APPROXIMATING FUNCTIONS ON STRATIFIED SETS

D. DRUSVYATSKIY AND M. LARSSON

ABSTRACT. We investigate smooth approximations of functions, with pre-scribed gradient behavior on a distinguished stratified subset of the domain. As an application, we outline how our results yield important consequences for a recently introduced class of stochastic processes, called the matrix-valued Bessel processes.

1. INTRODUCTION

Non-smoothness arises naturally in many problems of mathematical analysis. A conceptually simple way to alleviate the inherent difficulties involved is by smoothing. A classical result in this direction shows that any continuous function f on $\mathbf{R}^n$ can be uniformly approximated by a C^∞-smooth function g. See for example [14, Theorem 10.16]. In light of this, it is natural to ask the following question. Can we, in addition, guarantee that such an approximating function g satisfies "useful" properties on a distinguished non-smooth subset Q of $\mathbf{R}^n$? In the current work, we consider sets Q that are *stratified* into finitely many smooth manifolds $\{M_i\}$. The "useful" property we would like to ensure is that the gradient of the approximating function g at any point in Q is tangent to the manifold containing that point, that is, the inclusion

$$\nabla g(x) \in T_x M_i \text{ holds for any index } i \text{ and any point } x \in M_i.$$

This requirement can be thought of as a kind of Neumann boundary condition. In the language of stratification theory, we would like to approximate f by a smooth function g so that ∇g is a *stratified vector field*; see for example [19]. To the best of our knowledge, such a question has never been explicitly considered, and therefore we strive to make our development self-contained.

We provide an intuitive and transparent argument showing the existence of a C^1-smooth approximating function satisfying the tangency condition, provided that the partitioning manifolds yield a *Whitney (a)-regular C^2-stratification*. In particular, our techniques are applicable for all semi-algebraic sets—those sets that can be

written as a union of finitely many sets, each defined by finitely many polynomial inequalities. For more details on semi-algebraic geometry, see for example [3, 5]. Guaranteeing a higher order of smoothness for the approximating function, even when the partitioning manifolds are of class C^∞, seems fairly difficult, with the curvature of the manifolds entering the picture. Nevertheless, we identify a simple and easily checkable, though stringent, condition on the stratification —*normal flatness* (Definition 2.7)—that bypasses such technical difficulties and allows us to guarantee that whenever the partitioning manifolds are C^∞-smooth, so is the approximating function.

At first sight, the *normal flatness* condition is deeply tied to polyhedrality. However, we prove that this condition satisfies the so-called *Transfer Principle* investigated for example in [6–9, 15, 16, 18, 24]. Consequently this condition holds for a number of important subsets of matrix spaces, and our strongest results become applicable. This allows us to apply our techniques to the study of a class of stochastic processes called the *matrix-valued Bessel processes*, introduced in [13]. We give an informal outline of how our results constitute a key component needed to obtain a good description of the law of the process, and how they enable powerful uniqueness results to become available. Indeed, this was the original motivation for the current work. The main approximability results of the current paper are stated in terms of the uniform metric. In Section 5, we extend these results to weighted Sobolev norms, which is perhaps more natural given the boundary-value feel of the problem at hand.

The outline of the manuscript is as follows. In Section 2, we record basic notation that we will use throughout and state the main results of the paper. Section 3 contains the proofs of the main results. In Section 4, we discuss the *Transfer Principle* and how it relates to stratifications. In Section 5 we prove that, under reasonable conditions, smooth functions satisfying the tangency condition are dense in appropriate weighted Sobolev spaces. Finally in Section 6, we outline an application of our results to matrix-valued Bessel processes.

2. Basic notation and summary of main results

Throughout, the symbol $|\cdot|$ will denote the standard Euclidean norm on $\mathbf{R}^n$ and the absolute value of a real number, while $\|\cdot\|$ will denote the induced operator norm on the space of linear operators on $\mathbf{R}^n$. A function $f\colon Q \to \mathbf{R}$, defined on a set $Q \subset \mathbf{R}^n$, is called κ-*Lipschitz continuous* (for some real $\kappa \geq 0$) if the inequality

$$|f(x) - f(y)| \leq \kappa|x - y| \quad \text{holds for all } x, y \in Q.$$

The infimum of $\kappa \geq 0$ satisfying the inequality above is the *Lipschitz modulus* of f, and we denote it by $\operatorname{lip} f$. Thus we have

$$\operatorname{lip} f = \sup_{x,y\in Q} \frac{|f(x) - f(y)|}{|x - y|}.$$

Whenever $\operatorname{lip} f$ is finite, we will say that f is Lipschitz continuous. For notational convenience, 1-Lipschitz continuous functions will be called *non-expansive*. A function $f\colon Q \to \mathbf{R}$ is said to be *locally Lipschitz continuous* if around each point $x \in Q$, there exists a neighborhood U so that the restriction of f to $Q \cap U$ is Lipschitz continuous.

Given any set $Q \subset \mathbf{R}^n$ and a mapping $f \colon Q \to \widetilde{Q}$, where $\widetilde{Q} \subset \mathbf{R}^m$, we say that f is C^p-*smooth* if for each point $\bar{x} \in Q$, there is a neighborhood U of $\bar{x}$ and a C^p-smooth mapping $\widehat{f} \colon \mathbf{R}^n \to \mathbf{R}^m$ that agrees with f on $Q \cap U$. Throughout the manuscript, it will always be understood that p lies in $\{1, 2, \ldots, \infty\}$.

The following definition is standard.

Definition 2.1 (Smooth manifold)**.** Consider a set $M \subset \mathbf{R}^n$. We say that M is a C^p *manifold* (or "embedded submanifold") *of dimension* r if for each point $\bar{x} \in M$, there is an open neighborhood U around $\bar{x}$ such that $M \cap U = F^{-1}(0)$ for some C^p-smooth map $F \colon U \to \mathbf{R}^{n-r}$, with the derivative $\mathrm{D}F(\bar{x})$ having full rank.

The tangent space of a manifold M at a point $x \in M$ will be denoted by $T_x M$, while the normal space will be denoted by $N_x M$. We will consider $T_x M$ and $N_x M$ as embedded subspaces of $\mathbf{R}^n$.

In the current work, we will be interested in subsets of $\mathbf{R}^n$ that can be decomposed into finitely many smooth manifolds satisfying certain compatibility conditions. Standard references on stratification theory are [10, 21].

Definition 2.2 (Stratifications)**.** A C^p-*stratification* $\mathscr{A}$ of a set $Q \subset \mathbf{R}^n$ is a partition of Q into finitely many non-empty C^p manifolds (not necessarily connected), called *strata*, satisfying the following compatibility condition.

Frontier condition: For any two strata L and M, the implication

$$L \cap \operatorname{cl} M \neq \emptyset \quad \Longrightarrow \quad L \subset (\operatorname{cl} M) \setminus M \quad \text{holds.}$$

A C^p-stratification $\mathscr{A}$ is said to be *Whitney (a)-regular*, provided that the following condition holds.

Whitney condition (a): For any sequence of points x_k in a stratum M converging to a point $\bar{x}$ in a stratum L, if the corresponding normal vectors $v_k \in N_{x_k} M$ converge to a vector v, then the inclusion $v \in N_{\bar{x}} L$ holds.

Remark 2.3. We should emphasize that though we require stratifications to be comprised of finitely many manifolds (strata), each such manifold may have infinitely many connected components. This being said, it is worth noting that sometimes in the literature the term stratum (unlike our convention) refers to each connected component of the partitioning manifolds.

It is reassuring to know that many important sets fall in this class: every subanalytic set admits a Whitney (a)-regular C^p-stratification, for any finite p, as does any definable set in an arbitrary o-minimal structure. In particular, this is true for semi-algebraic sets. For a discussion, see for example [26].

Given a stratification $\mathscr{A}$, the frontier condition induces a (strict) partial order $\prec$ on $\mathscr{A}$, defined by

$$L \prec M \quad \Longleftrightarrow \quad L \subset (\operatorname{cl} M) \setminus M.$$

A stratum $M \in \mathscr{A}$ is *minimal* if there is no stratum $L \in \mathscr{A}$ with $L \prec M$. The *depth* of $\mathscr{A}$ is the maximal integer m such that there exist strata $M_1, \ldots, M_m$ with $M_1 \prec M_2 \prec \cdots \prec M_m$.

We are now ready to state the main results of this paper. Their proofs are given in Section 3.

Theorem 2.4 (Approximation on stratified sets). *Consider a closed set $Q \subset \mathbf{R}^n$ along with a C^2-stratification $\mathscr{A}$ of Q, and let $f : \mathbf{R}^n \to \mathbf{R}$ and $\varepsilon : \mathbf{R}^n \to (0, \infty)$ be continuous functions. Then there exists a function $g : \mathbf{R}^n \to \mathbf{R}$ satisfying the following properties:*

Uniform closeness: *The inequality $|f(x) - g(x)| < \varepsilon(x)$ holds for all $x \in \mathbf{R}^n$.*

Differentiability and Lipschitzness: *The function g is locally Lipschitz continuous and differentiable. Moreover, if f is Lipschitz continuous, then so is g and the estimate $\operatorname{lip} g \leq 11^{m+1} \operatorname{lip} f$ holds, where m is the depth of $\mathscr{A}$.*

Tangency condition: *For any stratum $M \in \mathscr{A}$ and any point $x \in M$, the inclusion*

$$\nabla g(x) \in T_x M \qquad holds.$$

Support: *Given a neighborhood V_1 of the support of f, we may choose g so that its support is contained in V_1.*

Agreement on full-dimensional strata: *If f is C^1-smooth, then given a neighborhood V_2 of the set*

$$\bigcup_{M \in \mathscr{A}, \ \dim M < n} M,$$

we may choose g so that it coincides with f outside of V_2.

If $\mathscr{A}$ is a Whitney (a)-regular C^2-stratification, then we may in addition to the above properties ensure that g is C^1-smooth.

Remark 2.5. Note that in the above theorem, f and g are defined on all of $\mathbf{R}^n$, as opposed to only on Q. In particular, any Lipschitz property should be understood with this in mind. Furthermore, the bound $\operatorname{lip} g \leq 11^{m+1} \operatorname{lip} f$ will be important for us because the multiplier 11^{m+1} *only* depends on the depth of $\mathscr{A}$.

It is also worth mentioning that the hypothesis on ε can be weakened to lower semicontinuity. Indeed, it is easy to show that any strictly positive, lower semicontinuous function on $\mathbf{R}^n$ can be bounded below by a strictly positive continuous function.

Remark 2.6. Theorem 2.4 immediately implies that any Whitney (a)-regular C^2-stratification $\mathscr{A}$ of a set $Q \subset \mathbf{R}^n$, where $\mathscr{A}$ contains at least one non-zero dimensional stratum, admits a vector field that is not identically zero, continuous, conservative, and stratified. This nicely complements the development in [23, Chapter 1].

Observe that even under the Whitney condition (a), the approximating function g, guaranteed to exist by Theorem 2.4, is only C^1-smooth. To guarantee a higher order of smoothness (under minimal conditions), it seems that one needs to impose stronger requirements, both on the stratification and on the curvature of the strata. Below we identify a simple, though stringent, condition which bypasses such technical difficulties. We begin with some notation.

Given a set $Q \subset \mathbf{R}^n$ and a point $x \in \mathbf{R}^n$, the *distance* of x to Q is

$$d(x, Q) := \inf_{y \in Q} |x - y|,$$

and the *projection* of x onto Q is

$$P_Q(x) := \{y \in Q : |x - y| = d(x, Q)\}.$$

Note that $P_Q(x)$ may be empty, a singleton, or it may contain multiple points. When $P_Q(x)$ is a singleton, we will abuse notation slightly and write $P_Q(x)$ for the point it contains. A crucial fact for us will be that any C^{p+1} manifold M (for $p = 1, \ldots, \infty$) admits a neighborhood on which the projection mapping P_M is single-valued and C^p-smooth; see for example [17, Lemma 4] or [22]. Indeed, this is the reason why throughout the manuscript we will be concerned with stratifications by manifolds that are at least of class C^2. See Subsection 3.1 for more details.

Definition 2.7 (Normally flat stratification). A C^{p+1}-*stratification* of a subset Q of $\mathbf{R}^n$ is said to be *normally flat* if for any two strata $L, M \in \mathscr{A}$ with $L \prec M$ there are neighborhoods V of L and U of M so that the equality

$$P_L(x) = P_L \circ P_M(x) \quad \text{holds for all} \quad x \in V \cap U.$$

The normal flatness condition is of course quite strong. In particular, normally flat stratifications automatically satisfy Whitney's condition (a); see Proposition 4.1. Nonetheless, this condition does hold in a number of important situations. Section 4 contains a detailed analysis, in particular showing that polyhedral sets and *spectral lifts* of *symmetric polyhedra* admit normally flat C^∞-stratifications (Proposition 4.4 and Theorem 4.8).

The next result is a strengthened version of Theorem 2.4, under the normal flatness condition.

Theorem 2.8 (Approximation on normally flat stratifications). *Consider a closed set $Q \subset \mathbf{R}^n$ along with a normally flat C^{p+1}-stratification $\mathscr{A}$ of Q, and let $f : \mathbf{R}^n \to \mathbf{R}$ and $\varepsilon : \mathbf{R}^n \to (0, \infty)$ be continuous functions. Then there exists a function $g : \mathbf{R}^n \to \mathbf{R}$ satisfying the following properties:*

Uniform closeness: *The inequality $|f(x) - g(x)| < \varepsilon(x)$ holds for all $x \in \mathbf{R}^n$.*
Smoothness and Lipschitzness: *The function g is C^p-smooth. Moreover, if f is Lipschitz continuous, then so is g and the estimate $\operatorname{lip} g \leq 11^{m+1} \operatorname{lip} f$ holds, where m is the depth of $\mathscr{A}$.*
Enhanced tangency condition: *Each stratum $M \in \mathscr{A}$ has a neighborhood U so that the equality*

$$g(x) = g \circ P_M(x) \quad \text{holds for all } x \in U.$$

Support: *Given a neighborhood V_1 of the support of f, we may choose g so that its support is contained in V_1.*
Agreement on full-dimensional strata: *If f is C^p-smooth, then given a neighborhood V_2 of the set*

$$\bigcup_{M \in \mathscr{A}, \ \dim M < n} M,$$

we may choose g so that it coincides with f outside of V_2.

In particular, the enhanced tangency condition of Theorem 2.8 directly implies that for any stratum $M \in \mathscr{A}$, any point $x \in M$, and any normal direction $v \in N_x M$, the function $\mathbf{R} \ni t \mapsto g(x + tv)$ is constant in a neighborhood of the origin. Consequently, we have

$$\mathrm{D}^{(k)} g(x)[v, \ldots, v] = 0, \quad \text{for any } k = 1, \ldots, p,$$

where $\mathrm{D}^{(k)} g(x)[v, \ldots, v]$ is the k'th order directional derivative of g at x in direction v. This is a significant strengthening of the tangency condition in Theorem 2.4.

3. THE ITERATIVE CONSTRUCTION AND PROOFS OF MAIN RESULTS

Our method for proving results such as Theorems 2.4 and 2.8 relies on an inductive procedure, where the original function f (possibly after some initial smoothing) is first modified on the minimal strata, then on the strata whose frontiers consist of minimal strata, and so on. At each step, the resulting function approximates f and has the desired properties on all strata preceding (in the sense of the partial order $\prec$) the stratum currently under consideration. This allows the induction to continue. When no more strata remain, the desired properties hold on the entire stratified set. In this section we give a detailed description of this iterative construction, in particular the induction step. The reason for bringing out some of the details, as opposed to hiding them inside proofs, is that the same general approach can sometimes be used in specific situations to obtain further properties of the approximating function. An example of this will be discussed in Section 6.

The induction step has two crucial ingredients, namely (1) finding a suitable *tubular neighborhood* of the stratum where the current function is to be modified, and (2) an *interpolation function* constructed using this tubular neighborhood. These two objects are described in Subsections 3.1 and 3.2. Then, in Subsection 3.3, the induction step is described in detail, laying the groundwork for the proofs of the main results, given in Subsection 3.4.

3.1. **The tubular neighborhood.** In this subsection, we follow the notation of [14, Section 10]. We should stress that this subsection does not really contain any new results. Its only purpose is to record a number of observations needed in the latter parts of the manuscript.

For a C^{p+1} manifold $M \subset \mathbf{R}^n$, the *normal bundle* of M, denoted by NM, is the set

$$NM := \{(x, v) \in \mathbf{R}^n \times \mathbf{R}^n : x \in M, v \in N_x M\}.$$

It is well known that the normal bundle NM is itself a C^p manifold. Consider the mapping $E : NM \to \mathbf{R}^n$ defined by

$$E(y, v) = y + v.$$

A *tubular neighborhood* U of M is by definition a C^p diffeomorphic image under E of an open subset $V \subset NM$ having the form

$$(3.1) \qquad\qquad V = \{(y, v) \in NM : |v| < \delta(y)\},$$

for some continuous function $\delta \colon M \to (0, \infty)$. It is standard that tubular neighborhoods of M always exist; see for example [14, Theorem 10.19]. Furthermore, there always exists a tubular neighborhood on which the projection mapping P_M is single-valued and C^p-smooth; see [17, Lemma 4]. Then for any point $\bar{x}$ lying in M, the derivative of the projection $\mathrm{D}P_M(\bar{x})$ is simply the orthogonal projection onto the tangent space $T_{\bar{x}}M$. It is worth noting that whenever the manifold M is only C^1 smooth, the projector P_M may easily fail to be single-valued on any neighborhood of M; this is the main reason why throughout the paper we consider C^{p+1} manifolds for $p = 1, \ldots, \infty$.

The following observation shows that we may squeeze the tubular neighborhood inside any given open set containing the manifold.

Lemma 3.1 (Existence of a tubular neighborhood)**.** *Consider a C^{p+1} manifold $M \subset \mathbf{R}^n$, as well as an arbitrary neighborhood V of M. Then there exists a tubular*

neighborhood

$$U_\delta = \{y + v \in \mathbf{R}^n : y \in M, \ v \in N_y M, \ |v| < \delta(y)\}$$

satisfying $U_\delta \subset V$. Furthermore, for any real $\kappa > 0$, we may ensure that δ is κ-Lipschitz continuous.

Proof. By intersecting V with some tubular neighborhood of M, we may assume that V is the diffeomorphic image of some neighborhood of the zero-section $M \times \{0\}$ in NM. Now, for any point $y \in M$, define

$$\delta(y) := \sup\{0 \le \epsilon \le 1 : V_\epsilon(y) \subset V\},$$

where

$$V_\epsilon(y) := \{z + v \in \mathbf{R}^n : z \in M, \ v \in N_z M, \ |y - z| < \epsilon, \ |v| < \epsilon\}.$$

Clearly δ is strictly positive and the inclusion $U_\delta \subset V$ holds. We now show that δ is non-expansive. To see this, first note that whenever the inequality $|x - y| \ge \delta(x)$ holds, we have $\delta(x) - \delta(y) \le |x - y|$ trivially. On the other hand, if we have $|x - y| \le \delta(x)$, an application of the triangle inequality shows that the inclusion $V_\epsilon(y) \subset V_{\delta(x)}(x)$ is valid for $\epsilon = \delta(x) - |x - y|$. We deduce $\delta(y) \ge \epsilon$, and hence the inequality $\delta(x) - \delta(y) \le |x - y|$ is also valid in this case. Interchanging the roles of x and y gives the non-expansive property. Finally, replacing δ with $\kappa\delta$, if need be, ensures that δ is κ-Lipschitz continuous. $\qquad\square$

In fact, in Lemma 3.1 we may ensure that δ is C^∞-smooth. To see this, we need the following result, which has classical roots.

Lemma 3.2 (Approximation of Lipschitz functions)**.** *Consider any C^{p+1} manifold $M \subset \mathbf{R}^n$ and a function $f \colon M \to \mathbf{R}$ that is Lipschitz continuous. Then for any continuous function $\varepsilon \colon M \to (0, \infty)$ and a real $r > 0$, there exists a Lipschitz continuous, C^∞-smooth function $\widehat{f} \colon M \to \mathbf{R}$ satisfying*

$$|f(x) - \widehat{f}(x)| < \varepsilon(x) \text{ for all } x \in M,$$

with $\operatorname{lip} \widehat{f} \le \operatorname{lip} f + r$.

Proof. The proof proceeds by extending the functions f and ε to an open neighborhood of M, using standard approximation techniques on this neighborhood, and then restricting the approximating function back to M. To this end, for any real $\epsilon > 0$ there exists an open neighborhood U of M so that the projection P_M is $(1 + \epsilon)$-Lipschitz continuous on U. Consider now the functions $\widetilde{f} \colon U \to \mathbf{R}$ and $\widetilde{\varepsilon} \colon U \to (0, \infty)$, defined by $\widetilde{f}(x) = f(P_M(x))$ and $\widetilde{\varepsilon}(x) = \varepsilon(P_M(x))$. Observe that $\widetilde{f}$ agrees with f on M and $\widetilde{\varepsilon}$ agrees with ε on M, and furthermore the inequality $\operatorname{lip} \widetilde{f} \le (1 + \epsilon)\operatorname{lip} f$ holds. It is standard then that for any $\widetilde{r}$ (see for example [2, Theorem 1]) there exists a C^∞ function $\widehat{f}$ on U satisfying

$$|\widetilde{f}(x) - \widehat{f}(x)| < \widetilde{\varepsilon}(x) \text{ for all } x \in U,$$

with $\operatorname{lip} \widehat{f} \le \operatorname{lip} \widetilde{f} + \widetilde{r}$. In particular, we deduce $\operatorname{lip} \widehat{f} \le \operatorname{lip} f + \epsilon \operatorname{lip} f + \widetilde{r}$. Since ϵ and $\widetilde{r}$ can be chosen to be arbitrarily small, restricting $\widehat{f}$ to M yields the result. $\qquad\square$

Coming back to Lemma 3.1, an application of Lemma 3.2 with $f = \frac{1}{2}\delta$ and $\varepsilon = \frac{1}{4}\delta$ shows that there is no loss of generality in assuming that δ is C^∞-smooth and non-expansive. For ease of reference, we now record a version of Lemma 3.1,

where we impose a number of open conditions on the tubular neighborhood, which will be important for our later development.

Corollary 3.3. *Consider a C^{p+1} manifold $M \subset \mathbf{R}^n$, continuous functions ε : $\mathbf{R}^n \to (0,\infty)$ and $f\colon \mathbf{R}^n \to \mathbf{R}$, a neighborhood V of M, and the closed set $\Gamma :=$ $(\mathrm{cl}\, M) \setminus M$. Then there exists a tubular neighborhood*

$$U_\delta = \{y + v \in \mathbf{R}^n : y \in M,\ v \in N_y M,\ |v| < \delta(y)\},$$

with $U_\delta \subset V \cap \Gamma^c$, and satisfying

  (1) the width function $\delta\colon M \to (0,\infty)$ is C^∞-smooth and non-expansive,
  (2) the metric projection P_M is C^p-smooth on U_δ,

and for each $x \in U_\delta$ we have

  (3) $|f(x) - f \circ P_M(x)| < \varepsilon(x)$,
  (4) $|x - P_M(x)| < d(x, \Gamma)^2$,
  (5) $\|\mathrm{D}P_M(x) - \mathrm{D}P_M(y)\| < d(y, \Gamma)$, where $y = P_M(x)$,
  (6) $\|\mathrm{D}P_M(x)\| < 2$.

3.2. The interpolation function. Given a tubular neighborhood U_δ as in Corollary 3.3 (constructed based on a C^{p+1} manifold M, a neighborhood V, functions ε and f, and the closed set $\Gamma = (\mathrm{cl}\, M) \setminus M$), define a function $\phi\colon \mathbf{R}^n \to \mathbf{R}$ by setting

$$\phi(x) = \begin{cases} 0, & x \notin U_\delta, \\[2mm] \psi\left(\dfrac{|x - P_M(x)|}{\delta \circ P_M(x)}\right), & x \in U_\delta, \end{cases}$$

where $\psi : [0,\infty) \to [0,1]$ is a C^∞-smooth cut-off function that is 1 on $[0, 1/4]$, vanishes on $[3/4, \infty)$, and whose first derivative is bounded by $7/3$ in absolute value. See Figure 1 for an illustration. Consider also the "annulus" around M, defined by

$$U_0 = \left\{x \in U_\delta : \frac{1}{4}\delta \circ P_M(x) < |x - P_M(x)| < \frac{3}{4}\delta \circ P_M(x)\right\}.$$

We refer to the function ϕ as the *interpolation function* associated with U_δ, and its decisive properties are collected in the following lemma.
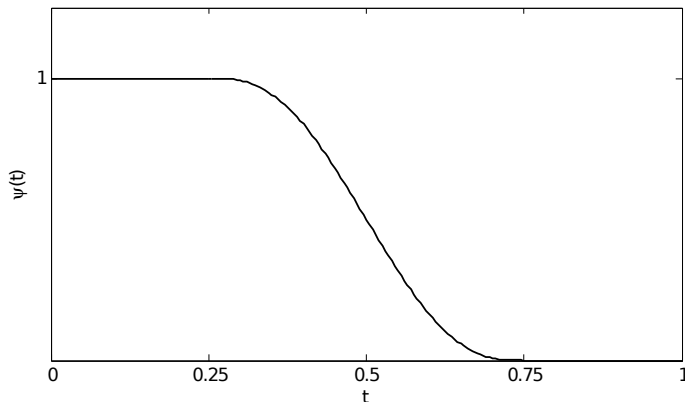

FIGURE 1. The function ψ.

Lemma 3.4 (Properties of the interpolation function). *The interpolation function ϕ satisfies the following properties:*

(1) $\phi = 1$ on $\{x \in U_\delta : |x - P_M(x)| \leq \frac{1}{4}\delta \circ P_M(x)\}$,

(2) $\phi = 0$ on $\{x \in U_\delta : |x - P_M(x)| \geq \frac{3}{4}\delta \circ P_M(x)\}$ and on U_δ^c,

(3) ϕ is C^p smooth on Γ^c, and for all $x \in U_0$ we have

$$\nabla\phi(x) = t\psi'(t)\left[\frac{x - P_M(x)}{|x - P_M(x)|^2} - \frac{\nabla(\delta \circ P_M)(x)}{\delta \circ P_M(x)}\right],$$

where $t = \frac{|x - P_M(x)|}{\delta \circ P_M(x)}$.

(4) For all $x \in U_\delta$ the inequality $|x - P_M(x)|\,|\nabla\phi(x)| \leq 8$ holds.

Proof. The first three properties follow directly from the definition, together with a calculation to compute the gradient. The only point that is somewhat delicate here is to verify the equality

$$(3.2) \qquad \nabla|\mathrm{Id} - P_M|(x) = \frac{x - P_M(x)}{|x - P_M(x)|},$$

for $x \in U_\delta \setminus M$. To this end, an application of the chain rule yields the equation

$$\langle\nabla|\mathrm{Id} - P_M|(x), u\rangle = \frac{\langle u - DP_M(x)[u], x - P_M(x)\rangle}{|x - P_M(x)|},$$

for any vector $u \in \mathbf{R}^n$. This together with the inclusions $DP_M(x)[u] \in T_{P_M(x)}M$ and $x - P_M(x) \in N_{P_M(x)}M$, establishes (3.2).

For the last property, first note that $\nabla\phi = 0$ holds outside U_0. Hence it is sufficient to consider only points $x \in U_0$. For such points, observe that the inequalities $|t\psi'(t)| \leq \frac{7}{4}$ and $|x - P_M(x)| < \delta \circ P_M(x)$ hold. Since δ is non-expansive, and Corollary 3.3 yields the bound $\|DP_M\| \leq 2$, we readily obtain the inequality $|\nabla(\delta \circ P_M)| \leq 2$. The claim now follows from the expression for $\nabla\phi(x)$. $\qquad\square$

3.3. The induction step.
We begin with the following straightforward lemma.

Lemma 3.5. *Consider a C^2-stratification $\mathscr{A}$ of a set $Q \subset \mathbf{R}^n$. Then there exists a family of sets $\{V_M\}_{M \in \mathscr{A}}$, where for each stratum $M \in \mathscr{A}$ the set V_M is a neighborhood of M, and the following properties hold:*

- *If we have $L \prec M$, then $L \cap V_M = \emptyset$.*
- *If L and M are incomparable, then $V_L \cap V_M = \emptyset$.*

Proof. Consider any pair of strata $L, M \in \mathscr{A}$. If we have $L \prec M$, then define $V_{L,M} := (\mathrm{cl}\,L)^c$. Similarly, if we have $M \prec L$, then define $V_{M,L} := (\mathrm{cl}\,M)^c$. If L and M are incomparable, then by the definition of a stratification, there exist disjoint neighborhoods $V_{L,M}$ of M and $V_{M,L}$ of L. Observe that for any strata $L, M \in \mathscr{A}$, the set $V_{L,M}$, if defined, is a neighborhood of M, and $V_{M,L}$, if defined, is a neighborhood of L.

Define now for each stratum M the neighborhood V_M to be the intersection of the (finitely many) sets $V_{L,M}$ where either $L \prec M$ or L and M are incomparable. If M is the only stratum, then simply define $V_M := \mathbf{R}^n$. It is easy to see that this yields a family of sets $\{V_M\}_{M \in \mathscr{A}}$ with the desired properties. $\qquad\square$

For the rest of this subsection, we fix the following objects:

- $\mathscr{A}$, a C^2-stratification of some closed set $Q \subset \mathbf{R}^n$. We emphasize that $\mathscr{A}$ is not assumed to satisfy the Whitney condition (a).

- $M \in \mathscr{A}$, a stratum.
- $\Gamma = (\mathrm{cl}\, M) \setminus M$, a closed set, possibly empty. Since Q is closed, the set Γ is the union of all strata $L \in \mathscr{A}$ satisfying $L \prec M$.
- $\varepsilon : \mathbf{R}^n \to (0, \infty)$, a continuous function.
- $f : \mathbf{R}^n \to \mathbf{R}$, a locally Lipschitz continuous and differentiable function. We emphasize that the gradient ∇f is not assumed to be continuous.
- U_δ, a tubular neighborhood of M. It is obtained by applying Corollary 3.3 with the functions ε, f, and Γ as above, and V taken to be the neighborhood V_M of M given by Lemma 3.5.
- ϕ, the interpolation function corresponding to U_δ.

Since Γ is a union of pairwise disjoint submanifolds in $\mathscr{A}$, we can unambiguously let $T_x\Gamma$ and $N_x\Gamma$, for $x \in \Gamma$, denote the tangent and normal spaces of the submanifold where x lies. Note that we have so far made no assumptions on the geometry of Q beyond the existence of a C^2-stratification.

Proposition 3.6 (Differentiable approximation). *Assume that the inclusion $\nabla f(x) \in T_x\Gamma$ holds for every $x \in \Gamma$, and define*

$$g = \phi f \circ P_M + (1 - \phi)f.$$

Then g satisfies the following properties:
- *g is locally Lipschitz continuous and differentiable.*
- *$\nabla g(x) = \nabla f(x)$ for all $x \in \Gamma$, and $\nabla g(x) \in T_x M$ for all $x \in M$.*
- *$|f(x) - g(x)| \leq \varepsilon(x)$ for all $x \in \mathbf{R}^n$.*
- *If f is Lipschitz continuous, then so is g, and the inequality $\mathrm{lip}\, g \leq 11\, \mathrm{lip}\, f$ holds.*

If $\mathscr{A}$ satisfies the Whitney condition (a) and f is C^1-smooth, then g is C^1-smooth as well.

Proof. It is convenient to split the proof into a number of steps.

Step 1 (Differentiability). We show that g is differentiable at each $\bar{x} \in \Gamma$ with $\nabla g(\bar{x}) = \nabla f(\bar{x})$. This is sufficient to establish the claim, since clearly g is differentiable on Γ^c. To this end, let v_k be any sequence of unit vectors in $\mathbf{R}^n$ converging to some vector $\bar{v}$, and let t_k be a sequence of positive reals converging to 0. We set $x_k := \bar{x} + t_k v_k$, and assuming without loss of generality $x_k \in U_\delta$ for each k, we define $y_k := P_M(x_k)$. Observe that $g(\bar{x}) = f(\bar{x})$, and consequently we have

$$(3.3) \qquad \frac{g(x_k) - g(\bar{x})}{t_k} = \frac{f(x_k) - f(\bar{x})}{t_k} + \phi(x_k)\frac{f(y_k) - f(x_k)}{t_k}.$$

The first term on the right-hand side converges to $\langle \nabla f(\bar{x}), \bar{v} \rangle$. For the second term, the Mean Value Theorem applied to the function $t \mapsto f(x_k + t(y_k - x_k))$ shows that there exists a vector z_k, lying on the line segment joining x_k and y_k, satisfying $f(y_k) - f(x_k) = \langle \nabla f(z_k), y_k - x_k \rangle$. Hence we obtain

$$\frac{|f(y_k) - f(x_k)|}{t_k} \leq |\nabla f(z_k)|\frac{|y_k - x_k|}{t_k} < |\nabla f(z_k)|t_k,$$

where we used the inequality $|y_k - x_k| < d(x_k, \Gamma)^2 \leq |x_k - \bar{x}|^2 = t_k^2$, which is a direct consequence of Corollary 3.3. Since ∇f is locally bounded due to the local Lipschitzness of f, we see that the second term on the right-hand side of (3.3) converges to zero, thereby establishing the claim.

Step 2 (Closeness, tangency condition, Lipschitz modulus). The closeness of g to f follows from the simple calculation

$$|f(x) - g(x)| = \phi(x)|f(x) - f \circ P_M(x)| \leq \varepsilon(x),$$

where the last inequality uses Corollary 3.3. Next, we verify the tangency condition. For any $x \in M$ and $v \in N_x M$ we have $g(x + tv) = f(x)$ for all t close to zero, since the equality $\phi = 1$ holds in a whole neighborhood of x. It follows that the directional derivative of g in the direction v vanishes at x, and hence we deduce $\nabla g(x) \in T_x M$, as desired. Finally, we prove that g is locally Lipschitz continuous and derive a bound on its Lipschitz modulus. For $x \in \Gamma^c$, we have

$$\nabla g(x) = (1 - \phi(x))\nabla f(x) + \nabla \phi(x)(f \circ P_M(x) - f(x)) + \phi(x)\nabla(f \circ P_M)(x),$$

and hence

$$|\nabla g(x)| \leq |\nabla f(x)| + |\nabla \phi(x)|\,|f \circ P_M(x) - f(x)| + \phi(x)|\nabla(f \circ P_M)(x)|.$$

For any $x \in U_\delta$ and $w \in \mathbf{R}^n$ we have

$$\langle \nabla(f \circ P_M)(x), w \rangle = \langle \nabla f(P_M(x)), \mathrm{D}P_M(x)w \rangle,$$

so that, due to Corollary 3.3(6), we have $|\nabla(f \circ P_M)(x)| \leq 2|\nabla f(P_M(x))|$. Moreover, by the Mean Value Theorem there exists a point z on the line segment joining x and $P_M(x)$, satisfying $f(x) - f(P_M(x)) = \langle \nabla f(z), x - P_M(x) \rangle$. Consequently we deduce

$$|\nabla g(x)| \leq |\nabla f(x)| + 2\phi(x)|\nabla f(P_M(x))| + |\nabla \phi(x)|\,|x - P_M(x)|\,|\nabla f(z)|$$

$$\leq \begin{cases} |\nabla f(x)|, & x \notin U_\delta, \\ |\nabla f(x)| + 2|\nabla f(P_M(x))| + 8|\nabla f(z)|, & x \in U_\delta, \end{cases}$$

where the second inequality follows in part from Lemma 3.4(4). Given any compact set $K \subset \mathbf{R}^n$, the points $P_M(x)$ and z remain inside some other compact set K' as x varies over $K \cap U_\delta$. This yields the inequality

$$\sup_{x \in K} |\nabla g(x)| \leq 11 \sup_{x \in K'} |\nabla f(x)|.$$

It easily follows that g is locally Lipschitz continuous. If f is globally Lipschitz continuous, then so is g, and we have $\mathrm{lip}\, g \leq 11\,\mathrm{lip}\, f$, as claimed.

Assume now that $\mathscr{A}$ satisfies the Whitney condition (a) and f is C^1-smooth.

Step 3 (Continuity of the gradient). We now show that the gradient mapping ∇g is continuous. It is enough to check continuity at each point $\bar{x} \in \Gamma$. For $x \in \Gamma^c$ we have

$$(3.4)\quad \nabla g(x) - \nabla f(x) = \nabla \phi(x)(f \circ P_M(x) - f(x)) + \phi(x)(\nabla(f \circ P_M)(x) - \nabla f(x)).$$

Since $\nabla f(\bar{x}) = \nabla g(\bar{x})$ holds and ∇f is continuous, it suffices to show that the right-hand side converges to zero as x tends to $\bar{x}$ with $x \in U_\delta$. We do this term by term. To deal with the first term, we define $y := P_M(x)$ and apply the Mean Value Theorem to obtain a vector z on the line segment joining x and y, satisfying $f(y) - f(x) = \langle \nabla f(z), y - x \rangle$. Assuming without loss of generality that $x \neq y$, we obtain

$$|f(y) - f(x)| = |y - x|\langle \nabla f(z), v \rangle,$$

where $v := \frac{y-x}{|y-x|} \in N_y M$. Whenever v converges to some vector $\bar{v}$, we have $\bar{v} \in N_{\bar{x}}\Gamma$ by the Whitney condition (a). Consequently, along any convergent subsequence we

obtain $\langle \nabla f(z), v \rangle \to \langle \nabla f(\bar{x}), \bar{v} \rangle = 0$, since ∇f is continuous and $\nabla f(\bar{x})$ lies in $T_{\bar{x}}\Gamma$. Finally, recalling from Lemma 3.4 that the quantity $|\nabla \phi(x)| |y - x|$ is bounded, we deduce that the first term on the right-hand side of (3.4) tends to zero.

For the second term, it suffices to show $|\nabla (f \circ P_M)(x) - \nabla f(x)| \to 0$ as $x \to \bar{x}$ with $x \in U_\delta$. To this end, apply the equality

$$\langle \nabla (f \circ P_M)(x), w \rangle = \langle \nabla f(y), \mathrm{D}P_M(x)w \rangle,$$

valid for any $w \in \mathbf{R}^n$, to obtain

$$\langle \nabla (f \circ P_M)(x) - \nabla f(x), w \rangle = \langle \nabla f(x), \mathrm{D}P_M(y)w - w \rangle$$
$$+ \langle \nabla f(y) - \nabla f(x), \mathrm{D}P_M(y)w \rangle$$
$$+ \langle \nabla f(y), \mathrm{D}P_M(x)w - \mathrm{D}P_M(y)w \rangle.$$

Observe that the inclusion $\mathrm{D}P_M(y)w - w \in N_y M$ holds, and so any limit point of such vectors lies in $N_{\bar{x}}\Gamma$. Hence along every convergent subsequence the first term vanishes. The second and third terms vanish due to the continuity of ∇f, boundedness of $\|\mathrm{D}P_M(y)\|$, and the inequality $\|\mathrm{D}P_M(x) - \mathrm{D}P_M(y)\| < d(y, \Gamma)$ established in Corollary 3.3. This concludes the proof of Step 3. $\qquad \square$

If we assume that the stratification $\mathscr{A}$ is normally flat, stronger approximation results can be obtained. First we record the following observation, which is a direct consequence of the finiteness of the number of strata and the defining property of normally flat stratifications.

Lemma 3.7. *Consider a normally flat C^p-stratification $\mathscr{A}$ of a set $Q \subset \mathbf{R}^n$ and a stratum M. Then there exists a neighborhood $\widehat{U}$ of M such that for each stratum $L \in \mathscr{A}$ with $L \prec M$, there is a neighborhood $\widehat{V}$ of L with $P_L(x) = P_L \circ P_M(x)$ for all $x \in \widehat{U} \cap \widehat{V}$.*

Proposition 3.8 (Arbitrarily smooth approximation). *Suppose that the stratification $\mathscr{A}$ is normally flat and that U_δ is contained in $\widehat{U}$ from Lemma 3.7. Assume that each stratum $L \in \mathscr{A}$ with $L \prec M$ has a neighborhood V with*

$$f(x) = f \circ P_L(x) \quad \text{for all} \quad x \in V.$$

Define a function $g = \phi f \circ P_M + (1 - \phi)f$. Then each stratum $L \in \mathscr{A}$ with $L \subset \mathrm{cl}\, M$ has a neighborhood W so that the equality

$$g(x) = g \circ P_L(x) \quad \text{holds for all} \quad x \in W.$$

Proof. We start with the case $L = M$. To this end, let W be the subset of U_δ where the equality $\phi = 1$ holds. Then W is a neighborhood of M, and for any point $x \in W$, we have

$$g(x) = f \circ P_M(x) = f \circ P_M \circ P_M(x) = g \circ P_M(x),$$

as required. Consider now a stratum $L \in \mathscr{A}$ with $L \prec M$. By hypothesis there is a neighborhood V of L in which the equality $f = f \circ P_L$ holds. Now pick a new neighborhood W of L satisfying

- $W \subset V$,
- $P_M(x) \in V$ for all $x \in W \cap U_\delta$,
- $P_L \circ P_M(x) = P_L(x)$ for all $x \in W \cap U_\delta$.

To see that such a neighborhood can be found, first note that it is easy to find W satisfying the first two properties, thanks to the continuity of P_M on U_δ: simply intersect V with a sufficiently small open set containing L. Then replace W by the smaller set $W \cap \widehat{V}$, where $\widehat{V}$ is as in Lemma 3.7. The resulting set, which we again denote by W, satisfies all three properties.

We now verify the equality $g = g \circ P_L$ on W. First consider a point $x \in W \setminus U_\delta$. Recall that we have $\phi = 0$ outside U_δ and $f = f \circ P_L$ on W. Hence we obtain

$$g(x) = f(x) = f \circ P_L(x) = g \circ P_L(x).$$

Now consider instead a point $x \in W \cap U_\delta$. The inclusion $P_M(x) \in V$, the equality $f = f \circ P_L$ on V, and the normal flatness condition yield

$$f \circ P_M(x) = f \circ P_L \circ P_M(x) = f \circ P_L(x).$$

Consequently, we deduce

$$g(x) = \phi(x) f \circ P_M(x) + (1 - \phi(x)) f(x) = f \circ P_L(x).$$

Finally, note that the right-hand side equals $g \circ P_L(x)$, since we have $\phi(P_L(x)) = 0$. This establishes the claim. $\qquad\square$

We end this subsection by recording the following remark.

Remark 3.9. In the notation of Proposition 3.8, observe that since the equality $\phi = 0$ holds on $\Gamma = (\mathrm{cl}\, M) \setminus M$, we have $g = f$ there. Consequently, for any stratum L contained in Γ, there exists a neighborhood V of L, so that for all $x \in V$ we have

$$g(x) = g \circ P_L(x) = f \circ P_L(x) = f(x).$$

Hence if f is C^p-smooth and the restriction of ϕ to Γ^c is C^p-smooth, then g is C^p-smooth as well.

3.4. **Proofs of main results.** Given the previous developments, the proofs of our main results are entirely straightforward.

Proof of Theorem 2.4. We first establish some notation. Given a differentiable function h we say that h is *tangential to* $M \in \mathscr{A}$ if the inclusion $\nabla h(x) \in T_x M$ holds for all $x \in M$. We let $\mathscr{T}_h$ be the collection of $M \in \mathscr{A}$ such that h is tangential to M, and we define m to be the depth of $\mathscr{A}$.

The proof proceeds by induction on the partial order $\prec$. In preparation for this, we set $f_0 := f$ if f is already differentiable and locally Lipschitz continuous. Otherwise we let f_0 be a C^1-smooth function that differs from f by at most $\varepsilon/(m+1)$, and whose Lipschitz modulus (when f is Lipschitz continuous) is at most 11 times that of f.

For the induction step, suppose a differentiable locally Lipschitz continuous function h is given, pick $M \in \mathscr{A}$ with $M \notin \mathscr{T}_h$, and assume we have $\{L \in \mathscr{A} : L \prec M\} \subset \mathscr{T}_h$ (this is the induction assumption). Proposition 3.6 then gives a new differentiable locally Lipschitz continuous function h_1 such that $\mathscr{T}_{h_1} = \mathscr{T}_h \cup \{M\}$ and $|h - h_1| < \varepsilon/(m + 1)$ hold.

Observe that for each minimal stratum M the family $\{L \in \mathscr{A} : L \prec M\}$ is empty, so the induction assumption is vacuously true for f_0. It then follows by induction on the partial order $\prec$ that we may find a differentiable locally Lipschitz continuous function g such that $\mathscr{T}_g = \mathscr{A}$.

At each application of Proposition 3.6, the modified function differs from the previous one by at most $\varepsilon/(m+1)$. However, the modifications associated with incomparable strata do not overlap. This is because the tubular neighborhood U_δ corresponding to $M \in \mathscr{A}$ is contained in V_M from Lemma 3.5. Hence the total error in the end is only $\varepsilon m/(m+1)$. Together with a possible initial error of $\varepsilon/(m+1)$ if f is not differentiable and locally Lipschitz continuous, this results in the claimed bound $|f - g| < \varepsilon$. Similarly, we lose a factor of 11^{m+1} in the Lipschitz modulus. To ensure that the support of g is contained in V_1, we simply ensure that the tubular neighborhood U_δ is contained in V_1 at each application of Proposition 3.6. The same can be done with V_2 whenever the proposition is applied with a stratum M such that $\dim M < n$. For M with $\dim M = n$, the tubular neighborhood U_δ coincides with M, and we have $P_M = \mathrm{Id}$ on M. Hence in this case Proposition 3.6 yields a function that is identical to the previous one. We deduce that if f was already C^1-smooth, so that no initial smoothing took place (i.e., $f_0 = f$), then $f = g$ holds outside V_2. Finally, in case $\mathscr{A}$ satisfies the Whitney condition (a), the C^1-smoothness of f_0 is preserved after each application of Proposition 3.6, and this yields the final assertion of Theorem 2.4. $\square$

Proof of Theorem 2.8. The result follows exactly as in the proof of Theorem 2.4, except that in each step we additionally apply Proposition 3.8. This can be done, provided we shrink U_δ if necessary. Note that the smoothness carries over in each step by Remark 3.9. The function g obtained in the end has the property that every stratum $M \in \mathscr{A}$ has a neighborhood on which $g = g \circ P_M$ holds. $\square$

4. NORMALLY FLAT STRATIFICATIONS

In this section we discuss important examples of sets admitting normally flat stratifications. First, however, we observe that this notion is strictly stronger than the Whitney condition (a).

Proposition 4.1 (Normally flat stratifications are Whitney (a)-regular)**.** *If a C^p-stratification of a set $Q \subset \mathbf{R}^n$ is normally flat, then for any strata L and M with $L \prec M$ there exists a neighborhood V of L so that the inclusion*

$$N_x M \subset N_{P_L(x)} L \quad \text{holds for all } x \in M \cap V.$$

Consequently, normally flat C^p-stratifications satisfy the Whitney condition (a).

Proof. Since the stratification is normally flat, there exist neighborhoods V of L and U of M so that the equality

$$P_L(x) = P_L \circ P_M(x) \quad \text{holds for all} \quad x \in V \cap U.$$

Consider a point $x \in M \cap V$, and let $v \in N_x M$ be an arbitrary normal vector. Define $y = x + tv$ for some $t > 0$ that is sufficiently small to guarantee the inclusion $y \in U$. Then we have $x = P_M(y)$, and consequently $P_L(x) = P_L \circ P_M(y) = P_L(y)$. We deduce

$$tv = y - x = (y - P_L(y)) + (P_L(x) - x) \in N_{P_L(x)} L,$$

as we had to show. The claim that normally flat C^p-stratifications satisfy the Whitney condition (a) is now immediate. $\square$

The following simple example confirms that the normal flatness condition is indeed strictly stronger than the Whitney condition (a).

Example 4.2 (Whitney (a)-regular stratification that is not normally flat). Consider the set
$$Q = \{(x, y, z) \in \mathbf{R}^3 : z = xy, x \geq 0\},$$
together with the stratification $\mathscr{A} = \{M_1, M_2\}$, where the strata are defined by
$$M_1 = \{0\} \times \mathbf{R} \times \{0\} \quad \text{and} \quad M_2 = \{(x, y, z) \in Q : x > 0\}.$$
Clearly $\mathscr{A}$ is a Whitney (a)-regular stratification. On the other hand, this stratification is not normally flat. To see this, first note the equivalence
$$P_{M_1}(x, y, z) = (0, 0, 0) \quad \Longleftrightarrow \quad y = 0.$$
On the other hand, $P_{M_2}(x, 0, z)$ is a singleton and has a non-zero y-coordinate whenever we have $x > 0$ and $z \neq 0$.

To record our first examples of normally flat stratifications, we need the following standard result.

Lemma 4.3 (Ordered affine subspaces). *Consider affine subspaces L and M in $\mathbf{R}^n$, satisfying the inclusion $L \subset M$. Then the equality*
$$P_L(x) = P_L \circ P_M(x) \quad \text{holds for all} \quad x \in \mathbf{R}^n.$$

Proof. Consider an arbitrary point $x \in \mathbf{R}^n$. Observe that for any point $y \in L$, we have
$$x - P_M(x) \in N_{P_M(x)}M \quad \text{and} \quad y - P_M(x) \in T_{P_M(x)}M.$$
Consequently, any two such vectors are orthogonal and we obtain
$$|x - y|^2 = |x - P_M(x)|^2 + |P_M(x) - y|^2 \geq |x - P_M(x)|^2 + |P_M(x) - P_L(P_M(x))|^2$$
$$= |x - P_L(P_M(x))|^2.$$
Since $y \in L$ is arbitrary, we deduce $P_L(x) = P_L \circ P_M(x)$, as claimed. $\square$

We denote the *affine hull* of any convex set $Q \subset \mathbf{R}^n$ by $\operatorname{aff} Q$.

Proposition 4.4 (Polyhedral stratifications are normally flat). *Consider a stratification of a set $Q \subset \mathbf{R}^n$, where each stratum is an open polyhedron. Then the stratification is normally flat.*

Proof. Consider strata L and M, with $L \prec M$. Clearly there exists a neighborhood V of M and U of L satisfying
$$P_L = P_{\operatorname{aff} L} \text{ on } V \quad \text{and} \quad P_M = P_{\operatorname{aff} M} \text{ on } U.$$
Furthermore observe that the inclusion, $\operatorname{aff} L \subset \operatorname{aff}(\operatorname{cl} M) = \operatorname{aff} M$, holds. The result is now immediate from Lemma 4.3. $\square$

It turns out that normally flat stratifications satisfy the so-called *Transfer Principle*, which we will describe below. This realization will show that many common subsets of matrices admit normally flat stratifications. We first introduce some notation.

- $\mathbf{M}^{n \times m}$ is the Euclidean space of all $n \times m$ real matrices, endowed with the trace inner product $\langle X, Y \rangle := \operatorname{tr}(X^\top Y)$ and Frobenius norm $\|X\|_F := \langle X^\top X \rangle^{1/2}$. For notational convenience, throughout we will assume $n \leq m$.
- $\mathbf{S}^n$ is the subspace (when $m = n$) of all symmetric matrices, endowed with the trace inner product and the Frobenius norm.

- $\mathbf{R}_{\geq}^n$ is the set $\{(x_1, \ldots, x_n) \in \mathbf{R}^n : x_1 \geq x_2 \geq \cdots \geq x_n\}$, and $\mathbf{R}_{+,\geq}^n :=$ $\mathbf{R}_{\geq}^{\overline{n}} \cap \mathbf{R}_+^n$.
- $\lambda : \mathbf{S}^n \to \mathbf{R}^n$ is the map taking X to its vector $(\lambda_1(X), \ldots, \lambda_n(X))$ of (real) eigenvalues, in decreasing order.
- $\sigma : \mathbf{M}^{n \times m} \to \mathbf{R}^n$ is the map taking X to its vector $(\sigma_1(X), \ldots, \sigma_n(X))$ of singular values, in decreasing order. (Recall throughout that we are assuming $n \leq m$.)
- $\operatorname{Diag} x \in \mathbf{M}^{n \times m}$, for a vector $x \in \mathbf{R}^n$, is a matrix that has entries all zero, except for its principle diagonal, which contains the entries of x.

For notational convenience, sets $Q \subset \mathbf{R}^n$ that are invariant under any permutation of coordinates will be called *permutation-invariant*, while sets that are invariant under any coordinatewise change of sign and permutation of coordinates will be called *absolutely permutation-invariant*.

Let $O(n)$ be the group of $n \times n$ orthogonal matrices. We can define an action of $O(n)$ on the space of symmetric matrices $\mathbf{S}^n$ by declaring

$$U.X = UXU^\top \text{ for all } U \in O(n) \text{ and } X \text{ in } \mathbf{S}^n.$$

We say that a subset of $\mathbf{S}^n$ is *spectral* if it is invariant under the action of $O(n)$. Equivalently, a subset of $\mathbf{S}^n$ is spectral if and only if it can be represented as $\lambda^{-1}(Q)$, for some permutation-invariant set $Q \subset \mathbf{R}^n$. Due to this invariance of Q, the spectral set can be written simply as the union of orbits

$$\lambda^{-1}(Q) = \bigcup_{x \in Q} O(n).(\operatorname{Diag} x).$$

Similarly, we can consider the Cartesian product $O(n) \times O(m)$, which we denote by $O(n, m)$, and its action on the space $\mathbf{M}^{n \times m}$ defined by

$$(U, V).X = UXV^\top \text{ for all } (U, V) \in O(n, m) \text{ and } X \text{ in } \mathbf{M}^{n \times m}.$$

We say that a subset of $\mathbf{M}^{n \times m}$ is *spectral* if it is invariant under the action of $O(n, m)$. Equivalently, a subset of $\mathbf{M}^{n \times m}$ is spectral if and only if it has the form $\sigma^{-1}(Q)$, for some absolutely permutation-invariant set $Q \subset \mathbf{R}^n$. In this situation, the spectral set is simply the union of orbits,

$$Q = \bigcup_{x \in Q} O(n, m).(\operatorname{Diag} x).$$

The mappings σ and λ have nice geometric properties, but are very badly behaved as far as, for example, differentiability is concerned. However, such difficulties are alleviated by the invariance assumptions on Q. The *Transfer Principle* asserts that many geometric (or more generally variational analytic) properties of the sets Q are inherited by the spectral sets $\sigma^{-1}(Q)$ and $\lambda^{-1}(Q)$. The collection of properties known to satisfy this principle is impressive: convexity [16], prox-regularity [7], Clarke-regularity [16, 18], smoothness [6, 8, 9, 15, 16, 24] and partial smoothness [6]. In this section, we will add the existence of normally flat stratifications to this list.

The following crucial lemma asserts that projections interact well with the singular value map (or eigenvalue map in the symmetric case). This result may be found in [17, Theorem A.1, Theorem A.2], and in [7, Proposition 2.3] for the symmetric case. For completeness, we record a short proof, adapted from [7, Proposition 2.3].

Lemma 4.5 (Projection onto spectral sets). *Let $Q \subset \mathbf{R}^n$ be an absolutely permutation-invariant set and consider a matrix $X \in \mathbf{M}^{n \times m}$ with singular value decomposition $X = U \operatorname{Diag} \sigma(X) \, V^\top$. Then the inclusion*

$$(4.1) \qquad P_{\sigma^{-1}(Q)}(X) \supset \{U \operatorname{Diag} z \, V^\top : z \in P_Q(\sigma(X))\} \quad \text{holds.}$$

Similarly if $Q \subset \mathbf{R}^n$ is only permutation-invariant and $X \in \mathbf{S}^n$ has an eigenvalue decomposition $X = U \operatorname{Diag} \lambda(X) \, U^\top$, then the analogous inclusion

$$P_{\lambda^{-1}(Q)}(X) \supset \{U \operatorname{Diag} z \, U^\top : z \in P_Q(\lambda(X))\} \quad \text{holds.}$$

Proof. The proof relies on the fact that the singular-value map σ is non-expansive; see Theorem 7.4.51 in [11]. Then for arbitrary $Y \in Q$ and $z \in P_Q(\sigma(X))$, we obtain

$$\|X - Y\|_F \geq |\sigma(X) - \sigma(Y)| \geq |\sigma(X) - z|,$$

where the second inequality follows from the equivalence $\sigma(Y) \in Q \iff Y \in \sigma^{-1}(Q)$. On the other hand, the matrix $Y = U \operatorname{Diag} z \, V^\top$ achieves equality and lies in $\sigma^{-1}(Q)$, and this yields the claim. The symmetric case follows in the same way since the eigenvalue map λ is also non-expansive. $\qquad \square$

Remark 4.6. Lemma 4.5 shows that in the particular case when $P_{\sigma^{-1}(Q)}(X)$ is a singleton, so is $P_Q(\sigma(X))$, and we have equality in (4.1). We furthermore have the appealing formula

$$\sigma \circ P_{\sigma^{-1}(Q)} = P_Q \circ \sigma \qquad (\text{or } \lambda \circ P_{\lambda^{-1}(Q)} = P_Q \circ \lambda),$$

which captures the gist of the result.

The proof of the following simple lemma may be found in [17, Lemmas A.1, A.2], though the statement of the result is somewhat different.

Lemma 4.7 (Permutations of projected points). *Consider an absolutely permutation invariant set $Q \subset \mathbf{R}^n$ and a point $x \in \mathbf{R}^n_{+,\geq}$. Then for any point y lying in $P_Q(x)$, there exists a signed permutation matrix A on $\mathbf{R}^n$ so that Ay lies in $\mathbf{R}^n_{+,\geq} \cap P_Q(x)$. The analogous statement holds in the symmetric case.*

Theorem 4.8 (Lifts of stratifications). *Consider a partition $\mathscr{A}$ of a set $Q \subset \mathbf{R}^n$ into finitely many C^p manifolds that are absolutely permutation-invariant. Then if $\mathscr{A}$ is a C^p-stratification of Q, the family*

$$\sigma^{-1}(\mathscr{A}) := \left\{ \sigma^{-1}(M) : M \in \mathscr{A} \right\}$$

is a C^p-stratification of $\sigma^{-1}(Q)$. The analogous statement holds for both Whitney (a)-regular and normally flat C^p-stratifications. The case of symmetric matrices is analogous as well.

Proof. First, we note that by [6, Theorem 5.1], each set $\sigma^{-1}(M)$ for $M \in \mathscr{A}$ is a C^p manifold. Throughout the proof, we let L and M be arbitrary strata in $\mathscr{A}$. We first claim that the implication

$$(4.2) \qquad L \subset \operatorname{cl} M \implies \sigma^{-1}(L) \subset \operatorname{cl} \sigma^{-1}(M)$$

holds. Indeed, consider a matrix $X \in \sigma^{-1}(L)$ and a singular value decomposition $X = U \operatorname{Diag} \sigma(X) \, V^T$. Then there exists a sequence $x_i \to \sigma(X)$ in M. Observe that since M is absolutely permutation-invariant, the matrices $X_i := U \operatorname{Diag} x_i \, V^T$ lie in $\sigma^{-1}(M)$ and converge to X. This establishes the validity of the implication.

Assume now that $\mathscr{A}$ is a C^p-stratification of Q. Clearly $\sigma^{-1}(\mathscr{A})$ is a partition of $\sigma^{-1}(Q)$. Now suppose $\sigma^{-1}(L) \cap \operatorname{cl} \sigma^{-1}(M) \neq \emptyset$. Then by continuity of σ, we have

$L \cap \mathrm{cl}\, M \neq \emptyset$. We deduce that the inclusion $L \subset \mathrm{cl}\, M$ holds. Then using (4.2), we obtain $\sigma^{-1}(L) \subset \mathrm{cl}\, \sigma^{-1}(M)$, thus verifying that $\sigma^{-1}(\mathscr{A})$ is a C^p-stratification of $\sigma^{-1}(Q)$.

Assume now that in addition, $\mathscr{A}$ is a Whitney (a)-regular C^p-stratification of Q. Consider a point $X \in M$. Then by [18, Theorem 7.1], we have the formula

$$N_X(\sigma^{-1}(M)) = \{U(\mathrm{Diag}\, N_{\sigma(X)}M)V^\top : U\, \mathrm{Diag}\, \sigma(X)V^\top = X\}.$$

Verification of the fact that $\sigma^{-1}(\mathscr{A})$ is a Whitney (a)-regular C^p-stratification of $\sigma^{-1}(Q)$ is now trivial.

Assume now that $\mathscr{A}$ is a normally flat C^p-stratification. We will show that $\sigma^{-1}(\mathscr{A})$ is a normally flat C^p-stratification of $\sigma^{-1}(Q)$. To this end, suppose that the inclusion $\sigma^{-1}(L) \subset \mathrm{cl}\, \sigma^{-1}(M)$ holds. Then there exist neighborhoods V of L and U of M, with $P_L = P_L \circ P_M$ on $V \cap U$. Define the neighborhoods

$$\widetilde{U} = \sigma^{-1}(U) \qquad \text{and} \qquad \widetilde{V} = \sigma^{-1}(V)$$

of M and L, respectively. Shrinking $\widetilde{U}$ and $\widetilde{V}$, we may suppose that the maps P_L, P_M, and the composition $P_L \circ P_M$ are all well defined and single-valued on $\widetilde{V} \cap \widetilde{U}$. Fix a matrix $X \in \widetilde{U} \cap \widetilde{V}$, with singular value decomposition $X = U\, \mathrm{Diag}\, \sigma(X)\, V^\top$. Applying Lemma 4.5 and Lemma 4.7 successively, we deduce

$$
\begin{aligned}
P_{\sigma^{-1}(L)} \circ P_{\sigma^{-1}(M)}(X) &= P_{\sigma^{-1}(L)}(U\, \mathrm{Diag}\, P_M(\sigma(X))\, V^\top) \\
&= U\, \mathrm{Diag}[P_L \circ P_M(\sigma(X))]\, V^\top \\
&= U\, \mathrm{Diag}\, P_L(\sigma(X))\, V^\top \\
&= P_{\sigma^{-1}(L)}(X),
\end{aligned}
$$

as claimed. The proof of the proposition in the case of symmetric matrices is similar. We leave the details to the reader. $\qquad\square$

5. APPLICATION: DENSITY IN SOBOLEV SPACES

In this section, we will always consider integration on $\mathbf{R}^n$ with respect to the Lebesgue measure. To make the notation consistent with existing literature, in contrast to previous sections, we will let p denote the order of Lebesgue spaces and we will let k denote the degree of smoothness of a function. Let $w\colon \mathbf{R}^n \to \mathbf{R}$ be a locally integrable function satisfying $w(x) > 0$ for almost all $x \in \mathbf{R}^n$; we will call w a *weight*. Consider an open subset Ω of $\mathbf{R}^n$. Then for $1 \leq p < \infty$, define $L_w^p(\Omega)$ as the set of measurable functions f on Ω satisfying

$$\|f\|_{L_w^p(\Omega)} := \left(\int_\Omega |f|^p w\, dx \right)^{\frac{1}{p}} < \infty.$$

Pairwise identifying functions in $L_w^p(\Omega)$ that are pairwise equal almost everywhere with respect to Lebesgue measure, the set $L_w^p(\Omega)$ becomes a Banach space with norm $\|\cdot\|_{L_w^p(\Omega)}$.

Consider a locally integrable function f on Ω. Then the *weak i'th partial derivative* of f (for $i = 1, \ldots, n$) is any other locally integrable function v_i on Ω satisfying

$$\int_\Omega f \frac{\partial \phi}{\partial x_i}\, dx = -\int_\Omega v_i \phi\, dx, \quad \text{for any } \phi \in C_c^\infty(\Omega).$$

Weak derivatives are unique up to measure zero. We will denote the weak i'th partial derivative of f by $D^i f$. The vector of weak partial derivatives $(D^1 f, \ldots, D^n f)$ will be denoted by Df.

We assume that $w^{-1/(p-1)}$ is locally integrable on Ω, which ensures that every $f \in L^p_w(\Omega)$ has weak partial derivatives; see [12, Theorem 1.5]. The *weighted Sobolev space* $W^{1,p}_w(\Omega)$ is defined by

$$W^{1,p}_w(\Omega) = \{f \in L^p_w(\Omega) : D^i f \in L^p_w(\Omega) \text{ for } i = 1, \ldots, n\},$$

and it becomes a Banach space when equipped with the norm

$$\|f\|_{W^{1,p}_w(\Omega)} := \left(\|f\|^p_{L^p_w(\Omega)} + \sum_{i=1}^{n} \|D^i f\|^p_{L^p_w(\Omega)} \right)^{\frac{1}{p}};$$

see [12, Theorem 1.11].

A basic question in the theory of Sobolev spaces is under which conditions the elements of $W^{1,p}_w(\Omega)$ can be approximated by more regular functions. In this direction Meyers and Serrin [20] (see also [1, Theorem 3.17]) famously showed that the set of C^1-smooth functions contained in $W^{1,p}_1(\Omega)$ is actually dense in $W^{1,p}_1(\Omega)$. More generally, if w satisfies *Muckenhoupt's A_p condition* (see [25, Definition 1.2.2]), it is true that C^∞-smooth functions contained in $W^{1,p}_w(\Omega)$ are dense in $W^{1,p}_w(\Omega)$ [25, Corollary 2.1.6]. Ensuring that the approximating functions can be extended to all of $\mathbf{R}^n$ in a smooth way requires extra conditions on the geometry of Ω. In particular, the following property holds in a variety of circumstances:

(5.1)   The set of restrictions to Ω of functions in $C^\infty_c(\mathbf{R}^n)$ is dense in $W^{1,p}_w(\Omega)$.

For example, in the unweighted case where $w = 1$, equation (5.1) holds whenever Ω satisfies the so-called *segment condition*, essentially stating that Ω is located on one side of its boundary. For more details, see [1, Definition 3.21 and Theorem 3.22]. In the weighted case and assuming that w satisfies the A_p condition, it is sufficient that Ω be an (ε, δ) *domain*; see [4].

Now let Q denote the closure of Ω, and assume it admits a Whitney (a)-regular C^2-stratification $\mathscr{A}$. Assume also that there is a subclass $\mathscr{A}_{\mathrm{bd}} \subset \mathscr{A}$ of strata whose union is equal to $\mathrm{bd}\, Q$. In particular, this is the case whenever $\mathscr{A}$ is a semi-algebraic stratification. Then, given that (5.1) holds, we can apply our previous results to identify a different class of functions that is also dense in $W^{1,p}(\Omega, w)$:

$$C^k_{\mathrm{Neu}}(\Omega) := \left\{ \begin{array}{l} \text{all restrictions to } \Omega \text{ of functions } g \in C^k_c(\mathbf{R}^n) \text{ with} \\ \nabla g(x) \in T_x M \text{ for all } M \in \mathscr{A}_{\mathrm{bd}} \text{ and all } x \in M \end{array} \right\}.$$

The elements of $C^k_{\mathrm{Neu}}(\Omega)$ satisfy a Neumann boundary condition, namely that its gradients are tangent to the boundary of Ω. In what follows, the p-norm on $\mathbf{R}^n$ will be denoted by $\|\cdot\|_p$.

Theorem 5.1. *Fix a real number $1 \leq p < \infty$ and let Q be the closure of an open set $\Omega \subset \mathbf{R}^n$. Assume that Q admits a Whitney (a)-regular C^2-stratification $\mathscr{A}$ such that $\mathrm{bd}\, Q$ consists of a collection $\mathscr{A}_{\mathrm{bd}} \subset \mathscr{A}$ of strata. If (5.1) holds, then $C^1_{\mathrm{Neu}}(\Omega)$ is dense in $W^{1,p}_w(\Omega)$.*

If in addition $\mathscr{A}$ is normally flat and a C^{k+1}-stratification, then $C^k_{\mathrm{Neu}}(\Omega)$ is dense in $W^{1,p}_w(\Omega)$.

Proof. We start with the first assertion. Since (5.1) holds, it suffices to approximate functions in $C_c^\infty(\mathbf{R}^n)$. Let f be such a function and let V_1 be a bounded open set containing its support. Additionally, let V_2 be an open set, to be specified later, containing bd Q. Consider now an arbitrary real number $\varepsilon > 0$. An application of Theorem 2.4 yields a C^1 function $g \in C_{\mathrm{Neu}}^1(\Omega)$ such that its support is contained in V_1, it coincides with f outside of V_2, the inequality $|f(x) - g(x)| < \varepsilon$ holds for all $x \in \mathbf{R}^n$, and we have the estimate lip $g \le c_1(\mathrm{lip}\, f)$ for a constant c_1 that only depends on the stratification $\mathscr{A}$. This yields

$$\int_\Omega |g(x) - f(x)|^p w(x)dx \le \varepsilon \int_{V_2} w(x)dx$$

and, since $\|\cdot\|_p \le c_2\|\cdot\|$ for some constant c_2 that only depends on p and n,

$$\begin{aligned}
\int_\Omega \|Dg(x) - Df(x)\|_p^p w(x)dx &\le 2^{p-1} \int_{V_2 \cap \Omega} \left(\|Dg(x)\|_p^p + \|Df(x)\|_p^p\right) w(x)dx \\
&\le 2^{p-1} c_2^p \int_{V_2 \cap \Omega} \left(\|Dg(x)\|^p + \|Df(x)\|^p\right) w(x)dx \\
&\le 2^{p-1} c_2^p (c_1^p + 1)(\mathrm{lip}\, f)^p \int_{V_2} w(x)dx.
\end{aligned}$$

Since w is locally integrable, $w(x)dx$ is a Radon measure on $\mathbf{R}^n$, hence outer regular. Since also bd Q is the union of finitely many C^2 manifolds of dimension at most $n - 1$, and hence a nullset, V_2 can be chosen so that we have bd $Q \subset V_2$ and

$$\int_{V_2} w(x)dx \le \min\{1,\ \varepsilon 2^{1-p} c_2^{-p}(c_1^p + 1)^{-1}(\mathrm{lip}\, f)^{-p}\}.$$

With this choice of V_2, the inequality $\|g - f\|_{W_w^{1,p}(\Omega)}^p \le 2\varepsilon$ holds. This finishes the proof of the first assertion. To prove the second assertion, we simply apply Theorem 2.8 instead of Theorem 2.4 in the above proof. This ensures that g is C^k-smooth. □

Remark 5.2. Unfortunately our results are not strong enough to obtain analogous results for higher order Sobolev spaces $W_w^{k,p}(\Omega)$, where $k > 1$. The reason is that the size of higher order derivatives of g cannot be controlled in terms of those of f using our current techniques.

6. APPLICATION: MATRICES WITH NON-NEGATIVE DETERMINANT

In this section, we outline how one may use Theorem 2.8 in a concrete situation. The full details may be found in [13]. Before proceeding though, we give a rather simple and very intuitive motivation for why our main results may have significant applicability. Consider a C^2-smooth function $f: \mathbf{R}^n \to \mathbf{R}$ and a C^1-smooth function $h: \mathbf{R}^n \to \mathbf{R}$. As is well known, for many subsets Q of $\mathbf{R}^n$, the integration-by-parts formula

$$(6.1) \qquad \int_Q h\Delta f\, dx = \int_{\mathrm{bd}\, Q} h\langle \nabla f, \hat{n} \rangle\, dS - \int_Q \langle \nabla h, \nabla f \rangle\, dx$$

holds, where $\hat{n}$ is a well-defined outward normal vector and dS denotes an appropriate surface area measure. The boundary term is often the troublesome term in the expression above. Theorem 2.8, on the other hand, implies that under suitable conditions on Q we may approximate f by another C^2-smooth function g, for which the corresponding boundary term vanishes.

Corollary 6.1 (Simplified integration-by-parts)**.** *Consider a subset $Q \subset \mathbf{R}^n$, a C^2-smooth function $f \colon \mathbf{R}^n \to \mathbf{R}$, and a C^1-smooth function $h \colon \mathbf{R}^n \to \mathbf{R}$. Suppose in addition that the equation* (6.1) *holds and that Q admits a normally flat C^{p+2}-stratification. Then for any $\epsilon > 0$, there exists a C^{p+1}-smooth function $g \colon \mathbf{R}^n \to \mathbf{R}$ with $|f - g| < \epsilon$, that satisfies the simplified integration-by-parts formula*

$$\int_Q h \Delta g \, dx = - \int_Q \langle \nabla h, \nabla g \rangle \, dx.$$

If in addition (5.1) *holds for some weight w, then we may ensure $\|f - g\|_{W_w^{1,q}(\Omega)} < \epsilon$ for any fixed $q \in [1, \infty)$.*

The corollary above is the driving force behind the specific application we consider in this section, though we defer most of the details to [13]. We now begin the development. To this end, consider the set

$$\mathbf{M}_+^{n \times n} := \left\{ X \in \mathbf{M}^{n \times n} : \det X \geq 0 \right\},$$

where $\mathbf{M}^{n \times n}$ is endowed with the trace product and the Frobenius norm. The reasons for studying this set are two-fold: firstly, it was one of the original motivations for developing the results in this paper. This is due to its role in a certain application involving stochastic processes, which we outline below. Secondly, and more importantly in the context of the present paper, it allows us to illustrate how the components of the iterative construction in Section 3 can be used more generally. Specifically, our goal is to establish an improved version of Theorem 2.8, discussed below.

We first need a few preliminaries. We begin by observing that Propositions 4.4 and Theorem 4.8 show that the collection of C^∞ manifolds

$$M_k := \{ X \in \mathbf{M}^{n \times n} : \mathrm{rk}\, X = k \} \quad \text{for } k = 0, \ldots, n$$

is a normally flat C^∞-stratification of $\mathbf{M}^{n \times n}$. Consequently, the collection

$$\mathscr{A}_+ := \{ M_0, \ldots, M_{n-1}, M_{n+} \},$$

where M_{n+} is the open set

$$M_{n+} := \{ X \in \mathbf{M}^{n \times n} : \det X > 0 \},$$

is a normally flat C^∞-stratification of $\mathbf{M}_+^{n \times n}$.

For a point $x \in \mathbf{R}^n$, define a new point $x^+ \in \mathbf{R}^n$ by setting $x_i^+ = \frac{1}{x_i}$ if $x_i \neq 0$, and $x_i^+ = 0$ otherwise. Then if a matrix $X \in \mathbf{M}^{n \times n}$ has a decomposition $X = U \operatorname{Diag} x\, V^\top$ with $U, V \in O(n)$, the *Moore-Penrose inverse* of X is given by $X^+ := V \operatorname{Diag} x^+\, U^\top$. The transpose of the Moore-Penrose inverse, which we denote by

$$X^\mp := (X^+)^\top = (X^\top)^+,$$

will play a role in our development. The main result of this section is the following.

Theorem 6.2 (Approximation on matrices with positive determinant)**.** *Consider the set of matrices $Q = \mathbf{M}_+^{n \times n}$, along with the C^∞-stratification $\mathscr{A}_+$, and let f and ε be as in Theorem 2.8. Then there is a function $g : \mathbf{R}^n \to \mathbf{R}$ that satisfies all the properties in Theorem 2.8, as well as the property that the mapping*

$$X \mapsto \langle X^\mp, \nabla g(X) \rangle$$

is continuous.

Remark 6.3. In the case $\det X > 0$, that is, when X is invertible, we have $X^{\mp} = (X^{-1})^{\top} = \nabla \ln \det(X)$.

The importance of Theorem 6.2 comes from its significance in the study of a class of stochastic processes called the *matrix-valued Bessel processes*, introduced in [13]. These are Markov processes whose infinitesimal generator arises as an extension of the differential operator

$$\mathscr{L} = \frac{1}{2}\Delta + \frac{\delta - 1}{2}\langle \nabla \ln \det(X), \nabla \rangle,$$

acting on a suitable class of functions on $\mathbf{M}_+^{n \times n}$. Here $\delta > 0$ is fixed. A consequence of Theorem 6.2 is that any compactly supported Lipschitz continuous function on $\mathbf{M}_+^{n \times n}$ can be approximated uniformly by some compactly supported C^2 function g that satisfies a Neumann boundary condition and has the crucial property that $\mathscr{L}g$ is bounded (in fact, the boundedness of $\mathscr{L}g$ is the incremental benefit that motivates Theorem 6.2). For such functions g, it is possible to obtain an integration-by-parts formula,

$$\int_Q h(X)\mathscr{L}g(X)w(X)dX = -\frac{1}{2}\int_Q \langle \nabla h(X), \nabla g(X) \rangle w(X)dX,$$

where $w(X) = (\det X)^{\delta - 1}$ can be shown to be locally integrable on $\mathbf{M}_+^{n \times n}$ and h varies over some class of test functions. This is perhaps not so surprising in light of Corollary 6.1. The right-hand side of the above equation extends to a *Dirichlet form* $\mathscr{E}(g, h)$ on the weighted space $L_w^2(\mathbf{M}_+^{n \times n})$. This allows one to apply standard existence results for the associated stochastic process, which we denote by $\mathbf{X} = (\mathbf{X}_t : t \geq 0)$. Furthermore, for a function g that satisfies the integration-by-parts formula, and for which $\mathscr{L}g$ is bounded, the process

$$g(\mathbf{X}_t) - g(\mathbf{X}_0) - \int_0^t \mathscr{L}g(\mathbf{X}_s)ds, \qquad t \geq 0,$$

is a martingale. Since, by Theorem 6.2, this holds for a large class of functions g, it is possible to obtain a good description of the probability law of $\mathbf{X}$. In particular, powerful uniqueness results become available. For a detailed discussion, see [13].

After this brief digression, let us proceed to the proof of Theorem 6.2. The following lemma establishes a key property of the Moore-Penrose inverse.

Lemma 6.4 (Projection of the Moore-Penrose inverse). *Consider a stratum $M \in \mathscr{A}_+$ and a matrix $X \in \mathbf{M}^{n \times n}$ that when projected onto M yields a single matrix Y. Then the equality*

$$P_{T_Y M}(X^{\mp}) = Y^{\mp} \quad \text{holds.}$$

Proof. Let k be the rank of the matrices in M, and note that since $P_M(X)$ is a singleton, the inequality $\mathrm{rk}\, X \geq k$ must hold. Let $X = U\Sigma V^{\top}$ be a singular value decomposition of X, and let Σ_k be the diagonal matrix obtained from Σ by setting all but the k largest entries to zero. Then we have the equality $Y = U\Sigma_k V^{\top}$, as is apparent from Lemma 4.5. Observe that

$$X^{\mp} = Y^{\mp} + \left(X^{\mp} - Y^{\mp}\right) = U\Sigma_k^+ V^{\top} + U\left(\Sigma^+ - \Sigma_k^+\right)V^{\top}.$$

The desired result follows once we show that the right-hand side is the decomposition of $X^{\mp}$ as a direct sum in $T_Y M \oplus N_Y M$. To do this, first note that for small t, the matrix

$$Y + tU\Sigma_k^+ V^{\top} = U(\Sigma_k + t\Sigma_k^+)V^{\top}$$

has rank k, and hence lies in M. Consequently, the inclusion $U\Sigma_k^+ V^\top \in T_Y M$ holds. Next, for small t we have

$$P_M \left(Y + tU \left(\Sigma^+ - \Sigma_k^+\right) V^\top\right) = P_M \left(U \left(\Sigma_k + t(\Sigma^+ - \Sigma_k^+)\right) V^\top\right) = Y,$$

thus verifying that the matrix $U \left(\Sigma^+ - \Sigma_k^+\right) V^\top$ lies in $N_Y M$, as required. $\qquad\square$

Next, we establish the validity of the following inductive step. We will use the well-known property that the Moore-Penrose inverse mapping $X \mapsto X^+$ is continuous on each stratum M of $\mathscr{A}_+$.

Proposition 6.5. *Consider the setup and notation given in Subsection 3.3, with $\mathscr{A} = \mathscr{A}_+$. Assume that the following properties hold.*

(1) U_δ is contained in the set $\widehat{U}$ from Lemma 3.7,
(2) f is C^1 smooth, and we have $\nabla f(X) \in T_X \Gamma$ for every $X \in \Gamma$,
(3) each stratum $L \in \mathscr{A}_+$ with $L \subset \Gamma$ has a neighborhood V such that

$$f(X) = f \circ P_L(X) \quad \text{for all} \quad X \in V,$$

(4) the map $X \mapsto \langle \nabla f(X), X^\mp \rangle$ is continuous at every $\overline{X} \in \Gamma$.

Define a function $g = \phi f \circ P_M + (1 - \phi)f$. Then the mapping

$$X \mapsto \langle \nabla g(X), X^\mp \rangle$$

is continuous at every $\overline{X} \in \Gamma \cup M$.

Proof. First note that, by Proposition 3.6, the function g is again C^1 smooth, with $\nabla f = \nabla g$ on Γ. Consider a matrix $\overline{X} \in M$, and write $Y = P_M(X)$ as usual for X near M. Since we have $g = f \circ P_M$ near M, we deduce

$$\langle \nabla g(X), X^\mp \rangle = \langle \nabla f(Y), DP_M(X)X^\mp \rangle.$$

Using Lemma 6.4 and the observation $DP_M(X)U = 0$ for $U \in N_Y M$, we obtain

$$DP_M(X)X^\mp = DP_M(X)P_{T_Y M}(X^\mp) = DP_M(X)Y^\mp.$$

From the convergence $X \to \overline{X}$, we deduce $Y \to \overline{X}$, and consequently $Y^\mp \to \overline{X}^\mp$, since the matrices Y and $\overline{X}$ have the same rank. We successively conclude that

$$\begin{aligned}
\lim_{X \to \overline{X}} \langle \nabla g(X), X^\mp \rangle &= \langle \nabla f(\overline{X}), DP_M(\overline{X})\overline{X}^\mp \rangle \\
&= \langle \nabla(f \circ P_M)(\overline{X}), \overline{X}^\mp \rangle \\
&= \langle \nabla g(\overline{X}), \overline{X}^\mp \rangle,
\end{aligned}$$

thereby verifying continuity at $\overline{X}$. Consider now a stratum $L \subset \Gamma$ and a point $\overline{X} \in L$. Since we have $\nabla f = \nabla g$ outside U_δ, we can assume $X \in U_\delta$. We obtain

$$(6.2) \quad \langle \nabla g(X), X^\mp \rangle - \langle \nabla f(X), X^\mp \rangle = \phi(X)\langle \nabla(f \circ P_M)(X) - \nabla f(X), X^\mp \rangle$$
$$+ (f \circ P_M(X) - f(X)) \langle \nabla \phi(X), X^\mp \rangle.$$

Observe that whenever X is close enough to $\overline{X}$, we have

$$f \circ P_M(X) = f \circ P_L \circ P_M(X) = f \circ P_L(X) = f(X),$$

due to the assumption on f and normal flatness of $\mathscr{A}_+$. Consequently, the right-hand side of (6.2) vanishes as soon as X gets sufficiently close to $\overline{X}$. The result follows. $\qquad\square$

Proof of Theorem 6.2. We simply carry out the same iterative procedure as in the proofs of Theorems 2.4 and 2.8 (see Subsection 3.4), while in addition applying Proposition 6.5 in each step. Note that each time this is done, the hypotheses (2) and (3) will be satisfied since Propositions 3.6 and 3.8 were applied in some earlier iteration (initially these hypotheses are vacuously true). The hypothesis (1) is easily achieved by shrinking U_δ if necessary, and (4) holds due to earlier applications of Proposition 6.5. □

Acknowledgements

References

[1] Robert A. Adams and John J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics (Amsterdam), vol. 140, Elsevier/Academic Press, Amsterdam, 2003. MR2424078 (2009e:46025)

[2] D. Azagra, J. Ferrera, F. López-Mesas, and Y. Rangel, *Smooth approximation of Lipschitz functions on Riemannian manifolds*, J. Math. Anal. Appl. **326** (2007), no. 2, 1370–1378, DOI 10.1016/j.jmaa.2006.03.088. MR2280987 (2007j:26010)

[3] S. Basu, R. Pollack, and M. Roy, *Algorithms in Real Algebraic Geometry (Algorithms and Computation in Mathematics)*, Springer-Verlag, Berlin, 2006. MR2248869

[4] Seng-Kee Chua, *Extension theorems on weighted Sobolev spaces*, Indiana Univ. Math. J. **41** (1992), no. 4, 1027–1076, DOI 10.1512/iumj.1992.41.41053. MR1206339 (94a:46035)

[5] M. Coste, *An introduction to semialgebraic geometry*, RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.

[6] A. Daniilidis, D. Drusvyatskiy, and A.S. Lewis, *Orthogonal invariance and identifiability*, SIAM J. Matrix Anal. Appl. **35** (2014), no. 2, 580–598. MR3205737

[7] Aris Daniilidis, Adrian Lewis, Jérôme Malick, and Hristo Sendov, *Prox-regularity of spectral functions and spectral sets*, J. Convex Anal. **15** (2008), no. 3, 547–560. MR2431411 (2009f:49013)

[8] A. Daniilidis, J. Malick, and H. S. Sendov, *Locally symmetric submanifolds lift to spectral manifolds*, arXiv:1212.3936 [math.OC] (2014).

[9] ⸻, *Spectral (isotropic) manifolds and their dimension*, Journal d'Analyse Mathématique (to appear) (2014).

[10] Mark Goresky and Robert MacPherson, *Stratified Morse theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 14, Springer-Verlag, Berlin, 1988. MR932724 (90d:57039)

[11] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985. MR832183 (87e:15001)

[12] Alois Kufner and Bohumír Opic, *How to define reasonably weighted Sobolev spaces*, Comment. Math. Univ. Carolin. **25** (1984), no. 3, 537–554. MR775568 (86i:46036)

[13] M. Larsson, *Matrix-valued Bessel processes*, arXiv:1212.4986 (2012).

[14] John M. Lee, *Introduction to smooth manifolds*, Graduate Texts in Mathematics, vol. 218, Springer-Verlag, New York, 2003. MR1930091 (2003k:58001)

[15] A. S. Lewis, *Derivatives of spectral functions*, Math. Oper. Res. **21** (1996), no. 3, 576–588, DOI 10.1287/moor.21.3.576. MR1403306 (97g:90141)

[16] A. S. Lewis, *Nonsmooth analysis of eigenvalues*, Math. Program. **84** (1999), no. 1, Ser. A, 1–24, DOI 10.1007/BF02925352. MR1687292 (2000d:49029)

[17] Adrian S. Lewis and Jérôme Malick, *Alternating projections on manifolds*, Math. Oper. Res. **33** (2008), no. 1, 216–234, DOI 10.1287/moor.1070.0291. MR2393548 (2009k:90110)

[18] Adrian S. Lewis and Hristo S. Sendov, *Nonsmooth analysis of singular values. I. Theory*, Set-Valued Anal. **13** (2005), no. 3, 213–241, DOI 10.1007/s11228-004-7197-7. MR2162512 (2006f:49022)

[19] John Mather, *Notes on topological stability*, Bull. Amer. Math. Soc. (N.S.) **49** (2012), no. 4, 475–506, DOI 10.1090/S0273-0979-2012-01383-6. MR2958928

[20] Norman G. Meyers and James Serrin, $H = W$, Proc. Nat. Acad. Sci. U.S.A. **51** (1964), 1055–1056. MR0164252 (29 #1551)

[21] Markus J. Pflaum, *Analytic and geometric study of stratified spaces*, Lecture Notes in Mathematics, vol. 1768, Springer-Verlag, Berlin, 2001. MR1869601 (2002m:58007)

[22] Jean-Baptiste Poly and Gilles Raby, *Fonction distance et singularités* (French, with English summary), Bull. Sci. Math. (2) **108** (1984), no. 2, 187–195. MR769927 (86d:32008)

[23] Marie-Hélène Schwartz, *Champs radiaux sur une stratification analytique* (French), Travaux en Cours [Works in Progress], vol. 39, Hermann, Paris, 1991. MR1096495 (92i:32041)

[24] Hristo S. Sendov, *The higher-order derivatives of spectral functions*, Linear Algebra Appl. **424** (2007), no. 1, 240–281, DOI 10.1016/j.laa.2006.12.013. MR2324386 (2008c:47035)

[25] Bengt Ove Turesson, *Nonlinear potential theory and weighted Sobolev spaces*, Lecture Notes in Mathematics, vol. 1736, Springer-Verlag, Berlin, 2000. MR1774162 (2002f:31027)

[26] Lou van den Dries and Chris Miller, *Geometric categories and o-minimal structures*, Duke Math. J. **84** (1996), no. 2, 497–540, DOI 10.1215/S0012-7094-96-08416-1. MR1404337 (97i:32008)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195-4350 — AND — DEPARTMENT OF COMBINATORICS AND OPTIMIZATION, UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA N2L 3G1
  E-mail address: `ddrusv@uw.edu`
  URL: `people.orie.cornell.edu/dd379/`

SWISS FINANCE INSTITUTE, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, SWITZERLAND
  E-mail address: `larsson@epfl.ch`
  URL: `sfi.epfl.ch/larsson`