Chebyshev Approximation and the Global Geometry of Model Predictions

Katherine N. Quinn, Heather Wilber, Alex Townsend, and James P. Sethna

Phys. Rev. Lett. **122**, 158302 — Published 18 April 2019

DOI: 10.1103/PhysRevLett.122.158302

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)

Complex nonlinear models are typically ill-conditioned or *sloppy*; their predictions are significantly affected by only a small subset of parameter combinations, and parameters are difficult to reconstruct from model behavior. Despite forming an important universality class and arising frequently in practice when performing a nonlinear fit to data, formal and systematic explanations of sloppiness are lacking. By unifying geometric interpretations of sloppiness with Chebyshev approximation theory, we rigorously explain sloppiness as a consequence of model smoothness. Our approach results in universal bounds on model predictions for classes of smooth models, capturing global geometric features that are intrinsic to their model manifolds, and characterizing a universality class of models. We illustrate this universality using three models from disparate fields (physics, chemistry, biology): exponential curves, reaction rates from an enzyme-catalysed chemical reaction, and an epidemiology model of an infected population.

Complex nonlinear models used to simulate and predict experimentally observed phenomena often exhibit a structural hierarchy: Perturbing a few model parameter combinations drastically impacts predictions, whereas most others can vary widely without effect. Such ill-conditioned models are called *sloppy*. Sloppy models appear to be common, arising in many areas of physics. In critical phenomena, this hierarchy of importance explains the parameter scaling with coarsening for diffusion and the Ising model of magnetism [1]. In accelerator physics, linear combinations of the multitude of tunable beam-line parameters exhibit a geometric hierarchy of importance [2]. Exponential curve fitting, a notoriously ill-conditioned problem, poses a significant challenge, *e.g.* finding correlators in lattice QCD [3, 4]. Sloppy models are not confined to physics, and in fact appear in systems biology [5–7], insect flight [8], power systems [9, 10], machine learning [11], and many other areas [12]. Understanding why sloppiness occurs can therefore connect models used across disparate fields.

There are many well-studied cases for insensitivity of model predictions to particular combinations of parameters. *Structural identifiability* describes systems for which parameters can be analytically exchanged for one another [13, 14]. Separation of scales, singular perturbations, and continuum limits can make the behavior at a particular time or distance region depend only on a subset of the underlying parameters [15–17]. Universal critical behavior can yield effective parameter compression on long length scales near continuous transitions [1]. However, these comprehensible sources of sloppiness do not explain the generality of the phenomenon, nor do they offer a rigorous framework by which to quantify the hierarchy of parameter importance. In this paper, we address the generic sloppiness of multiparameter nonlinear models in the absence of particular mechanisms or small parameters. We unify recently developed geometric descriptions of sloppiness [12] with classical ideas from polynomial approximation theory [18]. We posit that in many cases, sloppiness is fundamentally linked to the smoothness of the underlying model, and provide a rigorous description of this connection.

The hierarchy of parameter importance that characterizes sloppy models manifests geometrically. Given some model, the space of all possible predictions for all input parameters forms a geometric object know as the *model manifold* (Fig. 1(a)), whose metric is given by the Fisher Information (a measure of the distinguishability between predictions [19, Ch. 2] which sets a lower bound on the possible variance of parameter estimates for an unbiased prior through the Cramér–Rao bound). Studying the geometry of model manifolds yields fruitful information for several reasons: (1) the dominant components reflect emergent behavior of the models (how the microscopic interactions do or do not produce macroscopic behavior [1]), (2) the boundaries represent reduced-model approximations [20], and (3) knowledge of the manifold geometry leads to more efficient data fitting methods [21]. Model manifolds typically form striking *hyperribbons* [22], so-called because, like ribbons, successive widths follow a geometric decay: They are much longer than they are wide, much wider than they are thick, etc., yielding effective low-dimensional representations. Because directions along the model manifold correspond to specific parameter combinations, there is a direct connection between the hyperribbon nature of model man-

ifolds and the structural hierarchy of model parameters. Understanding why model manifolds form hyperribbons therefore leads to an understanding of why this structural hierarchy in parameter importance exists.

Consider a nonlinear model that depends continuously on K input parameters $\theta = (\theta^1, \ldots, \theta^K)$ to generate predictions $y_\theta(t)$. If we consider the model predictions at N fixed points, $\{t_0, \ldots, t_{N-1}\}$, then our predictions for parameters θ form an N-dimensional vector $Y(\theta) = (y_\theta(t_0), \ldots, y_\theta(t_{N-1}) = (Y_0, \ldots, Y_{N-1})$. We use $\mathcal{Y}$ to represent the model manifold, defined as the space of all possible predictions for all possible parameter combination (so all allowed $Y(\theta)$). Specifically, model manifold $\mathcal{Y}$ is a K-dimensional surface embedded in an N-dimensional prediction space.

To bound the model manifold $\mathcal{Y}$ and study its geometry, we consider polynomial approximations of model y_θ. Without loss of generality, we shift and rescale the points so that $\{t_k\}_{k=0}^{N-1} \subset [-1, 1]$. Let $\{\phi_j\}_{j=0}^{\infty}$ be a complete polynomial basis, and suppose that model $y_\theta(t)$ is decomposed into this basis: $y_\theta(t) = \sum_{j=0}^{\infty} b_j(\theta)\phi_j(t)$. Let $p_{N-1}(t;\theta)$ be the *truncated* series representing the polynomial approximation to model $y_\theta(t)$. Note that the truncation is set by the number of sampled points, N. We can view the coefficients $(b_0(\theta), \ldots, b_{N-1}(\theta))$ as a set of N parameters. Now, let $P(\theta) = (p_{N-1}(t_0), \ldots, p_{N-1}(t_{N-1})) = (P_0, \ldots, P_{N-1})$ define the polynomial manifold $\mathcal{P}$. Thus, we have model manifold $\mathcal{Y}$ and a polynomial manifold $\mathcal{P}$.

By definition, $P(\theta) = X\mathbf{b}$, where $X_{ij} = \phi_{j-1}(t_{i-1})$ and $\mathbf{b} = (b_0, \ldots, b_{N-1})^T$. Here, X forms a linear map from the space of polynomial coefficients to the space of possible predictions, and is determined by the chosen polynomial basis and fixed points t_i. The singular values of X can be used to understand the hyperribbon structure of the polynomial manifold $\mathcal{P}$. Suppose, for example, that $\|\mathbf{b}\|_2 < r$, so that the coefficient space is bounded in S, an n-sphere of radius r. The action of X on S distorts it into a hyperellipsoid H_P. If $\ell_j(H_P)$ is the diameter of the jth cross-section of hyperellipsoid H_P, then

$$\ell_j(H_P) = 2r\sigma_j(X), \tag{1}$$

where $\sigma_j(X)$ are the ordered singular values of X. When X has rapidly decaying singular values, H_P has a hyperribbon structure because there is a strict hierarchy in successive widths. Accounting for the polynomial approximation error $\|y_\theta - p_{N-1}\|_\infty$, where $\|\cdot\|_\infty$ is the L^∞ norm on $[-1, 1]$, we can define a hyperellipsoid H_Y that must enclose model manifold $\mathcal{Y}$, where the cross-sectional widths are given by

$$\ell_j(H_Y) = 2r\sigma_j(X) + 2\|y - p_{N-1}\|_\infty. \tag{2}$$

In this way, we find that *any* model manifold $\mathcal{Y}$ is bounded within a hyperribbon whenever $\sigma_j(X)$ decays geometrically and $\|y - p_{N-1}\|_\infty$ is small enough. A fundamental question is whether it matters which polynomial basis or which set of time points are chosen to define H_P and H_Y. The hyperribbon structure of $\mathcal{Y}$, of course, does not depend on our representation of y_θ, but rather on intrinsic properties of the model, such as its smoothness. For example, if for every $t_0 \in [-1, 1]$, the Taylor expansion of y_θ at t_0 has a large enough radius of convergence, any sequence of polynomial interpolants with N distinct interpolating points converges to y_θ at a geometric rate with N [18]. This fact underpins the qualitative observation in [12, 22] that certain analytic models have manifolds bounded within hyperribbons. Here we make that observation rigorous. We consider two such choices. First, we choose our basis functions $\{\phi_j\}_{j=0}^{\infty}$ as the Chebyshev polynomials. Truncated Chebyshev expansions converge to y_θ at an asymptotically optimal rate for polynomial approximation [18]. As we show below, this rate controls the magnitude of $\sigma_j(X)$ in Eq. (2), and can be used to explicitly bound the cross-sectional widths of H_Y. Our bounds deliver an outright description of a hyperribbon that must contain $\mathcal{Y}$.

We also analyze the case where $\{\phi_j\}_{j=0}^{\infty}$ are the monomials and p_{N-1} is the truncated Taylor series expansion of y_θ. In this case, we observe that the numerical computation of $\sigma_j(X)$ results in excellent practical and universal bounds on the prediction space for large classes of models.

Chebyshev expansions. Suppose that y_θ has a convergent Chebyshev expansion, so that it is given by $y_\theta(t) = \sum_{j=0}^{\infty} c_j(\theta)T_j(t)$, where $T_j(t) = \cos(j \arccos t)$ is the degree j Chebyshev polynomial [18, Ch. 3]. We can approximate y_θ with a degree $\leq N-1$ polynomial by truncating the Chebyshev series after N terms:

$$p_{N-1}(t;\theta) = \sum_{j=0}^{N-1} c_j(\theta)T_j(t). \tag{3}$$

Truncated Chebyshev expansions have near-best global approximation properties, and explicit bounds on $\|y_\theta - p_{N-1}\|_\infty$ are known when y_θ is sufficiently smooth.

We first consider the case where y_θ is analytic in an open neighborhood of $[-1, 1]$. Such a region contains a *Bernstein ellipse* E_ρ, defined as the image of the circle $|z| = \rho$ under the Joukowsky mapping $(z + z^{-1})/2$. It has foci at ± 1, and the lengths of its semi-major and semi-minor axes sum to ρ. The polynomial in Eq. (3) converges to y_θ as $N \to \infty$ at a rate determined by ρ:

Theorem 1. *Let $M > 0$ and $\rho > 1$ be constants and suppose that $y_\theta(t)$, $t \in [-1, 1]$, is analytically continuable to the region enclosed by the Bernstein ellipse E_ρ, with $|y_\theta| \leq M$ in E_ρ, uniformly in θ. Let $p_{N-1}(t;\theta)$ be as in*

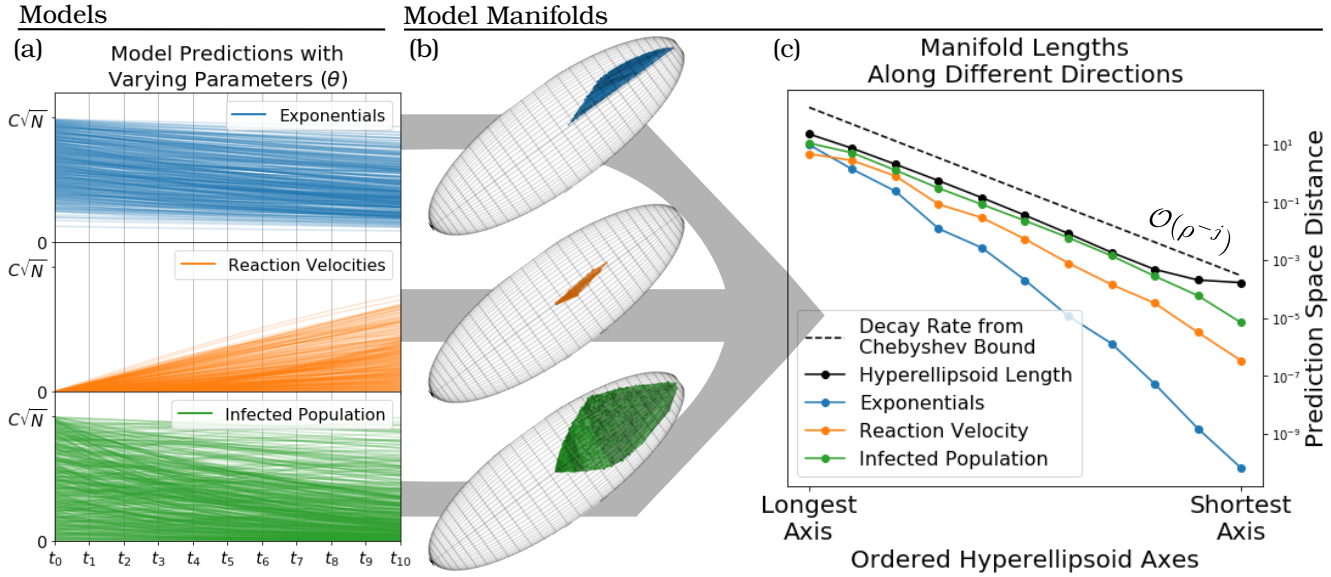

FIG. 1. **Model manifold** of three disparate models: (1) exponential curves, (2) reaction velocities of an enzyme-catalysed reaction, and (3) the infected population in an SIR model. The models are evaluated at 11 equally spaced points on $[0, 1]$, and obey the smoothness condition in Eq. (10), with $C = 1$ and $R = 2$. (a) An illustration of each model, where each line represents the respective model predictions with a different set of parameters. (b) The *model manifolds* are all bounded by the same hyperellipsoid, and so the two axes represent the first and second longest hyperellipsoid axes. Note that, in all three models, only values greater than 0 are physically significant. This constraint manifests itself geometrically through their *location* in the hyperellipsoid. (c) The lengths of each model manifold along the eleven axes of the hyperellipsoid H_P in Eq. (11). Black points are the numerically computed lengths of H_P, given by $2C\sqrt{N}\sigma_j(VD)$ in Eq. (11), and include the error term from Eq. (2) (note the kink at the second to last point), forming an upper bound on possible lengths of the manifolds. The explicit decay rate of the Chebyshev-based bound (black dotted line) is based on the fact that models obeying Eq. (10) are analytic in the ellipse $E_\rho(\zeta)$. (Here, $\rho(\zeta) \approx 3.81$.) It captures the decay rate of $\sigma_j(VD)$ for $j < 11$, and closely follows the true decay rate in the successive widths of the various manifolds.

Eq. (3). *Then,*

$$(i)\ \ \|y_\theta - p_{N-1}\|_\infty \leq \frac{2M\rho^{-N+1}}{\rho - 1}, \tag{4}$$

$$(ii)\ |c_0| \leq M,\ |c_j(\theta)| \leq 2M\rho^{-j}, \qquad j \geq 1. \tag{5}$$

Proof. For a proof, see Theorem 8.2 in [18]. □

To exploit the decay of the coefficients in Eq. (5), we define modified coefficients $\tilde{c}_j = \rho^j c_j$. We then have that polynomial predictions $P(\theta) = X\tilde{\mathbf{c}}$, where $X = JD$, $J_{ij} = T_{j-1}(t_{i-1})$, and D is diagonal with entries $D_{jj} = \rho^{-(j-1)}$. By (5), we have that $\|\tilde{\mathbf{c}}\|_2 < 4M\sqrt{4N-3}$. This implies that the polynomial manifold $\mathcal{P}$ is bound in a hyperellipsoid H_P. By Eq (1), we have that $\ell_j(H_P) = 8M\sqrt{4N-3}\sigma_j(X)$. To bound $\sigma_j(X)$ explicitly, we first prove a conjecture proposed in [23]:

Theorem 2. *Let $S \in \mathbb{R}^{N\times N}$ be symmetric and positive definite. Let $E \in \mathbb{R}^{N\times N}$ be diagonal with $E_{ii} = \epsilon^{i-1}$ and $0 < \epsilon < 1$. If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ are the ordered eigenvalues of ESE, then $\lambda_{m+1} = \mathcal{O}(\epsilon^{2m})$. Specifically,*

$$\lambda_{m+1} \leq \frac{\epsilon^{2m}}{1-\epsilon^2} \max_{1\leq j,k\leq N} |S_{jk}|, \qquad 1 \leq m \leq N-1. \tag{6}$$

Proof. [24] Consider the rank m matrix

$$S_m = S(:, 1:m)S(1:m, 1:m)^{-1}S(1:m, :), \tag{7}$$

where $1 \leq m \leq N$–1, and the notation $M(:, 1:m)$ denotes the submatrix of M consisting of its first m columns. Clearly, S_m is well-defined because $S(1:m, 1:m)$ is a principal minor of a positive definite matrix and is therefore invertible. Moreover, it can be verified that $(S - S_m)_{jk} = 0$ for $1 \leq j, k \leq m$.

Since ESE is positive definite and $\text{rank}(S_m) = m$, we know that $\lambda_{m+1} \leq \|E(S - S_m)E\|_2$, where $\|\cdot\|_2$ denotes the spectral matrix norm [25, Ch. 2]. Using $\|\cdot\|_F$ to denote the Frobenius norm, we have

$$\begin{aligned}
\lambda_{m+1}^2 &\leq \|E(S - S_m)E\|_2^2 \leq \|E(S - S_m)E\|_F^2 \\
&= \sum_{j=m+1}^{N} \sum_{k=m+1}^{N} \epsilon^{2(j-1)+2(k-1)} |S_{jk} - (S_m)_{jk}|^2 \\
&\leq \frac{\epsilon^{4m}}{(1-\epsilon^2)^2} \max_{1\leq j,k\leq N} |S_{jk} - (S_m)_{jk}|^2 \\
&\leq \frac{\epsilon^{4m}}{(1-\epsilon^2)^2} \max_{1\leq j,k\leq N} |S_{jk}|^2,
\end{aligned}$$

where the last inequality comes from the fact that the

block $S(m{+}1{:}N, m{+}1{:}N) - S_m(m{+}1{:}N, m{+}1{:}N)$ is the Schur complement of $S(1{:}m, 1{:}m)$ in S [25]. □

Applying Theorem 2 to $X^TX = DJ^TJD$, we have that for $j > 1$, $\sigma_j(X) \leq \sqrt{N}\rho^{-j+2}/\sqrt{\rho^2-1}$, where we have used the fact that $|T_k(t)| \leq 1$ for $k \geq 0$ and $-1 \leq t \leq 1$. It follows from Equations (2) and (4) that predictions for $y_\theta(t)$ are bounded by a hyperellipsoid H_Y, with

$$\ell_j(H_Y) \leq \frac{2M\sqrt{4N^2-3N}\rho^{-j+2}}{\sqrt{\rho^2-1}} + \frac{4M\rho^{-N+1}}{\rho-1}, \quad (8)$$

for $2 \leq j \leq N$, i.e.,

$$\ell_j(H_Y) = \mathcal{O}(\rho^{-j} + \rho^{-N}). \quad (9)$$

These bounds indicate that the hyperribbon structure of H_Y is controlled by ρ, a parameter characterizing the analyticity of the model. As ρ becomes larger, bounds on the widths of the successive cross-sections of H_Y must decay more rapidly: In principle, H_Y becomes successively thinner and more ribbon-like.

When y_θ is not analytic on an open neighborhood of $[-1, 1]$, the decay rate of $\sigma_j(JD)$ is instead controlled by the smoothness of y_θ on $[-1, 1]$. Furthermore, when we consider models with two experimental conditions (for instance, time and temperature) these bounds can be extended to the two-dimensional case. We provide more discussion of non-analytic and two-dimensional cases in the supplementary materials.

Taylor expansions. The degree $N{-}1$ truncated Taylor polynomial of y_θ is $p_{N-1}(t) = \sum_{k=0}^{N-1} a_k(\theta)(t-t_0)^j$, where $a_k(\theta) = y_\theta^{(k)}(t_0)/k!$. We describe the analyticity of y_θ using the following condition: For all $N \geq 1$,

$$\sum_{k=0}^{N-1} \left(\frac{R^k}{k!} \frac{d^k y_\theta(t)}{dt^k} \right)^2 < C^2 N, \quad (10)$$

where $C > 0$,$R > 1$ are constants in θ. A straightforward but tedious calculation outlined in the supplemental material [26] shows that the lengths of the resulting hyperellipsoid are given by

$$\ell_j(H_P) \leq \frac{2CN}{\sqrt{R^2-1}} R^{-j+2}. \quad (11)$$

To apply our results, we selected three models from quite disparate fields (physics, chemistry, biology). This was done deliberately, to illustrate the universal nature of our results. In all three cases, the context for model construction is different, and yet the underlying smoothness of each can be used to relate them to a single universal bound.

1. *Exponential curves*, such as for radioactive decay [12, 21] and calculating correlators in lattice QCD [3, 4]. Here, we set $y_\theta(t) = \sum_{\alpha=0}^{10} A_\alpha \exp(-\lambda_\alpha t)$, where model parameters are the amplitudes A_α and decay rates λ_α, and t represents time.
2. *Reaction velocities* of an enzyme-catalysed chemical reaction [27, 28]. This model can be expressed as $y_\theta(t) = (\theta_1 t^2 + \theta_2 t)/(t^2 + \theta_3 t + \theta_4)$ [12], where t represents the substrate concentration. This model stands in for steady-state behavior of complex chemical reaction networks in engineering and ecology [29].
3. *The infected fraction of a population* in an SIR epidemiology model [30]. This model predicts the size of a population that is susceptible to infection ($S(t)$), infected ($I(t)$), and recovered from infection ($R(t)$). These are expressed through three coupled differential equations: $\dot{S} = -\beta IS/N_{tot}$, $\dot{I} = \beta IS/N_{tot} - \gamma I$, and $\dot{R} = \gamma I$, where model parameters β and γ represent the rates of infection and recovery, and additional parameters include the total population N_{tot}, and initial infected and recovered population. At all times, $S(t)$, $I(t)$ and $R(t)$ sum to N_{tot}, and we set $y_\theta(t) = I(t)$. This model serves to represent classes of models involving numerical ODEs, which occur in power systems, *e.g.* for systems biology [5, 6] and power systems [31, 32].

The model manifolds for these three models are shown in Fig. 1. They are all contained within the *same* hyperellipsoid, as shown in Fig. 1(b), and so share the *same* universal bound. The hyperribbon structure of the manifolds is accurately captured by the numerical bound from Eq. (11), and the decay in successive manifold widths are clearly captured by the Chebyshev rate from Eq. (8). These three models were derived in very different contexts and exhibit what would appear to be fundamentally different properties, yet they all share a fundamental property: in all cases, there is a structural hierarchy in their model manifolds as determined by a *universal* bound. Because of the geometric decay in successive manifold widths, low-dimensional representations (as determined by the longest directions) capture the large variance in model predictions. This is because they are all part of the same universality class, that of sloppy models.

Our results explain a fundamental feature of the global geometry of sloppy models, and establish a rigorous framework that explains the role of model smoothness in the observation of sloppiness. An important implication of our results is that *any* model that satisfies the smoothness condition in Eq. 10 is *guaranteed* to be bounded in a manifold that exhibits this hierarchical structure. As such, it serves as a natural test of sloppiness. The implications of sharper bounds that depend on time-points are the focus of future work, as they open up far-ranging applications in optimizing the experimental design to focus data collection at time-points that maximize information extraction by minimizing the decay rate in hyperribbon widths. Furthermore, sloppy features appear in probabilistic models (such as the Ising Model of atomic spins in statistical physics and the dark energy cold dark matter ΛCDM cosmological predictions of the cosmic microwave

background) and so an extension of this approach is currently underway to explain all general, probabilistic models.

Acknowledgments. We thank Mark Transtrum for suggestions related to selecting models used in this letter, John Guckenheimer for suggesting that there could be a connection between the third and fourth author's research areas, and Peter Lepage for his expertise and insight into the connection with lattice QCD. KNQ was supported by a fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC), and JPS and KNQ were supported by the National Science Foundation (NSF) through grant DMR-1719490. AT was supported by NSF grant no. DMS-1818757, and HW was supported by NSF grant no. DGE-1650441.

[1] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, Science **342**, 604 (2013).
[2] R. Gutenkunst, *Sloppiness, modeling, and evolution in biochemical networks*, Ph.D. thesis, Cornell University (2007).
[3] G. Lepage, B. Clark, C. Davies, K. Hornbostel, P. Mackenzie, C. Morningstar, and H. Trottier, Nuclear Physics B - Proceedings Supplements **106-107**, 12 (2002), lATTICE 2001 Proceedings of the XIXth International Symposium on Lattice Field Theory.
[4] K. Hornbostel, G. P. Lepage, C. T. H. Davies, R. J. Dowdall, H. Na, and J. Shigemitsu (HPQCD collaboration), Phys. Rev. D **85**, 031504 (2012).
[5] K. S. Brown and J. P. Sethna, Phys. Rev. E **68**, 021904 (2003).
[6] K. Brown, C. Hill, C. Calero, C. Myers, K. Lee, J. Sethna, and R. A. Cerione, Phys. Biol. **1** (2004).
[7] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, PLOS Comput. Bio **3**, 1 (2007).
[8] G. Berman and Z. Wang, J. Fluid Mech. **582** (2007).
[9] M. K. Transtrum, A. T. Sari, and A. M. Stankovi, IEEE Transactions on Power Systems **32**, 2243 (2017).
[10] M. K. Transtrum, A. T. Sari, and A. M. Stankovi, IEEE Transactions on Power Systems **33**, 440 (2018).
[11] R. Pascanu, T. Mikolov, and Y. Bengio, in *International Conference on Machine Learning* (2013) pp. 1310–1318.
[12] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, J. Chem. Phys **143** (2015), 10.1063/1.4923066, 1501.07668.
[13] O.-T. Chis, J. R. Banga, and E. Balsa-Canto, PloS one **6**, e27755 (2011).
[14] D. V. Raman, J. Anderson, and A. Papachristodoulou, Phys. Rev. E **95**, 032314 (2017).
[15] A. Kavouras, C. Georgakis, C. T. Kelley, C. Siettos, and I. G. Kevrekidis, Aiche Journal **59**, 3308 (2013).
[16] C. J. Dsilva, R. Talmon, N. Rabin, R. R. Coifman, and I. G. Kevrekidis, Journal of Chemical Physics **139**, 184109 (2013).
[17] P. Liu, H. R. Safford, I. D. Couzin, and I. G. Kevrekidis, Computational Particle Mechanics **1**, 425 (2014).
[18] L. N. Trefethen, *Approximation Theory and Approximation Practice* (SIAM, 2013).
[19] S.-I. Amari, *Information geometry and its applications* (Springer).
[20] M. K. Transtrum and P. Qiu, Phys. Rev. Lett. **113**, 098701 (2014).
[21] M. K. Transtrum, B. B. Machta, and J. P. Sethna, Phys. Rev. E **83**, 036701 (2011).
[22] M. K. Transtrum, B. B. Machta, and J. P. Sethna, Phys. Rev. Lett. **104**, 060201 (2010).
[23] J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, PRL **97**, 150601 (2006).
[24] Previous proofs with weaker bounds were provided through private communications with Ari Turner and Yaming Yu.
[25] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1996).
[26] See Supplemental Material for bounds on non-analytical models, numerical results for high-dimensional manifolds, proof for the truncated Taylor series, extensions to two-dimensional experimental conditions, and a full description of how the visualizations of the model manifolds were achieved. The Supplemental Material includes Refs. [18, 33–36].
[27] B. Averick, J. M. Carter, and G. Xue, Preprint MCS-P153-0694, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois (1992).
[28] J. Kowalik and J. Morrison, Math. Biosci. **2** (1968).
[29] M. Mehta, R. O. Fox, and P. Pepiot, Industrial & Engineering Chemistry Research **54**, 5407 (2015), https://doi.org/10.1021/acs.iecr.5b00130.
[30] H. W. Hethcote, SIAM Review **42**, 599 (2000).
[31] M. K. Transtrum, A. T. Sari, and A. M. Stankovi, IEEE Transactions on Power Systems **33**, 440 (2018).
[32] M. K. Transtrum, A. T. Sari, and A. M. Stankovi, IEEE Transactions on Power Systems **32**, 2243 (2017).
[33] B. Beckermann and A. Townsend, SIAM J. Matrix Anal. & Appl. **38**, 1227 (2017).
[34] L. Demanet and A. Townsend, Found. Comput. Math. (2018), 10.1007/s10208-018-9384-1.
[35] J. Weideman and L. N. Trefethen, Numerische Mathematik **107**, 707 (2007).
[36] A. Townsend and H. Wilber, Lin. Alg. & Appl. **548**, 19 (2018).