

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 non-negative function as a sum of squares of functions which preserve some form of regularity. In the same way as decomposing non-negative polynomials as sum of squares of polynomials allows to derive methods in order to solve global optimization problems on polynomials, decomposing a regular function as a sum of squares allows 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 analysing 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 non-negative 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. Non-negative functions, manifolds, sum of squares, global optimization, second order conditions

1. Introduction. The relationship between non-negative 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 non-negative function in terms of sum of squares is the basis of important theoretical objects and properties: quadratic modules [15] in algebraic geometry, regularizing operators such as Laplacians or sub-Laplacians in (sub-)Riemannian geometry [4, 10], non-negative symbols in pseudo-differential calculus [11, 28]. From an applicative viewpoint, representing a non-negative function in terms of sum of squares allows to simplify the analysis of probability representations and optimization problems [14, 16]. Restricting to the case of non-negative polynomials this has been applied to global optimization and generalized methods of moments [8, 14]. More generally, the decomposition of non-negative p -times differentiable functions allowed to derive simple and fast optimization algorithms in the context of global optimization [24], the Kantorovich problem in optimal transport [29], some formulations of optimal control [3]. Moreover, it allowed to obtain an effective and concise representation for probability densities, with applications in probabilistic inference, sampling, machine learning [17, 23].

The importance of preserving regularity. In this work, we state sufficient conditions for a non-negative function f to be written as a sum of squares of functions f_i . Of course, if no other constraints are added, this is a trivial problem as writing $f = (\sqrt{f})^2$ would offer an immediate solution. What we want to understand here are sufficient conditions which allow to inherit a form of regularity of the function f in the sum of squares decomposition. This is not necessarily the case when taking the square root: for example, 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 $(0, 0)$.

*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 (ulyссе.marteau-feréy@inria.fr, francis.bach@inria.fr, alessandro.rudi@inria.fr).

In general, being able to decompose the function with a certain regularity is important. Of course, there is a complex interaction between the structural constraint of being a sum of squares and the original regularity of the function, and the two may not work very well together (see [Theorem 1.4](#)). However, for certain theoretical and applied problems, it is crucial to maintain some regularity. For example, in the setting introduced in [\[24\]](#), the speed of convergence of the presented algorithm of global optimization depends on the regularity of the sum of squares representation of the function $f - f_*$ (where f_* is the global minimum of f). In polynomial sum-of-squares (SoS) optimization, the running time depends on the degree in the sum of squares decomposition of $P - P_*$.

Abstractly, we can formulate the following generic question. If $f \in \mathcal{C}_1$ where \mathcal{C}_1 describes a form of regularity, can we write f as a sum of squares of functions of class \mathcal{C}_2 , where \mathcal{C}_2 inherits the regularity properties \mathcal{C}_1 as much as possible?

Problem setting. In this work, we will concentrate on the class C^p of p times differentiable functions with continuous p -th derivatives on \mathbb{R}^d (or any d -dimensional manifolds M). For simplicity, in this introduction, we will state the main results for functions on \mathbb{R}^d . We will show that under a certain condition on the set of zeros \mathcal{Z} of f , if f is a C^p non-negative function on \mathbb{R}^d , it can be decomposed as

$$(1.1) \quad f = \sum_{i \in I} f_i^2, \quad f_i \in C^{p-2}(\mathbb{R}^d),$$

where (f_i) is an at most countable family and has locally finite support. Two elements are important in [\(1.1\)](#): 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.

1.1. Intuition and previous results. Let us give an intuition as to how we obtain decompositions in the form [\(1.1\)](#). First, 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](#) (note that 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 [\(1.1\)](#) 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 non-negative, all such points are necessarily minimizers of f , hence the following *necessary second-order condition*:

$$(1.2) \quad \forall x_0 \in \mathcal{Z}, \nabla f(x_0) = 0, \nabla^2 f(x_0) \succeq 0.$$

Around any $x_0 \in \mathcal{Z}$, f can be approximated by a parabola since the eigenvalues of $\nabla^2 f(x_0)$ are non-negative: $f(x) = x^\top \nabla^2 f(x_0)x + o(\|x\|^2)$ using a Taylor expansion. Since any parabola can be written as the sum of at most d squares of linear functions (just write the eigen-decomposition 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.

It turns out that in the case where $\nabla^2 f(x_0) \succ 0$, that is when 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 [\[24\]](#), applied to the set $\mathcal{H} = C^{p-2}$. Precisely, it states

THEOREM 1.1 (Theorem 2 of [24]). *Let f be a non-negative function of class C^p for $p \geq 2$, and assume that the zeros \mathcal{Z} of f satisfy the strict Hessian condition:*

$$(1.3) \quad \forall x_0 \in \mathcal{Z}, \nabla^2 f(x_0) \succ 0.$$

If f has a finite number $m = |\mathcal{Z}|$ of zeros, then f satisfies (1.1) 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 1.1](#), 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.

Contribution. While the SHC condition (1.3) already offers a nice result in [Theorem 1.1](#), we see that there is a big difference with the necessary condition (1.2). Previous results in the literature show that (1.2) 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 1.4](#) in the background section for more details). On the other hand, (1.3) is very restrictive. In particular, it implies that the set \mathcal{Z} of zeros is discrete. However, in some situations such that of [29], the set of zeros has a natural structure, which can be a sub-manifold of \mathbb{R}^d (consider for instance the extreme case where $f = 0$). In this paper, we show that if the set \mathcal{Z} of zeros is a sub-manifold 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 (1.1) 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 1.2. *Let f be a non-negative function of class C^p for $p \geq 2$ and let \mathcal{Z} denote the set of zeros of f . If \mathcal{Z} is a sub-manifold of \mathbb{R}^d of class C^1 such that*

$$(1.4) \quad \forall x_0 \in \mathcal{Z}, \forall h \in \mathbb{R}^d \setminus T_{x_0} \mathcal{Z}, h^\top \nabla^2 f(x_0) h > 0,$$

then f satisfies (1.1), 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 sub-space of \mathbb{R}^d .

This theorem is proved as [Theorem 2.8](#) in [section 2](#), and the assumption (1.4) will be referred to as the normal Hessian condition (or NHC). Note that the NHC assumption encompasses that of the SHC assumption of [Theorem 1.1](#); in that case, the results presented in this paper make the result tighter by removing the assumption that \mathcal{Z} be finite and by needing only $d + 1$ squares to represent the function, and not $d|\mathcal{Z}| + 1$ (see the full version of [Theorem 2.8](#)).

The proof techniques used to prove this theorem differ from the proof of [24] 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 $\mathbb{T}^d \approx (S^1)^d$.

Background. 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 [7] for the original article, and [4] for the link with sum of squares decompositions, as well as [28]). In this context, the following result is proved (with $C_{loc}^{k,1}$ denoting the set of k times differentiable functions with locally Lipschitz k -th derivative):

THEOREM 1.3 (Fefferman-Phong [7], Theorem 1.1 of [6]). *Let Ω be an open set of \mathbb{R}^d , $d \geq 1$ and $f \in C_{loc}^{3,1}(\Omega)$ be a non-negative function. Then f can be written as a finite sum of squares of $C_{loc}^{1,1}(\Omega)$ functions.*

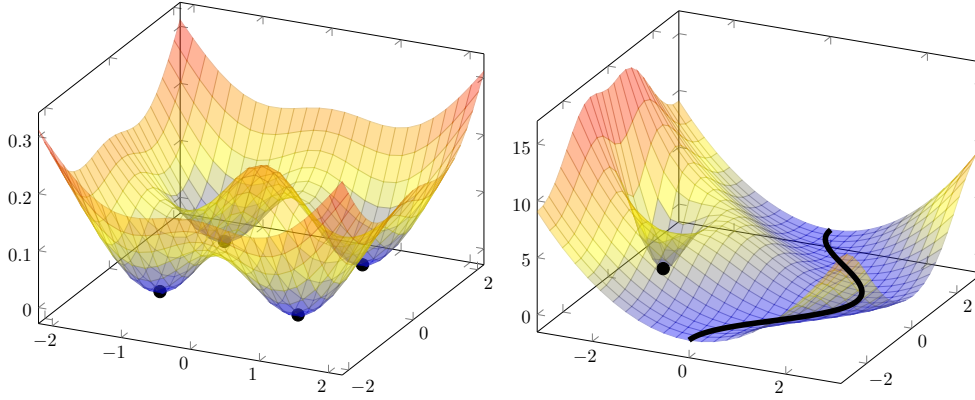


FIGURE 1. Plots of functions $z = f(x, y)$, where the zeros of f are highlighted in black. **left:** f satisfies the SHC, **right:** f satisfies the NHC but not the SHC.

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 square decomposition. In [5, 6], 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 1.4 (Theorem 2.1 of [6]). *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 holds:*

- 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 [5]: 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 [5]).

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.

Polynomials. Decomposing non-negative polynomials 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 [2] proved that any non-negative polynomial is a sum of squares of rational functions (that is formal fractions of polynomials $P(x)/Q(x)$). Moreover, Hilbert had earlier proved that there exist non-negative polynomials which cannot be written as sum of squares of polynomials in [9] (for more than 3 variables and with degree at least 6 for example). In algebraic geometry, the set of SoS polynomials has very interesting properties, and finding sufficient conditions for a non-negative polynomial or even a positive polynomial to be a sum of squares is an important question. More generally, one usually wishes to understand under which sufficient conditions a polynomial P which is non-negative

(or 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 SoS. The theoretical literature regroups these results under the name "Positivstellensatz". The most often seen in the SoS optimization literature are the Stengel [27], Schmügdén [25] and Putinar [22] 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. Indeed, on the one hand, the general impossibility results proved in [5, 6] (see [Theorem 1.4](#)) are obtained using Hilbert's theorem on the existence of non-negative polynomials which are not sum of squares. On the other hand, the fact that there is hope using our second-order assumptions is also due to the fact that second order non-negative polynomials can always be written as sums of squares.

Second, the certificates given by Positivstellensatz on the decomposability of certain non-negative polynomials can be algorithmically checked in some cases, using semi-definite programming. This has paved the way to so-called SoS hierarchies, and optimization of polynomial objective functions with polynomial constraints. These have been developed by Lasserre [14] (based on the Putinar Postivstellensatz [22]) and Parrillo [20] (based on the Stengel and Schmügdén Positivstellensatz [25, 27]). Using these theoretical results, they can provide certificates of lower bounds for certain optimization problems (or upper bound in the dual "moment problem", see [14]). Moreover, to have more interpretable results for these more applied settings, these works have motivated more practical Positivstellensatz, like that in [15], which provides a condition for writing a polynomial with a finite set of zeros as as sum of squares (this condition is actually a second order condition which greatly resembles ours in the polynomial setting, although it deals more with the constraints Q_i).

p-times differentiable functions. In the same spirit as the polynomial hierarchies, recent works [16, 23, 24] have developed models and methods based on sum of squares of regular functions. The computational properties of these methods are based on the fact that regular functions 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 [1] and can be adapted to the regularity. In order to obtain guarantees on these methods, it is crucial to have the equivalent of Positivstellensatz in the case of regular functions. Contrary to the case of algebraic geometry, where such results existed for other purposes, there is a need to build such results for regular functions from scratch. Certain results like [Theorem 1.1](#) have been presented. However, the aim of the present paper is to provide more general results, to be used in most situations.

Organisation of the work. In [section 2](#), we formalize the different notions needed to state [Theorem 1.2](#) in the case of non-negative 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 1.2](#) 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 . We start with a brief presentation of the notion of

sub-manifold of \mathbb{R}^d in [subsection 2.1](#). It is the key geometric object we use to represent the set of zeros \mathcal{Z} of the function f . In [subsection 2.2](#), we present the cornerstone result of this paper in [Theorem 2.3](#), as well as a sketch of its proof, which is done extensively in [section 4](#). It shows that as soon as a non negative function has positive Hessian in the orthogonal direction to its zeros at a given point, then it can be decomposed as a sum of squares around that point. Finally, in [subsection 2.3](#), we present [Theorem 2.8](#), which shows that given a function defined on an open subset Ω of the Euclidean space \mathbb{R}^d , and under conditions on the Hessian of f at its zeros, f can be decomposed as a locally finite sum of squares of functions defined on Ω .

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 $\phi : (x_0, M) \rightarrow (y_0, N)$ is a local map satisfying a property (P) if there exists an open neighborhood U of x_0 in M such that $\phi : U \rightarrow N$ is well defined, satisfies $\phi(x_0) = y_0$ and property (P) . We will say that $\phi : U \subset \mathbb{R}^d \rightarrow \mathbb{R}^e$ 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 $\phi : (x, \mathbb{R}^d) \rightarrow \mathbb{R}^e$ of class C^1 , we denote with $d\phi(x)$ its differential at x . It is an element of $\text{Hom}(\mathbb{R}^d, \mathbb{R}^e)$ 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 $[J_\phi]_{ij} = \frac{\partial \phi^i}{\partial x^j}(x)$.

2.1. Sub-manifolds of \mathbb{R}^d . One of the main assumptions in order to achieve our results will be that the set of zeros of the non-negative function f is a sub-manifold of \mathbb{R}^d . In this section, we restrict ourselves to introducing the definitions and results needed to state and prove those in this paper. For a more comprehensive introduction, see chapter 1 of [\[13\]](#), section 2.2 of [\[21\]](#) (in French) or [\[26\]](#). The notion of sub-manifold generalizes the notion of a curve in \mathbb{R}^d (a one dimensional manifold) or a surface in \mathbb{R}^d (a two dimensional manifold). Intuitively, a sub-manifold N is a subset of \mathbb{R}^d such that at each point $x \in N$, N “looks like” \mathbb{R}^{d_0} where d_0 is the dimension of the sub-manifold at x (one for a line, two for a surface, etc.). Another way to put this is that N can be locally parametrized by \mathbb{R}^{d_0} . To formalize this, we need the following definitions. We fix a subset $N \subset \mathbb{R}^d$.

A map $\phi : U \rightarrow \mathbb{R}^d$ defined on an open neighborhood U of 0 in \mathbb{R}^{d_0} is said to be a local parameterization 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 $\phi : U \rightarrow \phi(U)$ is a homeomorphism, i.e., it is bijective and has continuous inverse;
- (ii) its differential at 0 is injective (one to one), i.e., $d\phi(t_0) \in \text{Hom}(\mathbb{R}^{d_0}, \mathbb{R}^d)$ is injective.

The second condition guarantees that the local dimension of N is indeed d_0 , that ϕ is not an over-parameterization. N is said to be a sub-manifold of \mathbb{R}^d and of class C^k if there exists a local parameterization ϕ of class C^k around each point $x \in N$. Given a point $x \in N$, the dimension d_x of the local parametrization is independent of the parametrization (two local parametrizations will necessarily be of same dimension); it is called the dimension of N at x . Similarly, the subspace $T_x N := d\phi(x)\mathbb{R}^{d_x}$, which is a subspace of \mathbb{R}^d of dimension d_x is independent of the local parametrization: it is the linear approximation of N at x and is called the tangent space to N at x (see [Figure 2](#) and [Figure 3](#) for more visual representations).

A sub-manifold N of \mathbb{R}^d is said to be connected if it cannot be written as a union of disjoint open sets. Equivalently, it is connected if any two points in N can be connected by a continuous path $\gamma : [0, 1] \rightarrow N$. On a connected sub-manifold N , the dimension d_x is the same at every

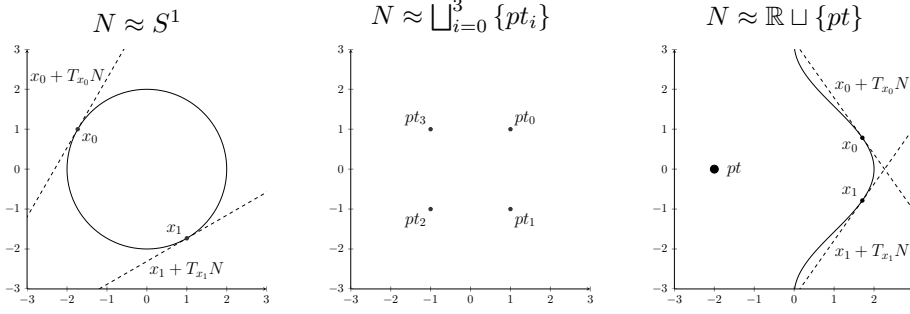


FIGURE 2. Examples of sub-manifolds of \mathbb{R}^2 ; points are denoted with pt . **Left:** connected sub-manifold of dimension 1 (a circle). **Center:** a sub-manifold of 4 connected components which are all points, i.e., of dimension 0 (their tangent space is not represented since it is reduced to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

point x , it is called the dimension of the connected sub-manifold N . This implies that all the tangent spaces $T_x N$ have the same dimension.

Example 2.1. All open sets of \mathbb{R}^d are sub-manifolds of \mathbb{R}^d . The d -dimensional sphere S^d is a sub-manifold 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 sub-manifold N of \mathbb{R}^d , the intersection of N with any open set of \mathbb{R}^d is a sub-manifold of \mathbb{R}^d .

If U_i is a family of disjoint open sets each containing a connected sub-manifold N_i of \mathbb{R}^d , it is clear the the disjoint union $\sqcup_{i \in I} N_i$ is also a sub-manifold. Conversely, any sub-manifold can be decomposed into its connected components N_i ; moreover, one can find a family of disjoint open sets U_i such that $N_i \subset U_i$ (see Lemma A.4).

These results, their proof and their broader context can be found in [13] in chapter 1.5. In particular, Theorem 1.21 presents equivalent definitions of a sub-manifold. Section 2.2 of [21] is also a good reference (in French).

2.2. Local decomposition as a sum of squares. In this section, f will always denote a non-negative function defined on an open set of \mathbb{R}^d . We will assume that f is of class C^p for $p \geq 2$. We will also denote with \mathcal{Z} the set of zeros of f , i.e., 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 .

We will denote with $d^2 f(x)$ 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 $d^2 f(x)[\xi, \eta]$ 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 $d^2 f(x)$ in the canonical basis of \mathbb{R}^d , and we have $d^2 f(x)[\xi, \eta] = \eta^\top \nabla^2 f(x) \xi$. For any vector sub-space 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 semi-definite 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 2.2), let $x_0 \in \mathbb{R}^d$ and $f : (x_0, \mathbb{R}^d) \rightarrow \mathbb{R}$ be a non-negative C^p function, for $p \geq 2$, such that $f(x_0) = 0$. We claim that if there is a sub-manifold 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 $d - d_0$

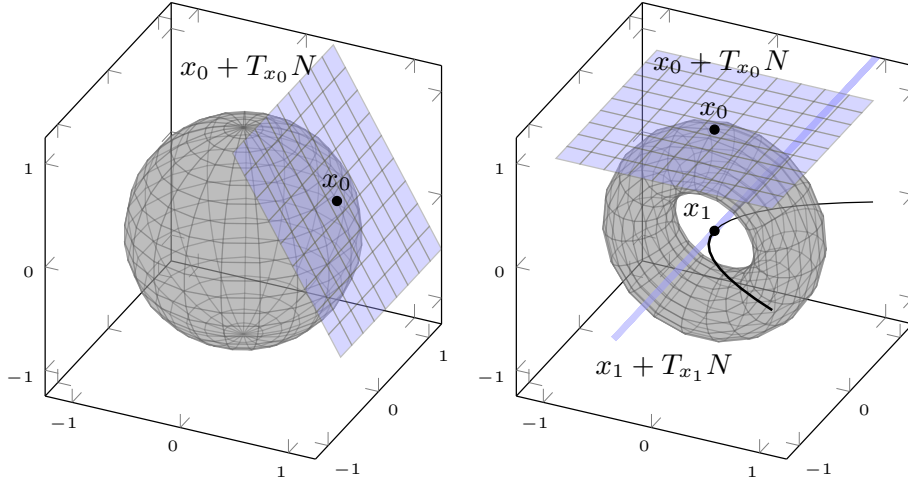


FIGURE 3. Two examples of sub-manifolds of \mathbb{R}^3 . The blue affine spaces represent tangent spaces. **Left:** connected sub-manifold of dimension 2 (the sphere S^1). **Right:** a sub-manifold of two connected components, one of dimension 2 (homeomorphic to the torus \mathbb{T}^2 on which lies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

(which we will call the normal Hessian condition), then it can be decomposed as a sum of squares as in (1.1).

DEFINITION 2.2 (normal Hessian condition). *Let \mathcal{Z} denote the set of zeros of f . We say that f satisfies the normal Hessian condition (NHC) at x_0 if there exists a dimension $0 \leq d_0 \leq d$ and a sub-manifold N of class C^k with $k \geq 1$ and of dimension d_0 such that $x_0 \in N \subset \mathcal{Z}$, and on one of the following equivalent conditions is satisfied:*

- (i) *the rank of $\nabla^2 f(x_0)$ at x_0 is $d - d_0$;*
- (ii) *the restriction of $d^2 f(x_0)$ to $T_{x_0}N^\perp$ is positive definite.*

The complete proof of the equivalence of these conditions as well as the proof of [Theorem 2.3](#) can be found in [section 4](#). To illustrate the definition of the normal Hessian condition, 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. *If f satisfies the NHC at x_0 ([Definition 2.2](#)) with regularity k and dimension d_0 , there exists an open neighborhood U of x_0 in \mathbb{R}^d on which f is defined and such that $U \cap \mathcal{Z}$ is a sub-manifold 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 $1 \leq i \leq d - d_0$ such that*

$$(2.1) \quad \forall x \in U, f(x) = \sum_{i=1}^{d-d_0} f_i^2(x).$$

Main steps of the proof. 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. Re-parametrizing f on a basis adapted to $T_{x_0}N^\perp \oplus T_{x_0}N$ as $f(x_\perp, x_\parallel)$, we apply the Morse lemma (see [Lemma B.3](#)), which decomposes the function f in the form

$$(2.2) \quad f(x_\perp, x_\parallel) = f(\varphi(x_\parallel), x_\parallel) + \frac{1}{2}d^2 f(x_0)|_{T_{x_0}N^\perp}[\xi(x_\perp, x_\parallel), \xi(x_\perp, x_\parallel)],$$

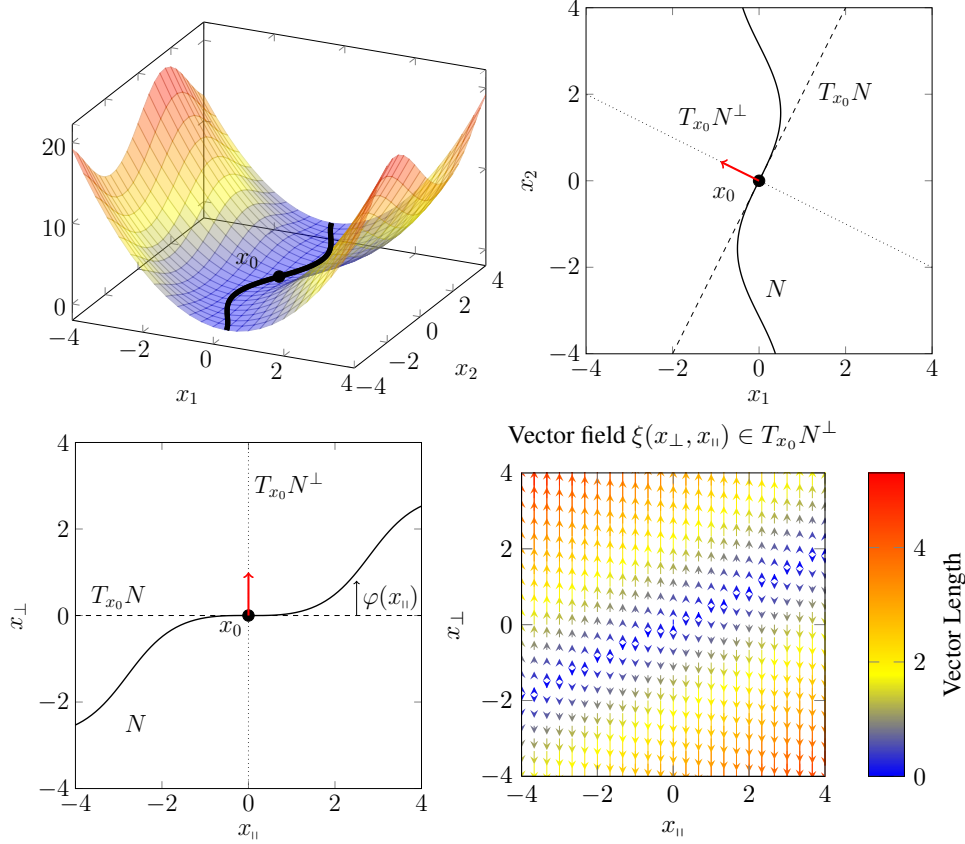


FIGURE 4. Local view of the function around a minimum lying on a 1-dimensional manifold. **Top left:** function around the minimum x_0 . **Top right:** decomposition of \mathbb{R}^2 at x_0 between tangent space and normal tangent space $T_{x_0}N + T_{x_0}N^\perp$, and positive eigen-vector of the Hessian in red. **Bottom left:** reparametrization in the right coordinate system, and representation of the map $\varphi(x)$ given by the Morse Lemma. **Bottom right:** vector field ξ given by the Morse lemma.

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

Step 3. We show that the second term of the right hand side of (2.2) can actually be seen as a sum of squares of $d - d_0$ functions of class C^{p-2} .

Step 4. We characterize the manifold of zeros around x_0 .

Step 5. We show that the first term of the result of the Morse lemma is equal to zero using the previous characterization, which shows (2.1). \square

Example 2.4 (case where $d_0 = 0$). When $d_0 = 0$, the NHC at x_0 is simply the SHC (1.3), 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 2.5 (Smoothing effect). Note that 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 satisfying the NHC, one sees that this manifold is actually of class C^{p-1} in a neighborhood of x_0 .

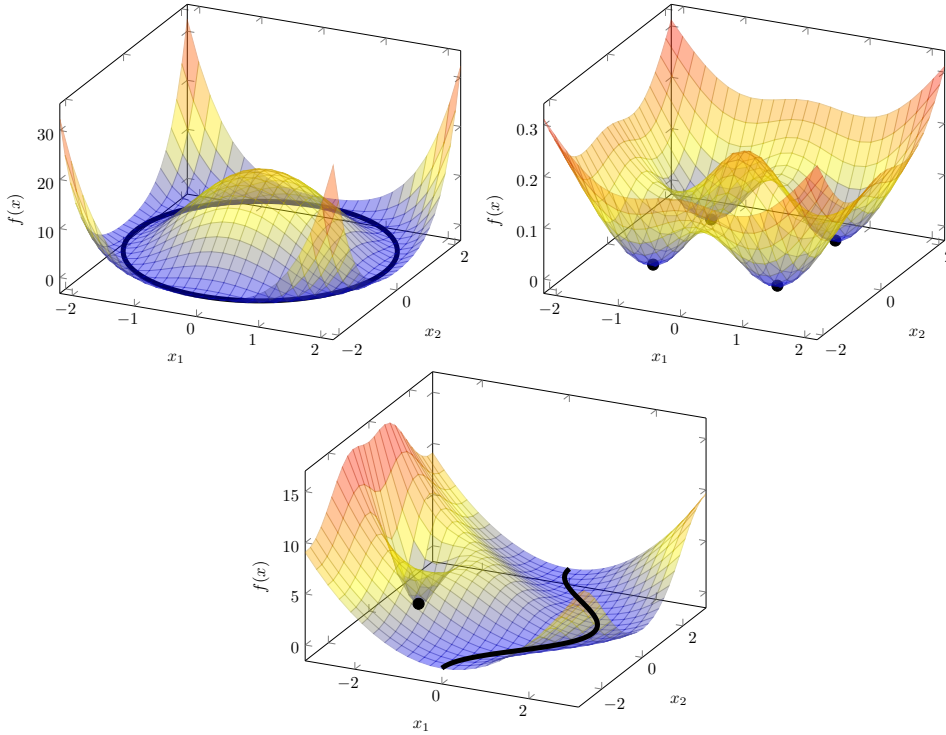


FIGURE 5. Example of functions f which satisfy the global normal Hessian condition, with sub-manifolds \mathcal{Z} of zeros corresponding to the sub-manifolds presented in Figure 2 in the same order

2.3. Global decomposition as a sum of squares for functions on \mathbb{R}^d . In this section, we fix f to be a non-negative C^p function defined on an open subset Ω of \mathbb{R}^d . Once again, we assume $p \geq 2$. The goal is to find conditions on f to be written as a sum of squares of functions defined on Ω . These conditions will be that the NHC holds at every $x_0 \in \mathcal{Z}$. We will start by reformulating this assumption in a more global and geometric way. We introduce the following definition of a manifold to which f is positively normal.

DEFINITION 2.6 (positive normally to a sub-manifold). *Let N be a sub-manifold of \mathbb{R}^d of class C^k for $k \geq 1$ and included in Ω . We say that f is positive normally to N if:*

- a) N is included in the set of critical points of f ($df(x) = 0$ for all $x \in N$);
- b) for any $x_0 \in N \cap \Omega$, if d_0 is the local dimension of N at x_0 , there exists a subspace $S \subset \mathbb{R}^d$ of dimension $d - d_0$ such that $d^2 f(x_0)|_S$ positive definite.

The intuition of this definition is that if f is positively normal to N , then f grows quadratically normally to N , which is a local minimum valley on which f is constant. Note that there can be more than one connected component in N : this will correspond to multiple local minima valleys (see second and third examples in Figure 5). We now reformulate the fact that the NHC holds at every point in \mathcal{Z} as a more geometric global assumption, using Definition 2.6.

LEMMA 2.7 (Global normal Hessian condition). *The following statements are equivalent and define the global NHC condition:*

- (i) for all $x_0 \in \mathcal{Z}$, f satisfies the NHC;

- (ii) \mathcal{Z} is a sub-manifold of \mathbb{R}^d (not necessarily connected) of class C^1 such that the Hessian of f is positive normally to \mathcal{Z} ;
- (iii) \mathcal{Z} is a sub-manifold of \mathbb{R}^d (not necessarily connected) of class C^{p-1} such that the Hessian of f is positive normally to \mathcal{Z} .

This equivalence is a direct consequence of the local description of \mathcal{Z} obtained in [Theorem 2.3](#) under the local NHC. Examples of functions satisfying the global normal Hessian condition can be found in [Figure 5](#). They have manifolds of zeros which are depicted in the same order in [Figure 2](#). Under this geometric condition, we will show in [Theorem 2.8](#) 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 [\[12\]](#) 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 $(\text{supp}(f_i))_{i \in I}$ is locally finite (recall that $\text{supp}(f_i) = \overline{\{x : f_i(x) \neq 0\}}$). 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 2.8. *If f satisfies the global normal Hessian condition in [Lemma 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*

$$(2.3) \quad \forall x \in \Omega, f(x) = \sum_{i \in I} f_i(x)^2.$$

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 an 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 $f_1 := \sqrt{f} : U_{>0} \rightarrow \mathbb{R}$ 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 of 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). \square

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 on the terminology and definitions related to manifolds, in [subsection 3.1](#), we will adapt the definitions of the local and global normal Hessian conditions, 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 as sum of squares together. Finally, in [subsection 3.3](#), we prove [Theorem 3.9](#), the equivalent of [Theorem 2.8](#) in the broader context of manifolds.

Additional definitions and notations for manifolds. In this section, we introduce the basic definitions we will need concerning manifold. For more formal introductions to manifolds, we refer to [[13](#), [21](#), [26](#)]. Informally, a manifold 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 topologically maps a neighborhood U of x to an open set of \mathbb{R}^d .

More generally, we define a chart on a topological space M as a map $\phi : U \rightarrow \mathbb{R}^d$ for some $d \in \mathbb{N}$, defined on an open set U of M , and which is a homeomorphism onto its image. We define a manifold M as a second-countable¹, Hausdorff² topological space equipped with a collection $\mathcal{A} = (\phi_i)_{i \in I}$ of charts such that

- (i) all transition maps $\phi_i \circ \phi_j^{-1} : \phi_j(U_j \cap U_i) \rightarrow \phi_i(U_i \cap U_j)$ are homeomorphisms;
- (ii) the charts cover M entirely, i.e. $M = \bigcup_{i \in I} U_i$.

The set \mathcal{A} is called an atlas. The manifold M is said to be of class C^k for $k \geq 0$ if all the transition maps $\phi_i \circ \phi_j^{-1}$ are of class C^k . It is said to be of dimension d if all its charts are in \mathbb{R}^d . As for sub-manifolds of \mathbb{R}^n , a manifold can always be decomposed as the union of its connected components, and the dimension is the same on each connected component. Note that with this definition, the restriction of a manifold to any open set is still a manifold (just by restricting the charts).

If M is a manifold of class at least C^1 , we can define at each point its tangent space $T_x M$. Informally, this set $T_x M$ is all the possible derivatives $\gamma'(0)$ of curves $\gamma : I \rightarrow M$ defined on an open interval I around 0 such that $\gamma(0) = x$. Of course $\gamma'(0)$ is not yet formally defined. Formally, $T_x M$ can be defined as the classes of C^1 curves defined on an open interval I around 0 such that $\gamma(0) = x$, where we identify two curves γ and $\tilde{\gamma}$ if $(\phi \circ \gamma)'(0) = (\phi \circ \tilde{\gamma})'(0)$ for a (or equivalently any) chart ϕ of M around x . We denote with $[\gamma] \in T_x M$ the equivalence class of a curve γ . It can be shown that $T_x M$ is a vector space (with the natural definition $\lambda[\gamma] + \mu[\tilde{\gamma}] = [\lambda\gamma + \mu\tilde{\gamma}]$) and that it is of dimension d where d is the dimension of M at x .

A map $g : M \rightarrow \mathbb{R}^p$ defined on a manifold M is said to be of class C^q if M is of class at least C^q and if for any chart ϕ , the map $g \circ \phi^{-1}$ is of class C^q . If $q \geq 1$, then we can define the differential

¹A 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 .

²A topological space is Hausdorff if for any two points $x \neq x'$, there exists two open sets U, V such that $x \in U$ and $x' \in V$ and $U \cap V = \emptyset$

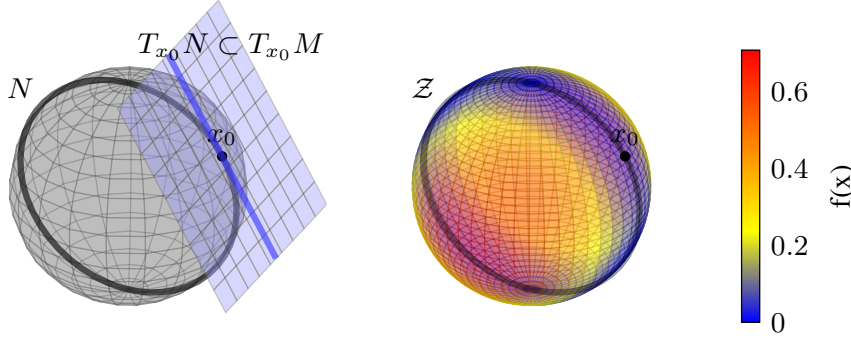


FIGURE 6. **Left:** Representation of the manifold $M = S^2$ as well as a sub-manifold N homeomorphic to a circle. The tangent spaces at a given point $x_0 \in N \subset M$ are represented as well. **Right:** Representation of a non-negative function on the sphere as a color map; it satisfies the NHC, and its null space Z is represented in black.

of g at any point $x \in M$ as $df(x)[\gamma] = (f \circ \gamma)'(0)$. Hence, $df(x) \in \text{Hom}(T_x M, \mathbb{R}^p)$.

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

For more precise definitions of topological spaces, atlases, charts, and details on the ‘‘Hausdorff second-countable’’ condition, see for instance [12, 13, 21, 26]. The main idea behind the introduction of manifolds as opposed to sub-manifolds 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.

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 (Lemma 2.7).

Fix $p \in \mathbb{N}$, $p \geq 2$, and a manifold M of regularity at least C^p and of dimension $d \in \mathbb{N}$. To start with, let $f : \Omega \rightarrow \mathbb{R}$ be a non-negative function defined on an open set of M and of class C^p . As before, define 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 to a symmetric bi-linear 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 non-negative function are critical points, this Hessian will be defined at all points in the set of zeros Z of f .

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

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

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 [18]. In order to formulate

the definition of the normal Hessian condition in the setting of manifolds, we further need a definition of what a sub-manifold of M is. A subset $N \subset M$ is said to be a sub-manifold of M of class C^k if M is of class C^k and if, for any $x \in N$ and any local chart $\phi : U \rightarrow \mathbb{R}^d$ defined on a neighborhood of x , $\phi(U \cap N)$ is a sub-manifold of \mathbb{R}^d of class C^k . In the literature, this is also called a proper sub-manifold (see on the left of [Figure 6](#) for an example).

DEFINITION 3.3 (Normal Hessian condition for a manifold). *Let $x \in \Omega$ be a point in the domain of f . We say that f satisfies the normal Hessian condition (NHC) at x if there exists a dimension $0 \leq d_0 \leq d$ and a sub-manifold N of M of class C^k with $k \geq 1$ and of dimension d_0 such that $x \in M \subset \mathcal{Z}$, and one of the following equivalent conditions is satisfied:*

- (i) *the rank of $H_f(x)$ is $d - d_0$;*
- (ii) *$H_f(x)[\xi, \xi] > 0$ for any vector $\xi \in T_x M \setminus T_x N$.*

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. *If f satisfies the NHC at x_0 with regularity k and dimension d_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 sub-manifold 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*

$$(3.1) \quad \forall x \in U, f(x) = \sum_{i=1}^{d-d_0} f_i^2(x).$$

Exactly in the same way as for the definition of the NHC for manifolds, we can similarly extend the definition of a function being positively normal to a sub-manifold in [Definition 2.6](#) as well as the global NHC in [Lemma 2.7](#). We will therefore say that $f : M \rightarrow \mathbb{R}$ which is non-negative satisfies the global NHC if it satisfies the local NHC at every point in its set of zeros \mathcal{Z} , or equivalently if it is positive normally to \mathcal{Z} which is a sub-manifold of M of class C^1 (or C^{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 2.3](#). The proof of these results can be found in [Appendix 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 , $g : U \rightarrow \mathbb{R}$ be a C^q function defined on U , and $\chi : M \rightarrow \mathbb{R}$ be a C^q function defined on the whole of M but with support included in U . Then the function $\chi g : U \rightarrow \mathbb{R}$ extended as 0 on $M \setminus U$, is of class C^q on the whole of M and has support included in $\text{supp}(\chi) \subset U$. 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 $\chi_i : M \rightarrow [0, 1]$ of class C^k with*

locally finite support, such that $\text{supp}(\chi_i) \subset U_i$ 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*

$$(3.2) \quad \forall x \in U, f(x) = \sum_{j \in J} f_j(x)^2.$$

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 $(f_{x,i})_{1 \leq i \leq d}$ 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 $\text{supp}(\chi_j) \subset U_j$ 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 . Note that since $\text{supp}(\tilde{f}_{j,i}) \subset \text{supp}(\chi_j)$ and since $1 \leq i \leq d$, 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 $\text{supp}(\chi_j) \subset U_j$ 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. \square

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 :*

$$(3.3) \quad \forall i \in I, \exists (f_{i,j})_{j \in J} \in (C^{p-2}(U_i))^J, \forall x \in U_i, f(x) = \sum_{j \in J} f_{i,j}(x)^2,$$

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

$$(3.4) \quad \forall x \in M, f(x) = \sum_{j \in J \cup \{\star\}} g_j(x)^2.$$

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 sub-manifold of M by [Lemma 2.7](#) (directly adapted to the manifold case). Hence, we can assume that the U_i are disjoint (consider instead $U_i \cap V_i$, the property still holds).

Define $U_\star = \{f > 0\}$. Note that since the U_i 's cover \mathcal{Z} , $U_\star \cup \bigcup_{i \in I} U_i$ covers M since f is non-negative: take $\chi_\star, (\chi_i)_{i \in I}$ to be a gluing family adapted to that covering given by [Lemma 3.6](#). Note that 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, $U_x \cap V_x \subset M \setminus \text{supp}(g_j)$ 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 χ_\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 [\(3.3\)](#), it holds

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

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 non-negative function defined on S^2 .

THEOREM 3.9. *Let M be a manifold and $f : M \rightarrow \mathbb{R}$ be a non-negative map of class C^p . Assume f satisfies the global normal Hessian condition. 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*

$$(3.5) \quad \forall x \in M, f(x) = \sum_{i \in I} f_i(x)^2.$$

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 [Lemma 3.7](#) and [Lemma 3.8](#). Note that the global NHC [Lemma 2.7](#) shows that \mathcal{Z} is a sub-manifold of M . Let N_i denote the connected components of \mathcal{Z} . By [Lemma A.4](#), we can find disjoint 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 $J = \{1, \dots, n\}$, 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 $J = \{1, \dots, d\}$, the result is proven with $d + 1$ functions.

Remark 3.10. Note that the difference between the number of functions in the SHC case is better than the one obtained in [\[24\]](#). 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. Note that the existence of the sub-manifold N of dimension d_0 around x_0 implies the existence of a local parametrization around x_0 (see for instance [Theorem 1.21](#) of [\[13\]](#)): 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 $\phi : \widetilde{W}_0 \rightarrow U_{x_0}$ of class C^k such that ϕ is a homeomorphism from \widetilde{W}_0 onto $U_{x_0} \cap N$. Since restricting N to $N \cap U_{x_0}$ does not change the assumptions of the theorem, we will assume that $N = \text{im}(\phi)$ for a C^k immersion $\phi : (0, \mathbb{R}^{d_0}) \rightarrow (x_0, \mathbb{R}^d)$. 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 $df(x) = 0$ and $d^2f(x) \succeq 0$ (or equivalently $\nabla^2 f(x) \succeq 0$). Moreover, note that if $A \in \mathbb{S}_+(\mathbb{R}^d)$ is a symmetric positive semi-definite matrix, if a vector $k \in \mathbb{R}^d$ satisfies $k^\top A k = 0$, 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 semi-definite, 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 :

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

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

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

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

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

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 $g : (x', y') \in \mathbb{R}^{d-d_0} \times \mathbb{R}^{d_0} \mapsto f(P_1 x' + P_2 y' + x_0) = f(\mathcal{A}(x', y'))$, where $\mathcal{A}(x', y') = P(x', y') + x_0$ is an isometry³ ($\mathcal{A}^{-1}x = P^\top(x - x_0)$). We have $\nabla_{x'} g(0, 0) = P_1^\top \nabla f(x_0) = 0$ and $\nabla_{x', x'}^2 g(0, 0) = P_1^\top \nabla^2 f(x_0) P_1 \succ 0$. 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 $\varphi : W \rightarrow V$ 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 $z : V \times W \rightarrow \mathbb{R}^{d-d_0}$ of class C^{p-2} such that

$$(4.1) \quad \forall (x', y') \in V \times W, g(x', y') = g(\varphi(y'), y') + \frac{1}{2} z(x', y')^\top H' z(x', y'),$$

where H' is the positive definite matrix $P_1^\top \nabla^2 f(x_0) P_1$.

Step 3: making the sum of squares appear. Since $H' \in \mathbb{S}_+(\mathbb{R}^{d-d_0})$ and $H' \succ 0$, we can decompose it using the spectral theorem: $H' = \sum_{i=1}^{d-d_0} \lambda_i u_i u_i^\top$ where the $\lambda_i > 0$. Defining $g_i(x', y') = \sqrt{\lambda_i/2} u_i^\top z(x', y')$, (4.1) can be rewritten as

$$(4.2) \quad \forall (x', y') \in V \times W, g(x', y') = g(\varphi(y'), y') + \sum_{i=1}^{d-d_0} g_i^2(x', y').$$

Note that the g_i are of class C^{p-2} since z is of class C^{p-2} . 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 4: characterizing \mathcal{Z} in a neighborhood of x_0 . Denote with $G_\varphi = \{(\varphi(y), y) : y \in W\}$ the graph of φ , and which is a sub-manifold of class C^{p-1} of $\mathbb{R}^{d-d_0} \times \mathbb{R}^{d_0}$ (see theorem 1.21, point (iv) of [13]). Since \mathcal{A} is an isometry, the set $\mathcal{A}(G_\varphi)$ is also a sub-manifold 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 $\mathcal{A}^{-1}(\phi(\widetilde{W})) \subset \mathcal{A}^{-1}(\mathcal{Z}) \cap (V \times W) \subset G_\varphi$.

³An isometry is simply a map which preserves distances, and can be defined as an orthogonal transformation plus an affine shift.

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 $\{0\} \times W$. If π_2 denotes the canonical projection $\pi_2 : \mathbb{R}^{d-d_0} \times \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_0}$, we see that $\pi_2 \circ \psi$ maps G_φ injectively onto $W \subset \mathbb{R}^{d_0}$.

Take $\Phi = \pi_2 \circ \psi \circ \mathcal{A}^{-1} \circ \phi : \widetilde{W} \rightarrow \mathbb{R}^{d_0}$, which is well defined by definition of \widetilde{W} , and C^1 by composition. Note that 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, 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 . Note that 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

$$(4.3) \quad \phi(\widetilde{W}') \subset \mathcal{Z} \cap U \subset \mathcal{A}(G_\varphi) \cap U \subset \phi(\widetilde{W}') \implies \phi(\widetilde{W}') = \mathcal{Z} \cap U = \mathcal{A}(G_\varphi) \cap U.$$

Step 5: conclusion. (4.3) 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 sub-manifold $\mathcal{A}(G_\varphi)$ 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 [13], 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 sub-manifold of class C^k . Thus, $U \cap \mathcal{Z}$ is a sub-manifold of \mathbb{R}^d of class $C^{\max(k, p-1)}$.

On the other, since $\mathcal{A}^{-1}(U) \subset V \times W$, (4.2) becomes

$$(4.4) \quad \forall u \in U, g(\mathcal{A}^{-1}(u)) = g(\varphi(y'), y') + \sum_{i=1}^{d-d_0} g_i^2(\mathcal{A}^{-1}(u)), \quad (x', y') = \mathcal{A}^{-1}(u).$$

Let $u \in U$ and write $(x', y') = \mathcal{A}^{-1}(u)$. First, note that $\mathcal{A}(\varphi(y'), y') \in \mathcal{A}(G_\varphi)$. Moreover, since $\mathcal{A}^{-1}u \in \psi^{-1}(\mathbb{R}^{d-d_0} \times W')$ by 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 (4.4), recalling that $g = f \circ \mathcal{A}$, and defining $f_i : u \in U \mapsto g_i(\mathcal{A}^{-1}u)$, we have

$$(4.5) \quad \forall u \in U, f(u) = \sum_{i=1}^{d-d_0} f_i^2(u).$$

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. \square

5. Discussion and possible extensions. In this work, we have provided second order sufficient conditions in order for a non-negative 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 [23, 24, 29], which rely on the smoothness of such decompositions. As these conditions are sufficient and not necessary, one main problem is understanding this gap. This seems a highly difficult, and while very interesting, we present three other more reachable subjects for future work.

The first is to have an explicit bound for the number of squares needed in the sum of squares decomposition in the compact case. We believe that using finer tools from differentiable geometry, we should be able to obtain a bound depending on meaningful quantities, and upper bounded by a constant n_d depending only on the dimension d of the manifold on which the function is defined.

The second is, as in the polynomial case, to handle functions f which are non-negative on a constrained set defined by inequalities $f_i \geq 0$. More precisely, we would like to show second order sufficient conditions to write $f = g + \sum_i g_i f_i$ where the g, g_i are sum of squares of regular functions when f and the f_i are regular. This would open up the field of constrained optimization for methods developed for functions, such as kernel sum of squares [24]. In the polynomial case, such conditions are given by so-called *Positivstellensätzen* ([22, 25, 27]), but usually assume the polynomial is positive. Second order conditions have been developed more specifically in [15] to deal with non-negative polynomials with zeros.

The third is to handle functions with conic outputs which are more general than the non-negativity one. For example, in order to represent functions which have values in a cone defined by linear inequalities [16] or in the PSD cone [19].

REFERENCES

- [1] N. ARONSZAJN, *Theory of reproducing kernels*, Transactions of the American Mathematical Society, 68 (1950), pp. 337–404.
- [2] E. ARTIN, *Über die Zerlegung definiter Funktionen in Quadrate*, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg, 5 (1927), pp. 100–115.
- [3] E. BERTHIER, J. CARPENTIER, A. RUDI, AND F. BACH, *Infinite-Dimensional Sums-of-Squares for Optimal Control*, working paper or preprint, Oct. 2021, <https://hal.archives-ouvertes.fr/hal-03377120>.
- [4] 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/. talk:3.
- [5] J.-M. BONY, *Sommes de carrés de fonctions dérivables*, Bulletin de la Société Mathématique de France, 133 (2005), pp. 619–639, http://bonyjm.perso.math.cnrs.fr/S2KR_bullSMF.pdf.
- [6] J.-M. BONY, F. BROGLIA, F. COLOMBINI, AND L. PERNAZZA, *Nonnegative functions as squares or sums of squares*, Journal of Functional Analysis, 232 (2006), pp. 137–147.
- [7] C. FEFFERMAN AND D. H. PHONG, *On positivity of pseudo-differential operators*, Proceedings of the National Academy of Sciences, 75 (1978), pp. 4673–4674, <https://doi.org/10.1073/pnas.75.10.4673>, <https://www.pnas.org/content/75/10/4673>, <https://arxiv.org/abs/https://www.pnas.org/content/75/10/4673.full.pdf>.

- [8] D. HENRION, M. KORDA, AND J. B. LASSERRE, *The Moment-SOS Hierarchy*, WORLD SCIENTIFIC (EUROPE), 2020, <https://doi.org/10.1142/q0252>, <https://www.worldscientific.com/doi/abs/10.1142/q0252>, <https://arxiv.org/abs/https://www.worldscientific.com/doi/pdf/10.1142/q0252>.
- [9] D. HILBERT, *Ueber die Darstellung definitiver Formen als Summe von Formenquadraten*, *Mathematische Annalen*, 32 (1888), pp. 342–350, <https://doi.org/10.1007/BF01443605>, <https://doi.org/10.1007/BF01443605>.
- [10] L. HÖRMANDER, *Hypoelliptic second order differential equations*, *Acta Mathematica*, 119 (1967), pp. 147–171.
- [11] L. HÖRMANDER, *The Analysis of Linear Partial Differential Operators III*, *Classics in Mathematics*, Springer, Berlin, 2007. Pseudo-differential operators, Reprint of the 1994 edition.
- [12] K. JÄNICH, *Topologie*, Springer-Verlag, Berlin-New York, 1980.
- [13] J. LAFONTAINE, *An introduction to differential manifolds*, Springer, 2015.
- [14] J.-B. LASSERRE, *Moments, Positive Polynomials and their Applications*, vol. 1, World Scientific, 2010.
- [15] M. MARSHALL, *Representations of non-negative polynomials having finitely many zeros*, *Annales de la Faculté des sciences de Toulouse : Mathématiques*, Ser. 6, 15 (2006), pp. 599–609, <https://doi.org/10.5802/afst.1131>, https://afst.centre-mersenne.org/item/AFST_2006_6_15_3_599_0/.
- [16] U. MARTEAU-FEREY, F. BACH, AND A. RUDI, *Non-parametric Models for Non-negative Functions*, *Advances in Neural Information Processing Systems*, 33 (2020).
- [17] U. MARTEAU-FEREY, F. BACH, AND A. RUDI, *Sampling from Arbitrary Functions via PSD Models*, 2021, <https://arxiv.org/abs/2110.10527>.
- [18] J. W. MILNOR, *Morse theory*, Based on lecture notes by M. Spivak and R. Wells. *Annals of Mathematics Studies*, No. 51. Princeton University Press, Princeton, N.J., 1963, <http://www.maths.ed.ac.uk/~aar/papers/milnmors.pdf>.
- [19] B. MUZELLEC, F. BACH, AND A. RUDI, *Learning psd-valued functions using kernel sums-of-squares*, 2022, <https://arxiv.org/abs/2111.11306>.
- [20] P. A. PARRILO, *Semidefinite programming relaxations for semialgebraic problems*, *Mathematical Programming*, 96 (2003), pp. 293–320, <https://doi.org/10.1007/s10107-003-0387-5>, <https://doi.org/10.1007/s10107-003-0387-5>.
- [21] F. PAULIN, *Géométrie différentielle élémentaire*, 2006, https://www.imo.universite-paris-saclay.fr/~paulin/notescours/cours_geodiff.pdf.
- [22] M. PUTINAR, *Positive Polynomials on Compact Semi-algebraic Sets*, *Indiana Univ. Math. J.*, 42 (1993), pp. 969–984.
- [23] A. RUDI AND C. CILIBERTO, *PSD Representations for Effective Probability Models*, 2021, <https://arxiv.org/abs/2106.16116>.
- [24] A. RUDI, U. MARTEAU-FEREY, AND F. BACH, *Finding Global Minima via Kernel Approximations*, 2020, <https://arxiv.org/abs/2012.11978>.
- [25] K. SCHMÜGDEN, *The k -moment problem for compact semi-algebraic sets*, *Mathematische Annalen*, (1991), pp. 203–206.
- [26] M. SPIVAK, *A Comprehensive Introduction to Differential Geometry*, vol. One, Publish or Perish, Inc., Houston, Texas, third ed., 1999.
- [27] G. STENGLE, *A Nullstellensatz and a Positivstellensatz in Semialgebraic Geometry.*, *Mathematische Annalen*, 207 (1974), pp. 87–98, <http://eudml.org/doc/162533>.
- [28] D. TATARU, *On the Jefferman–phong inequality and related problems*, *Communications in Partial Differential Equations*, 27 (2002), pp. 2101–2138, <https://doi.org/10.1081/PDE-120016155>, <https://doi.org/10.1081/PDE-120016155>.
- [29] A. VACHER, B. MUZELLEC, A. RUDI, F. BACH, AND F.-X. VIALARD, *A Dimension-free Computational Upper-bound for Smooth Optimal Transport Estimation*, in *Proceedings of Thirty Fourth Conference on Learning Theory*, M. Belkin and S. Kpotufe, eds., vol. 134 of *Proceedings of Machine Learning Research*, PMLR, 15–19 Aug 2021, pp. 4143–4173, <https://proceedings.mlr.press/v134/vacher21a.html>.

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 sub-manifold 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 [\[12\]](#). Main references for manifold can be found in [\[13, 21, 26\]](#). 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 transition 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 generate 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 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 [21] or can be found in Theorem 2.13 of [26].

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

Note that in [26], a manifold is defined to be a metric space locally like \mathbb{R}^d . In proposition 2.2 of [21], 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 [26] is however a bit more general; in fact, it allows a manifold M to be a union of a possible non-countable connected component (as theorem 2 of [26] 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 [21], 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, [21]). *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 $\chi_i : M \rightarrow [0, 1]$ of class C^k such that $\text{supp}(\chi_i) \subset U_i$ 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 $V = M \setminus \text{supp}(\chi)$, 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 $\text{supp}(\chi) \subset U$, the property holds. Moreover, since $\chi g = 0$ on V , $\text{supp}(\chi g) \subset \text{supp}(\chi) \subset U$. \square

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 $\tilde{\chi}_i : M \rightarrow [0, 1]$ of class C^k such that *i)* for all $i \in I$, $\text{supp}(\tilde{\chi}_i) \subset U_i$, *ii)* the support of $(\tilde{\chi}_i)$ is locally finite and *iii)* $\sum_i \tilde{\chi}_i = 1$.

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

A.2. Connected components. Connectedness is a key topological notion for manifolds, and allows to decompose 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 in classes with respect to that relation, which are called "connected components". Connected components are both open and closed⁴. On a connected component of a manifold, the dimension d of the charts $\phi : U \rightarrow \mathbb{R}^d$ is the same, and is called the dimension of that connected component (for more details, see any of the references on manifolds). Note that as a manifold M is assumed to be second-countable, it has at most a countable number of connected components. Recall that a sub-manifold is defined in the main text as follows (such a definition can be found in section 2.4.2 of [21]).

DEFINITION A.3. *Let M be a manifold of class $C^{k'}$, $k \leq k'$. N is a sub-manifold of M of class C^k if for any $x \in N$, and any chart $\phi : U \rightarrow \mathbb{R}^d$ defined around x and of class C^k , $\phi(U \cap N)$ is a sub-manifold (in the sense of \mathbb{R}^d , see [subsection 2.1](#)) of \mathbb{R}^d around $\phi(x)$. It is equivalent to ask the existence of one such chart per point x .*

Let N be a sub-manifold 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 iff $V = U \cap N$ 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 $\phi : U \rightarrow \mathbb{R}^d$ 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 $\mathbb{R}^{d'} \times \{0\} \approx \mathbb{R}^{d'}$). Note that 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 sub-manifold $N \subset M$ 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 sub-manifold 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 $\phi : U \rightarrow \mathbb{R}^d$ where U is an open neighborhood of x . But since $\phi(U \cap N)$ is a sub-manifold of \mathbb{R}^d of class C^k around $\phi(x)$, by [Theorem 2.5](#) of [21], there exists a C^k diffeomorphism $\psi : (\phi(x), V) \rightarrow (0, W)$ where V such that $\psi(\phi(U \cap N) \cap V) = W \cap (\mathbb{R}^{d'} \times \{0\})$ for some d' . Taking $\tilde{\phi} = \psi \circ \phi$ on $\tilde{U} = \phi^{-1}(V) \cap U$, we have a chart of class C^k around x such that $\tilde{\phi} : \tilde{U} \rightarrow W \subset \mathbb{R}^d$ such that $\tilde{\phi}(\tilde{U} \cap N) = W \cap (\mathbb{R}^{d'} \times \{0\})$. Now let r be a radius such that the closed ball $\overline{B}(0, r) \subset W$. Set $U_x = \tilde{\phi}^{-1}(B(0, r))$, which is an open neighborhood of x included in \tilde{U} . Note that $\overline{U_x} \subset \tilde{\phi}^{-1}(\overline{B}(0, r)) \subset \tilde{U}$ since $\tilde{\phi}$ is continuous. Since $\overline{U_x} \cap N = \tilde{\phi}^{-1}(\overline{B}(0, r) \cap (\mathbb{R}^{d'} \times \{0\}))$ 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 $i \in I, \alpha \in A_i$.

⁴For more details on connected components, see [12]

Step 3. For all $i \in I$, define $F_i = \bigcup_{j \in I \setminus \{i\}, \beta \in A_j} \bar{V}_{j,\beta}$ and for all $\alpha \in A_i$, consider 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 $\bar{V}_{j,\beta}$. Hence, $V_x \setminus F_i$ is still open. Hence $W_{i,\alpha}$ is open. The second condition comes from the fact that $\bar{V}_{i,\alpha} \cap N \subset N_i$, 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). \square

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 $\phi : U \subset \mathbb{R}^d \rightarrow V \subset \mathbb{R}^{d'}$ 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 [13]). *Let $f : (x_0, \mathbb{R}^d) \rightarrow \mathbb{R}^d$ 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 $f : U \rightarrow f(U)$ is a C^k diffeomorphism.*

THEOREM B.2 (Theorem 1.18 of [13]). *Let $f : (x_0, \mathbb{R}^{d_1}) \rightarrow (y_0, \mathbb{R}^{d_2})$ 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 $g : V \rightarrow U$ 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 [11], which is a generalization of the so-called Morse Lemma (see lemma 2.2 of [18]), 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 $\nabla_{xx}^2 f(x, y) \in \mathbb{R}^{d_1 \times d_1}$ 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 [11]). *Let $d_1, d_2 \in \mathbb{N}, p \in \mathbb{N}$ 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 C^p function defined on a neighborhood U_0 of $(0, 0)$. Assume that $\nabla_x f(0, 0) = 0$ and that $H := \nabla_{xx}^2 f(0, 0)$ is non-singular.*

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

$$(B.1) \quad \forall (x, y) \in V \times W, f(x, y) = f(\varphi(y), y) + \frac{1}{2} z(x, y)^\top H z(x, y).$$

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

LEMMA B.4. *Under the assumptions of Lemma B.3, there exists two open convex neighborhoods of zero $V_0 \subset \mathbb{R}^{d_1}$, $W_0 \subset \mathbb{R}^{d_2}$ and $\varphi : W_0 \rightarrow V_0$ 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 $\begin{pmatrix} H & \star \\ 0 & I_{d_2} \end{pmatrix}$. Since H is non-singular, this matrix is non-singular. 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 $\tilde{V}_0 \subset \mathbb{R}^{d_1}$, $\tilde{W}_0 \subset \mathbb{R}^{d_2}$ be open convex neighborhoods of 0 such that $\tilde{V}_0 \times \tilde{W}_0 \subset U_1 \cap \psi(U_1)$. Define $\varphi : w \in \tilde{W}_0 \mapsto \pi_1(\psi^{-1}(0, w)) \in \mathbb{R}^{d_1}$. 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 $(x, y) \in V_0 \times W_0 \subset U_1 \cap \psi(U_1)$, $\nabla_1 f(x, y) = 0$ iff $\psi(x, y) = (0, y) \in \psi(U_1)$, iff $(x, y) = \psi^{-1}(0, y) = (\varphi(y), y)$, iff $x = \varphi(y)$. \square

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 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

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

Defining $B : V_0 \times W_0 \rightarrow S(\mathbb{R}^{d_1})$, such that $B(x, y) := 2 \int_0^1 (1-t) \nabla_{xx}^2 f(x_t, y) dt$, 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 $G : R \in \mathbb{R}^{d_1 \times d_1} \mapsto R^\top H R \in S(\mathbb{R}^{d_1})$ which is C^∞ and whose differential in $\mathbf{I}_{\mathbb{R}^{d_1}}$ is surjective (see [11]). Theorem B.2 shows there exists an neighborhood \mathcal{O} of H in $S(\mathbb{R}^{d_1})$ and a C^∞ function $F : \mathcal{O} \rightarrow \mathbb{R}^{d_1 \times d_1}$ 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 $z(x, y) = (F \circ B)(x, y)(x - \varphi(y))$. z satisfies (B.1). \square