Active Manifolds:

A Geometric Approach to Dimension Reduction for Sensitivity Analysis

Anthony Gruber

Robert Bridges, Christopher Felder, Chelsey Hoff

Cyber & Information Security Research Group
Oak Ridge National Laboratory

August 1st, 2018

Outline

Outline

The Problem

- You are trying to recover the values of some unknown C^1 high-dimensional $f : \mathbb{R}^m \to \mathbb{R}$.
- You have a sampling of data points and their f-values. This cost you a lot of money to obtain, and obtaining more data is infeasible.
- How can you make the most of what you have? More precisely, how can you use your data to model f so that you can accurately approximate $f(p)$ for arbitrary $p \in \mathbb{R}^m$?

Historical Background

- Our method, Active Manifolds, is based on a method known as Active Subspaces, popularized by Paul Constantine in early 2010's.
- Active Subspaces uses techniques from Principle Component Analysis (PCA) to estimate what (linear combinations of) parameters change a function $f : \mathbb{R}^m \to \mathbb{R}$ the most "on average".
- The affine subspace spanned by the "largest" eigenvectors of a certain matrix is then used as an approximation to Domain(f).

The AS Algorithm

1. Sample ∇f at N random points $a_i \in [-1,1]^m$
2. Find the directions in which f changes the most on average, the *Active Subspace*. This is done by computing the eigenvalue decomposition of the matrix

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N} \nabla f_{a_i} \nabla f_{a_i}^T = \mathbf{W \Lambda W}^T$$

3. Manually inspect the set $\mathbf{\Lambda} = (\lambda_i)$ for "large" gaps. If there is a gap between λ_i and λ_{i+1}, we say $\text{Span}\left(\{w_1, ..., w_i\}\right)$ is the active subspace of dimension i.
4. Given an arbitrary point $p \in [-1,1]^M$, project p orthogonally to p' on the active subspace and obtain the value $f(p') \approx f(p)$.

AS Benefits

- It's fast, because it solves a linear problem.
- It's relatively insensitive to local data, because it takes an average of the rates of change.
- It has nice error estimates, because the arguments are linear algebraic in nature.

AS Drawbacks

- It isn't guaranteed to lower the dimension.
- It can be highly inaccurate (because it doesn't respect the geometry of f).
- Some functions don't even admit a well-defined Active Subspace e.g. $f(x, y) = x^2 + y^2$.

AS Drawbacks

- It isn't guaranteed to lower the dimension.
- It can be highly inaccurate (because it doesn't respect the geometry of f).
- Some functions don't even admit a well-defined Active Subspace e.g. $f(x, y) = x^2 + y^2$.

These problems all arise because the Active Subspace is restricted to be affine!

Outline

1. The Method and Algorithm
   - Background
   - Mathematical Justification
   - The AM Algorithm
2. Results
   - A Simple Example
   - MHD Data Example
   - Ebola Spread Example
3. Summary and Future Work

The Key Observation

The key observation to the success of AS as well as AM is the following classical result:

Theorem

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a C^1 function. Then ∇f is a continuous normal vector field on the level sets $\{f^{-1}(x)\}_{x \in \mathbb{R}}$ of f. In particular, the rate of change in f is everywhere maximized in the direction of ∇f.

This assures us that ∇f captures all of the change undergone by f.

Our Approach

We seek a nonlinear analogue of the Active Subspace defined by f. In particular, we consider a (1-D!) geometric object as follows:

Definition

Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is a given C^1 function. By an *active manifold defined by f* (hereby abbreviated AM) we mean an integral curve of ∇f. That is, a function $\gamma : [0, 1] \to \mathbb{R}^m$ such that $\gamma'(t) = \nabla f_{\gamma(t)}$ for all $t \in [0, 1]$.

Note: We will often want to assume $|\nabla f| = 1$ ((WLOG away from zeros), so that our AM is parametrized with unit speed.

Properties of Active Manifolds

The following illustrates some nice properties of AM's.

Proposition

Suppose df is nonvanishing, and let $\mathcal{F} = \{f^{-1}(x)\}_{x \in \mathbb{R}}$ denote the foliation of $\mathbb{R}^m$ by level sets of the C^1 submersion $f : \mathbb{R}^m \to \mathbb{R}$. Then, for every point $p \in \mathbb{R}^m$ there is a unique active manifold $\gamma(t)$ passing through p. Moreover, $\gamma(t)$ is everywhere transverse to $\mathcal{F}$ and intersects each leaf at most once.

Properties of Active Manifolds

The following illustrates some nice properties of AM's.

Proposition

Suppose df is nonvanishing, and let $\mathcal{F} = \{f^{-1}(x)\}_{x\in\mathbb{R}}$ denote the foliation of $\mathbb{R}^m$ by level sets of the C^1 submersion $f : \mathbb{R}^m \to \mathbb{R}$. Then, for every point $p \in \mathbb{R}^m$ there is a unique active manifold $\gamma(t)$ passing through p. Moreover, $\gamma(t)$ is everywhere transverse to $\mathcal{F}$ and intersects each leaf at most once.

Note for f not so nice: These properties remain valid away from critical points and singularities in the function domain.

Sketch of Proof

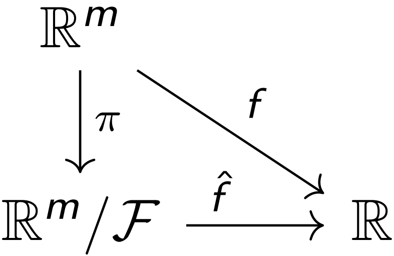
1. df nonvanishing $\implies \nabla f \neq 0 \implies$ every $p \in \mathbb{R}^m$ admits a maximal integral curve $\gamma(t)$ existing for all t.
2. $\ker df$ defines a codimension 1 distribution $\mathcal{D}$ in $\mathbb{R}^m$. Further, $\mathcal{D}$ is involutive, hence integrable by Frobenius. This yields the foliation of integral surfaces $\mathcal{F}$.
3. Transversality follows since the usual metric on $\mathbb{R}^m$ splits the short exact sequence

$$0 \longrightarrow T\mathcal{F} \xrightarrow{\iota} T\mathbb{R}^m \xrightarrow{df} T\mathbb{R} \longrightarrow 0.$$

   So we have $T\mathbb{R}^m \cong T\mathcal{F} \oplus T\mathbb{R}$ where $\gamma' \subset T\mathbb{R}$.
4. $\gamma(t)$ is globally transverse to $\mathcal{F}$, so parametrizes the leaf space $\mathbb{R}^m/\mathcal{F}$. Hence it intersects every leaf once (if the leaves are connected) or perhaps not at all.

Our Approach

It follows from this proposition (and a little more work) that we have the following commutative diagram



where the factorization map $\hat{f}$ is a C^1 diffeomorphism.

Outline

Algorithm Plan

This suggests the following scheme for our algorithm:

1. Build the active manifold $\gamma(t)$ through a given point p_0.
2. Approximate the projection map $\pi : \mathbb{R}^m \to \mathbb{R}^m/\mathcal{F} \cong \mathbb{R}$
3. Use $f(p) = \hat{f}([p]) = f(\gamma(t_0))$ to compute $f(p)$ for $\gamma(t_0) \in [p]$.

Preliminary Technical Annoyance

Note the following and then forget about it:

We will actually build an AM on $[-1,1]^m$, so we must map the input space $K \subset \mathbb{R}^m$ diffeomorphically through anisotropic scaling (point-slope linear fxn in m vars), and solve the problem

$$\tilde{\gamma}'(t) = (D^{-1}\gamma)'(t) = \frac{\nabla(D^* f)_{D^{-1}\gamma(t)}}{|\nabla(D^* f)_{D^{-1}\gamma(t)}|} = \frac{\nabla f_{\gamma(t)} \circ D_{D^{-1}\gamma(t)}}{|\nabla f_{\gamma(t)} \circ D_{D^{-1}\gamma(t)}|},$$

where $D : [-1,1]^m \to \text{Conv}(K)$ is the scaling function. In this way we associate the AM $\tilde{\gamma}(t)$ on $[-1,1]^m$ defined by $D^* f$ to the AM $\gamma(t)$ on $\text{Conv}(K)$.

Algorithm Details

Building γ given n samples $\{a_i\}$ of m-dimensional input data $A \in \mathbb{M}^{n \times m}$ and corresponding function values $\{f(a_i)\}$.

- Compute ∇f_{a_i} for each i (analytically or with F.D, A.D, etc.), map $\{a_i, \nabla f_{a_i}\}$ linearly to $[-1, 1]^m$ and normalize so that $|\nabla(D^* f)_{D^{-1} a_i}| = 1$.

WARNING: **Don't forget about the chain rule!** Be careful when normalizing to $[-1, 1]^m$, since we will actually be building an AM defined by $D^* f$ as in the previous slide. From now on, we drop the pullback notation and use simply $f, \nabla f$.

Algorithm Details (2)

Building the AM continued:

- Use gradient descent w/ nearest-neighbor search to construct the set $\{\gamma_i, f(p_i)\}_{i=1}^N$ where $p_i \in A$ is the nearest neighbor to γ_i.
- Scale $i \mapsto i/N$ so the domain of γ is $[0, 1]$ (uniform step size makes this okay). We now have $\{\gamma_i\}$ the discretized AM defined by f in $[-1, 1]^m$, as well as an approximation to $f(\gamma_i)$ for each i.
- Use $\{i, f(\gamma_i)\}_1^N$ to construct a piecewise-cubic Hermite interpolating function $\hat{f}(\gamma(t))$ defined on the entirety of $[0, 1]$. Note that $\hat{f}$ is C^1 by construction.

Algorithm Details (3)

Approximating the projection map: The idea is to walk the level set corresponding to given $p \in \mathbb{R}^m$ until it hits the AM.

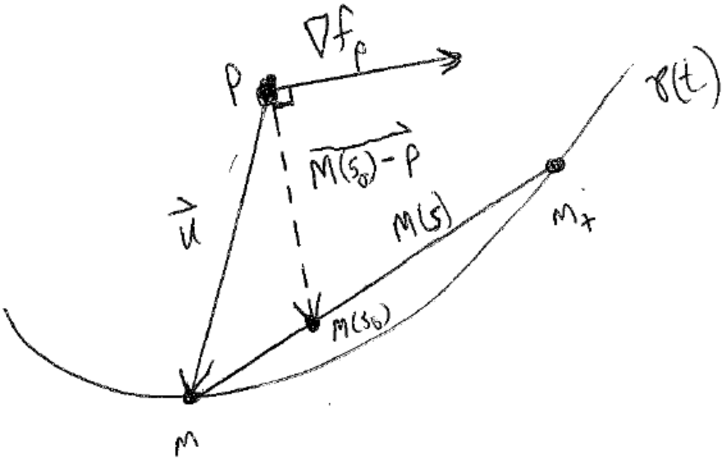
- Given $p \in \mathbb{R}^m$, compute $\tilde{p} = D^{-1}(p)$.
- Compute the closest point m on γ to $\tilde{p}$, by minimizing $|\tilde{p} - m|$ for $m \in \gamma$.
- Construct the vector $\mathbf{u} = m - \tilde{p}$ and compute the projection of $\mathbf{u}$ onto the tangent space $T\mathcal{F}|_{\tilde{p}}$, $\mathbf{v} = \mathbf{u} - \langle \nabla f_{\tilde{p}}, \mathbf{u} \rangle$.
- Let $\tilde{p} = \tilde{p} + \delta \mathbf{v}$ for specified δ.
- For specified $\epsilon > 0$, repeat steps $(1) - (3)$ while $|p - m| > \epsilon$.

Algorithm Details (4)

Approximating the projection map continued: (Assume $|\nabla f| = 1$).

- Parametrize the line segment $M(s) : [0,1] \to [m, m_+]$ between m and m_+, where m_+ is the next closest point on γ to $\tilde{p}$.
- Determine s_0 such that $M(s_0) - \tilde{p}$ is orthogonal to ∇f_p.
- Evaluate $\hat{f}(M(s_0)) \approx \hat{f}(\gamma(t_0)) = f(p)$.

Illustration of Projection Algorithm



Outline

Foliation by Ellipses

Let (x, y) be local coordinates on $\mathbb{R}^2$ and consider the function $L : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$L(x, y) = \frac{y^2}{1 - x^2} \qquad |x| < 1 \tag{1}$$

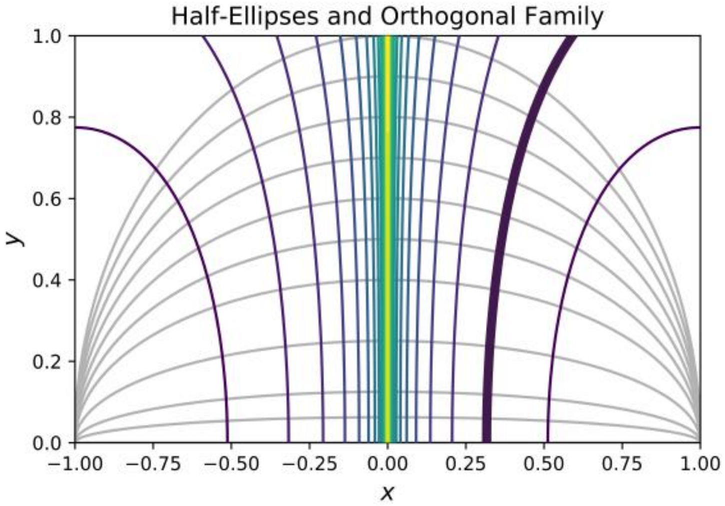
With differential given by

$$dL_{(x,y)}(\mathbf{v}) = \frac{2xy^2}{(1 - x^2)^2}\, dx(\mathbf{v}) + \frac{2y}{1 - x^2}\, dy(\mathbf{v}) \tag{2}$$

ker dL defines a codimension-one distribution on the open half-disk $\mathbb{D}_+ = \{(x, y) : x^2 + y^2 < 1\} \cap \{(x, y) : y > 0\}$ that integrates to a foliation of $\mathbb{D}_+$. Note that the orthogonal family of curves is the collection of level sets of

$$L^{\perp}(x, y) = x^2 + x^2 - \ln(x^2). \tag{3}$$

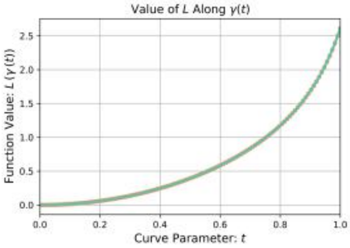
Visualization



The Experiment (L)

- A uniform grid of 10000 points over $(-1, 1) \times (0, 1]$ was created, and f-values / gradients were computed analytically at these points.
- 8000 of these points were sampled at random and chosen as the training points, the remaining 2000 were designated as testing points.
- Given a random point p, an AM γ through p was built from the training grid w/ stepsize 0.02.
- A p.w.-cubic Hermite interpolant $\hat{f}$ was fit to the f-values along γ.
- Each testing point q was projected to γ, and its value $\hat{f}(\gamma(t_q)) \approx f(q)$ was computed, where $\gamma(t_q) \in [q]$.
- The average absolute error and average ℓ^2 error in the approximated values $\{\hat{f}(t_q)\}_{q \in Q}$ was computed.

AM vs AS (L)



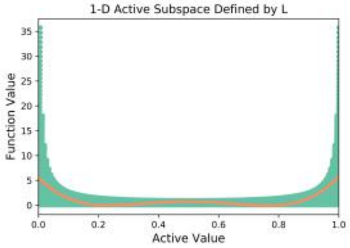


Figure: Cubic splines fit to the AM (left), and a degree 4 poly fit to the AS (right). It's clear that AM performs better in this example.

Method	**Average Absolute Error**	**Average ℓ^2 Error**
Active Subspaces	1.29	2.51
Active Manifolds	0.438	0.0507

At the Edge

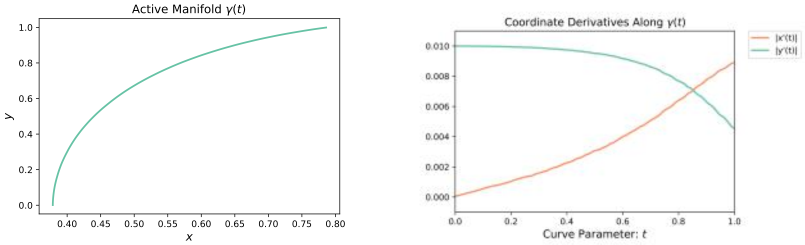
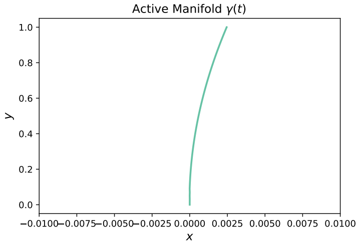


Figure: The AM $\gamma(t)$ (left) and derivatives (right) along the AM defined by L with starting value $(0.4, 0.3)$. Notice that the relative importance of the x-coordinate overtakes that of the y-coordinate as the manifold bends outward similar to the thick purple curve in the previous slide.

At the Center



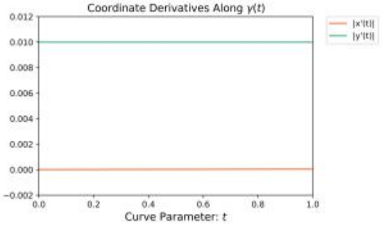


Figure: The AM $\gamma(t)$ (left) and derivatives (right) along the AM defined by L with starting value $(0, 0.1)$. Notice that the algorithm traces closely the line $x = 0$ (the thick yellow curve in the slide two previous).

Outline

AM on Real Data!

We now consider an example of a model for magnetohydrodynamic (MHD) power generation found in Glaws et al. [1]. This data is taken from a model for idealized 3-D duct flow through a MHD generator. We examine separately the average flow velocity u_{avg} and the induced magnetic field B_{ind}, each depending on 5 variable parameters.

We would like study the sensitivity of these functions to each parameter. The parameter variability ranges are found in the following table:

MHD Table

Variable	Notation	Range
Fluid Viscosity	$\log(\mu)$	$[\log(.001), \log(.01)]$
Fluid Density	$\log(\rho)$	$[\log(.1), \log(10)]$
Applied Pressure Gradient	$\log(\frac{\partial p_0}{\partial x})$	$[\log(.1), \log(.5)]$
Resistivity	$\log(\eta)$	$[\log(.1), \log(10)]$
Applied Magnetic Field	$\log(B_0)$	$[\log(.1), \log(1)]$
Magnetic Constant	μ_0	fixed at 1
Length	l	fixed at 1

Table: The parameters and ranges for the idealized MHD generator problem. We consider the first 5 to be variable, while the others are fixed at 1. Samples are drawn with respect to a uniform distribution on the specified ranges. This table is reproduced from Glaws et al. [2].

The Experiment (MHD)

We mimic the setup undertaken by Glaws et al. in [1]

- Using the same 483 samples of $\{p, u_{avg}(p), B_{ind}(p), \nabla u_{avg}(p), \nabla B_{ind}(p)\}$ as Glaws et al., we randomly partition 350/133 for training/testing and map the data to $[-1, 1]^5$.
- We run AM on the training data, stepsize 0.15, to construct γ passing through a random point in $[-1, 1]^5$, and fit the function data to a pw-cubic Hermite interpolant $\hat{f}$.
- We project the training points to γ, compute their f-value there, and compare to the exact value.
- We compute the average absolute error and average ℓ^2 error in our approximation.

Functions along Their Respective AM's (MHD)

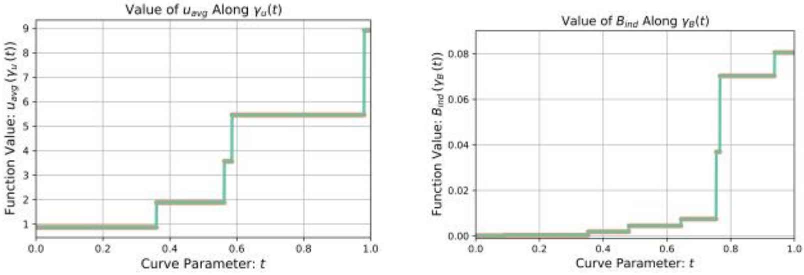


Figure: piecewise-cubic Hermite splines fit to the active manifolds defined by u_{avg} resp. B_{ind}. Data was taken and processed identically as in Glaws et al. [1]

Sensitivity Analysis (MHD)

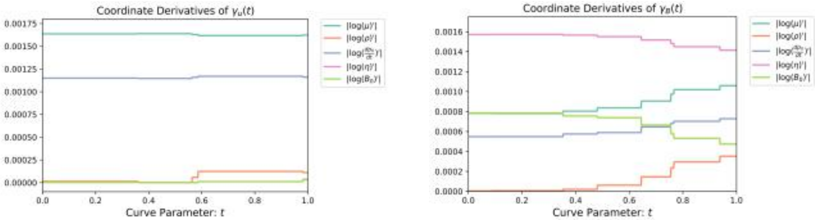


Figure: Change in the coordinate derivatives along γ_u (left) and γ_B (right). Note the nonlinear behavior as reflected in the errors. Once again, the increased resolution of AM makes a big difference.

Method	**Avg** abs/ℓ^2 **Errors u**	**Avg** abs/ℓ^2 **Errors B**
Active Subspaces	2.21/2.88	0.0151/0.0225
Active Manifolds	0.819/0.143	0.00396/0.000757

Outline

1. The Method and Algorithm
   - Background
   - Mathematical Justification
   - The AM Algorithm
2. Results
   - A Simple Example
   - MHD Data Example
   - Ebola Spread Example
3. Summary and Future Work

About the Model

We consider an 8-parameter model by Diaz et al. in [4] for modeling the spread of Ebola in Liberia and Sierra Leone, governed by the equations:

$$\frac{dS}{dt} = -\beta_1 SI - \beta_2 SR_I - \beta_3 SH, \tag{4}$$

$$\frac{dE}{dt} = \beta_1 SI + \beta_2 SR_I + \beta_3 SH - \delta E, \tag{5}$$

$$\frac{dI}{dt} = \delta E - \Gamma_1 I - \psi I, \tag{6}$$

$$\frac{dH}{dt} = \psi I - \Gamma_2 H, \tag{7}$$

$$\frac{dR_I}{dt} = \rho_1 \Gamma_1 I - \omega R_I, \tag{8}$$

$$\frac{dR_B}{dt} = \omega R_I + \rho_2 \Gamma_2 H, \tag{9}$$

$$\frac{dR_R}{dt} = (1-\rho_1)\Gamma_1 I + (1-\rho_2)\Gamma_2 H, \tag{10}$$

where the scalar quantity of interest is known as the basic reproduction number R_0, defined as (Diaz et al. [4])

$$R_0 = \frac{\beta_1 + \frac{\beta_2 \rho_1 \Gamma_1}{\omega} + \frac{\beta_3}{\Gamma_2}\psi}{\Gamma_1 + \psi}. \tag{11}$$

Parameters and Ranges (Ebola)

Parameter	Baseline (L)	Range (L)	Baseline (SL)	Range (SL)
β_1	.356	$(.1, .4)$	.251	$(.1, .4)$
β_2	.135	$(.1, .4)$	.395	$(.1, .4)$
β_3	.163	$(.1, .4)$	.079	$(.1, .4)$
ρ_1	.98	$(.41, 1)$	.76	$(.41, 1)$
Γ_1	.0542	$(.0276, .1702)$	.051	$(.0275, .1569)$
Γ_2	.174	$.081, .21$	.0833	$(.1236, .384)$
ω	.325	$(.25, .5)$	.370	$(.25, .5)$
ψ	.5	$(.0833, .7)$	.442	$(.0833, .7)$
ρ_2	.88	N/A	.74	N/A
δ	1/9	N/A	1/9	N/A

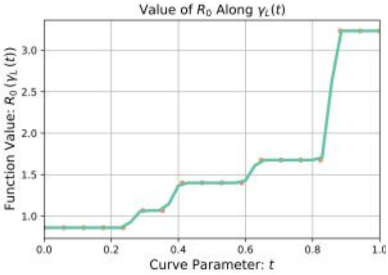
Table: The parameters and relevant data for the Ebola model. "L" denotes Liberia, and "SL" denotes Sierra Leone. The baselines were established by fitting the model to real data collected by the WHO. This table is amalgamated from tables in Diaz et al. [3].

The Experiment (Ebola)

The goal is again to perform sensitivity analysis. The following procedure was observed:

- A uniform grid over $[-1, 1]^8$ of 4^8 points was constructed.
- The data was mapped to the correct ranges, and the functions found in Diaz et al. [3] were used to compute gradients and function values.
- The input data was mapped back to the cube, being careful to adjust the gradients accordingly.
- The data was partitioned into 63536 testing points and 2000 training points, and AM was run with random starting point and step size 0.3.
- The data was fit and approximation errors were computed as usual.

R_0 Values Along Each AM



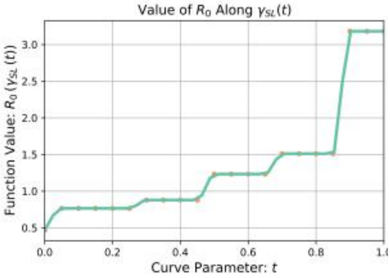


Figure: piecewise-cubic Hermite splines fit to the active manifolds γ_L (left) and γ_{SL} (right) from the Liberia and Sierra Leone data. Data was taken and processed identically to Diaz et al. in [4]

Sensitivity Analysis (Ebola)

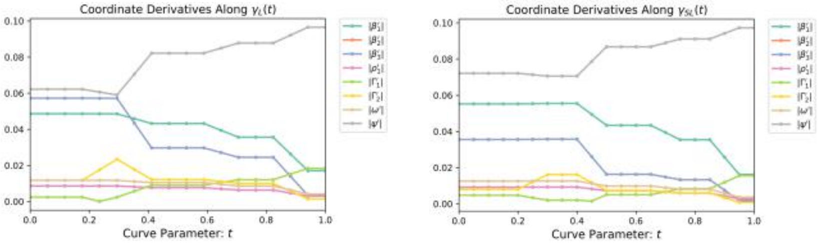


Figure: Change in the coordinate derivatives along γ_L (left) and γ_{SL} (right). Note that it is very important to consider where you are along the AM when discussing sensitivity.

Method	**Avg** abs/ℓ^2 **Errors L**	**Avg** abs/ℓ^2 **Errors SL**
Active Subspaces	0.751/0.994	0.710/0.965
Active Manifolds	0.321/0.0123	0.291/0.0121

Summary

- AM provides a slower but higher-resolution alternative to AS for dimension reduction.
- AM always reduces the dimension to 1, providing tractable and useful data visualization.
- AM is more flexible than AS due to its local (rather than global) nature.
- AM does a much better job of minimizing ℓ^2 error than AS.
- AM shows geometry is worth your money!!!

Future Work

- Critical points and singularities are still bad for both AS and AM. Adding a momentum parameter into the algorithm might help.
- AS has rigorous bounds on how much accuracy is lost with their method. It would be good to have the same for AM, though the estimates will be harder.
- Both the AM algorithm and its implementation in Python are naive and could be made much better/faster with more work.
- Most pressingly, we have two papers in progress over our current results that we need to finish.

End

Thank you for your attention!

Questions?

Glaws Andrew et al. "Dimension reduction in magnetohydrodynamics power generation models: Dimensional analysis and active subspaces". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 10.5 (2016), pp. 312–325. DOI: 10.1002/sam.11355. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sam.11355. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.11355.

Glaws Andrew et al. *Dimension Reduction in MHD Power Generation Models: Dimensional Analysis and Active Subspaces*. 2018. URL: http://nbviewer.jupyter.org/github/paulcon/as-data-sets/blob/master/MHD/MHD.ipynb#Dimension-Reduction-in-MHD-Power-Generation-Models:-Dimensional-Analysis-and-Active-Subspaces.

Paul Diaz et al. *A Modified SEIR Model for the Spread of Ebola in Western Africa and Metrics for Resource Allocation*. 2016. URL: http://nbviewer.jupyter.org/github/paulcon/as-data-sets/blob/master/Ebola/Ebola.ipynb#A-Modified-SEIR-Model-for-the-Spread-of-Ebola-in-Western-Africa-and-Metrics-for-Resource-Allocation.

Paul Diaz et al. "A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation". In: *Applied Mathematics and Computation* 324 (2018), pp. 141–155. ISSN: 0096-3003. DOI: https://doi.org/10.1016/j.amc.2017.11.039. URL: http://www.sciencedirect.com/science/article/pii/S0096300317308214.