JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Sober, Barak; Aizenbud, Yariv; Levin, David

Title: Approximation of functions over manifolds : A Moving Least-Squares approach

Year: 2021

Version: Accepted version (Final draft)

Copyright: © 2020 Elsevier B.V. All rights reserved.

Rights: CC BY-NC-ND 4.0

Rights url: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the original version:

Sober, B., Aizenbud, Y., & Levin, D. (2021). Approximation of functions over manifolds : A Moving Least-Squares approach. Journal of Computational and Applied Mathematics, 383, Article 113140. https://doi.org/10.1016/j.cam.2020.113140

Approximation of Functions over Manifolds: A Moving Least-Squares Approach

Barak Sober[1] Yariv Aizenbud[1,2] David Levin[1]

[1] School of Mathematical Sciences, Tel Aviv University, Israel

[2] Faculty of Information Technology, Jyväskylä University, Jyväskylä, Finland

29th September 2021

Abstract

We present an algorithm for approximating a function defined over a d-dimensional manifold utilizing only noisy function values at locations sampled from the manifold with noise. To produce the approximation we do not require any knowledge regarding the manifold other than its dimension d. We use the Manifold Moving Least-Squares approach of [33] to reconstruct the atlas of charts and the approximation is built on-top of those charts. The resulting approximant is shown to be a function defined over a neighborhood of a manifold, approximating the originally sampled manifold. In other words, given a new point, located near the manifold, the approximation can be evaluated directly on that point. We prove that our construction yields a smooth function, and in case of noiseless samples the approximation order is $\mathcal{O}(h^{m+1})$, where h is a local density of sample parameter (i.e., the fill distance) and m is the degree of a local polynomial approximation, used in our algorithm. In addition, the proposed algorithm has linear time complexity with respect to the ambient-space's dimension. Thus, we are able to avoid the computational complexity, commonly encountered in high dimensional approximations, without having to perform non-linear dimension reduction, which inevitably introduces distortions to the geometry of the data. Additionaly, we show numerical experiments that the proposed approach compares favorably to statistical approaches for regression over manifolds and show its potential.

keywords: Manifold learning, Regression over manifolds, Moving Least-Squares, Dimension reduction, High dimensional approximation, Out-of-sample extension
MSC classification: 65D99
(Numerical analysis - Numerical approximation and computational geometry)

1 Introduction

Approximating a function defined over an extremely large dimensional space from scattered data is a very challenging task. First, from the sample-set perspective, to achieve a constant sampling resolution, the number of points grows exponentially with respect to the number of dimensions. For example, a uniform grid on $[0,1]^n$ with resolution of 0.1 requires 10^n samples. Second, the high dimensionality of the domain introduces serious computational issues. Thus, the performance of both parametric and non-parametric approximations (or regressions) deteriorates sharply as the dimension increases [9, 14, 16]. These types of problems, sometimes

referred to as *the curse of dimensionality*, occur frequently in many scientific disciplines since data, originating from various sources and of various types, is becoming more and more available.

In the last three decades, there has been a rapid development of mathematical frameworks aiming to deal with complexity challenges, originating from high dimensional data. In many works, there exists an underlying assumption that the high dimensional domain of a sample set (i.e., point cloud) has a lower intrinsic dimension (e.g., [2, 8, 13, 17, 18, 19, 29, 30, 34]). In other words, the data points $\{r_i\}_{i=1}^N \subset \mathbb{R}^n$ are samples of a lower dimensional manifold $\mathcal{M}^d$, where d is the intrinsic dimension of $\mathcal{M}$ and $d \ll n$. Therefore, a natural way of reducing the effective number of parameters (in case of a parametric estimation) as well as computation complexity, would be to harvest this geometric relationship among the points. The framework of dimension reduction proposes to embed the data into a lower dimension Euclidean domain while maintaining some sort of local distances (for a survey see [21]). Then, the lower dimensional representations can be utilized to perform function approximation over the data. However, such methods inherently introduce distortion to the input data, as, for example, the curvature information is lost after performing such an embedding. In addition, performing out-of-sample extensions, in most of these methods, will require the re-computation of the embedding. Another effective framework, dealing with such problems is the Support Vector Machine based methods [32, 31], which in some sense are another way of performing non-linear dimension reduction prior to performing regression.

A somewhat different approach, designed to deal with a more general definition of low dimensionality, is the Geometric Multi-Resolution Analysis (GMRA), introduced in a series of papers [4, 11, 24, 25]. The GMRA utilizes a local affine representation of the data, in order to store the data in a multi-resolution dictionary. Thus, it does not project the data onto a lower dimensional Euclidean domain, but creates a tree-like representation of the original data based upon partitioning and performing local Singular Value Decomposition. This approach, leads to a faithful, locally sparse, representation of the input data in case the original tree was built from clean samples. Subsequently, these representations can be used to approximate functions over the original input data (e.g., [35]). However, this approach does not aim at yielding smooth or even continuous approximations.

In this work, we utilize a Moving Least-Squares (MLS) based framework. The MLS approximation was originally designed for the purpose of smoothing and interpolating scattered data, sampled from some multivariate function [20, 22, 26, 27]. Then, it evolved to deal with surfaces (i.e., $n-1$ dimensional manifolds in $\mathbb{R}^n$), which can be viewed as a function locally rather than globally [3, 23]. This has been generalized lately in [33] to the Manifold - Moving Least-Squares (M-MLS), which deals with manifolds of an arbitrary dimension d embedded in $\mathbb{R}^n$. This M-MLS framework, which will be described formally in Section 2, harvests an implicit construction of the manifold's atlas of charts. Explicitly, for each point $p \in \mathcal{M}$ a local coordinate chart (mapping a neighborhood of p into a Euclidean d-dimensional linear space) is constructed. Thus, the data is not being projected into a lower dimension Euclidean domain nor is it being compressed.

The contribution of the current paper is the approximation of a function defined over a manifold based upon discretely sampled data. We utilize the M-MLS atlas of charts to provide a Moving Least-Squares polynomial approximation of the function. Since we approximate the function through its atlas of charts, on a local level the approximation is defined from $\mathbb{R}^d$ to $\mathbb{R}$. Thus, our approximation framework avoids the curse of dimensionality without having to globally project the sample set into a lower dimensional Euclidean space. We show in Theorem 3.1 that our theoretical approximant is a smooth function defined on a neighborhood of the

manifold domain. In addition, in Theorem 3.2, we show that in case of clean samples the approximation yields an $\mathcal{O}(h^{m+1})$ approximation order, where h is the fill distance with respect to the manifold domain and m is the local polynomial degree. Our algorithmic approach has linear complexity with respect to the ambient space's dimension n, which makes the proposed method realizable in cases where n is extremely large. Furthermore, performing out-of-sample-extension with this framework is trivial and does not require any further computations.

The rest of the paper is organized as follows: in Section 2 we describe the MLS approximation framework; in Section 3 we describe the proposed approach of function approximation over manifolds; and in Section 4 we give some numerical examples showing the potential of the proposed method.

2 Preliminaries – the M-MLS framework

2.1 MLS For Function Approximation

The moving least-squares for function approximation was first presented by Mclain in [26] in order to approximate a function from noisy samples. Let $\{x_i\}_{i=1}^N$ be a set of distinct scattered points in $\mathbb{R}^d$ and let $\{f(x_i)\}_{i=1}^N$ be the corresponding sampled values of some function $f : \mathbb{R}^d \to \mathbb{R}$. Then, the mth degree moving least-squares approximation to f at a point $x \in \mathbb{R}^d$ is defined as $p_x(x)$, where

$$p_x = \underset{p\in\Pi_m^d}{\operatorname{argmin}} \sum_{i=1}^{N} (p(x_i) - f(x_i))^2 \theta(\|x - x_i\|), \tag{1}$$

$\theta(t)$ is a non-negative weight function rapidly decreasing as $t \to \infty$ (e.g. a Gaussian, or an indicator function on an interval around zero), $\|\cdot\|$ is the Euclidean norm and Π_m^d is the space of polynomials of total degree m in $\mathbb{R}^d$. Then, the MLS approximation is defined as,

$$\widetilde{f}(x) \triangleq p_x(x). \tag{2}$$

Notice, that if $\theta(t)$ is of finite support then the approximation is made local, and if $\lim_{t\to 0} \theta(t) = \infty$ the MLS approximation interpolates the data.

We wish to quote here two previous results regarding the resulting approximation presented in [22]. In Section 3 we will prove properties extending these theorems to the general case of approximation of functions over a d-dimensional manifold embedded in $\mathbb{R}^n$.

Theorem 2.1. *Let $\theta(t) \in C^\infty$ and let the distribution of the data points $\{x_i\}_{i=1}^N$ be such that the problem is well conditioned (i.e., the least-squares matrix of* (1) *is invertible). Then the MLS approximation is a C^∞ function.*

Remark 2.2. In case $\lim_{t\to 0} \theta(t) = \infty$, the the MLS approximant will interpolate the data points $\{f(x_i)\}_{i=1}^N$.

The second result, dealing with the approximation order with respect to the norm

$$\left\| \widetilde{f}(x) - f(x) \right\|_{\Omega,\infty} \triangleq \max_{x\in\Omega} \left| \widetilde{f}(x) - f(x) \right|,$$

necessitates the introduction of the following definition:

Definition 1. ***h-ρ-δ sets of fill distance h, density $\leq \rho$, and separation $\geq \delta$.*** *Let Ω be a d-dimensional domain in $\mathbb{R}^n$, and consider sets of data points in Ω. We say that the set $X = \{x_i\}_{i=1}^I$ is an h-ρ-δ set if:*

1. *h is the fill distance with respect to the domain Ω*

$$h = \sup_{x\in\Omega} \min_{x_i\in X} \|x - x_i\| . \tag{3}$$

2. $$\#\left\{X \cap \overline{B}_{kh}(y)\right\} \leq \rho \cdot k^d, \quad q \geq 1, \quad y \in \mathbb{R}^n. \tag{4}$$

   Here $\#Y$ denotes the number of elements in a given set Y, while $\overline{B}_r(x)$ is the closed ball of radius r around x.

3. *$\exists\delta > 0$ such that*

$$\|x_i - x_j\| \geq h\delta, \quad 1 \leq i < j \leq I. \tag{5}$$

Remark 2.3. Note, that in [22], the fill distance h was defined slightly different. However, the two definitions are equivalent.

Theorem 2.4. *Let f be a function in $C^{m+1}(\Omega)$ with an h-ρ-δ sample set. Then for fixed ρ and δ, there exists a fixed $k > 0$, independent of h, such that the approximant given by equation* (1) *is well conditioned for θ with a finite support of size $s = kh$. In addition, the approximant yields the following error bound:*

$$\left\|\widetilde{f}(x) - f(x)\right\|_{\Omega,\infty} < M \cdot h^{m+1} \tag{6}$$

for some M independent of h .

Remark 2.5. Although both Theorem 2.1 and Theorem 2.4 are stated in [22] assuming an interpolatory conditions (i.e., the weight function satisfies $\theta(0) = \infty$), the proofs articulated there are still valid taking any compactly supported non-interpolatory weight function. These proofs are based upon a representation of the solution to the minimization problem (1) through a multiplication of smooth matrices. These matrices remain smooth even when the interpolatory condition is dropped (see the proofs of Theorem 3.1 and Proposition 3.3, which utilize a similar proof technique).

Remark 2.6. Notice that the weight function θ in the definition of the MLS for function approximation is applied on the distances in the domain. In what follows, we will apply θ on the distances between points in $\mathbb{R}^n$ as we aim at approximating functions over manifold domains and as such each local coordinate chart should be affected by neighboring points in $\mathbb{R}^n$ (see Fig. 1). In order for us to be able to utilize Theorems 2.1 and 2.4, the distance in the weight function of equation (1) should be $\theta(\|(x, 0) - (x_i, f(x_i))\|)$ instead of $\theta(\|x - x_i\|)$ (see Figure 1). Nevertheless, as stated above, the proofs of both theorems as presented in [22] rely on the representation of the solution to the minimization problem as a multiplication of smooth matrices. These matrices will still remain smooth after replacing the weight, as the new weighting is still smooth. Moreover, as explained in [22] the approximation order remains the same even if the weight function is not compactly supported in case the weight function decays fast enough (e.g., by taking $\theta(r) \triangleq e^{-\frac{r^2}{h^2}}$).

Remark 2.7. The approximation order remains the same even if the weight function is not compactly supported in case the weight function decays fast enough (e.g., by taking $\theta(t) \triangleq e^{-\frac{t^2}{h^2}}$).

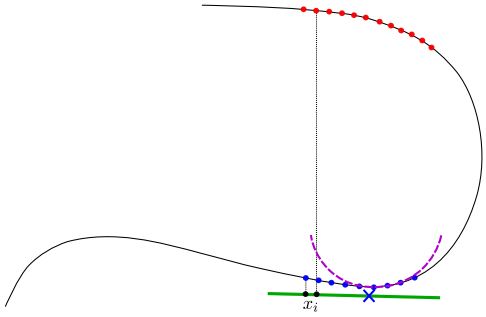

Figure 1: The effect of remote points when taking $\theta(\|(x,0)-(x_i,f(x_i))\|)$ instead of $\theta(\|x-x_i\|)$. Assuming that the green line represents a given coordinate system around the point x (marked by the blue $\times$), by taking the weights $\theta(\|x-x_i\|)$ the contribution of both the red and blue samples to the weighted cost function would be $\mathcal{O}(h^{m+1})$. Alternatively, by taking $\theta(\|(x,0)-(x_i,f(x_i))\|)$ with a fast decaying weight function the contribution of the red points would be negligible. Thus, the approximation (in purple) would fit the behavior of the blue points alone.

2.2 The M-MLS Projection

We now turn to the M-MLS projection procedure, introduced in [33], upon which we base the results of the current paper. Let $\mathcal{M}$ be a manifold of dimension d lying in $\mathbb{R}^n$, and let the samples of $\mathcal{M}$ hold the following conditions.

Noisy Sampling Assumptions

1. $\mathcal{M} \in C^2$ is a closed (i.e., compact and boundaryless) submanifold of $\mathbb{R}^n$.
2. $\tilde{R} = \{\tilde{r}_i\}_{i=1}^I \subset \mathcal{M}$ is an h-ρ-δ sample set with respect to the domain $\mathcal{M}$ (see Definition 1).
3. $R = \{r_i\}_{i=1}^I$ are noisy samples of R; i.e., $r_i = \tilde{r}_i + n_i$.
4. $\|n_i\| < \sigma$

Given a point r near $\mathcal{M}$ the Manifold Moving Least-Squares (M-MLS) projection of r is defined through two sequential steps:

1. Find a local d-dimensional affine space $H(r)$ around an origin $q(r)$ such that H approximates the sampled points. Explicitly, $H = q + Span\{e_k\}_{k=1}^d$, where $\{e_k\}_{k=1}^d$ is some orthonormal basis of $\mathbb{R}^d$. H will be used as a local coordinate system.
2. Similar to the function approximation described in Equation (1), define the projection of r using a local polynomial approximation $p : H \to \mathbb{R}^n$ of $\mathcal{M}$ over the new coordinate system. Explicitly, we denote by x_i the projections of r_i onto H and then define the samples of a function f by $f(x_i) = r_i$. Accordingly, the d-dimensional polynomial p is an approximation of the vector valued function f.

Remark 2.8. Since $\mathcal{M}$ is a differentiable manifold it can be viewed locally as a graph of a function from the tangent space $T_p\mathcal{M} \simeq \mathbb{R}^d$ to $(T_p\mathcal{M})^\perp \simeq \mathbb{R}^{n-d}$. Thus, we are looking for a coordinate system H and refer to the manifold $\mathcal{M}$ locally as a graph of some function $f : H \to \mathbb{R}^{n-d}$.

Remark 2.9. Throughout the paper, whenever we encounter an affine space

$$L = x + span\{e_k\}_{k=1}^{d},$$

we will denote its homogeneous part, which belongs to the Grassmannian (i.e., the linear space without the shift by x) as

$$\mathcal{G}L = span\{e_k\}_{k=1}^{d}.$$

In this paper, we intend to utilize the first step of the M-MLS to provide an atlas of charts for the manifold. This atlas would serve as the basis of our construction of function approximation as presented in Section 2.1. Therefore, we wish to present here formally just the first step in the M-MLS, and quote some results that will be useful to our analysis.

Step 1 - Formal Description

Let

$$J(r; q, H) = \sum_{i=1}^{I} d(r_i, H)^2 \theta(\|r_i - q\|) \tag{7}$$

be a cost function. We wish to Find a d-dimensional affine space $H(r)$, and a point $q(r)$ on $H(r)$, such that

$$q(r), H(r) = \underset{q,H}{\operatorname{argmin}}\, J(r; q, H) \tag{8}$$

under the constraints

1. $r - q \perp H$
2. $q \in B_\mu(r)$
3. $\#\left(R \cap B_{\sigma+h}(q)\right) \neq 0$,

where $d(r_i, H)$ is the Euclidean distance between the point r_i and the affine subspace H, $B_\eta(x)$ is an open ball of radius η around x, h is the fill distance from the h-ρ-δ set in the sampling assumptions.

We wish to give some motivation to the definition of the minimization problem portrayed above. Constraint 2 limits the search space to a neighboring part of the manifold, whereas constraint 3, narrows it further to the vicinity of the samples, and, thus, voids the possibility of achieving solutions with zero value of J (caused by the fact that there are no sample points in the support of θ); see Figure 2. The necessity of constraint 1 is less obvious though. First, minimizing $J(r; q, H)$ without this constraint will just yield a local PCA approximation around an unknown point q. Second, the added constraint links the approximation to the point r, which we aim to project onto $\mathcal{M}$, as well as generalizes the idea of the Euclidean projection onto a manifold. Explicitly, if we have a point r "close enough" to a given manifold $\mathcal{M}$ there exists a unique projection $P(r)$ of the point r onto $\mathcal{M}$. In addition, we know that this projection maintains $r - P(r) \perp T_{P(r)}\mathcal{M}$, which is echoed in constraint 1. This concept of a unique projection domain is better expressed by the definition of reach as introduced in [15] .

Definition 2 (Reach)**.** *The reach of a subset A of $\mathbb{R}^n$, is the largest τ (possibly ∞) such that for any $x \in \mathbb{R}^n$ that maintains $dist(A, x) \leq \tau$, there exists a unique point $P_A(x) \in A$, nearest to x. We denote $rch(A) = \tau$.*

Following this definition let

$$U_{reach} \triangleq \{x \in \mathbb{R}^n \mid dist(x, \mathcal{M}) < rch(\mathcal{M})\}. \tag{9}$$

In our context, we refer to manifolds with positive reach. Accordingly, for a point $r \in U_{reach}$, there exists a unique projection $P(r)$ onto the manifold $\mathcal{M}$. As shown in [33], the minimizers $q(r), H(r)$ of Equation (8) converge to $P(r), T_{P(r)}\mathcal{M}$ respectively as the fill distance h tends to zero (given some assumptions on the support of θ) for r in some fixed size neighborhood $U \subset U_{reach}$.

Therefore, in order to generalize the concept of a reach neighborhood (relevant for the limit case) to a domain where the procedure yields a unique approximation, we assume the existence of a *Uniqueness Domain*.

Assumption 2.10 (Uniqueness Domain)**.** *We assume that there exists an ϵ-neighborhood of the manifold*

$$U_{unique} \triangleq \{x \in \mathbb{R}^n \mid dist(x, \mathcal{M}) < \epsilon < rch(\mathcal{M})\}, \tag{10}$$

such that for any $r \in U_{unique}$ the minimization problem (8) *has a unique local minimum $q(r) \in B_\mu(r)$, for some constant $\mu < rch(\mathcal{M})/2$.*

In Lemma 4.4 of [33], it is shown that in the limit case, where $h \to 0$, there exists a uniqueness domain as described above for all closed manifolds. Note that in order to achieve a unique solution for a given r the decay of θ should be bounded from below, and μ should be large enough such that $P(r) \in B_\mu(r)$. Figure 2 illustrates a section of the reach neighborhood of a circle restricting r such that $d(r, \mathcal{M}) < rch(\mathcal{M})/4$ and setting $\mu = rch(\mathcal{M})/2$. To some extent, the circle example "bounds" the behaviour of the data in every 2D section of the manifold, as the reach bounds the sectional curvature of the manifold. An illustration of a uniqueness domain for a cleanly sampled curve embedded in $\mathbb{R}^3$ can be seen in Figure 3.

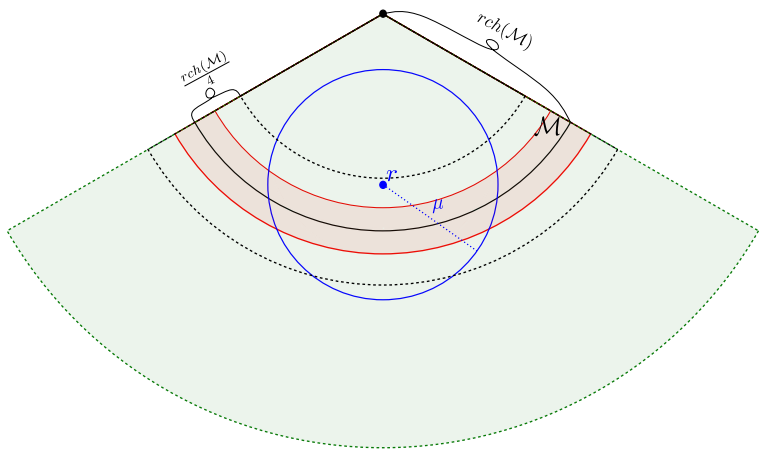

Figure 2: An illustration of a section of a uniqueness domain of a circle where we take r such that $d(r, \mathcal{M}) < rch(\mathcal{M})/4$ and set $\mu = rch(\mathcal{M})/2$. The black dot above is the center of the circle; the green region is the reach neighborhood of $\mathcal{M}$; the red region is the noisy region from which we sample the manifold (i.e., the support of the distribution of sample points); the blue ball is the search region defined in constraint 2 of Equation (8)*.*

As shown in [33], if we fix q, the affine space minimizing (8) is defined uniquely (see Lemma 4.5 there). Explicitly, we denote it by $H'(r; q)$ and we can thus reformulate (8) as

$$J^*(r; q) = \sum_{i=1}^{I} d(r_i, H'(r; q))^2 \theta(\|r_i - q\|), \tag{11}$$

We would use this representation in one of the Theorems we show below. Furthermore, we wish to quote here two results from [33], that will assist in the analysis of the function approximation in Section 3.

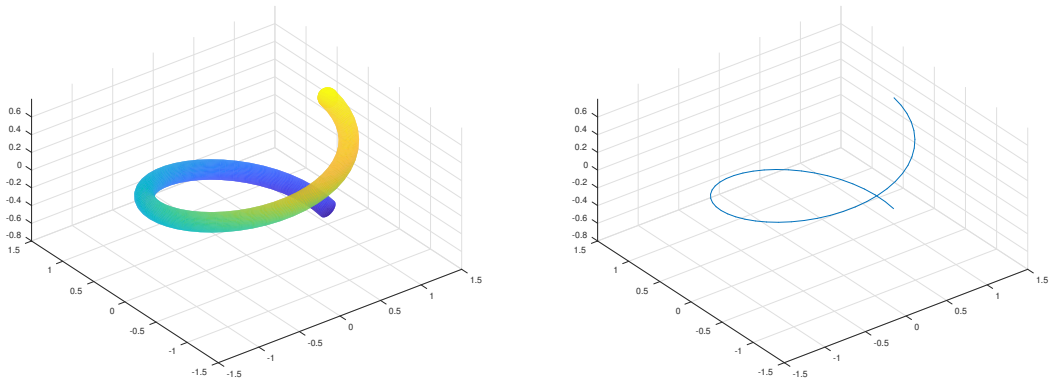

Figure 3: An illustration of a uniqueness domain. Right - a 1-dimensional manifold $\mathcal{M}$ embedded in $\mathbb{R}^3$. Left - a uniqueness domain U of $\mathcal{M}$.

Lemma 2.11 (Projection property)**.** *Let the Noisy Sampling Assumptions of Section 2.2 hold. Let r be in the uniqueness domain U_{unique} of assumption (2.10) and let $q(r)$ and $H(r)$ be the minimizers of $J(r; q, H)$ as defined above. Then for any point $\tilde{r} \in U_{unique}$ s.t. $\|\tilde{r} - q(r)\| < \mu$ and $\tilde{r} - q(r) \perp H(r)$ we get $q(\tilde{r}) = q(r)$ and $H(\tilde{r}) \equiv H(r)$*

Next, we need a notion of a smooth change for the coordinate system. In other words, we wish to define a smooth change of affine spaces.

Definition 3. *Let $H(r)$ be a parametric family of d-dimensional affine sub-spaces of $\mathbb{R}^n$ centered at $q(r)$. Explicitly,*

$$w = q(r) + \sum_{k=1}^{d} c_k e_k(r) \quad , \quad \forall w \in H(r),$$

where $\{e_k(r)\}_{k=1}^{d}$ is an orthonormal basis of the linear sub-space $\mathcal{G}H(r)$. We say that the family $(q(r), H(r))$ changes smoothly with respect to r if for any vector $v \in \mathbb{R}^n$ the function

$$w(r) = q(r) + \sum_{k=1}^{d} \langle v - q(r), e_k(r) \rangle e_k(r),$$

describing the Euclidean projections of v onto $H(r)$, vary smoothly with respect to r.

Remark 2.12. Definition 3 can be pronounced as a smooth function from $\mathbb{R}^n$ to $Gr_d(\mathbb{R}^n)$, where $Gr_d(\mathbb{R}^n)$ is the d-dimensional Grassmanian of $\mathbb{R}^n$. However, we believe that using the explicit definition pronounced above is clearer.

Accordingly, a *smoothly varying coordinate system* would be a family of affine sub-spaces which vary smoothly with respect to our parameter r, such that our manifold can be viewed locally as a graph of a function over it.

Theorem 2.13 (Smoothness of $q(r), H(r)$)**.** *Let the Noisy Sampling Assumptions of Section 2.2 hold. Let $\theta(x) \in C^\infty$, H be a d-dimensional affine space around an origin q and let U_{unique} be the uniqueness domain of Assumption 2.10. Let $q(r), H(r)$ be the minimizers of the constrained minimization problem* (8) *Let $J^*(r;q)$ be the function described in Equation* (11)*. Then for all $r' \in U_{unique}$ such that $\left(\frac{\partial^2 J^*}{\partial q_i \partial q_j}\right)_{ij} \in M_{n\times n}$ is invertible at $(r', q(r'))$ we get:*

1. *$q(r)$ is a smooth (C^∞) function in a neighborhood of r'.*
2. *The affine space $H(r)$ changes smoothly (C^∞) in a neighborhood of r'.*

Remark 2.14. Note that $(r, q(r))$ is a local minimum of J^* as r belongs to the uniqueness domain. As a consequence, the function J^* is locally convex at that point. Thus, the condition that the Hessian of J^* at $(r, q(r))$ is invertible implies that at this minimum there is no direction in which the second derivative vanishes. So, when the condition is not met, there should exist a sectional curve of J^* that has vanishing first and second derivatives. When the data is sampled at random, this seems to be unlikely. In any case, this condition can be verified numerically and in all of our experiments this condition is met.

3 Extending The M-MLS to Function Approximation

In the following section we use the M-MLS framework to address the problem of regression over manifolds. Let $\mathcal{M}^d \subset \mathbb{R}^n$ be a d-dimensional manifold, and $\psi : \mathcal{M}^d \to \mathbb{R}^{\tilde{n}}$ is a function sampled with noise at noisy locations. For the sake of clarity and simplicity of notations, in what follows we assume that $\psi : \mathcal{M}^d \to \mathbb{R}$ (i.e., scalar valued function). The extension to the multidimensional case is immediate.

Noisy Function Sampling Assumptions

1. $\mathcal{M} \in C^2$ is a closed (i.e., compact and boundaryless) submanifold of $\mathbb{R}^n$.
2. ψ is a function from $\mathcal{M}$ to $\mathbb{R}$
3. $\tilde{R} = \{\tilde{r}_i\}_{i=1}^I \subset \mathcal{M}$ is an (unknown) h-ρ-δ sample set with respect to the domain $\mathcal{M}$ (see Definition 1).
4. $R = \{r_i\}_{i=1}^I$ are noisy samples of R; i.e., $r_i = \tilde{r}_i + n_i$.
5. $\|n_i\| < \sigma_{\mathcal{M}}$
6. $\psi_i = \psi(\tilde{r}_i) + \delta_i$
7. $\|\delta_i\| < \sigma_\psi$
8. Accordingly, the sample-set at hand is $R_\psi = \{(r_i, \psi_i)\}_{i=1}^N$.

Given some point r adjacent to $\mathcal{M}$ (i.e., $r = \tilde{r} + \varepsilon$, where $\tilde{r} \in \mathcal{M}$ and $\varepsilon \in \mathbb{R}^n$) we wish to approximate $\psi(\tilde{r})$. Below we suggest an approximation framework and algorithm for this case, based upon the M-MLS procedure described in the preliminaries. The main theoretical results of this paper are the smoothness and approximation properties as portrayed in Theorems 3.1, 3.2. Following this, we describe how one can use our proposed framework to produce interpolatory approximation. We conclude the section with a concise description of the algorithm.

3.1 Constructing The Function Approximation

For the purpose of discussion let us assume for a moment that our samples of $\mathcal{M}$ are without noise, that is $\tilde{R}_\psi = \{(\tilde{r}_i, \psi(\tilde{r}_i))\}_{i=1}^N$. Then, the most natural way to obtain an approximation to a differentaible function defined over a manifold is through approximating the function's pull back to local parametrizations. More precisely, for any $r \in \mathcal{M}$, given some coordinate chart (V, ϕ), where $V \subset \mathcal{M}$ is an open neighborhood of r and $\phi : V \to \mathbb{R}^d$, we would have liked to approximate the following function:

$$g \triangleq \psi \circ \phi^{-1} : \mathbb{R}^d \to \mathbb{R},$$

at $x \in \mathbb{R}^d$ such that $\phi^{-1}(x) = r \in \mathcal{M}$. This way, instead of trying to approximate a function from $\mathbb{R}^n$ to $\mathbb{R}$ we can approximate, on a local level, a function from $\mathbb{R}^d$ to $\mathbb{R}$. Since we assume $\mathcal{M}$ to be a smooth manifold, it can be viewed, locally, as a graph of a differentiable function $\eta : T_r\mathcal{M} \to T_r\mathcal{M}^\perp$, where $T_r\mathcal{M} \simeq \mathbb{R}^d$ is the tangent space of $\mathcal{M}$ at r and $T_r\mathcal{M}^\perp \simeq \mathbb{R}^{n-d}$ is its orthogonal compliment. This gives us a valid option to produce a chart around r through taking $\phi \triangleq \eta^{-1} = P_{T_r\mathcal{M}}$ the projection onto $T_r\mathcal{M}$. Then, if we had $T_r\mathcal{M}$ we could generalize the standard MLS (described in Section 2.1) to a manifold domain in a natural way to:

$$p_r = \underset{p \in \Pi_m^d}{\operatorname{argmin}} \sum_{i=1}^N \|p(x_i) - g_i\|^2 \, \theta(\|x_i - r\|),$$

where $x_i = \phi(\tilde{r}_i)$ are the projections of $\tilde{r}_i$ onto $\mathcal{G}T_r\mathcal{M}$ (i.e., $\phi(r) = 0 \in \mathbb{R}^d$), and $g_i = \psi(\tilde{r}_i)$. And the approximating value of $\psi(r)$ would be

$$\psi(r) \approx \widetilde{\psi}(r) \triangleq p_r(0).$$

Unfortunately, to obtain the exact tangent space we need to have access to the manifold's atlas, or at least have infinite sampling resolution. Moreover, in our problem-setting, the points $\{r_i\}_{i=1}^N$ as well as r are sampled with noise. Thus, $r \notin \mathcal{M}$ and it is meaningless to have a tangent space around it. Nevertheless, taking an in-depth look at the two-step approximation method of the M-MLS (described in Section 2.2), we can utilize its first step to produce an alternative moving coordinate system for the manifold, or in other words an atlas of charts. Explicitly, for any given r near $\mathcal{M}$ we can apply step 1 of the M-MLS procedure to obtain an approximating affine space $H(r)$ around an origin $q(r)$. As shown below in Theorem 3.1, in case of clean samples ($\tilde{r}_i = r_i, \delta_i = 0$ for all i), $d(r_i, H(r)) = \mathcal{O}(h^2)$, and so $\{r_i\}_{i=1}^N$ can be viewed as samples of a function η defined over $H(r)$ (according to Lemma 4.4 in [33] this would still be the case if the noise of $\|\sigma_\mathcal{M}\| = \mathcal{O}(h)$). That is,

$$\eta(x) : H(r) \to \mathcal{M},$$

and $H(r) \simeq \mathbb{R}^d$. As have been stated above, instead of approximating ψ directly we can aim at approximating

$$g \triangleq \psi \circ \eta : \mathbb{R}^d \to \mathbb{R}, \tag{12}$$

at $x \in \mathbb{R}^d$ such that $\eta(x) = r$. Similar to what have been stated above, the generalization of the MLS for function approximation is:

$$p_r = \underset{p \in \Pi_m^d}{\operatorname{argmin}} \sum_{i=1}^N \|p(x_i) - g_i\|^2 \, \theta(\|r_i - q(r)\|), \tag{13}$$

where x_i are the projections of r_i onto $\mathcal{G}H(r)$ shifted around zero (i.e., $q(r) \mapsto 0 \in \mathbb{R}^d$) and $g_i = \psi_i$. The approximating value of $\psi(r)$ would then be

$$\psi(r) \approx \widetilde{\psi}(r) \triangleq p_r(0). \tag{14}$$

As explained in detail in Lemma 4.4 in [33], under some mild assumptions $H(r)$ is a valid moving coordinate system as it approximates $T_{\tilde{r}}\mathcal{M}$, where $\tilde{r}$ is the projection of r onto $\mathcal{M}$. Furthermore, $H(r)$ possesses the desired property of varying smoothly with respect to r (see Definition 3). In any case, any valid choice of a smoothly varying coordinate system should suffice for the approximation defined in Equations (13) and(14) to be smooth as well.

Theorem 3.1 (Smoothness of $\widetilde{\psi}(r)$)**.** *Let the Noisy Function Sampling Assumption of Section 3 hold. Let $\theta(t)$ be a C^∞ radial weight function, and let $r \in U_{unique}$. Assume that the data is spread such that the minimization problem of* (8) *is well conditioned (i.e., the least-squares matrix is invertible). Let $q(r), H(r)$ be the minimizers of* (8)*, and $J^*(r;q)$ be the function described in* (11)*. Assume that $\left(\frac{\partial^2 J^*}{\partial q_i \partial q_j}\right)_{ij} \in M_{n\times n}$ is invertible at $(r, q(r))$ for all $r \in U_{unique}$. Then,*

1. *$\widetilde{\psi}(r)$ derived from equations* (13) *and* (14) *is a C^∞ function from $U_{unique} \subset \mathbb{R}^n$ to $\mathbb{R}$.*
2. *For any points $r_0, r_1 \in U_{unique}$ such that $r_1 - q(r_0) \perp H(r_0)$, we have*
$$\tilde{\psi}(r_1) = \tilde{\psi}(r_0).$$

Proof. We begin with looking at the least-squares problem with respect to a fixed coordinate system (q, H)

$$p_r = \underset{p\in\Pi_m^d}{\operatorname{argmin}} \sum_{i=1}^{N} (p(x_i) - \psi_i)^2 \theta(\|r_i - q\|), \tag{15}$$

where x_i are the projections of r_i onto $\mathcal{G}H$. Let $\mathcal{B} = \{b_j(x)\}_{j=1}^J$ and $J = \binom{m+d}{d}$ be a basis of Π_m^d. As shown in [10] and later simplified in [22], the least-squares problem of Equation (15) has an equivalent representation due to the Backus-Gilbert theory [7, 6]. Namely, minimize

$$Q = \sum_{i=1}^{N} \frac{1}{\theta(\|q - r_i\|)} a_i^2, \tag{16}$$

under the set of constraints

$$\sum_{i=1}^{N} a_i b_j(x_i) = b_j(0), \ \forall j \le J. \tag{17}$$

Then, the approximating value of $p_r(0)$ is given by

$$p_r(0) = \sum_{i=1}^{N} a_i \psi_i = \bar{\sigma}^T \bar{a}, \tag{18}$$

where $\bar{\sigma}^T = (\psi_1, ..., \psi_N)$ and $\bar{a} = (a_1, ..., a_N)^T$. Furthermore, using Lagrange multipliers, this problem can be presented as the following set of equations

$$\begin{pmatrix} D & E \\ E^T & 0 \end{pmatrix} \begin{pmatrix} \bar{a} \\ \bar{z} \end{pmatrix} = \begin{pmatrix} \bar{c} \\ 0 \end{pmatrix}, \tag{19}$$

where $D = 2diag\{1/\theta(\|r_1 - q\|), ..., 1/\theta(\|r_N - q\|)\}$, $\bar{c} = (b_1(x), ..., b_J(x))^T$, $E_{i,j} = b_j(x_i)$, and $\bar{z}$ are Lagrange coefficients. Since D is invertible we can write $\bar{a}$ explicitly as

$$\bar{a} = D^{-1}E(E^T D^{-1} E)^{-1}\bar{c}. \tag{20}$$

Now, as $\theta \in C^\infty$ it follows that, in the case of a fixed coordinate system, the minimizing polynomials p_r will vary smoothly with respect to r. This result is articulated in Theorem 2.1.

In our case, the coordinate system $(q(r), H(r))$ depends on the parameter r. Thus, we now obtain

$$D(r) = 2diag\{1/\theta(\|r_1 - q(r)\|), ..., 1/\theta(\|r_N - q(r)\|)\},$$

and

$$E(r)_{i,j} = b_j(x_i(r)).$$

From Theorem 2.13 we get that $(q(r), H(r))$ vary smoothly with respect to r, and so we achieve that $x_i(r)$ vary smoothly as well. Combining this with the fact that $\theta \in C^\infty$ we get that the right hand side of Equation (20) will still vary smoothly with respect to r. Therefore, our local polynomial approximation p_r changes smoothly with respect to r and so does our MLS approximation given by:

$$\widetilde{\psi}(r) \triangleq p_r(0),$$

and statement 1 is proven. In addition by Lemma 2.11 we have that for all $r_0, r' \in U_{unique}$ such that $r' - q(r_0) \perp H(r_0)$ the values $q(r') = q(r_0)$ as well as $H(r') = H(r_0)$. Thus, we achieve in our case that

$$\widetilde{\psi}(r') = \widetilde{\psi}(r_0),$$

as required in 2. □

We wish to discuss further the assumption regarding Hessian of $J^*(r; q)$. Note that $(r, q^*(r))$ is a local minimum of J^* as r belongs to the uniqueness domain. As a consequence, the function J^* is locally convex at that point. Thus, the condition that the Hessian of J^* at $(r, q^*(r))$ is invertible implies that at this minimum there is no direction for which the second derivative vanishes. So, when the condition is not met, there should exist a sectional curve of the graph of J^* that has vanishing first and second derivatives. When the data is sampled at random, this seems to be unlikely. In any case, this condition can be verified numerically and in all of our experiments this condition is met.

After obtaining the smoothness property, we wish to show that the Moving Least-Squares approximation descried here obtains near optimal convergence rates with respect to the maximum norm on the manifold domain $\mathcal{M}$

$$\left\|\widetilde{\psi}(r) - \psi(r)\right\|_{\mathcal{M},\infty} \triangleq \max_{r\in\mathcal{M}} \left|\widetilde{\psi}(r) - \psi(r)\right|.$$

Explicitly, we show that $\left\|\widetilde{\psi}(r) - \psi(r)\right\|_{\mathcal{M},\infty} = \mathcal{O}(h^{m+1})$. To achieve such a result we need to restrict the behaviour of our weight function θ by the following conditions.

Conditions on (8) for approximation order

1. The function $\theta(t)$ is monotonically decaying and compactly supported with $supp(\theta) = c_1 h$, where c_1 is some constant greater than 2.

2. Suppose that $\theta(c_2 h) > c_3 > 0$, for some constants $c_2 \geq 3$ and $c_3 > 0$.

3. Set $\mu = rch(\mathcal{M})/2$ in constraint 2.

4. Let r be such that $d(r, \mathcal{M}) < rch(\mathcal{M})/4$

We note that under these conditions in case of clean samples Lemma 4.3 of [33] shows that the minimization problem (8) is well conditioned (i.e., there are enough data sites such that the least-squares problem can be solved). Furthermore, Lemma 4.4 of [33] shows that

$$(q(r), H(r)) \to (P(r), T_{P(r)}\mathcal{M})$$

as $h \to 0$, where $P(r)$ is the orthogonal projection of r onto $\mathcal{M}$.

Theorem 3.2 (Approximation order for clean samples)**.** *Let $\mathcal{M}$ be a C^{m+1} smooth submanifold of $\mathbb{R}^n$, let the Noisy Sampling Assumption of Section 3 hold with $\sigma_{\mathcal{M}} = 0$ and $\sigma_\psi = 0$. Let $r \in \mathcal{M}$ and assume that $\psi : \mathcal{M} \to \mathbb{R}$ and $\psi \in C^{m+1}(\mathcal{M})$. Let $(q(r), H(r))$ be the coordinate system resulting from the minimization of* (8)*. Then, for fixed ρ and δ, the approximant given by equations* (13)-(14)*, yields the following error bound, for any h small enough:*

$$\left\| \widetilde{\psi}(r) - \psi(r) \right\|_{\mathcal{M},\infty} < M \cdot h^{m+1} \tag{21}$$

Proof. The outline of the proof is as follows:

1. We show that in a neighborhood of $q(r)$

$$d(r_i, H(r)) = \mathcal{O}(h^2) \tag{22}$$

2. Using (22) we get that for small enough h the data projected onto $H(r)$ is still a $\tilde{h}$-$\tilde{\rho}$-$\tilde{\delta}$ set (for $\tilde{h} = \mathcal{O}(h)$, $\tilde{\rho} \approx \rho$, and $\tilde{\delta} \approx \delta$) in an $\mathcal{O}(h)$ neighborhood of $q(r)$.

3. Using a known bound regarding weighted least-squares polynomial approximation (Theorem 4 of [22]), we show that for any $r \in \mathcal{M}$ there exist $M(r)$ (independent of h) such that
$$\left| \widetilde{\psi}(r) - \psi(r) \right| < M(r) \cdot h^{m+1}.$$

4. We show that $M(r)$ is bounded by a constant M for all $r \in \mathcal{M}$ and get
$$\left| \widetilde{\psi}(r) - \psi(r) \right| < M \cdot h^{m+1}.$$

We first notice that $q = P(r)$, the projection of r onto $\mathcal{M}$, coupled with $H = T_{P(r)}\mathcal{M}$ maintain constraints 1-3 of Equation (8). Since the projection onto $\mathcal{M}$ keeps the condition

$$r - P(r) \perp T_{P(r)}\mathcal{M},$$

constraint 1 is met. By the fact that $P(r) = r \in \mathcal{M}$ and $\mu > 0$ we get that constraint 2 is met. In addition, since h is the fill distance, $\sigma_{\mathcal{M}} = 0$ and $supp(\theta) = ch$, there exists some $r_j \in R \subset \mathcal{M}$ such that $\|r_j - P(r)\| < h$. Therefore,

$$\#R \cap B_h(P(r)) \neq 0,$$

and constraint 3 is met as well.

Furthermore, since the tangent space is a first order approximation of a manifold $\mathcal{M} \in C^2$, the cost function is compactly supported, and the sampling is an h-ρ-δ set (see the definition of ρ in (4)), then for all $x \in \mathcal{M}$ (including $P(r)$) we have

$$J(r; x, T_x\mathcal{M}) = \sum_{i=1}^{N} d^2(r_i, T_x\mathcal{M})\theta(\|r_i - x\|) = \mathcal{O}(h^4),$$

and so

$$J(r; x, T_x\mathcal{M}) = \mathcal{O}(h^4), \text{ as } h \to 0. \tag{23}$$

Thus,

$$J(r; q(r), H(r)) = \sum_{i=1}^{N} d^2(r_i, H(r))\theta(\|r_i - q\|) = \mathcal{O}(h^4), \tag{24}$$

and since $\theta(c_2 h) > c_3 > 0$ as well as monotonically decreasing, we get that for $r_i \in B_{c_2 h}(q(r))$

$$d(r_i, H(r)) = \left\|P_{H(r)}(r_i) - r_i\right\| = \mathcal{O}(h^2). \tag{25}$$

Hence, we showed that (22) holds, and the sample set projected onto $H(r)$ is an $\tilde{h}$-$\tilde{\rho}$-$\tilde{\delta}$, where $\tilde{h} = \mathcal{O}(h)$. Furthermore, since $H(r) \to T_{P(r)}\mathcal{M}$ as $h \to 0$, for a small enough h we get that $\tilde{\rho} \approx \rho$ and $\tilde{\delta} \approx \delta$.

Next, we look at the approximated object, locally, as a function $g_r : H(r) \simeq \mathbb{R}^d \to \mathbb{R}$. Explicitly, let $V_r \subset \mathcal{M}$ be a neighborhood of r and let $\phi_r : V_r \to H(r) \simeq \mathbb{R}^d$ be our chart (i.e., the orthogonal projection onto $H(r)$). Then,

$$g_r \triangleq \psi \circ \phi_r^{-1}. \tag{26}$$

The minimization problem of (13)-(14) is a weighted least-squares polynomial approximation of degree $\leq m$ around a point $x \in \mathbb{R}^d$, with respect to the sample set $\{(x_i, g_r(x_i))\}_{i=1}^N$. We wish to give a bound for such local approximation for any $r \in \mathcal{M}$; i.e., show that statement 3 of the outline holds. To meet this goal, we wish to produce a bound similar in spirit to the one given in Theorem 2.4. Accordingly, we first verify that $g_r \in C^{m+1}$. Then, we utilize a general bound for weighted least-squares to find $M(r)$ such that

$$\left|\widetilde{\psi}(r) - \psi(r)\right| < M(r) \cdot h^{m+1}.$$

Since $\mathcal{M} \in C^{m+1}$, around any point r, $\mathcal{M}$ is locally a graph of a C^{m+1} function $\varphi : H(r) \to H(r)^\perp$. Therefore, ϕ_r^{-1} can be written in the coordinate system $H(r) \times H(r)^\perp$ as

$$\phi_r^{-1}(x) = (x, \varphi(x)) \ , \ \forall x \in \phi_r[V_r],$$

and thus $\phi_r^{-1}(x) \in C^{m+1}$. Since $\psi \in C^{m+1}$, and from (26), we have $g_r \in C^{m+1}$ as required.

Using the smoothness of g_r we can utilize a known bound for weighted least-squares. Without loss of generality, let $\phi_r(r) = 0$. Then, according to Theorem 4 from [22] we get that

$$|g_r(0) - \widetilde{g}_r(0)| \leq \left(1 + \sum_{i=1}^{N} |a_i^r|\right) E_{B_{ch}(0), \Pi_m^d}(g_r), \tag{27}$$

where a_i^r are the coefficients defined in (16),

$$E_{B_{ch}(0), \Pi_m^d}(g_r) \triangleq \inf_{p \in \Pi_m^d} \|g_r - p\|_{B_{ch}(0), \infty}$$

and $\|g_r - p\|_{B_{ch}(0),\infty}$ is the restriction of the infinity norm to the domain $B_{ch}(0)$. We wish to note that the term $\left(1 + \sum_{i=1}^{N} |a_i^r|\right)$ is independent of h. Due to the definition of the coefficients a_i^r in Equations (16)-(17), if θ is consistent across scales (i.e., scaling θ with h such that $\theta(th) = f(t)$), then the weights are scale invariant. Furthermore, if, for example, we take $\mathcal{B}$ to be the standard basis of Π_m^d (i.e., $\{1, x_1, ..., x_n, x_1^2, x_1x_2, ...\}$) then (17) are met regardless of the scale chosen (i.e., replacing x_i with κx_i will not change the fact that the equations hold). Thus, the coefficients for the local weighted least-squares problem are scale invariant. Since the problem (17) is independent of the basis choice for Π_m^d, we have that a_i^r are independent of the scale.

Taking the Taylor expansion as a possible polynomial approximation we get that

$$E_{B_{ch}(0),\Pi_m^d}(g_r) \leq R_m[g_r],$$

where $R_m[g_r]$ is the Taylor remainder of order m. That is,

$$R_m[g_r] = \sum_{|\alpha|=m+1} \partial^\alpha g_r(\xi) \cdot \frac{(ch)^{m+1}}{(m+1)!}, \tag{28}$$

where $\xi \in B_{ch}(0)$, α is a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$, $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and

$$\partial^\alpha g_r(x) = \frac{\partial^{\alpha_1} \cdots \partial^{\alpha_d}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} g_r(x). \tag{29}$$

Therefore, plugging (28) back into (27) we get

$$|g_r(0) - \widetilde{g}_r(0)| \leq M(r)h^{m+1}, \tag{30}$$

where

$$M(r) \triangleq \max_{x \in \bar{B}_{ch}(0)} \left(1 + \sum_{i=1}^{N} |a_i^r|\right) \frac{c^{m+1}}{(m+1)!} \sum_{|\alpha|=m+1} \partial^\alpha g_r(x) \tag{31}$$

Since we wish to bound $M(r)$ for all r, we note that

$$\sup_{r \in \mathcal{M}} M(r) = \sup_{r \in \mathcal{M}} \max_{x \in \bar{B}_{ch}(0)} \left(1 + \sum_{i=1}^{N} |a_i^r|\right) \frac{c^{m+1}}{(m+1)!} \sum_{|\alpha|=m+1} \partial^\alpha g_r(x), \tag{32}$$

and so

$$\begin{aligned} \sup_{r \in \mathcal{M}} M(r) &\leq \sup_{r \in \mathcal{M}, x \in \bar{B}_{ch}(0)} \left(1 + \sum_{i=1}^{N} |a_i^r|\right) \tfrac{c^{m+1}}{(m+1)!} \textstyle\sum_{|\alpha|=m+1} \partial^\alpha g_r(x) \\ &= \sup_{(r,x) \in \mathcal{M} \times \bar{B}_{ch}(0)} M(r,x) \end{aligned}, \tag{33}$$

Note that since $g_r(x)$ is continuous with respect to r so is $\partial^\alpha g_r(x)$. Furthermore, a_i^r are continuous in r as was mentioned in the proof of Theorem 3.1 due to their representation in (20). In addition, since $g_r(x)$ is in C^{m+1} with respect to x then $\partial^\alpha g_r(x)$ is smooth for $|\alpha| = m+1$. As a result, $M(r,x)$ is continuous and since the domain $\mathcal{M} \times \bar{B}_{ch}(0)$ is compact we get that its supremum is achieved and

$$|g_r(0) - \widetilde{g}_r(0)| \leq M(r)h^{m+1} \leq Mh^{m+1}, \tag{34}$$

where

$$M \triangleq \max_{(r,x) \in \mathcal{M} \times \bar{B}_{ch}(0)} M(r,x).$$

□

3.2 Interpolation Scheme

The above-mentioned mechanism can be used to provide an interpolatory function approximation. As mentioned above, in Section 2.1, when dealing with function approximation over a flat domain, if we wish the approximation to become interpolatory at the samples $\{r_i\}_{i=1}^N$ all we need is to choose a weight function θ such that $\theta(0) = \infty$. Evidently, the fact that $\lim_{t\to 0}\theta(t) = \infty$ forces the moving least-squares approximation $\widetilde{\psi}$ to reach $\lim_{r\to r_i}\widetilde{\psi}(r) = \psi(r_i)$. The next proposition shows the analog of this to the manifold case.

Proposition 3.3. *(Interpolation) Let Noisy Sampling Assumption of Section 3 hold with $\sigma_{\mathcal{M}} = 0$ and $\sigma_\psi = 0$. Let $\theta \in C^\infty(\mathbb{R}\setminus\{0\})$ be a decaying weight function satisfying $\lim_{t\to 0}\theta(t) = \infty$, and let $r \in U_{unique}$. We define the approximation of ψ at r as $\widetilde{\psi}(r)$ of Equation (14) for all $r \notin R$ (our sample set), and for $r = r_i \in R$ we define $\widetilde{\psi}(r_i) = \psi(r_i)$. Then, $\widetilde{\psi}$ is a smooth function interpolating our sample set.*

Proof. Let x_i be the projection of r_i onto the coordinate system $(q(r), H(r))$ solving the minimization of (8), then the condition $\lim_{t\to 0}\theta(t) = \infty$ enforces the local polynomial approximation of Equation (13) p_r to satisfy $\lim_{r\to r_j} p_r(0) = \psi(r_j)$. It is clear that the approximation is smooth around each $r \notin R$, as this is explained in Theorem 3.1 above. Assume that $r = r_i \in R$ for some specific i. Then, revisiting Equation (19), we can see that D is not invertible and the formula achieved for $\bar{a}$ does not hold. Nevertheless, the matrix

$$\begin{pmatrix} D & E \\ E^T & 0 \end{pmatrix}$$

is smooth and invertible, and by the Inverse Function Theorem we get that its inverse must be smooth as well. Thus, the coefficients vector $\bar{a}$ will be smooth still, and accordingly, so will the MLS approximation $\widetilde{\psi}(r)$. □

Remark 3.4. The extension of the interpolatory scheme to the multidimensional case is immediate.

3.3 Algorithm Description

As a result of the theoretical discussion, our procedure will go along the lines of the M-MLS procedure described above. For the sake of generality, we refer to the multidimensional case where $\psi : \mathcal{M}^d \to \mathbb{R}^{\tilde{n}}$. The two major steps of the procedure are as follows,

1. Find a local d-dimensional affine space $H(r)$ approximating the sampled points ($H \simeq \mathbb{R}^d$). This affine space will be used in the following step as a local coordinate system.

2. Approximate the function $g : H(r) \to \mathbb{R}^{\tilde{n}}$ through weighted least-squares, based upon the samples $\{(x_i, g_i)\}_{i=1}^N$, where x_i are the projections of r_i onto $H(r)$ and $g_i = \psi_i$.

In what follows, we discuss in more details both steps as well as their implementation. The implementation of step 2 is trivial as it is a standard least-squares problem. Thus, after explaining it, we give a short description of it in Algorithm 2. However, the implementation of step 1 is more complicated and will be discussed below in more details with a concise summary in Algorithm 1.

Step 1 - Finding The Local Coordinates

Find a d-dimensional affine space H, and a point q on H, such that the following constrained problem is minimized:

$$\begin{aligned} & J(r;q,H) = \sum_{i=1}^{N} d(r_i,H)^2\theta(\|r_i-q\|) \\ s.t. & \\ & r-q \perp H \quad i.e., \quad r-q \in H^{\perp} \end{aligned} \quad , \tag{35}$$

We find the affine space H by an iterative procedure. Assuming we have q_j and H_j at the jth iteration, we compute H_{j+1} by performing a linear approximation over the coordinate system H_j. In view of the constraint $r-q \perp H$, we define q_{j+1} as the orthogonal projection of r onto H_{j+1}. We initiate the process by taking $q_0 = r$ and choose d basis vectors $\{u_k^1\}_{k=1}^d$ randomly. This first approximation is denoted by H_1. Thence, we compute:

$$q_1 = \sum_{k=1}^{d} \langle r-q_0, u_k^1\rangle u_k^1 + q_0 = q_0.$$

Upon obtaining q_1, H_1 we continue with the iterative procedure as follows:

- Assuming we have H_j, q_j and its respective basis $\{u_k^j\}_{k=1}^d$ w.r.t the origin q_j, we project our data points r_i onto H_j and denote the projections by x_i. Then, we find a linear approximation of the samples $f_i^j = f^j(x_i) = r_i$:

$$\vec{l^j}(x) = \underset{\substack{\vec{p}=(p_1,..,p_n),\\ p_k\in\Pi_1^d}}{\operatorname{argmin}} \sum_{i=1}^{N} \|\vec{p}(x_i) - f_i^j\|^2\theta(\|r_i-q_j\|). \tag{36}$$

  Note, that this is a standard weighted linear least-squares as q_j is fixed! Thus, it involves a single inversion of a $(d+1)\times(d+1)$ dimensional matrix (applied n times for each coordinate in the ambient space), and the computational complexity $\mathcal{O}(nd^2)$.

- Given $\vec{l^j}(x)$ we obtain a temporary origin:

$$\widetilde{q}_{j+1} = \vec{l^j}(0).$$

  Then, around this temporary origin we build a basis $\hat{B} = \{v_k^{j+1}\}_{k=1}^d$ for H_{j+1} with:

$$v_k^{j+1} \triangleq \vec{l^j}(u_k^j) - \widetilde{q}_{j+1} \quad , \quad k = 1,...,d$$

  We then use the basis $\hat{B}$ in order to create an orthonormal basis $B = \{u_k^{j+1}\}_{k=1}^d$ through a Grham-Schmidt process, which costs $\mathcal{O}(nd^2)$ flops. Finally we derive

$$q_{j+1} = \sum_{k=1}^{d} \langle r-\widetilde{q}_{j+1}, u_k^{j+1}\rangle u_k^{j+1} + \widetilde{q}_{j+1}.$$

  This way we ensure that $r-q_{j+1} \perp H_{j+1}$. The complexity of computing this projection is $\mathcal{O}(nd)$.

Remark 3.5. In [33], the initialization of the basis vectors of H_1 is based upon a local PCA of the data points. In theory, this ought to improve the number of iterations needed for convergence, at the expense of significantly increasing the computational time. Alternatively, we could have utilized a low-rank approximation of the PCA such as proposed in [1] to do this more efficiently. Nevertheless, we found that in practice a random initialization requires a similar number of iterations, and thus reduces the computation time even further. For this reason, we used a random initialization of H_1.

See Figure 4 for the approximated local coordinate systems H obtained by Step 1 on noisy samples of a sphere.

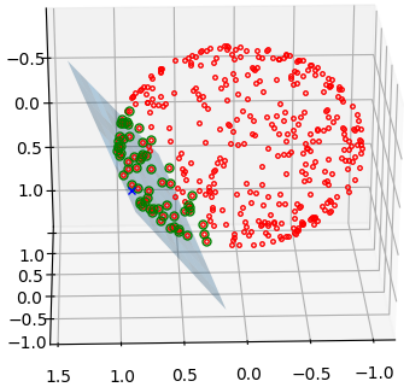

Figure 4: An approximation of the local coordinates $H(r)$ (blue plain) resulting from Step 1 implementation after three iterations, where r is marked by the blue $\times$. Marked in green are the points affecting the approximation.

Step 2 - The approximation of ψ

Let $\{e_k\}_{k=1}^d$ be an orthonormal basis of $H(r)$ (taking $q(r)$ as the origin), and let x_i be the orthogonal projections of r_i onto $H(r)$ (i.e., $x_i = q(r) + \sum_{k=1}^d \langle r_i - q(r), e_k \rangle e_k$). As before, we note that r is orthogonally projected to the origin q. Now we would like to approximate $g : \mathbb{R}^d \to \mathbb{R}^{\widetilde{n}}$, such that $g_i = g(x_i) = \psi_i$. The vector-valued approximation of g is performed by minimizing the weighted least-squares cost function, using a polynomial $\vec{p}_r(x) = (p_r^1(x), ..., p_r^{\widetilde{n}}(x))^T$ where $p_r^k(x) \in \Pi_m^d$, for $1 \leq k \leq \widetilde{n}$.

$$\vec{p}_r(x) = \underset{\substack{p^k \in \Pi_m^d \\ 1 \leq k \leq \widetilde{n}}}{\operatorname{argmin}} \sum_{i=1}^{N} \|\vec{p}(x_i) - \psi_i\|^2 \, \theta(\|r_i - q\|). \tag{37}$$

The approximation $\widetilde{\psi}(r)$ is then defined as:

$$\widetilde{\psi}(r) = \vec{p}_r(0) \tag{38}$$

Remark 3.6. The weighted least-squares approximation is invariant to the choice of an orthonormal basis of $\mathbb{R}^d$.

Remark 3.7. The requirement $r - q(r) \perp H(r)$ along with Assumption 2.10 implies that $\widetilde{\psi}(r)$ would be the same for all r such that $r - q(r) \perp H(r)$ and $r \in U_{unique}$ the uniqueness domain.

Remark 3.8. To save computation time, note that the normal equations of (37) are the same for all $\widetilde{n}$ coordinates just with a different right hand side. Thus, the least-squares matrix should be inverted only once.

Algorithm 1 Finding The Local Coordinate System $(H(r), q(r))$

Input: $\{r_i\}_{i=1}^N, r, \epsilon$
Output: q - an n dimensional vector ▷ $H = q + Span\{u_j\}_{j=1}^d$
U - an $n \times d$ matrix whose columns are $\{u_j\}_{j=1}^d$
Define R to be an $n \times N$ matrix whose columns are r_i
Initialize U with the first d principal components of the spatially weighted PCA
$q \leftarrow r$
repeat
  $q_{prev} = q$
  $\tilde{R} = R - repmat(q, 1, N)$
  $\tilde{R} = \tilde{R} \cdot \Theta$ ▷ Where $\Theta = diag(\sqrt{\theta(\|r_1 - q\|)}, \ldots, \sqrt{\theta(\|r_N - q\|)})$
  $X_{N\times d} = \tilde{R}^T U$ ▷ Find the representation of r_i in $Col(U)$
  Define $\tilde{X}_{N\times(d+1)} = \left[(1, ..., 1)^T, X\right]$
  Solve $\tilde{X}^T \tilde{X} \alpha = \tilde{X}^T \tilde{R}^T$ for $\alpha \in M_{(d+1)\times n}$ ▷ Solving the LS minimization of $\tilde{X}\alpha \approx \tilde{R}^T$
  $\tilde{q} = q + \alpha(1, :)^T$
  $Q, \hat{R} = qr(\alpha(2 : end, :)^T - repmat(\tilde{q}, 1, d))$ ▷ Where qr denotes the QR decomposition
  $U \leftarrow Q$
  $q = \tilde{q} + UU^T(r - \tilde{q})$
until $\|q - q_{\text{prev}}\| < \epsilon$

Algorithm 2 Function Approximation

Input: $\{(r_i, \psi_i)\}_{i=1}^N, r$
Output: $\widetilde{\psi}(r)$
Build a coordinate system H around r using $\{r_i\}_{i=1}^N$ (e.g., via Algorithm 1)
Project each $r_i \in \mathbb{R}^n$ onto $H \to x_i \in \mathbb{R}^d$
$\widetilde{\psi}(r)$ is the solution of the weighted least-squares problem using the samples $\{(x_i, \psi_i)\}_{i=1}^N$ around r

4 Numerical Examples

Generally, it is desirable for an algorithm to have a few, but not too many, parameters for tuning purposes. In our case, in order to fine tune the application of the algorithm, one needs to decide how to set the weight function θ of equations (35)-(37). In all of the examples bellow we have chosen to use the single parametric family of weight functions. For a given choice of the parameter k we define

$$\theta_k(t) \triangleq exp\left(\frac{-t^2}{(t - kh)^2}\right) \cdot \chi_{kh},$$

where χ_{kh} is an indicator function of the interval $[-kh, kh]$. This function is C^∞ and compactly supported. The minimal requirement for the support size is such that the local least-squares

matrix would be invertible. We chose k such that the support would contain about 3 times the minimal required amount of points. Intuitively, increasing the number of points or the support of θ will make the procedure more robust to noise, but, on the other hand, will add bias to the result.

4.1 A Function over a Helix

The approximant yielded by our algorithm is defined over a neighborhood of the sampled manifold. To show this numerically we have sampled a function over the helix

$$\begin{aligned} x &= \sin(t) \\ y &= \cos(t) \ , \\ z &= t \end{aligned}$$

for $t \in [-2\pi, 2\pi]$, and the function used is $\psi(x, y, z) = z$. We have added Gaussian noise to both domain $\mathcal{N}(\sigma_{domain}, 0)$ and target $\mathcal{N}(\sigma_{target}, 0)$. Figure 5 shows a case where the original data was sampled with $\sigma_{domain} = 0, \sigma_{target} = 6.25$ (Figure 5a). Then the evaluation is done for points sampled with $\sigma_{domain} = \sqrt{8 + z^2}$ (i.e., the noise varies w.r.t to the z value; Figure 5b and c). As can be seen, the noise is smoothed out in $\widetilde{\psi}$.

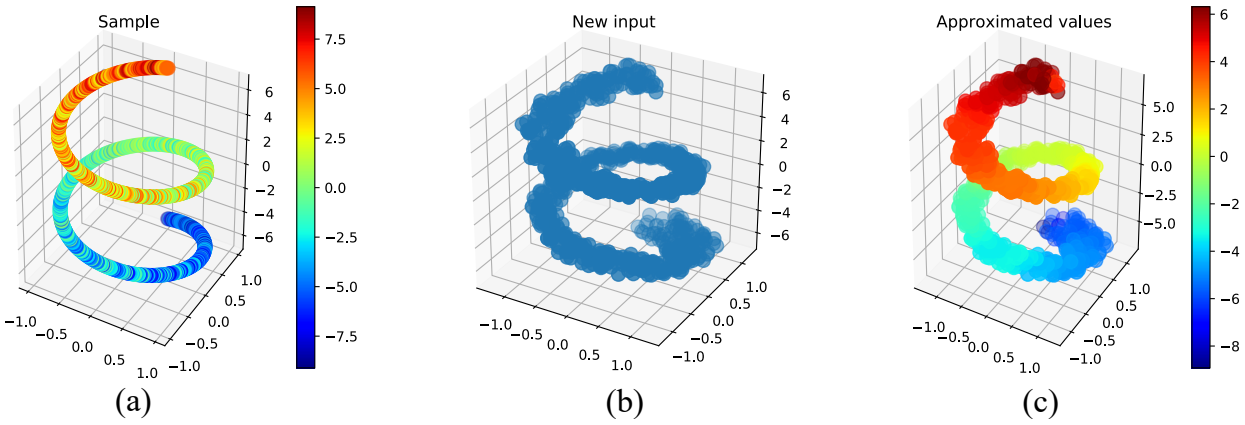

Figure 5: Approximation of the height function $f(x, y, z) = z$ defined over points on a helix: (a) a noisy sample; (b) new input; (c) approximation of f on the new input.

Figure 6 shows a case where the original data was sampled with $\sigma_{domain} = \sqrt{8 + z^2}, \sigma_{target} = 6.25$ (Figure 6b). Then the evaluation is done for the original samples locations in the domain. Figure 6c shows both the approximated projection onto the approximating manifold along with the value of $\widetilde{\psi}$. sampled with $\sigma_{domain} = \sqrt{8 + z^2}$ (i.e., the noise varies w.r.t to the z value; Figure 5b and c). As can be seen, the geometry as well as the behaviour of ψ are maintained in the approximant $\widetilde{\psi}$.

4.2 Approximation Order

In Theorem 3.2 we show that, given clean h-ρ-δ sample sets (for fixed ρ and δ), our function approximation scheme yields an approximation order of $\mathcal{O}(h^{m+1})$, where m is the total degree of the local polynomial. Denote the error of approximation of a point using a h_i-ρ-δ set by err_{h_i}. In this experiment, we show numerically that

$$err_h \approx M h^{m+1},$$

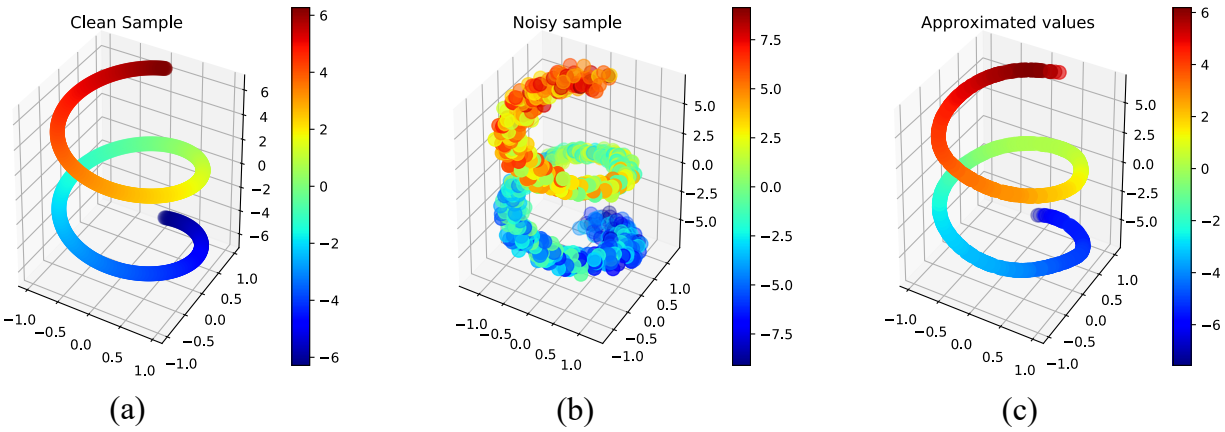

Figure 6: Approximation of the height function $f(x, y, z) = z$ sampled over points near a helix: (a) a clean helix with f as a color map; (b) the given sample set; (c) approximation of f evaluated on the points in the sample set, as well as the projection of the points (x, y, z) onto the approximating manifold $\mathcal{M}$ using the Manifold-MLS.

or, in other words, for h_1 and h_2:

$$\log \frac{err_{h_1}}{err_{h_2}} \approx (m+1) \log \frac{h_1}{h_2}. \tag{39}$$

In order to show that Equation (39) holds, we take N points on the unit sphere S^2 chosen on an equispaced grid in the spherical coordinate system (excluding the r coordinate), for $N = 20^2, 30^2, \ldots, 80^2$. Samples from this distribution is a good-enough approximation for h-ρ-δ sets with fixed ρ, δ parameters. The function that we approximate, $\psi : \mathbb{R}^3 \to \mathbb{R}^2$, match any point on the sphere with its spherical coordinates $(\phi, \theta) \in [0, 2\pi) \times [0, \pi)$. For any pair $\{N_i, N_j\} \subset \{20^2, 30^2, \ldots, 80^2\}$ we estimate h_k by $1/\sqrt{N_k}$. Then , in order to estimate the slope, we perform a least-squares linear fit using the points

$$\left(\log \frac{h_i}{h_j}, \log \frac{err_{h_i}}{err_{h_j}} \right).$$

In Figure 7 the small blue dots represent $\left(\log \frac{h_i}{h_j}, \log \frac{err_{h_i}}{err_{h_j}} \right)$ for an approximation using a first degree polynomial ($m = 1$), and similarly, the larger green dots correspond to $m = 3$. The dashed line and the full line are the linear fits for the $m = 1$ and $m = 3$ data points respectively. The slopes of the lines are 1.933 and 4.081, which is similar to $m + 1$ in both cases.

4.3 Large dimensional ambient space

The dataset in this experiment included a set of 72 gray-scale images of size 448×416 pixels. The images are taken from the unprocessed dataset of [28]. They are 2D projections of a 3D piggy bank obtained through rotating the object by 72 equispaced angles on a single axis. An example of the images is given in Figure 8. The approximated function ψ is the angle of rotation. Therefore our dataset consists of 72 samples of a 1-dimensional manifold embedded in $\mathbb{R}^{186368}$ ($186368 = 448 \times 416$) along with scalar values representing the angle of rotation.

In order to assess the presented algorithm, we used the leave-one-out cross-validation scheme. In each iteration, one image, chosen at random, is taken out of the dataset, and its angle is estimated using the angles of the other images.

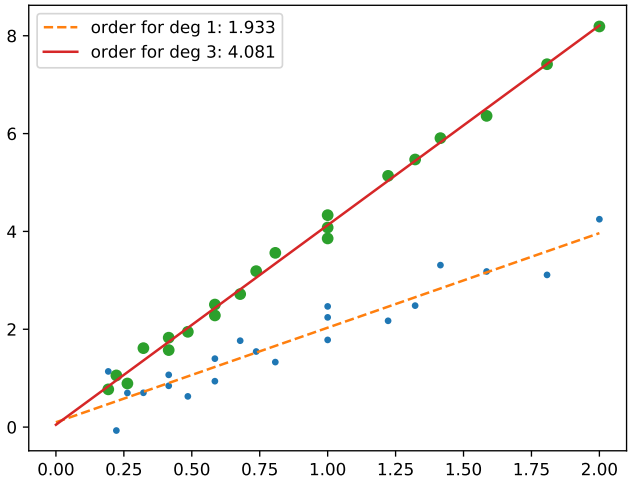

Figure 7: Estimating the approximation order for $m = 1, 3$. For any pair $\{N_i, N_j\} \subset \{20^2, 30^2, \ldots, 80^2\}$, we plotted the points $\left(\log \frac{h_i}{h_j}, \log \frac{err_{h_i}}{err_{h_j}}\right)$. The case of $m = 1$ is represented by the small blue dots and $m = 3$ is represented by the larger green dots. The slope of the linear fit, to each set of points, gives an estimate to the approximation order, which is nearly $m + 1$ in both cases.

Using $m = 1$, with 50 experiments, the average error is 0.0066 and the variance is $5.7 \cdot 10^{-5}$. However, when using $m = 3$ the average error is 0.06 and the variance is 0.02. This decrease in accuracy can be explained by the fact that higher order approximations require more data points, which means that the locality of the approximation is compromised.

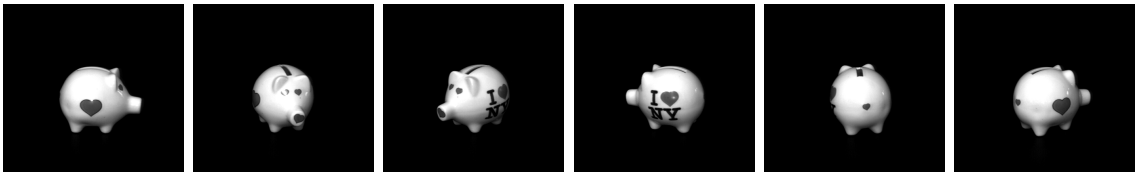

Figure 8: A part of the dataset of [28], consisting of 72 images of a rotating piggy bank. Each image is of size 448×416 pixels. Thus, we approximate a function $\psi : \mathbb{R}^{448\times416} \to [0, 2\pi)$, that returns the rotation angle for each given image.

4.4 Regression Over a Klein Bottle

In the following example we compare the algorithm presented here to the algorithms presented in [5, 12]. The setting is taken from section 5.1 in [12]. Let $\mathcal{M}$ be the Klein bottle, a two-dimensional closed and smooth manifold, embedded in $\mathbb{R}^4$, which is parametrized by $\phi_{Klein} : [0, 2\pi) \times [0, 2\pi) \to \mathbb{R}^4$ as

$$(u, v) \to ((2\cos v + 1)\cos u, (2\cos v + 1)\sin u, 2\sin v\cos(u/2), 2\sin v\sin(u/2)).$$

We sample n points (u_i, v_i) uniformly from $[0, 2\pi) \times [0, 2\pi)$ and obtain the corresponding points $p_i = \phi_{Klein}(u_i, v_i)$. Our sampled points are based on p_i with added noise. Explicitly, $r_i = p_i + \sigma_r \eta$ where η is a four dimensional normal random variable with zero-mean and identity covariance matrix, and σ_r is a parameter that changes in the experiment.

The function ψ that we approximate is defined as

$$\psi(p) = 7\sin(4u) + 5\cos^2(2v) + 6e^{-32((u-\pi)^2+(v-\pi)^2)},$$

Alg	$\sigma_r = 0$			
	$n = 1500$		$n = 1000$	
	snrdb = 5	snrdb = 2	snrdb = 5	snrdb = 2
Best performance from [12] (MALLER)	2.36 ± 0.68	2.64 ± 0.69	2.69 ± 0.69	2.94 ± 0.71
M-MLS ($m = 1$)	1.51 ± 0.34	1.21 ± 0.30	1.53 ± 0.41	1.77 ± 0.43
M-MLS ($m = 3$)	1.41 ± 0.35	1.05 ± 0.28	1.23 ± 0.35	1.56 ± 0.42
M-MLS ($m = 5$)	1.51 ± 0.37	1.07 ± 0.27	1.27 ± 0.33	1.73 ± 0.42

Alg	$\sigma_r = 0.2$			
	$n = 1500$		$n = 1000$	
	snrdb = 5	snrdb = 2	snrdb = 5	snrdb = 2
Best performance from [12] (MALLER)	3.85 ± 0.77	3.86 ± 0.77	4.00 ± 0.71	4.16 ± 0.78
M-MLS ($m = 1$)	3.11 ± 0.82	2.87 ± 0.75	3.02 ± 0.72	3.08 ± 0.78
M-MLS ($m = 3$)	2.97 ± 0.72	2.76 ± 0.78	2.88 ± 0.75	3.05 ± 0.88
M-MLS ($m = 5$)	2.87 ± 0.70	2.62 ± 0.61	2.95 ± 0.81	3.21 ± 0.79

Table 1: Accuracy of approximation - compared with the best performing algorithm out of 8 tested algorithms in [12]

where $(u, v) = \phi^{-1}_{Klein}(p)$. The samples that we have of ψ are $\psi_i = \psi(p_i) + \sigma(p)\epsilon$ where $\epsilon \sim \mathcal{N}(0, 1)$ and $\sigma(p) = \sigma_0(1 + 0.1\cos(u) + 0.1\sin(v))$, where σ_0 determines the signal-to-noise ratio, defined by:

$$\text{snrdb} = 10\log_{10}\left(\frac{\text{var } \psi(\{p_i | i = 1 \ldots n\})}{\sigma_0^2}\right)$$

We follow the same eight experiments done in [12] and compare the results to [5, 12]. The experiments parameters are: $n = 1000$ or 1500, snrdb = 5 or 2, and $\sigma_r = 0$ or 0.2, utilizing all the combinations. As reported in [12] the MALLER method yielded significantly better results than all the other tested algorithms. Thus, we show here a performance comparison of our approach and MALLER alone (Table 1). For more details regarding the performance of the algorithms designed in [5] see the original tables at [12].

It is easy to see that for $m = 1, 3$ and 5, the M-MLS algorithm achieves more accurate results (the results for $m = 2$ and 4 are similar).

For running time measurements of the M-MLS, we used a laptop with Intel i7 -6700HQ core with 16 GB RAM. We compared our timing against the fastest method reported in [12], which is NEDE [5] (Table 2). The timing of NEDE, quoted from [12], based upon a server with 96 GB of RAM, two Intel Xeon X5570 CPUs, each with four cores running at 2.93GHz.

Alg	$n = 1500$	$n = 1000$
Best runtime from [12] (NEDE)	6.04 ± 0.16	5.59 ± 0.15
M-MLS ($m = 1$)	2.18 ± 0.02	1.47 ± 0.02
M-MLS ($m = 3$)	2.86 ± 0.02	1.95 ± 0.02
M-MLS ($m = 5$)	3.73 ± 0.02	2.63 ± 0.03

Table 2: Time for computing the approximation - compared with the fastest performing algorithm out of 8 tested algorithms in [12]

5 Acknowledgments

This research was partially supported by the Israel Science Foundation (ISF 1556/17), Blavatink ICRC Funds, Fellowships from Jyväskylä University and the Clore Foundation.

References

[1] Yariv Aizenbud and Amir Averbuch. Matrix decompositions using sub-Gaussian random matrices. *Information and Inference: A Journal of the IMA*, 2018.

[2] Yariv Aizenbud, Amit Bermanis, and Amir Averbuch. PCA-based out-of-sample extension for dimensionality reduction. *arXiv preprint arXiv:1511.00831*, 2015.

[3] Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T Silva. Computing and rendering point set surfaces. *Visualization and Computer Graphics, IEEE Transactions on*, 9(1):3–15, 2003.

[4] William K Allard, Guangliang Chen, and Mauro Maggioni. Multi-scale geometric methods for data sets ii: Geometric multi-resolution analysis. *Applied and Computational Harmonic Analysis*, 32(3):435–462, 2012.

[5] Anil Aswani, Peter Bickel, and Claire Tomlin. Regression on manifolds: Estimation of the exterior derivative. *The Annals of Statistics*, pages 48–81, 2011.

[6] George Backus and Freeman Gilbert. The resolving power of gross earth data. *Geophysical Journal International*, 16(2):169–205, 1968.

[7] George E Backus and JF Gilbert. Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal International*, 13(1-3):247–276, 1967.

[8] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[9] Richard Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, USA, 1 edition, 1957.

[10] LP Bos and K Salkauskas. Moving least-squares are backus-gilbert optimal. *Journal of Approximation Theory*, 59(3):267–275, 1989.

[11] Guangliang Chen, Anna V Little, and Mauro Maggioni. Multi-resolution geometric analysis for data in high dimensions. In *Excursions in Harmonic Analysis, Volume 1*, pages 259–285. Springer, 2013.

[12] Ming-Yen Cheng and Hau-tieng Wu. Local linear regression on manifolds and its geometric interpretation. *Journal of the American Statistical Association*, 108(504):1421–1434, 2013.

[13] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[14] David L Donoho et al. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, pages 1–32, 2000.

[15] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

[16] G Hughes. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on*, 14(1):55–63, 1968.

[17] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[18] Peter W Jones, Mauro Maggioni, and Raanan Schul. Manifold parametrizations by eigenfunctions of the laplacian and heat kernels. *Proceedings of the National Academy of Sciences*, 105(6):1803–1808, 2008.

[19] Teuvo Kohonen. *Self-organizing maps*, volume 30. Springer Science & Business Media, 2001.

[20] Peter Lancaster and Kes Salkauskas. Surfaces generated by moving least squares methods. *Mathematics of computation*, 37(155):141–158, 1981.

[21] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.

[22] David Levin. The approximation power of moving least-squares. *Mathematics of Computation of the American Mathematical Society*, 67(224):1517–1531, 1998.

[23] David Levin. Mesh-independent surface interpolation. In *Geometric modeling for scientific visualization*, pages 37–49. Springer, 2004.

[24] Anna V Little, Mauro Maggioni, and Lorenzo Rosasco. Multiscale geometric methods for data sets i: Multiscale svd, noise and curvature. *Applied and Computational Harmonic Analysis*, 43(3):504–567, 2017.

[25] Mauro Maggioni, Stanislav Minsker, and Nate Strawn. Geometric multi-resolution analysis for dictionary learning. In *Wavelets and Sparsity XVI*, volume 9597, page 95971C. International Society for Optics and Photonics, 2015.

[26] Dermot H McLain. Drawing contours from arbitrary data points. *The Computer Journal*, 17(4):318–324, 1974.

[27] Andrew Nealen. An as-short-as-possible introduction to the least squares, weighted least squares and moving least squares methods for scattered data approximation and interpolation. *URL: http://www. nealen. com/projects*, 130:150, 2004.

[28] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al. Columbia object image library (coil-20). 1996.

[29] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[30] Lawrence K Saul and Sam T Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119–155, 2003.

[31] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.

[32] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[33] Barak Sober and David Levin. Manifolds' projective approximation using the moving least-squares (mmls). *arXiv preprint arXiv:1606.07104*, 2016.

[34] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[35] Yi Wang, Guangliang Chen, and Mauro Maggioni. High-dimensional data modeling techniques for detection of chemical plumes and anomalies in hyperspectral images and movies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9):4316–4324, 2016.