{z}(x,y)$, z satisfies (B.1). ∎

REFERENCES

[1] F. Acquistapace, F. Broglia, and J. F. Fernando, *Topics in Global Real Analytic Geometry*, Springer, Cham, Switzerland, 2022.

[2] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68 (1950), pp. 337–404.

[3] E. Artin, *Über die Zerlegung definiter Funktionen in Quadrate*, Abh. Math. Semin. Univ. Hambg., 5 (1927), pp. 100–115.

[4] F. Bach and A. Rudi, *Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials*, SIAM J. Optim., 33 (2023), pp. 2137–2159.

[5] E. Berthier, J. Carpentier, A. Rudi, and F. Bach, *Infinite-dimensional sums-of-squares for optimal control*, 2022 IEEE 61st Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2022, pp. 577–582.

[6] J.-M. Bony, *Sur l'inégalité de Fefferman-Phong*, Séminaire Équations aux dérivées partielles (Polytechnique) dit aussi "Séminaire Goulaouic-Schwartz," 1998–1999, http://www.numdam.org/item/SEDP_1998-1999____A3_0/.

[7] J.-M. Bony, *Sommes de carrés de fonctions dérivables*, Bull. Soc. Math. France, 133 (2005), pp. 619–639, http://bonyjm.perso.math.cnrs.fr/S2KR_bullSMF.pdf.

[8] J.-M. Bony, F. Broglia, F. Colombini, and L. Pernazza, *Nonnegative functions as squares or sums of squares*, J. Funct. Anal., 232 (2006), pp. 137–147.

[9] C. Fefferman and D. H. Phong, *On positivity of pseudo-differential operators*, Proc. Natl. Acad. Sci. USA, 75 (1978), pp. 4673–4674, https://doi.org/10.1073/pnas.75.10.4673.

[10] D. Henrion, M. Korda, and J. B. Lasserre, *The Moment-SOS Hierarchy*, World Scientific, Singapore, 2020, https://doi.org/10.1142/q0252.

[11] D. Hilbert, *Ueber die Darstellung definiter Formen als Summe von Formenquadraten*, Math. Ann., 32 (1888), pp. 342–350, https://doi.org/10.1007/BF01443605.

[12] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math., 119 (1967), pp. 147–171.
[13] L. Hörmander, *The Analysis of Linear Partial Differential Operators* III, Grundlehren Math. Wiss., Springer, Berlin, 2007.
[14] K. Jänich, *Topologie*, Springer, Berlin, 1980.
[15] J. Lafontaine, *An Introduction to Differential Manifolds*, Springer, Cham, Switzerland, 2015.
[16] J.-B. Lasserre, *Moments, Positive Polynomials and their Applications*, Vol. 1, World Scientific, London, 2010.
[17] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, Springer, Cham, Switzerland, 2015.
[18] M. Marshall, *Representations of non-negative polynomials having finitely many zeros*, Ann. Fac. Sci. Toulouse Math (6), 15 (2006), pp. 15–609, https://doi.org/10.5802/afst.1131.
[19] U. Marteau-Ferey, F. Bach, and A. Rudi, *Non-parametric models for non-negative functions*, Adv. Neural Inform. Process. Syst., 33 (2020), pp. 12816–12826.
[20] U. Marteau-Ferey, F. Bach, and A. Rudi, *Sampling from Arbitrary Functions via PSD Models*, Proc. Mach. Learn. Res. (PMLR), 151 (2022), pp. 2823–2861.
[21] J. W. Milnor, *Morse Theory*, Based on Lecture Notes, Ann. of Math. Stud. 51, M. Spivak and R. Wells, eds., Princeton University Press, Princeton, NJ, 1963, https://www.maths.ed.ac.uk/∼v1ranick/papers/milnmors.pdf.
[22] J. Nie, *Optimality conditions and finite convergence of Lasserre's hierarchy*, Math. Program., 146 (2014), pp. 97–121, https://doi.org/10.1007/s10107-013-0680-x.
[23] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Math. 1349, Springer, Berlin, 2006.
[24] P. A. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Math. Program., 96 (2003), pp. 293–320, https://doi.org/10.1007/s10107-003-0387-5.
[25] F. Paulin, *Géométrie Différentielle Élémentaire*, 2006, manuscript, https://www.imo.universite-paris-saclay.fr/∼frederic.paulin/notescours/cours_geodiff.pdf.
[26] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana Univ. Math. J., 42 (1993), pp. 969–984.
[27] A. Rudi and C. Ciliberto, *PSD representations for effective probability models*, Adv. Neural Inform. Process. Syst., 34 (2021), pp. 19411–19422.
[28] A. Rudi, U. Marteau-Ferey, and F. Bach, *Finding Global Minima via Kernel Approximations*, https://arxiv.org/abs/2012.11978, 2020.
[29] J. M. Ruiz, *On Hilbert's 17th problem and real nullstellensatz for global analytic functions*, Math. Z., 190 (1985), pp. 447–454, https://doi.org/10.1007/BF01215144.
[30] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Springer, Cham, Switzerland, 2015, https://doi.org/10.1007/978-3-319-20828-2.
[31] K. Schmügden, *The k-moment problem for compact semi-algebraic sets*, Math. Ann., (1991), pp. 203–206.
[32] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, 3rd ed., Vol. 1, Publish or Perish, Houston, TX, 1999.
[33] G. Stengle, *A Nullstellensatz and a Positivstellensatz in Semialgebraic Geometry*, Math. Ann., 207 (1974), pp. 87–98, http://eudml.org/doc/162533.
[34] D. Tataru, *On the Fefferman-Phong inequality and related problems*, Comm. Partial Differential Equations, 27 (2002), pp. 2101–2138, https://doi.org/10.1081/PDE-120016155.
[35] A. Vacher, B. Muzellec, A. Rudi, F. Bach, and F.-X. Vialard, *A dimension-free computational upper-bound for smooth optimal transport estimation*, Proc. Mach. Learn. Res. (PMLR), 134 (2021), pp. 4143–4173, https://proceedings.mlr.press/v134/vacher21a.html.