

# High-Order Methods for Parabolic Equations in Multiple Space Dimensions for Option Pricing Problems

**Dissertation**

zur Erlangung  
des akademischen Grades  
eines  
Doktors der Naturwissenschaften (Dr. rer. nat.)

der  
Fakultät 4 – Mathematik und Naturwissenschaften  
der  
Bergischen Universität Wuppertal (BUW)  
vorgelegt von

**Christian Hendricks**

geboren am 29.12.1984 in Recklinghausen

betreut von: Prof. Dr. Michael Günther (BUW)  
Prof. Dr. Matthias Ehrhardt (BUW)

Wuppertal, Dezember 2016

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20170704-143448-0

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3A468-20170704-143448-0>]

# Danksagung

An dieser Stelle gilt mein ganz besonderer Dank meinen Betreuern Prof. Dr. Matthias Ehrhardt und Prof. Dr. Michael Günther. Sie ermöglichten mir die Promotion in der Arbeitsgruppe Angewandte Mathematik und Numerische Analysis an der Bergischen Universität Wuppertal. Dabei standen sie immer für fachliche Diskussionen zur Verfügung und waren offen gegenüber meinen eigenen Ideen und Vorschlägen. Matthias begleitete mich dabei seit dem Beginn meiner Masterthesis und brachte mich auf das Thema der vorliegenden Dissertation.

Darüber hinaus möchte ich mich bei den anderen Mitgliedern, sowie den ehemaligen Mitgliedern der Arbeitsgruppe für die schönen gemeinsamen Jahre bedanken. Gerne erinnere ich mich an die vielen fachlichen und oftmals auch nicht fachbezogenen Gespräche zurück. Die gemeinsamen Ausflüge zu diversen Fußballspielen werden mir immer in besonderer Erinnerung bleiben.

Insbesondere bedanke ich mich bei Dr. Christof Heuer, der immer ein offenes Ohr für meine unzähligen Fragen zu high-order-compact-Verfahren hatte.

An dieser Stelle möchte ich auch Tilo Gehlhaar und Dr. Sissis Kamarianakis für das Korrekturlesen dieser Dissertation danken.

Zu guter Letzt richtet sich mein Dank an meine Familie. Sie ermutigte mich das Studium der (Wirtschafts-)Mathematik zu beginnen und unterstützte mich auch während des Studiums. Von ganzem Herzen danke ich Anja Reßler für ihre Liebe und Unterstützung.



# Acknowledgement

First and foremost I would like to thank my supervisors Prof. Dr. Matthias Ehrhardt and Prof. Dr. Michael Günther. They gave me the opportunity to write my PhD thesis at the chair of Applied Mathematics and Numerical Analysis at the Bergische University of Wuppertal. During my studies they always provided the room for fruitful discussions and were open-minded towards my own ideas and research topics. Matthias guided me since the beginning of my master's thesis and suggested the topic of this PhD thesis.

In addition I would like to thank all my colleagues and former colleagues of the working group for the wonderful years working together. I will always cherish the many topic- and often non-topic-related discussions. Our trips to various soccer games will always be a very special memory for me.

In particular I would like to thank Dr. Christof Heuer, who took the time and patience to answer my innumerable questions on high-order-compact methods.

I would also like to thank Tilo Gehlhaar and Dr. Sissis Kamarianakis for proofreading this thesis.

Finally, special thanks go out to my family. They encouraged me to study (business) mathematics and supported me during the time of my studies. I wholeheartedly thank Anja Reßler for her love and support.



# Foreword

The core of this thesis is based on the following publications which have been written during the time of my PhD studies from May 2013 till December 2016.

- C. Hendricks, M. Ehrhardt, and M. Günther. High-Order ADI Schemes for Diffusion Equations with Mixed Derivatives in the Combination Technique. *Appl. Numer. Math.*, 101:36–52, 2016.

In this article high-order-compact finite differences in combination with ADI time stepping methods are applied to multi-dimensional diffusion equations. Furthermore, the stability of these schemes is analyzed in the *von Neumann* framework. It can be shown that the stability regions of the proposed schemes coincide with their central second-order finite difference counterparts. Thus, one can obtain fourth-order accuracy in space without any restrictions regarding the stability. For higher dimensional problems grid based methods suffer from an exponential growth of the number of unknowns – *the curse of dimensionality*. In order to compute solutions with significantly less grid nodes while maintaining a high accuracy in space we use the *sparse grid combination technique*.

Parts of Sections 3.3.1–3.3.2 and the numerical experiments in Section 3.3.4 have been taken from this publication.

- C. Hendricks, C. Heuer, M. Ehrhardt, and M. Günther. High-Order ADI Finite Difference Schemes for Parabolic Equations in the Combination Technique with Application in Finance. *J. Comput. Appl. Math.*, to appear, 2016.

The idea of high-order ADI schemes is carried over to general convection-diffusion equations with space-dependent coefficient functions. The stability is investigated for the frozen coefficient case in a two-dimensional setting within the von Neumann framework. The stability analysis in three and four spatial dimensions is performed experimentally. In addition to this, numerical experiments with multi-dimensional basket option pricing problems in a sparse grid setting illustrate the theoretical findings.

Sections 2.2, 3.3.3 and the numerical experiments in Section 4.3.2 contain excerpts of this publication.

- C. Hendricks, M. Ehrhardt, and M. Günther. Error Splitting Preservation for High Order Finite Difference Schemes in the Combination Technique. *Numer. Math: Theory, Models and Appl.*, to appear, 2016.

In this work the error splitting structure of fourth-order finite difference schemes

is analyzed. The analysis follows the approach presented in the seminal article by Reisinger [79]. Furthermore, we show how to preserve the splitting structure with multivariate interpolation schemes. The multivariate interpolant is constructed as a tensor product of univariate interpolation schemes. Since in the sparse grid combination technique the discrete finite difference sub-solutions have to be extended via interpolation the preservation of the error splitting structure is especially important in order to maintain a high accuracy on the entire domain.

The derivation of the error splitting structure for finite difference schemes in Section 2.4.1 is based on this publication.

- C. Hendricks, M. Ehrhardt, and M. Günther. High Order Combination Technique for the Efficient Pricing of Basket Options. *Acta Math. Univ. Comen.*, 84(2):243–253, 2015.

Here, fourth-order finite difference schemes are used to price European basket options under the multi-dimensional *Black-Scholes model*. In order to attack the curse of dimensionality the sparse grid combination technique is applied. Furthermore, grid transformations are employed to simplify the partial differential equation. The adverse effects due to the non-smooth nature of the payoff function are tackled with the smoothing operator given by Kreiss et al. [62].

The numerical schemes in Section 4.3 are inspired by this work.

- C. Hendricks, M. Ehrhardt, and M. Günther. Hybrid Finite Difference / Pseudospectral Methods for the Heston and Heston-Hull-White PDE. *J. Comp. Fin.*, to appear, 2016.

In this article we propose a hybrid spatial finite difference / pseudo-spectral discretization for European option pricing problems under the Heston and Heston-Hull-White model. The scheme exploits the regularity of the solution in direction of certain coordinates by combining different discretization techniques. In the time direction we use ADI splitting methods. Due to the decomposition of the spatial discretization matrix the scheme can be shown to be very efficient. The computational complexity is analyzed and proven by numerical experiments. It turns out that the proposed method is able to beat its second-order benchmark method and provides robust and accurate results.

Sections 3.4, 3.5 and 4.4 are based on this article.

# Contents

<b>Foreword</b>	<b>V</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Mathematical Models . . . . .	3
1.2 Partial Differential Equation Methods in Computational Finance . . . . .	7
1.3 Literature Overview and Outline of this Thesis . . . . .	9
<b>2 Spatial Discretization</b>	<b>13</b>
2.1 Standard Finite Difference Methods . . . . .	15
2.2 High-Order-Compact Finite Difference Methods . . . . .	17
2.3 Pseudo-Spectral Methods . . . . .	21
2.4 The Curse of Dimensionality and the Sparse Grid Combination Technique	26
<b>3 Time Discretization - Alternating Direction Implicit Schemes</b>	<b>49</b>
3.1 Stability Considerations . . . . .	53
3.2 High-Order Finite Difference ADI Schemes . . . . .	55
3.3 Stability of HO-ADI Schemes . . . . .	58
3.4 (Hybrid) Pseudo-Spectral ADI Schemes . . . . .	71
3.5 Stability of Hybrid Finite Difference/Pseudo-Spectral ADI Schemes . . . . .	72
<b>4 Application to Financial Engineering</b>	<b>77</b>
4.1 Grid Transformation . . . . .	77
4.2 Non-Smooth Initial Data . . . . .	82
4.3 Basket Options . . . . .	84
4.4 Stochastic Volatility Models . . . . .	94
<b>5 Conclusions and Outlook</b>	<b>113</b>
5.1 Conclusions . . . . .	113
5.2 Outlook . . . . .	114
<b>References</b>	<b>115</b>



# 1

## Chapter 1

---

# Introduction

In today's financial markets derivatives are important products for traders and investors. They can be understood as some kind of insurance against the market risk of an uncertain development of future prices. Hence, they give market participants the possibility to manage their risk of a position in a portfolio of assets. Furthermore, derivatives can be actively traded to achieve speculative gains. A financial derivative is a right, whose price depends on the market price of other financial instruments. The dependent instrument is called *underlying*. Typical underlyings are

- securities, e.g. stocks, bonds,
- indices, e.g. DAX 30 (Deutscher Aktien Index), FTSE 100 (Financial Times Stock Exchange),
- interest rates, e.g. 3-months Euribor, 3-months Libor,
- currencies, e.g. Euro, US-Dollar,
- derivatives, e.g. options, futures, swaps.

We cite the definition of derivatives according to the German stock trading law §2 Abs. 2 WpHG, see [34]:

**Definition 1.** *Derivatives*

*Derivatives according to the law are*

1. *forward transactions or option contracts, whose price conditionally or unconditionally depends on*
  - a) *the exchange or market price of securities,*
  - b) *the exchange or market price of money market instruments,*
  - c) *interest rates or other earnings,*
  - d) *the exchange or market price of commodities or precious metals,*
2. *forward exchange transactions which are traded at an organized market, foreign exchange options, currency swaps, currency swap option transactions, currency future option transactions.*

Financial derivatives can roughly be divided into two categories: conditional and unconditional forward transactions. Unconditional derivatives force both contract partners to fulfill the contractual obligation, while conditional derivatives give the buyer the opportunity to either exercise his right or to let it expire. An obligation only applies to the

seller of the contract. Derivatives allow to separate market risks of underlyings, so that they can be traded individually. This enables market participants to use them for

- speculation:  
traders exploit price differences in the time lapse to earn money. Compared to a direct investment in the underlying, only a small fraction of capital has to be invested. The resulting *leverage effect* offers the opportunity to participate disproportionately in price moves. On the one hand this enables to earn large speculative gains, but on the other hand it leads to a high risk of loosing the invested money or even of running into debts.
- arbitrage:  
traders utilize price differences to earn money without or with hardly any risk. For example, they could buy a stock at one exchange and sell it at another, if the prices are different. Following the *efficient market hypothesis* with the *no-arbitrage principle* there should not be any opportunity for arbitrage.
- hedging:  
market participants hedge their portfolio to prevent financial risks. For example, an investor with long positions in stocks can buy financial derivatives to hedge his portfolio against sudden price crashes.

In this thesis we focus on conditional forward transactions in the form of options. We cite the definition of a financial option from [34]:

**Definition 2.** *Options*

*Options are standardized, exchange traded contracts, which give the buyer the right, but not the obligation*

- *to buy (Call) or sell (Put)*
- *a predefined amount of an underlying*
- *within a specified time period or at a fixed date,*
- *at a price  $K$ , specified at conclusion of the contract.*

In the following we give examples of the most common option styles:

- European option: the option can only be exercised at maturity.
- American option: the option can be exercised at any time during the term of the option.

These option types are often referred to as *plain vanilla options*. Beside them there is a huge variety of different option contracts, the so called *exotic options*, such as *Asian*, *Russian*, or *lookback* options, etc. Here the option price often depends on the value of the underlying during a certain time interval. Options with this feature are called *path-dependent*.

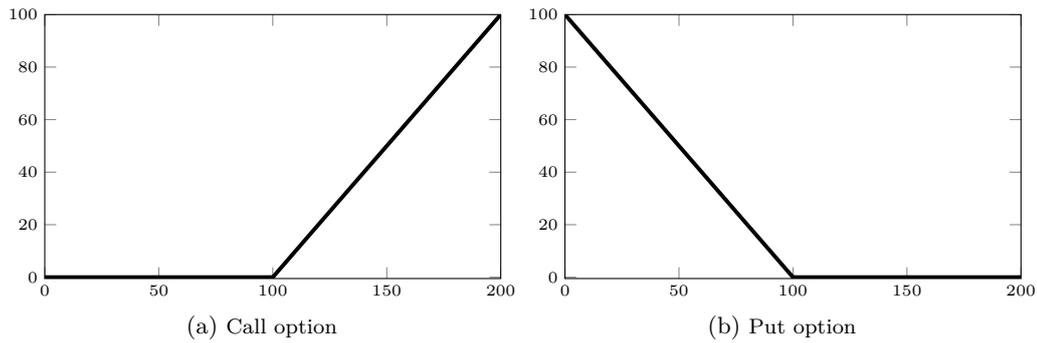


Figure 1.1: Payoff profile for plain vanilla call and put options with strike  $K = 100$ .

## 1.1 Mathematical Models

In the previous section we have discussed the importance of derivatives and options in modern financial markets. For traders and investors the question arises of what is the fair price of a financial option. The value of a plain vanilla option is determined by its payoff profile at maturity, but the value at contract conclusion is unknown. For example, the payoff of a plain vanilla call and of a put option is given by

$$\begin{aligned}\phi(s) &= \max\{s - K, 0\} =: (s - K)^+ && (\text{Call}), \\ \phi(s) &= \max\{K - s, 0\} = (K - s)^+ && (\text{Put}),\end{aligned}$$

where the asset price is denoted by  $s$  and the strike price is  $K \in \mathbb{R}^+$ . Figure 4.2 shows the payoff functions  $\phi$  for call and put options. In order to compute the fair price, mathematical models have to be applied. The price dynamics of the underlyings can be described via stochastic differential equations (SDEs). One very important classical model to describe the price movements of stocks is the *Black-Scholes model*.

**Definition 3.** *Black-Scholes model*

The  $d$ -dimensional Black-Scholes model consists of  $d$  assets with dynamics:

$$ds_i(t) = \mu_i s_i(t)dt + \sigma_i s_i(t)dW_i(t),$$

where  $\mu_i$  is the drift and  $\sigma_i$  the volatility of asset  $s_i$  for  $i = 1, 2, \dots, d$ . The Wiener processes  $W_i$  and  $W_j$  are correlated with  $dW_i dW_j = \rho_{ij} dt$ .

In the Black-Scholes framework the volatilities  $\sigma_i$  are assumed to be constant, which is one of the main restrictions of this model. Market data reveals that in practice volatilities are not constant. In equity markets one often observes a *volatility smile* for short maturities and a *volatility skew* for long maturities, see Figure 1.2. More sophisticated models assume a *local volatility* function or model the volatility by an additional stochastic process.

**Definition 4.** [19, 20] Dupire - Local volatility model

The local volatility model assumes the asset price to follow the SDE

$$ds(t) = \mu s(t)dt + \sigma(s(t), t)dW(t),$$

where  $\mu$  is the drift and the volatility  $\sigma$  depends on the spot price  $s$  and the time  $t$ .

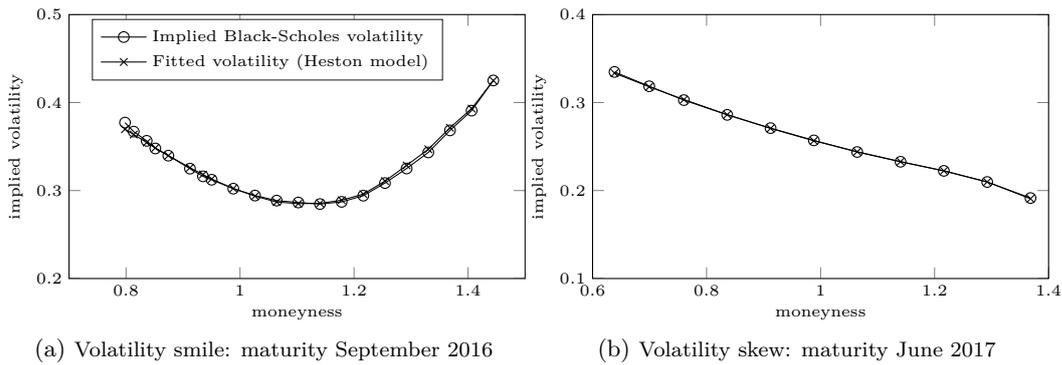


Figure 1.2: Volatility smile and skew pattern of Black-Scholes implied volatilities fitted to market prices and to prices of the Heston model. Historical option quotes of the Adidas<sup>®</sup> stock were taken from Eurex<sup>®</sup> 07.22.2016.

The local volatility function can be computed from the quoted call option prices  $c$  via

$$\sigma(K, t) = \frac{1}{K} \sqrt{2 \frac{\frac{\partial c}{\partial t}(K, t)}{\frac{\partial^2 c}{\partial K^2}(K, t)}}.$$

For details we refer to [19, 20]. Further local volatility models were introduced by Derman and Kani [15] and Rubinstein [81].

**Definition 5.** *Stochastic volatility models*

*In stochastic volatility models the dynamics of the stock price  $s$  as well as of the volatility  $v$  are determined by stochastic processes*

$$\begin{aligned} ds(t) &= \mu s(t)dt + \sqrt{v(t)}s(t)dW_1(t), \\ dv(t) &= \alpha(v(t), t)dt + \beta(v(t), t)dW_2(t), \end{aligned}$$

*with correlated Wiener processes  $dW_1dW_2 = \rho_{12}dt$ .*

Stochastic volatility models are frequently used in practice, as they can describe the volatility patterns which we observe in the markets. Figure 1.2 shows the Black-Scholes implied volatilities of Adidas<sup>®</sup> option quotes in comparison to the implied volatilities of a fitted stochastic volatility model (Heston). We observe that the stochastic model is capable of modeling the volatility structure much better than with a constant volatility as assumed by the Black-Scholes model.

We cite two well-known models: the Heston model and the SABR model.

**Definition 6.** [46] *Heston model*

*In the Heston model the spot price and volatility follow the processes*

$$\begin{aligned} ds(t) &= \mu s(t)dt + \sqrt{v(t)}s(t)dW_1(t), \\ dv(t) &= \kappa(\theta - v(t))dt + \sigma\sqrt{v(t)}dW_2(t), \end{aligned}$$

*with correlated Wiener processes  $dW_1dW_2 = \rho_{12}dt$ .*

**Definition 7.** *Stochastic Alpha, Beta, Rho (SABR) model [39]*

The SABR model assumes the forward price  $f(t)$  of the stock to follow the process

$$\begin{aligned}df(t) &= \alpha(t)f(t)^\beta dW_1(t), \\d\alpha(t) &= \sigma\alpha(t)dW_2(t),\end{aligned}$$

with volatility  $\alpha$  and correlated Wiener processes  $dW_1dW_2 = \rho_{12}dt$ .

The stochastic processes for the spot price  $s(t) = e^{-r(T-t)}f(t)$  can be derived with the help of the *Itô Lemma* (Lemma 9)

$$\begin{aligned}ds(t) &= rs(t)dt + e^{-r(T-t)(1-\beta)}\alpha(t)s(t)^\beta dW_1(t), \\d\alpha(t) &= \sigma\alpha(t)dW_2(t).\end{aligned}$$

The gap between SDEs, which describe the movement of the underlyings to the fair option value can be closed with the main principles of derivative pricing. We cite from [30]:

1. If a derivative security can be perfectly replicated (equivalently hedged) through trading in other assets, then the price of the derivative security is the cost of the replicating trading strategy.
2. Discounted (or deflated) asset prices are martingales under a probability measure associated with the choice of discount factor (or numeraire). Prices are expectations of discounted payoffs under such a martingale measure.
3. In a complete market, any payoff (satisfying modest regularity conditions) can be synthesized through a trading strategy and the martingale measure associated with a numeraire is unique. In an incomplete market there are derivative securities that cannot be perfectly hedged; the price of such a derivative is not completely determined by the prices of other assets.

Incorporating these principles, one can show that the option's price  $u$  is the discounted expected payoff under the *risk neutral* probability measure  $Q$

$$\begin{aligned}u(s(t), t) &= e^{-r(T-t)}\mathbb{E}_Q(\phi(s(T))|s(t)) \\ &= e^{-r(T-t)} \int_{\mathbb{R}} \phi(y)f(y|s(t))dy,\end{aligned}\tag{1.1}$$

where  $r$  is the risk free interest rate,  $T$  the maturity and  $f$  the density of the underlying's process. For a detailed derivation we refer to [30]. For most models the density  $f$  is not available in closed form. Therefore, numerical techniques have to be applied to compute the expectation. Three different classes of solvers are frequently used in computational finance:

- Monte Carlo methods

The system of SDEs is solved to draw samples from the density function. Let  $S^i$  be a realization of the stochastic process at point in time  $T$  with initial value  $s(t)$ , then it follows by the central limit theorem for independent and identically distributed

random numbers

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \phi(S^i) = \int_{\mathbb{R}} \phi(y) f(y|s(t)) dy,$$

where  $N$  is the number of realizations. The standard deviation of the difference between the sum truncated at fixed  $N$  and the integral can be shown to decrease with order  $\mathcal{O}(N^{-1/2})$ . The slow convergence is one of the main drawbacks of Monte Carlo methods. However, the accuracy can be improved by *variance reduction* techniques, such as *antithetic sampling*, *control variates*, etc, and by incorporating proper random number generators, e.g. *Fibonacci generators*, *Sobol sequences*, etc. On the other side Monte Carlo methods have two important advantages, see [83]: the order of accuracy holds independently of the dimension and the integrand does not have to be smooth as square integrability suffices. Furthermore, the computational complexity depends only weakly on the dimensionality. A detailed introduction to Monte Carlo methods can be found in many textbooks, e.g. [30, 83].

- Numerical integration

The integral in (1.1) is solved numerically. Typically, the density  $f$  is not known, but its characteristic function, given via

$$\varphi(x) = \int_{-\infty}^{\infty} f(y) e^{ixy} dy,$$

is available. From the definition it becomes clear that the characteristic function can be seen as the continuous *Fourier transform* of the density. By application of the inverse transform the density can be recovered. As the inverse can often not be derived analytically, numerical methods have to be used. The discrete transforms can be computed via the *Fast Fourier Transform* (FFT) algorithm with  $\mathcal{O}(N \log_2(N))$  operations, where  $N$  is the number of grid points within the numerical quadrature. In the literature several approaches have been discussed to efficiently compute the option values via Fourier transform techniques, e.g. in [12, 26, 35]. The advantage of this approach is its high accuracy with a very low computational effort. For example, the price of an European option under the Heston model can be computed with geometric accuracy in linear run-time with the COS method [26]. However, the technique relies on the characteristic function given in closed form. The covariance matrix of the system of SDEs has to fulfill certain affinity relations [18], such that the characteristic function can be derived. This restricts the applicability of the approach to certain problem classes.

- Partial differential equations (PDEs)

With the help of the *Feynman-Kac formula* or the *Itô Lemma* a link between the expected payoff of the price process and the solution of a PDE can be established. As in general an analytical solution to the arising PDE is unknown, numerical techniques have to be applied. Standard techniques rely on a discretization of the time-space domain, e.g. by finite differences [38, 51, 60], which results in a linear equation system with unknowns representing option values at certain grid nodes. Generally PDE techniques are more expensive compared to the integration methods described above and suffer, as a grid-based method, from the *curse of dimensionality*,

but are advantageous in terms of general applicability.

## 1.2 Partial Differential Equation Methods in Computational Finance

In this thesis we focus on PDE methods to compute the fair option value. The following Feynman-Kac formula provides a link between the model in form of SDEs and the PDE of the option price. For option pricing problems the payoff  $\phi$  is generally non-smooth and not twice differentiable, thus further techniques have to be employed to make it applicable. According to [73] the Feynman-Kac formula then applies to most payoffs of financial assets for log-normal models. Alternatively, one can use the Itô Lemma in combination with standard no-arbitrage arguments to derive the PDE formulation of the pricing problem, see e.g. [73, Section 3.3].

**Theorem 8.** [73, Proposition 2.5.1.] *Feynman-Kac formula*

Let the function  $g(x)$  be bounded, let  $\phi(x)$  be twice differentiable with compact support in  $K \subset \mathbb{R}$ , and let the function  $q(x)$  be bounded from below, and let  $X_t$  be given by the SDE

$$dX_t = a(X_t)dt + b(X_t)dW_t.$$

- For  $t \in [0, T]$ , the Feynman-Kac formula is

$$u(t, x) = \mathbb{E} \left( \int_t^T g(X_s) e^{-\int_t^s q(X_u) du} + e^{-\int_t^T q(X_s) ds} \phi(X_T) \mid X_t = x \right)$$

and is a solution of the following partial differential equation

$$\frac{\partial}{\partial t} u(t, x) + \frac{1}{2} b^2(x) \frac{\partial^2 u}{\partial x^2}(t, x) + a(x) \frac{\partial u}{\partial x}(t, x) - q(x) u(t, x) + g(x) = 0 \quad (1.2)$$

$$u(T, x) = \phi(x). \quad (1.3)$$

- If  $\omega(t, x)$  is a bounded solution to equations (1.2) and (1.3) for  $x \in K$ , then  $\omega(t, x) = u(t, x)$ .

**Lemma 9.** [73, Lemma 2.3.1.] *Itô Lemma*

Consider the process  $X_t$  with the SDE

$$dX_t = a(X_t)dt + b(X_t)dW_t.$$

For a function  $u(t, x)$  with at least one derivative in  $t$  and at least two derivatives in  $x$ , we have

$$du(t, X_t) = \left( \frac{\partial}{\partial t} + a(X_t) \frac{\partial}{\partial x} + \frac{b^2(X_t)}{2} \frac{\partial^2}{\partial x^2} \right) u(t, X_t) dt + b(X_t) \frac{\partial}{\partial x} u(t, X_t) dW_t.$$

Let a stock be governed by the Black-Scholes SDE (3) with  $d = 1$ , then either with the help of Theorem 8 or Lemma 9, one can show that the option value  $u$  has to fulfill the

following PDE in a risk-neutral setting

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma_1^2 s^2 \frac{\partial^2 u}{\partial s^2} + rs \frac{\partial u}{\partial s} - ru = 0, \quad (1.4)$$

Similar to this one dimensional example, a multivariate version of the Feynman-Kac formula or the Itô Lemma can be applied to derive PDEs of the multidimensional models:

**Multivariate Black-Scholes PDE** The multi-dimensional Black-Scholes PDE with  $d \in \mathbb{N}$  assets is given by

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{1}{2} \sum_{i,j=1}^d \sigma_i \sigma_j \rho_{i,j} s_i s_j \frac{\partial^2 u}{\partial s_i \partial s_j} + \sum_{i=1}^d r s_i \frac{\partial u}{\partial s_i} - ru = 0, \\ u(s_1, s_2, \dots, s_d, T) = \phi(s_1, s_2, \dots, s_d), \end{aligned} \quad (1.5)$$

in the space-time cylinder  $\Omega_d \times \Omega_t$  with  $\Omega_d = [0, \infty)^d$ ,  $\Omega_t = [0, T]$ . For a European basket option the payoff is typically given by

$$\begin{aligned} \phi(s_1, s_2, \dots, s_d) &= \left( \sum_{i=1}^d w_i s_i - K \right)^+ && (Call), \\ \phi(s_1, s_2, \dots, s_d) &= \left( K - \sum_{i=1}^d w_i s_i \right)^+ && (Put), \end{aligned}$$

with positive weights  $w_i$  for  $i = 1, 2, \dots, d$  summing up to one.

**Heston PDE** The Heston PDE [46] is given by

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{1}{2} s^2 v \frac{\partial^2 u}{\partial s^2} + \rho_{12} \sigma_1 s v \frac{\partial^2 u}{\partial s \partial v} + \frac{1}{2} \sigma_1^2 v \frac{\partial^2 u}{\partial v^2} + rs \frac{\partial u}{\partial s} + \kappa(\eta - v) \frac{\partial u}{\partial v} - ru = 0, \\ u(s, T) = \phi(s), \end{aligned} \quad (1.6)$$

for time  $t \in [0, T]$ , asset  $s \in [0, \infty)$  and volatility  $v \in [0, \infty)$ . The risk-less interest rate is denoted by  $r$  and the volatility of the volatility by  $\sigma_1$ . The long-term mean of  $v$  is given by  $\eta$ , while  $\kappa$  denotes the mean reversion rate of  $v$ . The correlation between the asset and the volatility is given by  $\rho_{12}$ .

**Heston-Hull-White PDE** The Heston-Hull-White (HHW) model is an extension of the Heston model, where the interest rate is assumed to follow a mean-reverting process. The option value is assumed to satisfy the PDE

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{1}{2} s^2 v \frac{\partial^2 u}{\partial s^2} + \frac{1}{2} \sigma_1^2 v \frac{\partial^2 u}{\partial v^2} + \frac{1}{2} \sigma_2^2 \frac{\partial^2 u}{\partial r^2} + \rho_{12} \sigma_1 s v \frac{\partial^2 u}{\partial s \partial v} + \rho_{13} \sigma_2 s \sqrt{v} \frac{\partial^2 u}{\partial s \partial r} \\ + \rho_{23} \sigma_1 \sigma_2 \sqrt{v} \frac{\partial^2 u}{\partial v \partial r} + rs \frac{\partial u}{\partial s} + \kappa(\eta - v) \frac{\partial u}{\partial v} + a_r (b_r - r) \frac{\partial u}{\partial r} - ru = 0, \\ u(s, T) = \phi(s), \end{aligned} \quad (1.7)$$

for time  $t \in [0, T]$ , asset  $s \in [0, \infty)$ , volatility  $v \in [0, \infty)$  and risk-free interest rate  $r \in (-\infty, \infty)$ . Compared to the Heston model, the HHW model has the following additional parameters: the volatility of the interest rate is  $\sigma_2$ ; the long-term mean of  $r$  is given by  $b_r$  and its mean reversion rate by  $a_r$ ; the correlation between  $s$  and  $r$  is denoted by  $\rho_{13}$  and between  $v$  and  $r$  by  $\rho_{23}$ . Although the Heston and HHW model were initially derived for one underlying asset, they can be extended to a basket of assets, where the volatility of each asset is driven by one stochastic process.

### 1.3 Literature Overview and Outline of this Thesis

The PDEs arising in financial option pricing problems, such as (1.5), (1.6), (1.7), are in general of convection-diffusion-reaction type

$$\begin{aligned} \frac{\partial u}{\partial t} &= Lu, & (\mathbf{x}, t) \in \Omega_d \times \Omega_t, \\ \text{with } Lu &= \sum_{i,j=1}^d a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^d c_i \frac{\partial u}{\partial x_i} + bu, \end{aligned} \quad (1.8)$$

on a rectangular domain  $\Omega_d \times \Omega_t$  with suitable initial and boundary data. This thesis is focused on the derivation of numerical methods to solve PDEs of this form in a financial option pricing setting.

In the literature several methods have been discussed to solve the PDEs arising in financial option pricing problems. The most common approach is to apply central finite differences to discretize the spatial domain with second-order accuracy, e.g. in [37, 38, 51, 60]. In a low dimensional setting Crank-Nicolson time marching is frequently used [4, 60, 90]. Even for a moderate number of spatial dimensions the resulting system of linear equations becomes expensive to solve. In't Hout et al. [37, 38, 51] applied dimensional splitting techniques, such as Alternating Direction Implicit (ADI) schemes, to derive efficient methods for the Heston and Heston-Hull-White PDEs. In addition to second-order accurate schemes higher-order discretizations were introduced and discussed by various researchers as well: Leentvaar and Oosterlee [66, 67] used standard fourth-order finite difference approximations, while Linde [69] employed broad sixth-order finite difference stencils. These schemes are generally more expensive from a computational point of view, since the discretization matrix is broadly banded. With the help of so called *high-order-compact* (HOC) schemes one can derive a fourth-order accurate approximation on the compact stencil [3, 21, 22, 47, 89]. Düring et al. [21, 22] constructed HOC schemes for stochastic volatility models with one underlying asset and one risk-factor as well as for basket options. In the time domain they applied Crank-Nicolson time stepping. Recently ADI splitting in combination with HOC discretizations has been introduced for convection-diffusion equations with mixed derivatives and constant coefficients by Düring et al. [23]. Their work was extended to PDEs arising in stochastic volatility models [25] and to the multivariate Black-Scholes model in [45].

Although the computational workload can be reduced significantly with ADI methods and the possibility to use HOC discretizations enables to achieve highly accurate solutions with fewer grid nodes, the methods still suffer from the exponential growth of the number of unknowns in a tensor based grid. For example, if 64 grid nodes are applied in each

coordinate direction, we have 4,096 degrees of freedom in a two-dimensional domain, but we already have 16,777,216 in a four-dimensional setting. Due to this immense growth the problems become very expensive to solve and the limit of finite available memory is reached quickly. In order to circumvent this *curse of dimensionality*, one can apply the *sparse grid combination technique* [33]. The method exploits the error structure of the underlying numerical scheme to reduce the number of grid nodes, while maintaining a high accuracy. In a financial setting the combination technique has been applied successfully to basket option pricing problems with second-order accuracy by Reisinger [78] and with standard fourth-order finite differences by Leentvaar and Oosterlee [66].

Beside finite differences other discretization techniques to solve PDEs in financial engineering applications have been applied as well, e.g. finite element-finite volume [100], multigrid [13] and spectral methods [44, 74, 99].

In this thesis we aim to derive efficient numerical schemes for option pricing in multiple space dimensions. Therefore, we combine several numerical approaches: in the spatial domain we consider high-order finite differences and spectral discretizations under variable transformations. The transformations allow to simplify the equation and to cluster grid points in a critical region where a high accuracy is desired. In the time domain we apply ADI time stepping. Based on the stability results for ADI schemes in [54, 55, 63, 71] we derive stability bounds for high-order finite difference ADI schemes applied to diffusion and convection-diffusion equations in the *von Neumann* framework.

Spectral methods with ADI splitting have rarely been used in the literature. To our best knowledge there only exists one article by Zeng et al. [97]. In order to motivate the use of ADI time stepping with a spectral discretization in space we discuss and numerically validate the computational effort. The stability of these methods is analyzed numerically.

Furthermore, we apply the sparse grid combination technique to compensate the effects of the curse of dimensionality. The technique is used for high-order finite differences as well as for pseudo-spectral methods. We investigate for which problem classes and regularity requirements sparse grids are suitable and superior to the full-grid approach.

After the methods have been introduced and their properties have been analyzed, we apply them to price basket options and options under stochastic volatility. As a test case for two- and three-factor models we consider the Heston and Heston-Hull-White PDE. We derive hybrid schemes, which use, depending on the regularity of the solution, different discretizations in the single coordinate directions.

### 1.3.1 Outline

In Section 2 we discuss approximations of the spatial operator  $L$ . First standard finite differences are introduced, which are frequently used in practice. Based on these finite difference stencils HOC discretizations are derived. They enable us to achieve fourth-order accuracy on the compact stencil. Hence, one can compute highly accurate solutions with the same computational effort as with second-order finite differences. In Section 2.3 we consider, in contrast to the local approximation approach of finite differences, the global ansatz of pseudo-spectral methods. These methods are highly accurate if the solution fulfills certain regularity requirements. In order to reduce the number of grid nodes within the spatial grid, we apply sparse grids for finite differences and pseudo-spectral

schemes.

Section 3 is devoted to the time discretization. Here we investigate the properties of Alternating Direction Implicit (ADI) splitting schemes for high-order (HO) spatial discretizations. In the case of HO finite difference methods we derive stability bounds for diffusion and convection-diffusion equations in the von Neumann framework. In the case of pseudo-spectral methods we are especially interested in the stability properties of hybrid methods, which use pseudo-spectral differentiation as well as finite differences depending on the spatial direction. The combination of different discretization methods allows to exploit special features of the solution of option pricing problems.

The methods introduced and discussed in Sections 2 and 3, are applied to basket options under the Black-Scholes model and to stochastic volatility models in Section 4.



# 2 Chapter 2

## Spatial Discretization

In this chapter we introduce the spatial discretization of the spatial operator  $L$

$$\begin{aligned}\frac{\partial u}{\partial t} &= Lu, & (\mathbf{x}, t) \in \Omega_d \times \Omega_t, \\ Lu &= \sum_{i,j=1}^d a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^d c_i \frac{\partial u}{\partial x_i} + bu,\end{aligned}$$

on a rectangular domain  $\Omega_d \times \Omega_t$  with suitable initial and boundary data. The spatial approximation of the PDE on the discrete grid  $\Omega_l$ , given in Definition 10, leads to a semi-discrete system of ordinary differential equations (ODEs) of the form

$$U'(t) = F(t)U(t), \quad t \geq 0, \quad (2.1)$$

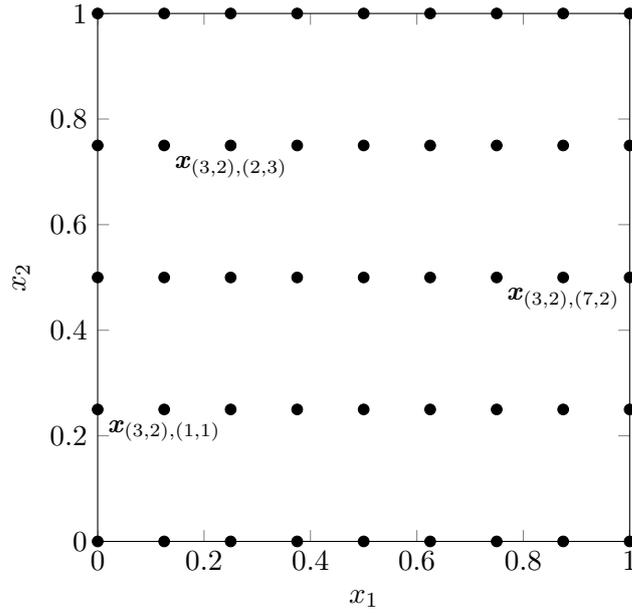
with initial value  $U(0) = U_0 \in \mathbb{R}^{(N_1+1) \cdot (N_2+1) \cdot \dots \cdot (N_d+1)}$  and discretization matrix  $F(t) \in \mathbb{R}^{(N_1+1) \cdot (N_2+1) \cdot \dots \cdot (N_d+1) \times (N_1+1) \cdot (N_2+1) \cdot \dots \cdot (N_d+1)}$  if  $L$  is discretized on the grid  $\Omega_l$ .

In the following we introduce several discretization techniques and state the resulting matrix  $F$ . First, we discuss the local approximation of derivatives with standard finite differences. The derivation is usually based on a truncation of the Taylor series. If a highly accurate approximation is desired, this generally leads to large finite difference stencils, resulting in a broadly banded matrix equation, which is expensive to solve. Furthermore, the discretization near the boundary might cause problems as it leads to ghost-points outside of the computational domain. In order to circumvent this problem to some extent, we employ HOC finite differences in Section 2.2. HOC schemes exploit the structure of the governing PDE to define a fourth-order discretization on the compact stencil. In Section 2.3 we consider a global approximation approach in form of pseudo-spectral methods. These methods have a geometric error decay if the solution fulfills certain regularity requirements. In the last part of this chapter we introduce sparse grids and the combination technique to reduce the number of grid nodes compared to a tensor based grid.

**Definition 10.** *Discrete grid*

We consider a  $d$ -dimensional domain  $\Omega_d$  in a continuous setting, where  $\mathbf{x} \in \Omega_d$  is given by  $\mathbf{x} = (x_1, x_2, \dots, x_d)$  and  $x_i$  is the position in the  $i$ -th coordinate direction for  $i = 1, 2, \dots, d$ . With the help of the multi-indices

$$\begin{aligned}\mathbf{l} &= (l_1, l_2, \dots, l_d) \in \mathbb{N}_0^d, \\ \mathbf{j} &= (j_1, j_2, \dots, j_d) \in \mathbb{N}_0^d, \\ \mathbf{N} &= (N_1, N_2, \dots, N_d) = (2^{l_1}, 2^{l_2}, \dots, 2^{l_d}),\end{aligned}$$

Figure 2.1: Sample grid  $\Omega_{(3,2)}$ .

we can define a tensor based grid  $\Omega_l$  with grid nodes

$$\mathbf{x}_{l,j} = (x_{l_1,j_1}, x_{l_2,j_2}, \dots, x_{l_d,j_d}) \quad \text{for } j_i = 0, 1, \dots, N_i.$$

The value  $x_{l_i,j_i}$  denotes the position in the  $i$ -th coordinate of the  $j_i$ -th node.

Figure 4.1 exemplary shows the grid  $\Omega_{(3,2)}$  on  $[0, 1]^2$  with uniform grid spacing  $h_1 = 2^{-3}$  in the first coordinate direction  $x_1$  and  $h_2 = 2^{-2}$  in the second coordinate direction  $x_2$ . Thus, we have  $N_1 = 8$  and  $N_2 = 4$  or equivalently 9 grid points in the first coordinate direction, which are numbered from 0 to 8 and 5 grid points in the second coordinate direction, which are numbered from 0 to 5. For example the grid node  $\mathbf{x}_{(3,2),(2,3)}$  is the third node in  $x_1$ - and the fourth node in  $x_2$ -direction and has the coordinates  $(0.25, 0.75) = (2 \cdot 2^{-3}, 3 \cdot 2^{-2}) = (j_1 \cdot 2^{-l_1}, j_2 \cdot 2^{-l_2})$  due to the uniform grid spacing.

In order to measure the accuracy of our numerical schemes, we introduce the following error norms in Definition 11. In a tensor based grid, such as  $\Omega_l$ , the higher-dimensional approximation of the derivatives can be computed by sequential numerical differentiation along each coordinate direction. This procedure can be written in a compact form with help of the Kronecker product notation in Definition 12 and will be used throughout this thesis.

**Definition 11.** *Error norms*

We consider the vector space  $\mathbb{R}^m$  with  $\hat{h} = 1/m$  and discrete norms  $\|v\|_2 = (\hat{h} \sum_{i=1}^m |v_i|^2)^{1/2}$ ,  $\|v\|_\infty = \max_{1 \leq i \leq m} |v_i|$  for  $v \in \mathbb{R}^m$ . We define the errors

$$err_2 = \|U_{ref} - U_{approx}\|_2, \quad (2.2)$$

$$err_2^{relative} = \frac{\|U_{ref} - U_{approx}\|_2}{\|U_{approx}\|_2}, \quad (2.3)$$

$$err_\infty = \|U_{ref} - U_{approx}\|_\infty, \quad (2.4)$$

where  $U_{ref}$  denotes a reference solution and  $U_{approx}$  its numerical approximation.

**Definition 12.** *Kronecker product*

The Kronecker product of the two matrices  $A$  of size  $k_1 \times l_1$  and  $B$  of size  $k_2 \times l_2$  is given by

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1l_1}B \\ \cdots & & \cdots \\ a_{k_11}B & \cdots & a_{k_1l_1}B \end{pmatrix}.$$

## 2.1 Standard Finite Difference Methods

In order to approximate the spatial derivatives one can use finite differences. Without loss of generality (w.l.o.g.) we restrict ourselves to the unit hypercube  $\Omega_d = [0, 1]^d$ . We consider a uniform grid spacing with mesh widths  $\mathbf{h} = (h_1, h_2, \dots, h_d) = (2^{-l_1}, 2^{-l_2}, \dots, 2^{-l_d})$ . Then we have  $x_{l_i, j_i} = j_i \cdot h_i$  for  $j_i = 0, 1, \dots, 2^{l_i}$ . With Taylor expansions under the assumption that  $u$  is sufficiently smooth, a second order approximation to the first and second derivative with respect to direction  $i$  at grid node  $\mathbf{x}_{l, j}$  is given by

$$\begin{aligned} \delta_i^0 u(\mathbf{x}_{l, j}) &= \frac{1}{2h_i} \left( u(\mathbf{x}_{l, j} + h_i e_i) - u(\mathbf{x}_{l, j} - h_i e_i) \right) = \frac{\partial u}{\partial x_i}(\mathbf{x}_{l, j}) + \mathcal{O}(h_i^2), \\ \delta_i^2 u(\mathbf{x}_{l, j}) &= \frac{1}{h_i^2} \left( u(\mathbf{x}_{l, j} + h_i e_i) - 2u(\mathbf{x}_{l, j}) + u(\mathbf{x}_{l, j} - h_i e_i) \right) = \frac{\partial^2 u}{\partial x_i^2}(\mathbf{x}_{l, j}) + \mathcal{O}(h_i^2), \end{aligned}$$

where  $e_i$  denotes the  $i$ -th unit vector. The mixed derivative can be approximated via a sequential application of the one-dimensional operators

$$\begin{aligned} \delta_i^0 \delta_j^0 u(\mathbf{x}_{l, j}) &= \frac{1}{4h_i h_j} \left( u(\mathbf{x}_{l, j} + h_i e_i + h_j e_j) - u(\mathbf{x}_{l, j} - h_i e_i + h_j e_j) \right. \\ &\quad \left. - u(\mathbf{x}_{l, j} + h_i e_i - h_j e_j) + u(\mathbf{x}_{l, j} - h_i e_i - h_j e_j) \right) \\ &= \frac{\partial^2 u}{\partial x_i \partial x_j}(\mathbf{x}_{l, j}) + \mathcal{O}(h_i^2) + \mathcal{O}(h_j^2) + \mathcal{O}(h_i^2 h_j^2), \end{aligned}$$

for  $i \neq j$  and  $i, j = 1, 2, \dots, d$ . In addition to these central finite differences upwind discretizations also are commonly used in practice. They rely on an approximation in direction of the propagation of information: especially in the case of convection equations or convection-dominated parabolic equations this is an important feature as it avoids spurious oscillations in the solution or its derivatives. For a more detailed discussion covering dispersion and the dissipative behavior of finite difference schemes, we refer to [83, 92]. If a higher accuracy is desired, more nodes can be added to the stencil to derive fourth-order accuracy. For example, to obtain a fourth-order accurate approximation of the first derivative in the  $i$ -th coordinate direction, we employ the ansatz

$$\begin{aligned} h_i \frac{\partial u}{\partial x_i}(\mathbf{x}_{l, j}) &= \alpha_{-2} u(\mathbf{x}_{l, j} - 2h_i e_i) + \alpha_{-1} u(\mathbf{x}_{l, j} - h_i e_i) + \alpha_0 u(\mathbf{x}_{l, j}) \\ &\quad + \alpha_1 u(\mathbf{x}_{l, j} + h_i e_i) + \alpha_2 u(\mathbf{x}_{l, j} + 2h_i e_i) + \mathcal{O}(h_i^4). \end{aligned}$$

Inserting the Taylor series for each term, one observes that the coefficients must satisfy the following linear system of equations

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \\ -8 & -1 & 0 & 1 & 8 \\ 16 & 1 & 0 & 1 & 16 \end{pmatrix} \begin{pmatrix} \alpha_{-2} \\ \alpha_{-1} \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Hence, the approximation reads

$$\frac{\partial u}{\partial x_i}(\mathbf{x}_{l,j}) = \frac{1}{12h_i} \left( u(\mathbf{x}_{l,j} - 2h_i e_i) - 8u(\mathbf{x}_{l,j} - h_i e_i) + 8u(\mathbf{x}_{l,j} + h_i e_i) - u(\mathbf{x}_{l,j} + 2h_i e_i) \right) + \mathcal{O}(h_i^4).$$

In an analogue way a fourth-order approximation of the second derivative can be derived

$$\frac{\partial^2 u}{\partial x_i^2}(\mathbf{x}_{l,j}) = \frac{1}{12h_i^2} \left( -u(\mathbf{x}_{l,j} - 2h_i e_i) + 16u(\mathbf{x}_{l,j} - h_i e_i) - 30u(\mathbf{x}_{l,j}) + 16u(\mathbf{x}_{l,j} + h_i e_i) - u(\mathbf{x}_{l,j} + 2h_i e_i) \right) + \mathcal{O}(h_i^4).$$

In order to streamline our notation, we introduce the following finite difference operators

$$\begin{aligned} \tilde{\delta}_{x_i}^0 u(\mathbf{x}_{l,j}) &= \frac{1}{12h_i} \left( u(\mathbf{x}_{l,j} - 2h_i e_i) - 8u(\mathbf{x}_{l,j} - h_i e_i) + 8u(\mathbf{x}_{l,j} + h_i e_i) - u(\mathbf{x}_{l,j} + 2h_i e_i) \right), \\ \tilde{\delta}_{x_i}^2 u(\mathbf{x}_{l,j}) &= \frac{1}{12h_i^2} \left( -u(\mathbf{x}_{l,j} - 2h_i e_i) + 16u(\mathbf{x}_{l,j} - h_i e_i) - 30u(\mathbf{x}_{l,j}) + 16u(\mathbf{x}_{l,j} + h_i e_i) - u(\mathbf{x}_{l,j} + 2h_i e_i) \right). \end{aligned}$$

The mixed derivative can be approximated similar to the second-order case via

$$\tilde{\delta}_{x_i}^0 \tilde{\delta}_{x_j}^0 u(\mathbf{x}_{l,j}) = \frac{\partial^2 u}{\partial x_i \partial x_j}(\mathbf{x}_{l,j}) + \mathcal{O}(h_i^4) + \mathcal{O}(h_j^4) + \mathcal{O}(h_i^4 h_j^4),$$

for  $i \neq j$  and  $i, j = 1, 2, \dots, d$ . Finite difference schemes with an accuracy higher than four are rarely used in practice: in financial engineering applications the non-smooth initial data deteriorates the theoretical accuracy since the truncation error is in general not bounded for non-smooth data. Furthermore, larger difference stencils lead to a higher computational effort as the resulting linear equation system does not have a tridiagonal structure anymore.

The finite difference operators act on a single grid node. Applying the difference operators to each grid node the approximation of the derivatives can be written in matrix notation. In the following we denote the matrix formulation of the first derivative and second derivative second-order finite difference matrix in direction  $i$  with  $D_{FD_i}$ ,  $D_{FD_i}^2$ ,

respectively, where

$$D_{FD_i} = \frac{1}{2h_i} \begin{pmatrix} 0 & 1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & -1 & 0 \end{pmatrix},$$

$$D_{FD_i}^2 = \frac{1}{h_i^2} \begin{pmatrix} -2 & 1 & & 0 \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & -2 \end{pmatrix}.$$

The fourth-order matrices are denoted by  $\tilde{D}_{FD_i}$ ,  $\tilde{D}_{FD_i}^2$  and are given by

$$\tilde{D}_{FD_i} = \frac{1}{12h_i} \begin{pmatrix} 0 & 8 & -1 & & & 0 \\ -8 & \ddots & \ddots & \ddots & & \\ 1 & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & -1 \\ & & & \ddots & \ddots & 8 \\ 0 & & & & 1 & -8 & 0 \end{pmatrix},$$

$$\tilde{D}_{FD_i}^2 = \frac{1}{12h_i^2} \begin{pmatrix} -30 & 16 & -1 & & & 0 \\ 16 & \ddots & \ddots & \ddots & & \\ -1 & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & -1 \\ & & & \ddots & \ddots & 16 \\ 0 & & & & -1 & 16 & -30 \end{pmatrix}.$$

## 2.2 High-Order-Compact Finite Difference Methods

In the following we derive a fourth-order discretization on the compact stencil, such that the resulting linear system of equations has a tridiagonal structure. In the last decades lots of effort has been spent on the derivation of HOC schemes: starting with the early work by Gupta et al. [36]. Further effort has been spent on the derivation of HOC schemes, e.g. in [21, 22, 24, 43, 58, 72, 88, 89] to mention a few examples. In the field of computational finance HOC methods have been proposed by [21, 22, 24].

We sketch the derivation of a HOC finite difference scheme for the two-dimensional diffusion equation with constant coefficients

$$\frac{\partial u}{\partial t} = a_{11} \frac{\partial^2 u}{\partial x_1^2} + a_{22} \frac{\partial^2 u}{\partial x_2^2} \quad (2.5)$$

on  $\Omega_2$ . Inserting the finite difference approximations we obtain

$$\frac{\partial u}{\partial t} = a_{11}\delta_1^2 u - a_{11}\frac{h_1^2}{12}\frac{\partial^4 u}{\partial x_1^4} + a_{22}\delta_2^2 - a_{22}\frac{h_2^2}{12}\frac{\partial^4 u}{\partial x_2^4} + \mathcal{O}(h_1^4) + \mathcal{O}(h_2^4). \quad (2.6)$$

Due to the squared mesh width in front of the leading error terms, we can obtain fourth-order consistency if the truncation error is approximated with second-order accuracy. Applying standard second-order finite differences to approximate the fourth derivatives  $\frac{\partial^4 u}{\partial x_1^4}$ ,  $\frac{\partial^4 u}{\partial x_2^4}$ , respectively, yields a fourth-order discretization. However, the discretization is not defined on the compact stencil anymore. In order to obtain a second-order accurate approximation on the compact stencil we derive auxiliary relations for the fourth derivative by differentiating equation (2.5) with respect to  $x_1$  twice and solve for the fourth derivative

$$\frac{\partial^4 u}{\partial x_1^4} = \frac{1}{a_{11}}\frac{\partial^3 u}{\partial x_1^2 \partial t} - \frac{a_{22}}{a_{11}}\frac{\partial^4 u}{\partial x_1^2 \partial x_2^2}.$$

A relation for the fourth derivative with respect to  $x_2$  can be obtained analogously

$$\frac{\partial^4 u}{\partial x_2^4} = \frac{1}{a_{22}}\frac{\partial^3 u}{\partial x_2^2 \partial t} - \frac{a_{11}}{a_{22}}\frac{\partial^4 u}{\partial x_1^2 \partial x_2^2}.$$

Inserting these relations into (2.6) and replacing the derivatives by their central finite difference counterpart, we obtain

$$\left(1 + \frac{h_1^2}{12}\delta_1^2 + \frac{h_2^2}{12}\delta_2^2\right)\frac{\partial u}{\partial t} = a_{11}\delta_1^2 u + a_{22}\delta_2^2 u + \frac{h_1^2}{12}a_{22}\delta_1^2\delta_2^2 u + \frac{h_2^2}{12}a_{11}\delta_1^2\delta_2^2 u + \mathcal{O}(h_1^2 h_2^2). \quad (2.7)$$

This spatial discretization is defined on the compact stencil and has a leading error term of order  $\mathcal{O}(h_1^2 h_2^2)$ . Thus, we can expect fourth-order accuracy if  $h_1 = ch_2$  holds for some constant  $c$ . Düring and Heuer [24] derived a HOC discretization for  $d$  dimensional convection-diffusion equations with mixed derivatives and space and time dependent coefficients. Beside the analogue condition that all mesh widths have to be of the same order, they have the additional constraints, that either no mixed derivative is present or  $\frac{h_i^2}{h_j^2} = \frac{a_{ii}}{a_{jj}}$  has to be fulfilled to obtain fourth-order consistency for  $i, j = 1, 2, \dots, d$ .

In the sequel we introduce a discretization which does not require these constraints. The derivation is based on the work of Düring et al. [23] for constant coefficient problems and has been extended for non-constant coefficients in [25, 45]. We consider unidirectional contributions

$$a_{ii}(\mathbf{x}_{l,j})\frac{\partial^2 u}{\partial x_i^2}(\mathbf{x}_{l,j}) + c_i(\mathbf{x}_{l,j})\frac{\partial u}{\partial x_i}(\mathbf{x}_{l,j}) = g(\mathbf{x}_{l,j}) \quad (2.8)$$

for  $i = 1, \dots, d$  and some arbitrary smooth right hand side  $g$ . In the following we streamline our notation and write  $a_{ii} := a_{ii}(\mathbf{x}_{l,j})$ ,  $c_i := c_i(\mathbf{x}_{l,j})$ . Inserting the finite difference

operators we obtain

$$\begin{aligned} & a_{ii}\delta_i^2 u(\mathbf{x}_{l,j}) - a_{ii} \frac{h_i^2}{12} \frac{\partial^4 u}{\partial x_i^4}(\mathbf{x}_{l,j}) - a_{ii} \frac{h_i^4}{360} \frac{\partial^6 u}{\partial x_i^6}(\mathbf{x}_{l,j}) + c_i \delta_i^0 u(\mathbf{x}_{l,j}) - c_i \frac{h_i^2}{6} \frac{\partial^3 u}{\partial x_i^3}(\mathbf{x}_{l,j}) \\ & - c_i \frac{h_i^4}{120} \frac{\partial^5 u}{\partial x_i^5}(\mathbf{x}_{l,j}) + \mathcal{O}(h_i^6) = g(\mathbf{x}_{l,j}). \end{aligned} \quad (2.9)$$

Since the leading error term in (2.9) is of order two, we can derive a fourth-order compact approximation if the third and fourth derivative is approximated with second order accuracy on the compact stencil. In order to derive these approximations, we differentiate equation (2.8) once with respect to  $x_i$  and thus get

$$\frac{\partial a_{ii}}{\partial x_i} \frac{\partial^2 u}{\partial x_i^2} + a_{ii} \frac{\partial^3 u}{\partial x_i^3} + \frac{\partial c_i}{\partial x_i} \frac{\partial u}{\partial x_i} + c_i \frac{\partial^2 u}{\partial x_i^2} = \frac{\partial g}{\partial x_i}.$$

Hence, the third derivative is given by the auxiliary equation

$$\frac{\partial^3 u}{\partial x_i^3} = \frac{1}{a_{ii}} \frac{\partial g}{\partial x_i} - \left( \frac{1}{a_{ii}} \frac{\partial a_{ii}}{\partial x_i} + \frac{c_i}{a_{ii}} \right) \frac{\partial^2 u}{\partial x_i^2} - \frac{1}{a_{ii}} \frac{\partial c_i}{\partial x_i} \frac{\partial u}{\partial x_i}. \quad (2.10)$$

In a similar fashion we obtain an expression for the fourth derivative by differentiating (2.8) twice with respect to  $x_i$

$$\begin{aligned} \frac{\partial^4 u}{\partial x_i^4} &= \frac{1}{a_{ii}} \frac{\partial^2 g}{\partial x_i^2} - \left( \frac{c_i}{a_{ii}^2} + \frac{2}{a_{ii}^2} \frac{\partial a_{ii}}{\partial x_i} \right) \frac{\partial g}{\partial x_i} + \left( \frac{c_i^2}{a_{ii}^2} + \frac{3c_i}{a_{ii}^2} \frac{\partial a_{ii}}{\partial x_i} + \frac{2}{a_{ii}^2} \left[ \frac{\partial a_{ii}}{\partial x_i} \right]^2 \right) \frac{\partial^2 u}{\partial x_i^2} - \frac{2}{a_{ii}} \frac{\partial c_i}{\partial x_i} \\ & - \frac{1}{a_{ii}} \frac{\partial^2 a_{ii}}{\partial x_i^2} \frac{\partial^2 u}{\partial x_i^2} + \left( \frac{c_i}{a_{ii}^2} \frac{\partial c_i}{\partial x_i} + \frac{2}{a_{ii}^2} \frac{\partial a_{ii}}{\partial x_i} \frac{\partial c_i}{\partial x_i} - \frac{1}{a_{ii}} \frac{\partial^2 c_i}{\partial x_i^2} \right) \frac{\partial u}{\partial x_i}. \end{aligned} \quad (2.11)$$

The third and fourth derivative can then be approximated with second-order stencils via central difference operators. Replacing the truncation error in (2.9) leads to a fourth-order accurate approximation

$$\begin{aligned} & \left( a_{ii} + \frac{h_i^2}{12} \frac{\partial^2 a_{ii}}{\partial x_i^2} - \frac{h_i^2 c_i}{12 a_{ii}} \frac{\partial a_{ii}}{\partial x_i} - \frac{h_i^2}{6 a_{ii}} \left[ \frac{\partial a_{ii}}{\partial x_i} \right]^2 + \frac{h_i^2 c_i^2}{12 a_{ii}} + \frac{h_i^2}{6} \frac{\partial c_i}{\partial x_i} \right) \delta_i^2 u(\mathbf{x}_{l,j}) \\ & + \left( c_i - \frac{h_i^2}{6 a_{ii}} \frac{\partial a_{ii}}{\partial x_i} \frac{\partial c_i}{\partial x_i} + \frac{h_i^2 c_i}{12 a_{ii}} \frac{\partial c_i}{\partial x_i} + \frac{h_i^2}{12} \frac{\partial^2 c_i}{\partial x_i^2} \right) \delta_i^0 u(\mathbf{x}_{l,j}) + h_i^4 \tau_i \\ & = g(\mathbf{x}_{l,j}) + \frac{h_i^2}{12} \delta_i^2 g(\mathbf{x}_{l,j}) + \left( \frac{h_i^2 c_i}{12 a_{ii}} - \frac{h_i^2}{6 a_{ii}} \frac{\partial a_{ii}}{\partial x_i} \right) \delta_i^0 g(\mathbf{x}_{l,j}) \end{aligned} \quad (2.12)$$

on the compact stencil with

$$\begin{aligned}
\tau_i = & \left( -\frac{1}{36a_{ii}} \frac{\partial^3 c_i}{\partial x_i^3} \frac{\partial a_{ii}}{\partial x_i} + \frac{c_i}{72a_{ii}} \frac{\partial^3 c_i}{\partial x_i^3} + \frac{1}{144} \frac{\partial^4 c_i}{\partial x_i^4} \right) \frac{\partial u}{\partial x_i}(\mathbf{x}_{l,j}) \\
& + \left( \frac{1}{144} \frac{\partial^4 a_{ii}}{\partial x_i^4} + \frac{c_i}{72a_{ii}} \frac{\partial^3 a_{ii}}{\partial x_i^3} - \frac{1}{12a_{ii}} \frac{\partial a_{ii}}{\partial x_i} \frac{\partial^2 c_i}{\partial x_i^2} - \frac{1}{36a_{ii}} \frac{\partial a_{ii}}{\partial x_i} \frac{\partial^3 a_{ii}}{\partial x_i^3} \right. \\
& \qquad \qquad \qquad \left. + \frac{c_i}{24a_{ii}} \frac{\partial^2 c_i}{\partial x_i^2} + \frac{1}{36} \frac{\partial^3 c_i}{\partial x_i^3} \right) \frac{\partial^2 u}{\partial x_i^2}(\mathbf{x}_{l,j}) \\
& + \left( \frac{1}{36} \frac{\partial^3 a_{ii}}{\partial x_i^3} + \frac{c_i}{24a_{ii}} \frac{\partial^2 a_{ii}}{\partial x_i^2} - \frac{1}{18a_{ii}} \frac{\partial a_{ii}}{\partial x_i} \frac{\partial c_i}{\partial x_i} - \frac{1}{12a_{ii}} \frac{\partial a_{ii}}{\partial x_i} \frac{\partial^2 a_{ii}}{\partial x_i^2} \right. \\
& \qquad \qquad \qquad \left. + \frac{c_i}{36a_{ii}} \frac{\partial c_i}{\partial x_i} + \frac{1}{36} \frac{\partial^2 c_i}{\partial x_i^2} \right) \frac{\partial^3 u}{\partial x_i^3}(\mathbf{x}_{l,j}) \\
& + \left( \frac{5}{144} \frac{\partial^2 a_{ii}}{\partial x_i^2} + \frac{c_i}{48a_{ii}} \frac{\partial a_{ii}}{\partial x_i} - \frac{1}{72a_{ii}} \left[ \frac{\partial a_{ii}}{\partial x_i} \right]^2 + \frac{c_i^2}{144a_{ii}} + \frac{1}{72} \frac{\partial c_i}{\partial x_i} \right) \frac{\partial^4 u}{\partial x_i^4}(\mathbf{x}_{l,j}) \\
& + \frac{1}{80} c_i \frac{\partial^5 u}{\partial x_i^5}(\mathbf{x}_{l,j}) + \frac{1}{240} a_{ii} \frac{\partial^6 u}{\partial x_i^6}(\mathbf{x}_{l,j}) + \mathcal{O}(h_i^2). \tag{2.13}
\end{aligned}$$

From the truncation error  $\tau_i$  we see that besides the solution  $u$ , the coefficient functions  $a_{ii}$  and  $c_i$  have to be sufficiently smooth as well, so that their fourth derivative is bounded. Rewriting this scheme in terms of matrices or symbolic operators gives

$$A_{x_i} U = B_{x_i} G, \tag{2.14}$$

where  $A_{x_i}$  corresponds to the left hand side of (2.12) and  $B_{x_i}$  to its right hand side. The matrices can be expressed with the help of the Kronecker product, see Definition 12,

$$\begin{aligned}
A_{x_i} = & \text{diag} \left( a_{ii}(X) + \frac{h_i^2}{12} \frac{\partial^2 a_{ii}}{\partial x_i^2}(X) - \frac{h_i^2 c_i(X)}{12a_{ii}(X)} \frac{\partial a_{ii}}{\partial x_i}(X) - \frac{h_i^2}{6a_{ii}(X)} \left[ \frac{\partial a_{ii}}{\partial x_i}(X) \right]^2 + \frac{h_i^2 c_i^2(X)}{12a_{ii}(X)} \right. \\
& \left. + \frac{h_i^2}{6} \frac{\partial c_i}{\partial x_i}(X) \right) \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \\
& + \text{diag} \left( c_i(X) - \frac{h_i^2}{6a_{ii}(X)} \frac{\partial a_{ii}}{\partial x_i}(X) \frac{\partial c_i}{\partial x_i}(X) + \frac{h_i^2 c_i(X)}{12a_{ii}(X)} \frac{\partial c_i}{\partial x_i}(X) + \frac{h_i^2}{12} \frac{\partial^2 c_i}{\partial x_i^2}(X) \right) \\
& \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i} \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d}, \\
B_{x_i} = & I_{N_1 \cdot N_2 \cdot \dots \cdot N_d} + \frac{h_i^2}{12} \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \\
& + \text{diag} \left( \frac{h_i^2 c_i(X)}{12a_{ii}(X)} - \frac{h_i^2}{6a_{ii}(X)} \frac{\partial a_{ii}}{\partial x_i}(X) \right) \\
& \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i} \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d},
\end{aligned}$$

where  $\text{diag}$  is a diagonal matrix and each operation in the  $\text{diag}$  operator is understood component-wise. The matrix  $X \in \mathbb{R}^{(N_1+1) \cdot (N_2+1) \cdot \dots \cdot (N_d+1) \times d}$  contains all discrete grid

nodes of  $\Omega_l$  and is given by

$$X = \left( \begin{aligned} & [x_{l_1,0}, x_{l_1,1}, \dots, x_{l_1,N_1}]^\top \otimes e_{N_2} \otimes \dots \otimes e_{N_d}, \\ & e_{N_1} \otimes [x_{l_2,0}, x_{l_2,1}, \dots, x_{l_2,N_2}]^\top \otimes e_{N_3} \otimes \dots \otimes e_{N_d}, \dots, \\ & e_{N_1} \otimes \dots \otimes e_{N_{d-1}} \otimes [x_{l_d,0}, x_{l_d,1}, \dots, x_{l_d,N_d}]^\top \end{aligned} \right),$$

with all-ones vectors  $e_{N_i} = (1, 1, \dots, 1)^\top$  of size  $N_i + 1$  for  $i = 1, 2, \dots, d$ . The solution vector  $U$  and the right-hand side  $G$  are of size  $(N_1 + 1) \cdot (N_2 + 1) \cdot \dots \cdot (N_d + 1)$ , while the identity matrices of size  $N_i + 1 \times N_i + 1$  are denoted by  $I_{N_i}$  for  $i = 1, 2, \dots, d$ .

The semi-discrete scheme can then be written as

$$U'(t) = F_0 U + B_{x_1}^{-1} A_{x_1} U + \dots + B_{x_d}^{-1} A_{x_d} U + \mathcal{O}(h_1^4) + \dots + \mathcal{O}(h_d^4) + \sum_{i,j} \mathcal{O}(h_i^4 h_j^4). \quad (2.15)$$

The mixed derivatives are approximated via

$$a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} = a_{ij} \tilde{\delta}_i^0 \tilde{\delta}_j^0 u + h_i^4 \tilde{\tau}_i + h_j^4 \tilde{\tau}_j + h_i^4 h_j^4 \tau_{i,j}$$

with

$$\tilde{\tau}_i = a_{ij} \frac{1}{30} \frac{\partial^6 u}{\partial x_i^5 \partial x_j}, \quad \tilde{\tau}_j = a_{ij} \frac{1}{30} \frac{\partial^6 u}{\partial x_i \partial x_j^5}, \quad \tau_{i,j} = -a_{ij} \frac{1}{900} \frac{\partial^{10} u}{\partial x_i^5 \partial x_j^5} \quad (2.16)$$

for  $i \neq j$ ,  $i, j = 1, \dots, d$  and collected in the matrix  $F_0$ . Since the treatment of the mixed derivative requires to use a broad stencil, the spatial approximation is not defined on a compact stencil anymore. However, we will see in Section 3 that this is no drawback if ADI time stepping methods are used.

## 2.3 Pseudo-Spectral Methods

Contrary to the local approach of the approximation via finite differences, we discuss a global ansatz to compute the derivatives. For simplicity we restrict ourselves to the one-dimensional case in direction  $i$  for  $i \in \{1, 2, \dots, d\}$  on the discrete grid  $\Omega_l$ . The general higher-dimensional differentiation can be performed by sequential differentiation along each coordinate direction in a tensor based grid. The global approach can be summarized in two steps: first an interpolant of the data  $u_{l_i, j_i}$  at grid nodes  $x_{l_i, j_i}$  for  $j_i = 0, 1, \dots, N_i$  is computed. In a second step, the interpolant is differentiated to obtain an estimate of the second derivative. Let the interpolant  $P_{N_i}$  be given in Lagrange form

$$P_{N_i} u(x_i) = \sum_{j_i=0}^{N_i} u_{l_i, j_i} L_{l_i, j_i}(x_i), \quad (2.17)$$

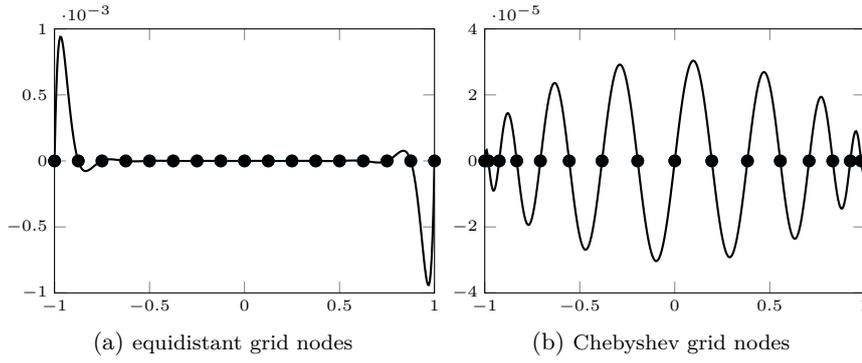


Figure 2.2: *Runge phenomenon*: Newton basis polynomial  $\prod_{i=0}^{16}(x - x_i)$  with equidistant nodes and Chebyshev nodes.

with Lagrange polynomials

$$L_{l_i, j_o}(x_i) = \prod_{\substack{j_i=0 \\ j_i \neq j_o}}^{N_i} \frac{x_i - x_{l_i, j_i}}{x_{l_i, j_o} - x_{l_i, j_i}}. \quad (2.18)$$

The Lagrange polynomials fulfill  $L_{l_i, j_o}(x_{l_i, j_i}) = \delta_{j_i j_o}$ , such that  $P_{N_i} u(x_{l_i, j_i}) = u_{l_i, j_i}$ . Then we can differentiate  $P_{N_i} u$  with respect to  $x_i$  to approximate the first derivative

$$(P_{N_i} u(x_i))' = \sum_{j_i=0}^{N_i} u_{l_i, j_i} L'_{l_i, j_i}(x_i). \quad (2.19)$$

Approximations of the higher derivatives can be derived in a similar fashion by sequential differentiation. Thus, for the  $p$ -th derivative we have

$$(P_{N_i} u(x_i))^{(p)} = \sum_{j_i=0}^{N_i} u_{l_i, j_i} L_{l_i, j_i}^{(p)}(x_i). \quad (2.20)$$

The derivative can be written in matrix form with  $(D_{SP_i}^p)_{j_i j_o} = L_{l_i, j_o}^{(p)}(x_{l_i, j_i})$  for  $j_i, j_o = 0, 1, \dots, N_i$ . Until now, we have left open the question of which grid points to use. The intuitive choice to employ an equidistant grid spacing leads to a strong oscillation at the boundaries. It can be shown that there exist functions, so that the interpolant diverges for  $N_i \rightarrow \infty$ . This is known as the *Runge phenomenon*. If the grid nodes are distributed according to the Chebyshev density  $\rho(x_i) = 1/(\pi\sqrt{1-x_i^2})$ , the oscillations at the boundaries are removed and the interpolant is of comparable size on the complete domain. In Figure 2.2 we demonstrate the effects of uniform and Chebyshev grid spacing. In the equidistant case we observe strong oscillations at the boundaries, while we have a similar magnitude in the Chebyshev case. Please also note the different scaling of the  $y$ -axis. For a detailed discussion of the *Runge phenomenon* we refer to [93].

In the following we restrict ourselves to Chebyshev-Gauss-Lobatto nodes  $x_{l_i, j_i} = \cos \frac{\pi j_i}{N_i}$  for  $j_i = 0, 1, \dots, N_i$  and  $i \in \{1, 2, \dots, d\}$ . Please note that the first and last node lie directly on the boundary of the domain,  $x_{l_i, 0} = 1$ ,  $x_{l_i, N_i} = -1$ , which is especially useful for the construction of numerical solvers for (initial) boundary value problems (IBVPs).

The function  $u$  can be expanded in a Chebyshev series

$$u(x_i) = \sum_{k=0}^{\infty} \hat{u}_k T_k(x_i), \quad (2.21)$$

with

$$\hat{u}_k = \frac{2}{\pi c_k} \int_{-1}^1 u(x_i) T_k(x_i) w(x_i) dx_i,$$

where

$$c_k = \begin{cases} 2, & k = 0, \\ 1, & k \geq 1. \end{cases}$$

and  $w(x_i) = 1/\sqrt{1-x_i^2}$ ,  $T_k(x_i) = \cos(k \cdot \arccos(x_i))$ . For computational purposes the series is truncated, such that we have

$$P_{N_i} u(x_i) = \sum_{k=0}^{N_i} \tilde{u}_k T_k(x_i), \quad (2.22)$$

with

$$\tilde{u}_k = \frac{2}{N_i \bar{c}_k} \sum_{j_i=0}^{N_i} \frac{1}{\bar{c}_{j_i}} u_{l_i, j_i} \cos \frac{\pi j_i k}{N_i} \quad \text{for } k = 0, 1, \dots, N_i,$$

and

$$\bar{c}_k = \begin{cases} 2, & k = 0, N, \\ 1, & k = 1, 2, \dots, N-1. \end{cases}$$

The accuracy of the Chebyshev interpolation depends on the regularity of the solution and is stated in Theorems 13 and 14. The convergence in Theorem 13 is what we call *algebraic* convergence with order  $m$ , whereas the convergence in Theorem 14 is known as *geometric* convergence. In order to make this more vivid, we consider the two test cases

$$u_1(x) = |x|^3, \quad u_2(x) = 1/(1+4x^2)$$

on  $\Omega_1 = [-1, 1]$  and compute the rate of convergence. The first function has a third derivative of bounded variation. Thus, we expect an algebraic convergence of third order. The second function is analytic, but has a convergence-limiting singularity at  $z_0 = x_0 + iy_0$  with  $x_0 = 0$  and  $y_0 = \pm \frac{1}{2}$ . The ellipse with foci  $\pm 1$  intersecting  $z_0$  has semimajor and semiminor axis lengths  $a = \sqrt{1 + \frac{1}{4}}$ ,  $b = \frac{1}{2}$  and thus we have  $r = a + b = \sqrt{1 + \frac{1}{4}} + \frac{1}{2} \approx 1.6180$ .

Figure 2.3 validates the theoretical considerations in a numerical experiment, where we compare the interpolant using  $N+1$  nodes to  $u_1$  and  $u_2$ , respectively. In Figure 2.3 (a) we observe an algebraic convergence of order 3.0045, which is close to the theoretical rate. In (b) we have a geometric convergence with factor 1.6173 in comparison to a theoretical one of 1.6180.

**Theorem 13.** [2] Let  $u, u', \dots, u^{(m-1)}$  be absolutely continuous for some  $m \geq 1$ , and

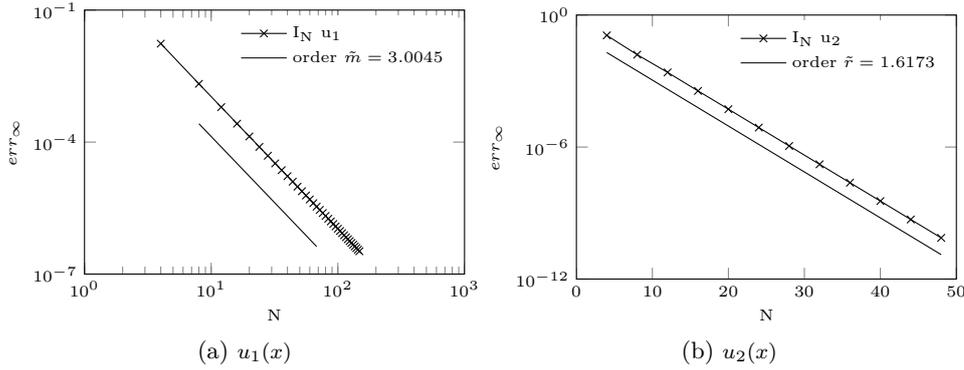


Figure 2.3: Maximal absolute interpolation error of  $u_1$ ,  $u_2$ , respectively. The straight lines indicate the experimental convergence factors  $\tilde{m}$  and  $\tilde{r}$  in Theorems 13 and 14.

let  $u^{(m)}$  be a function of bounded variation. Then

$$|u(x) - (P_N u)(x)| = \mathcal{O}(N^{-m})$$

as  $N_i \rightarrow \infty$  for all  $x \in [-1, 1]$ .

**Theorem 14.** [2] If  $u$  is analytic and bounded in the *Bernstein ellipse* of foci  $\pm 1$  with semimajor axis length  $a$  and semiminor axis lengths  $b = \sqrt{a^2 - 1}$  summing to  $r = a + b = a + \sqrt{a^2 - 1}$ , then the Chebyshev interpolant with  $N + 1$  Chebyshev-Gauss-Lobatto nodes fulfills

$$|u(x) - (P_N u)(x)| = \mathcal{O}(r^{-N})$$

as  $N \rightarrow \infty$  for all  $x \in [-1, 1]$ .

The derivative of the truncated Chebyshev series can either be computed in the physical space, e.g. by matrix vector multiplication, or in the transformed space. The matrix vector approach takes  $\mathcal{O}(N_i^2)$  operations as the Chebyshev derivative matrix is full, see Lemma 15. The differentiation in transformed space can be performed in linear run-time. Therefore, we consider the derivative of the Chebyshev series

$$u'(x_i) = \sum_{k=0}^{\infty} \hat{u}_k^{(1)} T_k'(x_i).$$

Exploiting the following recurrence relation of the Chebyshev polynomials to its derivatives

$$\begin{cases} T_0(x_i) = T_1'(x_i), \\ T_1(x_i) = \frac{1}{4} T_2'(x_i), \\ T_k(x_i) = \frac{1}{2} \left( \frac{1}{k+1} T_{k+1}'(x_i) - \frac{1}{k-1} T_{k-1}'(x_i) \right), \quad k \geq 2, \end{cases}$$

with  $T_0 = 1$ ,  $T_1 = x_i$ , and the alternative form of the derivative

$$u'(x_i) = \sum_{k=0}^{\infty} \hat{u}_k T_k'(x_i),$$

one observes

$$2k\hat{u}_k = c_{k-1}\hat{u}_{k-1}^{(1)} - \hat{u}_{k+1}^{(1)} \quad \text{for } k \geq 1, \quad (2.23)$$

holds. Thus, the derivative of the truncated series (2.22)

$$(P_{N_i}u(x_i))' = \sum_{k=0}^{N_i} \tilde{u}_k^{(1)} T_k(x_i)$$

can be computed in  $\mathcal{O}(N_i)$  operations via

$$\tilde{u}_k^{(1)} = \frac{1}{c_k} \left( 2(k+1)\tilde{u}_{k+1} + \tilde{u}_{k+2}^{(1)} \right) \quad \text{for } k = N_i - 1, N_i - 2, \dots, 0$$

with  $\tilde{u}_k^{(1)} = 0$  for  $k \geq N_i$ . A similar relation also holds for the higher derivatives and the generalization for the  $p$ -th derivative reads

$$\tilde{u}_k^{(p)} = \frac{1}{c_k} \left( 2(k+1)\tilde{u}_{k+1}^{(p-1)} + \tilde{u}_{k+2}^{(p)} \right) \quad \text{for } k = N_i - 1, N_i - 2, \dots, 0.$$

The mapping between physical  $U = (u_{i,0}, u_{i,1}, \dots, u_{i,N_i})^\top$  and transformed space  $\tilde{U} = (\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_N)^\top$  is given by  $\tilde{U} = CU$  with

$$(C)_{j_i j_o} = \frac{2}{N_i \bar{c}_{j_i} \bar{c}_{j_o}} \cos \frac{\pi j_i j_o}{N_i} \quad \text{for } j_i, j_o = 0, 1, \dots, N_i.$$

The backward transformation is given by  $C^{-1}\tilde{U} = U$  with

$$(C^{-1})_{j_i j_o} = \cos \frac{\pi j_i j_o}{N_i} \quad \text{for } j_i, j_o = 0, 1, \dots, N_i.$$

Using these mappings in form of a matrix vector multiplication results in a quadratic computational effort and is therefore expensive. However, it can be computed efficiently with the *Fast Fourier Transformation* (FFT) algorithm in  $\mathcal{O}(N_i \log_2 N_i)$  operations. Exploiting that  $U$  is a vector of real values, the effort to compute Chebyshev transform can be halved within the complex FFT algorithm. For a detailed discussion we refer to [11, 61].

**Lemma 15.** [11] *The Chebyshev differentiation matrices with the Chebyshev grid nodes  $x_{l_i, j_i} = \cos \frac{\pi j_i}{N_i}$  for  $j_i = 0, 1, \dots, N_i$  can be given in closed form.*

$$(D_{SP_i})_{j_i j_o} = \begin{cases} -\frac{\bar{c}_{j_i}}{2\bar{c}_{j_o}} \frac{(-1)^{j_i+j_o}}{\sin((j_i+j_o)\pi/2N_i) \sin((j_i-j_o)\pi/2N_i)}, & j_i \neq j_o, \\ -\frac{x_{j_i}}{2\sin^2(j_i\pi/N_i)}, & 1 \leq j_i = j_o \leq N_i - 1, \\ \frac{2N_i^2+1}{6}, & j_i = j_o = 0, \\ -\frac{2N_i^2+1}{6}, & j_i = j_o = N_i \end{cases} \quad (2.24)$$

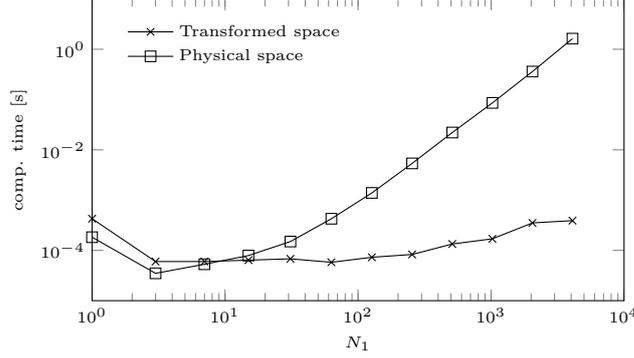


Figure 2.4: Computation time of the first derivative in 1d.

and

$$(D_{SP_i}^2)_{j_i j_o} = \begin{cases} \frac{(-1)^{j_i+j_o}}{\bar{c}_{j_o}} \frac{x_{j_i}^2 + x_{j_i} x_{j_o} - 2}{(1-x_{j_i}^2)(x_{j_i} - x_{j_o})^2}, & 1 \leq j_i \leq N-1, 0 \leq j_o \leq N_i, j_i \neq l, \\ -\frac{(N_i^2-1)(1-x_{j_i}^2)+3}{3(1-x_{j_i}^2)^2}, & 1 \leq j_i = j_o \leq N_i - 1, \\ \frac{2}{3} \frac{(-1)^{j_o}}{\bar{c}_i} \frac{(2N_i^2+1)(1-x_{j_o})-6}{(1-x_{j_o})^2}, & j_i = 0, 1 \leq j_o \leq N_i, \\ \frac{2}{3} \frac{(-1)^{(j_o+N_i)}}{\bar{c}_{j_o}} \frac{(2N_i^2+1)(1-x_{j_o})-6}{(1+x_{j_o})^2}, & j_i = N_i, 0 \leq j_o \leq N_i - 1, \\ \frac{N_i^4-1}{15}, & j_i = j_o = 0, j_i = j_o = N_i, \end{cases} \quad (2.25)$$

with

$$\bar{c}_{j_i} = \begin{cases} 2, & j_i = 0, N_i, \\ 1, & j_i = 1, 2, \dots, N_i - 1. \end{cases}$$

In Figure 2.4 we compare the computation time of both approaches: the differentiation in physical space corresponds to a matrix vector multiplication with quadratic effort, while the computation of the derivative in transformed space consists of the forward transform, the differentiation in Chebyshev space and a backward transformation. The transforms can be performed with complexity  $\mathcal{O}(N_i \log_2 N_i)$  and the differentiation with  $\mathcal{O}(N_i)$  operations. Thus, the total effort consists of  $\mathcal{O}(N_i \log_2 N_i)$  operations. In the numerical experiment the approach to compute the derivative in Chebyshev transformed space outperforms the matrix approach for  $N_i > 10$ . For small  $N_i$  it might be reasonable to apply the matrix approach, as here the overhead costs of the FFT algorithm dominate the overall run-time.

## 2.4 The Curse of Dimensionality and the Sparse Grid Combination Technique

Solving high dimensional equations numerically on a full tensor based grid with  $\mathcal{O}(N^d)$  grid points is an extensive work. Although, splitting schemes, such as ADI schemes, can reduce the computational workload significantly, there is a limit on the number of nodes due to the finite available memory. In the following we use so-called *sparse grids* to

reduce the impact of the curse of dimensionality. The approach was originally developed by Smolyak [86] for numerical integration. Zenger [98], Bungartz and Griebel [8] and Schiekofe [82] extended this idea and applied sparse grids to solve PDEs with finite element, finite volume and finite difference methods.

Schiekofe [82] introduced the framework to apply finite difference techniques to sparse grids. Since ordinary finite difference discretizations only act on a tensor based full grid in nodal basis, a special treatment is needed to design finite difference stencils for sparse grids. He proposes the use of a basis transform from nodal to hierarchical basis. The forward and backward basis transform leads to densely populated discretization matrices which are costly to solve. Further this approach in general requires hierarchical, tree-like data structures, which makes the data structure management more complicated than in the full grid case. These drawbacks can be circumvented with the help of the *sparse grid combination technique* [33]. Here, solutions on a tensor based grid are linearly combined to construct the sparse grid solution. Hence, standard full grid solvers can be used, which enables to compute sparse grid solutions with minimal additional effort in terms of implementation time. Since all sub-problems are independent the method can easily be parallelized and run on a cluster of computers.

In the following, we give a brief introduction to sparse grids and the combination technique. The introduction is based on the works [78, 79]. The method is based on the error splitting structure of the underlying numerical scheme. To make the basic idea clear, we consider a two-dimensional problem on the unit square  $\Omega_2 = [0, 1]^2$  and assume an numerical approximation  $u_{\mathbf{l}}$  on  $\Omega_{\mathbf{l}}$  with  $\mathbf{l} = (l_1, l_2) \in \mathbb{N}_0^2$ , with mesh widths  $\mathbf{h} = (h_1, h_2) = (2^{-l_1}, 2^{-l_2})$ , which satisfies an *error splitting* of the form

$$u - u_{\mathbf{l}} = h_1^2 w_1(h_1) + h_2^2 w_2(h_2) + h_1^2 h_2^2 w_{1,2}(h_1, h_2). \quad (2.26)$$

The crucial point is that the mesh sizes  $h_1, h_2$  are independent of one another and  $w_1$  only depends on the mesh width in the first coordinate direction, while  $w_2$  only depends on  $h_2$ . Furthermore, the functions  $w_1, w_2, w_{1,2}$  are assumed to be bounded. This structure can now be exploited by combining them in such a way that low order terms cancel out. Therefore, we introduce the *hierarchical surplus* of the numerical solution as

$$\delta(u_{\mathbf{l}}) = u_{\mathbf{l}} - u_{\mathbf{l}-e_1} - u_{\mathbf{l}-e_2} + u_{\mathbf{l}-e_1-e_2},$$

where  $e_1 = (1, 0)$  and  $e_2 = (0, 1)$ . Inserting the error splitting (2.26), we obtain

$$\begin{aligned} \delta(u - u_{\mathbf{l}}) &= h_1^2 w_1(h_1) + h_2^2 w_2(h_2) + h_1^2 h_2^2 w_{1,2}(h_1, h_2) \\ &\quad - 4 h_1^2 w_1(2h_1) - h_2^2 w_2(h_2) - 4 h_1^2 h_2^2 w_{1,2}(2h_1, h_2) \\ &\quad - h_1^2 w_1(h_1) - 4 h_2^2 w_2(2h_2) - 4 h_1^2 h_2^2 w_{1,2}(h_1, 2h_2) \\ &\quad + 4 h_1^2 w_1(2h_1) + 4 h_2^2 w_2(2h_2) + 16 h_1^2 h_2^2 w_{1,2}(2h_1, 2h_2) \\ &= h_1^2 h_2^2 w_{1,2}(h_1, h_2) - 4 h_1^2 h_2^2 w_{1,2}(2h_1, h_2) - 4 h_1^2 h_2^2 w_{1,2}(h_1, 2h_2) \\ &\quad + 16 h_1^2 h_2^2 w_{1,2}(2h_1, 2h_2) \\ &= \mathcal{O}(h_1^2 h_2^2) = \mathcal{O}(2^{-2l_1} 2^{-2l_2}) = \mathcal{O}(2^{-2|\mathbf{l}|_1}) \end{aligned}$$

The surplus can be interpreted as the information gain of the solution  $u_{\mathbf{l}}$ . Let the level of a numerical solution on the discrete grid  $\Omega_{\mathbf{l}}$  be given by  $|\mathbf{l}|_1$ , then we see that the solutions with the same level (with the same number of grid nodes) have the same surplus. This

motivates to combine all solutions with a high surplus or in other words, with a high information gain. In the following we define the combined sparse grid solution as the sum of all surpluses with  $|\mathbf{l}|_1 \leq n$  for  $n \in \mathbb{N}_0$

$$u_n^s = \sum_{|\mathbf{l}|_1 \leq n} \delta u_{\mathbf{l}}.$$

An upper error bound can be found by incorporating the surpluses of all sub-solutions, which are not used to compute  $u_n^s$ . We have

$$\|u_n^s - u\| \leq \sum_{|\mathbf{l}|_1 > n} \|\delta u_{\mathbf{l}}\| = \sum_{|\mathbf{l}|_1 > n} \mathcal{O}(2^{-2|\mathbf{l}|_1}) = \sum_{i > n} \mathcal{O}((i+1)2^{-2i}) = \mathcal{O}(n2^{-2n}).$$

The second equality holds since at each level  $l_1 + l_2 = i$  there are  $i + 1$  surpluses with an accuracy of  $\mathcal{O}(2^{-2i})$  involved. The latter equality follows by taking into account that the summands are a geometric series. Let  $h = 2^{-n}$ , we can rewrite the error bound to

$$\|u_n^s - u\| \leq \mathcal{O}(h^2 \log_2(h^{-1})).$$

For computational purposes one does not want to compute each single surplus, as this results in a large amount of unnecessary computations. To avoid this, we take a closer look at the used sub-solutions within the combined sparse grid solution

$$\begin{aligned} u_n^s &= \sum_{|\mathbf{l}|_1 \leq n} \delta u_{\mathbf{l}} = \sum_{q=0}^n \left( \sum_{|\mathbf{l}|_1=q} u_{\mathbf{l}} - 2 \sum_{\substack{|\mathbf{l}|_1=q-1 \\ q-1 \geq 0}} u_{\mathbf{l}} + \sum_{\substack{|\mathbf{l}|_1=q-2 \\ q-2 \geq 0}} u_{\mathbf{l}} \right) \\ &= \sum_{|\mathbf{l}|_1=n} u_{\mathbf{l}} - \sum_{|\mathbf{l}|_1=n-1} u_{\mathbf{l}}. \end{aligned}$$

It becomes clear that only the solutions at level  $n$  and  $n - 1$  are actually needed and all other sub-solutions cancel out. Please note that this is the standard formulation of the combination technique for  $d = 2$ , e.g. see [9].

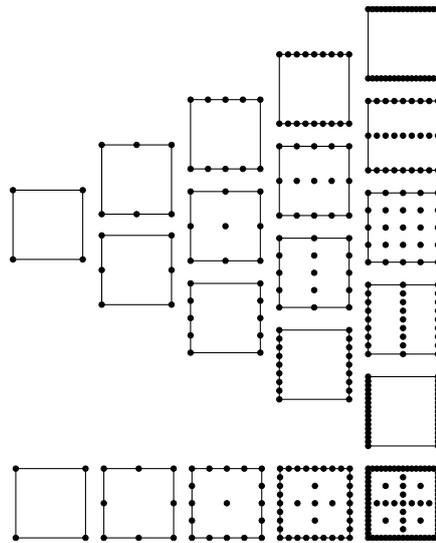


Figure 2.5: Sub-grids and sparse grid for  $n = 0, \dots, 4$ .

Figure 2.5 shows the two-dimensional grid hierarchy for level  $n = 0, \dots, 4$ . The sparse grid at level  $n$  consists of sub-grids, whose sum of refinement levels is equal to  $n$ . Hence the number of grid points on each sub-grid grows with  $\mathcal{O}(2^n)$ . As the number of grids increases with  $\mathcal{O}(n)$ , this leads to  $\mathcal{O}(n2^n)$  nodes in the sparse grid. Let  $h = 2^{-n}$ , then this results into  $\mathcal{O}(h^{-1} \log_2(h^{-1}))$  grid points compared to  $\mathcal{O}(h^{-2})$  nodes in the full grid.

The derivation of the combination technique in the general  $d$ -dimensional case for schemes with algebraic accuracy  $m$  follows the same steps. We consider the unit cube  $\Omega_d = [0, 1]^d$  and a family of Cartesian grids  $\Omega_{\mathbf{l}}$  with multi-index  $\mathbf{l} = (l_1, l_2, \dots, l_d) \in \mathbb{N}_0^d$  and grid spacing  $\mathbf{h} = (h_1, h_2, \dots, h_d) = (2^{-l_1}, 2^{-l_2}, \dots, 2^{-l_d})$ . The hierarchical surplus of  $u_{\mathbf{l}}$  is then given by

$$\delta u_{\mathbf{l}} = \delta_1 \delta_2 \cdots \delta_d u_{\mathbf{l}}, \quad (2.27)$$

with

$$\delta_i u_{\mathbf{l}} = \begin{cases} u_{\mathbf{l}} - u_{\mathbf{l}-\mathbf{e}_i}, & l_i > 0 \\ u_{\mathbf{l}}, & l_i = 0, \end{cases}$$

where  $\mathbf{e}_i$  is the  $i$ -th unit vector. Under the assumption of an error splitting

$$u - u_{\mathbf{l}} = \sum_{k=1}^d \sum_{\substack{\{j_1, \dots, j_k\} \\ \subseteq \{1, \dots, d\}}} w_{j_1, \dots, j_k}(\cdot; h_{j_1}, \dots, h_{j_k}) h_{j_1}^m \cdots h_{j_k}^m, \quad (2.28)$$

it holds

$$\delta u_{\mathbf{l}} = \mathcal{O}(h_1^m h_2^m \cdots h_d^m) = \mathcal{O}(2^{-m \|\mathbf{l}\|_1}).$$

As there are  $N = \binom{n+d-1}{d-1}$  possibilities to decompose a number  $n \in \mathbb{N}_0$  in  $d$  single non-negative summands, there are  $N = \mathcal{O}(n^{d-1})$  grids involved in the construction of the sparse grid solution. Each grid consists of  $\mathcal{O}(2^n)$  nodes, which leads to a total number of nodes  $\mathcal{O}(n^{d-1} 2^n) = \mathcal{O}(h^{-1} \log_2(h^{-1})^{d-1})$  in contrast to  $\mathcal{O}(h^{-d})$  in the full grid case for  $h = 2^{-n}$ .

In order to compute the accuracy, we proceed like in the two-dimensional case, such that we have

$$\|u - u_n^s\| \leq \sum_{\mathbf{l} \notin M_n} \|\delta u_{\mathbf{l}}\| = \sum_{i > n} \mathcal{O}(i^{d-1} 2^{-m \cdot i}) = \mathcal{O}(n^{d-1} 2^{-m \cdot n}).$$

The latter equality can be shown by taking into account that the summands form a

geometric series of the form

$$\begin{aligned}
\sum_{i>n} i^{d-1} 2^{-m \cdot i} &= \sum_{i=0}^{\infty} i^{d-1} 2^{-m \cdot i} - \sum_{i=0}^n i^{d-1} 2^{-m \cdot i} \\
&= \sum_{i=0}^{\infty} i^k q^i - \sum_{i=0}^n i^k q^i \quad q := 2^{-m}, k := d-1 \\
&= \underbrace{q \frac{d}{dq} \dots q \frac{d}{dq}}_{k \text{ times}} \left( \frac{1}{1-q} - \frac{1-q^{n+1}}{1-q} \right) \\
&= \mathcal{O}((n+1)^k q^{n+1}) = \mathcal{O}(n^k q^n) = \mathcal{O}(n^{d-1} 2^{-m \cdot n}).
\end{aligned}$$

Let  $h = 2^{-n}$ , then the error reads

$$\|u - u_n^s\| \leq \mathcal{O}(n^{d-1} 2^{-m \cdot n}) = \mathcal{O}(h^{-m} \log_2(h^{-1})^d).$$

Analogue to the two-dimensional case we exploit the cancellation of the sub-solutions to derive the sparse grid combination technique formula

$$\begin{aligned}
u_n^s &= \sum_{|l_1| \leq n} \delta u_l = \sum_{q=0}^N \left( \sum_{|l_1|=q} u_l - \binom{d}{1} \sum_{\substack{|l_1|=q-1 \\ q-1 \geq 0}} u_l + \binom{d}{2} \sum_{\substack{|l_1|=q-2 \\ q-2 \geq 0}} u_l \right. \\
&\quad \left. + \dots + (-1)^{d-1} \binom{d}{d-1} \sum_{\substack{|l_1|=q-d+1 \\ q-d+1 \geq 0}} u_l + (-1)^d \sum_{\substack{|l_1|=q-d \\ q-d \geq 0}} u_l \right) \\
&= \sum_{|l_1|=n} u_l + (-d+1) \sum_{|l_1|=n-1} u_l + \left( \binom{d}{2} - d+1 \right) \sum_{|l_1|=n-2} u_l \\
&\quad + \dots + \left( (-1)^{d-1} \binom{d}{d-1} + (-1)^{d-1} \binom{d}{d-2} + \dots - d+1 \right) \sum_{|l_1|=n-d+1} u_l \\
&\quad + \left( (-1)^d + (-1)^{d-1} \binom{d}{d-1} + (-1)^{d-1} \binom{d}{d-2} + \dots - d+1 \right) \sum_{|l_1|=n-d} u_l \\
&= \sum_{|l_1|=n} u_l + \sum_{q=0}^1 (-1)^q \binom{d}{q} \sum_{|l_1|=n-1} u_l + \sum_{q=0}^2 (-1)^q \binom{d}{q} \sum_{|l_1|=n-1} u_l \\
&\quad + \dots + \sum_{q=0}^{d-1} (-1)^q \binom{d}{q} \sum_{|l_1|=n-d+1} u_l + \sum_{q=0}^d (-1)^q \binom{d}{q} \sum_{|l_1|=n-d} u_l \\
&= \sum_{q=0}^{d-1} \binom{d-1}{q} \sum_{|l_1|=n-q} u_l.
\end{aligned}$$

The fourth equality follows from

$$\sum_{q=0}^d (-1)^q \binom{d}{q} = 0,$$

and

$$\sum_{q=0}^k (-1)^q \binom{d}{q} = (-1)^k \binom{d-1}{k} \quad \text{for } k \leq d-1.$$

**Definition 16.** *Sparse grid combination technique*

The sparse grid combination formula at level  $n \in \mathbb{N}$  is given by

$$u_n^s = \sum_{q=0}^{d-1} \binom{d-1}{q} \sum_{|\mathbf{l}|_1=n-q} u_{\mathbf{l}}.$$

Within the combination technique sub-solution with a strong imbalance of grid nodes in certain coordinate directions are computed. In practice this might cause problems as the numerical solution can be sensitive to it or physical features of the solution are not properly reflected by its numerical approximation. Griebel and Huber [32] suggest to only use a subset of all possible sub-grids. We follow their idea and neglect extremely distorted grids by only considering grids, where each  $l_i \geq l_{min}$  for  $i = 1, 2, \dots, d$ . This is especially useful for higher-order finite difference schemes with a large discretization stencil. If not mentioned otherwise we use  $l_{min} = 3$  throughout this thesis. Thus, each sub-grid has at least 9 grid points in each coordinate direction. The mesh width reported in the numerical experiments with the sparse grid combination technique is then given by  $h = 2^{-(n-(d-1) \cdot l_{min})}$ .

### 2.4.1 Error Splitting

The key assumption within the combination technique is that the underlying numerical method has an error splitting in of the form (2.28), so that the surpluses of the solutions with the same level are of comparable sizes. Thus, the question arises under which conditions and for which schemes an error splitting of the form (2.28) holds?

**Finite Difference Methods** In the case of linear finite difference schemes the splitting structure has been analyzed in [9, 79, 42]. Bungartz et al. [9] proved with help of Fourier series of discrete and semi-discrete solutions for the two-dimensional Laplace equation that a second-order central difference scheme exhibits this error structure. Reisinger [79] recently extended the framework to a wider class of equations and linear finite difference schemes.

He showed that a central second-order finite difference approximation of the Poisson equation with homogeneous Dirichlet data fulfills the desired splitting structure if the solution is sufficiently smooth, so that its mixed derivatives up to order four are bounded, see Lemma 17.

In the following let  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)$  and let further  $C_K^{(\alpha_1, \alpha_2, \dots, \alpha_d)}(\Omega)$  denote the function space, where all derivatives  $\frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_d^{\alpha_d}}$  for  $u \in C_K^{(\alpha_1, \alpha_2, \dots, \alpha_d)}(\Omega)$  are continuous and bounded by  $K$  in the supremum-norm and vanish at the boundary  $\partial\Omega$ .

**Lemma 17.** [79, Theorem 3.3] Let  $u \in C_K^{(4, \dots, 4)}(\Omega_d)$  be the solution to the Poisson equa-

tion and  $u_l$  be the finite difference approximation on a grid  $\Omega_l$ . Then

$$u(\mathbf{x}_{l,j}) - u_l = \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subset \{1, 2, \dots, d\}}} h_{j_1}^2 \cdot h_{j_2}^2 \cdot \dots \cdot h_{j_m}^2 w_{j_1, j_2, \dots, j_m}(\mathbf{x}_{l,j}; h_{j_1}, h_{j_2}, \dots, h_{j_m}),$$

for  $\mathbf{x}_{l,j} \in \Omega_l$  with

$$|w_{j_1, j_2, \dots, j_m}(\mathbf{x}_{l,j}; h_{j_1}, h_{j_2}, \dots, h_{j_m})| \leq m! 96^{-m} K.$$

In order to motivate the idea of the proof, we briefly review the derivation of the splitting of the approximation in the two dimensional case. The derivation for higher dimensional problems follows the same principles. For more details we refer to [79]. Therefore, we consider the Poisson equation on the unit hypercube

$$\begin{aligned} Lu &= f, & x \in \Omega_d &= [0, 1]^d \\ \text{with } L &= -\Delta, \\ u &= 0 & \text{on } \partial\Omega_d. \end{aligned} \tag{2.29}$$

Let the full discretization be given by

$$L_l u = -\delta_1^2 u - \delta_2^2 u,$$

demanding that the discrete solution  $u_l$  fulfills the system

$$L_l u_l = f_l,$$

where  $f_l$  is the restriction of  $f$  to the grid  $\Omega_l$ . Furthermore, one defines semi-discretizations

$$\begin{aligned} L_l^{(1)} u &= -\delta_1^2 u - \frac{\partial^2 u}{\partial x_2^2}, \\ L_l^{(2)} u &= -\frac{\partial^2 u}{\partial x_1^2} - \delta_2^2 u. \end{aligned}$$

The first semi-discretization is restricted to the first coordinate direction, such that one obtains a system of ODEs in the second coordinate. In the second semi-discretization the roles are reversed and one gets a system of ODEs in the first coordinate direction. The semi-discrete solutions fulfill the systems

$$\begin{aligned} L_l^{(1)} u_l^{(i)} &= f_l^{(1)}, \\ L_l^{(2)} u_l^{(i)} &= f_l^{(2)}, \end{aligned}$$

where  $f_l^{(i)}$  for  $i = 1, 2$  are the restrictions of  $f$  to the semi-discrete grid.

According to [79, Lemma 3.1] an upper bound holds for the solution

$$\|u\|_\infty \leq \frac{1}{8} \|f\|_\infty,$$

as well as for the (semi-)discrete counterparts

$$\begin{aligned}\|u_l\|_\infty &\leq \frac{1}{8}\|f\|_\infty, \\ \|u_l^{(1)}\|_\infty &\leq \frac{1}{8}\|f\|_\infty, \\ \|u_l^{(2)}\|_\infty &\leq \frac{1}{8}\|f\|_\infty.\end{aligned}\tag{2.30}$$

The bounds of the discrete solutions can be proven with help of the discrete maximum principle and the comparison principle.

The derivation of the splitting structure starts with the Taylor expansion of the finite difference scheme

$$L_l u(\mathbf{x}_{l,j}) - f(\mathbf{x}_{l,j}) = h_1^2 \tau_1(\mathbf{x}_{l,j}; h_1) + h_2^2 \tau_2(\mathbf{x}_{l,j}; h_2),$$

where  $\tau_i = \frac{1}{12} \frac{\partial^4 u}{\partial x_i^4}$  for  $i = 1, 2$ . The truncation errors of the finite difference scheme can be written in terms of semi-discrete auxiliary problems

$$\begin{aligned}L_l^{(1)} w_1 &= \tau_1, \\ L_l^{(2)} w_2 &= \tau_2.\end{aligned}$$

Due to the boundedness of the discretization operator we have

$$\|w_i\|_\infty \leq \frac{1}{8} \|\tau_i\|_\infty \leq \frac{1}{8} \frac{1}{12} K$$

for  $i = 1, 2$ . The last inequality follows since we consider  $u \in C_K^{(4,4)}(\Omega)$ . With the help of the auxiliary problems one concludes

$$\begin{aligned}&L_l [u(\mathbf{x}_{l,j}) - h_1^2 w_1(\mathbf{x}_{l,j}; h_1) - h_2^2 w_2(\mathbf{x}_{l,j}; h_2)] - f(\mathbf{x}_{l,j}) \\ &= h_1^2 (\tau_1(\mathbf{x}_{l,j}; h_1) - L_l w_1(\mathbf{x}_{l,j}; h_1)) + h_2^2 (\tau_2(\mathbf{x}_{l,j}; h_2) - L_l w_2(\mathbf{x}_{l,j}; h_2)) \\ &= h_1^2 (L_l^{(1)} - L_l) w_1(\mathbf{x}_{l,j}; h_1) + h_2^2 (L_l^{(2)} - L_l) w_2(\mathbf{x}_{l,j}; h_2).\end{aligned}\tag{2.31}$$

The difference between the semi- and full discretization yields

$$\begin{aligned}(L_l^{(1)} - L_l) w_1(\mathbf{x}_{l,j}; h_1) &= -\frac{\partial^2 w_1}{\partial x_2^2}(\mathbf{x}_{l,j}; h_1) + \frac{1}{h_2^2} \left( w_1(\mathbf{x}_{l,j} - e_2 h_2; h_1) \right. \\ &\quad \left. - 2w_1(\mathbf{x}_{l,j}; h_1) + w_1(\mathbf{x}_{l,j} + e_2 h_2; h_1) \right) \\ &= h_2^2 \frac{1}{12} \frac{\partial^4 w_1}{\partial x_2^4}(\xi; h_1) =: h_2^2 \sigma_{1,2},\end{aligned}\tag{2.32}$$

for some  $\xi$  in a neighborhood around  $\mathbf{x}_{l,j}$ . In an analogue way we obtain

$$(L_l^{(2)} - L_l) w_2(\mathbf{x}_{l,j}; h_2) =: h_1^2 \sigma_{2,1},\tag{2.33}$$

with  $\sigma_{2;1} := \frac{1}{12} \frac{\partial^4 w_2}{\partial x_1^4}$ . Inserting (2.32) and (2.33) into equation (2.31) we get

$$\begin{aligned} L_l [u(\mathbf{x}_{l,j}) - h_1^2 w_1(\mathbf{x}_{l,j}; h_1) - h_2^2 w_2(\mathbf{x}_{l,j}; h_2)] - f(\mathbf{x}_{l,j}) \\ = h_1^2 h_2^2 \sigma_{1;2}(\mathbf{x}_{l,j}; h_1, h_2) + h_1^2 h_2^2 \sigma_{2;1}(\mathbf{x}_{l,j}; h_1, h_2) \\ = h_1^2 h_2^2 \tau_{1,2}(\mathbf{x}_{l,j}; h_1, h_2), \end{aligned}$$

where

$$\begin{aligned} \|\tau_{1,2}\|_\infty &= \|\sigma_{1;2} + \sigma_{2;1}\|_\infty \leq \frac{1}{12} \left\| \frac{\partial^4 w_1}{\partial x_2^4} \right\|_\infty + \frac{1}{12} \left\| \frac{\partial^4 w_2}{\partial x_1^4} \right\|_\infty \\ &\leq \frac{1}{12} \frac{1}{8} \left\| \frac{\partial^4}{\partial x_2^4} \tau_1 \right\|_\infty + \frac{1}{12} \frac{1}{8} \left\| \frac{\partial^4}{\partial x_1^4} \tau_2 \right\|_\infty \\ &= \frac{1}{12^2} \frac{1}{8} \left\| \frac{\partial^8 u}{\partial x_1^4 \partial x_2^4} \right\|_\infty + \frac{1}{12^2} \frac{1}{8} \left\| \frac{\partial^8 u}{\partial x_1^4 \partial x_2^4} \right\|_\infty \leq 2 \frac{1}{12^2} \frac{1}{8} K. \end{aligned}$$

Multiplication with  $L_l^{-1}$  yields the desired splitting structure

$$u(\mathbf{x}_{l,j}) - u_l(\mathbf{x}_{l,j}) = h_1^2 w_1(\mathbf{x}_{l,j}; h_1) + h_2^2 w_2(\mathbf{x}_{l,j}; h_2) + h_1^2 h_2^2 w_{1,2}(\mathbf{x}_{l,j}; h_1, h_2),$$

where  $L_l^{-1} \tau_{1,2}(\mathbf{x}_{l,j}; h_1, h_2) = w_{1,2}(\mathbf{x}_{l,j}; h_1, h_2)$  and bounded errors

$$\begin{aligned} \|w_1\|_\infty &\leq \frac{1}{12} \frac{1}{8} K, \\ \|w_2\|_\infty &\leq \frac{1}{12} \frac{1}{8} K, \\ \|w_{1,2}\|_\infty &\leq 2 \frac{1}{12^2} \frac{1}{8^2} K. \end{aligned}$$

We see that the derivation relies on the form of the truncation error, the smoothness of the solution, so that the higher derivatives are bounded and on the bounds for the (semi-)discrete solutions. For general finite difference methods with an order of accuracy  $m$  Reisinger [79] notes the following key properties, which have to be fulfilled:

1. The scheme has a pointwise truncation error of the form

$$(L - L_l)u(\mathbf{x}_{l,j}) = \sum_{k=1}^d \sum_{\substack{\{j_1, \dots, j_k\} \\ \subseteq \{1, \dots, d\}}} \tau_{j_1, \dots, j_k}(\mathbf{x}_{l,j}; h_{j_1}, \dots, h_{j_k}) h_{j_1}^m \cdots h_{j_k}^m,$$

for  $\mathbf{x}_{l,j} \in \Omega_l$ ,

2. Stability of the discretization scheme.
3. Sufficiently smooth initial data and compatible boundary data, such that the mixed derivatives of required order are bounded.

In the case of higher-order schemes the same framework can be applied to investigate the regularity requirements. However, it is difficult to prove estimates of the form (2.30) as the higher-order discretization matrix is generally not a  $M$  matrix. For the standard

nine-points-cross finite difference Laplace operator of order four the maximum principle, usually referred to as *monotonicity*, has been proven by Price [77].

In the case of HOC schemes no theoretical results are available. However, one can derive a bound for the discrete solution. With the help of the *Lemma of Lax-Milgram* bounds in the 2-norm can be given

$$\|u_l\|_2 \leq \frac{1}{c} \|f\|_2,$$

where  $c$  is the smallest eigenvalue of the discretization matrix under the assumption that the matrix is symmetric positive semi-definite, such that  $c > 0$  holds. The constant  $c$  can be derived from the coercivity requirement of the bilinear form

$$\langle x, x \rangle_{L_l} = x^\top L_l x \geq c \|x\|_2^2 \quad \text{for } x \in \mathbb{R}^n \setminus \{0\}.$$

According to (2.12) we have

$$L_l = F_1 + F_2 + \dots + F_d,$$

with

$$\begin{aligned} F_i &= B_{x_i}^{-1} A_{x_i} \\ &= \left( I_{N_1 \cdot N_2 \cdot \dots \cdot N_d} + \frac{h_i^2}{12} \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \right)^{-1} \\ &\quad \cdot \left( -I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \right) \quad \text{for } i = 1, 2, \dots, d. \end{aligned}$$

The matrices  $A_{x_i}$  and  $B_{x_i}$  are commuting and symmetric for each  $i = 1, 2, \dots, d$ . Since the product of two symmetric matrices is symmetric iff both matrices commute, we conclude that  $L_l$  is symmetric. It remains to investigate the positive definiteness. According to Lemma 18 the eigenvalues of  $A_{x_i}$  and  $B_{x_i}$  of the HOC scheme to approximate (2.29) are given by

$$\begin{aligned} \tilde{z}_{i,j} &= 4N_i^2 \sin^2 \left( \frac{\pi j}{2N_i} \right), \quad \text{for } j = 1, 2, \dots, N_i - 1, \\ \bar{z}_{i,j} &= 1 - \frac{1}{3} \sin^2 \left( \frac{\pi j}{2N_i} \right), \quad \text{for } j = 1, 2, \dots, N_i - 1, \end{aligned}$$

for  $i = 1, 2, \dots, d$ . Let  $v_j$  denote the  $j$ -th eigenvector of  $A_{x_i}$ , then

$$\begin{aligned} A_{x_i} v_j &= \tilde{z}_{i,j} v_j \\ \Leftrightarrow \left( I_{N_1 \cdot N_2 \cdot \dots \cdot N_d} - \frac{h_i^2}{12} A_{x_i} \right) v_j &= \left( 1 - \frac{h_i^2}{12} \tilde{z}_{i,j} \right) v_j \\ \Leftrightarrow B_{x_i} v_j &= \bar{z}_{i,j} v_j. \end{aligned}$$

We observe that both matrices share the same set of eigenvectors and thus the discretization  $B_{x_i}^{-1} A_{x_i}$  has the eigenvalues

$$z_{i,j} := \frac{\tilde{z}_{i,j}}{\bar{z}_{i,j}}$$

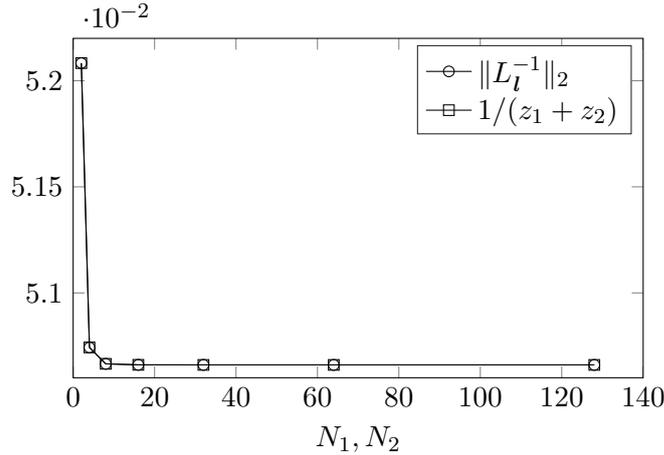


Figure 2.6: Theoretical bound and inverse discretization operator in the 2-norm.

with smallest eigenvalue

$$z_i := \frac{4N_i^2 \sin^2\left(\frac{\pi}{2N_i}\right)}{1 - \frac{1}{3} \sin^2\left(\frac{\pi}{2N_i}\right)}$$

for  $i = 1, 2, \dots, d$ . Since  $z_i > 0$  the matrix is positive definite.

**Lemma 18.** [83, Lemma 4.3]

$$\text{Let } G = \begin{pmatrix} \alpha & \beta & & 0 \\ \gamma & \ddots & \ddots & \\ & \ddots & \ddots & \beta \\ 0 & & \gamma & \alpha \end{pmatrix}.$$

be an  $N \times N$  matrix. Then the eigenvalues  $z_j$  and eigenvectors  $v_j$  are

$$z_j = \alpha + 2\beta \sqrt{\frac{\gamma}{\beta}} \cos\left(\frac{j\pi}{N+1}\right) \quad \text{for } j = 1, 2, \dots, N,$$

$$v_j = \left( \sqrt{\frac{\gamma}{\beta}} \sin\left(\frac{j\pi}{N+1}\right), \left(\sqrt{\frac{\gamma}{\beta}}\right)^2 \sin\left(\frac{2j\pi}{N+1}\right), \dots, \left(\sqrt{\frac{\gamma}{\beta}}\right)^N \sin\left(\frac{Nj\pi}{N+1}\right) \right)^\top.$$

Since the discrete matrix can be written as a Kronecker sum of the single  $B_{x_i}^{-1}A_{x_i}$ , the smallest eigenvalue of  $L_l$  is, according to [64, Theorem 13.16], given by

$$c = z_1 + z_2 + \dots + z_d.$$

In Figure 2.6 we compare the theoretical bound  $1/c$  with  $\|L_l^{-1}\|_2$ . One observes that the bound is sharp.

Lemma 19 yields the splitting for fourth-order schemes, which have a truncation error of

the form

$$(L - L_l)u(\mathbf{x}_{l,j}) = \sum_{k=1}^d \sum_{\substack{\{j_1, \dots, j_k\} \\ \subseteq \{1, \dots, d\}}} \tau_{j_1, \dots, j_k}(\mathbf{x}_{l,j}; h_{j_1}, \dots, h_{j_k}) h_{j_1}^4 \cdots h_{j_k}^4,$$

for  $\mathbf{x}_{l,j} \in \Omega_l$  and (semi-)discrete solutions, which are bounded by

$$\|u_l^{(i_1, i_2, \dots, i_n)}\|_\infty \leq \frac{1}{8} \|f\|_\infty.$$

Please note that one can derive a similar error structure for HOC schemes in the 2-norm and the bound  $c$  given above.

**Lemma 19.** *Let  $u \in C_K^{(6, \dots, 6)}(\Omega_d)$  be the solution to the Poisson equation and let  $u_l$  denote its fourth-order nine-points-cross finite difference solution of order four on the grid  $\Omega_l$ . Then the pointwise error is*

$$u(\mathbf{x}_{l,j}) - u_l = \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subseteq \{1, 2, \dots, d\}}} h_{j_1}^4 \cdot h_{j_2}^4 \cdots h_{j_m}^4 w_{j_1, j_2, \dots, j_m}(\mathbf{x}_{l,j}; h_{j_1}, h_{j_2}, \dots, h_{j_m}),$$

for  $\mathbf{x}_{l,j} \in \Omega_l$  with

$$|w_{j_1, j_2, \dots, j_m}(x; h_{j_1}, h_{j_2}, \dots, h_{j_m})| \leq m! 8^{-m} 90^{-m} K.$$

*Proof.* Consistency of finite difference approximation via Taylor expansion and application of the framework developed by Reisinger in [79].  $\square$

According to Lemmas 17, 19, respectively, the finite difference approximations have a pointwise error of the desired form. However, the error is only given for the discrete grid values. As the discrete solution of the finite difference scheme has to be extended via interpolation, it is necessary to preserve the error structure on the entire domain. Otherwise the accuracy of  $\mathcal{O}(h^m \log_2(h^{-1})^{d-1})$  of the sparse grid combination technique would only hold for grid points which belong to all sub-grids. In the interior of the domain this is just the node  $(0.5, 0.5, \dots, 0.5)$ . In order to demonstrate this deterioration of the rate of convergence, we solve the two-dimensional Poisson boundary problem

$$\begin{aligned} \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} &= (x_1^2 + x_2^2) \exp(x_1 x_2) && \text{on } \Omega_2 = (0, 1)^2, \\ u(x_1, x_2) &= \exp(x_1 x_2) && \text{on } \partial\Omega_2, \end{aligned} \quad (2.34)$$

with the combination technique and a standard fourth-order finite difference scheme. The sub-solutions are combined via multilinear interpolation. Figure 2.7 shows that the error decreases with  $\mathcal{O}(h^4 \log_2(h^{-1}))$  at the midpoint  $(0.5, 0.5)$ , while the error of the interpolation routine dominates in the maximum norm and leads to a reduced convergence order of  $\mathcal{O}(h^2 \log_2(h^{-1}))$ .

In order to establish a high rate of convergence, we require that the error structure is

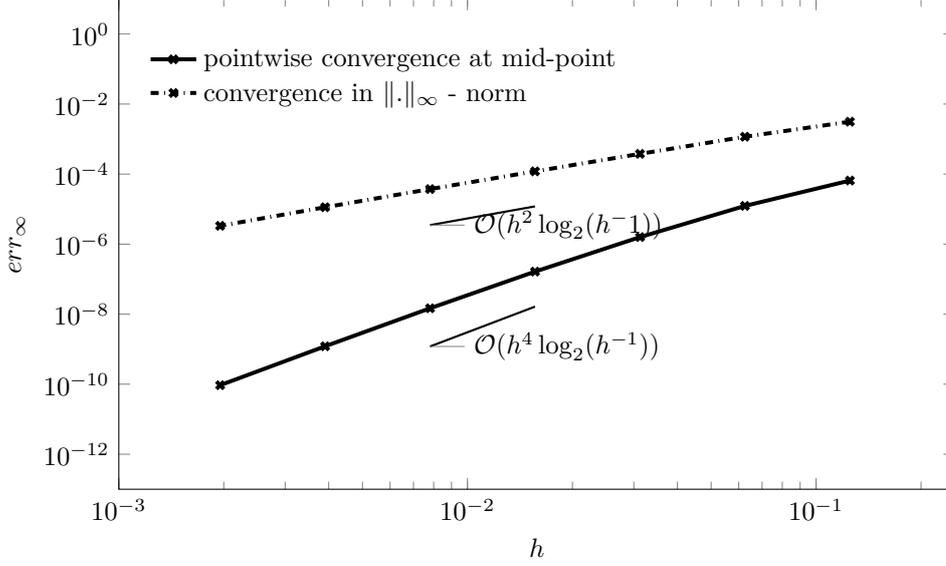


Figure 2.7: Convergence at the mid point and in the maximum norm.

preserved by the interpolation scheme, such that

$$u(\mathbf{x}) - (P_N u_I)(\mathbf{x}) = \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subset \{1, 2, \dots, d\}}} h_{j_1}^4 \cdot h_{j_2}^4 \cdot \dots \cdot h_{j_m}^4 \tilde{w}_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}) \quad (2.35)$$

holds for all  $\mathbf{x} \in \Omega_2$ . In order to analyze the error, we follow the approach of Reisinger [79] and split the error

$$u(\mathbf{x}) - (P_N u_I)(\mathbf{x}) = \underbrace{u(\mathbf{x}) - (P_N u)(\mathbf{x})}_I + \underbrace{(P_N(u - u_I))(\mathbf{x})}_{II}, \quad (2.36)$$

where  $P_N$  is the multi-dimensional tensor based interpolation operator. The error  $I$  is the interpolation error, while  $II$  is the interpolation of the error of the numerical solution. Lemma 20 states the interpolation error  $I$ , while Lemma 21 gives an expression of error term  $II$ . From  $I$  and  $II$  the desired error splitting (2.35) can be deduced, see Theorem 22.

**Lemma 20.** *Let  $u \in C_K^{(4, \dots, 4)}(\Omega_d)$  and a univariate cubic spline interpolation  $P_{N_i}$  along the  $i$ -th coordinate direction for  $i = 1, \dots, d$  with  $P_{N_i} u = u + h_i^4 R_{N_i} u$  and remainder operator  $R_{N_i}$  be given. Then the error of the tensor product interpolation  $P_N = P_{N_1} \otimes P_{N_2} \otimes \dots \otimes P_{N_d}$  is*

$$(P_N u)(\mathbf{x}) - u(\mathbf{x}) = \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subset \{1, 2, \dots, d\}}} h_{j_1}^4 \cdot h_{j_2}^4 \cdot \dots \cdot h_{j_m}^4 c_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}),$$

for all  $\mathbf{x} \in \Omega_d$  with  $\|c_{j_1, j_2, \dots, j_m}\|_\infty \leq \frac{5^m}{384^m} K$  for  $m = 1, 2, \dots, d$ .

*Proof.* The error of each univariate cubic spline interpolant is given by

$$P_{N_i}u = (I + h_i^4 R_{N_i})u, \quad \text{for } i = 1, 2, \dots, d,$$

where  $I$  is the identity. According to [40] the remainder term  $R_{N_i}u$  is bounded by

$$\|R_{N_i}u\|_\infty \leq \frac{5}{384} \left\| \frac{\partial^4 u}{\partial x_i^4} \right\|_\infty.$$

In the tensor based multivariate case we have

$$\begin{aligned} (P_N u)(\mathbf{x}) &= (P_{N_1} \otimes P_{N_2} \otimes \dots \otimes P_{N_d} u)(\mathbf{x}) \\ &= (I + h_1^4 R_{N_1}) \otimes (I + h_2^4 R_{N_2}) \otimes \dots \otimes (I + h_d^4 R_{N_d}) u(\mathbf{x}) \\ &= u(\mathbf{x}) + \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subset \{1, 2, \dots, d\}}} h_{j_1}^4 \cdot h_{j_2}^4 \cdot \dots \cdot h_{j_m}^4 R_{j_1} \otimes R_{j_2} \otimes \dots \otimes R_{j_m} u(\mathbf{x}). \end{aligned}$$

Defining  $c_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}) := R_{N_{j_1}} \otimes R_{N_{j_2}} \otimes \dots \otimes R_{N_{j_m}} u(\mathbf{x})$  we obtain the desired form. Thus, it holds

$$\|R_{N_{j_1}} \otimes R_{N_{j_2}} \otimes \dots \otimes R_{N_{j_m}} u\|_\infty \leq \frac{5^m}{384^m} \left\| \frac{\partial^{4m}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4} u \right\|_\infty \leq \frac{5^m}{384^m} K. \quad \square$$

Before we investigate the interpolation of the pointwise error, we derive bounds for the remainder terms  $R_{N_{j_1}} \otimes R_{N_{j_2}} \otimes \dots \otimes R_{N_{j_m}} w_{i_1, i_2, \dots, i_n}$ . Similar to the two-dimensional case one can compute bounds for the analytical and (semi-)discrete solution of the Poisson equation with homogenous Dirichlet boundary data

$$\begin{aligned} \|u\|_\infty &\leq \frac{1}{8} \|f\|_\infty, \\ \|u_l^{(i_1, \dots, i_m)}\|_\infty &\leq \frac{1}{8} \|f\|_\infty. \end{aligned} \quad (2.37)$$

Restricting ourselves to function spaces with vanishing derivatives of sufficiently high order at the boundary, we can also derive bounds for the derivatives of  $f$ ,  $u$  respectively. We cite from Reisinger [79] the auxiliary problem with solution  $w_{i_1, \dots, i_n}$

$$L_l^{(i_1, i_2, \dots, i_n)} w_{i_1, i_2, \dots, i_n} = \tau_{i_1, i_2, \dots, i_n} \quad (2.38)$$

and the definition of the terms  $\tau_{i_1, i_2, \dots, i_n}$

$$\tau_{i_1, i_2, \dots, i_n} := \sum_{\substack{z_1, z_2, \dots, z_{n-1}, z \\ \text{s.t. } \{z_1, z_2, \dots, z_{n-1}\} \cup \{z\} \\ = \{i_1, i_2, \dots, i_n\}}} \sigma_{z_1, \dots, z_{n-1}; z}. \quad (2.39)$$

Let us note that  $\tau_{i_1}$  for  $i_1 = 1, 2, \dots, d$  is the truncation error of the finite difference stencil in coordinate direction  $i_1$ . The functions  $\sigma_{z_1, z_2, \dots, z_n}$  are obtained via the expansion

$$\left( L_l^{(i_1, i_2, \dots, i_n)} - L_l \right) w_{i_1, i_2, \dots, i_n} = \sum_{\substack{k \in \{1, 2, \dots, d\} \\ k \notin \{i_1, i_2, \dots, i_n\}}} \frac{1}{90} h_k^4 \frac{\partial^6}{\partial x_k^6} w_{i_1, i_2, \dots, i_n} = \sum_{\substack{k \in \{1, 2, \dots, d\} \\ k \notin \{i_1, i_2, \dots, i_n\}}} h_k^4 \sigma_{i_1, i_2, \dots, i_n; k}. \quad (2.40)$$

The terms  $\sigma_{i_1, i_2, \dots, i_n; k}$  can be expressed as the truncation error of the semi-discrete and fully discrete problem from the above formula and thus

$$\|\sigma_{i_1, i_2, \dots, i_n; k}\|_\infty = \frac{1}{90} \left\| \frac{\partial^6}{\partial x_k^6} w_{i_1, i_2, \dots, i_n} \right\|_\infty$$

holds. Analogue to the procedure in the proof of Lemma 20, we have

$$\begin{aligned} & \left\| R_{N_{j_1}} \otimes R_{N_{j_2}} \otimes \dots \otimes R_{N_{j_m}} w_{i_1, i_2, \dots, i_n} \right\|_\infty \leq \frac{5^m}{384^m} \left\| \frac{\partial^{4m}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4} w_{i_1, i_2, \dots, i_n} \right\|_\infty \\ & \stackrel{(2.37), (2.38)}{\leq} \frac{5^m}{384^m} \frac{1}{8} \left\| \frac{\partial^{4m}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4} \tau_{i_1, i_2, \dots, i_n} \right\|_\infty \\ & \stackrel{(2.39)}{=} \frac{5^m}{384^m} \frac{1}{8} \left\| \frac{\partial^{4m}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4} \sum_{\substack{z_1, z_2, \dots, z_{n-1}, z \\ \text{s.t. } \{z_1, z_2, \dots, z_{n-1}\} \cup \{z\} \\ = \{i_1, i_2, \dots, i_n\}}} \sigma_{z_1, z_2, \dots, z_{n-1}; z} \right\|_\infty \\ & \stackrel{(2.40)}{=} \frac{5^m}{384^m} \frac{1}{8} \frac{1}{90} \left\| \frac{\partial^{4m}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4} \sum_{\substack{z_1, z_2, \dots, z_{n-1}, z \\ \text{s.t. } \{z_1, z_2, \dots, z_{n-1}\} \cup \{z\} \\ = \{i_1, i_2, \dots, i_n\}}} \frac{\partial^6}{\partial x_z^6} w_{z_1, z_2, \dots, z_{n-1}} \right\|_\infty. \end{aligned}$$

The sum has  $n$  terms and we repeat this procedure  $n - 1$  times until we can conclude the final result in the last step

$$\begin{aligned} & \leq \frac{5^m}{384^m} \frac{1}{8^{n-1}} \frac{1}{90^{n-1}} \left\| \frac{\partial^{4m}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4} (n-1)! \sum_{k=1}^n \frac{\partial^{6(n-1)}}{\partial x_{i_1}^6 \dots \partial x_{i_{l \neq k}}^6 \dots \partial x_{i_n}^6} w_k \right\|_\infty \\ & \leq \frac{5^m}{384^m} \frac{1}{8^n} \frac{1}{90^{n-1}} \left\| \frac{\partial^{4m}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4} (n-1)! \sum_{k=1}^n \frac{\partial^{6n}}{\partial x_{i_1}^6 \partial x_{i_2}^6 \dots \partial x_{i_n}^6} \tau_k \right\|_\infty \\ & \leq \frac{5^m}{384^m} \frac{1}{8^n} \frac{1}{90^n} n! \left\| \frac{\partial^{4m+6n}}{\partial x_{j_1}^4 \partial x_{j_2}^4 \dots \partial x_{j_m}^4 \partial x_{i_1}^6 \partial x_{i_2}^6 \dots \partial x_{i_n}^6} u \right\|_\infty \\ & \leq \frac{5^m}{384^m} \frac{1}{8^n} \frac{1}{90^n} n! K. \end{aligned} \tag{2.41}$$

Here we see that  $u \in C_K^{(10, \dots, 10)}(\Omega_d)$  has to be satisfied to ensure a bounded error.

**Lemma 21.** *Let  $u \in C_K^{(10, \dots, 10)}(\Omega_d)$  be the solution to the Poisson equation and let  $u_l$  denote its fourth-order nine-points-cross finite difference solution on the grid  $\Omega_l$ . Using a tensor product interpolation  $P_N$  with univariate cubic spline interpolation in each coordinate direction, then the interpolation of the pointwise error reads*

$$(P_N(u - u_l))(\mathbf{x}) = \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subset \{1, 2, \dots, d\}}} h_{j_1}^4 \cdot h_{j_2}^4 \cdot \dots \cdot h_{j_m}^4 \beta_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}),$$

for all  $\mathbf{x} \in \Omega$  with

$$|\beta_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m})| \leq K C_m,$$

for  $m = 1, 2, \dots, d$  and constant  $C_m \in \mathbb{R}$

$$\sum_{\substack{m, n \in \mathbb{N} \\ \text{s.t. } m, n \leq k \\ k \leq m+n}} \binom{k}{m} \binom{m}{n-(k-m)} \frac{5^m}{384^m} n! 8^{-n} 90^{-n} + k! 8^{-k} 90^{-k} =: C_k$$

*Proof.* Interpolation of the pointwise error gives

$$(P_N(u - u_I))(\mathbf{x}) = \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subset \{1, 2, \dots, d\}}} h_{j_1}^4 \cdot h_{j_2}^4 \cdot \dots \cdot h_{j_m}^4 (P w_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}))(\mathbf{x}), \quad (2.42)$$

where

$$\begin{aligned} (P_N w_{j_1, j_2, \dots, j_m}(\Omega_h; h_{j_1}, h_{j_2}, \dots, h_{j_m}))(\mathbf{x}) &= w_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}) \\ &+ \sum_{n=1}^d \sum_{\substack{\{l_1, l_2, \dots, l_n\} \\ \subset \{1, 2, \dots, d\}}} h_{l_1}^4 \cdot h_{l_2}^4 \cdot \dots \cdot h_{l_n}^4 R_{l_1} \otimes R_{l_2} \otimes \dots \otimes R_{l_n} w_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}). \end{aligned}$$

Condensing all terms in (2.42), which have the same leading step sizes, we define

$$\begin{aligned} h_{l_1}^4 \cdot h_{l_2}^4 \cdot \dots \cdot h_{l_k}^4 \beta_{l_1, l_2, \dots, l_k}(\mathbf{x}; h_{l_1}, h_{l_2}, \dots, h_{l_k}) &:= \\ \sum_{\substack{m, n \in \mathbb{N} \\ \text{s.t. } m, n \leq k \\ k \leq m+n}} \sum_{\substack{\{i_1, i_2, \dots, i_m\} \cup \{j_1, j_2, \dots, j_n\} \\ = \{l_1, l_2, \dots, l_k\}}} h_{i_1}^4 \cdot h_{i_2}^4 \cdot \dots \cdot h_{i_m}^4 \\ &\cdot R_{N_{i_1}} \otimes R_{N_{i_2}} \otimes \dots \otimes R_{N_{i_m}} h_{j_1}^4 \cdot h_{j_2}^4 \cdot \dots \cdot h_{j_n}^4 w_{j_1, j_2, \dots, j_n}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_n}) \\ &+ h_{l_1}^4 \cdot \dots \cdot h_{l_k}^4 w_{l_1, \dots, l_k}(\mathbf{x}; h_{l_1}, \dots, h_{l_k}). \end{aligned}$$

We already know that it holds (cf. (2.41))

$$\|R_{N_{i_1}} \otimes R_{N_{i_2}} \otimes \dots \otimes R_{N_{i_m}} w_{j_1, j_2, \dots, j_n}\|_\infty \leq \frac{5^m}{384^m} n! 8^{-n} 90^{-n} K.$$

The inner sum has  $\binom{k}{m} \binom{m}{n-(k-m)}$  elements and we obtain the estimate

$$\|\beta_{l_1, \dots, l_k}\|_\infty \leq K \sum_{\substack{m, n \in \mathbb{N} \\ \text{s.t. } m, n \leq k \\ k \leq m+n}} \binom{k}{m} \binom{m}{n-(k-m)} \frac{5^m}{384^m} n! 8^{-n} 90^{-n} + K k! 8^{-k} 90^{-k} =: K C_k. \quad \square$$

**Theorem 22.** Let  $u \in C_K^{(10, \dots, 10)}(\Omega_d)$  be the solution to the Poisson equation and let  $u_I$  denote its fourth-order nine-points-cross finite difference solution on the grid  $\Omega_I$ . Using a tensor product interpolation  $P_N$  with univariate cubic spline interpolation in each coordinate direction, the error between the analytical solution and the interpolation of the finite

difference solution is

$$u(\mathbf{x}) - (P_N u)(\mathbf{x}) = \sum_{m=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_m\} \\ \subset \{1, 2, \dots, d\}}} h_{j_1}^4 \cdot h_{j_2}^4 \cdot \dots \cdot h_{j_m}^4 \gamma_{j_1, j_2, \dots, j_m}(\mathbf{x}; h_{j_1}, h_{j_2}, \dots, h_{j_m}),$$

for  $\mathbf{x} \in \Omega_d$  with  $\|\gamma_{j_1, j_2, \dots, j_m}\|_\infty \leq (\frac{5^m}{384^m} + C_m)K$  for  $m = 1, 2, \dots, d$ .

The proof immediately follows from Lemma 20 *I* and Lemma 21 *II* and by definition of  $\gamma_{j_1, \dots, j_m} := \beta_{j_1, \dots, j_m} - c_{j_1, \dots, j_m}$ .  $\square$

$m$	1	2	3	4
$C_m$	0.0014	4.0599e-5	8.8699e-7	1.7416e-8

Table 2.1:  $C_m$  for different choices of  $m$ .

Table 2.1 states the constants  $C_m$  for the  $\beta$  functions in Lemma 3 for  $m = 1, \dots, 4$ .

**Pseudo-spectral Methods** Pseudo-spectral methods on sparse grids have been rarely used in the literature. To our knowledge there are only the works of Shen and Yu [84, 85], who construct a spectral sparse grid for elliptic problems. They use Smolyak's algorithms to build a sparse grid based on nested, spectrally accurate quadratures. We follow a different approach and use the combination technique to construct the sparse grid. Please recall that the combination technique is based on an appropriate error splitting structure of the underlying numerical full grid solver. At the current state, an analytical proof is missing and is part of future research. Here, the difficulty arises to prove a (semi-)discrete maximum principle to derive bounds for the coefficient functions in equation (2.28). In the following we investigate the splitting structure numerically and compute the hierarchical surpluses. We consider the four test problems given in [84] and solve the Poisson equation

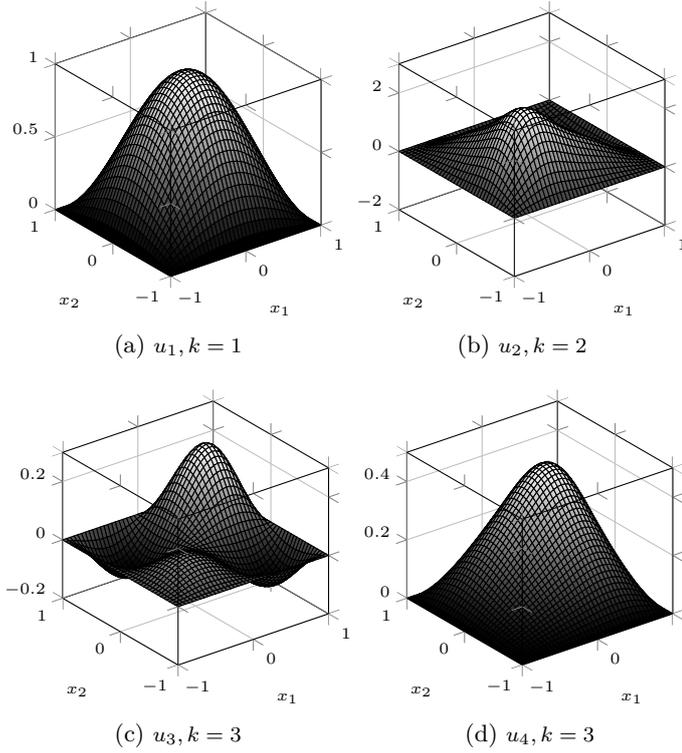
$$-\Delta u = f \quad \text{for } \mathbf{x} \in \Omega_d = [-1, 1]^d$$

with solutions

$$\begin{aligned} u_1(\mathbf{x}) &= \prod_{i=1}^d \sin(k\pi \frac{x_i+1}{2}), & u_2(\mathbf{x}) &= \sum_{i=1}^d \phi_k(x_i) \prod_{i \neq j} \sin(\pi \frac{x_i+1}{2}), \\ u_3(\mathbf{x}) &= \prod_{i=1}^d g_k(x_i), & u_4(\mathbf{x}) &= \prod_{i=1}^d (h_k(x_i) - \frac{x_i+1}{2}), \end{aligned}$$

where

$$\begin{aligned} \phi_k(x_i) &= e^{\sin(k\pi \frac{x_i+1}{2})} - 1 \\ g_k(x_i) &= (1 - x_i^2)(1 + x_i) \log(1 + x_i + 10^{-k}) \\ h_k(x_i) &= \begin{cases} 0, & x_i \leq 0 \\ x_i^k, & x_i > 0 \end{cases} \end{aligned}$$

Figure 2.8: Test cases for  $d = 2$ .

and  $k \in \mathbb{N}$ . The right hand sides are given by

$$f_1(\mathbf{x}) = d \frac{k^2 \pi^2}{4} u_1,$$

$$f_2(\mathbf{x}) = \sum_{i=1}^d \left( \frac{(d-1)\pi^2}{4} \phi_k(x_i) - \phi_k''(x_i) \right) \prod_{j \neq i} \sin \left( \pi \frac{x_j + 1}{2} \right),$$

$$f_3(\mathbf{x}) = - \sum_{i=1}^d \left( g_k''(x_i) \prod_{j \neq i} g_j(x_j) \right),$$

$$f_4(\mathbf{x}) = - \sum_{i=1}^d \left( h_k''(x_i) \prod_{j \neq i} \left( h_k(x_j) - \frac{x_j + 1}{2} \right) \right).$$

The first two functions are analytic without any singularity in the complex plane. In this case we expect supergeometric convergence ( $r = \infty$ ). The third function is also analytic, but has a singularity on the real axis at  $z_0 = -1 - 10^{-k}$ . According to Theorem 14 the factor  $r$  is given by  $r = 1 + 10^{-k} + \sqrt{(1 + 10^{-k})^2 - 1}$ . For smaller  $k$  values the distance of  $z_0$  to the domain  $[-1, 1]$  becomes larger and leads to a faster convergence. We assume a numerical error of the form

$$u(\mathbf{x}) - (P_N u_l)(\mathbf{x}) = \sum_{k=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_k\} \\ \subset \{1, 2, \dots, d\}}} r^{-N_{j_1}} \cdot r^{-N_{j_2}} \cdot \dots \cdot r^{-N_{j_k}} \gamma_{j_1, j_2, \dots, j_k}(\mathbf{x}; N_{j_1}, N_{j_2}, \dots, N_{j_k}),$$

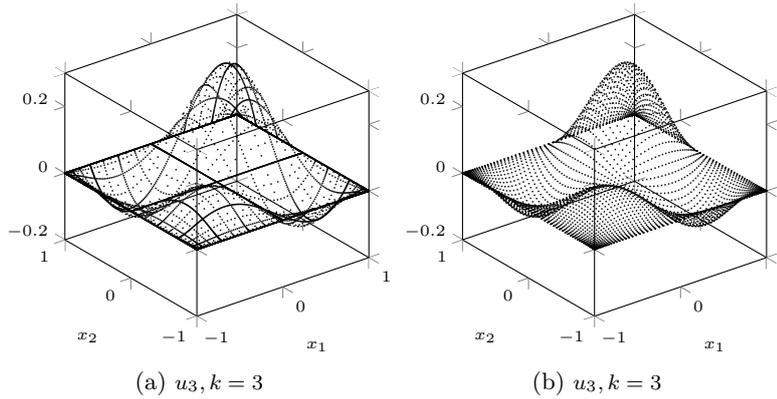


Figure 2.9: Sparse grid and full grid solution.

with bounded functions  $\gamma$ . Under this assumption we can expect a hierarchical surplus of order

$$\delta u_l(\mathbf{x}) = \mathcal{O}(r^{-N_1} \cdot r^{-N_2} \dots r^{-N_d}) = \mathcal{O}(r^{-\sum_{i=1}^d N_i}) = \mathcal{O}(r^{-\sum_{i=1}^d 2^{l_i}})$$

for all  $\mathbf{x} \in \Omega_d$ . We see that this splitting structure is not appropriate for the combination technique, as the sub-grids at the same level have a different order of accuracy, e.g. let  $d = 2$  and  $n = 6$ , then the surplus of  $\delta u_{(3,3)} = \mathcal{O}(r^{-16})$ , while  $\delta u_{(2,4)} = \mathcal{O}(r^{-20})$ . The fourth example has a  $k$ -th derivative of bounded variation and according to Theorem 13 we expect an algebraic convergence. Let the error splitting be given by

$$u(\mathbf{x}) - (P_N u_l)(\mathbf{x}) = \sum_{k=1}^d \sum_{\substack{\{j_1, j_2, \dots, j_k\} \\ \subset \{1, 2, \dots, d\}}} N_{j_1}^m \cdot N_{j_2}^m \cdot \dots \cdot N_{j_k}^m \gamma_{j_1, j_2, \dots, j_k}(\mathbf{x}; N_{j_1}, N_{j_2}, \dots, N_{j_k}),$$

with bounded functions  $\gamma$  and  $m$  denotes the order of algebraic convergence. Then, analogue to the finite difference case, the hierarchical surplus is of order

$$\delta u_l(\mathbf{x}) = \mathcal{O}(N_{j_1}^m N_{j_2}^m \dots N_{j_d}^m) = \mathcal{O}(2^{-m|l|_1}).$$

If not otherwise mentioned, we have chosen  $l_{min} = 1$  in the following numerical experiments. Figures 2.10 and 2.11 state the accuracy versus the number of grid nodes or computation time, respectively. In the analytic case with supergeometric convergence of cases 1 and 2, the sparse grid combination technique performs similar or even worse than the full grid method. Due to the high rate of convergence there is no benefit in using the combination technique. In the third case we observe geometric convergence and the combined solution outperforms the full grid solution in the high accuracy region. This result is quite surprising, as the surpluses at one level are of different order, see the discussion above. However, for  $r$  close to 1 the difference in the surpluses is small and the benefit of the cancellation of low order error terms outweighs the unnecessary computational workload. Tables 2.2 and 2.3 show the hierarchical surpluses in the two-dimensional case. We observe that in the analytic case (Table 2.2) the surpluses at the same level are of different size, while they are of comparable size in the algebraic case (Table 2.3). This leaves room to improve the numerical method in the analytic case: since the boundary grids (grids with low level in one direction) are of higher order than the center grids,

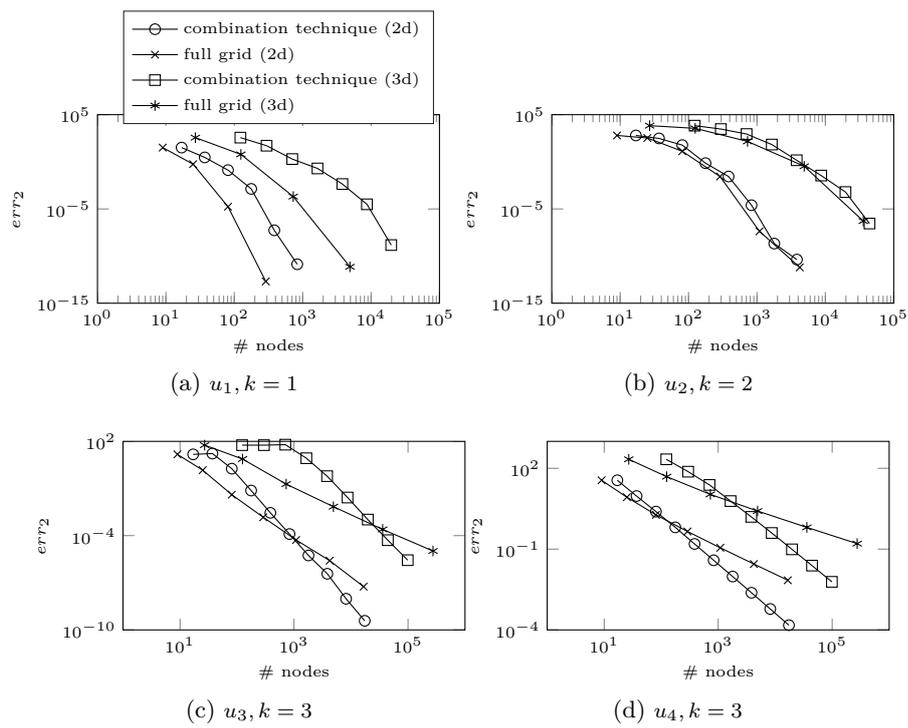


Figure 2.10: Convergence of the full grid approach versus the sparse grid combination technique.

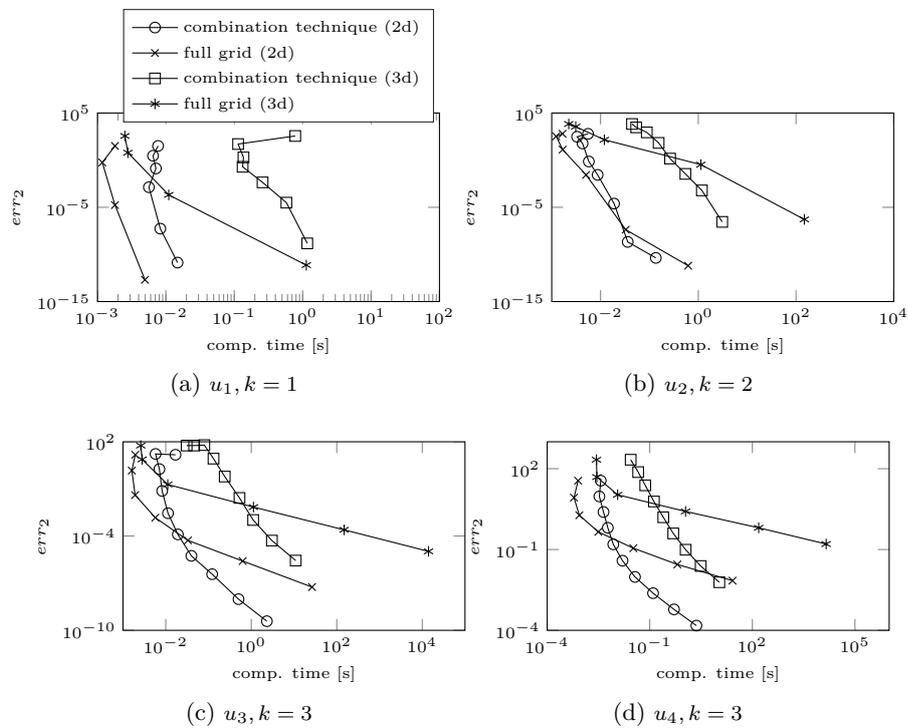


Figure 2.11: Computation time versus accuracy of the full grid approach and the sparse grid combination technique.

$l_1, l_2$	1	2	3	4	5
1	0.05320007109	3.01091817426	0.67948036244	0.00090058675	0.00000425864
2	3.01091817426	5.81390129215	1.20139326430	0.02748897172	0.00101213360
3	0.67948036244	1.20139326430	0.06600992745	0.00176267546	0.00006409891
4	0.00090058675	0.02748897172	0.00176267546	0.00007605425	0.00000253873
5	0.00000425864	0.00101213360	0.00006409891	0.00000253873	0.00000017619

Table 2.2: Hierarchical surplus of the spectral method for case 3 with  $k = 3$  and  $d = 2$ .

$l_1, l_2$	1	2	3	4	5
1	0	15.20688403880	3.15596490099	0.71863369303	0.17283592883
2	15.20688403880	4.89281826359	0.53018892670	0.10161973130	0.02346582984
3	3.15596490099	0.53018892670	0.16327550147	0.02257005970	0.00528871129
4	0.71863369303	0.10161973130	0.02257005970	0.00068325977	0.00010098876
5	0.17283592883	0.02346582984	0.00528871128	0.0001009887	0.00000821986

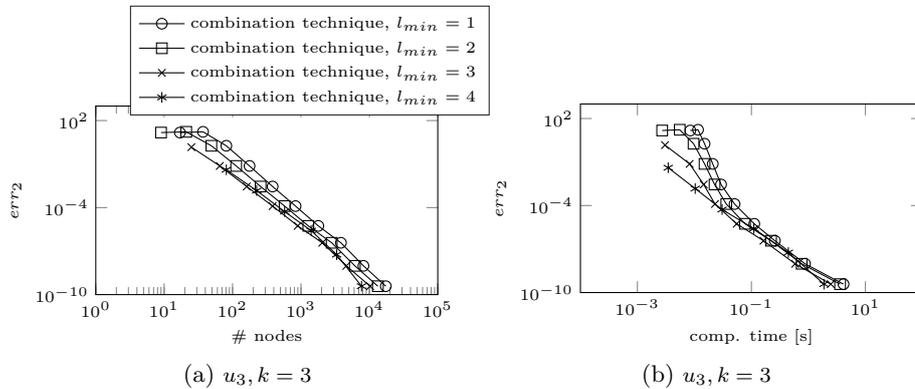
Table 2.3: Hierarchical surplus of the spectral method for case 4 with  $k = 3$  and  $d = 2$ .

they do not contribute any reasonable additional information and can therefore be neglected. In Figure 2.12 we show the improvement by exclusion of the boundary grids. By neglecting solutions with little information gain, we reduce the number of grid nodes and therefore the computation time without any loss of accuracy. Persuading this idea further and removing more boundary grids, the combined solution converges to the full grid solution. This explains the optimality of the full grid solution in the analytic case as in the asymptotic

$$\mathcal{O}(r^{-2^{n/d} - \dots - 2^{n/d}}) \gg \mathcal{O}(r^{-2^{n/d} - \dots - 2^{n/d-1} - \dots - 2^{n/d+1} - \dots - 2^{n/d}})$$

holds, where  $n$  is the level of the sparse grid combination technique w.l.o.g.  $n$  being a multiple of  $d$ . In other words, the surplus of the solution with  $N_i = 2^{n/d}$  nodes in each direction is significantly more important than all other surpluses at the same level, so that it gives no additional accuracy if we add them to our solution. In the first two experiments this effect can clearly be seen as we have supergeometric convergence with  $r = \infty$ , while in the third experiment the asymptotically optimal region of the full grid method is not reached and the sparse grid combination technique performs better.

Figure 2.13 shows the benefit of the combination technique for different rates of geometric convergence. The closer the value  $r$  is to 1 the better the combination technique performs, which reflects the previous discussion.

Figure 2.12: Modified combination technique ( $d = 2$ ).

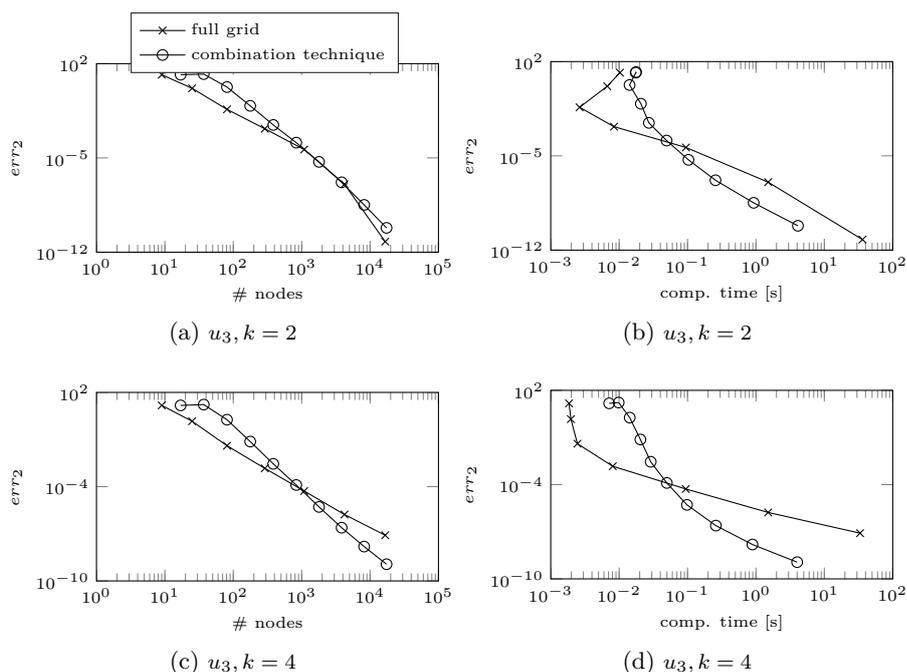


Figure 2.13: Combination technique versus full grid in experiment 3 with two spatial dimensions and  $k = 2, 4$ , such that  $r \approx 1.1518$ ,  $r \approx 1.0142$ , respectively.

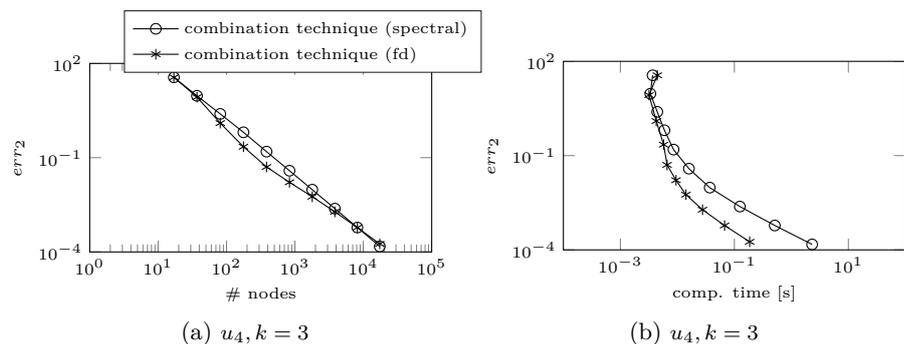


Figure 2.14: Combination technique with pseudo-spectral and finite difference discretization ( $d = 2$ ).

In the fourth case the convergence reduces to the algebraic case and the combination technique clearly outperforms the full grid method. Shen and Yu in [84] claim that this is the ideal case for their sparse grid method. However, the low accuracy due to the regularity constraints of  $u_4$  (for moderate  $k$  values) is the right setting for finite difference methods as they are much cheaper due to their sparse matrix structure and should therefore be preferred to spectral methods, see Figure 2.14.



# 3

## Chapter 3

# Time Discretization - Alternating Direction Implicit Schemes

In the previous chapter we have discussed the approximation of the spatial derivatives via finite differences or pseudo-spectral differentiation. The application of both approaches to time-dependent problems leads to the semi-discrete system of ordinary differential equations (2.1)

$$U'(t) = F(t)U(t), \quad t \geq 0,$$

supplied with suitable initial and boundary data. In the next step, a suitable time discretization method can be chosen. Standard techniques are the explicit or implicit Euler method, which exhibit first order accuracy in time. Second-order accuracy can be achieved with the Crank-Nicolson (CN) method. The three schemes are given by

$$U_{n+1} = U_n + (1 - \theta)\Delta_t F(n\Delta_t)U_n + \theta F((n+1)\Delta_t)U_{n+1},$$

where  $\Delta_t$  is the step size in time,  $U_n \sim U(n\Delta_t)$  and  $\theta > 0$  is a real parameter, called *implicitness parameter*. If  $\theta = 1$  the resulting scheme is the implicit Euler method. In the case of  $\theta = 0$  the implicit terms vanish and one obtains the explicit Euler method, while we get the Crank-Nicolson scheme for  $\theta = \frac{1}{2}$ . The Euler schemes suffer from a low accuracy and are therefore rarely used in financial applications. The second-order accuracy makes the CN scheme an interesting method to consider and explains its widespread popularity. However, if high dimensional PDEs shall be solved, the method becomes computationally too expensive. For example, if a standard central second-order finite difference discretization is applied to a problem without mixed derivatives, the matrix  $F$  has up to  $2^d + 1$  entries in each row. Since in each time step the linear system  $I - \theta\Delta_t F(n\Delta_t)$  is involved and leads to an undesirable large computational effort if the arising linear system of equations is solved exactly, e.g. via LU-decomposition.

In Figure 3.1 we demonstrate the increase in the run-time with Crank-Nicolson time stepping. We solve the two-dimensional heat equation

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{\partial u^2}{\partial x_1^2} + \frac{\partial u^2}{\partial x_2^2} \quad \text{on } \Omega_2 = (0, 1)^2, \\ u(x_1, x_2) &= 0 \text{ for } (x_1, x_2) \in \partial\Omega_2. \end{aligned}$$

on a grid with  $\mathbf{N} = (N, N)$  nodes. We observe a slope of order 1.7012 for the last four data points, which corresponds to a computational effort of  $\mathcal{O}(N^{2 \cdot 1.7012})$ . This strong increase of the computation time seems to be surprising as there are only a maximum of five entries per row in the system of equations. It can be explained with the large fill-in within the  $L$  and  $U$  matrices, see Figure 3.2. Although the matrix  $I - \theta\Delta_t F(n\Delta_t)$  with 1377 non zero entries is sparse, the  $L$  and  $U$  matrices suffer from a large degree of fill-in

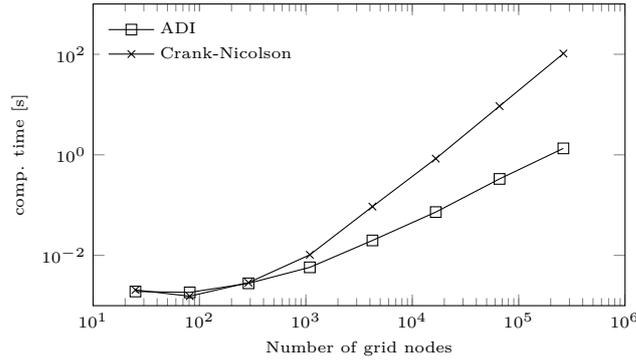


Figure 3.1: Number of grid points ( $N^2$ ) versus time consumption.

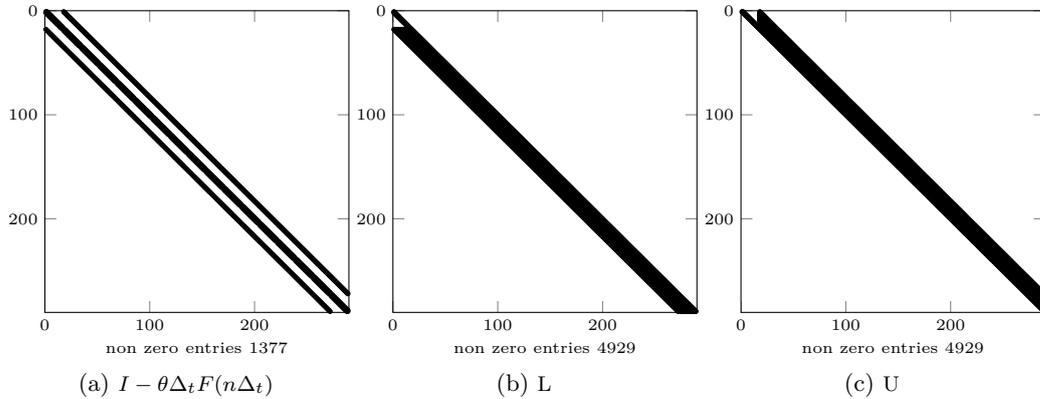


Figure 3.2: Fill-in of discretization matrices in the Crank-Nicolson scheme with  $N = 17$  nodes in each coordinate direction.

and have 4929 non zero entries.

This problem can be circumvented by applying Alternating Direction Implicit (ADI) schemes. Here, the spatial discretization matrix is decomposed into

$$F(t) = F_0(t) + F_1(t) + \dots + F_d(t),$$

where  $F_0$  stems from all mixed derivatives and  $F_i$  from each unidirectional contribution of coordinate direction  $i = 1, \dots, d$ . With the help of ADI time stepping the equation system can be solved as a sequence of one-dimensional problems, which significantly reduces the run-time compared to implicit Euler or Crank-Nicolson time marching. In the following we consider four well known ADI schemes.

*Douglas scheme (DO):*

$$\begin{cases} Y_0 &= U_n + \Delta_t F(t) U_n, \\ Y_i &= Y_{i-1} + \theta \Delta_t (F_i(t) Y_i - F_i(t) U_n) \text{ for } i = 1, \dots, d \\ U_{n+1} &= Y_d. \end{cases} \quad (3.1)$$

*Craig-Sneyd scheme (CS):*

$$\begin{cases} Y_0 &= U_n + \Delta_t F(t)U_n, \\ Y_i &= Y_{i-1} + \theta \Delta_t (F_i(t)Y_i - F_i(t)U_n) \text{ for } i = 1, \dots, d \\ \tilde{Y}_0 &= Y_0 + \frac{1}{2} \Delta_t (F_0 Y_d - F_0 U_n) \\ \tilde{Y}_i &= \tilde{Y}_{i-1} + \theta \Delta_t (F_i(t)\tilde{Y}_i - F_i(t)U_n) \text{ for } i = 1, \dots, d \\ U_{n+1} &= \tilde{Y}_d. \end{cases} \quad (3.2)$$

*Modified Craig-Sneyd scheme (MCS):*

$$\begin{cases} Y_0 &= U_n + \Delta_t F(t)U_n, \\ Y_i &= Y_{i-1} + \theta \Delta_t (F_i(t)Y_i - F_i(t)U_n) \text{ for } i = 1, \dots, d \\ \hat{Y}_0 &= Y_0 + \theta \Delta_t (F_0 Y_d - F_0 U_n) \\ \tilde{Y}_0 &= \hat{Y}_0 + (\frac{1}{2} - \theta) \Delta_t (F(t)Y_d - F(t)U_n) \\ \tilde{Y}_i &= \tilde{Y}_{i-1} + \theta \Delta_t (F_i(t)\tilde{Y}_i - F_i(t)U_n) \text{ for } i = 1, \dots, d \\ U_{n+1} &= \tilde{Y}_d. \end{cases} \quad (3.3)$$

*Hundsdorfer-Verwer scheme (HV):*

$$\begin{cases} Y_0 &= U_n + \Delta_t F(t)U_n, \\ Y_i &= Y_{i-1} + \theta \Delta_t (F_i(t)Y_i - F_i(t)U_n) \text{ for } i = 1, \dots, d \\ \tilde{Y}_0 &= Y_0 + \frac{1}{2} \Delta_t (F(t)Y_d - F(t)U_n) \\ \tilde{Y}_i &= \tilde{Y}_{i-1} + \theta \Delta_t (F_i(t)\tilde{Y}_i - F_i(t)Y_d) \text{ for } i = 1, \dots, d \\ U_{n+1} &= \tilde{Y}_d, \end{cases} \quad (3.4)$$

where  $\Delta_t$  is the step size in time,  $U_n \sim U(n\Delta_t)$  and  $\theta > 0$  is a real parameter. In the first scheme (3.1) an explicit Euler step is followed by a stabilizing correction step in each of the spatial directions. In the case of  $\theta = \frac{1}{2}$  the method is known as the Douglas [17] and Brian [5] scheme. The value  $\theta = 1$  has been considered by Douglas in [16]. If no mixed derivatives are present in the PDE and  $\theta = \frac{1}{2}$  the scheme is of order two in time and of order one otherwise. As mixed derivatives are quite common in finance PDEs this is a severe restriction. The Craig-Sneyd scheme [14] was originally introduced as an extension of the Douglas scheme, where a second explicit step is followed by  $d$  implicit stabilizing steps. It exhibits order two if  $\theta = \frac{1}{2}$ . The modified Craig-Sneyd scheme (3.3) can be seen as an extension of the iterated scheme in the article by Craig and Sneyd [14] and was defined by in't Hout et al. in [54]. This scheme has order two for arbitrary  $\theta > 0$ . The Hundsdorfer-Verwer scheme was derived in [48] and possesses like the modified Craig-Sneyd scheme order two for any  $\theta > 0$ .

Figure 3.1 shows the run-time of the second-order finite difference ADI CS scheme with  $\theta = \frac{1}{2}$  and  $\mathbf{N} = (N, N)$  grid nodes applied to the two-dimensional heat equation. Due to the decomposition of the discretization matrix  $N$  one-dimensional problems have to be solved in each implicit leg. Since the discretization matrices  $I - \theta \Delta_t F_1$  and  $I - \theta \Delta_t F_2$ , respectively, have only a maximum of three entries per row, each one-dimensional problem can be solved in linear run-time. Thus, we obtain a total computational effort proportional to  $N\mathcal{O}(N) = \mathcal{O}(N^2)$  or, in other words, the computational complexity rises linearly with the number of unknowns. The fillings of the matrices  $L_1 U_1 = I - \theta \Delta_t F_1$  and

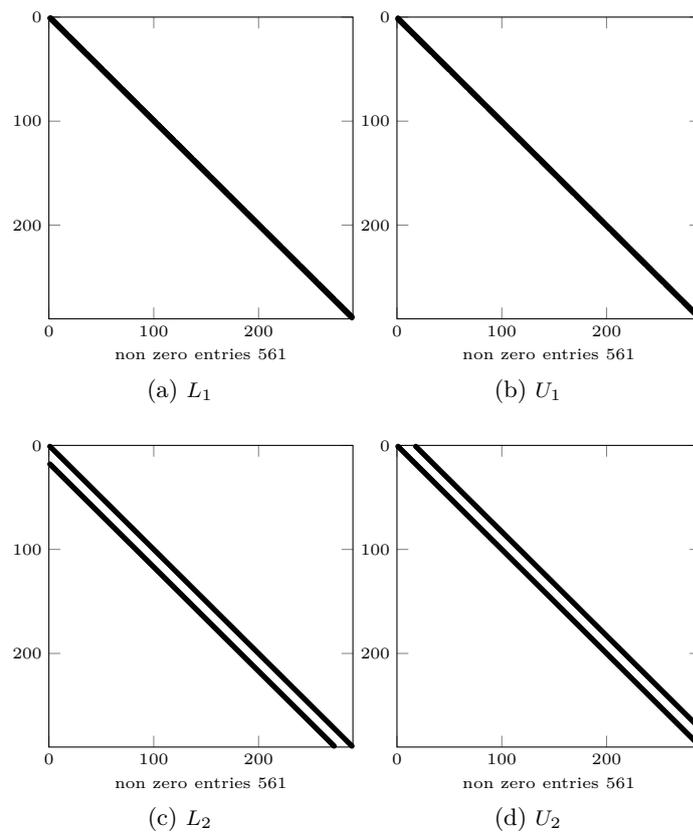


Figure 3.3: Discretization matrices in an ADI scheme with  $N = 17$  nodes in each coordinate direction.

$L_2U_2 = I - \theta\Delta_t F_2$  are exemplary given in Figure 3.3. In contrast to the CN scheme the number of non zero entries is low and each matrix has only up to two entries per row.

### 3.1 Stability Considerations

In the ADI schemes low  $\theta$  values in general lead to more accurate results, but might cause instabilities if chosen too small. Hence, the value has to be determined carefully to maintain a stable scheme and to get accurate results. Therefore, much effort has been spent on the derivation of bounds on  $\theta$  ensuring unconditional stability. The recent literature provides theoretical results for standard second-order finite differences applied to multi-dimensional diffusion and convection-diffusion equations with and without mixed derivatives [52, 53, 54, 55, 63, 71]. The results are derived in the *von Neumann* framework under the simplifying assumption of constant coefficients and periodic boundary conditions (BCs). We consider the general convection-diffusion equation with constant coefficients

$$\frac{\partial u}{\partial t} = \text{div}(A\nabla u) + c \cdot \nabla u \quad (3.5)$$

with  $A = (a_{ij})$ ,  $c = (c_1, c_2, \dots, c_d)^\top$  and supplemented with periodic BCs. The diffusion coefficient matrix  $A$  is assumed to be symmetric and positive semi-definite, which is in line with the parabolicity of the problem. Let the numerical scheme to solve this equation be given in the one-step form

$$U_{n+1} = RU_n,$$

where  $R$  denotes the iteration matrix. The iteration matrices are of the following form

$$R_{DO} = I + P^{-1}Z, \quad (3.6)$$

$$R_{CS} = I + P^{-1}Z + \frac{1}{2}P^{-1}Z_0P^{-1}Z, \quad (3.7)$$

$$R_{MCS} = I + P^{-1}Z + P^{-1}(\theta Z_0 + (\frac{1}{2} - \theta)Z)P^{-1}Z, \quad (3.8)$$

$$R_{HV} = (I + P^{-1}Z)^2 - P^{-1}(I + \frac{1}{2}Z)P^{-1}Z, \quad (3.9)$$

with  $P = \prod_{i=1}^d (I - \theta\Delta_t F_i)$ ,  $Z_0 = \Delta_t F_0$ ,  $Z = \Delta_t \sum_{i=0}^d F_i$ . In the case of standard central finite difference schemes the discretization matrices  $D_{FD_i}$ ,  $D_{FD_i}^2$  for  $i = 1, 2, \dots, d$  are normal and commuting. Due to the construction of the operators  $F_i$  for  $i = 0, 1, \dots, d$  via the Kronecker product of normal and commuting matrices the normality is sustained and the operators commute, see e.g. [64, Chapter 13]. Thus, they can be written in a more condensed form

$$R_{DO} = I + P^{-1}Z, \quad (3.10)$$

$$R_{CS} = I + P^{-1}Z + \frac{1}{2}P^{-2}Z_0Z, \quad (3.11)$$

$$R_{MCS} = I + P^{-1}Z + \theta P^{-2}Z_0Z + (\frac{1}{2} - \theta)P^{-2}Z^2, \quad (3.12)$$

$$R_{HV} = I + 2P^{-1}Z - P^{-2}Z + \frac{1}{2}P^{-2}Z^2. \quad (3.13)$$

Furthermore, due to the normality the discretization matrix  $F$  is simultaneously diagonalizable  $F = V\Lambda V^{-1}$  with unitary matrix of eigenvectors  $V$  and diagonal matrix  $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_d)$ , where  $\lambda_i$  is the eigenvalue of operator  $F_i$ . The stability can then be

analyzed by considering the linear scalar test equation

$$U'(t) = (\lambda_0 + \lambda_1 + \dots + \lambda_d) U(t), \quad (3.14)$$

with complex values  $\lambda_i$  for  $i = 0, 1, \dots, d$ . This is equivalent to the *von Neumann* stability analysis. For a detailed introduction and discussion we refer to [49, Chapter 1], [70, Chapter 1]. Then the iteration matrix  $R$  reduces to the scalar factor  $r(z_0, z_1, \dots, z_d)$  with  $z_i = \Delta_t \lambda_i$  for  $i = 0, 1, \dots, d$  and the numerical scheme is stable iff

$$|r| \leq 1$$

is fulfilled. Defining

$$z = z_0 + z_1 + \dots + z_d, \quad p = (1 - \theta z_1) \cdot (1 - \theta z_2) \cdot \dots \cdot (1 - \theta z_d)$$

the scalar functions  $r$  for the ADI schemes (3.1)–(3.4) are given by

$$r_{DO}(z_0, z_1, \dots, z_d) = 1 + \frac{z}{p}, \quad (3.15)$$

$$r_{CS}(z_0, z_1, \dots, z_d) = 1 + \frac{z}{p} + \frac{1}{2} \frac{z_0 z}{p^2}, \quad (3.16)$$

$$r_{MCS}(z_0, z_1, \dots, z_d) = 1 + \frac{z}{p} + \theta \frac{z_0 z}{p^2} + \left(\frac{1}{2} - \theta\right) \frac{z^2}{p^2}, \quad (3.17)$$

$$r_{HV}(z_0, z_1, \dots, z_d) = 1 + 2\frac{z}{p} - \frac{z}{p^2} + \frac{1}{2} \frac{z^2}{p^2}. \quad (3.18)$$

Each  $z_i$  can be derived by inserting Fourier modes into the discretization of the derivatives, see [53].

Conditions on  $\theta$  ensuring unconditional stability have been derived in the case of pure diffusion with two and three spatial dimensions [53, 54]:

$$\begin{aligned} \text{DO:} & \quad \theta \geq \frac{1}{2} \quad (d=2), \theta \geq \max \left\{ \frac{1}{2}, \frac{2(2\gamma+1)}{9} \right\} \quad (d=3), \\ \text{CS:} & \quad \theta \geq \frac{1}{2} \quad (d=2, 3) \\ \text{MCS:} & \quad \theta \geq \max \left\{ \frac{1}{4}, \frac{\gamma+1}{6} \right\} \quad (d=2), \theta \geq \max \left\{ \frac{1}{4}, \frac{2}{13}(2\gamma+1) \right\} \quad (d=3), \\ \text{HV:} & \quad \theta \geq \max \left\{ \frac{1}{4}, \frac{\gamma+1}{4+2\sqrt{2}} \right\} \quad (d=2), \theta \geq \max \left\{ \frac{1}{4}, \frac{2\gamma+1}{4+2\sqrt{3}} \right\} \quad (d=3), \end{aligned}$$

for a parameter  $\gamma \in [0, 1]$ , which describes the relative size of the mixed derivative coefficient

$$|a_{ij}| \leq \gamma \sqrt{a_{ii} a_{jj}} \quad \text{for all } i \neq j. \quad (3.19)$$

In [53] necessary lower bounds for higher dimensional problems were given. In case of the HV scheme it has been proven in [54] that these conditions are sufficient. For the other three schemes it is not clear if the lower bounds are sufficient.

$$\begin{aligned} \text{DO: } \theta & \geq \max \left\{ \frac{1}{2}, \frac{1}{2} \left(1 - \frac{1}{d}\right)^{d-1} ((d-1)\gamma + 1) \right\}, & \text{CS: } \theta & \geq \max \left\{ \frac{1}{2}, \frac{1}{2} \left(1 - \frac{1}{d}\right)^d d\gamma \right\}, \\ \text{MCS: } \theta & \geq \max \left\{ \frac{1}{4}, \frac{1}{2} \frac{(d-1)\gamma + 1}{1 + \left(1 + \frac{1}{d-1}\right)^{d-1}} \right\}, & \text{HV: } \theta & \geq \max \left\{ \frac{1}{4}, \frac{1}{2} a_d ((d-1)\gamma + 1) \right\}, \end{aligned}$$

where  $a_d \in (0, \frac{1}{2})$  is the unique solution of the equation  $2a_d(1 + \frac{1-a_d}{d-1})^{d-1} - 1 = 0$ .

In [52, 55, 63] the stability of these four schemes is analyzed for convection-diffusion equations. In these works the following conditions for two-dimensional problems have been derived:

Assume the conditions

$$\operatorname{Re}(z_1) \leq 0, \quad \operatorname{Re}(z_2) \leq 0 \quad \text{and} \quad |z_0| \leq 2\sqrt{\operatorname{Re}(z_1) \cdot \operatorname{Re}(z_2)} \quad (3.20)$$

hold for  $z_0, z_1, z_2 \in \mathbb{C}$  and  $\theta \geq \frac{1}{2}$ , then  $|r_{DO}| \leq 1$  and  $|r_{CS}| \leq 1$  holds. In 't Hout and Mishra [52] showed for  $F_0 = 0$  that for the modified Craig-Sneyd scheme it holds  $|r_{MCS}(0, z_1, z_2)| \leq 1$  for all  $z_1, z_2 \in \mathbb{C}$  with negative real part  $\operatorname{Re}(z_1) \leq 0, \operatorname{Re}(z_2) \leq 0$  if and only if  $\theta \geq \frac{1}{4}$ . For  $z_0 \in \mathbb{R}$  and  $z_1, z_2 \in \mathbb{C}$  fulfilling (3.20), they derived the necessary stability condition  $\theta \geq \frac{2}{5}$ . However, the scheme has been applied successfully to convection-diffusion equations with mixed derivatives in [51] even for  $\theta \geq \frac{1}{3}$ . An experimental analysis of this observation can be found in [52]. Recently it has been shown by Mishra [71] that the MCS scheme is unconditionally stable for  $\theta > \frac{1}{3}$  if the modulus of the mixed derivative coefficient is at most  $\frac{2+\sqrt{10}}{6}$ .

In [38] the stability for convection-diffusion problems with three spatial dimensions was analyzed experimentally. The bounds derived for pure diffusion equations turned out to lead to a stable behavior in case of the DO, CS and MCS scheme. For the HV scheme with  $\theta = \frac{1}{2} + \frac{1}{6}\sqrt{3}$  the error decayed monotonically with  $\Delta_t$ . This  $\theta$ -value was derived for two dimensional convection-diffusion equations without mixed derivatives in [63].

## 3.2 High-Order Finite Difference ADI Schemes

In this section we combine high-order compact finite differences with ADI time stepping. This approach goes back to Mitchell and Fairweather [72], who considered problems without mixed derivatives. Karaa and Zhang [58] revisited this idea and solved convection-diffusion equations with a D'Yakonov splitting scheme. Düring et al. [23] then derived high-order (HO-)ADI methods for convection-diffusion equations with mixed derivatives. Recently their work was extended in [43], where multi-dimensional diffusion equations in a sparse grid setting are solved, and by Düring and Miles [25], who considered an HO-ADI discretization for stochastic volatility models.

According to Section 2.2 we can discretize the one-dimensional problem (2.8) with fourth-order accuracy on the compact stencil. In matrix notation this leads to a system of the form (2.14)

$$A_{x_i}U = B_{x_i}G \quad \text{for } i = 1, 2, \dots, d.$$

Inserting this into the semi-discrete formulation one obtains

$$U'(t) = F_0(t)U + B_{x_1}^{-1}A_{x_1}U + B_{x_2}^{-1}A_{x_2}U + \dots + B_{x_d}^{-1}A_{x_d}U.$$

The mixed derivative can be approximated via standard fourth-order stencils. Application of the ADI time stepping yields

*HO Douglas scheme (HDO):*

$$\begin{cases} Y_0 & = U_n + \Delta_t (F_0(U_n) + B_{x_1}^{-1} A_{x_1} U_n + \dots + B_{x_d}^{-1} A_{x_d} U_n), \\ (B_{x_i} - \theta \Delta_t A_{x_i}) Y_i & = B_{x_i} Y_{i-1} - \theta \Delta_t A_{x_i} U_n \text{ for } i = 1, \dots, d \\ U_{n+1} & = Y_d. \end{cases} \quad (3.21)$$

*HO Craig-Sneyd scheme (HCS):*

$$\begin{cases} Y_0 & = U_n + \Delta_t (F_0(U_n) + B_{x_1}^{-1} A_{x_1} U_n + \dots + B_{x_d}^{-1} A_{x_d} U_n), \\ (B_{x_i} - \theta \Delta_t A_{x_i}) Y_i & = B_{x_i} Y_{i-1} - \theta \Delta_t A_{x_i} U_n \text{ for } i = 1, \dots, d \\ \tilde{Y}_0 & = Y_0 + \frac{1}{2} \Delta_t (F_0(Y_d) - F_0(U_n)) \\ (B_{x_i} - \theta \Delta_t A_{x_i}) \tilde{Y}_i & = B_{x_i} \tilde{Y}_{i-1} - \theta \Delta_t A_{x_i} U_n \text{ for } i = 1, \dots, d \\ U_{n+1} & = \tilde{Y}_d. \end{cases} \quad (3.22)$$

*HO modified Craig-Sneyd scheme (HMCS):*

$$\begin{cases} Y_0 & = U_n + \Delta_t (F_0(U_n) + B_{x_1}^{-1} A_{x_1} U_n + \dots + B_{x_d}^{-1} A_{x_d} U_n), \\ (B_{x_i} - \theta \Delta_t A_{x_i}) Y_i & = B_{x_i} Y_{i-1} - \theta \Delta_t A_{x_i} U_n \text{ for } i = 1, \dots, d \\ \hat{Y}_0 & = Y_0 + \theta \Delta_t (F_0(Y_d) - F_0(U_n)) \\ \tilde{Y}_0 & = \hat{Y}_0 + (\frac{1}{2} - \theta) \Delta_t (F(Y_d) - F(U_n)) \\ (B_{x_i} - \theta \Delta_t A_{x_i}) \tilde{Y}_i & = B_{x_i} \tilde{Y}_{i-1} - \theta \Delta_t A_{x_i} U_n \text{ for } i = 1, \dots, d \\ U_{n+1} & = \tilde{Y}_d. \end{cases} \quad (3.23)$$

*HO Hundsdorfer-Verwer scheme (HHV):*

$$\begin{cases} Y_0 & = U_n + \Delta_t (F_0(U_n) + B_{x_1}^{-1} A_{x_1} U_n + \dots + B_{x_d}^{-1} A_{x_d} U_n), \\ (B_{x_i} - \theta \Delta_t A_{x_i}) Y_i & = B_{x_i} Y_{i-1} - \theta \Delta_t A_{x_i} U_n \text{ for } i = 1, \dots, d \\ \tilde{Y}_0 & = Y_0 + \frac{1}{2} \Delta_t (F(Y_d) - F(U_n)) \\ (B_{x_i} - \theta \Delta_t A_{x_i}) \tilde{Y}_i & = B_{x_i} \tilde{Y}_{i-1} - \theta \Delta_t A_{x_i} Y_d \text{ for } i = 1, \dots, d \\ U_{n+1} & = \tilde{Y}_d. \end{cases} \quad (3.24)$$

In order to avoid the inverse of the operators  $B_{x_i}$  for  $i = 1, \dots, d$  one can rewrite the schemes. To do this, we introduce new artificial variables. The variables are defined as

$$\begin{aligned} Z_i &:= \prod_{j=i+1}^d B_{x_j} Y_i, \\ \tilde{Z}_i &:= \prod_{j=i+1}^d B_{x_j} \tilde{Y}_i, \\ \hat{Z}_0 &:= \prod_{j=1}^d B_{x_j} \hat{Y}_0, \end{aligned}$$

for  $i = 0, 1, \dots, d$ .

*HO Douglas scheme (HDO):*

$$\begin{cases} Z_0 & = \prod_{j=1}^d B_{x_j} U_n + \Delta t \left( \prod_{j=1}^d B_{x_j} F_0(U_n) + \sum_{i=1}^d \prod_{j=1, j \neq i}^d B_{x_j} A_{x_i} U_n \right) \\ (B_{x_i} - \theta \Delta t A_{x_i}) Z_i & = Z_{i-1} - \theta \Delta t \prod_{j=i+1}^d B_{x_j} A_{x_i} U_n \text{ for } i = 1, \dots, d \\ U_{n+1} & = Z_d, \end{cases} \quad (3.25)$$

*HO Craig-Sneyd scheme (HCS):*

$$\begin{cases} Z_0 & = \prod_{j=1}^d B_{x_j} U_n + \Delta t \left( \prod_{j=1}^d B_{x_j} F_0(U_n) + \sum_{i=1}^d \prod_{j=1, j \neq i}^d B_{x_j} A_{x_i} U_n \right) \\ (B_{x_i} - \theta \Delta t A_{x_i}) Z_i & = Z_{i-1} - \theta \Delta t \prod_{j=i+1}^d B_{x_j} A_{x_i} U_n \text{ for } i = 1, \dots, d \\ \tilde{Z}_0 & = Z_0 + \frac{1}{2} \Delta t \left( \prod_{j=1}^d B_{x_j} F_0(Z_d) - \prod_{j=1}^d B_{x_j} F_0(U_n) \right) \\ (B_{x_i} - \theta \Delta t A_{x_i}) \tilde{Z}_i & = \tilde{Z}_{i-1} - \theta \Delta t \prod_{j=i+1}^d B_{x_j} A_{x_i} U_n \text{ for } i = 1, \dots, d \\ U_{n+1} & = \tilde{Z}_d. \end{cases} \quad (3.26)$$

*HO modified Craig-Sneyd scheme (HMCS):*

$$\begin{cases} Z_0 & = \prod_{j=1}^d B_{x_j} U_n + \Delta t \left( \prod_{j=1}^d B_{x_j} F_0(U_n) + \sum_{i=1}^d \prod_{j=1, j \neq i}^d B_{x_j} A_{x_i} U_n \right) \\ (B_{x_i} - \theta \Delta t A_{x_i}) Z_i & = Z_{i-1} - \theta \Delta t \prod_{j=i+1}^d B_{x_j} A_{x_i} U_n \text{ for } i = 1, \dots, d \\ \hat{Z}_0 & = Z_0 + \theta \Delta t \left( \prod_{j=1}^d B_{x_j} F_0(Z_d) - \prod_{j=1}^d B_{x_j} F_0(U_n) \right) \\ \tilde{Z}_0 & = \hat{Z}_0 + \left( \frac{1}{2} - \theta \right) \Delta t \left( \prod_{j=1}^d B_{x_j} F(Z_d) - \prod_{j=1}^d B_{x_j} F(U_n) \right) \\ (B_{x_i} - \theta \Delta t A_{x_i}) \tilde{Z}_i & = \tilde{Z}_{i-1} - \theta \Delta t \prod_{j=i+1}^d B_{x_j} A_{x_i} U_n \text{ for } i = 1, \dots, d \\ U_{n+1} & = \tilde{Z}_d. \end{cases} \quad (3.27)$$

*HO Hundsdorfer-Verwer scheme (HHV):*

$$\begin{cases} Z_0 & = \prod_{j=1}^d B_{x_j} U_n + \Delta t \left( \prod_{j=1}^d B_{x_j} F_0(U_n) + \sum_{i=1}^d \prod_{j=1, j \neq i}^d B_{x_j} A_{x_i} U_n \right) \\ (B_{x_i} - \theta \Delta t A_{x_i}) Z_i & = Z_{i-1} - \theta \Delta t \prod_{j=i+1}^d A_{x_i} U_n \text{ for } i = 1, \dots, d \\ \tilde{Z}_0 & = Z_0 + \frac{1}{2} \Delta t \left( \prod_{j=1}^d B_{x_j} F(Z_d) - \prod_{j=1}^d B_{x_j} F(U_n) \right) \\ (B_{x_i} - \theta \Delta t A_{x_i}) \tilde{Z}_i & = \tilde{Z}_{i-1} - \theta \Delta t \prod_{j=i+1}^d B_{x_j} A_{x_i} Z_d \text{ for } i = 1, \dots, d \\ U_{n+1} & = \tilde{Z}_d. \end{cases} \quad (3.28)$$

Please note that this formulation is slightly different to the ones given in the literature, e.g. in [23, 25], where the HOC discretization is applied to the implicit legs and standard fourth-order discretizations are used in the explicit legs.

### 3.3 Stability of HO-ADI Schemes

We investigate the stability of the HO-ADI schemes (3.21)–(3.24) within the von Neumann framework. Therefore, we consider the general convection-diffusion PDE (3.5)

$$\frac{\partial u}{\partial t} = \operatorname{div}(A\nabla u) + c \cdot \nabla u$$

with constant coefficients and supplemented with periodic BCs. Please note that this causes all derivatives of the coefficients in our scheme to vanish. In order to be able to investigate the stability with the help of the linear test equation (3.14) to prove stability the operators  $F_i$  need to be normal and commuting for  $i = 0, 1, \dots, d$ . Due to the periodic BCs the discretization matrices  $D_{FD_i}$ ,  $D_{FD_i}^2$ ,  $\tilde{D}_{FD_i}$  for  $i = 1, 2, \dots, d$  are circulant and commuting. Thus, the HOC matrices  $B_{x_i}$ ,  $A_{x_i}$  are as a Kronecker product normal and commute with each other. Therefore, they can be diagonalized simultaneously by an unitary matrix.

#### 3.3.1 Stability in 2 or 3 Dimensions for Diffusion Equations

First, we consider diffusion equations and let  $c = (0, 0, \dots)^\top$ . The matrices in (2.14) are then given by

$$\begin{aligned} A_{x_i} &= a_{ii} I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d}, \\ B_{x_i} &= I_{N_1 \cdot N_2 \cdot \dots \cdot N_d} + \frac{h_i^2}{12} I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \end{aligned}$$

for  $i = 1, 2, \dots, d$ . Inserting Fourier modes into each discretization operator, we obtain the eigenvalues

$$\begin{aligned} \tilde{z}_i &= -2a_{ii} \frac{1}{h_i^2} (1 - \cos \phi_i) && \text{for } i = 1, \dots, d, \\ \bar{z}_i &= 1 - \frac{1}{6} (1 - \cos \phi_i) && \text{for } i = 1, \dots, d, \\ z_0 &= - \sum_{i \neq j} a_{ij} \frac{4}{144} \frac{\Delta_t}{h_i h_j} (8 \sin \phi_i - \sin 2\phi_i) (8 \sin \phi_j - \sin 2\phi_j) && . \end{aligned} \quad (3.29)$$

The eigenvalues  $\tilde{z}_i$  stem from  $a_{ii}\delta_i^2$ ,  $A_{x_i}$  respectively,  $\bar{z}_i$  from  $1 + \frac{h_i^2}{12}\delta_i^2$ ,  $B_{x_i}$  respectively, and  $z_0$  from all mixed derivatives. The angles  $\phi_i$  are integer multiples of  $2\pi/m_i$ , where  $m_i$  denotes the dimension of the grid in  $x_i$ -direction for  $i = 1, \dots, d$ . Hence we have the scalar factor

$$\begin{aligned} p &= \prod_i (1 - \theta \Delta_t \tilde{z}_i / \bar{z}_i) \\ z &= z_0 + \Delta_t \sum_i \frac{\tilde{z}_i}{\bar{z}_i}. \end{aligned}$$

Defining

$$z_i := \Delta_t \tilde{z}_i / \bar{z}_i, \quad (3.30)$$

we obtain

$$p = \prod_i (1 - \theta z_i),$$

$$z = z_0 + z_1 + \dots + z_d.$$

By introduction of the quantities

$$y_i := \frac{1}{6} \frac{\sqrt{\Delta t}}{h_i} (8 \sin \phi_i - \sin 2\phi_i),$$

$$q_i := (1 - c_i)(1 - 3c_i)(1 + 2c_i)^2,$$

with  $c_i = \frac{1}{6}(1 - \cos \phi_i)$  for  $i = 1, \dots, d$ , we can rewrite the eigenvalues to

$$z_0 = - \sum_{i \neq j} a_{ij} y_i y_j \tag{3.31}$$

$$z_i = - \frac{a_{ii}}{q_i} y_i^2 = -a_{ii} y_i^2 - \delta_i \tag{3.32}$$

for  $i = 1, 2, \dots, d$  with  $\delta_i = -z_i(1 - q_i)$ . One directly observes  $0 \leq q_i \leq 1$ ,  $z_i \leq 0$  and  $\delta_i \geq 0$  for  $i = 1, 2, \dots, d$ .

The stability matrices reduce to the stability functions

$$r_{HDO}(z_0, z_1, \dots, z_d) = 1 + \frac{z}{p},$$

$$r_{HCS}(z_0, z_1, \dots, z_d) = 1 + \frac{z}{p} + \frac{1}{2} \frac{z_0 z}{p^2},$$

$$r_{HMCS}(z_0, z_1, \dots, z_d) = 1 + \frac{z}{p} + \theta \frac{z_0 z}{p^2} + \left(\frac{1}{2} - \theta\right) \frac{z^2}{p^2},$$

$$r_{HHV}(z_0, z_1, \dots, z_d) = 1 + 2\frac{z}{p} - \frac{z}{p^2} + \frac{1}{2} \frac{z^2}{p^2}.$$

Note that the stability functions of the HO-ADI schemes have the same structure as for the 'standard' ADI schemes (3.15)–(3.18). This gives rise to the assumption that both approaches share similar stability properties. Nevertheless, the eigenvalues  $z_i$  for  $i = 0, 1, \dots, d$  stem from high-order discretizations compared to the standard central second-order discretizations in the literature. In the following we want to use the results from the literature as far as applicable, cf. [53, 54]. Therefore, we formulate a Lemma which states properties regarding the eigenvalues. We will use these results later in the proofs.

**Lemma 23.** *Let  $z_0, z_1, \dots, z_d$  be given by (3.31), (3.32), respectively. Further let  $A$  be symmetric positive semi-definite, then*

$$\text{all } z_i \text{ are real,} \tag{3.33}$$

$$z_i \leq 0 \text{ for } i=1, \dots, d \tag{3.34}$$

$$z \leq 0, \tag{3.35}$$

$$|z_0| \leq \sum_{i \neq j} \sqrt{z_i z_j}. \tag{3.36}$$

*Proof.* The properties (3.33), (3.34) are clear. It remains to prove (3.35), (3.36). We first

show  $z \leq 0$ . It holds

$$\begin{aligned} z &= z_0 + z_1 + \dots + z_d = - \sum_{i \neq j} a_{ij} y_i y_j + \sum_{i=1}^d (-a_{ii} y_i^2 - \delta_i), \\ &= -y^\top A y - \sum_{i=1}^d \delta_i, \end{aligned}$$

with  $y = (y_1, \dots, y_d)^\top$ . Due to the positive semi-definiteness of the symmetric coefficient matrix  $A$  it follows  $-y^\top A y \leq 0$ . Due to  $z_i \leq 0$  and  $q_i \leq 1$  it follows  $\delta_i \geq 0$ , which directly proves the desired inequality  $z \leq 0$ .

In the following we prove condition (3.36).

By the positive semi-definiteness of  $A$  it follows  $|a_{ij}| \leq \sqrt{a_{ii} a_{jj}}$ . Hence, we can conclude

$$|z_0| \leq \sum_{i \neq j} |a_{ij}| |y_i| |y_j| \leq \sum_{i \neq j} \sqrt{a_{ii} a_{jj}} |y_i| |y_j| \leq \sum_{i \neq j} \sqrt{z_i z_j}.$$

The last inequality follows due to  $\delta_i \geq 0$  for  $i = 1, 2, \dots, d$ .  $\square$

Theorem 2.3 in [53] states the parameter values of  $\theta$ , so that the ADI schemes (3.1)–(3.4) using second-order central finite difference stencils are unconditionally stable, when applied to the PDE (3.5) without convection term. Since only the conditions in Lemma 23 are used in the proof, the same stability conditions also hold for the HO-ADI schemes.

**Theorem 24.** *Consider the equation (3.5) with  $c = (0, 0, \dots, 0)^\top$ , periodic BCs and symmetric positive semi-definite coefficient matrix  $A$  in two or three spatial dimensions. Then the HO-ADI schemes (3.21)–(3.24) are unconditionally stable with the following lower bound on  $\theta$ :*

HO Douglas scheme (3.21)

$$\theta \geq \frac{1}{2} \text{ if } d = 2 \qquad \theta \geq \frac{2}{3} \text{ if } d = 3$$

HO Craig-Sneyd scheme (3.22)

$$\theta \geq \frac{1}{2} \text{ if } d = 2, 3$$

HO modified Craig-Sneyd scheme (3.23)

$$\theta \geq \frac{1}{3} \text{ if } d = 2 \qquad \theta \geq \frac{6}{13} \text{ if } d = 3$$

HO Hundsdorfer-Verwer scheme (3.24)

$$\theta \geq \frac{1}{2 + \sqrt{2}} \text{ if } d = 2 \qquad \theta \geq \frac{3}{4 + 2\sqrt{3}} \text{ if } d = 3$$

*Proof.* As  $A$  is symmetric positive semi-definite, Lemma 23 concludes that the inequalities (3.33)–(3.36) hold. The proof directly follows the analogue steps as in [53, Theorem 2.3],

since only conditions (3.33)–(3.36) are used there.  $\square$

### 3.3.2 Stability in Arbitrary Dimensions for Diffusion Equations

In this section we want to derive necessary conditions on  $\theta$  for the HO-ADI schemes. The stability condition  $|r| \leq 1$  can be rewritten to:

*HO Douglas scheme (3.21)*

$$2p + z \geq 0, \quad z \leq 0, \quad (3.37)$$

*HO Craig-Sneyd scheme (3.22)*

$$p + \frac{1}{2}z_0 \geq 0, \quad 2p^2 + pz + \frac{1}{2}z_0 z \geq 0, \quad (3.38)$$

*HO modified Craig-Sneyd scheme (3.23)*

$$p - \theta(z - z_0) + \frac{1}{2}z \geq 0, \quad 2p + pz + \theta z_0 z + \left(\frac{1}{2} - \theta\right)z^2 \geq 0, \quad (3.39)$$

*HO Hundsdorfer-Verwer scheme (3.24)*

$$2p - 1 + \frac{1}{2}z \geq 0, \quad 2p^2 + (2p - 1)z + \frac{1}{2}z^2 \geq 0. \quad (3.40)$$

**Theorem 25.** *Let  $d \geq 2$ . Then the HO-ADI schemes (3.21)–(3.24) applied to equations (3.5) with  $c = (0, 0, \dots, 0)^\top$ , symmetric positive semi-definite coefficient matrix  $A$  and periodic BCs, need to fulfill the following lower bound on  $\theta$  for unconditional stability:*

*HO Douglas scheme (3.21)*

$$\theta \geq \frac{1}{2}d\left(1 - \frac{1}{d}\right)^{d-1},$$

*HO Craig-Sneyd scheme (3.22)*

$$\theta \geq \max \left\{ \frac{1}{2}, \frac{1}{2}d\left(1 - \frac{1}{d}\right)^d \right\},$$

*HO modified Craig-Sneyd scheme (3.23)*

$$\theta \geq \frac{1}{2} \frac{d}{1 + \left(\frac{d}{d-1}\right)^{d-1}},$$

*HO Hundsdorfer-Verwer scheme (3.24)*

$$\theta \geq \frac{1}{2}da_k,$$

where  $a_k$  is the unique solution  $a \in (0, \frac{1}{2})$  of  $2a \left(1 + \frac{1-a}{d-1}\right)^{d-1} - 1 = 0$ .

*Proof.* We consider a coefficient matrix  $A$  with  $a_{ij} = 1$  for  $1 \leq i, j \leq d$ . Note, that for this choice  $A$  is positive semi-definite. In the following we assume equal step sizes in all coordinate directions  $h = h_1 = \dots = h_d$  and choose equal angles  $\phi = \phi_1 = \dots = \phi_d$  for all

$z_i$ . Hence, the eigenvalues are given by

$$\begin{aligned} z - z_0 &= -d \frac{1}{q} y^2, \\ z_0 &= -d(d-1)y^2, \end{aligned}$$

with  $q = q_1 = \dots = q_d$  and  $y = y_1 = \dots = y_d$ . The stability conditions (3.37)–(3.40) yield

$$(3.37) : \quad \theta \geq -\theta \frac{z}{2p} = \frac{\theta}{2} \frac{d(d-1)y^2 + d \frac{1}{q} y^2}{(1 + \theta \frac{1}{q} y^2)^d}$$

$$(3.38) : \quad \theta \geq -\theta \frac{z_0}{2p} = \frac{\theta}{2} \frac{d(d-1)y^2}{(1 + \theta \frac{1}{q} y^2)^d}$$

$$(3.39) : \quad \theta \geq -\frac{1}{2} \theta \frac{z}{p - \theta(z - z_0)} = \frac{\theta}{2} \frac{d(d-1)y^2 + d \frac{1}{q} y^2}{(1 + \theta \frac{1}{q} y^2)^d + \theta d \frac{1}{q} y^2}$$

$$(3.40) : \quad \theta \geq -\frac{1}{2} \theta \frac{z}{2p - 1} = \frac{\theta}{2} \frac{d(d-1)y^2 + d \frac{1}{q} y^2}{2(1 + \theta \frac{1}{q} y^2)^d - 1}.$$

Defining  $\alpha := \theta \frac{1}{q} y^2$ , we obtain

$$(3.37) : \quad \theta \geq \frac{\alpha d (d-1)q + 1}{2 (1 + \alpha)^d}$$

$$(3.38) : \quad \theta \geq \frac{\alpha d (d-1)q}{2 (1 + \alpha)^d}$$

$$(3.39) : \quad \theta \geq \frac{\alpha d (d-1)q + 1}{2 (1 + \alpha)^d + \alpha d}$$

$$(3.40) : \quad \theta \geq \frac{\alpha d (d-1)q + 1}{2 (2(1 + \alpha)^d - 1)}.$$

Taking the supremum of  $q(\phi)$  on the interval  $\phi \in (0, 2\pi)$ , one obtains

$$(3.37) : \quad \theta \geq \frac{\alpha d^2}{2(1 + \alpha)^d}$$

$$(3.38) : \quad \theta \geq \frac{\alpha d(d-1)}{(1 + \alpha)^d}$$

$$(3.39) : \quad \theta \geq \frac{d}{2} \frac{\alpha d}{(1 + \alpha)^d + \alpha d}$$

$$(3.40) : \quad \theta \geq \frac{d}{2} \frac{\alpha d}{2(1 + \alpha)^d - 1}.$$

Please note that we have taken the supremum on the open interval since  $q$  and  $\alpha$  both depend on  $\phi$  and are therefore not independent on the closed interval ( $\alpha = 0$  for  $\phi = 0$ ). Maximization regarding the parameter  $\alpha > 0$  completes the proof.

In case of the HO Craig-Sneyd scheme we also consider the case  $\phi_i = 0$  for all  $i > 1$ , such that  $z_2 = \dots = z_d = 0$  and  $z_0 = 0$ . The stability criterion (3.38) reduces to

$$2(1 - \theta z_1)^2 + (1 - \theta z_1) z_1 \geq 0 \quad \Leftrightarrow \quad 2 + (1 - 2\theta) z_1 \geq 0,$$

which leads to the lower bound  $\theta \geq \frac{1}{2}$ .  $\square$

The necessary condition on  $\theta$  coincides with the sufficient condition in two or three spatial dimensions. From the inequalities (3.39), (3.40) and  $z_i, z \leq 0$  for  $i = 1, \dots, d$  follows  $r_{HHV} \leq r_{HMCS}$  since  $2p - 1 + \frac{1}{2}z = p + \prod(1 - \theta z_i) - 1 + \frac{1}{2}z \geq p - \theta \sum z_i + \frac{1}{2}z = p - \theta(z - z_0) + \frac{1}{2}z$ . Thus, the necessary condition on  $\theta$  for unconditional stability is weaker for the HHV than for the HMCS scheme.

### 3.3.3 Stability for Convection-Diffusion Equations

In the case of convection-diffusion equations the condition (3.20) yields stability results for the four introduced ADI schemes in a two-dimensional setting. In the following we prove that the conditions (3.20) hold. In the higher dimensional case stability for problems with strong convection-dominance cannot be guaranteed. Therefore, we investigate the stability behavior experimentally for problems with different magnitudes of convection in three and four spatial dimensions.

The discretization matrices are given by

$$\begin{aligned} A_{x_i} &= \left( a_{ii} + \frac{h_i^2 c_i^2}{12 a_{ii}} \right) I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \\ &\quad + c_i I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i} \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d}, \\ B_{x_i} &= I_{N_1 \cdot N_2 \cdot \dots \cdot N_d} + \frac{h_i^2}{12} I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \\ &\quad + \frac{h_i^2}{12} \frac{c_i}{a_{ii}} I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i} \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \end{aligned}$$

for  $i = 1, 2, \dots, d$ . Inserting Fourier modes into the discretization operators we obtain the eigenvalues

$$\begin{aligned} \tilde{z}_i &= 2 \left( a_{ii} + \frac{h_i^2 c_i^2}{12 a_{ii}} \right) \frac{1}{h_i^2} (\cos \phi_i - 1) + c_i \frac{1}{h_i} \sqrt{-1} \sin \phi_i && \text{for } i = 1, \dots, d, \\ \bar{z}_i &= 1 - \frac{1}{6} (1 - \cos \phi_i) + \frac{c_i}{12 a_{ii}} h_i \sqrt{-1} \sin \phi_i && \text{for } i = 1, \dots, d, \\ z_0 &= - \sum_{i \neq j} a_{ij} \frac{1}{36} \frac{\Delta_t}{h_i h_j} (8 \sin \phi_i - \sin 2\phi_i) (8 \sin \phi_j - \sin 2\phi_j). \end{aligned} \quad (3.41)$$

The eigenvalues  $\tilde{z}_i$  stem from  $A_{x_i}$  whereas  $\bar{z}_i$  from  $B_{x_i}$  and  $z_0$  from all cross derivatives. The angles  $\phi_i$  are integer multiples of  $2\pi/m_i$  with  $m_i$  being the dimension of the grid in  $x_i$ -direction for  $i = 1, \dots, d$ . Similar to the pure diffusion case, we obtain the scalar stability functions (3.15)–(3.18) with

$$z_i = \Delta_t \tilde{z}_i / \bar{z}_i. \quad (3.42)$$

Thus, the proof of stability reduces to the analysis of the eigenvalues. Lemma 26 ensures the unconditional stability of the HDO (3.21), HCS (3.22) schemes and the necessary condition on the lower bound for  $\theta$  in the HMCS scheme (3.23). For the HHV scheme (3.24) this leads to unconditional stability if  $z_0 = 0$ .

**Lemma 26.** *Let  $d = 2$  and HO-ADI schemes (3.21)–(3.24) be applied to the convection-diffusion problem (3.5) with symmetric positive semi-definite coefficient matrix  $A$ . Then*

it holds for the eigenvalues, defined according to (3.41), (3.42), respectively,

$$\operatorname{Re}(z_1) \leq 0, \quad \operatorname{Re}(z_2) \leq 0 \quad \text{and} \quad |z_0| \leq 2\sqrt{\operatorname{Re}(z_1) \cdot \operatorname{Re}(z_2)}.$$

*Proof.* The positive semi-definiteness of  $A$  is equivalent to

$$a_{11} \geq 0, \quad a_{22} \geq 0, \quad (a_{12} + a_{21})^2 \leq 4a_{11}a_{22}.$$

We first compute the real part of the eigenvalues  $z_i$  for  $i = 1, 2$  and obtain

$$\operatorname{Re}(z_i) = \Delta_t \frac{2 \left( \frac{a_{ii}}{h_i^2} + \frac{c_i^2}{12a_{ii}} \right) (\cos \phi_i - 1) \left( 1 - \frac{1}{6}(1 - \cos \phi_i) \right) + \frac{c_i^2}{12a_{ii}} \sin^2 \phi_i}{\left( 1 - \frac{1}{6}(1 - \cos \phi_i) \right)^2 + \left( \frac{c_i h_i}{12a_{ii}} \right)^2 \sin^2 \phi_i}.$$

With

$$\begin{aligned} \alpha_i &:= \frac{\cos \phi_i - 1}{1 - \frac{1}{6}(1 - \cos \phi_i)}, \\ \beta_i &:= h_i^2 \frac{c_i^2}{a_{ii}^2}, \\ \gamma_i &:= \sin^2 \phi_i / \left( 1 - \frac{1}{6}(1 - \cos \phi_i) \right)^2, \end{aligned}$$

it holds

$$\operatorname{Re}(z_i) = \Delta_t \frac{a_{ii}}{h_i^2} 2\alpha_i \frac{1 + \frac{1}{24}\beta_i(2 + \gamma_i/\alpha_i)}{1 + \frac{1}{144}\beta_i\gamma_i}.$$

Please note that

$$\frac{1 + \frac{1}{24}\beta_i(2 + \gamma_i/\alpha_i)}{1 + \frac{1}{144}\beta_i\gamma_i} \geq 1, \quad (3.43)$$

which can be verified by straightforward calculus

$$\frac{1 + \frac{1}{24}\beta_i(2 + \gamma_i/\alpha_i)}{1 + \frac{1}{144}\beta_i\gamma_i} \geq 1 \quad \Leftrightarrow \quad \frac{1}{6} \leq \frac{2\alpha_i + \gamma_i}{\alpha_i\gamma_i} = \frac{5 + \cos \phi_i}{18 \cos^2 \phi_i / 2} \in \left[ \frac{1}{3}, \infty \right).$$

Furthermore, the inequality

$$\frac{1}{36} (8 \sin \phi_i - \sin 2\phi_i)^2 \leq -2\alpha_i \quad (3.44)$$

is fulfilled. Due to  $\alpha_i \leq 0$  and inequality (3.43) we directly observe that the real parts of the eigenvalues  $z_i$  lie on the left hand side of the complex plane. It remains to show that  $|z_0| \leq 2\sqrt{\operatorname{Re}(z_1) \cdot \operatorname{Re}(z_2)}$ . Due to the positive semi-definiteness of  $A$  we obtain

$$|z_0|^2 \leq 4a_{11}a_{22} \left( \frac{1}{36} \right)^2 \frac{\Delta_t^2}{h_1^2 h_2^2} (8 \sin \phi_1 - \sin 2\phi_1)^2 (8 \sin \phi_2 - \sin 2\phi_2)^2.$$

Exploiting conditions (3.44) and (3.43), we obtain

$$\begin{aligned}
 |z_0|^2 &\leq 16a_{11}a_{22}\frac{\Delta_t^2}{h_1^2h_2^2}\alpha_1\alpha_2 \\
 &\leq 16a_{11}a_{22}\frac{\Delta_t^2}{h_1^2h_2^2}\alpha_1\alpha_2\frac{1+\frac{1}{24}\beta_1(2+\gamma_1/\alpha_1)}{1+\frac{1}{144}\beta_1\gamma_1}\frac{1+\frac{1}{24}\beta_2(2+\gamma_2/\alpha_2)}{1+\frac{1}{144}\beta_2\gamma_2} \\
 &= 4Re(z_1) \cdot Re(z_2).
 \end{aligned}$$

□

In Figures 3.4 and 3.5 we plot the stability regions of the HO-ADI schemes for three and four spatial dimensions. The part in dark gray shows the stability region for the special choice  $z_0 = 0, z_1 = z_2 = \dots = z_d$ . As in [43, 53] we consider the choice  $z_1 = z_2 = \dots = z_d$  to derive a necessary stability condition. The part in light gray shows the position of the eigenvalues  $z_i$  given by the equation (3.42). The sample points have been computed for the parameter set  $h_i = 10^{-1}, \Delta_t = 1, c_i = 1/2$  and  $a_{ii} = c_i \cdot \hat{p}$ . This case is rather conservative as it considers a large parabolic mesh ratio. The parameter  $\hat{p}$  determines the ratio between convection and diffusion and can be seen as the non-scaled reciprocal of the *Péclet number* [83]; the smaller  $\hat{p}$ , the stronger is the convection-dominance. Note, in the case of  $z_0 = 0$  the stability functions of the Douglas and the Craig-Sneyd scheme coincide. Therefore, we omit to plot the regions for the HCS scheme. The  $\theta$  values have been chosen according to the results for ADI schemes with second-order spatial discretization applied to diffusion equations without mixed derivative terms from the literature, e.g. [53]. In Sections 3.3.1 and 3.3.2 it was shown that these bounds are also valid for HO-ADI schemes applied to pure diffusion problems. Hence, we expect the HO-ADI schemes for convection-diffusion equations to have similar stability properties as their second-order counterpart. Both in the three (Figure 3.4) and four-dimensional (Figure 3.5) case the eigenvalues (light gray) lie within the stability region (dark gray) in all plots except in Figure 3.5 (i). For strong convection-dominance in plot (i) the HHV scheme becomes unstable.

### 3.3.4 Numerical Experiments with Diffusion Equations

In this section we validate our theoretical results with numerical experiments. We consider diffusion equations in up to four spatial dimensions on the unit hypercube  $\Omega_d = [0, 1]^d$ . Experiments with convection-diffusion equations can be found in Section 4. In addition to the temporal error decay, we also investigate the order of convergence in the spatial domain. We compare the accuracy of the numerical solution  $U_{\text{approx}}$  with a reference solution  $U_{\text{ref}}$ . If  $U_{\text{approx}}$  is the combined sparse grid solution, we extend the sparse grid approximation  $u_n^s$  to the full grid via multi-variate cubic spline interpolation in order to make its accuracy comparable to the full grid reference solution. In those figures which show the spatial error decay, we only state the numerical results of the HDO scheme. In all the other discussed HO-ADI schemes we apply the same spatial discretization and thus the same spatial error occurs.

In a first experiment we use the following symmetric positive definite coefficient matrix

$$A = 0.025 \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}.$$

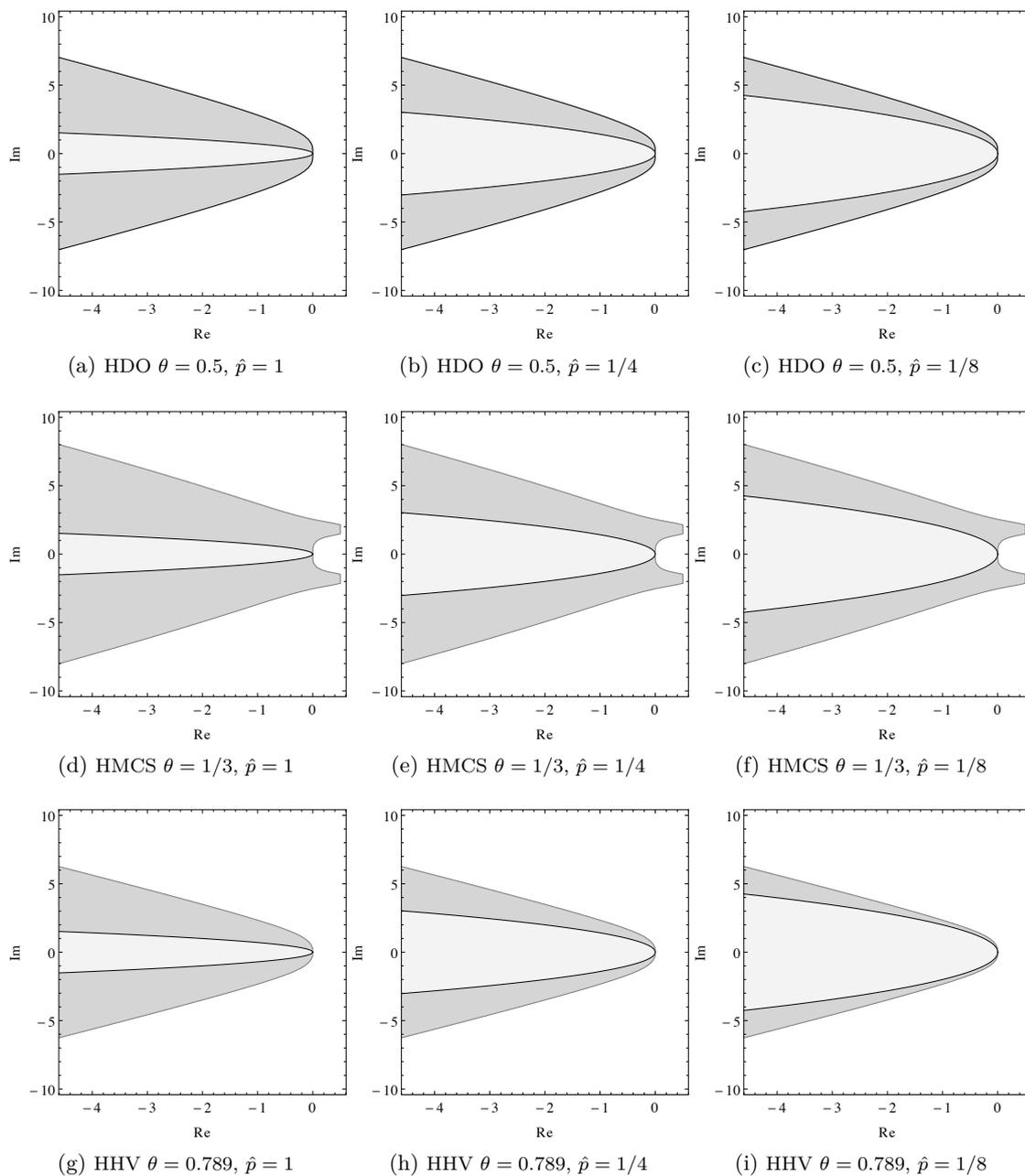


Figure 3.4: 3d: stability region (dark gray) for  $z_1 = z_2 = z_3$  and eigenvalues  $z_i$  (light gray) for special parameter choices.

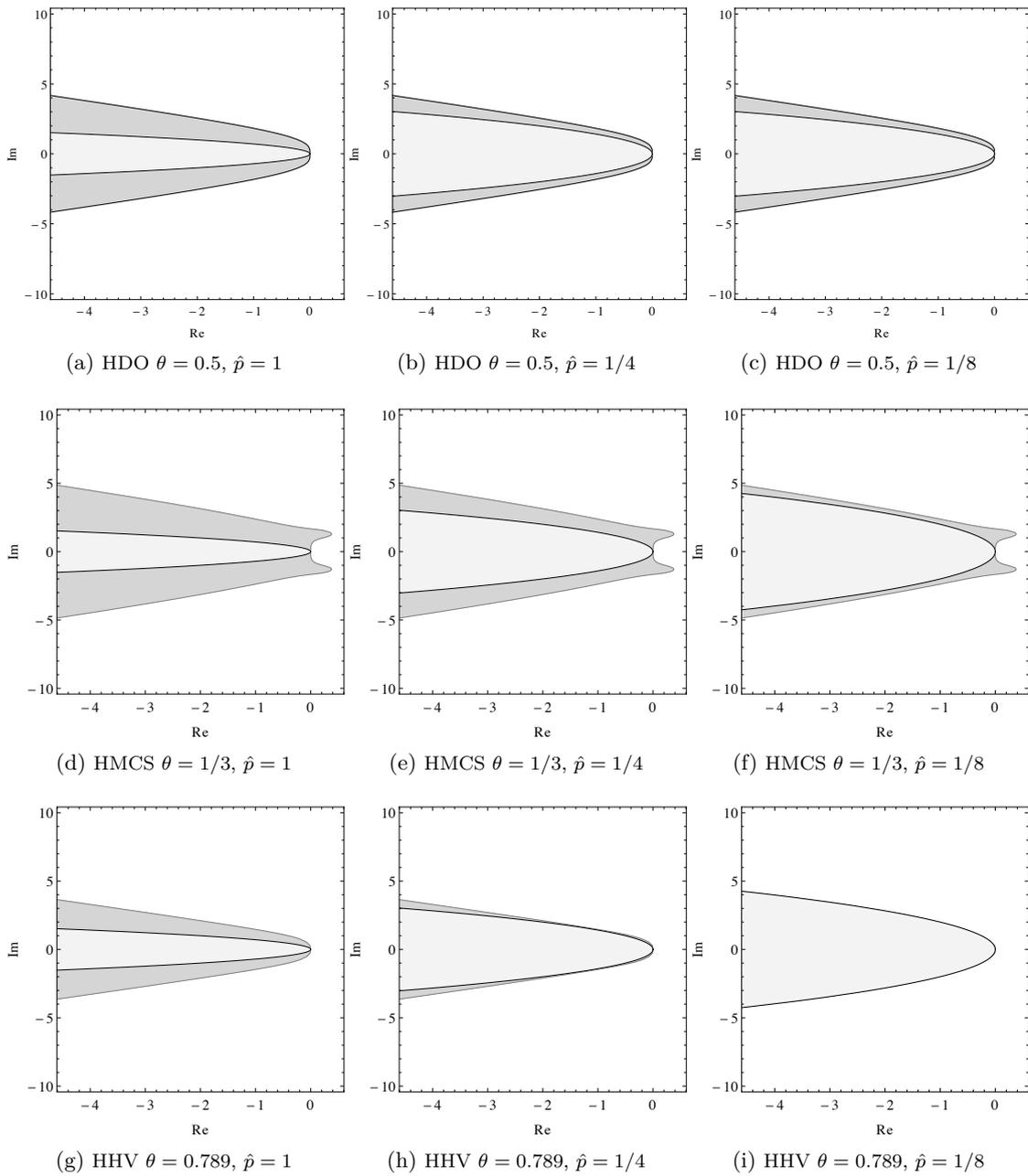


Figure 3.5: 4d: stability region (dark gray) for  $z_1 = z_2 = z_3 = z_4$  and eigenvalues  $z_i$  (light gray) for special parameter choices.

The initial value is chosen by

$$u(x_1, x_2, 0) = e^{-4(\sin^2(\pi x_1) + \cos^2(\pi x_2))}.$$

We apply periodic BCs, such that

$$u(x_1 \pm 1, x_2 \pm 1, t) = u(x_1, x_2, t).$$

Figure 3.6 shows the initial value, the solution as well as the sparse grid solution at the time  $t = 1$ .

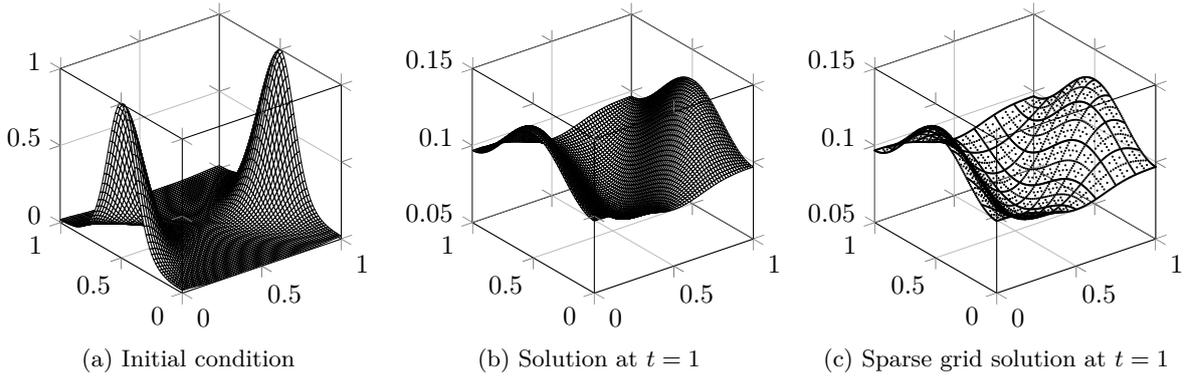


Figure 3.6: Initial condition, solution and sparse grid solution at  $t = 1$ .

To measure the order of convergence in time, we fix the spatial mesh width at  $h_1 = h_2 = 2^{-8}$  and compute the error (2.2) at  $t = 1$ . In Figure 3.6 (a) we compare schemes (3.25)–(3.28) to a reference solution with  $\Delta_t = 2^{-14}$ . In the left plot  $\theta$  is set to the lowest possible value ensuring unconditionally stability according to Theorem 24. All schemes show a stable behavior. The Douglas scheme exhibits order one in time, while the others show second-order convergence.

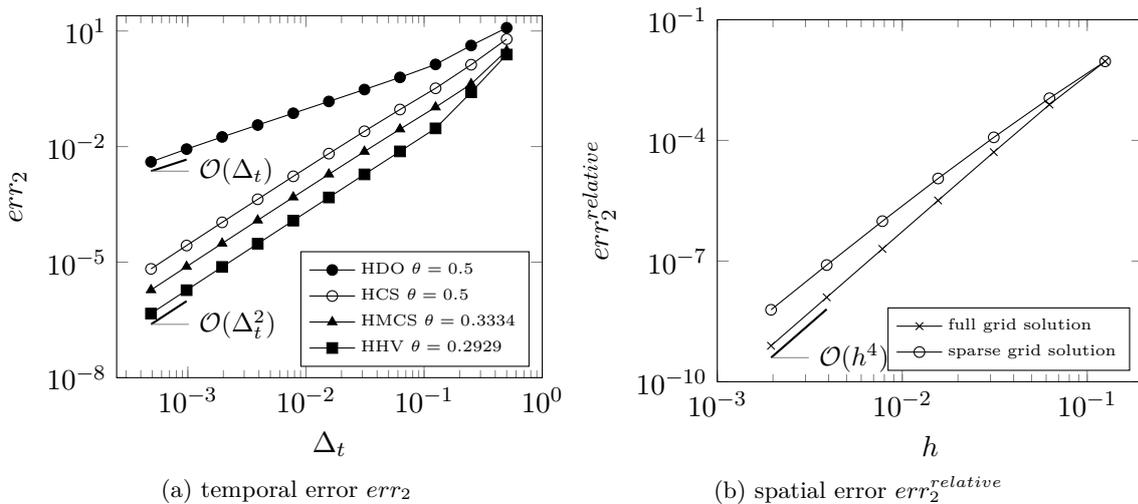


Figure 3.7: Numerical convergence plots  $d = 2$ .

Next we compute the spatial error of the sparse and full grid solution in Figure 3.7

(b). The time step is fixed at  $\Delta_t = 1/100$  and the reference solution is computed with mesh width  $h_1 = h_2 = 2^{-11}$ . The rate of convergence is slightly lower than the full grid convergence. Nevertheless, one has to keep in mind, that the sparse grid consists of significantly less grid points than the full grid ( $\mathcal{O}(h^{-1} \log_2(h^{-1})^{d-1})$  vs.  $\mathcal{O}(h^{-d})$ ). Thus, a better efficiency is achieved when taking both order of convergence and computational effort into considerations. This can be seen in Figure 3.8, where we compare the number of nodes of the sparse and full grid to the achieved accuracy: the sparse grid approach outperforms the full grid solution.

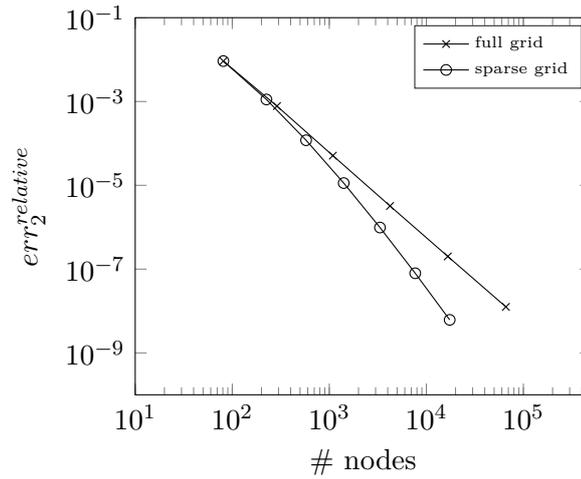


Figure 3.8: Number of grid points versus accuracy  $d = 2$ .

In the three-dimensional case we choose the coefficient matrix

$$A = 0.025 \begin{pmatrix} 2 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

as well as the initial value

$$u(x_1, x_2, x_3, 0) = e^{-(\cos^2(\pi x_1) + \cos^2(\pi x_2) + \cos^2(\pi x_3))}.$$

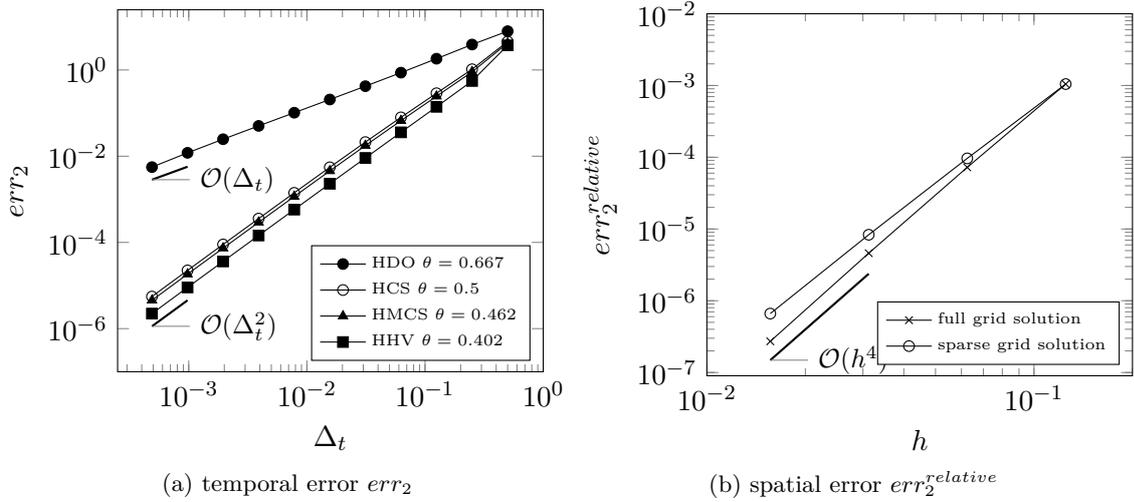
For the four-dimensional problem we use

$$A = 0.025 \begin{pmatrix} 2 & 2 & 1 & 1 \\ 2 & 4 & 2 & 1 \\ 1 & 2 & 3 & 2 \\ 1 & 1 & 2 & 3 \end{pmatrix}$$

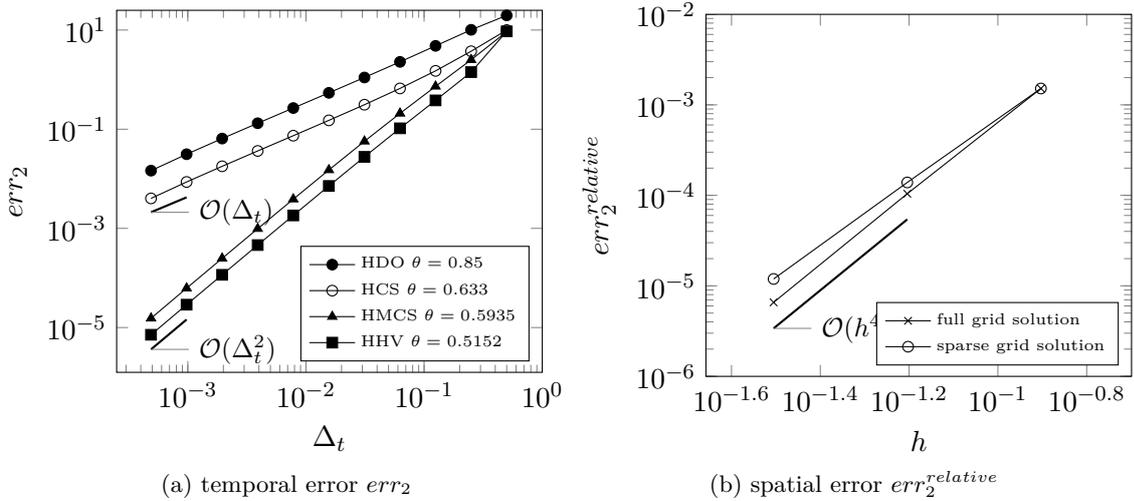
and

$$u(x_1, x_2, x_3, x_4, 0) = e^{-(\cos^2(\pi x_1) + \cos^2(\pi x_2) + \cos^2(\pi x_3) + \cos^2(\pi x_4))}.$$

In both cases we apply periodic BCs. Figure 3.9 shows the temporal and spatial error for  $d = 3$ . In Figure 3.9 (a) the mesh width is set to  $h_1 = h_2 = h_3 = 2^{-6}$  and the reference solution is computed with step size  $\Delta_t = 2^{-14}$ . The  $\theta$  values are again chosen according to Theorem 24. All schemes show the desired stable behavior and rate of convergence. In the spatial domain (Figure 3.9 (b)) the reference solution is computed at  $h_1 = h_2 = h_3 = 2^{-7}$ . The other parameters remain unchanged to the two dimensional examples. The numerical

Figure 3.9: Numerical convergence plots  $d = 3$ .

results reflect the theoretical considerations and we observe a rate of convergence of order  $\mathcal{O}(h^4 \log_2(h^{-1})^2)$  for the sparse grid and  $\mathcal{O}(h^4)$  in case of the full grid. Figure 3.10 shows

Figure 3.10: Numerical convergence plots  $d = 4$ .

the performance of the HO-ADI schemes in a four-dimensional spatial domain. In the time domain we compute the reference solution with  $\Delta_t = 2^{-14}$  and  $h_1 = h_2 = h_3 = h_4 = 2^{-5}$ . With four spatial dimensions the curse of dimensionality shows its effect very quickly and the memory consumption becomes very large even for moderate mesh width in each spatial dimension. Therefore, we use the combined sparse grid solution at level 16 as a reference solution evaluated at a discrete grid with size  $h_1 = h_2 = h_3 = h_4 = 2^{-6}$ . The numerical results in Figure 3.10 are in line with the theoretical findings. The necessary condition in Theorem 25 seems to be sufficient to yield a stable behavior. In contrast to the previous examples the HCS scheme exhibits order one in time since  $\theta \neq \frac{1}{2}$ . However, it performs better than the HDO scheme due to the lower  $\theta$  value. The HMCS and HHV scheme both show second-order convergence. In Figure 3.10 (b) one observes an error decay of order

four in case of the full grid and a slightly lower decay in case of the sparse grid. In Figure 3.11 the number of grid points of the full and sparse grid is shown versus the accuracy. Similar to the two-dimensional case the sparse grid achieves a higher accuracy per grid node compared to the full grid solution.

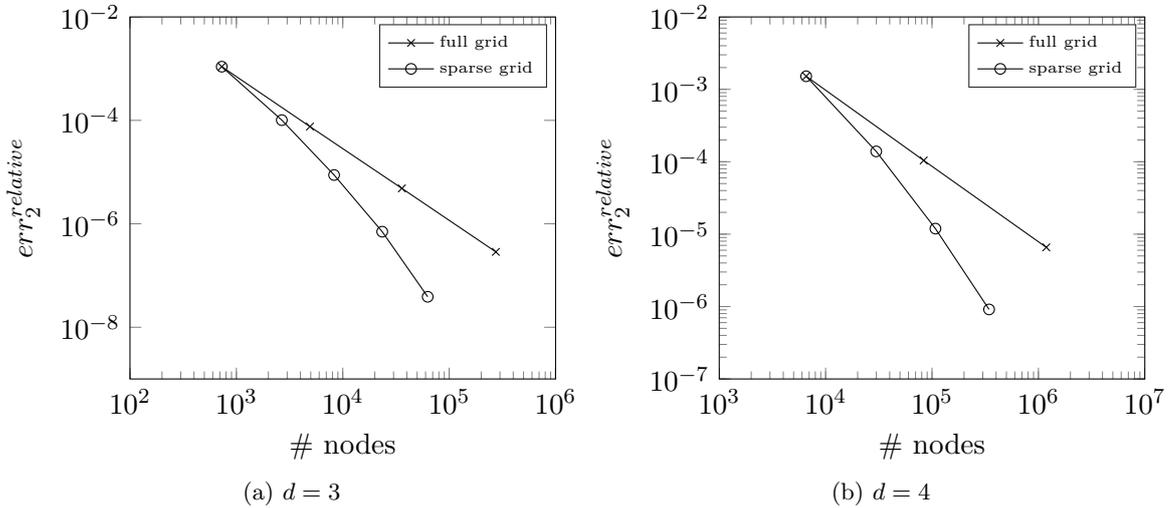


Figure 3.11: Number of grid points versus accuracy  $d = 3, 4$ .

### 3.4 (Hybrid) Pseudo-Spectral ADI Schemes

The idea of dimensional splitting can be carried over to the case of spectral or pseudo-spectral discretizations in space. As far as we know, this approach has only been discussed in the literature by Zeng et al. [97], where an ADI scheme based on the factorization of a Crank-Nicolson discretization is derived for the two-dimensional Riesz space fractional non-linear reaction-diffusion equation.

Analogue to finite difference approximations the computational workload is significantly reduced due to the splitting. For the following discussion we consider a spatial discretization on a grid with an equal number of nodes in each coordinate direction  $\mathbf{N} = (N, N, \dots, N)$ . In the explicit steps without mixed derivatives the computational complexity is proportional to  $N^{d-1} \cdot \mathcal{O}(N \log_2(N)) = \mathcal{O}(N^d \log_2(N))$  operations if FFT differentiation is used. If mixed derivatives are present the complexity increases to  $N^{d-2} \cdot \mathcal{O}(N^2 \log_2^2(N)) = \mathcal{O}(N^d \log_2^2(N))$ . In each of the implicit legs  $N^{d-1}$  one-dimensional problems have to be solved, with each having a complexity of  $\mathcal{O}(N^3)$  if Gaussian elimination is used to solve the system. If the coefficients in the PDE are time-independent the system can be solved via a LU decomposition in a startup phase, such that only one forward and backward substitution with quadratic effort has to be done to solve each one-dimensional problem. This leads to a complexity  $N^{d-1} \cdot \mathcal{O}(N^2) = \mathcal{O}(N^{d+1})$ . In comparison a fully implicit scheme has an effort of  $\mathcal{O}(N^{2d})$  if the system of equations is solved with the help of a precomputed LU decomposition via forward and backward substitution.

In the case of time-dependent parameters the discretization matrices change in each time iteration and therefore the updated system has to be solved in each time step. In order

to reduce the complexity, an iterative solver can be applied, e.g. the Biconjugate gradient stabilized method [94]. This method can handle the non-symmetry of the pseudo-spectral differentiation matrices. Each iteration of the solver can be performed in  $\mathcal{O}(N^2)$ , such that the total effort is  $\mathcal{O}(N_{iter}N^2)$ . If initial values with a small residual in a suitable norm are available, the number of iterations needed to achieve the desired accuracy can be bounded by a small constant, which leads to a quadratic effort. A good candidate as an initial value is the artificial solution of the previous leg within the ADI procedure. In our numerical experiments this choice led to  $N_{iter} < 3$ , cf. Section 4.

As the one-dimensional system can be solved in  $\mathcal{O}(N^2)$  operations, either via forward and backward substitution (time-independent case) or via an iterative solver (time-dependent case), the total effort for bounded  $N_{iter}$  is  $\mathcal{O}(N^{d+1})$ .

### 3.5 Stability of Hybrid Finite Difference/Pseudo-Spectral ADI Schemes

In many financial applications it is reasonable to combine pseudo-spectral differentiation and finite difference approximation to exploit the regularity of the solution in direction of certain coordinates, cf. Section 4. In the following we refer to methods, which use pseudo-spectral schemes in a subset of the coordinate directions and finite differences in the other directions as *hybrid schemes*. We investigate the stability of a method with finite differences in the first direction and pseudo-spectral differentiation in the second direction for the general convection-diffusion equation with fixed coefficients

$$\frac{\partial u}{\partial t} = \operatorname{div}(A\nabla u) + c \cdot \nabla u \quad \text{on } \Omega_2 = [0, 1] \times [-1, 1], \quad t > 0,$$

where  $A = (a_{ij})$  is a symmetric positive semi definite matrix and  $c = (c_1, c_2, \dots, c_d)^\top$  the vector of convection coefficients. Let the mesh be given by a tensor based discretization of directions  $x_1, x_2$  via  $x_{l_1, i} = i \cdot 1/N_1$ ,  $i = 0, 1, \dots, N_1$ ,  $N_1 = 2^{l_1}$  and  $x_{l_2, i} = \cos(\frac{\pi \cdot i}{N_2})$  for  $i = 0, 1, \dots, N_2$ ,  $N_2 = 2^{l_2}$ . The spatial discretization of the hybrid scheme can be written in matrix notation

$$\begin{aligned} F = & a_{11}D_{FD_1}^2 \otimes I_{N_2} + (a_{12} + a_{21})D_{FD_1} \otimes D_{SP_2} + a_{22}I_{N_1} \otimes D_{SP_2}^2 \\ & + c_1D_{FD_1} \otimes I_{N_2} + c_2I_{N_1} \otimes D_{SP_2}, \end{aligned}$$

where  $I_{N_1}, I_{N_2}$  denote the identity matrix of size  $N_1 + 1, N_2 + 1$ , respectively. Then we decompose the system via

$$\begin{aligned} F_0 &= (a_{12} + a_{21})D_{FD_1} \otimes D_{SP_2}, \\ F_1 &= a_{11}D_{FD_1}^2 \otimes I_{N_2} + c_1D_{FD_1} \otimes I_{N_2}, \\ F_2 &= a_{22}I_{N_1} \otimes D_{SP_2}^2 + c_2I_{N_1} \otimes D_{SP_2}. \end{aligned}$$

In a next step the ADI time discretization can be applied. For purposes of the stability investigations we rewrite methods (3.1)–(3.4) into the one-step form

$$U_{n+1} = RU_n,$$

with an iteration matrix  $R$  given by (3.6)–(3.9). The method is stable if  $\|R\| \leq 1$  holds. One crucial property for stability of the ADI schemes is that the eigenvalues of the operators  $F_1$  and  $F_2$  have negative real parts. For central second-order finite differences this is clearly fulfilled, see [55]. In the case of Chebyshev spectral methods it was shown by [31] that the second derivative matrix has negative and distinct real valued eigenvalues, which are bounded by  $\mathcal{O}(N_2^4)$ . They prove this result for Dirichlet, Neumann and Robin BCs. In [11, Section 7.3.2] the eigenvalues of convection-diffusion operators are analyzed for Dirichlet BCs. Following their proof, one directly observes that  $Re(\lambda) \leq -a_{22} \frac{\pi^2}{4}$  and that the spectral radius is bounded by  $\mathcal{O}(N_2^4)$  due to the second derivative matrix. Numerical tests in [11] reveal that these bounds are sharp. In the case of convection-diffusion problems with Neumann BCs we numerically compute the eigenvalues of the generalized problem

$$QU = \lambda BU, \quad (3.45)$$

where  $Q$  is a  $(N_2 + 1) \times (N_2 + 1)$  matrix, which consists of the matrix  $D_{SP_2}^{(2)} + D_{SP_2}$  at the inner nodes and the first and last row are identical to the first and last row of the differentiation matrix  $D_{SP_2}$  due to the homogeneous Neumann BCs. The  $B$  matrix is identical to the identity matrix of size  $(N_2 + 1) \times (N_2 + 1)$  except for the first and last entry, which is set to zero. Figure 3.12 shows the eigenvalues of problem (3.45), which has been solved using the QZ algorithm provided by the Matlab<sup>®</sup> routine  $eig(., .)$ . One observes that the results for Dirichlet also hold for Neumann BCs: except for one zero eigenvalue, all eigenvalues lie on the left-hand side of the complex plane and the spectral radius grows with  $\mathcal{O}(N_2^4)$ . The zero eigenvalue is associated with the eigenvector  $u = c \cdot (1, \dots, 1)^T$  for an arbitrary constant  $c$ . These results ensure the stability if no mixed derivatives are present. Since in financial engineering mixed derivative terms naturally arise due to the correlation structure between assets and/or risk factors it is important to include them in our stability considerations. Thus, we numerically compute the eigenvalues of the problem

$$RU = \lambda BU. \quad (3.46)$$

If Dirichlet BCs are applied, then  $R$  is of size  $(N_1 - 1)(N_2 - 1) \times (N_1 - 1)(N_2 - 1)$ . In the second coordinate direction, where the Chebyshev collocation method is used, the first and last row as well as the first and last column are removed due to the BC. The matrix  $B$  is the identity matrix of appropriate size. If a homogeneous Neumann BC is used in the second coordinate direction, we proceed according to the problem (3.45) to construct the differentiation matrices, which are employed to compute  $P$  (contained in  $R$ ). This matrix stems from all implicitly treated terms in the ADI method. For the explicit parts, namely  $Z_0$  and  $Z$ , we proceed as follows: we compute the solution at the interior nodes and determine the boundary values in such a way that they satisfy the BC by solving the system

$$\begin{aligned} d_{00}u_{k,0} + d_{0N_2}u_{k,N_2} &= - \sum_{j=1}^{N_2-1} d_{0j}u_{k,j}, \\ d_{N_20}u_{k,0} + d_{N_2N_2}u_{k,N_2} &= - \sum_{j=1}^{N_2-1} d_{N_2j}u_{k,j}, \end{aligned}$$

for  $k = 1, \dots, N_1$  and  $D_{SP_2} = (d_{ij})$ . Let  $\tilde{D}$  denote the matrix which forces the boundary nodes in the  $x_2$ -direction to fulfill the BC according to the system above, then we can

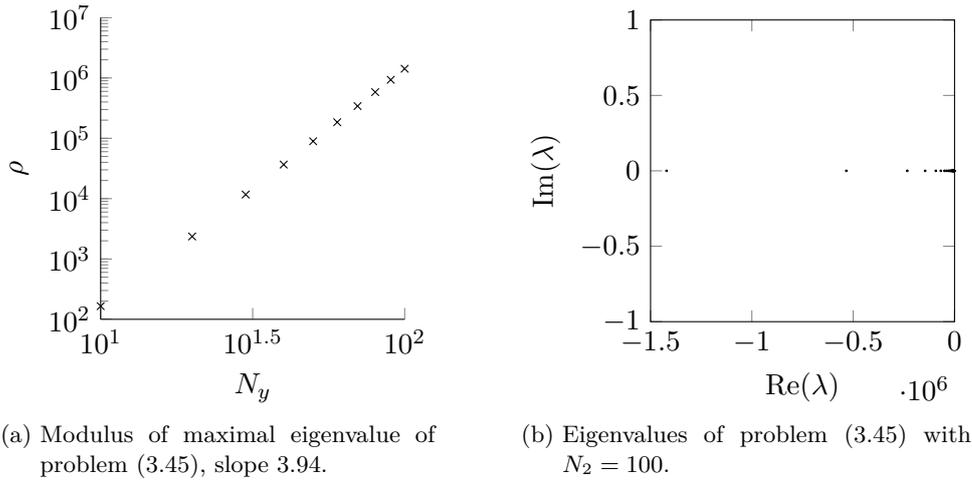


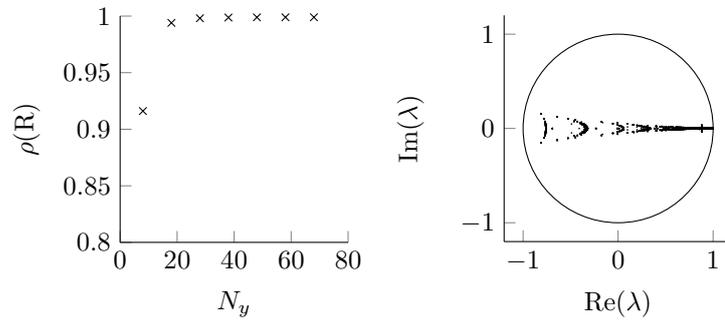
Figure 3.12: Neumann boundary conditions.

compute the matrix stemming from the explicit time-stepping via  $Z := \tilde{D}Z$  and  $Z_0 := \tilde{D}Z_0$ . Similar to the problem (3.45) the matrix  $B$  is the identity matrix with zeros on the diagonal for each grid node lying on the boundary of  $y$ .

In the following we numerically investigate the stability of the ADI schemes. Therefore, we compute the spectrum of the iteration matrices  $R$  given by equations (3.6)–(3.9). If  $\rho(R) < 1$  is fulfilled the numerical scheme is stable. For our numerical experiment we consider the diffusion coefficient matrix and the convection vector

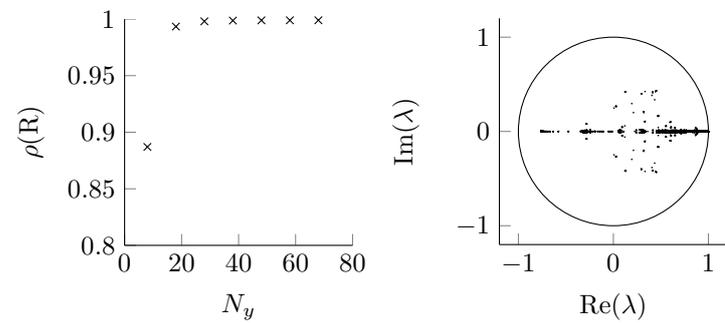
$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The matrix  $A$  is symmetric positive semi-definite with the largest possible relative size of the mixed derivative coefficient  $\gamma = 1$  defined in (3.19). This choice can be seen as a worst case scenario in terms of the stability since the evolution of the solution in one variable is completely determined by the variable in the other coordinate direction. The ratio between convection and diffusion is equal to one for this parameter choice. In case of the MCS scheme we choose  $\theta = 0.42$  since the eigenvalues of the approximation of the mixed derivative term are complex and since  $\gamma = 1$ . Further we let  $\Delta_t = 0.1$  for our numerical evaluations. The  $\theta$  value is chosen according to the values given in Section 3.1 which is derived for finite difference schemes. Figures 3.13 and 3.14 show the largest modulus of eigenvalue and the location of all eigenvalues in the complex plane of problem (3.46) with Dirichlet and Neumann BCs. For an increasing number of grid nodes, the spectral radius for both problems approaches one from below. Thus, we expect a stable behavior of the hybrid scheme, even for problems with large correlations. Please note that one obtains similar results for problems with strong convection-dominance.



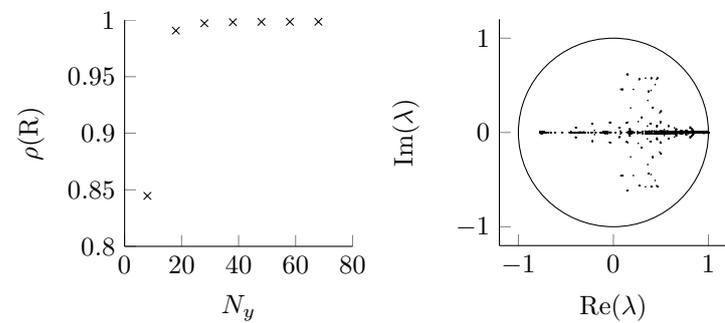
(a) DO,  $\theta = 0.5$ ,  $N_x = 100$

(b) DO,  $\theta = 0.5$ ,  $N_x = 100$



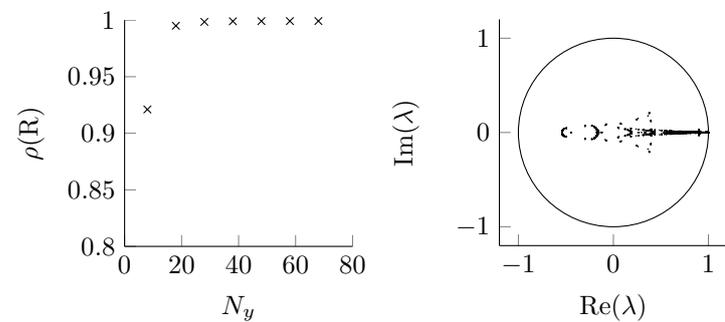
(c) CS,  $\theta = 0.5$ ,  $N_x = 100$

(d) CS,  $\theta = 0.5$ ,  $N_x = 100$



(e) MCS,  $\theta = 0.42$ ,  $N_x = 100$

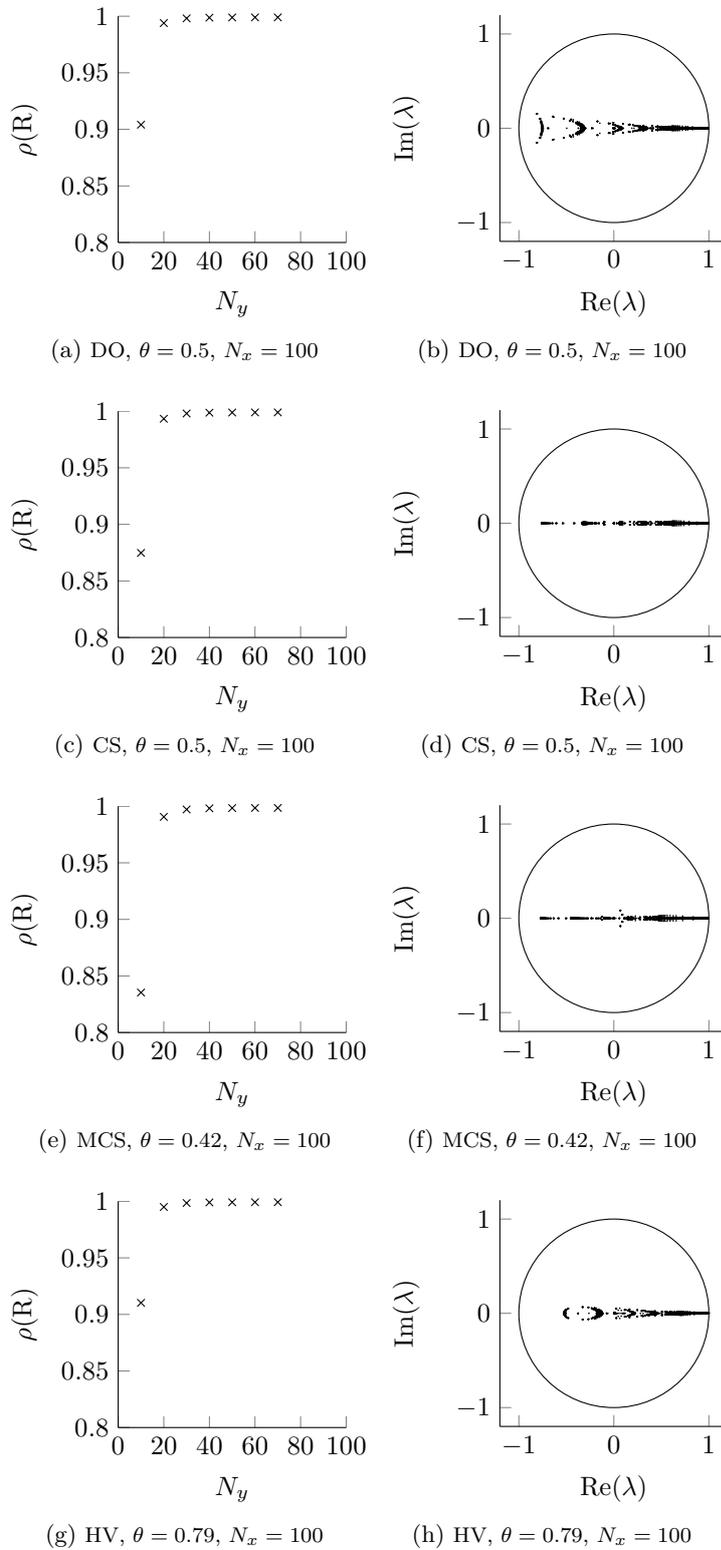
(f) MCS,  $\theta = 0.42$ ,  $N_x = 100$



(g) HV,  $\theta = 0.79$ ,  $N_x = 100$

(h) HV,  $\theta = 0.79$ ,  $N_x = 100$

Figure 3.13: Dirichlet BCs at  $\partial\Omega$ .

Figure 3.14: Neumann BCs at the boundary in  $y$ -direction, Dirichlet BCs in  $x$ .

# 4 Chapter 4

---

## Application to Financial Engineering

This chapter is devoted to the application of the methods presented in the previous chapters to PDEs arising in the field of financial engineering. These PDEs are often defined on an open interval, e.g. in the interest rate direction in the HHW model (1.7) or on one sided open intervals, like in the asset direction of the Black-Scholes model (1.5). For computational purposes the domain has to be truncated and boundary conditions have to be imposed. Often the exact boundary value is unknown and boundary conditions, which mimic the true solution or its behavior, are used to close the numerical scheme. This introduces a numerical error. In order to keep this error and its influence on the accuracy in the interior of the domain small, the computational domain has to be chosen quite large. For example in [38, 44, 51] the asset direction in stochastic volatility models is truncated at 14 to 20 times the strike price. On the one hand this minimizes the error due to the boundary treatment, but on the other hand it leads to a very large computational domain, however, practitioners are usually only interested in an accurate solution close to the strike price. Therefore, it is quite natural to ask for non-uniform meshes, which allow to cluster grid points in a region of interest, where a high accuracy is desired. In Section 4.1 we give examples of grid transformations for clustering.

As the option value is known at maturity, the option pricing PDE is solved backwards in time from maturity  $T$  to the actual point in time  $t = 0$ . Equivalently one can transform the problem via  $\tau = T - t$ . The terminal value becomes an initial value and the PDE can be solved forward in time as an IBVP. For many financial options the initial condition exhibits discontinuities, e.g. for European options the first derivative in asset direction is discontinuous, while for Digital options the payoff profile is discontinuous. In ADI time stepping methods this leads to a large numerical error if the time steps are chosen too large, see [38, 51]. This problem can be circumvented with a Rannacher time stepping. Here, the first time steps are computed with an implicit scheme to damp out high frequency modes. This approach has been analyzed in [29, 56] for CN and CS time marching. Furthermore, the effects are illustrated by numerical examples in Section 4.4. For high-order spatial discretizations the non-smooth payoff leads to a reduction of the order of accuracy. The derivation of the spatial approximation generally relies on a certain regularity of the solution. If the regularity assumption is not fulfilled, the numerical scheme will not exhibit its theoretical order of convergence. In Section 4.2 we discuss how to cope with this issue and how to recover the high rate of convergence in practice.

### 4.1 Grid Transformation

The accuracy of grid-based methods can be improved by concentrating grid nodes near critical points, such as the strike price. For simplicity we consider a one-dimensional time-independent transformation first. Let the old coordinate be given by  $x$ , then we can

express the new variable  $y$  via its smooth transform function

$$y = \psi(x).$$

By the chain rule and applying the ansatz  $u(x) = \tilde{u}(y)$  we can express the derivatives of  $u$  with derivatives of  $\tilde{u}$  depending on the new variable  $y$

$$\begin{aligned} \frac{\partial u(x)}{\partial x} &= \frac{\partial \tilde{u}(y)}{\partial x} = \frac{\partial \tilde{u}(y)}{\partial y} \frac{\partial \psi(x)}{\partial x} \\ &= \frac{\partial \tilde{u}(y)}{\partial y} \frac{\partial \psi(\psi^{-1}(y))}{\partial x}, \\ \frac{\partial^2 u(x)}{\partial x^2} &= \frac{\partial}{\partial x} \left( \frac{\partial \tilde{u}(y)}{\partial y} \frac{\partial \psi(x)}{\partial x} \right) \\ &= \frac{\partial^2 \tilde{u}(y)}{\partial y^2} \left( \frac{\partial \psi(\psi^{-1}(y))}{\partial x} \right)^2 + \frac{\partial \tilde{u}(y)}{\partial y} \frac{\partial^2 \psi(\psi^{-1}(y))}{\partial x^2}. \end{aligned}$$

In the following we give examples of transformations, which are frequently used in computational finance, e.g. see [21, 22, 24, 38, 51, 90]. The log-transform is given by

$$\psi(x) = \log(x). \quad (4.1)$$

Thus, it holds for the derivatives

$$\begin{aligned} \frac{\partial u(x)}{\partial x} &= \frac{\partial \tilde{u}(y)}{\partial y} e^{-y}, \\ \frac{\partial^2 u(x)}{\partial x^2} &= \frac{\partial^2 \tilde{u}(y)}{\partial y^2} e^{-2y} - \frac{\partial \tilde{u}(y)}{\partial y} e^{-2y}. \end{aligned}$$

If applied to the Black-Scholes equation the PDE in the new variable has constant coefficients, which leads to a flatter distribution of the eigenvalues of the discretization matrix. This makes this transformation especially interesting for explicit schemes, see [90]. Please note that in the domain of the  $x$  variable the lower bound has to be truncated at  $\epsilon > 0$ . A second transformation is given in [51, 90]

$$\psi(x) = \frac{c_1 + \sinh^{-1}\left(\frac{K-x}{\alpha}\right)}{c_1 - c_2} \quad (4.2)$$

where

$$\begin{aligned} c_1 &= \sinh^{-1}\left(\frac{a-K}{\alpha}\right), \\ c_2 &= \sinh^{-1}\left(\frac{b-K}{\alpha}\right). \end{aligned}$$

The transformation maps the arbitrary interval  $[a, b]$  with  $a, b \in \mathbb{R}$ ,  $a < b$  to  $[0, 1]$  and clusters grid points around the strike price  $K$ , which is the region of highest interest from a perspective of practitioners. Small  $\alpha$ -values lead to a highly non-uniform grid, while large values lead to a uniform distribution of grid nodes.

In the case of Chebyshev spectral differentiation one first needs to map the domain to the

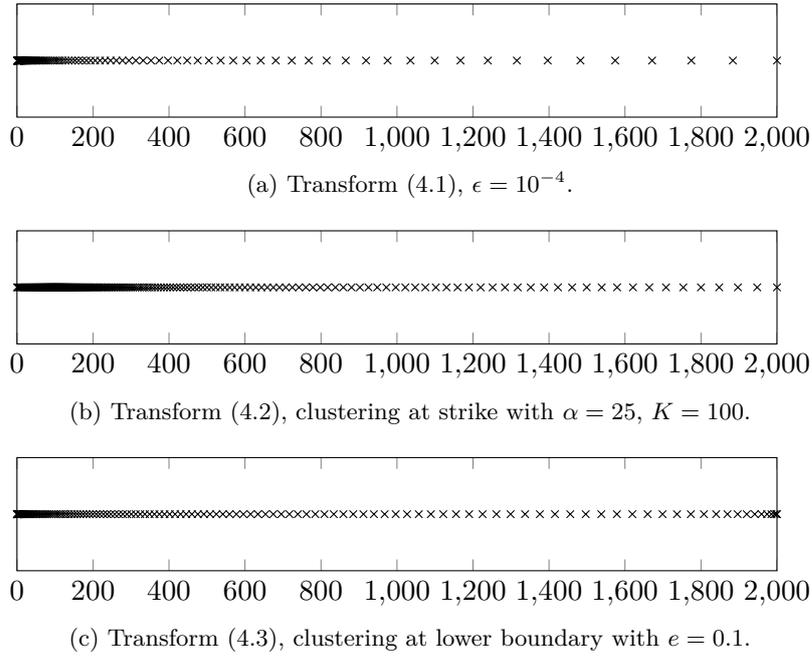


Figure 4.1: Sample mesh.

unit intervall  $[-1, 1]$ . This can be done by the linear transformation

$$\psi_1(x) = \frac{2}{b-a}x + \frac{a+b}{a-b}.$$

In a second step the clustering can be done. Tee et al. [91] and Pindza et al. [74] propose to use

$$\psi_2(x) = e \sinh \left( \frac{1}{2}(x-1) \left( \sinh^{-1} \left( \frac{1-d}{e} \right) + \sinh^{-1} \left( \frac{d+1}{e} \right) \right) + \sinh^{-1} \left( \frac{1-d}{e} \right) \right) + d,$$

where the parameter  $d \in [-1, 1]$  determines the region of clustering and  $e > 0$  the degree of non-uniformity of the grid spacing. Smaller  $e$  values lead to a stronger clustering. The complete transformation is then given by the composition

$$\psi = \psi_2 \circ \psi_1. \quad (4.3)$$

Figure 4.1 shows the transformed grids. In the first two cases we have used a uniform mesh in the new variable and transformed it back to the original variable via the inverse of  $\psi$ . In Figure 4.1(c) we applied a Chebyshev grid and plotted the backward-transformed grid.

In the higher-dimensional case one can apply a factorization of one-dimensional transformations to achieve a clustering in various dimensions. The construction of such a factorization is a straightforward sequentially application of the one-dimensional transformation given above. However, for some problems it might be advantageous to apply a general multi-dimensional transformation. Let the new variable be described via

$$y_i = \psi_i(x_1, x_2, \dots, x_d) \quad \text{for } i = 1, 2, \dots, d.$$

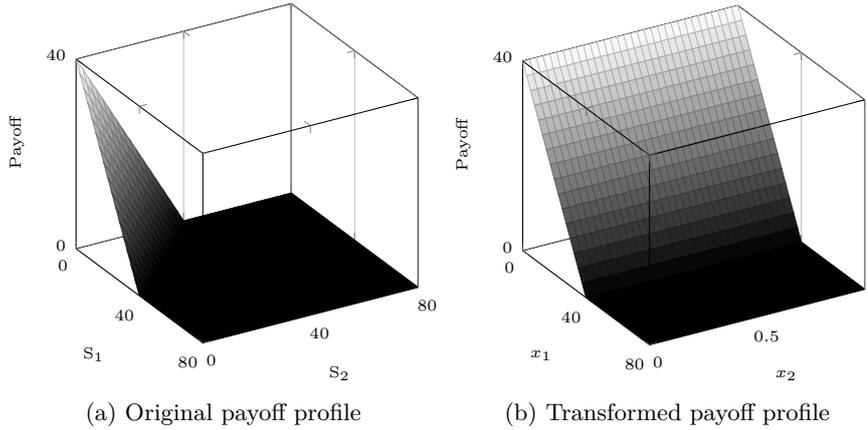


Figure 4.2: European 2-d basket put payoff without and with transformation.

Analogously to the one-dimensional case we use the ansatz

$$u(x_1, x_2, \dots, x_d) = \tilde{u}(y_1, y_2, \dots, y_d).$$

Application of the multidimensional chain rule yields

$$\begin{aligned} \frac{\partial u(x_1, x_2, \dots, x_d)}{\partial x_i} &= \frac{\partial \tilde{u}(y_1, y_2, \dots, y_d)}{\partial x_i} = \frac{\partial \tilde{u}}{\partial y_1} \frac{\partial \psi_1}{\partial x_i} + \frac{\partial \tilde{u}}{\partial y_2} \frac{\partial \psi_2}{\partial x_i} + \dots + \frac{\partial \tilde{u}}{\partial y_d} \frac{\partial \psi_d}{\partial x_i} \\ &= \sum_{k=1}^d \frac{\partial \psi_k}{\partial x_i} \frac{\partial \tilde{u}}{\partial y_k}, \\ \frac{\partial^2 u(x_1, x_2, \dots, x_d)}{\partial x_j \partial x_i} &= \frac{\partial}{\partial x_j} \left( \sum_{k=1}^d \frac{\partial \psi_k}{\partial x_i} \frac{\partial \tilde{u}}{\partial y_k} \right) \\ &= \sum_{k=1}^d \sum_{l=1}^d \frac{\partial \psi_k}{\partial x_i} \frac{\partial \psi_l}{\partial x_j} \frac{\partial^2 \tilde{u}}{\partial y_k \partial y_l} + \sum_{k=1}^d \frac{\partial^2 \psi_k}{\partial x_i \partial x_j} \frac{\partial \tilde{u}}{\partial y_k}, \end{aligned}$$

for  $i, j = 1, 2, \dots, d$ . With the help of this multi-dimensional transformation one can not only concentrate grid points in a region of interest, but it also allows to eliminate terms, e.g. mixed derivatives, or to align the new coordinate system to the direction of greatest variation of the initial condition. In the case of basket options one can transform the grid, such that the discontinuity only occurs in one coordinate direction. Examples of multi-dimensional transformations can be found in [41, 66, 78, 90]. In this thesis we restrict ourselves to a non-linear transformation [66, 78] for European plain vanilla basket options.

The transformation is given by

$$x_1 := \sum_{i=1}^d w_i s_i, \quad x_j := \frac{w_{j-1} s_{j-1}}{\sum_{i=j-1}^d w_i s_i} \text{ for } j = 2, 3, \dots, d. \quad (4.4)$$

The payoff transforms to  $\phi(x_1) = (x_1 - K)^+$ ,  $\phi(x_1) = (K - x_1)^+$  respectively. Figure 4.2 shows the alignment of the discontinuity to the first coordinate direction. The PDE (1.5)

transforms to

$$\frac{\partial u}{\partial t} + \sum_{i,j=1}^d \alpha_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^d \beta_i \frac{\partial u}{\partial x_i} - ru = 0 \quad (4.5)$$

in  $\tilde{\Omega}_d \times \Omega_t$  with  $\tilde{\Omega}_d = [0, x_1^{\max}] \times [0, 1]^{d-1}$ ,  $\Omega_t = [0, T]$ . The coefficient functions are given by

$$\begin{aligned} \alpha_{11} &= x_1^2 \sum_{k,l=1}^d \hat{\rho}_{kl} f_{1,k} f_{1,l}, \\ \alpha_{1j} &= x_1 x_j (1 - x_j) \sum_{k,l=1}^d (\hat{\rho}_{k,l-1} - \hat{\rho}_{kl}) f_{1k} f_{1l}, \\ \alpha_{ij} &= x_i (1 - x_i) x_j (1 - x_j) \sum_{k,l=1}^d (\hat{\rho}_{kl} - \hat{\rho}_{i-1,l} - \hat{\rho}_{k,j-1} + \hat{\rho}_{i-1,j-1}) f_{ik} f_{jl}, \end{aligned}$$

and

$$\begin{aligned} \beta_1 &:= x_1 \sum_{k=1}^d (r - q_k) f_{1k}, \\ \beta_i &:= x_i (1 - x_i) \sum_{k,l=1}^d \left( -2\hat{\rho}_{i-1,i-1} x_i + (2x_i - 1)(\hat{\rho}_{k,i-1} + \hat{\rho}_{l,i-1}) \right. \\ &\quad \left. + 2(1 - x_i) \hat{\rho}_{k,l} \right) f_{ik} f_{il} \\ &\quad + x_i (1 - x_i) \left( r - q_1 - \sum_{k=1}^d ((r - q_k) f_{ik}) \right), \end{aligned}$$

with

$$f_{il} := \begin{cases} x_{l+1} \prod_{j=i+1}^l (1 - x_j) & i < l < d \\ \prod_{j=i+1}^l (1 - x_j) & i < l = d \\ x_{l+1} & i = l < d \\ 1 & i = l = d \\ 0 & i > l \end{cases}$$

and  $\hat{\rho}_{ij} = \frac{1}{2} \rho_{ij} \sigma_i \sigma_j$  for  $i, j = 1, \dots, d$ . We see that (4.5) possesses the same structure as the original PDE (1.5) but with different coefficient functions. Since it holds  $\alpha_{ij} = 0$ ,  $\beta_j = 0$  for  $x_j = \{0, 1\}$  with  $i \geq 1$  and  $j > 1$  and  $\alpha_{i,1} = 0$ ,  $\beta_i = 0$  for  $x_1 = 0$ , we do not need to prescribe any boundary conditions in these cases. Only at the upper limit of the truncated domain in the first coordinate direction,  $x_1 = x_1^{\max}$ , a boundary condition has to be specified. In the case of a put option the option's value can be set to zero and for call options the second derivative can be considered as zero.

## 4.2 Non-Smooth Initial Data

The analysis of consistency for numerical methods typically relies on smoothness assumptions of the initial data. In practice, and especially in financial option pricing, the payoff function usually exhibits discontinuities at the strike price. This leads to a maximal error in the at-the-money region in the numerical solution. Since the option values close to the strike price are in general of the highest interest this is a severe problem. In [76] several methods have been discussed to overcome this issue and to recover a high rate of accuracy.

An intuitive approach is to place more grid points in the region of interest. For example Linde [69] solves a sub-problem with  $\mathcal{O}(h^{-2})$  grid nodes around the strike price to gain sixth order accuracy in space. Kreiss et al. [62] propose to smooth the initial condition. With this averaging a high rate of convergence can be recovered, while the initial condition converges to the original initial condition as the grid spacing goes to zero. This approach was successfully applied to option pricing problems in one dimension [76], two dimensions [24] and three dimensions [45]. An additional method to cope with the non-smooth initial data was given by Wahlbin [96], where the initial payoff profile undergoes an  $L_2$  projection onto a set of basis functions. Besides these techniques, in [90] grid shifting is suggested. Here the grid is sequentially shifted, such that the discontinuity falls midway between two successive grid nodes. The discrete payoff for the shifted grid reveals that this method can be interpreted as a kind of smoothing.

The last three approaches have been investigated in [76] for one-dimensional option pricing problems in the case of second-order accuracy: in the numerical tests all techniques showed the desired order of convergence. Furthermore, they give a brief outlook how to apply these methods to higher-dimensional problems: at the current state it is not clear if grid shifting is possible for higher dimensions since the grid is not allowed to coincide with the discontinuity. In the case of the projection Wahlbin reports technical difficulties if the discontinuities do not match with the grid nodes.

We restrict ourselves to smooth the initial condition according to [62] via a convolution operator. This approach can be easily extended for arbitrary dimension via a tensor product of one-dimensional convolutions. Compared to the first method it can be computed during an offline phase and the smoothed data can be reused to price options with different parameter sets. We briefly review the approach given in [62]:

Let  $\eta_p(\sin \omega)$  be the polynomial of lowest degree  $p$ , such that

$$\eta_p(\sin \omega) = \omega^p + \mathcal{O}(\omega^{2p}), \quad \text{for } \omega \rightarrow 0, \quad (4.6)$$

holds and define

$$\hat{\Phi}_p(\omega) = \frac{\eta_p(\sin \frac{1}{2}\omega)}{(\frac{1}{2}\omega)^p}.$$

Then one can construct a smoothing operator  $S_{h_i}^{(p)}$  along coordinate direction  $i$  for  $i = 1, 2, \dots, d$  and grid point  $x_{l_i, j_i}$  via

$$S_{h_i}^{(p)} g = h_i^{-1} \Phi_p(h_i^{-1} x_{l_i, j_i}) * \phi,$$

where  $\phi$  is the initial condition and  $\Phi_p$  denotes the Fourier inverse of  $\hat{\Phi}_p$ . The higher-

dimensional smoothing operator can be formed as a tensor product of one-dimensional operators

$$S_h^{(p)} = \prod_{i=1}^d S_{h_i}^{(p)}.$$

If the solution shall be smooth enough to achieve fourth-order accuracy we choose  $p = 4$ . Thus, we have

$$\hat{\Phi}_4(\omega) = \frac{\sin^4 \frac{1}{2}\omega + \frac{2}{3} \sin^6 \frac{1}{2}\omega}{(\frac{1}{2}\omega)^4}.$$

The inverse Fourier transform  $\Phi_4$  is a polynomial of degree three with support  $[-3, 3]$ . Hence the smoothed initial condition at grid point  $\mathbf{x}_{l,j}$  is given by

$$S_h^{(4)}\phi(x_{l_1,j_1}, \dots, x_{l_d,j_d}) = h_1^{-1} \dots h_d^{-1} \int_{-3h_1}^{3h_1} \dots \int_{-3h_d}^{3h_d} \Phi_4(h_1^{-1}\tilde{x}_1) \dots \Phi_4(h_d^{-1}\tilde{x}_d) \cdot \phi(x_{l_1,j_1} - \tilde{x}_1, \dots, x_{l_d,j_d} - \tilde{x}_d) d\tilde{x}_1 \dots d\tilde{x}_d. \quad (4.7)$$

Since  $\Phi_p$  are polynomials the smoothed initial value can be computed analytically for many problems, such as for plain-vanilla options.

In the following we apply the smoothing procedure to the initial condition of an European put option under the Black-Scholes model (1.4). This problem serves as a good test problem as the payoff function exhibits a discontinuity in the first derivative and an analytic reference solution is available. The smoothing functions for  $p = 6, 8, 10$  are given by

$$\begin{aligned} \hat{\Phi}_6(\omega) &= \frac{\sin^6 \frac{1}{2}\omega + \sin^8 \frac{1}{2}\omega + \frac{13}{15} \sin^{10} \frac{1}{2}\omega}{(\frac{1}{2}\omega)^6}, \\ \hat{\Phi}_8(\omega) &= \frac{\sin^8 \frac{1}{2}\omega + \frac{4}{3} \sin^{10} \frac{1}{2}\omega + \frac{62}{45} \sin^{12} \frac{1}{2}\omega + \frac{1244}{945} \sin^{14} \frac{1}{2}\omega}{(\frac{1}{2}\omega)^8}, \\ \hat{\Phi}_{10}(\omega) &= \frac{\sin^{10} \frac{1}{2}\omega + \frac{5}{3} \sin^{12} \frac{1}{2}\omega + 2 \sin^{14} \frac{1}{2}\omega + \frac{134}{63} \sin^{16} \frac{1}{2}\omega + \frac{2021}{945} \sin^{18} \frac{1}{2}\omega}{(\frac{1}{2}\omega)^{10}}. \end{aligned}$$

The parameters of the put option are chosen to be:  $\sigma = 0.25$ ,  $r = 0$ ,  $T = 1$ ,  $K = 100$ . The computational domain is truncated at  $s_{\max} = 4K$ . In the case of finite differences we use a uniform mesh and in the spectral case a Chebyshev grid under a linear coordinate transformation, which maps the interval  $[0, s_{\max}]$  to  $[-1, 1]$ . In the non-uniform Chebyshev grid we use the mesh distance to the next grid point as the step size  $h_i$  in the corresponding direction  $i$  for  $i = 1, 2, \dots, d$  within the smoothing procedure (4.7). For the time discretization we employ a Crank-Nicolson scheme and time steps  $\Delta_t = 10^{-4}$ . Figure 4.3 shows the difference between the smoothed and non-smoothed initial condition. Not surprisingly the effect of the smoothing is strongest at the discontinuity  $K$ . In Table 4.1 we compare the accuracy and convergence for different rates of smoothing. The error is computed according to (2.4), where the numerical and the analytical solution are evaluated at the discrete points  $s = 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150$ . The error ratio is the quotient of two consecutive errors. With the original initial condition

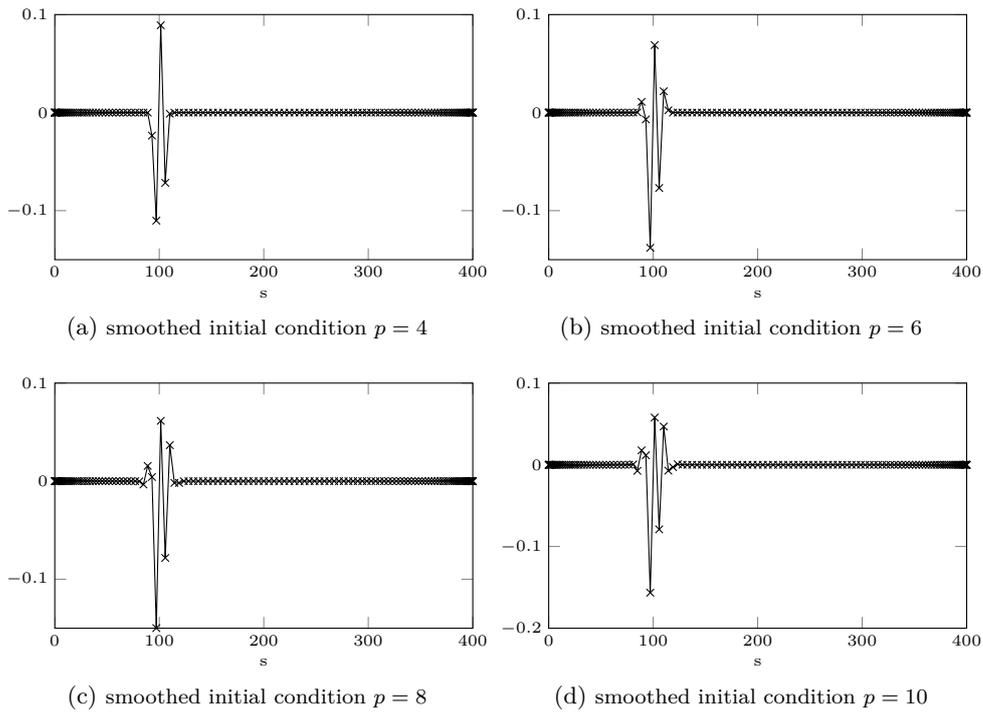


Figure 4.3: Smoothed initial condition with 129 grid nodes.

the rate of convergence is limited to an order of two. By application of the smoothing, a higher rate of convergence can be recovered. In the case of  $p = 4$  and  $p = 6$  the theoretical rate of order 4 and 6, respectively, is reached in the experiment. For the higher order cases the saturation due to the temporal error is reached before the scheme exhibits its theoretical error. However, we observe an error decrease with quotients of 199.23 and 607.01 in the peak. If an accuracy of up to two decimals is desired the finite difference method needs 257 nodes, whereas the pseudo-spectral scheme with  $p = 10$  only needs 33 nodes. Please note that the accuracy can be further improved if one additionally applies coordinate stretching to cluster grid nodes in the region of interest.

### 4.3 Basket Options

This section is devoted to the computation of basket options in the Black-Scholes framework. We restrict ourselves to European basket put options, which pay

$$\phi(s_1, s_2, \dots, s_d) = \left( K - \sum_{i=1}^d s_i \right)^+$$

at the maturity  $T$ .

In the literature basket options or options on the spread between two assets have been solved by Villeneuve and Zanette [95] with second-order finite differences and ADI splitting. Furthermore, also HOC discretizations were employed in a two and three-dimensional setting [24, 47]. Higher-dimensional problems with up to five spatial dimensions were

	Finite Differences		Pseudo-spectral		Pseudo-spectral $p = 4$	
N	error	error ratio	error	error ratio	error	error ratio
5	17.3658		10.4843		13.2215	
9	4.3857	3.9597	3.6256	2.8917	3.8886	3.400
17	1.5200	2.8853	0.5885	6.1607	0.5609	6.9322
33	0.3209	4.7371	0.1420	4.1436	0.0578	9.7073
65	0.0770	4.1690	0.0319	4.4502	0.0040	14.4204
129	0.0191	4.0353	0.0081	3.9396	0.0003	15.7405
257	0.0048	4.0085	0.0020	4.0991	1.6e-5	16.0330
513	0.0012	4.0021	0.0005	3.9871	1.0e-6	16.0132
1025	0.0003	4.0005	0.0001	3.9449	6.5e-8	15.3057

	Pseudo-spectral $p = 6$		Pseudo-spectral $p = 8$		Pseudo-spectral $p = 10$	
N	error	error ratio	error	error ratio	error	error ratio
5	12.3261		11.9710		11.7843	
9	3.4025	3.6227	3.1946	3.7473	3.0796	3.8266
17	0.4095	8.3082	0.3529	9.0522	0.3239	9.5099
33	0.0226	18.124	0.0132	26.701	0.0093	35.001
65	0.0005	41.333	0.0001	100.06	4.4e-5	211.77
129	9.4e-6	57.742	6.7e-7	199.23	7.2e-8	607.01
257	1.5e-7	62.154	5.4e-9	125.58	2.1e-8	3.3566
513	5.3e-9	28.937	3.1e-9	1.6855	2.5e-8	0.8553
1025	3.2e-9	1.6522	3.3e-9	0.9644	2.7e-8	0.9300

Table 4.1: Numerical error for second-order finite differences and pseudo-spectral differentiation with smoothed initial data.

solved by Reisinger [78], Reisinger and Wittum [80] as well as by Leentvaar and Oosterlee [66] with the help of the sparse grid combination technique. In their works the underlying schemes were based on standard central second- or fourth-order finite differences and Crank-Nicolson time marching. For problems with even more underlying assets usually Monte Carlo methods are applied as they do not suffer from the curse of dimensionality like grid-based methods, e.g. in [68]. Furthermore, dimension reduction techniques can be used to reduce the size of the problem. For more details we refer to [10, 75, 78].

We propose two approaches: firstly, we use a hybrid scheme with Chebyshev differentiation and finite differences. The problem is transformed into a more beneficial PDE, such that the structure allows to use a spectral approximation in the direction of greatest variation. Secondly, we derive a high-order compact scheme and employ the combination technique to derive a fourth-order accurate solution on the sparse grid.

### 4.3.1 Hybrid ADI Scheme

In Section 4.2 it could already be seen that the smoothing of the initial data allows to recover a high rate of convergence. Further variable transformation (4.4) enables to align the discontinuity to one coordinate direction. This is beneficial in two ways: firstly, since the discontinuity only occurs in one coordinate direction the smoothing procedure reduces to a one-dimensional problem. This can be solved easily online or during an offline-phase of the algorithm. Secondly, the other coordinate directions have only a small influence on the overall solution, such that significantly fewer grid nodes can be employed in these directions, see [90]. These observations motivate to use a high-order method in the first coordinate direction and standard finite difference discretizations in the other directions.

In order to be able to concentrate grid nodes in a region of interest we additionally apply

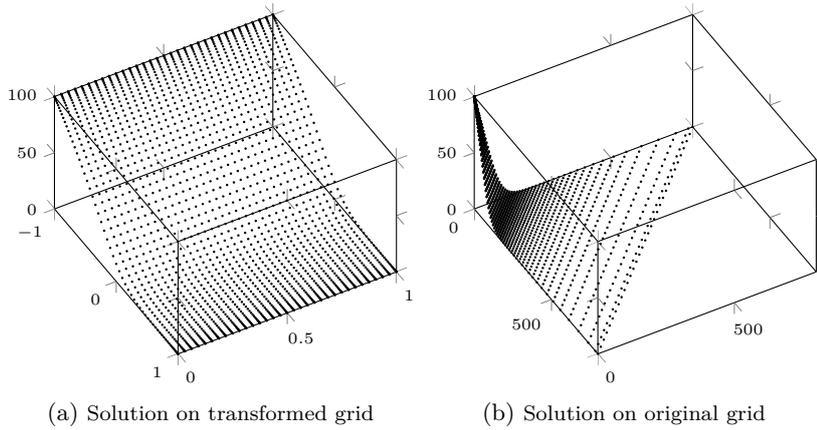


Figure 4.4: 2d basket put option on transformed and original grid.

the transformation (4.3) for the Chebyshev discretization. We cluster grid nodes in the first coordinate direction and are satisfied with a uniform mesh in the other directions. A pure finite difference discretization serves as a benchmark method. Here we apply transformation (4.2). Applying these transformations to equation (4.5), we obtain

$$\begin{aligned} \frac{\partial u}{\partial t} + \alpha_{11} \left[ \psi'(x_1)^2 \frac{\partial^2 u}{\partial y_1^2} + \psi''(x_1) \frac{\partial u}{\partial y_1} \right] + \sum_{i=2}^d (\alpha_{1i} + \alpha_{i1}) \psi'(x_1) \frac{\partial^2 u}{\partial y_1 \partial y_i} \\ + \sum_{\substack{i=2, j=2 \\ i \neq j}}^d \alpha_{ij} \frac{\partial^2 u}{\partial y_i \partial y_j} + \sum_{i=2}^d \beta_i \frac{\partial u}{\partial y_i} + \beta_1 \psi'(x_1) \frac{\partial u}{\partial y_1} - ru = 0, \end{aligned} \quad (4.8)$$

where  $\psi$  is either given by (4.2) or (4.3) and  $x_1 = \psi^{-1}(y_1)$ ,  $x_i = y_i$  for  $i = 2, 3, \dots, d$ . In the Chebyshev case with transformation (4.3) the domain  $\tilde{\Omega}_d$  transforms to  $\tilde{\tilde{\Omega}}_d = [-1, 1] \times [0, 1]^{d-1}$  and under (4.2) to  $\tilde{\tilde{\tilde{\Omega}}}_d = [0, 1] \times [0, 1]^{d-1}$ . By replacing the derivatives with their discrete counterpart we obtain the spatial discretization. In the time domain we use ADI schemes (3.1) - (3.4), where  $F_i$  is the spatial discretization of all terms stemming from the coordinate direction  $i$  for  $i = 1, 2, \dots, d$ . The reaction term is distributed equally over the operators  $F_i$ . All approximations of the mixed derivatives are collected in  $F_0$ . In the explicit legs of the method it is possible to differentiate in Chebyshev transformed space to gain some speedup compared to a differentiation in physical space, cf. Figure 2.4. Figure 4.4 shows the solution of the equation (4.8) in the coordinate system  $y_1, y_2$  and in the original coordinates  $s_1, s_2$ .

**Numerical Experiments** In order to test the proposed numerical method we compute a European put option on the domain

$$\tilde{\tilde{\tilde{\Omega}}}_d = [0, 8K] \times [0, 1]^{d-1}.$$

This choice ensures that the errors arising due to the domain truncation in the first coordinate direction are negligibly small. The numerical solution is evaluated in the

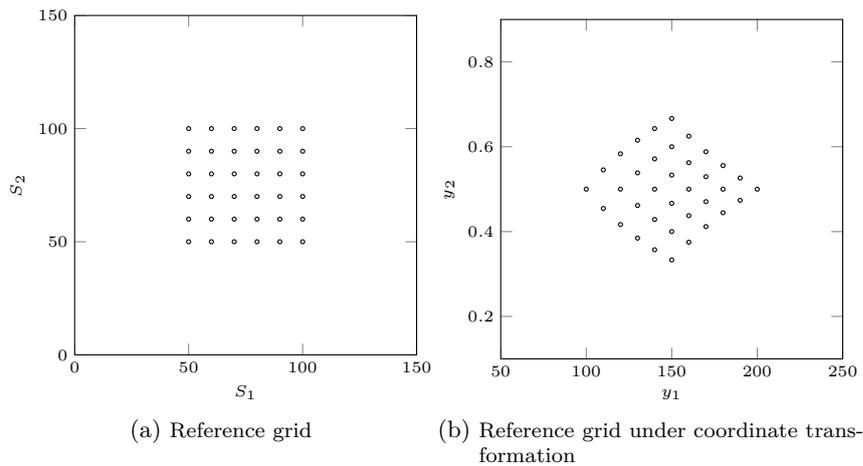


Figure 4.5: Reference grid.

region of interest, defined in the original coordinate system  $s_1, s_2, \dots, s_d$

$$[1/d \cdot K, K]^d.$$

Figure 4.5 shows the region of interest in the original coordinate system and under the coordinate transformation in a two-dimensional setting. The option parameters are chosen to be

$$K = 100, \quad T = 1, \quad r = 0, \quad \sigma_i = 0.3 \text{ for } i = 1, 2, \dots, d,$$

with correlation matrix

$$\rho = \begin{pmatrix} 1 & 0.5 & 0.4 & 0.6 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.4 & 0.5 & 1 & 0.5 \\ 0.6 & 0.5 & 0.5 & 1 \end{pmatrix}.$$

In the hybrid method we cluster grid nodes at  $y_1 = -1$  and use the non-uniformity parameter  $e = 0.01$ . In the finite difference benchmark method we concentrate grid nodes at  $K$  and choose  $\alpha = 150$ . We compute the error (2.4), where the reference solution  $U_{\text{ref}}$  is computed with the hybrid scheme using  $\Delta_t = 10^{-3}$ ,  $\mathbf{N} = (257, 17, \dots, 17)$  and a smoothed initial condition ( $p = 10$ ). In Figures 4.6 - 4.8 we compare the accuracy of the hybrid method (Chebyshev pseudo-spectral (CPS)) to its benchmark method (finite differences (FD)). In comparison to the one-dimensional example in Section 4.2 we observe a higher accuracy due to the concentration of grid points in the critical region for both schemes. The hybrid method clearly outperforms the finite difference scheme in terms of number of grid nodes versus accuracy. However, the high accuracy of the spectral method does not come for free. The discretization matrices of the second-order finite difference ADI scheme are of tridiagonal structure, which results in a linear computational effort. In contrast to this, the Chebyshev differentiation matrices are full and due to the ADI splitting the total effort is  $\mathcal{O}(N_1^2)$  if the number of nodes in the other directions is kept constant. Thus, we observe an increase of the computation time by a factor of four, while we see a doubled computation time for the benchmark method, when

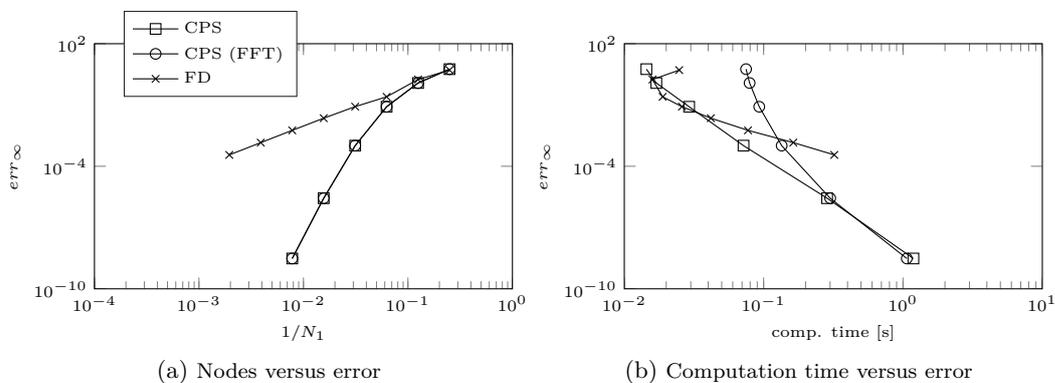


Figure 4.6: Basket with two underlying assets.

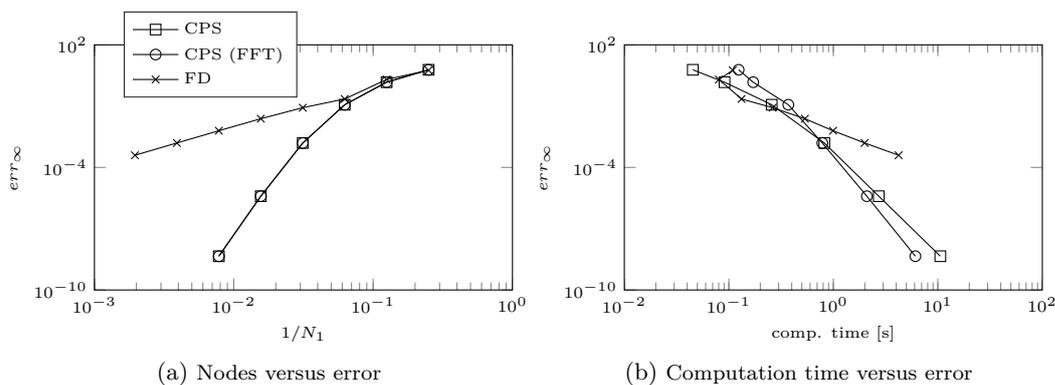


Figure 4.7: Basket with three underlying assets.

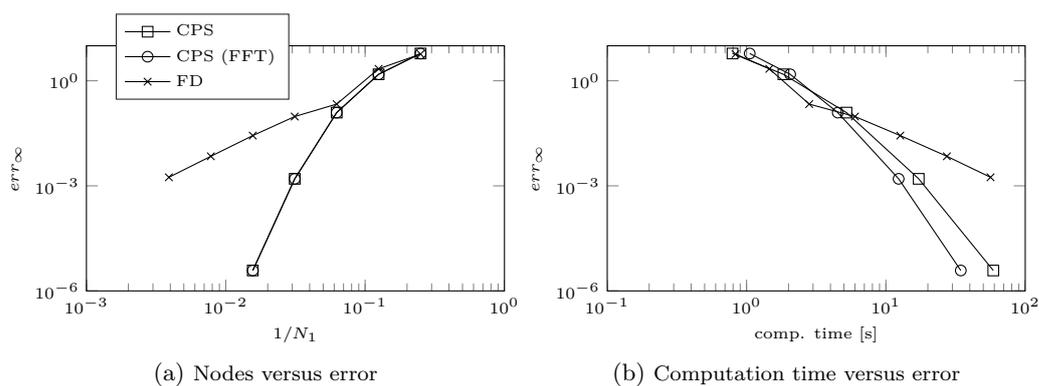


Figure 4.8: Basket with four underlying assets.

the number of nodes doubles. The high accuracy of the hybrid method outweighs this drawback and outperforms the pure finite difference scheme in the high accuracy region. For large problems in a higher-dimensional setting the run-time can be further reduced by differentiating in Chebyshev transformed space. For our numerical test we used the FFT algorithm to compute the mixed derivatives containing the first coordinate direction. Especially in the higher accuracy region of the three- and four-dimensional basket options we see a significant speedup.

### 4.3.2 HO-ADI Schemes

In this section we derive a high-order finite difference approximation of the Black-Scholes PDE (1.5). We apply the logarithmic transformation (4.1)  $x_i = \log(s_i)$  for  $i = 1, \dots, d$ ,  $\tau = T - t$ ,  $u = e^{r\tau} u$  and obtain

$$\frac{\partial u}{\partial \tau} - \frac{1}{2} \sum_{i,j=1}^d \rho_{ij} \sigma_i \sigma_j \frac{\partial^2 u}{\partial x_i \partial x_j} - \sum_{i=1}^d \left( r - \frac{1}{2} \sigma_i^2 \right) \frac{\partial u}{\partial x_i} = 0.$$

The payoff transforms to  $\phi(x_1, \dots, x_d) = \left( K - \sum_{i=1}^d e^{x_i} \right)^+$ . The HOC formulation of  $F_i$  for  $i = 1, \dots, d$  according to Section 2.2 can be derived by inserting  $a_{ii} := \frac{1}{2} \sigma_i^2$  and  $c_i := r - \frac{1}{2} \sigma_i^2$  into equation (2.12). Thus, we obtain the following discretization matrices

$$\begin{aligned} A_{x_i} &= \left( \frac{1}{2} \sigma_i^2 + \frac{h_i^2 (r - 1/2 \sigma_i^2)^2}{6 \sigma_i^2} \right) \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \\ &\quad + \left( r - 1/2 \sigma_i^2 \right) \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i} \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d}, \\ B_{x_i} &= I_{N_1 \cdot N_2 \cdot \dots \cdot N_d} + \frac{h_i^2}{12} \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i}^2 \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d} \\ &\quad + \left( \frac{h_i^2 (r - 1/2 \sigma_i^2)}{6 \sigma_i^2} \right) \cdot I_{N_1} \otimes \dots \otimes I_{N_{i-1}} \otimes D_{FD_i} \otimes I_{N_{i+1}} \otimes \dots \otimes I_{N_d}. \end{aligned}$$

As the spatial domain in each direction is truncated at  $[\epsilon, \log(s_i^{\max})]$ , boundary values have to be prescribed. At  $s_i = 0$  the PDE reduces to a lower dimensional PDE and one can solve this lower dimensional PDE at each lower boundary. This is called the *natural boundary condition* [66, 78]. In the numerical scheme under the logarithmic transformation we also use the *natural boundary condition*. If  $\epsilon > 0$  is chosen sufficiently small, the error, which is introduced due to the domain truncation at the lower boundary, is negligibly small. At the upper boundary we impose Dirichlet boundary conditions and set the option value to zero. This means that the computational domain has to be chosen large enough, such that the option is far out of the money at the upper boundary.

**Numerical Experiments** In numerical experiments we test the proposed HO-ADI methods. Besides the accuracy of the full grid solver we are also interested in whether the usage of sparse grids is beneficial. Here we are especially interested in whether the higher regularity requirements of the combination technique are a restriction in practice. In the

full grid case the derivatives in the truncation error need to be bounded, namely

$$\frac{\partial^6 u}{\partial x_i^6}, \quad \frac{\partial^{10} u}{\partial x_i^5 \partial x_j^5}, \quad \frac{\partial^6 u}{\partial x_i^5 \partial x_j} \quad \text{for } i, j = 1, \dots, d \text{ and } i \neq j$$

arising in (2.13) and (2.16), respectively. According to the results in Section 2 the combination technique requires the mixed derivatives

$$\frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \quad \text{with } \alpha_i \in \{0, 1, \dots, 10\}$$

to be bounded to achieve a fourth-order approximation to the Poisson equation on the entire domain. These derivatives arise due to the anisotropic splitting and the extension of the discrete solution via cubic spline interpolation within the combination technique and do not stem from mixed derivatives in the PDE. In order to relate the option parameters to the expected smoothness of the solution, we compute the mixed fourth derivative  $\|D_{FD_1}^2 \otimes D_{FD_2}^2 U_{\text{approx}}\|_{\infty}$  for a decreasing mesh width as a measure of smoothness. Figure 4.9 suggests that the solution at the final time level becomes smoother for large diffusion and positive correlation. If strong diffusion is present already a resolution of 65 grid points in each dimension seems to be sufficient to capture the maximum of the derivative. For smaller diffusion a higher resolution is required to capture the maximum. If one assumes that analogue results hold for the higher derivatives, arising in the truncation errors of the full grid solver as well as in the combination technique, the theoretical asymptotic rate of convergence will be reached much faster for high diffusion with positive correlation than for small diffusion and negative correlation. Please note that a longer time to maturity also leads to a smoother solution at  $t = 0$ .

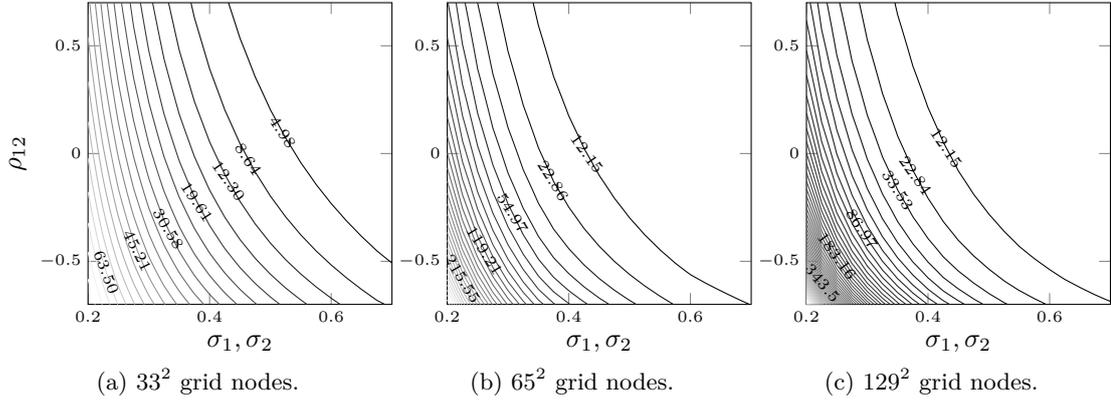
In order to measure the accuracy of the numerical approximation we compute the errors (2.2) and (2.4), where  $U_{\text{ref}}$  is given by a highly accurate reference solution with step size  $\Delta_t$  in time and spatial mesh width  $h$ . The numerical approximation  $U_{\text{approx}}$  is either the full grid solution with time step size  $\tilde{\Delta}_t$  and mesh width  $\tilde{h}$  or the sparse grid solution  $u_n^s$  with time step size  $\tilde{\Delta}_t$  and finest mesh width  $\tilde{h} = 2^{-(n-(d-1)\cdot 3)} \cdot \text{gridlength}$ . The solutions are compared at the final time level on the spatial grid of the reference solution. Therefore, we extend the approximation via multivariate cubic spline interpolation to the grid of the reference solution. Note that the interpolant is fourth-order accurate and hence does not have any negative effect on the rate of convergence. For the experiments in the spatial domain we employ the full grid solver to compute the reference solution in the two-dimensional case. In the higher-dimensional case the combination technique is used to compute the reference solution and then extended to the full grid. The experiments in the temporal domain are always performed on the full grid.

In the following we consider three different test cases given in Table 4.2. In case A we choose a parameter set with large diffusion and a positive correlation. In case B the diffusion coefficients are reduced. Compared to A the correlation in C is decreased, which leads, according to Figure 4.9, to a larger mixed fourth derivative. Thus, case A seems to be the most favorable for the combination technique, while the other two test cases are expected to be more difficult due to the reduced smoothness in these cases. If not mentioned otherwise, we choose  $T = 1$ ,  $K = 20$  and  $r = 0.025$  in the numerical experiments.

In a first experiment we investigate the spatial accuracy of the uniform full grid and of the

	$\sigma_1$	$\sigma_2$	$\sigma_3$	$\rho_{12}$	$\rho_{13}$	$\rho_{23}$
A	0.6	0.6	0.6	0.2	0.2	0.2
B	0.4	0.4	0.4	0.2	0.2	0.2
C	0.6	0.6	0.6	-0.5	0.5	-0.25

Table 4.2: Parameter sets for numerical experiments

Figure 4.9: Maximum of the mixed fourth derivative for a decreasing mesh width  $h$  in the two-dimensional case.

sparse grid, for a basket put with two underlying assets. In Figure 4.10 (a) - (c) the full grid solution exhibits an accuracy close to order four. However, compared to case A one observes a slightly lowered rate of convergence in the more difficult cases B and C. Due to the stronger regularity requirements of the sparse grid combination technique, the rate of convergence shows a greater sensitivity to the smoothness of the solution. Nevertheless, the sparse grid has a higher accuracy per grid node than the full grid in two of the three test cases: only in case C the sparse grid is outperformed.

In Figure 4.11 all numerical schemes show the desired rate of convergence in the time domain. The HDO scheme exhibits order one in time, while the HCS, HMCS and HHV scheme show order two. According to the stability results in Section 3.3.3 we see a stable behavior.

In the following we numerically analyze the performance for basket options with three assets. The full grid solution in Figure 4.12 states a convergence rate close to order four. Similar to the two-dimensional problems, the rate of convergence of the sparse grid

Scenario		A	B	C	Scenario		A	B	C
$err_2$	sparse grid	3.70	3.34	2.34	$err_2$	sparse grid	3.55	3.15	2.52
	full grid	3.85	3.71	3.77		full grid	3.69	3.48	3.64
$err_\infty$	sparse grid	3.39	2.98	2.05	$err_\infty$	sparse grid	3.32	2.68	2.01
	full grid	3.79	3.62	3.52		full grid	3.57	3.28	3.44

(a) 2d

(b) 3d

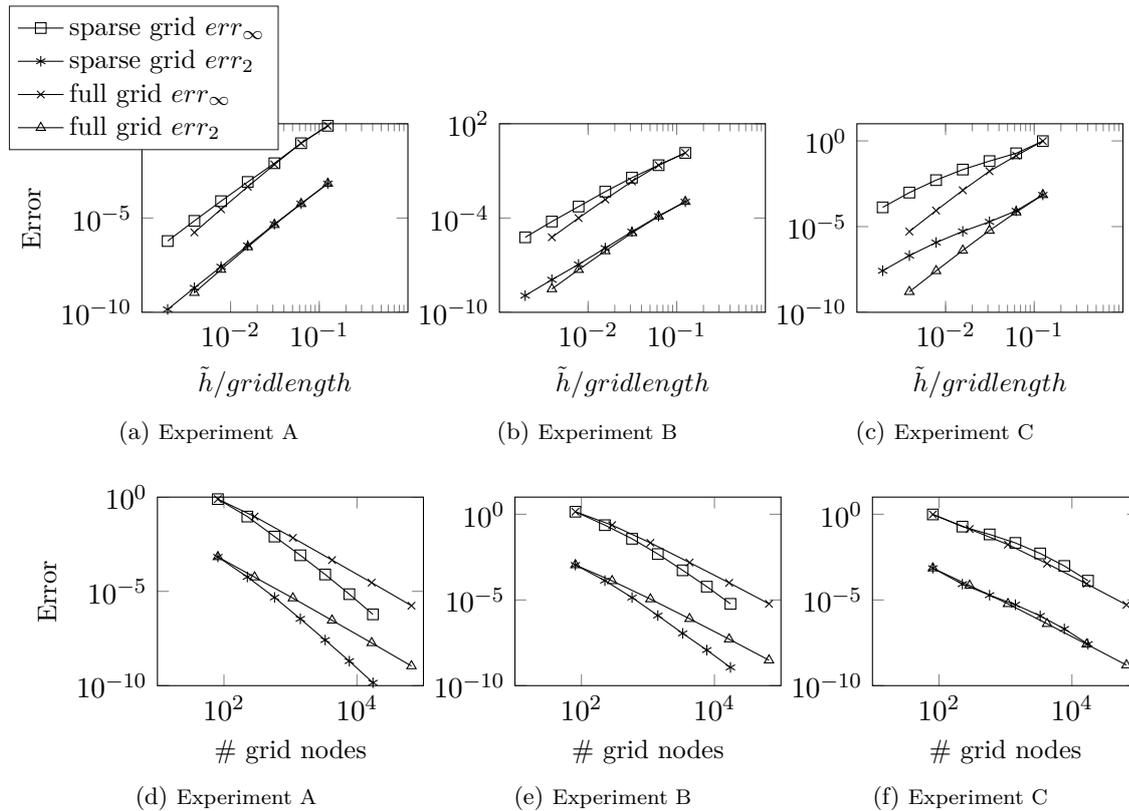
Table 4.3: Estimated order of convergence in space (cf. Figures 4.10 and 4.12).

Scenario		A	B	C	Scenario		A	B	C
$err_2$	HDO	1.18	1.14	1.06	$err_2$	HDO	1.09	1.02	1.04
	HCS	2.15	2.09	2.15		HCS	2.07	2.01	2.07
	HMCS	2.02	2.02	2.03		HMCS	2.07	2.01	2.07
	HHV	2.16	2.14	1.97		HHV	2.10	2.08	1.95
$err_\infty$	HDO	1.15	1.09	1.07	$err_\infty$	HDO	1.02	1.02	1.11
	HCS	2.28	2.15	2.19		HCS	2.14	2.02	2.05
	HMCS	2.06	2.05	2.10		HMCS	2.13	2.02	2.08
	HHV	2.14	2.07	2.05		HHV	2.15	2.03	1.97

(a) 2d

(b) 3d

Table 4.4: Estimated order of convergence in time (cf. Figures 4.11 and 4.14).

Figure 4.10: 2d spatial error for  $\tilde{h} \rightarrow h$ .

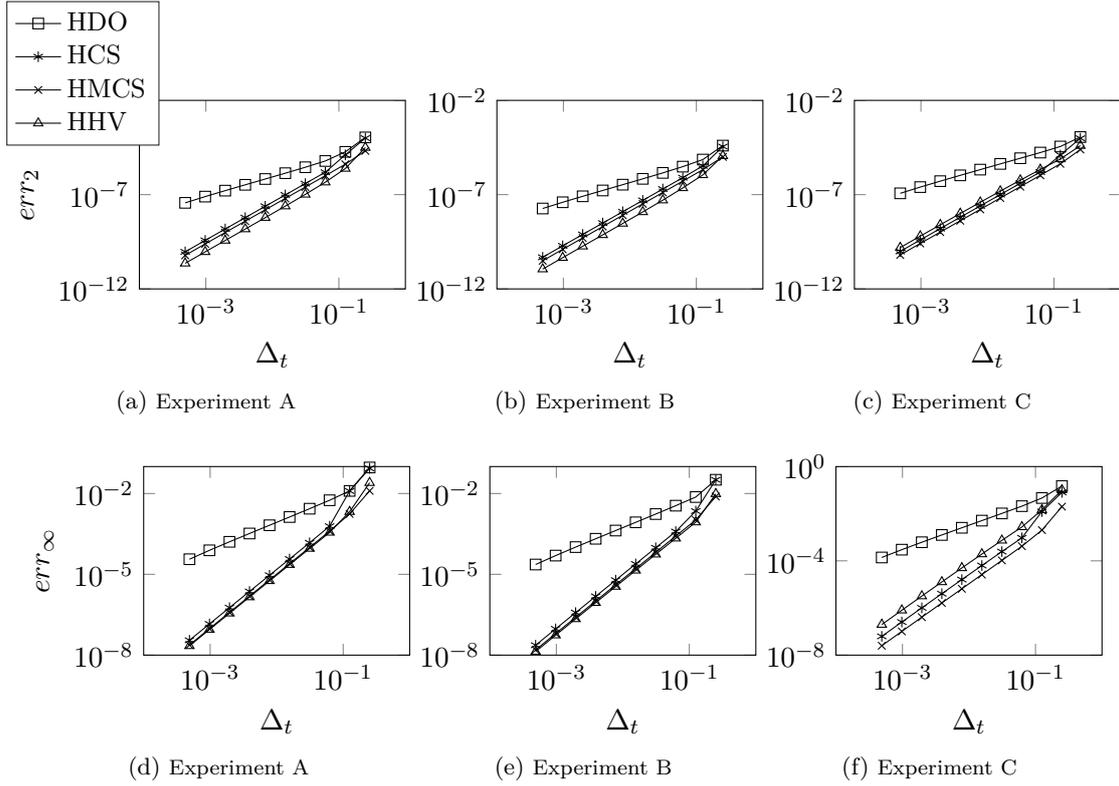


Figure 4.11: 2d temporal error for  $\tilde{\Delta}_t \rightarrow \Delta_t = 2^{-14}$ . The following values of  $\theta$  were used: HDO  $\theta = 0.5$ , HCS  $\theta = 0.5$ , HMCS  $\theta = 0.334$ , HHV  $\theta = 0.79$ . The spatial discretization is computed on a grid with 129 nodes in both coordinate directions.

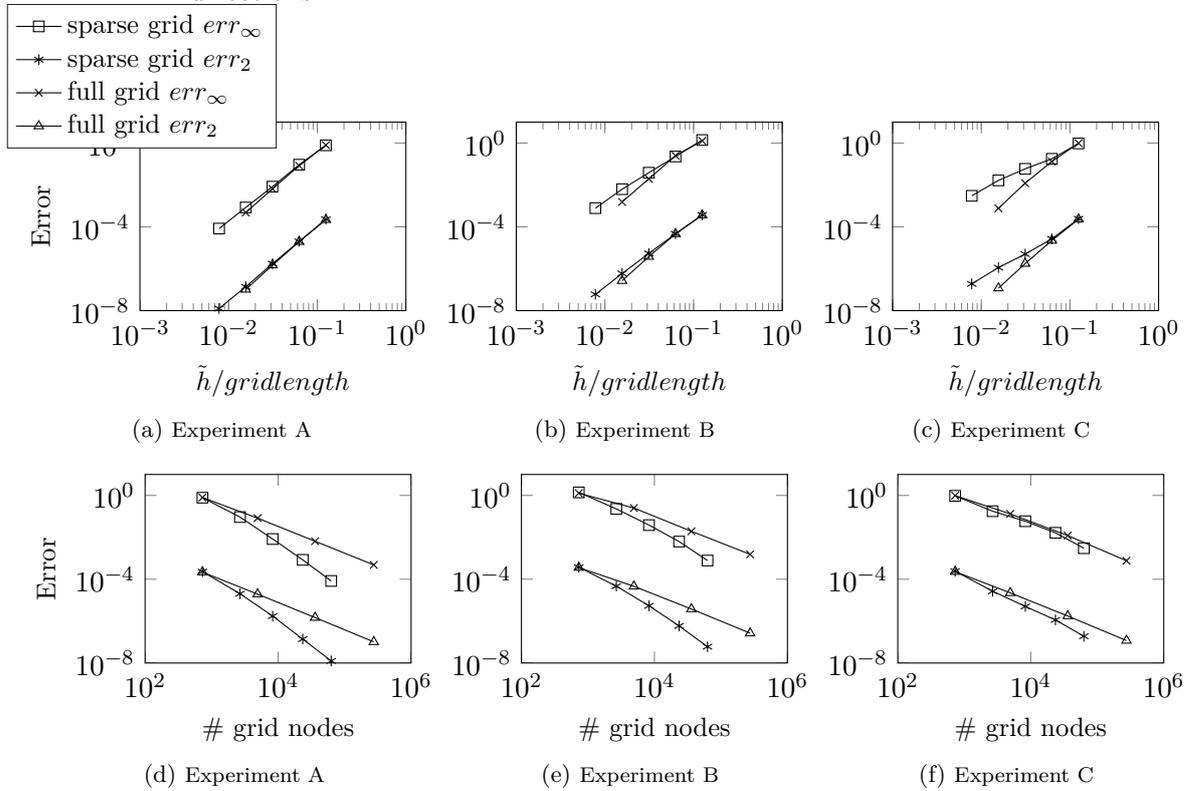


Figure 4.12: 3d spatial error for  $\tilde{h} \rightarrow h$ . The sparse grid solution at level 14 is used as a reference solution and evaluated on a full grid with  $129^3$  grid points.

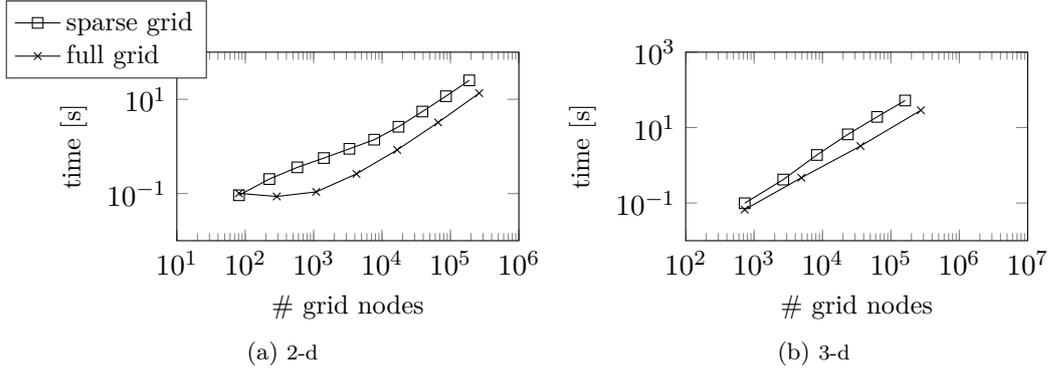


Figure 4.13: Number of grid nodes versus computation time in the case of 100 time steps in the HCS scheme.

solution is more sensitive to the parameter changes in cases B and C than the full grid solver. If we consider the accuracy per grid node, the sparse grid error is lower than the full grid error. Figure 4.13 indicates that the run-time per node is only slightly higher for the combination technique than for the full grid solver and thus the sparse grid is more efficient in the asymptotic. Both plots show that the time increases with an order of approximately one as would be expected due to the ADI time stepping.

For the parameter choices A-C there is no convection dominance ( $|\hat{p}| \geq 36/31$ ). Hence, we expect from the stability region plots in Figure 3.4 a stable behavior. The numerical experiments in Figure 4.14 confirm this and the error decreases monotonically with a rate of accuracy according to the theory.

*Remark:*

During our numerical tests the initial condition is smoothed according to the convolutions described above with  $p = 4$ . Düring and Heuer [24] suggest to only smooth the grid points around the discontinuity to reduce the computational workload. Doing so the full grid performed according to the theoretical results, but the sparse grid solution showed oscillations near the discontinuity in our numerical experiments. This issue could be cured by smoothing all grid points. In the case of the above given payoff under the logarithmic transformed variables an analytical solution to the integral (4.7) is available if the domain of integration does not intersect the discontinuity. Thus, smoothing the initial condition on the entire grid does not introduce a large additional computational effort. In order to smooth the initial condition for our numerical experiments, we either solve the integral analytically if possible or use the Matlab<sup>®</sup> built-in routine *integral2* and *integral3*, respectively.

## 4.4 Stochastic Volatility Models

In this section we derive numerical schemes for stochastic volatility models. As test cases for two- and three-factor models we consider the Heston PDE (1.6)

$$\frac{\partial u}{\partial t} = \frac{1}{2} s^2 v \frac{\partial^2 u}{\partial s^2} + \rho_{12} \sigma_1 s v \frac{\partial^2 u}{\partial s \partial v} + \frac{1}{2} \sigma_1^2 v \frac{\partial^2 u}{\partial v^2} + r s \frac{\partial u}{\partial s} + \kappa(\eta - v) \frac{\partial u}{\partial v} - r u,$$

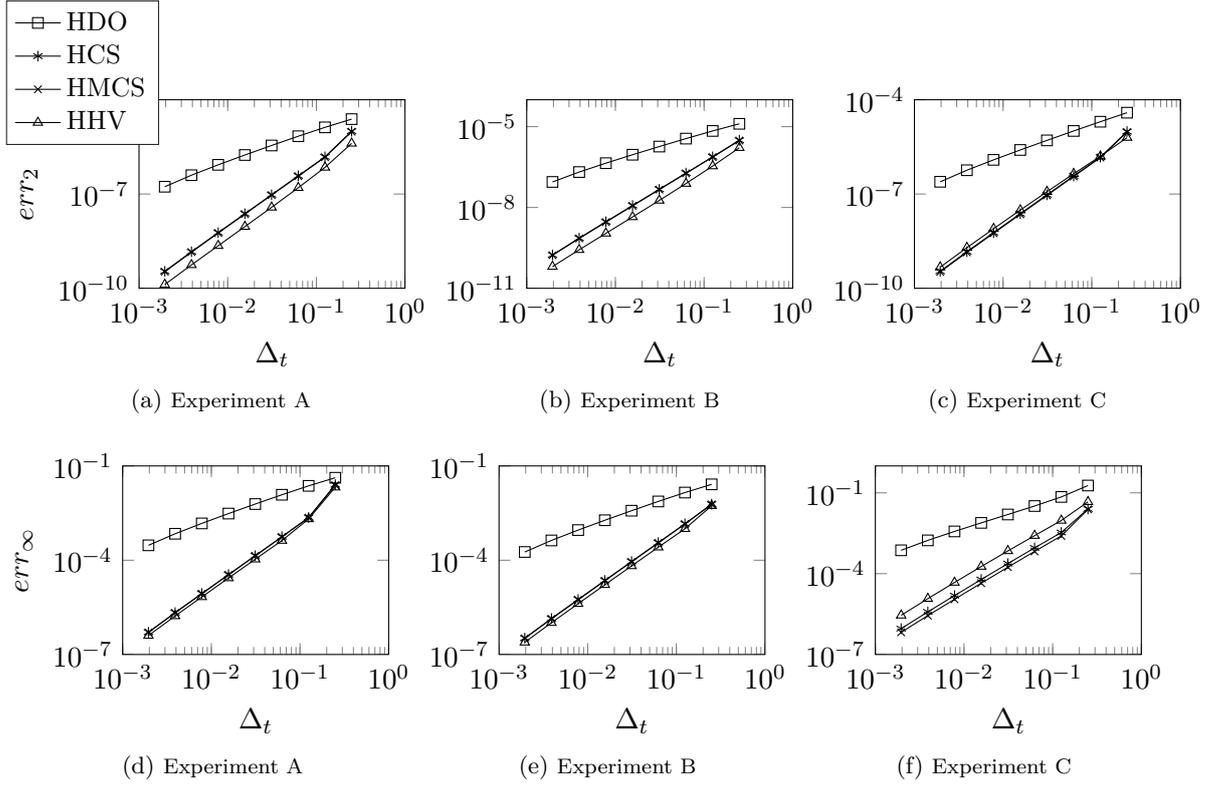


Figure 4.14: 3d temporal error for  $\tilde{\Delta}_t \rightarrow \Delta_t = 2^{-11}$ . The following values of  $\theta$  were used: HDO  $\theta = 0.67$ , HCS  $\theta = 0.5$ , HMCS  $\theta = 0.462$ , HHV  $\theta = 0.79$ . The spatial discretization is computed on a grid with 65 nodes in all coordinate directions.

as well as the HHW PDE (1.7)

$$\begin{aligned} \frac{\partial u}{\partial t} = & \frac{1}{2}s^2v \frac{\partial^2 u}{\partial s^2} + \frac{1}{2}\sigma_1^2v \frac{\partial^2 u}{\partial v^2} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 u}{\partial r^2} \\ & + \rho_{12}\sigma_1sv \frac{\partial^2 u}{\partial s\partial v} + \rho_{13}\sigma_2s\sqrt{v} \frac{\partial^2 u}{\partial s\partial r} + \rho_{23}\sigma_1\sigma_2\sqrt{v} \frac{\partial^2 u}{\partial v\partial r} \\ & + rs \frac{\partial u}{\partial s} + \kappa(\eta - v) \frac{\partial u}{\partial v} + a_r(b_r - r) \frac{\partial u}{\partial r} - ru, \end{aligned}$$

for inverse time  $t \in [0, T]$ , asset  $s \in [0, \infty)$ , volatility  $v \in [0, \infty)$  and risk-free interest rate  $r \in (-\infty, \infty)$ . We solve the pricing problem for European plain-vanilla put options with the terminal value

$$u(s, v, r, 0) = (K - s)^+.$$

In the literature several methods have been discussed to solve problems (1.6) and (1.7) for vanilla option pricing problems. They range from semi-closed approximations [46, 50], Fourier-cosine [26, 35] and tree approaches [6, 7, 27, 28] to finite difference methods [38, 51, 60]. Kluge [60] has solved the Heston PDE via second-order finite differences. In [38, 51, 60] Alternating Direction Implicit (ADI) time stepping has been used to efficiently deal with the mixed derivative term. High-order compact finite differences were proposed by [21, 22]. Spectral methods were used in [74].

The discontinuity of the payoff profile for option pricing problems in general occurs in the direction of the underlying asset, while in the directions of the other risk factors the solution is smooth. We want to exploit this structure and propose a hybrid scheme which uses a second-order central finite difference approximation in the direction where the discontinuity occurs. In the other spatial dimensions we employ a high-order Chebyshev spectral approximation.

#### 4.4.1 Hybrid Finite Difference/Pseudo-Spectral Method

Similar to the previous methods we apply a coordinate transformation to cluster grid points in the region of interest. We apply the transformation (4.2)

$$\psi_s(s) = \frac{c_1 + \sinh^{-1}\left(\frac{K-s}{\alpha}\right)}{c_1 - c_2}$$

where

$$\begin{aligned} c_1 &= \sinh^{-1}\left(\frac{s_{\min}-K}{\alpha}\right), \\ c_2 &= \sinh^{-1}\left(\frac{s_{\max}-K}{\alpha}\right). \end{aligned}$$

The transformation maps  $[s_{\min}, s_{\max}]$  to  $[0, 1]$  and clusters grid points around the strike price  $K$ . In our numerical tests we use  $\alpha = K/4$  and a uniform grid spacing in  $[0, 1]$ .

In the coordinate direction of the volatility and interest rate we apply transformation (4.3). We denote with  $\psi_j$  the transform in direction  $j$  for  $j \in \{v, r\}$ . The inverse of the transformations are denoted by  $f_s = \psi_s^{-1}$ ,  $f_j = \psi_j^{-1}$ , respectively. Numerical tests revealed that a clustering at the upper boundary of the domain and the choice  $e_1 = 10\sigma_1^2/(\kappa\eta)$  and  $e_2 = 10\sigma_2^2/(a_r b_r)$  yields good results. In the numerical scheme we use a Chebyshev-Gauss-Lobatto grid in the transformed intervals  $[-1, 1]$  stemming from the  $v$  and  $r$  directions, respectively.

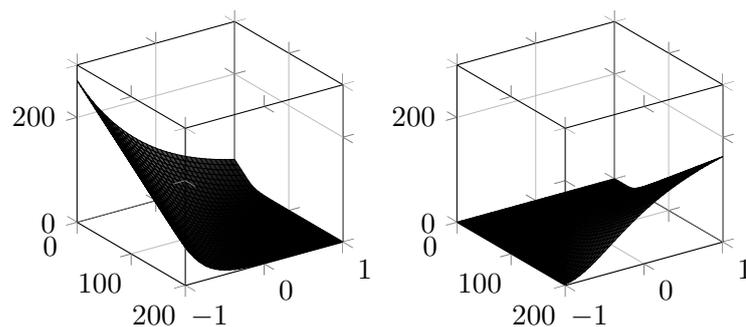
The PDEs (1.6) and (1.7) transform to

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2}s^2v \left[ \psi'_s(s)^2 \frac{\partial^2 u}{\partial x_1^2} + \psi''_s(s) \frac{\partial u}{\partial x_1} \right] + \rho_{12}\sigma_1sv\psi'_s(s)\psi'_v(v) \frac{\partial^2 u}{\partial x_1\partial x_2} + \frac{1}{2}\sigma_1^2v \left[ \psi'_v(v)^2 \frac{\partial^2 u}{\partial x_2^2} + \psi''_v(v) \frac{\partial u}{\partial x_2} \right] \\ &\quad + rs\psi'_s(s) \frac{\partial u}{\partial x_1} + \kappa(\eta - v)\psi'_v(v) \frac{\partial u}{\partial x_2} - ru \end{aligned} \quad (4.9)$$

and

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2}s^2v \left[ \psi'_s(s)^2 \frac{\partial^2 u}{\partial x_1^2} + \psi''_s(s) \frac{\partial u}{\partial x_1} \right] + \frac{1}{2}\sigma_1^2v \left[ \psi'_v(v)^2 \frac{\partial^2 u}{\partial x_2^2} + \psi''_v(v) \frac{\partial u}{\partial x_2} \right] + \frac{1}{2}\sigma_2^2 \left[ \psi'_r(r)^2 \frac{\partial^2 u}{\partial x_3^2} + \psi''_r(r) \frac{\partial u}{\partial x_3} \right] \\ &\quad + \rho_{12}\sigma_1sv\psi'_s(s)\psi'_v(v) \frac{\partial^2 u}{\partial x_1\partial x_2} + \rho_{13}\sigma_2s\sqrt{v}\psi'_s(s)\psi'_r(r) \frac{\partial^2 u}{\partial x_1\partial x_3} + \rho_{23}\sigma_1\sigma_2\sqrt{v}\psi'_v(v)\psi'_r(r) \frac{\partial^2 u}{\partial x_2\partial x_3} \\ &\quad + rs\psi'_s(s) \frac{\partial u}{\partial x_1} + \kappa(\eta - v)\psi'_v(v) \frac{\partial u}{\partial x_2} + a_r(b_r - r)\psi'_r(r) \frac{\partial u}{\partial x_3} - ru, \end{aligned} \quad (4.10)$$

where  $s = f_s(x_1)$ ,  $v = f_v(x_2)$  and  $r = f_r(x_3)$  with  $(x_1, x_2) \in \tilde{\Omega} = [0, 1] \times [-1, 1]$  and  $(x_1, x_2, x_3) \in \tilde{\Omega} = [0, 1] \times [-1, 1]^2$ . At the boundary we impose the following conditions



(a) European put option at  $v = 0.125$  (b) European call option at  $v = 0.125$

Figure 4.15: Reference solution of the HHW model (computed with semi closed-form pricing formula).

for the European put option under the Heston model

$$\begin{aligned} u(0, v, t) &= Ke^{-rT}, \\ u(s_{\max}, v, t) &= 0, \\ \frac{\partial u}{\partial v}(v_{\max}, s, t) &= 0, \end{aligned}$$

and under the HHW model

$$\begin{aligned} u(0, v, r, t) &= Ke^{p(r,t)}, \\ u(s_{\max}, v, r, t) &= 0, \\ \frac{\partial u}{\partial v}(s, v_{\max}, r, t) &= 0, \\ \frac{\partial u}{\partial r}(s, v, r_{\max}, tu) &= 0, \end{aligned}$$

with the discounting factor

$$\begin{aligned} p(r, t) &= -\frac{r}{a_r} (1 - e^{-a_r t}) - \frac{1}{a_r} \int_t^T b_r (1 - e^{-a_r(T-s)}) ds \\ &+ \frac{\sigma_2^2}{2a_r^2} \left( t + \frac{2}{a_r} e^{-a_r t} - \frac{1}{2a_r} e^{-2a_r t} - \frac{3}{2a_r} \right). \end{aligned}$$

If the asset price is zero the option price is given by the discounted strike price. For sufficiently large  $s$  the probability that the put option ends up in-the-money tends to zero and therefore also the option value. In the direction of the volatility we only imply a homogeneous Neumann boundary condition at  $v_{\max}$  as suggested in [51]. At the boundary in the direction of the interest rate, we propose a homogeneous Neumann boundary condition at  $r_{\max}$ . One might argue that such a condition should be applied at both boundaries of  $r$  since  $\rho$  in the Black-Scholes pricing formula vanishes for extreme values of  $r$ , but the reference solution in Figure 4.15 indicates that this does not hold for the HHW model.

The application of the spatial discretization of Section 2 to (4.9), (4.10) yields

$$\begin{aligned}
F_{Heston} = & \frac{1}{2} \text{diag}(S^2 \psi'_s(S)^2 \otimes V) D_{FD_1}^2 \otimes I_{N_2} + \frac{1}{2} \sigma_1^2 \text{diag}(e_{N_1} \otimes V \psi'_v(V)^2) I_{N_1} \otimes D_{SP_2}^2 \\
& + \rho_{12} \sigma_1 \text{diag}(S \psi'_s(S) \otimes V \psi'_v(V)) D_{FD_1} \otimes D_{SP_2} \\
& + \left[ \frac{1}{2} \text{diag}(S^2 \psi''_s(S) \otimes V) + r \text{diag}(S \psi'_s(S) \otimes e_{N_2}) \right] D_{FD_1} \otimes I_{N_2} \\
& + \left[ \frac{1}{2} \sigma_1^2 \text{diag}(e \otimes (V \psi''_v(V) + \kappa(\eta - V) \psi'_v(V))) \right] I_{N_1} \otimes D_{SP_2} - r
\end{aligned}$$

and

$$\begin{aligned}
F_{HFW} = & \frac{1}{2} \text{diag}(S^2 \psi'_s(S)^2 \otimes V \otimes e_{N_3}) D_{FD_1}^2 \otimes I_{N_2} \otimes I_{N_3} \\
& + \frac{1}{2} \sigma_1^2 \text{diag}(e_{N_1} \otimes V \psi'_v(V)^2 \otimes e_{N_3}) I_{N_1} \otimes D_{SP_2}^2 \otimes I_{N_3} \\
& + \frac{1}{2} \sigma_2^2 \text{diag}(e_{N_1} \otimes e_{N_2} \otimes \psi'_r(R)^2) I_{N_1} \otimes I_{N_2} \otimes D_{SP_3}^2 \\
& + \rho_{12} \sigma_1 \text{diag}(S \psi'_s(S) \otimes V \psi'_v(V) \otimes e_{N_3}) D_{FD_1} \otimes D_{SP_2} \otimes I_{N_3} \\
& + \rho_{13} \sigma_2 \text{diag}(S \psi'_s(S) \otimes \sqrt{V} \otimes \psi'_r(r)) D_{FD_1} \otimes I_{N_2} \otimes D_{SP_3} \\
& + \rho_{23} \sigma_1 \sigma_2 \text{diag}(e_{N_1} \otimes \sqrt{V} \psi'_v(V) \otimes \psi'_r(R)) I_{N_1} \otimes D_{SP_2} \otimes D_{SP_3} \\
& + \left[ \frac{1}{2} \text{diag}(S^2 \psi''_s(S) \otimes V \otimes e_{N_3}) + \text{diag}(S \psi'_s(S) \otimes e_{N_2} \otimes R) \right] D_{FD_1} \otimes I_{N_2} \otimes I_{N_3} \\
& + \left[ \text{diag}(e_{N_1} \otimes \left( \frac{1}{2} \sigma_1^2 V \psi''_v(V) + \kappa(\eta - V) \psi'_v(V) \right) \otimes e_{N_3}) \right] I_{N_1} \otimes D_{SP_2} \otimes I_{N_3} \\
& + \left[ \text{diag}(e_{N_1} \otimes e_{N_2} \otimes \left( \frac{1}{2} \sigma_2^2 \psi''_r(R) + a_r(b_r - R) \psi'_r(R) \right)) \right] I_{N_1} \otimes I_{N_2} \otimes D_{SP_3} \\
& - \text{diag}(e_{N_1} \otimes e_{N_2} \otimes R),
\end{aligned}$$

where  $\text{diag}$  is a diagonal matrix and each operation in the  $\text{diag}$  operator is understood component-wise.  $e_{N_i} = (1, 1, \dots, 1)^\top$  for  $i = 1, 2, 3$  denotes the all-ones vector of size  $N_1 + 1$  in  $x_1$ -direction, of size  $N_2 + 1$  in  $x_2$  and of size  $N_3 + 1$  in  $x_3$  direction. The spatial grid vector in  $s$ -direction is given by  $S \in \mathbb{R}^{N_1+1}$ , while  $V \in \mathbb{R}^{N_2+1}$  and  $R \in \mathbb{R}^{N_3+1}$  are the grid vectors in the  $v$ - $r$  directions, respectively.

The spatial discretization is now decomposed into one-dimensional problems according to the ADI splitting:  $F_1$  takes all terms, which only stem from the  $x_1$ -direction,  $F_2$  all terms from the  $x_2$ -direction and  $F_3$  all terms from the  $x_3$ -direction. The reaction term is distributed equally over the operators  $F_i$  for  $i = 1, 2, 3$ . The mixed derivative terms are collected in  $F_0$ . The arising linear system of equations can be solved with the help of a LU decomposition in the startup phase if the coefficients of the PDEs (1.6) and (1.7) do not depend on time. In each time step the major computational effort then consists of performing one forward and one backward substitution for each leg of the ADI scheme.

In order to evaluate the performance of the scheme, we compare it to a scheme using second-order finite differences in all coordinate directions given in the articles [38, 51], but with a transformed coordinate system instead of a non-uniform grid. The following transformation is employed in  $v$  and  $r$  direction in the benchmark method

$$\psi_j(w) = \sinh^{-1}(d_j^{-1}(w - c_j)),$$

with the critical point  $c_j$  and the strength of smoothing determined by  $d_j$  for  $j \in \{v, r\}$ . According to [38, 51] we use  $c_v = 0$ ,  $c_r = b_r$ ,  $d_v = v_{\max}/500$  and  $d_r = r_{\max}/500$ .

	Case 1	Case 2	Case 3	Case 4
$K$	100	100	100	100
$T$	1	1	3	0.5
$\sigma_1$	0.3	0.04	0.2928	0.5
$\rho_{12}$	-0.9	-0.6	-0.7571	-0.5
$\kappa$	1.5	3	0.6067	2
$\eta$	0.04	0.12	0.0707	0.02
$r$	0.025	0.04	0.03	0.01
$a_r$	0.00883	0.2	0.05	0.15
$b_r$	0.025	0.05	0.055	0.101
$\sigma_2$	0.00631	0.06	0.03	0.1
$\rho_{13}$	0.6	0.2	0.6	-0.3
$\rho_{23}$	-0.7	0.4	-0.2	0.2

Table 4.5: Scenarios for numerical tests.

**Numerical Experiments** In this section we test the hybrid method and compare it to the second-order benchmark method. In order to gain realistic performance results and to test the robustness of the method we consider four different scenarios given in Table 4.5. The parameters for the Heston model used in the first three scenarios stem from [51]. Compared to the original parameters we use a negative correlation  $\rho_{12}$  in scenario 2 since a positive correlation is not realistic and can lead to an explosion of the moments, cf. [1, 59]. The additional parameters for the HHW model have been taken from [87] in the case of scenario 1 and the parameters in the scenarios 2 and 3 are taken from [38]. In scenario 4 we have chosen the parameters in such a way that the Feller condition  $2\kappa\eta > \sigma_1^2$  is violated.

We investigate the accuracy in the time domain as well as the spatial error. Therefore, we compute the error (2.4), where  $U_{\text{approx}}$  is the numerical approximation on the discrete grid with time step  $\Delta_t$  and  $\mathbf{N} = (N_1, N_2) = (2^{l_1}, 2^{l_2})$ ,  $\mathbf{N} = (N_1, N_2, N_3) = (2^{l_1}, 2^{l_2}, 2^{l_3})$ , respectively, grid nodes in the spatial domain. The error is always computed at the final time slice. For the sake of simplicity and to streamline our notation we choose  $N_2 = N_3$  in all numerical experiments and write  $\mathbf{N} = (N_1, N_2)$  instead of  $\mathbf{N} = (N_1, N_2, N_3)$ . The numerical solution is computed on a grid of size  $[0, 20K] \times [0, 1.5] \times [-0.5, 0.5]$  while the error is computed in the region of interest, which is defined as  $[0, 2K] \times [0, 1] \times [0, 0.125]$ . This choice ensures that the error due to the domain truncation in the asset direction and the error stemming from the homogeneous Neumann boundary conditions are negligibly small and one does not observe any saturation effects in the numerical convergence plots. The hybrid HHW (Heston in brackets) CS method takes about 0.10 (0.015), 0.71 (0.05), 4.65 (0.24), 85 (1.9) seconds for the following number of grid points and time steps  $\mathbf{N} = (65, 9)$ ,  $T/\Delta_t = 25$ ,  $\mathbf{N} = (129, 13)$ ,  $T/\Delta_t = 50$ ,  $\mathbf{N} = (257, 17)$ ,  $T/\Delta_t = 100$ ,  $\mathbf{N} = (512, 33)$ ,  $T/\Delta_t = 100$ , respectively, to compute the solution. The MCS and HV scheme have approximately the same run-time, while the Douglas method has just one implicit sweep and therefore the computation takes about half of the run-time.

In our first numerical experiment we investigate the error decay for  $\Delta_t \rightarrow 0$ . Here the reference solution is given by a highly accurate numerical approximation with  $\Delta_t = 2^{-13}$  and  $\mathbf{N} = (129, 33)$  grid nodes. The  $\theta$  value within the ADI procedure is always chosen according to the lowest possible value ensuring unconditional stability, given in Section

	DO	CS	MCS	HV
Heston	0.5	0.5	0.34	0.79
HHW	0.67	0.5	$\max\{\frac{1}{3}, \frac{2}{13}(2\gamma + 1)\}$	0.79

Table 4.6:  $\theta$ -values used in the ADI methods within the numerical experiments, where  $\gamma = \max\{|\rho_{12}|, |\rho_{13}|, |\rho_{23}|\}$ .

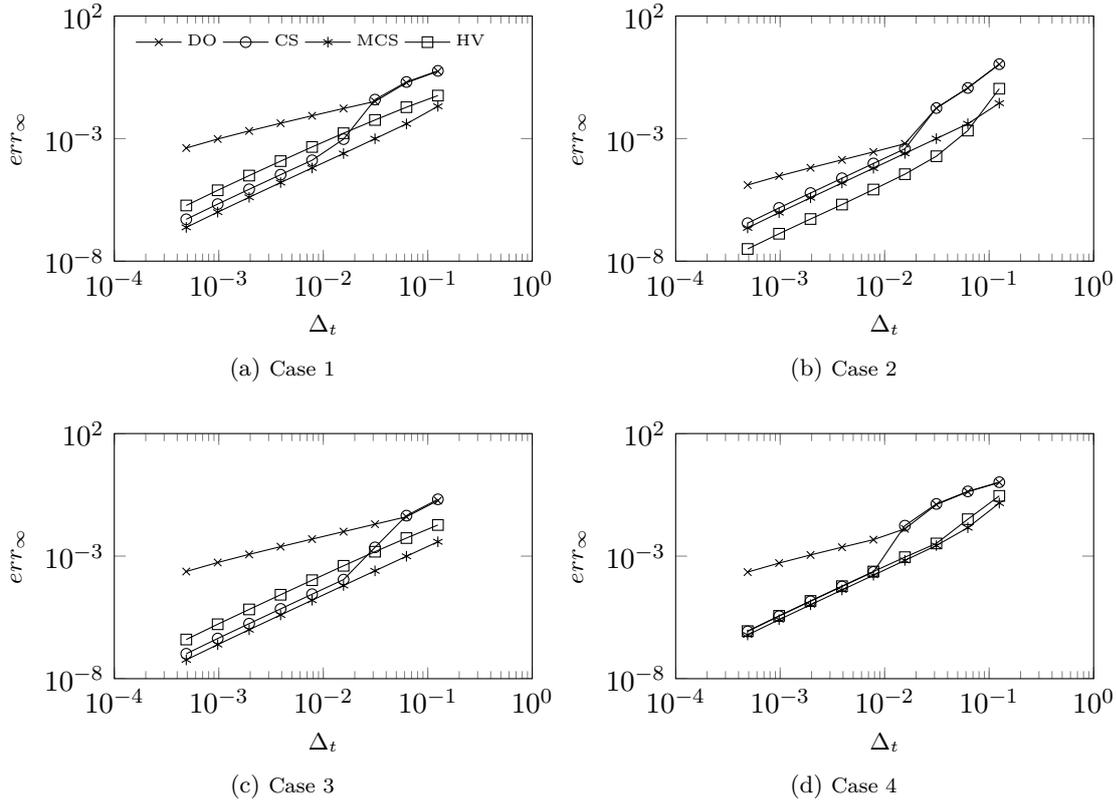


Figure 4.16: Heston: convergence in time. 129 grid points in  $S$  and 33 in  $v$ -direction.

3.3, see Table 4.6. Please note that we choose  $\theta = 0.34$  in case of the MCS scheme for the Heston model as it holds for the correlation  $|\rho_{12}| \leq 0.96$  for all test scenarios. Although these bounds have been derived for finite difference schemes in the von Neumann framework, the positive results of Section 3.5 encourage that these are also valid for the Chebyshev spectral method. Figures 4.16 and 4.18 show that the error decays monotonically both for the Heston and the HHW model. The DO scheme exhibits order one, while the errors of the CS, MCS and HV scheme decrease with second-order. If the time step  $\Delta_t$  is very large the schemes show an undesirable high error. Especially the DO and CS scheme suffer from a large error in all four test scenarios. In order to improve the results we employ a kind of Rannacher startup [29] and perform four steps with  $\Delta_t/4$  and  $\theta = 1$  to damp high frequency errors, which arise due to the non-smooth initial condition. The Figures 4.17 and 4.19 show that the startup procedure is capable to smooth the error and thus leads to a much smaller error for large time steps.

For the experiments in the spatial domain we use the semi closed-form solution to the Heston and Heston-Hull-White PDE from [46], [50] as a reference solution. In the case of

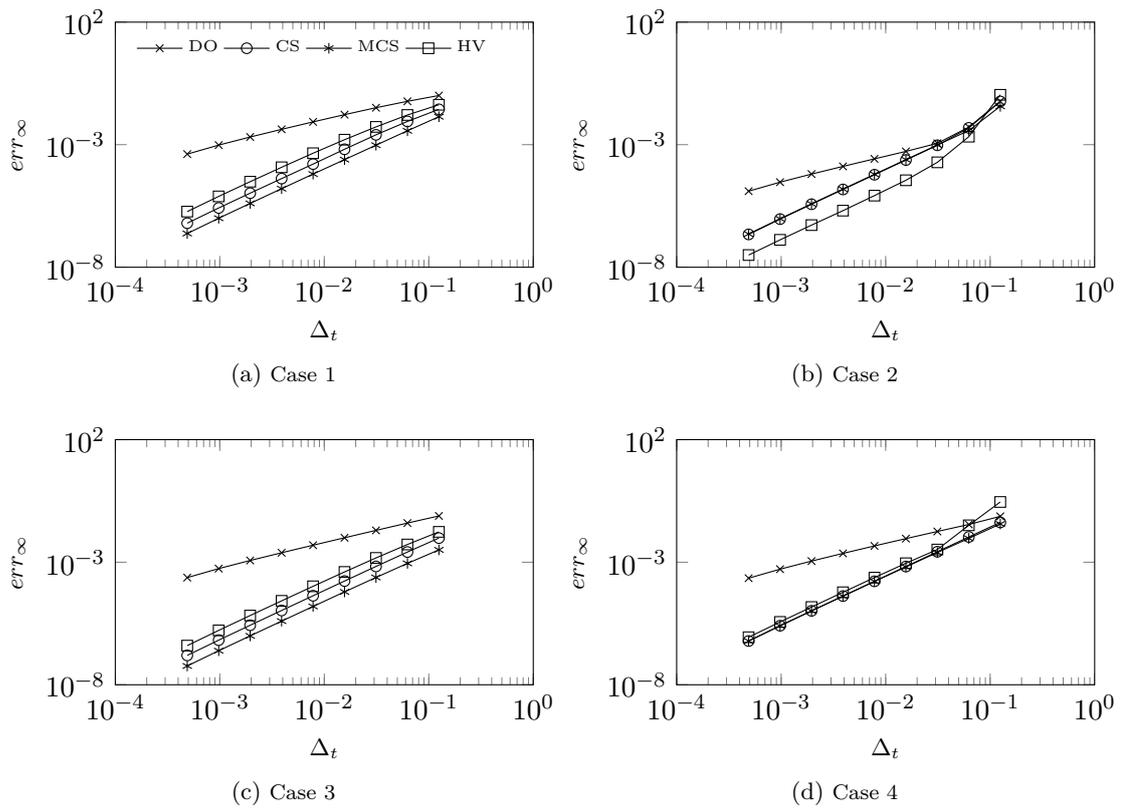


Figure 4.17: Heston: convergence in time with four initial steps using  $\theta = 1$ . 129 grid points in  $S$  and 33 in  $v$ -direction.

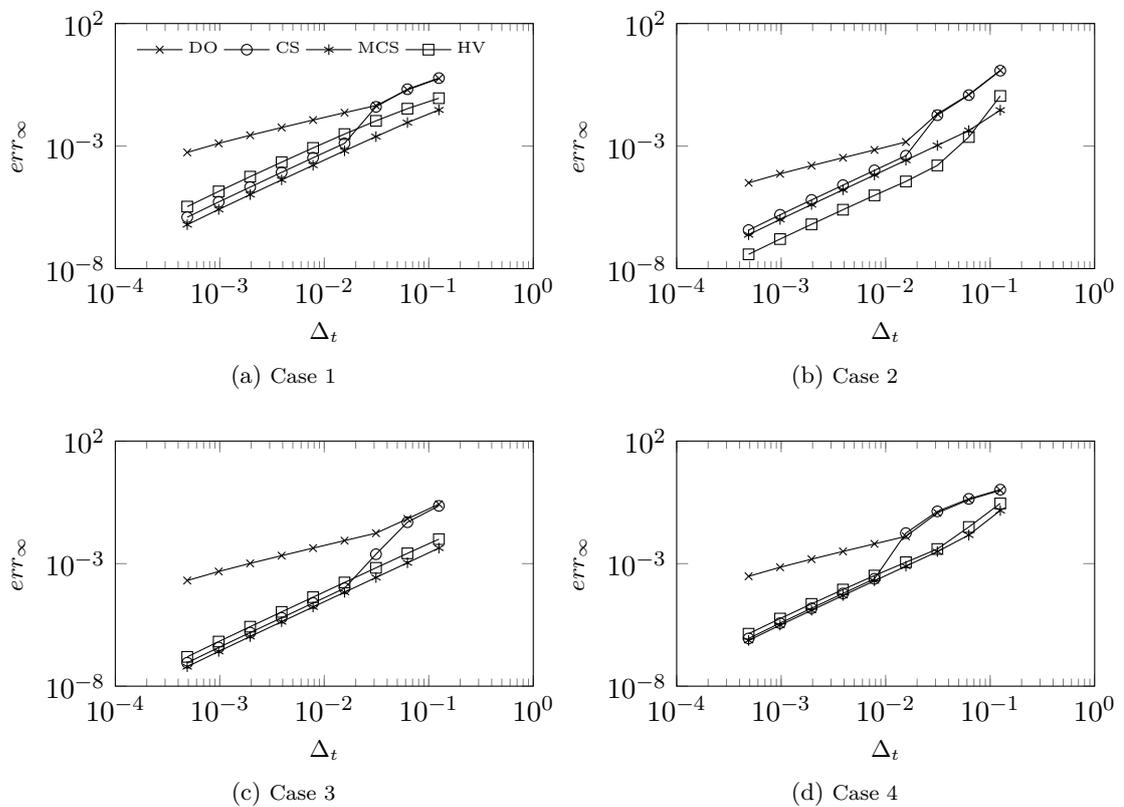


Figure 4.18: Heston-Hull-White: convergence in time. 129 grid points in  $S$  and 33 in  $v$  and  $r$  directions.

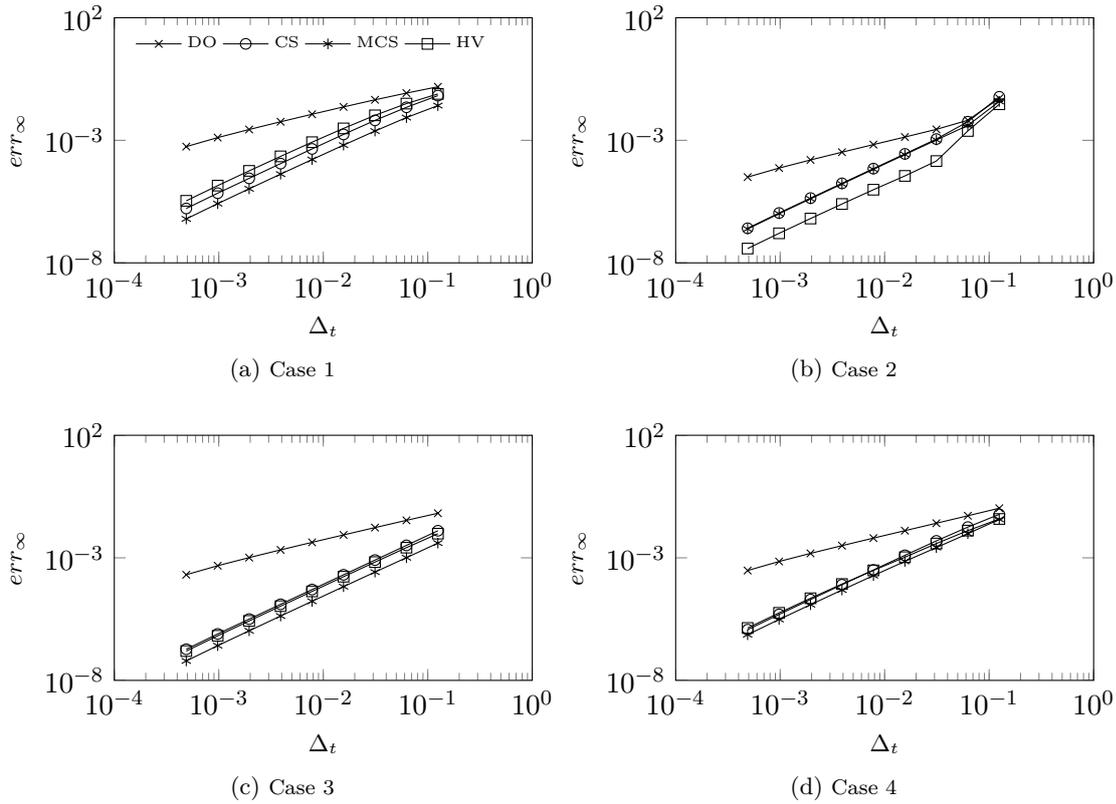


Figure 4.19: Heston-Hull-White: convergence in time with four initial steps using  $\theta = 1$ . 129 grid points in  $S$  and 33 in  $v$  and  $r$  directions.

the Heston-Hull-White model the pricing formula is available under the assumption that  $\rho_{13} = \rho_{23} = 0$ . Hence, we set these correlation values to zero in the following numerical experiments. It is well known that the complex logarithm in the pricing formula faces discontinuities, which is why we follow the approach by Kahl and Jäckel [57] and apply a rotation count correction algorithm to both pricing formulas. The experiments are performed with the CS ADI scheme with  $\theta = 0.5$ . The other schemes have the same spatial discretization and thus lead to the same results except for roundoff errors.

Figures 4.20 and 4.21 show the convergence in the direction of the underlying asset. In order to ensure that the error stemming from the first coordinate direction is dominant, we place 45 grid points in the directions of volatility, interest rate, respectively. This choice leads to negligible small errors in  $v$  and  $r$ . In the time discretization we use  $\Delta_t = 10^{-3}$ . We observe an error decay with the desired order close to two.

In Figures 4.22 and 4.23 we compare the convergence of the spectral approximation to the second-order finite difference approximation. The plots reveal that the spectral accuracy allows to use significantly fewer grid points than in the finite difference discretization to achieve the same accuracy, but at the cost of densely filled discretization matrices. Thus, it is of highest interest whether the spectral accuracy can offset this drawback. Let  $N_{SP} := N_{SP_2} = N_{SP_3}$  denote the number of grid points in each direction of the Chebyshev discretization and  $N_{FD} := N_{FD_2} = N_{FD_3}$  the number of nodes for the finite difference scheme in the directions  $v$  and/or  $r$ . Please note that we neglect the influence of discretization of the asset direction in the following discussion. As it can be seen in Figure 4.26 (a) the computational effort of the ADI scheme for the Chebyshev and FD dis-

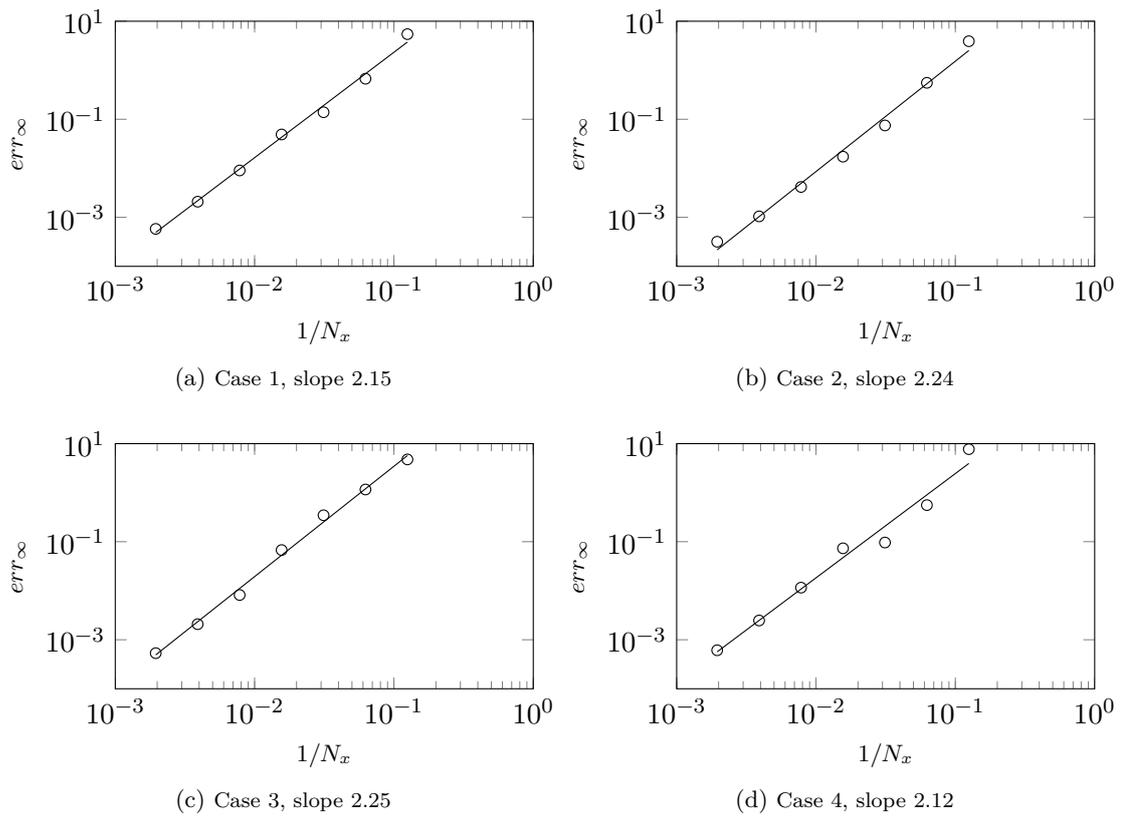


Figure 4.20: Heston: convergence in the direction of the underlying asset (hybrid CS ADI scheme),  $N_2 = 45$ .

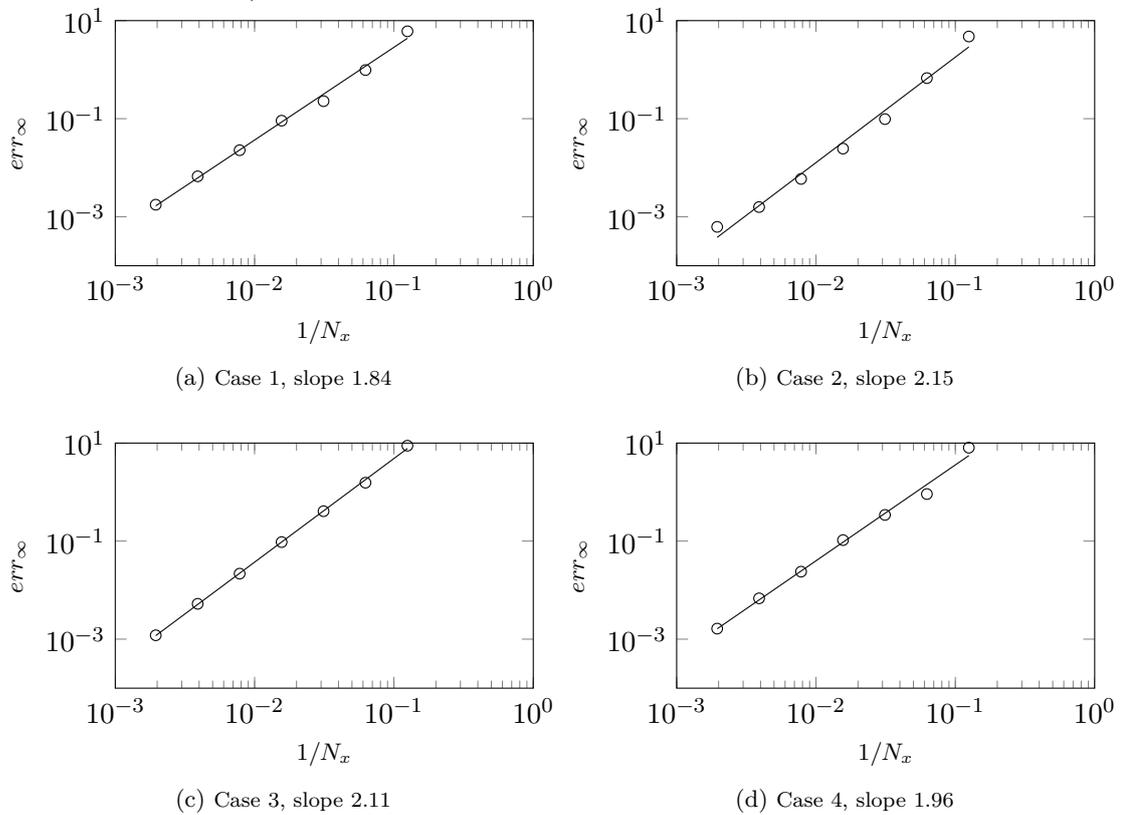


Figure 4.21: Heston-Hull-White: convergence in the direction of the underlying asset (hybrid CS ADI scheme),  $N_2 = N_3 = 45$ .

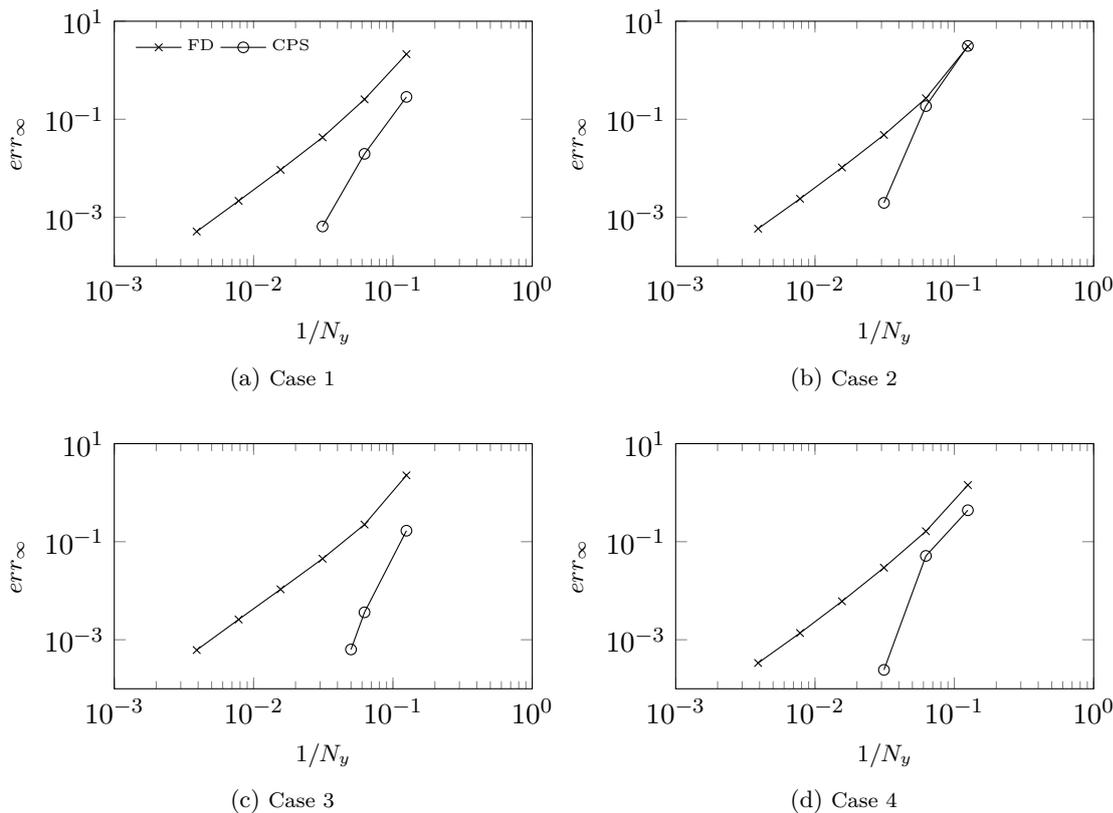


Figure 4.22: Heston: convergence in the direction of volatility with 1025 grid nodes in the direction of the asset and  $\Delta_t = 10^{-3}$ .

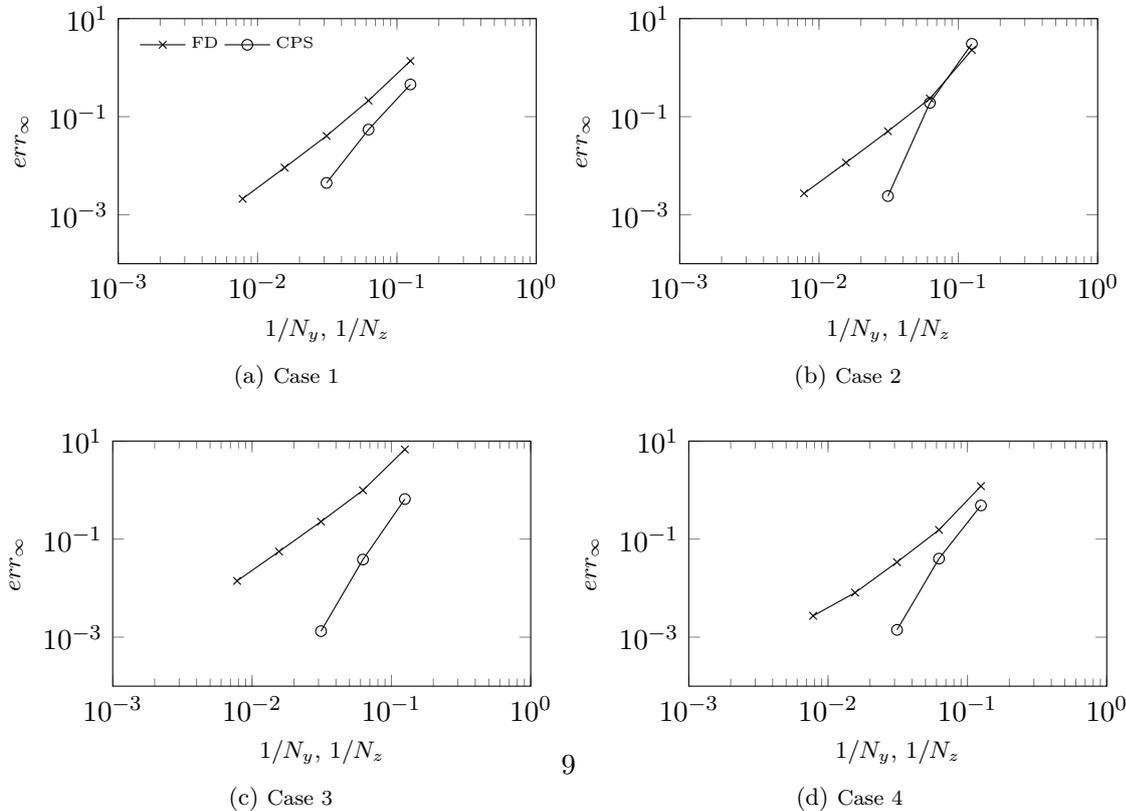


Figure 4.23: Heston-Hull-White: convergence in the direction of volatility/interest rate with 513 grid nodes in the direction of the asset and  $\Delta_t = 10^{-3}$ .

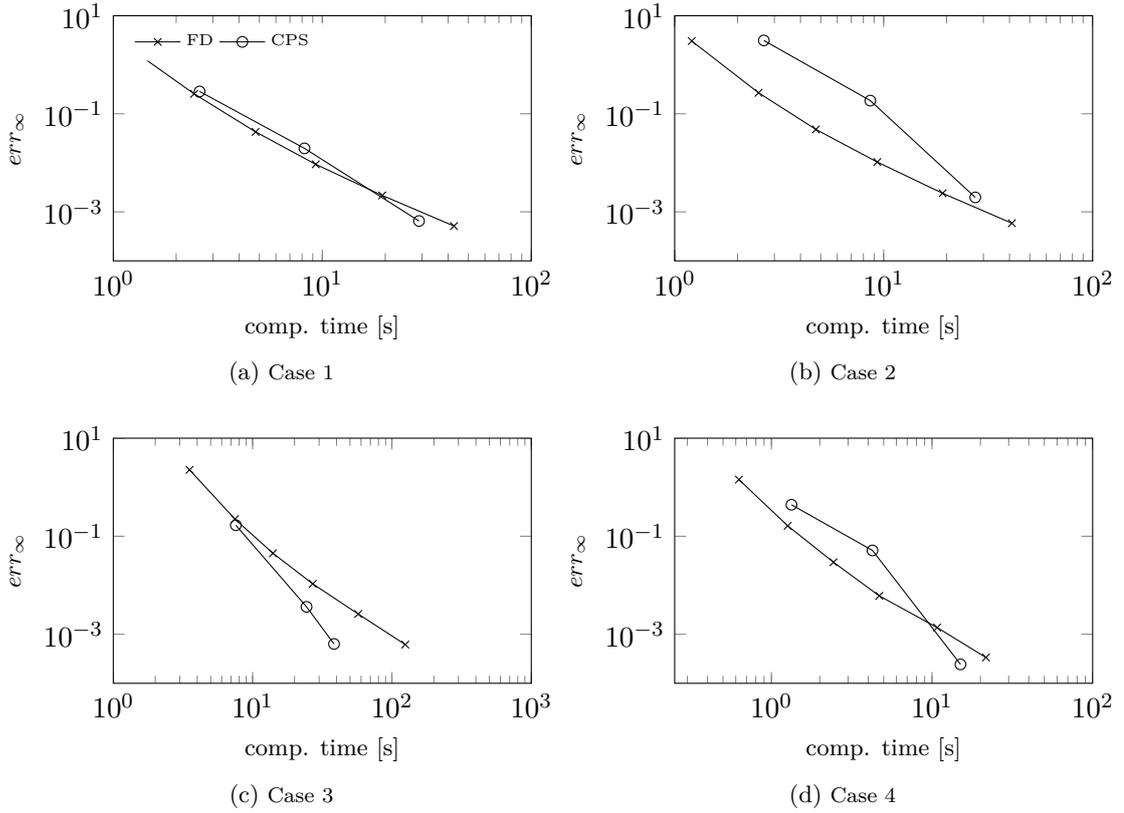


Figure 4.24: Heston: accuracy versus computation time in the direction of volatility with 1025 grid nodes in the direction of the asset and  $\Delta_t = 10^{-3}$ .

cretizations for the Heston model grow with  $\mathcal{O}(N_{SP}^2)$  and  $\mathcal{O}(N_{FD})$ . As mentioned before, the major workload consists of performing a forward and backward substitution to solve the linear system of equations in each leg of the ADI scheme after the LU decomposition has been computed during a startup phase, see Section 3.4. For a full quadratic matrix of size  $N$  this consists of  $N^2$  operations compared to  $2N$  operations for the forward and  $3N$  operations for the backward substitution in case of a tridiagonal matrix. Thus, the run-time for both methods is equivalent if  $N_{SP}^2 \approx 5N_{FD}$ . In the three-dimensional case of the HHW model  $N_{FD_1} \cdot N_{SP}$  one-dimensional problems have to be solved with an effort of  $\mathcal{O}(N_{SP}^2)$  and thus the run-time grows with  $\mathcal{O}(N_{SP}^3)$ . The analogous arguments lead to a growth of  $\mathcal{O}(N_{FD}^2)$  for the FD discretization. Figure 4.26 (b) underlines this result and we see a slope of 3.06 (excluding the first two data points), 2.03 respectively. Both methods have the same run-time if  $N_{SP}^3 \approx 5N_{FD}^2$ . In the general  $d$ -dimensional case, under the assumption that the solution is sufficiently smooth in  $d - 1$  coordinate directions, such that  $d - 1$  Chebyshev approximations can be applied, we expect a growth of  $\mathcal{O}(N_{SP}^d)$ , while we expect  $\mathcal{O}(N_{FD}^{d-1})$  for the FD method. Hence, the computational effort is approximately the same if  $N_{SP}^{d/(d-1)} 5^{-1/(d-1)} \approx N_{FD}$  holds. As the left-hand side is monotonically decreasing for growing  $d$  the hybrid approach with ADI time splitting becomes more efficient compared to the FD discretization for higher-dimensional problems. The dashed line in Figure 4.26 underlines that the theoretical result  $N_{SP}^{d/(d-1)} 5^{-1/(d-1)} \approx N_{FD}$  holds in practice. Please note that the explicit treatment of the mixed derivative term  $\frac{\partial^2 u}{\partial x_2 \partial x_3}$  in (4.10) via matrix-vector multiplication has a computational effort proportional to  $\mathcal{O}(N_{SP}^4)$  since both Chebyshev differentiation matrices are full. Hence, we perform the computation of the mixed derivative by sequential differentiation in  $x_2$ -,  $x_3$ -direction

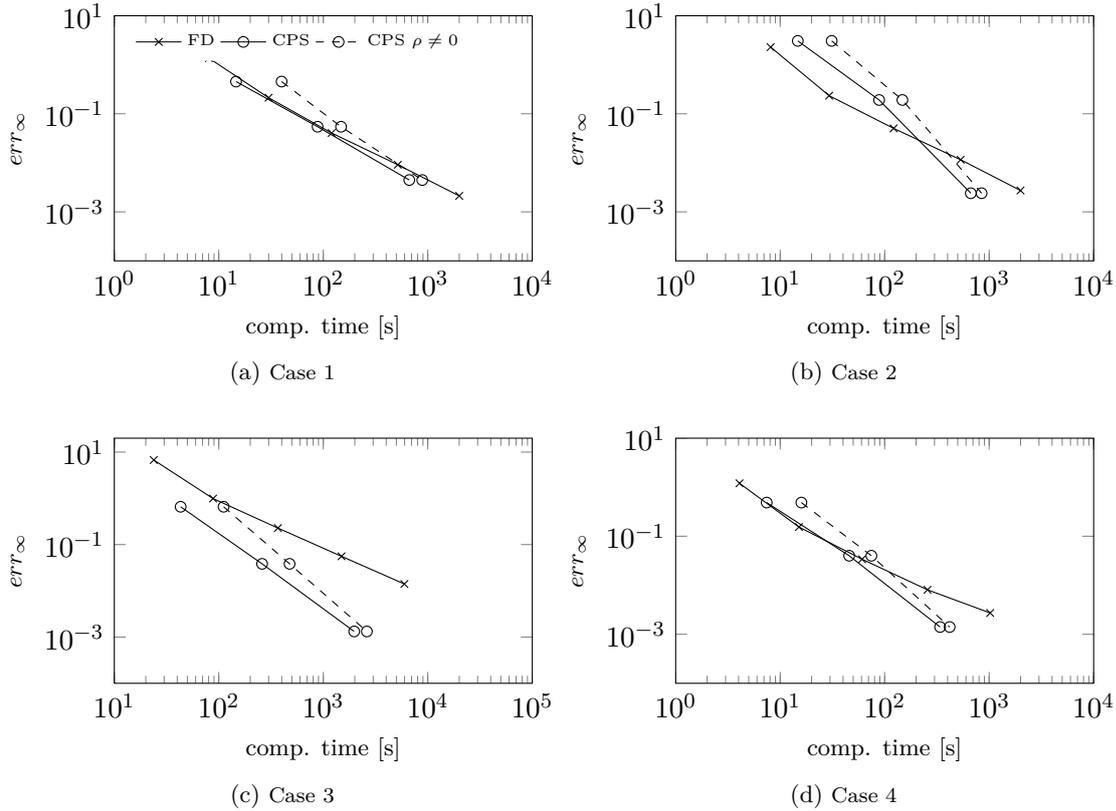


Figure 4.25: Heston-Hull-White: accuracy versus computation time in the direction of volatility/interest rate with 513 grid nodes in the direction of the asset and  $\Delta_t = 10^{-3}$ .

in transformed space to reduce the computational effort to  $\mathcal{O}(N_{SP}^2 \log_2^2(N_{SP}))$ . In contrast, with central second-order finite differences the derivative can be computed in linear run-time due to the sparse structure ( $\leq 4$  entries per row) of the discretization matrix.

Figures 4.24 and 4.25 show the accuracy versus the computation time. The hybrid method is able to outperform the FD method in the majority of the test scenarios - only in case 2 of the Heston model, the FD method yields more accurate result. In the three-dimensional case of the HHW model the spectral discretization achieves a higher accuracy than the FD scheme even for small run-times in test cases 1, 3, 4 and even in case 2 the method is able to beat its benchmark in the high accuracy region. These results can be explained by the argumentation given above. The geometric error decay of the Chebyshev approximation in combination with ADI time stepping shows its strength if highly accurate results for higher-dimensional problems are desired. Here, the fast convergence compensates the disadvantage of full discretization matrices and the second-order FD discretization is clearly outperformed. The dashed lines in Figure 4.25 show the run-time if  $\rho_{13} \neq 0$ ,  $\rho_{23} \neq 0$ . Since in this case no reference solution is available, we are content with the error of the uncorrelated case ( $\rho_{13} = \rho_{23} = 0$ ). For the sake of readability we omit to plot the run-time for non-zero correlation in case of the pure finite difference method and assume the favorable case that the run-time does not increase.

In an additional experiment we investigate the run-time properties for time-dependent parameters. Therefore, we consider a time-dependent long term mean in the direction of

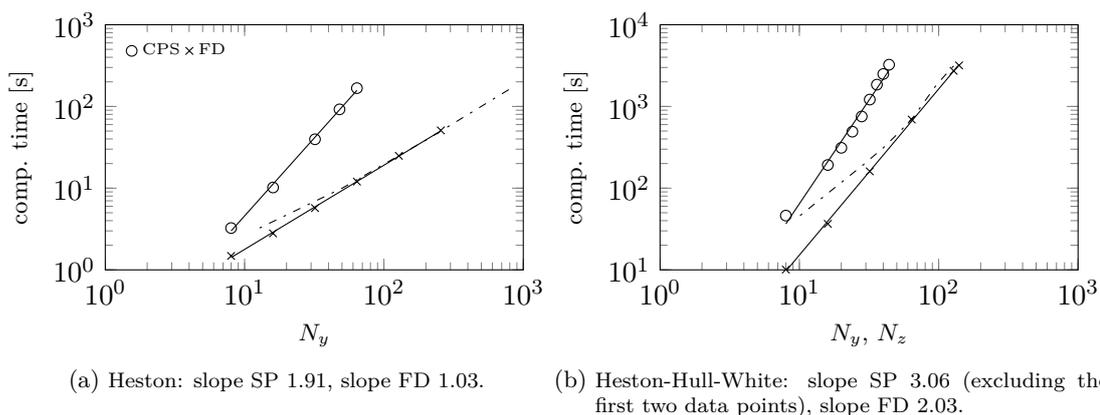


Figure 4.26: Heston and HHW run-time scaling for growing number of grid nodes in  $v$  and  $r$  directions with fixed number of grid nodes in the asset direction (Heston:  $N_1 = 1025$ , correlation  $\neq 0$ , HHW:  $N_1 = 513$ , all correlations  $\neq 0$ ,  $\frac{\partial^2 u}{\partial y \partial z}$  via FFT differentiation) and 1000 time steps. The dotted line shows the shifted Chebyshev run-time curve:  $N_{SP}^{d/(d-1)} 5^{-1/(d-1)}$  for  $d = 2, d = 3$  respectively.

the interest rate within the HHW model. Similar to [38] we use

$$b_r(t) = c_0 - c_1 e^{-c_2 \cdot t},$$

with constants  $c_0, c_1$  and  $c_2$ . The third implicit leg in the ADI schemes changes to

$$Y_3 = Y_2 + \theta \Delta_t (F_3((n+1)\Delta_t) Y_3 - F_3(n\Delta_t)) U_n$$

and the computation of  $\tilde{Y}_3$  has to be modified in an analogue way. In the explicit steps  $F, F_0$  depend on point in time  $n\Delta_t$ . Due to the time-dependency the implicit system to compute  $Y_3, \tilde{Y}_3$  has to be solved in each time iteration and cannot be solved via an LU decomposition in a startup phase. In order to reduce the computational effort we employ an iterative solver, cf. Section 3.4. For our numerical example we choose the Matlab<sup>®</sup> builtin solver *bicgstab* with an effort of  $\mathcal{O}(N^2)$  per iteration for a quadratic matrix of size  $N$ . Thus, if the number of iterations is small, one can apply similar arguments like in the previous analysis to show that both methods have approximately the same run-time if  $\mathcal{O}(N_{SP}^{d/(d-1)}) = \mathcal{O}(N_{FD})$  is fulfilled. In Figure 4.27 we compare the run-time of the hybrid to the finite difference method with the parameters of scenario 4 and  $c_0 = 0.101$ ,  $c_1 = 0.003$ ,  $c_3 = 1$ . As the starting value for the iterative solver we use the solution of the previous leg  $Y_2, \tilde{Y}_2$ , respectively. For this choice the solver converges in less than three iterations in our experiment. Similar to the time-independent case the Chebyshev method is able to outperform the benchmark method. The sparse linear system of equations within the finite difference method can be solved very efficiently and hence it is not necessary to use an iterative solver. In our numerical experiment this even led to a longer computation time.

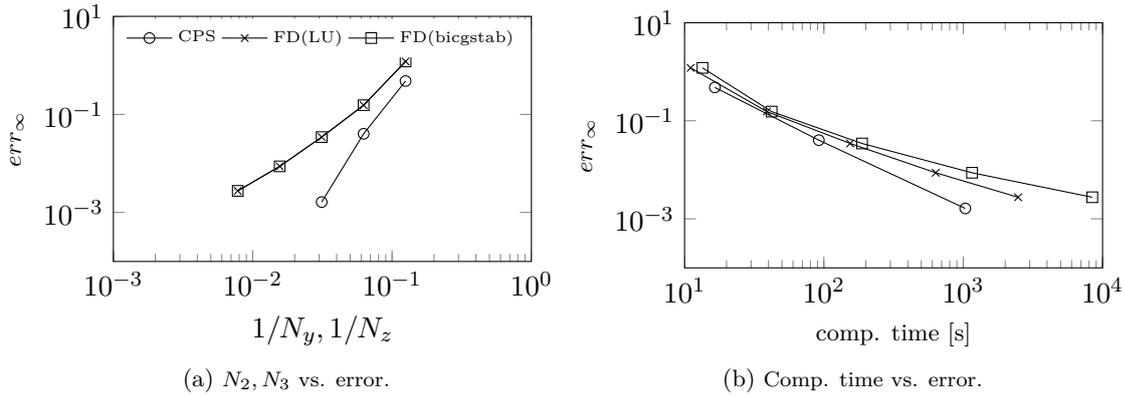


Figure 4.27: HHW: accuracy in the direction of volatility/interest rate with 513 grid nodes in the direction of the asset,  $\rho_{13} = \rho_{23} = 0$ ,  $\Delta_t = 10^{-3}$  and time dependent parameter  $b_r$ .

#### 4.4.2 High-Order Finite Difference Discretization in the Asset Direction

The second-order accuracy in the asset direction of the hybrid scheme is a major bottleneck since a large number of grid nodes needs to be used to achieve an accurate solution. Thus, it seems reasonable to apply higher-order discretizations. In the following we propose a high-order-compact finite difference approximation. According to Section 2.2 we can achieve fourth-order accuracy for the unidirectional problem (asset direction) on the compact stencil with discretization (2.12). Based on equation (4.9) and (4.10), respectively, we have

$$\begin{aligned}
 a_{11}^{\text{Heston}}(x_1, x_2) &= \frac{1}{2} f_s(x_1)^2 f_v(x_2) \psi'_s(f_s(x_1))^2, \\
 c_1^{\text{Heston}}(x_1, x_2) &= \frac{1}{2} f_s(x_1)^2 f_v(x_2) \psi''_s(f_s(x_1)) + r f_s(x) \psi'_s(f_s(x_1)), \\
 a_{11}^{\text{HHW}}(x_1, x_2, x_3) &= \frac{1}{2} f_s(x_1)^2 f_v(x_2) \psi'_s(f_s(x_1))^2, \\
 c_1^{\text{HHW}}(x_1, x_2, x_3) &= \frac{1}{2} f_s(x_1)^2 f_v(x_2) \psi''_s(f_s(x_1)) + f_r(x_3) f_s(x_1) \psi'_s(f_s(x_1)),
 \end{aligned}$$

with  $(x_1, x_2) \in \tilde{\Omega} = [0, 1] \times [-1, 1]$ ,  $(x_1, x_2, x_3) \in \tilde{\Omega} = [0, 1] \times [-1, 1]^2$ . Inserting these into equation (2.12) one obtains the fourth-order discretization, which is defined on the compact stencil. Analogous to Section 3.2 we treat the mixed derivatives with a broad standard fourth-order stencil. Since the reciprocal of  $a_{11}^{\text{Heston/HHW}}$  needs to be computed, one has to truncate the domain at  $v > 0$  to avoid a division by zero. We therefore use the computational domain  $[0, 20K] \times [0.001, 1.5] \times [-0.5, 0.5]$  and the region of interest  $[\frac{1}{2}K, 2K] \times [0.01, 1] \times [0, 0.125]$  in the following numerical experiments. The initial condition is smoothed with  $p = 4$ . Figures 4.28 and 4.30 show the accuracy in the asset direction. In Table 4.7 we report the estimated order of convergence. The high-order method clearly outperforms the second-order scheme. Since the computational effort remains within the same complexity class, due to the tridiagonal structure of the discretization matrices in the implicit legs, like the second-order method, the run-time is approximately the same. Thus, in Figures 4.29 and 4.31 the accuracy of the proposed fourth-order scheme is significantly higher for a fixed computation time.

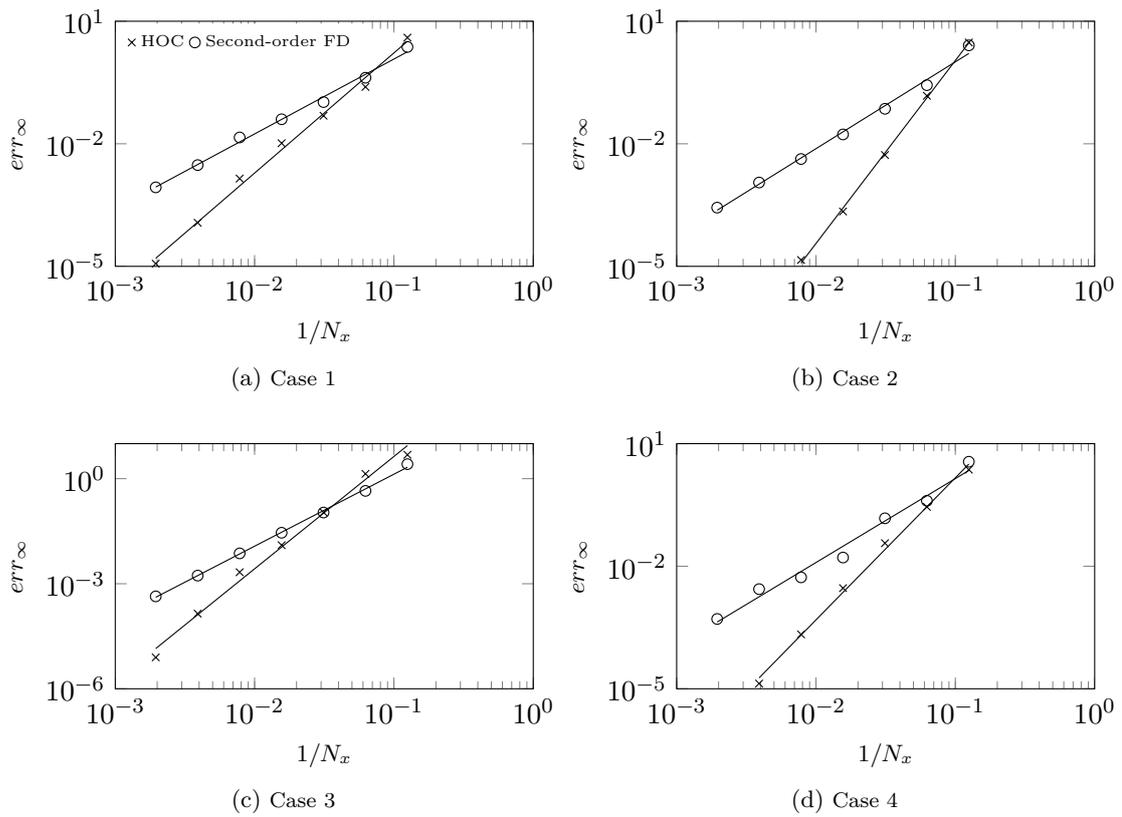


Figure 4.28: Heston: convergence in the direction of the underlying asset,  $N_2 = 45$ .

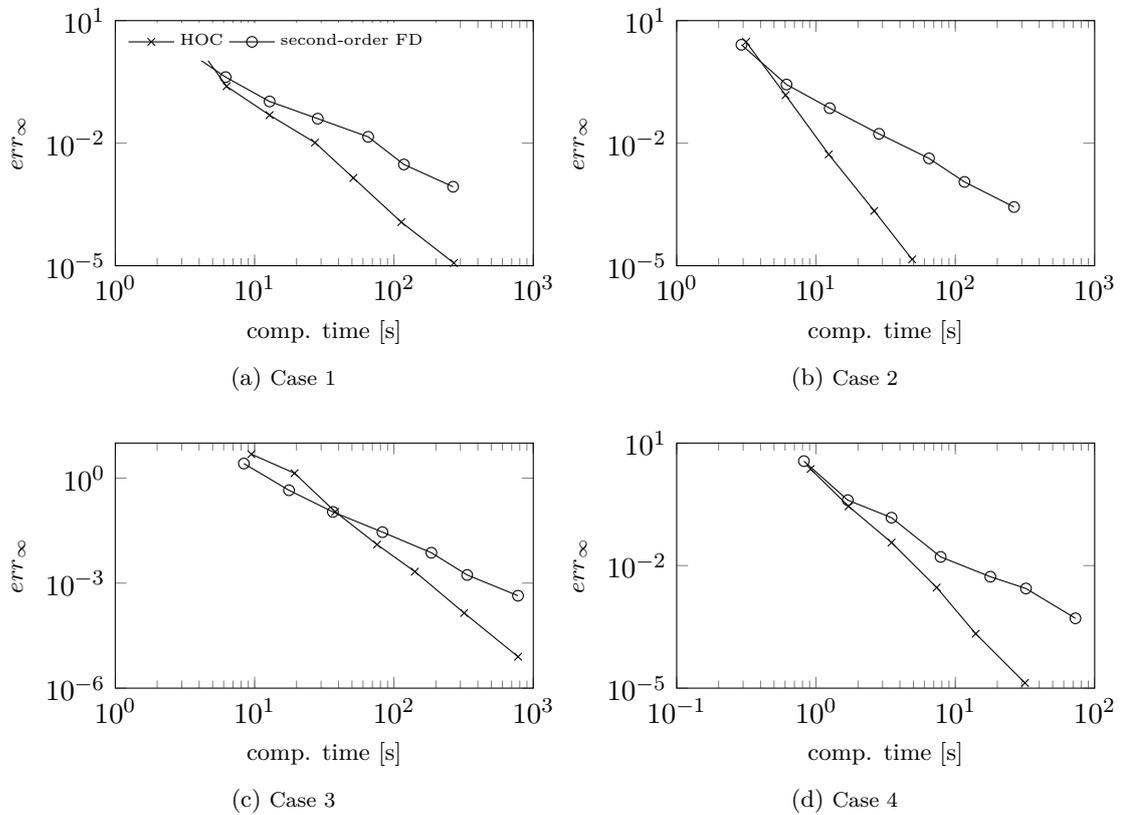


Figure 4.29: Heston: accuracy versus computation time in the direction of the asset with 45 grid nodes in the direction of the volatility and  $\Delta_t = 10^{-3}$ .

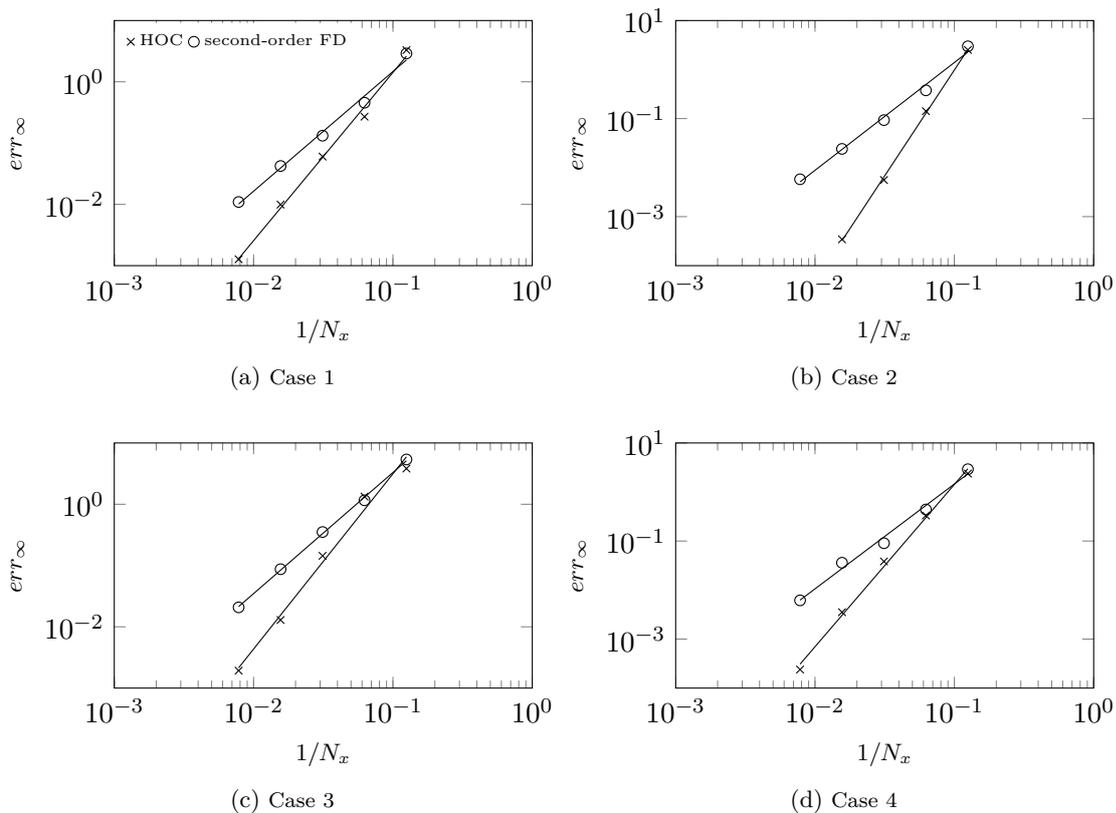


Figure 4.30: HHW: convergence in the direction of the underlying asset,  $N_2 = N_3 = 45$ .

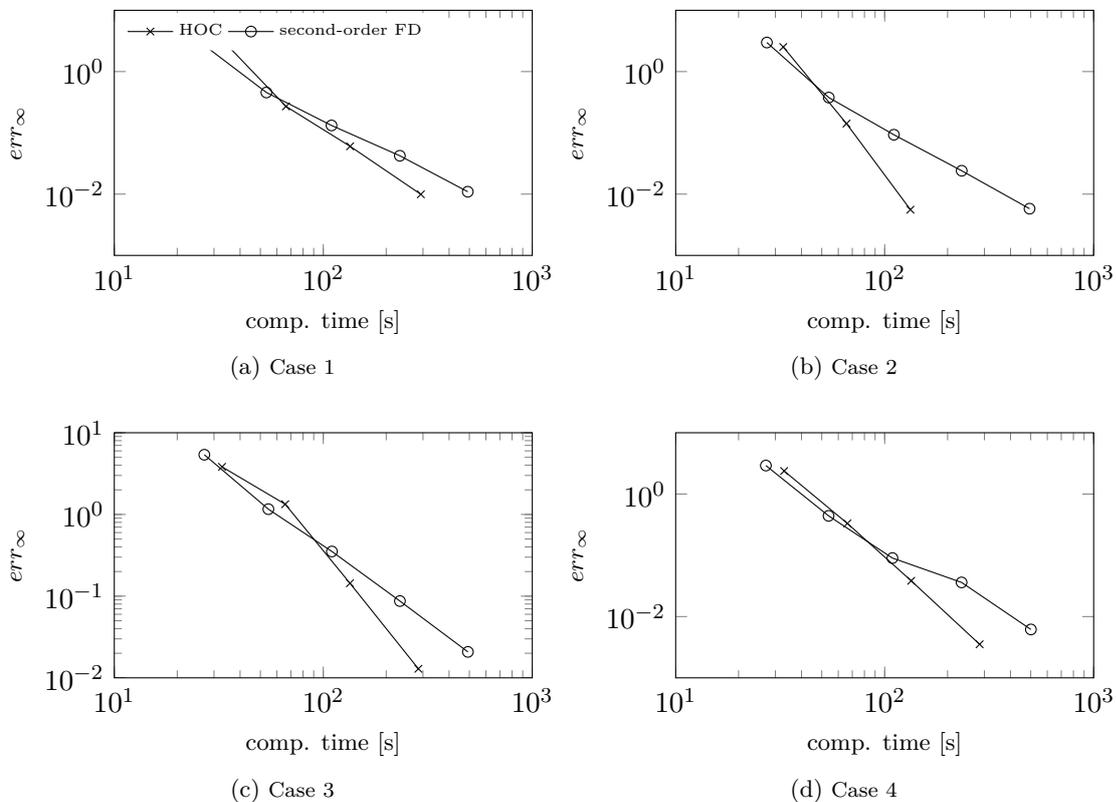


Figure 4.31: HHW: accuracy versus computation time in the direction of the asset with 45 grid nodes in the direction of the volatility/interest rate and  $\Delta_t = 10^{-3}$ .

Scenario		1	2	3	4
$\text{err}_\infty$	second-order FD	1.83	2.13	2.06	2.05
	HOC FD	2.94	4.48	3.21	3.48

(a) Heston

Scenario		1	2	3	4
$\text{err}_\infty$	second-order FD	1.95	2.20	1.98	2.14
	HOC FD	2.74	4.32	2.86	3.31

(b) Heston-Hull-White

Table 4.7: Estimated order of convergence in the direction of the underlying asset (cf. Figures 4.28 and 4.30).



# 5

## Conclusions and Outlook

### 5.1 Conclusions

In this thesis we examined high-order methods in a financial engineering setting. The PDEs arising in finance are in general of convection-diffusion-reaction type with mixed derivatives present. In the time domain we decomposed the spatial discretization matrix with ADI splitting methods, which are able to treat the mixed derivative efficiently via explicit time stepping. We considered four prominent schemes: the Douglas, the Craig-Sneyd, the modified Craig-Sneyd and the Hundsdorfer-Verwer scheme. In the case of high-order-compact finite difference discretizations in space we proved bounds on  $\theta$  for unconditional stability based on the existing results from literature in the von Neumann framework. It turned out that the stability bounds of central second-order finite difference ADI and HO-ADI schemes coincide. Since in general a lower  $\theta$  value leads to a smaller numerical error, HO-ADI schemes do not suffer any loss in accuracy in the time domain due to stability restrictions. In addition to finite differences we considered pseudo-spectral approximations. In numerical experiments we investigated the stability properties. Therefore, we computed the eigenvalues of the iteration matrices for Dirichlet and Neumann boundary conditions. The positive outcome of the experiments encouraged us to apply ADI time stepping with pseudo-spectral differentiation in space to basket options and to stochastic volatility models.

In order to reduce the number of grid nodes of the spatial discretization we employed sparse grids. We used the combination technique to construct the sparse grid solution and analyzed its regularity requirements. For fourth-order finite difference schemes  $u \in C^{(10,10,\dots,10)}$  was actually needed to obtain an error decay of order  $\mathcal{O}(h^4 \log_2(h^{-1})^d)$  on the entire spatial domain. For pseudo-spectral methods two cases have to be distinguished: the bounded variation and the analytic case. If the solution has a derivative of bounded variation the numerical scheme exhibits an algebraic error decay and thus the error splitting structure is well suited for the combination technique. If the true solution is even analytic the numerical solution will converge with geometric accuracy. Here the hierarchical surpluses at the same level are of different size and therefore the splitting structure is not appropriate for the combination technique. However, for problems with a convergence-limiting singularity close to  $[-1, 1]$  the surpluses are almost of the same size and the sparse grid outperforms the full grid approach.

In the last part of this thesis we applied the proposed numerical techniques to financial engineering PDEs. Within the Black-Scholes model we used HOC finite differences and derived a hybrid scheme using pseudo-spectral as well as finite differences. Since the initial condition in option pricing problems is generally non-smooth, we investigated the impact of smoothed initial data on the rate of convergence obtained in financial applications for both approaches. For fourth-order finite differences a smoothing according to Kreiss et al. of order four recovered the fourth-order accuracy, while higher-order smoothing

significantly increased the performance of the hybrid pseudo-spectral / finite difference method. Furthermore, we related the option parameters in the multivariate Black-Scholes model to the performance of HOC schemes in the sparse grid combination technique. The experiments revealed that one can expect a smoother solution for large volatility and positive correlation between the assets and thus also a faster error decay. In comparison the full grid solver did not show such a sensitivity on the parameters, which can be explained by the less restrictive regularity requirements.

For the Heston and Heston-Hull-White PDE we proposed a hybrid method using finite differences in the direction of the underlying and pseudo-spectral methods in the direction of the other risk-factors. The method was able to outperform the second-order finite difference benchmark method. Both the numerical results and the theoretical complexity discussion showed that especially in the higher-dimensional case the method is very efficient due to the dimensional splitting. In the asset direction we additionally applied a fourth-order finite difference discretization. As the arising matrix is of tridiagonal structure in the implicit legs of the asset direction the method is capable of computing highly accurate solutions with only a very low additional computational effort. Thus, the benchmark method is clearly outperformed.

## 5.2 Outlook

In future research it would be interesting to extend the presented numerical methods to American option pricing problems. The early-exercise feature adds an additional level of complexity since the treatment of the early-exercise boundary is highly non-trivial if one wants to maintain fourth-order or even higher accuracy. Leentvaar [65] has already discussed this topic for standard fourth-order finite differences. An adaption of his approach, to enforce a smooth pasting condition, to HOC discretizations will be of great interest. But other approaches could be considered as well, such as penalty methods [100] or solving the partial differential complementarity problem [37].

At the current state the temporal discretization is of second-order. To make the schemes more efficient higher-order methods could be used. If one wants to stay within the ADI framework fourth-order can be obtained via Richardson extrapolation. Alternatively, other time discretization techniques can be applied, e.g. backward difference formula (BDF) methods.

## References

- [1] L. B. G. Andersen and V. Piterbarg. Moment Explosions in Stochastic Volatility Models. *Fin. Stoch.*, 11(1):29–50, 2007.
- [2] Z. Battles and L. N. Trefethen. An Extension of Matlab to Continuous Functions and Operators. *SIAM J. Sci. Comp.*, 25:1743–1770, 2004.
- [3] G. Berikelashvili, M. Gupta, and M. Mirianashvili. Convergence of Fourth Order Compact Finite Difference Schemes for Three-dimensional Convection-Diffusion equations. *SIAM J. Numer. Anal.*, 45:443–455, 2007.
- [4] V. Bhansali. *Pricing and Managing Exotic and Hybrid Options for Free*. McGraw-Hill Companies, 1998.
- [5] P. Brian. A Finite-difference Method of High-Order Accuracy for the Solution of Threedimensional Transient Heat Conduction Problems. *AIChE Journal*, 7:367–370, 1961.
- [6] M. Briani, L. Caramellino, and A. Zanette. A Hybrid Approach for the Implementation of the Heston Model. *IMA J. of Manag. Math.*, published online, 2015.
- [7] M. Briani, L. Caramellino, and A. Zanette. A Hybrid Tree/Finite-Difference Approach for Heston-Hull-White Type Models. *J. Comp. Fin.*, to appear, 2016.
- [8] H. Bungartz and M. Griebel. Sparse grids. *Cambridge University Press*, pages 1–123, 2004.
- [9] H.-J. Bungartz, M. Griebel, D. Rösche, and C. Zenger. Pointwise Convergence of the Combination Technique for Laplace’s Equation. *East-West J. Numer. Math.*, 2:21–45, 1994.
- [10] O. Burkovska, B. Haasdonk, J. Salomon, and B. Wohlmuth. Reduced Basis Methods for Pricing Options with the Black–Scholes and Heston Models. *SIAM J. Fin. Math.*, 6(1):685–712, 2015.
- [11] C. Canuto, M. Hussaini, A. Quarteroni, and T. Zang. *Spectral Methods: Fundamentals in Single Domains*. Scientific Computation. Springer, Berlin, 2006.
- [12] P. Carr and D. B. Madan. Option Valuation using the Fast Fourier Transform. *J. Comp. Fin.*, 2:61–73, 1999.
- [13] N. Clarke and K. Parrott. Multigrid for American Option Pricing with Stochastic Volatility. *Appl. Math. Fin.*, 6(3):177 – 195, 1999.
- [14] I. Craig and A. Sneyd. An Alternating-Direction Implicit Scheme for Parabolic Equations with Mixed Derivatives. *J. Comp. Math. Appl.*, 16(4):341–350, 1988.

- 
- [15] E. Derman and I. Kani. Riding on a Smile. *Risk*, 2:32–39, 1994.
- [16] J. Douglas and H. Rachford. On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.
- [17] J. J. Douglas. Alternating Direction Methods for Three Space Variables. *Numer. Math.*, 4(1):41–63, 1962.
- [18] D. Duffie, J. Pan, and K. Singleton. Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68, 2000.
- [19] B. Dupire. Pricing and Hedging with Smiles. In *Proceedings of AFFI Conference*, 1993.
- [20] B. Dupire. Pricing with a Smile. *Risk Magazine*, pages 18–20, 1994.
- [21] B. Düring and M. Fournié. High-Order Compact Finite Difference Scheme for Option Pricing in Stochastic Volatility Models. *J. Comput. Appl. Math.*, 236(17):4462–4473, 2012.
- [22] B. Düring, M. Fournié, and C. Heuer. High-Order Compact Finite Difference Schemes for Option Pricing in Stochastic Volatility Models on Non-uniform Grids. *J. Comput. Appl. Math.*, 271:247–266, 2014.
- [23] B. Düring, M. Fournié, and A. Rigal. High-Order ADI Schemes for Convection-Diffusion Equations with Mixed Derivative Terms. In *Spectral and High Order Methods for Partial Differential Equations - ICOSAHOM 2012*, volume 95, pages 217–226. Springer International Publishing, 2014.
- [24] B. Düring and C. Heuer. High-Order Compact Schemes for Parabolic Problems with Mixed Derivatives in Multiple Space Dimensions. *SIAM J. Numer. Anal.*, 53(5):2113–2134, 2015.
- [25] B. Düring and J. Miles. High-Order ADI Scheme for Option Pricing in Stochastic Volatility Models. *Preprint*, 2015.
- [26] F. Fang and C. W. Oosterlee. A Novel Pricing Method for European Options based on Fourier-Cosine Series Expansions. *SIAM J. Sci. Comput.*, 31(2):826–848, 2008.
- [27] I. Florescu and V. Frederi. A Binomial Tree Approach to Stochastic Volatility Driven Model of the Stock Price. *An. of Univ. Craiova, Math. Comp. Sci.*, 32:126–142, 2005.
- [28] I. Florescu and V. Frederi. Stochastic Volatility: Option Pricing Using a Multinomial Recombining Tree. *Appl. Math. Fin.*, 15(2):151–181, 2005.
- [29] M. B. Giles and R. Carter. Convergence Analysis of Crank-Nicolson and Rannacher Time-marching. *J. Comp. Fin.*, 9(4):89–112, 2006.
- [30] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Applications of Mathematics. Springer, New York, 2004.
- [31] D. Gottlieb and L. Lustman. The Spectrum of the Chebyshev Collocation Operator

- for the Heat Equation. *SIAM J. Numer. Anal.*, 20(5):909–921, 1983.
- [32] M. Griebel and W. Huber. Turbulence Simulation on Sparse Grids using the Combination Method. In S. N., J. Periaux, and A. Ecer, editors, *New Algorithms and Applications*, pages 75–84, 1995.
- [33] M. Griebel, M. Schneider, and C. Zenger. A Combination Technique for the Solution of Sparse Grid Problems. *IMACS Elsevier, Iter. Meth. Lin. Alge.*, 16:263–281, 1992.
- [34] W. Grill and H. Perczynski. *Wirtschaftslehre des Kreditwesens*. Deutscher Sparkassen Verlag, 2003.
- [35] L. A. Grzelak and C. W. Oosterlee. On the Heston Model with Stochastic Interest Rates. *SIAM J. Fin. Math.*, 2(1):255–286, 2011.
- [36] M. Gupta, R. Manohar, and J. Stephenson. A Single Cell High Order Scheme for the Convection-Diffusion Equation with Variable Coefficients. *Int. J. Numer. Methods Fluids*, 4:641–651, 1984.
- [37] T. Haentjens and K. J. in t Hout. ADI Schemes for Pricing American Options under the Heston Model. *Appl. Math. Fin.*, 22(3):207–237, 2015.
- [38] T. Haentjens and K. J. in’t Hout. ADI Finite Difference Schemes for the Heston-Hull-White PDE. *J. Comp. Fin.*, 16:83–110, 2012.
- [39] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing Smile Risk. *Wilmott Magazine*, 1(1):84–108, September 2002.
- [40] C. A. Hall. On Error Bounds for Spline Interpolation. *J. Approx. Theo.*, 1:209–218, 1968.
- [41] C. Hendricks, M. Ehrhardt, and M. Günther. High Order Combination Technique for the Efficient Pricing of Basket Options. *Acta Math. Univ. Comen.*, 84(2):243–253, 2015.
- [42] C. Hendricks, M. Ehrhardt, and M. Günther. Error Splitting Preservation for High Order Finite Difference Schemes in the Combination Technique. *Numer. Math: Theory, Models and Appl.*, to appear, 2016.
- [43] C. Hendricks, M. Ehrhardt, and M. Günther. High-Order ADI Schemes for Diffusion Equations with Mixed Derivatives in the Combination Technique. *Appl. Numer. Math.*, 101:36–52, 2016.
- [44] C. Hendricks, M. Ehrhardt, and M. Günther. Hybrid Finite Difference / Pseudospectral Methods for the Heston and Heston-Hull-White PDE. *J. Comp. Fin.*, to appear, 2016.
- [45] C. Hendricks, C. Heuer, M. Ehrhardt, and M. Günther. High-Order ADI Finite Difference Schemes for Parabolic Equations in the Combination Technique with Application in Finance. *J. Comput. Appl. Math.*, to appear, 2016.
- [46] S. L. Heston. A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Rev. Finan. Stud.*, 6(2):327–343, 1993.

- 
- [47] C. Heuer. *High-Order Compact Finite Difference Schemes for Parabolic Partial Differential Equations with Mixed Derivative Terms and Applications in Computational Finance*. PhD thesis, University of Sussex, 2014.
- [48] W. Hundsdorfer. Accuracy and Stability of Splitting with Stabilizing Corrections. *Appl. Numer. Math.*, 42(1-3):213–233, 2002.
- [49] W. Hundsdorfer and J. Verwer. *Numerical Solution of Time-dependent Advection-Diffusion-Reaction Equations*, volume 33. Springer, 2003.
- [50] K. in't Hout, J. Bierkens, A. P. C. van der Ploeg, and J. in't Panhuis. A Semi Closed-form Analytic Pricing Formula for Call Options in a Hybrid Heston–Hull–White Model. *Proceedings of the 58th European Study Group Mathematics with Industry*, eds. R.H. Bisseling et. al., Utrecht, pages 101–105, 2007.
- [51] K. J. in't Hout and S. Foulon. ADI Finite Difference Schemes for Option Pricing in the Heston Model with Correlation. *Int. J. Numer. Anal. Mod.*, 7:303–320, 2010.
- [52] K. J. in't Hout and C. Mishra. Stability of the Modified Craig-Sneyd Scheme for Two-dimensional Convection-Diffusion Equations with Mixed Derivative Terms. *Math. Comp. Simul.*, 81:2540–2548, 2011.
- [53] K. J. in't Hout and C. Mishra. Stability of ADI Schemes for Multidimensional Diffusion Equations with Mixed Derivative Terms. *Appl. Numer. Math.*, 74:83–94, 2013.
- [54] K. J. in't Hout and B. Welfert. Unconditional Stability of Second-order ADI Schemes applied to Multi-Dimensional Diffusion Equations with Mixed Derivative Terms. *Appl. Numer. Math.*, 59(3-4):677–692, 2009.
- [55] K. J. in't Hout and B. D. Welfert. Stability of ADI Schemes applied to Convection-Diffusion Equations with Mixed Derivative Terms. *Appl. Numer. Math.*, 57(1):19–35, Jan. 2007.
- [56] K. J. in't Hout and M. Wyns. Convergence of the Modified Craig-Sneyd Scheme for Two-dimensional Convection-Diffusion Equations with Mixed Derivative Term. *J. Comput. Appl. Math.*, 296(C):170–180, 2016.
- [57] C. Kahl and P. Jäckel. Not-so-complex Logarithms in the Heston Model. *Wilmott Magazine*, pages 94–103, September 2005.
- [58] S. Karaa and J. Zhang. High-Order ADI Method for Solving Unsteady Convection-Diffusion Problems. *J. Comput. Phys.*, 198(1):1–9, 2004.
- [59] M. Keller-Ressel. Moment Explosion and Long-term Behavior of Affine Stochastic Volatility Models. *Math. Fin.*, 21:73–98, 2011.
- [60] T. Kluge. Pricing Derivatives in Stochastic Volatility Models Using the Finite Difference Method. Master's thesis, Technische Universität Chemnitz, 2002.
- [61] D. A. Kopriva. *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers*. Springer Publishing, 1st edition, 2009.

- 
- [62] H. O. Kreiss, V. Thomée, and O. Widlund. Smoothing of Initial Data and Rates of Convergence for Parabolic Difference Equations. *Commun. Pure Appl. Math.*, 23(2):241–259, 1970.
- [63] D. Lanser, J. Blom, and J. Verwer. Time Integration of the Shallow Water Equations in Spherical Geometry. *J. Comp. Phys.*, 171:373–393, 2001.
- [64] A. Laub. *Matrix Analysis for Scientists and Engineers*. SIAM, 2004.
- [65] C. Leentvaar and C. Oosterlee. American Options With Discrete Dividends Solved by Highly Accurate Discretizations. In *Progress in Industrial Mathematics at ECMI 2014*, pages 427–431. Springer, 2004.
- [66] C. Leentvaar and C. Oosterlee. On Coordinate Transformation and Grid Stretching for Sparse Grid Pricing of Basket Options. *J. Comput. Appl. Math.*, 222(1):193–209, 2008.
- [67] C. C. W. Leentvaar and C. W. Oosterlee. Pricing Multi-Asset Options with Sparse Grids and Fourth Order Finite Differences. In *Numer. Math. Advan. Appl.*, pages 975–983. Springer, 2006.
- [68] A. Leitao and C. Oosterlee. GPU Acceleration of Stochastic Grid Bundling Method for Early-Exercise Options. *Int. J. Comput. Math.*, 92:2433–2454, 2015.
- [69] G. Linde. High-Order Adaptive Space Discretizations for the Black-Scholes Equation. Master’s thesis, University of Uppsala, 2005.
- [70] C. Mishra. *Stability of Alternating Direction Implicit Schemes with Application to Financial Option Pricing Equations*. PhD thesis, Univ. Antwerp, 2014.
- [71] C. Mishra. A New Stability Result for the Modified Craig-Sneyd Scheme Applied to Two-dimensional Convection-Diffusion Equations with Mixed Derivatives. *Appl. Math. Comput.*, 285(C):41–50, 2016.
- [72] A. R. Mitchell and G. Fairweather. Improved Forms of the Alternating Direction; Methods of Douglas, Peaceman, and Rachford for Solving Parabolic and Elliptic Equations. *Numer. Math.*, 6(1):285–292, 1964.
- [73] A. Papanicolaou. Introduction to Stochastic Differential Equations (SDEs) for Finance. *Lecture Notes*, 2013.
- [74] E. Pindza, K. Patidar, and E. Ngounda. Implicit-Explicit Predictor-Corrector Methods Combined with Improved Spectral Methods for Pricing European Style Vanilla and Exotic Options. *Electr. Trans. Numer. Anal.*, 40:269–293, 2013.
- [75] O. Pironneau. Calibration of Options on a Reduced Basis. *J. Comput. Appl. Math.*, 232(1):139–147, 2009.
- [76] D. M. Pooley, K. R. Vetzal, and P. A. Forsyth. Convergence Remedies for Non-smooth Payoffs in Option Pricing. *J. Comp. Fin.*, 6(4):25–40, 2003.
- [77] H. Price. Monotone and Oscillation Matrices Applied to Finite Difference Approximations. *Math. Comput.*, 22(103):489–516, 1968.

- 
- [78] C. Reisinger. *Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben*. PhD thesis, Ruprecht-Karls-Universität Heidelberg, 2004.
- [79] C. Reisinger. Analysis of Linear Difference Schemes in Sparse Grid Combination Technique. *IMA J. Numer. Anal.*, 33(2):544–581, 2013.
- [80] C. Reisinger and G. Wittum. Efficient Hierarchical Approximation of High-dimensional Option Pricing Problems. *SIAM J. Sci. Comput.*, 2007.
- [81] M. Rubinstein. Implied Binomial Trees. *J. Fin.*, 49(3):771–818, 1994.
- [82] T. Schiekofer. *Die Methode der Finiten Differenzen auf dünnen Gittern zur Lösung elliptischer und parabolischer partieller Differentialgleichungen*. PhD thesis, Universität Bonn, 1999.
- [83] R. Seydel. *Tools for Computational Finance*. Springer, 2006.
- [84] J. Shen and H. Yu. Efficient Spectral Sparse Grid Methods and Applications to High-Dimensional Elliptic Problems. *SIAM J. Sci. Comput.*, 32(6):3228–3250, 2010.
- [85] J. Shen and H. Yu. Efficient Spectral Sparse Grid Methods and Applications to High-Dimensional Elliptic Equations II. Unbounded Domains. *SIAM J. Sci. Comput.*, 34(2):A1141–A1164, 2012.
- [86] S. Smolyak. Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions. *Dokl. Akad. Nauk SSSR*, 148:1042–1045, 1963.
- [87] K. Spanderen. Finite Difference Schemes for the Heston-Hull-White Model. <https://hpcquantlib.wordpress.com/2011/09/11/finite-difference-schemes-for-the-heston-hull-white-model/>, 2011.
- [88] W. F. Spitz. *High-Order Compact Finite Difference Schemes for Computational Mechanics*. PhD thesis, University of Texas, Austin, 1995.
- [89] W. F. Spitz and G. F. Carey. High-Order Compact Scheme for the Steady Streamfunction Vorticity Equations. *Int. J. Numer. Meth. Engin.*, 38:3497–3512, 1995.
- [90] D. Tavella and C. Randall. *Pricing Financial Instruments: The Finite Difference Method*. Wiley, New York, 2000.
- [91] T. W. Tee and L. N. Trefethen. A Rational Spectral Collocation Method with Adaptively Transformed Chebyshev Grid Points. *SIAM J. Sci. Comput.*, 28(5):1798–1811, 2006.
- [92] J. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*. Springer, 1995.
- [93] L. N. Trefethen. *Spectral Methods in MatLab*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.
- [94] H. van der Vorst. Bi-CGSTAB: a Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems. *SIAM J. Sci. Statist. Comput.*, 13:631–644, 1992.

- 
- [95] S. Villeneuve and A. Zanette. Parabolic ADI Methods for Pricing American Options on Two Stocks. *Math. Oper. Res.*, 27(1):121–149, 2002.
- [96] L. B. Wahlbin. A Remark on Parabolic Smoothing and the Finite Element Method. *SIAM J. Numer. Anal.*, 17(1):33–38, 1980.
- [97] F. Zeng, F. Liu, C. Li, K. Burrage, I. W. Turner, and V. Anh. A Crank-Nicolson ADI Spectral Method for a Two-dimensional Riesz Space Fractional Nonlinear Reaction-Diffusion Equation. *SIAM J. Numer. Anal.*, 52(6):2599–2622, 2014.
- [98] C. Zenger. Sparse Grids. Technical Report, Institut für Informatik, Technische Universität München, October 1990.
- [99] W. Zhu and D. Kopriva. A Spectral Element Approximation to Price European Options with One Asset and Stochastic Volatility. *J. Sci. Comput.*, 42(3):426–446, 2010.
- [100] R. Zvan, P. Forsyth, and K. R. Vetzal. Penalty Methods for American Options with Stochastic Volatility. *J. Comput. Appl. Math.*, 91(2):199–218, 1998.