Oscillatory differentiation dynamics fundamentally restricts the resolution of pseudotime reconstruction algorithms

Huy K. Vo, Jonathan H.P. Dawes, and Robert N. Kelsh

5 January 2024

Guide to the Electronic Supplementary Material

This ESM for this paper consists of 7 files containing code. Three of the code files are in *MATLAB* and four are in *R*. Below they are discussed in (approximately) the natural order in which they generate data and figures as presented in the paper. Further details and comments are given in each of the code files.

Streams of random numbers are needed both for the generation of solutions to the stochastic differential equation and for initialisation of the k-means clustering algorithms. These are initiated with random seeds in the relevant codes.

In several cases the file paths will need to be adjusted to suit local conditions.

MATLAB Version number: 9.13.0.2126072 (R2022b) Update 3
R Version number: 4.2.2

1. code_1_generate_grn_data.m

This generates the sample trajectories from the stochastic differential equation (SDE) (5) in the paper. It relies on an external package for the Euler-Maruyama method implemented to solve the SDE.

Output: data file (.csv format) containing 400 points sampled at random from a realisation of the SDE, plus a further 400 points each from two symmetric copies of that realisation, i.e. 1200 points in total.

The code also contains commands to plot realisations of the SDE, e.g. to generate Figures 2, 3 and 9 in the paper.

2. code_2_count_lineages.R

This code takes as input a file of random sampled points generated by code_1 and computes a set of clusterings (for k=1 … 40 clusters) using the k-means algorithm. Then, for each clustering, the slingshot routines are used to compute lineages. The number of lineages in each case that is 'short' is calculated, and these summary statistics are written to an output datafile in .csv format.

Note that this code requires functions from the **slingshot** package.

3. code_3_analyse_lineages_1000_realisations.m

This code takes as input the files generated by code_2 and produces as output the heatmap plots shown in Figure 11 in the paper.

4. code_4_labora.R

code_5_labora_analysis.R

The first of these two R code files computes clusters, lineages and numbers of long lineages, and produces data files from which the second R code generates the stacked bar charts shown in Figures 7,8 and 10. Note that these codes require functions from the **slingshot** package.

5. figure_13_code_plot_regimes.m

This code generates the plots shown in Figure 13 in the paper which summarises the analytical results of section 5.

6. r_plots.R

Generates 3D and 2D plots (2D ones in barycentric coordinates) as shown in Figures 4, 5 and 6 in the paper. Note that this code requires functions from the **slingshot** package.