Classifying and quantifying parameter nonlinearity in biological models

Alex Shumway[1], Mark K. Transtrum[1]

Models of biological systems typically include many parameters. Furthermore, the model behavior often responds to changes in these parameters in highly nonlinear ways. This nonlinear response is responsible for many of the unique emergent behaviors of biological systems. We discuss how model nonlinearity can be both quantified and classified. We consider the tensor of model second derivatives, i.e., the vector of Hessian matrices for each model prediction or alternatively the Jacobian of the Jacobian matrix. We use a higher-order singular value decomposition to identify the principal parameter combinations exhibiting the greatest nonlinearity (generalizations of singular vectors) and quantify this nonlinearity using generalizations of singular values. We further classify types of nonlinearity by decomposing the second derivative tensor into geometrically motivated components, including extrinsic, intrinsic, and parameter-effects nonlinearity. We discuss applications to model interpretation as well as for numerical methods.

MODELS of biological systems, such as those describing dynamics of protein signaling, gene regulation, and other cellular activity typically include a large number parameters. In many cases these parameters are unknown and must be estimated from data. The response of the model behavior to changes in these parameters is often highly nonlinear. The nonlinearity in the model parameters leads to challenges for numerical methods, such as data fitting [1]. Furthermore, parameter nonlinearity makes it difficult to interpret the model. In particular, the nonlinearity makes it challenging to identify the particular parameter or parameter combination that controls a feature of the model behavior. Consequently, highly nonlinear models often exhibit nontrivial, emergent behavior that is obscured by this nonlinear parameter response [2].

Understanding the role of parameter nonlinearity in models is important for a host of modeling activities including numerical algorithms (such as data fitting or Bayesian posterior sampling), model interpretation, model construction, and experimental design. We present a theoretical and computational framework for understanding the effect of nonlinear parameters in complex biological models that uses techniques from differential geometry, information theory, and linear algebra.

The response of a model to *small* changes in parameters can be studied using a local linearization of the model: characterized by either a Jacobian matrix (derivatives of model predictions with respect to each parameter) or through the closely related Fisher Information Matrix (FIM). Considerable effort has been devoted to understanding how these objects characterize the model, including "sloppy" model research [3] with applications to experimental design and numerical methods.

We extend these methods by considering the tensor of second derivatives, i.e., the Jacobian of the Jacobian matrix. We use the higher-order singular value decomposition [4] of this tensor to quantify the nonlinearity and identify the principal parameter directions corresponding to this nonlinearity. We further construct other measure of nonlinearity motivated by the information geometric interpretation of the model, including intrinsic curvature, i.e., Riemann and Ricci tensors, extrinsic curvature, and parameter-effects curvature, i.e., the connection coefficients [5].

We find that for many systems biology models, most of the parameter nonlinearity is dominated by a few parameter combinations. This result is analogous to "sloppy" models in which the Jacobian matrix has an exponential distribution of singular values. The nonlinearity quantifies how principal parameter combinations "rotate" into one another and lead to compensatory effects. Using this, we identify groups of parameters that act as effective control knobs for model behavior. We compare these groups of parameters with the functional relationships identified by the manifold boundary approximation method [6] and discuss how these insights can guide the interpretation of the model and improve numerical methods for data fitting and posterior sampling.

[1]Department of Physics & Astronomy, Brigham Young University, Provo Utah. E-mail: mktranstrum@byu.edu

REFERENCES

[1] Transtrum, Mark K., Benjamin B. Machta, and James P. Sethna. "Why are nonlinear fits to data so challenging?." *Physical review letters* 104.6 (2010): 060201.

[2] Machta, Benjamin B., et al. "Parameter space compression underlies emergent theories and predictive models." *Science* 342.6158 (2013): 604-607.

[3] Waterfall, Joshua J., et al. "Sloppy-model universality class and the Vandermonde matrix." *Physical review letters* 97.15 (2006): 150601.

[4] De Lathauwer, Lieven, Bart De Moor, and Joos Vandewalle. "A multilinear singular value decomposition." *SIAM journal on Matrix Analysis and Applications* 21.4 (2000): 1253-1278.

[5] Transtrum, Mark K., Benjamin B. Machta, and James P. Sethna. "Geometry of nonlinear least squares with applications to sloppy

[6] Transtrum, Mark K., and Peng Qiu. "Model Reduction by Manifold Boundaries."*Physical review letters* 113.9 (2014): 098701.models and optimization."*Physical Review E* 83.3 (2011): 036701.