Technische Universität Bergakademie Freiberg

Institut für Numerische Mathematik und Optimierung

Numerical Modelling of Uncertainty in Porous Media Transport by the Stochastic Finite Element Method

Diploma Thesis

Supervisor:
Priv.-Doz. Dr. O. Ernst

written by
Susanna Kube, Matr.-Nr. 40945

August 25, 2004

corrected: November 23, 2004

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Partial differential equations play an important role in applied mathematics because such equations occur in many fields in natural science and engineering. For example, they can be used to describe transport processes in groundwater or oil reservoirs. The principal variable of interest is the concentration of a fluid. In reservoir simulation it indicates how much of the reservoir is swept by a solvent, or equivalently, how much oil is recovered. In subsurface contaminant transport, it models the movement of a solute in groundwater porous media flows. In general, the spreading of an invading fluid in groundwater or the transport of contaminations depend on the velocity of the resident fluid. The velocity field has to be known in order to simulate the spreading. Hence, the first step in modelling transport processes consists in calculating an approximate velocity field. Secondly, the transport can be simulated. However, this procedure is only possible if transport and flow are decoupled. This is often possible in practice, especially when the concentration of the invading substance is very small. This means density and viscosity of the resident fluid are not influenced. In the other case, flow and transport have to be determined simultaneously, which becomes more challenging.

The transport process is driven by advection, which takes place when a velocity field is present, by diffusion, which is due to concentration gradients, and by dispersion, which is caused by heterogeneities in the porous medium [15], [9]. The governing advection diffusion equation results from Fick's law and conservation of mass [4]. Only in a few cases, an analytical solution is available. Hence, the equation must be solved numerically. Several numerical methods are available, such as Finite Difference Methods (FDM) or Finite Element Methods (FEM) [3]. However, it turns out that there arise numerical difficulties when the process is dominated by advection [6]. Often, one can observe oscillations or strong damping of the numerical solution. Many methods have been developed to improve the numerical simulation, for example by using stabilization techniques. Another approach is based on the characteristic method which follows the particles with time instead of considering a fixed control volume.

Our first task is to examine some numerical methods for solving unsteady advection-diffusion equations. Dispersion analysis, truncation errors, and stability results are useful tools to analyze different schemes. Their performance is compared with regard to some typical test problems in one and two dimensions. Special emphasis is put on the Eulerian Lagrangian Localized Adjoint Method (ELLAM), described in [2] and [23], which overcomes many disadvantages other numerical methods suffer from.

Many equations arise directly from engineering applications. Often, boundary conditions and coefficients are determined by physical measurements. These measurements suffer from uncertainties. Therefore, it seems reasonable to model the data as stochastic processes instead of treating them as deterministic variables [29], [11]. The stochastic finite element method [1] will be used to solve the stochastic advection-diffusion equation and to examine the propagation of uncertainties to the simulation output.

This work is arranged as follows. Chapter 2 deals with the modelling of transport processes in porous media, following the approaches given in [9], [4] and [24]. Chapter 3 gives an overview of numerical methods for the advection-diffusion equation. The first section deals with the steady equation and the second section deals with the unsteady process. One-dimensional examples illustrate the basic idea of the Galerkin finite element method and point out the numerical difficulties. Chapter 4 introduces the concept of ELLAM. It shows how the method can be implemented in one and two dimensions. The main aspects are implementation of boundary conditions and conservation of mass. Chapter 5 presents numerical results to compare different methods for the unsteady advection-diffusion equation in one and two dimensions. The topic of Chapter 6 is the solution of the stochastic advection-diffusion equation. It starts with the theoretical aspects of the stochastic Galerkin finite element method such as proper function spaces and the variational formulation. Then the stochastic advection-diffusion equation with stochastic velocity and stochastic diffusion is solved. It is shown how uncertainties in the data influence the solution.

Chapter 2

Modelling of Subsurface Flow

2.1 Modelled Quantities

Holzbecher [9] describes flow and transport in porous media as follows. There are holes of different shape and size between rocks and soil. These holes, known as "pores", are filled with fluid (generally water) and/or gas (often air). Concerning the aggregate state, there are three different phases: rock/soil (solid phase), water (fluid phase), and air (gas phase). Here only the saturated state is considered, which means the gas phase is neglected. Furthermore, the solid phase will not be taken into consideration. Hence, only the fluid phase in the pore space will be considered. In reservoir simulation of oil recovery processes, the fluid phase indeed consists of two phases—oil and water. However, in all cases presented in this paper, the mixed fluid is assumed to form a single phase.

For the mathematical description of processes in groundwater flow, variables defined in space and time are required, e.g. concentration or velocity. Some variables are not defined on the continuum, e.g. the porosity must be given by averaged values. Before dealing with the governing equations, some basic concepts will be explained.

2.2 Transport Processes

Advection: Advection is the transport with the flow. In a uniform velocity field, a moving "cloud" does not change its shape. It is only shifted in the direction of the velocity field with the corresponding speed. When the velocity is not constant, the cloud is deformed. In natural systems, pure advection does not occur. Advection and convection are often used as synonyms even though they do not mean the same. Convection denotes a system where flow and transport influence each other. Convection is due to density and viscosity gradients. Advection plays a role in convective processes, while advection itself is an independent process.

Molecular Diffusion: According to [15] and [7], this is the spreading caused by the random molecular motion and collisions of the particles themselves. This type of motion is driven by concentration gradients. In the case of pure diffusion (without flow), the concentration spreads equally in all space directions. Around sources, concentric circles (in 2D) or spheres (3D) of equal concentration appear. Molecular diffusion is present whether or not the fluid is moving.

Kinematic (or **Mechanical**) **Dispersion**: This is the spreading caused by the variability of the complex, microscopic velocities through the pores in the medium. It is linked to the heterogeneities in the medium and is present only if there is flow. Thus, it adds a spreading effect to the diffusion. It is observed in higher dimensions that the spreading caused by dispersion is greater in the direction of flow than in the transverse directions.

2.3 Analytical Description

Fluid flow in porous media is governed by the fundamental laws of conservation of mass, momentum, and energy. In addition, rate equations and equations of state must be specified. In order to obtain a complete mathematical model, a set of initial conditions and a set of appropriate boundary conditions must be given.

2.3.1 Governing Equations

In many cases, the thickness of the medium is significantly smaller than its length and width. Hence, it is reasonable to average the medium properties vertically and to assume the flow to take place in a region $\Omega \subset \mathbb{R}^2$ with a nonuniform local elevation.

The following equations are taken from [24] and [4]. The most widely used relation in analytical models of flow in porous media is Darcy's law, which is a basic relationship between the flow rate and the pressure gradient. It states that the volumetric flow rate Q of a fluid through a porous medium is proportional to the pressure gradient and to the cross-sectional area A (normal to the direction of flow) and inversely proportional to the viscosity μ of the fluid. The law defines the permeability $\boldsymbol{K}(\boldsymbol{x})$ of the rock, which quantifies its ability to transmit fluid. The Darcy velocity $\boldsymbol{u}(\boldsymbol{x}, t)$ is given by

$$\boldsymbol{u} = -\frac{\boldsymbol{K}}{\mu(c)}(\nabla p - \rho g \nabla Z), \qquad \boldsymbol{x} \in \Omega,\ t \in [0, T] \tag{2.1}$$

where $p(\boldsymbol{x}, t)$ is the fluid pressure, ρ is the fluid density, g is the acceleration due to gravity, the depth $Z = Z(\boldsymbol{x})$ is a vector function pointing in the direction of gravity, and $\boldsymbol{K}(\boldsymbol{x})$ is an absolute permeability tensor with units of length squared. Often $\boldsymbol{K}$ is assumed to be a special diagonal tensor. Generally the viscosity μ depends on the concentration $c(\boldsymbol{x}, t)$ of an invading fluid

$$\mu(c) = \mu(0)[(1 - c) + M^{\frac{1}{4}} c]^{-4},$$

where M is the mobility ratio between the resident and injected fluids, and $\mu(0)$ is the viscosity of the resident fluid (oil). The concentration c is the mass fraction of the concentration of the invading substance relative to the total mass of the fluid.

The mass conservation equation for the fluid mixture leads to the equation

$$\nabla \cdot \boldsymbol{u} = q, \qquad \boldsymbol{x} \in \Omega,\ \ t \in [0, T] \tag{2.2}$$

with an external source/sink term q that accounts for the effect of e.g. injection and production wells.

Following Holzbecher [9], the differential form of a general conservation equation can be written as

$$-\nabla \cdot \boldsymbol{j} + q = \frac{\partial \theta}{\partial t},$$

where $\boldsymbol{j}$ denotes the vector of mass flow, q is a source/sink term, and θ denotes the total change of mass. Generally, exchange processes between different phases have to be taken into account in q. However, the total fluid mixture is assumed to form a single phase which flows as one fluid.

Here, θ is given by $\theta = \Phi\rho c$, where $\Phi(\boldsymbol{x})$ is the porosity of the medium, the portion of the volume available for flow. It is the fraction of the size of the pore space relative to the total volume. The pore space is the space where flow and transport can take place. It is defined as an average value. The pore space is assumed to be saturated, i.e. it is completely filled with fluid.

The mass flow $\boldsymbol{j}$ can be decomposed into an advective and a diffusive-dispersive part

$$\boldsymbol{j} = \rho c \boldsymbol{u} + \boldsymbol{j}_c,$$

where the diffusive-dispersive mass flow $\boldsymbol{j}_c = \rho \boldsymbol{u}_d$ in a porous medium is given by Fick's law

$$\boldsymbol{u}_d = -\boldsymbol{D}\nabla c.$$

$\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{u})$ is the diffusion-dispersions tensor

$$\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{u}) := \Phi(\boldsymbol{x}) d_m \boldsymbol{I} + \frac{d_l}{|\boldsymbol{u}|} \begin{pmatrix} u_x^2 & u_x u_y \\ u_x u_y & u_y^2 \end{pmatrix} + \frac{d_t}{|\boldsymbol{u}|} \begin{pmatrix} u_y^2 & -u_x u_y \\ -u_x u_y & u_x^2 \end{pmatrix}, \tag{2.3}$$

where d_m is the molecular diffusion coefficient, $\boldsymbol{I}$ is the identity tensor, and d_l and d_t are the longitudinal and transverse dispersivities. For very low velocity flows, molecular diffusion dominates. For larger velocities, dispersion dominates, and the transverse dispersivity is roughly an order of magnitude smaller than the longitudinal dispersivity.

For constant ρ and Φ the governing equation for the miscible displacement of one incompressible fluid by another in a porous medium can be derived from conservation of mass and is given by

$$\Phi \frac{\partial c}{\partial t} + \nabla \cdot (\boldsymbol{u}c - \boldsymbol{D}\nabla c) = \overline{c} q = f, \qquad \boldsymbol{x} \in \Omega, \;\; t \in [0, T]. \tag{2.4}$$

$\overline{c}(\boldsymbol{x}, t)$ is either the specified concentration of the injected fluid at injection wells, or $\overline{c}(\boldsymbol{x}, t) = c(\boldsymbol{x}, t)$ is the resident concentration at production wells.

Equations (2.1) to (2.4) present a system of coupled nonlinear PDEs. In a first step we consider the simplified problem

$$\begin{aligned} \nabla \cdot \boldsymbol{u} &= q, && \boldsymbol{x} \in \Omega, \;\; t \in [0, T] \\ \boldsymbol{u} &= -\boldsymbol{K}\nabla p, && \boldsymbol{x} \in \Omega, \;\; t \in [0, T] \end{aligned} \tag{2.5}$$

$$\Phi \frac{\partial c}{\partial t} + \nabla \cdot (\boldsymbol{u}c - \boldsymbol{D}\nabla c) = \overline{c} q = f, \qquad \boldsymbol{x} \in \Omega, \;\; t \in [0, T]. \tag{2.6}$$

The viscosity is assumed to be independent of the concentration and is included in $\boldsymbol{K}$. Furthermore, the velocity field is assumed to be stationary. This allows to solve (2.5)

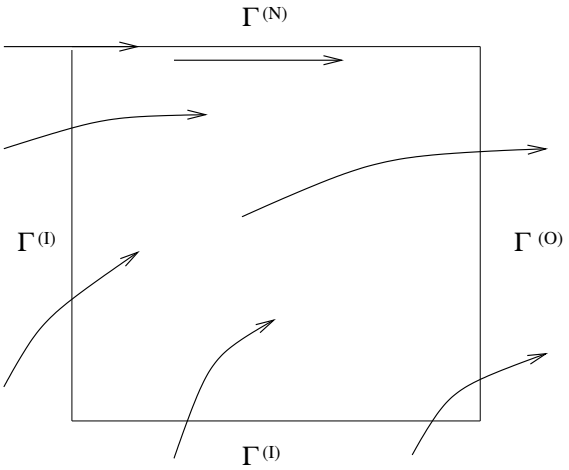


Figure 2.1: Boundary notation

independently of (2.6). Mixed finite element methods (MFEMs) approximate both p and $\boldsymbol{u}$ from the system (2.5) simultaneously. It is assumed that $\boldsymbol{u}$ and p are already known from such computations. Consequently, (2.6) can be solved for $c(\boldsymbol{x}, t)$. (2.6) is an advection-diffusion equation. In general, the magnitude of the diffusion-dispersion tensor is much smaller than the Darcy velocity $\boldsymbol{u}$. Hence, equation (2.6) is an advection-dominated PDE.

In the following chapters we also consider the one dimensional analogue to (2.6)

$$\Phi c_t - D c_{xx} + u c_x + u_x c = f, \qquad x \in \Omega, \;\; t \in [0, T], \tag{2.7}$$

where D is assumed to be constant. If u is independent of x the last term vanishes.

2.3.2 Initial and Boundary Conditions

It is assumed that there exists a fixed spatial boundary $\partial\Omega := \Gamma$. For steady state problems this boundary can be decomposed into

- an inflow boundary $\Gamma^{(I)} := \{\boldsymbol{x} \in \Gamma | \boldsymbol{u}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) < 0\}$
- an outflow boundary $\Gamma^{(O)} := \{\boldsymbol{x} \in \Gamma | \boldsymbol{u}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) > 0\}$
- a noflow boundary $\Gamma^{(N)} := \{\boldsymbol{x} \in \Gamma | \boldsymbol{u}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) = 0\}$

where $\boldsymbol{n} = \boldsymbol{n}(\boldsymbol{x})$ is the outward unit normal. Figure 2.1 illustrates this notation.

For time-dependent problems the velocity can change with time. Thus, the space-time boundary $\Gamma \times [0, T]$ can be decomposed into

- an inflow boundary $S^{(I)} := \{(\boldsymbol{x}, t) \in \Gamma \times [0, T] | \boldsymbol{u}(\boldsymbol{x}, t) \cdot \boldsymbol{n}(\boldsymbol{x}) < 0\}$
- an outflow boundary $S^{(O)} := \{(\boldsymbol{x}, t) \in \Gamma \times [0, T] | \boldsymbol{u}(\boldsymbol{x}, t) \cdot \boldsymbol{n}(\boldsymbol{x}) > 0\}$
- a noflow boundary $S^{(N)} := \{(\boldsymbol{x}, t) \in \Gamma \times [0, T] | \boldsymbol{u}(\boldsymbol{x}, t) \cdot \boldsymbol{n}(\boldsymbol{x}) = 0\}$

In the following it is assumed that the type of boundary for a fixed spatial location $\boldsymbol{x}$ does not change with time. For example, this is the case if the velocity field is steady. Then there holds

- $S^{(I)} = \Gamma^{(I)} \times [0,T]$ with $\Gamma^{(I)} := \{\boldsymbol{x} \in \Gamma | \boldsymbol{u}(\boldsymbol{x},t) \cdot \boldsymbol{n}(\boldsymbol{x}) < 0 \ \ \forall t \in [0,T]\}$
- $S^{(O)} = \Gamma^{(O)} \times [0,T]$ with $\Gamma^{(O)} := \{\boldsymbol{x} \in \Gamma | \boldsymbol{u}(\boldsymbol{x},t) \cdot \boldsymbol{n}(\boldsymbol{x}) > 0 \ \ \forall t \in [0,T]\}$
- $S^{(N)} = \Gamma^{(N)} \times [0,T]$ with $\Gamma^{(N)} := \{\boldsymbol{x} \in \Gamma | \boldsymbol{u}(\boldsymbol{x},t) \cdot \boldsymbol{n}(\boldsymbol{x}) = 0 \ \ \forall t \in [0,T]\}$

Along $S^{(I)}$ and $S^{(O)}$, one of Dirichlet, Neumann, or Robin (flux) boundary conditions may be imposed

$$\begin{aligned} c(\boldsymbol{x},t) &= g_1^{(i)}(\boldsymbol{x},t), && (\boldsymbol{x},t) \in S^{(i)} \\ -\boldsymbol{D}\nabla c(\boldsymbol{x},t) \cdot \boldsymbol{n} &= g_2^{(i)}(\boldsymbol{x},t), && (\boldsymbol{x},t) \in S^{(i)} \\ (\boldsymbol{u}c - \boldsymbol{D}\nabla c(\boldsymbol{x},t)) \cdot \boldsymbol{n} &= g_3^{(i)}(\boldsymbol{x},t), && (\boldsymbol{x},t) \in S^{(i)}, \ \ i \in \{I,O\} \end{aligned} \tag{2.8}$$

while only a zero flux boundary condition is possible at the impermeable boundary $S^{(N)}$:

$$(\boldsymbol{u}c - \boldsymbol{D}\nabla c(\boldsymbol{x},t)) \cdot \boldsymbol{n} = 0, \qquad (\boldsymbol{x},t) \in S^{(N)}. \tag{2.9}$$

In addition, an initial condition

$$c(\boldsymbol{x},0) = c_0(\boldsymbol{x}) \tag{2.10}$$

is needed to close (2.6).

Chapter 3

Numerical Methods for the Advection-Diffusion Equation

In many industrial applications the advection-diffusion equation is often discretized by finite difference methods (FDM) or finite element methods (FEM). However, when advection dominates the transport process, these methods suffer from serious numerical difficulties. At first, some qualitative aspects will be discussed. Then some numerical difficulties arising in steady advection-diffusion problems will be considered before discussing numerical methods for the unsteady equation.

3.1 Qualitative Discussion

The description of the following concepts follows [6], pp.123-131. Sometimes rapid, high-frequency—typically node-to-node (or time-step-to-time-step)—oscillations, called **wiggles**, appear in the numerical solution. Wiggles are non-physical oscillations that can be caused when the gradient of the dependent variable in the flow direction is too large to be resolved by the mesh. One way to understand wiggles is to study the eigenproblems associated with the spatial operators (matrices) in that wiggles are excitations of high-frequency eigenmodes, i.e. the amplitude coefficient of one or more of the oscillatory eigenvectors become relatively "too large". For example, wiggles can occur in one of the following cases:

- flow toward a boundary with a Dirichlet boundary condition that generates a boundary layer that is "too thin" relative to the mesh;
- advection of a wave form that cannot be resolved by the chosen mesh, e.g. propagation of a steep front through the domain in a transient problem; in this case, the location where mesh refinement is desired changes in every time step;
- poorly resolved or rough initial conditions.

If wiggles occur they indicate that there exists a deficiency in the mesh design or in the problem specification. They can be circumvented by creating a finer mesh at the price of higher computational cost or by the use of numerical methods that use various stabilization techniques.

While wiggles also appear in stationary problems, the following concepts apply to the unsteady equation. A general solution to the unsteady advection-diffusion equation with constant coefficients consists of the superposition of several modes

$$c = a\cos(kx - \omega t),$$

with frequency ω determined by the particular system and wavenumber k (wavelength $\lambda = 2\pi/k$). The system is said to be **dispersive** if the phase speed $\omega(k)/k$ is not a constant but depends on k. Then the modes will propagate at different speeds, they will disperse. Concerning the constant-coefficient advection-diffusion equation, all grid based numerical solutions are dispersive, while continuum solutions are not.

Another notion is **dissipation**. It denotes the loss of energy in the solution with time where the energy is measured in a special norm which depends on the problem. The pure advection equation with constant coefficients and periodic boundary conditions is not dissipative. The goal is to use numerical schemes that preserve this property. Often, this is not the case. When the pure advection equation is solved by a numerical approximation method that reduces the amplitude and changes the shape of the initial wave in a way analogous to a diffusive process, the method is said to suffer from dissipation. A dissipative scheme will decrease the energy in the wave.

Many numerical methods were designed that display as little dispersion as possible, because dispersion generates wiggles. For this reason, numerical diffusion is sometimes added to an otherwise non-dissipative scheme. Hence, the interaction of dissipation and dispersion must be taken into account.

An extended discussion of these concepts illustrated with many examples can be found in [6].

3.2 Numerical Methods for the Steady Advection-Diffusion Equation

The steady advection-diffusion equation can be written as

$$\nabla \cdot (\boldsymbol{u}c - \boldsymbol{D}\nabla c) = f.$$

For $\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{u}) \equiv D$ independent of $\boldsymbol{u}$ and $\boldsymbol{x}$, it reduces to

$$-D\Delta c + \boldsymbol{u} \cdot \nabla c + \nabla \cdot \boldsymbol{u}c = f. \tag{3.1}$$

The first term on the left hand side is the diffusion-dispersion term. It contains the highest order differential operator (second order), which determines the type of the PDE (elliptic) and the type of boundary conditions one has to specify. Often, only the second term on the left hand side is referred to as advection term while the third is included in the reactive term. Therefore (3.1) is also called advection-diffusion-reaction equation.

$\boldsymbol{u}$ is a prescribed velocity field. If D is a characteristic diffusivity, L a characteristic length scale, u a characteristic speed, and c_0 a characteristic concentration, one can estimate the size of the two terms by

$$|D\Delta c| \sim \frac{Dc_0}{L^2}, \quad |\boldsymbol{u} \cdot \nabla c| \sim \frac{uc_0}{L}$$

and their relative size as

$$\frac{|\text{advection term}|}{|\text{diffusion term}|} \sim \frac{uL}{D} =: \alpha, \tag{3.2}$$

which is the non-dimensional Peclet number. For $\alpha \ll 1$ diffusion dominates, and for $\alpha \gg 1$ convection dominates.

The purpose of this subsection is to give a survey of some numerical methods for the stationary scalar linear advection-dominated advection-diffusion equation. Details can be found in [19], [10] and [12]. Throughout this subsection consider the problem

$$-\varepsilon\Delta c + \boldsymbol{u}(\boldsymbol{x}) \cdot \nabla c + r(\boldsymbol{x})c = f \text{ in } \Omega, \;\; c = g \text{ on } \Gamma, \tag{3.3}$$

where Ω is a bounded domain in $\mathbb{R}^d$ with boundary Γ. $\boldsymbol{u} = (u_1, \ldots, u_d)$ and r are smoothly varying coefficients with $|\boldsymbol{u}| \sim 1$ and $\varepsilon > 0$ is a small constant. In contrast to (3.1), $\nabla \cdot \boldsymbol{u}$ has been replaced by $r(\boldsymbol{x})$ to indicate that the underlying advection-diffusion equation can be considered as a special case of the more general advection-diffusion-reaction equation. Furthermore, we use ε instead of D to emphasize that it is small.

The solution c of this problem is in general not globally smooth even for smooth data f and g. c can vary rapidly in a layer of width $O(\varepsilon)$ (exponential boundary layer) at the outflow boundary $\Gamma^{(O)}$. In the limiting case $\varepsilon = 0$ (pure advection) the boundary data g can be prescribed only on the inflow boundary $\Gamma^{(I)}$. Furthermore, c will be discontinuous across the characteristic curve (streamline) $\boldsymbol{x}(s)$ given by $\frac{d\boldsymbol{x}}{ds} = \boldsymbol{u}(\boldsymbol{x})$, $\boldsymbol{x}(0) = \boldsymbol{x}_0 \in \Gamma^{(I)}$ whenever g is discontinuous at $\boldsymbol{x}_0$. For $\varepsilon > 0$ such a discontinuity is spread out over a layer around the characteristic $\boldsymbol{x}(s)$ of width $O(\sqrt{\varepsilon})$ (parabolic boundary layer).

The aim is to construct a numerical method for (3.3) which (i) is higher-order accurate and (ii) has good stability properties without requiring h to be smaller than ε. Conventional schemes can be divided into two categories. The first class consists of formally higher-order accurate methods such as the standard Galerkin method or finite difference methods based on centered differences for the approximation of the advective term $\boldsymbol{u} \cdot \nabla c$. These methods will produce severely oscillating solutions unless $h \leq \varepsilon$ or the exact solution happens to be globally smooth. The second class of methods includes classical monotone upwind schemes obtained by adding an artificial diffusion (or viscosity) term of the form $h\Delta c$. These methods satisfy (ii) and produce non-oscillating solutions but are only first-order accurate. The streamline diffusion method (SDM) satisfies both the conditions (i) and (ii) stated above. This method is a Petrov-Galerkin modification of the standard Galerkin method where artificial diffusion in the streamline direction is introduced by modifying the test functions.

3.2.1 Galerkin Finite Element Methods and Artificial Diffusion

3.2.1.1 Stability Results

The variational formulation to the boundary value problem (3.3) with $g \equiv 0$ is stated as: Find $c \in X = H_0^1(\Omega)$ such that for all $w \in X$,

$$B(c, w) = L(w),$$

where

$$\begin{aligned} B(c,w) &= \varepsilon(\nabla c, \nabla w) + (\boldsymbol{u}\cdot\nabla c, w) + (rc, w), \\ L(w) &= (f, w). \end{aligned}$$

$(\cdot,\cdot)$ denotes the L^2 inner product. The functions $\boldsymbol{u}, r$ and f are assumed to be sufficiently smooth with $r - \frac{1}{2}\nabla\cdot\boldsymbol{u} > \alpha^* > 0$. This assumption is not restrictive, because it could be achieved by a change of variable $c \mapsto e^{\sigma x}v$ for a suitable chosen σ [19]. Stability is established as follows:

$$\begin{aligned} (\boldsymbol{u}\cdot\nabla w, w) &= (\nabla\cdot(\boldsymbol{u}w), w) - (w(\nabla\cdot\boldsymbol{u}), w) \\ &= (w\boldsymbol{u}\cdot\boldsymbol{n}, w)_\Gamma - (\boldsymbol{u}w, \nabla w) - (w(\nabla\cdot\boldsymbol{u}), w) \\ &= -(w, \boldsymbol{u}\cdot\nabla w) - (w(\nabla\cdot\boldsymbol{u}), w). \end{aligned}$$

Hence

$$(\boldsymbol{u}\cdot\nabla w, w) = -\frac{1}{2}(w(\nabla\cdot\boldsymbol{u}), w),$$

and

$$B(w,w) = \varepsilon(\nabla w, \nabla w) - \frac{1}{2}(w(\nabla\cdot\boldsymbol{u}), w) + (rw, w) \geq \varepsilon||\nabla w||^2 + \alpha^*||w||^2, \tag{3.4}$$

with $||v|| = ||v||_{L^2(\Omega)}$. Consider a triangulation $\mathcal{T}_h$ of Ω and let T denote any triangle of $\mathcal{T}_h$. Let $X_h \subset X$ be a conforming finite element space that consists of continuous piecewise polynomials of degree k, i.e.

$$X_h := \{v_h \in X : v_h|_T \in \mathscr{P}_k(T) \text{ for all } T \in \mathcal{T}_h\}.$$

The classical Galerkin method becomes

$$B(c^h, w^h) = L(w^h) \qquad \forall w^h \in X^h.$$

Suppose the mesh is constructed such that there exists a constant d so that the length of any side of a triangle $T \in \mathcal{T}_h$ is bounded below by dh. By the usual approximation theory there holds [12]: Given a function $c \in H^{k+1}(\Omega)$ there exists an interpolant $\tilde{c}^h \in X_h$ such that

$$\begin{gathered} ||c - \tilde{c}^h|| + h||c - \tilde{c}^h||_1 \leq Ch^{k+1}||c||_{k+1}, \\ |c - \tilde{c}^h| \leq Ch^{k+\frac{1}{2}}||c||_{k+1}, \end{gathered}$$

where $|v| = \left(\int_\Gamma v^2|\boldsymbol{n}\cdot\boldsymbol{u}|ds\right)^{\frac{1}{2}}$ and $||v||_s = ||v||_{H^s(\Omega)}$. As a consequence of this estimate and the stability property it is possible to prove the following error estimate [12]:

$$||c - c^h|| + |c - c^h| \leq Ch^k||c||_{k+1}.$$

Since ε may be arbitrarily small one cannot rely on the term $\varepsilon||\nabla w||^2$ for the stability in (3.4). Thus, only L^2-stability is guaranteed. The standard Galerkin method does not perform satisfactorily if the exact solution is non-smooth, i.e. if large gradients occur.

The solution can be stabilized by introducing additional terms in the streamline direction (upstream-weighting). This can be achieved by modifying the test functions or the original equation. However, while classical space-upwinded schemes can greatly suppress the oscillations, they tend to generate solutions with severe damping.

3.2.1.2 A one-dimensional Example

Consider the model problem with constant coefficients,

$$-Dc_{xx} + uc_x = 0, \quad \Omega = (0,1), \qquad c(0) = 1,\ c(1) = 0. \tag{3.5}$$

Dividing both sides by u results in

$$-\varepsilon c_{xx} + c_x = 0 \tag{3.6}$$

where $\varepsilon = D/u$ denotes a length. Let L be the length of the domain (here $L = 1$). For $\varepsilon \ll L$ there holds $\alpha = \frac{L}{\varepsilon} \gg 1$. This means the process is dominated by advection. Assume that $u = 1$ and $D = 0.05$. In this case ε is small compared to L, the length of the domain, and there will be a boundary layer of thickness ε near $x = 1$. Outside this thin layer the solution is approximately equal to one. The analytical solution for this case is given by

$$c(x) = \frac{e^{\alpha x} - e^{\alpha}}{1 - e^{\alpha}}.$$

The weak formulation of (3.5) is: Find $c \in X = \{v \in H^1(\Omega) | v(0) = 1, v(1) = 0\}$ such that

$$\underbrace{\int_0^1 (Dc_x v_x + uc_x v)dx}_{=B(c,v)} = \underbrace{\int_0^1 fv dx}_{L(v)}, \ \ \forall\ v \in Y = H_0^1(\Omega).$$

A finite dimensional approximation results in: Find $c_h \in X_h \subset X$ such that

$$B(c_h, v) = L(v), \ \ \forall\ v \in Y_h \subset Y,$$

which leads to a system of equations

$$A_h c_h = F_h.$$

Take a uniform triangulation $\mathcal{T} = \{T_h^1, T_h^2, \ldots, T_h^N\}$ and linear finite elements, i.e.

$$X_h = \{v \in X : v|_{T_h^k} \in \mathscr{P}_1(T_h^k),\ k = 1, \ldots, N\}.$$

The element Matrix A_h^k associated with element T_h^k has the entries

$$A_h^k = \frac{D}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{u}{2}\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} = \underbrace{A^k}_{\text{diffusion}} + \underbrace{B^k}_{\text{advection}},$$

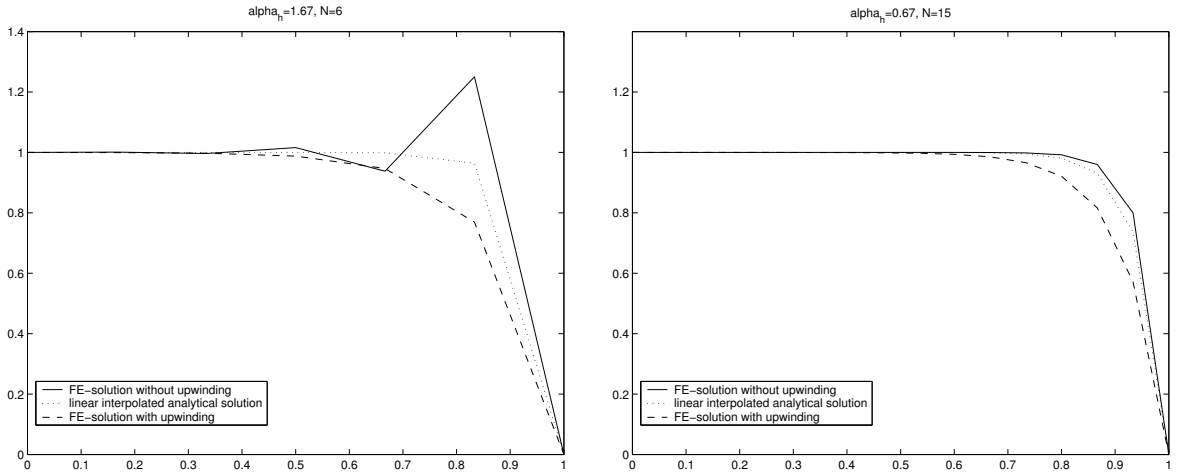
where h is the mesh size. A typical equation in the finite element formulation will thus read

$$-\frac{D}{h}(c_{i-1} - 2c_i + c_{i+1}) + \frac{u}{2}(c_{i+1} - c_{i-1}) = 0,$$

which is identical to a second order finite difference scheme on a uniform mesh.

The (local) grid Peclet number is defined as

$$\alpha_h = \frac{uh}{2D}.$$



oscillatory solution resolved boundary layer

Figure 3.1: Galerkin FE-solution with and without upwinding for different grid sizes.

It gives the ratio of the mesh size h relative to the boundary layer thickness ε. In terms of α_h the scheme can be written as

$$(1-\alpha_h)c_{i+1} - 2c_i + (1+\alpha_h)c_{i-1} = 0, \qquad c_0 = 1,\ c_N = 0$$

with the solution

$$c_j = \frac{\left(\frac{1+\alpha_h}{1-\alpha_h}\right)^N - \left(\frac{1+\alpha_h}{1-\alpha_h}\right)^j}{\left(\frac{1+\alpha_h}{1-\alpha_h}\right)^N - 1}, \qquad j = 0, \ldots, N.$$

It can be observed that the numerical solution will oscillate if the grid Peclet number is greater than one, $\alpha_h > 1$. For $\alpha_h < 1$ there are no oscillations but the boundary layer is not resolved. One possibility to circumvent this problem is to use upwinding. In the finite element context this is achieved by adding a controlled amount of diffusion in the upstream direction. In particular, in 1D, instead of using the physical diffusivity D, one uses the modified quantity

$$\tilde{D} = D + u\frac{h}{2} = D(1+\alpha_h).$$

This corresponds to a finite difference scheme in which a first order upwind difference is used instead of a second order central difference for approximating the advection term. The boundary layer thickness will thus be $\tilde{\varepsilon} = \varepsilon + \frac{h}{2}$. For a fixed discretization the numerical solution will have a boundary layer of no less than $\frac{h}{2}$ even if $\varepsilon \to 0$. In this case, the numerical error will be $\mathcal{O}(h)$ near the boundary layer. On the other hand, the solution is stable. If f varies slowly, the solution outside the boundary layer will typically be well resolved. The results with $u = 1$ and $D = 0.05$ are presented in Figure 3.1.

Instead of solving a modified problem it appears natural to consider the use of a nonuniform grid where the element size is smallest near the boundaries and biggest in the middle of the domain. Thus the boundary layer can be resolved in a proper way.

3.2.2 The Streamline Diffusion Method (SDM)

The streamline diffusion method was developed by Hughes, Johnson and Nävert in the 1980s, see [12]. It combines good global stability properties with high accuracy in subdomains that exclude boundary layers. It can be interpreted as a Petrov-Galerkin method, so it is also known as the **Streamline Upwind Petrov-Galerkin Method (SUPG)** as in [10]. A Petrov-Galerkin method is characterized by the use of distinct trial and test spaces. The SDM has improved stability properties but includes an additional parameter δ which has dimension of time. The streamline diffusion method achieves stability by adding diffusion to the problem only in the direction of the velocity. This produces much less diffusion in the numerical solution than the artificial diffusion method.

Consider a triangulation $\mathcal{T}_h$ of Ω and let T denote any triangle of $\mathcal{T}_h$. As before, consider the finite element space that consists of piecewise polynomials of degree k, i.e.,

$$X_h := \{v_h \in H_0^1(\Omega) : v_h|_T \in \mathscr{P}_k(T) \text{ for all } T \in \mathcal{T}_h\}.$$

Let $c \in H^{k+1}(T)$ with $k \geq 1$.

The idea is to replace the standard FE test functions by modified test functions

$$w \to w + \delta \boldsymbol{u} \cdot \nabla w.$$

Assume that the solution c of (3.3) is regular in the sense that

$$-\varepsilon \Delta c + \boldsymbol{u} \cdot \nabla c + rc = f \qquad \text{in } L^2(T) \ \ \forall T.$$

Then the streamline diffusion method would read: Find $c^h \in X_h$ such that

$$B_{SDM}(c_h, w_h) = L_{SDM}(w_h) \qquad \forall w \in X_h \qquad \text{with} \tag{3.7}$$

$$B_{SDM}(c, w) := \varepsilon(\nabla c, \nabla w) + (\boldsymbol{u} \cdot \nabla c + rc, w) + \sum_{T \in \mathcal{T}_h} \delta_T(-\varepsilon \Delta c + \boldsymbol{u} \cdot \nabla c + rc, \boldsymbol{u} \cdot \nabla w)_T,$$

$$L_{SDM}(w) := (f, w) + \sum_{T \in \mathcal{T}_h} \delta(f, \boldsymbol{u} \cdot \nabla w)_T.$$

$(\cdot, \cdot)$ denotes the inner product in $L^2(T)$. Since in general $\Delta c_h \notin L^2(\Omega)$, but $\Delta c_h \in L^2(T)$ for each T, Δc_h is calculated element by element. The SDM is a residual method, i.e. (3.7) is satisfied if c_h is replaced by c. Hence

$$B_{SDM}(c - c_h, w_h) = 0 \qquad \forall\, w_h \in X_h.$$

A finite element method that satisfies this projection property is said to be consistent.

Errors and stability are measured in the following norm that is related to the discrete bilinear form B_{SDM}:

$$|||v|||_{SDM} := \left(\varepsilon |v|_1^2 + \sum_{T \in \mathcal{T}_h} (\delta_T ||\boldsymbol{u} \cdot \nabla v||_{0,T}^2 + r_0 ||v||_{0,T}^2) \right)^{\frac{1}{2}}.$$

Here, $||\cdot||_{k,T}$ denotes the H^k-norm restricted to T, and $|\cdot|_{k,T}$ denotes the semi-norm, respectively. The constants r_T and r_0 have to satisfy

$$r_T := \max_{\boldsymbol{x}\in T} |r(\boldsymbol{x})| \qquad \forall\, T \in \mathcal{T}_h, \qquad r - \frac{1}{2}\nabla\cdot\boldsymbol{u} \geq r_0 > 0 \qquad \text{on } \Omega.$$

Furthermore, one can prove the local inverse inequality [19]

$$||\Delta v_h||_{0,T} \leq \mu h_T^{-1} |v_h|_{1,T} \qquad \forall\, v_h \in X_h, \tag{3.8}$$

where the constant μ is independent of T and h. Following [19], a proof of the basic stability estimate is given, which reveals that the presence of the term $-\varepsilon\delta(\Delta v_h, \boldsymbol{u}\cdot\nabla w_h)$ does not degrade the extra stability introduced by the term $\delta(\boldsymbol{u}\cdot\nabla v_h, w_h)$.

Lemma 3.1 *Let the SDM-parameter δ_T satisfy*

$$0 < \delta_T \leq \frac{1}{2}\min\left(\frac{r_0}{r_T^2}, \frac{h_T^2}{\varepsilon\mu^2}\right) \qquad \forall\, T \in \mathcal{T}_h.$$

Then the discrete bilinear form is coercive, i.e.,

$$B_{SDM}(v_h, v_h) \geq \frac{1}{2}|||v_h|||_{SDM}^2 \qquad \forall\, v_h \in X_h.$$

Proof: For each $v_h \in X_h$, there holds

$$B_{SDM}(v_h, v_h) \geq \varepsilon|v_h|_1^2 + r_0||v_h||_{0,T}^2 + \sum_{T\in\mathcal{T}_h} \delta_T||\boldsymbol{u}\cdot\nabla v_h||_{0,T}^2 + \sum_{T\in\mathcal{T}_h} \delta_T(-\varepsilon\Delta v_h + r v_h, \boldsymbol{u}\cdot\nabla v_h)_T.$$

Using the local inverse inequality (3.8), the fact that $|(a-b)c| \leq |\frac{1}{2}(a-b)^2 + \frac{1}{2}c^2| \leq a^2 + b^2 + \frac{1}{2}c^2$ and the assumption on δ_T, one obtains

$$\begin{aligned}
&|\sum_{T\in\mathcal{T}_h} \delta_T(-\varepsilon\Delta v_h + r v_h, \boldsymbol{u}\cdot\nabla v_h)_T| \\
&\quad\leq \sum_T \varepsilon^2\delta_T||\Delta v_h||_{0,T}^2 + \sum_T r_T^2\delta_T||v_h||_{0,T}^2 + \frac{1}{2}\sum_T \delta_T||\boldsymbol{u}\cdot\nabla v_h||_{0,T}^2 \\
&\quad\leq \frac{\varepsilon}{2}|v_h|_1^2 + \frac{r_0}{2}||v_h||_0^2 + \frac{1}{2}\sum_T \delta_T||\boldsymbol{u}\cdot\nabla v_h||_{0,T}^2.
\end{aligned}$$

♯

Lemma 3.1 implies the a priori estimate [19]

$$|||c_h|||_{SDM} \leq C(||f||_0^2 + \sum_T \delta_T||f||_{0,T}^2)^{\frac{1}{2}}.$$

Moreover, it gives the stability inequality

$$|||c_h|||_{SDM} \leq 2|||A_h c_h|||_*$$

for the discrete operator $A_h : X_h \mapsto X_h^*$ defined by

$$\langle A_h v_h, w_h \rangle := B(v_h, w_h)_{SDM} \qquad \forall\, v_h,\, w_h \in X_h,$$

with the norm

$$|||g|||_* := \sup_{w_h \in X_h} \frac{\langle g_h, w_h \rangle}{|||w_h|||_{SDM}} \qquad \text{for} \qquad g_h \in X_h^*.$$

Compared with the standard finite element method, for which $\delta_T = 0$, the SDM allows additional control over the derivatives in the streamline direction. This is one of the main results of this subsection.

The mesh Peclet number is defined here by

$$\alpha_T := \frac{h_T ||\boldsymbol{u}||_{L_\infty,T}}{2\varepsilon}.$$

In [19], a proof for the following error estimate can be found.

Theorem 3.1 *Let the assumptions of Lemma 3.1 be fulfilled and let δ_T be specified by*

$$\delta_T = \begin{cases} \delta_0 h_T & \text{if } \alpha_T > 1 \text{ (advection-dominated case)} \\ \delta_1 h_T^2/\varepsilon & \text{if } \alpha_T \le 1 \text{ (diffusion-dominated case)} \end{cases},$$

with appropriate positive constants δ_0 and δ_1. Then the solution c_h of the SDM satisfies the global error estimate

$$|||c - c_h|||_{SDM} \le C(\varepsilon^{\frac{1}{2}} + h^{\frac{1}{2}}) h^k |c|_{k+1}.$$

In the advection-dominated case with a small ε, the global estimate becomes

$$||c - c_h||_0 + (\sum_T \delta_T ||\boldsymbol{u} \cdot \nabla(c - c_h)||_{0,T}^2)^{\frac{1}{2}} \le C h^{k+\frac{1}{2}} |c|_{k+1}.$$

Compared with the error between c_h and the interpolant c_I from X_h to the exact solution c,

$$||c - c_I||_0 \le C h^{k+1} |c|_{k+1} \qquad \text{and} \qquad |c - c_I|_1 \le C h^k |c|_{k+1},$$

one can see that for the above choice of δ_T the L^2 error of the derivative in the streamline direction is optimal, but the bound on $||c - c_h||_0$ is order $\frac{1}{2}$ less than optimal. This is the price for the improved stability property.

In the special case of piecewise linear elements, the expression $\sum_T \varepsilon \delta_T (\Delta c_h, \boldsymbol{u} \cdot \nabla w_h)_T$ vanishes and hence can be omitted from $B_{SDM}(c_h, w_h)$. The bound for δ_T can be relaxed to

$$0 < \delta_T \le \frac{r_0}{r_T^2}$$

which is independent of the discretization constant μ.

Numerical calculations have shown that the SDM does not fulfill the discrete maximum principle which states that

$$Lc(x) = 0 \;\forall\, x \in \Omega \Rightarrow \min_{x \in \Gamma}(c(x), 0) \le c(x) \le \max_{x \in \Gamma}(c(x), 0) \qquad \forall\, x \in \overline{\Omega}. \tag{3.9}$$

The maximum principle admits the following physical interpretation. Suppose that $r = 0$ and that c denotes the density of a substance where no source is present. The maximum principle states that the greatest density occurs on the boundary Γ and that the density never takes negative values. Since SDM does not possess this property, oscillations can be observed near sharp layers. A modification of the SDM that satisfies (3.9) exists, but, according to [3], the method is nonlinear, even in the constant coefficient case.

So far, the error estimates were useful for solutions where the norm $|c|_{k+1}$ is of moderate size. This norm will be large if boundary or interior layers are present in the solution. However, local error estimates can be derived. These results state that effects are propagated in the discrete problem approximately as in the continuous problem, i.e. along the characteristics. More precisely, the influence of a source in the discrete problem decays with the distance d to the source like $\exp(-Cd/h)$ in any direction with a positive component in the upwind direction $-\boldsymbol{u}(x)$ and like $\exp(-Cd/\sqrt{h})$ in directions orthogonal to the streamlines (crosswind directions). Alternatively, these results can be phrased as local error estimates of the form

$$|||c - c_h|||_{\Omega'} \leq Ch^{k+\frac{1}{2}} \left[||c||_{H^{k+1}(\Omega'')} + ||f||_{L^2(\Omega)} \right],$$

where $\Omega' \subset \Omega''$. The distance from a point $x \in \Omega'$ to $\Omega \setminus \Omega''$ is $O(h \log(1/h))$ in directions not orthogonal to $\boldsymbol{u}(x)$ and $O(\sqrt{h} \log(1/h))$ in crosswind directions. Moreover, Ω'' does not allow 'upstream cut off', i.e. all points upstream a point in Ω'' also belong to Ω''. A proof for the local error estimate can be found in [12].

The standard Galerkin method for (3.3) does not allow such a local estimate. In this case effects may propagate in crosswind or even upwind directions with little damping.

3.2.3 The Quadratic Petrov-Galerkin Finite Element Method (QPG)

In the SDM, test functions were modified by a linear term. Now, consider test functions $w(x)$ constructed by adding an asymmetric quadratic perturbation to the original piecewise linear hat function:

$$w = \phi + \nu\psi_i,$$

where ν is a real parameter, ϕ is the standard piecewise linear "hat" function, and ψ_i is piecewise quadratic. Figure 3.2 illustrates the test functions given by

$$w_i(x) = \begin{cases} \dfrac{x - x_{i-1}}{\Delta x} + \nu \dfrac{(x - x_{i-1})(x_i - x)}{(\Delta x)^2}, & x \in [x_{i-1}, x_i] \\ \dfrac{x_{i+1} - x}{\Delta x} - \nu \dfrac{(x - x_i)(x_{i+1} - x)}{(\Delta x)^2}, & x \in [x_i, x_{i+1}] \\ 0, & \text{otherwise.} \end{cases} \tag{3.10}$$

For one-dimensional two point boundary value problems with constant coefficients, one can choose ν so that the QPG method yields solutions that coincide with the exact solution at the nodal points. In accordance with [16] and [23] ν is chosen as $\nu = 3[\coth(\frac{V\Delta x}{2D}) - \frac{2D}{V\Delta x}]$. Note that $\nu = 3$ corresponds to full upwinding. However, compared to pure upwinding, QPG leads to better results in two dimensions. While a standard upwind scheme cannot eliminate the artificial dissipation in the cross flow direction, it an be confined to the

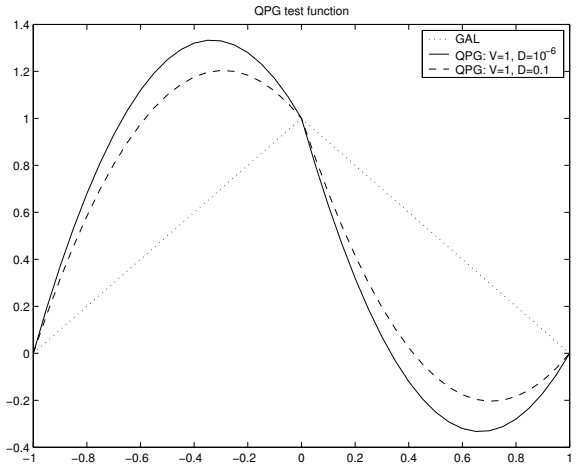


Figure 3.2: QPG test function

direction of flow with QPG provided that $\nu_x/\nu_y = \tan\theta$ where θ is the angle between flow velocity and x-axis [16]. For time-dependent problems, no value of α will guarantee a nodally exact solution.

3.3 Numerical Methods for the Unsteady Advection-Diffusion Equation

The unsteady advection-diffusion equation is given by

$$\Phi\frac{\partial c}{\partial t} + \nabla\cdot(\boldsymbol{u}c - \boldsymbol{D}\nabla c) = f, \qquad \boldsymbol{x}\in\Omega,\ \ t\in[0,T]. \tag{3.11}$$

This equation is characterized by a non-dissipative (hyperbolic) advective transport component and a dissipative (parabolic) diffusive component. Its solution typically has moving steep fronts that need to be resolved accurately.

Two general classes of numerical methods can be identified: Eulerian methods and characteristic methods. Characteristic methods are based on the Lagrangian point of view where a typical control volume is transported with the velocity field. In contrast, in the Eulerian description locally fixed control volumes are considered and the flow passes by. The differential equations arising from theses two approaches differ in the advection term which does not occur in the Lagrangian description.

3.3.1 Eulerian Methods

Eulerian methods carry out the temporal discretization in the time direction. They require small time steps, either for reasons of stability (for explicit methods) or accuracy (for implicit methods) because the time truncation error depends on high-order time derivatives of the solution which are large when a sharp front passes by.

3.3.1.1 Method of Lines

This method is also referred to as semi-discrete method. First, a Galerkin finite element method is applied with respect to the space variable.

$$\int_\Omega \Phi \dot{c} w d\boldsymbol{x} + \int_\Omega (\boldsymbol{D}\nabla c)\cdot\nabla w d\boldsymbol{x} - \int_\Omega \boldsymbol{u}c\cdot\nabla w d\boldsymbol{x} + \int_{\partial\Omega}(\boldsymbol{u}c - \boldsymbol{D}\nabla c)\cdot\boldsymbol{n} w d\boldsymbol{x} = \int_\Omega f w d\boldsymbol{x}$$

Discretization results in a system of ordinary differential equations with respect to time:

$$M\dot{c} + Bc + Ac = F,$$

where M is the mass matrix, A is the matrix that corresponds to the discrete diffusion operator, and B is the matrix associated with advection.

Standard ODE solvers can be applied to solve these equations. The numerical examples given in Chapter 5 are based on the scheme

$$\frac{Mc^n - Mc^{n-1}}{\Delta t} + \lambda(Bc^n + Ac^n) + (1-\lambda)(Bc^{n-1} + Ac^{n-1}) = \lambda F^n + (1-\lambda)F^{n-1}.$$

$\lambda \in [0,1]$ is a weighting parameter between time levels t^{n-1} and t^n. $\lambda = 1$ corresponds to backward Euler, and $\lambda = 0.5$ yields the Crank-Nicolson scheme. Both schemes are implicit. $\lambda = 0$ yields forward Euler, which is an explicit method. Fully implicit methods allow large time steps but the temporal and the spatial error add together. Explicit methods require small time steps for stability but temporal and spatial error cancel each other in that they have opposite signs. However, space and time step have to be reduced simultaneously which increases computational and storage costs. This can be explained by considering the following model problem:

$$c_t + Vc_x = 0, \qquad V > 0 \text{ constant.}$$

Discretization with linear finite elements results in a scheme which is similar to a finite difference approximation. Some conclusions can be derived from the local truncation error of different schemes which are presented in Table 3.1 (spatial discretization with and without upwinding, combined with forward Euler (FE), backward Euler (BE) or Crank-Nicolson (CN) time discretization). $Cr := \frac{V\Delta t}{\Delta x}$ denotes the Courant number. It denotes the number of elements over which the information is transported during one time step. For a fixed spatial grid, the upwinded explicit scheme becomes second order accurate if $Cr = 1$. Smaller or larger time steps will not lead to further improvement of the solution. In contrast, the simple explicit scheme will produce more accurate solutions with finer time steps. The same holds for the implicit schemes. It should also be noted that the Crank-Nicolson scheme is second order accurate in time while the Euler schemes are only first order accurate.

The advective and diffusive part of the equation behave quite differently. Some methods overcome these difficulties by using an operator splitting approach where the advective part is treated explicitly and the diffusive part is treated implicitly. In the following, one-dimensional examples introduced in [18] are considered where the two parts are discussed separately in order to demonstrate the numerical difficulties arising in the solution of unsteady advection-diffusion equations.

λ	scheme	local truncation error
scheme without upwinding: $\frac{c_i^n - c_i^{n-1}}{\Delta t} + \lambda V \frac{c_{i+1}^n - c_{i-1}^n}{2\Delta x} + (1-\lambda) V \frac{c_{i+1}^{n-1} - c_{i-1}^{n-1}}{2\Delta x} = 0$		
0	FE	$\frac{\Delta t}{2} \frac{\partial^2 c}{\partial t^2}(x_i, t_{n-1}) + \mathcal{O}((\Delta x)^2 + (\Delta t)^2)$
$\frac{1}{2}$	CN	$-\frac{\Delta t^2}{4} \frac{\partial^3 c}{\partial t^3}(x_i, t_n) + \mathcal{O}((\Delta x)^2 + (\Delta t)^2)$
1	BE	$-\frac{\Delta t}{2} \frac{\partial^2 c}{\partial t^2}(x_i, t_n) + \mathcal{O}((\Delta x)^2 + (\Delta t)^2)$
scheme with upwinding: $\frac{c_i^n - c_i^{n-1}}{\Delta t} + \lambda V \frac{c_i^n - c_{i-1}^n}{\Delta x} + (1-\lambda) V \frac{c_i^{n-1} - c_{i-1}^{n-1}}{\Delta x} = 0$		
0	FE	$\frac{1}{2} V \Delta x (Cr - 1) \frac{\partial^2 c}{\partial x^2}(x_i, t_{n-1}) + \mathcal{O}((\Delta x)^2 + (\Delta t)^2)$
1	BE	$-\frac{1}{2} V \Delta x (Cr + 1) \frac{\partial^2 c}{\partial x^2}(x_i, t_n) + \mathcal{O}((\Delta x)^2 + (\Delta t)^2)$

Table 3.1: Local truncation errors

One-dimensional Examples

Unsteady Diffusion. Consider the unsteady diffusion equation

$$c_t - Dc_{xx} = f, \quad \text{in } (\Omega = (0,1)) \times [0,T],$$

with boundary conditions

$$c(0,t) = c(1,t) = 0$$

and initial condition

$$c(x,0) = c_0(x).$$

Define the function spaces

$$X = H_0^1(\Omega),$$

and

$$Y(X) = \{v | v(\cdot,t) \in X \ \forall t \in [0,T], \int_0^T ||v||^2_{H^1(\Omega)} dt < \infty\} = L^2\left([0,T], H_0^1(\Omega)\right).$$

The weak form can be expressed as: Find $c \in Y(X)$ such that

$$(c_t, v) + a(c,v) = L(v), \qquad \forall\, v \in X,$$

where

$$a(w,v) = \int_0^1 D w_x v_x dx, \qquad \forall\, w, v \in X,$$

$$(w,v) = \int_0^1 wv dx, \qquad \forall\, w, v \in X.$$

The semi-discrete formulation is obtained by first discretizing in space: Find $c_h \in Y(X_h)$ such that

$$(\dot{c}_h, v) + a(c_h, v) = L(v), \qquad \forall\, v \in X_h,$$

with $X_h \subset X$ and $\dim(X_h) = N < \infty$.
Using a nodal basis for $X_h = \text{span}\{\varphi_1, \ldots, \varphi_N\}$, one can set

$$c_h(x,t) = \sum_{j=1}^{N} c_{hj}(t)\varphi_j(x).$$

The choice of test functions $v(x) = \varphi_i(x)$ results in the system of ordinary differential equations

$$M_h \dot{c}_h = -A_h c_h + F_h.$$

The next step is to discretize the system in time. This is usually done by finite difference approximations. For example, a backward Euler discretization results in the following system of algebraic equations

$$M_h \left(\frac{c_h^{n+1} - c_h^n}{\Delta t} \right) = -A_h c_h^{n+1} + F_h^{n+1}.$$

c_h^n denotes c_h at time $t^n = n\Delta t$ with time step Δt. Rearranging of terms leads to

$$(A_h + \frac{1}{\Delta t} M_h) c_h^{n+1} = \frac{1}{\Delta t} M_h c_h^n + F_h^{n+1}.$$

This system of equations has to be solved at every time step. One should note that the matrix $A_h + \frac{1}{\Delta t} M_h$ is symmetric and positive definite. Backward Euler represents an implicit scheme which is stable for all choices of Δt. However, because it is a first order scheme, accuracy is reduced severely when the time step is increased. Higher order methods are recommended, for example a 3rd order Adams Bashforth scheme, in order to obtain good approximations even with larger time steps.

Using a forward Euler temporal discretization scheme results in

$$M_h \left(\frac{c_h^{n+1} - c_h^n}{\Delta t} \right) = -A_h c_h^n + F_h^n$$

which can be rewritten as

$$c_h^{n+1} = [I - \Delta t M_h^{-1} A_h] c_h^n + \Delta t M_h^{-1} F_h^n.$$

Although the Euler forward scheme is explicit, one still needs to solve a system of linear equations with M_h at each time step. However, this matrix is well conditioned. Furthermore, the time step has to be chosen small enough for stability. The numerical solution is stable if and only if the eigenvalues of the matrix $[I - \Delta t M_h^{-1} A_h]$ lie within the unit circle ($|\sigma| \leq 1$). Thus, the stability restriction becomes

$$\Delta t < \frac{2}{\lambda_{\max}(M_h^{-1} A_h)}.$$

For linear finite elements there holds

$$\lambda_{\max}(M_h^{-1} A_h) \sim \mathcal{O}(h^{-2})$$

which enforces the restriction

$$\Delta t \leq \mathcal{O}(h^2).$$

Due to the severe time step restriction associated with explicit methods, implicit schemes are preferred in practice. It should be noted that the accuracy of the temporal scheme and of the spatial FEM discretization should be compatible. The accuracy of the solution also depends on how good the initial condition c_{h0} can be represented as the interpolant $I_h c_0(x)$ with respect to the finite element nodes.

Unsteady Advection. Consider the one-dimensional unsteady advection equation

$$c_t + V c_x = 0, \qquad \text{in } (\Omega = (0, L)) \times [0, T],$$

with a periodic boundary condition

$$c(0, t) = c(L, t),$$

and an initial condition

$$c(x,0) = c_0(x).$$

With the function spaces

$$X = H^1_{\#}(\Omega) = \{v \in H^1(\Omega) | v(0) = v(L)\}, \qquad \text{and}$$
$$Y(X) = L^2\left([0,T]; H^1(\Omega)\right)$$

the weak formulation can be expressed as: Find $c \in Y(X)$ such that

$$(c_t, v) + b(c,v) = 0, \qquad \forall\, v \in X$$

where

$$b(w,v) = \int_0^1 V w_x v dx.$$

For the model problem with periodic Dirichlet boundary conditions, the bilinear form b is skew-symmetric, $b(w,v) = -b(v,w)$. It follows that $b(c,c) = 0$ and hence

$$\frac{d}{dt}(c,c) = 0,$$

which is the same as saying $\frac{d}{dt}||c||^2_{L^2(\Omega)} = 0$. This can be interpreted as an energy conservation property

$$E(t) = ||c(\cdot,t)||^2_{L^2(\Omega)} = \text{constant}.$$

In the absence of any temporal discretization error, the quantity $||c||^2_{L^2(\Omega)}$ is conserved. No energy is leaving or entering the system. In the presence of temporal errors, the energy will decrease with time. The system of fully discrete equations is then said to be dissipative.

The solution of the model problem is of the functional form

$$c(x,t) = g(x - Vt).$$

The initial condition will just propagate at a constant speed V without any change in shape. This is a consequence of the fact that there is no dispersion.

The eigenvalues λ_n and eigenfunctions Ψ_n of the one-dimensional convection operator associated with the eigenvalue problem $V\Psi_x = \lambda\Psi,\ \Psi(0) = \Psi(L)$ are given by

$$\Psi_n(x) = \exp(ik_n x), \quad \lambda_n = iVk_n,$$

where $k_n = \frac{2\pi n}{L}$ and $n \in \mathbb{N}$ is the wave number. The solution can be expanded in terms of eigenfunctions

$$c(x,t) = \sum_{n=-\infty}^{\infty} a_n(0) \exp(ik_n(x - Vt)),$$

where $a_n(0)$ are the coefficients from the expansion of $c_0(x)$.

A first discretization in space leads to the semi-discrete formulation

$$M_h \dot{c}_h + B_h c_h = 0$$

which can be rewritten as the system of ordinary differential equations

$$\dot{c}_h = -M_h^{-1} B_h c_h.$$

The discrete eigenvalues of the matrix are given by

$$\lambda_h (M_h^{-1} B_h)_n = i V_n k_n$$

with a discrete velocity V_n which is wave number dependent. It can be shown that the lower half of the spectrum for the discrete problem agrees well with the corresponding spectrum for the continuous problem. On the other hand, the eigenvalues to the highest wave numbers agree poorly. This results in dispersion errors in the time dependent problem. The eigenvectors for this particular model problem are the same as the finite continuous eigenfunctions sampled at the finite element nodes. Hence, the basis coefficients of the numerical solution can be written as

$$c_{hj}(t) = c_h(x_j, t) = \sum_{n \leq N/2} a_n(0) \exp(i k_n (x_j - V_n t)).$$

Unlike the exact solution, different wave numbers in the initial condition are propagated at different speeds V_n. The numerical scheme is called dispersive. This is independent of the choice of the temporal scheme. If in the initial condition most of the energy is located in the lower part of the discrete spectrum, there will be small dispersion errors. If that is not the case, oscillations and damping will occur in the numerical solution.

With an implicit backward Euler scheme one has to solve

$$(B_h + \frac{1}{\Delta t} M_h) c_h^{n+1} = \frac{1}{\Delta t} M_h c_h^n$$

at each time step which is a non-symmetric system of algebraic equations. On the other hand, one has to be very careful if an explicit scheme is used for time integration. Since the eigenvalues of $M_h^{-1} B_h$ are purely imaginary, the stability region of the numerical scheme must enclose parts of the imaginary axis. Therefore, the explicit forward Euler scheme is useless for this pure advection model problem because its stability region is the unit circle with center -1 and radius 1. The stability condition is often written as

$$Cr = \frac{V \Delta t}{h} < C \tag{3.12}$$

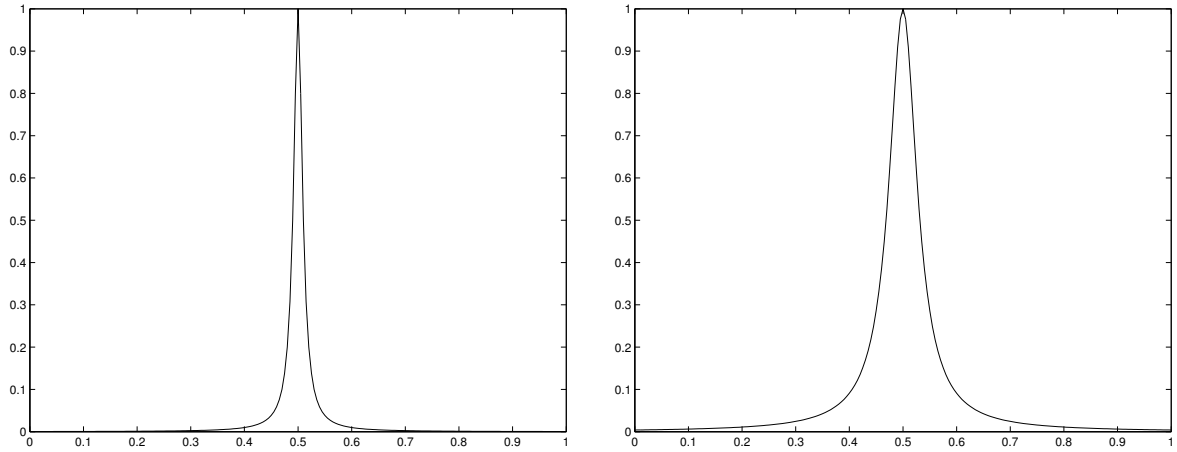
and it is referred to as CFL-condition (Courant-Friedrich-Levy).

Figures 3.3 to 3.5 present the numerical solution of the model problem with initial condition

$$c_0(x) = \frac{1}{1 + a(x - x_0)^2}$$

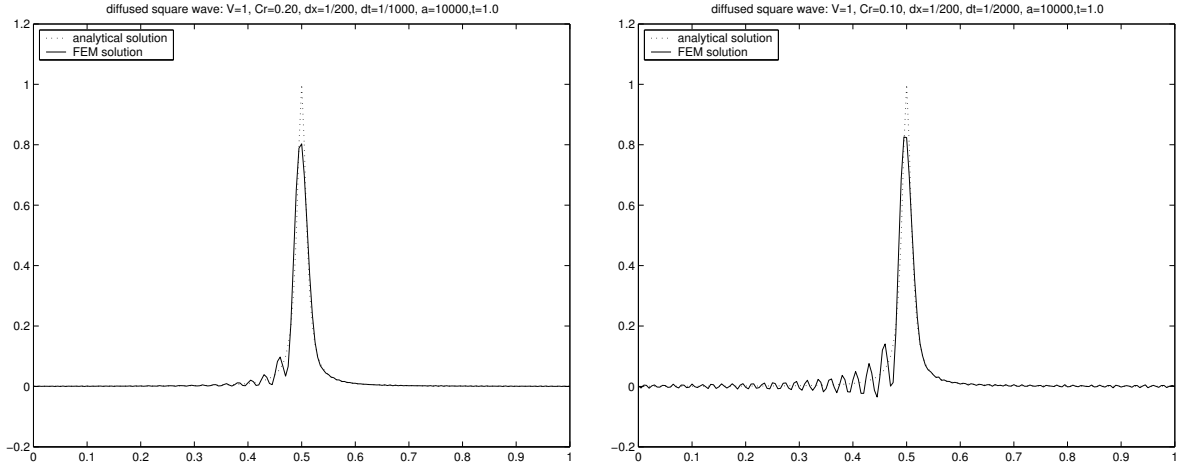
and constant velocity $V = 1$. A third order Adams Bashforth scheme is used for the temporal discretization which is only slightly dissipative.

Larger time steps cause higher temporal errors but introduce more artificial diffusion which in turn will dampen out the high wave number signals and reduce oscillations. Smaller time steps improve the energy conservation property but dispersion errors are



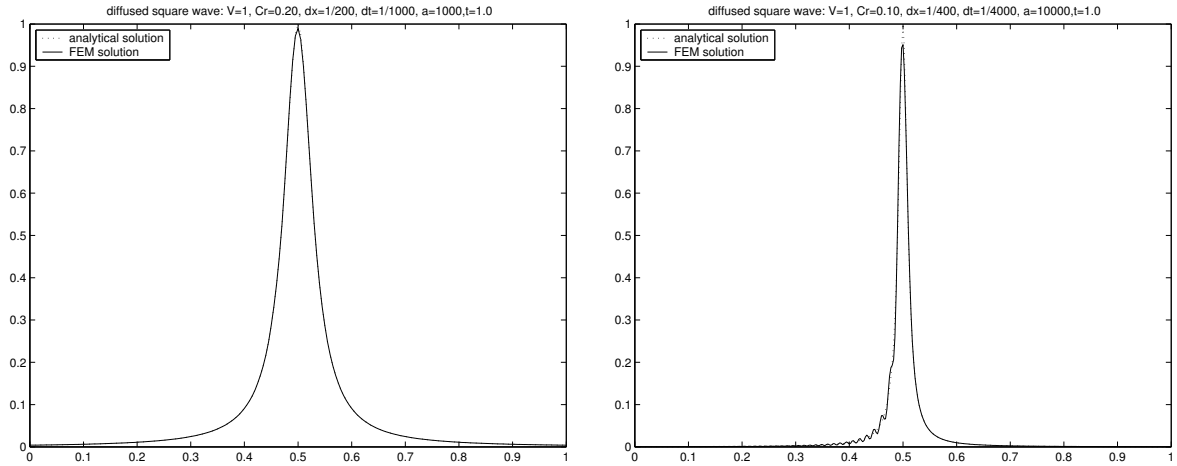
$a = 10000$ $a = 1000$

Figure 3.3: Initial data for the 1D advection equation



$Cr = 0.2$ $Cr = 0.1$

Figure 3.4: FE-solution for $a = 10^5$ with two different time steps



solution for $a = 1000$ finer grid, $a = 10000$

Figure 3.5: FE-solution on a finer grid and for a less severe initial condition

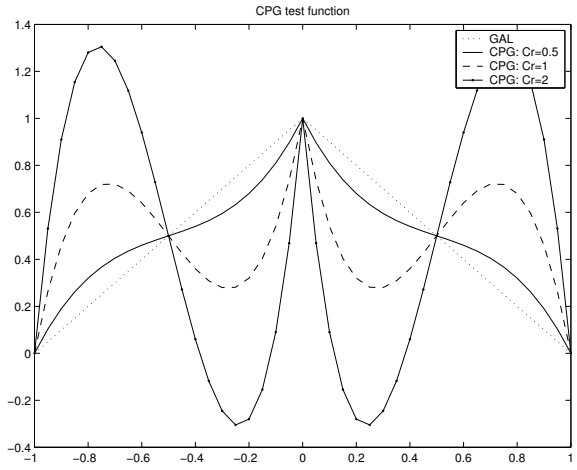


Figure 3.6: CPG test function

more noticeable. The numerical solution can be improved by using a finer grid, which pushes the most energetic modes towards the lower half of the spectrum. Furthermore, a less severe initial condition with a smaller gradient leads to better numerical results.

The particular tests and numerical results obtained here illustrate some of the issues related to solving time-dependent advection problems.

The Cubic Petrov-Galerkin Finite Element Method (CPG) If one tries to use quadratically modified test functions for transient problems as in the QPG-method mentioned above, low order truncation error terms are introduced which successfully suppress oscillations but the general solution accuracy degrades. In order to improve the solution to the time dependent problem, a symmetric cubic perturbation is added to the original piecewise linear hat functions. Figure 3.6 shows the test functions for different values of the parameter γ which are given by

$$w_i(x) = \begin{cases} \dfrac{x - x_{i-1}}{\Delta x} + \gamma \dfrac{(x - x_{i-1})(x_i - x)(x_{i-1} + x_i - 2x)}{(\Delta x)^3}, & x \in [x_{i-1}, x_i] \\ \dfrac{x_{i+1} - x}{\Delta x} - \gamma \dfrac{(x - x_i)(x_{i+1} - x)(x_i + x_{i+1} - 2x)}{(\Delta x)^3}, & x \in [x_i, x_{i+1}] \\ 0, & \text{otherwise.} \end{cases} \tag{3.13}$$

Since the weighting functions are symmetric, they lead to flow direction invariant upwind coefficients for the one-dimensional problem. The difference between QPG and CPG becomes more obvious when combined quadratic and cubic weighting functions of the form $w = \psi \pm \nu F_q(x) \pm \gamma F_c(x)$ are introduced. The elemental matrices can be found in the appendix. The time discretization is implemented through the use of a Crank-Nicolson scheme. The truncation error $\mathcal{T}$ is computed by a Taylor series expansion of the nodal difference equation around C_j^n, using the underlying differential equation

$c_t + uc_x - Dc_{xx} = 0$ and its derivatives [27]. This results in

$$\begin{aligned}\mathcal{T} &= h^2\left[(2Cr^2 - \frac{2}{5}\gamma) - \frac{1}{\alpha}4\nu\right]\frac{u}{24}\frac{\partial^3 C_j^n}{\partial x^3} \\ &+h^3\left[-Cr(Cr^2 - \frac{\gamma}{5}) + \frac{\nu}{3}(1-Cr^2) + \frac{1}{\alpha}(2 - 6Cr^2 + \frac{2\gamma}{5} + 2\nu Cr)\right]\frac{u}{24}\frac{\partial^4 C_j^n}{\partial x^4} \\ &+h^4\Bigg[(-\frac{2}{15} + \frac{1}{3}Cr^2 + \frac{3}{10}Cr^4 - \frac{1}{30}\gamma - \frac{1}{15}Cr^2\gamma) - \nu\frac{Cr}{6}(1-Cr^2) \\ &+\frac{1}{\alpha}(-Cr + 3Cr^2 - \frac{2}{5}Cr\gamma - \frac{2}{3}\nu) + \frac{1}{\alpha^2}(6Cr^2 - 2\nu Cr)\Bigg]\frac{u}{24}\frac{\partial^5 C_j^n}{\partial x^5} + \mathcal{O}(h^5)\end{aligned}$$

where $h = \Delta x$, $\alpha = uh/D$ is the grid Peclet number and $Cr = u\Delta t/h$ the Courant number.

In agreement with [27] and [23], the parameter γ is chosen as $\gamma = 5Cr^2$. This would eliminate the leading order truncation term for pure advection ($D = 0,\ \alpha = \infty$) or if $\nu = 0$. For $D = 0$ this also forces the non-ν-dependent portion of the $\mathcal{O}(h^3)$ term to vanish. For Courant numbers other than unity, on can eliminate that term by selecting $\nu = 0$. Then the solution is $\mathcal{O}(h^4)$ accurate. $Cr = 1$ yields a solution which is at least $\mathcal{O}(h^5)$. In general, the above choice of γ will considerably decrease the magnitude of the coefficient in the $\mathcal{O}(h^2)$ and $\mathcal{O}(h^3)$ term, especially for large values of α. More properties of the quadratic and cubic upwind schemes can be found by Fourier analysis. Westerink and Shea [27] have shown that both QPG and CPG can improve phase behavior for advection dominated problems, but only CPG does not introduce excessive artificial damping. In sum, CPG improves both spatial and temporal accuracy, and the highest order of consistency is obtained at $Cr = 1$.

Dispersion Analysis Dispersion analysis can be a useful tool to predict whether there will be oscillations or damping in the numerical solution. All grid based numerical approximations are dispersive. The simplest case of dispersion is when an initial wave form is placed on the grid, and the pure advection solution is sought. After a certain time it will break up into a trail of wiggles. This case will be examined now, following the analysis given in [6].

Consider the advection equation

$$T_t + \boldsymbol{u} \cdot \nabla T = 0.$$

The following variables will be needed.

- wave vector $\boldsymbol{k}$, direction of the wave, normal to the lines of constant phase
- wave number $k = |\boldsymbol{k}|$
- velocity $\boldsymbol{u}$
- angular frequency ω of a plane wave $T = a\cos(\boldsymbol{k} \cdot \boldsymbol{x} - \omega t)$
- phase velocity $\boldsymbol{c} = \omega \boldsymbol{k}/k^2$

- phase speed $c = |\boldsymbol{c}| = \omega/k$
- group velocity $\boldsymbol{G}(k) = \nabla_k \omega$

For the simple advection equation one obtains $\omega = \boldsymbol{u} \cdot \boldsymbol{k}$ and $\boldsymbol{G} = \boldsymbol{u}$.

With any numerical method both phase speed and group velocity will differ from those above and will exhibit dispersion by not being independent of the wavelength. In contrast to other methods, e.g. truncation error analysis, dispersion analysis does not require $h \to 0$ or $(h/\lambda \to 0)$ which converges slowly for short waves. This can be considered as an advantage of dispersion analysis. For simplicity consider the one-dimensional advection equation

$$T_t + uT_x = 0, \qquad \text{on} \quad 0 \le x \le L = 1$$

with constant u and periodic boundary conditions

$$T(0,t) = T(L,t).$$

As initial condition take a single Fourier mode

$$T(x,0) = T_0(x) = e^{ikx}.$$

The exact solution is

$$T(x,t) = T_0(x-ut) = e^{ik(x-ut)} = e^{i(kx-\omega_c t)}.$$

Due to the periodic boundary conditions, the allowable wave numbers are

$$k = k_n = 2\pi n, \qquad n = 0, 1, 2, \ldots$$

Hence the general solution of the discrete system of ODEs is sought in the form

$$T(x_j,t) = T(jh,t) := T_j(t) = e^{i(kjh-\omega t)}. \tag{3.14}$$

The discrete periodic boundary conditions are

$$T_0 = T_N \qquad \text{and} \qquad T_1 = T_{N+1}.$$

ω is to be determined by inserting (3.14) into the specific discrete system and is hopefully close to $2\pi nu$. Introduce $\theta = kh$ as a dimensionless wave number. Table 3.2 lists the continuous and discrete equations together with the values for the frequency for the standard Galerkin method, the simple upwind scheme and a streamline upwind Petrov Galerkin scheme. Figure 3.8 illustrates the phase speed relative to u for different schemes.

First consider the standard GFEM. The approximate frequency ω is only a good approximation to the true frequency $\omega_c = uk$ for "small" values of $\theta = \theta_n = k_n h = 2\pi n/N$, $n = 1, 2, \ldots, N$, see Figure 3.7. The upper half of the spectrum ($n > N/2$) corresponds to waves with wavelength $\lambda = 2\pi/k$ shorter than $2h$. Each of these waves is connected to the lower half of the spectrum by $T_j^{N-n}(t) = \overline{T}_j^n(t)$. Thus one can focus on the resolvable lower half of the spectrum. One refers to "short" waves as waves with wave numbers k_n between $N\pi/2$ and $N\pi$ (wavelengths between $2h$ and $4h$). For these waves the phase speed $c_n = \omega/k_n$ agrees poorly with the true phase speed. This will cause upstream moving

GFEM	continuous weak formulation	$(T_t, w) + u(T_x, w) = 0$
	discrete eq.	$\frac{h}{6}(\dot{T}_{j-1} + 4\dot{T}_j + \dot{T}_{j+1}) + \frac{u}{2}(T_{j+1} - T_{j-1}) = 0$
	ω	$uk\frac{\sin\theta}{\theta}\frac{3}{2+\cos\theta}$
	G	$u\frac{1+2\cos\theta}{2+\cos\theta}\frac{3}{2+\cos\theta}$
pure upwinding	continuous weak formulation	$(T_t, w) + u(T_x, w) - \frac{uh}{2}(T_{xx}, w) = 0$
	discrete eq.	$\frac{h}{6}(\dot{T}_{j-1} + 4\dot{T}_j + \dot{T}_{j+1}) + \frac{u}{2}(T_{j+1} - T_{j-1}) - \frac{uh}{2}\frac{T_{j-1} - 2T_j + T_{j+1}}{h} = 0$
	ω	$\frac{uk}{\theta}[\sin\theta - i(1-\cos\theta)]\frac{3}{2+\cos\theta}$
SUPG	continuous weak formulation	$(T_t, w + \beta h w_x) + u(T_x, w + \beta h w_x) = 0$
	discrete eq.	$\frac{h}{6}((1+3\beta)\dot{T}_{j-1} + 4\dot{T}_j + (1-3\beta)\dot{T}_{j+1}) + \frac{u}{2}(T_{j+1} - T_{j-1}) - \beta uh\frac{T_{j-1} - 2T_j + T_{j+1}}{h} = 0$
	ω^*	$\frac{uk\frac{\sin\theta}{\theta}\left[\frac{2+\cos\theta}{3} + 2\beta^2(1-\cos\theta)\right]}{\left(\frac{2+\cos\theta}{3}\right)^2 + \beta^2\sin^2\theta}$

$*$frequency corresponding to (3.15)

Table 3.2: Results of dispersion analysis for the 1D advection equation with constant coefficients

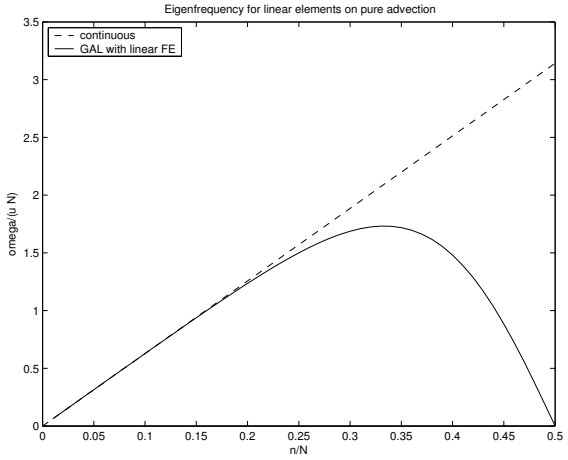


Figure 3.7: Approximate frequency of the Galerkin method with linear elements

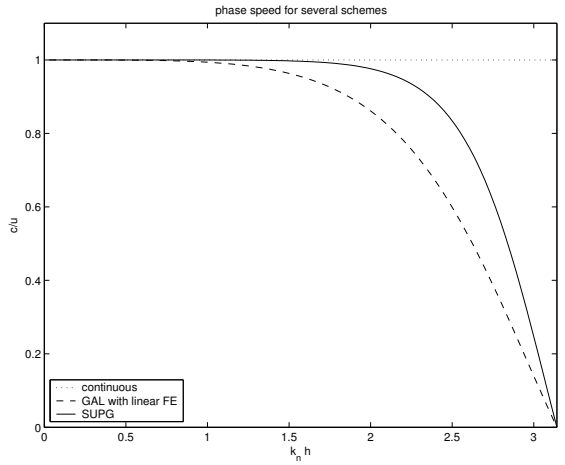


Figure 3.8: Phase speed for different schemes

wiggles, i.e. the linear combination of all modes, each of which is moving to the right, causes the resulting total wave form to display leftward moving wiggles. This is indicated by the group velocity which is smaller than the continuous one and even negative for large values of θ.

Regarding the upwind scheme, an imaginary part appears in the frequency. Hence the discrete solution (3.14) could be rewritten as

$$T_j(t) = e^{-[ut(1-\cos\theta)/h][3/(2+\cos\theta)]} e^{ik(x_j-ct)} \xrightarrow{\theta\to 0} e^{-k^2hut/2} e^{ik(x_j-ct)}$$

with the same c as in the GFEM. Now there is also dissipation of the wave with diffusivity $\kappa = uh/2$. The upwinded advection equation reacts as if it were solving the advection-diffusion equation $T_t + uT_x - \kappa T_{xx} = 0$ with analytic solution $T(x,t) = e^{-k^2\kappa t} e^{ik(x-ut)}$. On the other hand, short waves ($\theta \to \pi$) are damped faster than others. Consequently, wiggles are suppressed.

The frequency for the SUPG method was derived directly for the solution of the form

$$T_j(t) = e^{-k^2\kappa t} e^{ik(x_j-ct)}, \tag{3.15}$$

where

$$\kappa = \frac{2\beta u}{k^2 h} \frac{(1-\cos\theta)\left(\dfrac{2+\cos\theta}{3}\right) - \sin^2\theta}{\left(\dfrac{2+\cos\theta}{3}\right)^2 + \beta^2 \sin^2\theta}.$$

For $\theta \to 0$ the phase speed becomes

$$c = u\left[1 - \left(\frac{1}{180} - \frac{\beta^2}{12}\right)\theta^4 + \mathcal{O}(\theta^6)\right],$$

which is sixth-order accurate if $\beta = 1/\sqrt{15}$. For this choice the diffusivity becomes

$$\kappa = \frac{\sqrt{15}}{180} u k^2 h^3 + \mathcal{O}(h^5),$$

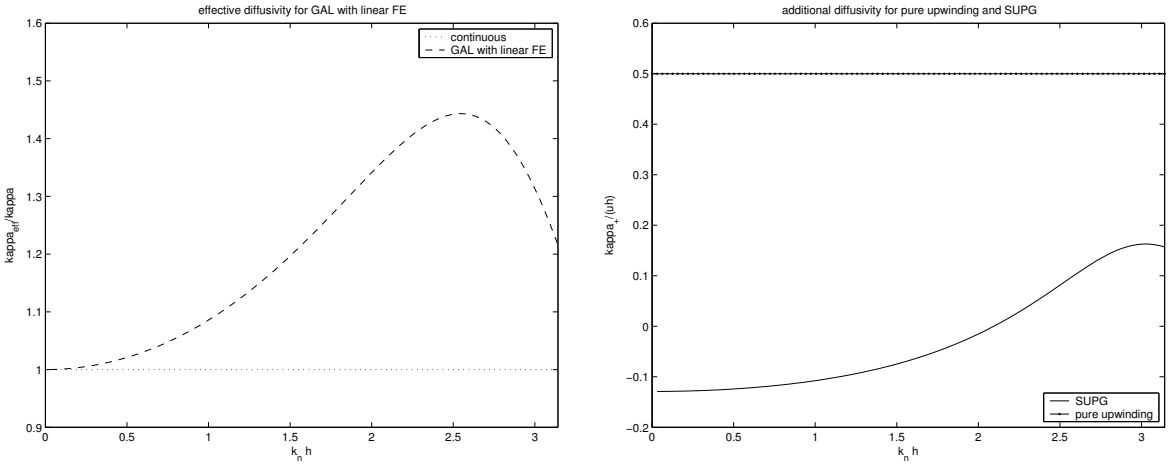


Figure 3.9: Effective diffusivity for GAL and additional diffusivity with pure upwinding and SUPG

which is actually a higher order diffusive term. Hence, SUPG is a big improvement over simple upwinding. Finally note that β is related to the parameter δ used in the SUPG-method described above by $\delta = h\beta/u$.

The results achieved so far can be generalized to the advection-diffusion equation with constant coefficients

$$T_t + uT_x - \kappa T_{xx} = 0.$$

A solution is sought of the form (3.15). GFEM results in the discrete equation

$$\frac{h}{6}(\dot{T}_{j-1} + 4\dot{T}_j + \dot{T}_{j+1}) + \frac{u}{2}(T_{j+1} - T_{j-1}) - \kappa\frac{T_{j-1} - 2T_j + T_{j+1}}{h} = 0,$$

which gives the solution

$$T_j(t) = e^{-k^2\kappa_{eff}t}e^{i(kjh-\omega t)},$$

with ω given in Table 3.2 and

$$\kappa_{eff} = 2\kappa\frac{(1-\cos\theta)}{\theta^2}\frac{3}{2+\cos\theta}.$$

Hence, the advective part of the solution is unchanged compared to the pure advection equation, while the solution is over-damped, especially for short waves which can be considered as an advantage because it removes wiggles. In contrast, with simple upwinding all diffusion is numerical for large Peclet numbers. This is also true for general boundary conditions and higher dimensions. Therefore, simple upwinding should not be used for the advection-diffusion equation. Figure 3.9 illustrates the effective diffusivity of the standard Galerkin method and the additional diffusivity κ_+ with $\kappa_{eff} = \kappa + \kappa_+$ of pure upwinding and SUPG with $\beta = 1/\sqrt{15}$ or $\delta = h/(\sqrt{15}u)$ respectively.

In two dimensions, the dispersion error extends to directional error as well as translational error. Dispersion analysis can be used to examine whether there will occur cross-directional "pollution" factors in the phase speed and group velocity. However, calculations are more difficult than in one dimension. Further results for general boundary conditions and other schemes, e.g. lumped mass, can be found in [6].

3.3.1.2 Discontinuous Galerkin Methods (DG)

Discontinuous Galerkin methods were originally developed to solve ordinary differential equations as in [21]. For partial differential equations, a Galerkin method applied to the space variable results in a system of ordinary differential equations. In contrast to the method of lines where finite difference methods were applied to the resulting ODE, in the DG approach a Galerkin method is also applied to the time variable. The time and space variables are thus treated similarly. The trial and test functions are defined as tensor products of the temporal and spatial test functions. They are continuous in space, but discontinuous in time at time discretization points t^n. Therefore, it suffices to consider the equations in a time interval $J_n = (t^{n-1}, t^n]$. The numerical scheme becomes

$$\int_{t_{n-1}}^{t_n} \int_\Omega \left[\dot{c} + \nabla \cdot (\boldsymbol{u}c - \boldsymbol{D}\nabla c)\right] w d\boldsymbol{x} dt + \int_\Omega c_{n-1}^+ w_{n-1}^+ d\boldsymbol{x} = \int_{t_{n-1}}^{t_n} \int_\Omega f w d\boldsymbol{x} dt + \int_\Omega c_{n-1}^- w_{n-1}^+ d\boldsymbol{x} \tag{3.16}$$

with $w_{n-1}^+ = \lim_{t \to t_{n-1}, t > t_{n-1}} w(\boldsymbol{x}, t)$ and $c_{n-1}^- = \lim_{t \to t_{n-1}, t < t_{n-1}} c(\boldsymbol{x}, t)$.

Continuity is enforced weakly. The discontinuous Galerkin method is implicit. It couples the unknown nodal values at the beginning and the end of the time interval with the nodal values at the end of the previous time interval.

Introduce the following notations:

$$(c, w) = \int_\Omega c w d\boldsymbol{x}$$

for the L^2 inner product on Ω, and

$$B(c, w) = \int_\Omega \nabla \cdot (\boldsymbol{u}c - \boldsymbol{D}\nabla c) w d\boldsymbol{x}$$

as a bilinear form. Assume homogeneous Dirichlet boundary conditions. Then

$$B(c, w) = ((\nabla \cdot \boldsymbol{u})c, w) + (\boldsymbol{u} \cdot \nabla c, w) + (\boldsymbol{D}\nabla c, \nabla w).$$

Now, (3.16) can be rewritten as the discrete scheme

$$\int_{J_n} ((\cdot c, w) + B(c, w)) dt + (c_+^{n-1}, w_+^{n-1}) = (c_-^{n-1}, w_+^{n-1}) + \int_{J_n} (f, w) dt \tag{3.17}$$

with initial value $c^0 = s$. This can also be considered as an approximating scheme for the underlying ODE

$$\dot{c} + Ac = f, \qquad \text{for } t > 0, \text{ with } c(0) = s, \tag{3.18}$$

where $B(c, w) = (Ac, w)$.

Assume that $B(c, w)$ is coercive. Then the local problem (3.17) has a unique solution on J_n because the corresponding homogeneous equation only has the trivial solution $c \equiv 0$.

In the following, some a priori error bounds are given. Let the approximating functions be piecewise polynomials of degree $q - 1$ in time. Here and below $k_n = t_n - t_{n-1}$, $c^{(q)} = (d/dt)^q c$, and $|| \cdot || = || \cdot ||_{L^2}$. The numerical solution is denoted by c_k while c denotes the exact solution. Then the following error estimate holds, see [21].

Theorem 3.2 *For the solutions of (3.17) (with $q \geq 1$) and (3.18) there holds*

$$||c_k^N - c(t_N)|| \leq C \left(\sum_{n=1}^{N} k_n^{2q} \int_{J_n} |c^{(q)}|_1^2 dt \right)^{\frac{1}{2}}, \text{ for } t_N \geq 0.$$

This error estimate concerns only nodal values, but estimates of the same optimal order may be derived in the interior of the intervals J_n. Here and below, $||\varphi||_{J_n} = \sup_{t \in J_n} ||\varphi(t)||$.

Theorem 3.3 *For the solutions of (3.17) (with $q \geq 1$) and (3.18) there holds*

$$||c_k - c||_{J_n} \leq ||c_k^n - c(t_n)|| + C||c_k^{n-1} - c(t_{n-1})|| + Ck_n^q ||c^{(q)}||_{J_n}.$$

The error bound at the nodal points can be improved for $q \geq 2$. It is actually of order $O(k^{2q-1})$.

Theorem 3.4 *For the solutions of (3.17) (with $q \geq 2$) and (3.18) there holds*

$$||c_k^N - c(t_N)|| \leq Ck^{q-1} \left(\sum_{n=1}^{N} k_n^{2q} \int_{J_n} |c^{(q)}|_{2q-1}^2 dt \right)^{\frac{1}{2}}, \text{ for } t_N \geq 0,$$

where $k = \max_n k_n$.

However, in applications to partial differential equations, many assumptions are required to force the solution $c^{(q)}(t)$ to be in H^{2q-1} for $t > 0$. Hence, the above estimate is not always possible.

The L^2-type bound in time of the error bound can be replaced by a maximum-norm.

Theorem 3.5 *Assume that $k_{n+1}/k_n \geq d > 0$ for $n \geq 0$. Then, for the solutions of (3.17) (with $q \geq 1$) and (3.18) there holds*

$$||c_k - c||_{J_n} \leq CL_n \max_{n \leq N} (k_n^q ||c^{(q)}||_{J_n}), \text{ where } L_N = (\log \frac{t_N}{k_N})^{\frac{1}{2}} + 1.$$

This result suggests, e.g. that to keep the error uniformly small, one should choose the time steps inversely proportional to $||c^{(q)}||_{J_n}^{1/q}$. For $q = 1$, a posteriori error bounds can be found in [21], as well as applications to the heat equation in a bounded domain.

In sum, the discontinuous Galerkin method provides a very flexible tool to obtain schemes which have a good accuracy with respect to time. However, introduction of the time t as a new variable in the numerical scheme increases the dimension by one and doubles the number of unknowns compared to standard Eulerian or characteristic methods.

The Streamline Diffusion Method (SDM) The SDM can also be applied to time dependent problems. It can be considered as a combination of the streamline diffusion method in space and a discontinuous Galerkin procedure in time. The SDM uses continuous and piecewise polynomial trial and test functions in space as in standard FEM but a discontinuous Galerkin approximation in time. It adds numerical diffusion only in the direction of characteristics (streamlines) to suppress the oscillations and does not introduce any crosswind diffusion. Therefore, this numerical method possesses many advantages other Eulerian methods do not have. However, this method contains an undetermined parameter in the test functions that needs to be chosen very carefully to obtain good numerical results. An optimal choice is not clear and heavily problem dependent. The numerical scheme is the following:

$$\int_{t_{n-1}}^{t_n}\int_\Omega [\Phi\dot{c} + \boldsymbol{u}\cdot\nabla c - \nabla\cdot(\boldsymbol{D}\nabla c)][w + \delta(w_t + \boldsymbol{u}\cdot\nabla w)]d\boldsymbol{x}dt + \int_\Omega c_{n-1}^+ w_{n-1}^+ d\boldsymbol{x}$$
$$= \int_{t_{n-1}}^{t_n}\int_\Omega f[w + \delta(w_t + \boldsymbol{u}\cdot\nabla w)]d\boldsymbol{x}dt + \int_\Omega c_{n-1}^- w_{n-1}^+ d\boldsymbol{x}.$$

δ is typically chosen to be $\mathcal{O}(h)$ where h is the diameter of the space partition on Ω. Its choice has significant effects on the accuracy of the numerical solutions. If δ is too small, the numerical solutions will exhibit oscillations. If δ is too big, the solution will be damped seriously. According to [23], δ is chosen as

$$\delta = \begin{cases} \frac{Kh}{\sqrt{1+|\boldsymbol{u}|^2}}, & \text{for } |\boldsymbol{u}|h > |\boldsymbol{D}| \\ 0, & \text{otherwise.} \end{cases} \tag{3.19}$$

K is typically 1 or 0.5.

Similarly to the steady case, when steep fronts or jump discontinuities (shocks) appear in the exact solution, the SDM may develop over- and under-shoots. A modified SDM with shock-capturing properties exists, but it leads to a nonlinear scheme and will not be considered here.

In [19] some analytical results are presented for the one-dimensional equation

$$Lc(x,t) := -\varepsilon c_{xx}(x,t) + u(x,t)c_x(x,t) + r(x,t)c(x,t) + c_t(x,t) = f(x,t), \tag{3.20}$$

where $(x,t) \in Q := (0,1)\times(0,T]$, and

$$\begin{aligned} c(x,0) &= s(x) \text{ on } S_x := \{(x,0) : 0 \le x \le 1\}, \\ c(0,t) &= q_0(t) \text{ on } S_0 := \{(0,t) : 0 < t \le T\}, \\ c(1,t) &= q_1(t) \text{ on } S_1 := \{(1,t) : 0 < t \le T\}. \end{aligned}$$

ε satisfies $0 < \varepsilon \ll 1$. Without loss of generality, one may assume $r(x,t) \ge r_0 > 0$ and $u(x,t) \ge u_0 > 0$ on $\overline{Q}$. Furthermore, one may assume homogeneous boundary conditions which is equivalent to solving (3.20) for the unknown function $c(x,t) - (1-x)q_0(t) - xq_1(t)$. There should also hold

$$r(x,t) - \frac{1}{2}u_x(x,t) \ge \gamma > 0 \qquad \text{on } Q.$$

As already mentioned in the 1D case, this assumption is not restrictive.

$[0,T]$ is partitioned by an equidistant mesh $\{t_j : j = 0,\ldots,N\}$ with $t_j = \frac{jT}{N} = j\tau$. The trial functions are continuous on each space time strip $Q_j := [0,1] \times (t_{j-1}, t_j)$. The interval $[0,1]$ is divided by the equidistant mesh $\{x_i : i = 0,\ldots,M\}$ with $x_i = \frac{i}{M} = ih$. X_j denotes the space of standard piecewise bilinear functions on this triangulation of Q_j that vanish at $x = 0$ and $x = 1$. The solution satisfies $c_{h,\tau}|_{Q_j} \in X_j$ for each j. Set $\hat{c}^j = c_{h,\tau}|_{Q_j}$. The test functions are obtained from a corresponding trial function w by the mapping $w \mapsto w + \delta(w_t + uw_x)$, where δ is a sufficiently small constant. In the following w_β denotes $w_\beta = (w_t + uw_x)$. The complete streamline diffusion formulation is

$$\begin{aligned} B_{SDM}(\hat{c}^j, w) &= \varepsilon(\hat{c}^j_x, w_x)_{Q_j} - \varepsilon \sum_{T \subset Q_j} (\hat{c}^j_{xx}, \delta w_\beta)_T + (\hat{c}^j_\beta + r\hat{c}^j, w + \delta w_\beta)_{Q_j} + \langle \hat{c}^j_+, w_+ \rangle_{j-1} \\ &= (f, w + \delta w_\beta)_{Q_j} + \langle \hat{c}^{j-1}_-, w_+ \rangle_{j-1}. \end{aligned} \tag{3.21}$$

$\langle \cdot, \cdot \rangle_j$ denotes the $L^2(0,1)$ inner product at $t = t_j$, and $z_\pm(x,t_j) := \lim_{k \to 0^+} z(x, t_j \pm k)$. The initial condition is incorporated by $\hat{c}^0_- = s$. Let $w_{m,k}$ be the canonical basis functions that satisfy

$$w_{m,k}(x_i, t_l) = \delta_{mi}\delta_{kl} \qquad \text{for } i = 0,\ldots,M \text{ and } l = j-1, j.$$

When $\hat{c}^j$ is piecewise linear, the contribution from $\hat{c}^j_{xx}$ is zero. For this case, it will be shown that (3.21) has a unique solution. This is a consequence of the coercivity of B_{SDM} [19].

Lemma 3.2 *Under the above assumptions the bilinear form from (3.21) is coercive.*

Proof: It holds

$$(w_t, w)_{Q_j} = \frac{1}{2}\langle w, w \rangle|^{t_j}_{t_{j-1}},$$

and

$$(uw_x, w)_{Q_j} = ((uw)_x, w)_{Q_j} - (u_x w, w)_{Q_j} = -(uw, w_x)_{Q_j} - (u_x w, w)_{Q_j}.$$

Hence

$$(w_t + uw_x + rw, w)_{Q_j} = \frac{1}{2}\langle w_-, w_- \rangle_j - \frac{1}{2}\langle w_+, w_+ \rangle_{j-1} - \frac{1}{2}(u_x w, w)_{Q_j} + (rw, w)_{Q_j}.$$

Consequently

$$\begin{aligned} B_{SDM}(w,w) &= \varepsilon||w_x||^2_{L^2(Q_j)} + (w_\beta + rw, w)_{Q_j} + \delta||w_\beta||^2_{L^2(Q_j)} \\ &\quad + \delta(rw, w_\beta)_{Q_j} + \langle w_+, w_+ \rangle_{j-1} \\ &\geq \varepsilon||w_x||^2_{L^2(Q_j)} + \gamma||w||^2_{L^2(Q_j)} + \frac{1}{2}\langle w_-, w_- \rangle_j \\ &\quad + \delta||w_\beta||^2_{L^2(Q_j)} + \delta(rw, w_\beta)_{Q_j} + \frac{1}{2}\langle w_+, w_+ \rangle_{j-1}. \end{aligned}$$

Furthermore

$$|\delta(rw, w_\beta)_{Q_j}| \leq \frac{\delta}{2}||r||^2_{L_\infty(Q_j)}||w||^2_{L^2(Q_j)} + \frac{\delta}{2}||w_\beta||^2_{L^2(Q_j)}.$$

Hence, if $\delta \leq \frac{\gamma}{||r||^2_{L_\infty(Q_j)}}$, then

$$B_{SDM}(w,w) \geq \varepsilon||w_x||^2_{L^2(Q_j)} + \frac{\gamma}{2}||w||^2_{L^2(Q_j)} + \frac{\delta}{2}||w_\beta||^2_{L^2(Q_j)} + \frac{1}{2}\langle w_-, w_- \rangle_j + \frac{1}{2}\langle w_+, w_+ \rangle_{j-1}.$$

♯

Coercivity can also be shown for higher order piecewise polynomial trial spaces, but with different stability requirements on δ and ε.

As in the steady case, global and local error estimates for the streamline diffusion solution $c_{h,\tau}$ can be given, see [19]. For each set $\hat{Q}$ which is the closure of a union of open triangles T, and each v that lies in $H^1(T)$ for all $T \subset \hat{Q}$, define

$$|||v|||_Q := \left\{ \varepsilon \sum_{T \subset \hat{Q}} ||\nabla v||^2_{L^2(T)} + \sum_{T \subset \hat{Q}} ||v_\beta||^2_{L^2(T)} + ||v||^2_{L^2(\hat{Q})} \right\}^{\frac{1}{2}}. \tag{3.22}$$

Theorem 3.6 *(Global error bound) Assume $\tau \leq Ch$, $\varepsilon \leq h$ and $\delta \leq C'h$ for some sufficiently small constant C'. Then for all sufficiently small h (independently of ε),*

$$|||c - c_{h,\tau}|||_Q \leq Ch^{\frac{3}{2}}|c|_{H^2(Q)}.$$

Theorem 3.7 *(Local error bound) Assume the hypotheses of theorem 3.6. Let (x_i, t_j) be a node of the triangulation. Suppose that u is constant. Define Q_2 by*

$$Q_2 := \{(x,t) \in Q : x + ut \leq x_i + ut_j + C_2 h \log \frac{1}{h}, |x - ut - (x_i - ut_j)| \leq C_2\sqrt{h} \log \frac{1}{h}\},$$

where C_2 is a fixed constant chosen in the proof. Then

$$|||c - c_{h,\tau}|||_{Q_2} \leq C\{h^{\frac{3}{2}}|c|_{H^2(Q_2)} + h^2||f||_{L^2(Q)} + h^2||s||_{L^2(0,1)}\}.$$

Q_2 is a layer around the sub-characteristic defined by $uw_x + rw + w_t = 0$ that passes through (x_i, t_j). Q_2 has width $C_2\sqrt{h}\log\frac{1}{h}$ and extends from the upstream boundary to a distance $C_2 h \log\frac{1}{h}$ downstream of (x_i, t_j). The global bound does not guarantee that $c_{h,\tau}$ is close to c, since typically $|c|_{H^2(Q)}$ is $\mathcal{O}(\varepsilon^{-\frac{3}{2}})$ and in practice $\varepsilon < h$. Although it is harder to prove, the local bound is much more useful. It ensures that $c_{h,\tau}$ is an $\mathcal{O}(h^{\frac{3}{2}})$ L^2 approximation to c on regions in Q that are not "near" layers. The SDM is not uniformly convergent. In particular it fails to converge inside layers. This can be improved by using an adapted mesh.

3.3.2 Characteristic Methods

In a characteristic (or Lagrangian) method, the transport of the fluid is referred to a Lagrangian coordinate system that moves with the fluid velocity. One tracks the movement of a fluid particle and the coordinate system follows the movement of the fluid. The time derivative along the characteristics of the advection-diffusion equation is expressed as

$$\frac{Dc}{Dt} = \frac{\partial c}{\partial t} + \frac{\boldsymbol{u}}{\Phi} \cdot \nabla c$$

and often referred to as total derivative. Consequently, the advection-diffusion equation can be rewritten as a parabolic diffusion-reaction PDE in a Lagrangian system

$$\Phi \frac{Dc}{Dt} - \nabla \cdot (\boldsymbol{D} \nabla c) = f,$$

where the advection has seemingly disappeared. In a Lagrangian coordinate system one would not see the effect of the advection or moving steep fronts. Hence, the solutions are much smoother along the characteristics than they are in the time direction. Therefore, characteristic methods allow large time steps to be used in a numerical simulation while still maintaining stability and accuracy. However, Lagrangian methods often raise extra and nontrivial analytical, numerical, and implementational difficulties, which require very careful treatment. The following classification follows [3] and [23].

- **The Classical Eulerian-Lagrangian Method** This is a finite difference method based on the forward tracking of particles in cells. Traditional forward tracking or moving mesh methods advance the grid following the characteristic and greatly reduce temporal errors. But they often distort the involved grids and greatly complicate the solution procedures.

- **The Modified Method of Characteristics (MMOC)** This method is based on a backward tracking from a fixed grid at the current time step to a point at the last time step. Hence MMOC avoids grid distortion problems as in forward tracking methods. However, MMOC and many other characteristic methods fail to conserve mass and have difficulties in treating general boundary conditions.

- **Eulerian Lagrangian Localized Adjoint Method (ELLAM)** ELLAM was introduced by Celia, Russel, and Herrera, see [2] and [8]. for solving one-dimensional constant-coefficient convection-diffusion equations. The ELLAM approach provides a general characteristic solution procedure for advection-dominated problems and a consistent framework for conserving mass and treating general boundary conditions. The ELLAM scheme generates accurate numerical solutions even if large time steps are used. It eliminates oscillations, numerical dispersion, and grid orientation problems.

Chapter 4

The Eulerian Lagrangian Localized Adjoint Method (ELLAM)

This chapter presents an ELLAM scheme for the numerical solution of the one- and two-dimensional linear advection-diffusion equations with all possible combinations of boundary conditions. Unlike other characteristic methods, ELLAM retains the Eulerian form of the transport equation and defines test functions to satisfy the homogeneous adjoint equation. It combines the classical method of characteristics with a Galerkin finite element approximation. A significant advantage of this method is that the CFL restriction is relaxed. Compared to Eulerian schemes, the temporal discretization error is reduced because spatial and temporal discretization are carried out separately. The temporal discretization is performed on the total derivative by tracking fictitious fluid particles during each time step. Thus, the time truncation error is proportional to the total derivative of the solution with respect to time, which is typically smaller than the local time derivatives [13]. Hence, much larger time steps than in Eulerian methods can be used without loss of accuracy. Unlike other characteristic methods, ELLAM is based on a forward tracking algorithm which has no effect on the solution grid or data structure of the discrete system. In contrast to other characteristic-based methods, this approach allows to treat any combination of boundary conditions. Furthermore, it has been proved that ELLAM globally conserves mass [2].

4.1 Localized Adjoint Methods

The general approach of localized adjoint methods (LAMs) is based on an algebraic theory of numerical methods presented by Herrera [8]. Let $\mathcal{L}$ be the operator of the governing differential equation

$$\mathcal{L}c(\boldsymbol{x}) = f(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega,$$

where c is the dependent variable and $\boldsymbol{x}$ is the vector of independent variables. The weak form is written as

$$\int_\Omega (\mathcal{L}c) w(\boldsymbol{x}) d\boldsymbol{x} = \int_\Omega f(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x} \tag{4.1}$$

with a test function w. Generally, the domain Ω is discretized into a number of subdomains Ω_i, $(i = 1, 2, \ldots, E)$. Depending on the regularity of c and w, simple integration by parts, the theory of distributions, or the general Green's formulas are applied to (4.1) in order to write the equation as a sum of elemental boundary integrals and integrals over the interior of each element. The resulting interior integrals involve an integrand that includes the adjoint $\mathcal{L}^*$ of $\mathcal{L}$ acting on w. The LAM procedure defines the test functions to satisfy the homogeneous adjoint equation

$$\mathcal{L}^* w = 0$$

within each element. Therefore, all interior elemental integrals are eliminated and only boundary integrals remain to be evaluated. The key to LAM algorithms is the definition of test functions that locally satisfy the homogeneous adjoint equation. For transient problems, the LAM approach has been applied in space to achieve a semi-discrete system for which standard time-marching algorithms were used. Unfortunately, this optimal spatial method suffers from large time truncation errors. However, the LAM approach need not be restricted to semi-discrete formulations. The approximations can be applied to the full space-time operator. That is the starting point for the ELLAM scheme.

4.2 The ELLAM Scheme for the 1D Advection-Diffusion Equation

In order to explain the concepts of a general two-dimensional ELLAM-scheme, it is useful to start with a one-dimensional equation with constant coefficients as in [2].

$$\begin{aligned} \mathcal{L}c = c_t + uc_x - Du_{xx} &= f(x,t), \quad x \in \Omega = [0,l], \quad t \in J = [0,\infty) \\ c(x,0) &= c_I(x) \\ c(0,t) &= c_0(t) \\ c_x(l,t) &= q_l(t). \end{aligned} \tag{4.2}$$

The LAM-approach is initiated by writing the weak form of (4.2). Let $w(x,t)$ be a test function which is chosen from the solution space of the homogeneous adjoint equation

$$\mathcal{L}^* w = -w_t - uw_x - Dw_{xx} = 0. \tag{4.3}$$

Different choices of test functions lead to different numerical schemes including optimal spatial methods and general characteristics methods. To derive a general family of characteristic methods one considers solutions of the two homogeneous sub-equations

$$w_t + uw_x = 0 \qquad \text{and} \qquad Dw_{xx} = 0.$$

The second constraint implies linear functions in x, while the first one implies that w is constant along lines $x - x_0 = u(t - t_0)$. These are the characteristics.

ELLAM is based upon the discontinuous Galerkin method in time. A partition of the time interval $J = [0, T]$ is defined by

$$0 = t_0 < t_1 < t_2 < \ldots < t_n < \ldots < t_{N_t - 1} < t_{N_t} = T.$$

Each space-time slab $\Sigma_n \equiv \Omega \times J_n$, $J_n \equiv (t_{n-1}, t_n]$, is discretized by space-time finite elements. The finite element spaces consist of piecewise linear functions in x and t. They

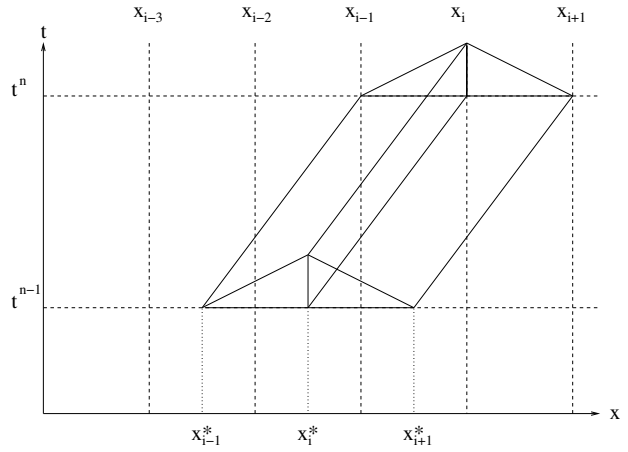


Figure 4.1: Interior test function w_i^n

are continuous in x, but discontinuous across time slabs. This means the trial and test functions w have support on $\Sigma_n \equiv \Omega \times J_n$ with $J_n \equiv (t_{n-1}, t_n]$ and vanish outside. This allows decoupling of the ELLAM scheme in time and permits to focus on the current time interval $(t_{n-1}, t_n]$.

For simplicity, assume a decomposition of the domain Ω into E constant spatial steps Δx. Then test functions can be defined in space and time as follows

$$w_i^n = \begin{cases} \dfrac{x - x_{i-1}}{\Delta x} + u\dfrac{t^n - t}{\Delta x}, & (x,t) \in \Omega_1^i \\ \dfrac{x_{i+1} - x}{\Delta x} - u\dfrac{t^n - t}{\Delta x}, & (x,t) \in \Omega_2^i \\ 0, & \text{otherwise.} \end{cases}$$

Figures 4.1 and 4.2 illustrate the shape of the test functions and the definition of $\Omega_{1,2}^i$. The subscript i denotes the spatial location $i\Delta x$, and $t^n = n\Delta t$ is the time level for a constant time step Δt.

The point x_i^* denotes the spatial location at time t^{n-1} on the characteristic that passes through x_i at time t^n. This point is also called 'foot' of the characteristic. The characteristics x_l^i, x_c^i, and x_r^i are lines of spatial derivative discontinuities of the above defined test functions.

For the calculation it is important to know which nodes are associated with the inflow and outflow boundaries. This information is contained in the Courant number which is defined locally as

$$Cr = \frac{u\Delta t}{\Delta x}. \tag{4.4}$$

IC denotes its next higher integer, $IC = [Cr] + 1$. The Courant number is a measure for the number of spatial intervals a particle passes during one time step. In the following, the Courant number will be restricted to $1 \leq Cr < 2$, and hence $IC = 2$, which is only for demonstration purpose. Figure 4.2 shows the support of different test functions.

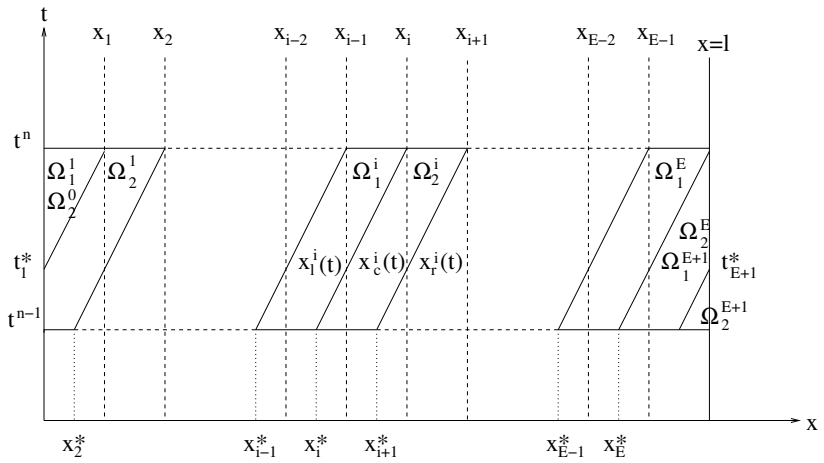


Figure 4.2: Geometric definition of test functions

IC also denotes the number of degrees of freedom on the boundary. There are $IC+1$ test functions that incorporate boundary terms. The characteristic curve that passes through $x = x_1$ at time t^n intersects the inflow boundary at $x = x_0 = 0$ at time $t_1^* \geq t^{n-1}$. For the case $IC = 2$, equations associated with w_0^n, w_1^n and w_2^n will involve boundary terms. Similarly, the characteristic with foot between x_{E-1} and x_E at time t^{n-1} will intersect the outflow boundary at $x = x_E = l$ at time $t^{n-1} < t_{E+1}^* < t^n$. Therefore, boundary terms at $x = x_l$ will appear in equations for w_E^n, w_{E+1}^n, and w_{E+2}^n.

The integrals in the weak formulation

$$\int_0^\infty \int_0^l (\mathcal{L}c - f)w(x,t)dxdt = 0$$

can be written as a sum of elemental integrals, where 'elements' are defined as the regions Ω_1^i and Ω_2^i. The integrals are evaluated using integration by parts.

$$\begin{aligned}
&\int_0^\infty \int_0^l \left[c_t + uc_x - Dc_{xx} - f(x,t)\right] w_i^n(x,t)dxdt = \\
&\quad - \int_0^\infty \int_0^l c\frac{\partial w_i^n}{\partial t}(x,t)dxdt + \underbrace{\int_{x_{i-1}}^{x_{i+1}} c(x,t^n)w_i^n(x,t^n)dx}_{I_1} - \underbrace{\int_{x_{i-1}^*}^{x_{i+1}^*} c(x,t^{n-1})w_i^n(x,t^{n-1})dx}_{I_2} \\
&\quad - \int_0^\infty \int_0^l uc\frac{\partial w_i^n}{\partial x}(x,t)dxdt + \underbrace{\int_{t^{n-1}}^{t^n} u\, cw_i^n(x,t)|_{\partial\Omega_l^i}^{\partial\Omega_r^i} dt}_{I_3} \\
&\quad - \int_0^\infty \int_0^l Dc\frac{\partial^2 w_i^n}{\partial x^2}(x,t)dxdt + \underbrace{\int_{t^{n-1}}^{t^n} D\, c\frac{\partial w_i^n}{\partial x}(x,t)|_{\partial\Omega_l^i}^{\partial\Omega_r^i} dt}_{I_4} - \underbrace{\int_{t^{n-1}}^{t^n} D\, c_x w_i^n(x,t)|_{\partial\Omega_l^i}^{\partial\Omega_r^i} dt}_{I_5} \\
&\quad - \underbrace{\int_{\Omega_1^i \cup \Omega_2^i} f(x,t)w_i^n(x,t)dxdt}_{I_6} = 0
\end{aligned} \tag{4.5}$$

The sum of the first terms in lines 2 - 4 equals zero because w satisfies the homogeneous adjoint equation. Taking into account the spatial derivative discontinuities of w results in

$$\begin{aligned}
&\int_{t^{n-1}}^{t^n} D\,c\frac{\partial w_i^n}{\partial x}(x,t)|_{\partial\Omega_l^i}^{\partial\Omega_r^i}dt = \\
&\quad\int_{t^{n-1}}^{t^n} D\,c\,\frac{\partial w_i^n}{\partial x}(x_c^i,t)dt - \int_{t^{n-1}}^{t^n} D\,c\,\frac{\partial w_i^n}{\partial x}(x_l^i,t)dt + \\
&\quad\int_{t^{n-1}}^{t^n} u\,c\frac{\partial w_i^n}{\partial x}(x_r^i,t)dt - \int_{t^{n-1}}^{t^n} u\,c\,\frac{\partial w_i^n}{\partial x}(x_c^i,t)dt = \\
&\quad D\frac{1}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_c^i(t),t)dt - D\frac{1}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_l^i(t),t)dt \\
&\quad +D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_r^i(t),t)dt - D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_c^i(t),t)dt.
\end{aligned}$$

For test functions w_i^n that do not intersect $\partial\Omega$ during the time interval $(t^{n-1},t^n]$, I_3 and I_5 become zero because w equals zero on $\partial\Omega^i$.

The integrals in (4.5) can be approximated in many different ways. For example, piecewise linear spatial interpolation of c at time levels t^{n-1} and t^n, coupled with a one-point implicit approximation to the temporal integrals at $t = t^n$, leads to the modified method of characteristics (MMOC). In all cases, the integrals are approximated in terms of nodal values of c at the discrete time levels t^{n-1} and t^n, $\{C_0^n, C_1^n, \ldots, C_E^n\}$,

$$c(x,t^n) = \sum_{j=0}^{E} C_j^n \phi_j(x),$$

where $\phi_j(x) = w_j^n(x,t^n)$.

Given the definition of the test functions and the assumption of constant Δx and u, the integrals can be evaluated exactly. For this purpose one has to determine the locations of the points x_i^*. There holds

$$x_i - x_i^* = u\Delta t = Cr\Delta x = IC\Delta x - \underbrace{(IC - Cr)}_{=:\alpha}\Delta x.$$

Hence

$$x_{i-1} - x_i^* = (x_i - x_i^*) - (x_i - x_{i-1}) = (IC - \alpha)\Delta x - \Delta x = (1-\alpha)\Delta x.$$

Similarly

$$t^n - t_1^* = \frac{\Delta x}{u}.$$

The discrete approximation for $i = 3, \ldots, E-1$ results in

$$\begin{aligned}
&\frac{\Delta x}{6}C_{i-1}^n + \frac{2\Delta x}{3}C_i^n + \frac{\Delta x}{6}C_{i+1}^n - \Delta x[\beta_1 C_{i-3}^{n-1} + \beta_2 C_{i-2}^{n-1} + \beta_3 C_{i-1}^{n-1} + \beta_4 C_i^{n-1}] \\
&-\frac{D\Delta t}{\Delta x}[C_{i-1}^n - 2C_i^n + C_{i+1}^n] = 0,
\end{aligned}$$

with

$$\beta_1 = \frac{1}{6}(1-\alpha)^3,\ \beta_2 = \frac{2}{3} - \alpha^2 + \frac{\alpha^3}{2},\ \beta_3 = \frac{1}{6}(1+\alpha)^3 - \frac{2}{3}\alpha^3,\ \beta_4 = \frac{\alpha^3}{6}.$$

The equations for test functions that intersect the boundary are more complicated. They appear in the appendix.

It is important to recognize that at the inflow boundary both $c(0,t)$ and $\frac{\partial c}{\partial x}(0,t)$ are present in the equations, as well as $c(l,t)$ and $\frac{\partial c}{\partial x}(l,t)$ at the outflow boundary. The reason is that the space-time LAM elements of Figure 4.2 are not parallel to the time axis.

4.2.1 Conservation of Mass

Summation of all equations associated with test functions w_0^n through w_{E+2}^n results in the expression

$$\begin{aligned}
&\int_{x_0}^{x_E} c(x,t^n)dx - \int_{x_0}^{x_E} c(x,t^{n-1})dx \\
&- \int_{t^{n-1}}^{t^n} [u\,c(0,t) - Dc_x(0,t)]dt + \int_{t^{n-1}}^{t^n} [u\,c(l,t) - D\,c_x(l,t)]dt \\
&= \int_{t^{n-1}}^{t^n}\int_{x_0}^{x_E} f(x,t)dxdt.
\end{aligned}$$

Hence, mass is conserved globally. We used the fact that $\Omega_1^k = \Omega_2^{k-1}$ and $w_k^n + w_{k-1}^n = 1$. Furthermore

$$\sum_{i=0}^{E+2} w_i^n(x,t^n) = 1 \qquad (0 \le x \le l) \qquad \text{and} \qquad \sum_{i=0}^{E+2} w_i^n(0,t) = 1 \qquad (t^{n-1} \le t \le t^n).$$

However, use of all ELLAM equations might over-specify the system. Different strategies can be applied depending on the type of boundary conditions.

4.2.2 Implementation of Boundary Conditions in 1D

The equations for $w_0^n, w_1^n, \ldots, w_{E+2}^n$ form a set of $E+3$ equations for the $E+3$ unknowns $(\partial C_0^n/\partial x), C_0^n, C_1^n, \ldots, C_{E-1}^n, C_E^n, (\partial C_E^n/\partial x)$. Incorporation of boundary conditions decreases the number of unknowns by two. Hence, only $E+1$ equations are needed but conservation of mass must be maintained. Especially, the ELLAM equation associated with w_{E+2}^n is not needed to solve for the nodal unknowns of interest, because the known values from the previous time level at node E supersede this equation. However, this final equation is needed to enforce global mass conservation.

4.2.2.1 Inflow Boundary

Approximation of the inflow boundary integral may eliminate some terms. For example, with a one-point integration quadrature at $t = t^n$ the term

$$\int_{t^{n-1}}^{t^n} c_x(0,t) w_i^n(0,t)dt, \qquad i > 0$$

is zero because $w_i^n(0,t^n) = 0$ for all $i > 0$.

Dirichlet Boundary Conditions If a Dirichlet boundary condition is specified, equations associated with $i = 0, 1, \ldots, E+1$ should be written. The first three of these include the diffusive flux $D\frac{\partial c}{\partial x}(0,t)$ which is unknown. With a one-point quadrature at $t = t^n$ for $i = 1$ (or equivalently at $t = t_1^*$ for $i = 2$) this term becomes zero because $w_i^n(0,t^n) = 0$ for all $i > 0$ (and $w_2^n(0,t_1^*) = w_2^n(x_1,t^n) = 0$). Then the flux integral only appears in the first equation which is uncoupled from the others in this case. It can be used to calculate the inflow boundary flux, if desired. This can be done by replacing w_1^n with the sum $w_0^n + w_1^n$.

Neumann Boundary Conditions For a Neumann boundary condition, $(\partial c/\partial x)(0,t)$ is known and $c(0,t)$ must be determined. It is stated in [2] that elimination of the integrals involving $c(0,t)$ in the equations for $i = 1$ and $i = 2$ will cause large quadrature errors. Hence, there is one more degree of freedom and the equation for w_0^n must be used, independent of the boundary integration method.

Flux Boundary Conditions In this case $\frac{\partial c}{\partial x}(0,t)$ can be expressed by $c(0,t)$. Hence, the unknown $(\partial C_0^n/\partial x)$ can be eliminated and the equation for w_0^n must be used to determine C_0^n.

4.2.2.2 Outflow Boundary

Dirichlet Boundary Conditions The outflow boundary is similar to the inflow boundary in that no boundary equations are required when a Dirichlet condition is specified. Boundary equations associated with w_E^n, w_{E+1}^n and w_{E+2}^n are only necessary to calculate the outflow boundary flux $\partial C_E^n/\partial x$. If desired, evaluation of boundary flux terms can be done by simple interpolation between time levels t^{n-1} and t^n. This corresponds to the use of a modified test function $w^{**} = w_E^n + w_{E+1}^n + w_{E+2}^n$.

Neumann Boundary Conditions In this case $c(l,t)$ is unknown for $t^{n-1} < t \leq t^n$ and the additional boundary equation associated with w^{**} must be written to solve for C_E^n.

Flux Boundary Conditions As in the inflow case, $\frac{\partial c}{\partial x}(l,t)$ can be expressed by $c(l,t)$ and the previous tools can be applied. If more refined information is desired at the boundary, individual equations may be written with additional nodal values on the boundary.

4.2.2.3 Overview

Table 4.1 sums up which unknowns are involved for different types of boundary conditions at the inflow and outflow boundary and which test functions are required to solve for these unknowns. The terms in parentheses are optional. They are only needed if detailed information at the boundary is required.

outflow inflow	Dirichlet	Neumann,flux
Dirichlet: unknowns	$\left(\frac{\partial C_0^n}{\partial x}\right), C_1^n, \ldots, C_{E-1}^n, \left(\frac{\partial C_E^n}{\partial x}\right)$	$\left(\frac{\partial C_0^n}{\partial x}\right), C_1^n, \ldots, C_{E-1}^n, C_E^n$
equations	$(w_0^n), w_1^n, \ldots, w_{E-1}^n,$ $(w_E^n + w_{E+1}^n + w_{E+2}^n)$	$(w_0^n), w_1^n, \ldots, w_{E-1}^n,$ $w_E^n + w_{E+1}^n + w_{E+2}^n$
Neumann,flux: unknowns	$C_0^n, C_1^n, \ldots, C_{E-1}^n, \left(\frac{\partial C_E^n}{\partial x}\right)$	$C_0^n, C_1^n, \ldots, C_{E-1}^n, C_E^n$
equations	$w_0^n, w_1^n, \ldots, w_{E-1}^n,$ $(w_E^n + w_{E+1}^n + w_{E+2}^n)$	$w_0^n, w_1^n, \ldots, w_{E-1}^n,$ $w_E^n + w_{E+1}^n + w_{E+2}^n$

Table 4.1: Unknowns and test functions for 1D-Ellam

4.3 The ELLAM Scheme for the 2D Advection-Diffusion Equation

In the past few years, Wang et al. developed ELLAM schemes for multi-dimensional advection-diffusion and advection-reaction equations, see [26] and [23]. His work also includes convergence analysis and error estimates [22], [25]. However, the aim of this section is to demonstrate how the ideas from one dimension can be generalized to higher dimensions. The aspects of error analysis will be skipped here.

4.3.1 Problem Description

Recall the governing differential equation

$$\Phi \frac{\partial c}{\partial t} + \nabla \cdot (\boldsymbol{u}c - \boldsymbol{D}\nabla c) = f, \qquad \boldsymbol{x} \in \Omega, \;\; t \in [0, T] \tag{4.6}$$

where Ω is the spatial domain with boundary $\Gamma := \partial\Omega$, and $J := [0, T]$ is the time interval. For simplicity, assume $\Phi = 1$ and $\boldsymbol{D} = D\boldsymbol{I}$.

The following assumptions are imposed on D and f by Wang [23]:

1. $D(\boldsymbol{x}, t), f(\boldsymbol{x}, t) \in W^{1,\infty}(\Omega \times (0, T))$ and $\boldsymbol{u}(\boldsymbol{x}, t) \in (W^{1,\infty}(\Omega \times (0, T)))^2$
2. There exist positive constants $D_{\min}$, $D_{\max}$, such that
$$0 < D_{\min} \leq D(\boldsymbol{x}, t) \leq D_{\max} < \infty \qquad \forall\, (\boldsymbol{x}, t) \in \overline{\Omega} \times [0, T].$$
3. The solution $c(\boldsymbol{x}, t) \in L^\infty(0, t; W^{2,\infty}(\Omega))$ and $c_t(\boldsymbol{x}, t) \in L^2(0, t; H^2(\Omega))$.

4.3.2 Variational Formulation

As in one dimension, the trial and test functions w have support on $\Sigma_n \equiv \Omega \times J_n$ with $J_n \equiv (t_{n-1}, t_n]$ and vanish outside. Multiplication of (4.6) with such a test function leads to the space-time variational formulation:

$$\int_{t_{n-1}}^{t_n} \int_\Omega [c_t + \nabla \cdot (\boldsymbol{u}c - D\nabla c)] w(\boldsymbol{x}, t) d\boldsymbol{x} dt = \int_{t_{n-1}}^{t_n} \int_\Omega f w(\boldsymbol{x}, t) d\boldsymbol{x} dt.$$

Integration by parts yields

$$\begin{aligned} &\int_\Omega (cw)(\boldsymbol{x}, t_n) d\boldsymbol{x} + \int_{t_{n-1}}^{t_n} \int_\Omega \nabla w \cdot (D\nabla c) d\boldsymbol{x} dt + \\ &\int_{t_{n-1}}^{t_n} \int_\Gamma (\boldsymbol{u}c - D\nabla c) \cdot \boldsymbol{n} w(\boldsymbol{x}, t) ds dt - \int_{t_{n-1}}^{t_n} \int_\Omega c(\Phi w_t + \boldsymbol{u} \cdot \nabla w) d\boldsymbol{x} dt \qquad (4.7) \\ &= \int_\Omega [c(\boldsymbol{x}, t_{n-1}) w(\boldsymbol{x}, t_{n-1}^+) d\boldsymbol{x} + \int_{t_{n-1}}^{t_n} \int_\Omega f w(\boldsymbol{x}, t) d\boldsymbol{x} dt \end{aligned}$$

with $w(\boldsymbol{x}, t_{n-1}^+) = \lim_{t \to t_{n-1}, t > t_{n-1}} w(\boldsymbol{x}, t)$.
Note that $w(\boldsymbol{x}, t_n) = w(\boldsymbol{x}, t_n^-) = \lim_{t \to t_n, t < t_n} w(\boldsymbol{x}, t)$. Continuity of the solution across time slabs is enforced weakly.

4.3.3 Test Functions

Following the concept of localized adjoint methods, test functions are chosen from the solution space of the homogeneous adjoint equation of (4.6).

$$-w_t - \boldsymbol{u} \cdot \nabla w - \nabla \cdot (D\nabla w) = 0, \qquad \boldsymbol{x} \in \Omega, \;\; t \in J_n \qquad (4.8)$$

To reduce the space of possible test functions, they are restricted to the solution space of the following system of equations:

$$\begin{aligned} w_t + \boldsymbol{u} \cdot \nabla w &= 0 \\ \nabla \cdot (D\nabla w) &= 0. \end{aligned} \qquad (4.9)$$

The first equation implies that the test functions should be constant along the characteristics $\boldsymbol{y} = \boldsymbol{r}(\theta, \boldsymbol{x}, t)$ defined by the initial value problem

$$\frac{d\boldsymbol{r}}{d\theta} = \boldsymbol{u}(\boldsymbol{r}, \theta), \text{ and } \boldsymbol{r}(\theta, \boldsymbol{x}, t)|_{\theta = t} = \boldsymbol{x} \qquad (4.10)$$

which reflects the hyperbolic nature of equation (4.6) and assures Lagrangian treatment of advection. The second equation is an elliptic PDE, so standard FEM approximations can be chosen for the spatial configuration of test functions. They are defined to be standard FEM basis functions on the spatial domain $\overline{\Omega}$ at time t_n and are extended constant into the space-time slab $\overline{\Omega} \times (t_{n-1}, t_n]$.

4.3.4 Trial Functions

The trial functions are based on standard FEM basis functions defined at time t^n. They coincide with the test function w_{ij}^n at time t^n.

$$c(x, t^n) = \sum_{j=0}^{M} C_j^n \phi_j(x),$$

where $\phi_j(x) = w_j^n(x, t^n)$.

4.3.5 Characteristics Tracking

Only in a few cases can characteristics be tracked exactly. In general one has to calculate approximate characteristics such that the adjoint term becomes small enough to be dropped. In practice, this can be achieved with a one-point Euler quadrature or a Runge Kutta quadrature.

In the following, $S_n^{(i)}$ denotes $S^{(i)}|_{(t_{n-1},t_n]} = \Gamma^{(i)} \times (t_{n-1}, t_n]$. When a particle moves along a characteristic from one time level to another, one has to distinguish four different cases, see Figure 4.3:

1. The starting point is $(\boldsymbol{x}, t_n) \in (\Omega \times t_n) \setminus S_n^{(O)}$.

   (a) The characteristic can be backtracked to time t_{n-1} without intersecting $S_n^{(I)}$. One defines the time step

   $$\Delta t^{(I)}(\boldsymbol{x}) = t_n - t^*(\boldsymbol{x}) = t_n - t_{n-1}.$$

   (b) The characteristic backtracks to the inflow boundary $S_n^{(I)}$ at a time $t^*(\boldsymbol{x}) \in (t_{n-1}, t_n]$. One defines the time step

   $$\Delta t^{(I)}(\boldsymbol{x}) = t_n - t^*(\boldsymbol{x}).$$

2. The starting point is $(\boldsymbol{x}, t) \in S_n^{(O)}$.

   (a) The characteristic does not intersect $S_n^{(I)}$ during $(t_{n-1}, t_n]$. The time step is defined by

   $$\Delta t^{(O)}(\boldsymbol{x}, t) = t - t^*(\boldsymbol{x}, t) = t - t_{n-1}.$$

   (b) The characteristic backtracks to the inflow boundary $S_n^{(I)}$ at a time $t^*(\boldsymbol{x}, t) \in (t_{n-1}, t]$. One defines the time step

   $$\Delta t^{(O)}(\boldsymbol{x}, t) = t - t^*(\boldsymbol{x}, t).$$

These time steps will be used to approximate the space-time volume integrals in (4.7).

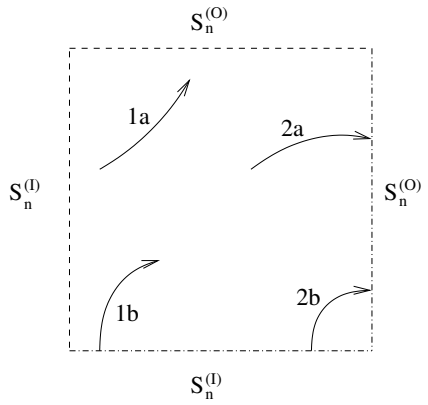


Figure 4.3: Characteristics tracking

4.3.6 Practical Integration

The aim of this subsection is to evaluate the integrals $\int_{t_{n-1}}^{t_n} \int_\Omega \ldots d\boldsymbol{x}dt$ in (4.7). For this purpose, the domain Ω is decomposed into the set $\Omega^{(O)}(\theta) \subset \Omega$ containing the points that will flow out of Ω during the time interval $(t_{n-1}, t_n]$ and the rest $\Omega \setminus \Omega^{(O)}$. The set $\Omega^{(O)}$ is defined by

$$\Omega^{(O)}(\theta) := \{\boldsymbol{x} \in \Omega | \exists \gamma \in [\theta, t_n] : \boldsymbol{r}(\gamma; \boldsymbol{x}, \theta) \in \Gamma\} \tag{4.11}$$

The two subsets can be characterized as follows:

- For any $(\boldsymbol{y}, \theta) \in (\Omega \setminus \Omega^{(O)}(\theta))$ there exits $\boldsymbol{x} \in \Omega$ such that $\boldsymbol{x} = \boldsymbol{r}(t_n; \boldsymbol{y}, \theta)$ which can be inverted to obtain $\boldsymbol{y} = \boldsymbol{r}(\theta; \boldsymbol{x}, t_n)$.

- For any $(\boldsymbol{y}, \theta) \in \Omega^{(O)}(\theta)$ there exists an $(\boldsymbol{x}, t) \in S_n^{(O)}$ such that $\boldsymbol{x} = \boldsymbol{r}(t; \boldsymbol{y}, \theta)$ which can be inverted to obtain $\boldsymbol{y} = \boldsymbol{r}(\theta; \boldsymbol{x}, t)$.

The characteristics are approximated by an one-step Euler formula:

$$\boldsymbol{r}(\theta; \boldsymbol{x}, t_n) := \boldsymbol{x} - \boldsymbol{u}(\boldsymbol{x}, t_n)(t_n - \theta), \quad \theta \in [t^*(\boldsymbol{x}), t_n] \tag{4.12}$$

or, respectively

$$\boldsymbol{r}(\theta; \boldsymbol{x}, t) := \boldsymbol{x} - \boldsymbol{u}(\boldsymbol{x}, t)(t - \theta), \quad \theta \in [t^*(\boldsymbol{x}, t), t]. \tag{4.13}$$

The Jacobian determinant of the transformation from $\boldsymbol{r}$ to $\boldsymbol{x}$ is given by:

$$
\begin{aligned}
|\boldsymbol{J}(\theta;\boldsymbol{x},t)| &:= \left|\frac{\partial \boldsymbol{r}(\theta;\boldsymbol{x},t)}{\partial \boldsymbol{x}}\right| \\
&= \begin{vmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} \end{vmatrix} \\
&= \begin{vmatrix} 1-\frac{\partial u_1}{\partial x_1}(t-\theta) & -\frac{\partial u_1}{\partial x_2}(t-\theta) \\ -\frac{\partial u_2}{\partial x_1}(t-\theta) & 1-\frac{\partial u_2}{\partial x_2}(t-\theta) \end{vmatrix} \\
&= 1-\frac{\partial u_1}{\partial x_1}(t-\theta)-\frac{\partial u_2}{\partial x_2}(t-\theta)+\frac{\partial u_1}{\partial x_1}\frac{\partial u_2}{\partial x_2}(t-\theta)^2-\frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1}(t-\theta)^2 \\
&= 1-(t-\theta)\nabla\cdot\boldsymbol{u}+\mathcal{O}((t-\theta)^2) \\
&= 1+\mathcal{O}(t-\theta). \qquad (4.14)
\end{aligned}
$$

Now, one is in the position to evaluate the integrals, keeping in mind that w is constant along the characteristics.

$$
\begin{aligned}
&\int_{t_{n-1}}^{t_n}\int_\Omega fw(\boldsymbol{y},\theta)d\boldsymbol{y}d\theta = \\
&\qquad \int_{t_{n-1}}^{t_n}\int_{\Omega\setminus\Omega^{(O)}} fw(\boldsymbol{r}(\theta;\boldsymbol{x},t_n),\theta)d\boldsymbol{r}d\theta+\int_{t_{n-1}}^{t_n}\int_{\Omega^{(O)}} fw(\boldsymbol{r}(\theta;\boldsymbol{x},t),\theta)d\boldsymbol{r}d\theta \\
&= \int_\Omega\int_{t^*(\boldsymbol{x})}^{t_n} fw(\boldsymbol{r}(\theta;\boldsymbol{x},t_n),\theta)|\boldsymbol{J}(\theta;\boldsymbol{x},t_n)|d\boldsymbol{x}d\theta+ \\
&\qquad \int_{S_n^{(O)}}\int_{t^*(\boldsymbol{x},t)}^{t} fw(\boldsymbol{r}(\theta;\boldsymbol{x},t),\theta)|\boldsymbol{J}(\theta;\boldsymbol{x},t)|\boldsymbol{u}(\boldsymbol{x},t)\cdot\boldsymbol{n}(\boldsymbol{x})d\theta d\boldsymbol{x}dt \\
&= \int_\Omega \Delta t^{(I)}(\boldsymbol{x})fw(\boldsymbol{x},t_n)d\boldsymbol{x}+ \\
&\qquad \int_{S_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x},t)fw(\boldsymbol{x},t)\boldsymbol{u}(\boldsymbol{x},t)\cdot\boldsymbol{n}(\boldsymbol{x})d\boldsymbol{x}dt+E(f,w) \qquad (4.15)
\end{aligned}
$$

In the last step, backward Euler quadrature was applied at the upper time boundary. E denotes the resulting truncation error. In a similar way, the diffusion-dispersion term can be rewritten as

$$
\begin{aligned}
&\int_{\Sigma_n}\nabla w\cdot(D\nabla c)(\boldsymbol{y},\theta)d\boldsymbol{y}d\theta \\
&\quad = \int_\Omega \Delta t^{(I)}(\boldsymbol{x})\nabla w\cdot(D\nabla c)(\boldsymbol{x},t_n)d\boldsymbol{x}+ \\
&\qquad \int_{S_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x},t)\nabla w\cdot(D\nabla c)\boldsymbol{u}\cdot n(\boldsymbol{x},t)d\boldsymbol{x}dt+E(D,c,w).
\end{aligned}
$$

4.3.7 A Reference Equation

Substituting the above equations into (4.7) results in a reference equation, written here for a special test function w_{ij}:

$$\begin{aligned}
&\int_\Omega c_n w_{ij,n} d\boldsymbol{x} + \int_\Omega \Delta t^{(I)}(\boldsymbol{x}) \nabla w_{ij,n} \cdot (D_n \nabla c_n) d\boldsymbol{x} + \\
&\quad \int_{S_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x}, t) \nabla w_{ij} \cdot (D \nabla c) \boldsymbol{u} \cdot n \, dsdt + \int_{t_{n-1}}^{t_n} \int_\Gamma (\boldsymbol{u} c - D \nabla c) \cdot \boldsymbol{n} w_{ij} \, dsdt \\
= &\int_\Omega c_{n-1} w^+_{ij,n-1} d\boldsymbol{x} + \int_\Omega \Delta t^{(I)}(\boldsymbol{x}) f_n w_{ij,n} d\boldsymbol{x} + \\
&\int_{S_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x}, t) f w_{ij} \boldsymbol{u} \cdot n \, dsdt + E(D, f, c, w)
\end{aligned} \tag{4.16}$$

4.3.8 Conservation of Mass

From (4.16) it can be seen that all the test functions should sum to one on $\overline{\Omega}$ at time t_n and on the space-time outflow boundary $S_n^{(O)}$ in order to conserve mass. Then, summation of (4.16) for all test functions results in

$$\begin{aligned}
&\int_\Omega c_n d\boldsymbol{x} + \int_{t_{n-1}}^{t_n} \int_\Gamma (\boldsymbol{u} c - D \nabla c) \cdot \boldsymbol{n} ds dt \\
&= \int_\Omega c_{n-1} d\boldsymbol{x} + \int_\Omega \Delta t^{(I)}(\boldsymbol{x}) f_n d\boldsymbol{x} + \int_{S_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x}, t) f \boldsymbol{u} \cdot n \, dsdt \\
&= \int_\Omega c_{n-1} d\boldsymbol{x} + \int_{t_{n-1}}^{t_n} \int_\Omega f \, d\boldsymbol{x} dt.
\end{aligned} \tag{4.17}$$

In other words: The difference in mass between times t_{n-1} and t_n equals the rate of mass flow across the boundary of Ω during the time interval $(t_{n-1}, t_n]$ plus the amount of mass entering Ω via sources or sinks.

4.3.9 Incorporation of Boundary Conditions

It is assumed that the type of boundary (inflow, outflow, or noflow) does not change during the time interval J_n. Substituting the boundary conditions (2.8) and (2.9) into (4.16) yields

$$
\begin{aligned}
&\int_\Omega c_n\, w_{ij,n} d\boldsymbol{x} + \int_\Omega \Delta t^{(I)}(\boldsymbol{x}) \nabla w_{ij,n} \cdot (D_n \nabla c_n) d\boldsymbol{x} + \\
&\qquad \int_{S_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x},t) \nabla w_{ij} \cdot (D\nabla c)\boldsymbol{u} \cdot n\, ds dt + \\
&\qquad \int_{t_{n-1}}^{t_n} \int_{\Gamma_N^{(I,O)}} \boldsymbol{u} \cdot \boldsymbol{n} c\, w_{ij}\, ds dt - \int_{t_{n-1}}^{t_n} \int_{\Gamma_D^{(I,O)}} D\nabla c \cdot \boldsymbol{n} w_{ij}\, ds dt \\
= &\int_\Omega c_{n-1} w_{ij,n-1}^{+} d\boldsymbol{x} + \int_\Omega \Delta t^{(I)}(\boldsymbol{x}) f_n w_n d\boldsymbol{x} + \\
&\qquad \int_{\Gamma_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x},t) f w_{ij} \boldsymbol{u} \cdot n ds dt - \int_{t_{n-1}}^{t_n} \int_{\Gamma_F^{(I,O)}} g_3 w(\boldsymbol{x},t) ds dt \\
&\qquad - \int_{t_{n-1}}^{t_n} \int_{\Gamma_N^{(I,O)}} g_2 w(\boldsymbol{x},t) ds dt - \int_{t_{n-1}}^{t_n} \int_{\Gamma_D^{(I,O)}} \boldsymbol{u} \cdot \boldsymbol{n} g_1 w(\boldsymbol{x},t) ds dt \qquad (4.18)
\end{aligned}
$$

where $\Gamma_{D,N,F}$ denote boundaries with given Dirichlet, Neumann or flux boundary conditions.

4.3.10 Implementation

The numerical scheme can be derived for a general domain Ω with a quasi-uniform triangular partition. Let $\mathcal{T} = \{T_1, T_2, \ldots, T_N\}$ be a triangulation of Ω with mesh spacing parameter $h = \max_{T\in\mathcal{T}} \mathrm{diam}(T)$. As mentioned above, the Courant number is a measure for the number of elements a particle passes during one time interval. From the Courant number in the normal direction at the boundary,

$$
Cr^{(I,O)} := \max_{(\boldsymbol{x},t)\in S_n^{(I,O)}} \left\{ \boldsymbol{u}(\boldsymbol{x},t) \cdot \boldsymbol{n} \frac{\Delta t}{h} \right\},
$$

one can conclude that particles with a distance to the boundary greater than $h\,Cr$ will not be influenced by it. On the other hand, the number of spatial degrees of freedom crossing the outflow boundary $S_n^{(O)}$ during the time interval $(t_{n-1}, t_n]$ is given by the Courant number in the normal direction. To preserve this information, the outflow boundary is refined in time by

$$
\mathcal{T}^t : \quad t_{n_i} := t_n - \frac{i\Delta t}{IC}, \qquad i = 0, 1, \ldots, IC.
$$

Let $\phi(\boldsymbol{x})$ be any piecewise bilinear function defined on $\overline{\Omega}$ with the partition $\mathcal{T}$, and $\phi(\boldsymbol{x},t)$ be any piecewise-bilinear function on the space-time outflow boundary $S_n^{(O)}$ with the partition $\mathcal{T}\times\mathcal{T}^t$. Then the test functions are defined on $\overline{\Omega}\times[t_{n-1},t_n]$ by

$$
\begin{aligned}
w(\boldsymbol{r}(\theta;\boldsymbol{x},t_n),\theta) &:= \phi(\boldsymbol{x}), \qquad \theta \in [t^*(\boldsymbol{x}),t_n], \quad \boldsymbol{x}\in\overline{\Omega} \\
w(\boldsymbol{r}(\theta;\boldsymbol{x},t),\theta) &:= \phi(\boldsymbol{x},t), \qquad \theta \in [t^*(\boldsymbol{x},t),t], \quad (\boldsymbol{x},t)\in S_n^{(O)}.
\end{aligned}
$$

In the discrete case the nodal basis functions defined at t_n are denoted by $\phi_{ij,n}$, satisfying $\phi_{ij,n}(\boldsymbol{x}_{kl}) = \delta_{ik}\delta_{jl}$, and the basis functions on $S_n^{(O)}$ are ϕ_{ij,n_k}, satisfying $\phi_{ij,s}(\boldsymbol{x}_{kl}, t_{n_m}) =$

$\delta_{ik}\delta_{jl}\delta_{sm}$. The weak formulation (4.16) solves for $c(\boldsymbol{x}, t_n)$ on Ω at time t_n and for $c(\boldsymbol{x}, t)$ on $S_n^{(O)}$.

If one is not interested in exact modelling of the outflow boundary, the problem can be simplified. The test functions that are defined on $S_n^{(O)}$ and intersect the boundary in the same space strip but at different times $t \in (t_{n-1}, t_n]$ are replaced by their sum

$$\hat{\phi}_{ij} = \sum_{k=0,1,\ldots,IC} \phi_{ij,n_k}.$$

This results in a new test function which coincides with the original test function $\phi_{ij,n}$ at time t_n and represents a constant extension of $\phi_{ij,n}$ along $(t_{n-1}, t_n]$. Summation of test functions preserves conservation of mass.

4.3.10.1 Evaluation of Integrals

Evaluation of $\int_\Omega c_{n-1} w_{n-1}^+ d\boldsymbol{x}$ In this term the value $c(\boldsymbol{x}, t_{n-1})$ is known from the solution at time t_{n-1} but the test function w is only defined at t_n.

The idea of many characteristic methods is to rewrite the integral as an integral at time t_n, to evaluate $w_{ij}(\boldsymbol{x}, t_n)$, and to use a backtracking algorithm to evaluate $c(\boldsymbol{x}^*, t_{n-1})$ with $\boldsymbol{x}^* = \boldsymbol{r}(t_{n-1}; \boldsymbol{x}, t_n)$. However, this algorithm requires significant effort for implementation.

A practical approach based on a forward tracking algorithm was proposed by Russel and Trujillo [23]. It uses the fact that

$$w_{n-1}^+ = \lim_{t \to t_{n-1}, t > t_{n-1}} w(\boldsymbol{x}, t) = w(\tilde{\boldsymbol{x}}, t_n), \text{ with } \tilde{\boldsymbol{x}} = \boldsymbol{r}(t_n; \boldsymbol{x}, t_{n-1})$$

and works as follows:

1. Choose quadrature points (e.g. Gaussian point) $\boldsymbol{x}_p$ on a fixed spatial grid at time t_{n-1}.
2. Evaluate c at these points.
3. Use a forward tracking algorithm to determine the head of the characteristic $\tilde{\boldsymbol{x}}_p = \boldsymbol{r}(t_n; \boldsymbol{x}_p, t_{n-1})$ at time t_n.
4. Determine which test functions w_k are nonzero at $(\tilde{\boldsymbol{x}}_p, t_n)$.
5. Add $c_{n-1}(\boldsymbol{x}_p) w_k(\tilde{\boldsymbol{x}}_p)$ at the corresponding position in the right-hand side vector in the global discrete linear algebraic system.
6. Exception: If the characteristic intersects $\Gamma_n^{(O)}$ at a time $\theta < t_n$ one has to evaluate w at the point of intersection.

This forward tracking algorithm does not influence the solution grid or the data structure of the discrete system. Hence, it does not suffer from grid distortion.

Remark: For problems with a given steady velocity field, the quadrature points at time t_{n-1} need to be tracked only once. Their corresponding end points of the tracking operation at time t_n should be stored in a separate data structure together with the values of the basis functions defined at these points.

Evaluation of $\int_\Omega \Delta t^{(I)}(\boldsymbol{x}) \ldots d\boldsymbol{x}$ For elements far away from the inflow boundary this integral is a standard FEM integral because $\Delta t^{(I)}(\boldsymbol{x})$ simplifies to $\Delta t = t_n - t_{n-1}$ and can be put outside the integral. In the other case, quadrature and backtracking is used to evaluate the term.

1. Choose quadrature points $\boldsymbol{x}_p$ on the fixed spatial grid at time t_n.
2. Evaluate the integrand.
3. Track the characteristic $\boldsymbol{r}(\theta; \boldsymbol{x}_p, t_n),\ \theta \in J_n$, to determine whether it reaches the boundary $\Gamma^{(I)}$ or not.
4. If so, calculate the time of intersection $t^*(\boldsymbol{x}_p)$.
5. Set $\Delta t^{(I)}(\boldsymbol{x}_p) = t_n - t^*(\boldsymbol{x}_p)$ or $\Delta t^{(I)}(\boldsymbol{x}_p) = t_n - t_{n-1}$, otherwise.

Here the backward tracking algorithm is only used to calculate $\Delta t^{(I)}(\boldsymbol{x})$ and does not effect mass conservation.

Evaluation of Inflow Boundary Integrals

$\int_{S_n^{(I)}} g^{(I)} w(\boldsymbol{x}, t) ds dt$ While the function g is defined on the space-time boundary $S_n^{(I)}$, the test function $w(\boldsymbol{x}, t)$ must be determined by $w(\boldsymbol{x}, t) = w(\tilde{\boldsymbol{x}}, t_n)$ where $\tilde{\boldsymbol{x}} = \boldsymbol{r}(t_n; \boldsymbol{x}, t)$ is the point at the head of the characteristic. Therefore one evaluates the integral as follows.

1. Choose quadrature points $(\boldsymbol{x}_p, t_q)$ at $S_n^{(I)}$.
2. Evaluate $g(\boldsymbol{x}_p, t_q)$ (also $\boldsymbol{u} \cdot n$ for a Dirichlet boundary condition).
3. Use forward tracking to determine $\tilde{\boldsymbol{x}} = \boldsymbol{r}(t_n; \boldsymbol{x}_p, t_q)$.
4. Evaluate the nonzero test functions $w_k(\tilde{\boldsymbol{x}}, t_n)$.
5. Add the products $g(\boldsymbol{x}_p, t_q) w_k(\tilde{\boldsymbol{x}}, t_n)$ with the corresponding quadrature weights at the correct position in the right-hand side vector in the global discrete linear algebraic system.

$\int_{S_n^{(I)}} (D \nabla c) \cdot \boldsymbol{n} w_{ij}(\boldsymbol{x}, t) ds dt$ To circumvent the difficulty of evaluating the unknown diffusive boundary flux, one approximates $\nabla c(\boldsymbol{x}, t)$ by $\nabla c(\boldsymbol{r}(t_n; \boldsymbol{x}, t), t_n)$. The introduced error is small because it is along the characteristics [23]. Hence, the algorithm works as follows.

1. Choose quadrature points $(\boldsymbol{x}_p, t_q)$ at $S_n^{(I)}$.
2. Evaluate $D(\boldsymbol{x}_p)$ and $\boldsymbol{n}(\boldsymbol{x}_p)$.
3. Use forward tracking to determine $\tilde{\boldsymbol{x}} = \boldsymbol{r}(t_n; \boldsymbol{x}_p, t_q)$.
4. Evaluate the nonzero trial functions $\nabla c(\tilde{\boldsymbol{x}}, t_n)$ and test functions $w_k(\tilde{\boldsymbol{x}}, t_n)$

5. Add $(D(\boldsymbol{x}_p)\nabla c(\tilde{\boldsymbol{x}}, t_n))\cdot\boldsymbol{n}(\boldsymbol{x}_p)w_k(\tilde{\boldsymbol{x}}, t_n)$ multiplied with the corresponding quadrature weights at the correct position in the left-hand side matrix in the global discrete linear algebraic system.

This term introduces non-symmetry to the coefficient matrix at nodes near the inflow boundary.

$\int_{S_N^{(I)}}(\boldsymbol{u}\cdot\boldsymbol{n})c\, w_{ij}(\boldsymbol{x}, t)dsdt$ This term can be evaluated in the same way as the expression above.

1. Choose quadrature points $(\boldsymbol{x}_p, t_q)$ at $S_n^{(I)}$.
2. Evaluate $\boldsymbol{u}\cdot\boldsymbol{n}(\boldsymbol{x}_p)$.
3. Use forward tracking to determine $\tilde{\boldsymbol{x}} = \boldsymbol{r}(t_n; \boldsymbol{x}_p, t_q)$.
4. Evaluate the nonzero trial functions $c(\tilde{\boldsymbol{x}}, t_n)$ and test functions $w_k(\tilde{\boldsymbol{x}}, t_n)$.
5. Add $\boldsymbol{u}\cdot\boldsymbol{n}(\boldsymbol{x}_p)cw_k(\tilde{\boldsymbol{x}}, t_n)$ multiplied with the corresponding quadrature weights at the correct position in the left-hand side matrix in the global discrete linear algebraic system.

It is also possible to use one-point quadrature at time t_n. That results in

$$\Delta t \int_{\Gamma_N^{(I)}} (\boldsymbol{u}\cdot\boldsymbol{n})c\, w_{ij}(\boldsymbol{x}, t_n)ds$$

which can be handled as a standard FEM integral.

Evaluation of Outflow Boundary Integrals The integrals defined on $S_n^{(O)}$ are standard since both the trial function c and the test functions w_{ij} are defined on $S_n^{(O)}$. They are evaluated by integration quadrature.

$\int_{S_n^{(O)}} \Delta t^{(O)}(\boldsymbol{x}, t)\ldots dsdt$ The factor $\Delta t^{(O)}(\boldsymbol{x}, t)$ is in general $t - t_{n-1}$, except near the corner where the inflow and outflow boundary intersect. For those elements one has to use a backtracking algorithm to determine the time $t^*(\boldsymbol{x}, t)$ when the characteristic $\boldsymbol{r}(\theta; \boldsymbol{x}, t)$ intersects $S_n^{(I)}$.

1. Choose quadrature points $(\boldsymbol{x}_p, t_q)$ at $S_n^{(O)}$.
2. Evaluate the integrand.
3. Track the characteristic $\boldsymbol{r}(\theta; \boldsymbol{x}_p, t_q)$, $\theta \in (t_{n-1}, t_q]$ to determine whether it reaches the boundary $\Gamma^{(I)}$ or not.
4. If so, calculate the time of intersection $t^*(\boldsymbol{x}_p, t_q)$.
5. Set $\Delta t^{(O)}(\boldsymbol{x}_p, t_q) = t_q - t^*(\boldsymbol{x}_p, t_q)$ or $\Delta t^{(O)}(\boldsymbol{x}_p, t_q) = t_q - t_{n-1}$, otherwise.

$\int_{S_n^{(O)}} g^{(O)} w(\boldsymbol{x}, t) ds dt$ As mentioned above, basis functions were defined on $S_n^{(O)}$ to be independent of time. The integral can be rewritten as

$$\int_{S_n^{(O)}} g^{(O)} w(\boldsymbol{x}, t) ds dt = \int_{\Gamma_n^{(O)}} w(\boldsymbol{x}, t_n) \int_{t_{n-1}}^{t_n} g^{(O)}(\boldsymbol{x}, t) dt ds.$$

It can be evaluated by integration quadrature or by a simplified one-point approximation for g and a standard FEM integral

$$\int_{S_n^{(O)}} g^{(O)} w(\boldsymbol{x}, t) ds dt \approx \Delta t g^{(O)}(\overline{\boldsymbol{x}}, \bar{t}) \int_{\Gamma_n^{(O)}} w(\boldsymbol{x}) d\boldsymbol{x},$$

where $\overline{\boldsymbol{x}}$ and $\bar{t}$ are interpolation points, e.g $\overline{\boldsymbol{x}}$ is the center of gravity and $\bar{t} = \frac{t_n + t_{n-1}}{2}$.

Chapter 5

Numerical Examples

The aim of this chapter is to compare the performance of some of the methods presented so far, particularly

- the standard Galerkin method with backward Euler (BE-GAL) and Crank-Nicolson (CN-GAL) time discretization,
- the cubic Petrov-Galerkin method with backward Euler (BE-CPG) and Crank-Nicolson (CN-CPG) time discretization,
- the streamline diffusion method (SDM),
- the ELLAM scheme.

The first two examples are taken from [23] and are standard test problems. Their exact solution is available. The third was examined in [20] and involves different boundary conditions.

5.1 Transport of a Diffused Square Wave

We consider the one-dimensional advection-diffusion equation

$$c_t + Vc_x - Dc_{xx} = f, \qquad x \in (a,b) = (0,2), \;\; t \in [0,T] = [0,1]$$

with the initial condition

$$c_0(x) = \begin{cases} 1, & \text{if } x \in [x_l, x_r] \subset (a,b) \\ 0, & \text{otherwise} \end{cases}, \qquad x_l = 0.2, x_r = 0.7. \tag{5.1}$$

Figure 5.1 illustrates the problem.

Homogeneous flux boundary conditions are specified at $x = a$ and $x = b$. Furthermore, the equation has the constant coefficients $V = 1$, $D = 10^{-4}$ and $f = 0$, so that the analytic solution can be given in closed form as long as the square wave does not intersect the outflow boundary during the time interval $[0,T]$:

$$c(x,t) = \frac{1}{2}\left[\operatorname{erf}\left(\frac{x - Vt - x_l}{\sqrt{4Dt}}\right) - \operatorname{erf}\left(\frac{x - Vt - x_r}{\sqrt{4Dt}}\right)\right], \;\; \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-s^2)ds.$$

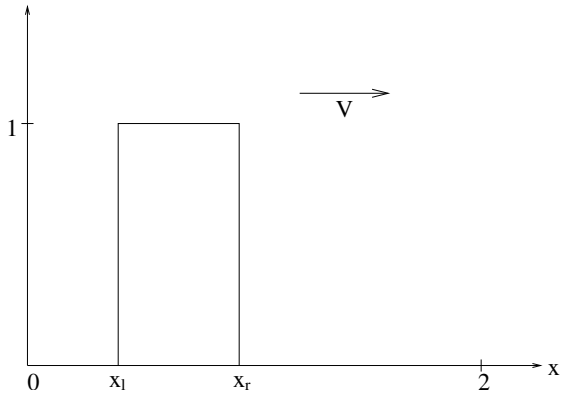
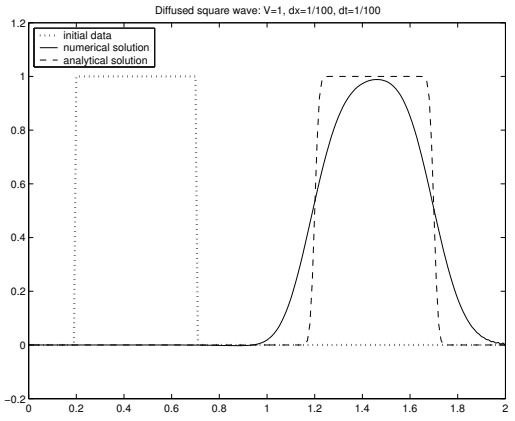
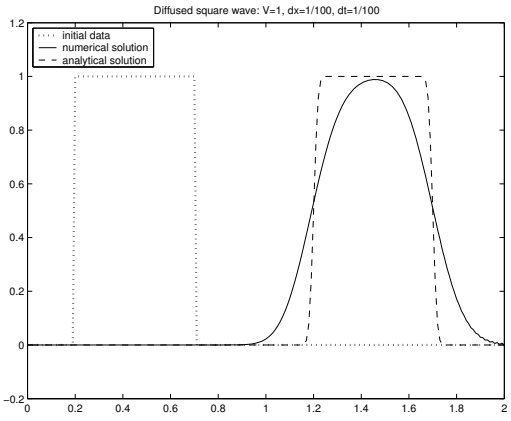


Figure 5.1: Initial data at $t = 0$



BE-GAL



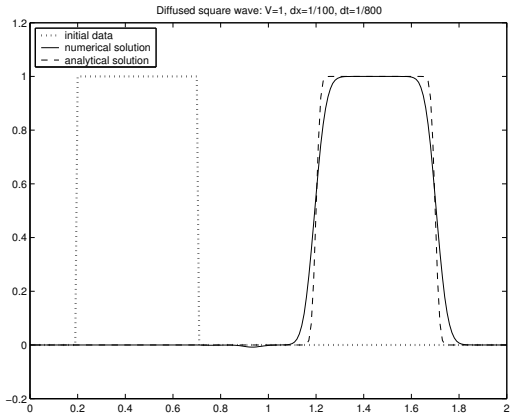
BE-CPG,

Figure 5.2: BE-GAL and BE-CPG solution with time step $\Delta t = \frac{1}{100}$

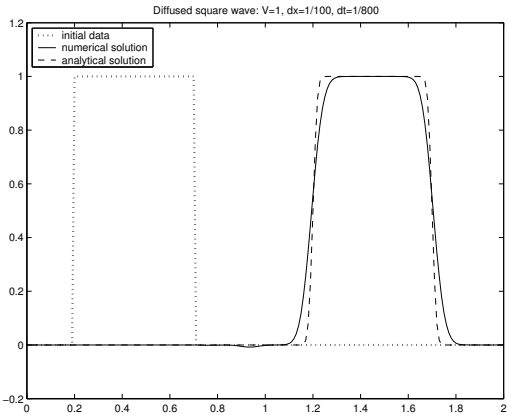
The grid size $\Delta x = 1/100$ is chosen so that the analytical solution can be represented properly. Linear finite elements were used. Figures 5.2 to 5.7 present the numerical solutions.

It can be observed that BE-GAL and BE-CPG generate almost identical numerical solutions which are over-damped for larger time steps. When the time step is decreased, the numerical diffusion is reduced and the numerical solutions become close to the analytical one. The reason for this behavior is that BE is so stable that it will even drive to zero the numerical solution of some ODEs with an exponentially growing solution for sufficiently large time steps. For example, for the simple linear test equation $\dot{y} = \lambda y$, $\lambda > 0$, BE yields $y_n = y_0/(1 - \lambda\Delta t)^n$, and hence the solution will be damped for $\lambda\Delta t > 1$, i.e. if Δt is sufficiently large. BE is the most dissipative method [6].

With the CN-GAL and CN-CPG schemes, the numerical diffusion was reduced considerably but they generate solutions with overshoots and undershoots. CN-CPG performs best for $\Delta t = \frac{1}{100}$ and $Cr = 1$ and cannot be improved with smaller time steps. On the other hand, the wiggles in GAL can be reduced as the time step decreases. In contrast to BE which is first order accurate, the CN-scheme is a second-order scheme. It is also

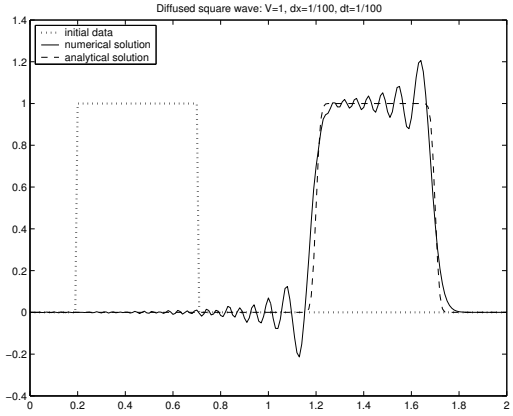


BE-GAL

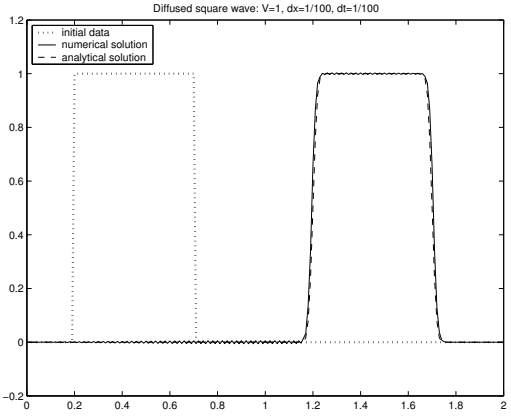


BE-CPG,

Figure 5.3: BE-GAL and BE-CPG solution with time step $\Delta t = \frac{1}{800}$

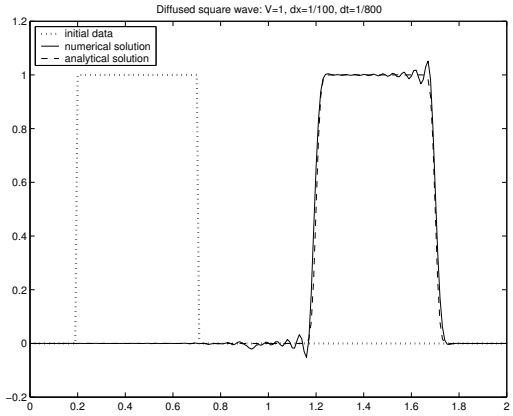


CN-GAL

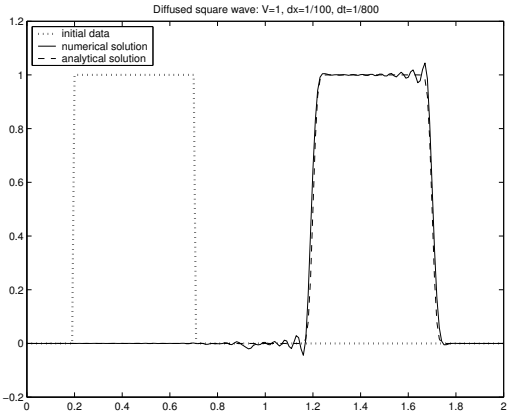


CN-CPG,

Figure 5.4: CN-GAL and CN-CPG solution with time step $\Delta t = \frac{1}{100}$

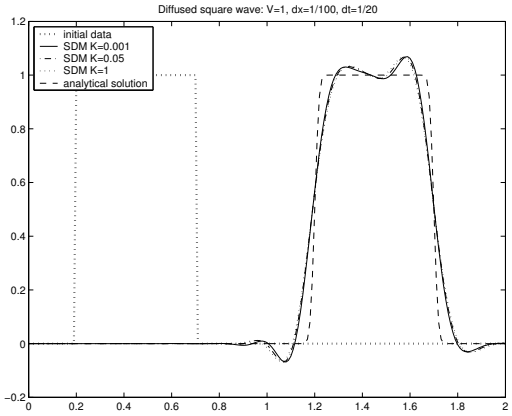


CN-GAL

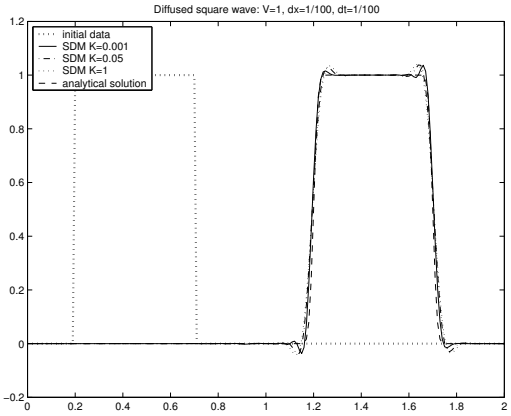


CN-CPG,

Figure 5.5: CN-GAL and CN-CPG solution with time step $\Delta t = \frac{1}{800}$



$\Delta t = \frac{1}{20}$



$\Delta t = \frac{1}{100}$

Figure 5.6: SDM solution with different time steps

A-stable, but it displays no spurious damping. However, its neutral stability is sometimes regarded as a disadvantage because too large time steps can lead to oscillations. To see this, consider the ODE $\dot{y} = -\lambda y$, $\lambda > 0$, with a monotonically decaying solution. The Crank-Nicolson scheme

$$y_{n+1} = y_n + \frac{\Delta t}{2}(\dot{y}_n + \dot{y}_{n+1})$$

gives for $\lambda \Delta t \gg 1$

$$y_{n+1} = \frac{1 - \lambda\Delta t/2}{1 + \lambda\Delta t/2} y_n = \frac{\frac{2}{\lambda\Delta t} - 1}{\frac{2}{\lambda\Delta t} + 1} y_n \approx -(1 - \frac{2}{\lambda\Delta t}) y_n \approx (-1)^{n+1}(1 - \frac{2}{\lambda\Delta t})^{n+1} y_0 \approx (-1)^{n+1} y_0.$$

The solution oscillates while decaying only very slowly.

The SDM approximates the analytical solution well even for large time steps. The smearing of the solution depends on the choice of the parameter K in (3.19). When the time step decreases, the damping is reduced but some wiggles appear near the steep fronts. Therefore, the modified SDM which avoids over- and undershoots near sharp fronts (shock capturing property) should generate better solutions for this example.

We observe that the ELLAM scheme generates accurate numerical solutions even for a quite large time step. Here, the largest allowable time step is $\Delta t = \frac{1}{51}$ because the one-dimensional ELLAM scheme was implemented for $1 \leq Cr < 2$. It is obvious that the ELLAM outperforms all other numerical solutions discussed before.

5.2 A Gaussian Pulse in 2D

Consider the two-dimensional advection-diffusion equation

$$\frac{\partial c}{\partial t} + \nabla \cdot (\boldsymbol{u} c - D \nabla c) = f, \qquad x \in \Omega = [-0.5, 0.5] \times [-0.5, 0.5], \;\; t \in [0, \pi/2]$$

with initial condition

$$c_0(x) = \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{2\sigma^2} \right)$$

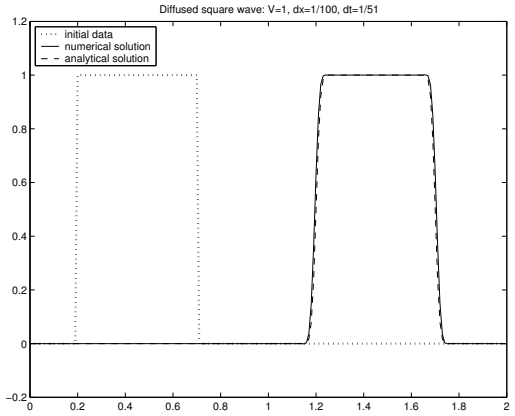


Figure 5.7: ELLAM solution with $\Delta t = \frac{1}{51}$

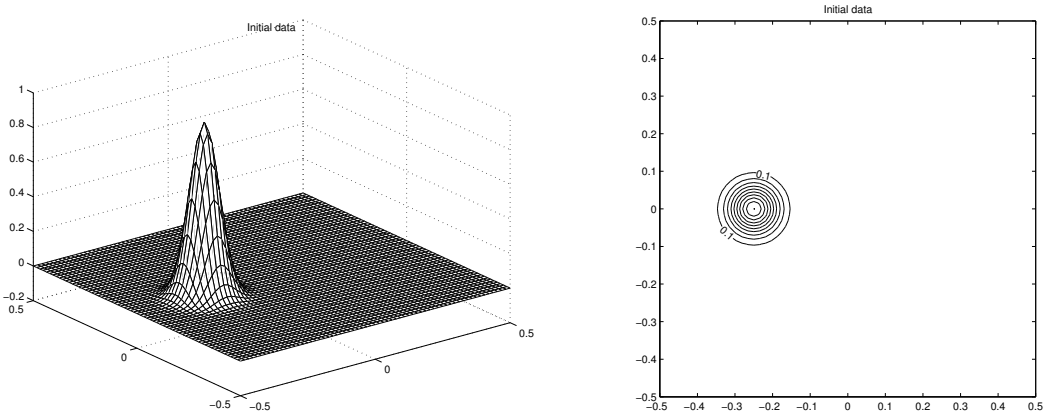


Figure 5.8: Initial condition

with $x_c = -0.25$, $y_c = 0$, $\sigma^2 = 0.002$, see Figure 5.8. Hence, the pulse is centered at (x_c, y_c) with a maximum value 1 and minimum 0. The velocity field is imposed as $\boldsymbol{u} = (-4y, 4x)$. The remaining coefficients are given by $D = 10^{-4}$ and $f = 0$. Then, the analytic solution is

$$c(x, y, t) = \frac{2\sigma^2}{2\sigma^2 + 4Dt} \exp\left(-\frac{(\overline{x} - x_c)^2 + (\overline{y} - y_c)^2}{2\sigma^2 + 4Dt}\right),$$

where $\overline{x} = x\cos(4t) + y\sin(4t)$, $\overline{y} = -x\sin(4t) + y\cos(4t)$. It is centered at $(\overline{x}, \overline{y})$. After the time $\frac{\pi}{2}$, which corresponds to one complete rotation of the pulse, the analytic solution has a maximum value 0.8642, see Figure 5.9.

This example is a standard test for numerical schemes for advection-diffusion problems [23]. The problem changes from diffusion dominance near the origin of the pulse to advection dominance in the rest of the domain. This often arises in many applications. In the numerical simulation, a uniform rectangular grid with $\Delta x = \Delta y = 1/64$ and bilinear finite elements were used.

The BE-GAL solutions are shown in Figures 5.10 and 5.11. BE-GAL requires small time steps in order to approximate the analytical solution. However, even with a time

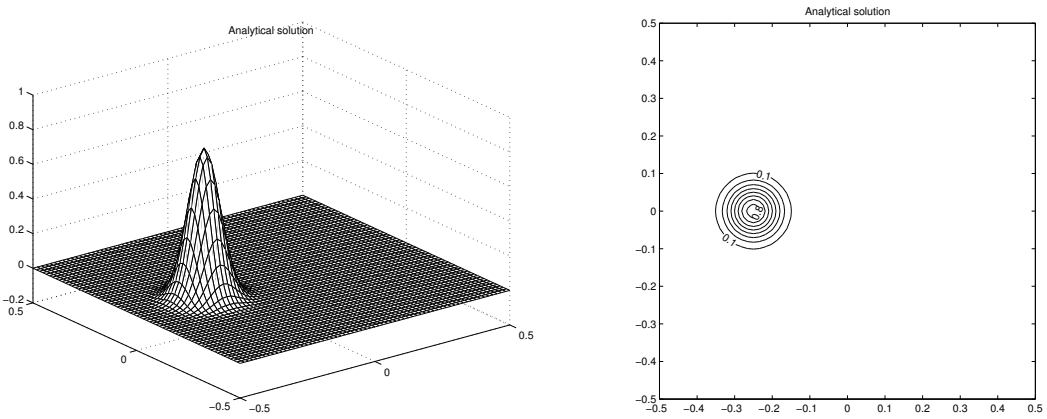


Figure 5.9: Analytical solution at $T = \frac{\pi}{2}$

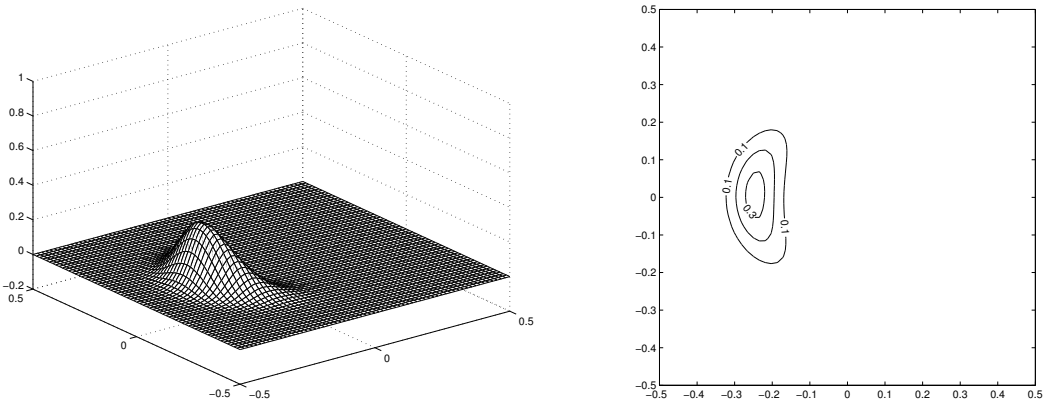
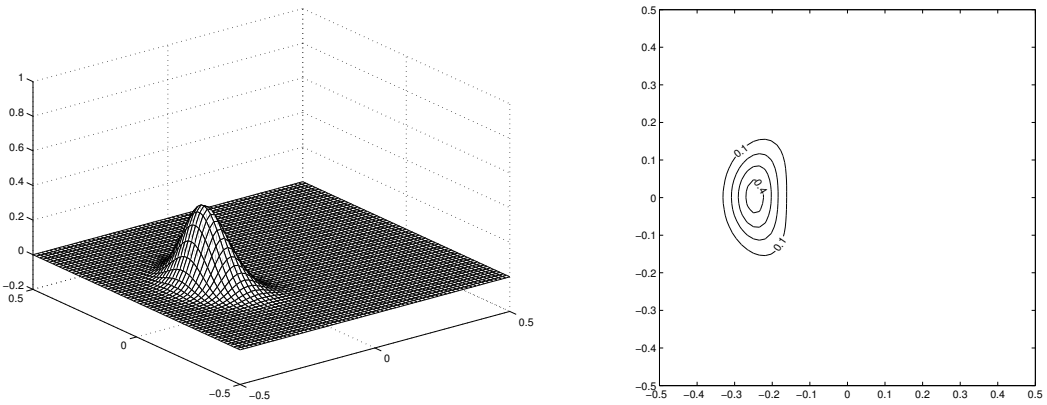


Figure 5.10: BE-GAL solution, $\Delta t = \frac{\pi}{400}$, max $= 0.3441$



mesh contour

Figure 5.11: BE-GAL solution, $\Delta t = \frac{\pi}{800}$, max $= 0.4517$

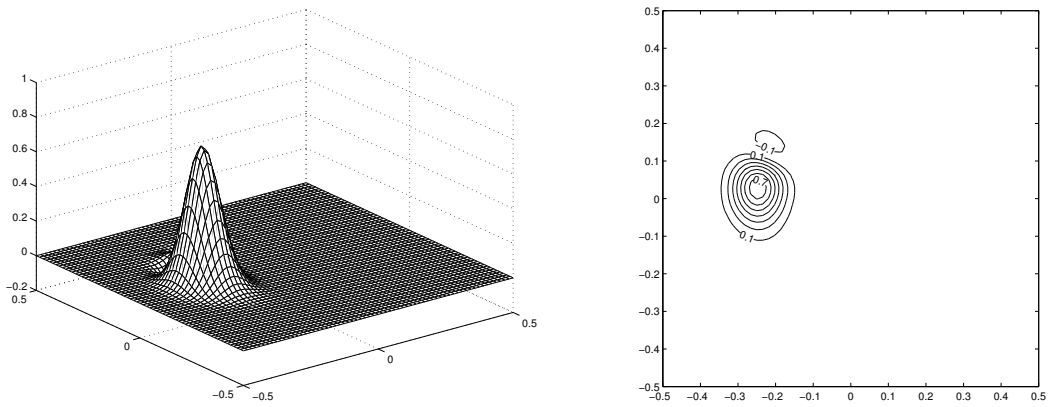


Figure 5.12: CN-GAL, $\Delta t = \frac{\pi}{200}$, max = 0.7861,min = -0.1564.

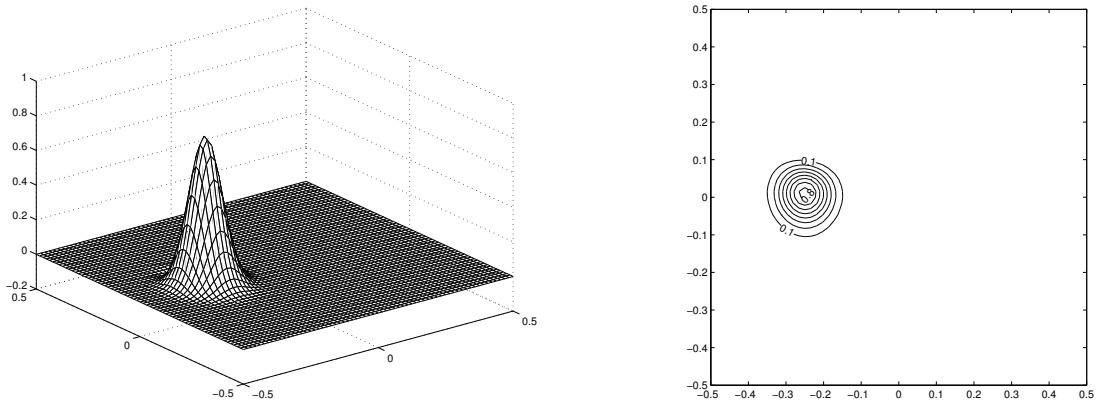


Figure 5.13: CN-GAL, $\Delta t = \frac{\pi}{400}$, max = 0.8438 min = -0.0159

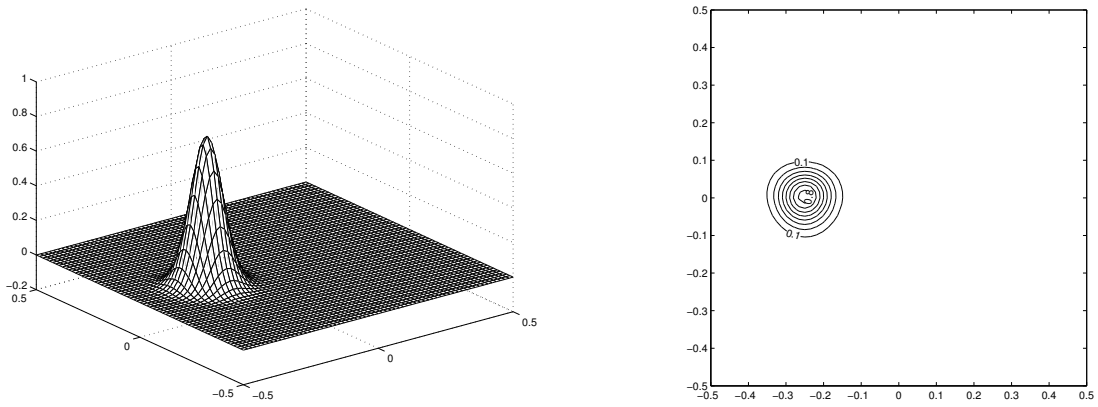


Figure 5.14: CN-CPG, $\Delta t = \frac{\pi}{400}$, max = 0.8557, min = -0.0002.

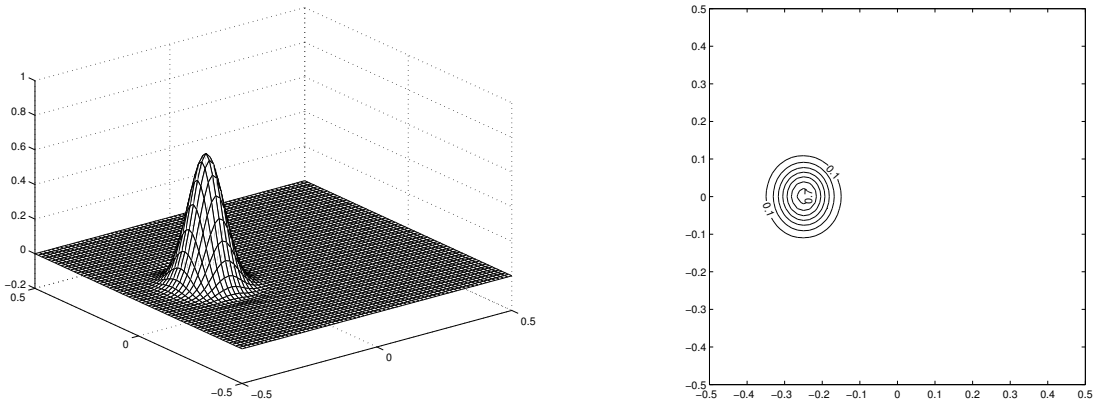


Figure 5.15: SDM, $K = 0.5$, $\Delta t = \frac{\pi}{200}$,max $= 0.7089$, min $= -0.0102$

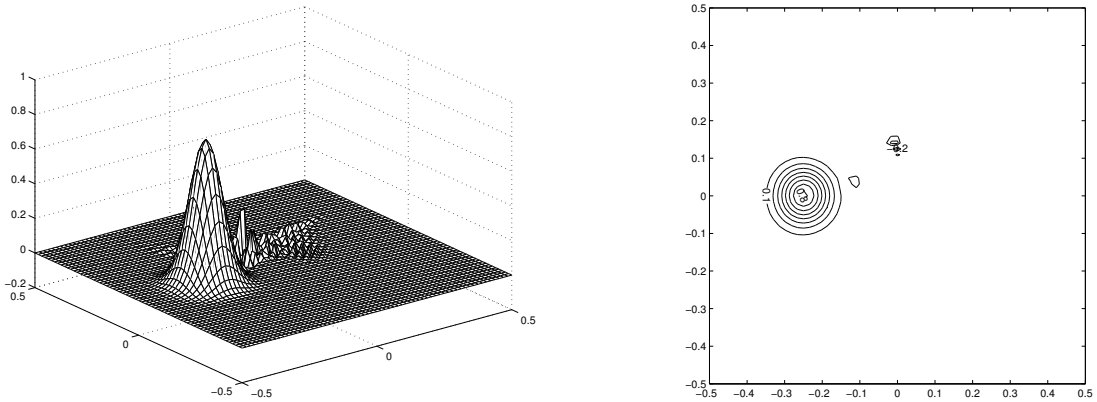


Figure 5.16: SDM, $K = 0.01$, $\Delta t = \frac{\pi}{200}$, max $= 0.8265$, min $= -0.2277$

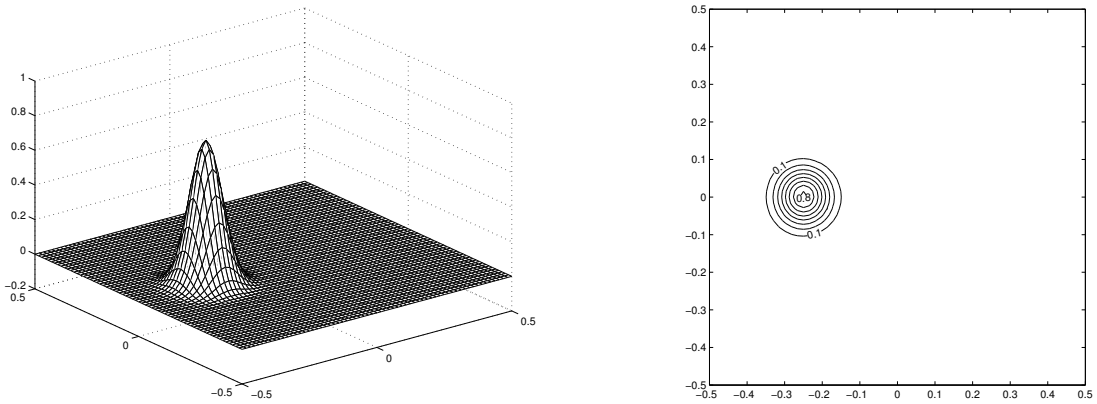


Figure 5.17: SDM, $K = 0.001$, $\Delta t = \frac{\pi}{200}$, max $= 0.8283$, min $= -0.0019$

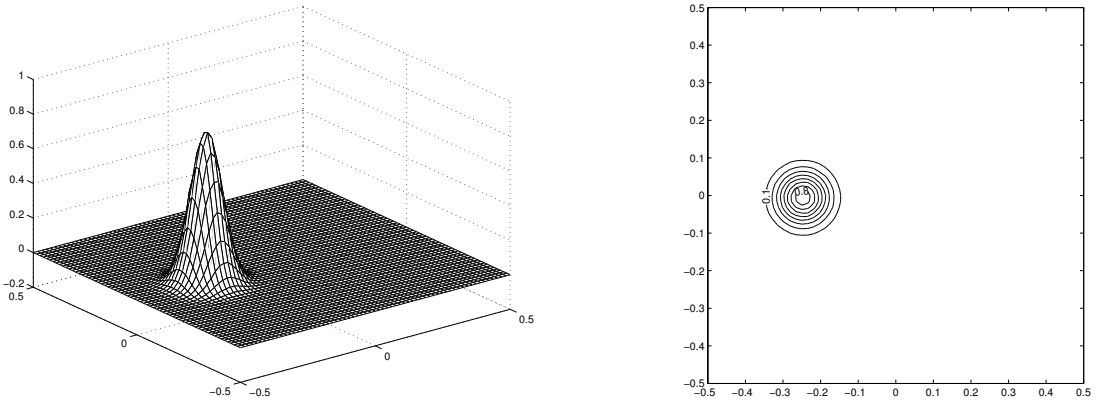


Figure 5.18: ELLAM with backward Euler, $\Delta t = \frac{\pi}{40}$, $\Delta t_f = \frac{\Delta t}{80}$,max $= 0.8329$.

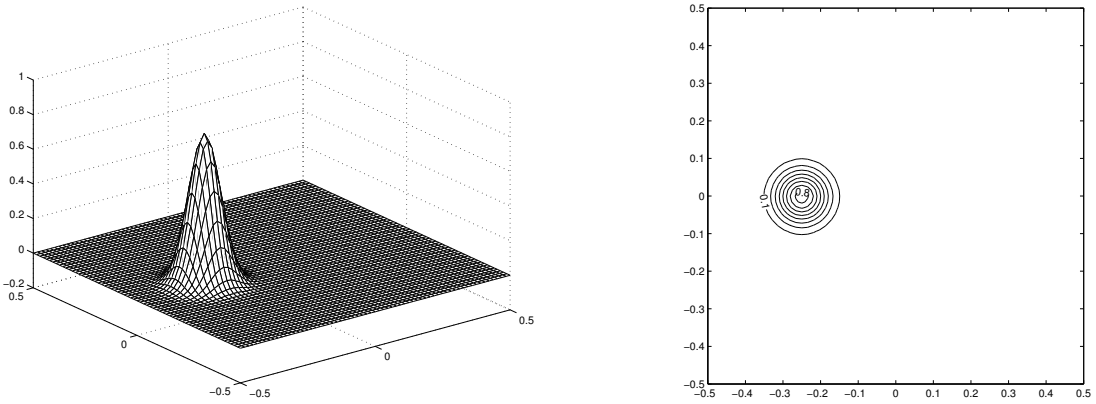


Figure 5.19: ELLAM with 2^{nd} order Runge-Kutta, $\Delta t = \frac{\pi}{40}$, $\Delta t_f = \frac{\Delta t}{4}$, max $= 0.8487$.

step $\Delta t = \frac{\pi}{400}$ the numerical solution is excessively over-damped and considerably deformed. Since the temporal error dominates the overall error, the numerical solution can be improved by reducing the time step but it still cannot compete with the ELLAM solutions which are shown in Figures 5.18 and 5.19. Hence, even though the backward Euler temporal discretization is unconditionally stable, extremely small time steps have to be used for the purpose of comparative accuracy. Consequently, the efficiency of the simulation is reduced significantly.

The CN-GAL solution for $\Delta t = \frac{\pi}{200}$ is presented in Figure 5.12. The CN-discretization yields more accurate results than the BE schemes due to its higher order temporal accuracy. However, some undershoot is observed. The reason for this was discussed in the one-dimensional case. When the time step is reduced to $\Delta t = \frac{\pi}{400}$, the numerical solution is improved considerably, see Figure 5.13. For this time step, the CN-CPG solution is shown in Figure 5.14.

Figures 5.15 to 5.17 show the results of the SDM simulation with different values for the undetermined parameter K from (3.19). The SDM solutions are more accurate than the Galerkin solutions but require more CPU time because the number of unknowns doubles. Furthermore, the parameter K has to be chosen very carefully. $K = 0.01$ generates oscillatory solutions, while other values, smaller as well as larger than 0.01, perform well.

For the ELLAM simulation, two different methods were applied to track the characteristics. First, a time step $\Delta t = \frac{\pi}{40}$ and backward Euler with a micro-time step $\Delta t_f = \frac{\Delta t}{80}$ for the characteristics was used. Then, a second order Runge-Kutta (RK) method with $\Delta t_f = \frac{\Delta t}{4}$ was used which reduces the CPU time significantly. However, both methods yield accurate numerical solutions with slightly less damping for the RK-method.

5.3 A Rotated Inflow Profile

The following example is taken from [20]. The requirement is to solve

$$c_t + \boldsymbol{u} \cdot \nabla c - D\Delta c = 0, \qquad \text{in } \Omega = [-1,1] \times [0,1],\ t > 0$$

with the velocity field specified as

$$u = 2y(1-x^2), \quad v = -2x(1-y^2),$$

where u and v are along the x and y coordinate directions, respectively. With exception of the outflow boundary, c is specified as

$$c = 1 + \tanh[(2x+1)\alpha], \qquad \text{for } y = 0,\ -1 \le x \le 0,$$

$$c = 1 - \tanh(\alpha), \text{ for } \begin{cases} x = -1, & 0 \le y \le 1 \\ y = 1, & -1 \le x \le 1 \\ x = 1, & 0 \le y \le 1 \end{cases}$$

Figure 5.20 illustrates the problem. The value of α determines the sharpness of the climb from 0 to 2 halfway along the inflow boundary. The smaller α, the smoother the front. The outflow boundary condition is specified by

$$\nabla c \cdot \boldsymbol{n} = 0, \qquad \text{for } y = 0,\ 0 < x \le 1.$$

The initial condition is given by

$$c(x, t = 0) = 0.$$

In [20], the problem was posed as a stationary problem. Here, an unsteady analogue will be considered as well. Figures 5.21 and 5.22 contain the numerical solutions of the steady equation, while Figures 5.23 and 5.24 present the results for the unsteady equation.

For $\alpha = 10$ the sharp front can be resolved and the standard Galerkin method performs well, but for $\alpha = 100$ the solution exhibits upstream moving wiggles which are also present on a finer grid. These wiggles can be removed by using upwinded schemes such as QPG or SDM, which deliver similar results. The solution becomes smoother in regions outside the steep front, but over- and undershoots appear near the front.

For the time dependent simulation, the final time was set to $t = 2$ which is the time needed to transport all information from the inflow to the outflow boundary. However, at this time the numerical solutions with GAL and CPG exhibit wiggles on top and at the bottom of the steep front, see Figure 5.23. Note that CPG requires a time step $\Delta t < \Delta x$ but an improvement compared to GAL is not visible. However, if the simulation is continued until a larger final time, for example $t = 4$, this has a smoothing influence on the

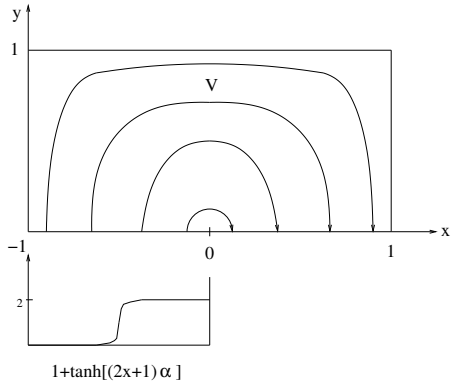
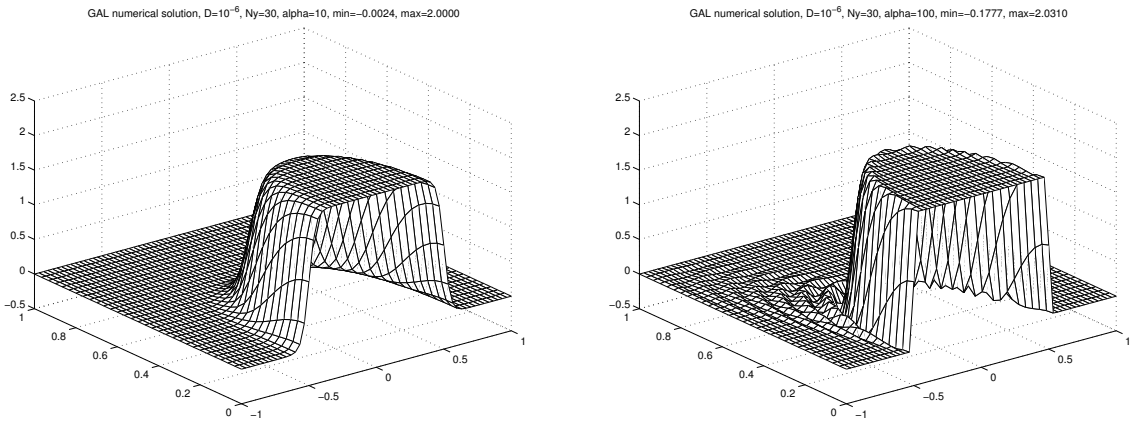
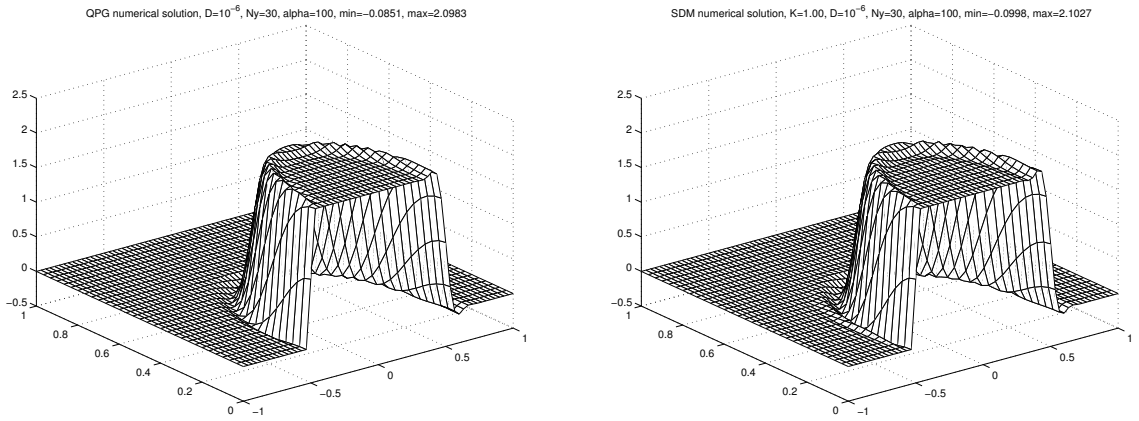


Figure 5.20: Initial data of the inflow problem



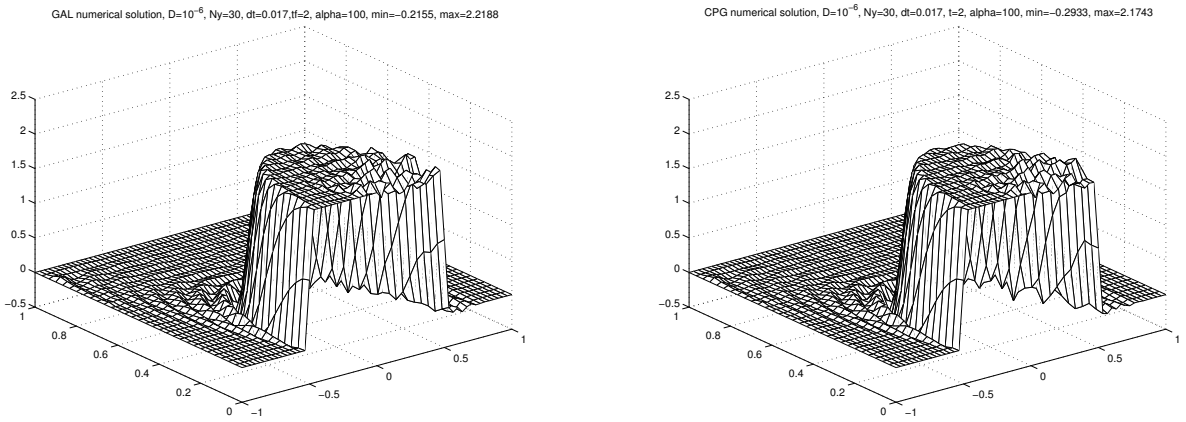
$\alpha = 10$ $\alpha = 100$

Figure 5.21: CN-GAL solution of the steady advection-diffusion equation for different initial conditions



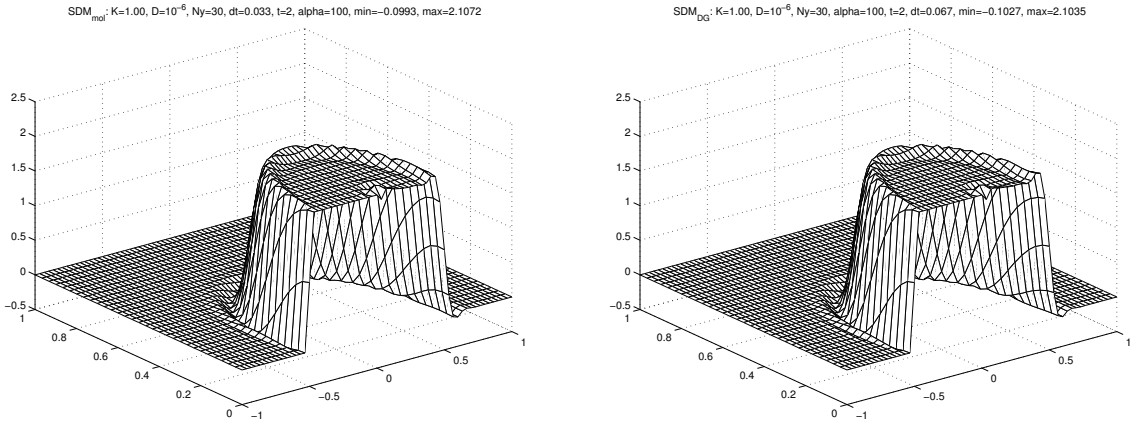
QPG SDM

Figure 5.22: QPG and SDM solution of the steady advection-diffusion equation



GAL CPG

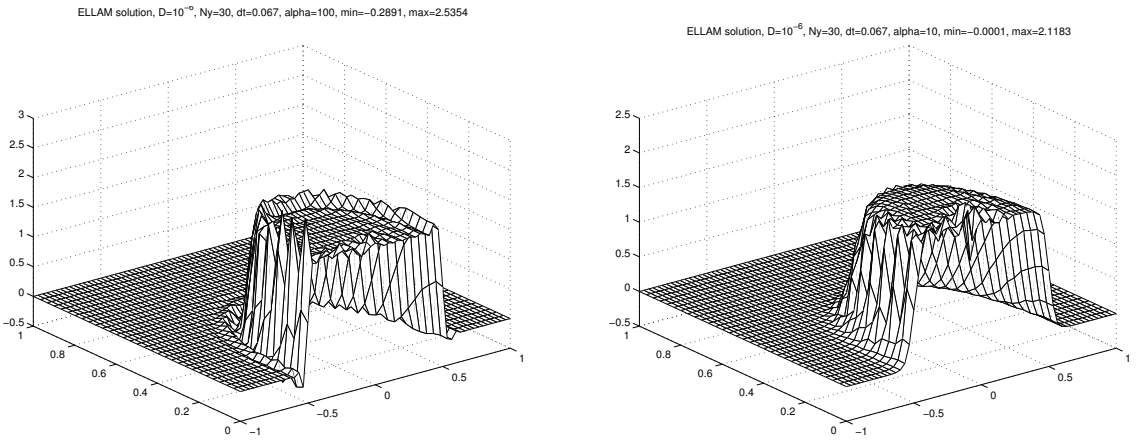
Figure 5.23: GAL and CPG solution of the unsteady advection-diffusion equation



method of lines discontinuous Galerkin solution

Figure 5.24: SDM solution of the unsteady advection-diffusion equation

numerical solution and the wiggles on top of the front will disappear. On the other hand, SDM is able to remove the wiggles in the method of lines ($w_{SDM} = w + \delta \boldsymbol{u} \cdot \nabla w$) as well as in combination with a discontinuous Galerkin method in time ($w_{SDM} = w + \delta(w_t + \boldsymbol{u} \cdot \nabla w)$) where even a larger time step leads to numerical solutions which are smooth outside the layer, see Figure 5.24. The SDM parameter $K = 1$ was chosen so that over- and undershoot have nearly the same magnitude. The ELLAM solution was calculated with a time step $\Delta t = 2\Delta x$ and is shown in Figure 5.25. Larger time steps cause oscillations along the moving front of the concentration profile. In order to damp the wiggles along the discontinuity, a smaller time step is required. One can observe that ELLAM performs better if the exact solution is smoother ($\alpha = 10$).



$\alpha = 100$ $\alpha = 10$

Figure 5.25: ELLAM solution of the unsteady advection-diffusion equation

5.4 Conclusion

In many cases, ELLAM generates very accurate numerical solutions compared with the other methods considered, even though a much larger time step is used. On the other hand, ELLAM requires the most CPU time per time step and is the most difficult to implement. The last example has shown that ELLAM has difficulties with discontinuities in the solution, but other methods do so as well. Often, the Galerkin finite element method with sophisticated upwinding like QPG/CPG or SDM yields good numerical results. CPG is especially useful if the velocity is constant. In this case, the time step can be chosen such that the truncation error becomes small. This concept fails if the velocity varies rapidly. Therefore, CPG did not perform better than GAL in the last example of this chapter but it could improve the numerical solution in the second example because the difference in the velocity between the top and the bottom of the pulse was not too big. On the other hand, the SDM always delivers solutions which are smooth outside boundary layers but suffer from over- and undershoot if steep fronts appear in the solution. The shock-capturing SDM should be applied in these cases. In sum, it depends on the problem which method will be preferred.

Chapter 6

The Stochastic Advection-Diffusion Equation

The algorithms presented so far have assumed complete knowledge of the problem data, such as boundary conditions, initial data, coefficients or source terms. However, in many cases the available information to solve the problem is limited, for example when the coefficients depend on material properties that are known only to some accuracy or at selected points in the domain. The more general question arises of how to incorporate uncertainty, and how to formulate algorithms in order to reflect the propagation of uncertainty to the simulation output. By uncertainty we mean variability of physical quantities which is interpreted as randomness. Hence, probability theory can be applied. A probabilistic description is used for the coefficient variability which leads to the study of stochastic differential equations. Here, by stochastic partial differential equations (SPDEs) we refer to equations where the involved functions are modelled as stochastic processes parameterized by points in space and/or time. The information from the stochastic input can be used to compute approximations to the distribution function of the stochastic solution or at least statistical moments of the solution, in particular the expected value.

One obvious approach is the Monte-Carlo method. This generates a set of independent identically distributed realizations of the solution by sampling the coefficients of the equation and solving the resulting deterministic problem by conventional finite element methods. Statistical properties of the solution of the SPDE can then be computed by postprocessing the resulting population of solution realizations. However, this turns out to be computationally expensive because the accuracy for the postprocessing depends on the sample size.

Another method was developed by Ghanem and Spanos [5] in the context of finite element methods in which the random field is discretized directly by a polynomial chaos expansion. This expansion is based on the homogeneous chaos by Wiener [28] and is essentially a Fourier expansion of the random variables. The finite dimensional approximation of the stochastic coefficients turns the original stochastic problem into a deterministic parametric problem.

Here we follow a similar approach suggested by Babuška et al. [1]. It is similar to that of Ghanem and Spanos but differs in the choice of the approximating function spaces by using piecewise polynomials instead of a Hermite expansion. Double orthogonal polynomials are used to compute efficiently the solution by the stochastic Galerkin finite element method.

6.1 Theoretical Aspects of the Stochastic Galerkin Finite Element Method

6.1.1 Notation and Function Spaces

Let D be a convex bounded polygonal domain in $\mathbb{R}^d$, $[0,T]$ a time interval in $\mathbb{R}$ and $(\Omega, \mathcal{F}, P)$ a complete probability space. Here, Ω is the set of outcomes, $\mathcal{F} \subset 2^\Omega$ is a σ-algebra of events and $P : \mathcal{F} \to [0,1]$ is a probability measure. In addition, set $\mathcal{D} = D \times [0,T]$.

In the following, we concentrate on a stochastic velocity field and a stochastic right hand side. Consider the stochastic boundary value problem: Find a random field $u : \overline{\mathcal{D}} \times \Omega \to \mathbb{R}$, such that P-almost everywhere in Ω (almost surely), there holds:

$$\begin{aligned} \frac{\partial u(\cdot,\omega)}{\partial t} + \nabla \cdot (\boldsymbol{V}(\cdot,\omega)u(\cdot,\omega) - D\nabla u(\cdot,\omega)) &= f(\cdot,\omega), \quad \text{on } \mathcal{D} \\ u(\boldsymbol{x}, t=0, \omega) &= g(\boldsymbol{x}), \quad \text{on } D \\ u(\cdot, t, \omega) &= 0, \quad \text{on } \partial D \times [0,T] \end{aligned} \tag{6.1}$$

In contrast to the deterministic case, all functions involved are modelled as random fields.

As in the non-stochastic case define

$$\mathcal{V} = L^2\left([0,T]; H_0^1(D)\right).$$

Furthermore

$$L_P^q(\Omega) = \{Y : \Omega \mapsto \mathbb{R} | \int_\Omega Y^q(\omega) dP(\omega) < \infty\}, \;\; 1 \le q < \infty.$$

Let $\boldsymbol{\xi}$ be a $\mathbb{R}^M$-valued random variable and assume that $\boldsymbol{\xi} \in L_P^1(\Omega)$ has a density function $\rho_{\boldsymbol{\xi}} : \mathbb{R}^M \to [0,\infty)$. Then its expected value is denoted by

$$E[\boldsymbol{\xi}] = \int_\Omega \boldsymbol{\xi}(\omega) dP(\omega) = \int_{\mathbb{R}^M} \boldsymbol{\xi} \rho_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}.$$

Whenever $\xi_i \in L_P^2(\Omega)$ for $i = 1, \ldots, M$, the covariance matrix $\mathrm{cov}[\boldsymbol{\xi}] \in \mathbb{R}^{M \times M}$ of $\boldsymbol{\xi}$ is defined by

$$\mathrm{cov}(\xi_i, \xi_j) = E[(\xi_i - E[\xi_i])(\xi_j - E[\xi_j])], \qquad i,j = 1, \ldots, M.$$

$V, f : \mathcal{D} \times \Omega \to \mathbb{R}$ are random fields. They are defined to be elements of the space

$$\tilde{\mathcal{V}} = \mathcal{V} \otimes L_P^2(\Omega).$$

6.1.2 Weak Formulation

Define the bilinear form $\mathcal{B} : \tilde{\mathcal{V}} \times \tilde{\mathcal{V}} \to \mathbb{R}$

$$\mathcal{B}(v,w) \equiv E\left[\int_0^T \int_D (v_t w - (\boldsymbol{V} v - D\nabla v) \cdot \nabla w) d\boldsymbol{x} dt\right] \tag{6.2}$$

and the linear functional

$$\mathcal{L}(w) \equiv E\left[\int_0^T \int_D f w \, d\boldsymbol{x} dt\right] = \int_\Omega \int_0^T \int_D f(\boldsymbol{x}, t, \omega) w(\boldsymbol{x}, t, \omega) d\boldsymbol{x} dt \, dP(\omega), \quad \forall w \in \tilde{\mathcal{V}} \tag{6.3}$$

Then, the weak formulation can be written as:

Find $u \in \tilde{\mathcal{V}}$, such that

$$\mathcal{B}(u, w) = \mathcal{L}(v), \qquad \forall \, w \in \tilde{\mathcal{V}}. \tag{6.4}$$

6.1.3 Finite Dimensional Approximation of the Stochastic Coefficients

Assume that the input random fields $\boldsymbol{V}$ and f can be approximated using just a small number of mutually uncorrelated, possibly mutually independent, random variables $\{\xi_i\}_{i=1}^M$ and that they depend, aside from ω, either only on $\boldsymbol{x}$ or only on t. Consider the case of space-dependent random fields

$$\boldsymbol{V}(\boldsymbol{x}, \omega) = \boldsymbol{V}(\boldsymbol{x}, \xi_1(\omega), \ldots, \xi_M(\omega)) \qquad \text{and} \qquad f(\boldsymbol{x}, \omega) = f(\boldsymbol{x}, \xi_1(\omega), \ldots, \xi_M(\omega)).$$

For example, this is the case if $\boldsymbol{V}$ and f can be approximated by a truncated Karhunen-Loève expansion [1], [14].

Whenever a numerical method is applied to (6.1), it is assumed that $\{\xi_i\}_{i=1}^M$ are real random variables with mean value zero and unit variance which are mutually independent. Their images $\Gamma_i \equiv \xi_i(\Omega)$ are intervals in $\mathbb{R}$. Moreover, assume that each ξ_i has a density function $\rho_i : \Gamma_i \to \mathbb{R}^+$.

In the following, the notation

$$\rho(\boldsymbol{\xi}) = \prod_{i=1}^M \rho_i(\xi_i) \qquad \forall \, \boldsymbol{\xi} \in \Gamma$$

is used for the joint probability density of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)$ and

$$\Gamma \equiv \Gamma_1 \times \Gamma_2 \times \ldots \times \Gamma_M \subset \mathbb{R}^M$$

for the support of the joint density.

If $\boldsymbol{V}$ and f fulfill these assumptions, then the same holds for the solution u. Now, the stochastic variational formulation (6.4) has a deterministic equivalent:

Find $u \in \mathcal{V} \otimes L^2_\rho(\Gamma)$ such that

$$\begin{aligned} &\int_\Gamma \rho(\boldsymbol{\xi}) \int_0^T \int_D [u_t w + (\boldsymbol{V} u - D\nabla u) \cdot \nabla w](\boldsymbol{x}, t, \xi) d\boldsymbol{x} dt \, d\boldsymbol{\xi} \\ &\quad = \int_\Gamma \rho(\boldsymbol{\xi}) \int_0^T \int_D f(\boldsymbol{x}, \boldsymbol{\xi}) w(\boldsymbol{x}, t, \boldsymbol{\xi}) d\boldsymbol{x} dt \, d\boldsymbol{\xi}, \qquad \forall \, w \in \mathcal{V} \otimes L^2_\rho(\Gamma) \end{aligned} \tag{6.5}$$

with

$$L^2_\rho(\Gamma) = \left\{ w : \Gamma \to \mathbb{R} | \int_\Gamma \rho(\boldsymbol{\xi}) w^2(\boldsymbol{\xi}) d\boldsymbol{\xi} < \infty \right\}.$$

The corresponding strong formulation is a partial differential equation containing the M-dimensional parameter $\boldsymbol{\xi}$, i.e.

$$\begin{aligned} \frac{\partial u(\cdot,\boldsymbol{\xi})}{\partial t} + \nabla\cdot(\boldsymbol{V}(\cdot,\boldsymbol{\xi})u(\cdot,\boldsymbol{\xi}) - D\nabla u(\cdot,\boldsymbol{\xi})) &= f(\cdot,\boldsymbol{\xi}), \quad \forall\,(\boldsymbol{x},t,\boldsymbol{\xi}) \in \mathcal{D}\times\Gamma \\ u(\boldsymbol{x},t=0,\boldsymbol{\xi}) &= g(\boldsymbol{x}), \quad \forall\,(\boldsymbol{x},\boldsymbol{\xi}) \in D\times\Gamma \\ u(\cdot,t,\boldsymbol{\xi}) &= 0, \quad \forall\,(\boldsymbol{x},t,\boldsymbol{\xi}) \in \partial D\times[0,T]\times\Gamma. \end{aligned} \tag{6.6}$$

This allows the use of finite element techniques to approximate the solution of the deterministic problem.

6.1.3.1 Karhunen-Loève Expansion

Let $a(\boldsymbol{x},\omega)$ be a stochastic process with a continuous covariance function $\mathrm{cov}[a] : \overline{D}\times\overline{D} \to \mathbb{R}$. Let $\{(\lambda_i, b_i(\boldsymbol{x}))\}_{i=1}^{\infty}$ denote the sequence of eigenpairs associated with the compact self adjoint operator that maps

$$f \in L^2(D) \mapsto \int_D \mathrm{cov}[a](\boldsymbol{x},\cdot)f(\boldsymbol{x})d\boldsymbol{x} \in L^2(D),$$

i.e. the λ_i and b_i satisfy the eigenvalue equation

$$\int_D \mathrm{cov}[a](\boldsymbol{x}_1,\boldsymbol{x}_2)b(\boldsymbol{x}_1)d\boldsymbol{x}_1 = \lambda b(\boldsymbol{x}_2).$$

The non-negative eigenvalues satisfy $\sum_{i=1}^{\infty}\lambda_i = \int_D \mathrm{var}[a](\boldsymbol{x})d\boldsymbol{x}$. The corresponding eigenfunctions are orthogonal, and we assume them to be normalized, such that

$$\int_D b_i(\boldsymbol{x})b_j(\boldsymbol{x})d\boldsymbol{x} = \delta_{ij}.$$

The truncated Karhunen-Loève expansion of the stochastic process a is

$$a_M(\boldsymbol{x},\omega) = E[a](\boldsymbol{x}) + \sum_{i=1}^{M}\sqrt{\lambda_i}b_i(\boldsymbol{x})\xi_i(\omega)$$

where the real random variables $\{\xi_i\}_{i=1}^{\infty}$ are mutually uncorrelated, have mean zero and unit variance. They are uniquely determined by

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}}\int_D (a(\boldsymbol{x},\omega) - E[a](\boldsymbol{x}))b_i(\boldsymbol{x})d\boldsymbol{x}.$$

Mercer's theorem [17] states that

$$\sup_{\boldsymbol{x}\in D} E[(a-a_M)^2](\boldsymbol{x}) = \sup_{\boldsymbol{x}\in D}\sum_{i=M+1}^{\infty}\lambda_i b_i^2(\boldsymbol{x}) \to 0 \qquad \text{for } M\to\infty.$$

6.1.4 Finite Element Spaces

In this work, we apply the $p \times h$ method proposed by Babuška [1]. This means the finite element spaces use global polynomials of degree p in $\boldsymbol{\xi}$ and piecewise polynomials in $\boldsymbol{x}$.

For the approximation of $\mathcal{V}$ consider any finite dimensional subspace $\mathcal{V}_h \subset \mathcal{V}$ appropriate for the non-stochastic case.

To discretize functions defined on $\Gamma \subset \mathbb{R}^M$, consider the subspace $Z^{\boldsymbol{p}} \subset L^2_\rho(\Gamma)$,

$$Z^{\boldsymbol{p}} = \bigotimes_{i=1}^{M} Z_i^{p_i}, \qquad \dim Z^{\boldsymbol{p}} = \prod_{i=1}^{M}(1+p_i),$$

where $Z_i^{p_i}$ denotes the space of polynomials of degree p_i in ξ_i, i.e.

$$Z_i^{p_i} = \{v : \Gamma_i \to \mathbb{R} \mid v \in \mathscr{P}_{p_i}(\xi_i)\}, \qquad i = 1, \dots, M.$$

For simplicity, consider $\boldsymbol{p} = \{p, \dots, p\}$, i.e. $p_1 = p_2 = \ldots = p_M =: p$ and $\dim Z^p = (p+1)^M := P$.

6.1.5 Discrete Formulation

The discrete problem can be formulated as follows:

Find $u_h^p \in \mathcal{V}_h \otimes Z^p$, such that

$$\begin{aligned}\int_\Gamma \rho(\boldsymbol{\xi}) \int_0^T \int_D \left(\frac{\partial u_h^p}{\partial_t} w + (V u_h^p - D\nabla u_h^p) \cdot \nabla w(x,t,\xi) \right) d\boldsymbol{x} dt d\boldsymbol{\xi} \\ = \int_\Gamma \rho(\boldsymbol{\xi}) \int_0^T \int_D f(\boldsymbol{x},\boldsymbol{\xi}) w(\boldsymbol{x},t,\boldsymbol{\xi}) d\boldsymbol{x} dt d\boldsymbol{\xi}, \qquad \forall\, w \in \mathcal{V}_h \otimes Z^p. \end{aligned} \tag{6.7}$$

Let $\{\psi_j(\boldsymbol{\xi})\}$ be a basis of Z^p and $\{\varphi_i(\boldsymbol{x},t)\}$ a basis of $\mathcal{V}_h$. The approximating solution can be written as

$$u_h^p(\boldsymbol{x},t,\boldsymbol{\xi}) = \sum_i \sum_j u_{ij} \psi_j(\boldsymbol{\xi}) \varphi_i(\boldsymbol{x},t).$$

With the test functions $w(\boldsymbol{x},t,\boldsymbol{\xi}) = \psi_k(\boldsymbol{\xi})\varphi_l(\boldsymbol{x},t)$ one obtains

$$\begin{aligned}&\sum_{i,j} \Bigg[\int_\Gamma \Big(\underbrace{\int_0^T \int_D \frac{\partial \varphi_i(\boldsymbol{x},t)}{\partial t} \varphi_l(\boldsymbol{x},t) + (\boldsymbol{V}(\boldsymbol{x},\boldsymbol{\xi})\varphi_i(\boldsymbol{x},t) - D\nabla\varphi_i(\boldsymbol{x},t)) \cdot \nabla\varphi_l(\boldsymbol{x},t) d\boldsymbol{x} dt}_{=:[K(\boldsymbol{\xi})]_{i,l}} \Big) \\ &\rho(\boldsymbol{\xi})\psi_k(\boldsymbol{\xi})\psi_j(\boldsymbol{\xi}) d\boldsymbol{\xi} \Bigg] u_{ij} \\ &= \int_\Gamma \rho(\boldsymbol{\xi})\psi_k(\boldsymbol{\xi}) \underbrace{\int_0^T \int_D f(\boldsymbol{x},\boldsymbol{\xi})\varphi_l(\boldsymbol{x},t) d\boldsymbol{x} dt}_{[f(\boldsymbol{\xi})]_l} d\boldsymbol{\xi}, \quad \forall\, k,l. \end{aligned} \tag{6.8}$$

Now, insert the KL-expansion of $\boldsymbol{V}$

$$\boldsymbol{V}(\boldsymbol{x},\boldsymbol{\xi}) = \overline{\boldsymbol{V}} + \sum_{s=1}^{M} \boldsymbol{b}_s(\boldsymbol{x})\xi_s$$

into (6.8). This results in

$$[K(\boldsymbol{\xi})]_{i,l} := \tag{6.9}$$
$$\int_0^T \int_D \frac{\partial \varphi_i(\boldsymbol{x},t)}{\partial t}\varphi_l(\boldsymbol{x},t) + ((\overline{\boldsymbol{V}} + \sum_{s=1}^{M} \boldsymbol{b}_s(\boldsymbol{x})\xi_s)\varphi_i(\boldsymbol{x},t) - D\nabla\varphi_i(\boldsymbol{x},t)) \cdot \nabla\varphi_l(\boldsymbol{x},t)d\boldsymbol{x}dt$$
$$= [K^{(0)}]_{i,l} + \sum_{s=1}^{M} \xi_s [K^{(s)}]_{i,l}$$

with

$$[K^{(0)}]_{i,l} := \int_0^T \int_D \frac{\partial \varphi_i(\boldsymbol{x},t)}{\partial t}\varphi_l(\boldsymbol{x},t) + (\overline{\boldsymbol{V}}\varphi_i(\boldsymbol{x},t) - D\nabla\varphi_i(\boldsymbol{x},t)) \cdot \nabla\varphi_l(\boldsymbol{x},t)d\boldsymbol{x}dt,$$

$$[K^{(s)}]_{i,l} := \int_0^T \int_D \boldsymbol{b}_s(\boldsymbol{x})\varphi_i(\boldsymbol{x},t) \cdot \nabla\varphi_l(\boldsymbol{x},t)d\boldsymbol{x}dt.$$

Furthermore, we introduce the notations

$$\begin{aligned} [G^{(0)}]_{k,j} &:= \langle \psi_k \psi_j \rangle = E[\psi_k(\boldsymbol{\xi})\psi_j(\boldsymbol{\xi})], \qquad k,j = 1,\ldots,P = \prod_{s=1}^{M}(1+p_s) \\ [G^{(s)}]_{k,j} &:= \langle \xi_s \psi_k \psi_j \rangle = E[\xi_s \psi_k(\boldsymbol{\xi})\psi_j(\boldsymbol{\xi})]. \end{aligned}$$

Then the global matrix in the linear system of equations $A\boldsymbol{u} = \boldsymbol{f}$ can be written as

$$A = G^{(0)} \otimes K^{(0)} + \sum_{s=1}^{M} G^{(s)} \otimes K^{(s)}. \tag{6.10}$$

The same procedure applied to the right hand side with

$$f(\boldsymbol{x},\boldsymbol{\xi}) = \overline{f} + \sum_{s=1}^{M} c_s(\boldsymbol{x})\xi_s,$$

$$[f^{(0)}]_l := \int_0^T \int_D \overline{f}\varphi_l(\boldsymbol{x},t)d\boldsymbol{x}dt, \quad [f^{(s)}]_l := \int_0^T \int_D c_s(\boldsymbol{x})\varphi_l(\boldsymbol{x},t)d\boldsymbol{x}dt,$$

$$[q^{(0)}]_k := \langle \psi_k \rangle = E[\psi_k(\boldsymbol{\xi})], \quad [q^{(s)}]_k := \langle \xi_s \psi_k \rangle = E[\xi_s \psi_k(\boldsymbol{\xi})]$$

results in

$$\boldsymbol{f} = [q^{(0)}] \otimes [f^{(0)}] + \sum_{s=1}^{M} [q^{(s)}] \otimes [f^{(s)}].$$

6.1.5.1 Choice of Stochastic Basis Functions

To decouple the integral over Γ with respect to the Γ_s, the basis functions $\psi_k \in Z^{\boldsymbol{p}}$ are represented as

$$\psi_k(\boldsymbol{\xi}) = \prod_{s=1}^{M} \psi_{k_s}^s(\xi_s),$$

where $\psi_{k_s} : \Gamma_s \to \mathbb{R}$ is a basis function of Z^{p_s}, i.e. a polynomial in ξ_s.

The question arises of how to choose the functions $\psi_{k_s}^s$ so that the calculation becomes efficient. Babuška proposed the use of double-orthogonal polynomials, i.e. for $s = 1, \ldots, M$ they must satisfy

$$\int_{\Gamma_s} \rho_s(\xi_s)\psi_k^s(\xi_s)\psi_j^s(\xi_s)d\xi_s = \delta_{kj}, \qquad j,k = 0,\ldots,p \tag{6.11}$$

$$\int_{\Gamma_s} \xi_s\rho_s(\xi_s)\psi_k^s(\xi_s)\psi_j^s(\xi_s)d\xi_s = c_k^s\delta_{kj}, \qquad j,k = 0,\ldots,p. \tag{6.12}$$

This implies the decoupling of (6.8) with respect to the ξ_s. (6.11) yields

$$[G^{(0)}]_{k,j} = \prod_{s=1}^{M} \int_{\Gamma_s} \rho_s(\xi_s)\psi_{k_s}^s(\xi_s)\psi_{j_s}^s(\xi_s)d\xi_s = \delta_{kj}, \tag{6.13}$$

and (6.12) yields

$$\begin{aligned}[G^{(s)}]_{k,j} &= \int_{\Gamma_s} \xi_s\rho_s(\xi_s)\psi_{k_s}^s(\xi_s)\psi_{j_s}^s(\xi_s)d\xi_n \prod_{m\neq s, m=1}^{M} \int_{\Gamma_m} \rho_m(\xi_m)\psi_{k_m}^m(\xi_m)\psi_{j_m}^m(\xi_m)d\xi_m \\ &= c_{k_s}^s\delta_{kj}. \end{aligned} \tag{6.14}$$

Thus the matrices $G^{(0)}$ and $G^{(s)}$ become diagonal and (6.10) simplifies to

$$A = ([K^{(0)}] + \sum_{s=1}^{M} c_{k_s}^s[K^{(s)}])\delta_{kj}. \tag{6.15}$$

The structure of the system that determines u_{ij} thus becomes block diagonal where the number of blocks corresponds to the number of basis functions or the dimension of $Z^{\boldsymbol{p}}$ respectively.

To find the polynomials that fulfill (6.11) and (6.12) one has to solve M eigenproblems, each of them with size $(1 + p_s) = (1 + p)$. Regarding (6.11) and (6.12), the idea is to choose the basis functions as linear combinations of orthogonal polynomials,

$$\psi_k^s(\xi_s) = \sum_{i=0}^{p} \sigma_{ki}^s P_i^s(\xi_s), \tag{6.16}$$

with polynomials $P_i(\xi)$ of degree $i \leq p$. Define

$$S^{(s)} = \{\sigma_{ij}^s\}_{i,j=0}^{p},$$

$$M_1^{(s)}(i,l) := \int_{\Gamma_s} \xi_s \rho(\xi_s) P_i^s(\xi_s) P_l^s(\xi_s) d\xi_s,$$
$$M_2^{(s)}(i,l) := \int_{\Gamma_s} \rho(\xi_s) P_i^s(\xi_s) P_l^s(\xi_s) d\xi_s.$$

The orthogonality relations (6.11) and (6.12) thus become

$$\sum_{i=0}^{p}\sum_{l=0}^{p} \sigma_{ki}^s \sigma_{jl}^s \int_{\Gamma_s} \rho(\xi_s) P_i^s(\xi_s) P_l^s(\xi_s) d\xi_s = \delta_{kj} \quad \Leftrightarrow \quad S^T M_2 S = I, \tag{6.17}$$
$$\sum_{i=0}^{p}\sum_{l=0}^{p} \sigma_{ki}^s \sigma_{jl}^s \int_{\Gamma_s} \xi_s \rho(\xi_s) P_i^s(\xi_s) P_l^s(\xi_s) d\xi_s = c_k^s \delta_{kj} \quad \Leftrightarrow \quad S^T M_1 S = C. \tag{6.18}$$

This corresponds to the generalized eigenvalue problem

$$M_1 S = M_2 S C.$$

This problem simplifies further if we choose the polynomials which are orthogonal with respect to the density function. This concept is associated with the generalized polynomial chaos or the Askey-chaos and was proposed by Xiu and Karniadakis [29]. In the original polynomial chaos, $\{\psi_n\}$ are the Hermite polynomials and $\boldsymbol{\xi}$ are the Gaussian random variables. However, this can be generalized to other random variables, see Table 6.1. Thus, M_2 becomes the identity and M_1 will be tridiagonal due to the three-term recursion formula which yields for the polynomials.

	random variables $\boldsymbol{\xi}$	orthogonal polynomials	support
continuous	Gaussian	Hermite	$(-\infty, \infty)$
	Gamma	Laguerre	$[0, \infty)$
	Beta	Jacobi	$[a, b]$
	Uniform	Legendre	$[a, b]$

Table 6.1: Correspondence of orthogonal polynomials and random variables

Example: Consider a Gaussian random variable ξ with mean zero and variance σ. Its density function is

$$\rho(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{\xi^2}{2\sigma^2}).$$

The Hermite polynomials $\tilde{H}_n(\xi)$ of degree n with highest coefficient 1 are given by:

$$\tilde{H}_n(\xi) = (-1)^n e^{\frac{\xi^2}{2}} \frac{d^n}{dx^n}(e^{-\frac{\xi^2}{2}}).$$

They fulfill an orthogonality relation and a three term recursion formula:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \tilde{H}_i(\xi) \tilde{H}_l(\xi) d\xi = (i!)\delta_{il},$$

$$\tilde{H}_{i+1}(\xi) - \xi\tilde{H}_i(\xi) + i\tilde{H}_{i-1}(\xi) = 0.$$

Choose

$$H_i(\xi) = \frac{\tilde{H}_i(\xi/\sigma)}{\sqrt{i!}},$$

and set

$$\psi_k^s(\xi_s) = \sum_{i=0}^{p} \sigma_{ki}^s H_i^s(\xi_s). \tag{6.19}$$

In the sequel, the index s is omitted. For the matrices M_1 and M_2 there holds:

$$\begin{aligned}
M_2(i,l) &= \int_{-\infty}^{\infty} \rho(\xi) H_i(\xi) H_l(\xi) d\xi \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\xi^2}{2\sigma^2}} H_i(\xi) H_l(\xi) d\xi \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \frac{\tilde{H}_i(\xi)}{\sqrt{i!}} \frac{\tilde{H}_l(\xi)}{\sqrt{l!}} \, \sigma \, d\xi \\
&= \frac{1}{\sqrt{i!}} \frac{1}{\sqrt{l!}} \, i! \, \delta_{il} \\
&= \delta_{il}
\end{aligned}$$

$$\begin{aligned}
M_1(i,l) &= \int_{-\infty}^{\infty} \xi\rho(\xi) H_i(\xi) H_l(\xi) d\xi \\
&= \int_{-\infty}^{\infty} \xi \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\xi^2}{2\sigma^2}} H_i(\xi) H_l(\xi) d\xi \\
&= \int_{-\infty}^{\infty} \sigma\xi \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \frac{\tilde{H}_i(\xi)}{\sqrt{i!}} \frac{\tilde{H}_l(\xi)}{\sqrt{l!}} \sigma d\xi \\
&= \sigma \frac{1}{\sqrt{i!}} \frac{1}{\sqrt{l!}} \int_{-\infty}^{\infty} \xi \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \tilde{H}_i(\xi) \tilde{H}_l(\xi) d\xi \\
&= \sigma \frac{1}{\sqrt{i!}} \frac{1}{\sqrt{l!}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} [\tilde{H}_{i+1}(\xi) + i\tilde{H}_{i-1}(\xi)] \tilde{H}_l(\xi) d\xi \\
&= \sigma \frac{1}{\sqrt{i!}} \frac{1}{\sqrt{l!}} [i!\delta_{l,i-1} + (i+1)!\delta_{l,i+1}] \\
&= \sigma[\sqrt{i}\delta_{l,i-1} + \sqrt{i+1}\delta_{l,i+1}]
\end{aligned}$$

□

Note that the outcome of the eigenvalue problem can be used to compute $[q^{(0)}]$ and $[q^{(s)}]$. Since $P_0(\xi) \equiv 1$, there holds

$$[q^{(0)}]_k = \prod_{s=1}^{M} \int_{\Gamma_s} \rho(\xi_s) \left(\sum_{i=0}^{p} \sigma_{k_s i}^s P_i^s(\xi_s) \right) P_0^s(\xi_s) d\xi_s = \prod_{s=1}^{M} \sigma_{k_s 0}^s M_2^{(s)}(0,0),$$

$$
\begin{aligned}
[q^{(s)}]_k &= \left(\int_{\Gamma_s} \xi_s \rho(\xi_s) \left(\sum_{i=0}^{p} \sigma^s_{k_s i} P^s_i(\xi_s) \right) P^s_0(\xi_s) d\xi_s \right) \\
&\quad \times \prod_{m=1, m\neq s}^{M} \int_{\Gamma_m} \rho(\xi_m) \left(\sum_{i=0}^{p} \sigma^m_{k_m i} P^m_i(\xi_m) \right) P^m_0(\xi) d\xi_m \\
&= \sigma^s_{k_s 1} M^{(s)}_1(1,0) \prod_{m=1, m\neq s}^{M} \sigma^m_{k_m 0} M^{(m)}_2(0,0).
\end{aligned}
$$

6.1.5.2 Incorporation of Initial Data

In order to start the calculation, the values of u at time $t = 0$ must be known. There holds

$$
g(\boldsymbol{x}) = u(\boldsymbol{x}, t = 0, \boldsymbol{\xi}) = \sum_i \sum_j u_{ij} \psi_j(\boldsymbol{\xi}) \varphi_i(\boldsymbol{x}, t = 0) =: \sum_i \sum_j u^0_{ij} \psi_j(\boldsymbol{\xi}) \varphi_i(\boldsymbol{x}).
$$

With nodal basis functions with respect to $\boldsymbol{x}$, this expression can be written as

$$
g(x_l) = u(\boldsymbol{x} = \boldsymbol{x}_l, t = 0, \xi) = \sum_j u^0_{lj} \psi_j(\boldsymbol{\xi}).
$$

It follows

$$
g(\boldsymbol{x}_l) \int_\Gamma \rho(\boldsymbol{\xi}) \psi_k(\boldsymbol{\xi}) d\boldsymbol{\xi} = \sum_j u^0_{lj} \int_\Gamma \rho(\boldsymbol{\xi}) \psi_j(\boldsymbol{\xi}) \psi_k(\boldsymbol{\xi}) = u^0_{lk},
$$

or, using the above notations,

$$
g(\boldsymbol{x}_l) E[\psi_k(\boldsymbol{\xi})] = g(\boldsymbol{x}_l)[q^{(0)}]_k = u^0_{lk}, \qquad \text{for } \; k = 1, \ldots, P.
$$

Boundary conditions can be treated in the same way. In contrast to the deterministic case, the nodal values must be multiplied with the mean value of the corresponding stochastic basis function.

6.2 The Advection-Diffusion Equation with a Stochastic Velocity Field

In this section we consider the one-dimensional linear advection equation with a random transport velocity as a prototype problem. This problem is taken from [11]. Under certain conditions it is possible to derive an exact solution for the mean and the variance of the stochastic solution. The stochastic input is represented by a Karhunen-Loève expansion which is derived from a given covariance kernel.

Consider the stochastic advection equation

$$
\frac{\partial u}{\partial t} + V(x, t, \omega) \frac{\partial u}{\partial x} = 0, \qquad \forall (x, t) \in (D = [-1, 1]) \times [0, T] \tag{6.20}
$$

with initial condition

$$u(x, t = 0) = g(x) = \sin(\pi(x + 1)) \tag{6.21}$$

and periodic boundary conditions

$$u(-1, t) = u(1, t) \qquad \forall t \in [0, T].$$

6.2.1 Exact Solution of the Stochastic Advection Equation

Assume that the transport velocity $V(t, \omega)$ is independent of x and strictly positive. In this case the random field V can be removed from (6.20) by a change of variables. Define

$$\tau = \tau(t, \omega) = \int_0^t V(s, \omega) ds, \quad \tau(t = 0) = 0.$$

Then, (6.20) can be rewritten as

$$\frac{\partial u}{\partial \tau} + \frac{\partial u}{\partial x} = 0, \quad u(x, 0) = g(x).$$

The solution to this equation is given by

$$u(x, t) = g(x - \tau).$$

Assume further that τ can be expressed as the sum of a deterministic mean value and a perturbation which depends on a single normal random variable ξ, i.e.

$$\tau(t, \xi) =: \overline{V}t + \sigma\xi, \quad \xi \sim N(0, 1), \quad \rho(\xi) = \frac{1}{\sqrt{2\pi}} \exp(-\xi^2/2). \tag{6.22}$$

The mean value of the solution $u(x, t) = g(x - \tau(t))$ can be calculated by

$$\begin{aligned} E[u(x, t)] &= \int_{-\infty}^{\infty} \rho(\xi) g(x - \tau(t, \xi)) d\xi \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\xi^2/2) g(x - \tau(t, \xi)) d\xi. \end{aligned} \tag{6.23}$$

A change of variables $x_0 := x - \tau(t, \xi)$ leads to

$$\begin{aligned} E[u(x, t)] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp(-\frac{(x - x_0 - \overline{V}t)^2}{2\sigma^2}) g(x_0) dx_0 \\ &= \sin(\pi(x + 1 - \overline{V}t)) e^{-\pi^2\sigma^2/2}. \end{aligned} \tag{6.24}$$

It is important to notice that the stochastic solution equals the deterministic solution $\sin(\pi(x+1-\overline{V}t))$ multiplied with a damping factor $e^{-\pi^2\sigma^2/2}$. Hence, uncertainties have a smoothing influence on the advection process. The bigger the variance σ of the stochastic input, the stronger the damping. A similar calculation yields

$$\text{var}[u(x, t)] = \frac{1}{2}(1 - e^{-\pi^2\sigma^2})[1 + \cos(2\pi(x + 1 - \overline{V}t)) e^{-\pi^2\sigma^2}].$$

We point out that σ should depend on the time variable t. There should hold

$$\lim_{t\to 0}\sigma(t) = 0.$$

Otherwise, the expected value of the solution will not fulfill the initial condition. The special form of σ is determined by the representation of the velocity field V.

6.2.2 Representation of Stochastic Input

There are several possibilities to obtain the representation for τ given in (6.22). A first approach is based on the Karhunen-Loève expansion which can be derived from a given covariance structure. A second approach is based on a representation of τ as a stochastic process.

6.2.2.1 Covariance Kernel

Assume the random variables ξ_k in the Karhunen-Loève expansion of τ are standard normally distributed, i.e.

$$V(t,\omega) = \overline{V} + \sum_{k=1}^{M} \sqrt{\lambda_k} b_k(t)\xi_k(\omega), \qquad \xi_k \sim N(0,1). \tag{6.25}$$

In the following, set $b_k(t) := \sqrt{\lambda_k} b_k(t)$. Hence,

$$\tau(t) = \overline{V}t + \sum_{k=1}^{M} \xi_k \int_0^t b_k(s)ds = \overline{V}t + \sum_{k=1}^{M} \xi_k a_k(t)$$

with

$$a_k(t) := \int_0^t b_k(s)ds.$$

For example, when the functions $b_k(t)$ are determined numerically from the covariance kernel as piecewise linear functions, the integral can be approximated by

$$\int_0^{i\Delta t} b_k(s)ds = \Delta t \left[\frac{1}{2}(b_k(0) + b_k(i\Delta t)) + \sum_{j=1}^{i-1} b_k(j\Delta t)\right].$$

The sum $\sum_{k=1}^{M} \xi_k a_k$ forms a new Gaussian random field which has also mean value zero but variance $\sum_{k=1}^{M} a_k^2(t)$. With the definition

$$\sigma(t) := \sqrt{\sum_{k=1}^{M} a_k^2(t)}$$

τ can be rewritten as

$$\tau(t) = \overline{V}t + \sigma(t)\xi, \qquad \xi \sim N(0,1).$$

Indeed, σ fulfills $\lim_{t\to 0}\sigma(t) = 0$.

The Karhunen-Loève expansion of V can be calculated from its covariance kernel.

Example: The 1D Exponential Covariance Eigenproblem Consider the exponential covariance function

$$\begin{aligned} k : [-a,a] \times [-a,a] &\to \mathbb{R}, \\ (x,y) &\mapsto k(x,y) = e^{-c|x-y|} \end{aligned}$$

with correlation length $b = \frac{1}{c} > 0$ and $a > 0$. The associated covariance operator $C : C[-a,a] \to C[-a,a]$ with kernel k is defined by

$$(Cu)(x) = \int_{-a}^{a} k(x,y)u(y)\,dy, \qquad x \in [-a,a].$$

The eigenvalue problem

$$Cu = \lambda u,$$

may be rewritten as

$$\int_{-a}^{x} e^{-c(x-y)}u(y)\,dy + \int_{x}^{a} e^{c(x-y)}u(y)\,dy = \lambda u(x). \tag{6.26}$$

Differentiating (6.26) with respect to x yields

$$-c\int_{-a}^{x} e^{-c(x-y)}u(y)\,dy + c\int_{x}^{a} e^{c(x-y)}u(y)\,dy = \lambda u'(x), \tag{6.27}$$

and once more results in

$$(c^2\lambda - 2c)u = \lambda u'' \ . \tag{6.28}$$

Introduce the new variable

$$\omega^2 = \frac{2c - c^2\lambda}{\lambda} \tag{6.29}$$

($\lambda = 0$ is excluded as this would imply $u \equiv 0$ since $c > 0$), and (6.28) becomes

$$u''(x) + \omega^2 u(x) = 0, \qquad x \in [-a,a]. \tag{6.30}$$

Boundary conditions for (6.30) can be derived by evaluating (6.26) and (6.27) at $x = -a$ and $x = a$, yielding

$$\begin{aligned} \int_{-a}^{a} e^{c(-a-y)}u(y)\,dy &= \lambda u(-a), & c\int_{-a}^{a} e^{c(-a-y)}u(y)\,dy &= \lambda u'(-a), \\ \int_{-a}^{a} e^{-c(a-y)}u(y)\,dy &= \lambda u(a), & -c\int_{-a}^{a} e^{-c(a-y)}u(y)\,dy &= \lambda u'(a), \end{aligned}$$

from which one concludes, since $\lambda \neq 0$, the conditions

$$cu(-a) - u'(-a) = 0, \tag{6.31a}$$

$$cu(a) + u'(a) = 0. \tag{6.31b}$$

The general solution $u(x) = A\cos\omega x + B\sin\omega x$ of (6.30) satisfies the boundary conditions (6.31) for nontrivial values of A and B if and only if ω is a solution of the equation

$$\left(\tan\omega a - \frac{c}{\omega}\right)\left(\tan\omega a + \frac{\omega}{c}\right) = 0. \tag{6.32}$$

The solution can be obtained by a Newton iteration, for example.
For $n = 0, 1, 2, \ldots$, denote by

$$\begin{array}{lll} \omega_{2n} & \text{the solutions of} & c - \omega \tan \omega a = 0, \text{ and} \\ \omega_{2n+1} & \text{the solutions of} & \omega + c \tan \omega a = 0, \end{array}$$

both arranged in increasing order. The associated eigenfunctions scaled to have unit L^2-norm are then given by

$$u_{2n}(x) = A_{2n} \cos \omega_{2n} x, \qquad A_{2n} = \left(a + \frac{\sin(2\omega_{2n} a)}{2\omega_{2n}} \right)^{-1/2},$$

$$u_{2n+1}(x) = B_{2n+1} \sin \omega_{2n+1} x, \qquad B_{2n+1} = \left(a - \frac{\sin(2\omega_{2n+1} a)}{2\omega_{2n}} \right)^{-1/2}$$

with eigenvalues, in view of (6.29),

$$\lambda_n = \frac{2c}{\omega_n^2 + c^2}, \qquad n = 0, 1, \ldots \ .$$

The solution of the same eigenvalue problem on an interval $[0, L]$, $L > 0$ can be obtained by the change of variables

$$\xi = \xi(x) = \frac{L}{2} \left(\frac{x}{a} + 1 \right) \in [0, L], \qquad x = x(\xi) = a \left(\frac{2\xi}{L} - 1 \right) \in [-a, a].$$

The kernel function in the new variables is

$$\tilde{k}(\xi, \eta) = k(x, y) = e^{-c|x-y|} = e^{-\frac{2ac}{L}|\xi - \eta|},$$

i.e., the reciprocal correlation length scales as $\tilde{c} = 2ac/L$. Similarly, from

$$\begin{aligned} \tilde{\lambda} \tilde{u}(\xi) = (\tilde{C}\tilde{u})(\xi) &= \int_0^L \tilde{k}(\xi, \eta) \tilde{u}(\eta) \, d\eta = \frac{L}{2a} \int_{-a}^{a} k(x, y) u(y) \, dy \\ &= \frac{L}{2a} (Cu)(x) = \frac{L}{2a} \lambda u(x) = \frac{L}{2a} \lambda \tilde{u}(\xi), \end{aligned}$$

we see that the eigenvalues transform as $\tilde{\lambda} = \lambda L/2a$, and therefore

$$\tilde{\lambda}_n = \frac{L}{2a} \lambda_n = \frac{L}{2a} \frac{2c}{\omega_n^2 + c^2} = \frac{2\tilde{c}}{\tilde{\omega}_n^2 + \tilde{c}^2}$$

with $\tilde{\omega}_n = 2a\omega_n/L$. The corresponding normalized eigenfunctions are thus

$$\tilde{u}_{2n}(\xi) = \tilde{A}_{2n} \cos \left(\tilde{\omega}_{2n} \left(\xi - \frac{L}{2} \right) \right),$$

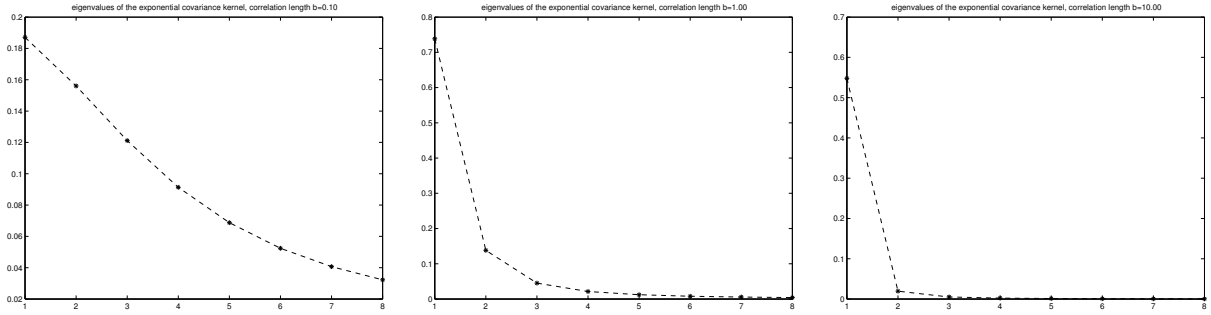
$$\tilde{u}_{2n+1}(\xi) = \tilde{B}_{2n+1} \sin \left(\tilde{\omega}_{2n+1} \left(\xi - \frac{L}{2} \right) \right)$$

with

$$\tilde{A}_{2n} = \left[\frac{1}{2} \left(L + \frac{\sin(\tilde{\omega}_{2n} L)}{\tilde{\omega}_{2n}} \right) \right]^{-1/2}, \quad \tilde{B}_{2n+1} = \left[\frac{1}{2} \left(L - \frac{\sin(\tilde{\omega}_{2n+1} L)}{\tilde{\omega}_{2n+1}} \right) \right]^{-1/2}.$$

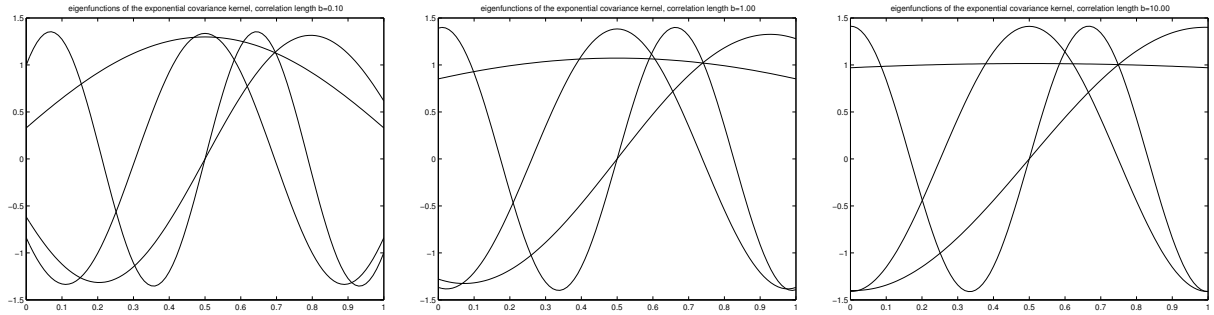
Figures 6.1 to 6.2 present the eigenvalues and eigenfunctions of the exponential covariance kernel on the interval $[0, 1]$.

□



$b = 0.1$ $b = 1$ $b = 10$

Figure 6.1: Eigenvalues of the exponential covariance kernel for different values of the correlation length



$b = 0.1$ $b = 1$ $b = 10$

Figure 6.2: Eigenfunctions of the exponential covariance kernel for different values of the correlation length

6.2.2.2 Dynamical System Approach

Assume that V is given as a stochastic process

$$\begin{aligned} V(t_0) &= v_0 = \overline{V}, \\ V(t) &= \overline{V} + v_1, \quad v_1 = a\xi_1, \quad \text{for } t_0 < t \le t_1 \\ V(t) &= \overline{V} + v_i, \quad v_i = cv_{i-1} + af\xi_i, \quad \text{for } t_{i-1} < t \le t_i, \quad i = 2, 3, \ldots, Q = \frac{T}{\Delta t}, \end{aligned}$$

where $\{\xi_1, \xi_2, \ldots, \xi_Q\}$ is a set of independent standard normal random variables. Moreover, $f^2 = 1 - c^2$, $0 \le c \le 1$, so that each v_j has variance a^2. The process is autoregressive of order one and corresponds to a Markov process [11].

Then τ is given by

$$\tau(t_Q) = \int_0^{t_Q} V(s)ds = \Delta t \sum_{i=1}^{Q} V(t_i) = \overline{V} t_Q + \Delta t \sum_{j=1}^{Q} v_j, \qquad Q = t_Q/\Delta t,$$

where the set of random variables v_j corresponds to the number of time increments. This representation of τ corresponds to the formula for τ given in [11]. For $c = 1$ (fully correlated input) one obtains

$$\sum_{j=1}^{Q} v_j = aQ\xi_1 := aQ\xi, \qquad \xi \sim N(0,1),$$

and $c = 0$ (mutually independent input) results in

$$\sum_{j=1}^{Q} v_j = a \sum_{j=1}^{Q} \xi_j := a\sqrt{Q}\xi, \qquad \xi \sim N(0,1).$$

Here, the fact is used that

$$\sum_{j=1}^{Q} k_j\xi_j \sim N(\sum_{j=1}^{Q} k_j\mu_j, \sum_{j=1}^{Q} k_j^2\sigma_j^2) \qquad \text{if} \quad \xi_j \sim N(\mu_j, \sigma_j^2).$$

The evaluation of the sum of random variables for $0 < c < 1$ (partially correlated input) results in

$$\begin{aligned} \sum_{j=1}^{Q} v_j &= a\xi_1 + (ca\xi_1 + af\xi_2) + (c^2a\xi_1 + caf\xi_2 + af\xi_3) + \ldots + af\xi_Q \\ &= \underbrace{\frac{c^Q - 1}{c-1}a}_{a_1}\xi_1 + \underbrace{\frac{c^{Q-1} - 1}{c-1}af}_{a_2}\xi_2 + \underbrace{\frac{c^{Q-2} - 1}{c-1}af}_{a_3}\xi_3 + \ldots + \underbrace{\frac{c^1 - 1}{c-1}af}_{a_Q}\xi_Q. \end{aligned}$$

The sum $\sum_{k=1}^{Q} a_k\xi_k$ forms a new random variable with variance

$$\begin{aligned}
\sum_{k=1}^{Q} a_k^2 &= a^2 \left[\left(\frac{c^Q - 1}{c - 1} \right)^2 + \frac{1+c}{1-c} \left((c^{Q-1} - 1)^2 + (c^{Q-2} - 1)^2 + \ldots + (c-1)^2 \right) \right] \\
&= a^2 \frac{1+c}{1-c} \left[\frac{(c^Q - 1)^2}{1 - c^2} + \sum_{k=1}^{Q-1} (c^2)^k - 2 \sum_{k=1}^{Q-1} c^k + Q - 1 \right] \\
&= a^2 \frac{1+c}{1-c} \left[\frac{(c^Q - 1)^2}{1 - c^2} + \frac{c^2 - c^{2Q}}{1 - c^2} - 2 \frac{c - c^Q}{1 - c} + Q - 1 \right] \\
&= a^2 \left[\frac{1+c}{1-c} Q - \frac{2c}{(1-c)^2} (1 - c^Q) \right].
\end{aligned}$$

Hence τ can be rewritten in terms of a new standard normal random variable ξ,

$$\tau(t) = \overline{V} t + \sigma(t) \xi, \qquad \xi \sim N(0,1)$$

with

$$\sigma^2(t) = \begin{cases} a^2 t^2 & \text{for} \quad c = 1, \\ a^2 \left[\dfrac{1+c}{1-c} Q - \dfrac{2c}{(1-c)^2} (1 - c^Q) \right] & \text{for} \quad 0 < c < 1, \\ a^2 (\Delta t) t & \text{for} \quad c = 0. \end{cases}$$

Although the exact solution can be determined in this case, V can also be approximated by its Karhunen-Loève expansion and a SGFEM solution can be calculated. For this purpose one has to construct the covariance kernel. There holds

$$v_n = c^{n-1} a \xi_1 + af \sum_{i=2}^{n} \xi_i c^{n-i}.$$

Consequently,

$$\begin{aligned}
\operatorname{cov}[V(t_n), V(t_m)] &= E[(c^{n-1} a \xi_1 + af \sum_{i=2}^{n} \xi_i c^{n-i})(c^{m-1} a \xi_1 + af \sum_{i=2}^{m} \xi_i c^{m-i})] \\
&= a^2 c^{n+m-2} + a^2 f^2 \sum_{i=2}^{\min(n,m)} c^{n+m-2i} \\
&= a^2 \left[\sum_{i=1}^{\min(n,m)} c^{n+m-2i} - c^2 \sum_{i=2}^{\min(n,m)} c^{n+m-2i} \right] \\
&= a^2 c^{n+m-2\min(n,m)} \\
&= a^2 c^{|n-m|}, \qquad n, m = 1, \ldots, Q,
\end{aligned}$$

and $\operatorname{cov}[V(t_n), V(t_m)] = 0$ if $n = 0$ or $m = 0$.

6.2.3 Stochastic Galerkin Finite Element Approximation

In the following, the formulas are presented for one dimension. It is first assumed that the velocity $V = V(x, \omega)$ is independent of time and given as a random process with mean value one and bounded variance, that is

$$E[V(x,\omega)] = 1, \qquad \text{and} \qquad E[V(x,\omega)^2] < \infty.$$

Assume V is given by the Karhunen-Loève expansion

$$V(x,\omega) = E[V] + \sum_{s=1}^{M} \sqrt{\lambda_s} b_s(x)\xi_s(\omega) =: \sum_{s=0}^{M} g_s(x)\xi_s(\omega), \tag{6.33}$$

where ξ_s are random variables with mean value zero and bounded variance.

We will consider the pure advection equation. However, the diffusive part can be added in a straight forward manner analogous to the deterministic case. Results of some numerical computations will be presented for the advection equation as well as for the advection-diffusion equation.

6.2.3.1 Method of Lines

The solution u is written as a tensor product of basis functions

$$u(x,t,\omega) = \sum_{k=1}^{N}\sum_{i=1}^{P} \alpha_{ik}(t)\varphi_k(x)\psi_i(\boldsymbol{\xi}(\omega)). \tag{6.34}$$

Substituting (6.33) and (6.34) in (6.20) yields

$$\frac{\partial}{\partial t}\sum_{k=1}^{N}\sum_{i=1}^{P} \alpha_{ik}(t)\varphi_k(x)\psi_i(\boldsymbol{\xi}) + \left(\sum_{s=0}^{M} g_s(x)\xi_s\right)\frac{\partial}{\partial x}\sum_{k=1}^{N}\sum_{i=1}^{P} \alpha_{ik}(t)\varphi_k(x)\psi_i(\boldsymbol{\xi}) = 0.$$

Now a standard Galerkin projection is applied. The above equation is multiplied with a test function $w(x,\omega) = \varphi_l(x)\psi_j(\boldsymbol{\xi}),\ (l = 1,\dots,N,\ j = 1,\dots,P)$, and the expected value is computed of both sides:

$$\sum_{k=1}^{N}\sum_{i=1}^{P} I_{kl}\frac{\partial \alpha_{ik}}{\partial t}\langle\psi_i(\boldsymbol{\xi})\psi_j(\boldsymbol{\xi})\rangle + \sum_{k=1}^{N}\sum_{i=1}^{P}\sum_{s=0}^{M} I_{skl}\alpha_{ik}(t)\langle\xi_s\psi_i(\boldsymbol{\xi})\psi_j(\boldsymbol{\xi})\rangle = 0, \tag{6.35}$$

with

$$I_{kl} = \int_D \varphi_k(x)\varphi_l(x)dx,\ I_{skl} = \int_D g_s(x)\frac{\partial\varphi_k(x)}{\partial x}\varphi_l(x)dx,$$

and

$$\langle\psi_i(\boldsymbol{\xi})\psi_j(\boldsymbol{\xi})\rangle = \int_\Gamma \rho(\boldsymbol{\xi})\psi_i(\xi)\psi_j(\boldsymbol{\xi})d\boldsymbol{\xi} = \delta_{ij},$$

$$\langle \xi_s \psi_i(\boldsymbol{\xi}) \psi_j(\boldsymbol{\xi}) \rangle = \int_\Gamma \rho(\boldsymbol{\xi}) \xi_s \psi_i(\boldsymbol{\xi}) \psi_j(\boldsymbol{\xi}) d\xi = c^s_{i_s} \delta_{ij}.$$

Hence, (6.35) simplifies to

$$\sum_{k=1}^{N} I_{kl} \frac{\partial \alpha_{jk}}{\partial t} + \sum_{k=1}^{N} \sum_{s=0}^{M} I_{skl} \alpha_{jk}(t) c^s_{j_s} = 0 \tag{6.36}$$

(6.36) represents a set of P ordinary differential equations which can be solved for α_{jk}.

A similar approach can be made if $V = V(t, \omega)$ is independent of the space variable x. In this case the final equation reads

$$\sum_{k=1}^{N} \sum_{i=1}^{P} I_{kl} \frac{\partial \alpha_{ik}}{\partial t} \langle \psi_i(\boldsymbol{\xi}) \psi_j(\boldsymbol{\xi}) \rangle + \sum_{k=1}^{N} \sum_{i=1}^{P} \sum_{s=0}^{M} J_{kl} g_s(t) \alpha_{ik}(t) \langle \xi_s \psi_i(\boldsymbol{\xi}) \psi_j(\boldsymbol{\xi}) \rangle = 0 \tag{6.37}$$

with

$$J_{kl} = \int_D \frac{\partial \varphi_k(x)}{\partial x} \varphi_l(x) dx.$$

In contrast to the former case, the coefficients in the ODE are now time dependent.

If the use of Gaussian random variables is not reasonable, other distributions can be applied, and a different basis of orthogonal polynomials has to be used.

Example 1: Figures 6.3 to 6.5 present the results of the stochastic Galerkin FEM for (6.20) with $V = E[V] + \xi$, $E[V] = 1$, and $\xi \sim N(0, \sigma^2)$. The pictures on the left hand side contain the expected value $E[u_h^p]$ of the numerical solution, the expected value $E[u]$ of the exact solution, and the Monte Carlo solution u_{MC}. The right hand side shows the error dependent on the polynomial degree p of the stochastic basis functions. p denotes the highest degree of the polynomials that form the basis of $Z^{\boldsymbol{p}}$. Here, only one stochastic variable ξ is used, and hence $\dim(Z^{\boldsymbol{p}}) = 1 + p$. The expected value of the solution is given by

$$E[u] = \sin(\pi(x + 1 - \overline{V}t)) e^{-\pi^2 \sigma^2 t^2/2}. \tag{6.38}$$

Here and from now, u_{MC} refers to the Monte Carlo solution that was obtained by taking the average of the analytical solution with different values of V. The Crank-Nicolson scheme was used for solving the ODEs. The grid sizes are given by $N = 100$ spatial intervals and a time step $\Delta t = \frac{1}{50}$. The solutions are plotted for the final time $T = 1$ and different variances σ of the stochastic variable. As expected from formula (6.38), the higher σ, the stronger the damping of the deterministic solution. If σ is large, higher order polynomials have to be used in order to model the uncertainties. Furthermore, many more Monte Carlo simulations are necessary for a good approximation of the stochastic solution. On the other hand, for smaller values of σ it does not make sense to use higher order polynomials in ξ because the stochastic error is small compared to the spatial and temporal error. For all values of σ, exponential convergence in p is observed, i.e. there holds $\max_{t \in [0,T]} ||E[u - u_h^p]||_{L^2(D)} \propto e^{-Cp}$. In the numerical calculations, we used the value of p for which no further convergence could be achieved for the given temporal and spatial discretization.

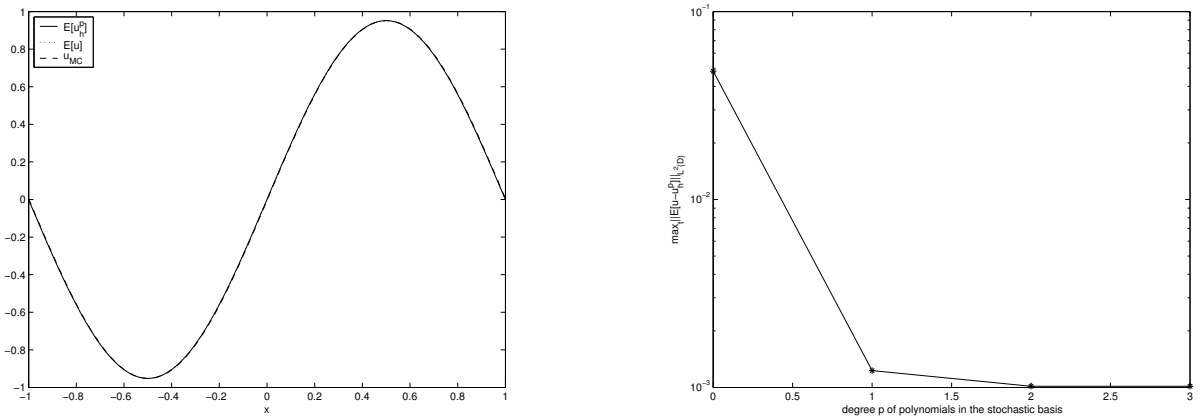


Figure 6.3: Mean solutions for $\sigma = 0.1$, $p = 3$, $N = 100$, $\Delta t = \frac{1}{50}$, $\#MC = 10000$, Crank-Nicolson time discretization

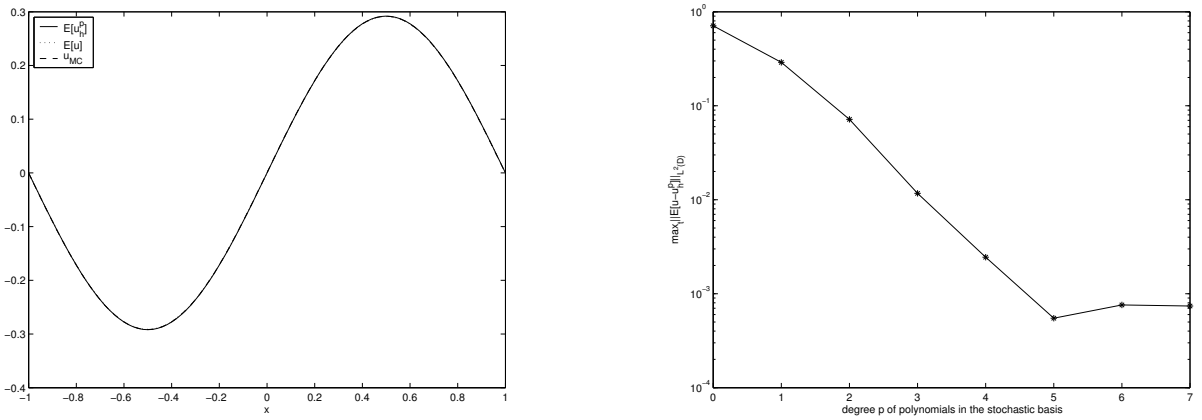


Figure 6.4: Mean solutions for $\sigma = 0.5$, $p = 7$, $N = 100$, $\Delta t = \frac{1}{50}$, $\#MC = 100000$, Crank-Nicolson time discretization

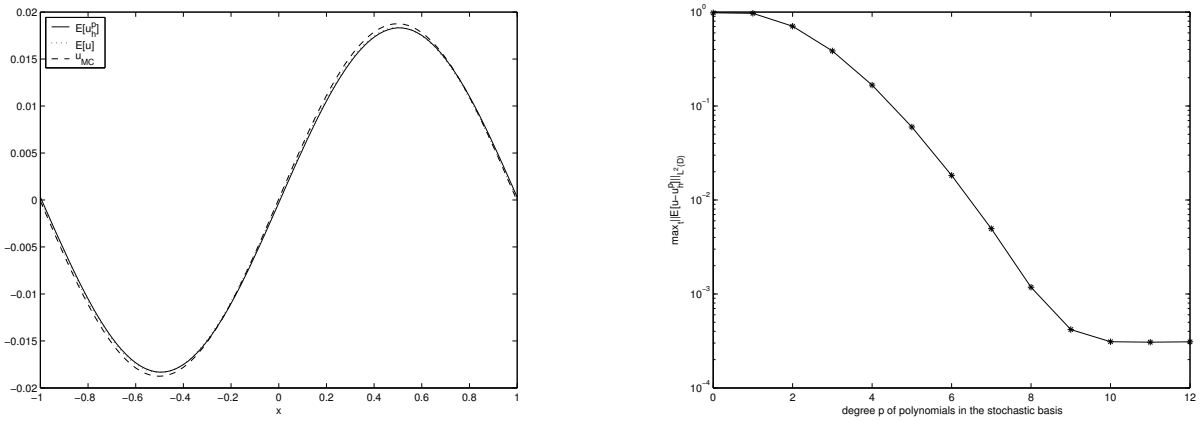


Figure 6.5: Mean solutions for $\sigma = 0.9$, $p = 12$, $N = 100$, $\Delta t = \frac{1}{50}$, $\#MC = 100000$, Crank-Nicolson time discretization

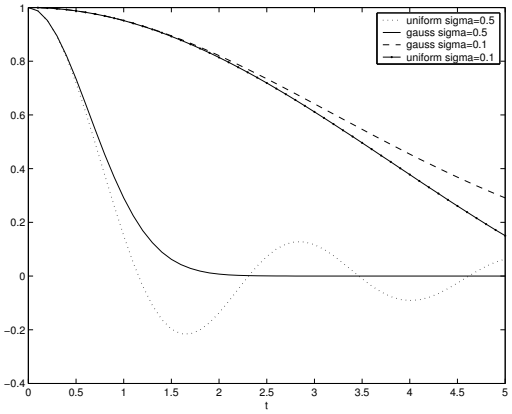


Figure 6.6: Damping factor for uniformly and normally distributed ξ for $t \in [0, 5]$

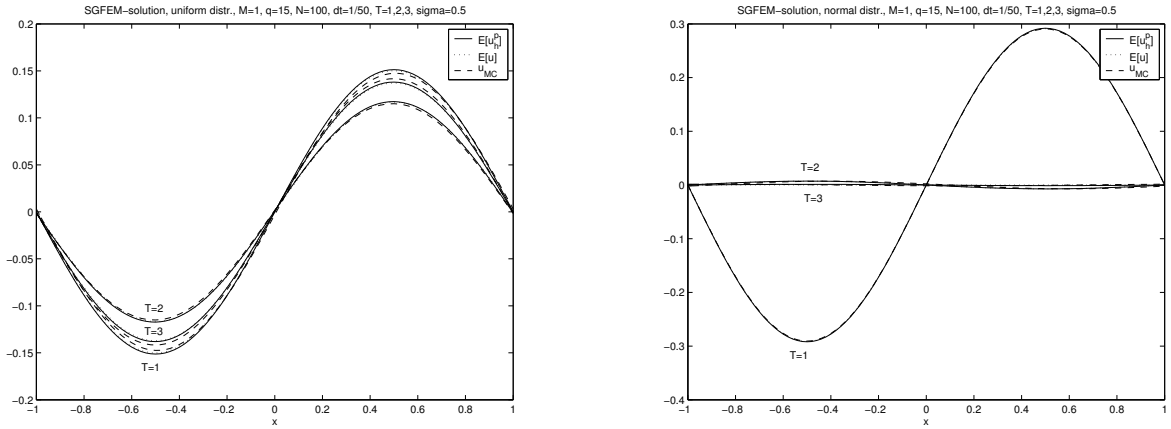


Figure 6.7: Mean solutions with $\sigma = 0.5$ for a uniformly (left) and normally (right) distributed ξ, $\#MC = 10000$

Remark: If the perturbation ξ is uniformly distributed, i.e.

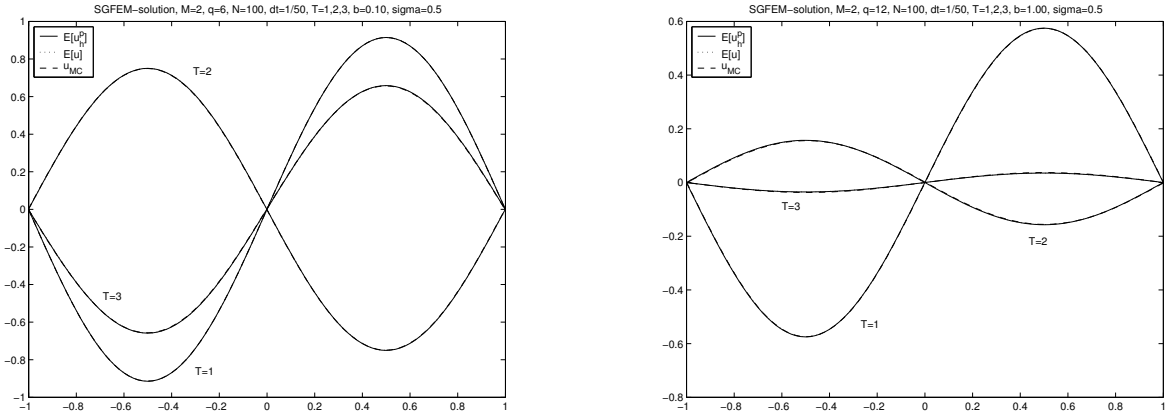
$$\tau(t) = \int_0^t V(s)ds := \overline{V}t + \sigma\xi t, \quad \xi \sim U[-\sqrt{3}, \sqrt{3}], \quad \rho(\xi) = \frac{1}{2\sqrt{3}}, \tag{6.39}$$

then the expected value of the solution becomes

$$E[u] = \sin(\pi(x + 1 - \overline{V}t))\frac{\sin(\pi\sigma t\sqrt{3})}{\pi\sigma t\sqrt{3}}.$$

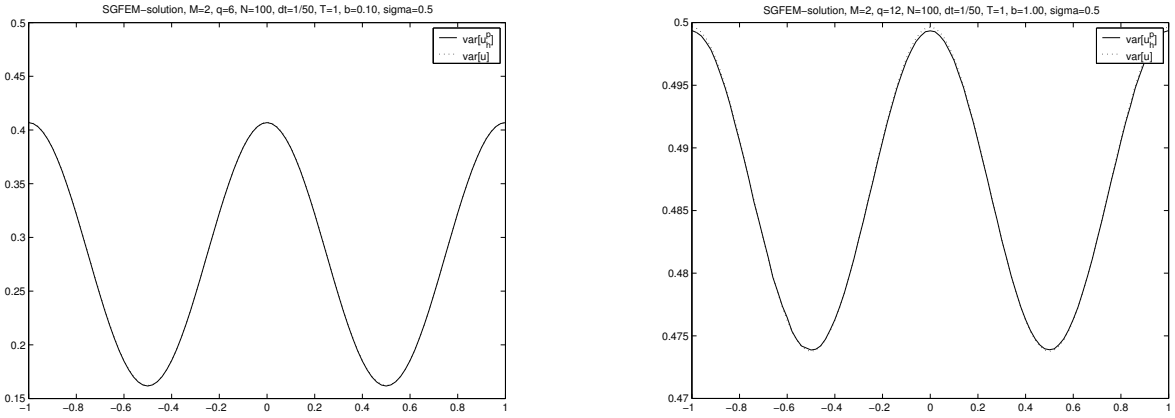
Hence, the damping factor is different from the Gaussian case. The result can be seen in Figures 6.6 and 6.7. For a normally as well as for a uniformly distributed ξ we observe that the expected value of the solution tends to zero as $t \to \infty$. The larger σ, the faster $E[u_h^p]$ reaches zero.

□



$b = 0.1$ $b = 1$

Figure 6.8: Mean solutions with $\sigma = 0.5$ for different correlation lengths, $\#MC = 100000$



$b = 0.1$ $b = 1$

Figure 6.9: Variance of the solutions with $\sigma = 0.5$ for different correlation lengths

Example 2: Consider (6.20) with $\overline{V} = 1$ where $V = V(t, \boldsymbol{\xi}) = \overline{V} + \sum_{s=1}^{M} b_s(t)\xi_s$ is approximated by a Karhunen-Loève expansion from the exponential covariance kernel

$$\mathcal{C}(s,t) = \exp\{-\frac{|t-s|}{b}\} \tag{6.40}$$

by

$$\int_0^T \sigma^2 \mathcal{C}(s,t)b(s)ds = \lambda b(t).$$

The solution is presented in Figures 6.8 to 6.9. The correlation length b influences the solution. A small correlation length results in less damping and a smaller variance of the solution.

□

Now, we turn to the advection-diffusion equation

$$c_t + Vc_x - Dc_{xx} = 0, \qquad x \in (a,b), \quad t \in [0,T] = [0,0.8] \tag{6.41}$$

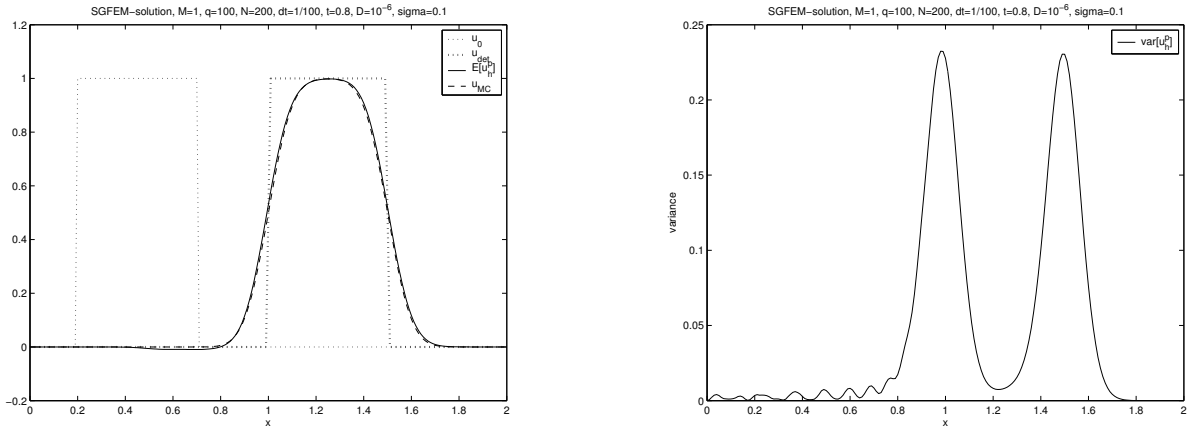


Figure 6.10: Expected value and variance of the solution with $D = 10^{-6}$

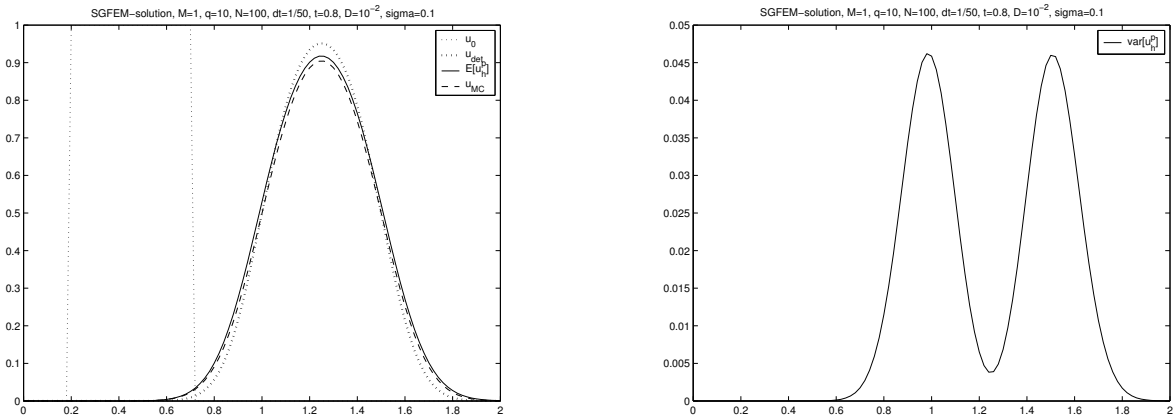


Figure 6.11: Expected value and variance of the solution with $D = 10^{-2}$

with the initial condition

$$c_0(x) = \begin{cases} 1, & \text{if } x \in [x_l, x_r] \subset (a, b) \\ 0, & \text{otherwise} \end{cases}, \qquad x_l = 0.2,\ x_r = 0.7. \tag{6.42}$$

The final time was set $T = 0.8$ to exclude boundary effects. Homogeneous Dirichlet boundary conditions are specified at $x = a$ and $x = b$. Furthermore, assume $D =$ constant and $E[V] = \overline{V}$. Then the analytic deterministic solution can be given in closed form as long as the square wave does not intersect the outflow boundary during the time interval $[0, T]$:

$$c(x,t) = \frac{1}{2}\left[\operatorname{erf}\left(\frac{x - \overline{V}t - x_l}{\sqrt{4Dt}}\right) - \operatorname{erf}\left(\frac{x - \overline{V}t - x_r}{\sqrt{4Dt}}\right)\right].$$

From the previous calculations, one can conclude that a stochastic velocity field acts like diffusion in that the solution is damped. If diffusion dominates, the influence of uncertainties in the velocity is reduced, i.e. the variance of the solution is much smaller than in the advection dominated case.

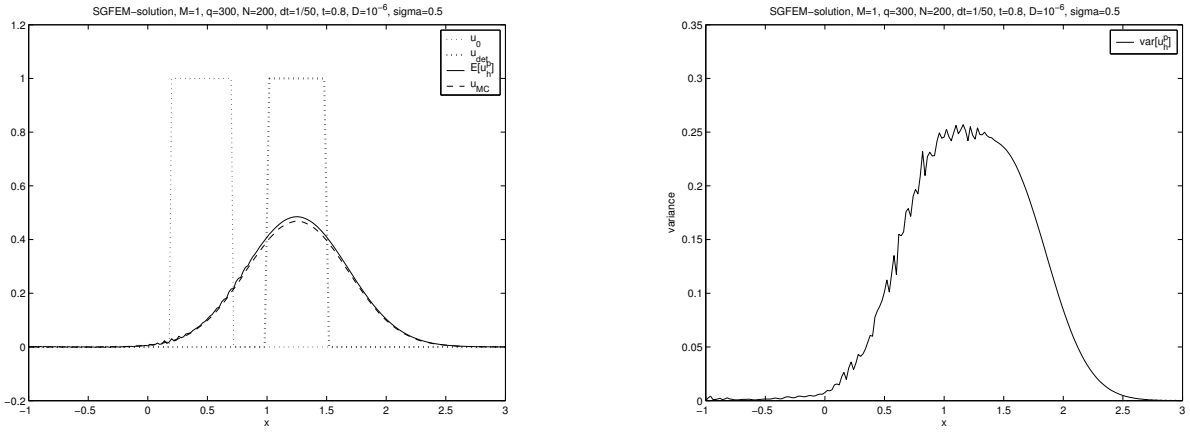


Figure 6.12: Expected value and variance of the solution with $\sigma = 0.5$, $\#MC = 100000$

Example 3: Figures 6.10 and 6.11 show the results for a given velocity field $V = 1+\xi,\ \xi \sim N(0, \sigma^2),\ \sigma = 0.1$. u_{det} denotes the deterministic mean solution, i.e. the solution which is obtained from the stochastic equation in which the coefficients are replaced by their mean values. 100000 Monte Carlo simulations were performed to compare the results. With a smaller diffusion, the effect of uncertainties is more visible. This is due to the fact that a stochastic velocity adds extra diffusion to the system which could be neglected if the original diffusion is big. We observe that the variance attains the highest values in the regions of the steep front. Furthermore, we point out that for $D = 10^{-6}$ a smaller value of p would cause more wiggles in the variance of the solution. That is why we used the large value $p = 100$ in the calculations. This becomes more severe for higher values of σ.

□

Example 4: Consider the previous example with $D = 10^{-6}$ and $\sigma = 0.5$ on a domain $[-1, 3]$. Even with a high polynomial degree $p = 300$ the numerical solution exhibits some oscillations. Figure 6.12 shows the results.

□

In the following, we want to examine how the numerical solution can be improved by using different temporal and spatial discretizations.

6.2.3.2 Discontinuous Galerkin Method

The discrete solution u is written as a tensor product of basis functions

$$u(x,t,\omega) = \sum_{k=1}^{N}\sum_{i=1}^{P}\sum_{j=1}^{T} U_{ijk}^{t_n}\varphi_k(x)\alpha_j(t)\psi_i(\boldsymbol{\xi}).$$

With the KL-expansion $V(t,\omega) = \overline{V} + \sum_{s=1}^{M}\sqrt{\lambda_s}b_s(t)\xi_s(\omega)$ and test functions $w(x,t,\omega) = \varphi_n(x)\alpha_m(t)\psi_l(\boldsymbol{\xi}),\ (n = 1,\dots,N,\ m = 1,2,\ l = 1,\dots,P)$, the numerical scheme for the

homogeneous advection equation can be written as

$$\begin{aligned}
&\sum_{i,j,k} U_{ijk}^{t_n}[(\varphi_k\varphi_n)(\dot{\alpha}_j\alpha_m)\langle\psi_i\psi_l\rangle + \overline{V}(\varphi_k'\varphi_n)(\alpha_j\alpha_m)\langle\psi_i\psi_l\rangle \\
&+\sum_{s=1}^{M}(\varphi_k'\varphi_n)(\sqrt{\lambda_s}b_s(t)\alpha_j\alpha_m)\langle\xi_s\psi_i\psi_l\rangle + \sum_{i,k} U_{i1k}^{t_n}(\varphi_k\varphi_n)\delta_{m1}\langle\psi_i\psi_l\rangle \\
&= \sum_{i,k} U_{i2k}^{t_{n-1}}(\varphi_k\varphi_n)\delta_{m1}\langle\psi_i\psi_l\rangle
\end{aligned} \tag{6.43}$$

with

$$(\varphi_k\varphi_n) = \int_D \varphi_k(x)\varphi_n(x)dx,$$

$$(\alpha_j\alpha_m) = \int_0^T \alpha_j(t)\alpha_m(t)dt,$$

and the meaning of $\langle\cdot,\cdot\rangle$ as in (6.35). φ' denotes the derivative of φ with respect to x, and $\dot{\alpha}$ denotes the derivative with respect to t.

Using the notations

$$K_{knjm}^0 = (\varphi_k\varphi_n)(\dot{\alpha}_j\alpha_m) + \overline{V}(\varphi_k'\varphi_n)(\alpha_j\alpha_m),$$

$$K_{knjm}^s = (\varphi_k'\varphi_n)(\sqrt{\lambda_s}b_s(t)\alpha_j\alpha_m)$$

and the orthogonality of $\{\psi_i\}$, (6.43) simplifies to

$$\sum_{j,k} U_{ijk}^{t_n}[K_{knjm}^0 + \sum_{s=1}^{M} c_{i_s}^2 K_{knjm}^s] + \sum_k U_{i1k}^{t_n}(\varphi_k\varphi_n)\delta_{m1} = \sum_k U_{i2k}^{t_{n-1}}(\varphi_k\varphi_n)\delta_{m1}.$$

Hence, one has to solve P linear systems of equations with $K_{knjm} \in \mathbb{R}^{2(N+1)\times 2(N+1)}$, and $U_{ijk} \in \mathbb{R}^{2(N+1)}$.

Example 1: Consider again (6.41) with $D = 10^{-6}$, $V = 1+\xi$, $\xi \sim N(0,\sigma^2)$, $\sigma = 0.1$. 100000 Monte Carlo simulations were performed. Figure 6.13 shows the numerical solution obtained with a standard semi discrete Galerkin approach, and Figure 6.14 contains the solution from the discontinuous Galerkin method. While the GAL solution exhibits oscillations for the polynomial degree $p = 4$, the DG solution leads to better numerical results. The numerical solution can also be improved by using the semi discrete approach and a higher polynomial degree $p = 20$ in the approximation of the stochastic space, see Figure 6.15. However, this can be very expensive if several random variables are involved.

□

For a larger variance of the input random variable, stabilization techniques such as the streamline diffusion method can be used to obtain good numerical results.

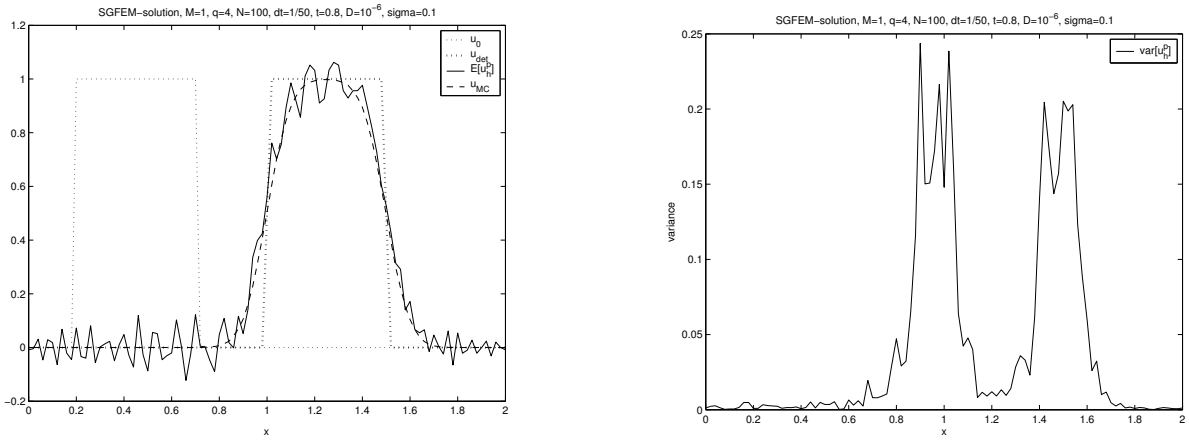


Figure 6.13: Expected value and variance of the solution for $p = 4$

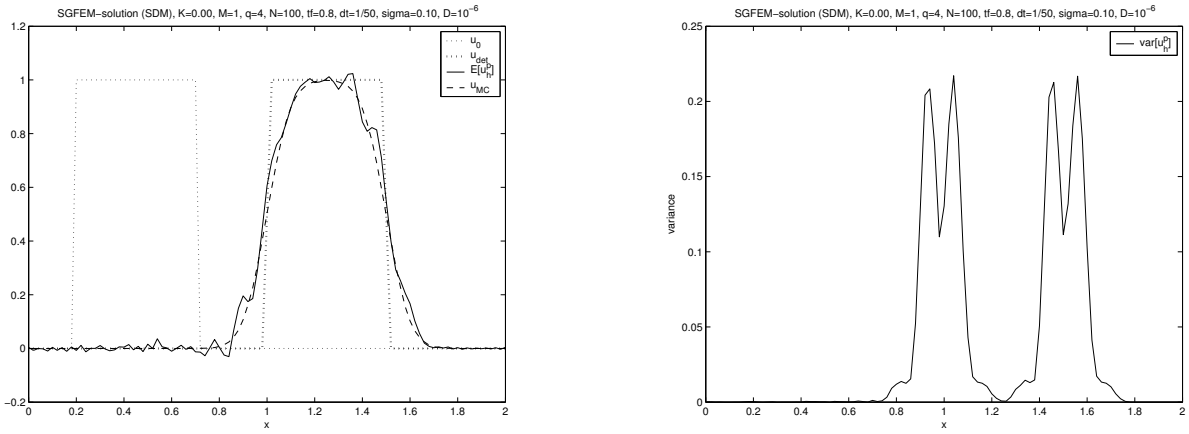


Figure 6.14: Expected value and variance of the solution for $p = 4$, DG approach

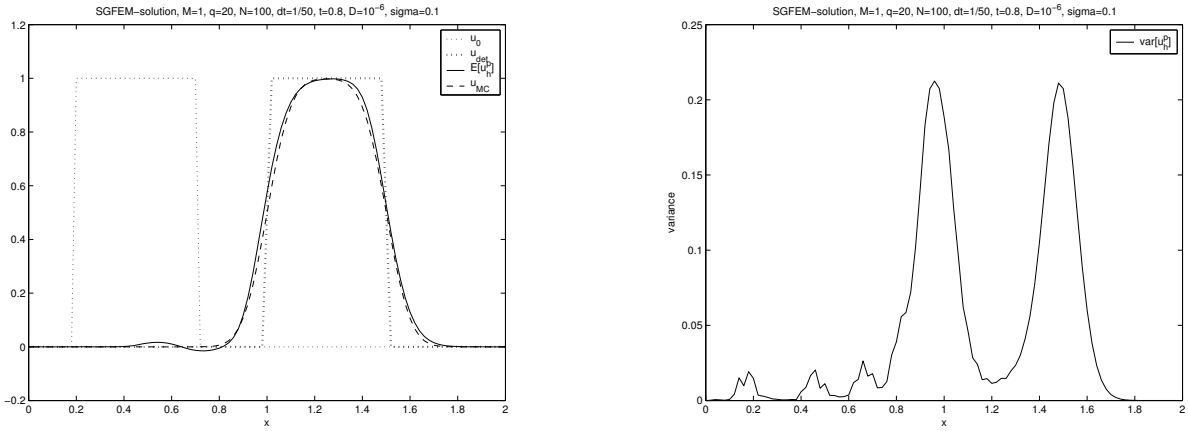


Figure 6.15: Expected value and variance of the solution for $p = 20$

Streamline Diffusion Method With a modified test function $w + \delta(w_t + Vw_x)$ the numerical scheme becomes more complicated:

$$\begin{aligned}
&\sum_{i,j,k} U_{ijk}^{t_n}[(\varphi_k\varphi_n)(\dot\alpha_j\alpha_m)\langle\psi_i\psi_l\rangle + \delta(\varphi_k\varphi_n)(\dot\alpha_j\dot\alpha_m)\langle\psi_i\psi_l\rangle \\
&+\delta\overline{V}(\varphi_k\varphi_n')(\dot\alpha_j\alpha_m)\langle\psi_i\psi_l\rangle + \delta\sum_{s=1}^{M}(\varphi_k\varphi_n')(\sqrt{\lambda_s}b_s(t)\dot\alpha_j\alpha_m)\langle\xi_s\psi_i\psi_l\rangle \\
&+\overline{V}(\varphi_k'\varphi_n)(\alpha_j\alpha_m)\langle\psi_i\psi_l\rangle + \sum_{s=1}^{M}(\varphi_k'\varphi_n)(\sqrt{\lambda_s}b_s(t)\alpha_j\alpha_m)\langle\xi_s\psi_i\psi_l\rangle \\
&+\delta\overline{V}(\varphi_k'\varphi_n)(\alpha_j\dot\alpha_m)\langle\psi_i\psi_l\rangle + \delta\sum_{s=1}^{M}(\varphi_k'\varphi_n)(\sqrt{\lambda_s}b_s(t)\alpha_j\dot\alpha_m)\langle\xi_s\psi_i\psi_l\rangle \\
&+\delta\overline{V}^2(\varphi_k'\varphi_n')(\alpha_j\alpha_m)\langle\psi_i\psi_l\rangle + 2\delta\overline{V}\sum_{s=1}^{M}(\varphi_k'\varphi_n')(\sqrt{\lambda_s}b_s(t)\alpha_j\alpha_m)\langle\xi_s\psi_i\psi_l\rangle \\
&+\delta\sum_{s=1}^{M}\sum_{r=1}^{M}(\varphi_k'\varphi_n')(\sqrt{\lambda_s}b_s(t)\sqrt{\lambda_r}b_r(t)\alpha_j\alpha_m)\langle\xi_s\xi_r\psi_i\psi_l\rangle] \\
&+\sum_{i,k} U_{i1k}^{t_n}(\varphi_k\varphi_n)\delta_{m1}\langle\psi_i\psi_l\rangle = \sum_{i,k} U_{i2k}^{t_{n-1}}(\varphi_k\varphi_n)\delta_{m1}\langle\psi_i\psi_l\rangle.
\end{aligned}$$

The following notations are introduced:

$$\begin{aligned}
K_{knjm}^0 &= (\varphi_k\varphi_n)(\dot\alpha_j\alpha_m) + \delta(\varphi_k\varphi_n)(\dot\alpha_j\dot\alpha_m) \\
&\quad +\delta\overline{V}(\varphi_k\varphi_n')(\dot\alpha_j\alpha_m) + \overline{V}(\varphi_k'\varphi_n)(\alpha_j\alpha_m) \\
&\quad +\delta\overline{V}(\varphi_k'\varphi_n)(\alpha_j\dot\alpha_m) + \delta\overline{V}^2(\varphi_k'\varphi_n')(\alpha_j\alpha_m),
\end{aligned}$$

$$\begin{aligned}
K_{knjm}^s &= \delta(\varphi_k\varphi_n')(\sqrt{\lambda_s}b_s(t)\dot\alpha_j\alpha_m) + (\varphi_k'\varphi_n)(\sqrt{\lambda_s}b_s(t)\alpha_j\alpha_m) \\
&\quad +\delta(\varphi_k'\varphi_n)(\sqrt{\lambda_s}b_s(t)\alpha_j\dot\alpha_m) + 2\delta\overline{V}\varphi_k'\varphi_n')(\sqrt{\lambda_s}b_s(t)\alpha_j\alpha_m),
\end{aligned}$$

$$K_{knjm}^{sr} = \delta\sum_{s=1}^{M}\sum_{r=1}^{M}(\varphi_k'\varphi_n')(\sqrt{\lambda_s}b_s(t)\sqrt{\lambda_r}b_r(t)\alpha_j\alpha_m).$$

Now, the system of equations can be rewritten as

$$\begin{aligned}
\sum_{j,k} U_{ijk}[K_{knjm}^0 + \sum_{s=1}^{M} c_{i_s}^s K_{knjm}^s + \sum_{s=1}^{M}\sum_{r=1}^{M} c_{i_s}^s c_{i_r}^r K_{knjm}^{sr}] + \sum_k U_{i1k}^{t_n}(\varphi_k\varphi_n)\delta_{m1} \\
= \sum_k U_{i2k}^{t_{n-1}}(\varphi_k\varphi_n)\delta_{m1}.
\end{aligned}$$

The case of the standard discontinuous Galerkin method is included by setting $\delta = 0$. Note that the scheme would be more complicated with non-orthogonal basis functions or

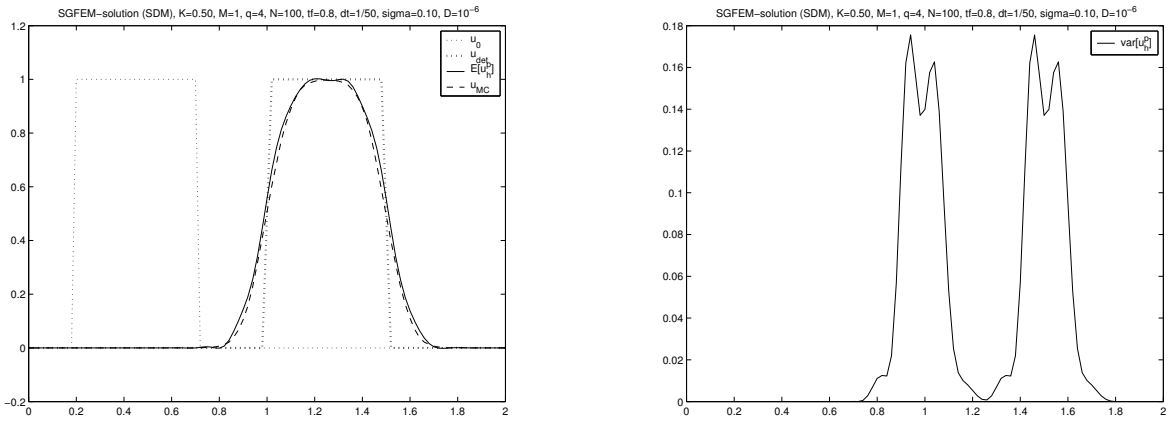


Figure 6.16: Expected value and variance of the solution: SDM ($K = 0.5$), $p = 4$, $\sigma = 0.1$

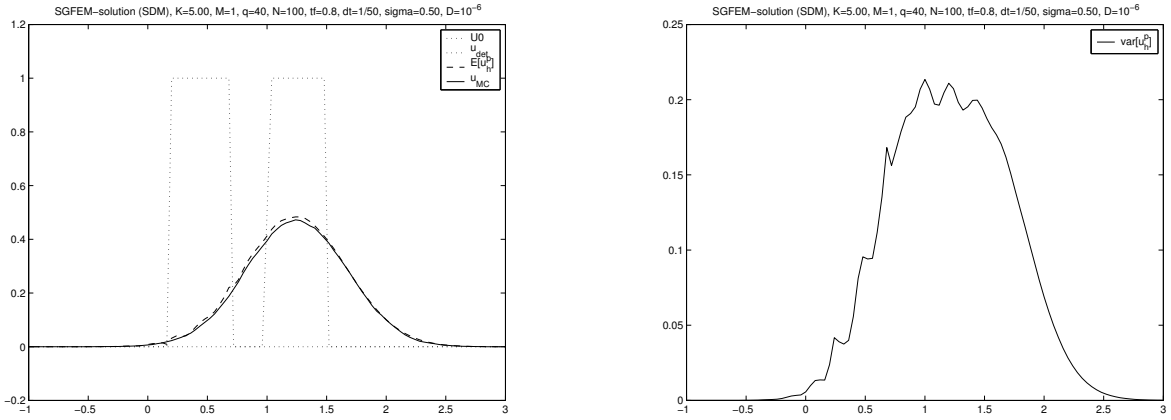


Figure 6.17: Expected value and variance of the solution: SDM ($K = 5$), $p = 40$, $\sigma = 0.5$

a generalized polynomial chaos expansion of V due to the presence of the term $\langle \xi_s \xi_r \psi_i \psi_l \rangle$.

Example 1: Consider again the previous example, i.e. the 1D advection diffusion equation with $D = 10^{-6}$ and $V = 1 + \xi$, $\xi \sim N[0, \sigma^2]$. 100000 Monte Carlo simulations were performed. The numerical solution with $p = 4$ and $\sigma = 0.1$ can be improved further by using the SDM, see Figure 6.16. For $\sigma = 0.5$, a much higher polynomial degree $p = 40$ and a larger parameter $K = 5$ which corresponds to a higher numerical diffusion are necessary before convergence to the Monte-Carlo solution is achieved. Figure 6.17 shows the results.

□

So far, V was chosen to depend on a random variable ξ but being independent of x and t. Now, as for the sine-wave it is examined how the solution behaves if V depends on the location or on time.

Example 2: This time, V is represented by its Karhunen-Loève expansion derived

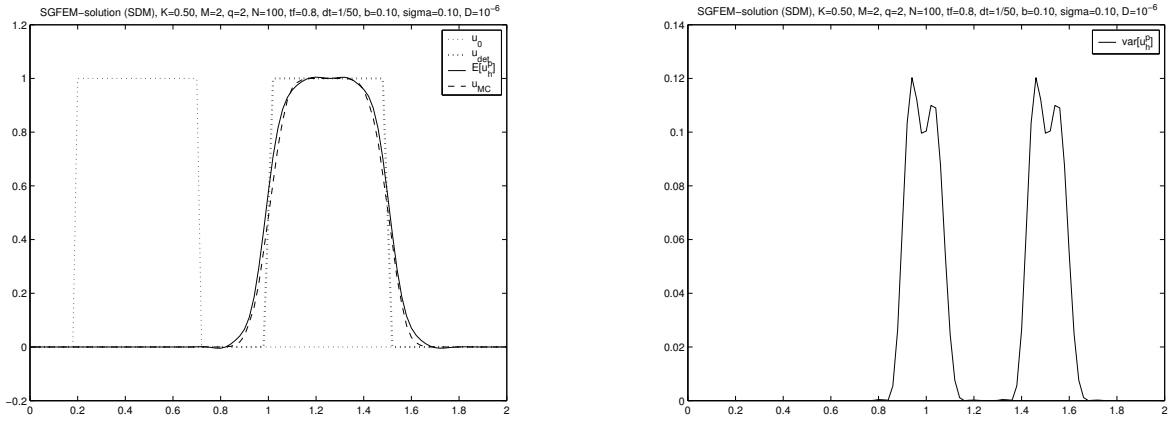


Figure 6.18: Expected value and variance of the solution: KL-expansion in time

from the exponential covariance kernel with respect to the time variable

$$\mathcal{C}(s,t) = \sigma^2 \exp\{-\frac{|t-s|}{b}\},$$

where $b = 0.1$ and $\sigma = 0.1$. 100000 Monte Carlo simulations were performed. The SDM was used for the numerical solution, see Figure 6.18. In contrast to the previous examples, the coefficients in front of the random variables ξ_s depend on t now. Consequently, the stochastic solution is similar to the deterministic one but has an additional time-dependent damping factor.

□

Example 3: Now, V is represented by its Karhunen-Loève expansion derived from the exponential covariance kernel with respect to the space variable $x \in [0, 3]$. As before, $b = 0.1$ and $\sigma = 0.1$. A high polynomial degree is necessary until convergence is achieved, see Figure 6.19. Unfortunately, no analytical solution is available, but it can be observed that the solution is damped more than in the time-dependent case.

□

Example 4: At last, the results for two different fields $V = 1 + x\xi$ and $V = 1 + t\xi$ are compared. For the time dependent V, the initial condition is damped symmetrically while in the other case the initial concentration is spread more into the flow direction, see Figures 6.20 and 6.21.

□

The last two examples illustrate that it is more difficult to model space dependent stochastic velocity fields.

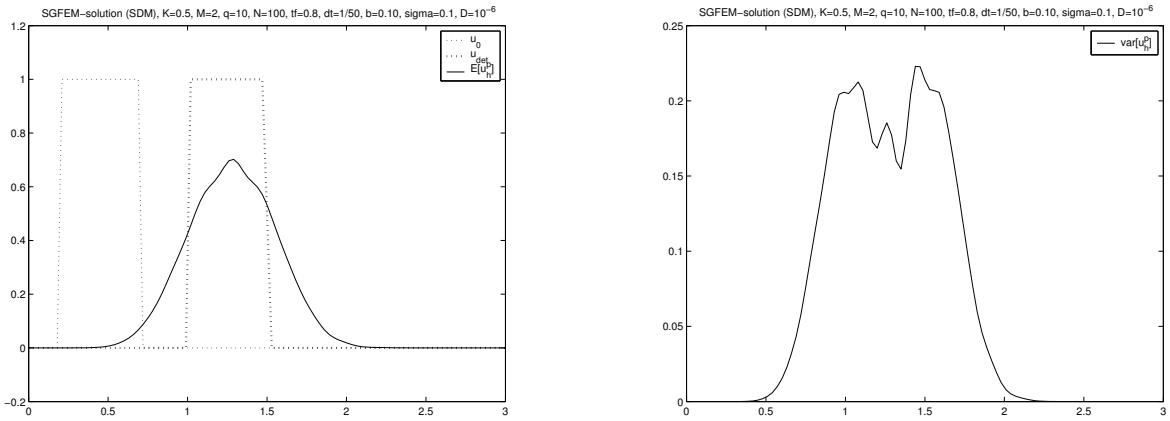


Figure 6.19: Expected value and variance of the solution: KL-expansion in space

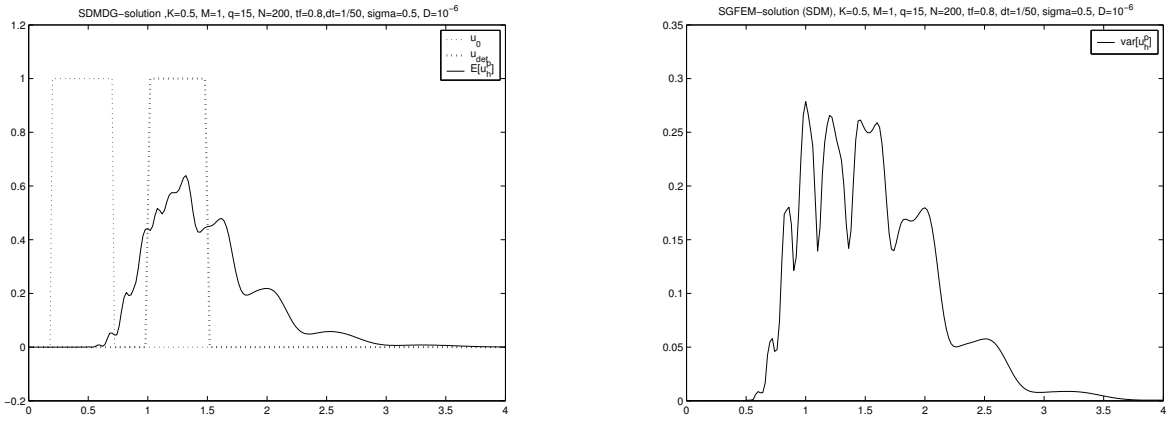


Figure 6.20: Expected value and variance of the solution: SDM, $V = \overline{V} + x\xi$

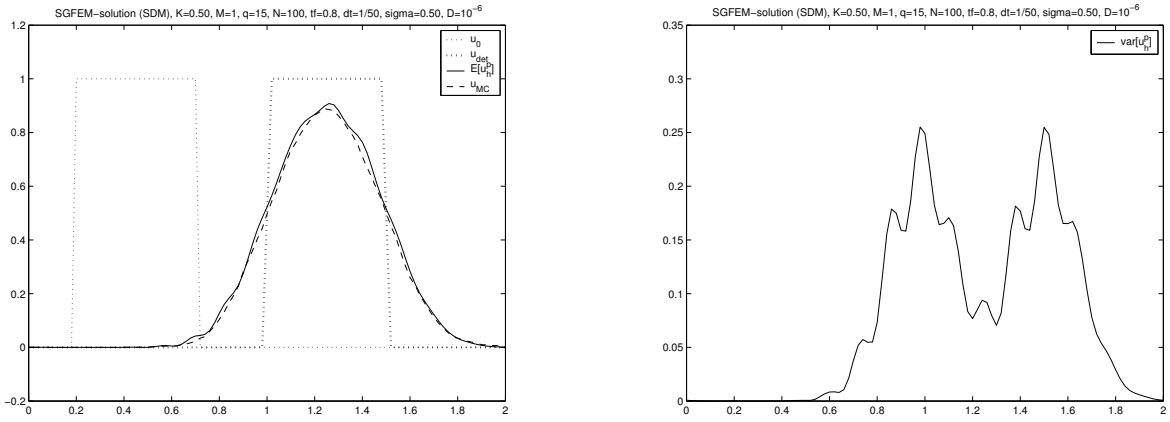


Figure 6.21: Expected value and variance of the solution: SDM, $V = \overline{V} + t\xi$

6.3 The Advection-Diffusion Equation with Stochastic Diffusion

6.3.1 Stochastic Molecular Diffusion

As in the previous examples, only molecular diffusion is considered in this subsection, i.e. the diffusion-dispersion tensor $\boldsymbol{D}$ simplifies to a diffusion coefficient D which is independent of the velocity. Although the main topic of this work are advection-dominated equations, the diffusion coefficient D must be allowed to take larger values now, in order to illustrate the influence of stochastic data. Furthermore, one has to assure that D takes only positive values. This must be taken into account when a distribution function is chosen. The Gamma distribution $\Gamma(p, \lambda)$ which is given by the density function

$$f(\xi) = \frac{\lambda^p}{\Gamma(p)} \xi^{p-1} e^{-\lambda \xi}, \ \xi > 0, \qquad E[\xi] = \frac{p}{\lambda}, \qquad \text{var}[\xi] = \frac{p}{\lambda^2}$$

seems to be an appropriate choice.

The matrix M_1 which is needed to compute the basis coefficients, is

$$M_1(i,l) = \frac{1}{\lambda}[-\sqrt{i(i+p-1)}\delta_{i,l+1} + (2i+p)\delta_{i,l} - \sqrt{(i+1)(i+p)}\delta_{i,l-1}], \ i,l = 0,\ldots,p.$$

In the following, consider the one-dimensional advection-diffusion equation from Chapter 5

$$c_t + Vc_x - Dc_{xx} = f, \qquad x \in (a,b), \quad t \in [0,T] = [0, 0.8]$$

with the initial condition

$$c_0(x) = \begin{cases} 1, & \text{if } x \in [x_l, x_r] \subset (a,b) \\ 0, & \text{otherwise} \end{cases}, \qquad x_l = 0.2, \ x_r = 0.7 \tag{6.44}$$

Homogeneous flux boundary conditions are specified at $x = a$ and $x = b$. Furthermore, set $V = 1$, $f = 0$, $a = -1$, $b = 3.5$, and assume $D = \overline{D} + \xi$ with a certain random variable ξ. Then the analytic deterministic solution can be given in closed form as long as the square wave does not intersect the outflow boundary during the time interval $[0, T]$:

$$c(x,t) = \frac{1}{2}\left[\text{erf}\left(\frac{x - Vt - x_l}{\sqrt{4E[D]t}}\right) - \text{erf}\left(\frac{x - Vt - x_r}{\sqrt{4E[D]t}}\right)\right].$$

Example 1: Figures 6.22 and 6.23 present the results for $\overline{D} = 0.01$ and $E[\xi] = 0.1$. In the first figure, $\text{var}[\xi] = 0.1$, and in the second figure, $\text{var}[\xi] = 0.01$. The pictures on the left show the expected value of the solutions, and the pictures on the right hand side illustrate the variances. A small variance means that D takes more values close to $\overline{D}$ while also larger values for D are possible if the variance is bigger. The smaller the variance, the closer the stochastic solution is to the deterministic solution with $D = \overline{D} + E[\xi]$. However, the stochastic solution is damped less than the deterministic one.

□

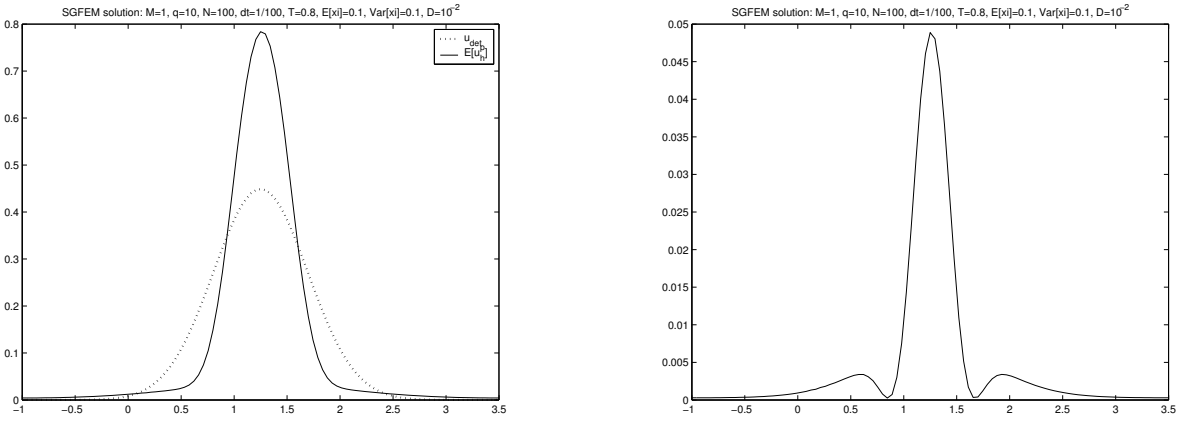


Figure 6.22: Expected value and variance of the solution: $\xi \sim \Gamma(p, \lambda),\ p = 10\overline{D},\ \lambda = 1$

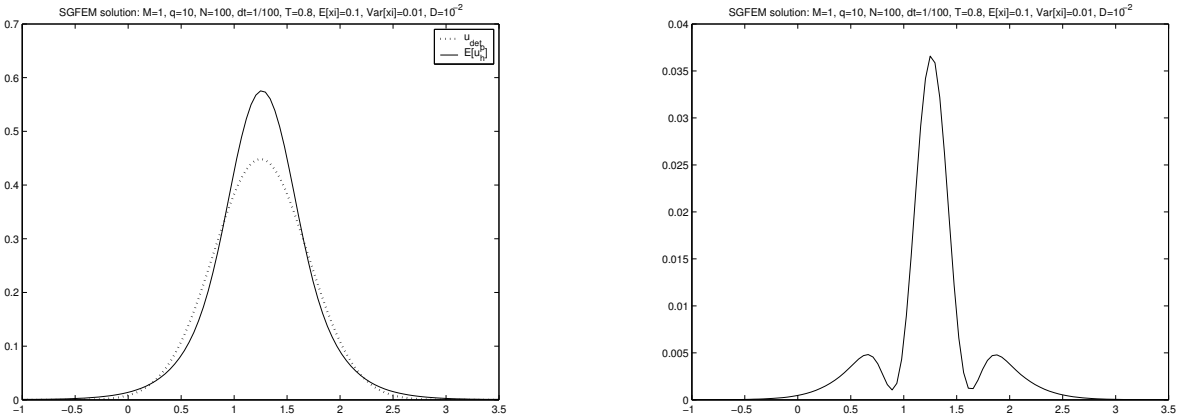


Figure 6.23: Expected value and variance of the solution: $\xi \sim \Gamma(p, \lambda),\ p = 100\overline{D},\ \lambda = 10$

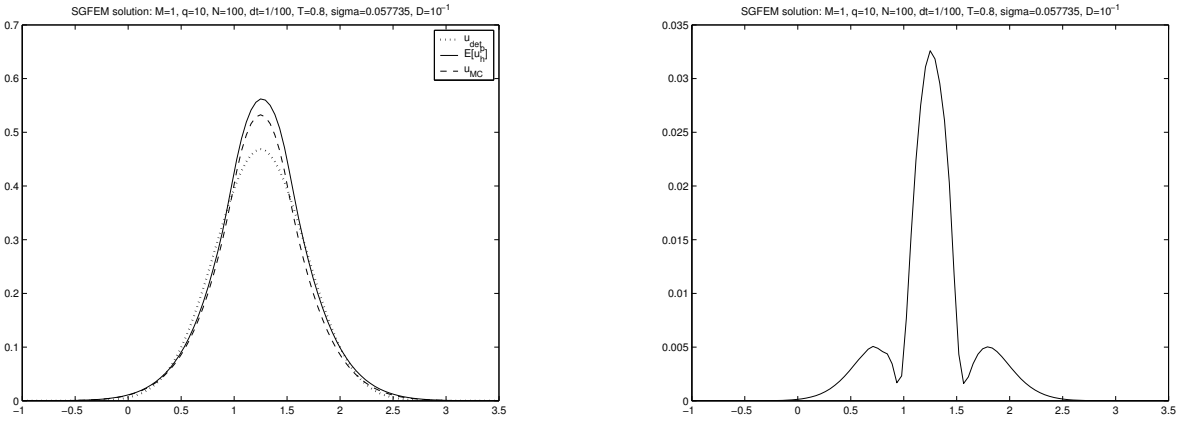


Figure 6.24: Expected value and variance of the solution: $\xi \sim U[-\sqrt{3}\sigma, \sqrt{3}\sigma],\ \sigma = 0.1/\sqrt{3}$

It is also possible to choose $\xi \sim U[-\sqrt{3}\sigma, \sqrt{3}\sigma]$ so that $E[\xi] = 0$ and $\text{var}[\xi] = \sigma^2$. Note that σ has to be chosen such that $\overline{D} - \sqrt{3}\sigma \geq 0$. This is fulfilled for $\sigma = s\overline{D}/\sqrt{3}$ with $0 \leq s \leq 1$.

Example 2: Results for $\overline{D} = 0.1$ can be seen in Figure 6.24. It can be observed that for the large variance $\sigma = \overline{D}/\sqrt{3}$ ($s = 1$) the stochastic solution is damped less than in the deterministic case even though $E[D] = \overline{D}$. Such results were already obtained for the stationary elliptic diffusion equation in [29]. With a smaller amount of diffusion or a smaller variance, the random effect is nearly invisible. The SGFEM solution will be very close to the deterministic one.

□

6.3.2 Stochastic Dispersion

In practice, dispersion plays a more important role than molecular diffusion. Assume that the velocity was determined from Darcy's law with a stochastic permeability coefficient. This stochastic velocity field is used as input for the transport equation. Since dispersion depends on the velocity, it will also be modelled as a random process. The equation treated in this subsection is given by

$$c_t + \nabla \cdot (\boldsymbol{V} c - \boldsymbol{D} \nabla c) = 0,$$

with $\boldsymbol{V} = (u, v)^\top$ and

$$\boldsymbol{D} = D_m I + \frac{d_L}{|\boldsymbol{V}|} \begin{pmatrix} u^2 & uv \\ uv & v^2 \end{pmatrix} + \frac{d_T}{|\boldsymbol{V}|} \begin{pmatrix} v^2 & -uv \\ -uv & u^2 \end{pmatrix}.$$

Transversal dispersion d_T and molecular diffusion D_m will not be taken into account in this first approach in order to keep the equations as short as possible. Besides, they are much smaller than the longitudinal dispersion d_L. Furthermore, we assume no-flow boundary conditions throughout this subsection.

Assume $\boldsymbol{V}$ is given in the form

$$\boldsymbol{V} = \overline{\boldsymbol{V}} + \sum_{s=1}^{M} \boldsymbol{b}_s(\boldsymbol{x}) \xi_s$$

with $\boldsymbol{b}_s = (b_s^1, b_s^2)^\top$, $\overline{\boldsymbol{V}} = (u, v)^\top$, $\nabla \cdot \boldsymbol{V} = 0$ and certain random variables ξ_s. In the following, the term $d_L/|\boldsymbol{V}|$ is approximated by $D_L = d_L/|\overline{\boldsymbol{V}}|$. Let $\{\varphi_k(\boldsymbol{x})\}_{k=1}^N$ and $\{\psi_i(\boldsymbol{\xi})\}_{i=1}^P$ be the basis functions for the finite dimensional deterministic and stochastic spaces, respectively. The solution c is represented as

$$c(\boldsymbol{x}, t, \boldsymbol{\xi}) = \sum_k \sum_i \alpha_{ik}(t) \varphi_k(\boldsymbol{x}) \psi_i(\boldsymbol{\xi}).$$

With test functions $w_{lm} = \varphi_l(\boldsymbol{x})\psi_m(\boldsymbol{\xi})$, the weak formulation

$$(c_t, w) + (\boldsymbol{V} \cdot \nabla c, w) + (\boldsymbol{D} \nabla c, \nabla w) = 0$$

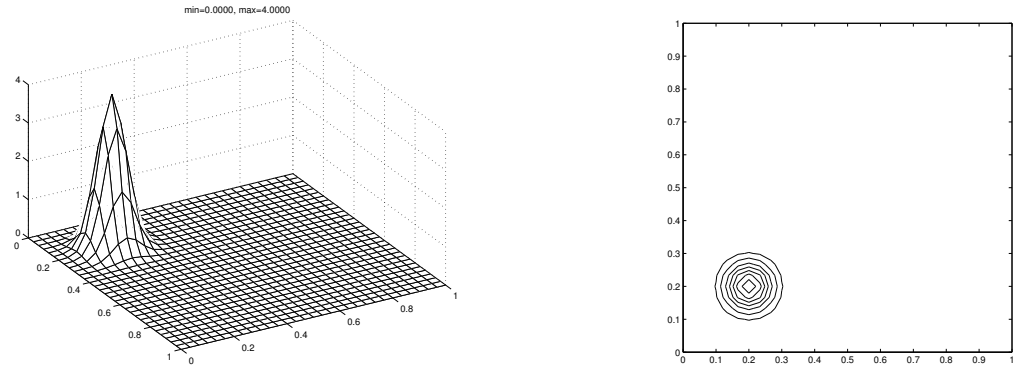


Figure 6.25: Initial value for the advection-diffusion-dispersion problem

can be rewritten as

$$
\begin{aligned}
0 = \;& \sum_k \sum_i \dot{\alpha}_{ik}(t)(\varphi_k, \varphi_l)\langle \psi_i, \psi_m \rangle + \sum_k \sum_i \alpha_{ik}(t)(\overline{\boldsymbol{V}} \cdot \nabla \varphi_k, \varphi_l)\langle \psi_i, \psi_m \rangle \\
& + \sum_k \sum_i \alpha_{ik}(t) D_L \left[(u^2 \varphi_{kx}, \varphi_{lx}) + (uv\varphi_{ky}, \varphi_{lx}) + (uv\varphi_{kx}, \varphi_{ly}) + (v^2 \varphi_{ky}, \varphi_{ly}) \right] \langle \psi_i, \psi_m \rangle \\
& + \sum_s \sum_k \sum_i \alpha_{ik}(t) D_L \langle \xi_s \psi_i, \psi_m \rangle [(\boldsymbol{b} \cdot \nabla \varphi_k, \varphi_l) + 2(u b_s^1(\boldsymbol{x})\varphi_{kx}, \varphi_{lx}) + (u b_s^2(\boldsymbol{x})\varphi_{ky}, \varphi_{lx}) \\
& \quad + (v b_s^1(\boldsymbol{x})\varphi_{ky}, \varphi_{lx}) + 2(v b_s^2(\boldsymbol{x})\varphi_{ky}, \varphi_{ly}) + (u b_s^2(\boldsymbol{x})\varphi_{kx}, \varphi_{ly}) + (v b_s^1(\boldsymbol{x})\varphi_{kx}, \varphi_{ly})] \\
& + \sum_t \sum_s \sum_k \sum_i \alpha_{ik}(t) D_L \langle \xi_t \xi_s \psi_i, \psi_m \rangle [(b_t^1(\boldsymbol{x}) b_s^1(\boldsymbol{x})\varphi_{kx}, \varphi_{lx}) \\
& \quad + (b_s^1(\boldsymbol{x}) b_t^2(\boldsymbol{x})\varphi_{ky}, \varphi_{lx}) + (b_t^2(\boldsymbol{x}) b_s^2(\boldsymbol{x})\varphi_{ky}, \varphi_{ly}) + (b_s^1(\boldsymbol{x}) b_t^2(\boldsymbol{x})\varphi_{kx}, \varphi_{ly})].
\end{aligned}
$$

As in the streamline diffusion method, the equation is nonlinear in $\boldsymbol{V}$ and the term $\langle \xi_t \xi_s \psi_i, \psi_m \rangle$ appears. Thanks to the orthogonality requirement and the assumption of independent random variables, this term does not destroy the block-diagonal structure.

Example 1: Consider the problem given above on a rectangular domain $\Omega = [0,1] \times [0,1]$ in the time interval $T = [0, 0.4]$ with $D_m = 10^{-6}$, $d_L = 0.005$, $d_T = 0.0005$, $\boldsymbol{V} = (1,1)' + (\xi, \xi)$, $\xi \sim N(0, \sigma^2)$, $\sigma = 0.2$ and initial condition

$$
u_0(x) = 2 \exp\left(-\frac{(x - 0.2)^2 + (y - 0.2)^2}{2(0.05)^2} \right).
$$

For the numerical computations it was set $N = 30$ and $\Delta t = \Delta x$. The results can be seen in Figures 6.25 to 6.28. Compared to the deterministic solution, it can be observed that the contour lines in the stochastic solution are denser in upstream direction and that the maximum value is slightly larger (2.1437 compared to 1.9272). This means the invading fluid is swept out faster than in the deterministic case.

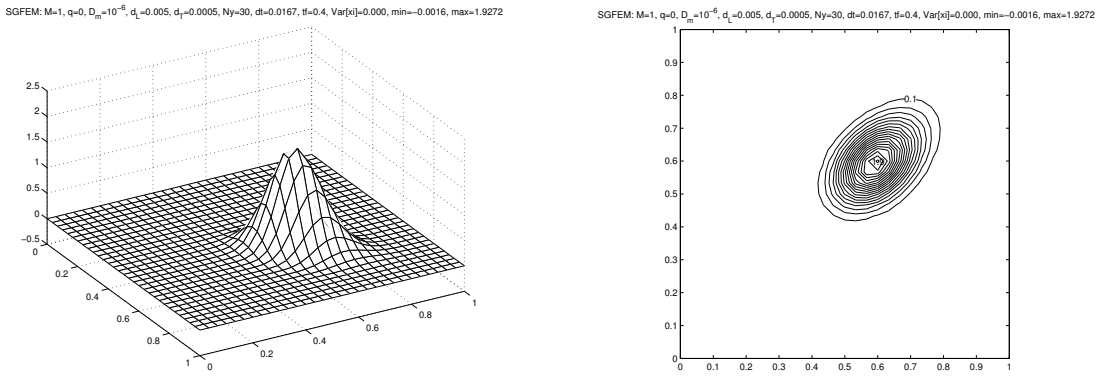


Figure 6.26: GAL solution of the deterministic advection-diffusion-dispersion problem

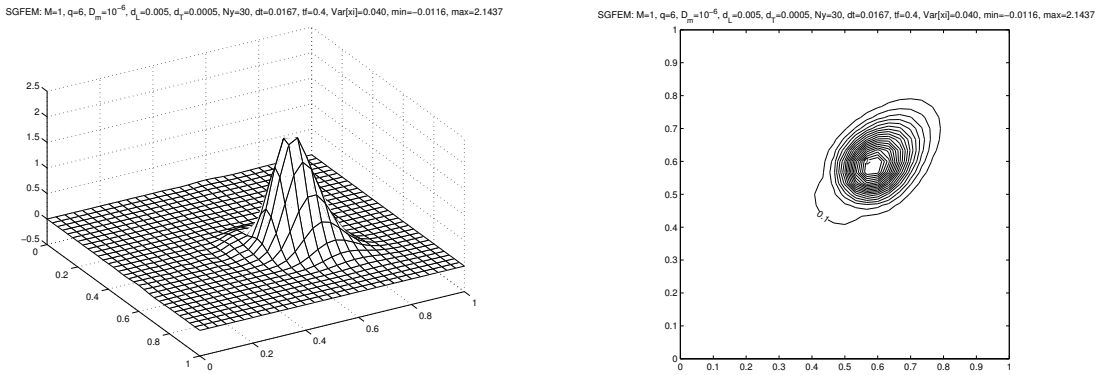


Figure 6.27: Mean solution of the stochastic advection-diffusion-dispersion problem with $\sigma = 0.2$

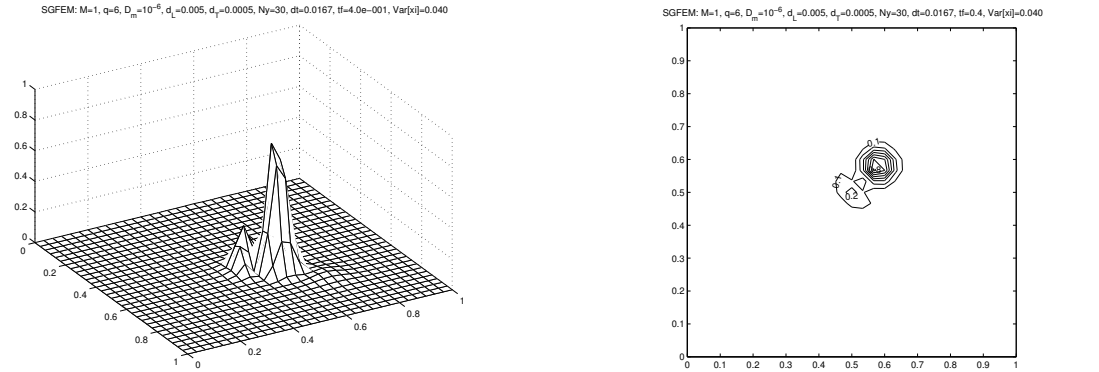


Figure 6.28: Variance of the SGFEM solution with $\sigma = 0.2$

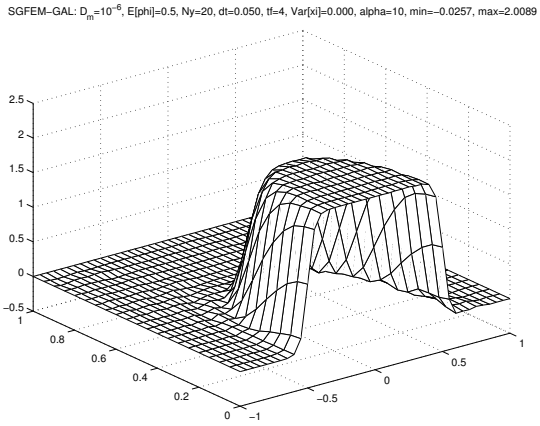


Figure 6.29: GAL solution for $\Phi = 0.5$

6.4 The Advection-Diffusion Equation with Stochastic Porosity

In this last section, we examine the influence of a stochastic porosity on the solution. Regarding (2.3), the diffusion coefficient D depends on the porosity via $D = \Phi d_m$. Hence, also D will become a stochastic process. We consider the equation

$$\Phi c_t + \boldsymbol{u} \cdot \nabla c - D\Delta c = 0, \qquad \text{in } \Omega = [-1,1] \times [0,1],\ t > 0$$

with $\boldsymbol{u} = (u, v)^\top$,

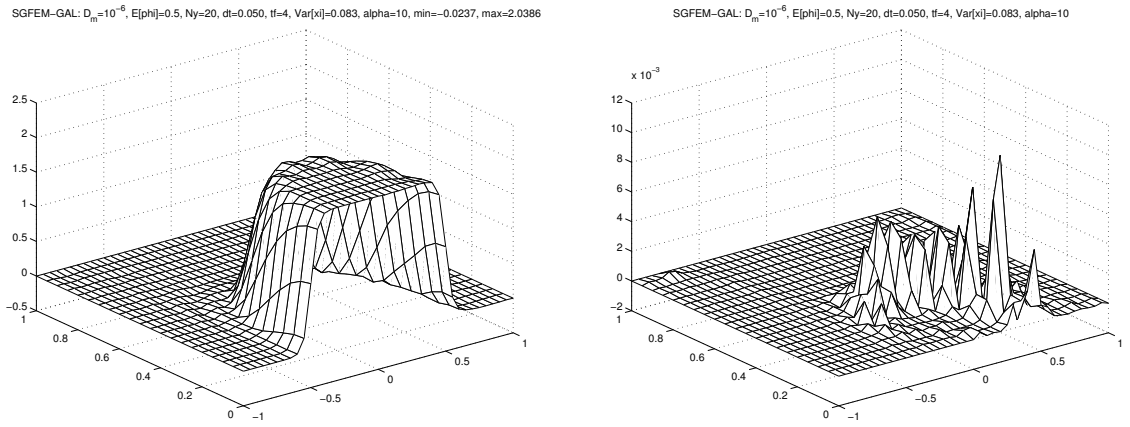
$$u = 2y(1 - x^2), \quad v = -2x(1 - y^2).$$

This time, the porosity is assumed to take all possible values $0 < \Phi \leq 1$. We use the data $d_m = 10^{-6},\ h = 1/20,\ t = 4,\ \alpha = 10$ and a time step $\Delta t = 1/20$. Figure 6.29 contains the result for the deterministic case with $\Phi = 0.5$, and Figure 6.30 shows the result for $\Phi = \overline{\Phi} + \xi,\ \xi \sim U[-0.5, 0.5]$ with $\overline{\Phi} = 0.5$. We observe that the variance of the solution attains the highest values in the region along the sharp front of the solution. Besides, we could not observe any significant difference in the behavior of the expected value of the stochastic solution and the deterministic solution. (Since this example is not relevant in practice, it will not be carried out further now.)

6.5 Conclusion

In this chapter, the influence of uncertainties in the porosity Φ, the velocity $\boldsymbol{V}$, and the Diffusion D was examined. Emphasis was put on the one dimensional equation due to the additional dimension which is introduced by randomness.

As expected, the stochastic solution differs from the mean solution. However, one cannot state, for example, that uncertainties always have a damping influence on the solution. The behavior of the solution depends on the underlying distribution function. We have observed that the solution with a stochastic diffusion coefficient D is damped less than the corresponding deterministic solution with the mean value of D. In contrast, a



expected value variance

Figure 6.30: Expected value and variance of the solution: $\Phi = \overline{\Phi} + \xi,\ \xi \sim U[-0.5, 0.5]$

stochastic velocity field acts like an artificial diffusion and damps the solution. In the last example with a stochastic dispersion, both effects influence each other and it is difficult to predict the behavior of the stochastic solution. Sometimes, an analytic solution can be derived if the exact solution of the deterministic problem is given and integration with respect to the random variable can be carried out. This is, of course, not always possible. Furthermore, we observed that a smaller correlation length in the covariance kernel of the stochastic velocity field results in less damping and a smaller variance of the solution.

On the other hand, there are some results which were to be expected and can be generalized. For example, the smaller the variance of the stochastic input, the closer the solution gets to the deterministic result.

There are also some open questions. It turned out that modelling space dependent stochastic processes requires much more effort than time dependent processes. An analytical deterministic solution for a space-dependent velocity would be desirable in order to examine this behavior. Moreover, only Eulerian methods were used to solve the stochastic partial differential equations. Since characteristic methods are preferred for hyperbolic advection-dominated problems, it is desirable to apply them to stochastic equations as well. Further work is required to examine how randomness can be incorporated into these methods.

Chapter 7

Summary

The mathematical modelling of transport in porous media results in advection diffusion equations. They are mostly dominated by advection and can be classified as parabolic equations of essentially hyperbolic character. Their numerical treatment often presents severe numerical and analytical difficulties due to exponential or parabolic boundary layers. Standard finite element techniques tend to generate numerical solutions with severe nonphysical over- and undershoot.

The standard techniques for dealing with oscillations stemming from the spatial discretization involve adding artificial diffusion. They arose from the study of the steady-state case. For Galerkin finite element methods this is accomplished by using modified forms of the standard test function as weighting functions. These methods, also known as upwinding, produce non-oscillating solutions and have good stability properties but are less accurate. However, improvements can be made by a careful choice of the upwind parameter as in the quadratic Petrov Galerkin method. The streamline diffusion method improves the stability and reduces numerical diffusion. It differs from the original upwind techniques in that the test functions are no longer modified by a higher degree polynomial but by a lower degree function than the element shape function.

In transient problems, the time dependence introduces new numerical difficulties such as numerical dispersion and the inaccurate representation of phase speed. The techniques and formula derived for the steady-state problems generally produce overly diffusive solutions. However, the use of test functions two degrees higher than the trial functions, as the cubic Petrov Galerkin method, improves both spatial and temporal accuracy. Unlike the quadratic upwinding, it does not add any artificial damping to the solution. The advantages of the streamline diffusion method for steady-state problems can be transferred to transient problems by using space-time finite elements and including the time dependency in the test functions. This yields solutions which are smooth outside internal layers but over- and undershoots appear at sharp fronts.

Due to the hyperbolic character of the underlying equation, characteristic methods are recommended. They are more accurate even for large time steps because the temporal discretization follows the characteristics. However, many characteristic methods fail to conserve mass or to treat general boundary conditions. The ELLAM scheme, which is based on a forward tracking approach, overcomes these difficulties. The numerical results show that the ELLAM scheme yields better results than the Eulerian methods but it is the most time consuming method.

Uncertainties in the data of the underlying equations can be modelled as stochastic processes. They are approximated by finite dimensional noise, for example by a truncated Karhunen-Loève expansion, under the assumption that the random variables are mutually independent. The stochastic equation can be transferred into a parametric deterministic problem. At this point the finite element method can be used to compute an approximate solution. When the finite dimensional stochastic subspace consists of polynomials which are orthogonal with respect to the given density function, then the system of equations is decoupled with respect to the random variables. The dimension of the stochastic subspace corresponds to the number of linear systems which have to be solved. However, this number can be very large in the case of poorly correlated input. The outcome of the theory is an expansion of the physical size of interest in terms of the random variables. This expansion can be used to generate realizations, or to compute statistics of the solution process. Exponential convergence can be observed when the stochastic input is represented accurately, for example if the stochastic process is already given as a sum of terms that are linear with respect to the random variables. In general, the stochastic solution differs from the solution which is obtained if the mean values of the stochastic coefficients are used in the computations. It is not easy to predict whether the stochastic solution will be damped more or less than the deterministic one without further calculations, and sometimes it is difficult to interpret the results. We hope that some effort in this direction can be made if the stochastic finite element method gains acceptance in science and engineering.

Appendix A

ELLAM Equations in one Dimension

This appendix contains the ELLAM equations for a Courant number $1 \leq Cr < 2$, constant mesh spacing Δx and constant velocity u.

A.1 Equations for the Inflow Boundary

The following notations are used:

$$
\begin{aligned}
H_1 &= \int_{t_1^*}^{t^n} c(0,t)dt \approx (t^n - t_1^*)C_0^n, \\
H_2 &= \int_{t_1^*}^{t^n} t\, c(0,t)dt \approx t^n(t^n - t_1^*)C_0^n, \\
H_3 &= \int_{t^{n-1}}^{t_1^*} c(0,t)dt \approx (t_1^* - t^{n-1})C_0^{n-1}, \\
H_4 &= \int_{t^{n-1}}^{t_1^*} t\, c(0,t)dt \approx t^{n-1}(t_1^* - t^{n-1})C_0^{n-1}, \\
G_1 &= \int_{t_1^*}^{t^n} c_x(0,t)dt \approx (t^n - t_1^*)\frac{\partial C_0^n}{\partial x}, \\
G_2 &= \int_{t_1^*}^{t^n} t\, c_x(0,t)dt \approx t^n(t^n - t_1^*)\frac{\partial C_0^n}{\partial x} \\
G_3 &= \int_{t^{n-1}}^{t_1^*} c_x(0,t)dt \approx t^{n-1}(t_1^* - t^{n-1})\frac{\partial C_0^{n-1}}{\partial x}, \\
G_4 &= \int_{t^{n-1}}^{t_1^*} t\, c_x(0,t)dt \approx t^{n-1}(t_1^* - t^{n-1})\frac{\partial C_0^{n-1}}{\partial x}
\end{aligned}
$$

A.1.1 w_0^n

$$
\begin{aligned}
I_1 &= \int_{x_0}^{x_1} c(x,t^n) w_0^n(x,t^n)dx = \frac{1}{6}\Delta x C_0^n + \frac{1}{3}\Delta x C_1^n, \\
I_2 &= 0, \\
I_3 &= u\int_{t_1^*}^{t^n} c(0,t) w_0^n(0,t)dt = u(1-u\frac{t^n}{\Delta x})H_1 + u^2\frac{1}{\Delta x}H_2, \\
I_4 &= D\frac{(-1)}{\Delta x}\int_{t_1^*}^{t^n} c(x_r^0(t),t)dt - D\frac{(-1)}{\Delta x}\int_{t_1^*}^{t^n} c(0,t)dt = -\frac{D}{\Delta x}(t^n - t_1^*)C_1^n + \frac{D}{\Delta x}H_1, \\
I_5 &= D\int_{t_1^*}^{t^n} c_x(0,t) w_0^n(0,t)dt = \frac{D}{\Delta x}(\Delta x - ut^n)G_1 + \frac{Du}{\Delta x}G_2, \\
I_6 &= \int_{\Omega_2^0} f(x,t) w_0^n(x,t)dxdt = \int_{\Omega_2^0} f(x,t)(\frac{\Delta x - x}{\Delta x} - u\frac{t^n - t}{\Delta x})dxdt.
\end{aligned}
$$

A.1.2 w_1^n

$$
\begin{aligned}
I_1 &= \int_{x_0}^{x_2} c(x,t^n) w_1^n(x,t^n)dx = \frac{1}{6}\Delta x C_0^n + \frac{2}{3}\Delta x C_1^n + \frac{1}{6}\Delta x C_2^n \\
I_2 &= \int_{x_0}^{x_2^*} c(x,t^{n-1}) w_1^n(x,t^{n-1})dx = \Delta x\left[\left(\frac{\alpha^2}{2} - \frac{\alpha^3}{6}\right)C_0^{n-1} + \frac{\alpha^3}{6}C_1^{n-1}\right] \\
I_3 &= u\int_{t^{n-1}}^{t^n} c(0,t) w_1^n(0,t)dt = u(2-u\frac{t^n}{\Delta x})H_3 + \frac{u^2}{\Delta x}H_4 + \frac{u^2 t^n}{\Delta x}H_1 - \frac{u^2}{\Delta x}H_2 \\
I_4 &= D\frac{1}{\Delta x}\int_{t_1^*}^{t^n} c(x_c^1(t),t)dt - D\frac{1}{\Delta x}\int_{t_1^*}^{t^n} c(0,t)dt + \\
&\quad D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_r^1(t),t)dt - D\frac{(-1)}{\Delta x}\int_{t_1^*}^{t^n} c(x_c^1(t),t)dt - D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t_1^*} c(0,t)dt \\
&= \frac{2D}{u}C_1^n - \frac{D\Delta x}{\Delta t}C_2^n - \frac{D}{\Delta x}H_1 + \frac{D}{\Delta x}H_3 \\
I_5 &= D\int_{t^{n-1}}^{t^n} c_x(0,t) w_1^n(0,t)dt \\
&= \frac{D}{\Delta x}(2\Delta x - ut^n)G_3 + \frac{Du}{\Delta x}G_4 + \frac{Dut^n}{\Delta x}G_1 - \frac{Du}{\Delta x}G_2 \\
I_6 &= \int_{\Omega_1^1 \cup \Omega_2^1} f(x,t) w_1^n(x,t)dxdt \\
&= \int_{\Omega_1^1} f(x,t)(\frac{x}{\Delta x} + u\frac{t^n - t}{\Delta x})dxdt + \int_{\Omega_2^1} f(x,t)\left(\frac{2\Delta x - x}{\Delta x} - u\frac{t^n - t}{\Delta x}\right)dxdt
\end{aligned}
$$

A.1.3 w_2^n

$$
\begin{aligned}
I_1 &= \int_{x_1}^{x_3} c(x,t^n) w_2^n(x,t^n)dx = \frac{1}{6}\Delta x C_1^n + \frac{2}{3}\Delta x C_2^n + \frac{1}{6}\Delta x C_3^n \\
I_2 &= \int_{x_0}^{x_3^*} c\, w_2^n(x,t^{n-1})dx \\
&= \Delta x \left[\left(\frac{\alpha^3}{6} - \alpha^2 + \frac{\alpha}{2} + \frac{1}{3} \right) C_0^{n-1} + \left(-\frac{\alpha^3}{2} + \frac{\alpha^2}{2} + \frac{\alpha}{2} + \frac{1}{6} \right) C_1^{n-1} + \frac{\alpha^3}{6} C_2^{n-1} \right] \\
I_3 &= u \int_{t^{n-1}}^{t_1^*} c(0,t) w_2^n(0,t)dt = -u(1 - \frac{ut^n}{\Delta x})H_3 - \frac{u^2}{\Delta x} H_4 \\
I_4 &= D\frac{1}{\Delta x} \int_{t^{n-1}}^{t^n} c(x_c^2(t),t)dt - D\frac{1}{\Delta x} \int_{t_1^*}^{t^n} c(x_l^2(t),t)dt - D\frac{1}{\Delta x} \int_{t^{n-1}}^{t_1^*} c(0,t)dt \\
&\quad + D\frac{(-1)}{\Delta x} \int_{t^{n-1}}^{t^n} c(x_r^2(t),t)dt - D\frac{(-1)}{\Delta x} \int_{t^{n-1}}^{t^n} c(x_c^2(t),t)dt \\
&= -\frac{D}{u} C_1^n + 2\frac{D\Delta t}{\Delta x} C_2^n - \frac{D\Delta t}{\Delta x} C_3^n - \frac{D}{\Delta x} H_3 \\
I_5 &= D \int_{t^{n-1}}^{t_1^*} c_x(0,t) w_2^n(0,t)dt = -D(1 - \frac{ut^n}{\Delta x})G_3 - \frac{Du}{\Delta x} G_4 \\
I_6 &= \int_{\Omega_1^2 \cup \Omega_2^2} f(x,t) w_2^n(x,t)dxdt \\
&= \int_{\Omega_1^2} f(x,t)(\frac{x - \Delta x}{\Delta x} + u\frac{t^n - t}{\Delta x})dxdt + \int_{\Omega_2^2} f(x,t) \left(\frac{3\Delta x - x}{\Delta x} - u\frac{t^n - t}{\Delta x} \right) dxdt
\end{aligned}
$$

A.2 Equations for the Outflow Boundary

A.2.1 w_E^n

$$
\begin{aligned}
I_1 &= \int_{x_{E-1}}^{x_E} c(x,t^n) w_E^n(x,t^n)dx \\
I_2 &= \int_{x_{E-1}^*}^{x_{E+1}^*} c(x,t^{n-1}) w_E^n(x,t^{n-1})dx \\
I_3 &= u\int_{t_{E+1}^*}^{t^n} c(l,t) w_E^n(l,t)dt \\
I_4 &= D\frac{1}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_c^E(t),t)dt - D\frac{1}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_l^E(t),t)dt + D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t_{E+1}^*} c(x_r^E(t),t)dt \\
&\quad + D\frac{(-1)}{\Delta x}\int_{t_{E+1}^*}^{t^n} c(l,t)dt - D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_c^E(t),t)dt \\
I_5 &= D\int_{t_{E+1}^*}^{t^n} c_x(l,t) w_E^n(l,t)dt \\
I_6 &= \int_{\Omega_1^E \cup \Omega_2^E} f(x,t) w_E^n(x,t)dxdt
\end{aligned}
$$

A.2.2 w_{E+1}^n

$$
\begin{aligned}
I_1 &= 0 \\
I_2 &= \int_{x_E^*}^{x_E} c(x,t^{n-1}) w_{E+1}^n(x,t^{n-1})dx \\
I_3 &= u\int_{t^{n-1}}^{t^n} c(l,t) w_E^n(l,t)dt \\
I_4 &= D\frac{1}{\Delta x}\int_{t^{n-1}}^{t_{E+1}^*} c(x_c^{E+1}(t),t)dt + D\frac{1}{\Delta x}\int_{t_{E+1}^*}^{t^n} c(l,t)dt - D\frac{1}{\Delta x}\int_{t^{n-1}}^{t^n} c(x_l^{E+1}(t),t)dt \\
&\quad + D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t_{E+1}^*} c(l,t)dt - D\frac{(-1)}{\Delta x}\int_{t^{n-1}}^{t_{E+1}^*} c(x_c^{E+1}(t),t)dt \\
I_5 &= D\int_{t^{n-1}}^{t^n} c_x(l,t) w_{E+1}^n(l,t)dt \\
I_6 &= \int_{\Omega_1^{E+1} \cup \Omega_2^{E+1}} f(x,t) w_{E+1}^n(x,t)dxdt
\end{aligned}
$$

A.2.3 w^n_{E+2}

$$
\begin{aligned}
I_1 &= 0 \\
I_2 &= \int_{x^*_{E+1}}^{x_E} u(x,t^{n-1}) w^n_{E+2}(x,t^{n-1})dx \\
I_3 &= u\int_{t^{n-1}}^{t^*_{E+1}} c(l,t) w^n_{E+2}(l,t)dt \\
I_4 &= D\frac{1}{\Delta x}\int_{t^{n-1}}^{t^*_{E+1}} c(l,t)dt - D\frac{1}{\Delta x}\int_{t^{n-1}}^{t^*_{E+1}} c(x_l^{E+2}(t),t)dt \\
I_5 &= D\int_{t^{n-1}}^{t^*_{E+1}} c_x(l,t) w^n_{E+2}(l,t)dt \\
I_6 &= \int_{\Omega_1^{E+2}} f(x,t) w^n_{E+2}(x,t)dxdt
\end{aligned}
$$

A.2.4 Discrete Equation associated with $w_E + w_{E+1} + w_{E+2}$ for $f = 0$

$$
\begin{aligned}
0 &= \frac{1}{3}\Delta x C^n_{E-1} + \frac{1}{6}\Delta x C^n_E - \Delta x[\frac{1}{6}(1-\alpha)^3 C^{n-1}_{E-3} + \left(\alpha^2 - \frac{3}{2}\alpha + \frac{2}{3}\right) C^{n-1}_{E-2} \\
&\quad + \left(\frac{1}{6}\alpha^3 - \frac{3}{2}\alpha^2 + \alpha + 1\right) C^{n-1}_{E-1} + \frac{1}{2}C^{n-1}_E] - \frac{D\Delta t}{\Delta x}\left[C^n_{E-1} - C^n_E\right]
\end{aligned}
$$

Appendix B

Element Matrices for Several Numerical Schemes

The following discrete equations correspond to the underlying differential equation

$$c_t + \nabla \cdot (Vc - D\nabla c) = f$$

with constant coefficients.

B.1 One dimension

B.1.1 Method of Lines

The global system of ordinary differential equations can be written as:

$$M\dot{c} + Ac - Bc = F,$$

where c is the vector of nodal unknowns, M is the mass matrix, B is the advection matrix, A is the diffusion matrix, and F is the vector containing boundary conditions and sources/sinks. In the following, the elemental matrices which combine to form the global matrices are listed for linear trial functions and different test functions on a uniform mesh. V and D denote the velocity and the diffusion on the corresponding element even though the index e is sometimes missing. Regarding the matrices, lines correspond to test functions, and columns to trial functions. These are the one-dimensional matrices. In higher dimensions, they can be combined with the Kronecker-product because of the tensor product approach.

B.1.1.1 Standard Galerkin Method

$$M^e = \frac{h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad B^e = \frac{V}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad A^e = \frac{D}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

B.1.1.2 QPG

$$M^e = \frac{h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \frac{h\nu}{12}\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix},$$

$$B^e = \frac{V}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} + \frac{V\nu}{6}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, \quad A^e = \frac{D}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

B.1.1.3 CPG

$$M^e = \frac{h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \frac{h\nu}{60}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, \quad B^e = \frac{V}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad A^e = \frac{D}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

B.1.2 SDM, Discontinuous Galerkin Method in Time

At the end, one has to solve a global linear system of equations $Gc = F$. Introduce

$$M_t = \frac{\Delta t}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad B_t = \frac{1}{2}\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \quad A_t = \frac{1}{\Delta t}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and

$$M_h = \frac{h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad B_h = \frac{1}{2}\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \quad A_h = \frac{1}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The elemental left hand side matrix has the following form:

$$\begin{aligned} G^e &= (c_t, w)_e + \delta[(c_t, w_t)_e + V(c_t, w_x)_e] + V(c_x, w)_e + \delta V[(c_x, w_t)_e + V(c_x, w_x)_e] \\ &\quad + D(c_x, w_x)_e + (c, w)_e^{n-1} \end{aligned}$$

$$\begin{aligned} G^e &= B_t' \otimes M_h + \delta[A_t \otimes M_h + VB_t' \otimes B_h] + VM_t \otimes B_h' + \delta V[B_t \otimes B_h' + VM_t \otimes A_h] \\ &\quad + DM_t \otimes A_h + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes M_h \end{aligned}$$

B.2 Two Dimensions

B.2.1 Method of Lines

The system of equations can again be written as $M\dot{c} + Ac - Bc = F$. In the sequel M_x denotes M_h with $h = \Delta x$, $M_y = M_h$ with $h = \Delta y$ and $M_t = M_h$ with $h = \Delta t$. Then

$$M^e = M_x \otimes M_y, \; A^e = D^e[A_x \otimes M_y + M_x \otimes A_y], \; B^e = V_x^e B_x \otimes M_y + V_y^e M_x \otimes B_y.$$

For CPG one has to use the corresponding elemental matrices from the one dimensional CPG method.

B.2.2 SDM, Discontinuous Galerkin Method in Time

The elemental left hand side matrix has the following form:

$$\begin{aligned} G^e &= (c_t, w)_e + \delta[(c_t, w_t)_e + (c_t, V \cdot \nabla w)_e] + (V \cdot \nabla c, w)_e \\ &\quad + \delta[(V \cdot \nabla c, w_t)_e + (V \cdot \nabla c, V \cdot \nabla w)_e] + D(\nabla c, \nabla w)_e + (c, w)_e^{n-1} \end{aligned}$$

Assume a uniform triangulation with $\Delta x = \Delta y = h$. Then

$$\begin{aligned} G^e &= B'_t \otimes (M_h \otimes M_h) + \delta[A_t \otimes (M_h \otimes M_h) + B'_t \otimes (V_x^e(B_h \otimes M_h) + V_y^e(M_h \otimes B_h))] \\ &\quad + M_t \otimes (V_x(B'_h \otimes M_h) + V_y(M_h \otimes B'_h)) + \delta B_t \otimes (V_x(B'_h \otimes M_h) + V_y(M_h \otimes B'_h)) \\ &\quad + \delta[M_t \otimes (V_x^2(A_h \otimes M_h) + V_y^2(M_h \otimes A_h) + V_x V_y((B_h \otimes B'_h) + (B'_h \otimes B_h)))] \\ &\quad + DM_t \otimes ((A_h \otimes M_h) + (M_h \otimes A_h)) + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes (M_h \otimes M_h). \end{aligned}$$

Bibliography

[1] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800 – 825, 2004.

[2] M. A. Celia, T. F. Russel, I. Herrera, and R. Ewing. An Eulerian-Lagrangian localized adjoint method for the advection-diffusion equation. *Advances in Water Resources*, 13:187 – 206, 1990.

[3] R. Ewing and H. Wang. A summary of numerical methods for time-dependent advection-dominated partial differential equations. *Journal of Computational and Applied Mathematics*, 128:423 – 445, 2001.

[4] R. E. Ewing, editor. *The Mathematics of Reservoir Simulation*, volume 1 of *Frontiers in Applied Mathematics*. SIAM, 1983.

[5] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag New York, 1991.

[6] P.M. Gresho, R.L. Sani, and M.S. Engelmann. *Incompressible Flow and the Finite Element Method*, volume 1, Advection-Diffusion. John Wiley and Sons, 2000.

[7] F. Häfner, D. Sames, and H.-D. Voigt. *Wärme- und Stofftransport*. Springer-Verlag Heidelberg, 1992.

[8] I. Herrera, R. E. Ewing, M. A. Celia, and T. F. Russell. Eulerian-Lagrangian localized adjoint method: The theoretical framework. *Numerical Methods for Partial Differential Equations*, 9:431 – 457, 1993.

[9] E. Holzbecher. *Modellierung Dynamischer Prozesse in der Hydrologie*. Springer-Verlag Berlin Heidelberg, 1996.

[10] T. J. R. Hughes, L. P. France, and G. M. Hulbert. A new finite element formulation for computational fluid dynamics. *Computer Methods in Applied Mechanics and Engineering*, VIII. The Galerkin/Least-Squares Method for Advective-Diffusive Equations:173 – 189, 1989.

[11] M. Jardak, C.-H. Su, and G. E. Karniadakis. Spectral polynomial chaos solutions of the stochastic advection equation. Technical report, Division of Applied Mathematics, Brown University, 2001.

[12] C. Johnson, U. Nävert, and J. Pitkäranta. Finite element methods for linear hyperbolic problems. *Computer Methods in Applied Mechanics and Engineering*, 45:285 – 312, 1984.

[13] M. R. Kaazempur-Mofrad and C. R. Ethier. An efficient characteristic galerkin scheme for the advection equation in 3-d. *Computer Methods in Applied Mechanics and Engineering*, 191:5345 – 5363, 2002.

[14] M. Loève. *Probability Theory*. Springer-Verlag, fourth edition, 1977.

[15] J. D. Logan. *Transport Modeling in Hydrogeochemical Systems*, volume 15 of *Springer Series in Interdisciplinary Applied Mathematics*. Springer-Verlag New York, 2001.

[16] A. R. Mitchell and D. F. Griffiths. Upwinding by Petrov-Galerkin methods in convection-diffusion problems. *Journal of Computational and Applied Mathematics*, 6(3):219 – 228, 1980.

[17] F. Riesz and B. Sz. Nagy. *Functional Analysis*. Dover Publications Inc. New York, 1990.

[18] E. M. Rønquist. Numerical solution of partial differential equations by element methods. Lecture notes, Department of Mathematical Sciences, NTNU Trondheim, 2002.

[19] H.-G. Roos, M. Stynes, and L. Tobiska. *Numerical Methods for Singularly Perturbed Differential Equations*. Springer Series in Computational Mathematics. Springer-Verlag Berlin Heidelberg, 1996.

[20] R. M. Smith and A. G. Hutton. The numerical treatment of advection: A performance comparison of current methods. *Numerical Heat Transfer*, 5:439 – 461, 1982.

[21] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. Springer-Verlag Berlin Heidelberg, 1997.

[22] H. Wang. An optimal-order error estimate for an ELLAM scheme for two-dimensional linear advection-diffusion equations. *SIAM Journal on Scientific Computing*, 37(4):1338 – 1368, 2000.

[23] H. Wang, H. K. Dahle, R. Ewing, M.S. Espedal, R. C. Sharpley, and S. Man. An ELLAM scheme for advection-diffusion equations in two dimensions. *SIAM Journal on Scientific Computing*, 20(6):2160 – 2194, 1999.

[24] H. Wang, D. Liang, R. E. Ewing, S. L. Lyons, and G. Qin. An approximation to miscible flows in porous media with point sources and sinks by an Eulerian-Lagrangian localized adjoint method and mixed finite element methods. *SIAM Journal on Scientific Computing*, 22(2):561 – 581, 2000.

[25] H. Wang, X. Shi, and R. E. Ewing. An ELLAM scheme for multi-dimensional advection-reaction equations and its optimal-order error estimate. *SIAM Journal on Numerical Analysis*, 36(6):1846 – 1885, 2001.

[26] R. E. Wang, H.and Ewing, G. Qin, S. L. Lyons, and S. Man. A family of Eulerian-Lagrangian localized adjoint methods for multi-dimensional advection-reaction equations. 152:120 – 163, 1999.

[27] J. J. Westerink and D. Shea. Consistent higher degree Petrov-Galerkin methods for the solution of the transient convection-diffusion equation. *International Journal for Numerical Methods in Engineering*, 28:1077 – 1101, 1989.

[28] N. Wiener. The homogeneous chaos. *Amererican Journal of Mathematics*, 60:897 – 936.

[29] D. Xiu and G. E. Karniadakis. Modelling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, 191:4927 – 4948, 2002.