Numerical Analysis of Methods for Simulating Clostridium Thermocellum

by

Eric M. Jalbert

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Master of Science
in
Applied Mathematics

Guelph, Ontario, Canada

Contents

List of Figures

List of Tables

Chapter 1

Model

1.1 History

The tradition biofilm model has been continually developed over many iterations since 1980. Rittmann and McCarty (1980) formulated the steady-state biofilm model, developed using the concept that biofilm growth would be the result of a steady flux from substrate. Since then the model have evolved to include modelling three-dimensional growth of multispecies anaerobic biofilms (Noguera et al. (1999)) and spatially heterogeneous biofilm structures (Eberl et al. (2001)).

The modelling of *Clostridium Thermocellum* is unique because this celluloytic anaerobic bacteria does not generate an extracellular polymeric substance. This uncharacteristic behaviour means that the mathematical model based on the work of Eberl and Demaret (2007) cannot be used as is. They modelled the biomass density and nutrient concentrations as a two-PDE-coupled system. Recently, Wang et al. (2011), used a cellular automata based model for simulating the growth of *Clostridium Thermocellum*. From this, better results were thought to derive from a continuous differential equation based model. Here the spatial diffusion of the substrait concentration is removed to mimic the carbon substrait that is consumed by *Clostridium Thermocellum*. This results in a PDE-ODE-coupled system. This is based on the work done by Dumitrache (2014), where this same coupling was used and formulated.

1.2 Model Description

The model used for simulations is based on the deterministic model developed in Eberl et al. (2001), which was designed to simulate the development of spatially heterogenous biofilm structures. They modelled the biomass density and nutrient concentration as a two-PDE-coupled system. Here the spatial diffusion of the nutrient concetration is removed to mimic the carbon substrait that *C.Thermocellum* consumes in growth. This makes a PDE-ODE-coupled system purposed as,

$$M = \nabla_x \left(d(M)\nabla_x M\right) + f(C)M \tag{1.1}$$

$$C = -g(C)M \tag{1.2}$$

where

$$d(M) = d\frac{M^\alpha}{(1-M)^\beta} \tag{1.3}$$

$$f(C) = u\frac{C}{k+C} - n \tag{1.4}$$

$$g(C) = y\frac{C}{k+C} \tag{1.5}$$

Here we have a pair of equations, (1.1) and (1.2), that represent the biomass density and substrait concentration respectivly. The spatial diffusion of the biofilm is modelled with density-dependent diffusion, represented by (1.3), and the growth rate of biomass is given by (1.4). The growth rate is simple Monod kinetic growth with a constant death rate. In (1.2) there is only a consumption term from the bacteria consuming the carbon substrait. This term is based on the growth of the biomass, differing only by a scalar multiplier.

The dimensions of the parameters and variables are in Tabel 1.1.

Variable/Parameter	Dimensions
t	$[days]$
x	$[meters]$
M	$[\frac{grams}{meters^3}]$
C	$[\frac{grams}{meters^3}]$
d	$[\frac{meters^2}{days}]$
α	$[-]$
β	$[-]$
u	$[days^{-1}]$
k	$[\frac{grams}{meters^3}]$
y	$[\frac{C}{M}]$
n	$[\frac{grams}{meters^3 \cdot days}]$

Table 1.1: List of parameters and their dimensions

1.3 Nondimensionalization

To help facilitate the analyses of this system, the full removal of all physical units is preferred. This process of nondimensionalization involves using known parameters to create substitutions with physical units cancelling. Here the parameters used are: the biomass growth rate, u; the length of the region, L; and the maximum density for biomass and substrait, M_∞ and C_∞. From using the following parameter changes, the system can be made unitless.

$$\chi = \frac{x}{L} \implies Ld\chi = dx \tag{1.6}$$

$$\tau = ut \implies \frac{1}{u}d\tau = dt \tag{1.7}$$

$$\mathcal{M} = \frac{M}{M_\infty} \tag{1.8}$$

$$\mathcal{C} = \frac{C}{C_\infty} \tag{1.9}$$

$$\delta = \frac{1}{uL^2}d \tag{1.10}$$

$$\kappa = \frac{k}{C_\infty} \tag{1.11}$$

$$\nu = \frac{n}{uC_\infty} \tag{1.12}$$

$$\gamma = \frac{M_\infty}{C_\infty}y \tag{1.13}$$

Using these, (1.1) and (1.2) can be simplified and nondimensionalized into,

$$\mathcal{M}_\tau = \nabla_\chi \left(D(\mathcal{M})\nabla_\chi \mathcal{M}\right) + F(\mathcal{C})\mathcal{M} \tag{1.14}$$

$$\mathcal{C}_\tau = -G(\mathcal{C})\mathcal{M}, \tag{1.15}$$

where,

$$\begin{aligned} D(\mathcal{M}) &= \delta \frac{\mathcal{M}^\alpha}{(1-\mathcal{M})^\beta} \\ F(\mathcal{C}) &= \frac{\mathcal{C}}{\kappa + \mathcal{C}} - \nu \\ G(\mathcal{C}) &= \gamma \frac{\mathcal{C}}{\kappa + \mathcal{C}}. \end{aligned} \tag{1.16}$$

with only $\delta, \kappa, \nu, \gamma$ as model parameters.

1.4 Parameters

Each of the dimensionless parameters in (1.16) have a biological representation based on the transformations done. The parameter δ is the dimensionaless constant for diffusion. It affects the change in biomass from adjacent biomass sources, a greater δ results in a greater change. The parameter κ is the half-saturation point, it is exactly the value for which substrait concentration results in 0.5-optimum growth rate. Parameter ν is the death rate of the biomass. Specifically, it is the ratio of biomass growth to death, representing the fraction of biomass density that perishs from natural causes or a lack of substrait. Lastly, γ is the yield ratio. It signifies the ratio of substrait consumed to biomass growth. Here, a larger γ value results in more substrait being consumed to produce the same amount of biomass.

With (1.14) being reduced to four parameters the numerical analysis become more simiplified while still retaining the same significance in results.

Chapter 2

Numerics

2.1 Discritization

In order to find the solution for (1.14) spatial and temporal discritizations must be made. First the equations are discritzied by time,

$$\frac{M^{k+1} - M^k}{\Delta t} = \nabla_x (D(M^{k+1}) \nabla_x M^{k+1}) + F(C^{k+1}) M^{k+1}, \tag{2.1}$$

$$\frac{C^{k+1} - C^k}{\Delta t} = \frac{h}{2} (G(C^{k+1}) M^{k+1} + G(C^k) M^k). \tag{2.2}$$

Here, (2.1) follows the ideas of the Backwards Euler Method; (2.2) follows Trapezoidal Rule. The index variable k has also been introduced in (2.1 - 2.2) as time step counting variable.

Now, only (2.1) requires spatial considerations since, according to the biology of our system, the substrate does not diffuse across the region. The spatial discritization will be through the Finite Difference Method as described in Saad (2003). Here a grid will be created over the region and the solution of (2.1) will be approximated at each grid point using a five-point stencil. This results in,

$$\frac{M_{i,j}^{k+1} - M_{i,j}^k}{\Delta t} = \frac{1}{\Delta x^2} \sum_{(s,r) \in \mathbb{A}} \left(D(M_{i+\frac{s}{2}, j+\frac{r}{2}}^k) \cdot (M_{i+s,j+r}^{k+1} - M_{i,j}^{k+1}) \right) + F(C_{i,j}^k) M_{i,j}^{k+1} \tag{2.3}$$

where $\mathbb{A} = \{(0, \pm 1), (\pm 1, 0)\}$. Now there are two additional indexing variables, i and j. These count out the respective grid point for the spatial discritization. On a $n \times m$ grid, the x_1 and x_2 dimensions would be represented by $i \in (0, n)$ and $j \in (0, m)$ respectivly. One item to note is that the index on D is halved because it is the arithmetic average of two adjacent values.

Now (2.2) and (2.3) can be solved as a fixed-point-iteration method. In a single time step, the solutions for M and C can be solved using the previous time step solution in the follow manner:

$$\frac{M_{i,j}^{p+1} - M_{i,j}^{k}}{\Delta t} = \frac{1}{\Delta x^2} \sum_{(s,r)\in\mathbb{A}} \left(D(M_{i+\frac{s}{2},j+\frac{r}{2}}^{p}) \cdot (M_{i+s,j+r}^{p+1} - M_{i,j}^{p+1}) \right) + F(C_{i,j}^{p}) M_{i,j}^{p+1} \tag{2.4}$$

$$\frac{C^{p+1} - C^{k}}{\Delta t} = \frac{-1}{2}(G(C^{p+1})M^{p+1} + G(C^{k})M^{k}) \tag{2.5}$$

where $p \in (0, P)$, and for $p = 0$ we have $M^0 = M^k, C^0 = C^k$, and for $p = P$ we have $M^P = M^{k+1}, and C^P = C^{k+1}$.

2.2 Solving Method

The basic idea to get a solution for (1.14) is to solve (2.4 - 2.5) as a fixed-point-iteration at each time step. In the current form, the equations can be rearrange and solved by conventional methods.

For (2.4), a linear system of equations can be created following Saad (2003). For each grid point (i, j) a linear system exists, defined as:

$$\begin{aligned} \frac{M_{i,j}^{k}}{\Delta t} = & \sum_{(i,j)\in\mathbb{A}} \left(\frac{D(M_{i+\frac{s}{2},j+\frac{r}{2}}^{p})}{\Delta x^2} \cdot M_{i+s,j+r}^{p+1} \right) \\ & + \left(\sum_{(i,j)\in\mathbb{A}} \left(\frac{D(M_{i+\frac{s}{2},j+\frac{r}{2}}^{p})}{\Delta x^2} \right) - F(C_{i,j}^{p}) + \frac{1}{\Delta t} \right) M_{i,j}^{p+1}. \end{aligned} \tag{2.6}$$

From (2.6), a five-diagonal matrix can be created defined as,

$$A = \begin{pmatrix} M_{i,j} & M_{i+1,j} & & M_{i,j+1} & & & & & \\ M_{i-1,j} & \ddots & \ddots & & \ddots & & & & \\ & \ddots & \ddots & \ddots & & \ddots & & & \\ M_{i,j-1} & & M_{i-1,j} & M_{i,j} & M_{i+1,j} & & M_{i,j+1} & & \\ & \ddots & & \ddots & \ddots & \ddots & & \ddots & \\ & & M_{i,j-1} & & M_{i-1,j} & M_{i,j} & M_{i+1,j} & & M_{i,j+1} \\ & & & \ddots & & \ddots & \ddots & \ddots & \\ & & & & \ddots & & \ddots & \ddots & M_{i+1,j} \\ & & & & & M_{i,j-1} & & M_{i-1,j} & M_{i,j} \end{pmatrix} \tag{2.7}$$

where each $M_{i,j}$ is the coefficient based on (2.6).

Solving (2.7) can be done by use of a linear solver. According to Barrett et al. (1987), if A is positive definite and symmetric then it is best solved using the Conjugate Gradient method.

Proposition 2.2.1. *The matrix A, defined in (2.7) is positive definite and symmetric.*

Proof. Matrix A is positive definite if all the eigenvalues are positive. Using the Circle theorem described by Geršgorin (1931), the eigenvalues can be shown to be positive if, independently on all rows, the sum of the off-diagonals values is less then the diagonal value. This can be verified. We have,

$$\sum_{(i,j)\in\mathbb{A}} \left(\frac{D(M^p_{i+\frac{s}{2},j+\frac{r}{2}})}{\Delta x^2} \right) < \left(\sum_{(i,j)\in\mathbb{A}} \left(\frac{D(M^p_{i+\frac{s}{2},j+\frac{r}{2}})}{\Delta x^2} \right) - F(C^p_{i,j}) + \frac{1}{\Delta t} \right) \tag{2.8}$$

which simplifies to,

$$F(C^p_{i,j}) < \frac{1}{\Delta t} \tag{2.9}$$

which is true given the values of Δt that are used. Therefore we have that A is positive definite.

The symmetry of A can be trivially shown if one considers the formation of the diagonals. On a single row, each element corresponds to the adjacent grid points of grid i, j. As the grid ordering

counts along, the elements that are equidistance from the diagonal are actually reference to the same grid point. Therefore we have symmetry. □

Given that A is positive definite and symmetric, the conjugate gradiant method can be used to compute the solution. As an added property, A also happens to be diagonally dominate. This means that it could be solved using Bi-Conjugate Gradient Method. However the Conjugate Gradient method has a faster computation time then Bi-Conjugate Gradiant method for this problem and is used for this reason Barrett et al. (1987).

For solving (2.5), the equation can be rearranged into a quadratic form, substituting in $G(C)$ from (1.16)

$$\left(C^{p+1}\right)^2 + \left(\kappa - C^k + \frac{h}{2}\gamma M^{p+1} + \frac{h}{2}\frac{\gamma C^k M^k}{\kappa + C^k}\right) C^{p+1} + \left(-\kappa C^k + \frac{h}{2}\frac{\gamma \kappa C^k M^k}{\kappa + C^k}\right) = 0. \tag{2.10}$$

Using the quadratic equation results in,

$$C^{p+1} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{2.11}$$

for which,

$$\begin{aligned} a &= 1 \\ b &= \kappa - C^k + \frac{h}{2}\gamma M^{p+1} + \frac{h}{2}\frac{\gamma C^k M^k}{\kappa + C^k} \\ c &= -\kappa C^k + \frac{h}{2}\frac{\gamma \kappa C^k M^k}{\kappa + C^k} \end{aligned} \tag{2.12}$$

To determine which branch of (2.11) to use, a physical situation is used. Specifically the case where there exist no biomass, $M = 0$. The expected outcome is that no substrate is consumed and thus the substrait concentration will remain constant as a function of τ. When the equations in (2.12) are evaluated at $M = 0$, the result it,

$$a = 1, \quad b = \kappa - C^k, \quad c = -\kappa C^k, \tag{2.13}$$

which can be used to evaluate (2.11) as,

$$\begin{aligned} C^{p+1} &= \frac{-(\kappa - C^k) \pm \sqrt{(\kappa - C^k)^2 - 4(-\kappa C^k)}}{2} \\ &= \frac{1}{2}\left(C^k - \kappa \pm \sqrt{\kappa^2 + 2\kappa C^k + (C^k)^2}\right) \\ &= \frac{1}{2}\left(C^k - \kappa \pm (\kappa + C^k)\right). \end{aligned} \tag{2.14}$$

Now, if the positive branch is used the above equation evaluates to $C^{p+1} = C^k$. This means that between any two distinct times, the substrait concentration will remain constants, which was expected. To further this confirmation, the negative branch results in $C^{p+1} = -\kappa$, a non-postive substrate concentration, which is not physically relavent.

$$C^{p+1} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{2.15}$$

where a, b, and c are defined in (2.12).

Now that computable solutions for M and C at a single time step have been found, an algorithm to solve for the next time step can be esstablished. Algorithm 1 shows the organizations of solving (2.5 - 2.4). Note that Algorithm 1 actually describes both a fully- and semi- implicite method for solving

Data: M^k and C^k is previous timestep solutions.
M^p and C^p are temporary solutions defined such that
$M^p \to M^{k+1}$ and $C^p \to C^{k+1}$ as $p \to P$.
ϵ_{sol} is a tolerance set for a desired accuracy.
begin
  while $C_{diff} + M_{diff} > \epsilon_{sol}$ **do**
    Solve for M^{p+1} using C^p and M^k;
    Solve for C^{p+1} using M^{p+1}, C^k, and M^k;
    Let $C_{diff} = (C^{p+1} - C^p)$;
    Let $M_{diff} = (M^{p+1} - M^p)$;
    Let $C^p = C_{p+1}$;
    Let $M^p = M_{p+1}$;
    Let $p = p + 1$;
  end
end

Algorithm 1: Algorithm for the fully-implicit solving of (1.14)

(1.14). If $P = 1$ then only a single iteration of the algorithm is applied. This would result in a change similar to how the Gauss-Seidal method changes the Jacobi method; the values used would no longer be updated in a single timestep when $P = 1$.

2.3 Implementation

The implementation of Algorithm 1 was done with Fortran. The matrix system was converted into a 1D array by use of a bijective mapping defined as:

$$\pi : \{0, \ldots, n\} \times \{0, \ldots, m\} \to \{1, \ldots, (n+1)(m+1)\}(i, j) \quad \to \pi(i, j) \tag{2.16}$$

The matrix was stored in diagonal format since A is a five-diagonal matrix.

All the computations were run on a custom built workstation with an Intel Xeon CPU E5-2650 (1.2 GHz, 20MB cache size) and 32 GB RAM under Red Hat Enterprise Linux Server release 6.5 (Santiago). Running the computations with OpenMP, took advantage of 6 out of the 16 processors of the Intel Xeon CPU, each with 2 threads. The GNU Fortran compiler, version 4.4.7, was used for all computations; the compiler arguments were

```
-O3 -fdefault-real-8 -fopenmp
```

References

R. Barrett, M. Berry, Chan T.F., J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. Society for Industrial and Applied Mathematics, 1st edition, 1987.

J.C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2nd edition, June 2008.

A Dumitrache. *Understanding Biofilms of Anaerobic, Thermophilic and Cellulolytic Bacteria: A Study towards the Advancement of Consolidated Bioprocessing Strategies*. PhD thesis, University of Toronto, 2014.

HJ Eberl and L Demaret. A finite difference scheme for a doubly degenerate diffusionreaction equation arising in microbial ecology. *Electron. J. Differential Equations*, page 7795, 2007.

HJ Eberl, DF Parker, and MCM van Loosdrecht. A new deterministic spatio-temporal continuum model for biofilm development. *Journal of Theoretical Medicine*, pages 161–175, 2001.

S. Geršgorin. über die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, pages 749–754, 1931.

DR Noguera, G Pizarro, DA Stahl, and BE Rittmann. Simulation of multispecies biofilm development in three dimensions. *Water Science and Technology*, 39:123–130, 1999.

BE Rittmann and PL McCarty. Model of steady-state biofilm kinetics. *Biotechnology and Bioengineering*, 22:2343–2357, 1980.

Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematic, 2nd edition, 2003.

S Sirca and M Horvat. *Computational Methods for Physicistsl: Compendium for Students*. Springer, 2012.

Z-W Wang, S-H Lee, JG Elkins, and JL Morrell-Falvey. Spatial and temporal dynamics of cellulose degradation and biofilm formation by caldicellulosiruptor obsidiansis and clostridium thermocellum. *AMB Express*, 1:30, 2011.

205