



Bounding HyperRibbons

Basic Training in Condensed Matter
02/02/2024

Katherine Quinn
kq57@georgetown.edu



Outline

- 1) Information Geometry and the Model Manifold
- 2) Approximation Theory and Manifold Bounds
- 3) Simple Examples

Sloppiness on the model manifold arises from an underlying analyticity in the models we construct to understand phenomena.

Information Geometry

Math field that fuses together statistics and information theory with differential geometry.

Translate the problem of finding structure in models and data to one of understanding the geometric properties of high dimensional objects.

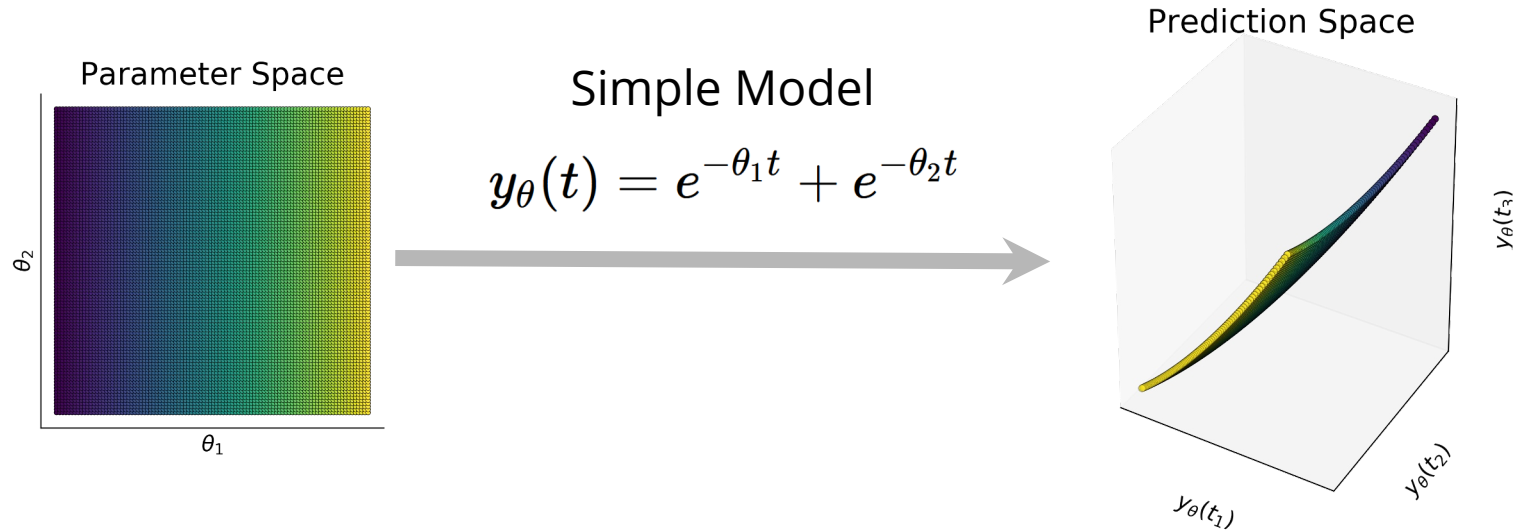
Two main objects we care about:

1. Hessian/Fisher Information Matrix (local metric)
2. Model Manifold (full model range)

Model Manifold

A model as a *mapping* from *parameter space* to *prediction space*.

The region occupied in the prediction space is our *model manifold*.



Simple Construction of a Model Manifold

Sums of Exponentials: $y_{\theta}(t) = e^{-\theta_1 t} + e^{-\theta_2 t}$

Repeat as needed

Select Evaluation Points (t)

Select a set of points at which to evaluate the model. Here we'll pick 3 values:
($t_0 = 0$, $t_1 = 0.5$, $t_2 = 1$)

Pick a Parameter Combination (θ)

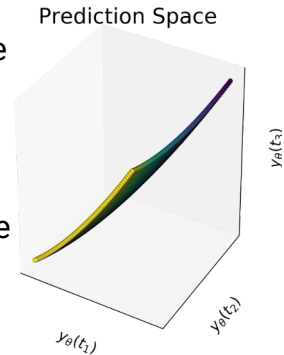
Here we have 2 parameters, θ_1 and θ_2 . Pick initial values for these, say $\theta_1 = 0$ and $\theta_2 = 0.9$.

Evaluate Model Predictions ($y_{\theta}(t)$)

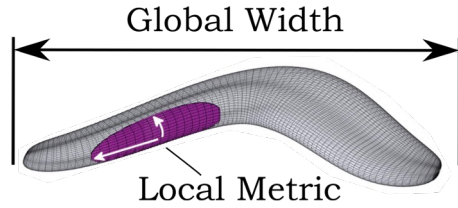
Evaluate the three points of interest with the parameter values from the previous step, to get (2, $1 + e^{-0.45}$, $1 + e^{-0.9}$)

Visualize Model Predictions

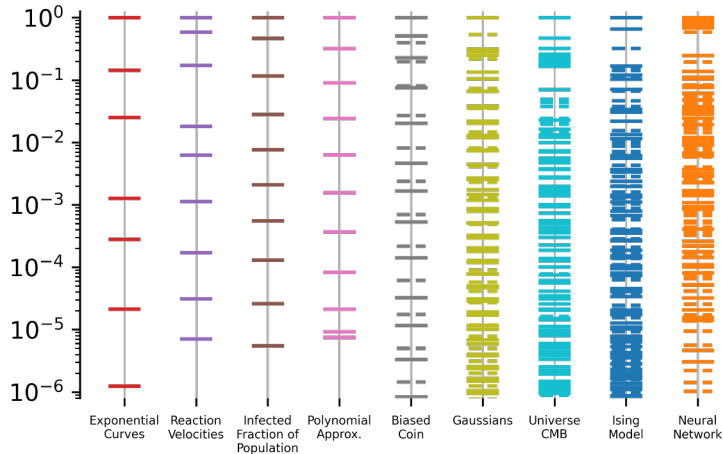
Visualized the set of 3D vectors with your favorite method (more on this next week).



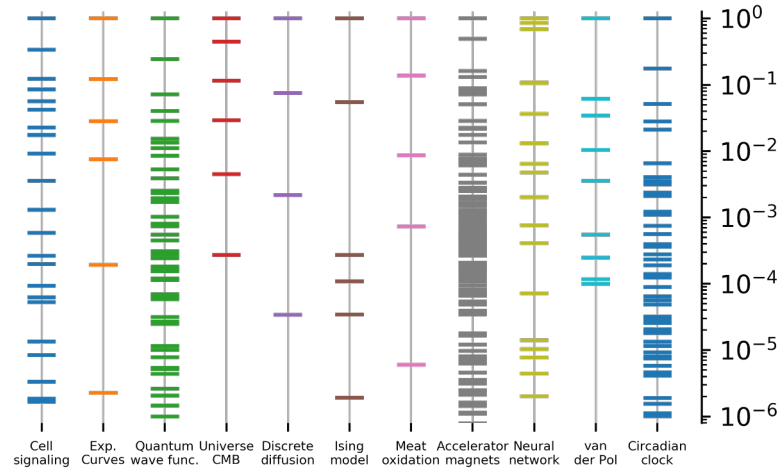
Local vs. Global Structure



Local Structure characterized by the local metric.



Global Structure characterized by manifold lengths.



Effect of Re-Parametrization

Local Regions: Re-parametrization impacts the local metric, it is possible to eliminate the hierarchy of parameter importance.

Global Regions: Re-parametrization does not change the region occupied in prediction space so the model manifold remains the same.

Universal Features of the Model Manifold

1) Hierarchies

Local and global hierarchies pop up.

1) Flatness

For the most part, model manifolds don't curl up on themselves in prediction space.

Why Hierarchies and Flatness?

There are bounds on the model manifold which constrain its shape.

Approximation Theory

- 1) Approximate complex, nonlinear models with simpler functional forms (e.g. polynomials).

$$\underbrace{y_{\theta}(t)}_{\text{model}} = \sum_{j=0}^{\infty} \underbrace{a_j(\theta)}_{\text{coeff.}} \underbrace{\phi_j(t)}_{\text{basis}} \rightarrow \underbrace{p_{N-1}(t)}_{\text{approx.}} = \sum_{j=0}^{N-1} a_j(\theta) t^j$$

- 2) Find bounds on the error.

Polynomials Create Linear Mappings

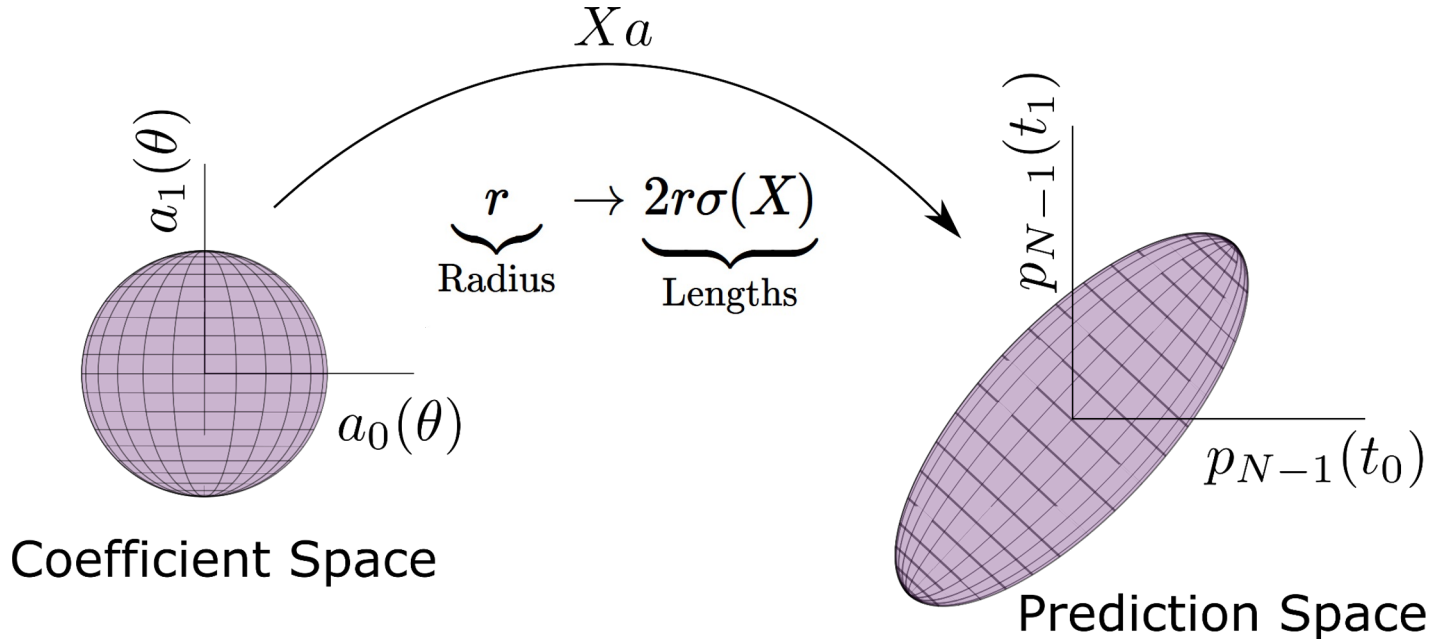
Model predictions are *vectors*, and by separating the model parameters from the evaluation points we can create a *linear relationship* between the two.

Example: Truncated Taylor Series

$$\underbrace{y_\theta(t)}_{\text{model}} = \sum_{j=0}^{\infty} \underbrace{a_j(\theta)}_{\text{coeff.}} \underbrace{\phi_j(t)}_{\text{basis}} \rightarrow \underbrace{p_{N-1}(t)}_{\text{approx.}} = \sum_{j=0}^{N-1} a_j(\theta) t^j$$
$$\begin{pmatrix} y_\theta(t_0) \\ \dots \\ y_\theta(t_{N-1}) \end{pmatrix} \rightarrow \begin{pmatrix} p_{N-1}(t_0) \\ \dots \\ p_{N-1}(t_{N-1}) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & t_0 & t_0^2(\theta) & \dots & t_0^{N-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & t_{N-1} & t_{N-1}^2(\theta) & \dots & t_{N-1}^{N-1} \end{pmatrix}}_{\text{linear map, } X} \begin{pmatrix} a_0(\theta) \\ \dots \\ a_{N-1}(\theta) \end{pmatrix}$$

Linear Maps Rotate and Warp Hyperspheres

The new axes lengths of the resulting *hyperellipsoid* are determined by the radius of the hypersphere and singular values of the linear mapping.



Bounding the Model Manifold

Approximation theory gives us an error term of $\|y_{\theta}(t) - p_N(t)\|_{\infty}$.

The bounds on the lengths of the **original model manifold** are given by:

$$2r\sigma_j(X) + 2\|y - p_N\|_{\infty}$$

Hypersphere radius
of polynomial
coefficients.

Singular values
from linear
mapping.

Error term from
approximation theory.

Finding Singular Values and Errors

- 1) Without loss of generality, we shift and rescale our model so that points $(t_0, t_1, \dots, t_{N-1})$ are all between -1 and 1. Center on $t=0$.
- 2) All polynomial approximations are of the order $N-1$ where N is the number of points t .

To actually find the singular values and error terms, we consider two different cases:

- 1) Taylor Series Approximation
- 2) Chebyshev Polynomial Approximation

1) Taylor Approximation

$$p_{N-1}(t) = \sum_{k=0}^{N-1} a_k(\theta)(t - t_0)^k$$

The points t in this expansion take the shape of the *Vandermonde Matrix*.

Find constants C and R such that, for our truncated Taylor series and for all θ :

$$\sum_{k=0}^{N-1} \left(\frac{R^k}{k!} \frac{d^k y_\theta(t)}{dt^k} \right)^2 < C^2 N$$

Meaning that the Taylor series has a radius of convergence of at least R .

Analyticity of the Model and Hyperellipsoid Bounds

The linear mapping which goes from a hypersphere in polynomial coefficient space to a hyperellipsoid in prediction space takes the shape of $X = VD$ where:

Vandermonde $\rightarrow V_{ij} = t_{i-1}^{j-1}$ $D = \text{diag}(R^0, \dots, R^{-(N-1)})$

The singular values of X are therefore given by

$$\sigma_k(X) \leq \frac{\sqrt{N}}{\sqrt{R^2 - 1}} R^{-k+2}$$

Bounds decay exponentially

Error Term for Taylor Approximation

As a sanity check

We see that the error on the model approximation is given by:

$$\|y - p_{N-1}\|_{\infty} \leq \frac{C(NR - N + R)}{(1 - R)^2} R^{-N+1}$$

On the order of the
size of the last bound

2) Chebyshev Approximation

$$p_{N-1}(t; \theta) = \sum_{j=0}^{N-1} c_j(\theta) T_j(t)$$

Chebyshev expansions have near-best global approximation properties.

Improve on Taylor Series constraints by factors of 2^j (i.e. tighter bounds).

Analyticity of the original model $y_\theta(t)$ is determined by its region of analyticity in the complex plane, called the *Bernstein Ellipse*, categorized by ρ . We find a constant M such that for all θ ,

$$|y_\theta| \leq M$$

Analyticity of the Model and Hyperellipsoid Bounds

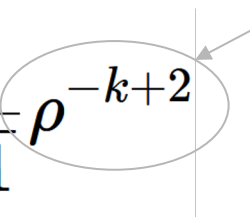
The linear mapping which goes from a hypersphere in polynomial coefficient space to a hyperellipsoid in prediction space takes the shape of $X = J D$ where:

$$J_{ij} = T_{j-1}(t_{i-1}) \quad D_{jj} = \rho^{-(j-1)}$$

Chebyshev Polynomial 

The singular values of X are therefore given by

$$\sigma_k(X) \leq \frac{\sqrt{N}}{\sqrt{\rho^2 - 1}} \rho^{-k+2}$$

 Bounds decay exponentially

Error Term for Chebyshev Approximation

As a sanity check

We see that the error on the model approximation is given by:

$$\|y_\theta - p_{N-1}\|_\infty \leq \frac{2M\rho^{-N+1}}{\rho - 1}$$

On the order of the size of the last bound

Plotting Bounds with Examples

To illustrate the universality of these bounds, we consider 3 models and generate their model manifolds, then see how they compare to the extracted bounds from the hyperellipsoid from the Chebyshev approximation method.

Same evaluation points (t) with same analyticity.



Sums of
Exponentials



Reaction
Velocities



Infected
Population



Sums of Exponentials

$$\sum_{\alpha} A_{\alpha} \exp(-\lambda_{\alpha} t)$$

Parameter Sampling

Pick a Parameter Combination (θ)

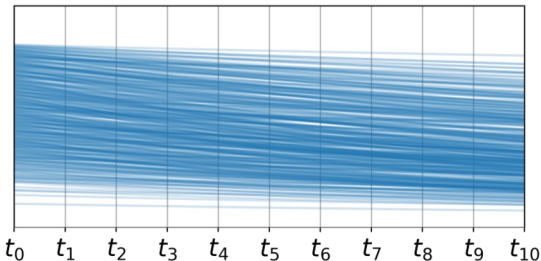
Evaluate Model Predictions ($y_{\theta}(t)$)

Visualize Model Predictions

Parameters represent decay rates and relative starting fractions for radioactive materials.

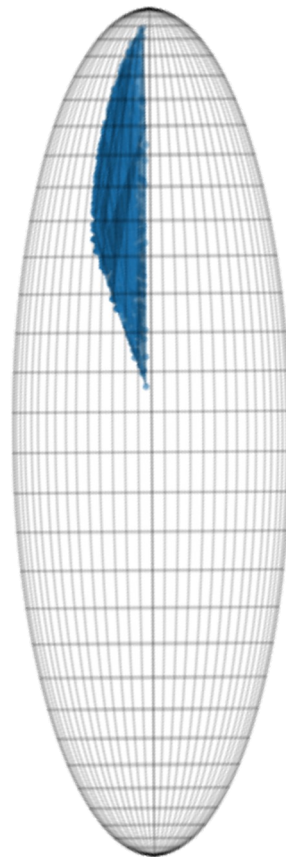
Here we have 10 “radioactive materials” in our model, and vary these parameters which are always positive.

Each set of predictions is a line from t_0 to t_{10} . All lines are slowly decaying with varied rates.



We visualize an 11-dimensional object in a 2D plane, so we orient the vectors along the axes of the bounding hyperellipsoid.

Mesh is the hyperellipsoid.





Reaction Velocities

$$\frac{\theta_1 t^2 + \theta_2 t}{t^2 + \theta_3 t + \theta_4}$$

Parameter Sampling

Pick a Parameter Combination (θ)

Evaluate Model Predictions ($y_\theta(t)$)

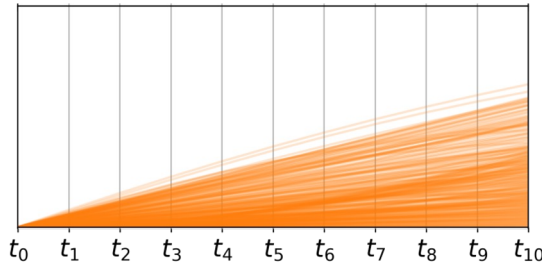
Visualize Model Predictions

Model to predict the reaction velocity of enzyme-catalyzed chemical reaction.

Here we have 5 parameters in our model.

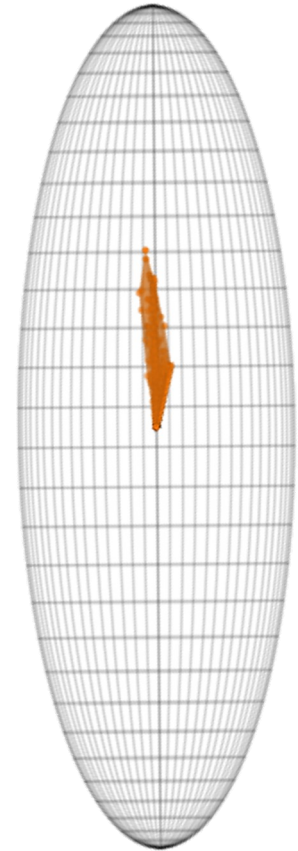
Substrate concentration is represented by t .

Each set of predictions is a line from t_0 to t_{10} . All lines start at 0 and then increase with substrate concentration.



We visualize an 11-dimensional object in a 2D plane, so we orient the vectors along the axes of the bounding hyperellipsoid.

Mesh is the hyperellipsoid.





Infected Population

Parameter Sampling

$$\dot{S} = -\beta \frac{IS}{N_{tot}}$$
$$\dot{I} = \beta \frac{IS}{N_{tot}} - \gamma I$$
$$\dot{R} = \gamma I,$$

Pick a Parameter Combination (θ)

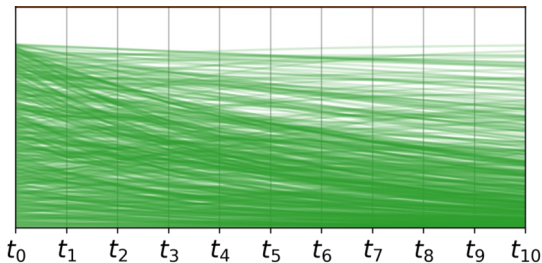
Evaluate Model Predictions ($y_{\theta}(t)$)

Visualize Model Predictions

Vanilla model in epidemiology that has infection (β) and recovery (γ) rates for a pandemic.

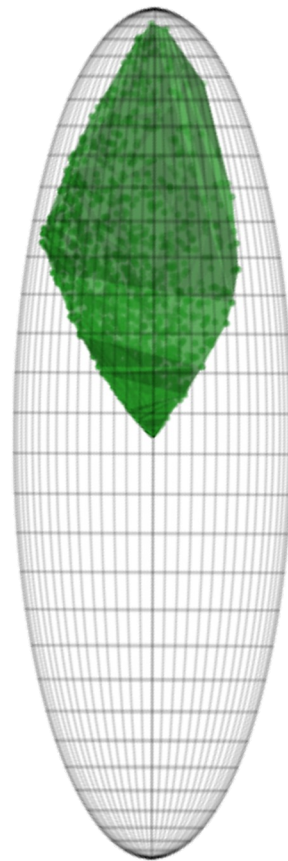
Susceptible (S), Infected (I) and Recovered (R) fractions of the population.

Each set of predictions for the **Infected** is a line from t_0 to t_{10} . All lines start at 0 and then vary in time.



We visualize an 11-dimensional object in a 2D plane, so we orient the vectors along the axes of the bounding hyperellipsoid.

Mesh is the hyperellipsoid.

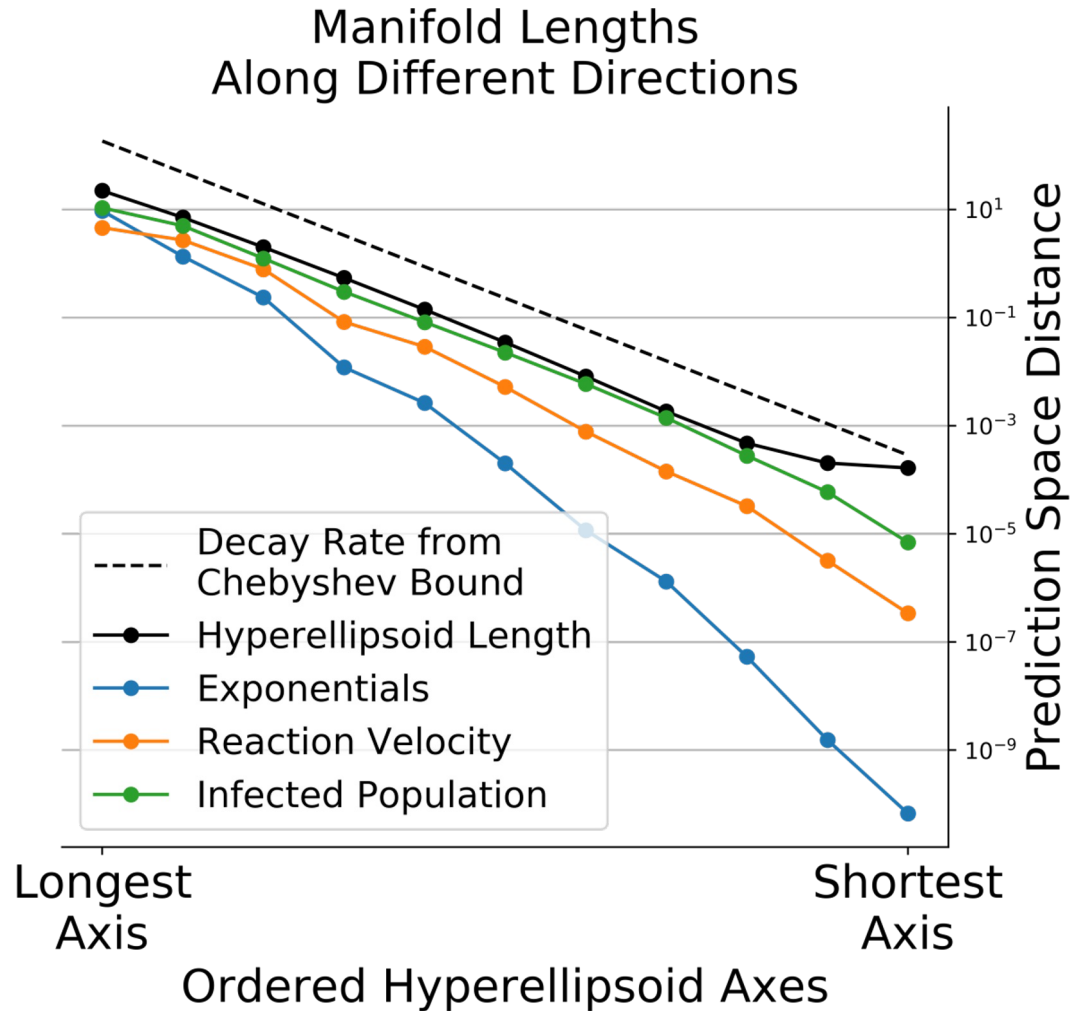


Manifold Lengths

Dotted line shows universal decay rate for our models.

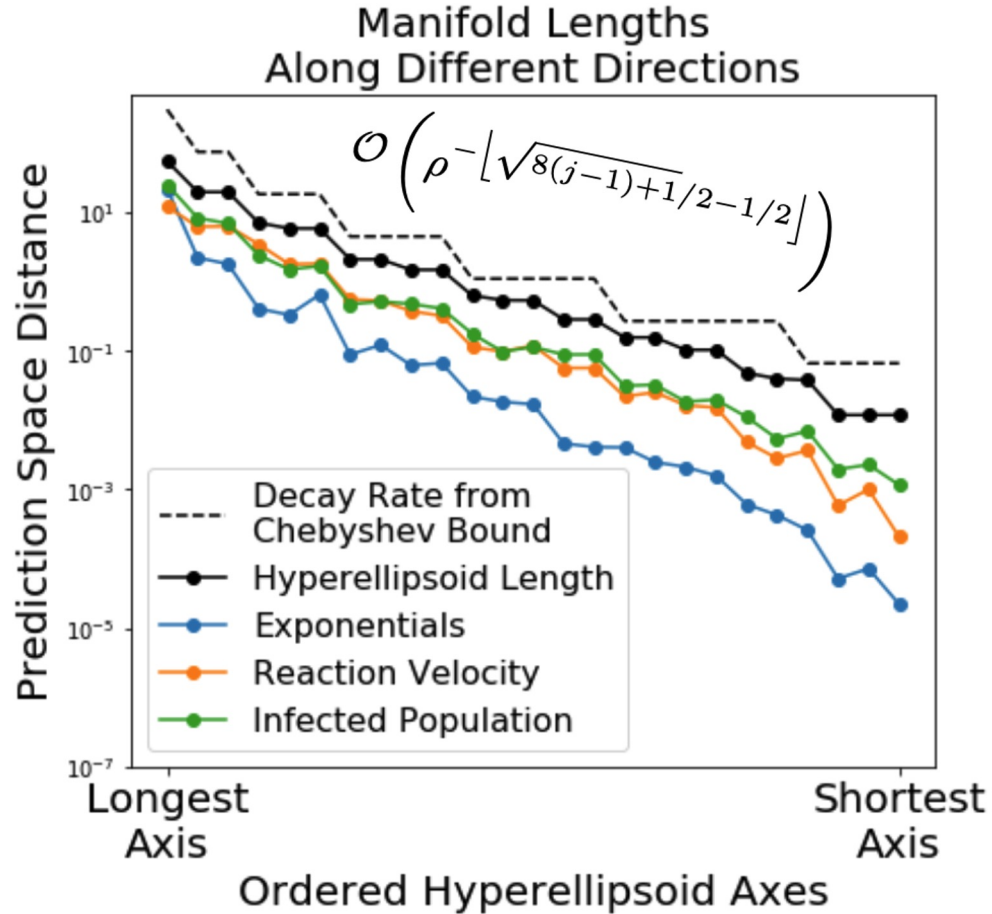
Black line are hyperellipsoid bounds (note uptick for error terms).

In all cases manifold lengths are within bounds.



Multidimensional Cases

For models that predict outcomes with more than one experimental condition (e.g. time and temperature) we still see a hyperribbon structure and can bound the model manifold by a hyperellipsoid.



Summary

- 1) Models have geometric objects called *model manifolds* which exhibit structural hierarchies in lengths and widths (sloppy), and are relatively flat.
- 2) Using approximation theory, bounds on the model manifolds can be obtained which show a geometric decay in widths, thus constraining manifolds and giving them their characteristics.