TOC: Transcriptomics Optimal Control

Simulating and controlling diseased and healthy brain cells

Ionelia Buzatu, Ryan Cubero, Manuel Del Verme

Introduction

Recent advances in understanding RNA dynamics[1] has allowed us to simulate single cell transcriptomes[2]. We use these ideas to develop a transcriptomic optimal control formulation where we control the state of the cell with a single shooting control system.

Method

1.Create the GRN from the single cell RNA sequencing data with PyScenic[3].

2.Train a neural network classifier to learn the cell states, healthy and unhealthy. The expert is part of the cost function.

3.Simulate the cell fate with SERGIO[2] implemented in JAX[7].

4.Auto-differentiate with respect to the cost function with a single-shooting control optimization.

Objective

The goal is to find the best actions, or the genes whose expressions need to be altered in such a way to make the unhealthy cell healthy. We have actions for the unspliced transcript and actions for the spliced transcript, as the spliced transcript depends on the unspliced too.

Expert

The expert uses the gene expression profile to classify the cell state at time step *t*. Our objective is to be healthy or 0 in the cost function. The t-SNE below helps visualize the trajectory in the real data and distinguish well their state.

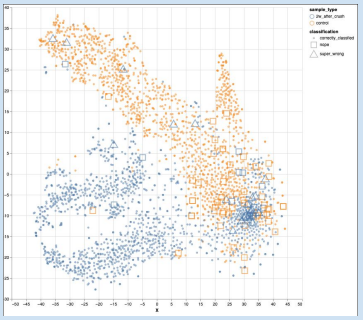

Figure 1: t-SNE plots showing retinal ganglion cells in healthy conditions (yellow) and 2 weeks after the optic nerve crush (blue). Each point is a cell. In the t-SNE projection, we mark the misclassified cells (square and triangle points).

Dynamics

Generating unlimited amount of biological data is a common concern. Thanks to the recent advances into the understanding of the transcriptome dynamics, we generate synthetic datasets with SERGIO, which statistically resemble the experimental data, with stochastic differential equations (SDEs). Simulating steady state or unchanged cell state overtime is done with (Equation 1):

$$\frac{dx_i}{dt} = P_i(t) - \lambda_i x_i(t) + q_i \left(\sqrt{P_i(t)}\alpha + \sqrt{\lambda_i x_i(t)}\beta \right) \quad \text{Eq.1}$$

Instead, the cell state changing over time is done with Equation 2 and 3:

$$\frac{du}{dt} = P_i(t) - (\lambda_i + \mu_i)u_i(t) - q_i^u \left(\sqrt{P_i(t)}\alpha + \sqrt{(\lambda_i + \mu)u_i(t)}\beta \right) \quad \text{(Eq.2)}$$

$$\frac{ds_i}{dt} = \mu_i u_i(t) - \gamma_i s_i(t) + q_i^s \left(\sqrt{\mu_i u_i(t)}\phi) + \sqrt{\gamma_i s_i(t)}\omega \right) \quad \text{(Eq.3)}$$

where P is the production rate of pre-mRNA that includes regulatory interactions, λ and μ are the degradation and splicing rate respectively of pre-mRNA and q is the noise of the transcription of pre-mRNA. Also, γ is the degradation rate of spliced mRNA. α, β, ϕ and γ are independent Gaussian white noise processes.

Cost Function

To clarify the notation here, we keep *u* for the unspliced term given by the simulator and for the control we use a^u and a^s for the unspliced transcript actions and spliced transcript actions respectively.

$$c_t(x_i | a_t^u, a_t^s) = \phi_\theta(x_t) + f(a_t^u) + f(a_t^s) \quad \text{Eq.4}$$

$$\text{s.t. } x_t = g(x_{t-1}, a_{t-1}^u, a_{t-1}^s)$$

where $\phi_\theta(x_t)$, the classifier expert output, is 1 if the cell is still unhealthy and 0 if the cell has reached a healthy state.

Discussion

The project is still ongoing. We expect to see specific interactions in the GRN or specific genes or transcription factors playing a major role more than others in the cell state shifting. Future results and development will be visible at: https://github.com/ioneliabuzatu/TOC.

References

[1] Bergen et al. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol 38, 1408–1414 (2020).

[2] Dibaeinia, Payam, and Saurabh Sinha. "SERGIO: a single-cell expression simulator guided by gene regulatory networks." Cell Systems 11.3, 252-271 (2020).

[3] Van de Sande, Bram, et al. "A scalable SCENIC workflow for single-cell gene regulatory network analysis." Nature Protocols 15.7 (2020): 2247-2276.

[4] Tran, Nicholas M., et al. "Single-cell profiles of retinal ganglion cells differing in resilience to injury reveal neuroprotective genes." Neuron 104.6 (2019): 1039-1055.

[5] Liu, Yunlong, Hui Bin Sun, and Hiroki Yokota. "Regulating gene expression using optimal control theory." Third IEEE Symposium on Bioinformatics and Bioengineering, 2003. Proceedings.. IEEE, 2003.

[6] Imani, Mahdi, and Ulisses M. Braga-Neto. "Control of gene regulatory networks with noisy measurements and uncertain inputs." IEEE Transactions on Control of Network Systems 5.2 (2017): 760-769.

[7] Frostig, Roy, Matthew James Johnson, and Chris Leary. "Compiling machine learning programs via high-level tracing." Systems for Machine Learning (2018).