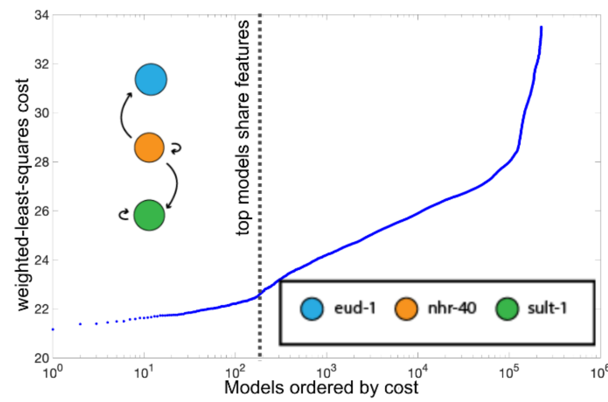
National Institute for Theory and Mathematics in Biology
Funded by the U.S. National Science Foundation and the Simons Foundation

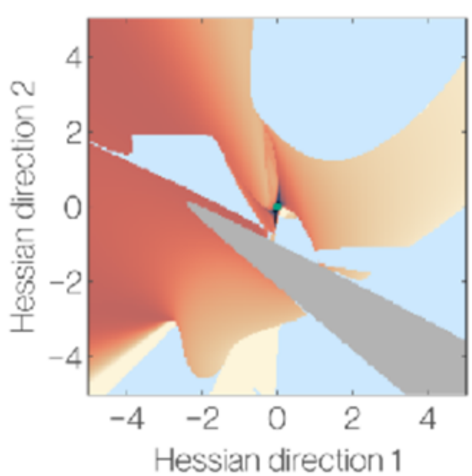
Discovering uncertain model structures

Cody FitzGerald, Alasdair Hastewell, Christina Catlett, Yifan Zhang, Niall Mangan

Inferring causal interactions between biological species in nonlinear dynamical networks is a much-pursued problem across biological length scales, from gene regulation to ecosystems. We are especially interested in whether models that capture varying interactions can be distinguished given data and approach this question using A) a computationally intensive combinatorial approach to ensemble model fitting and 2) sparse optimization, which selects a subset of terms from a library using an L1 regularization or thresholding.



Combinatorial Approach. We attempt to model a gene regulatory network involved in a developmental decision in the nematode *P. pacificus* in the low data limit. From sparse, noisy measurements of mRNA expression over the course of development, we apply a combinatorial search across approximately 250,000 model variants and extract an effective model of the regulatory network. The ensemble of low-cost or best-fit models shares a set of characteristics: one gene, *nhr-40,* is a master regulatory of the network, and *sult-1* likely undergoes autoregulation (Right Fig.). Given the data quality, we cannot specify any further interactions of the network or further resolve the type of regulation (positive or negative). This study provides biological insight and acts as a generalized sensitivity analysis across model structures, illuminating what we can *honestly* say about biological systems from typical experimental datasets.



Sparse Optimization: Suitable algorithms for sparse model selection depend on many factors such as model class, dynamical behavior, and data availability (noise present and sampling frequency) and computational resources. Our group has experience with regression, forward-simulation-based optimization (Left Fig. Blue/Grey are unstable dynamics, dark->light is low->high cost), and data-assimilation strategies, all of which have been used in sparse-model selection. To date, the algorithm is chosen by trial and error or intuition of the practitioner. We seek to develop robust metaheuristics to choose or swap between different model selection formulations to improve algorithmic efficiency, through benchmarking and analyzing the cost functions associated with different formulations, including weak versus strong model constraints; numerical discretization; and optimization procedure. Initial results suggest that using different formulations as better estimates of model parameters become available can improve robustness, and sparsifying too early during model selection can lead to errors.

This research was supported in part by grants from the NSF (DMS-2235451) and Simons Foundation (MP-TMPS-00005320).