UPPSALA
UNIVERSITET

PROJECT REPORT

Modelling of biochemical reactions in cells

Jonathan Andersson, Agnes Johansson

Project in Computational Science: Report

January 2014

INSTITUTIONEN FÖR INFORMATIONSTEKNOLOGI

Abstract

Three different hybrid methods using explicit time steps are used for simulation of stochastic-deterministic models for the biochemistry in cells. They were compared by investigating their accuracy in different systems. The methods are the Euler forward method, the Lie-Trotter method and the Strang method, and they are used to simulate three different model systems: a simple example with two chemical species (irreversible heterodimerization), metabolites controlled by enzymes and the circadian rhythm.

The results showed that the Strang method seemed to be the best among all tested approaches. The Lie-Trotter method did not behave as expected and more research on that needs to be done. The value of the parameter coupling the species in the simple example had a great impact on the error, while changing parameters which only affects one of the species directly had a smaller impact.

Keywords: Hybrid method, Chemical reactions, Euler forward, Strang, Lie-Trotter

Table of Contents

1 Introduction

Each cell has a complicated biochemical network where chemical species move in space by diffusion, and react with each other to form new compounds. One cell may consist of many different chemical species and the modelling of the reactions between them is therefore very complex [1]. The massive acquisition of data in molecular and cellular biology has made it possible to use computational simulations of biological systems. Computational simulations are very important in the work to understand and predict quantitative behaviour of complex biological systems [2].

1.1 Models for biochemical reactions

Most computational models for the biochemical reactions in cells are based on the deterministic reaction rate equations that form a system of ordinary differential equations (ODEs) [3] [4]. This is a popular approximation of the mechanisms how the concentrations of species change in a well-mixed system. Each equation in the set typically represents the rate of change of a species' continuous concentration [1] [2] [4]. The underlying assumption is that the medium in which the reactions take place is spatially homogeneous and that the process is isothermal and isometric [4].

This deterministic approach is best suited to capturing the behaviour of systems where species are abundant and reaction events frequent. In that case the species concentrations are acceptably approximated as varying continuously and predictably [2]. When this is met, the deterministic approach is a powerful tool in the understanding of the dynamical behaviour of the chemical system [5].

However, molecular interactions are intrinsically random and cellular behaviour itself seems to reflect this randomness [2]. In living cells the number of molecules of the chemically active species such as proteins is often low [1]. Species that only exist in low copy number are subject to random fluctuations which cannot be neglected, and in many cases these fluctuations have a great impact on the behaviour of the system [5]. Examples of this kind of system can be found in gene regulation. Liu and Jia [7] have shown that noise plays important role in the genetic regulatory switch processes. When a gene controls its own expression, a simple model consists of a single gene transcribed to mRNA which in turn is translated to a transcription factor regulating the gene [4]. Typically the genes are present in one or two

copies and the copy number of mRNA is small [1]. Owing to such small numbers, the transcription is rather a stochastic than a deterministic process [5] [6].

A stochastic process approach can identify key physiological control parameters to which the behaviour of specific genetic regulatory systems is particularly sensitive [7]. One way to model coupled chemical reactions stochastically is to use the Stochastic Simulation Algorithm (SSA) introduced by Gillespie in 1976 [8]. The SSA is an exact method for numerically calculating the time evolution of any spatially homogeneous mixture of molecular species which interact through a specified set of coupled chemical reaction channels [8].

The SSA yields a correct realization of the process in the cell. However, computational complexity increases compared to the reaction rate equations [6]. Most systems are not analytically tractable, and the only way to analyse such processes is to simulate individual trajectories. If the number of reactions is large or if some molecular species appear in large numbers numerical computations become very time consuming [5].

In networks with many reactions or with reactants occurring in high numbers, the SSA becomes inefficient. On the other hand, small numbers of some of the members of a genetic network often makes an approximation by a purely deterministic approach inaccurate [5]. To be able to simulate realistic chemical reactions, without the computational complexity associated with SSA, the hybrid method is introduced [6].

The idea behind the hybrid method is to split the set of species Z into two sets $\mathrm{Z} \rightarrow (\mathrm{X}, \mathrm{Y})$. The species with low statistical variation (Y) are well represented by their mean values, and can be approximated by equations related to the reaction rate equations. The species with low copy number or large variation (X) are treated stochastically and need to be simulated with the SSA [6]. The stochastic set (X) and the deterministic set (Y) are simulated separately but are synchronized at some time t determined by a time stepping method.

1.2 Project definition

The aim of this project is to compare different methods for time stepping in the hybrid method by investigating their accuracy in different systems. Three different methods for time stepping are investigated: the well-known Euler forward method and two methods using operator splitting: due to Lie-Trotter and Strang. The methods chosen are not unique for stochastic-deterministic splitting but can also be used in different situations.

To analyse and compare the different methods they are used to simulate a simple example with two chemical species (irreversible heterodimerization). Selected methods are then used to simulate metabolites controlled by enzymes and the circadian rhythm. Each system will be described in detail in Section 3.

To measure accuracy, a reference solution is computed with the Piecewise Deterministic Markov Process (PDMP). The solution obtained from PDMP is chosen as a reference solution since it has the same stochastic-deterministic splitting as the other methods and gives an exact realization of the hybrid system. Therefore the other methods are expected to converge to the result of the PDMP method as the time step goes to zero. PDMP is described closer in Section 2.3 .

2 Method

In the following section the three time stepping methods and the method for calculating the reference solution are described.

2.1 Hybrid method

A chemical system with N molecular species $\mathrm{Z}_i, i = 1, \dots, N$, is characterised by the state vector $\boldsymbol{z} \in \mathbb{Z}_+^N$. The component z_i is the number of molecules of Z_i [4]. A reaction r, $r = 1, \dots, R$ in the system is a transition from a state $\boldsymbol{z_r}$ to $\boldsymbol{z}$ so that $\boldsymbol{z_r} = \boldsymbol{z} + \boldsymbol{n_r}$, where $\boldsymbol{n_r} \in \mathbb{Z}^N$ is the state-change vector. The probability that a reaction r occur per unit time is the non-negative propensity $w_r(\boldsymbol{z_r}, t)$. The change in the state vector caused by the reaction r can be written:

$$\boldsymbol{z_r} \xrightarrow{\boldsymbol{w_r(z_r,t)}} \boldsymbol{z}, \ \boldsymbol{n_r} = \boldsymbol{z_r} - \boldsymbol{z} \tag{1}$$

When splitting the species $\mathbf{Z}^T \rightarrow (\mathbf{X}^T, \mathbf{Y}^T)$ there is a corresponding split of the state vector $\mathbf{z}^T \rightarrow (\mathbf{x}^T, \mathbf{y}^T)$, such that $\mathbf{x} \in \mathbb{Z}_+^m$, and $\mathbf{y} \in \mathbb{R}^n$ with $N = m + n$, and a split of the transition vector $\boldsymbol{n}_r^T \rightarrow (\boldsymbol{m}_r^T, \boldsymbol{n}_r^T)$, with $\boldsymbol{m}_r \in \mathbb{Z}^m, \boldsymbol{n}_r \in \mathbb{Z}^n$ for $r = 1, \dots, R$. [6].

The different sets are then simulated separately and synchronized at some time t. The time is chosen with different time stepping methods which are described in the next section.

When using hybrid methods part of the system is simulated using ODEs and parts is simulated using SSA. The ODEs were solved using the fourth order Runge-Kutta method

(RK4) with time steps small enough for the errors to become negligible. The algorithm for SSA [8] is presented below.

Algorithm SSA

Set $t = 0, z = z_0$

Repeat

Select $\Delta t \sim \text{Exp}(w_0(z))$, where Exp is the exponential distribution

Update z with a reaction chosen randomly, the chance of a reaction r being chosen is w_r/w_0

Set $t = t + \Delta t$

Until done

Note: The sum of all propensities is denoted w_0.

2.2 Time stepping

The time stepping is conducted with the Euler forward, Lie-Trotter and Strang methods. Each method and algorithm will be explained below.

The error in the mean of the trajectories is expected to be

$$e = C * \Delta t^p + \frac{\sigma}{\sqrt{k}} \quad (2)$$

Where C is a constant depending on the time stepping method and the system being simulated. p is the convergence rate of the method, which is expected to be 1 for the Euler forward and Lie-Trotter methods and 2 for the Strang method. Since the simulations are partly stochastic we get a stochastic error with expected value $\frac{\sigma}{\sqrt{k}}$, where σ is the standard deviation and k is the number of trajectories. Since we are interested in finding p we need a large number of simulations to reduce the stochastic error.

2.2.1 The Euler forward method

The Euler forward method is a first-order numerical method for solving equations with a given initial value.

Algorithm Euler forward

Choose time step Δt, set $t = 0$, $\boldsymbol{x} = \boldsymbol{x_0}$ and $\boldsymbol{y} = \boldsymbol{y_0}$

Repeat

Advance $\boldsymbol{x}$ from t to t + Δt with SSA with $\boldsymbol{y}$ frozen at t

Solve ODE system for $\boldsymbol{y}$ from t to t + Δt with $\boldsymbol{x}$ frozen at t

set $t = t + \Delta t$

Until $t = t_{end}$

2.2.2 The Lie- Trotter method

The Lie-Trotter method (LT) is also a first-order numerical method. It can be implemented in two different ways, one where the deterministic part is solved before the stochastic, henceforth called LT-ds, and one where the stochastic part is solved before the deterministic one, henceforth called LT-sd. Where the d stands for deterministic and s stands for stochastic and the order is meant to represent the order in which the deterministic and the stochastic part is solved.

Algorithm LT-ds

Choose time step Δt, set $t = 0$, $\boldsymbol{x} = \boldsymbol{x_0}$ and $\boldsymbol{y} = \boldsymbol{y_0}$

Repeat

Solve ODE system for $\boldsymbol{y}$ from t to $t + \Delta t$ with $\boldsymbol{x}$ frozen at t

Advance $\boldsymbol{x}$ from t to $t + \Delta t$ with SSA with $\boldsymbol{y}$ frozen at $t + \Delta t$

set $t = t + \Delta t$

Until $t = t_{end}$

Algorithm LT-sd

Choose time step Δt, set $t = 0$, $\boldsymbol{x} = \boldsymbol{x_0}$ and $\boldsymbol{y} = \boldsymbol{y_0}$

Repeat

Advance $\boldsymbol{x}$ from t to $t + \Delta t$ with SSA with $\boldsymbol{y}$ frozen at t

Solve ODE system for $\boldsymbol{y}$ from t to $t + \Delta t$ with $\boldsymbol{x}$ frozen at $t + \Delta t$

set $t = t + \Delta t$

Until $t = t_{end}$

2.2.3 The Strang method

The Strang method is a second-order numerical method [9].

Algorithm Strang method

Choose time step Δt, set $t = 0$, $\boldsymbol{x} = \boldsymbol{x_0}$ and $\boldsymbol{y} = \boldsymbol{y_0}$

Repeat

Solve ODE system for $\boldsymbol{y}$ from t to $t + \frac{\Delta t}{2}$ with **x** frozen at t

Advance $\boldsymbol{x}$ from t to $t + \Delta t$ with SSA with $\boldsymbol{y}$ frozen at $t + \frac{\Delta t}{2}$

Solve ODE system for $\boldsymbol{y}$ from $t + \frac{\Delta t}{2}$ to $t + \Delta t$ with $\boldsymbol{x}$ frozen at $t + \Delta t$

set $t = t + \Delta t$

Until $t = t_{end}$

2.3 Piecewise deterministic Markov Processes

The Piecewise Deterministic Markov Process (PDMP) was formalized by [10] and has since then found many applications in various areas, from finance to biology [11]. It can be thought of as a mixture of a deterministic and a stochastic approach [5].

The PDMP constitute a subclass of Markov processes, by which events can be modelled in a general way with a split set of chemical species(X, Y). Up to a random jump time the process develops deterministically with constant stochastic variables. At the jump time the process moves randomly to a new state [5]. To define a PDMP we follow the construction given by Davis in [10]. The PDMP is implemented as in [5].

Algorithm Pricewise Deterministic Markov Process

Set $t = 0$, $\boldsymbol{x} = \boldsymbol{x_0}$ and $\boldsymbol{y} = \boldsymbol{y_0}$

Repeat

Choose $u \sim$ unif[0,1], where unif is the continuous uniform distribution

Set $F = 0$

Solve the coupled differential equations system:

$$\begin{cases} \frac{d}{dt}\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{y}) \\ \frac{d}{dt}F = w_0(\boldsymbol{x}, \boldsymbol{y})(1 - F) \end{cases}$$

Until $F = u$

Add the time taken for F to become equal to u to t

Update $\boldsymbol{x}$ using one reaction chosen with SSA

Until done

Note: $\boldsymbol{y}$ is updated in the coupled differential equations system and $\boldsymbol{x}$ is constant while $\boldsymbol{y}$ is updated. The sum of all propensities for reactions involving the stochastically treated species is denoted w_0.

3 Systems

In this section we describe the three systems that have been analysed and the result when the different methods were applied.

3.1 Irreversible heterodimerization

Irreversible heterodimerization is used as a simple example, which can be seen as a part of bigger and more complex systems. The size of the system in the simple example makes it suitable for numerical experiments.

The system consists of two chemical species X and Y. The chemical reactions between the species X and Y are written:

$$\begin{array}{ll} \boldsymbol{X} + \boldsymbol{Y} \xrightarrow{\boldsymbol{K_{xy}xy}} \emptyset & \\ \boldsymbol{X} \xrightarrow{\boldsymbol{K_x x}} \emptyset & \boldsymbol{Y} \xrightarrow{\boldsymbol{K_y y}} \emptyset \\ \emptyset \xrightarrow{\boldsymbol{c_x}} \boldsymbol{X} & \emptyset \xrightarrow{\boldsymbol{c_y}} \boldsymbol{Y} \end{array} \quad (3)$$

The reaction constants are given in Table 1 and a deterministic simulation of the system is shown in Figure 1.

Table 1. The reaction constants for the simple example.

K_{xy}	K_x	K_y	C_x	C_y
1	0.2	0.2	3.0	2.08

Table 2. The starting values for the simple example.

X	Y
0	0

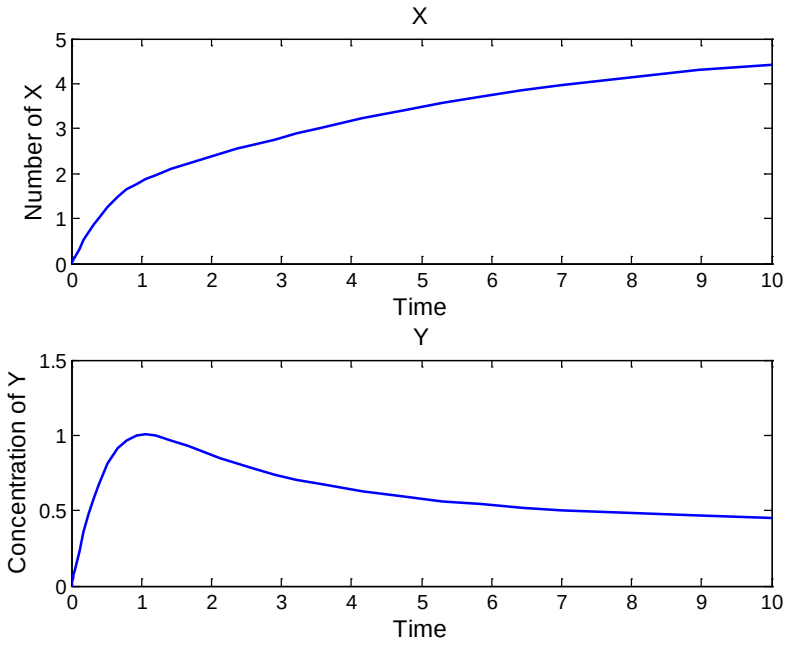

Figure 1. ODE simulation of the system in simple example, X and Y are the chemical species.

As seen in Figure 1, the system has no oscillations and quickly converges to a steady state. In the experiments presented below the end time is therefore set to t = 4, before the system has reached the steady state.

3.1.1 Numerical result and discussion of the Irreversible heterodimerization

The solid blue line in Figure 2, Figure 5, Figure 8, and Figure 10 represent the expected size of the stochastic noise of the simulations. It is calculated as two times the expected stochastic error of the PDMP simulations. It is included to give a feeling of how important the stochastic noise is in the plots. Since we are interested in how well the different methods work, we are

not interested in situations where the stochastic error dominates. Instead we are interested in minimizing the systematic error caused by the different methods.

Variation of the time step

To investigate what effect the time step has on the different methods 1,000,000 trajectories are simulated for each method and time step, with time steps from $1/12$ to 4. The result is presented in convergence plots of the mean value error (Figure 2) and the variance error compared with PDMP (Figure 3) and the Kolmogorov distance to the distribution when using PDMP (Figure 4). The Kolmogorov distance is defined as

$$D_{kolm} = \sup_x |F_1(x) - F_2(x)| \qquad (4)$$

where F_1 and F_2 are cumulative distribution functions.

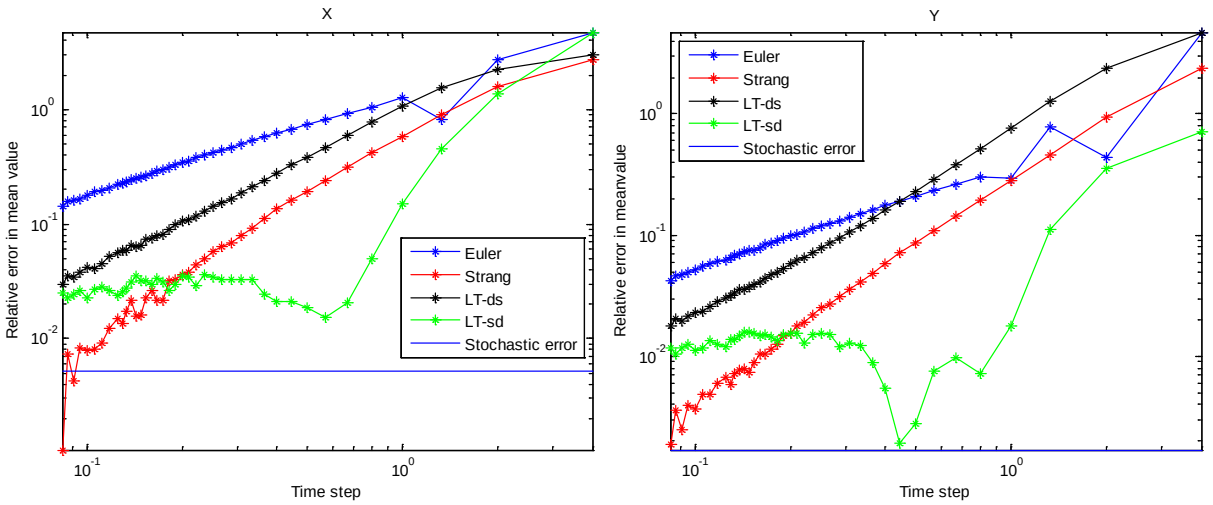

Figure 2. Relative error in mean value for the stochastic species (X) and the deterministic species (Y) between the different methods and PDMP.
Measured convergence rates, X: Euler: 0.86, Strang: 1.85, LT-ds: 1.52.
Measured convergence rates, Y: Euler: 0.87, Strang: 1.86, LT-ds: 1.41.

Figure 2 shows the convergence plot of the mean value of the trajectories for the different methods for the stochastic (X) and the deterministic species (Y). The plots for the stochastic and deterministic sets look quite similar for all methods. The convergence rate for the mean values of all the trajectories, seen as the slope in the figures, is a little lower than expected for Euler forward and the Strang method. The Lie-Trotter method is a first order method and is expected to have a convergence rate of 1. However for LT-ds the measured convergence rate is significantly larger than expected. LT-sd converges very quickly but not to the value obtained by PDMP. None of the two Lie-Trotter methods behaves as expected, and they are

not behaving like each other. These results are unexpected and the cause of the discrepancy with theory is unknown.

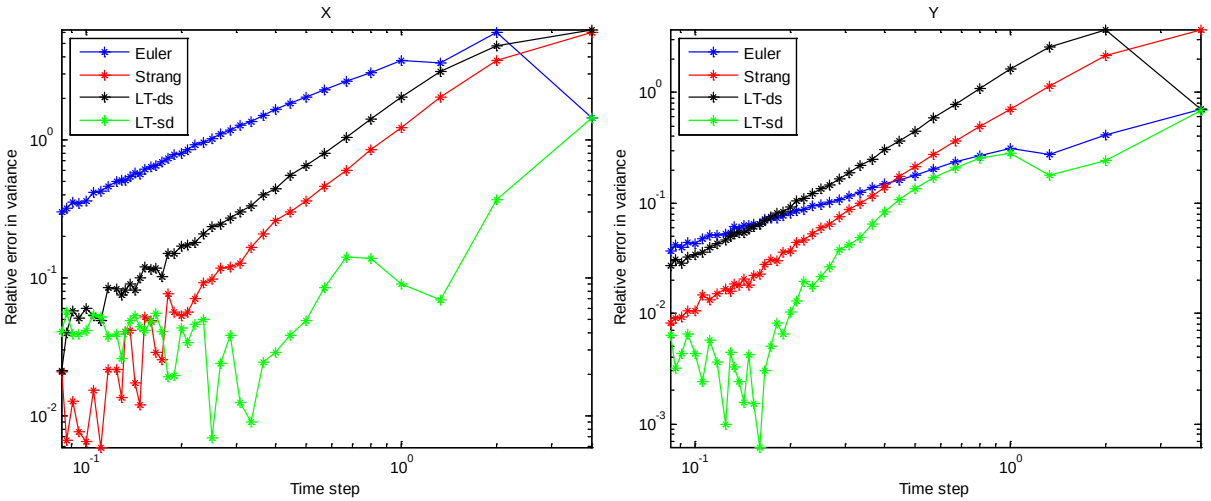

Figure 3. Convergence plot of the variance for the stochastic species (X) and the deterministic species (Y).
Measured convergence rates, X: Euler: 1.03, Strang: 1.78, LT-ds: 1.54.
Measured convergence rates, Y: Euler: 0.86, Strang: 1.81, LT-ds: 1.63.

In Figure 3 it is seen how the error in variance, σ^2 in equation (2), varies with different time steps. We see that the behaviour of the variance is close to the behaviour of the mean value in Figure 2. The main difference seems to be that LT-sd converges slower. The convergence rates are close to the values for the mean value and LT-sd does not seem to converge to the correct value for the variance either.

For the Strang method we see some fluctuations in the error for small time steps in all cases, especially for the variance. This is because the error in that case lies close to the stochastic error which means that the stochastic error has a greater impact on the total error which we can see in Figure 2.

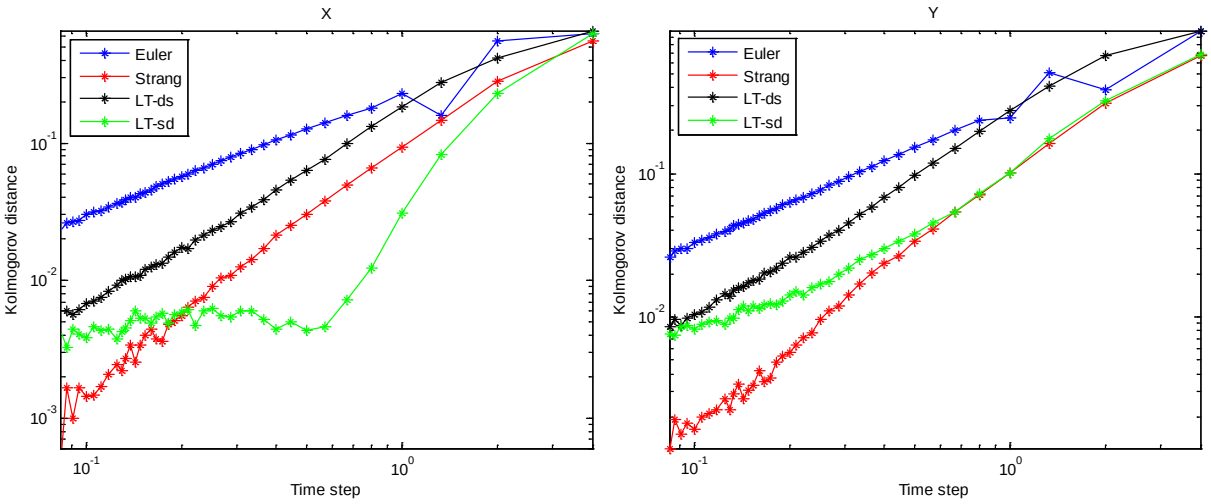

Figure 4. Kolmogorov distance between the different methods and PDMP for different time steps for the stochastic species (X) and the deterministic species (Y).
Measured convergence rates, X: Euler: 0.90, Strang: 1.78, LT-ds: 1.40.
Measured convergence rates, Y: Euler: 0.96, Strang: 1.81, LT-ds: 1.39.

Figure 4 shows the Kolmogorov distance for the different methods. The Kolmogorov distance is used to compare a sample of the different methods with the reference solution, the PDMP. When the time step is large the difference between the different methods is small. However, the Strang method has a higher convergence rate than the other methods, and for small time steps the Strang method has a much smaller Kolmogorov distance than the other methods. Again the Lie-Trotter method does not behave as expected. The convergence rates are close to what was observed for the mean value. The LT-sd behaves inexplicably as before.

Variation of parameter K_{xy}

To get an idea of how different aspects of the system affected the result of the simulations the parameters K_{xy}, K_x, and K_y in (3) are changed. K_{xy} is the coupling factor that couples the two species together. A constant time step of 0.25 is used and K_{xy} is varied from 1 to 10 while the other parameters retain their values from Table 1.

The result is presented in plots of the mean value error (Figure 5) and the variance error compared with PDMP (Figure 6) and the Kolmogorov distance to the distribution when using PDMP (Figure 7).

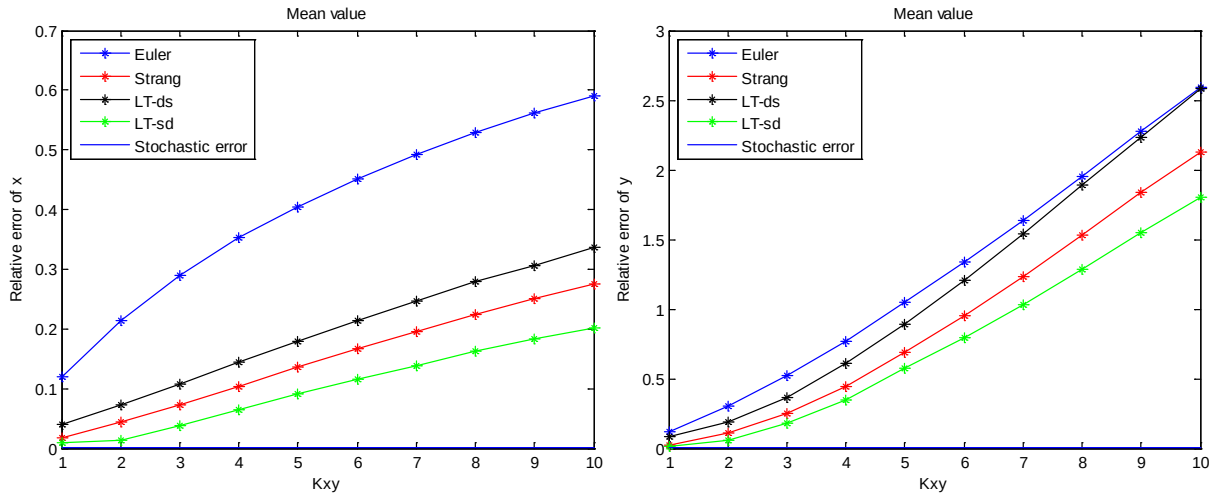

Figure 5. Relative error in mean value for stochastic species X and the deterministic species Y between the different methods and PDMP, for different values of the coupling constant K_{xy}.

As we see in Figure 5 the error increases when the coupling constant increases. This is expected since a larger coupling factor makes the chemical species more interdependent. Figure 5 show that the Strang method is better than the Euler forward method regardless of the size of the coupling constant.

The error for the deterministic species grows faster than the error for the stochastic species in and quickly becomes far too large for the result of the simulations to be of practical use. It can be seen in Figure 5 that at least for smaller K_{xy} the error for the stochastic species grows faster for the Euler forward method than for the other methods.

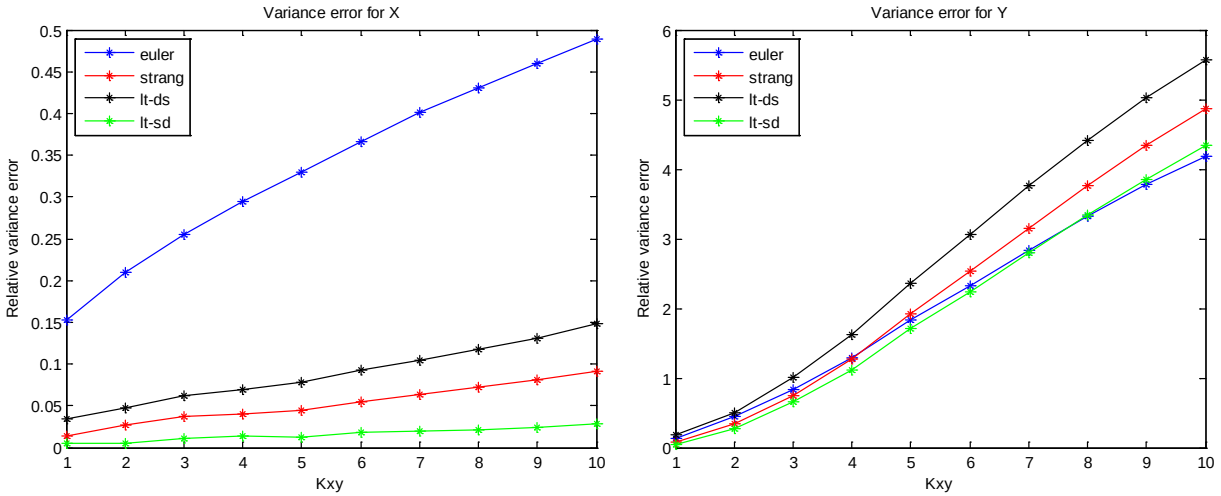

Figure 6. Relative error in variance for the stochastic species X and the deterministic species Y between the different methods and PDMP, for different value of the coupling constant K_{xy}.

As seen in Figure 6 the error in the variance increases in ways similar to the error in the mean values.

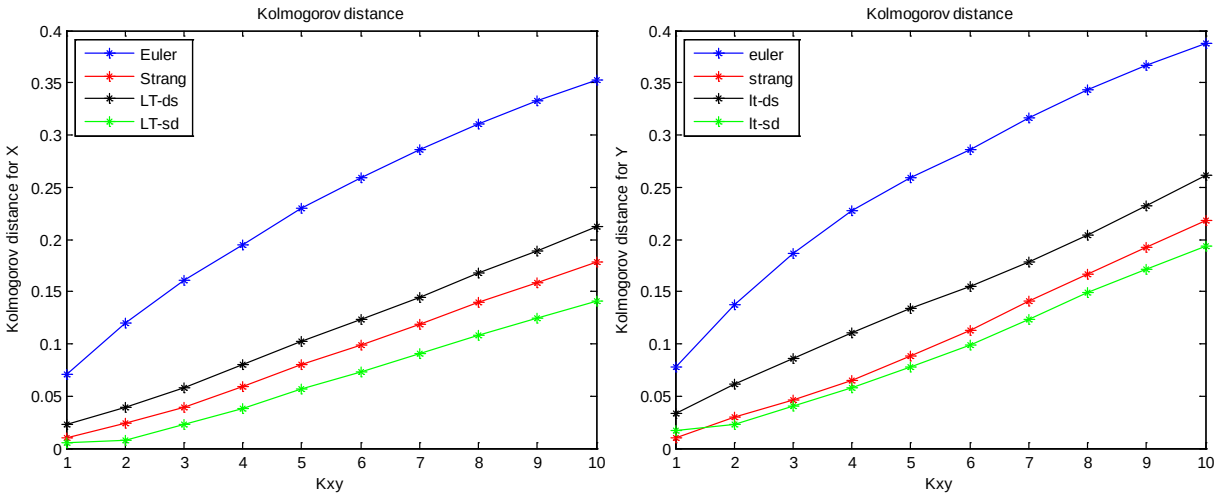

Figure 7. Kolmogorov distance between the different methods and PDMP, and different coupling constants K_{xy} for the stochastic species (X) and the deterministic species (Y).

The Kolmogorov distance for different values of the coupling factor is found in Figure 7. As in Figure 5, the error increases with larger values of the coupling factor. For smaller K_{xy} the Kolmogorov distance increases faster for the Euler forward method.

Variation of parameter K_x

To get an idea of how different aspects of the system affected the result of the simulations the parameters K_{xy}, K_x, and K_y in (3) are changed. K_x is the rate at which the stochastic species (X) vanish. A constant time step of 0.25 is used and K_x is varied from 0.2 to 2 while the other parameters retain their values from Table 1.

The result is presented in plots of the mean value error (Figure 8) and the Kolmogorov distance to the distribution when using PDMP (Figure 9). The variance behaves similar to the mean value, as for the case with K_{xy} in Figure 6, and will therefore not be presented.

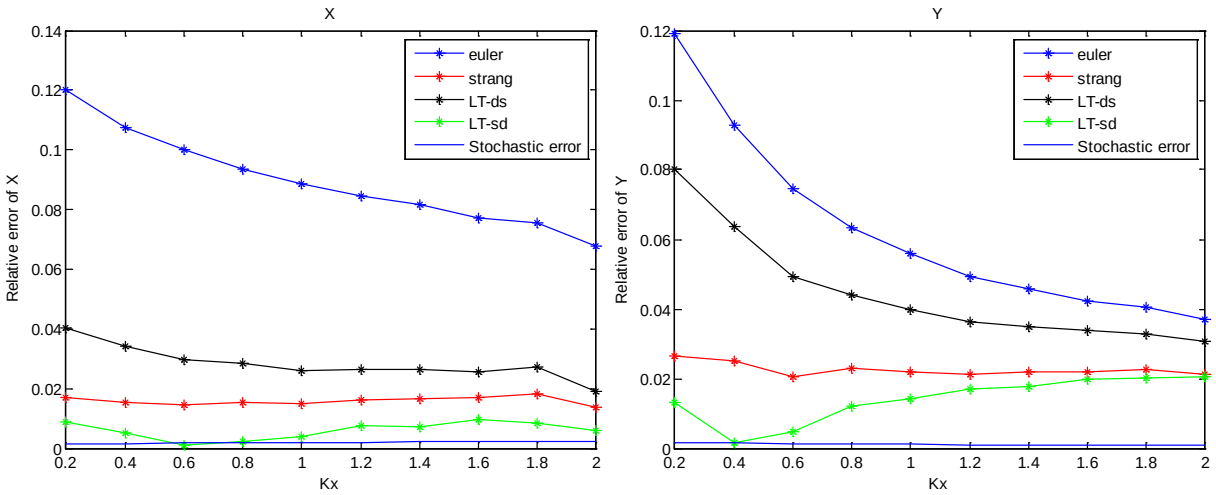

Figure 8. Relative error in mean value for the stochastic species (X) and the deterministic species (Y) between the different methods and PDMP, for different size of K_x.

From Figure 8 we find that the error for the deterministic species does not vary much for any method except the Euler forward method. The relative error of the mean value for the deterministic species decreases as K_x increases, for both Euler forward and LT-ds. The explanation may be that it is closer to its steady state since the system becomes faster. It seems that the rate in which the stochastic species (X) vanishes does not affect the error that much for the Strang method in this case.

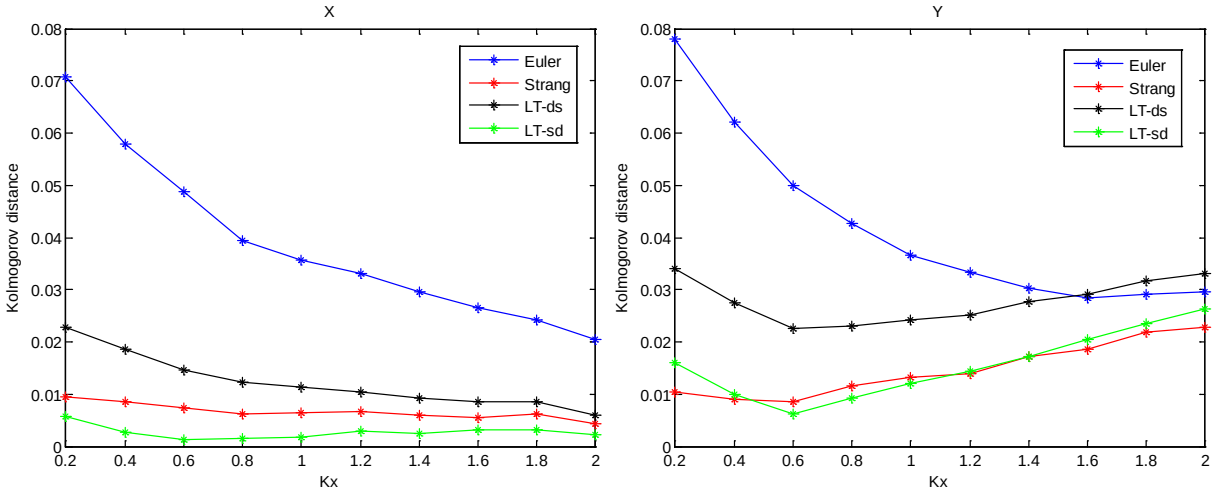

Figure 9. Kolmogorov distance for the stochastic species (X) and the deterministic species (Y) between the different methods and PDMP, for different K_x.

The Kolmogorov distance for the different methods, for different K_x is seen in Figure 9. For the stochastic species (X) the error decreases a little when the parameter value increases. This is probably because the solution for the stochastic set gets close to steady-state faster when the

rate with which one of the sets goes to zero increases. For the deterministic set (Y) the difference between the different methods decrease as K_x increase. In that case the errors for the Strang method, LT-sd and LT-ds increase while the error for Euler forward decreases as K_x increases.

Variation of K_y

To get an idea of how different aspects of the system affected the result of the simulations the parameters K_{xy}, K_x, and K_y in (3) are changed. K_y is the rate at which the deterministic species (Y) vanish. A constant time step of 0.25 is used. K_y is varied from 0.2 to 2 while the other parameters retain their values from Table 1.

The result is presented in plots of the mean value error (Figure 10) and the Kolmogorov distance to the distribution when using PDMP (Figure 11). The variance behaves similar to the mean value, as for the case with K_{xy} in Figure 6, and will therefore not be presented.

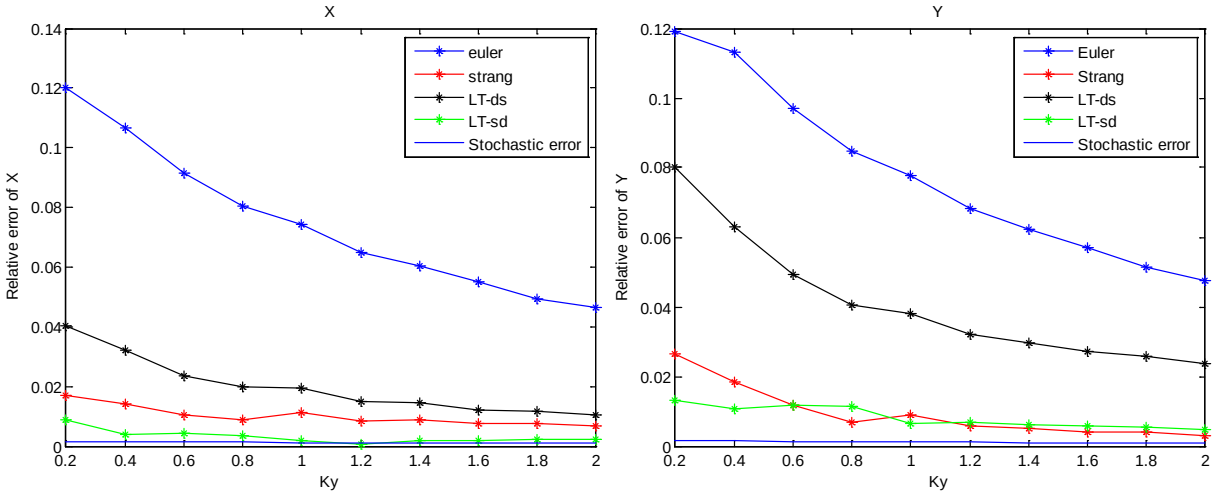

Figure 10. Relative error in mean value for the stochastic species (X) and the deterministic species (Y) between the different methods and PDMP, for different size of K_y.

Figure 10 depicts how the relative error in the mean value between the different methods and PDMP decreases as the rate (K_y) in which the deterministic set (Y) goes to the empty set increases. As before, the explanation may be that the solution at the final time is closer to the steady state.

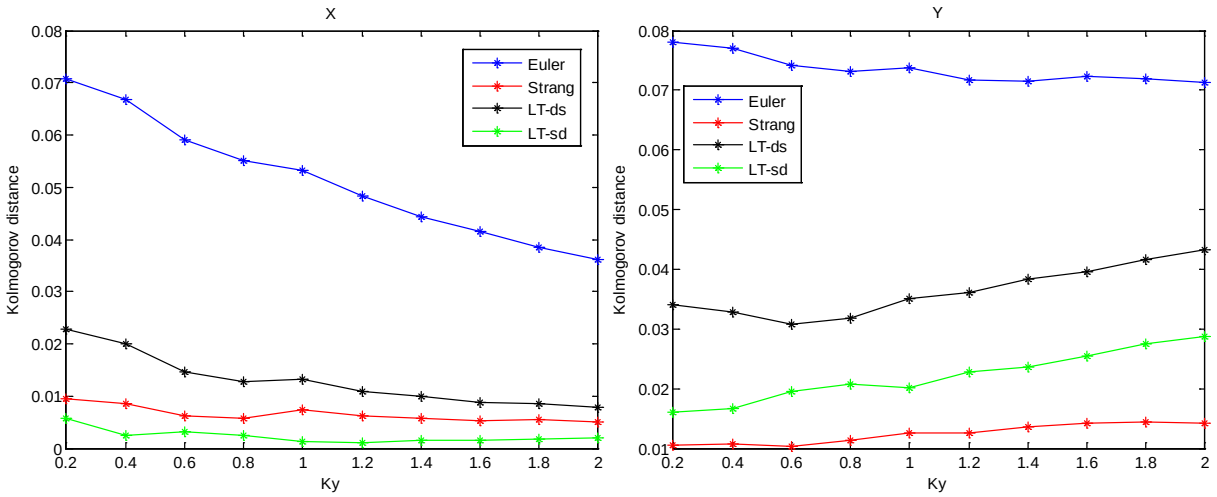

Figure 11. Kolmogorov distance for the stochastic species (X) and the deterministic species (Y) between the different methods and PDMP, for different K_y.

The Kolmogorov distance for the different methods and for different K_y is found in Figure 11. For the stochastic species (X) in Figure 11 the result is similar to the case with K_x in Figure 9, the error decreases as K_y increases.

For the deterministic species (Y) the increase of K_y makes the errors increase for all methods except for Euler forward, for which the error even decreases a little. There is a big difference between the errors for the different methods, no matter the size of K_y.

3.1.2 Summary of the Irreversible heterodimerization

The simple example with two chemical species shows that the Strang method and Euler forward converge as expected. The Lie-Trotter method however does not behave as expected. Instead of converging with a rate of 1, as expected since the method is of order 1, it converges with a rate close to 1.5 or quickly converged to some incorrect value depending on if the stochastic part was calculated before or after the deterministic part. It could be of interest to further investigate the cause of this behaviour.

The simple example also shows that the larger the value of the coupling factor is, the larger the error get for all hybrid methods tested. The error grows larger faster for the deterministic variable and the coupling factor does not need to be increased much to make the result practically useless.

The factors that control the rate with which the species vanish do not affect the error as much as the coupling factor. But an increased disappearance rate makes the error smaller in some

cases, even if the difference is not as big as for the other factors. The reason may be that the system is closer to the steady state in these cases at the final time. However the error could also increase and the error is far from the stochastic error in most cases, so there might be some other explanation.

3.2 Metabolites controlled by enzymes

This is a simple generic model with two metabolites A and B and two enzymes E_A and E_B as in [6] and [13]. The system is small and the behaviour of different parts of the algorithm are therefore easily investigated. This might provide insight into some of the limitations of the hybrid algorithm, and give guidelines for how to choose some of the parameters [6].

The production of the metabolites, A and B, is regulated by the enzymes. The reactions are:

$$
\begin{array}{ll}
\boldsymbol{A + B \xrightarrow{k_2 ab} \emptyset} & \\
\emptyset \xrightarrow{\frac{k_a e_A}{1+\frac{a}{K_i}}} \boldsymbol{A} & \emptyset \xrightarrow{\frac{k_b e_B}{1+\frac{b}{K_i}}} \boldsymbol{B} \\
\boldsymbol{A \xrightarrow{\mu a} \emptyset} & \boldsymbol{B \xrightarrow{\mu b} \emptyset} \\
\emptyset \xrightarrow{\frac{k_{eA}}{1+\frac{a}{K_r}}} \boldsymbol{E_A} & \emptyset \xrightarrow{\frac{k_{eB}}{1+\frac{b}{K_r}}} \boldsymbol{E_B} \\
\boldsymbol{E_A \xrightarrow{\mu e_A} \emptyset} & \boldsymbol{E_B \xrightarrow{\mu e_B} \emptyset}
\end{array}
\qquad (5)
$$

The reaction constants are given in Table 2. The system is partitioned as in [6] so that the metabolites A and B are treated as stochastic variables and the enzymes E_A and E_B are treated deterministically.

Table 2. The reaction constants for the metabolite-enzyme model.

k_2	k_a	k_b	K_i	μ	k_{eA}	k_{eB}	K_r
0.001	0.3	0.3	60	0.002	0.02	0.02	30

Table 3. Starting values for the metabolite-enzyme model.

A	B	E_A	E_B
33	33	6	6

3.2.1 Numerical result and discussion of the Metabolites controlled by enzymes

The probability density function for the metabolites and the enzymes, calculated from 8000 trajectories of PDMP, is plotted in Figure 12.

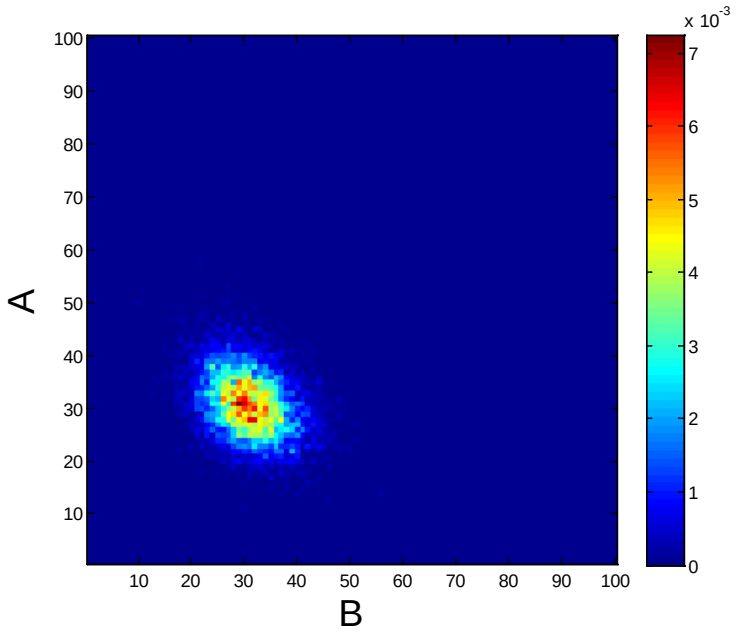

Figure 12. Probability density function at t=1000 for A and B computed with PDMP.

The PDMP does not capture the characteristic shape of the marginal distribution of A and B sufficiently well compared to the figure plotted with SSA in [6]. The result is the same as for the hybrid method in [6]. The result is expected since the simplifying assumptions in the derivation of the underlying equations are violated because of the significant correlation between the species in the stochastic and deterministic set. A good splitting should ideally keep highly correlated species within the stochastic subset [6]. This is not the case here and since our reference solution does not give a realistic simulation the other hybrid methods are not tested.

3.3 Circadian rhythm model

A wide range of organisms use circadian clocks to keep an internal sense of daily time and regulate their behaviour accordingly. Most of these clocks use intracellular genetic networks based on positive and negative regulatory elements [12]. Oscillator models are control systems which assume periodic oscillators of certain molecular species to appear in order to establish a circadian rhythm in the organism [6].

The model used is the same as in [12] and [6]. It has nine variables. It involves two genes, D_a and D_r, an activator A and a repressor R, which are transcribed into mRNA and subsequently translated into protein, M_a and M_r. The activator and repressor can associate and form a complex C, in which the activator A is degraded. The variables $D_a{}'$ and $D_r{}'$ are the genes D_a and D_r with a bound activator. In the model it is assumed that there is only one gene coding for the repressor and the activator. Thus $D_a + D_a' = 1$, and the same holds true for the repressor gene. The reactions for the nine molecular species are:

$$\left.\begin{array}{c} \boldsymbol{D'_a \xrightarrow{\theta_a D'_a} D_a} \\ \boldsymbol{D_a + A \xrightarrow{\gamma_a D_a A} D'_a} \\ \boldsymbol{D'_r \xrightarrow{\theta_r D'_r} D_r} \\ \boldsymbol{D_r + A \xrightarrow{\gamma_r D_r A} D'_r} \end{array}\right\} \left.\begin{array}{c} \boldsymbol{\emptyset \xrightarrow{\alpha'_a D'_a} M_a} \\ \boldsymbol{\emptyset \xrightarrow{\alpha_a D_a} M_a} \\ \boldsymbol{M_a \xrightarrow{\delta_{ma} M_a} \emptyset} \end{array}\right\} \left.\begin{array}{c} \boldsymbol{\emptyset \xrightarrow{\beta_a M_a} A} \\ \boldsymbol{\emptyset \xrightarrow{\theta_a D'_a} A} \\ \boldsymbol{\emptyset \xrightarrow{\theta_r D'_r} A} \\ \boldsymbol{A \xrightarrow{\delta_a A} \emptyset} \\ \boldsymbol{A + R \xrightarrow{\gamma_c AR} C} \end{array}\right\} \left.\begin{array}{c} \boldsymbol{\emptyset \xrightarrow{\alpha'_r D'_r} M_r} \\ \boldsymbol{\emptyset \xrightarrow{\alpha_r D_r} M_r} \\ \boldsymbol{M_r \xrightarrow{\delta_{mr} M_r} \emptyset} \end{array}\right\} \left.\begin{array}{c} \boldsymbol{\emptyset \xrightarrow{\beta_r M_r} R} \\ \boldsymbol{R \xrightarrow{\delta_r R} \emptyset} \\ \boldsymbol{C \xrightarrow{\delta_a C} R} \end{array}\right\} \quad (6)$$

The reaction constants are found in Table 4. The partitioning is the same as in [6], i.e. A and R are treated as stochastic variables, while the rest of the species are treated deterministically.

Table 4. Constants for the Circadian rhythm model.

α_A	α'_a	α_r	α'_r	β_a	β_r	δ_{ma}	δ_{mr}	δ_a	δ_r	γ_a	γ_r	γ_c	Θ_a	Θ_r
50	500	0.01	50	50	5	10	0.5	1.0	0.2/0.08	1.0	1.0	2.0	50	100

Table 5. Starting values for the Circadian rhythm model.

D_a	D_r	D'_a	D'_r	M_a	A	M_r	R	C
0.2	0.2	0	0	0	0	0	0	10

In this type of control system there is a high correlation between many of the species. Even if the species A and R have large copy numbers most of the time, at critical time intervals they drop near zero [6]. Two different values of δ_r are used, 0.2 and 0.08. The typical behaviour of the circadian rhythm with δ_r = 0.2 when PDMP is used is shown in Figure 13.

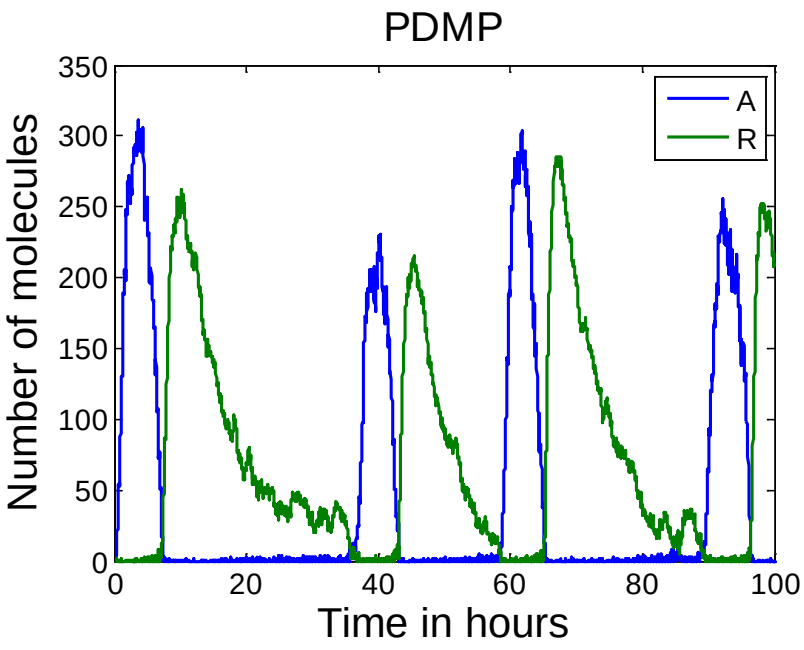

Figure 13. Typical behavior of circadian rhythm simulated with PDMP when δ_r = 0.2.

3.3.1 Numerical result and discussion of the Circadian Rhythm

Since the Strang method is the best method for the simple example (irreversible heterodimerization), the circadian rhythm is only simulated with the Strang method and the PDMP as a reference solution.

To investigate how well the Strang method simulates the circadian rhythm for different size of the time step the circadian rhythm is simulated for 1000 trajectories for both values of δ_r. Each one of the trajectories is simulated for 1000 hours. The result for two different time steps, 0.01 h and 0.32 h when δ_r is 0.2, is shown in Figure 14.

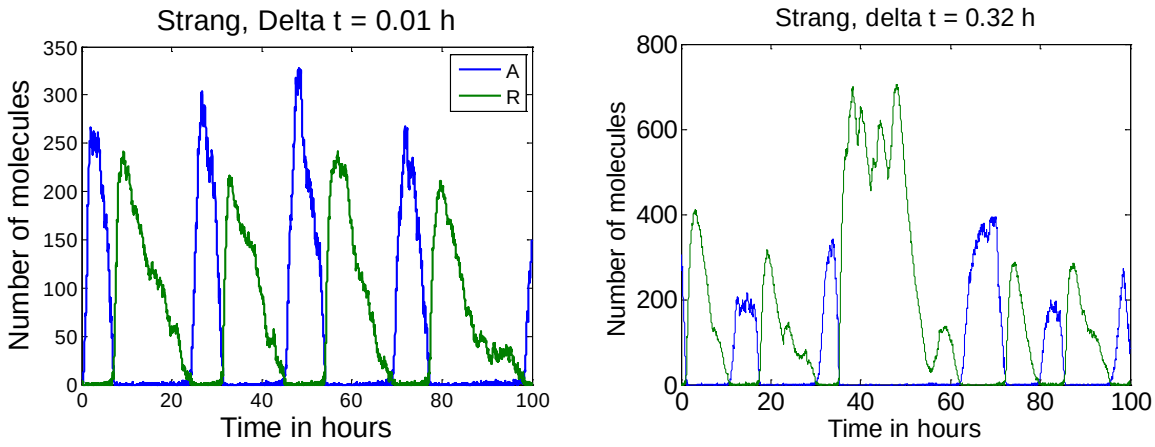

Figure 14. Typical behaviour for Circadian Rhythm simulated with the Strang method with Δt *= 0.01 h and* Δt *= 0.32 h when* δ_r *= 0.2.*

In Figure 14 it is seen that the Strang method with the time step 0.01 h gives a simulation that looks quite similar to the one with the PDMP in Figure 13. However, when the time step increases to 0.32 h the amplitudes of the peaks as well as the length of the periods become more random. To see this better the autocorrelation for the PDMP and for the Strang method with time steps 0.01 h and 0.32 h is plotted in Figure 15. The autocorrelation is defined as

$$A_k = \frac{\sum_{i=1}^{N-k} Y_i \cdot Y_{i+k}}{N-k} \qquad (7)$$

Where Y_n is the measurement at point n, k is the distance between the points and N is the total amount of points.

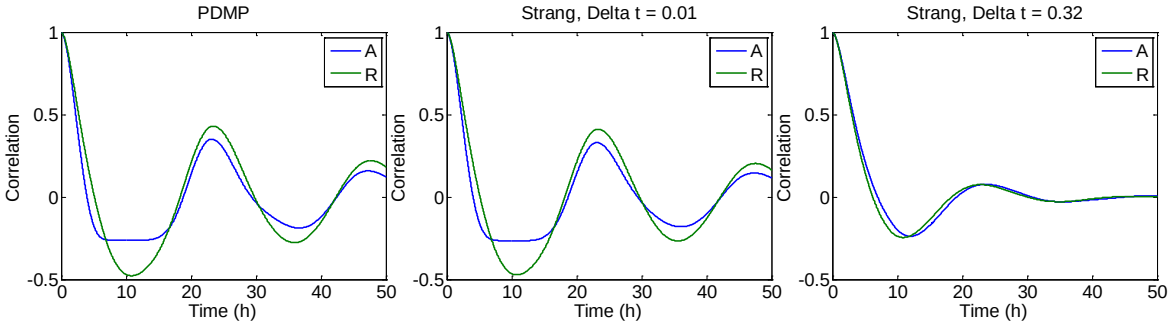

Figure 15. Autocorrelation of the circadian rhythm using PDMP and the Strang method using different Δt *for* $\delta_r = 0.2$.

As we see in Figure 15 the autocorrelation plots flattens out as the time step increases due to the increased randomness. To measure convergence the 2-norm of the difference between simulations using different time steps and PDMP was measured. When calculating the

autocorrelation measurements of a maximal distance of 50 hours is used when $\delta_r = 0.2$ and a maximal distance of 100 hours when $\delta_r = 0.08$ since the system has a longer period for that value.

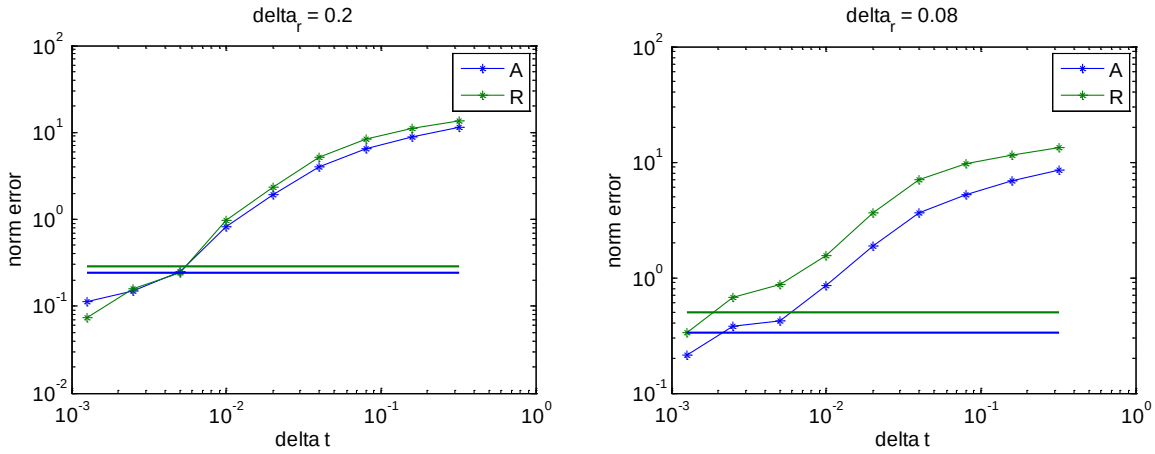

Figure 16. Convergence plots for both values of δ_R, the straight lines are the stochastic error of the PDMP method multiplied by 2.

As we see in Figure 16 the Strang method converges for both values of δ_R for our choice of measuring the convergence. It is noted that for $\delta_R = 0.2$ there is a stochastic reaction approximately every 0.0024 h when using PDMP while for $\delta_R = 0.08$ there is one approximately every 0.0034 h. These values could be interesting to compare to the size of the time step in the Strang method to find out which method is the fastest given some error tolerance.

3.3.2 Notes about the ODE solver for the Circadian Rhythm

RK4 is used as the numerical solver which leads to exponential growth of the number of molecules with time even for very small time steps, an example can be seen in Figure 17.

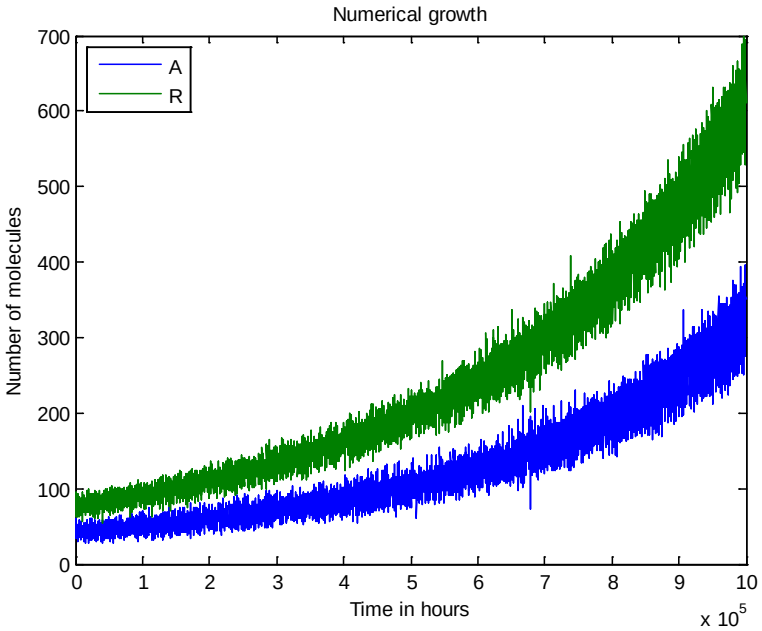

Figure 17. Mean value of A and R for windows of 100 hours for one trajectory when $\delta_R = 0.2$, 160,000 ODE time steps per hour is used. The hybrid method used is the PDMP.

The same problem is found when using a pure deterministic simulation of the system. The problem is probably due to the stiffness of the system, and when using MATLAB's ode15s solver no growth is found. It would probably be a good idea to use an implicit ODE solver instead of an explicit one or using adaptive time steps, especially when making long time series.

All results for the Circadian Rhythm uses 160,000 ODE time steps per hour of simulated time, and this part of the calculation takes most of the time. However the growth is expected to have little impact on the results since the length of the time span we simulated was quite short (1000 hours as mentioned above).

4 Conclusion

The aim of this project was to compare three different methods for time stepping in the hybrid method by investigating their accuracy in different systems. The time stepping methods we have looked at are the Euler forward method, the Lie-Trotter method and the Strang method, and the systems we have simulated were a simple example with two chemical species, the circadian rhythm and metabolites controlled by enzymes.

The results show that the Strang method is the best method, with smallest error and fastest convergence rate for the simple example. However when using big time steps the difference between the methods is small, although the error is then too great for the result to be of any practical use.

The coupling factor plays an important role for how well the simulation works in the simple example. The greater the coupling factor gets the greater the error becomes for all methods. The value of the factors only affecting one of the species have some, but not a large, impact on the error.

Metabolites controlled by the enzymes is not possible to simulate with hybrid methods. The system has too many highly correlated species, and the splitting of the chemical species is done with incorrect assumptions.

Surprising is that the Lie-Trotter method does not behave as we expected. The result differs a lot depending on if the stochastic or the deterministic part is simulated first. But none of the methods converge as expected. To find out why more research is needed.

5 Future research

This project has improved the understanding for how different hybrid methods works, and when they encounter problems. However, more research can be done to increase the knowledge.

Especially more research need to be done on the Lie-Trotter method. Researchers should look into why and when LT-ds converge faster than expected and why and when LT-sd converges to the wrong values. As an example the simple example could be simulated with both species being treated stochastically or both species being treated deterministically to see if the methods still behave unexpectedly.

Research needs to be done to find a way to a priori find which hybrid method is the most efficient given a chemical system and an error tolerance.

6 References

[1] Engblom, S., Ferm, L., Hellander, A. & Lötstedt, P., 2009. Simulation of Stochastic Reaction-diffusion processes on unsctructed meshes. *SIAM Journal of Scientific Computing*, 31(3), pp. 1774-1797.

[2] Di Ventura, B., Lemerle, C., Michalodimitrakis, K., & Serrano, L., 2006. From in vivo to in silico biology and back. *Nature*, 443, pp. 527-533.

[3] Elowitz, M. B., Levine A. J, Siggia E. D. & Swain, P. S., 2002. Stochastic gene expression in a single cell. *Science*, 297(5584), pp.1183-1186.

[4] Hegland, M., Hellander, A. & Lötstedt, P., 2008. Sparse grids and hybrid methods for the chemicak master equation. *BIT Numerical Mathematics*, 48, pp. 265-283.

[5] Zeiser, S., Franz, U., Wittich, O. & Liebscher, V., 2007. Simulation of genetic networks modelled by piecewise deterministic Markov processes. *IET Systems Biology*, 2(3), pp. 113-135.

[6] Hellander, A. & Lötstedt, P., 2007. Hybrid method for the chemical master equation. *Journal of Computational Physics* 227, pp.100-122.

[7] Liu, Q. & Jia, Y., 2004. Fluctuations-induced switch in the gene transcriptional regulatory system. *Physical Review E*, 70 (041907).

[8] Gillespie, D., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22, pp. 403-434.

[9] Strang, G., 1968. On the construction and comparison of difference schemes. *SIAM Journal of Numerical Analysis*, 5(3), pp.506-517.

[10] Davis, M. H. A.,1984. Piecewise-Deterministic Markov Processes: A General Class of Non-diffusion Stochastic Models. *Journal of the Royal Statistical Society, Series B*, 46(3), pp. 353-388.

[11] Crudu, A., Debussche, A., & Radulescu, O., 2009. Hybrid stochastic simplifications for multiscale gene networks. *BMC Systems Biology*, 3(89).

[12] Vilar, J. M. G., Kueh, H. Y., Barkai, N. & Leibler, S., 2002. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Science USA*, 99(9), pp. 5988-5992.

[13] Lötstedt, P. & Ferm, L., 2006. Dimensional Reduction of the Fokker-Planck equation for stochastic chemical reactions. *Multiscale Modeling and Simulations*, 5(2), pp. 593-614.