

Selecting simple, transferable models with the supremum principle

Cody Petrie¹,* Christian Anderson¹, Casie Maekawa, Travis Maekawa¹, and Mark K. Transtrum¹†

Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA



(Received 23 September 2021; revised 11 February 2022; accepted 1 June 2022; published 23 September 2022)

We consider how mathematical models enable predictions for conditions that are qualitatively different from the training data. We propose techniques based on information topology to find models that can apply their learning in regimes for which there is no data. The first step is to use the manifold boundary approximation method to construct simple, reduced models of target phenomena in a data-driven way. We consider the set of all such reduced models and use the topological relationships among them to reason about model selection for new, unobserved phenomena. Given minimal models for several target behaviors, we introduce the *supremum principle* as a criterion for selecting a new, transferable model. The supremal model, i.e., the least upper bound, is the simplest model that reduces to each of the target behaviors. We illustrate how to discover supremal models with several examples; in each case, the supremal model unifies causal mechanisms to transfer successfully to new target domains. We use these examples to motivate a general algorithm that has formal connections to theories of analogical reasoning in cognitive psychology.

DOI: [10.1103/PhysRevResearch.4.L032044](https://doi.org/10.1103/PhysRevResearch.4.L032044)

One of the first important tasks in modeling data is selecting the form for a mathematical model. The form of the model defines the types of predictions a model can make and therefore—accurately or not—creates a type of “hypothesis space” called *inductive bias* [1]. In this Letter, we use the geometric and topological relationships among candidate models to reason about inductive bias and model selection. Of particular interest are predictions for qualitatively different conditions than those on which a model was trained, such as predicting a time series outside of the range of sampled time points, predicting under different experimental conditions, or applying insights from two populations to a third. A model’s ability to make such *out-of-domain* predictions is sometimes known as *transferability*, which is stronger than generalization, i.e., predicting data generated for inputs similar to those on which it was trained [2]. We propose a general principle of model selection, the *supremum principle*, that encodes a preference for simplicity with respect to target quantities of interest while enabling model transferability and whose construction uses topological relationships formally equivalent to models of human analogical reasoning.

One of the key struggles of model selection is balancing inductive bias against model flexibility. Consider, for example, explaining the change in a cell’s state (e.g., healthy to cancerous) in terms of the proteome. A potential hypothesis space could include all possible interactions between all $\sim 25\,000$

known proteins. This has very little bias since the correct explanation is somewhere in this space; however, it would require an unreasonable amount of data to learn all the parameters of such a complex model. Furthermore, it would be even more difficult to interpret the model afterwards as most of the interactions are merely explanatory noise relative to the phenomenon of interest [3]. Therefore, we seek to restrict the hypothesis space to the one that minimally includes our behavioral regimes of interest. Such a model does not fit both states with one set of parameters, rather, it fits either set of data independently, i.e., some parameters could be unidentifiable to data from either state. In addition to describing cells in either state, this model *predicts* a mechanism for switching between them.

A minimal criterion for a useful predictive model is that it reproduces the training data within statistical noise, that is, a kind of coarse interpolation. Common statistical practices such as holdout, jackknife, and cross-validation reinforce this intuition. Sloppy models [4–7], a class of overparametrized models, further formalize the relation between prediction and interpolation using information geometry [8,9]. The predictions of sloppy models are controlled by only a few *stiff* parameter combinations and so are said to have a *low effective dimensionality* [9,10]. Effective dimensionality is quantified in terms of widths of a model manifold, rigorous bounds for which are given by theorems from interpolation theory [9,11]. Indeed, it has been suggested that predictive models are generalized interpolation schemes [12].

However, there is a sense that more than simple interpolation ought to be possible [13,14]. Human cognition is driven by understanding, rather than mere pattern mimicry. When we reason about molecular bonds as if they were balls and springs, we use analogical reasoning to identify abstract relationships and transfer insights among superficially different systems. Can machines similarly analogize to make

*codypetrie89@gmail.com

†mktranstrum@byu.edu

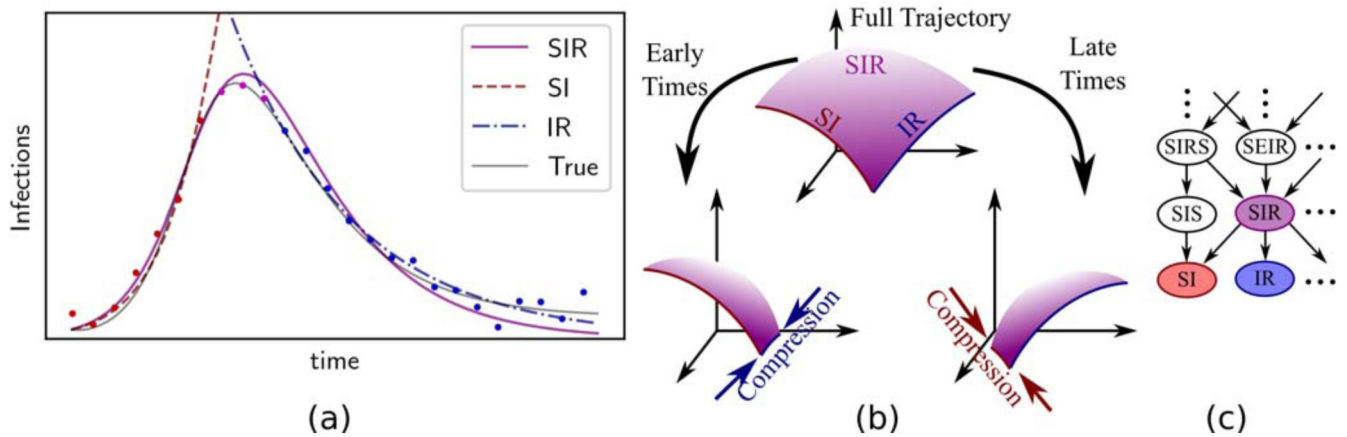


FIG. 1. (a) Infections vs time for a hypothetical epidemic. Data taken at different times—early (red), late (blue), and intermediate (purple)—exhibit qualitatively different types of behaviors. These data carry information about different aspects of the generating process. The “True” curve represents the model from which data was generated, without the corruption of Gaussian noise. (b) Information geometry quantifies how data identify a model’s parameters. The full trajectory completely identifies an SIR model, manifest by a model manifold that is not compressed along either direction. Restricting to data at early times compresses the directions related to a recovery rate so that the model is well approximated by an effective SI model. Similarly, restricting to late times compresses information about infection, leading to an effective IR model. (c) Candidate models of varying complexity can be arranged hierarchically in a directed graph. The SIR model is the *supremum* of the SI and IR models, i.e., the simplest model that combines information from both early and late times. It can use the information from data in two of the regimes to accurately predict a third.

predictions of a qualitatively different nature than those on which they were trained?

To explore this question, we use information geometry to assess parameter identifiability and predictive performance for models fit to data from different regimes and reason about the hypotheses they encode. The Fisher information matrix (FIM) is information geometry’s fundamental object, a Riemannian metric on a manifold of models using parameters as coordinates [9,15]. Model manifolds are often thin, and boundaries correspond to simplified models, i.e., having fewer parameters [16]. Distances measured by the FIM typically *compress* the model manifold into a few relevant directions [7] so that the manifold is thin and well approximated by a low-dimensional, simplified model that resides on the boundary. Given training data, the manifold boundary approximation method (MBAM) explicitly finds limiting approximations to give a minimal, reduced model that encodes the information in the data. MBAM is an enabling technology for our approach and is described in detail elsewhere [16,17].

Given several reduced models for target quantities of interest, we next seek a single model that unifies their simplified explanations. To choose an appropriate model, we introduce the *supremum principle*: Select the simplest model that is reducible to each of the target behaviors. One of the primary contributions of this Letter is to show that this intuitive idea can be given a rigorous definition using the formalism of information topology. We call this model the *supremal model* and give an algorithm below for constructing it. The supremum principle formally encapsulates a preference for simplicity akin to Occam’s razor, motivated by the assumption that abstract models that explain multiple behaviors are more likely to transfer accurately to novel behaviors than models developed for a single phenomenon.

As a motivating example, consider modeling infection trajectories during an epidemic. Figure 1(a) shows data

generated from an MSEIR (Maternally-derived immunity, Susceptible, Exposed, Infectious, Recovered) model with birth and death rates (six parameters, fifth-order dynamics) and corrupted by Gaussian noise. We partition the data into three qualitatively distinct regimes—early (red), intermediate (purple), and late (blue)—and ask the following: Which subsets of the data are informative for predicting data in another regime?

To illustrate the key principles, consider fitting the data with a simple SIR (Susceptible, Infectious, Recovered) model (two parameters, second-order dynamics),

$$\frac{dS}{dt} = -\beta I \frac{S}{N}, \quad \frac{dI}{dt} = -\gamma I + \beta I \frac{S}{N}, \quad \frac{dR}{dt} = \gamma I. \quad (1)$$

When fitting to qualitatively different data, the two-dimensional SIR model manifold is compressed depending on the informativity of the available data. The compression determines which parameters are identifiable from data and leads to an appropriate reduced model.

We focus on two reduced models on the boundary of the SIR model, shown in Fig. 1(b). The first boundary segment, corresponding to $\gamma \rightarrow 0$, is the model with no recovery compartment, i.e., an “SI” (Susceptible, Infectious) model. Similarly, the “IR” (Infectious, Recovered) model with $\beta \rightarrow \infty$ has a very fast infection rate. Consider only data from early times (red in Fig. 1). The FIM compresses the model manifold along the SI boundary segment, rendering the recovery rate γ irrelevant. The approximate SI model (red dashed line in Fig. 1) has an effective infection rate that fits the early exponential growth [18]. However, recovery data at later stages (blue) render β irrelevant and are well approximated by the IR model.

The SI and IR models interpolate in their respective domains, but fail to transfer beyond those domains. The SIR model is the simplest that can interpolate all three regimes.

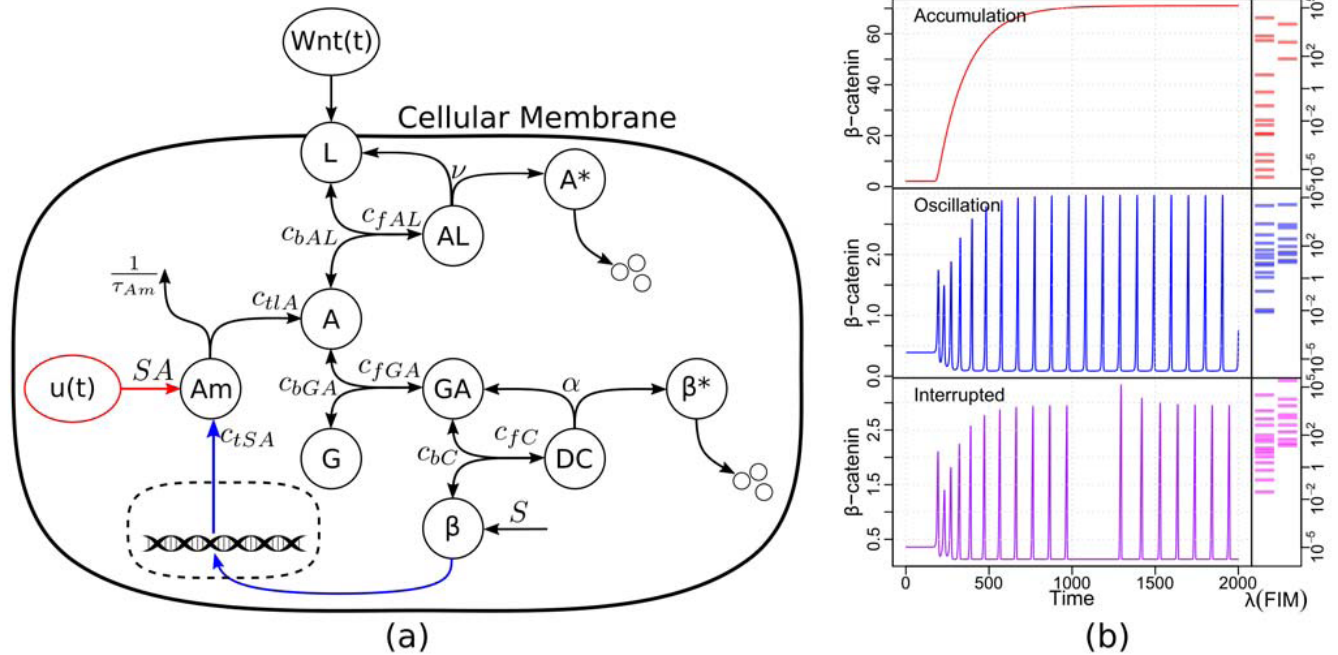


FIG. 2. (a) A network diagram showing the “full” model of the Wnt pathway. The red indicates mechanisms that are unique to the accumulation model, and blue indicates mechanisms unique to the oscillation model. (b) Three possible behaviors of β -catenin in response to Wnt. The accumulation [19] and oscillation [20] behaviors are known to occur naturally, while behaviors similar to the interrupted behavior have been reported in Refs. [21–25], but not previously expressed mathematically.

Formally, the hierarchy of potential models forms a graded partially ordered set (POSet). A POSet generalizes the concept of order within a set. Real numbers are completely ordered, that is $\forall a, b \in \mathbb{R}$, with $a \neq b$, either $a > b$ or $a < b$. POSets additionally allow for elements to be *incomparable*, i.e., neither $a > b$ nor $a < b$. Discrete POSets can be represented by a directed graph known as a Hasse diagram [26] as in Fig. 1(c). In this formalism, the SI and IR models are incomparable; there is no path in the directed graph connecting them. The SIR model is the *supremum* (i.e., least upper bound) of the SI and IR models as it is the simplest model connected to both the SI and IR models within the Hasse diagram. The topological relationships (the adjacency relationships summarized in the Hasse diagram) among candidate models enable reasoning about the mechanisms at play in diverse contexts and inform the construction of the supremal model which minimally merges model elements. The resulting supremal model is more expressive than either of its children, and so enables predictions under qualitatively different conditions than either training set. This is because the supremal model’s additional parameters have been identified by the reduction steps as meaningful, and so by definition must create at least one novel behavior in combination.

This simple example suggests the possibility of an algorithm for finding supremal models. The next example will introduce mathematical concepts necessary for a general algorithm, but the conceptual steps in the process are already clear. First, select a hypothesis space, i.e., pick a function form for a model that describes the behaviors of interest, the MSEIR model in this case. Second, reduce the model via MBAM to find minimal models that described each behavior of interest,

the IR and SI models. Finally, find the reductions that are common to each of the child models and apply them to the full model. In this case, the SI model removed all parameters except β while the IR model removed all parameters except γ so the supremal model is the SIR model, the simplest model to include both parameters. In generic scenarios, some of the reductions may combine parameters in the reduced models in ways that obscure which are the common approximations. The example below illustrates this possibility and introduces a formalism to deal with it.

The supremum principle is applicable to any hierarchical family of models. In this Letter, we focus specifically on hierarchies generated by MBAM, which includes things as diverse as power systems [27–30], systems biology [17,31–33], materials science [34], biogeochemistry [35], nuclear physics [36], neuroscience [37], and others [38–41]. To better illustrate the general algorithm, we demonstrate the construction of supremal models in the Supplemental Material [42] with a simple network spin model, and in a more complex biological system below.

The Wnt signaling pathway induces cell division in animals, and is one of the best studied in all of biology [43–45] [see Fig. 2(a) and the Supplemental Material [42]]. Via a multistep process, an extracellular Wnt molecule causes a change in intracellular levels of the transcription factor β -catenin. *In vivo*, β -catenin either “accumulates” to a new, higher equilibrium [19,46], or “oscillates” between a low baseline and periodic spikes of high concentration [20], as illustrated in Fig. 2(b).

The hypothesis space, i.e., functional form, we chose to model these two phenomena is a slight adaptation of that proposed by Jensen *et al.* [20], summarized in Fig. 2(a). We

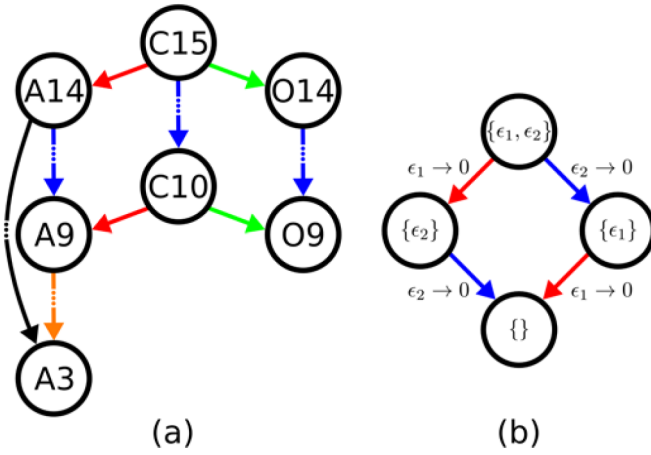


FIG. 3. (a) The Hasse diagram of key reduced models for the Wnt system. The nodes (models) are labeled by “A” for accumulation, “O” for oscillation, “C” for combined, as well as the number of parameters each model contains. The black line represents the original sequence of reduced accumulation models and the blue line represents the new sequence with parameter limits reordered to match those in the oscillation models. The central, blue arrow shows that the original model can be reduced to the supremal model using the same set of limits used to arrive at A9 and O9. Details about the model equations and sequence of limits can be found in the Supplemental Material [42]. (b) A general example of the diamond property. Consider a model with N parameters containing two parameters ϵ_1 and ϵ_2 . The diamond property states that the order in which parameters can be removed commutes. The same model with $N - 2$ parameters can be reached by first taking either ϵ_1 or ϵ_2 to zero, and then taking the other to zero. The diamond property is used to reorder the limits of a sequence, to build a supremal model.

adapt this model to the accumulation phase by removing the negative feedback loop and replacing it with a controllable activation of Axin2 [as *in vivo* by ubiquitin-specific protease 7 (USP7) [47] among others] denoted by $u(t)$ in Fig. 2. Figure 2(b) presents a characteristic time series for each of these two models. In each case, the system begins in steady state and a Wnt stimulus is introduced at 200 min. In the first case, β -catenin accumulates and equilibrates at a new steady state [19,46]. In the second case, the negative feedback loop triggers a Hopf bifurcation leading to sustained oscillations [20].

Each model has 14 parameters. A sloppy model analysis [12,48] reveals many small eigenvalues [right panel in Fig. 2(b)] in the respective FIM, indicating that many parameters are unidentifiable. We remove irrelevant parameters using the manifold boundary approximation method (MBAM) [16], as summarized in Fig. 3(a). The accumulation behavior is minimally described by three parameters while the oscillation phenomenon requires nine.

The MBAM reductions are not black boxes. Although the reduced models do abstract away many specific details, they retain vestiges of the full mechanisms, analogous to the SI and IR models in our epidemiology example. To relate these behaviors, we now seek the *supremum* of these two minimal representations. However, unlike the simple example, the min-

imal Wnt models contain partially overlapping combinations of parameters, so the construction is nontrivial.

Each reduction can be rewritten as a single parameter taken to zero. For example, consider an equilibrium approximation, i.e., $c_f, c_b \rightarrow \infty$. This can be rewritten as a time constant going to zero $\tau_f = 1/c_f \rightarrow 0$ and a nonzero equilibrium constant $K_D = c_b/c_f$. This form, however, is not unique as we could also have chosen $\tau_b = 1/c_b \rightarrow 0$. For all equilibrium approximations we adopt the first as a standard form.

Next, we observe that the same reduced models could be derived by applying the same approximations in different orders. Commuting the order of reductions creates a diamond motif in the Hasse diagram, as in Fig. 3(b). Because of the ambiguity in how reductions are labeled, consecutive limits including the same parameters can obscure this commutation relation. For example, consider the consecutive limits of an equilibrium approximation ($c_{bGA} \rightarrow \infty, c_{fGA} \rightarrow \infty$ with c_{bGA}/c_{fGA} constant and finite) followed by an irreversible approximation ($c_{bGA}/c_{fGA} \rightarrow 0, c_{fAL} \rightarrow \infty$). Reparametrizing as $\epsilon_1 = 1/c_{fGA}$, $\epsilon_2 = 1/c_{fAL}$, and $\phi = c_{fAL}c_{bGA}/c_{fGA}$ makes the diamond property apparent. The two limits of the diamond property can now be written as $\epsilon_1 \rightarrow 0$ and $\epsilon_2 \rightarrow 0$.

Writing all of the reductions in a standard form allows us to identify the approximations common to both reduced models. Applying these common approximations to the original full model constructs the supremal model, as illustrated by the blue line connecting C15 to C10 in Fig. 3. This process motivates a general algorithm for finding supremal models, and is given below.

(1) Define the hypothesis space by selecting a complex, multiparameter model to describe all desired behaviors.

(2) Perform MBAM to find reduced models that minimally describe each behavior.

(3) Reparametrize the models to detangle any conflated limits and find the approximations common to the both reduced models.

(4) Apply those common approximations to the original, full model to obtain the supremal model.

Each of these steps is illustrated in Fig. 3. The original hypothesis space is represented by C15 [step (1)]. Models A3 and O9 minimally describe the accumulation and oscillation behaviors, respectively [step (2)]. Using the diamond property, illustrated in Fig. 3(b), we reparametrize and find the approximations common to each reduced model, represented by the blue arrows [step (3)]. Applying these common approximations to C15 gives the supremal model, C10 [step (4)]. This process is described in more detail in the Supplemental Material [42], including a discussion on fitting the supremal model parameters and application to the Wnt model. Additionally, the supplemental material presents a second algorithm that exploits a general duality inherent in POSets.

Since the supremum has more parameters than either of the reduced models, the original data sets cannot individually constrain all of the supremal parameters. However, since each parameter is constrained by one data set of the other (e.g., SA is constrained by the accumulation data but not the oscillation data), fitting the supremal model to both data sets simultaneously does identify each parameter. By including both the feedback loop and external control, and their associated parameters, the supremum enables the accumulation

and oscillation phenomena, as well as additional behaviors neither A14 nor O14 can produce. Figure 2(b) demonstrates one such example, the “interrupted” behavior, in which the external control modulates the phase of the oscillation. Regular oscillatory behavior in the Wnt pathway is well known *in vivo* from the segmentation clock in vertebrate embryos along the anterior-posterior axis to establish, for example, the repeating pattern of vertebrae and ribs [49]. These regular oscillations can have their period and phase modified, stopped, or restarted through manipulation of “dorsalizing” or “ventralizing” molecular regulators, much as the interrupted behavior we see in the supremal model [21–25]. To validate our model, we apply MBAM to the full model using data for the interrupted regime. This reduction gives the supremal model constructed by our algorithm, indicating that the supremum is the model that would have been selected had observations been available for this behavior.

The supremum principle shows promise for transferring predictability to truly new domains. For example, the SI and IR models fail in the intermediate regime, and the accumulation and oscillation models fail in the interrupted regime, but the supremal models in each case are able to embrace all three behaviors. It does this by including key modeling elements (e.g., feedback and external control) that are missing from the reduced models. Since the supremal model combines distinct modeling elements, it enables new behaviors in regimes in which those modeling elements are all necessary. With a different starting model, couched in a different hypothesis space, the supremal model will be different, but it will still transfer according to the given hypothesis. This is more than the simple generalization of, e.g., multitask learning (MTL) [50]. Supremal models apply in a more global way; they aim to improve the transferability to data in a completely new regime.

Classical psychological theories use geometric constructions to represent analogical relationships. Most notably, in the parallelogram model [51], an analogy such as man:king::woman:queen is represented as four corners of a parallelogram with analogical relationships forming parallel sides [52]. Such constructions are widespread in artificial intelligence (AI) applications ranging from recommender systems [53] to natural language processing [54]. The key property, however, is the topological relationship between analogous elements [55] that for parallelograms form the same diamond motif as in Fig. 3. The analogical relationships among words are the same as those between reduced models. Kings are subsets of men just as reduced models are restricted cases of more general models, and classifications based on royalty analogize across genders just as approximations trans-

fer across models. Thus, the supremum construction identifies the mathematical “analogies” between models by teasing out the common mechanisms or analogous reductions (see the colored arrows in Fig. 3). The approximations in linking model C15 to model C10 are the same as those connecting model A14 to model A9, i.e., C15:C10::A14:A9. The colored arrows indicate the many other possible analogies that could be drawn among the models.

The supremum principle is applicable to any hierarchical family of models, and so there are some inherent limitations and potential extensions. First, the algorithm we present here is specific to hierarchies generated by MBAM, but future work could consider other families. Next, the result depends on the hierarchy one uses, for example, our Wnt study used the hierarchy generated by the model of Jensen *et al.* [20]. Given different hierarchies, supremal models are a principled way of reasoning about the implications of those hypothesis. Future work may use supremal models to guide experimental design for hypothesis testing or parameter estimation. Finally, one could consider models that are derived independently of a hierarchical family. Future work could explore how to most naturally embed such models within a hierarchy to enable transferability.

Beyond the appeal of elegant, simplified models, we expect supremal models to be of broad practical use, for example, in systems that need a controller to move between two behavioral states, but is difficult to fully model and a reduced model is needed. Such systems include shifting from diseased to healthy states in medical contexts, failing to stable power grids in electrical engineering, ductile to brittle structures in material science, and collapsed to restored resources in ecosystem-based management. Supremal models are also designed for maximal simplicity while retaining some transferability, i.e., attempting to predict in regimes not yet examined, such as in climate modeling, prosperous non-growth-based economics, and human behavior during a pandemic. Practitioners from a wide variety of fields will find supremal modeling a powerful addition to their toolboxes.

This work was supported by the U.S. National Science Foundation under Awards No. DMR-1753357 (C.P., C.A., M.K.T.), No. DMR-1834332 (C.P., M.K.T.), No. ECCS-1710727 (C.A., M.K.T.), and No. PHY-1757998 (C.M.). This work was also supported by the BYU College of Physical and Mathematical Sciences (T.M.). We thank Sean Warnick, Kolten Barfuss, and Alex Stankovic for helpful conversations. We thank Ben Francis, Dan Karls, Ellad Tadmor, and Ryan Elliott and two anonymous reviewers for comments on the manuscript.

- [1] J. Baxter, A model of inductive bias learning, *J. Artif. Intell. Res.* **12**, 149 (2000).
- [2] K. Weiss, T. M. Khoshgoftaar, and D. Wang, A survey of transfer learning, *J. Big Data* **3**, 9 (2016).
- [3] R. W. Batterman, *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence* (Oxford University Press, New York, 2002).

- [4] K. S. Brown and J. P. Sethna, Statistical mechanical approaches to models with many poorly known parameters, *Phys. Rev. E* **68**, 021904 (2003).
- [5] K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione, The statistical mechanics of complex signaling networks: Nerve growth factor signaling, *Phys. Biol.* **1**, 184 (2004).

- [6] J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, Sloppy-Model Universality Class and the Vandermonde Matrix, *Phys. Rev. Lett.* **97**, 150601 (2006).
- [7] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, Parameter space compression underlies emergent theories and predictive models, *Science* **342**, 604 (2013).
- [8] S.-i. Amari and H. Nagaoka, *Methods of Information Geometry*, Vol. 191 (American Mathematical Society, Providence, RI, 2007).
- [9] M. K. Transtrum, B. B. Machta, and J. P. Sethna, Why are Nonlinear Fits to Data so Challenging?, *Phys. Rev. Lett.* **104**, 060201 (2010).
- [10] C. H. LaMont and P. A. Wiggins, Correspondence between thermodynamics and inference, *Phys. Rev. E* **99**, 052140 (2019).
- [11] K. N. Quinn, H. Wilber, A. Townsend, and J. P. Sethna, Chebyshev Approximation and the Global Geometry of Model Predictions, *Phys. Rev. Lett.* **122**, 158302 (2019).
- [12] M. K. Transtrum, B. B. Machta, and J. P. Sethna, Geometry of nonlinear least squares with applications to sloppy models and optimization, *Phys. Rev. E* **83**, 036701 (2011).
- [13] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, Building machines that learn and think like people, *Behav. Brain Sci.* **40**, e253 (2017).
- [14] T. Webb, Z. Dulberg, S. Frankland, A. Petrov, R. O'Reilly, and J. Cohen, Learning representations that support extrapolation, in *Proceedings of the 37th International Conference on Machine Learning*, edited by H. Daumé III and A. Singh, Proceedings of Machine Learning Research, Vol. 119 (PMLR, 2020), pp. 10136–10146.
- [15] A. F. Brouwer and M. C. Eisenberg, The underlying connections between identifiability, active subspaces, and parameter space dimension reduction, [arXiv:1802.05641](https://arxiv.org/abs/1802.05641).
- [16] M. K. Transtrum and P. Qiu, Model Reduction by Manifold Boundaries, *Phys. Rev. Lett.* **113**, 098701 (2014).
- [17] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, Perspective: Sloppiness and emergent theories in physics, biology, and beyond, *J. Chem. Phys.* **143**, 010901 (2015).
- [18] Although this approximation is constructed by taking $\gamma \rightarrow 0$, it does not require the “true” value of γ to be small. Rather, the role of the recovery mechanism can be compressed into a simpler model with an effective infection rate, similar to the effective electron mass in a condensed matter system.
- [19] L. Goentoro and M. W. Kirschner, Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling, *Mol. Cell* **36**, 872 (2009).
- [20] P. B. Jensen, L. Pedersen, S. Krishna, and M. H. Jensen, A Wnt oscillator model for somitogenesis, *Biophys. J.* **98**, 943 (2010).
- [21] I. H. Riedel-Kruse, C. Müller, and A. C. Oates, Synchrony dynamics during initiation, failure, and rescue of the segmentation clock, *Science* **317**, 1911 (2007).
- [22] S. Gibb, A. Zagorska, K. Melton, G. Tenin, I. Vacca, P. Trainor, M. Maroto, and J. K. Dale, Interfering with Wnt signalling alters the periodicity of the segmentation clock, *Dev. Biol.* **330**, 21 (2009).
- [23] A. Goldbeter and O. Pourquié, Modeling the segmentation clock as a network of coupled oscillations in the Notch, Wnt and FGF signaling pathways, *J. Theor. Biol.* **252**, 574 (2008).
- [24] C. Gomez, E. M. Özbudak, J. Wunderlich, D. Baumann, J. Lewis, and O. Pourquié, Control of segment number in vertebrate embryos, *Nature (London)* **454**, 335 (2008).
- [25] Y. Rui, Z. Xu, B. Xiong, Y. Cao, S. Lin, M. Zhang, S. C. Chan, W. Luo, Y. Han, Z. Lu, Z. Ye, H. M. Zhou, J. Han, A. Meng, and S. C. Lin, A β -catenin-independent dorsalization pathway activated by Axin/JNK signaling and antagonized by Aida, *Dev. Cell* **13**, 268 (2007).
- [26] M. K. Transtrum, G. Hart, and P. Qiu, Information topology identifies emergent model classes, CoRR, [arXiv:1409.6203](https://arxiv.org/abs/1409.6203).
- [27] V. G. Švenda, M. K. Transtrum, B. L. Francis, A. T. Sarić, and A. M. Stanković, State estimation model reduction through the manifold boundary approximation method, *IEEE Trans. Power Syst.* **37**, 272 (2022).
- [28] A. T. Sarić, A. A. Sarić, M. K. Transtrum, and A. M. Stanković, Symbolic regression for data-driven dynamic model refinement in power systems, *IEEE Trans. Power Syst.* **36**, 2390 (2020).
- [29] B. L. Francis, J. R. Nuttall, M. K. Transtrum, A. T. Sarić, and A. M. Stanković, Network reduction in transient stability models using partial response matching, in *2019 North American Power Symposium (NAPS)* (IEEE, New York, 2019), pp. 1–6.
- [30] M. K. Transtrum, A. T. Sarić, and A. M. Stanković, Measurement-directed reduction of dynamic models in power systems, *IEEE Trans. Power Syst.* **32**, 2243 (2016).
- [31] J. E. Jeong, Q. Zhuang, M. K. Transtrum, E. Zhou, and P. Qiu, Experimental design and model reduction in systems biology, *Quant. Biol.* **6**, 287 (2018).
- [32] M. K. Transtrum and P. Qiu, Bridging mechanistic and phenomenological models of complex biological systems, *PLoS Comput. Biol.* **12**, e1004915 (2016).
- [33] B. K. Mannakee, A. P. Ragsdale, M. K. Transtrum, and R. N. Gutenkunst, Sloppiness and the geometry of parameter space, in *Uncertainty in Biology* (Springer, Berlin, 2016), pp. 271–299.
- [34] Y. Kurniawan, C. L. Petrie, K. J. Williams, M. K. Transtrum, E. B. Tadmor, R. S. Elliott, D. S. Karls, and M. Wen, Bayesian, frequentist, and information geometry approaches to parametric uncertainty quantification of classical empirical interatomic potentials, *J. Chem. Phys.* **156**, 214103 (2022).
- [35] G. L. Marschmann, H. Pagel, P. Kügler, and T. Streck, Equifinality, sloppiness, and emergent structures of mechanistic soil biogeochemical models, *Envir. Modell. Software* **122**, 104518 (2019).
- [36] T. Nikšić, M. Imbrišak, and D. Vretenar, “Sloppy” nuclear energy density functionals. II. Finite nuclei, *Phys. Rev. C* **95**, 054304 (2017).
- [37] J. Rasband, Two reduced models of nerve behavior, Bachelor’s thesis, Brigham Young University, 2021.
- [38] P. E. Paré, D. Grimsman, A. T. Wilson, M. K. Transtrum, and S. Warnick, Model boundary approximation method as a unifying framework for balanced truncation and singular perturbation approximation, *IEEE Trans. Autom. Control* **64**, 4796 (2019).
- [39] T. Gerach, D. Weiß, O. Dössel, and A. Loewe, Observation guided systematic reduction of a detailed human ventricular cell model, in *2019 Computing in Cardiology (CinC)* (IEEE, New York, 2019), p. 1.
- [40] D. M. Lombardo and W.-J. Rappel, Systematic reduction of a detailed atrial myocyte model, *Chaos* **27**, 093914 (2017).
- [41] P. E. Paré, A. T. Wilson, M. K. Transtrum, and S. C. Warnick, A unified view of balanced truncation and singular perturbation

- approximations, in *2015 American Control Conference* (IEEE, New York, 2015), pp. 1989–1994.
- [42] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.L032044> for model equations and more detailed descriptions and examples of the supremum algorithm.
- [43] R. Nusse and H. Clevers, Wnt/ β -catenin signaling, disease, and emerging therapeutic modalities, *Cell* **169**, 985 (2017).
- [44] C. Y. Logan and R. Nusse, The Wnt signaling pathway in development and disease, *Annu. Rev. Cell Dev. Biol.* **20**, 781 (2004).
- [45] Y. Ding, S. Su, W. Tang, X. Zhang, S. Chen, G. Zhu, J. Liang, W. Wei, Y. Guo, L. Liu, Y.-G. Chen, and W. Wu, Enrichment of the β -catenin–TCF complex at the S and G2 phases ensures cell survival and cell cycle progression, *J. Cell Sci.* **127**, 4833 (2014).
- [46] E. Lee, A. Salic, R. Krger, R. Heinrich, and M. W. Kirschner, The roles of APC and Axin derived from experimental and rheoretical analysis of the Wnt pathway, *PLoS Biol.* **1**, e10 (2003).
- [47] L. Ji, B. Lu, R. Zamponi, O. Charlat, R. Aversa, Z. Yang, F. Sigoillot, X. Zhu, T. Hu, J. S. Reece-Hoyes, C. Russ, G. Michaud, J. S. Tchorz, X. Jiang, and F. Cong, USP7 inhibits Wnt/ β -catenin signaling through promoting stabilization of Axin, *Nat. Commun.* **10**, 4184 (2019).
- [48] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, Universally sloppy parameter sensitivities in systems biology models, *PLoS Comput. Biol.* **3**, e189 (2007).
- [49] O. Pourquie, The segmentation clock: Converting embryonic time into spatial pattern, *Science* **301**, 328 (2003).
- [50] R. Caruana, Multitask learning, *Mach. Learn.* **28**, 41 (1997).
- [51] D. E. Rumelhart and A. A. Abrahamson, A model for analogical reasoning, *Cogn. Psychol.* **5**, 1 (1973).
- [52] J. C. Peterson, D. Chen, and T. L. Griffiths, Parallelograms revisited: Exploring the limitations of vector space models for simple analogies, *Cognition* **205**, 104440 (2020).
- [53] C. Musto, Enhanced vector space models for content-based recommender systems, in *Proceedings of the Fourth ACM Conference on Recommender Systems* (ACM Press, New York, 2010), pp. 361–364.
- [54] J. N. Reid and A. N. Katz, Vector space applications in metaphor comprehension, *Metaphor Symb.* **33**, 280 (2018).
- [55] D. Gentner, Structure-mapping: A theoretical framework for analogy, *Cogn. Sci.* **7**, 155 (1983).