Applied Stochastic Processes

Multi-scale modelling and simulation of a gene regulatory circuit

Miquel Barcelona (NIU: 1359817)

February 2020

Contents

Introduction

The dynamics of a gene regulatory system (GRN), which is described in Figure 1, can be simulated using two different stochastic algorithms. This system acts as a circuit of self-regulation of the transcription of the gene X_1. In this case, the gene product forms a dimmer and it works as its own transcription factor which binds the promoter region of the gene thus triggering gene transcription. In this model, X_2 denotes the number of bound promoter sites in the gene promoter region while X_1 will be the number of transcription factor molecules. So X_1 and X_2 are the two variables of the described stochastic system which is modelled using the rates from Table 1.

Transcription rate	r	Event
$W_1(\mathbf{X}) = \hat{R} + k_1 X_2$	$r_1 = (1, 0)$	Synthesis of the transcription factor
$W_2(\mathbf{X}) = k_2 X_1$	$r_2 = (-1, 0)$	Degradation of the transcription factor
$W_3(\mathbf{X}) = b_{11} X_1 (X_1 - 1)(E - X_2)$	$r_3 = (-2, +1)$	Dimer binding to the gene promoter region
$W_4(\mathbf{X}) = u_{11} X_2$	$r_4 = (+2, -1)$	Unbinding from the gene promoter region

Table 1: Transitions rates for the GRN system where E denotes the total number of binding sites within the gene's regulatory system.

The aim of this project consists on simulating the dynamics of the GRN model using the Gillespie Algorithm and the Quasi Steady State Approximation Algorithm and comparing both performance results. For that reason, the first part of this report is based on writing the corresponding equations that described the behaviour of the system depending on different scales. Once this is done, we present the main results of both algorithms and we compare them.

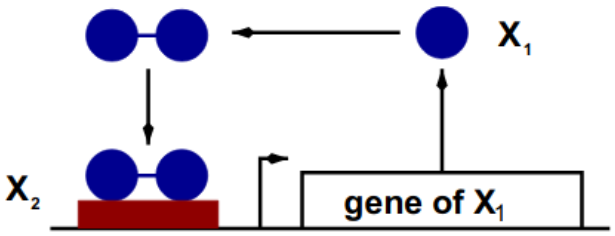


Figure 1: Scheme of the GRN system. Image taken from [1].

Dynamics of the GRN system

First of all, the evolution of the two variables X_1 and X_2 can be described in terms of the corresponding Poisson representation. For that reason, the state of the system at time t can be written as

$$X_i(t) = X_i(0) + \sum_{j=0}^{M} r_j Z_j(t), \forall i = 1, \ldots, N. \tag{1}$$

In this case, N is the number of chemical species, M is the number of reactions, r_j are the array of reaction and Z_j are the number of times that reaction j occurs in the interval $[0, t)$, [2, 3]. The term Z_j is a counting process with intensity W_j that can be calculated using a Poisson distribution since we model $\mathbf{X}$ as a continuous time Markov chain. Then, we can use the following expression

$$Z_j(t) \sim \text{Poiss}\left(\int_0^t W_j(\mathbf{X}(s))ds\right) \equiv Y_j\left(\int_0^t W_j(\mathbf{X}(s))ds\right), \tag{2}$$

where W_j will be the transition rates extracted from Table 1. In this case, Y_j are independent unit Poisson processes and W_j is the reaction rate at which the j-th reaction occurs.

Considering these assumptions and using Equation 1 and Table 1, one can easily formulate the dynamics of the GRN system using the following expression:

$$\begin{aligned} X_1(t) =& X_1(0) + Y\left(\int_0^t \left(\hat{R} + k_1 X_2\right) ds\right) + 2Y\left(\int_0^t (u_{11} X_2)\, ds\right) \\ & - Y\left(\int_0^t (k_2 X_1)\, ds\right) - 2Y\left(\int_0^t (b_{11} X_1(X_1 - 1)(E - X_2)\, ds\right), \\ X_2(t) =& X_2(0) + Y\left(\int_0^t (b_{11} X_1(X_1 - 1)(E - X_2)\, ds\right) - Y\left(\int_0^t (u_{11} X_2)\, ds\right). \end{aligned} \tag{3}$$

This expression, which is equivalent to the Master equation of the system, will be used later to simulate the dynamics using the Gillespie algorithm.

Multiscale Approximation

In order to be able to obtain some good representative results from Equation 3, one has to consider a different approach to deriving a scaling limit approximation for the GRN model [3]. In this case, we perform a separation of the time scales by using the following re-scaling variables,

$$n_1^S = \frac{X_1}{S}, \quad n_2^E = \frac{X_2}{E}, \tag{4}$$

and the time dimensionless variable,

$$\tau = b_{11} S E t, \tag{5}$$

assuming that $\epsilon = E/S \leqq 1$. Equation 4 uses the new scale factor S which is the characteristic number of protein transcripts.

Using Equations 4 and 5 in Equation 3, we obtain

$$\begin{aligned} n_1^S(\tau) =& n_1^S(0) + S^{-1}Y\left(S\int_0^\tau \left(R + \kappa_1 n_2^E\right) d\sigma\right) + 2S^{-1}Y\left(S\int_0^\tau \left(\nu_{11} n_2^E\right) d\sigma\right) \\ & - S^{-1}Y\left(S\int_0^\tau \left(\kappa_2 n_1^S\right) d\sigma\right) - 2S^{-1}Y\left(S\int_0^\tau \left(\beta_{11} n_1^S (n_1^S - \frac{1}{S})(1 - n_2^E)\right) d\sigma\right), \\ n_2^E(\tau) =& n_2^E(0) + E^{-1}Y\left(\frac{E}{\epsilon}\int_0^\tau \left(\beta_{11} n_1^S (n_1^S - \frac{1}{S})(1 - n_2^E)\right) d\sigma\right) - E^{-1}Y\left(\frac{E}{\epsilon}\int_0^\tau \left(\nu_{11} n_2^E\right) d\sigma\right), \end{aligned} \tag{6}$$

for the following re-scaled parameters:

$$\begin{aligned} R = \frac{\hat{R}}{b_{11}S^2E}, \quad \nu_{11} = \frac{u_{11}}{S^2 b_{11}}, \quad \beta_{11} = 1, \\ \kappa_1 = \frac{k_1}{b_1 1 S^2}, \quad \kappa_2 = \frac{k_2}{b_{11}ES}. \end{aligned} \tag{7}$$

The Equation 6 allows us to separate the dynamics of the GRN model in the slow variable and the fast variable. In this case, as we are assuming that $\epsilon \leqq 1$, the re-scaled rates from the equation of the fast variable n_2^E are much larger than those associated with the slow variable n_1^S [4].

Inner Solution

In order to get the evolution of the fast variable under a frozen state of the slow variable we proceed by taking the second re-scaling factor given by $T = \epsilon - 1\tau$. Applying this change to Equation 6, we obtain:

$$\begin{aligned} n_1^S(T) =& n_1^S(0) + S^{-1}Y\left(S\epsilon\int_0^T \left(R + \kappa_1 n_2^E\right) du\right) + 2S^{-1}Y\left(S\epsilon\int_0^T \left(\nu_{11} n_2^E\right) du\right) \\ & - S^{-1}Y\left(S\epsilon\int_0^T \left(\kappa_2 n_1^S\right) du\right) - 2S^{-1}Y\left(S\epsilon\int_0^T \left(\beta_{11} n_1^S (n_1^S - \frac{1}{S})(1 - n_2^E)\right) du\right), \\ n_2^E(T) =& n_2^E(0) + E^{-1}Y\left(E\int_0^T \left(\beta_{11} n_1^S (n_1^S - \frac{1}{S})(1 - n_2^E)\right) du\right) - E^{-1}Y\left(E\int_0^T \left(\nu_{11} n_2^E\right) du\right). \end{aligned} \tag{8}$$

Then, by taking the law of large numbers, we can see how the dynamics of the slow variable n_1^S is $\mathcal{O}(\epsilon)$ [3]. This means that the fast variable n_2^E will evolve under a constant value of the slow variable.

At this point, considering that the variable n_1^S is constant, we can obtain an expression for the quasi steady state distribution of X_2. For that, we start by writing the Master Equation for the evolution of X_2 with constant n_1^S and the re-scaled rate constants given by Equation 7. This can be expressesed by:

$$\begin{aligned} \frac{\partial P(X_2, T)}{\partial T} =& E\left\{n_1^S(n_1^S - \frac{1}{S})(E - X_2 + 1)P(X_2 - 1, T) + \nu_{11}(X_2 + 1)P(X_2 + 1, T)\right\} \\ & - E\left\{n_1^S(n_1^S - \frac{1}{S})(E - X_2)P(X_2, T) + \nu_{11}P(X_2, T)\right\}. \end{aligned} \tag{9}$$

Using the last Master Equation, we can write the corresponding PDE for the generating function of the variable X_2. By taking $G(q, T) = \sum_{n=0}^{\infty} q^n P(n, T)$, we obtain:

$$\frac{\partial G(q, T)}{\partial T} = E\left\{G(q, T)(q - 1)\left[n_1^S(n_1^S - \frac{1}{S})E\right] + \frac{\partial G(q, T)}{\partial q}(1 - q)\left[q n_1^S(n_1^S - \frac{1}{S}) + \nu_{11}\right]\right\}. \tag{10}$$

If we take the limit of $T \to \infty$, then $\partial G(q,T)/\partial T = 0$ and we can get an expression for the generating function $G(q,T)$ by solving the following equation:

$$\frac{\partial G(q,T)}{\partial T} = \frac{n_1^S(n_1^S - \frac{1}{S})EG(q,T)}{qn_1^S(n_1^S - \frac{1}{S}) + \nu_{11}}. \tag{11}$$

Hence, the generating function for the variable X_2 when $T \to \infty$ will be

$$G(q,T) = \left(\frac{q\Theta + \nu_{11}}{\Theta + \nu_{11}}\right)^E, \tag{12}$$

where

$$\Theta = n_1^S(n_1^S - \frac{1}{S}).$$

Taking as $p_S = \frac{\Theta}{\Theta+\nu_{11}}$ and $p_E = \frac{\nu_{11}}{\Theta+\nu_{11}}$, Equation 12 can be expressed as

$$G(q,T) = (qp_S + p_E)^E \tag{13}$$

where the new definitions satisfies $p_S + p_E = 1$. Then, the generating function 12 corresponds to the generating function of a binomial variable. In this way, the variable X_2 can be expressed at its quasi steady state as a random variable that satisfies the following distribution

$$X_2 \sim \text{Bin}(p_S, E). \tag{14}$$

This property is used in the next section while applying the QSSA Algorithm to sample the variable X_2.

Results of the simulations

In this section we present the results of simulating a GRN model by using he Gillespie Algorithm applied to Equation 3 and the QSSA Algorithm using the distribution given by 14. The corresponding code to generate these results can be found in the folder `src`. This folder contains a `Makefile` that can be run using the following command

```
make superall
```

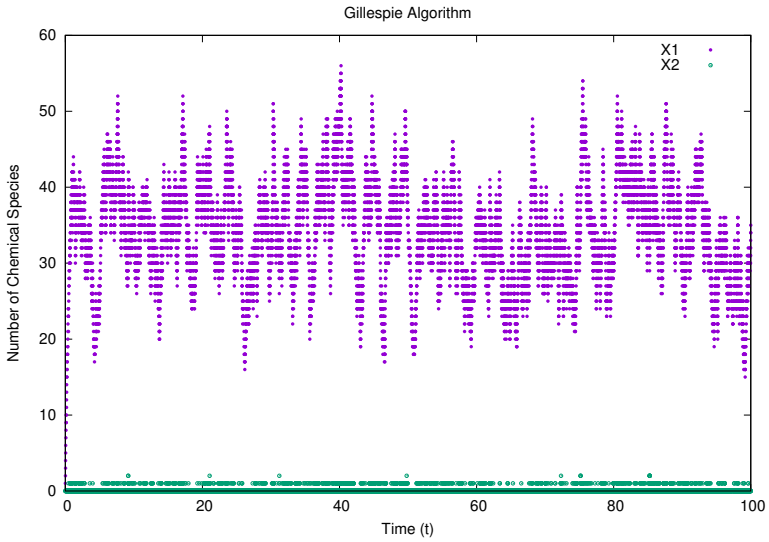
that will compile and execute both algorithms using a maximum time simulation of 50000 minutes. This command also generates the corresponding files with the obtained results for both simulation methods.

These two methods are used applying the parameter values from Table 2 and initial conditions $X_1(0) = X_2(0) = 0$. Considering this and applying both methods, we obtain that the evolution of the number of chemical species for a GRN model can be described as it is shown in Figure 2.

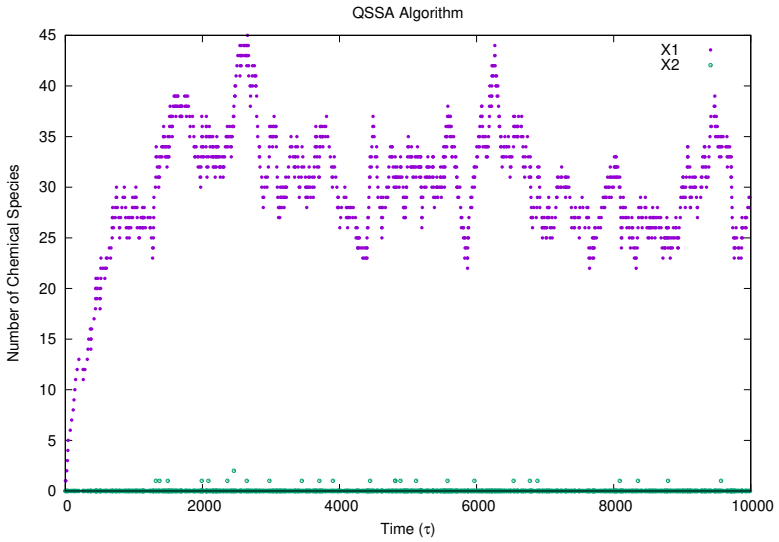
As we can see, the X_1 variable reaches an steady state while the number of bound promoter sites in the gene promoter region X_2 perform sporadic jumps from 0 to 1. This is due to the high value of the rate W_4 which is the largest one of the four when X_2 is different from 0. Meanwhile, the transition rates that describes the binding to the gene promoter region is the smallest one. Then, the system evolves by performing the synthesis and the degradation of the transcription factor X_1.

Rescaled parameter	Parameter	Units
$\kappa_1 = \frac{a}{k_{deg}\sqrt{K_d}}$	$K_d = 10$	nM
$\kappa_2 = 1$	$k_{deg} = 2$	min^{-1}
$\nu_{11} = 1$	$r = 0.4$	$\text{nM} \cdot \text{min}^{-1}$
$R = \frac{r}{k_{deg}\sqrt{K_d}}$	$S = 500$	
$b_{11}ES = k_{deg}$	$E = 5$	
$a = 5$		

Table 2: Parameter values used in simulations of the stochastic self-activating single gene regulatory system.



(a) Gillespie Algorithm



(b) QSSA Algorithm

Figure 2: Dynamics of a GRN model applying both methods of Gillespie and QSSA.

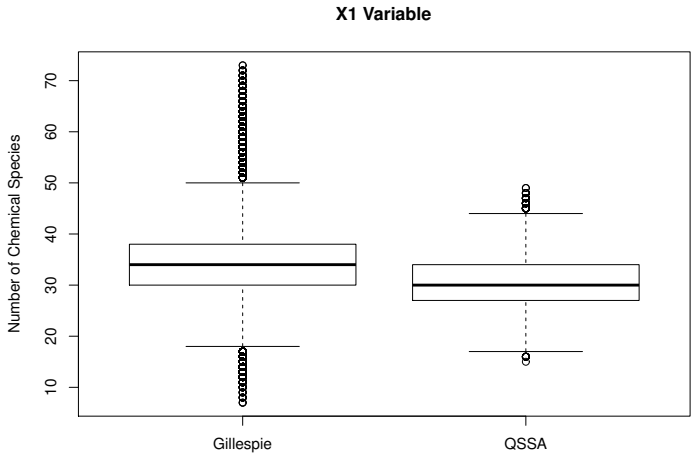


Figure 3: Boxplot corresponding to the simulation results for the X_1 variable and using both methods.

The simulation results obtained with both methods are presented right now. Considering the variable X_1, we have obtained that the mean of the transcription factor is $\langle X_1 \rangle = 34.18$ and its standard deviation is $\sigma_{X_1} = 6.34$. In that case, by using QSSA method, the mean of the X_1 variable was $\langle X_1 \rangle = 30.71$ and the standard deviation, $\sigma_{X_1} = 5.35$. These results are shown in Figure 3 by using a boxplot chart. As we can see, there is fraction of the most common values of the X_1 variable that coincides in both methods.

Even though both methods gives us an accurate image of the dynamics of the system, QSSA allows us to obtain a more detailed picture from the steady state. While, the Gillespie method performs an algorithm step every 7.0×10^{-3} minutes in average, the QSSA algorithm performs an step every 3.8 minutes in average. That drives to a faster performance for the QSSA method. On one hand, Gillespie takes almost 60 seconds to simulate 50000 minutes. And, on the other hand, QSSA spends 0.14 seconds. Which is a huge and important difference between both methods considering that both results gives us consistent results.

Conclusions

To conclude, throughout this project, we could derive the dynamics of the GRN system in its steady state by applying different rescalating parameters. This has allowed us to use the QSSA method to simulate the dynamics of the system. This approximation method, as we have seen above, generates very consistent results that can be compared with an exact method as it is the Gillespie Algorithm. In addition, the main advantage of this algorithm is its performance time which is lamost 120 times better than the execution time obtained with the Gillespie algorithm.

References

[1] Roberto De La Cruz et al. "The effects of intrinsic noise on the behaviour of bistable cell regulatory systems under quasi-steady state conditions". In: *Journal of Chemical Physics* 143.7 (2015).

[2] Hye Won Kang and Thomas G. Kurtz. "Separation of time-scales and model reduction for stochastic reaction networks". In: *Annals of Applied Probability* 23.2 (2013), pp. 529–583.

[3] Karen Ball et al. *Asymptotic analysis of multiscale approximations to reaction networks*. Vol. 16. 4. 2006, pp. 1925–1961.

[4] Núria Folguera-Blasco et al. "Supplemental Information: A multiscale model of epigenetic heterogeneity-driven cell fate decision-making". In: *PLoS Computational Biology* 15.4 (2019), pp. 1–23.