Supplemental Material for: Chebyshev approximation and the global geometry of model predictions

Katherine N. Quinn,[1] Heather Wilber,[2] Alex Townsend,[3] and James P. Sethna[4]

[1] *Physics Department, Cornell University, Ithaca, NY 14853-2501, United States. (knq2@cornell.edu)*
[2] *Center for Applied Mathematics, Cornell University, Ithaca, NY 14853-3801, United States. (hdw27@cornell.edu)*
[3] *Mathematics Department, Cornell University, Ithaca, NY 14853-4201, United States. (townsend@cornell.edu)*
[4] *Physics Department, Cornell University, Ithaca, NY 14853-2501, United States. (sethna@lassp.cornell.edu)*

(Dated: March 31, 2019)

In Section I, we discuss how our bounds can be extended to models that are not analytically continuable, but are k-times continuously differentiable on the interval of approximation. In Section II, we give numerical results for high-dimensional manifolds and discuss the behavior of the singular values of the matrix VD associated with truncated Taylor expansions. In Section III, we extend the 1D models described in the main text to include two experimental conditions, and show that their manifolds exhibit a hyperribbon-like structure that is captured by our bounds. Finally, in Section IV, we show how the visualizations of the model manifolds in the main text were generated.

I. NON-ANALYTIC MODELS

In the main text, we considered models $y_\theta(t)$, $t \in [-1,1]$, that are continuously dependent on parameters $\theta = (\theta_1, \ldots, \theta_K)$ and analytic in an open neighborhood of $[-1,1]$. We bounded the model manifold $\mathcal{Y}$ of model predictions by considering the truncated Chebyshev approximation

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

where T_j is the Chebyshev polynomial of degree j. When y_θ is not analytic on $[-1,1]$, the convergence of Eq. (1) to y_θ as $N \to \infty$ is still controlled by the smoothness of y_θ. A standard result supplied in [1, Ch. 7] states that if y_θ has $\nu - 1 \geq 0$ derivatives that are absolutely continuous on $[-1,1]$, with the νth derivative of total bounded variation $V < \infty$, then

$$(i)\ \|y_\theta - p_{N-1}\|_\infty \leq \frac{2V}{\pi\nu}(N-1-\nu)^{-\nu}, \quad N > \nu+1,$$

$$(ii)\ |c_j| \leq \frac{2V}{\pi}(j-\nu)^{-(\nu+1)}, \qquad j \geq \nu+1.$$

To bound $\mathcal{P}$, the model manifold of $p_{N-1}(\mathbf{t})$, we note that $p_{N-1}(\mathbf{t}) = X\tilde{\mathbf{c}}$ for $\mathbf{t} = (t_0, \ldots, t_{N-1})^T$, where $X = JD$, with $J_{ij} = T_{j-1}(t_{i-1})$, $D_{jj} = (j-1-\nu)^{-(\nu+1)}$ for $j \geq \nu+2$, with $D_{jj} = 1$ otherwise. Likewise, we set $\tilde{\mathbf{c}} = (\tilde{c}_0, \ldots, \tilde{c}_{N-1})^T$, where $\tilde{c}_j = (j-\nu)^{(\nu+1)} c_j$ for $j \geq \nu+1$, and $\tilde{c}_j = c_j$ otherwise. The singular values of X decay at, at least, an algebraic rate that increases with ν (see Fig. 1). As in the analytic case, one can use X as a linear map and construct a hyperellipsoid H_Y that bounds the model manifold associated with $y_\theta(\mathbf{t})$. Its cross sections are controlled by the singular values of X and typically shrink algebraically fast.

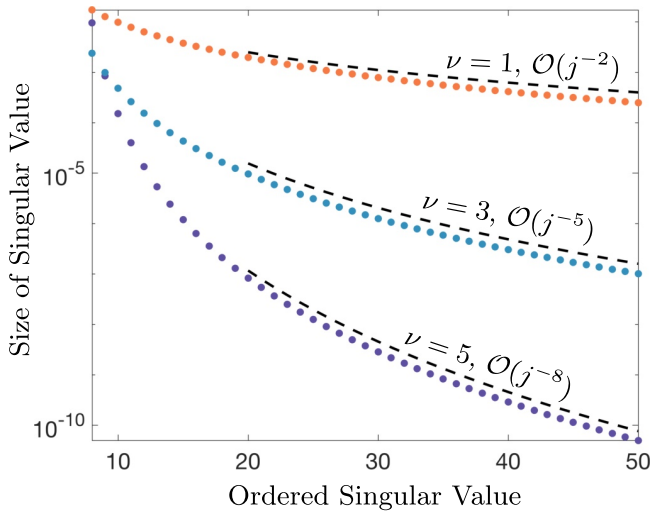

FIG. 1. **The singular values** $\sigma_j(X)$, where X is described in Section I, are plotted on a log scale against the index j for three models of the form $y_\theta(t) = f(\theta)|t|^\nu$: $\nu = 1$ (orange), $\nu = 3$ (blue), and $\nu = 5$ (purple). For simplicity, we assume f is smooth and independent of t. In each case, the model y_θ is ν-times differentiable on $[-1,1]$. The asymptotic decay of the singular values (dotted black lines) is algebraic, with stronger decay rates as ν becomes larger. This suggests that continuously differentiable models have manifolds with (fat) hyperribbon structures, since a ν-times differentiable model y_θ has a manifold enclosed in H_Y, with $\ell_j(H_Y) \approx 2r\sigma_j(X)$ for some constant $r > 0$.

As a question of nomenclature, we suggest that an object with an algebraic decay of widths should also be described as a hyperribbon.

II. NUMERICAL OBSERVATIONS FOR HIGH DIMENSIONAL MANIFOLDS

In the main text, we bounded model predictions $y_\theta(\mathbf{t})$ evaluated at N points $\mathbf{t} = (t_0, \ldots, t_{N-1})^T$ by approximating y_θ with its degree $\leq N$–1 truncated Taylor expansion, which we denote by $p_{N-1}(t;\theta)$. The manifold associated with p_{N-1} is bounded within a hyperellipsoid H_P. The cross-sectional diameters of H_P are defined in terms of the singular values of the column-scaled Vandermonde matrix $X = VD$, where $(VD)_{ij} = t_{i-1}^{j-1} R^{-(j-1)}$. Here, we show how the bound on the hyperellipsoid was obtained, and provide numerical observations for high dimensional manifolds comparing the Chebyshev and monomial (Taylor expansion) bases.

In deriving the bound for the monomial basis, we use the bound on the derivatives from the main text,

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

where where $C > 0$, $R > 1$. It follows that $\|d^k y_\theta / dt^k / k!\|_\infty < C\sqrt{k+1} R^{-k}$ for $k \geq 0$. The Taylor series for y_θ expanded about any point $t \in [-1, 1]$ has a radius of convergence of at least R. One can use the Cauchy integral formula to show that the assumptions in Theorem 1 from the main text implies that the derivative bound in Eq. (2) holds for some C and R dependent on M and ρ. If $t_0 = 0$ and $R > 1$, then we find by simple estimates that

$$\|y - p_{N-1}\|_\infty \leq \frac{C(NR - N + R)}{(1-R)^2} R^{-N+1}. \tag{3}$$

As with the Chebyshev coefficients, we define $\tilde{a}_k = R^k a_k$, and express the polynomial predictions as $P(\theta) = VD\tilde{\mathbf{a}}$, where $V_{ij} = t_{i-1}^{j-1}$ and $D = \mathrm{diag}(R^0, \ldots, R^{-(N-1)})$. While explicit bounds on the singular values of VD can be derived using its displacement structure [2], we require bounds that are characterized by the analyticity of y_θ. For this reason, we instead apply Theorem 2 from the main text to DV^TVD, so that $\sigma_j(VD)$ is bounded in terms of R. By applying the constraint from Eq. (2) to p_{N-1}, we see that $\|\tilde{\mathbf{a}}\|_2 < C\sqrt{N}$. It follows that the polynomial manifold $\mathcal{P}$ is bounded in a hyperellipsoid H_P, where for $j \geq 2$,

$$\ell_j(H_P) = 2C\sqrt{N}\sigma_j(VD) \leq \frac{CN}{\sqrt{R^2-1}} R^{-j+2}, \tag{4}$$

One can conclude the manifold associated with $y_\theta(\mathbf{t})$, is bounded in a hyperellipsoid H_Y with cross-sectional widths obeying

$$\ell_j(H_Y) \leq \ell_j(H_P) + 2\|y_\theta - p_{N-1}\|_\infty.$$

How do these bounds compare to the Chebyshev-based results? The constraint in Eq. (2) implies that y_θ is analytic in the region $\mathcal{R}$ of the complex plane of distance $< R$ from $[-1, 1]$. It can be shown that y_θ must also be analytic and bounded by a function $M(\zeta)$ on any Bernstein ellipse $E_{\rho(\zeta)}$ in $\mathcal{R}$, with $\rho(\zeta) = \zeta + \sqrt{\zeta^2 + 1}$ [3]. The largest such ellipse is given by $\rho_{\max} = R + \sqrt{R^2+1}$, suggesting that Chebyshev-based bounds can improve (4) by nearly a factor of 2^j. However, $M(\zeta)$ is unbounded as $\zeta \to R$, so one must select $0 < \zeta < R$ to minimize the Chebyshev bound. Even when ζ is selected carefully, the conversion from Eq. (2) to a constraint involving $E_{\rho(\zeta)}$ may introduce an unphysically large constant into the bound.

One expects that the decay rate $\mathcal{O}(R^{-j})$ in Eq. (4) is weak as an upper bound on the ordered widths of the underlying hyperribbon $\mathcal{Y}$. This is related to the fact that unlike truncated Chebyshev expansions, truncated Taylor polynomials do not converge to y_θ at a rate that is asymptotically optimal for polynomial approximants (see [1, Ch. 12–16]).

However, we find that the singular values $\sigma_j(VD)$ behave in a surprising way: For small to moderate j, the magnitude of $\sigma_j(VD)$ decays at a rate close to the limit predicted by Chebyshev approximation: $\mathcal{O}(\rho_{\max}^{-j})$, where $\rho_{\max} = R + \sqrt{R^2+1}$. It is only when j is sufficiently large that $\sigma_j(VD)$ appears to decay at the predicted rate $\mathcal{O}(R^{-j})$. We do not yet fully understand why the singular values of VD decay at two distinct rates, but speculate that it may be related to the kink observed in error plots for Clenshaw–Curtis quadrature on analytic functions [4].

Due to this phenomenon, we find that using $\sigma_j(VD)$ directly results in good bounds on model prediction spaces for low dimensions (the larger axes of the hyperellipsoid H_Y). At higher dimensions (shorter hyperellipsoid axes), the Taylor-based bounds become suboptimal, and it is beneficial to instead convert the constraint in (4) to one involving Bernstein ellipses, and then use the Chebyshev-based bounds from Eq. (8) in the main text. The conversion of the constraint can result in bounds that are inflated by a large unphysical constant, but the decay rate in the new bound, close to $\mathcal{O}(\rho_{\max}^{-j})$, is nearly double the rate $\mathcal{O}(R^{-j})$. When viewed together, the Chebyshev-based bounds and numerical Taylor-based bounds describe the successive lengths of the model manifold across two regimes (low vs. high dimension). We illustrate this observation using a high-dimensional manifold ($N = 100$) in Fig. 2.

III. TWO-DIMENSIONAL EXTENSION OF MODEL PREDICTIONS

In this section, we extend the three models used in the main text to the 2D setting. We do this by adding an extra experimental condition, denoted by s, to each model. In Fig. 3, we construct the model manifolds for all three. Just as before, the model manifold is bounded by a hyperellipsoid H_Y with a hierarchy of widths that form a hyperribbon structure.

1. For *exponentials*, we consider temperature depen-

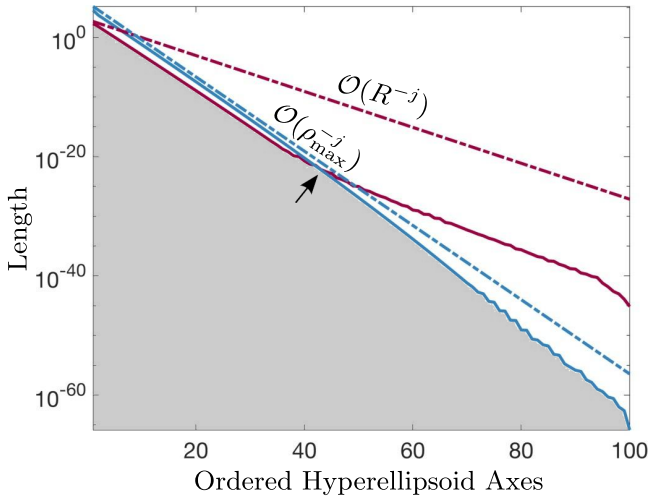

FIG. 2. **Bounds on the hyperellipsoid lengths** $\ell_j(H_P)$ using truncated Taylor (dotted purple) and truncated Chebyshev (dotted blue) expansions are plotted on a log scale against the dimension index j. These form a universal bound on the ordered manifold widths of the prediction space for models y_θ that satisfy Eq. (4). In this example, $C = 1$, $R = 2$, $N = 100$, and $\rho_{\max} \approx 4.2$. The solid lines show the actual computed hyperellipsoid cross-sectional lengths (on a log scale) $\ell_j(H_P) = 2r\sigma_j(X)$, where $X = VD$ for the Taylor-based bounds, and $X_{ij} = T_{j-1}(t_{i-1})\rho_{\max}^{-(j-1)}$ for the Chebyshev-based bounds. The largest 40 Taylor-based hyperellipsoid lengths decay at the rate predicted by the Chebyshev-based bounds. Then, a kink occurs (indicated by a black arrow) and the lengths decay at the rate predicted by the bound in Eq. (4). For the smaller dimensions, the Chebyshev-based results produce tighter bounds. Model manifold lengths outside of the shaded region cannot occur.

dent decay rates,

$$\lambda_\alpha \to \lambda_\alpha \exp\left(-E_\alpha s\right), \tag{5}$$

$$y(t) \to y(t,s) = \sum_\alpha A_\alpha \exp\left(-\lambda_\alpha \exp(-E_\alpha s)t\right), \tag{6}$$

where $s = 1/T$ is inverse temperature.

2. For the model of *reaction velocities*, we consider temperature dependent parameters,

$$\theta_\alpha \to \theta_\alpha \exp\left(-E_\alpha s\right), \tag{7}$$

where again $s = 1/T$ is inverse temperature.

3. Finally, for the *infected population* in an SIR model, we introduce infection and recovery rates that vary continuously with an infection parameter s by introducing

$$\beta \to \beta \exp\left(-E_\beta s\right), \tag{8}$$

$$\gamma \to \gamma \exp\left(-E_\gamma s\right). \tag{9}$$

In all cases, E_α, E_β and E_γ represent activation energies in the respective models. Fig. 3 shows the model manifolds of all three example models, illustrating their hyperribbon structures. To generate these figures, we consider models that obey an analyticity constraint analogous to Eq. (4). Specifically, we assume that for all $0 \le j+k \le N-1$, the following condition holds uniformly in θ for a given $2D$ model $y_\theta(t,s)$:

$$\sum_{j+k\le N-1} \left(\frac{R^{j+k}}{j!k!}\frac{d^{j+k}y_{\boldsymbol{\theta}}(t,s)}{dt^j\,ds^k}\right)^2 < C^2 n. \tag{10}$$

where $R > 1, C > 0$ are constants, and $n = N(N+1)/2$. Under this constraint, it makes sense to bound the prediction space using truncated Taylor expansions of total degree $\le N-1$ for small to moderate N (see the discussion in Section II). This choice results in an $n \times n$ linear system of the form $y_\theta(\mathbf{t},\mathbf{s}) \approx X\tilde{\mathbf{a}}$, where X is a column-scaled 2D Vandermonde matrix, and $\|\tilde{\mathbf{a}}\|_2 < C\sqrt{n}$. The structure of X can be exploited to bound its singular values explicitly [5]. Alternatively, one can apply the 2D analogue to Theorem 2 from the main text to find explicit bounds in terms of R. In Fig. 3, we simply use the relation $\ell_j(H_Y) = \ell_j(H_P) + 2\|y_\theta - p_{N-1}\|_\infty$, and compute $\ell_j(H_P) = 2r\sigma_j(X)$ numerically.

We compare this with the Chebyshev-based bound (established in the following section),

$$\ell_j(H_P) \le \sqrt{N}\frac{3\sqrt{C_2}}{2} n \rho^{-\left\lfloor\sqrt{8(j-1)+1}/2-1/2\right\rfloor}, \tag{11}$$

where ρ is a characteristic length related to the analyticity of the model, $C_2 = (1+\rho^{-2}+\rho^{-4})/(1-\rho^{-2})^3$, and $\lfloor\cdot\rfloor$ represents the floor function. This bound captures the subgeometric decay rate of the model manifold lengths for all three examples, illustrated through the dashed line in Fig. 3.

A. Bounds on the 2D Extension

We can again use polynomial approximation to constrain the geometry of the resulting model manifold $\mathcal{Y}$. In this case, we assume without loss of generality that $(t,s) \in [-1,1]^2$, and we assume y_θ can be expressed as a 2D Chebyshev expansion: $y_\theta(t,s) = \sum_{j=0}^\infty \sum_{k=0}^\infty c_{jk}(\theta)T_{jk}(t,s)$, where $T_{jk}(t,s) = T_j(t)T_k(s)$. The following 2D polynomial of total degree $N-1$ approximates y_θ:

$$p_{N-1}(t,s;\theta) = \sum_{0\le j+k\le N-1} c_{jk}(\theta)T_{jk}(t,s). \tag{12}$$

Let $\rho > 1$ and $M > 0$ be constants. For all fixed choices of $s = s^*$, suppose that the 1D function of t, $y_\theta(t,s^*)$, is analytic in t and bounded $\le M$ uniformly with respect to both s and θ, and that an analogous condition holds for $y_\theta(s,t^*)$. A result similar to Theorem 1 from the main

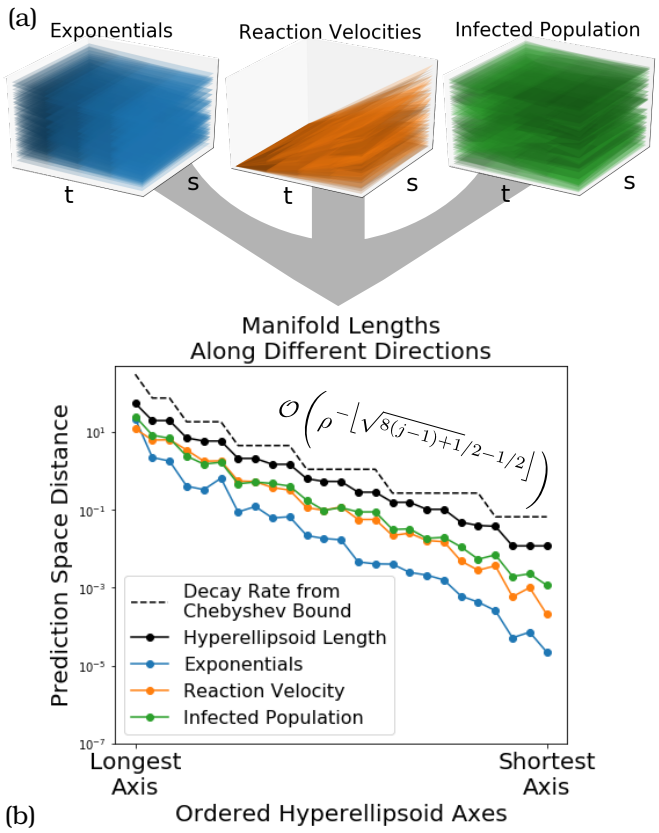

FIG. 3. **Model manifold** of three models with two experimental conditions: (1) exponential decay with temperature dependent decay rates, (2) reaction velocities of an enzyme-catalysed reaction with temperature dependent reaction rates, and (3) the infected population in an SIR model with infection and recovery rates that vary with parameter s. (a) The models are evaluated at 25 equally spaced points $(t_i, s_i) \in [0,1]^2$ (shifted and rescaled from the interval $[-1,1]^2$) with different model parameters. All models obey the analyticity condition in Eq. (10) with C=1 and $R = 2$. (b) The explicit lengths of the three models are shown along the twenty-five axes of the hyperellipsoid H_P. The upper bounds on the possible lengths (black dots) are given by $\ell_j(H_P) = 2C\sqrt{n}\sigma_j(X)$, where X is described in Section III. They exhibit subgeometric decay, with a rate that is captured by the bound in Eq. (15) (dashed line) with $\rho \approx 4.1$. The hierarchy of widths coming from the explicit bounds suggests that the manifolds are hyperribbons.

text can be proven by adapting the ideas in [1, Ch. 8] to the 2D setting. Specifically, we have that

$$(i)\ \|y - p_{N-1}\|_\infty \leq 4MNC_1\rho^{-N+1}, \tag{13}$$

$$(ii)\ |c_{jk}(\theta)| \leq 4M\rho^{-(j+k)}, \tag{14}$$

where $C_1 = (2\rho-1)/(1-\rho)^2$.

As in the 1D case, we study the model manifold $\mathcal{P}$ associated with p_{N-1} as an approximation to $\mathcal{Y}$, the manifold for y_θ. We parameterize $\mathcal{P}$ using a vector of blocks, $P(\theta) = (B_0, \ldots, B_{N-1})^T$, where $B_j = (P_{0j}, P_{1(j-1)}, \ldots, P_{j0})$ and $P_{jk} = p_{N-1}(t_j, s_k; \theta)$. Since each block B_j has $j+1$ entries, $P(\theta)$ is of length $n = N(N+1)/2$. Corresponding vectors of sample locations $\mathbf{t}$ and $\mathbf{s}$ are defined so that $P(\theta) = p_{N-1}(\mathbf{t}, \mathbf{s}; \theta)$.

As before, we exploit the decay of the bounds in Eq. (14) to show that P lies in the range of a matrix with strongly decaying singular values. To see this, define $\tilde{\mathbf{c}}$ as an appropriately ordered $n \times 1$ vector of the scaled coefficients $\tilde{c}_{jk} = \rho^{-(j+k)}c_{jk}$, and form the linear map $P(\theta) = X\tilde{\mathbf{c}}$. Here, $X = [X_{B^0}|\cdots|X_{B^{N-1}}]$, where X_{B^j} is a block of $j+1$ columns scaled by ρ^{-j}. Specifically, $X_{B^j} = \rho^{-j}[T_0(\mathbf{t})T_j(\mathbf{s})\,|T_1(\mathbf{t})T_{j-1}(\mathbf{s})\,|\cdots|\,T_j(\mathbf{t})T_0(\mathbf{s})]$. Since $\tilde{\mathbf{c}}$ is constrained to lie in an n-sphere of radius $4M\sqrt{n}$, the manifold $\mathcal{P}$ is contained in a hyperellipsoid H_P with cross-sectional widths characterized by the singular values of X. One can show that the singular values of X must decay at, at least, a subgeometric rate. An argument similar to the one used in Theorem 2 from the main text shows that for $2 \leq j \leq n$,

$$\sigma_j(X) \leq \frac{3\sqrt{C_2}}{2} n\rho^{-\left\lfloor\sqrt{8(j-1)+1}/2-1/2\right\rfloor}, \tag{15}$$

where $C_2 = (1+\rho^{-2}+\rho^{-4})/(1-\rho^{-2})^3$ and $\lfloor\cdot\rfloor$ represents the floor function. One can use H_P and Eq. (13) to explicitly construct a hyperellipsoid H_Y that must contain $\mathcal{Y}$. We expand our three previous models to the 2D setting in the supplementary materials to illustrate our bounds. While our results are stated in terms of Chebyshev expansions, a similar argument can be made using 2D Taylor expansions, and all of these ideas extend naturally to the multidimensional case.

IV. GENERATING MODEL MANIFOLDS

Here, we provide a detailed description of how data for the 1D models used in the main text were generated. Data for the 2D models in Section III were computed in a similar way. In order to generate the model manifolds, a Monte Carlo sampling was performed on the parameter space of all three models. The model predictions for the randomly selected parameters were accepted or rejected based on whether or not they satisfied the constraint on the derivative from Eq. (4), where we set $C = 1$ and $R = 2$. Since we consider eleven equally spaced points in the main manuscript, in all example models the derivative constraint was applied up to the eleventh derivative.

1. For *exponentials*, the model is of the form

$$y_\theta(t) = \sum_{\alpha=0}^{10} A_\alpha \exp\left(-\lambda_\alpha t\right), \tag{16}$$

and the derivative constraint from Eq. (4) can be expressed as

$$\sum_{k=0}^{N-1}\left(\sum_{\alpha=0}^{10}\frac{R^k A_\alpha}{k!}(-\lambda_\alpha)^k \exp\left(-\lambda_\alpha t\right)\right)^2 < C^2 N \tag{17}$$

for all $-1 \leq t \leq 1$. From a Monte Carlo sampling, 42,000 valid samples were randomly generated. A histogram of parameters used to generate the model manifold is shown in Fig. 4(a).

2. The model of *reaction velocities* is given by

$$y_\theta(t) = \frac{\theta_1 t^2 + \theta_2 t}{t^2 + \theta_3 t + \theta_4}, \tag{18}$$

where t is the substrate concentration. The derivative constraint can be expressed as

$$\sum_{k=1}^{N} \left(\frac{R^k}{k!} \frac{d^k}{dt^k} \left(\frac{\theta_1 t^2 + \theta_2 t}{t^2 + \theta_3 t + \theta_4} \right) \right)^2 < C^2 N, \tag{19}$$

for all $-1 < t < 1$. We generated 24,000 valid parameter combinations, and a histogram of the different parameter values is shown in Fig. 4(b).

3. Finally, for the *infected population* in an SIR model, the number of people susceptible (S), infected (I), and recovered (R) are determined through three coupled differential equations:

$$(i)\ \dot{S} = -\beta \frac{IS}{N_{tot}},$$
$$(ii)\ \dot{I} = \beta \frac{IS}{N_{tot}} - \gamma I,$$
$$(iii)\ \dot{R} = \gamma I,$$

where β is the infection rate, γ is the recovery rate, and N_{tot} is the total size of the population. If we let the model predictions be the infected population, then we have $y_\theta(t) = I(t)$. To find the kth derivative of such a model, we note that $\dot{S} = f_1(S, I)$ and $\dot{I} = g_1(S, I)$. The subsequent derivatives can therefore be found recursively, by $\ddot{y}_\theta = \ddot{I} = \frac{dg_1}{dS}\dot{S} + \frac{dg_1}{dI}\dot{I} = g_2(S, I)$ and so on. From a Monte Carlo sampling, we obtained 20,000 valid parameter combinations. A histogram of parameter values used to generate the model manifold is shown in Fig. 4(c).

In all three models, the smallest physically meaningful prediction is $y_\theta(t) = 0$. For exponentials and the SIR model, the largest physically meaningful prediction allowed by Eq. (4) is $y_\theta(t) = C\sqrt{N}$, and so the longest manifold distance possible is CN. With this sampling method, we obtained manifold lengths that are within 1.5% of this maximally allowed distance, and so while more refined sampling methods could be used to resolve the manifold boundaries, they are unnecessary for our purposes.

Once a sampling of the possible parameter combinations is obtained for a model, we visualize it. Each parameter combination is evaluated at eleven equally spaced points. The space spanned by the model predictions at these points forms the model manifold $\mathcal{Y}$.

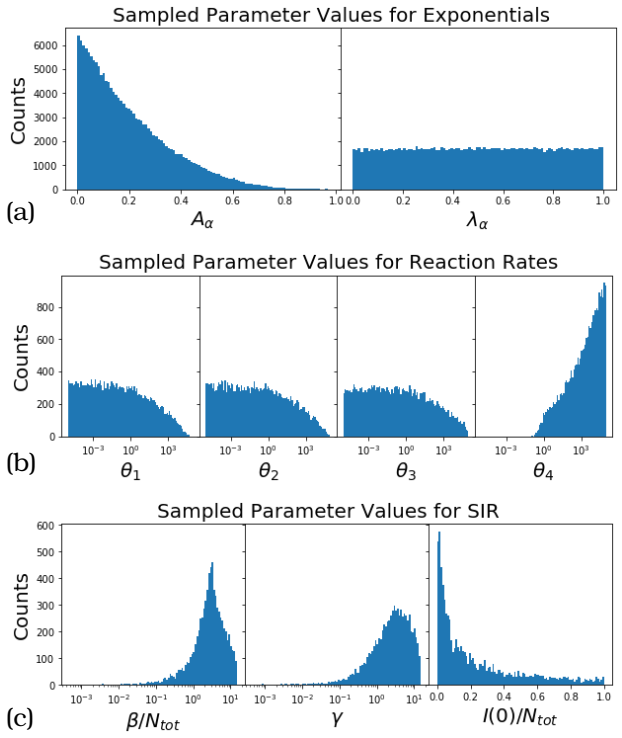

FIG. 4. **Histograms of valid parameter values** used to generate the model manifolds. In all the models, a Monte Carlo sampling was performed, with parameters accepted or rejected based on whether or not they satisfied the derivative condition from Eq. (4). (a) Parameter values for exponentials, showing the distributions for the amplitudes A_α and decay rates λ_α. (b) Parameter values for the reaction velocities, for each θ_1, θ_2, θ_3 and θ_4. (c) Parameter values for the SIR epidemiology model, showing the distribution of infection rates β/N_{tot}, recovery rates γ and initial infected population.

To visualize $\mathcal{Y}$, it is rotated into the basis given by the hyperellpsoid axes constructed from the space of allowed polynomials predictions, $\mathcal{P}$. Let $\{\phi_j\}_{j=0}^{\infty}$ be a complete polynomial basis, and let $P(\mathbf{b}) = (P_0, \ldots, P_{N-1})$ define the model manifold $\mathcal{P}$ of $p_{N-1}(t) = \sum_{j=0}^{N-1} b_j \phi_j(t)$. Polynomial predictions are given by $P_k = p_{N-1}(t_k)$. By definition, $P(\mathbf{b}) = X\mathbf{b}$, where $X_{ij} = \phi_{j-1}(t_{i-1})$ and $\mathbf{b} = (b_0, \ldots, b_{N-1})^T$. To find the rotation matrix used to visualize the model manifold $\mathcal{Y}$, we perform a singular value decomposition on X,

$$X = U\Sigma V^T, \tag{20}$$

to extract the rotation matrix U. The data points on the model manifold are then rotated using this matrix, and visualized in Fig. I(b) in the main text where we set $X = VD$ to be the column-scaled Vandermonde matrix.

[1] L. N. Trefethen, *Approximation Theory and Approximation Practice* (SIAM, 2013).
[2] B. Beckermann and A. Townsend, SIAM J. Matrix Anal. & Appl. **38**, 1227 (2017).
[3] L. Demanet and A. Townsend, Found. Comput. Math. (2018), 10.1007/s10208-018-9384-1.
[4] J. Weideman and L. N. Trefethen, Numerische Mathematik **107**, 707 (2007).
[5] A. Townsend and H. Wilber, Lin. Alg. & Appl. **548**, 19 (2018).