Monotonicity Conditions for Discretization of Parabolic Conservation Laws

Master of Science Thesis in Applied Mathematics

Hilde Kristine Hvidevold

Department of Mathematics
University of Bergen

June 2, 2009

Preface

First of all I want to thank my supervisor Jan Martin Nordbotten for his brilliant mind and his genuine enthusiasm. Thanks to all my friends at the math department for two fantastic years. I specially want to thank Hanne Christine for five enjoyable year in Bergen. We did it! Finally I want to thank my family, friends, and my boyfriend Fredrik.

Contents

Introduction

Modelling of flow in porous media is an important topic of research within applied mathematics. Understanding flow in porous media is vital not only in the oil industry, but for instance also in ground water hydrology and in geothermal energy extractions. The models are built upon two physical principles; mass conservation and Darcy's law, which relates the force on the fluid and the fluid velocity. These two laws result in the time dependent parabolic conservation equation and the time independent elliptic conservation equation, which are the governing equations for flow in porous media.

In general we have to utilize numerical methods to solve these equations. We will study control volume methods, and in particular one MultiPoint Flux Approximation (MPFA) method. Control volume methods are designed to preserve the conservation property, and MPFA methods in particular handle complex geological structures in the reservoir. The MPFA method was derived independently and simultaneously by Aavatsmark et al. [4] and Edwards and Rogers [14].

When we discretize an equation we want the discrete system to maintain the same properties as the continuous problem. A discretization may not capture all structures of the equations, and we must emphasize on those properties which seems to be significant for our system. Maximum principles are important properties of elliptic and parabolic differential equations. Discretization methods which satisfy discrete analogues of these maximum principles are called *monotone*. Control volume methods are in general not monotone, and it is desirable to establish conditions under which discrete maximum principles are satisfied.

In the recent years monotonicity of control volume methods for elliptic equations has been studied. A discrete maximum principle is established in Keilegavlen et al. [18], and a set of monotonicity conditions on general quadrilateral grids has been derived in Nordbotten et al. [23]. Monotonicity criteria for parabolic equations have not yet been studied. We will therefore in this thesis extend the already existing monotonicity conditions for elliptic equations to a set of conditions for parabolic equations. These conditions

is derived under the assumption that the discrete maximum principle for parabolic equations is the same as the principle for elliptic problem. It turns out that these conditions are stricter than the elliptic conditions.

Since the maximum principle for the time discrete parabolic equation is different from the principle for the elliptic equation, it may be necessary to reformulate the discrete maximum principle. It is not obvious how this shall be done. We will therefore discuss various formulations of a time discrete maximum principle together with numerical examples.

In Chapter 1 we give an introduction to reservoir mechanics, and the two model equations which we will use throughout this thesis. In Chapter 3 we give a detailed derivation of the two point flux approximation and the MPFA method. The maximum principles and discrete analogous are proposed in Chapter 4. In Chapter 5 we derive monotonicity criteria on general quadrilateral grids. Then we especially study these for a specific and simple case. Analytical and numeral results are illustrated in Chapter 6. Here we also discuss alternative formulations of the discrete maximum principle with together with numerical examples. Finally we summarize, conclude and propose further work in Chapter 7.

Chapter 1

Reservoir Mechanics

In this chapter we will look at relevant theory of reservoir mechanics. The presentation is mainly based upon the book of Pettersen [24] and the lecture note of Aavatsmark [3]. We start by introducing some basic reservoir and fluid properties. Next, we consider equations describing flow in porous media; the principle of mass conservation and Darcy's law. Based upon these equations we formulate two models for single phase flow, which we will use throughout the thesis.

1.1 Porous Media

In reservoir mechanics we study the flow of fluids in porous media. A reservoir is a porous geological structure with fluids, which can be either gas or liquid, filling the void space. Most reservoir rocks and formations are composed of compressed minerals. These rocks may be considered as solid, but in reality consist of a fine structure of pores and grains. The void space in these rocks is a complex structure of connected and isolated pores, as illustrated in Figure 1.1. We will refer to such materials as porous media.

The interconnected system of pores forms an irregular lattice of pore bodies (junctions) and pore throats (connections). The radii of the pore bodies and pore throats may vary over many length scales and is typically in the range of 1 mm - 1μ m. Hence, a model can be derived where the flow in each of the pores is given by equations from fluid mechanics. However, the complex structure and the fact that such a detailed knowledge about the microscopic structure is unknown make this an almost impossible problem to solve. So, when modelling flow in porous media we consider a macroscopic model. By this we mean that the irregularities of the porous medium network can be considered as random variations with a well defined average. Thus

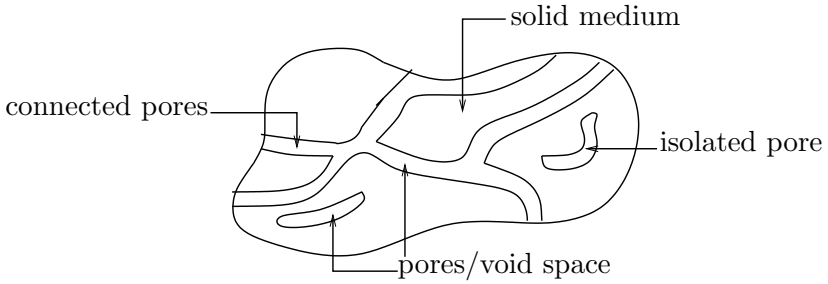


Figure 1.1: *The structure of a porous media, it consists of solid medium and void space. The void space is a complex structure of connected pores and isolated pores.*

we may define velocity, pressure and other quantities as averages over a sufficiently large reference volume. In the literature, e.g. Bear [11], such a reference volume is denoted a representative elementary volume (REV). By introducing a REV the geometrical properties of different porous media are characterised by their *porosity* and *permeability*.

1.1.1 Porosity

Porosity is a geometric property of the solid medium, and indicates the amount of volume available for fluid flow. It is defined by the ratio

$$\phi = \frac{V_{\text{Pores}}}{V_{\text{Total}}},$$

where V_{Pores} is the volume of connected pores and V_{Total} is the total or bulk volume of the material, including the solid and void components. We only consider the connected pore channels since these are the only part of the void space which contributes to the flow.

1.1.2 Permeability

Permeability, $\boldsymbol{K}$, is a measure of the ability of a material to transmit fluids, and can be interpreted as the conductivity of the porous media. It is of great importance when determining the flow of hydrocarbons in reservoirs, and of groundwater in aquifers. The permeability is a property of the porous media only, not the fluid. The unit for permeability is Darcy, and one Darcy is approximately $0.987 \cdot 10^{-12}$ m^2.

The porous media we consider are typically generated by sedimentation processes. In such processes fine grain particles are oriented in ways that

make the porous media anisotropic, i.e. the permeability is directional dependent. To account for this we need to represent the permeability by a second order tensor, $\boldsymbol{K} = \{k_{ij}\}$. It may be shown that $\boldsymbol{K}$ in general has to be symmetric, $\boldsymbol{K} = \boldsymbol{K}^T$. If the permeability varies with spatial location, $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{x})$, we say that the medium is heterogeneous. In its simplest form the medium is isotropic and homogeneous, and the permeability is constant.

1.2 Equations for Flow in Porous Media

In this section we will introduce the governing equations for flow and transport of one single fluid through a porous media. A good understanding of single phase flow is essential for handling multiphase flow. In general, more than one fluid can co-existing within an REV, and they may be distinguished at the pore scale by fluid-fluid interfaces. Despite the fact that single phase models are rarely god enough to describe a typically reservoir problem, the solution techniques for single phase flow are very important because they act as building blocks for solving multiphase problems. We will not consider multiphase flow in this thesis, and refer to e.g. Bear [11] for more information.

1.2.1 Conservation of Mass

Equations concerning flow in porous media are based on a conservation law. Conservation of mass is a fundamental physical principle, and can be formulate by looking at an arbitrary fixed geometrical volume, Ω, inside a reservoir and require that the following equation is valid:

$$\{\text{accumulation}\} + \{\text{outflow}\} = \{\text{source/sink}\}.$$

Let ϕ be the porosity and $\boldsymbol{v}$ the volumetric flow velocity of the fluid inside Ω. The volumetric flow velocity is the rate of volume flow across a unit area $[m^3/(s \cdot m^2) = m/s]$. Then the fluid concentration is $\phi\rho$, and the momentum of the fluid through a surface, which is a measure of the volume flowing through a surface per time, is $\rho\boldsymbol{v}$. The mass conservation equation then takes the form

$$\frac{\partial}{\partial t}\int_\Omega (\phi\rho)\, dt + \int_{\partial\Omega} \rho\boldsymbol{v}\cdot\boldsymbol{n}\, d\sigma = \int_\Omega Q\, d\tau, \tag{1.1}$$

where $\partial\Omega$ is the boundary to the volume Ω, and $\boldsymbol{n}$ is outer unit normal vector to the boundary, as illustrated in Figure 1.2. The source term Q represents a production or injection well in the reservoir.

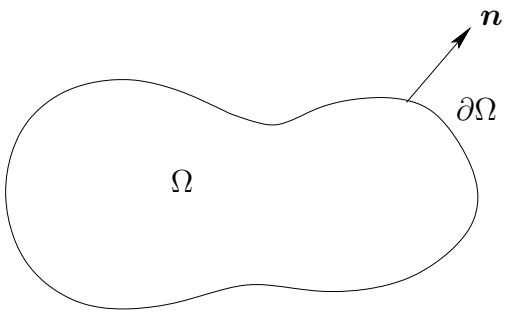


Figure 1.2: *An arbitrary fixed geometrical volume* Ω *with boundary* $\partial\Omega$ *and outer unit normal* $\boldsymbol{n}$.

Since the geometrical volume Ω is said to be fixed, it is independent of time. If $\boldsymbol{v}$ is continuously differentiable we can use Gauss theorem to get the equation

$$\int_\Omega \left[\frac{\partial}{\partial t}(\phi\rho) + \nabla \cdot \rho\boldsymbol{v} - Q \right] d\tau = 0. \tag{1.2}$$

When the volume Ω is arbitrary we may write

$$\frac{\partial}{\partial t}(\phi\rho) + \nabla \cdot \rho\boldsymbol{v} - Q = 0.$$

This is the differential form of the continuity equation. The latter equation is only valid if the terms involved are sufficiently smooth. If not, the two integral forms of the continuity equation (1.2) and (1.1) are still valid.

1.2.2 Darcy's law

In porous media, flow of fluids through pores is strongly affected by friction between the fluid and the pore walls. A mathematical equation that relates the viscous forces to applied potentials was first formulated by Henry Darcy in the 19th century. Hence, the volumetric flow velocity $\boldsymbol{v}$ is referred to as the Darcy velocity and is determined by Darcy's law. Darcy conducted experiments with flow of water vertically through different types of sand. He concluded that the flow through the sand was proportional to the potential difference between the top and bottom pressure.

In modern form, Darcy's law may be expressed by:

$$\boldsymbol{v} = -\frac{1}{\mu}\boldsymbol{K}(\nabla p - \rho g \boldsymbol{k}).$$

Here p is the pressure, ρ is the density, g is the gravity constant and $\boldsymbol{K}$ is the permeability. The viscosity, μ, is a property of the fluid and is the

internal resistance to flow in the fluid. As an example water has low viscosity while syrup has high. The equation above states that the fluid will move from regions of high pressure to regions of low pressure, and the velocity is dependent on the medium and phase conductivity.

We define the conductivity $\boldsymbol{K} = \frac{\boldsymbol{K}}{\mu}$, which expresses the ease with which a fluid is transported through a porous media, and depends on both the solid and fluid properties, e.g. [11]. For single phase flow, the viscosity effect on the conductivity will be insignificant with respect to flow characteristics. The interesting feature for us is the effect of media anisotropy and heterogeneity in reservoirs, therefore the term permeability will be used instead of conductivity.

1.2.3 Equations of state

We shall only study isothermal processes. An isothermal process is a thermodynamic process in which the temperature of the system remains constant. The heat transfer into or out of the system typically must happen at such a slow rate that the thermal equilibrium is maintained. When this is fulfilled, we have pressure and density as functions of each other

$$\rho = \rho(p). \tag{1.3}$$

This relation is referred to as an equation of state. Viscosity and pressure are also functions of each other, $\mu = \mu(p)$.

Later it will turn out to be convenient with an equation that expresses the density with respect to pressure. By using the compressibility we obtain such an expression. Compressibility is a measure of the relative volume change of a fluid or solid as a response to a pressure change. The compressibility is defined by

$$c = -\frac{1}{V}\frac{dV}{dp} = \frac{1}{\rho}\frac{d\rho}{dp}. \tag{1.4}$$

Now the basic reservoir properties is discussed and the different equation are set up, we can summarize the model and derive different formulations.

1.3 Models for Single Phase Flow

To summarize our model we have; the continuity equation

$$\frac{\partial}{\partial t}(\phi\rho) + \nabla \cdot \rho\boldsymbol{v} - Q = 0, \tag{1.5}$$

n component equations

$$\boldsymbol{v} = -\boldsymbol{K}(\nabla p - \rho g \boldsymbol{k}), \tag{1.6}$$

and one equation of state (1.3). These $n+2$ equations possess $n+2$ unknowns, namely $\boldsymbol{v}$, p and ρ, hence our system is closed.

1.3.1 Incompressible Fluids

For an incompressible fluid the density does not change with pressure, hence ρ is constant and from equation (1.4) $c = 0$. Assume further that the solid medium is incompressible, ϕ is constant. Substituting for the Darcy velocity (1.6) in equation (1.5) then gives

$$-\nabla \cdot (\boldsymbol{K}(\nabla p - \rho g \boldsymbol{k})) = \frac{Q}{\rho}. \tag{1.7}$$

Here we can introduce the flow potential defined by

$$\psi = p - \rho g z, \quad \text{such that} \quad \nabla \psi = \nabla p - \rho g \boldsymbol{k}.$$

The potential ψ is the pressure which exceeds the hydrostatic pressure. By introducing the potential, equation (1.7) takes the form

$$-\nabla \cdot (\boldsymbol{K} \nabla \psi) = \frac{Q}{\rho}. \tag{1.8}$$

This is an elliptic differential equation with the potential ψ as dependent variable. The equation state that the flow over the boundary of a given domain must be balanced by possible sources or since inside the domain, hence the accumulation is zero.

1.3.2 Compressible Fluids

Further in our study, we will only consider solutions of our models in two dimensions, e.g. horizontal flow. We will therefore neglect gravity and assume that the solid medium is incompressible. From equation (1.4) it follows that $d\rho = c\rho dp$, and if we use the property $\nabla \cdot (uv) = u\nabla \cdot v + \nabla u \cdot v$ we get the conservation equation (1.5) included the Darcy velocity (1.6) without the gravity term

$$\begin{aligned} 0 &= \phi \rho c \frac{\partial p}{\partial t} - \nabla \cdot (\rho \boldsymbol{K} \nabla p) - Q \\ &= \phi \rho c \frac{\partial p}{\partial t} - \rho \nabla \cdot (\boldsymbol{K} \nabla p) - \rho c \nabla p^T \boldsymbol{K} \nabla p - Q. \end{aligned} \tag{1.9}$$

Equation (1.9) is a non-linear equation. It may be convenient to linearise the equation, then we obtain an equation for weakly compressible fluids.

Weakly Compressible Fluids

We assume that

$$c\nabla p^T \boldsymbol{K}\nabla p \ll |\nabla \cdot (\boldsymbol{K}\nabla p)|\,.$$

This can be a reasonable assumption for two reasons; if the changes in pressure is small, it makes sense to assume that $(\nabla p)^2 < \nabla \cdot \nabla p$. In addition, if we assume that the compressibility is a small constant, $c \ll 1$, then $c(\nabla p)^2 \ll \nabla \cdot \nabla p$.

When this inequality is satisfied we may neglect the smallest term, and equation (1.9) simplifies to

$$\phi c\frac{\partial p}{\partial t} - \nabla \cdot (\boldsymbol{K}\nabla p) = \frac{Q}{\rho}. \tag{1.10}$$

This is a linear parabolic differential equation in the pressure p . Flow in confined aquifers is described by an equation on this form. In Bear [12] a *confined aquifer* is defined as "an aquifer bounded from above and from below by impervious formations, formations which is incapable of transmitting significant quantities under ordinary field conditions". If gravity is not neglected, the arguments above may still be used and an additional gravity term will appear, but the flow characteristic in the reservoir will be the same. The analysis and results throughout this thesis will therefore not collapse if gravity is included.

Now that we have defined our model equations (1.10) and (1.8), we will in the next chapter go through a discretization technique for solving these equations.

Chapter 2

Control Volume Methods

In this chapter we are going to consider control volume methods as discretizations of the parabolic equation (1.10). The discretization techniques both for two point flux- and multipoint flux- approximation will be explained, and the composition of the system matrix for the discrete system will be given. Our presentation of the control volume method is mainly based on Aavatsmark [2].

2.1 Discretization

Most equations describing physical phenomena can not be solved analytically. When analytic results are not available, simplifications and numerical methods are needed. To be able to solve a mathematical problem numerically, we need a discrete representation of the continuous problem. This process is called discretization. To discretize a particular problem we need information about the domain in which the equations are going to be solved, as well as initial and boundary values. When this information is given, we can start by defining a finite number of points on the domain where our solution shall be given a discrete representation. This is solved by defining a grid on our domain.

A grid is created by defining neighbourhoods between given points. For each point in the grid, connection lines are drawn between the point and its neighbouring points. The connection line dividing two cells will be termed interface or edge. The neighbourhood between the points must be chosen such that the edges between the points do not cross each other. Such a grid is shown in Figure 2.1. Those subdomains which is separated by the connection lines in the grid are called grid cells. We use the geometrical midpoint, which is the average of the coordinates of the four vertices that

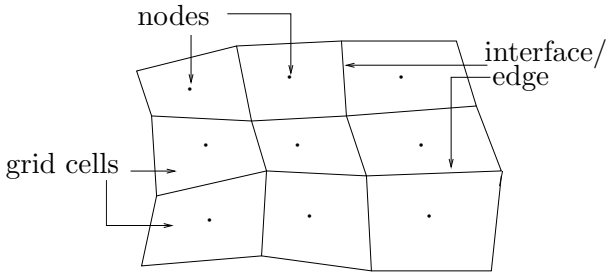


Figure 2.1: *A cell-centred grid with nine cells and nine nodes located in the centre of each cell..*

define the cell, as a node point. Such a grid is called a cell-centred grid. The grid in Figure 2.1 has nine quadrilateral cells with nine corresponding nodes. Nodes may be located on cell interfaces or corners as well as inside the domain. The location of the nodes is dependent on the grid type and the variable which is discretized.

In each cell we assemble a linear set of equations for a finite number of values for the dependent variables being approximated. To derive the linear set of equations, we use the nodes to give a discrete representation of a function defined on the whole grid. Let us denote the nodes in a grid x_i, where $i = 1, \cdots, n$ and let a function $g(x)$ be defined on the grid. A discrete representation of the function g is then given by the vector

$$\boldsymbol{g} = [g(x_1), \cdots, g(x_n)]^T.$$

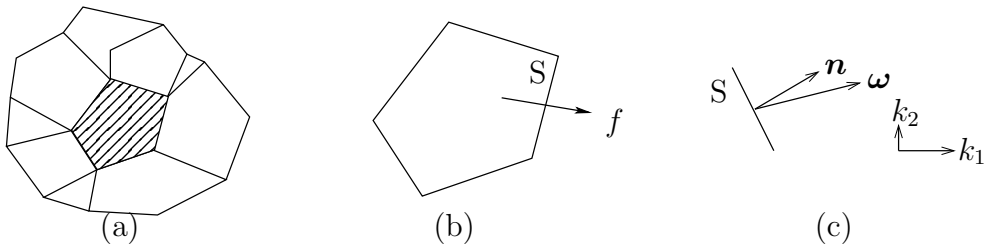
The grid cells can have different shape and size. We will only use uniform grid in our numerical computations. A **regular grid** is a grid where all the cells have the same shape, e.g. square, parallelogram or rectangle. A **uniform grid** is a regular grid where all the cells have the same size.

A number of techniques are derived to discretize partial differential equations, each of them having advantages and disadvantages. We are going to discretize a conservation equation, so it is favourable that the conservation principle is preserved.

Definition 2.1. A discretization is said to be locally conservative if

i) the flow normal to an interface is the same on each side of the interface,

ii) the outflow of a cell equals the source in the cell subtracted the accumulation in that cell.



(a) (b) (c)

Figure 2.2: (a) *A polygon grid,* (b) *A polygon cell illustrating the flux* f *through an interface,* S*, and* (c) *The vectors* $\boldsymbol{\omega}$ *and* $\boldsymbol{n}$ *for an edge* S*.*

2.2 Control Volume Formulation

Control volume methods are based on the integral formulation of the differential equation which is to be discretized. These methods preform spatial discretizations and therefore only the elliptic part of equation (1.10) will be considered here, i.e. equation (1.8). We will come back to discretization of the time dependent term in the following chapters.

The discretization is carried out by first defining a grid over the domain of the equation and then make use of the integral formulation on each cell in the grid. The integral formulation is initially conservative when Definition 2.1 is fulfilled. In addition, all control volume methods yields an explicit expression for the flux, in contrast mixed finite element methods do not have this property. This is valuable since it enables fully implicit multiphase flow simulations [19]. Control volumes may also be referred to as finite volumes in some literatures.

The cells are denoted control volumes since the principle of mass conservation, which initially was set up for an arbitrary volume, is used on the grid cells. Figure 2.2(a) illustrates an arbitrary polygon grid. Let us consider the shaded cell, and denote this Ω_j. Equation (1.8) for this cell is

$$\int_{\partial\Omega_j} \boldsymbol{f} \cdot \boldsymbol{n} \, d\sigma = \int_{\Omega_j} q \, d\tau, \tag{2.1}$$

where $\boldsymbol{f} = -(\boldsymbol{K}\nabla p)$, $q = Q/\rho$, and $\boldsymbol{n}$ is the outer unit normal to the interface S as illustrated in Figure 2.2 (c). The permeability $\boldsymbol{K}$ is assumed to be symmetric and positive definite. The flux, f, over the interface, S, is defined through the following equation,

$$f = \int_S \boldsymbol{f} \cdot \boldsymbol{n} \, d\sigma = -\int_S \boldsymbol{n}^{\boldsymbol{T}} \boldsymbol{K} \nabla p \, d\sigma = -\int_S \boldsymbol{\omega} \nabla p \, d\sigma, \tag{2.2}$$

where the vector $\boldsymbol{\omega}$ is defined as $\boldsymbol{\omega} = \boldsymbol{n}^T \boldsymbol{K}$. The flux over the over the interface S is illustrated in Figure 2.2 (b).

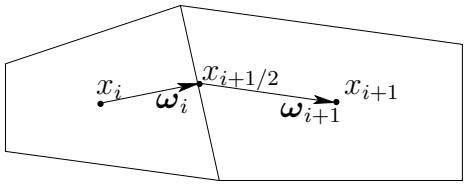


Figure 2.3: *The directions $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_{i+1}$ connect $x_{i+1/2}$ with x_i and x_{i+1}, respectively.*

The control volume formulation for equation (2.1) can be stated as

$$\sum_i f_i^{(j)} = q_j, \tag{2.3}$$

The term q_j is the source integrated over cell number j and $f_i^{(j)}$ is the flux from cell j through interface number i. The flux terms $f_i^{(j)}$ are functions of the variable p. For the elliptic case, i.e. incompressible flow, the variable p is a potential (see Section 1.3.1). When we include the compressibility, p denotes the pressure. For simplicity, we use the phrase pressure in both cases.

If we assume that the source q is evenly distributed within each cell, it is trivial to compute the right-hand side in equation (2.1). Consequently, since the pressure is the unknown, the challenge in the control volume formulation is to calculate the flux over the edges.

We shall assume that the permeability $\boldsymbol{K}$ is constant in each cell, and that the interfaces are straight lines. The vector $\boldsymbol{\omega}$ is then constant on each interface in a cell. Then, $\boldsymbol{n}\cdot\boldsymbol{\omega} = \boldsymbol{n}^T\boldsymbol{K}\boldsymbol{n} > 0$, since $\boldsymbol{K}$ is positive definite, and hence $\boldsymbol{\omega}$ points in the same cell as $\boldsymbol{n}$. However, since $\boldsymbol{K}$ can have different values in the two cells sharing the same edge, it is possible for the vector $\boldsymbol{\omega}$ to have different direction on each side of an edge.

2.2.1 Two Point Flux Approximation (TPFA)

The flux (2.2) may be approximated by a two point flux approximation

$$f_{i+1/2} \approx t_{i+1/2}(p_i - p_{i+1}). \tag{2.4}$$

The coefficients t_i resemble conductances and will be called the transmissibilities. Here $t_{i+1/2}$ is the transmissibility of interface $(i+1/2)$, and p_i and p_{i+1} are the pressure at the cell centres of the adjacent cells x_i and x_{i+1}. The flux, $f_{i+1/2}$, over interface $(i+1/2)$ will for simplicity be denoted f.

According to equation (2.2), the flux is expressed as the directional derivatives of p in $\boldsymbol{\omega}$-direction multiplied by the length of $\boldsymbol{\omega}$ integrated over the edge, S. We shall assume that ∇p is constant on each side of an interface. The flux on each side of S can be approximated with a pressure difference between two points if and only if the connection line between the two points is parallel with $\boldsymbol{\omega}$. Only in this case the pressure difference between the points can give an approximation to the directional derivative along $\boldsymbol{\omega}$.

Figure 2.3 illustrates two cells in a two dimensional grid. The cell centre in each cell is x_i and x_{i+1}. Notice that it is the directions of $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_{i+1}$ which are important. The directions must be such that the line connecting $x_{i+1/2}$ with x_i coincide with $\boldsymbol{\omega}_i$, and line connecting $x_{i+1/2}$ with x_{i+1} coincide with $\boldsymbol{\omega}_{i+1}$, as illustrated in Figure 2.3. The flux from cell i through interface S can now be expressed by approximating the directional derivative of $\boldsymbol{\omega}_i$ with the pressure difference between x_i and x_{i+1}

$$p_i - \bar{p}_{i+1/2} = \frac{f}{\Gamma_{i+1/2}} \frac{||x_{i+1/2} - x_i||_2}{||\boldsymbol{K}_i \boldsymbol{n}||_2}. \tag{2.5}$$

Here $\Gamma_{i+1/2}$ is the length of the interface, and $\bar{p}_{i+1/2}$ is the value of p in at the interface. The same procedure can be used to find the flux from cell $(i+1)$ through interface S,

$$\bar{p}_{i+1/2} - p_{i+1} = \frac{f}{\Gamma_{i+1/2}} \frac{||x_{i+1} - x_{i+1/2}||_2}{||\boldsymbol{K}_{i+1} \boldsymbol{n}||_2}. \tag{2.6}$$

Since the flux is continuous, f is the same on each side of S. By adding together the two equations (2.5) and (2.6), we get

$$p_i - p_{i+1} = \frac{f}{\Gamma_{i+1/2}} \left(\frac{||x_{i+1/2} - x_i||_2}{||\boldsymbol{K}_i \boldsymbol{n}||_2} + \frac{||x_{i+1} - x_{i+1/2}||_2}{||\boldsymbol{K}_{i+1} \boldsymbol{n}||_2} \right). \tag{2.7}$$

A flux approximation like this is called a two point flux. To be able to derive formula (2.7), it is necessary that $x_{i+1/2}$ is connected with x_i and x_{i+1} with lines that runs along $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_{i+1}$, respectively. A grid with this property is called $\boldsymbol{K}$-orthogonal. A grid is $\boldsymbol{K}$-orthogonal if and only if the flux through all the edges can be approximated in a consistent way with two point flux. TPFA is only consistent for $\boldsymbol{K}$-orthogonal grids. So if we do not have a $\boldsymbol{K}$-orthogonal grid the assumptions for the method are no loner valid. An example of the how erroneous a TPFA might be for non-orthogonal grids can be found in for instance Aavatsmark et al. [7] or Reme [26].

$\boldsymbol{K}$-orthogonality

$\boldsymbol{K}$-orthogonal grids are important, because the discretization on such grids is simple and the resulting method is consistent. We will give a sufficient

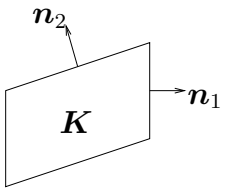


Figure 2.4: *The vectors* $\mathbf{n}_1$ *and* $\mathbf{n}_2$ *in a parallelogram cell.*

criterion for a parallelogram grid to be $\boldsymbol{K}$-orthogonal. Consider the parallelogram grid in Figure 2.4. The permeability $\boldsymbol{K}$ is constant and the unit normal for the interfaces is $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$. The grid is $\boldsymbol{K}$-orthogonal if $\boldsymbol{K}\boldsymbol{n}_i$ is parallel with interface j, $j \neq i$, consequently if

$$\boldsymbol{n}_2^T \boldsymbol{K} \boldsymbol{n}_1 = 0, \tag{2.8}$$

for all cells.

It is often not possible to achieve $\boldsymbol{K}$-orthogonal grids when dealing with reservoir simulations. Therefore it is necessary to consider another flux approximation when $\boldsymbol{K}$-orthogonality fails. For this we introduce the multipoint flux approximation.

2.2.2 Multipoint Flux Approximation (MPFA)

MPFA methods are designed to manage non-orthogonal grids, a strong feature of the method is the ability to handle media inhomogeneity and anisotropy, as well as irregular grid cells.

As the name suggests, the MPFA discretization is a control volume formulation where more than two pressure values are used in the flux approximation. Hence, the flux will be approximated by a multipoint flux approximation expression

$$f_i \approx \sum_{j \in J} t_{i,j} p_j. \tag{2.9}$$

Here the coefficient $t_{i,j}$ is the transmissibility coefficient, it is a conductance term of interface i and control volume j. We have that $\sum_{j \in J} t_{i,j} = 0$, since the flux must be zero when $\{p_j\}$ is a constant vector, $\nabla p = 0$. The quantity p_j is the pressure value at the centre of cell j. The set J depends on the grid and will be discussed later. For the two-dimensional quadrilateral grid, J consists of six cells.

MPFA methods can be arranged in different ways, and be applied on triangle- , quadrilateral-, and general polygon grids, e.g. Aavatsmark et al [6], [5]. We will now go through the MPFA O(0)-method on a quadrilateral grid. The grid in Figure 2.1 shows an example of a quadrilateral grid.

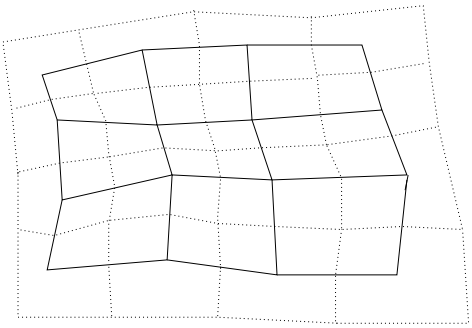


Figure 2.5: *A general quadrilateral grid with interaction regions.*

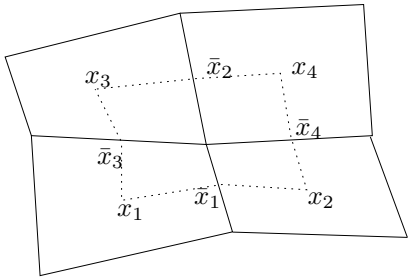


Figure 2.6: *An enlargement of the four upper left cells in Figure 2.5 with a interaction region illustrated by the dotted lines.*

Let us consider the four quadrilaterals with a common vertex as shown in Figure 2.6. For each cell we denote the cell centre x_j and the midpoint of the cell interface $\bar{x}_i$. We introduce a dual grid by drawing lines connection the cell centres with the midpoints of the cell surfaces. The cells of the dual grid is termed interaction regions. A quadrilateral grid with interaction regions is illustrated in Figure 2.5. The interaction region on the Figure 2.6 is the polygon with vertices x_1, $\bar{x}_1$, x_2, $\bar{x}_4$, x_4, $\bar{x}_2$, x_3, and $\bar{x}_3$. Each interaction region divide the cell interfaces in two parts, and each part will be termed a subinterface. One interaction region contains four subinterfaces.

The method is constructed such that the transmissibility coefficients for all the subinterfaces inside an interaction region are determined by the local interaction between the cells of the interaction region. These transmissibilities determine the flux over the subinterface. When the flux is determined for the four subinterfaces in a interaction region, we can repeat this for neighbouring interaction regions until we have determined the flux for all the subinterfaces in a grid. When the flux over two subinterfaces is known, we can add up in order to get an expression for the flux over the entire interface.

Inside an interaction region we apply the same principles as for the TPFA: continuity in flux over the sum of subinterfaces in the interaction region and continuity of pressure on the midpoint of an interface. We assume that the pressure is described by a linear function in each cell j in the interaction region. Hence it can be written

$$p(\boldsymbol{x}) = \nabla p \cdot (\boldsymbol{x} - \bar{\boldsymbol{x}}_i) + p_j, \tag{2.10}$$

where p_j is the value at the centre $\boldsymbol{x}_j$ of cell j, $p_j = p(\boldsymbol{x}_j)$. The continuity points $\bar{\boldsymbol{x}}_i$ and cell centre $\boldsymbol{x}_j$ are shown in Figure 2.7.

Each linear function is described by three coefficients, of which one of them is determined beforehand by the pressure value at the cell centre. Summing up, we have eight unknown coefficients in the linear functions. These are determined through the eight equations of the flux through the four subinterfaces, two equations for each subinterface, one for each of the two cells shearing the same subinterface. These equations are reduced to four because of continuity of flux over the subinterfaces.

The permeability in each cell j is denoted by $\boldsymbol{K}^{(j)}$. To evaluate the expression for the flux

$$f_i^{(j)} = -\boldsymbol{n_i^T} \boldsymbol{K}^{(j)} \nabla p_j, \tag{2.11}$$

through subinterface i seen from cell j one need to compute the gradient ∇p_j, and the normal vector $\boldsymbol{n}_i$ with length equal to the area of the subinterface. The gradient is determined by the value of the pressure at each of the continuity points, $\bar{\boldsymbol{x}}_i$, of the interface i of cell j .

In the two dimensional case the gradient ∇p has two components which are constant in each cell. Let $\bar{p}_i = p(\bar{\boldsymbol{x}}_i)$, $i = 1, 2$ be the pressure at the continuity points. From equation (2.10) we then have that

$$\nabla p \cdot (\bar{\boldsymbol{x}}_i - \boldsymbol{x}_j) = (\bar{p}_i - p_j), \qquad i = 1, 2. \tag{2.12}$$

This system of equations can be written

$$\boldsymbol{X} \nabla p = \begin{bmatrix} \bar{p}_1 - p_j \\ \bar{p}_2 - p_j \end{bmatrix},$$

where

$$\boldsymbol{X} = \begin{bmatrix} (\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_j)^T \\ (\bar{\boldsymbol{x}}_2 - \boldsymbol{x}_j)^T \end{bmatrix}.$$

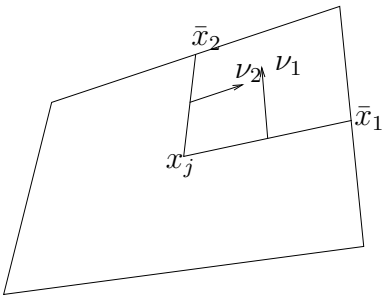


Figure 2.7: *Cell centre x_k and continuity point $\bar{x}_i$. The continuity points and the cell centre describe a triangle in a two dimensional cell*

We introduce the rotation matrix

$$\boldsymbol{R} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Then for any pair of two dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the expression $\boldsymbol{a}^T\boldsymbol{R}\boldsymbol{b}$ is the third component of the cross product between $\boldsymbol{a}$ and $\boldsymbol{b}$. It follows that $\boldsymbol{a}^T\boldsymbol{R}\boldsymbol{a} = 0$. The determinant of $\boldsymbol{X}$ is

$$F = \det\boldsymbol{X} = (\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_j)^T\boldsymbol{R}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x}_j)^T.$$

F is twice the area of the triangle spanned by the points $\boldsymbol{x}_j, \bar{\boldsymbol{x}}_1$ and $\bar{\boldsymbol{x}}_2$. To express the inverse of $\boldsymbol{X}$, we introduce the vectors $\boldsymbol{\nu}_i$, where $i = 1, 2$, given by

$$\boldsymbol{\nu}_1 = \boldsymbol{R}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x}_j), \quad \text{and} \quad \boldsymbol{\nu}_2 = -\boldsymbol{R}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_j).$$

The vector $\boldsymbol{\nu}_i$ is the inner normal vector to the triangle edge joining the points $\boldsymbol{x}_j$ and $\bar{\boldsymbol{x}}_i$, having length equal to the length of this edge, see figure (2.7). The inverse of the matrix $\boldsymbol{X}$ is given by

$$\boldsymbol{X}^{-1} = \frac{1}{F}[\boldsymbol{\nu}_1, \boldsymbol{\nu}_2].$$

It follows that

$$\nabla p = \frac{1}{F}\sum_{i=1}^{2}\boldsymbol{\nu}_i(\bar{p}_i - p_j), \tag{2.13}$$

and hence the flux through subinterface i seen from cell j is

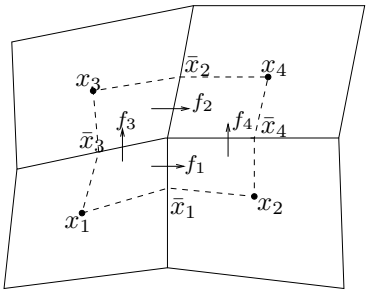
$$f_i^{(j)} = -\boldsymbol{n_i^T}\boldsymbol{K}^{(j)}\nabla p_j = \frac{1}{F}\sum_{i=1}^{2} -\boldsymbol{n_i^T}\boldsymbol{K}^{(j)}\boldsymbol{\nu}_i(\bar{p}_i - p_j). \tag{2.14}$$

The flux in the cell illustrated in Figure 2.8(b), is now expressed as

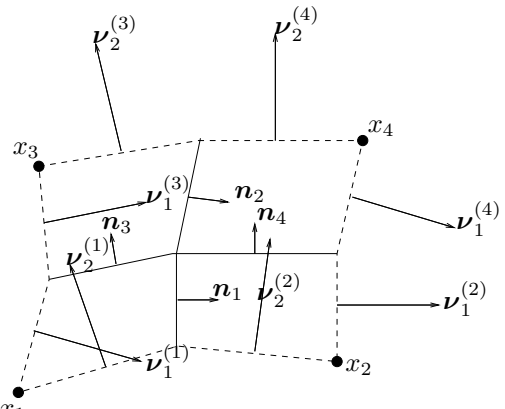
$$\begin{aligned}\begin{bmatrix} f_1^{(j)} \\ f_2^{(j)} \end{bmatrix} &= -\begin{bmatrix} \Gamma_1\boldsymbol{n}_1^T \\ \Gamma_2\boldsymbol{n}_2^T \end{bmatrix}\boldsymbol{K}^{(j)}\nabla p = -\frac{1}{F}\begin{bmatrix} \Gamma_1\boldsymbol{n}_1^T \\ \Gamma_2\boldsymbol{n}_2^T \end{bmatrix}\boldsymbol{K}^{(j)}\begin{bmatrix}\boldsymbol{\nu}_1^{(j)}\boldsymbol{\nu}_2^{(j)}\end{bmatrix}\begin{bmatrix} \bar{p}_1 - p_j \\ \bar{p}_2 - p_j \end{bmatrix} \\ &= -\boldsymbol{G}^{(j)}\begin{bmatrix} \bar{p}_1 - p_j \\ \bar{p}_2 - p_j \end{bmatrix},\end{aligned}$$

where Γ_i is the length of subinterface i. The matrix $\boldsymbol{G}^{(j)} = \{g_{i,k}^{(j)}\}_{i=1,2;k=1,2}$ is defined as

$$\boldsymbol{G}^{(j)} = \frac{1}{F}\begin{bmatrix} \Gamma_1\boldsymbol{n}_1^T \\ \Gamma_2\boldsymbol{n}_2^T \end{bmatrix}\boldsymbol{K}^{(j)}\begin{bmatrix}\boldsymbol{\nu}_1^{(j)}\boldsymbol{\nu}_2^{(j)}\end{bmatrix}.$$



(a) *A interaction region defined by the dotted lines. The fluxes through the subinterfaces is also shown.*



(b) *Normal vectors with local numbering in an interaction volume.*

Figure 2.8: *The figures illustrates a interaction region with 2.8(a) fluxes and 2.8(b) normal vectors.*

Now consider the interaction volume illustrated in Figure 2.8(b). The matrix $\boldsymbol{G}^{(j)}$ is defined through all the normal vectors in the figure, thus the flux through the subinterfaces is defined through the following eight equations

$$\begin{bmatrix} f_1^{(1)} \\ f_3^{(1)} \end{bmatrix} = \boldsymbol{G}^{(1)} \begin{bmatrix} \bar{p}_1 - p_1 \\ \bar{p}_3 - p_1 \end{bmatrix}, \quad \begin{bmatrix} f_1^{(2)} \\ f_4^{(2)} \end{bmatrix} = \boldsymbol{G}^{(2)} \begin{bmatrix} \bar{p}_1 - p_2 \\ \bar{p}_4 - p_2 \end{bmatrix}, \tag{2.15}$$

$$\begin{bmatrix} f_2^{(3)} \\ f_3^{(3)} \end{bmatrix} = \boldsymbol{G}^{(3)} \begin{bmatrix} \bar{p}_2 - p_3 \\ \bar{p}_3 - p_3 \end{bmatrix}, \quad \begin{bmatrix} f_2^{(4)} \\ f_4^{(4)} \end{bmatrix} = \boldsymbol{G}^{(4)} \begin{bmatrix} \bar{p}_2 - p_4 \\ \bar{p}_4 - p_4 \end{bmatrix}. \tag{2.16}$$

With respect to cell 1, is the directions of $\boldsymbol{\nu}_1^{(2)}$, $\boldsymbol{\nu}_2^{(3)}$, $\boldsymbol{\nu}_1^{(4)}$ and $\boldsymbol{\nu}_2^{(4)}$ reversed as visualized in Figure 2.8(b). That is why the expressions $\bar{p}_1 - p_2$, $\bar{p}_3 - p_3$, $\bar{p}_2 - p_4$ and $\bar{p}_4 - p_4$ are included in equation (2.15) and (2.16) with opposite sign.

The continuity assumptions in the flux now give

$$\begin{aligned} f_1 = f_1^{(1)} = f_1^{(2)}, \quad f_2 = f_2^{(4)} = f_2^{(3)}, \\ f_3 = f_3^{(3)} = f_3^{(1)}, \quad f_4 = f_4^{(2)} = f_4^{(4)}. \end{aligned}$$

Using expression (2.15) and (2.16) these equations become

$$\begin{aligned}
f_1 &= -g^{(1)}_{1,1}(\bar{p}_1 - p_1) - g^{(1)}_{1,2}(\bar{p}_3 - p_1) = g^{(2)}_{1,1}(\bar{p}_1 - p_2) - g^{(2)}_{1,2}(\bar{p}_4 - p_2),\\
f_2 &= g^{(4)}_{1,1}(\bar{p}_2 - p_4) + g^{(4)}_{1,2}(\bar{p}_4 - p_4) \;\; = -g^{(3)}_{1,1}(\bar{p}_2 - p_3) + g^{(3)}_{1,2}(\bar{p}_3 - p_3),\\
f_3 &= -g^{(3)}_{2,1}(\bar{p}_2 - p_3) + g^{(3)}_{2,2}(\bar{p}_3 - p_2) = -g^{(1)}_{2,1}(\bar{p}_1 - p_1) - g^{(1)}_{2,2}(\bar{p}_3 - p_1),\\
f_4 &= g^{(2)}_{2,1}(\bar{p}_1 - p_2) - g^{(2)}_{2,2}(\bar{p}_4 - p_2) \;\; = g^{(4)}_{2,1}(\bar{p}_2 - p_4) + g^{(4)}_{2,2}(\bar{p}_4 - p_4).
\end{aligned} \tag{2.17}$$

These equations contain the interface points $\bar{p}_1,\ \bar{p}_2,\ \bar{p}_3$ and $\bar{p}_4$. We have used the same expression for the interface points on each side of a surface, thus we have implicitly required continuity of the pressure in the points $\bar{x}_1,\ \bar{x}_2,\ \bar{x}_3$ and $\bar{x}_4$. The continuity in flux has left us with four equations (2.17).

We define the vectors $\tilde{\boldsymbol{f}} = [f_1,\ f_2,\ f_3,\ f_4]$, $\boldsymbol{p} = [p_1,\ p_2,\ p_3,\ p_4]^T$ and $\boldsymbol{v} = [\bar{p}_1,\ \bar{p}_2,\ \bar{p}_3,\ \bar{p}_4]$. The expressions on each side of the left equality sign in system (2.17) can now be written

$$\tilde{\boldsymbol{f}} = \boldsymbol{C}\boldsymbol{v} + \boldsymbol{F}\boldsymbol{p}. \tag{2.18}$$

The expressions on each side of the right equality sign of the system (2.17), can after a reorganization, be expressed as

$$\boldsymbol{A}\boldsymbol{v} = \boldsymbol{B}\boldsymbol{p}. \tag{2.19}$$

Now we may eliminate $\boldsymbol{v}$ by solving equation (2.19) with respect to $\boldsymbol{v}$ and by setting $\boldsymbol{v} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{p}$ in equation (2.18). The flux expression is then

$$\tilde{\boldsymbol{f}} = \boldsymbol{T}\boldsymbol{p}, \tag{2.20}$$

where

$$\boldsymbol{T} = \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} + \boldsymbol{F}. \tag{2.21}$$

Equation (2.20) gives the flux over all the subinterfaces expressed with the pressure value in the cell centre for an interaction volume. Note that we have used the same principles as in the derivation of equation (2.7): Continuity in flux and pressure.

When all the subfluxes are found, one can sum up and find the fluxes over all the interfaces, arranged in a matrix system

$$\sum_{i,j} t_{i,j} p_j = \boldsymbol{A}\boldsymbol{p} \tag{2.22}$$

where $\boldsymbol{p}$ is a discretization of the function p at the cell centre. The matrix $\boldsymbol{A}$ will then be a discretization of the elliptic operator, L_{E}, defined as $L_{\mathrm{E}}p =$

$-\nabla \boldsymbol{K} \nabla p$. Note also that the transmissibility coefficients for single phase flow only depend on the grid geometry and the permeability. Hence, the calculation of the transmissibility coefficients may be done in advance.

We will only consider homogeneous Dirichlet boundary conditions in this thesis. These boundary conditions are handled by defining a set of inactive ghost cells around the grid. In these cells the corresponding right hand side is set to zero.

The above choice of continuity conditions is known as the O(0)-method. It is possible to chose other points of continuity e.g. Edwards and Rogers [15]. Other continuity point of the pressure will yield O(η)-methods, where the parameter η decides the location of this point on the subinterface. Also other continuity conditions can be used, see for instance Aavatsmark et al. [7] or Nordbotten et al. [8]. Such methods will not be further discussed here.

So far the theoretical convergence properties of MPFA methods are not well understood [20], although some convergence proofs can be shown. Convergence proofs of the MPFA O(0)-method is given in Klausen and Winther [19] on general quadrilateral grid, and Aavatsmark et al. [9]. For rough grids Klausen and Winther [20] have shown, that when a condition, which is sufficient for L^2 convergence of pressure and flow density, is exceeded, the rate of converges is reduced and ultimately for large violations lost. Thus, unfortunately, for bad enough grids, MPFA does not converge.

2.2.3 The system matrix

Now we want to set up the system matrix $\boldsymbol{A}$ defined by the transmissibilities as shown in equation (2.22). This section and illustrations is inspired by Aadland [1]. To calculate the flux out of a single grid cell with MPFA O(0)-

5	4	3
6	1	2
7	8	9

Figure 2.9: *Local cell numbering in a grid. The figure illustrates the nine cells needed to calculate the flux out of cell* 1.

$i+N_x-1$	$i+N_x$	$i+N_x+1$
$i-1$	i	$i+1$
$i-N_x-1$	$i-N_x$	$i-N_x+1$

Figure 2.10: *Global cell numbering in a $N_x \times N_y$ grid.*

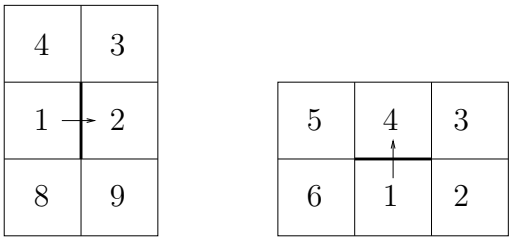


Figure 2.11: *Illustrates the local numbering of the set J. For the flux from cell 1 to cell 2 $J = \{1, 2, 3, 4, 8, 9\}$, see the figure to the left. To the right we see that the set $J = \{1, 2, 3, 4, 5, 6\}$ gives the flux from cell 1 to cell 4.*

method in 2D, a six point flux stencils and nine point cell stencils are used. This means that six cells is needed to calculate the flux over one interface in a grid cell. So, if you have a grid cell with four interfaces, nine cells is needed to calculate the flux over all four interfaces. The local cell numbering is illustrated in Figure 2.9. For a grid with dimension $N_x \times N_y$ a corresponding global cell numbering is illustrated in Figure 2.10. For the flux from cell 1 to cell 2, the six point flux stencil will consist of cell 1, 2, 3, 4, 8 and 9.

Suppose that the transmissibility matrix $\boldsymbol{T}$, defined in equation (2.21), for both interaction regions that contributes to the flux is given. The flux through an interface, see Figure 2.11, with number k is

$$f_k = \sum_{j \in J} t_{kj} p_j. \tag{2.23}$$

Here p_j is the pressure in the centre of the cells j that contribute to the flux. The set J consists of the local numbers for the six cells which contribute to the flux. This is illustrated for the flux from cell 1 to cell 2 and from cell 1 to cell 4 in Figure 2.11. As always t_{kj} is the transmissibility coefficient for interface k in control volume j.

The subinterfaces is in an interaction region is numbered from 1 to 4, starting with the vertical subinterfaces counting first the lower then the upper, then the horizontal subinterfaces is counted starting with the rightmost. This is illustrated in Figure 2.12. In the upper interaction region, the subinterface has local number 1, while in the lower interaction region, the subinterface has local number 2. The coefficients $\{t_{kj}\}_{j \in J}$ for the flux from cell 1 to cell 2 can now be expressed as

$$\begin{aligned} t_{k1} &= t_{11} + t_{21}, & t_{k2} &= t_{12} + t_{22}, \\ t_{k3} &= t_{13}, & t_{k4} &= t_{14}, \\ t_{k8} &= t_{28}, & t_{k9} &= t_{29}. \end{aligned} \tag{2.24}$$

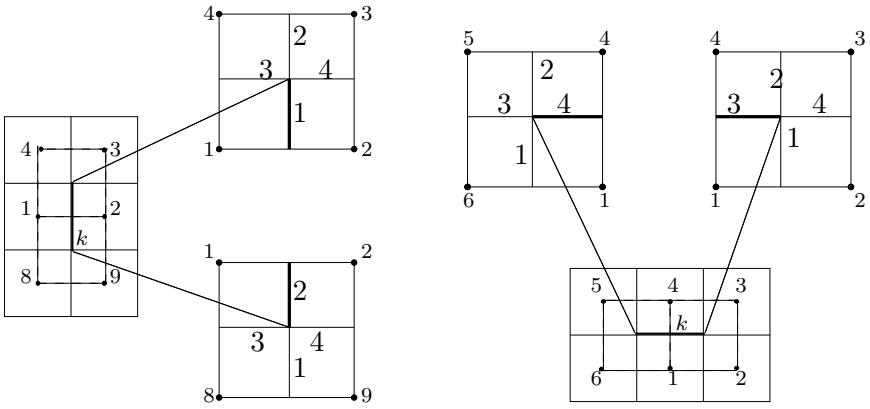


Figure 2.12: *To the left is an illustration of all the local numbers which are needed to calculate the flux over interface* k *from cell* 1 *to* 2*. On the right hand side we see the local numbers for the flux form* 1 *to* 4*.*

Here, the first index for the coefficients at the right hand side is the local number of the subinterface in question, while the second index is the local cell number. The local numbering of the subinterfaces and the cell centres is illustrated in Figure 2.12. Similarly, we get the transmissibilities for the flux from cell 1 to cell 4.

$$\begin{aligned} t_{k1} &= t_{41} + t_{31}, & t_{k2} &= t_{32}, \\ t_{k3} &= t_{33}, & t_{k4} &= t_{44} + t_{34}, \\ t_{k5} &= t_{45}, & t_{k6} &= t_{46}. \end{aligned} \tag{2.25}$$

When all the transmissibilities t_{kj} for the interfaces k is known, we use them to set up the system-matrix $\boldsymbol{A}$. The matrix is constructed such that each cell in the grid has its own column. The flux f of cell i, which is the sum of the fluxes over the edges of cell i, is given by row number i in the matrix. As illustrated in Figure 2.13, the fluxes over the lower horizontal edge and left vertical edge calculated from (2.23), (2.24), and (2.25), contribute to the influx, hence the signs must be change to find the outflux,

$$f = f_1^{(k)} + f_2^{(k)} - f_3^{(k)} - f_4^{(k)}. \tag{2.26}$$

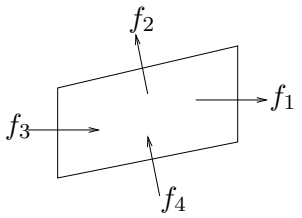


Figure 2.13: *The fluxes through the interfaces in a cell.*

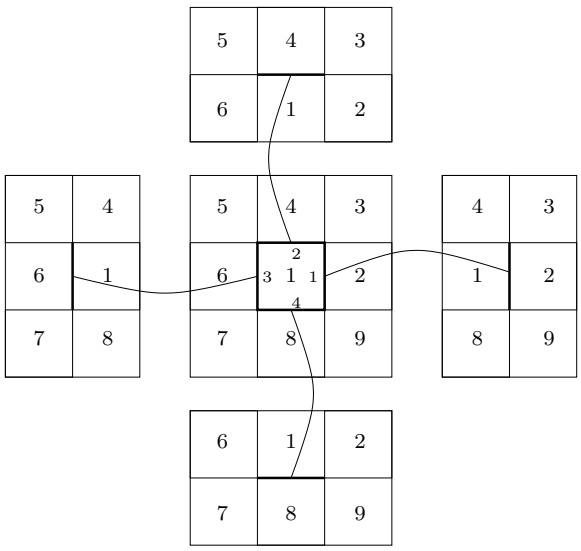


Figure 2.14: *The local numbering of the cell molecule and the flux molecule.*

Let C_i be the set of the global cell numbers that are needed to calculate the flux (2.26) for a cell. According to Figure 2.10:

$$C_i = \{i-N_x-1, i-N_x, 1-N_x+1, i-1, i, i+1, i+N_x-1, i+N_x, i+N_x+1\}.$$

If $a_{i,j}$ is element j in row i of matrix A, then we have that $a_{i,j} = 0$ for $j \notin C_i$ in row i. Now, the elements $\{a_{i,j}\}_{j \in C_i}$ are given by summing the transmissibility coefficients for cell i that are concerned with the pressure in cell C_i:

$$\begin{aligned}
a_{i,i-N_x-1} &= -t_{37} - t_{47} &&= a_{i,7},\\
a_{i,i-N_x} &= t_{18} - t_{34} + t_{48} &&= a_{i,8},\\
a_{i,i-N_x+1} &= t_{19} - t_{49} &&= a_{i,9},\\
a_{i,i-1} &= t_{26} - t_{36} - t_{46} &&= a_{i,6},\\
a_{i,i} &= t_{11} + t_{21} - t_{31} - t_{41} &&= a_{i,1},\\
a_{i,i+1} &= t_{12} + t_{22} - t_{42} &&= a_{i,2},\\
a_{i,i+N_x-1} &= t_{25} - t_{35} &&= a_{i,5},\\
a_{i,i+N_x} &= t_{14} + t_{24} - t_{34} &&= a_{i,4},\\
a_{i,i+N_x+1} &= t_{13} + t_{23} &&= a_{i,3}.
\end{aligned} \tag{2.27}$$

Here t_{mn} is the transmissibility coefficient, where n is the local number of the cells and m denotes the local interface. Notice the difference from the numbering of the subinterfaces. Now the transmissibilities from two subinterfaces

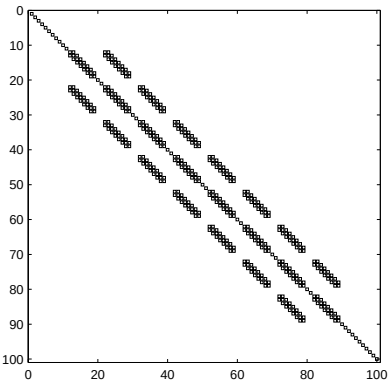


Figure 2.15: *Shows the elements in a system matrix for a* 8×8*, ghost cells are included.*

that create one interface are added and numbered independently from the previous number, see Figure 2.14. These transmissibilities are dependent on the global numbering in the grid. On the right hand side in the equations the elements $\{a_{i,j}\}_{j=1}^{9}$ are written on local form, where i is the global cell number and j is the local number for the nine point flux stencil. Figure 2.14 illustrates the local numbering in a flux molecule, the global numbering is shown in Figure 2.10.

We see from the equations (2.27) that each row i for a cell not at the boundary will have nine elements. The resulting system matrix has nine diagonals, and is illustrated in Figure 2.15. We only consider homogeneous Dirichlet boundary conditions in this thesis, so under the discretization we include ghost cells around the grid, and set the pressure values in these cells equal to zero. This expand our original $N_x \times N_y$ grid to a $(N_x+2) \times (N_y+2)$ grid, which again expand the system matrix. The rows in the system matrix that correspond to a ghost cell are set zero everywhere except on the diagonal which is set to 1. The system matrix for a 8×8 grid where ghost cells are included is illustrated in Figure 2.15.

The system matrix for a two point flux is constructed in a similar manner, but it has only five diagonals. This is because only five cells contribute to the flux, the cells which are not directly in contact with the interfaces the flux is calculated for, do not contribute to the flux. So if the local numbering of the cells are the same as for MPFA, see Figure 2.9, only cells $1, 2, 4, 6$ and 8 contribute. So in equation (2.27) $a_{i,i-N_x-1} = a_{i,i-N_x+1} = a_{i,i+N_x-1} = a_{i,i+N_x+1} = 0$.

Now that we have a discrete expression for the flux, it is appropriate to study the behaviour of the continuous solution and how these can be capture

through the discrete system. The next chapter will therefore consider an important property of the continuous solution and give a motivation for why it is important that this property is remained in the discrete solution. We will also define an analogous discrete property, to the continuous property, which we want our discrete system to hold.

Chapter 3

Maximum Principles

In this chapter we will discuss properties of the elliptic and parabolic differential equations (1.8) and (1.10), and analogous properties for the discrete equations. First we will go through this theory for elliptic equations, and introduce a similar property for the discrete system. Then we will move on to study the same principle for parabolic equations. For more detailed information on maximum principles and proofs we recommend Protter and Weinberger [25] or Evans [16]. Finally we characterise monotone matrices and derive a class of matrices that have this property.

The essential property of a discrete solution is of course whether and how it converges to the solution the continuous problem. For this to be fulfilled it is important that the discrete representation possesses the same properties as the continuous problem. One important property of the equations we study is the maximum principle. Since the control volume method already captures the essential conservation principle, we now want our discrete system to satisfy a discrete maximum principle.

A violation of the discrete maximum principle may lead to unphysical oscillations, which may have serious consequences in multiphase flow simulations. If the computed pressure lies below the bubble point pressure of the mixture, and the actual pressure lies above it, artificial gas may be liberated, yielding a strongly diverging solution. Since the flux is a function of the derivative of the pressure, circulations in the velocity field may occur for oscillating pressure, which is unphysical in a curl free velocity field.

Before we consider the maximum principles, we define two important properties in the following theory.

Definition 3.1. Monotone Matrix
Let $\boldsymbol{O}$ be the zero matrix. A matrix $\boldsymbol{A}$ with the sign property

$$\boldsymbol{A} \geq \boldsymbol{O}, \quad \text{i.e.} \quad \{a_{ij}\} \geq 0 \quad \forall \quad i,j \tag{3.1}$$

is called monotone. A monotone matrix is also referred to as a nonnegative matrix.

Consider a class of matrices defined by

$$Z = \{\boldsymbol{A} = \{a_{i,j}\} \quad | \quad a_{i,j} \leq 0 \quad \text{for} \quad i \neq j\}. \tag{3.2}$$

For matrices in Z we define the following

Definition 3.2. M-matrix
A non-singular matrix $\boldsymbol{A} \in Z$ with a nonnegative inverse, i.e. $\boldsymbol{A}^{-1} \geq \boldsymbol{O}$, is called an M-matrix.

3.1 Elliptic Equations

Here we will give a presentation of the maximum principle for elliptic equations. The theory concerning elliptic equations is included in this chapter to give a basis for comparison with the parabolic equations. We follow the presentations done in Nordbotten et al. [23] and Aavatsmark [3].

Consider the boundary value problem

$$L_{\mathrm{E}} p = q \quad \text{in} \quad U, \tag{3.3}$$
$$p = 0 \quad \text{on} \quad \partial U, \tag{3.4}$$

where U is an open, bounded subset of $\mathbb{R}^n$ with the boundary ∂U. Here p is the unknown and the source q is given. Assume that L_{E} exists and denotes a second order elliptic partial differential operator having the form

$$L_{\mathrm{E}} p = -\nabla \cdot (\boldsymbol{K} \nabla p). \tag{3.5}$$

Assume that both $\boldsymbol{K}$ and its inverse $\boldsymbol{K}^{-1}$ are bounded.

The differential equation (3.3) satisfies as mentioned an important property which is well known in physics. Equation (3.3) can for instance be used as a model for stationary heat transfer. Then p is temperature, $\boldsymbol{K}$ is heat conductivity and q represents the density of the heat source. If there are only nonnegative heat sources in U, that is if $q \geq 0$, then the temperature can have no local minima in the interior of U. A local minimum would imply a heat sink, but since $q \geq 0$ this is impossible.

The property explained above is called the *maximum principle*, and the version that excludes local minima is known as Hopf's Lemma. Hopf stated that a subsolution p cannot attain its minimum at an interior point of a connected region unless p is constant. He proved this in [17] under the condition that $\boldsymbol{K}$ is continuously differentiable.

Theorem 1. ***Strong maximum principle***
Let p satisfy the equation (3.3), where q and the elements of $\boldsymbol{K}$ are continuously differentiable. Suppose that $L_E p = q \geq 0$ in U, and that p attains its minimum over $\bar{U}$ at an interior point. The pressure p is then constant within U.

The theorem with proof may be found in Evans [16]. We are going to use a weaker form of this lemma, which we formulate as.

Property 3.1. *If $L_E p = q \geq 0$ in U, there is no point $\boldsymbol{x}_0 \in U$ such that $p(\boldsymbol{x}_0) \leq p(\boldsymbol{x})$ for all other points $\boldsymbol{x}$ in a neighbourhood of $\boldsymbol{x}_0$.*

Since Hopf's lemma can be stated for any subdomain in U, it also means that if $q \geq 0$ in U, p can have no local minima in U.

As in [23] we may apply Green's function for boundary value problems with homogeneous Dirichlet boundary conditions to derive an appropriate equivalent property. Suppose that on some domain $D \subset \Omega$ the potential p satisfies equation (3.3) with homogeneous Dirichlet boundary conditions

$$p = 0 \quad \text{on} \quad \partial D. \tag{3.6}$$

If the tensor $\boldsymbol{K}$ and the boundary ∂D are sufficiently smooth, the solution is given by

$$p(\boldsymbol{x}) = \int_D G_D(\boldsymbol{x}, \boldsymbol{\xi}) q(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{3.7}$$

where $G_D(\boldsymbol{x}, \boldsymbol{\xi})$ is Green's function for the given boundary value problem. Thus, $G_D(\boldsymbol{x}, \boldsymbol{\xi})$ is the solution of (3.3), (3.6) on D with $q(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{\xi})$. Applying Green's formula with delta functionals as source terms, it follows that Green's function is symmetric, $G_D(\boldsymbol{x}, \boldsymbol{\xi}) = G_D(\boldsymbol{\xi}, \boldsymbol{x})$, e.g. [16]. Below, we assume that $\boldsymbol{K}$ and ∂D are sufficiently smooth to make Green's function continuous at all points but $\boldsymbol{\xi}$. Then the following inequality holds

$$G_D(\boldsymbol{x}, \boldsymbol{\xi}) \geq 0 \quad \text{for} \quad \boldsymbol{x}, \boldsymbol{\xi} \in D. \tag{3.8}$$

Inequality (3.8) and its significance follow from Theorem 2 below. An immediate consequence of (3.7) and (3.8) is that

$$q \geq 0 \quad \Rightarrow \quad p \geq 0 \quad \text{in} \quad D. \tag{3.9}$$

Theorem 2. *Property 3.1 holds if and only if inequality (3.8) holds for all $D \in \Omega$, where $G_D(\boldsymbol{x}, \boldsymbol{\xi})$ is Green's function with homogeneous Dirichlet boundary conditions on ∂D.*

When a discrete maximum principle is proposed, we will make use of the monotonicity property (3.9). In addition the result in Theorem 2 will be important.

3.1.1 The Discrete Maximum Principle 1

We will now introduce a discrete version of the maximum principle, in the next section a second discrete maximum principle for parabolic equations will be introduced. A discrete maximum principle for elliptic equations is proposed in Nordbotten et al. [23]. This year a more precise definition was suggested by Keilegavlen et al. in [18].

The control volume method for the boundary value problem (3.3) and (3.4) can be stated as

$$\sum_i f_i^{(j)} = q_j,$$

where q_j is the source integrated over the cell with index j. The derivation of the flux, $f_i^{(j)}$ through interface i seen from cell j is discussed in Chapter 2. It is shown how multipoint approximation is used to derive the discrete flux expression for flow over subinterfaces,

$$\boldsymbol{f} = \sum_{i,j} t_{ij} p_j = \boldsymbol{A}\boldsymbol{p}$$

where the system matrix $\boldsymbol{A}$ is a discretization of the differential operator L_{E}.

The discretization of (3.3) with homogeneous Dirichlet boundary conditions on Ω leads to a system of equations

$$\boldsymbol{A}\boldsymbol{p} = \boldsymbol{q}, \tag{3.10}$$

where $\boldsymbol{A}$ is a $n \times n$ matrix and $\boldsymbol{p}$ and $\boldsymbol{q}$ are n-vectors. The j-th row of $\boldsymbol{A}$ is an approximation to the differential operator L_{E} integrated over the grid cell j. The j-th component of $\boldsymbol{q}$ is an approximation to q integrated over the grid cell j, and the j-th component of $\boldsymbol{p}$ is an approximation the pressure value in the cell centre x_j.

If each element of $\boldsymbol{A}^{-1}$ is nonnegative then the discrete system satisfies the same monotonicity property as the continuous system:

$$\boldsymbol{q} \geq 0 \quad \Rightarrow \quad \boldsymbol{p} \geq 0. \tag{3.11}$$

Note that the matrix $\boldsymbol{A}^{-1}$ is a discrete version of the integral operator which uses Green's function as kernel.

However, while the sign property (3.1) excludes negative solutions for nonnegative source terms and homogeneous Dirichlet boundary conditions, it does not exclude local minima. As Theorem 2 shows, a natural discrete maximum principle is achieved by requiring that the monotonicity property (3.1) holds when $\boldsymbol{A}$ is constructed for any subset of the grid points. This is the motivation for the maximum principle introduced in [14,17]:

The Discrete Maximum Principle 1. *For a given grid in D, let (3.10) be a discretization of (3.3) with homogeneous Dirichlet boundary conditions on any subgrid. Then for any subgrid bounded by a closed Jordan curve the discretization must yield a system matrix that satisfies the monotonicity property $\boldsymbol{A}^{-1}$.*

Simple closed curves that lie in a plane are usually called Jordan curves. Every Jordan curve C decomposes the plane into two disjoint open connected sets having the curve C as their common boundary, e.g. [10].

Remark 1. As pointed out in [23] "The Discrete Maximum Principle 1 relies on Theorem 2 which was derived under sufficient smoothness conditions on $\boldsymbol{K}$ and ∂D. However, the Discrete Maximum Principle 1 should be valid for any $\boldsymbol{K}$ acceptable for the discretization, which in practice means any piecewise constant $\boldsymbol{K}$."

Remark 2. For the continuous case we have that Hopf's lemma implies no local minima in the solution. An essential object of the proof is that it is possible to define closed curves inside the domain. Technical details are given in for example [16]. In the discrete case it is not possible to define a continuous boundary that possesses function values at each point since our variables only are defined at discrete points. Therefore we do not have enough information to state anything about a local minima. We only hope that the Discrete Maximum Principle 1 helps us capture some of the same qualitative behaviour as its continuous solution.

3.2 Parabolic equations

Now we move on to discuss the maximum principle for parabolic equations. For a given time in space the maximum principle for the continuous solution weakens, i.e. Hopf's lemma is no longer valid. This may require a reformulation of a discrete maximum principle.

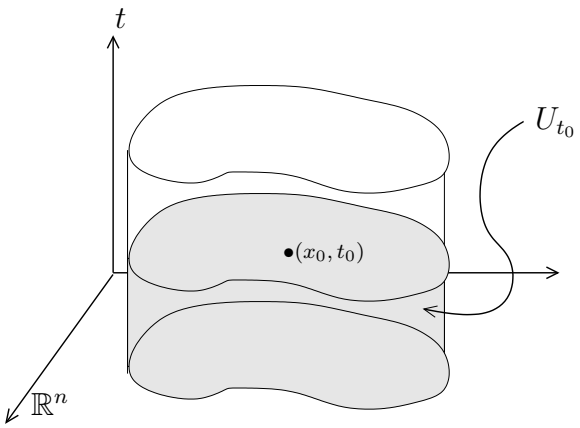


Figure 3.1: *A domain $\bar{U}_{t_N}$, the shaded region illustrates U_{t_0}.*

Throughout this section we assume U to be an open, bounded subset of $\mathbb{R}^n$ with the boundary ∂U. We denote $U_{t_N} = U \times (0, t_N]$ for some fixed time t_N. Consider the boundary-initial value problem

$$\begin{aligned} L_{\mathrm{P}} p &= q \quad \text{in} \quad U_{t_N} \\ p &= 0 \quad \text{on} \quad \partial U \times [0, t_N] \\ p &= g \quad \text{on} \quad U \times \{t = 0\}, \end{aligned} \tag{3.12}$$

where q and g are given, and $p = p(\boldsymbol{x}, t)$ is the unknown. Assume that the operator L_{P} exists, and denotes a second order parabolic partial differential operator having the form

$$L_{\mathrm{P}} p = \phi c \frac{\partial}{\partial t} p - \nabla \cdot (\boldsymbol{K} \nabla p) = \phi c \frac{\partial}{\partial t} p + L_{\mathrm{E}} p. \tag{3.13}$$

Now we state the strong maximum principle for parabolic equations, the theorem with proof may be found in [16].

Theorem 3. ***Strong maximum principle***
Let p satisfy the equation (3.12). Suppose that $q \geq 0$ in U_{t_N}. If p attains its minimum over $\bar{U}_{t_N}$ at a point $(x_0, t_0) \in U_{t_N}$, then p is constant on U_{t_0}.

So if p attains a minimum at an interior point $(x_0, t_0) \in U_{t_N}$, then p is constant at all earlier times, Figure 3.1 gives an illustration of a domain in time and space. There is no direct analogous of property 3.1 for parabolic equations, because the solution at a given time step may have local minima. Imagine a cold surface with constant temperature T_0, and place a hot ring in the middle of the domain as illustrated in Figure 3.2. As the time evolves

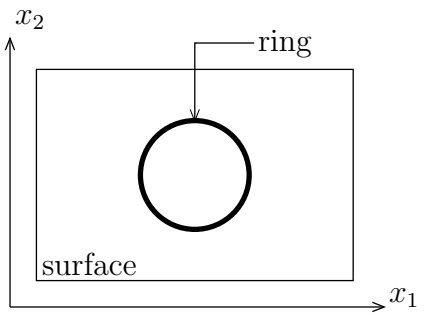


Figure 3.2: *A two dimensional surface with a hot ring placed in the middle, at time* $t = t_0$.

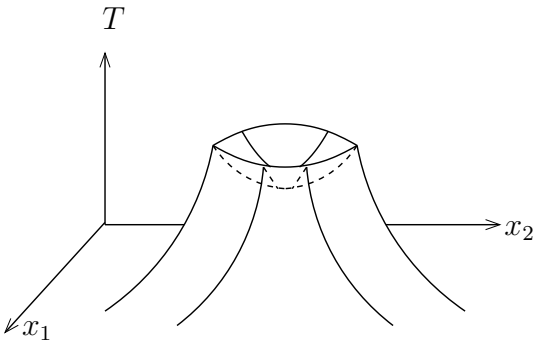


Figure 3.3: *The temperature* T *at a time* $t = t_n$. *Notice the local minima in the middle.*

the ring will transmit heat and eventually the domain will achieve a constant temperature, $T_N \geq T_0$, given no external forces acting on the boundary. If we look at the temperature distribution at time t_n, in which $t_0 \leq t_n \leq t_N$, it may look like the illustration in Figure 3.3. Hence it is a local minimum inside the domain. Notice that Theorem 3 is not violated. Although we have a local minimum in space, it is not a minimum in time and space, because in the area where we have a local minima in space the temperature increase in time.

Theorem 3 is valid for any parabolic operator L_{P} (for a general definition of L_{P} see for instance [25]). Hence the theorem also applies for the non-linear parabolic equation 1.9.

A valuable consequence of the maximum principle is the comparison principle, also known as the *monotonicity* principle, e.g. Smoller [27]. Assume that the continuity condition given above holds and consider the following conditions:

$$\begin{aligned} L_{\mathrm{P}}p - L_{\mathrm{P}}v \geq 0 \quad &\text{in} \quad U_T, \\ p = v \quad &\text{on} \quad \partial U \times [0,T], \\ p \geq v \quad &\text{on} \quad U \times \{t = 0\}. \end{aligned}$$

[2] If the three conditions above hold then $p(\boldsymbol{x},t) - v(\boldsymbol{x},t) \geq 0$ for all $(\boldsymbol{x},t) \in \bar{U}_T$. So if $v = 0$ we simply have:

Property 3.2. *Let* $L_P p \geq 0$ *with homogeneous Dirichlet boundary conditions , if* $p(\boldsymbol{x},0) \geq 0$ *then* $p(\boldsymbol{x},t) \geq 0$ *for all* $(\boldsymbol{x},t) \in \bar{U}_T$.

Hence, if the pressure is greater than zero at the start then the pressure must be greater than zero at all later times.

3.2.1 The Discrete Maximum Principle 2

Discretization in time

The time discretization for equation (3.12) can be stated as

$$\phi c \left(\frac{p^n - p^{n-1}}{\Delta t^n} \right) + L_{\mathrm{E}} p^m = q^n. \tag{3.14}$$

The superscripts n and m denotes the time level and Δt^n denotes the time step. Here q^n is the source term at time n and L_{E} is the elliptic operator defined in equation (3.5). An implicit formulation is achieved if $m = n$, with $m = n - 1$ we get an explicit formulation.

The implicit time discretization leads to the equation

$$\frac{\phi c}{\Delta t} p^n + L_{\mathrm{E}} p^n = \tilde{q}^{\,n}, \tag{3.15}$$

where

$$\tilde{q}^{\,n} = \frac{\phi c}{\Delta t} p^{n-1} + q^n.$$

This an elliptic equation and the maximum principle read,

Theorem 4. *Let p satisfy the equation $p + L_E p = q$ in U. Suppose U is connected. If $q \geq 0$ in U and p attains a* ***nonpositive*** *minimum over $\bar{U}$ at an interior point then p is constant within U.*

For theorem with proof see for instance [16].

Remark 3. Note that this theorem allows for positive minima inside U, hence p may attain local minima within U as long as they are greater than or equal to zero. Theorem 4 is consistence with our interpretation of Theorem 3. Hopf's lemma is no longer satisfied.

Discretization of the parabolic equation

The control volume formulation for the initial boundary value problem (3.12) can be stated as

$$\phi_j c_j \delta_j \left(\frac{p_j^n - p_j^{n-1}}{\Delta t^n} \right) + \sum_i f_i^{(j)m} = q_j^n, \tag{3.16}$$

where δ_j is the volume of cell j, and ϕ_j and c_j are the porosity and compressibility in cell j, respectively. The superscript n denotes the time level and Δt^n denotes the time step. Derivation of the flux, $f_i^{(j)}$, out of cell j through

interface i is done by multipoint flux approximation. We define the scalar $\alpha = \phi c\delta/\Delta t^n$, and assume that it is homogeneous, i.e. constant and equal in each cell.

As previously mentioned, an implicit formulation is achieved if $m = n$, and with $m = n - 1$ we get an explicit formulation. Each formulation leads to a system of equations for the solution vector $\boldsymbol{p}^n$

$$\boldsymbol{p}^n = \boldsymbol{B}_{\mathrm{E}}\boldsymbol{p}^{n-1} + (\alpha^{-1}\boldsymbol{q}^n) \quad (m = n-1), \tag{3.17}$$

$$\boldsymbol{B}_{\mathrm{I}}\boldsymbol{p}^n = \boldsymbol{p}^{n-1} + (\alpha^{-1}\boldsymbol{q}^n) \quad (m = n). \tag{3.18}$$

The expressions $\boldsymbol{B}_{\mathrm{E}} = (\boldsymbol{I} - \alpha^{-1}\boldsymbol{A})$ for the explicit matrix and $\boldsymbol{B}_{\mathrm{I}} = (\boldsymbol{I} + \alpha^{-1}\boldsymbol{A})$ for the implicit matrix, which imply that $\boldsymbol{B}_{\mathrm{E}} + \boldsymbol{B}_{\mathrm{I}} = 2\boldsymbol{I}$. Here $\boldsymbol{A}$ is as before a discretization of the differential operator L_{E}. With appropriate continuity and boundary conditions imposed, the solution p of a conservation law like (3.12) satisfies the monotonicity Principle 3.2 discussed above. Hence the discrete system should ideally satisfy the same principle as the continuous system, i.e.,

$$\boldsymbol{p}^{n-1} \geq 0 \quad \Rightarrow \quad \boldsymbol{p}^n \geq 0. \tag{3.19}$$

The implication above is fulfilled for the schemes (3.17) and (3.18) if and only if the matrices $\boldsymbol{B}_{\mathrm{E}}$ and $\boldsymbol{B}_{\mathrm{I}}$ satisfy the inequalities

$$\boldsymbol{B}_{\mathrm{E}} \geq \boldsymbol{O}, \tag{3.20}$$

$$\boldsymbol{B}_{\mathrm{I}}^{-1} \geq \boldsymbol{O}. \tag{3.21}$$

The first inequality, explicit case, is satisfied if and only if all the elements in the matrix $\boldsymbol{B}_{\mathrm{E}}$ are greater than or equal to zero. If this is fulfilled then $\boldsymbol{B}_{\mathrm{I}}$ is an M-matrix, since $\boldsymbol{B}_{\mathrm{E}} + \boldsymbol{B}_{\mathrm{I}} = 2\boldsymbol{I}$, hence inequality $\boldsymbol{B}_{\mathrm{I}}^{-1} \geq \boldsymbol{O}$ is satisfied. Thus, inequality (3.20) is satisfied if inequality (3.21) holds. In addition a implicit time discretization is in general more stable than explicit ones, e.g. [21]. Therefore, the implicit time discretization is the better choice.

Now, assume that the discretization of (3.12) with homogeneous Dirichlet boundary conditions on Ω leads to a system of equations

$$\boldsymbol{B}\boldsymbol{p}^n = \boldsymbol{q}^n + \alpha\boldsymbol{p}^{n-1}, \tag{3.22}$$

where $\boldsymbol{B} = \alpha\boldsymbol{I} + \boldsymbol{A}$ is a $m \times m$ matrix and $\boldsymbol{p}^n$ and $\boldsymbol{q}^n$ are m-vectors. The j-th row of $\boldsymbol{B}$ is an approximation to the differential operator L_{P} integrated over the grid cell j. The term $\alpha\boldsymbol{I}$ in $\boldsymbol{B}$ will from now be referred to as the compressibility term. The j-th component of $\boldsymbol{q}^n$ is an approximation to q integrated over the grid cell j for the given time step. The vector $\boldsymbol{p}^n$ is a discretization of the pressure function p^n for the given time step n.

When we solve the discrete system (3.22) it should be solved in a sequence. By this we mean that for the discrete system to be a solution to the continuous equation (3.12) we must carry out the numerical computation for more than one time step. If this is not done we lose the effect of the right hand side $\boldsymbol{p}^{n-1}$. Fortunately, the matrix $\boldsymbol{B}$ can be calculated beforehand because it only dependent upon predetermined variables. It is not obvious how the discrete version of the maximum principle should be formulated. Since the continuous problem satisfies Theorem 3, it may be natural to require the same of the system matrix $\boldsymbol{B}$ as we did for the system matrix in the elliptic problem.

The Discrete Maximum Principle 2. *For a given grid let (3.22) be a discretization of (3.12) with homogeneous Dirichlet boundary conditions. Then for any subgrid the matrix of coefficients* $\boldsymbol{B}$ *must satisfy the monotonicity property* $\boldsymbol{B}^{-1} \geq \boldsymbol{O}$.

Unfortunately, compressibility does not improve the monotonicity of our discrete operator, just as it weakens the maximum principle for the continuous solution at a given point in time. This is slightly counter-intuitive from a numerical analysis perspective, since compressibility usually stabilizes numerical methods. This aspect will be discussed in depth in the next chapters. One consequence of the more nuanced view on monotonicity in the presence of compressibility, is that we are motivated to consider two approaches to maximum principles:

- Require restrictions on $\boldsymbol{B}$.
- Require restrictions on both $\boldsymbol{A}$ and $\boldsymbol{B}$.

Within these options we have several choices on the matrices. We can require monotonicity on the inverse of the matrices or demand that the Discrete Maximum Principle 1 or 2 shall hold. Requiring the same on $\boldsymbol{B}$ as we did for $\boldsymbol{A}$ in section 3.1.1 may be to strong since the time discretized equation (3.14) does not satisfy Theorem 3. Again requiring only monotonicity of $\boldsymbol{B}^{-1}$ and no restrictions on subdomains, can be to weak since this may allow for unphysical oscillations. This subject and reasonable choice of discrete maximum principle will be discussed together with numerical examples in Chapter 5.

The definitions of the Discrete Maximum Principle 1 and 2 lead to our definition of a monotone method.

Definition 3.3. Monotone Method
A method which defines a discretization satisfying the Discrete Maximum Principle 1 or 2 is said to be monotone.

Remark 4. A monotone method must not be confused with a monotone matrix. A monotone matrix is defined in Definition 3.1. It is possible that a discretization leads to a monotone inverse of the system matrix even though the discretization method is not monotone.

3.3 Monotone Matrices

Monotone matrices are as we have seen of great importance in the discretization of differential equations. We now move on to characterise matrices with nonnegative inverse. It is in general difficult to give a generalized characteristic of such matrices. (Notice that the matrices that are mentioned in this section are arbitrary $n \times n$ matrices.)

We have already introduced a family of matrices with nonnegative inverses, namely the M-matrices. M-matrices in the class Z often appear under discretization of differential equation, and they have been comprehensively studied, see for instance Berman and Plemmons [13]. If TPFA is used when we discretize the elliptic equation (3.3) the resulting system matrix will always be an M-matrix.

An M-matrix is a matrix with nonpositive off-diagonal elements, which can be seen from the class Z. In addition all the diagonal elements are positive. Notice that a matrix $\boldsymbol{A} \in Z$ where all the diagonal elements are positive is not necessarily an M-matrix. If we have given an M-matrix and denote it $\boldsymbol{M}$ then we are quarantined that a matrix $\boldsymbol{B} = \boldsymbol{D} + \boldsymbol{M}$, where $\boldsymbol{D}$ is a nonnegative diagonal matrix, is also an M-matrix [13].

Now recall the system matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, in which $\boldsymbol{A}$ is the discretization of the elliptic operator L_{E}, and $\boldsymbol{B} = \alpha\boldsymbol{I} + \boldsymbol{A}$, $\alpha \geq 0$. From the statement above we see that if $\boldsymbol{A}$ is an M-matrix we are guaranteed that $\boldsymbol{B}$ is an M-matrix. Hence adding compressibility will never weaken the monotonicity of discretizations which yield M-matrices. Therefore it seems reasonable to assume that a discretization of the parabolic equation will yield at least the same monotonicity properties as the elliptic equation.

When using MPFA methods the conditions for M-matrices are in general not satisfied. Since a matrix can have nonnegative inverse without being an M-matrix, this does not rule out that implicit MPFA methods can be monotone. Matrices that are monotone without being M-matrices are derived in [23] and we will reproduce this theory. Through the properties of the splitting of the matrix, we show that it has a monotone inverse. This splitting can be applied to the system matrix, which is achieved under the discretization, and leads to local monotonicity criteria for our discrete system.

A splitting $\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{C}$ is called weakly regular if $\boldsymbol{B}$ is non-singular, $\boldsymbol{B}^{-1} \geq \boldsymbol{O}$, and $\boldsymbol{B}^{-1}\boldsymbol{C} \geq \boldsymbol{O}$, see [13]. A matrix $\boldsymbol{A}$ with a weakly regular splitting $\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{C}$ has a monotone inverse, $\boldsymbol{A}^{-1} \geq O$, if and only if the spectral radius ρ of $\boldsymbol{B}^{-1}C$ satisfies the inequality

$$\rho(\boldsymbol{B}^{-1}\boldsymbol{C}) < 1. \tag{3.23}$$

The spectral radius of a $n \times n$ matrix is the eigenvalue of the matrix with greatest absolute value. When $\boldsymbol{A}$ stems from a discretization which yields nonpositive off-diagonal elements, the matrix $\boldsymbol{A}$ is often irreducibly diagonally dominant. This property can be utilized to demonstrate the inequality (3.23). Matrices whose off-diagonal elements have different signs, cannot be expected to be diagonally dominant. However, for matrices with a weakly regular splitting, the inequality

$$\forall i : \sum_j a_{ij} \geq 0 \tag{3.24}$$

can be utilized to prove (3.23).

Theorem 5. *Suppose that the matrix $\boldsymbol{A}$ has a weakly regular splitting $\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{C}$, and suppose that inequality (3.24) holds. Then $\|\boldsymbol{B}^{-1}\boldsymbol{C}\|_\infty \leq 1$. Assume, in addition, that either the inequality (3.24)) is strict for all i or $\boldsymbol{B}^{-1}\boldsymbol{C}$ is irreducible and the inequality (3.24) is strict for at least one i. Then the inequality (3.23) holds.*

Proof. Let $\boldsymbol{e} = [1, \cdots, 1]^T$. Then inequality (3.24) can be written $\boldsymbol{A}\boldsymbol{e} \geq \boldsymbol{0}$. Thus,

$$\boldsymbol{C}\boldsymbol{e} = \boldsymbol{B}\boldsymbol{e} - \boldsymbol{A}\boldsymbol{e} \leq \boldsymbol{B}\boldsymbol{e}.$$

Hence, utilizing $\boldsymbol{B}^{-1} \geq \boldsymbol{O}$,

$$\boldsymbol{0} \leq \boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{e} \leq \boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{e} = \boldsymbol{e}.$$

Hence, it follows that $\|\boldsymbol{B}^{-1}\boldsymbol{C}\|_\infty \leq 1$.

If inequality (3.24) is strict for all i, then $\|\boldsymbol{B}^{-1}\boldsymbol{C}\|_\infty \leq 1$ and the theorem is proved. To prove the rest of the second part of the theorem, assume that $\sum_j a_{ij} > 0$ for $i = k$, i.e., $[\boldsymbol{A}\boldsymbol{e}]_k > 0$. Then

$$[\boldsymbol{C}\boldsymbol{e}]_k = [\boldsymbol{B}\boldsymbol{e}]_k - [\boldsymbol{A}\boldsymbol{e}]_k < [\boldsymbol{B}\boldsymbol{e}]_k.$$

Since $\boldsymbol{B}^{-1}$ is non-singular, the k-th column in $\boldsymbol{B}^{-1}$ must have at least one positive element. Let us assume that $[\boldsymbol{B}^{-1}]_{l,k} > 0$. Then

$$[\boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{e}]_l < [\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{e}]_l = 1.$$

The case when $\|\boldsymbol{B}^{-1}\boldsymbol{C}\|_\infty \leq 1$ for all i has already been considered, so assume that $max_i[\boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{e}]_i = 1$. With the assumption that $\boldsymbol{B}^{-1}\boldsymbol{C}$ is irreducible, we can apply the Perron-Frobenius theorem [13] on the matrix $\boldsymbol{B}^{-1}\boldsymbol{C}$. It follows that the spectral radius $\rho(\boldsymbol{B}^{-1}\boldsymbol{C})$ satisfies (3.23). This proves the theorem. □

Notice that choosing $\boldsymbol{B}$ as the diagonal part of $\boldsymbol{A}$, Theorem 5 gives sufficient conditions for $\boldsymbol{A}$ to be an M-matrix. If the transpose of a matrix $\boldsymbol{A}$ has a weakly regular splitting $\boldsymbol{A}^T = \boldsymbol{B}^T - \boldsymbol{C}^T$, then $\boldsymbol{B}^T \geq \boldsymbol{O}$ and $\boldsymbol{B}^T\boldsymbol{C}^T \geq \boldsymbol{O}$. Hence, $\boldsymbol{B}^{-1} \geq \boldsymbol{O}$ and $\boldsymbol{C}\boldsymbol{B}^{-1} \geq O$. With the inequality

$$\forall j : \sum_i a_{ij} \geq 0 \tag{3.25}$$

the theory of weakly regular splitting now yields the

Corollary 3.1. Suppose that the matrix $\boldsymbol{A}$ has a splitting $\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{C}$, such that $\boldsymbol{B}$ is non-singular, $\boldsymbol{B}^{-1} \geq \boldsymbol{O}$, and $\boldsymbol{C}\boldsymbol{B}^{-1} \geq \boldsymbol{O}$, and suppose that inequality (3.25) holds. Assume further that either the inequality (3.25) is strict for all j or $\boldsymbol{C}\boldsymbol{B}^{-1}$ is irreducible and the inequality (3.25) is strict for at least one j. Then $\boldsymbol{A}$ is non-singular with a monotone inverse, i.e., $\boldsymbol{A}^{-1} \geq \boldsymbol{O}$.

By splitting our system matrix $\boldsymbol{B}$ we may now establish, by using Corollary 3.1, under which conditions the matrix in monotone. This will be done in the next chapter.

Chapter 4

Monotonicity Criteria

In this chapter we want to determine conditions under which a control volume discretization of the parabolic equation (3.12) fulfils the Discrete Maximum Principle 2. This will give a set of local conditions on the system matrix $\boldsymbol{B}$, which is the discretization of the parabolic operator L_{P}. The paper of Nordbotten and Aavatsmark [22] was the first to considered local monotonicity conditions of MPFA-methods beyond the traditional M-matrix conditions. These conditions were derived on uniform parallelogram grids in homogeneous media. In Nordbotten et al. [23] such a set of conditions is derived on general quadrilateral grids for elliptic equations. Our presentation will follow the derivation in [23] closely, it turns out that only small adjustments are needed to extend the theory for elliptic equation to parabolic equations. We also show that these criteria are necessary. Then we will derive these conditions analytically for a homogeneous medium and grid uniform. As pointed out in Keilegavlen [18] "It follows from this that nine-point control volume methods cannot be constructed that satisfies a discrete maximum principle for all media and quadrilateral grids".

4.1 General Quadrilateral Grids

We are only studying discretization on general quadrilateral grids. To make the notation more readable we now denote the global numbering of the cells by the indices (i, j), where i is the column number and j is the row number in a $N_x \times N_y$ grid. The global cell numbering in a stencil is shown in Figure 4.2. The elements in the stencil of cell (i, j) are denoted by $m_k^{i,j}$, $k = 1, \cdots, 9$, where k is the local index of Figure 4.1.

The discrete system of equation (3.12) is

$$\boldsymbol{B}\boldsymbol{p}^n = \boldsymbol{q}^n + \alpha \boldsymbol{p}^{n-1}. \tag{4.1}$$

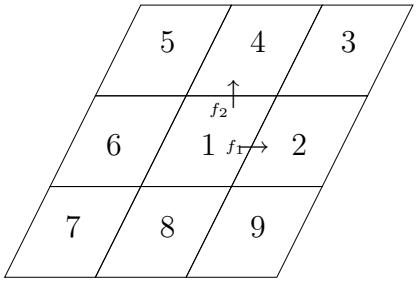


Figure 4.1: *Local cell numbering in a nine point stencil.*

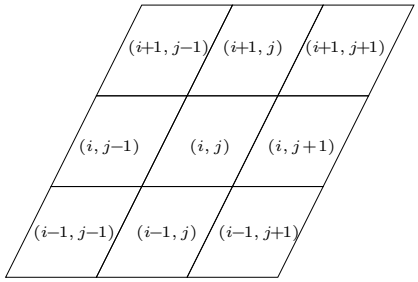


Figure 4.2: *Global cell numbering in a in a nine point stencil.*

Now, if the global numbering of the system matrix in Chapter 2 is to be compared with the matrix $\boldsymbol{B} = \{b_{m,l}\}$ then row m in the system matrix is given by the global indices (i, j) as $m = N_y(i-1) + j$. In addition we now only use the local number k on the cells which contributes to the flux through a single cell, see Figure 4.1, whereas in Chapter 2 we mostly employed global numbering.

The cell stencil approximates the integral of the expression over the cell (i, j), i.e., the outflux out of cell (i, j),

$$-\int_{\Omega(i,j)} \nabla \cdot (\boldsymbol{K}\nabla p)\, d\Omega \approx \sum_{k=1}^{9} m_k^{i,j} p_k. \tag{4.2}$$

The accumulation in cell (i, j) is

$$\int_{\Omega(i,j)} \phi c \frac{\partial}{\partial t} p \, d\Omega \approx \alpha(p_1^n - p_1^{n-1}), \tag{4.3}$$

since it is only the pressure in the present cell that contributes to the accumulation. Here is $\alpha = \frac{\phi_{i,j} c_{i,j} \Omega_{i,j}}{\Delta t}$, and for simplicity we set $p_1^{n-1} = 0$. Each equation of the system (4.1) has the form

$$\alpha p_1 + \sum_{k=1}^{9} m_k^{i,j} p_k = \int_{\Omega(i,j)} q \, d\Omega. \tag{4.4}$$

If p_k are constant, there should be no flow, therefore, for each cell (i, j) not at the boundary we have

$$\sum_{k=1}^{9} m_k^{i,j} = 0.$$

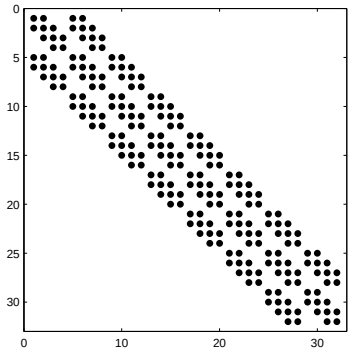


Figure 4.3: *Shows the block tridiagonal structure in the system matrix* $\boldsymbol{B}$ *for a* 4×8 *grid.*

For the system matrix $\boldsymbol{B} = \boldsymbol{A} + \alpha \boldsymbol{I}$ we have that inequality (3.24) is strict for all inner cells. Because

$$\forall m: \qquad \sum_l b_{m,l} = \alpha + \sum_{k=1}^{9} m_k^{i,j} > 0. \tag{4.5}$$

If we assume that inequality (3.24) is fulfilled for the boundary cells, we may use the theory of weakly regular splitting, which was derived in Section 3.3, to establish conditions which ensure monotonicity of $\boldsymbol{B}^{-1}$.

For any splitting $\boldsymbol{B} = \boldsymbol{D} - \boldsymbol{C}$, we may now establish monotonicity of $\boldsymbol{B}^{-1}$ in two ways: either by determining the conditions under which $\boldsymbol{D}^{-1} \geq \boldsymbol{0}$ and $\boldsymbol{C}\boldsymbol{D}^{-1} \geq \boldsymbol{0}$, or by determining the conditions under which $\boldsymbol{D}^{-1} \geq \boldsymbol{0}$ and $\boldsymbol{D}^{-1}\boldsymbol{C} \geq \boldsymbol{0}$. In the following, we will apply the first approach.

We choose the natural ordering of the unknowns, first counting the unknowns in column i of row 1, and then proceeding by counting the unknowns of each subsequent row j. In a grid with m columns and l rows, the matrix $\boldsymbol{B}$ has a block-tridiagonal structure, each block being an $m \times m$ tridiagonal matrix. $\boldsymbol{B}$ has n diagonal blocks and $l - 1$ blocks in the upper and lower block diagonal. Figure 4.3 illustrate this for a 4×8 grid.

Let $\boldsymbol{D}$ consist of the diagonal blocks of $\boldsymbol{B}$, and let $\boldsymbol{C} = \boldsymbol{D} - \boldsymbol{B}$. The diagonal blocks of $\boldsymbol{D}$ will be denoted $\boldsymbol{D}_j$, $j = 1, \cdots, l$. The blocks of the first lower block diagonal of $\boldsymbol{C}$ will be denoted $\boldsymbol{C}_j^L$, $j = 2, \cdots, l$, while the blocks of the first upper block diagonal of $\boldsymbol{C}$ will be denoted $\boldsymbol{C}_j^U$, $j = 1, \cdots, l-1$. The matrices $\boldsymbol{D}_j$, $\boldsymbol{C}_j^L$ and $\boldsymbol{C}_j^U$ are tridiagonal. These matrices are shown by displaying the nonzero elements around the i th diagonal element:

$$\boldsymbol{D}_j = \begin{pmatrix} \ddots & m_2^{i-1,j} & \\ m_6^{i,j} & m_1^{i,j}+\alpha & m_2^{i,j} \\ & m_6^{i+1,j} & \ddots \end{pmatrix},$$

$$\boldsymbol{C}_j^L = -\begin{pmatrix} \ddots & m_9^{i-1,j} & \\ m_7^{i,j} & m_8^{i,j} & m_9^{i,j} \\ & m_7^{i+1,j} & \ddots \end{pmatrix},$$

$$\boldsymbol{C}_j^U = -\begin{pmatrix} \ddots & m_3^{i-1,j} & \\ m_5^{i,j} & m_4^{i,j} & m_3^{i,j} \\ & m_5^{i+1,j} & \ddots \end{pmatrix}.$$

Thus, $\boldsymbol{D}_j \geq \mathbf{0}$ if

A:	$m_1^{i,j} + \alpha > 0,$
B1:	$m_2^{i,j} < 0,$
B3:	$m_6^{i,j} < 0,$
C:	$m_1^{i,j} + m_2^{i,j} + m_6^{i,j} + \alpha > 0.$

In fact these conditions ensure that $\boldsymbol{D}_j$, and thereby $\boldsymbol{D}$, are M-matrices.

To derive the conditions which also guarantee that $\boldsymbol{C}\boldsymbol{D}^{-1} \geq \mathbf{0}$, we introduce the matrices

$$\boldsymbol{F}_j = \boldsymbol{D}_j^{-1}, \qquad \boldsymbol{E}_{j+1}^L = \boldsymbol{C}_{j+1}^L \boldsymbol{F}_j, \qquad \boldsymbol{E}_{j-1}^U = \boldsymbol{C}_{j-1}^U \boldsymbol{F}_j. \tag{4.6}$$

The matrices $\boldsymbol{E}_j^L$ and $\boldsymbol{E}_j^U$ are nonzero blocks of $\boldsymbol{C}\boldsymbol{D}^{-1}$. Thus we have to derive conditions which ensure that $\boldsymbol{E}_j^L \geq \boldsymbol{O}$ and $\boldsymbol{E}_j^U \geq \mathbf{0}$. Let $\boldsymbol{F}_j = \{f_{i,k}^j\}$, $\boldsymbol{E}_j^L = \{e_{i,k}^{j,L}\}$ and $\boldsymbol{E}_j^U = \{e_{i,k}^{j,U}\}$. From the equation $\boldsymbol{D}_j\boldsymbol{F}_j = \boldsymbol{D}_j\boldsymbol{D}_j^{-1} = \boldsymbol{I}$. i.e. from

$$m_6^{i,j} b_{i-1,k}^j + (m_1^{i,j} + \alpha) f_{i,k}^j + m_2^{i,j} f_{i+1,k}^j = \delta_{i,k}, \tag{4.7}$$

it follows that

$$f_{i,k}^j = \frac{\delta_{i,k}}{(m_1^{i,j}+\alpha)} - \frac{m_6^{i,j}}{(m_1^{i,j}+\alpha)} f_{i-1,k}^j - \frac{m_2^{i,j}}{(m_1^{i,j}+\alpha)} f_{i+1,k}^j. \tag{4.8}$$

The definition of $\boldsymbol{E}_{j+1}^{L}$ and equation (4.8) yield

$$\begin{aligned} e_{i,k}^{j+1,L} &= -m_7^{i,j+1} f_{i-1,k}^j + m_8^{i,j+1} f_{i,k}^j - m_9^{i,j+1} f_{i+1,k}^j \\ &= \frac{m_8^{i,j+1}}{(m_1^{i,j}+\alpha)}\delta_{i,k} + \left(\frac{m_6^{i,j}}{(m_1^{i,j}+\alpha)} m_8^{i,j+1} - m_7^{i,j+1}\right) f_{i-1,k}^j \\ &+ \left(\frac{m_2^{i,j}}{(m_1^{i,j}+\alpha)} m_8^{i,j+1} - m_9^{i,j+1}\right) f_{i+1,k}^j. \end{aligned} \tag{4.9}$$

Likewise. the definition of $\boldsymbol{E}_{j-1}^{U}$ and equation (4.8) yield

$$\begin{aligned} e_{i,k}^{j-1,U} &= -m_5^{i,j-1} f_{i-1,k}^j + m_4^{i,j-1} d_{i,k}^j - m_3^{i,j-1} f_{i+1,k}^j \\ &= \frac{m_4^{i,j-1}}{(m_1^{i,j}+\alpha)}\delta_{i,k} + \left(\frac{m_6^{i,j}}{(m_1^{i,j}+\alpha)} m_4^{i,j-1} - m_5^{i,j-1}\right) f_{i-1,k}^j \\ &+ \left(\frac{m_2^{i,j}}{(m_1^{i,j}+\alpha)} m_4^{i,j-1} - m_3^{i,j-1}\right) f_{i+1,k}^j. \end{aligned} \tag{4.10}$$

since $d_{i,k}^j \geq 0$, it follows that $e_{i,k}^{j,L} \geq 0$ and $e_{i,k}^{j,U} \geq 0$ if all the terms in the expressions (4.9) and (4.10) are non-negative. Hence, $\boldsymbol{C}\boldsymbol{D}^{-1} \geq \boldsymbol{0}$ if

$$\begin{aligned} &\text{B2:} && m_4^{i,j} < 0, \\ &\text{B4:} && m_8^{i,j} < 0, \\ &\text{D1:} && m_2^{i,j} m_4^{i,j-1} - m_3^{i,j-1}(m_1^{i,j}+\alpha) > 0, \\ &\text{D2:} && m_6^{i,j} m_4^{i,j-1} - m_5^{i,j-1}(m_1^{i,j}+\alpha) > 0, \\ &\text{D3:} && m_2^{i,j} m_8^{i,j+1} - m_9^{i,j+1}(m_1^{i,j}+\alpha) > 0, \\ &\text{D4:} && m_6^{i,j} m_8^{i,j+1} - m_7^{i,j+1}(m_1^{i,j}+\alpha) > 0. \end{aligned}$$

Lemma 4.1. *The inverse of the matrix $\boldsymbol{B} = c\boldsymbol{I} - \boldsymbol{A}$ arising from a locally conservative nine-point discretization in 2D is monotone if conditions A through D defined above holds for all pairs(i, j).*

To sum up we have the monotonicity criteria for the compressible case:

$$
\begin{aligned}
&\text{A:} && m_1^{i,j} + \alpha > 0,\\
&\text{B1:} && m_2^{i,j} < 0,\\
&\text{B2:} && m_4^{i,j} < 0,\\
&\text{B3:} && m_6^{i,j} < 0,\\
&\text{B4:} && m_8^{i,j} < 0,\\
&\text{C:} && m_1^{i,j} + m_2^{i,j} + m_6^{i,j} + \alpha > 0,\\
&\text{D1:} && m_2^{i,j} m_4^{i,j-1} - m_3^{i,j-1}(m_1^{i,j} + \alpha) > 0,\\
&\text{D2:} && m_6^{i,j} m_4^{i,j-1} - m_5^{i,j-1}(m_1^{i,j} + \alpha) > 0,\\
&\text{D3:} && m_2^{i,j} m_8^{i,j+1} - m_9^{i,j+1}(m_1^{i,j} + \alpha) > 0,\\
&\text{D4:} && m_6^{i,j} m_8^{i,j+1} - m_7^{i,j+1}(m_1^{i,j} + \alpha) > 0.
\end{aligned}
$$

For the incompressible case $\alpha = 0$. If we compare the monotonicity criteria for a compressible and a incompressible medium we notice:
Criterion A: Since $\alpha > 0$ we have that the compressibility will improve the monotonicity region.
Criteria B: The criteria B are the same for both cases, hence compressibility does not have any affect in this case.
Criterion C: The same statement as for A holds.
Criteria D: Take for instance criterion D1,

$$
m_3^{i,j-1} < \frac{m_2^{i,j} m_4^{i,j-1}}{(m_1^{i,j} + \alpha)} < \frac{m_2^{i,j} m_4^{i,j-1}}{m_1^{i,j}}. \tag{4.11}
$$

Here we see that the compressibility will have an negative effect on the monotonicity since $m_2^{i,j} m_4^{i,j-1}$ must be positive and $m_1^{i,j} + \alpha > m_1^{i,j}$.

We also note that the Criteria D allow for positive contribution from the cells 3, 5, 7, and 9. Thus the conditions A through D are less restrictive than the classical M-matrix conditions.

Remark 5. If the compressibility term α goes towards infinity we see form equation (4.11) that $m_3^{i,j-1}$ approach zero. Hence the conditions D approach the M-matrix conditions when $\alpha \to \infty$. Clearly the conditions will approach the elliptic conditions when $\alpha \to 0$ the .

4.2 Most Sufficient Criteria are Necessary

Keilegavlen, Nordbotten and Aavatsmark have in [18] showed that most of the sufficient criteria for monotone control volume derived in [23] are also

necessary. We will expand this, and show that it also applies for parabolic conservation laws. This is straight forward and we will follow [18] for this.

We will go systematically through criteria A, B and D and show subdomains for which these criteria are necessary criteria for monotonicity.

Criterion A

To show that Criterion A is necessary, consider a grid containing a single cell. Then the local matrix for this problem is

$$(m_1^{i,j} + \alpha)p^{i,j} = q^{i,j} \qquad \forall \qquad (i,j). \tag{4.12}$$

The maximum principle states that $q^{i,j} > 0 \Rightarrow p^{i,j} > 0$. Suppose that $m_1^{i,j} + \alpha < 0$, if $q^{i,j} > 0$ then $p^{i,j} < 0$ if equation (4.12) should hold, which is a contradiction to the maximum principle. It follows that Criterion A is necessary for a discrete maximum principle.

Criterion B

The compressibility does not have any effect on these criteria, hence the proof is the same as in [18].

Criterion D

To show the necessity of Criterion D3, consider a domain consisting of three cells shaped like a capital L. We refer to these cells as cell $(i,j), (i+1,j)$ and $(i,j+1)$. Let $q^{i,j} = 0, q^{i,j+1} = 0$, while $q^{i+1,j}$ be positive, with a magnitude such that $p^{i+1,j} = 1$. Then

$$\begin{bmatrix} m_1^{i,j} + \alpha & m_4^{i,j} \\ m_8^{i,j+1} & m_1^{i,j+1} + \alpha \end{bmatrix} \begin{bmatrix} p^{i,j} \\ p^{i,j+1} \end{bmatrix} = - \begin{bmatrix} m_2^{i,j} \\ m_9^{i,j+1} \end{bmatrix}.$$

We invert the left hand side to obtain

$$\begin{bmatrix} p^{i,j} \\ p^{i,j+1} \end{bmatrix} = -\frac{1}{D} \begin{bmatrix} m_1^{i,j+1} + \alpha & -m_4^{i,j} \\ -m_8^{i,j+1} & m_1^{i,j} + \alpha \end{bmatrix} \begin{bmatrix} m_2^{i,j} \\ m_9^{i,j+1} \end{bmatrix}.$$

where $D = (m_1^{i,j} + \alpha)(m_1^{i,j+j} + \alpha) - m_4^{i,j} m_8^{i,j+1}$. The positivity of the determinant D is immediate from criteria A and B. To ensure that $p^{i,j+1}$ is positive, Criterion D3 is necessary.

Criteria D1, D2, and D4 are necessary by reflections of the domain considered above.

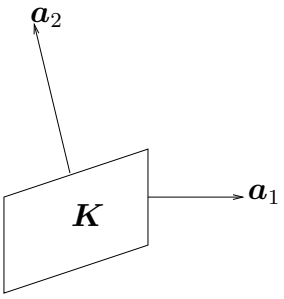


Figure 4.4: *The vectors* $\boldsymbol{a}_1$ *and* $\boldsymbol{a}_2$.

4.3 Homogeneous Media and Uniform Grid

In this section, we consider the case where the medium is homogeneous and the grid is uniform. In this special case, the conditions simplify. If the medium is homogeneous we have that the permeability is constant throughout the medium. Homogeneity combined with a uniform grid, gives that $m_k^{i,j} = m_k$, independent of i and j. Due to symmetry we also have, $m_k = m_{k+4}$ for $k = 2, \cdots, 5$. In this case, the conditions A through D can be stated as

$$\begin{aligned}
&\text{E1:} && m_1^{i,j} + \alpha > 0 \\
&\text{E2:} && \max\{m_2, m_4\} < 0 \\
&\text{E3:} && m_1 + \alpha + 2\max\{m_2, m_4\} > 0 \\
&\text{E4:} && m_2 m_4 - \max\{m_3, m_5\} \cdot (m_1 + \alpha) > 0
\end{aligned}$$

On a uniform grid where the medium is homogeneous it is possible to find the system matrix analytical. Consider the fluxes across the interfaces in Figure 4.1 which separate cell 1 form cell 2 and cell 1 from cell 4, respectively. Fluxes can be written as the weighted sum of potentials:

$$f_1 = \sum_{k=1,2,3,4,8,9} t_{1,k} p_k \qquad f_2 = \sum_{k=1}^{6} t_{2,k} p_k \tag{4.13}$$

When we look at the special case of uniform parallelogram grids on homogeneous media, the local grid symmetry implies that

$$\begin{aligned}
&t_{1,1} = -t_{1,2}, && t_{1,3} = -t_{1,8}, && t_{1,4} = -t_{1,9}, \\
&t_{2,1} = -t_{2,4}, && t_{2,2} = -t_{2,5}, && t_{2,3} = -t_{2,6}.
\end{aligned}$$

Equation (4.13) can now be written as

$$f_1 = t_{1,1}(p_1 - p_2) + t_{1,3}(p_3 - p_8) + t_{1,4}(p_4 - p_9) \tag{4.14}$$

$$f_2 = t_{2,1}(p_1 - p_4) + t_{2,2}(p_2 - p_5) + t_{2,3}(p_3 - p_6) \tag{4.15}$$

For any grid cell, let $\boldsymbol{a}_i$, $i = 1, 2$, be the normal vector of edge i, having length equal to the length of the edge, see Figure 4.4. Further, let F be the area of the parallelogram cell, and define the quantities a, b and c by

$$\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \frac{1}{F}[\boldsymbol{a}_1 \quad \boldsymbol{a}_2]^T \boldsymbol{K} [\boldsymbol{a}_1 \quad \boldsymbol{a}_2]$$

These quantities cannot attain arbitrary values since the positive definiteness of $\boldsymbol{K}$ implies that $a > 0,\ b > 0$, and

$$|c| < \sqrt{ab} \tag{4.16}$$

Notice that if $c = 0$ the criteria for M-matrices are fulfilled. In case of a linear potential field, the fluxes can be expressed by

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} p_1 - p_2 \\ p_1 - p_4 \end{bmatrix} = \begin{bmatrix} a(p_1 - p_2) + c(p_1 - p_4) \\ c(p_1 - p_2) + b(p_1 - p_4) \end{bmatrix}. \tag{4.17}$$

We may apply linear potential fields in two independent directions to determine the transmissibility coefficient in equation (4.14) and (4.15). To determine the coefficient of (4.15) we choose the pressure field such that either $f_2 = 0$ or $\nabla p \parallel \boldsymbol{a}_2$. When $f_2 = 0$ equation (4.17) yields

$$p_1 - p_2 = \frac{b}{c}(p_1 - p_4). \tag{4.18}$$

From Figure 4.1 it follows that for linear potential fields,

$$p_2 - p_5 = \ p_1 - p_4 - 2(p_1 - p_2) \ = (2\frac{b}{c} + 1)(p_1 - p_4), \tag{4.19}$$

$$p_3 - p_6 = -(p_1 - p_4) - 2(p_1 - p_2) = (2\frac{b}{c} - 1)(p_1 - p_4). \tag{4.20}$$

Applying the expressions (4.18), (4.19), and (4.20) in equation (4.17), it follows that

$$f_2 = \left[t_{21} + \left(2\frac{b}{c} + 1 \right) t_{22} + \left(2\frac{b}{c} - 1 \right) t_{23} \right] (p_1 - p_4) = 0. \tag{4.21}$$

In the case when $\nabla p \parallel \boldsymbol{a}_2$,

$$p_1 - p_4 = p_2 - p_5 = -(p_3 - p_6),$$

and thus from (4.15),

$$f_2 = [t_{21} + t_{22} - t_{23}](p_1 - p_4) = b(p_1 - p_4). \tag{4.22}$$

The equations (4.21) and (4.22) yield the pair of equations

$$\begin{aligned} ct_{21} + (2b + c)t_{22} + (2b - c)t_{23} &= 0, \\ t_{21} + t_{22} - t_{23} &= b. \end{aligned} \tag{4.23}$$

Similarly, we may determine the transmissibility coefficients of (4.14) by choosing the pressure field such that either $f_1 = 0$ or $\nabla p \parallel \boldsymbol{a}_1$. These cases yield the following pair of equations,

$$\begin{aligned} ct_{11} + (2a - c)t_{13} + (2a + c)t_{14} &= 0, \\ t_{11} - t_{13} + t_{14} &= a. \end{aligned} \tag{4.24}$$

The pair of equations (4.23) determines the coefficients t_{21}, t_{22}, and t_{23} up to one undetermined parameter β. Likewise, the pair of equations (4.24) determines the coefficients t_{11}, t_{13}, and t_{14} up to one undetermined parameter α. The solutions read

$$\begin{aligned} t_{1,1} &= a - \alpha, & t_{2,1} &= b - \beta, \\ t_{1,3} &= -\frac{c}{4} - \frac{\alpha}{2}, & t_{2,2} &= -\frac{c}{4} + \frac{\beta}{2}, \\ t_{1,4} &= -\frac{c}{4} + \frac{\alpha}{2}, & t_{2,3} &= -\frac{c}{4} - \frac{\beta}{2}. \end{aligned} \tag{4.25}$$

These transmissibilities can be combined into a nine-point stencil given by

$$\sum_{k=1}^{9} m_k^{i,j} p_k = f_1 + f_2 - f_3 - f_4, \tag{4.26}$$

where the fluxes f_1, f_2, f_3 and f_4 are shown in Figure . Introducing the parameter $\gamma = \alpha + \beta$, the element of the nine point stencil read

$$\begin{aligned} m_2 = m_6 &= t_{1,2} + t_{2,2} - t_{4,2} = -t_{1,1} + t_{2,2} - t_{2,3} = -a + \gamma \\ m_3 = m_7 &= t_{1,3} + t_{2,3} = t_{1,3} + t_{2,3} = -\frac{c}{2} - \frac{\gamma}{2} \\ m_4 = m_8 &= t_{1,4} + t_{2,4} - t_{3,4} = +t_{1,4} - t_{1,3} - t_{2,1} = -b + \gamma \\ m_5 = m_9 &= t_{2,5} + t_{3,4} = -t_{2,2} - t_{1,4} = \frac{c}{2} - \frac{\gamma}{2} \\ m_1 &= -\sum_{k=2}^{9} m_k = 2(t_{1,1} + t_{2,1}) = 2(a + b - \gamma) \end{aligned} \tag{4.27}$$

The parameter γ defines all possible conservative nine-point discretization on uniform parallelogram grids in homogeneous media, where the discretization method has an explicit flux representation which is exact for linear

potential fields. All MPFA-methods have this property. To determine the monotonicity region for a given method we need to calculate the expression for γ for the given method.

From the sign property $m_k \leq 0$ for $k = 2, \cdots, 9$ we have that

$$|c| \leq \gamma \leq \min\{a, b\}.$$

Further the inequality imply that

$$|c| \leq \min\{a, b\}. \tag{4.28}$$

Therefore the coefficients yield an M-matrix if and only if the inequality (4.28) is fulfilled. The following lemma applies.

Lemma 4.2. *For control-volume methods with local flux approximations which yield exact solutions of linear potential fields, it is impossible to define a nine-point scheme resulting in an M-matrix, on grids violating inequality (4.28).*

The conditions E yield a wider class of methods than those resulting in an M-matrix. In the next chapter we will investigate implications of the conditions E on the MPFA O(0)-method.

Chapter 5

Results and Discussion

The intention with introducing monotonicity conditions for parabolic conservation laws was that it would improve the conditions compared with elliptic laws. This seemed to be a natural hope since compressibility usually stabilizes numerical methods, and adding compressibility to M-matrices do not weaken the monotonicity region. The results in this chapter shows that this is not the fact when the Discrete Maximum Principle 2 defines the monotonicity region for the discretized parabolic equation.

5.1 Analytical Results

We will now investigate the implications of the conditions E, which are defined by (4.27), on the MPFA O(0)-method derived in Chapter 2. We will compare the results for the elliptic and parabolic case, which will show us that the compressibility term does not improve the monotonicity region.

Conditions E1-E4 inserted the coefficients (4.27) are

$$\begin{aligned}
&\text{E1:} && \gamma < a + b + \frac{\alpha}{2} \\
&\text{E2:} && \gamma < \min\{a, b\} \\
&\text{E3:} && 2(a + b - \gamma) + \alpha + 2\max\{-a + \gamma, -b + \gamma\} > 0 \\
&\text{E4:} && (\gamma - a)(\gamma - b) - (\gamma - |c|)(\gamma - (a + b) - \frac{\alpha}{2}) > 0
\end{aligned}$$

For the MPFA O(0) method, γ is given by

$$\gamma = \frac{c^2(a + b)}{2ab}. \tag{5.1}$$

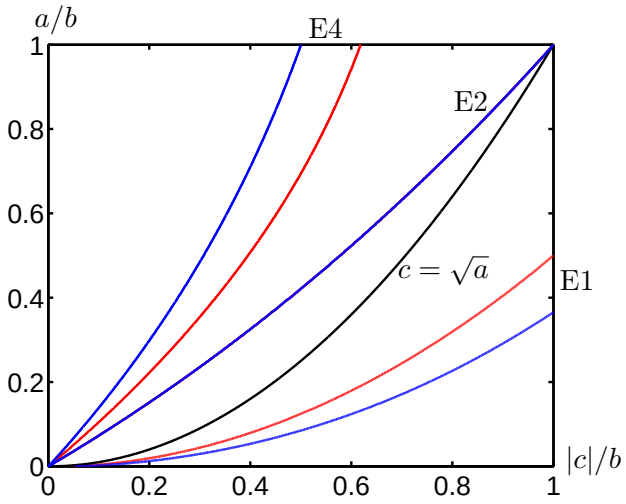


Figure 5.1: *Analytically computed monotonicity regions of the conditions E. The red curves are the monotonicity conditions for* $\alpha = 0$ *and the blue are for* $\alpha = 1$*. Condition E2 is the same for both cases. The black curve is the elliptic bound* $c = \sqrt{a}$*. The monotonicity region are above the curves in question.*

The derivation of γ for the MPFA O(0)-method may be found in Nordbotten et al. [23].

The condition E1 is satisfied given condition E2, and E3 is obvious since $a, b, \alpha \geq 0$. The conditions E2 and E4 restrict the parameter γ to

$$\frac{|c|(a + b + \frac{\alpha}{2}) - ab}{\frac{\alpha}{2} + |c|} < \gamma < \min\{a, b\}. \tag{5.2}$$

Look at the limit of large α, then the left-hand term in equation (5.2) will be equal to $|c|$. If $\alpha \to 0$ the same will apply. Therefore, inequality (5.2) imply that

$$|c| \leq \min\{a, b\}. \tag{5.3}$$

Since the conditions E is necessary there is no way to construct a nine-point control volume method that is exact for linear pressure field, which is monotone for the values a, b and c between E2 and the elliptic bound, plotted in Figure 5.1. Therefore we may expand Lemma 4.2 to monotone methods, [18],

Lemma 5.1. *For control-volume methods with local flux approximations which yield exact solutions of linear potential fields, it is impossible to define a nine-point scheme resulting in an monotone method, on grids violating inequality (4.28).*

Since γ is defined by equation (5.1), the monotonicity region for the O(0) method is

$$\frac{|c|(a+b+\frac{\alpha}{2})-ab}{\frac{\alpha}{2}+|c|} < \frac{c^2(a+b)}{2ab}. \tag{5.4}$$

Notice that the monotonicity criteria E for a homogeneous media on uniform grid are necessary for the discrete maximum principle to be fulfilled, which again is sufficient to assure monotonicity. Even though the criteria are necessary for the discrete maximum principle to be fulfilled they are not necessary for the inverse of the system matrix to be monotone.

The monotonicity region can be defined by criteria E4, as stated in equation (5.4). An investigation of the monotonicity area for all four conditions verifies this. Figure 5.1 illustrates the four conditions for both the elliptic and parabolic case. We fixed $b = 1$ under the computation of the curves. Since the monotonicity region is above the curve in question it is obvious that criterion E4 decides the monotonicity region. For the homogeneous medium on uniform grid E4 corresponds to the D criteria derived in the previous chapter. As already noticed, see equation (4.11), compressibility has a negative effect on the monotonicity region for these criteria.

5.2 Numerical Results

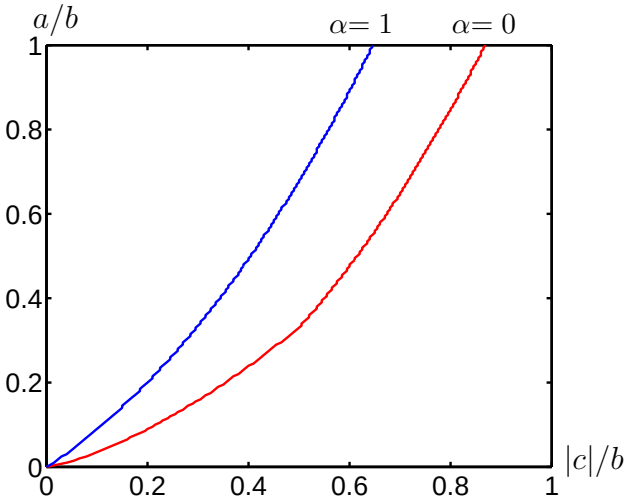
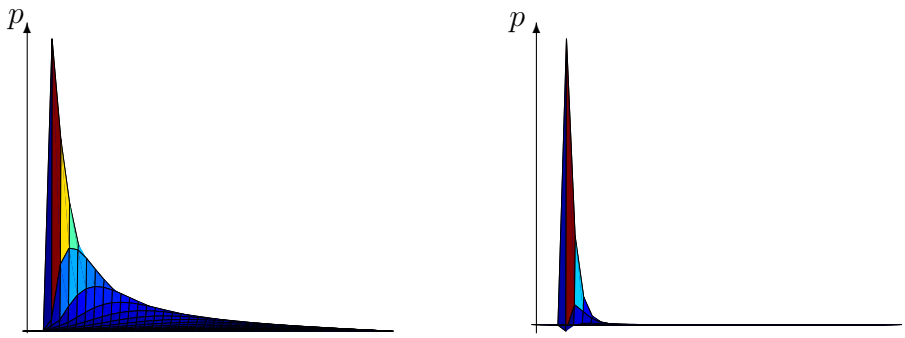


Figure 5.2: *Numerically computed monotonicity regions for the MPFA O-(0) method, the red curve illustrates the region when* $\alpha = 0$ *while the blue is for* $\alpha = 1$.



(a) Incompressible medium (b) Compressibility= 1

Figure 5.3: *Pressure distribution on a* 39×39 *grid, in the region where the incompressible scheme is monotone while the compressible is not:* $a = 0.1$ *and* $c = 0.2$*, the source is in block* $(1, 1)$.

Now we want to investigate the monotonicity regions numerically. This is done by defining the system matrix from the coefficients (4.27) and checking the inverse for non-positive terms. Hence we only check if the system matrix has a monotone inverse. Figure 5.2 shows the numerically computed monotonicity regions for the elliptic and parabolic MPFA O(0) method. Again we see that the elliptic problem has better monotonicity properties.

The article by Keilegavlen et al [18] concludes that if all subdomains are tested for monotonicity then the numerical results will coincide with the analytical result. However, the numerical results that only test for monotone inverse yields valuable information. Even though the criteria defined in (4.27) is necessary to fulfil the discrete maximum principles, it is not necessarily given that they have to be fulfilled to yield correct solutions. Computations of the pressure on uniform grids and homogeneous medium seem to be reasonable, even though the numerical curves in Figure 5.2 is used as reference for monotonicity.

The figures 5.3(a) and 5.3(b) show a case where the incompressible case is monotone while the compressible is not. The test case is a 39×39 grid, with the source term equal to one in block $(1, 1)$. The source term is set to one in all further computations where a source is included. We have chosen an area inside the monotonicity region for the incompressible equation and outside the compressible monotonicity region, $a = 0.1$ and $c = 0.2$. In Figure 5.3(a) the pressure is monotone, but for the compressible case 5.3(b) we can see that the solution is negative in some areas of the domain, i.e. monotonicity is lost. The same tendency will be the case when the mid-block is used as a source term, but it is harder to visualise because of very small oscillations.

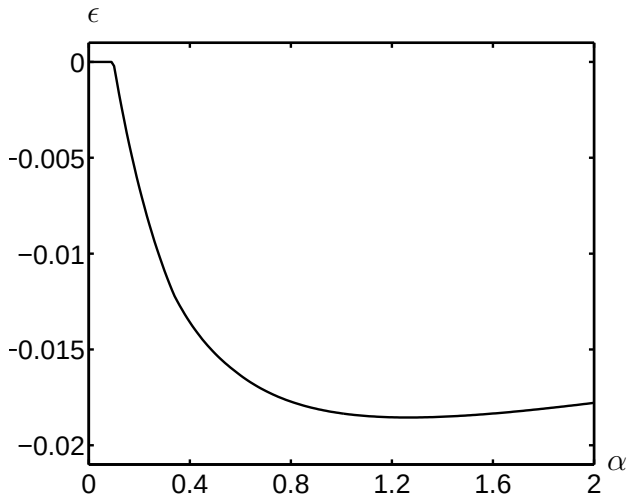


Figure 5.4: *Function (5.5) for varying compressibility on a* 39×39 *grid,* $a = 0.1$ *and* $c = 0.2$

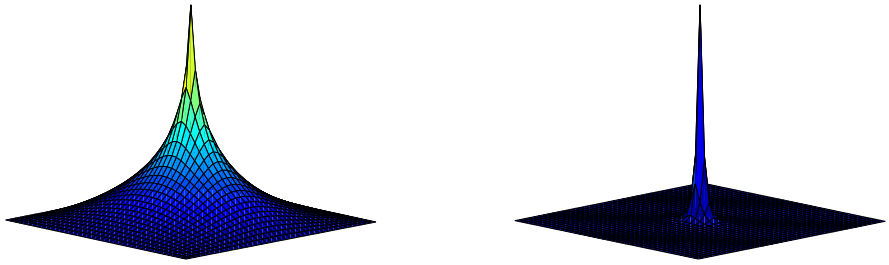
Computations show that as the compressibility term increases, the monotonicity region decreases. This is in accordance with *Remark* 5. So if the time step is small compared to the area of the grid cell we can risk that most of the monotonicity region is lost. Therefore, when $\Delta t \to 0$ the monotonicity criteria will approach the M-matrix criteria. The compressibility term decreases with decreasing grid size, which is in accordance with our intuitive interpretation of the accumulation; the smaller grid cell, the smaller accumulation.

It may be interesting to investigate how the oscillation changes with the compressibility term. For this we introduce the quantity ϵ defined as

$$\epsilon = \frac{\min[A^{-1}]_{i,j}}{\max[A^{-1}]_{i,j}}. \tag{5.5}$$

The quantity ϵ is nonnegative for monotone methods and negative for non-monotone cases. It is computed for different values of the compressibility term. In Figure 5.4 we have computed ϵ for the case when $a = 0.1$ and $c = 0.2$. The elliptic case, i.e. $\alpha = 0$, is inside the monotonicity region, i.e. $\epsilon = 0$. As the compressibility increases up to a certain point we see that the monotonicity criteria are violated and that the oscillations increase. Beyond this point the oscillations decrease slightly. Computation of similar curves for varying a and c indicate that the oscillations are bigger for small values of a and c.

The fact that the monotonicity region decreases with increasing compressibility is easily seen from equation (4.11). Physically it is not necessarily intuitive. The compressibility term may be interpreted as a delayer of the



(a) *Incompressible* (b) *Compressible,* $\alpha = 1$

Figure 5.5: *The pressure distribution for a homogeneous media on uniform grid with a source in the middle.*

flow, i.e accumulation of mass. When it is set to zero, i.e. incompressible flow, its effect will vanish. This is illustrated in Figure 5.5, which shows the pressure distribution in a 39×39 grid with a source term in the middle. In Figure 5.5(b) where the compressible term is set to one, we see that the slope is much steeper than in Figure 5.5(a). When the additional compressibility term is included, the solution has steeper gradients, and it may be more difficult to capture the behaviour and the changes in the pressure.

In reservoir simulation a typical grid sizes can vary from ten to hundreds of metres with a time step from several hours to a few days. With these values the compressibility term is relatively small, which is promising with respect to its impact on the monotonicity during reservoir simulations.

5.3 Formulations of the Discrete Maximum Principle

Recall our discrete system

$$\boldsymbol{B}\boldsymbol{p} = \boldsymbol{q}, \tag{5.6}$$

where $\boldsymbol{B} = \boldsymbol{A} + \alpha \boldsymbol{I}$. Here $\boldsymbol{A}$ is the discretization of the elliptic operator (3.5) such that $\boldsymbol{B}$ is the discretization of the parabolic operator (3.13).

A problem with using the same discrete maximum principle for the parabolic and the elliptic equation is discussed in Chapter 2. For a time discrete solution of parabolic equation its maximum property differs from the elliptic maximum property, and the parabolic equation may allow for local minima. The question then is if we must lower our requirements on the system matrix $\boldsymbol{B}$. It is natural to consider four options:

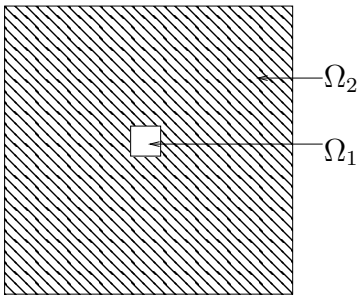


Figure 5.6: *The permeability field in a heterogeneous media. The source is in the middle of domain* Ω_1.

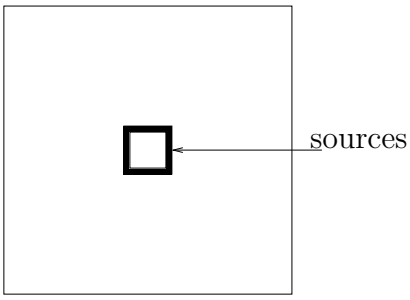


Figure 5.7: *The location of the sources in a homogeneous media.*

1. Monotonicity of $\boldsymbol{B}^{-1}$.
2. Monotonicity of both $\boldsymbol{A}^{-1}$ and $\boldsymbol{B}^{-1}$.
3. Monotonicity of $\boldsymbol{B}^{-1}$ and the Discrete Maximum Principle 1 on $\boldsymbol{A}$.
4. The Discrete Maximum Principle 2 and 1 on $\boldsymbol{B}$ and $\boldsymbol{A}$, respectively.

Now we will discuss each of these options.

Alternative 1: If the system matrix $\boldsymbol{B}$ has a monotone inverse then the discretization shares the property of the continuous system.
Consider a uniform quadrilateral grid Ω consisting of two subdomains Ω_1 and Ω_2 as illustrated in Figure 5.6. The subdomain Ω_1 has permeability $\boldsymbol{K}_1$ and consists of a 5×5 grid placed in the middle of Ω. The permeability in Ω_2 is $\boldsymbol{K}_2$. The grid has dimension 41×41, including ghost cells, and the source term is set to one in the mid-block $(21, 21)$. For this example we have chosen $k^1_{1,1} = 0.05$, $k^1_{1,2} = 0.215$, $k^1_{2,2} = 1$, $k^2_{1,1} = 0.5$, $k^2_{1,2} = 0$, $k^2_{2,2} = 0.5$, and $\alpha = 0.0014$. The system matrix is monotone for this choice of permeabilities, i.e. $\boldsymbol{B}^{-1} > 0$.

In Figure 5.8 we see four different plots of the pressure distribution in the grid. Figure 5.8(a) illustrates the pressure distribution on the 2D-domain. From the figure it is not obvious that there are local minima in the pressure, but there is no absolute minimum inside the domain. However in the magnified the critical area around the source, see Figure 5.8(c), it is easy to see that the solution has local minima, and the maximum principle is violated. In Figure 5.8(b) the column vector of the pressure, which goes through the source, is shown. To underscore that it is a local minimum we also illustrate the row vector, see Figure 5.8(d), which intersects the second minimum in Figure 5.8(d) at the point 0.5.

The local minima which is shown in this figures are unphysically, when there is only one source in the middle it is no reasonable explanation of how such local minima may appear. Hence, demanding monotonicity of $\boldsymbol{B}$ is not enough to assure a physically reasonable solution.

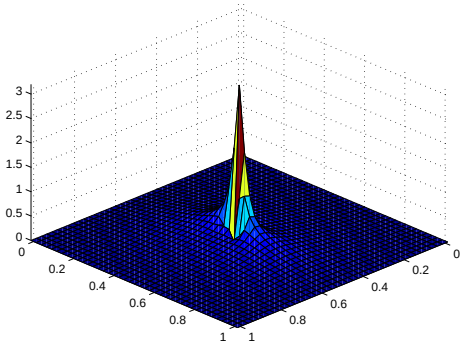
Alternative 2: If the system matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ have monotone inverses then discretization satisfies the property of the continuous system.
From the criteria derived in the previous chapter we know that if $\boldsymbol{B}^{-1} \geq 0$ then $\boldsymbol{A}^{-1} \geq 0$. Hence, the same case as for statement 1 can be used to disprove the statement. For M-matrices this is opposite: $\boldsymbol{A}^{-1} \geq 0$ implies that $\boldsymbol{B}^{-1} \geq 0$. Since we have one example illustrating that oscillations occur when both $\boldsymbol{A}^{-1} \geq 0$ and $\boldsymbol{B}^{-1} \geq 0$, we have shown that the statement is not enough to assure a physically reasonable solution.

Alternative 3: If the Discrete Maximum Principle 1 holds for $\boldsymbol{A}$, and $\boldsymbol{B}$ has a monotone inverse, then the discretization satisfies the property of the continuous system.
If we again look at the example in Figure 5.8. Now we investigate whether the Discrete Maximum Principle 1 holds for the matrix $\boldsymbol{A}$, i.e. for all possible subdomains $\boldsymbol{A}_{sub}^{-1} \geq 0$. To do this you can for instance take out a single row, which corresponds to the pressure in one cell, of the system matrix $\boldsymbol{A}$ and check if the criteria A through D holds. For the case we find that these are violated, hence the matrix $\boldsymbol{A}$ does not satisfy the discrete maximum principle. Therefore the discrete maximum principle is stronger than monotonicity. Note that we do not have a homogeneous medium such that criteria E are no longer valid.

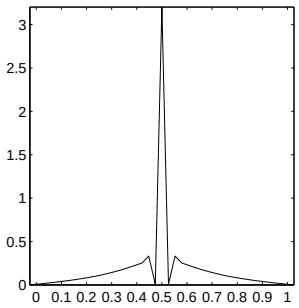
In Figure 5.2 we see where the inverse of the matrix $\boldsymbol{B}$ is monotone for $\alpha = 1$. Plotting this together with the curve of the Discrete Maximum Principle 1 for $\boldsymbol{A}$ results in Figure 5.9. We see roughly that if $\alpha > 1$ then the monotonicity of $\boldsymbol{B}^{-1}$ will determine the monotonicity region and if $\alpha < 1$ then the Discrete Maximum Principle 1 will determine the region. These criteria will be less restrictive than the criteria E.

Now the interesting question is whether Alternative 3 is good enough to prevent unwanted oscillations. We have not been able to come up with a way to verify or disprove the alternative.

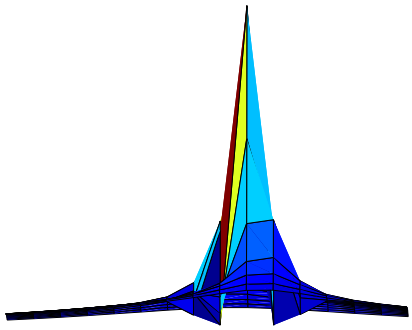
Alternative 4: Discrete maximum principle for both $\boldsymbol{A}$ and $\boldsymbol{B}$.
This is the case which we derived monotonicity criteria for in the previous chapter. We know that these criteria are good enough to prevent unwanted oscillations, the question is if they are to strong. An alternative is to test the example with the heat ring on a cold surface, which we discussed in Chapter 3. The same case will be valid for the pressure. Given a homogeneous media with permeability values inside the monotonicity region for $\alpha = 1$, and include a source square in the middle, as illustrated in figure 5.7. For a



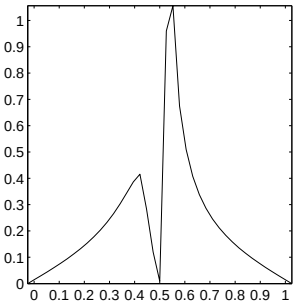
(a) *Surface of the pressure in the grid.*



(b) *Column vector* 21 *of the pressure.*



(c) *Close up of Figure 5.8(a)*



(d) *A plot of the row vector* 22 *correspond to the right local minima in Figure 5.8(b).*

Figure 5.8: 41×41 *grid with a positive source in the middle,* $[21, 21]$. *The permeabilities are* $k^1_{1,1} = 0.05$, $k^1_{1,2} = 0.215$ *and* $k^1_{2,2} = 1$ *in the source cell and in the eight grid cells which surround it. In the other grid cells* $k^2_{1,1} = 0.5$, $k^2_{1,2} = 0$ *and* $k^2_{2,2} = 1$.

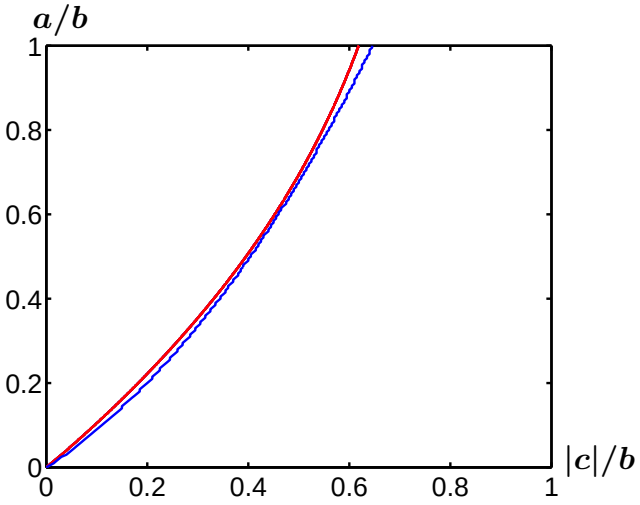


Figure 5.9: *The figure illustrates the monotonicity region for Alternative* 3. *The blue curve is when the matrix* $\boldsymbol{B}^{-1}$ *is monotone for* $\alpha = 1$ *and the red curve is the curve for when the Discrete Maximum Principle 1 is fulfilled.*

compressible fluid there shall be a local minimum between the sources, if the time step is small enough. In an incompressible fluid there is no room for local minima in. The pressure distribution for two fluids, the incompressible and the compressible, is illustrated in Figure 5.10 and Figure 5.11, respectively. We see that the pressure behaves as expected, which imply that the Discrete Maximum Principle 2 allows for physically local minima.

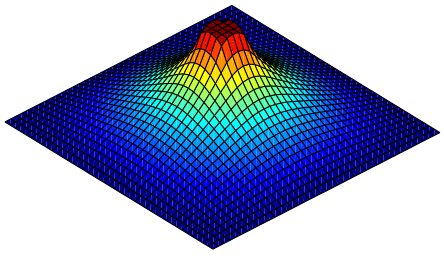


Figure 5.10: *Incompressible medium.*

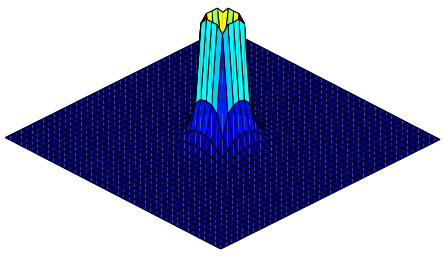


Figure 5.11: *Compressible medium,* $\alpha = 1$.

L-method

Last year a new, promising MPFA method was introduced. The method is referred to as the L-method and was derived by Aavatsmark et al. [8]. It was actually first mentioned in [23] as the seven-point method. One of the reason why this method is promising, is that For local grid refinements, test runs indicate that the new method yield almost optimal solutions, and that the solution is considerably better than the solutions obtained by the O(η)-methods. With respect to the conditions E, the method has optimal monotonicity properties, i.e. $\gamma = |c|$. Thus,

$$|c| \leq \min\{a, b\}, \tag{5.7}$$

is the monotonicity region. In addition the system matrix is always an M-matrix when the method is monotone. Which indicates that the compressibility will not have any affect on the compressibility. Thus, the weakening of the monotonicity when introducing the compressibility term on an MPFA discretization will not necessary be a problem for all MPFA discretization.

Chapter 6

Summary, Conclusion and Further Work

In this thesis we have expanded a set of monotonicity criteria for elliptic conservation laws to be valid for parabolic conservation laws. In Chapter 1 we went through some theory of reservoir mechanics and derive models for single phase flow. Two control volume methods, with particular emphasis on the MPFA O(0)-method, were explained in Chapter 2.

In Chapter 3 we went through the maximum principle for the elliptic and parabolic equations. Since the elliptic and parabolic equations have different maximum principles in space, we discussed whether it was necessary to introduce a new discrete maximum principle for the parabolic equation. We proposed to maintain the already existing discrete maximum principle, and postponed the discussion to Chapter 5 in which numerical examples were available.

In Chapter 4 we used the theory of weakly regular splitting, introduced in Chapter 3, to derive a set of monotonicity conditions on general quadrilateral grids for a discrete parabolic conservation law. This was preformed on general quadrilateral grid for the discrete parabolic system satisfying the Discrete Maximum Principle 2. We deduced these criteria analytically for the case of homogeneous media and uniform grids. Both analytical and numerical results were discussed in Chapter 5. We saw that these criteria lead to a stronger set of criteria than those resulting form an elliptic conservation law, both analytically and numerically. In addition we studied four different formulations of the discrete maximum principle and their implications. Two of these formulations allowed for unphysical oscillations. The two others, in which the Discrete Maximum Principle 2 is included, seem more promising.

The Discrete Maximum Principle 2 is well established and tested for elliptic problems, so it may be natural to assume that it is the right choice

for preventing unwanted oscillations in the solution. It seems like the principle also allows for physically local minima for the time discrete parabolic equation. Although this principle guarantees no unwanted oscillations, we cannot claim that other formulations of a discrete maximum principle is wrong. There may be solutions of the discrete problem which yield physically correct solutions without fulfilling the Discrete Maximum Principle 2. Therefore, the third alternative in Chapter 5 may be useful in further investigation of monotonicity of parabolic equations.

When we look at the discretization, the pressure in the accumulation term is actually an average pressure over the cell, while the pressure calculated from the flux expression is an approximation to the pressure in the cell centre. For a pressure that is changing, these values will not be the same. In a cell with a maximum the average pressure will be lower than the pressure in the cell centre, which is the maximum value. In our computations we have not taken this into consideration. Therefore it may be interesting to use TPFA or MPFA to smoothen the pressure value from the compressibility term, such that it actually will contribute as a average pressure.

Since the time discrete parabolic equation allows for local minima, it may be difficult to know what we can expect from our solution. Therefore it may be an idea to use a second order method on the time discretization. If more than one time step is included in the method we may be able capture more information about the changes in time, and thus utilize monotonicity restrictions in the time domain.

The motivation for using monotone methods is to prevent spurious oscillations, which may lead to circulations in the velocity or for instance artificial liberation of gas due to wrongly computed pressure. Studying the fluxes it is natural to imagine that the monotonicity of the compressible problem is necessary to prevent unwanted circulations in the velocity. Circulations in a curl free velocity field is unphysical and the resulting streamlines may be useless. Since lost of monotonicity in theory can have serious effects, it may be interesting to investigate if this actually is a problem in a reservoir simulation, and if compressibility make it even worse. If this is the case it may be necessary to consider other discretization techniques for compressible problems. This may mean that methods which always yield M-matrices for monotone methods, e.g. the L-method, is the right choice when compressible problems are to be solved. Due to this, it may be more important to favour further analysis of such methods.

Bibliography

[1] T. Aadland. Flukssplitting som prekondisjonering for iterativ løsning av en elliptisk differensialligning diskretisert ved mpfa o-metoden. Master's thesis, Universitetet i Bergen, 2003.

[2] I. Aavatsmark. An introduction to multipoint flux approximation for quadrilateral grids. *Computational Geosciences*, 6:405–432, 2002.

[3] I. Aavatsmark. *Bevarelsesmetoder for elliptiske differensialligninger*. Matematisk Institutt, Universitetet i Bergen, 2004.

[4] I. Aavatsmark, T. Barkve, Ø. Bøe, and T. Mannseth. Discretization on non-orthogonal, curvilinear grids for multi-phase flow. In *Proc. of the 4th European Conf. on the Mathematics of Oil Recovery, Røros, Norway,*, volume D, 1994.

[5] I. Aavatsmark, T. Barkve, Ø. Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. part 1: Derivation of the methods. *SIAM Journal on Scientific Computing*, 19(5):1700–1716, 1998.

[6] I. Aavatsmark, T. Barkve, Ø. Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. part 2: Discussion and numerical results. *SIAM Journal on Scientific Computing*, 19(5):1717–1736, 1998.

[7] I. Aavatsmark, T. Barkve, Ø. Bøe, and T.Mannseth. Discretization on non-orthogonal, quadrilateral grids for inhomogeneous, anisotropic media. *Journal of computational physics*, 127:2–14, 1996.

[8] I. Aavatsmark, G.T. Eigestad, B.T. Mallison, and J.M. Nordbotten. A compact multipoint flux approximation method with improved robustness. *Numerical Methods for Partial Differential Equations*, 24:1329–1360, 2008.

[9] I. Aavatsmark, T. Eigestad, R. A. Klausen, M.F. Wheeler, and I. Yotov. Convergence of a symmetric mpfa method on quadrilateral grids. *Computational geosciences*, 11:333–345, 2007.

[10] T.M. Apostol. *Calculus, Volume II*. John Wiley and sons, 1969. ISBN 0-471-00007-8.

[11] J. Bear. *Dynamics of Fluids in Porous Media*. Dover Publications, Inc., 1972.

[12] J. Bear. *Hydraulics of Groundwater*. New York : McGraw-Hill, 1979. ISBN 0-07-004170-9.

[13] A. Breman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, 1979. ISBN 0-12-092250-9.

[14] M.G. Edwards and C.F. Rogers. A flux continuous scheme for the full tensor pressure equation. In *Proc. of the 4th European Conf. on the Mathematics of Oil Recovery, Røros, Norway,*, 1994.

[15] M.G. Edwards and C.F. Rogers. Finite volume discretization with imposed flux continuity for the general tensor pressure equation. *Computational Geosciences*, 2(4):259–290, 1998.

[16] L.C. Evans. *Partial Differential Equations*. American Mathematical Society, 1998. ISBN 0-8218-0772-2.

[17] E. Hopf. Elementare bemerkungen uber die losungen partieller differentialgeleichungen zweitet ordnung vom elliptischen typus. *Sitzungesberichte der Preussischen Akademie der Wissenschaften*, 19:147–152, 1927.

[18] E. Keilegavlen, I. Aavatsmark, and J.M. Nordbotten. Sufficient criteria are necessary for monotone control volume methods. *Applied Mathematics Letters*, 2009.

[19] R.A. Klausen and R. Winther. Convergence of multipoint flux approximations on quadrilateral grids. *Numerical Methods for Partial Differential Equations*, 22:1438–1454, 2006.

[20] R.A. Klausen and R. Winther. Robust convergence of multi point flux approximation on rough grids. *Numerische Mathematik*, 104:317–337, 2006.

[21] R.J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhauser, 1992. ISBN 3-7643-2723-5.

[22] J.M. Nordbotten and I. Aavatsmark. Monotonicity conditions of control volume methods on uniform parallelogram grids in homogeneous media. *Computational Geosciences*, 9:61–72, 2005.

[23] J.M. Nordbotten, I. Aavatsmark, and G.T. Eigestad. Monotonicity of control volume methods. *Numerische Mathematik*, 106:255–288, 2007.

[24] Ø. Pettersen. *Grunnkurs i reservoarmekanikk*. Matematisk Institutt, Universitetet i Bergen, 1990.

[25] M.H. Protter and H.F. Weinberger. *Maximum Principles in Differential Equations*. Prentice-Hall, Inc, 1967.

[26] H. Reme. *The Preconditioning and Multi-Point Flux Approximation Methods for solving Secondary Oil Migration and Upscaling Problems*. PhD thesis, University of Bergen, 1999.

[27] J. Smoller. *Shock waves and reaction-diffusion equations*. New York : Springer, 1994. ISBN 0-387-94259-9.