One-dimensional Consolidation Models

Mariana Dorobantu

NADA - KTH
Linstedtsvägen 3
S - 10044 Stockholm - Sweden
e-mail: mariana@nada.kth.se

November 14, 1997

Abstract

In a 1D case the state of flocculated suspension is described by the volume fraction and the pressure of the solids. Under the influence of gravity the flocs and liquid form an elasto-plastic structure at a certain stress. The sedimentation process is modeled by a parabolic conservation law with inter-phase velocity according to Darcy's law. In fact the equations are nearly hyperbolic and the solution develops near-shocks. These are important when e.g. simulating high centrifugal forces or filtering under pressure.

We present two approaches to the same model: A fixed grid two-phase problem and an inter-phase tracking scheme. Since the solution has different time scales an efficient numerical method must use implicit time stepping and allow for big variations in the time-step.

Using conservative discretization of space and the method of lines, a system of differential-algebraic equations (DAE) is built.

In the two-phase approach, a modified Gauss-Seidel iteration is implemented for solving a nonlinear two-phase problem at each time step.

For the inter-phase tracking method a modified BDF(3) solver is implemented. The initial singularity is removed by building an approximation to the solution at a time $t_0 > 0$.

Keywords: Consolidation simulation, 2-phase flow, flocculated suspensions, variational formulation.

Contents

Acknowledgments

I thank my advisors, *Prof. Björn Engquist* and *Prof. Jesper Oppelstrup*, for their guidance throughout this project. As project leader, Prof. Jesper Oppelstrup has introduced me to this fascinating problem, and provided invaluable advice and constant encouragement. *Jon Eiken* of Alfa Laval has constantly supported this project. *Dr. Jacob Yström* has on several occasions discussed topics related to this work. *Prof. Germund Dahlquist* and *Dr. Lennart Edsberg* kindly shared with me their excellent BDF solver. This work would not have been possible without the wonderful staff of NADA. Last, but not least, I thank my husband *Mihai Dorobantu* for his moral support, rewarding discussions and advice.

The research behind this paper was funded by the Parallel Scientific Computing Institute (PSCI) research program on CFD and NADA, to whom I am grateful.

1 Introduction and Overview

This paper deals with modeling and simulation of one-dimensional consolidation of flocculated suspensions. The 1-D case of sedimentation in a container with impervious bottom, is important in applications for parameter identification purposes. It is also a good model for sedimentation in simple containers.

The paper is organized as follows: We start with a brief presentation of the physical background and then derive two formulations (partial differential equations) that describe the same model. Each PDE is then analyzed. For each formulation we give a numerical scheme and present an analysis of its properties. Finally, we give numerical experiments and compare the performance, advantages, and drawbacks of the schemes.

2 Background

The sedimentation process of flocculated suspensions is an example of non-Newtonian fluid dynamics. The suspension consists of liquid and small particles forming larger aggregates, the flocs. The balance between the attractive forces and the electrostatic repulsion determines the spontaneous volume fraction of solids, known also as gel forming concentration. Under the influence of gravitational forces, the flocs and liquid form an elasto-plastic structure at a certain stress. When this stress is increased and exceeds a critical value, P_{crit}, the flocs form a solid but flimsy structure which is compressed while liquid is squeezed out (see Figure 1) and the concentration of solid particles is increased. This process is called consolidation.

We distinguish three zones:

1. The compression zone with consolidation and increasing solids concentration.

2. The overburden where the stress is smaller than the critical value, with constant concentration.

3. The clear liquid zone.

The three zones are of course coupled together. Let $\phi = \phi(x,t)$ denote the volume fraction of flocs as a function of the space variable x (height) and time t. At the beginning of the consolidation process ($t = 0$), we assume that $\phi = \phi_0$ is constant. As the consolidation takes place we expect a "cake" to develop at the bottom of the vessel in which the concentration is greater then ϕ_0. Let x_c be the top edge of the cake. The remaining flocs above x_c are still at concentration ϕ_0 forming a column up to a height h (the overburden). Above h we have clear liquid, see Figure 1. The process of consolidation continues until the solids stress balances the gravity and buoyancy. At $t = \infty$ there will still be a thin overburden above x_c, i.e. there will always be uncompressed flocs. The concentration profile at steady state is governed by the model parameters such as gravitational forces g, initial and maximum concentrations ϕ_0, ϕ_{max}, initial stress value σ_0, liquid and solid densities, etc.

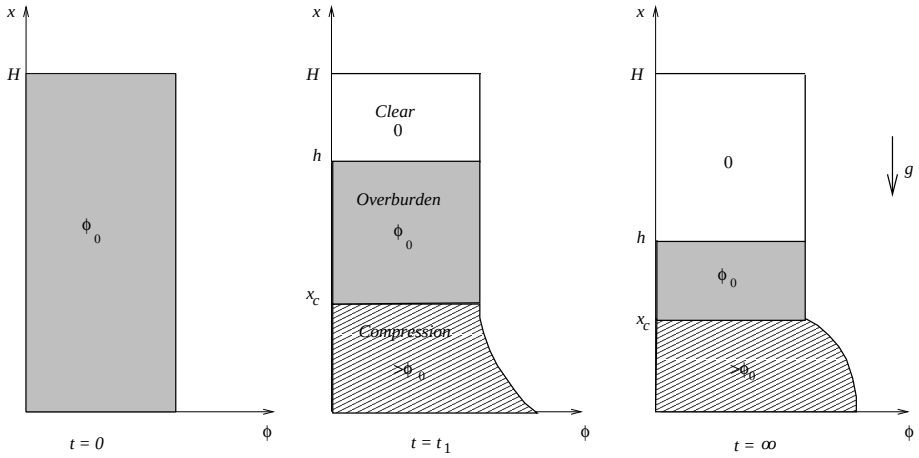

Figure 1: Distributions of volume fraction at three typical times: initially, during consolidation and at steady state.

We are concerned with the simulation of the consolidation process under natural gravity (1g) and centrifugal forces.

The mathematical model is a parabolic conservation law. We only consider forces in the x-direction. In the compression zone, the stress σ_x equals the

yield pressure $p_s(\phi)$. The velocity is assumed Darcian, i.e.,

$$v = -\frac{k}{\eta}\left(\frac{\partial p_l}{\partial x} + \rho_l g\right), \tag{1}$$

where p_l, ρ_l and η are the liquid pressure, density and viscosity. $k = k(\phi)$ is the permeability.

In the cake the total pressure is $p_s + p_l$ and the gravity force is $-g(\Delta\rho\phi + \rho_l)$. ρ_s is the solid density and $\Delta\rho = \rho_s - \rho_l$, g is the gravitational acceleration. Finally using force balance and the continuity equations of solids and liquid we find the solid phase velocity

$$u = D\left(\frac{\partial p_s}{\partial x} + g\Delta\rho\phi\right) \tag{2}$$

in the cake. The effective Darcy coefficient is $D = D(\phi) = \dfrac{k}{\eta}$.

The motion is governed by the continuity equation, known also as the Kynch equation

$$\frac{\partial \phi}{\partial t} + \frac{\partial (u(\phi)\phi)}{\partial x} = 0. \tag{3}$$

The important assumption due to Kynch [1], that the sedimentation velocity depends on ϕ only, is satisfied by our model, and thus the Kynch equation (3) becomes:

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x}\left(D\phi\left(\frac{\partial p_s}{\partial x} + g\Delta\rho\phi\right)\right) \tag{4}$$

subject to boundary conditions.

The equation is of the type

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x}\left(A \cdot \left(\frac{\partial \phi}{\partial x} + Pe \cdot \phi\right)\right),$$

where Pe, the Peclet number, is the ratio of gravity forces to solids stress forces. In the presence of large forces, such as during centrifugation, the Peclet number is large and thus the equations are near-hyperbolic and the concentration profiles can develop near-shocks. In this case the numerical solution requires shock-capturing [25] or shock-fitting schemes.

There are several practical situations which we would like to model:

- Sedimentation due to gravity is the simplest case.
- Centrifugal sedimentation is similar to gravity, but the force is variable. In this case the forces can be several orders of magnitude greater than in the gravitational case.
- Filter pressing requires modified boundary conditions.

It is possible to solve the problem with specific techniques of computational fluid dynamics [25, 26]. Shock–solutions are studied in e.g. [10].

In this paper we present two mathematical formulations accompanied by two numerical schemes,

- The compression zone scheme treats the the inter-phases as coupling equations between smooth solutions. We follow Auzerais *et al.* [1] in solving the parabolic equation in the "cake" coupled to an ODE for the overburden height h. Mass conservation or flux continuity yield an equation for the inter-phase x_c. This idea is also used by Eiken [15] for 1g sedimentation, and then extended in [11]. This formulation leads to a parabolic PDE with moving boundaries. The model is discussed in Chapter 5 and the numerical implementation is presented in Chapter 6.
- The mass-pressure approach resembles the enthalpy formulation of the two-phase heat equation. Such problems are described in [16], and we follow their ideas to prove existence of solutions and implement a modified Gauss–Seidel solver.

  There are two unknowns, the volume fraction ϕ and the pressure p, coupled by a differential equation

$$\frac{\partial \phi}{\partial t} = \Delta \rho g \frac{\partial}{\partial x}\left(D\phi^2\right) + \frac{\partial}{\partial x}\left(D\phi \frac{\partial p}{\partial x}\right) \tag{5}$$

  and a nonlinear relation between ϕ and p, see Figure 2. Theoretical considerations around the well-posedness of this approach are found in Chapter 3, while the numerical issues are presented in Chapter 4.

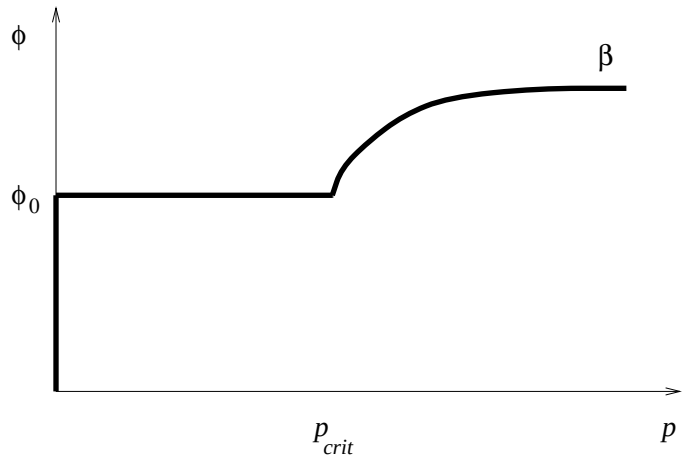

Figure 2: Relation between volume fraction ϕ and pressure p

3 The Mass–Pressure Formulation

The classical approach to sedimentation problems is to derive a two-phase problem similar to the Stefan problem (two phase heat equation) in unknown ϕ and keep track of moving inter-phases. In the enthalpy method the process is described by two unknowns, the temperature T and the enthalpy H. In the enthalpy form, the Stefan problem is:

$$\frac{\partial H}{\partial t} = \frac{\partial^2 T}{\partial x^2}. \tag{6}$$

The inter-phase does not appear explicitly, it is hidden in a non linear relation, $(H(x,t), T(x,t)) \in \beta$, exhibiting a jump of size ρL where L is the latent heat of the phase change, see Figure 3.

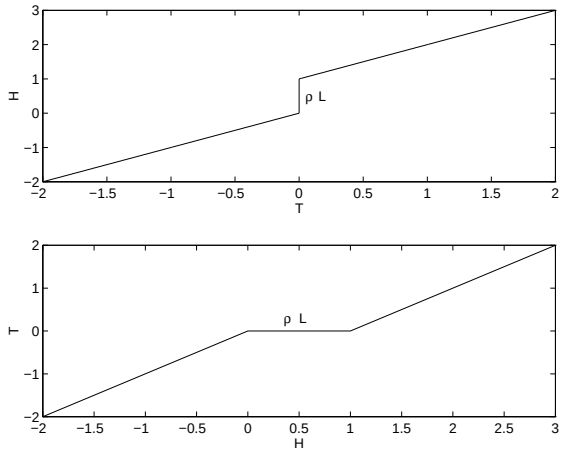

Figure 3: Relation β between enthalpy H and temperature T (top). β is not a function, the inverse graph β^{-1} (bottom) is a well-defined function.

For details on enthalpy formulation of the Stefan problem we refer to [16].

The basic observation is that the sedimentation problem has the same structure as equation (6) with the unknowns volume fraction ϕ and pressure p. Indeed, the continuity equation (4) can be written as:

$$\frac{\partial \phi}{\partial t} = \Delta\rho g \frac{\partial}{\partial x}\left(D\phi^2\right) + \frac{\partial}{\partial x}\left(D\phi\frac{\partial p}{\partial x}\right). \tag{7}$$

The relation between ϕ and p is modeled by a graph β, and the equation (7) is supplemented by the non-linear restriction $(\phi(x,t), p(x,t)) \in \beta \quad \forall x, t.\ p(x,t)$ must be continuous as function of x, and ϕ must satisfy $0 \leq \phi(x,t) \leq 1$, but can be discontinuous.

The essential features of the $\beta-$ curve are, as seen in Figure 2:

- The particles have a maximal packing density, hence the existence of $\phi_{\max}$ horizontal asymptote in the $p - \phi$ plane.
- Around ϕ_0 the particles come in contact and p jumps from 0 to a critical value p_{crit}. This is modeled by the horizontal line assuming that the particles are "hard", i.e. there is no elasticity. If we assume an elastic contact, p increases linearly around ϕ_0 and the horizontal line becomes slightly tilted with slope ε, as in Figure 4.

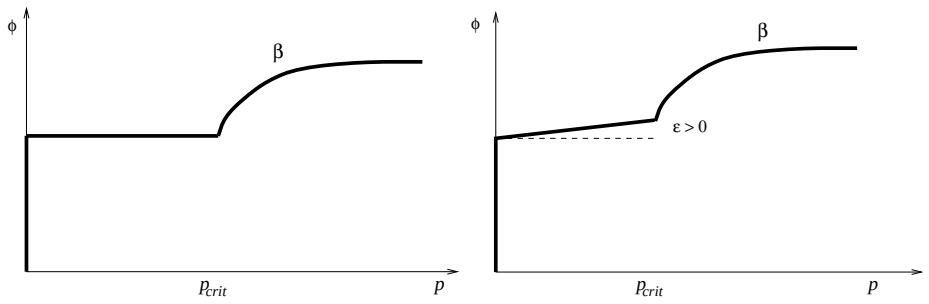

Figure 4: Two different graphs for β. The $\varepsilon = 0$ case (left) and the $\varepsilon > 0$ case (right)

- For $\phi < \phi_0$ there is no contact between particles and thus $p = 0$, hence the vertical line in the $p - \phi$ plane.

3.1 Qualitative behavior of the solution

The behavior of solutions to equation (7) is determined by $D(\phi)$. In our case $D(\phi)\phi$ behaves like a constant ($\neq 0$) in the vicinity of $\phi = 0$. The

property $\lim_{\phi \to 0} D(\phi)\phi = C \neq 0$ comes from the dilute suspension limit case $\phi \to 0_+$,because the fall speed of a single particle is

$$0 < u_0 = \Delta\rho g D(\phi)\phi_{|_{\phi=0}}.$$

Let us make the simplification $D\phi = \overline{k} = constant$ and the equation (7) becomes:

$$\frac{\partial \phi}{\partial t} = g\Delta\rho\overline{k}\frac{\partial \phi}{\partial x} + \overline{k}\frac{\partial^2 p}{\partial x^2}, \tag{8}$$

which is like a convection-diffusion equation. This assumption permits exact integration of the equation (8) at certain stages, which simplifies the analysis.

3.1.1 Initial profiles

Usually we know the initial value of the volume fraction $\phi(x,0) = \phi_0$ as seen in Figure 5. Since $(\phi, p) \in \beta$ and $\phi \equiv \phi_0$ we need $p \leq p_{crit}$, which is equivalent to $\max_x p(x,0) \leq p_{crit}$. Also all ϕ derivatives vanish and the PDE (8) is reduced to the equation

$$\overline{k}\frac{\partial^2 p}{\partial x^2} = 0$$

Two boundary conditions are derived:

$$\begin{cases} p(H,0) = 0, & \text{from } (\phi, p) \in \beta \\ \dfrac{\partial p}{\partial x}(0,0) = -\Delta\rho g\phi_0, & \text{from the zero flux boundary condition at the container bottom} \end{cases}$$

This means that $p(x,0) = \Delta\rho g\phi_0(H - x)$. In this situation the maximum condition on $p(x,0)$ implies

$$\Delta\rho g\phi_0 H \leq p_{crit}$$

This solution is in fact not interesting for sedimentation problems, as seen from the following theorem.

Theorem 3.1 *If* $\Delta\rho g\phi_0 H \leq p_{crit}$, *then*

$$\begin{cases} \phi(x,t) \equiv \phi_0 \\ p(x,t) = \Delta\rho g\phi_0(H-x) \end{cases}$$

is a stationary solution to the equation (8) with the nonlinear constraint $(\phi, p) \in \beta$.

In other words if $\Delta\rho g\phi_0 H \leq p_{crit}$, there is no sedimentation.

In the opposite situation we cannot construct an initial solution which is consistent with all the boundary conditions, and a transient process takes place.

Theorem 3.2 *If* $\Delta\rho g\phi_0 H > p_{crit}$ *and* $\phi(x,0) \equiv \phi_0$, *there is no function* $p(x)$ *such that*

$$\begin{cases} (\phi, p) \in \beta \\ \left(\Delta\rho g\phi_0 + \dfrac{\partial p}{\partial x}\right) = 0 \quad at \quad x = 0 \end{cases}$$

$p(.,0)$ is not necessary as an initial value for our equation, see (14), and Theorems 3.6 and 3.7. A good guess of $p(.,0)$ is needed to start the implicit scheme with a large time step, as seen in Section 4.2.

In this case, the zero-flux boundary condition is violated, at $t = 0$.

$$\begin{cases} \phi(x,0) = \phi_0 \\ p(x,0) = \dfrac{a}{H}(H-x) \quad \text{where } 0 \leq a \leq p_{crit} \end{cases}$$

Note also that $\max_x p(x,0) = a \leq p_{crit}$. Because of the nonlinear constraint, ϕ cannot grow until p becomes greater than p_{crit}. Choosing $a = p_{crit}$ permits the growth of $\phi(0,t)$ for any $t \geq 0$.

3.1.2 Evolution of the solution

In the beginning, for small t, the equation is like a transport equation with constant velocity $-g\Delta\rho\overline{k} = u_0$ and we have a contact discontinuity at the

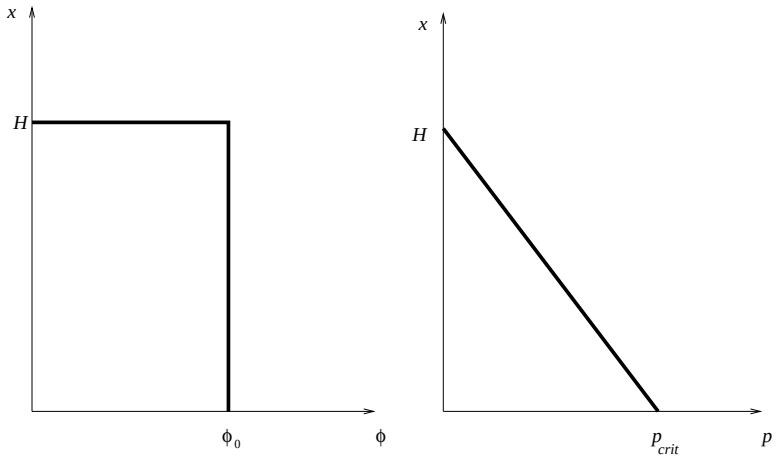

Figure 5: Initial profiles of ϕ and p at $t = 0$ that satisfy the PDE and the nonlinear constraint.

top of the graph of ϕ. In this case $\dfrac{\partial^2 p}{\partial x^2} = 0$.

The zero flux boundary condition $u(0,t) = 0$, produces a reflection and mass conservation yields $\phi \geq \phi_0$. Let

$$x_c(t) = \sup_x\{\phi(x,t) > \phi_0\} \quad \text{and} \quad h(t) = \inf_x\{\phi(x,t) = 0\}$$

where x_c and h are defined in the background section. On the interval $(x_c(t) \quad h(t))$ the pressure is linear because $\phi = \phi_0$ (thus $\frac{\partial \phi}{\partial t} = 0$ and the PDE yields $\frac{\partial^2 p}{\partial x^2} = 0$). This situation is illustrated in Figure 6. The continuity of p and the relation β between ϕ and p forces $p(x_c(t)) = p_{crit}$, and this determines the slope of p. Note also that $p = 0$ on $[h(t) \quad H]$ because $\phi = 0$ on that interval. On the interval $(0 \quad x_c(t))$, p has the same monotonicity as ϕ because β is increasing. The inter-phases $h(t)$ and $x_c(t)$ descend respective rise. The velocity $\dot{h}(t)$ is given by the advection part $-g\Delta\rho\overline{k}$ in equation (8) but also by the thickness of the overburden. The latter decreases in time such that the problem has a stationary solution for $t \to \infty$. Note in Figure 6

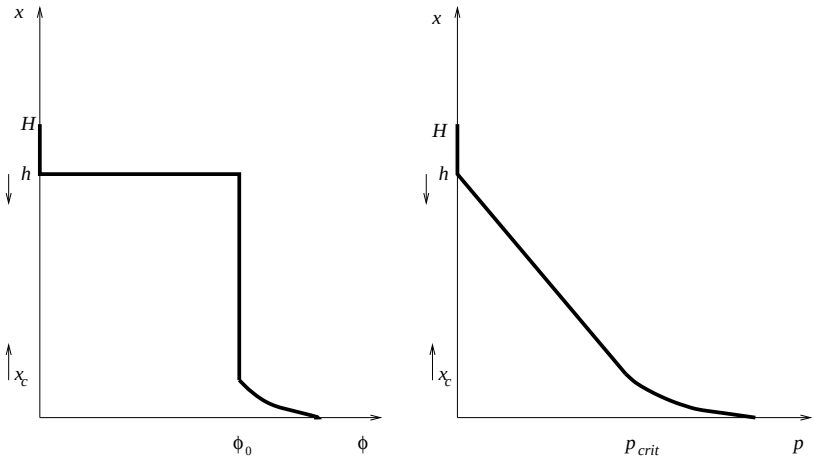

Figure 6: The profiles of ϕ and p for small $t = t_1$

that

$$\frac{\partial p}{\partial x} = \begin{cases} 0 & H > x > h \\ \\ -\dfrac{p_{crit}}{h - x_c} & x_c < x < h \end{cases} \tag{9}$$

has a discontinuity at $x = h(t)$ and thus $\frac{\partial^2 p}{\partial x^2}$ exists only in the sense of distributions (i.e. p is a weak solution, as defined later on). Using (9) for any smooth function v, $\quad v(H) = v(x_c) = 0$ we have:

$$\begin{aligned} \int_{x_c}^{H} \frac{\partial^2 p}{\partial x^2} v dx &= \int_{x_c}^{h} \frac{\partial^2 p}{\partial x^2} v dx + \int_{h}^{H} \frac{\partial^2 p}{\partial x^2} v dx \\ &= \left[\frac{\partial p}{\partial x} v\right]_{x_c}^{h} + \left[\frac{\partial p}{\partial x} v\right]_{h}^{H} = -\frac{p_{crit}}{h - x_c} v(h), \end{aligned} \tag{10}$$

which shows that

$$\frac{\partial^2 p}{\partial x^2} = -\frac{p_{crit}}{h - x_c} \delta(x - h).$$

Thus the equation (8) becomes:

$$\frac{\partial \phi}{\partial t} = g \Delta \rho \overline{k} \frac{\partial \phi}{\partial x} - \overline{k} \frac{p_{crit}}{h - x_c} \delta(x - h), \quad x_c < x < H \tag{11}$$

We interpret equation (11) as a linear advection equation with a source term at the moving discontinuity which slows the velocity $\dot{h}(t)$. Note that the slowing effect becomes more important as the overburden thickness $h - x_c$ decreases.

To compute $\dot{h}(t)$ we look at the mass balance around $h(t)$ over a time interval Δt. $\phi_0 \dot{h} \Delta t$ must balance the advected mass $g \Delta \rho \overline{k} \phi_0 \Delta t$ and the source $-\overline{k} \frac{p_{crit}}{h - x_c} \Delta t$ and thus

$$\dot{h} = \overline{k} \left(g \Delta \rho + \frac{p_{crit}}{h - x_c} \right)$$

The process ends when $\dot{h} = 0$, i.e. $h - x_c = \dfrac{p_{crit}}{\phi_0 g \Delta \rho}$, confirmed by the numerical experiments.

Our reasoning is based on the assumption that $x_c(t) < h(t)$ at all times t. This may contradict the intuition from the parabolic equation $\frac{\partial \phi}{\partial t} = \frac{\partial^2 \phi}{\partial x^2}$, where the response of a growth of $\phi(x = 0)$ is felt immediately throughout the domain, thus $\dot{x_c} = \infty$. In our case the nonlinear restriction β forbids such behavior. Otherwise, if $x_c(t) = h(t)$, then p cannot be continuous which is unphysical.

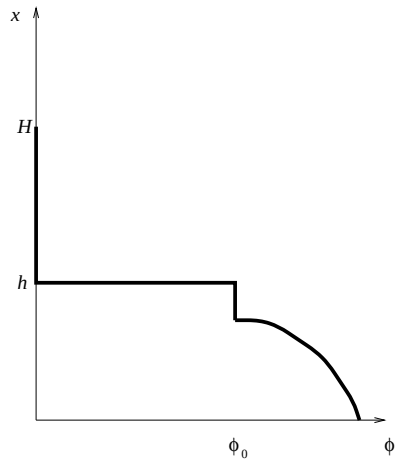

Figure 7: The profile of ϕ at $t = \infty$ (steady state)

3.1.3 Shock solutions

In the $D(\phi)\phi = constant$ case there are no shocks, just a contact discontinuity. Shocks are possible if $D(\phi)\phi$ is not constant (but $\lim_{\phi \to 0} D(\phi)\phi = C > 0$) as seen in the beginning of the chapter.

It is intuitively natural to assume that u decreases with growing ϕ because dense packing of the particles should lead to larger flow resistance. In the case $\frac{\partial u}{\partial \phi} < 0_{|\phi=0}$, the solution exhibit shocks as given by the following theorem:

Theorem 3.3 *Define* $f(\phi) = \Delta\rho g \phi D(\phi)$. *If* $\dfrac{\partial f}{\partial \phi}(0) < 0$ *then equation (8) develops shocks.*

Proof: Consider the two states $\phi = 0$ and $\phi = \varepsilon$ $(0 < \varepsilon < \phi_0)$. Because ϕ is either 0 or $\varepsilon < \phi_0$, the constraint $(\phi, p) \in \beta$ implies $p \equiv 0$ and then

$$u\phi = \Delta\rho g D(\phi)\phi^2 = f(\phi)\phi.$$

and the equation (8) becomes:

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x}(u\phi) = \frac{\partial}{\partial x}(f(\phi)\phi) = \left(\frac{\partial f}{\partial \phi}\phi + f\right)\frac{\partial \phi}{\partial x}.$$

The slope of the characteristics in the $x - t$ plane is $-\left(f + \phi\dfrac{\partial f}{\partial \phi}\right)$. A shock is formed if the slope for $\phi = 0$ is grater than that of $\phi = \varepsilon$, as seen in Figure 8, which is equivalent to

$$\begin{aligned} C &> f(\varepsilon) + \varepsilon \frac{\partial f}{\partial \phi}_{|\phi=\varepsilon} \\ &= f(0) + \varepsilon \frac{\partial f}{\partial \phi}_{|\phi=0} + \mathcal{O}(\varepsilon^2) + \varepsilon\left(\frac{\partial f}{\partial \phi}_{|\phi=0} + \mathcal{O}(\varepsilon)\right) \qquad (12) \\ &= C + 2\varepsilon \frac{\partial f}{\partial \phi}_{|\phi=0} + \mathcal{O}(\varepsilon^2) \end{aligned}$$

The inequality (12) holds for sufficiently small $\varepsilon > 0$, when $f(\phi)$ is a decreasing function, proving the theorem. □

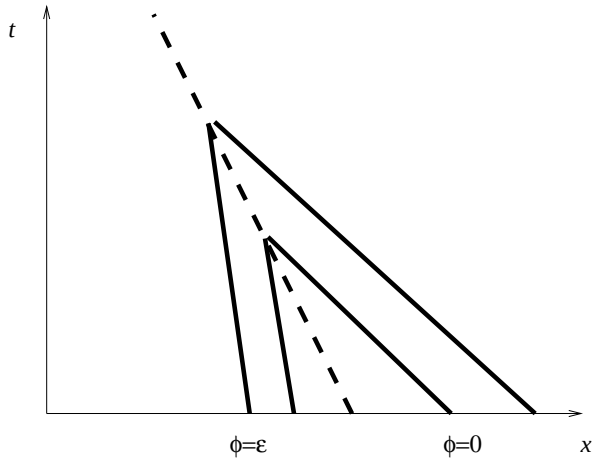

Figure 8: Shock forming at the inter-phase $\phi = 0$ and $\phi > 0$

3.2 Existence and uniqueness

The existence and uniqueness of the solution of the sedimentation problem will be studied on a simplified model problem. The simplified equation is $\frac{\partial \phi}{\partial t} = \frac{\partial^2 p}{\partial x^2}$, and $(\phi, p) \in \beta$ with β as in Figure 4. This model problem for $\varepsilon > 0$ is very similar to the Stefan problem and the analysis follows [16]. The case $\varepsilon = 0$ is much more difficult.

3.2.1 The simplified model problem

We start with the equation (8) and we want to eliminate the $\frac{\partial \phi}{\partial x}$-term. We change variables: $\tau = t$ and $y = g\Delta\rho\overline{k}t + x$ and obtain the equation:

$$\frac{\partial \phi}{\partial t} = \overline{k}\frac{\partial^2 p}{\partial y^2} \tag{13}$$

The variable change makes the "advection" front stationary since the new variables follow the transport. This equation is coupled by a nonlinear relation between ϕ and p, as seen in Figure 4.The shape of β is not affected by the variable change because β does not depend explicitly on x and t. Note the similarity to the Stefan problem, the only difference is the form of the

enthalpy relation β.

A weak solution of equation (13) is to be a pair of functions (ϕ, p) such that for all v

$$\begin{cases} \displaystyle\int_0^T \int_{-\infty}^{\infty} \left(\overline{k} p \frac{\partial^2 v}{\partial x^2} - \phi \frac{\partial v}{\partial t} \right) dx dt = - \int_{-\infty}^{\infty} \phi(x,0) v(x,0) dx \\ \\ (\phi(x,t), p(x,t)) \in \beta \end{cases} \tag{14}$$

where the test functions v have continuous first time and second space derivatives and vanish as $|x| \to \infty$ and at $t = T$. Note that the only initial data is $\phi(.,0)$ and not $p(.,0)$. In fact $p(x,0)$ should be defined as $p(x,0) = \lim_{t \to 0_+} p(x,t)$. Note also that this formulation defines a Cauchy problem posed on the whole real axis and $\phi \in L_2(\mathbb{R})$.

An alternative weak formulation of equation (13), which we will use in the following, is:

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty} \left(\frac{\partial \phi}{\partial t} v + \overline{k} \frac{\partial p}{\partial x} \frac{\partial v}{\partial x} \right) dx = 0 \\ \\ (\phi, p) \in \beta \quad \forall v \in H^1(\mathbb{R}) \end{cases} \tag{15}$$

Using the monotonicity of β we can prove the following L_2 estimates:

Theorem 3.4 *If ϕ and p are solutions of the weak formulation (15) and β is increasing then:*

$$||\phi(.,t)|| \leq ||\phi(.,0)|| \tag{16}$$

$$\left|\left| \frac{\partial p}{\partial x}(.,t) \right|\right| \leq \left|\left| \frac{\partial p}{\partial x}(.,0) \right|\right| \tag{17}$$

Proof: First note that the monotonicity of β implies $\frac{\partial \phi}{\partial x} \frac{\partial p}{\partial x} \geq \varepsilon (\frac{\partial p}{\partial x})^2$. This is so because whenever $(\phi(x_1), p(x_1)) \in \beta$ and $(\phi(x_2), p(x_2)) \in \beta$, then

$$(\phi(x_1) - \phi(x_2))(p(x_1) - p(x_2)) \geq \varepsilon (p(x_1) - p(x_2))^2.$$

This implies, by a limiting process, $\frac{\partial \phi}{\partial x}\frac{\partial p}{\partial x} \geq \varepsilon(\frac{\partial p}{\partial x})^2$ whenever the derivations exist. In a similar fashion we also have $\frac{\partial \phi}{\partial t}\frac{\partial p}{\partial t} \geq \varepsilon(\frac{\partial p}{\partial t})^2$. Thus we have

$$\frac{1}{2}\frac{d}{dt}\left|\left|\phi\right|\right|^2 = \left(\frac{\partial \phi}{\partial t}, \phi\right) = -\overline{k}\left(\frac{\partial p}{\partial x}, \frac{\partial \phi}{\partial x}\right) \leq -\overline{k}\varepsilon\left|\left|\frac{\partial p}{\partial x}\right|\right|^2 \leq 0,$$

using $v = \phi$ in (15). This yields the estimate (16).

Next we take $v = \frac{\partial p}{\partial t}$ in (15) and obtain:

$$\left(\frac{\partial \phi}{\partial t}, \frac{\partial p}{\partial t}\right) = -\overline{k}\left(\frac{\partial p}{\partial x}, \frac{\partial^2 p}{\partial t \partial x}\right) = -\frac{\overline{k}}{2}\frac{d}{dt}\left|\left|\frac{\partial p}{\partial x}\right|\right|^2.$$

Using again the consequence of monotonicity of β we obtain

$$\left(\frac{\partial \phi}{\partial t}, \frac{\partial p}{\partial t}\right) \geq \varepsilon\left|\left|\frac{\partial p}{\partial t}\right|\right|^2 \geq 0$$

and combining the last two inequalities yields the estimate (17). □

Unfortunately the estimates of Theorem 3.4 can not be used to prove well-posedness in the case $\varepsilon = 0$: We can build initial data for ϕ such that $\left|\left|\frac{\partial p}{\partial x}(.,0)\right|\right|$ is arbitrarily large.

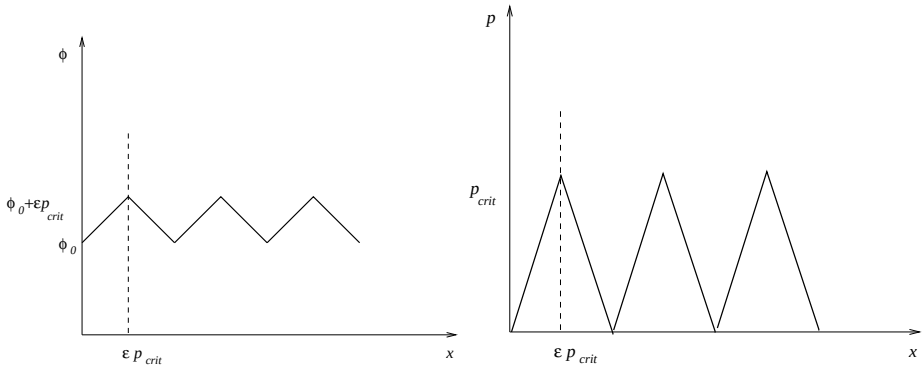

Figure 9: Initial data $\phi_\varepsilon(.,0)$ (left) is a small perturbation of ϕ_0. The corresponding $p_\varepsilon(.,0)$ (right) is strongly oscillatory.

For any $\varepsilon > 0$ choose β as in Figure 4 (right) and $\phi_\varepsilon(.,0)$ as in Figure 9 (left). The constraint $(\phi_\varepsilon, p_\varepsilon) \in \beta$ yields the initial pressure $p_\varepsilon(.,0)$ as in Figure 9 (right). Obviously, $\phi_\varepsilon(.,0) \to \phi_0$ but $\left\| \frac{\partial p_\varepsilon}{\partial x}(.,0) \right\| = 1/\varepsilon \to \infty$, when $\varepsilon \to 0$.

As illustrated by numerical experiments, even though $p_\varepsilon(.,0)$ has a wild behavior, there is a strong smoothing effect in both $\phi_\varepsilon(.,0)$ and $p_\varepsilon(.,0)$. Figure 10 displays the smoothing effect.

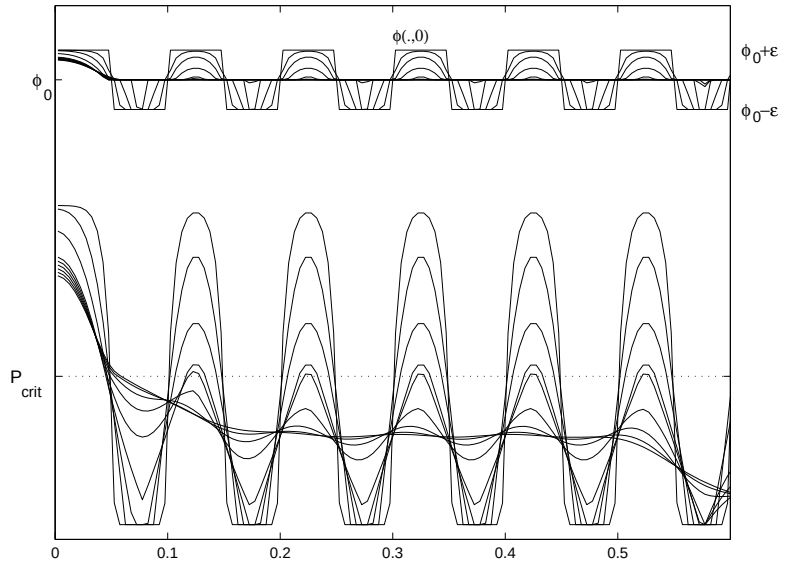

Figure 10: Solution $\phi(.,t)$, $p(.,t)$ at times $t = 0$ and $t > 0$. The mild variations of $\phi(.,0)$ (ε=20%) induce the strong variations of $p(.,0)$. All coefficients are scaled to 1 and ϕ and p are computed at times $t_n = n\Delta t$ where $\Delta t = 5 \cdot 10^{-6}$ using $\Delta x = 5 \cdot 10^{-3}$. Plots at n = 2, 5, 11, 17,18,19,20,21,22. Note that p becomes smooth by the time $t_{21} = 10^{-4}$.

3.2.2 Existence using Finite Elements for $\varepsilon \neq 0$

We follow the ideas of Elliott [16] to prove existence and uniqueness of the weak solution for (13) and $(\phi, p) \in \beta$. This technique works for $\varepsilon > 0$ only.

Let V^h denote the space of piece-wise linear and continuous functions on a mesh determined by the points $x_i = ih$ that lie in $H^1(\mathbb{R})$ such that $v(.,T) = 0$. Let $T = m\Delta t$ and n denote the time index. A finite element discretization using continuous piece-wise linear elements in space and discontinuous piece-wise constant elements in time for (15) is:
Find ϕ^{n+1}, p^{n+1} *such that* $p^{n+1} \in V^h$ *and*

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty} \left(\frac{\phi^{n+1} - \phi^n}{\Delta t} v + \overline{k} \frac{\partial p^{n+1}}{\partial x} \frac{\partial v}{\partial x} \right) dx = 0 \quad \forall v(., t_{n+1}) \in V^h \\ \\ (\phi^{n+1}(x_i), p^{n+1}(x_i)) \in \beta \quad \forall i \in \mathbb{Z}; \quad n = 0, 1, \ldots, m-1 \end{cases} \tag{18}$$

This formulation is equivalent to approximating $\phi(x,t) \approx \phi^n(x)$ for $t \in [t_n, t_{n+1}]$ and using backward Euler time discretization. Note that the non-linear constraint holds only in the nodes x_i.

Writing $\phi^n = \sum_i \phi_i^n \psi_i$ and $p^n = \sum_i p_i^n \psi_i$ the implicit scheme (18) is a system of nonlinear algebraic equations:

$$\begin{cases} \underline{\phi} + \Delta t C_h \underline{p} = \underline{b} \\ \\ \phi_i \in \beta(p_i) \end{cases} \tag{19}$$

where C_h is a symmetric positive definite matrix usually called the stiffness matrix. For the Cauchy problem C_h is an infinite matrix and $\underline{\phi}, \underline{p}, \underline{b} \in \ell^2$. ψ_i are the standard continuous piece-wise linear Finite Element basis on the grid x_i, such that $\psi_i(x_j) = \delta_{ij}$.

Elliott [16] proves existence of the solution in the following way:

1. The discrete system (19) has a unique solution. Let $p_{h,\Delta t}, \phi_{h,\Delta t}$ denote the discrete solution, e.g. $p_{h,\Delta t}(x,t) = p^n(x)$ if $t \in [t_n\ t_{n+1}]$.

2. A stability estimate of the type:

$$\varepsilon \int_0^T \left\| \frac{\partial p_{h,\Delta t}}{\partial t}(.,t) \right\|^2 dt + \left\| \frac{\partial p_{h,\Delta t}}{\partial x}(.,T) \right\|^2 \leq C$$

3. Bounds for $p_{h,\Delta t}$ and $\phi_{h,\Delta t}$ independent of h and Δt.

4. Existence of a limit (ϕ, p) when $h, \Delta t \to 0$.

5. The limit (ϕ, p) satisfies the differential equation (13) in variational form and the nonlinear constraint.

Step 1 Solving the system (19) is equivalent to minimizing the non-differentiable convex functional:

$$J(\underline{p}) = \sum_i \Phi(p_i) + \frac{\Delta t}{2} \underline{p}^T C_h \underline{p} - \underline{p}^T \underline{b}. \tag{20}$$

where $\Phi(p) = \int_0^p \phi(\tau) d\tau$ is a primitive of the volume fraction ϕ as a function of p. This functional is strictly convex because:

- Φ is strictly convex, its derivative is increasing as a result of the monotonicity of $(\phi, p) \in \beta$.
- $\underline{p}^T C_h \underline{p}$ is strictly convex because C_h is positive definite.
- $\underline{p}^T \underline{b}$ is convex being a linear function.

We also have that $\lim_{||\underline{p}|| \to \infty} J(\underline{p}) = \infty$ because:

- Φ is increasing, its derivative $\phi(p)$ is positive.
- The growth of $\underline{p}^T C_h \underline{p}$ is like $\left|\left|\underline{p}\right|\right|^2$ and dominates $\underline{b}^T \underline{p}$.

The following optimization theorem yields the existence and uniqueness result for (19).

Theorem 3.5 *(Theorem III.3 in [16]) If J is a strictly convex functional on a Hilbert space H such that $\lim_{||\underline{p}|| \to \infty} J(\underline{p}) = \infty$, then there exists a unique minimum and minimizer of J.*

Step 2 The stability estimate Theorem III.5 in [16] becomes:

$$\varepsilon \sum_{n=1}^{m} \Delta t \left|\left| \frac{1}{\Delta t}(\underline{p}^n - \underline{p}^{n-1}) \right|\right|^2 + \left|\left| \underline{p}^m \right|\right|_{C_h}^2 \leq C \left|\left| \underline{p}^0 \right|\right|_{C_h}^2 \tag{21}$$

where $||v||_{C_h} = (v^T C_h v)^{1/2}$. This estimate is obtained by scalar multiplying the equation $\underline{\phi}^{n+1} + \Delta t C_h \underline{p}^{n+1} = \underline{\phi}^n$ with $\underline{p}^{n+1} - \underline{p}^n$, using monotonicity of β and summing.

Step 3 The discrete system (19) is started with linear interpolation $\underline{\phi}^0$ of the initial data $\phi(.,0)$. Since the approximation $\underline{\phi}^0$ converges to $\phi(.,0)$ as $h \to 0$, we have that the approximate initial data is uniformly bounded in $L^2(\mathbb{R})$. Again using monotonicity of β, in the case $\varepsilon > 0$ this yields similar bounds also for $\underline{p}^0$.

The stability estimate (21) and the bounds on $\underline{\phi}^0, \underline{p}^0$ are used to derive uniform bounds for p and $\hat{p}$ (the linear interpolant in time for p) in $L^\infty(0,T;H^1(\mathbb{R}))$ and for ϕ in $L^\infty(0,T;L^2(\mathbb{R}))$. It can be shown that $\hat{p}$ is bounded in $H^1(\mathbb{R} \times [0\ T])$.

Step 4 Bounded sequences in Hilbert spaces contain convergent subsequences. Thus the Finite Element solution of (19) converges weakly (perhaps a subsequence only) to some functions ϕ, p. We find that $\phi \in L^2(\mathbb{R} \times [0\ T])$ and $p \in L^2(0,T;H^1(\mathbb{R}))$.

Step 5 We take a smooth test function $v = v(x,t)$ and project it onto V^h (linear interpolation in space). Obviously (18) holds for any h and Δt, and the interpolant of v converges strongly to v as $h, \Delta t \to 0$. Thus the weak form is also satisfied in the limit case. The nonlinear constraint is satisfied as a consequence of the maximal monotonicity of β, see Brezis [5].

The Steps 1–5 prove the following result:

Theorem 3.6 *(Theorem III.6 in [16]) If $\varepsilon > 0$, and $\phi(.,0) \in L^2(\mathbb{R})$, then there exists a weak solution to (13) and the standard Finite Element approximation converges: p converges strongly in $L^2(\mathbb{R} \times [0\ T])$ and weakly in $L^2(0,T;H^1(\mathbb{R}))$ and ϕ converges weakly in $L^2(\mathbb{R} \times [0\ T])$.*

For $\varepsilon = 0$ the above existence proof fails in two places: In Step 3 we cannot bound $\underline{p}^0$ from the bounds of the initial data $\phi(.,0)$, see also the remark after

Theorem 3.4. Secondly, the stability estimate (21) becomes:

$$\left\|\underline{p}^m\right\|_{C_h}^2 \leq C \left\|\underline{p}^0\right\|_{C_h}^2$$

and cannot be used to bound $p(.,T)$, as noted above.

3.2.3 The $\varepsilon = 0$ case

In the case $\varepsilon = 0$, similar problems were studied for the Stefan problem with zero "specific heat", see Crank [8] for a comprehensive list of examples.

An interesting existence result is due to Visintin [23]. In this case we need to modify β as in Figure 11, such that it is defined on the whole real axis. Let Φ be a primitive function of β as defined before, and Ψ the primitive function of β^{-1} (β^{-1} is the inverse relation).

Essentially the same kind of proof as for Theorem 3.6 is given. However, Visintin gives a different stability estimate based on the monotonicity of β^{-1} rather than β. In the weak formulation (15) we discretize time using backward Euler to obtain the continuous analog of (18–19):
Find ϕ^{n+1}, p^{n+1} *such that*

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty} \left(\frac{\phi^{n+1} - \phi^n}{\Delta t} v + \overline{k} \frac{\partial p^{n+1}}{\partial x} \frac{\partial v}{\partial x} \right) dx = 0 \quad \forall v \in H^1(\mathbb{R}) \\ \\ (\phi^{n+1}(x), p^{n+1}(x)) \in \beta \quad \forall x, \quad n = 0, 1, \ldots, m-1 \end{cases} \tag{22}$$

or in the operator notation:

$$\begin{cases} \phi + \Delta t A p = b \\ \\ \phi \in \beta(p) \end{cases} \tag{23}$$

The existence is proved in the following way:

Step 1 The semi-discrete equation (23) has a unique solution. This uses the continuous analog of (20) where:

$$J(p) = \int_{\mathbb{R}} \Phi(p(x)) + \frac{\Delta t}{2} \left| \frac{\partial p}{\partial x} \right|^2 - p(x) b(x) \; dx. \tag{24}$$

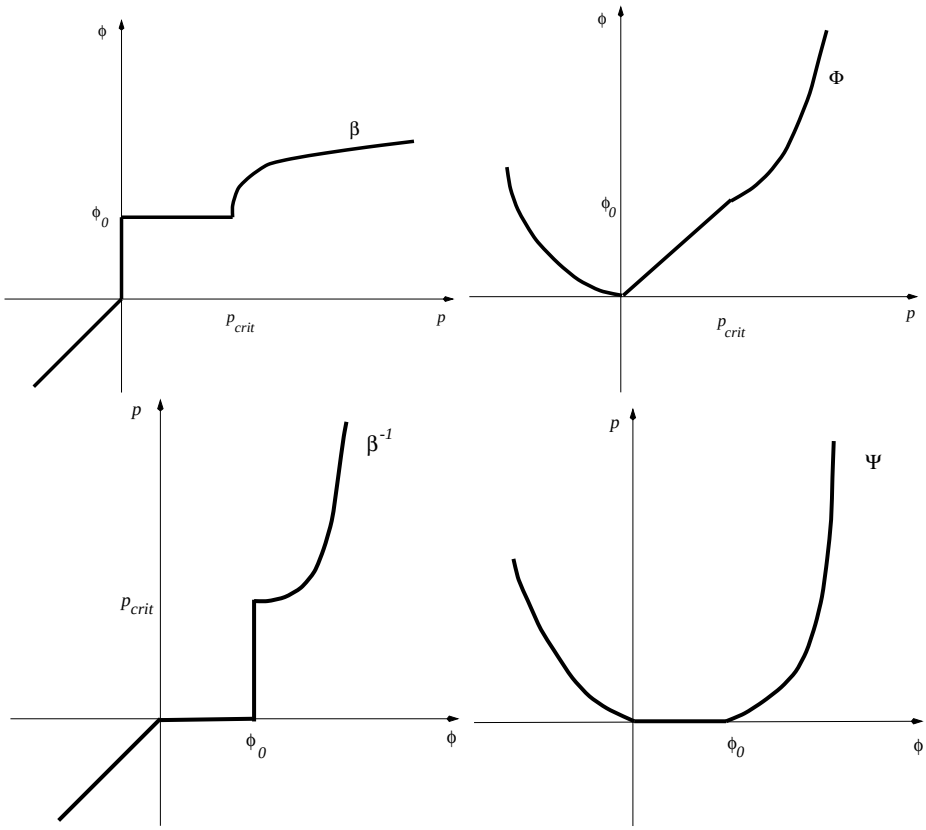

Figure 11: Modified graph of β (upper left) and the primitive function Φ (upper right) defined on the whole real axis. The inverse relation β^{-1} (lower left) and its primitive function Ψ (lower right) have a vertical asymptote at $\phi = \phi_{\max}$.

and the same convexity arguments to apply Theorem 3.5.

Step 2 A different stability estimate is derived using the convexity properties of Ψ (i.e. the monotonicity of β^{-1} rather than β). Summing over the time index we have:

$$\sum_{n=1}^{m} \int_{\mathbb{R}} \left(\Delta t \left(\frac{\phi^n - \phi^{n-1}}{\Delta t} \right) p^n \right) \geq \sum_{n=1}^{m} \int_{\mathbb{R}} \left(\Psi(\phi^n) - \Psi(\phi^{n-1}) \right) = \int_{\mathbb{R}} \left(\Psi(\phi^n) - \Psi(\phi^0) \right) \tag{25}$$

Step 3 To bound ϕ and p we need two assumptions:

$$\exists a, b \geq 0 \text{ such that: } |p| > a \text{ and } \phi \in \beta(p) \Rightarrow |\phi| \leq b\,|p| \tag{26}$$

as seen in Figure 12.

$$\inf \Psi \geq 0 \tag{27}$$

which is true in our case, as seen in Figure 11.

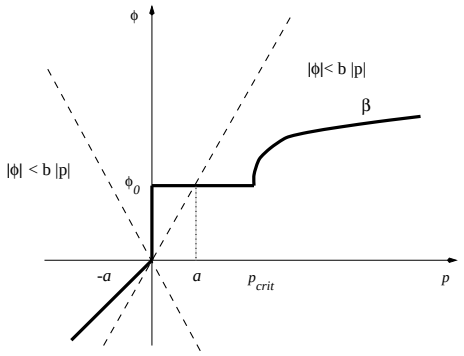

Figure 12: The graph β lies within the region determined by $|\phi| \leq b\,|p|$ away from the origin.

The first assumption yields that

$$\Psi(\phi) \geq \tilde{a}\,|\phi|^2 - \tilde{b}, \quad \tilde{a}, \tilde{b} > 0.$$

Together with the second assumption and the stability estimate they yield:

$$\frac{T}{m}\sum_{n=1}^{n}||p^n||^2_{H^1(\mathbb{R})} \leq C$$

$$\frac{T}{m}\sum_{n=1}^{n}||\phi^n||^2_{L^2(\mathbb{R})} \leq C$$

which give uniform bounds for $p_{\Delta t}$ and $\phi_{\Delta t}$ in $L^2(0,T;H^1(\mathbb{R}))$ and $L^2(\mathbb{R} \times [0\ T])$ respectively. $p_{\Delta t}$, $\phi_{\Delta t}$ are the semi-discrete solutions $p_{\Delta t} = p^n$ if $t \in [t_n\ t_{n+1}]$.

Step 4 As in Elliott's proof, as $\Delta t = \frac{T}{m} \to 0$, we find convergence of subsequences using the same compactness arguments.

Step 5 Using the convexity of Φ and interpolating a smooth test function v at the time $t_n = n\Delta t$, the weak formulation is satisfied in the limit, like in Elliott's proof.

The Steps 1–5 prove the following result:

Theorem 3.7 *(Theorem 2 in [23]) Assume that* $\phi(.,0) \in L^2(\mathbb{R})$*, the relation (26) holds,* $\Psi(\phi^0) \in L^1(\mathbb{R})$ *and* $\inf \Psi = 0$*. Then there exists a weak solution to (13) such that* $\phi \in L^2(\mathbb{R} \times [0\ T])$*.*

Remarks By using the properties of β^{-1} in the stability estimate, the bounds depend on $\phi(.,0)$ rather than $p(.,0)$. This allows for $\varepsilon = 0$ case because $\phi(.,0)$ can be controlled (which is not the case of $p(.,0)$).

The graph β is modified as in Figure 11 to permit the existence result in Step 2. This is not a restriction, because the solution (ϕ, p) is always non-negative. Gauss-Seidel iterations applied to the discretization presented in Chapter 4 produce non-negative discrete solutions for any choice of $\Delta t, \Delta x$, see Theorem 4.1, and it follows that the same properties holds for the exact solution.

4 Simulation in the Mass–Pressure Formulation

In this section, we construct a numerical method for equation (5), relying on the minimization of a convex functional. In the numerical experiment, we have the Darcy coefficient $D = \dfrac{k}{\eta}$ and for the permeability k we use an empirical relation [6] depending on volume fraction ϕ and particle radius a.

$$k = \frac{2a^2}{9\phi} \frac{(2-3\phi)^2}{3\phi + 4 + 3\sqrt{8\phi - 3\phi^2}} \tag{28}$$

For simplification, we make the notation $D\phi = \tilde{k}$ and the equation (4) becomes:

$$\frac{\partial \phi}{\partial t} = \Delta\rho g \frac{\partial}{\partial x}\left(\tilde{k}\phi\right) + \frac{\partial}{\partial x}\left(\tilde{k}\frac{\partial p}{\partial x}\right) \tag{29}$$

It is essentially a convection diffusion equation. The convection is given by the terms

$$\frac{\partial \phi}{\partial t} = \Delta\rho g \frac{\partial}{\partial x}\left(\tilde{k}\phi\right)$$

The diffusion term is

$$\frac{\partial}{\partial x}\left(\tilde{k}\frac{\partial p}{\partial x}\right)$$

and it is active only in the region where $\phi(x) > \phi_0$. When we choose the discretization we must account for:

- The equation (5) is a conservation law. A conservative discretization will guarantee convergence to the correct weak solution (the correct shock speed), which is important because in this approach we do not have an explicit equation for the front $x_c(t)$ or the front's speed.
- The p-term cannot be treated explicitly because there is no time derivative on p, and $p(t + \Delta t)$ cannot be computed from $\phi(t + \Delta t)$ because p is not a proper function of ϕ, in contrast to the Stefan problem case where the temperature T is a function of the enthalpy H.

- The scheme has to be implicit in p because $p(.,0)$ is not an initial value for this problem, see the formulation (14) and Theorems 3.6 and 3.7.
- The parabolic term will restrict the time step in explicit schemes. We need to use stable, implicit schemes, as discussed in [21].

4.1 The discretization

We have employed a Finite Volume Method (FVM) of Godunov-type [22, 19], with implicit treatment of $\frac{\partial^2 p}{\partial x^2}$-term. Thus we have the conservative form (correct shock speed) and upwinding (no oscillations at the shock). Integrate the equation (5) on interval $[x_{i-1/2}\ x_{i+1/2}] \times [t_n\ t_{n+1}]$ and find:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \int_{t_n}^{t_{n+1}} \frac{\partial \phi}{\partial t} dt dx = \Delta \rho g \int_{t_n}^{t_{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{\partial}{\partial x} \left(\tilde{k}(\phi)\phi + \tilde{k}(\phi)\frac{\partial p}{\partial x} \right) dx dt$$

The following notations are used: cell centers are $x_i = i\Delta x$ and $x_{i+1/2} = (i+\frac{1}{2})\Delta x$. Time instances are $t_n = n\Delta t$. Integrate the left-hand side in time and the right-hand side in space and then:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} (\phi(x,t_{n+1}) - \phi(x,t_n))dx = \Delta \rho g \int_{t_n}^{t_{n+1}} ((\tilde{k}(\phi)\phi)_{|_{x_{i+1/2}}} - (\tilde{k}(\phi)\phi)_{|_{x_{i-1/2}}})dt$$

$$+ \int_{t_n}^{t_{n+1}} ((\tilde{k}(\phi)\frac{\partial p}{\partial x})_{|_{x_{i+1/2}}} - (\tilde{k}(\phi)\frac{\partial p}{\partial x})_{|_{x_{i-1/2}}})dt \tag{30}$$

We define

$$\overline{\phi}_i^n = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \phi(x,t_n)dx, \quad i = 1, \ldots, N$$

and we represent at time $t = t_n$, ϕ in each cell $[x_{i-1/2}\ x_{i+1/2}]$ by its average $\overline{\phi}_i^n$.

The four flux terms in equation (30) are computed using local Riemann solvers. We approximate the terms $\frac{\partial p}{\partial x}$ with finite differences from the neighboring cells at time $t = t_{n+1}$. This choice makes the parabolic term implicit.

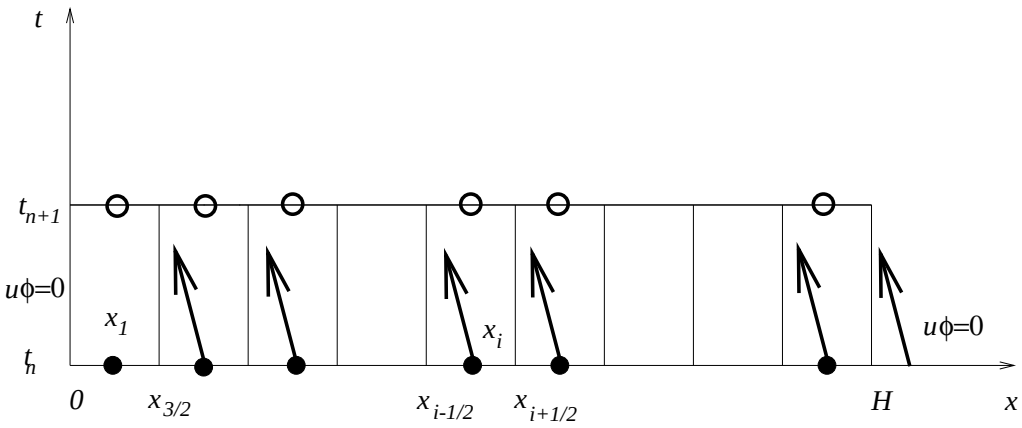

Figure 13: The discretization scheme: The fluxes on the vertical segments are computed using the characteristics and the cell-averages $\overline{\phi}$ at time $t = t_n$. Pointwise values for p at time $t = t_{n+1}$ and cell centers. Boundary conditions are incorporated by specifying the fluxes at $x = 0$ and $x = H$.

Since $-\tilde{k}(\phi) < 0$, we pick values from the right side of each vertical edge in Figure 13, to compute the fluxes:

$$
\begin{aligned}
(\tilde{k}(\phi)\phi)_{|_{x_{i+1/2}}} &= \tilde{k}(\overline{\phi}^n_{i+1})\overline{\phi}^n_{i+1}, \\
(\tilde{k}(\phi)\phi)_{|_{x_{i-1/2}}} &= \tilde{k}(\overline{\phi}^n_{i})\overline{\phi}^n_{i}, \\
(\tilde{k}(\phi)\tfrac{\partial p}{\partial x})_{|_{x_{i+1/2}}} &= \tilde{k}(\overline{\phi}^n_{i+1})\tfrac{p^{n+1}_{i+1}-p^{n+1}_{i}}{\Delta x}, \\
(\tilde{k}(\phi)\tfrac{\partial p}{\partial x})_{|_{x_{i-1/2}}} &= \tilde{k}(\overline{\phi}^n_{i})\tfrac{p^{n+1}_{i}-p^{n+1}_{i-1}}{\Delta x}.
\end{aligned}
\tag{31}
$$

In all cells, except the first and last, we have

$$\overline{\phi}_i^{n+1} = \overline{\phi}_i^n + \Delta\rho g \Delta t D_+(\tilde{k}(\overline{\phi}_i^n)\overline{\phi}_i^n) + \Delta t D_+(\tilde{k}(\overline{\phi}_i^n) D_- p_i^{n+1}) \qquad (32)$$
$$i = 2, ..., N-1$$

Using now the zero flux boundary condition for $x = 0$, we obtain the first equation:

$$\overline{\phi}_1^{n+1} = \overline{\phi}_1^n + \frac{\Delta t}{\Delta x}\tilde{k}(\overline{\phi}_2^n)\left(\Delta\rho g \overline{\phi}_2^n + \frac{p_2^{n+1} - p_1^{n+1}}{\Delta x}\right).$$

At the top we consider a ghost cell with $\phi \equiv 0$. This gives $\phi_{N+1}^{n+1} = 0$. For p in the ghost cell we can choose $p \equiv 0$ (and get Dirichlet boundary conditions: $p_{N+1}^{n+1} = 0$) or $\frac{\partial p}{\partial x} = 0$, (and get the Neumann boundary conditions: $p_{N+1}^{n+1} = p_N^{n+1}$). Since $\phi \equiv 0$ around $x = H$ after a small time, we get $p \equiv 0$ in that region and the two boundary conditions are equivalent. We choose to implement the Neumann boundary conditions because of the simplicity of the difference scheme.

With these assumptions we find the last equation:

$$\overline{\phi}_N^{n+1} = \overline{\phi}_N^n - \frac{\Delta t}{\Delta x}\tilde{k}(\overline{\phi}_N^n)\left(\Delta\rho g \overline{\phi}_N^n + \frac{p_N^{n+1} - p_{N-1}^{n+1}}{\Delta x}\right).$$

From now on ϕ_i^n denote $\overline{\phi}_i^n$. The matrix A is given by $-\Delta t D_+(\tilde{k}(\phi^n) D_- p^{n+1})$ and is symmetric positive definite.

We write all the $2N$ equation as a system with $2N$ unknowns:

$$\underline{\phi} + A\underline{p} = \underline{f}, \quad (\phi_i, p_i) \in \beta, \quad i = 1, ..., N \qquad (33)$$

where the right-hand side $\underline{f}$ and the system matrix A are evaluated using the old values at $t = t_n$, while the vectors $\underline{\phi}$ and $\underline{p}$ represent the discrete solution

at $t = t_{n+1}$. The tridiagonal matrix A has the following entries:

$$a_{i,i-1} = -\frac{\Delta t}{\Delta x^2}\tilde{k}(\phi_i^n) \qquad i = 2, ..., N$$

$$a_{i,i} = \frac{\Delta t}{\Delta x^2}\begin{cases} \tilde{k}(\phi_2^n) & i = 1 \\ \tilde{k}(\phi_i^n) + \tilde{k}(\phi_{i+1}^n) & i = 2, \ldots, N-1 \\ \tilde{k}(\phi_N^n) & i = N \end{cases} \tag{34}$$

$$a_{i,i+1} = -\frac{\Delta t}{\Delta x^2}\tilde{k}(\phi_{i+1}^n) \qquad i = 1, ..., N-1$$

and the right-hand side is

$$f_i = \phi_i^n + \Delta\rho g\frac{\Delta t}{\Delta x}\begin{cases} \tilde{k}(\phi_2^n)\phi_2^n & i = 1 \\ \tilde{k}(\phi_{i+1}^n)\phi_{i+1}^n - \tilde{k}(\phi_i^n)\phi_i^n & i = 2, \ldots, N-1 \\ -\tilde{k}(\phi_N^n)\phi_N^n & i = N \end{cases} \tag{35}$$

4.2 The non-linear Gauss–Seidel solver

Newton type algorithms require differentiability. The non-linear system (33) is non-differentiable because of the form of β. Thus other solution algorithms must used.

The correspondence with the minimization problem (20) makes a relaxation scheme natural. Gauss–Seidel and SOR iteration have been used for the Stefan problem [16]. We will try to determine $(\underline{\phi}, \underline{p})$ component-wise from equation (33) using an iteration

$$\phi^{(j)}, p^{(j)} \to \phi^{(j+1)}, p^{(j+1)} \quad j = 0, 1, 2, ...$$

In each iteration j we require that $(\phi_i^{(j)}, p_i^{(j)}) \in \beta$ is satisfied exactly on all components. The iteration is stopped when the residual $\underline{\phi}^{(j)} + A\underline{p}^{(j)} - \underline{f}$ is small enough, and thus the approximate solution $(\underline{\phi}, \underline{p})$ satisfies the requirement $(\phi_i, p_i) \in \beta$ exactly.

These ideas are used by Elliott [16] for solving the heat equation in enthalpy formulation and Glowinski [18] in applications to variational inequality problems of solid mechanics, e.g. plasticity problems.

Since p_i is also coupled to p_{i-1} and p_{i+1} by the matrix A, we use the Gauss–Seidel idea, i.e. use the latest updates for the values of p_{i-1} and p_{i+1}. Thus, for each component i we solve:

$$\phi_i^{(j+1)} + a_{i,i} p_i^{(j+1)} = f_i - \begin{cases} a_{1,2} p_2^{(j)} & i = 1 \\ a_{i,i-1} p_{i-1}^{(j+1)} + a_{i,i+1} p_{i+1}^{(j)} & i = 2, \ldots, N-1 \\ a_{N,N-1} p_{N-1}^{(j+1)} & i = N \end{cases} \tag{36}$$

$$\left(\phi_i^{(j+1)}, p_i^{(j+1)}\right) \in \beta \qquad \text{for} \quad i = 1, ..., N$$

The superscript j is the non-linear iteration index. For each component i, equation (36) is a non-linear system with two scalar equations:

$$\begin{cases} y + a_{i,i} x = b, \\ (y, x) \in \beta \end{cases} \tag{37}$$

The non-linear constraint yields:

$$\begin{cases} x = 0 \quad \text{if} \quad 0 \le b \le \phi_0 \quad (I), \\ y = \phi_0 \quad \text{if} \quad \phi_0 < b < \phi_0 + a_{i,i} p_{crit} \quad (II), \\ y = \beta(x) \quad \text{if} \quad b \ge \phi_0 + a_{i,i} p_{crit} \quad (III). \end{cases} \tag{38}$$

which can be solved analytically for cases (I) and (II) and by Newton iteration for case (III).

The monotonicity of β and $a_{i,i} > 0$ imply that the system (37) has a (unique) solution, as seen in Figure 14.

Theorem 4.1 *If the CFL-condition $\Delta t \le C \Delta x$, (when $C = 1/\min(\Delta \rho g \tilde{k}(\phi))$, is satisfied and the initial values $\phi^0, p^0 \ge 0$, then the solutions ϕ^n, p^n of the discrete equation (33) are non-negative $\forall n \ge 0$.*

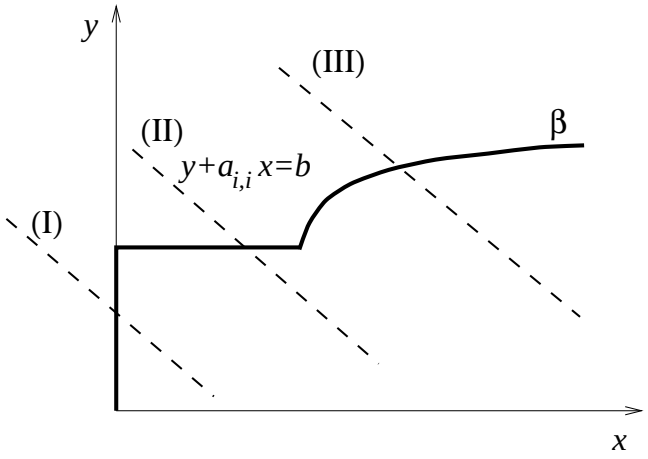

Figure 14: Existence of a solution to each scalar equation in the Gauss–Seidel sweep.

Proof: Let $\phi^{(j+1)}, p^{(j+1)}$ be the new Gauss–Seidel approximation of ϕ^{n+1}, p^{n+1}. Using (35) in (36), we have that the right hand side of (37) is

$$b = \phi_i^n \left(1 - \Delta\rho g \frac{\Delta t}{\Delta x}\tilde{k}(\phi_i^n)\right) + \Delta\rho g \frac{\Delta t}{\Delta x}\tilde{k}(\phi_{i+1}^n)\phi_{i+1}^n - a_{i,i-1}p_{i-1}^{(j+1)} - a_{i,i+1}p_{i+1}^{(j)}.$$

Assume that $\phi^n, p^n \geq 0$ (solution at previous time t_n), $\phi^{(j)}, p^{(j)} \geq 0$ (previous G-S approximation j) and $\phi_k^{(j+1)}, p_k^{(j+1)} \geq 0$ for $k = 1, 2, \ldots, i-1$ (first $i-1$ components of current approximation).

Using $a_{i,i\pm1} < 0$ and the above hypothesis, we find $b \geq 0$ if we take $\Delta t \leq \Delta x/(\Delta\rho g\tilde{k}(\phi_i^n))$. This is satisfied if we take $C = 1/(\Delta\rho g \min_{0\leq\phi\leq\phi_{max}} \tilde{k}(\phi))$.

Since $b \geq 0$, the equation (37) has non-negative solutions $y = \phi_i^{(j+1)}$, $x = p_i^{(j+1)}$ given by (38).

We have $\phi^0, p^0 \geq 0$ and then induction yields that $\phi^{(j+1)}, p^{(j+1)} \geq 0$ for all j. Thus $\phi^{n+1}, p^{n+1} \geq 0$. □

The iteration (36) must be started with an initial guess. For this we pick $\phi^{(0)} = \phi^n$ and $p^{(0)} = p^n$, where n refers to the solution at time $n\Delta t$. This is a good guess provided Δt is small enough. The first iteration is special

because it uses the initial guess $\phi(.,0)$ and $p(.,0)$. Note that the only initial data is $\phi(.,0)$, see Section 3.2. In general $p(.,0)$ is unknown, but in certain cases we can guess $p(.,0)$ such that $(\phi,p) \in \beta$.

4.3 Numerical experiments

A simulation program for the mass–pressure formulation is implemented in Matlab. The relation β is modeled by the empirical form

$$p = \sigma_0 \frac{\phi^3}{\phi_{\max} - \phi}, \quad \phi > \phi_0, \quad p > p_{crit}.$$

The following model parameters from [3, 4] have been used: The yield stress constant $\sigma_0 = 6000 N/m^2$, maximum packing volume fraction $\phi_{\max} = 0.595$, initial volume fraction of solids $\phi_0 = 0.15$, the height of the container $H = 0.2m$, density of the pore liquid $\rho_l = 877 kg/m^3$, density of the colloidal particles $\rho_s = 3960 kg/m^3$, viscosity of the pore liquid $\eta = 2.6 \cdot 10^{-3} kg/ms$ and $p_{crit} = 45$. These data are relevant for experiments made by L. Bergström [3, 4] which we have used for validation of the mathematical model against a real case with well-defined properties. Figure 15 is a plot of the volume fraction $\phi(.,t)$ at different times up to the steady-state.

All the experiments are done in the $\varepsilon = 0$ case. A detail of the solution is given in Figure 16. Note that there is only one grid point in the shock at h. Note also the correct representation of the overburden, where $\phi = \phi_0$ and p is linear.

Simulations where run with $N = 10, 20, 40, 80$ grid points up to $t = 126$ days. The Gauss–Seidel iterations converged rapidly except for the first two time steps, when convergence was slow. The following table gives the number of time steps and the Gauss–Seidel iteration count and their evolution is seen in Figure 17.

N	Time steps	G-S iterations at			
		t_1	t_2	t_3	all times
10	37	16	8	5	205
20	74	38	18	7	590
40	284	66	47	10	1577
80	778	98	74	17	3598

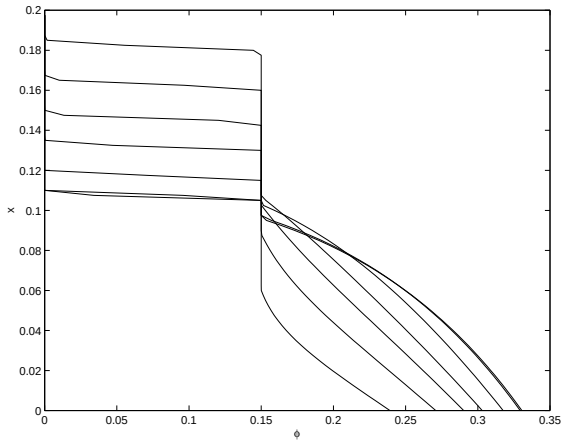

Figure 15: Numerical solutions using $N = 80$ space points at times $t = 6, 13, 20, 27, 39, 75, 126$ days.

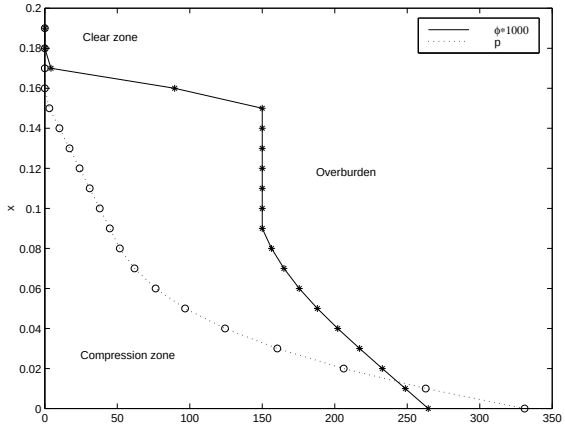

Figure 16: Solution profiles at $t = 13$ days using $N = 20$ grid points.

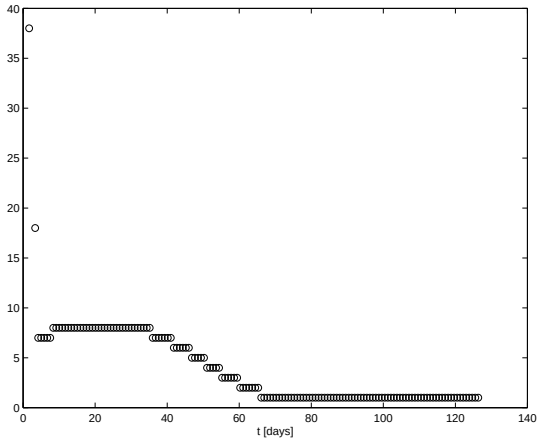

Figure 17: Evolution of the number of G-S iterations using $N = 20$ grid points and time $0 < t < 126$.

The time step is controlled by the following criteria:

- The initial time step is chosen to satisfy a CFL-type condition $\Delta t = \Delta x/(2\Delta\rho g k(\phi_0)\phi_0)$, motivated by the initial speed of the front $\dot{h}(0) = -u_0$.
- If the Newton solver exceeds a certain number of iterations in any component, the time step is reduced by a factor 2.
- If the Gauss–Seidel iteration converges very fast, the time step is increased by a factor 1.2.

We tried to improve the convergence of Gauss–Seidel iterations using an overrelaxation technique (the component i is relaxed with a factor ω if ϕ_i^n and ϕ_i^{n+1} are in the same region of β), as described in [16, 18] for the heat equation. Numerical experiments did not indicate improvement because:

- No ω-value improves the initial steps (the problem is essentially hyperbolic for small t).
- Very few iterations $O(5)$ at larger times.

5 The Compression Zone Formulation

The basic idea of this approach is to scale the compression zone and find equations for the inter-phase $x_c(t)$ and the total height $h(t)$. This approach is used by Auzerais, *et al.* [1], and in a slightly different formulation, by Eiken [15] and then in our earlier papers [13, 11].

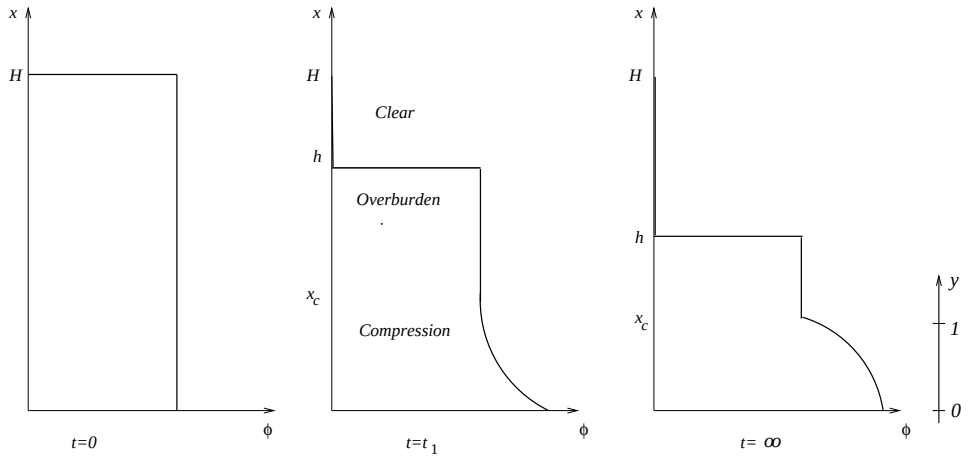

Figure 18: Distributions of volume fraction at three typical times: initially, during consolidation and the steady state. The compression zone is always scaled to [0 1].

In the compression zone, let $M = \dfrac{dp_s}{d\phi}$ and then the flux (2) becomes:

$$u = D\left(M\frac{\partial\phi}{\partial x} + g\Delta\rho\phi\right). \tag{39}$$

Using the notations of Figure 18, the conditions that govern the flow are:

- Zero flux on the bottom: $u(0,t) = 0, \quad t \geq 0.$
- Constant concentration in the overburden:

$$\phi(x,t) = \phi_0, \quad t \geq 0, \quad x_c \leq x \leq h.$$

- Continuity of ϕ and u at the compression zone height x_c.
- Conservation of mass: $\int_0^h \phi(x,t)dx = \text{ constant } = \phi_0 H$.

5.1 Non-dimensional variables

It is advantageous to use non-dimensional variables since it helps finding the "small" and "large" unknowns of the problem. The scalings used by us are:

$$\begin{cases} u^* = \dfrac{u}{u_0} & \text{velocity} \\ D^* = \dfrac{D}{D(\phi_0)} & \text{Darcy coefficient} \\ p_s^* = \dfrac{p_s}{p_{max}} & \text{pressure} \\ x^* = \dfrac{x}{H} & \text{space} \\ h^* = \dfrac{h}{H} & \text{height} \\ t^* = \dfrac{tu_0}{H} & \text{time} \end{cases}$$

u_0 is the initial velocity of a free falling "plug" of concentration ϕ_0. Note that

$$\frac{dh}{dt} = -u_0 \quad \text{at} \quad t = 0$$

p_{max} is the maximum pressure and $D(\phi_0)$ is the effective Darcy coefficient in the overburden. $D(\phi_0)$ is related to u_0 by:

$$u_0 = D(\phi_0)\Delta\rho g\phi_0.$$

There exists a formula for the permeability k due to Kozeny-Carman which uses the unknown radii of the particles. A mean radius is sometimes used. We prefer to determine $D(\phi_0)$ from experimental data. Note that the non-dimensional permeability $k^* = k/k(\phi_0)$ coincides with the non-dimensional effective Darcy coefficient D^*

By scaling $\phi^* = \dfrac{\phi}{\phi_0}$ and $M^* = \dfrac{\phi_0}{p_{max}} M$, and dropping the * notation we find:

$$u = \begin{cases} D\left(\dfrac{\partial p_s}{\partial x} + 1\right) & x \geq x_c \\ \\ D\left(M\dfrac{\partial \phi}{\partial x} + \phi\right) & x \leq x_c \end{cases}$$

in the continuity equation (4).

5.2 Scaling the compression zone

Since the only activity is in the compression zone, which varies from very small at $t \approx 0$ to $O(1)$ at steady state, we rewrite the equations after rescaling $y = x/x_c$. To avoid a moving boundary problem we change $x \in [0, x_c(t)]$ to a fixed boundary problem for $y \in [0, 1]$:

$$\begin{cases} \dfrac{\partial \phi}{\partial t} = \dfrac{y}{x_c}\dfrac{dx_c}{dt}\phi_y + \dfrac{1}{x_c}\dfrac{\partial}{\partial y}\left(D\phi\left(\phi + \dfrac{M}{x_c}\phi_y\right)\right) \\ \\ \phi(1,t) = 1, \qquad \phi(0,t) + \dfrac{M}{x_c}\phi_y(0,t) = 0 \end{cases} \tag{40}$$

For h and x_c we have the equations

$$\begin{cases} \dfrac{dh}{dt} = \dfrac{p_s(1)}{h - x_c} - 1 \\ \\ 1 = h - x_c + x_c \displaystyle\int_0^1 \phi(y,t)dy \end{cases} \tag{41}$$

We interpret the splitting of equation (4) into the parabolic problem (40) and the system (41) as a method for dealing with the particular kind of discontinuities developed by the solution. At the inter-phase $x = h$, ϕ has a jump while at $x = x_c$ a discontinuity of $\dfrac{\partial \phi}{\partial x}$ appears.

The major difficulties encountered when solving the equations numerically using this approach are:

- The equations have a singularity at $t = 0$ which needs special treatment.
- There are different time-scales that govern the sedimentation process and a reliable time-step control is needed, otherwise a huge number of steps are taken.

5.3 Equivalence of formulations

We show that the Eiken and Auzerais formulations described in [15, 1] are mathematically equivalent. Identifying notations we have

$$\text{Auzerais} = \begin{cases} \phi_t = y\dfrac{\dot{x_c}}{x_c}\phi_y + \dfrac{1}{x_c}\left(u(\phi)\phi\right)_y \\ \\ \dot{h} = \dfrac{p_s(1)}{h - x_c} - 1 \end{cases} \tag{42}$$

subject to the following boundary conditions:

$$u(\phi(0,t)) = 0, \qquad \phi(1,t) = 1, \qquad u \text{ continuous at } \phi = 1,$$

i.e.

$$M(1)\phi_y = -\frac{p_s(1)}{h - x_c}.$$

The Eiken formulation is:

$$\text{Eiken} = \begin{cases} \phi_t = y\dfrac{\dot{x_c}}{x_c}\phi_y + \dfrac{1}{x_c}\left(u(\phi)\phi\right)_y \\ \\ \dot{h} = \dfrac{p_s(1)}{h - x_c} - 1 \\ \\ h - x_c + x_c \displaystyle\int_0^1 \phi(y,t)dy = 1 \end{cases} \tag{43}$$

with the conditions:

$$u(\phi(0,t)) = 0, \qquad \phi(1,t) = 1.$$

In brief, Eiken replaced the continuity of u at $\phi = 1$ by conservation of mass.

The two formulations are equivalent mathematically. Multiply the equation for ϕ_t by x_c and integrate:

$$\int_0^1 x_c \phi_t dy = \int_0^1 y\dot{x_c}\phi_y dy + \int_0^1 (u\phi)_y dy.$$

Integrate by parts and use the conditions $u(\phi(0,t)) = 0$ and $\phi(1,t) = 1$:

$$x_c(t) \int_0^1 \phi_t(y,t)dy = \dot{x}_c(t) - \dot{x}_c(t) \int_0^1 \phi(y,t)dt + u(1).$$

If $\dot{h} = -u(1)$ (continuity of u), so

$$\frac{d}{dt}\left(x_c \int_0^1 \phi dy\right) + \dot{h} = \dot{x_c}.$$

Integrating this equation gives

$$h - x_c + x_c \int_0^1 \phi dy = const,$$

i.e. the mass is constant. Conversely, if mass is conserved then $\dot{h} = -u(1)$, which proves that the two formulations are equivalent. The constant is found at $t = 0$ when $h = 1$ and $x_c = 0$.

5.4 The steady-state solution

An interesting consequence of the system (41) is that the overburden region $h \geq y \geq x_c$ never vanishes. Indeed, at steady state we have $h - x_c = p_s(1) > 0$. This means that there are always uncompressed particles "floating" above the compression zone.

At steady state the equations are simplified and reduce to $u = 0$, i.e., zero-flux throughout the compression zone. In particular, we obtain an ODE:

$$M\frac{\partial \phi}{\partial x} + \phi = 0.$$

Numerical integration of this ODE is the basis for parameter identification presented in [11].

6 Simulation in the Compression Zone Formulation

Using conservative discretization of space and the method of lines, we build a differential-algebraic equation (DAE). For large Peclet numbers, the numerical scheme must be upwind. This is done by adding artificial compressibility.

Since the solution has different time scales, we need a good time-step control and a numerical method that allows big variations in the time-step. We use implicit BDF methods. In our tests we used Backward Euler and modified a third order BDF code developed by Dahlquist and Edsberg for DAE [14].

The Eiken model (43) leads to a differential-algebraic equation (DAE) when we use the method of lines to discretize the space variables. Algebraic equations arise from the boundary conditions

$$u(\phi(0,t)) = 0, \qquad \phi(1,t) = 1, \quad t \geq 0 \tag{44}$$

and the mass conservation:

$$h - x_c + x_c \int_0^1 \phi(y,t)dy - 1 = 0, \quad t \geq 0. \tag{45}$$

The direct discretization of the "algebraic" equation (45) of the DAE is a very ill-conditioned operation. This problem is treated using standard conservation law techniques. It also turns out that a linearized model of the DAE has the differential index 2, which are in general difficult to solve numerically. We desire that our numerical treatment of the equations has the following properties:

- The space discretization should be conservative. Such a discretization guarantees e.g. exact evaluation of mass at any fixed time.
- The space discretization should be upwind. For large Peclet numbers the equations are nearly hyperbolic and upwind character is desirable.
- The time integration should be efficient. Because the problem is nonlinear each time integration is an expensive operation. The different

time scales make explicit time integration unacceptably slow. An implicit linear multistep solver with automatic time-step control for stiff problems was modified to handle the differential-algebraic system, after reducing the index to 1 by differentiation.

6.1 Conservative discretizations

To reduce the index we differentiate the equation (45) with respect to time, substitute ϕ_t using the PDE (43), and finally integrate by parts to obtain:

$$\dot{h} - \dot{x_c} + \dot{x_c}\phi(1,t) + u(\phi(1,t))\phi(1,t) = 0.$$

The zero-flux condition at $y = 0$ was used.

We denote the cell average of a quantity f_i by $\overline{f}_i = (f_i + f_{i-1})/2$ and discretize the last equation using the divided differences operator $D_- f_k = (f_k - f_{k-1})/\Delta y$ for approximating the derivative ϕ_y. Using the form of u (39) we obtain the discrete differential equation:

$$\dot{h} - \dot{x}_c + \dot{x}_c\overline{\phi}_n + \overline{k}_n\overline{\phi}_n\left(\frac{\overline{M}_n}{x_c}D_-(\overline{\phi}_n) + \overline{\phi}_n\right) = 0. \tag{46}$$

Now we discretize the PDE (43) and boundary conditions. The space derivative $\frac{\partial}{\partial y}$ is discretized using forward divided differences $D_+ f_k = (f_{k+1} - f_k)/\Delta y$

on cell averages thus obtaining:

$$\begin{cases} \phi_1 + \dfrac{\overline{M}_1}{x_c} D_+(\phi_1) = 0 \\ \\ x_c \dot{\overline{\phi}}_i - y_k \dot{x}_c D_+(\overline{\phi}_i) - D_+ \left(\overline{k}_i \overline{\phi}_i (\overline{\phi}_i + \dfrac{1}{x_c} \overline{M}_i D_-(\phi_i)) \right) = 0, \\ \qquad\qquad 2 \le i \le n-1 \\ \\ \phi_n = 1 \\ \\ \dot{h} = \dfrac{p_s(1)}{h - x_c} - 1 \\ \\ \dfrac{p_s(1)}{h - x_c} - 1 - \dfrac{\dot{x}_c}{2}(1 - \phi_{n-1}) + \overline{k}_n \overline{\phi}_n^2 + \dfrac{\overline{k}_n \overline{\phi}_n \overline{M}_n}{x_c} D_-(\overline{\phi}_n) = 0 \end{cases} \tag{47}$$

Summing over $i = 2, \ldots, n-1$ in (47) we obtain after some manipulations $\frac{d}{dt}(h - x_c + x_c \Delta y \sum_{i=1}^{n} \overline{\phi}_i) = 0$. This relation together with $h(0) = 1$ and $x_c(0) = 0$ give the corresponding discrete mass conservation principle

$$1 = h - x_c + x_c \Delta y \sum_{i=1}^{n} \overline{\phi}_i \approx h - x_c + x_c \int_0^1 \phi(y) dy.$$

This is a nonlinear function in the variables h, x_c and $\overline{\phi_i}$ and thus the last equation of the system (47) in fact no guarantee for mass conservation after time discretization. Thus each time iteration is followed by an l_2 projection onto the manifold

$$h - x_c + x_c \Delta y \sum_{i=1}^{n} \overline{\phi}_i = 1. \tag{48}$$

The equation can be written in the form $a^T z = b$. In this context, z is a vector containing all the unknowns and a is a vector containing the weights of the quadrature rule multiplied by x_c followed by -1 and 1 corresponding to x_c and h. If the time stepping yields a solution z, we form a and then solve the constrained minimization problem in the l_2 norm

$$\min_{a^T y = b} ||z - y||^2$$

to find the corrected solution. The solution to this simple minimization problem is obvious:

$$y = z - \frac{z^T a - b}{a^T a} a.$$

6.2 Upwind discretizations

For large Peclet numbers, the "advection" term ϕ_x dominates the "diffusion" term ϕ_{xx}. The exact solution develops steep fronts and boundary-layers, while the numerical solution requires very small Δx for stability. Essentially the PDE is

$$\phi_t = A\phi_x + B\phi_{xx}.$$

When $B = 0$, the equation describes flow in the direction given by $-A$. It is natural that the solution at a given point (x_0, t_0) should be influenced only by the upstream values from the past, i.e. only by $x < x_0$ if $A > 0$ or $x > x_0$ if $A < 0$. Our discretization should mimic this behavior in the limit case $B \to 0$. This is not the case of central differences which use both upstream and downstream values of the past to compute the present. The alternative is to switch from central differences to one-sided differences when needed (i.e. only when $A\Delta x$ dominates B). This is conveniently done by adding a diffusive term to the equation since

$$D_+ f_i = \frac{f_{i+1} - f_i}{\Delta x} = \frac{f_{i+1} - f_{i-1}}{2\Delta x} + \frac{\Delta x}{2} \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta x^2} = (D_0 + \frac{\Delta x}{2} D_+ D_-) f_i.$$

Note that the artificial diffusion needs to be added only at those points where $A\Delta x$ dominates B. In this case it is necessary to add only the amount that gives stability. Thus the equation is modified to

$$\phi_t = A\phi_x + B\phi_{xx} + C\phi_{xx}$$

where

$$C = \begin{cases} \dfrac{|A|\,\Delta x}{2} - B, & \text{when } B < \dfrac{|A|\,\Delta x}{2} \\ 0, & \text{otherwise} \end{cases}$$

6.3 Index 2 DAE

The direct discretization of equation (43) is of the form:

$$A\dot{y} + By = f(t), \quad y \in \mathbb{R}^n$$

If A is non-singular this is a standard system of ODE which has a unique solution through any initial point $(0, y_0)$. But in our case A is singular, there are algebraic relations which must be satisfied by solutions, and must be satisfied by the initial values also. The problem is to find the algebraic relations. For a linear system, this is trivial, but a very hard problem for non-linear systems. If the matrix $sA + B$ is non-singular for some s, there are solutions, but if it is singular for all s, there may be no unique solutions at all. One usually wants to transform DAEs into ODEs to be assured of existence of solutions. The algebraic relations are differentiated and turned into ODEs.

Definition 6.1 *The differentiation index of the system is the number of differentiations necessary to obtain a system with a non-singular A.*

We refer to [24] for details. The essential point is that while index 1 problems can be integrated with a wide class of algorithms, index 2 problems need special treatment. Now we need to determine the index of the DAE. We introduce the new variable $z = \int_0^1 \phi(y,t)dy$ and integrate the PDE (43) to obtain:

$$\begin{cases} \dot{h} = \dfrac{p}{h - x_c} - 1 \\ \\ h - x_c + x_c z = 1 \\ \\ x_c\dot{z} + \dot{x_c}z - \dot{x_c} = 1 + \dfrac{M}{x_c}\phi_y(1,t) \approx 1 - mx_c(z-1) \end{cases} \tag{49}$$

which we will use to demonstrate some properties of the system:

Theorem 6.1 *The system (49) has differential index 2.*

Proof: The linearized version of (49) with the notation $x_c z = v$ is

$$\begin{cases} \dot{h} + a(h - x_c) = 0 \\ h - x_c + v = 1 \\ \dot{v} - \dot{x_c} - cv = 0 \end{cases} \tag{50}$$

or

$$A \begin{pmatrix} \dot{h} \\ \dot{x}_c \\ \dot{v} \end{pmatrix} + B \begin{pmatrix} h \\ x_c \\ v \end{pmatrix} = f,$$

with

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} a & -a & 0 \\ 1 & -1 & 1 \\ 0 & 0 & -c \end{pmatrix}, \quad f = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

The algebraic equation is $h - x_c + v = 1$. We differentiate and the new matrices are:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 1 \\ 0 & -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} a & -a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -c \end{pmatrix}, \quad f = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Note that A is again a singular matrix. Subtracting the first and last equation from the second we obtain

$$\begin{cases} \dot{h} + a(h - x_c) = 0 \\ -\dot{x_c} + \dot{v} - ah + ax_c = 0 \\ ah - ax_c - cv = 0 \end{cases}$$

which also contains an algebraic equation. Differentiating again we obtain:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ a & -a & -c \end{pmatrix},$$

which is not singular unless $c + a = 0$. Hence, the index of our system is 2. □

6.4 BDF methods

The BDF (Backward Difference) methods are popular for solving DAE. Explicit methods can only be used if the DAE is transformed into an ODE e.g. by singular perturbation methods, see [24] and then many evaluations of a non-linear function are needed. In contrast, implicit linear multistep methods (such as BDF) are implemented directly in the DAE form. BDF have good stability properties which make them favorites also for solving stiff problems where the time-scales have different orders of magnitude.

A BDF method is based on the approximation

$$\Delta t \dot{y}_{m+k} \approx \left(\nabla + \frac{1}{2}\nabla^2 + \cdots + \frac{1}{k}\nabla^k \right) y_{m+k},$$

where $\nabla y_n = y_n - y_{n-1}$ is the backward difference operator. A BDF(k) method for solving the equation $\dot{y} = f(y)$ is

$$\left(\nabla + \frac{1}{2}\nabla^2 + \cdots + \frac{1}{k}\nabla^k \right) y_{m+k} = \Delta t f(y_{m+k}).$$

Such discretizations are zero-stable for $k \leq 6$ and their order of consistency is k, one less than optimal order of k-step methods. The stability regions for BDF(k), k=1,2,3,4, are given in Figure 19.

The implementation of the BDF method is easily modified to solve an ODE in the implicit form $A\dot{y} = f(y)$ in which equation (47) can be written. At each step we need to solve a nonlinear equation of the type:

$$A \left(\nabla + \frac{1}{2}\nabla^2 + \cdots + \frac{1}{k}\nabla^k \right) y_{m+k} = \Delta t f(y_{m+k}). \tag{51}$$

The unknown is the vector $y_{m+k} = (x_c, h, \phi_1, \ldots, \phi_n)^T$, representing the solution at time $t = (m+k)\Delta t$. Equation (51) is solved using a Newton-type iteration for minimizing the residual $||A\dot{y} - f(y)||$. The implementation makes use of the exact Jacobians A and $\frac{\partial f}{\partial y}$ (with the permeability computed at the last time-step) to obtain fast convergence. The minimization is constrained by the mass-conservation equation (48).

We use BDF(3), which is "almost" A-stable and has an acceptable consistency order. The implementation of BDF(3) in [14] computes the solution in

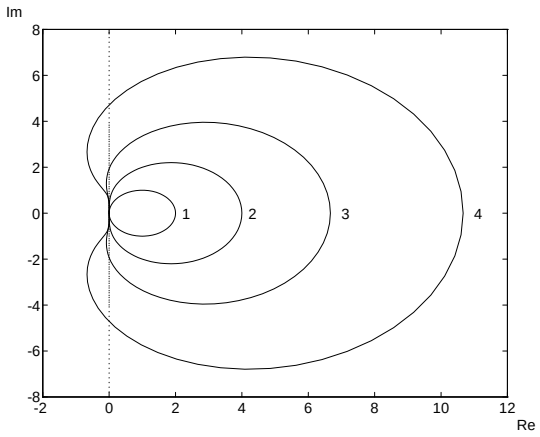

Figure 19: Stability regions for BDF(k) methods, k=1,2,3,4. The stability regions lie outside the curves

blocks of three steps, keeping Δt and the Jacobian constant. A new Δt and a new Jacobian is computed after each block.

Note that BDF(3) has to be started with a lower-order method. We use 3 steps with BDF(1) with a very small constant time-step Δt as a starter. Due to a robust time-step control, a suitable Δt (i.e. adapted to the time-scale of the solution) is found after a few iterations as seen in Figure 20.

The integrator uses a relatively low number of time-steps, in our tests ≈ 100 time-steps were used to reach steady-state. If the solution is needed at a priori given times, cubic interpolation is performed.

6.5 The initial profile

Scaling the "cake" has the disadvantage of introducing a singularity at $t = 0$. Since $x_c(0) = 0$ the scaling

$$y = \frac{x}{x_c}$$

is impossible at $t = 0$. The way out of this artificial Big Bang phenomenon is to start integration at $t = t_0 > 0$. The problem now is that the volume

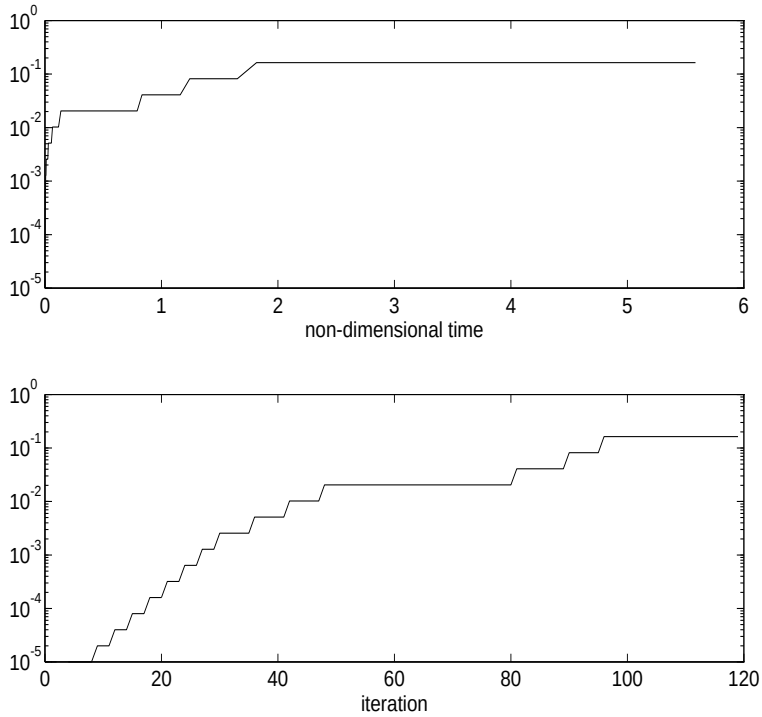

Figure 20: Step-size Δt used by BDF(3) as a function of non-dimensional time (top) and of the iteration (bottom)

fraction is no longer constant ϕ_0 and we must determine

$$h(t_0), \quad x_c(t_0), \quad \text{and} \quad \phi(y, t_0), \quad 0 < y < 1.$$

The relation between $h(t_0)$ and t_0 is given by the initial velocity u_0. For small t_0 and $h(t_0) \approx 1$ we use

$$\frac{dh}{dt} = -u_0,$$

and thus

$$h(t_0) = 1 - t_0 u_0.$$

We approximate $y \mapsto \phi(y, t_0)$ by a quadratic which we determine along with $x_c(t_0)$. Using $\phi(1, t_0) = 1$ we have

$$\phi(y) = a(y-1)^2 + b(y-1) + 1, \quad \phi_y = 2a(y-1) + b.$$

Conservation of mass (45) yields

$$\frac{a}{3} - \frac{b}{2} - \frac{1-h}{x_c} = 0. \tag{52}$$

The zero-flux condition at the bottom $y = 0$ in (40) gives

$$\frac{M}{x_c}(-2a + b) + a - b + 1 = 0. \tag{53}$$

Finally, the continuity of the flux at the top of the "cake" $y = 1$ is

$$\frac{p_s(1)}{h - x_c} + \frac{M}{x_c} b = 0. \tag{54}$$

We assume constant $M = M(1)$ since ϕ is always close to 1 and M is smooth as seen in Figure 21.

We solve the slightly non-linear system (52–54) in the unknowns a, b, x_c with a few Newton iterations.

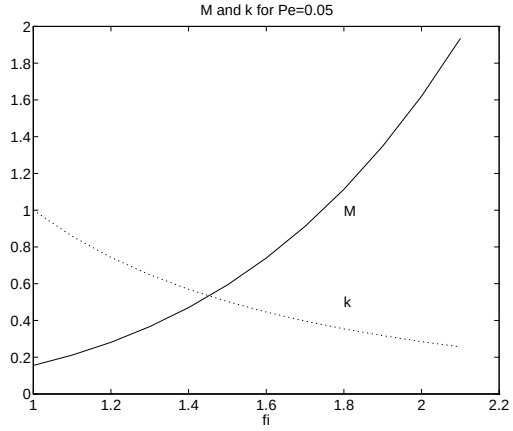

Figure 21: Smooth variation of k and M

6.6 Numerical experiments

A Matlab based program package [11, 12] was developed for both simulation and identification problems. Apart from the solvers, the package contains a number of utility functions for handling graphics, experimental data, etc. It contains some 60 functions that manage a graphics interface, pre-processing of experimental data and post-processing of results. This package is in use with Alfa–Laval Separation A/S, Denmark.

The numerical experiments below were performed with the package using the parameters given in Section 4.3. The results of the simulation are shown in Figure 22 (right).

For comparison, the experimental data (volume fractions at different times), courtesy of Lennart Bergström, is displayed in Figure 22(left).

The Compression Zone method yields values for the two moving fronts $x_c(t)$ and $h(t)$, seen in Figure 23. As predicted by theory, the two fronts are separated at all times, see Section 5.4. At steady state, we have $h - x_c = 0.0097$ which is consistent with the simulated value of 0.0101 after $t = 126$ days. In this simulation the compression zone was discretized with $N = 20$ points. Note that "overshoot" of the compression zone at $t \approx 20$ days.

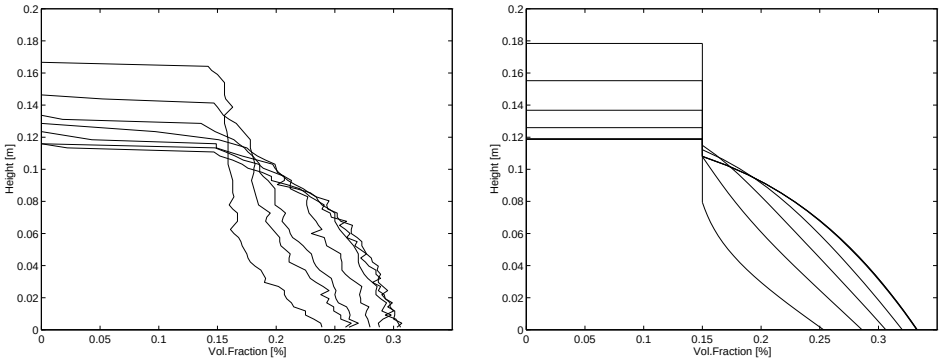

Figure 22: Experimental and simulated volume fractions $\phi(x,t)$ at times $t = 6, 13, 20, 27, 39, 75, 126$ days.

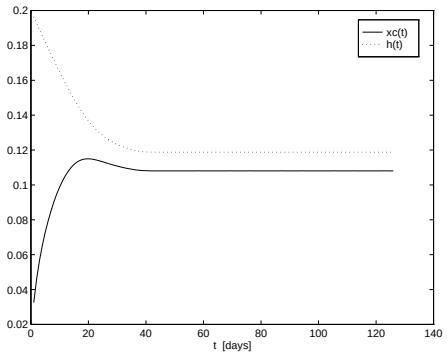

Figure 23: Evolution of the fronts $x_c(t)$ and $h(t)$ for $0 < t \leq 126$.

6.7 Conclusions

Extensive experimentation with various approaches show that:

- For large Peclet numbers, upwind discretizations are necessary.
- Conservative discretizations do not yield a mass-conserving solution since after scaling, mass is not a linear functional of the solution. This is corrected by projection on the constant mass manifold.
- The moving grid has the advantage of producing smooth solutions for overburden and sediment heights.
- The major disadvantage of the moving grid is the introduction of the artificial initial singularity, and of course the difficulty of generalizing to 2-3 D.
- The time-step control is important. Our tests show a variation of Δt from 10^{-5} to 10^{-2} after the transient and then a slow increase to 10^{-1} around steady state.

7 A Comparison of the Two Formulations

We compare the pros and cons of the numerical implementations of the two formulations. On the left side we have the mass–pressure formulation and on the right side the compression zone formulation. The two methods complement each other, so there is no absolute winner. The qualities of the first formulation are listed from good to bad, while the corresponding qualities of the second formulation are given in the reverse order.

The compression zone implementation has been refined to professional, commercial code standards and is in use in industry. The mass–pressure code is of research nature and we have a pilot implementation at this time.

$\oplus$ Can be extended to multiple dimensions	$\ominus$ Works only in 1-D
$\oplus$ Can handle the values $0 < \phi < \phi_0$ and produces good profiles in the overburden.	$\ominus$ Works for $\phi \geq \phi_0$ or $\phi = 0$ only
$\oplus$ Needs only $\phi(.,0)$ as initial values which is a measurable quantity.	$\ominus$ Cannot handle $\phi(.,0) = \phi_0$. In this case $x_c(0) = 0$ which introduces a singularity in the scaling at $t = 0$. Needs a guess for $x_c(0), h(0), \phi(.,0)$.
$\ominus$ Solve a $2n$ system: n linear but coupled equations and n nonlinear but uncoupled equations.	$\oplus$ Solve a n coupled and nonlinear equations.
$\ominus$ Must use many grid points (in space) to resolve discontinuities. Can over-resolve the smooth parts of the solution, in particular the clear-water region.	$\oplus$ Resolve the compression zone only with a fixed number of grid points.

References

[1] F.M. Auzerais, R. Jackson, and W.B. Russel. The resolution of shocks and the effects of compressible sediments in transient settling. *J. Fluid Mech.*, 195, 1988.

[2] J.D. Beasley and S. Torquato. New bounds on the permeability of a random array of spheres,. *J. Fluid Mech.*, 2, 1989.

[3] L. Bergström. Sedimentation of flocculated alumina suspensions: γ-ray measurements and comparison with model. *J. Chem. Soc.*, 1992.

[4] L. Bergström, C. Schilling, and I. Aksay. Consolidation behavior of flocculated alumina suspensions. *J. Am. Ceram. Soc.*, 1992.

[5] H. Brezis. *Operateurs Maximaux Monotones.* Mathematics Studies. North Holland, Amterdam, 1973.

[6] H.C. Brinkman. A calculation of the viscous force exerted by a flowing fluid on a dense swarm of particles. *Appl. Sci. Res.*, A1:27–34, 1947.

[7] R.W. Cottle, F. Giannesi, and J.L. Lions. *Variational Inequalities and Complementarity Problems.* John Wiley, New York, 1980.

[8] J. Crank. *Free and Moving Boundary Problems.* Clarendon Press, Oxford, 1984.

[9] G. Dahlquist and A. Björk. Numerical methods. 2-nd ed. in print, 1996.

[10] S. Diehl. *Conservation Laws with Application to Continuous Sedimentation.* PhD thesis, Lund Institute of Technology, 1995.

[11] M. Dorobantu. Numerical integration of 1d consolidation models. Master's thesis, NADA – Royal Institute of Technology, Stockholm, 1995.

[12] M. Dorobantu, J. Eiken, and J. Oppelstrup. 1d consolidation simulator. Technical report, Alfa-Laval Separation, Stockholm, 1994.

[13] M. Dorobantu, J. Eiken, and J. Oppelstrup. Numerical integration of 1d consolidation models. Technical report, Alfa-Laval Separation, Stockholm, 1994.

[14] L. Edsberg and P.A. Wedin. Diffpar a toolbox for parameter estimation in ode-systems. Technical report, NADA – Royal Institute of Technology, Stockholm, 1993.

[15] J. Eiken. Consolidation of flocculated suspensions. Technical report, Alfa-Laval Separation, Stockholm, 1993.

[16] C.M. Elliot and J.R. Ockendon. *Weak and Variational Methods for Moving Boundary Problems.* Research Notes in Mathematics. Pitman, 1982.

[17] G. Eriksson and G. Dahlquist. On an inverse non-linear diffusion problem. In Deuflhard P. and Hairer E., editors, *Progress in Scientific Computing 2*, chapter Numerical Treatment of Inverse Problems in Differential and Integral Equations. Birkhäuser, Basel, 1983.

[18] R. Glowinski, J.L. Lions, and R. Tremolieres. *Analyse Numerique des Inequations Variationnelles.* Bordas, Paris, 1976.

[19] E. Godlewski and P.A. Raviart. *Hyperbolic Systems of Conservation Laws.* Ellipses, Paris, 1991.

[20] A. Grace. Optimization toolbox. Manual: The MathWorks, Inc., 1992.

[21] H.O. Kreiss, B. Gustafsson, and J. Oliger. *Time Dependent Problems and Difference Methods.* John Wiley & Sons, inc, 1995.

[22] R. LeVeque. *Numerical Methods for Conservation Laws.* Birkhäuser, Basel, 1992.

[23] A. Visintin. The Stefan problem for a class of degenerate parabolic equations. In Fasano and Primicerio, editors, *Free Boundary Problems*, volume II of *Research Notes in Mathematics.* Pitman, 1983.

[24] G. Wanner, E. Hairer, and S.P. Nörsett. *Solving Ordinary Differential Equations.* Springer Verlag, 1991.

[25] J. Ystrom. On convective effects in sedimentation of concentrated suspensions with high effective gravity. Technical report, NADA – Royal Institute of Technology, Stockholm, 1993.

[26] J. Yström. *On the Numerical Modelind of Concentrated Suspensions and of Viscoelastic Fluids.* PhD thesis, Royal Institute of Technology, 1996.