Laura Viktoria Somorowsky

Vom Fachbereich IV für Wirtschafts- und Sozialwissenschaften, Mathematik und Informatikwissenschaften der Universität Trier zur Verleihung des akademischen Grades Doktorin der Naturwissenschaften (Dr. rer. nat.) genehmigte Dissertation

The Nonlocal Spatial Ramsey Model with Endogenous Productivity Growth

Betreuer und Berichterstatter: Prof. Dr. Ekkehard Sachs
Berichterstatter: Prof. Dr. Fredi Tröltzsch
Berichterstatter: Prof. Dr. Georg Müller-Fürstenberger

Datum der wissenschaftlichen Aussprache: 29. Juni 2018

Trier, 2018

Diese Arbeit entstand im Rahmen des Graduiertenkollegs Algorithmic Optimization (ALOP)

Betreuende PI: Prof. Dr. L. Frerick [1]

Prof. Dr. G. Müller-Fürstenberger [2]

Prof. Dr. E. Sachs [1]

[1] Mathematik, Universität Trier
[2] Volkswirtschaftslehre, Universität Trier

Contents

German Summary

Die Wachstumstheorie ist ein Zweig der Volkswirtschaftslehre, der untersucht, welche Faktoren wie und wie stark die wirtschaftliche Entwicklung einer Ökonomie beeinflussen. Ein viel beachtetes neoklassisches Wachstumsmodell ist das *Ramsey-Cass-Koopmans Modell*, welches untersucht, wie viel des Einkommens der konsumierende Sektor einer Ökonomie sparen sollte, um den maximalen Gesamtnutzen zu erzielen. In diesem Modell wurde ursprünglich unterstellt, dass sich Produktionsfaktoren nur durch die Zeit bewegen können, Kapital also durch die Sparentscheidungen eines Haushaltes in die Zukunft transferiert werden kann. Mit dem Aufkommen der *Geographical Economics* wurde diese Annahme dahingehend erweitert, dass auch Flüsse der Produktionsfaktoren durch (geographischen) Raum modelliert werden.

Diese Arbeit befasst sich mit der Entwicklung und Analyse einer neuen Kapitalakkumulationsgleichung für das räumliche Ramsey Modell. Kapitalflüsse durch den Raum werden durch einen nichtlokalen Diffusionsoperator modelliert, welcher Sprünge des Kapitalbestandes von einem Punkt zu einem anderen zulässt und die Umverteilung von Heterogenitäten verzögert. Darüber hinaus wird ein endogener Produktivitäts-Produktions-Operator vorgestellt, welcher den technologischen Fortschritt einer Ökonomie darstellt. Dabei ist dieser abhängig von der Zeit und der Verteilung des Kapitals im Raum.

Das resultierende mathematische Modell ist ein Optimalsteuerungsproblem unter einer parabolischen partiellen Integro-Differentialgleichung, Anfangs- und Randwertbedingungen sowie Boxconstraints an Zustands- und Steuerungsvariable. Im Rahmen dieser Arbeit wird dieses Modell zum einen auf einem unbeschränkten und zum anderen auf einem beschränkten Ortsgebiet untersucht. Beide Male wird ein endlicher Zeithorizont betrachtet. Die Hauptresultate dieser Arbeit sind die Existenzbeweise einer schwachen Lösung der partiellen Integro-Differentialgleichung unter einer gemischten lokal-nichtlokalen Diffusion, sowie einer optimalen Steuerung im Modell auf unbeschränktem Ortsgebiet und der Nachweis der Existenz einer schwachen Lösung der rein nichtlokalen Kapitalgleichung im nichtlokalen räumlichen Ramsey Modell mit endogenem Produktivitätswachstum auf beschränktem Ortsgebiet.

Die Arbeit endet mit der numerischen Umsetzung des neuen nichtlokalen Wachstumsmodells und einer ökonomischen Auswertung.

English Summary

'How much of it's income should a nation save?' (Ramsey, 1928, p.543)

Economic growth models are of huge interest in macroeconomic research. They aim to determine how long-run economic growth, which can be measured for example in the percentage change of various indicators such as gross domestic product (GDP) or GDP per capita, is generated, analyze how it can last, and in that way explain the observable differences in output levels and growth rates across different countries or different times. Especially in the *Neoclassical Growth Theory* which was developed in the last century, three factors are assumed to be responsible for economic growth: capital, labor, and technology. Most neoclassical growth models are equilibrium models and claim that adjusting the three growth driving factors appropriately, a temporary equilibrium can be achieved.
In the last decades, a new trend has significantly changed the economic view on economic growth, namely the *Geographic Economics*. In these models, the accumulation of growth driving factors is not considered as a purely time-depending process anymore, but also spatial agglomeration effects are taken into account.

In this thesis, we focus on one seminal neoclassical growth model, the Ramsey-Cass-Koopmans model. As a prime example of a neoclassical growth model, it has become a corner stone in macroeconomics since its development in 1928. Though originally only time-depending, it also has been spatialized in the previous years. This spatial Ramsey model analyzes which distribution of capital and labor over time and space maximizes the welfare of an economy.
Especially when production factors are not only mobile through time, but also through space, when the optimal saving decisions in the consuming sector of an economy are determined within the accumulation process itself, or when endogenous technology change is considered, these models become quite complex - from an economic and mathematical point of view.

In this monograph, we develop a new spatial Ramsey model, including nonlocal diffusion effects in the capital accumulation equation to describe the mobility of this production factor across space. Moreover, we consider an endogenous productivity growth which is mainly driven by spillover effects in time and the amount of capital in a location and the respective surrounding. The resulting Ramsey equilibrium is given as the solution of an optimal control problem with a convex objective function under a semilinear parabolic partial integro-differential equation. We consider the problem on bounded and unbounded spatial domains. Both cases are important for the application, since boundaries and interactions between different countries always affect economies and hence economic growth.

This thesis can be divided into three parts. In the first part, we start with an introduction to economic growth models in general in Chapter 2.1. Here, we give a short historical overview on economic growth models and differentiate the classical from the neoclassical growth theory. Afterwards, we introduce the Ramsey-Cass-Koopmans model in discrete time, which is one of the oldest versions of this model. As in many other equilibrium models, Ramsey considers an economy with two sectors, namely the producing sector, consisting of the firms, and the consuming sector, which is composed by the households. The seminal idea in Ramsey's work is the lifetime utility maximization approach in the consuming sector. Like other neoclassical growth models, also in his model firms are assumed to tend to maximize their gain in every point in time. Groundbreaking in economic growth theory was his idea of an endogenous saving rate that directly affects the capital distribution over time and that is determined via the consumers' ambition to maximize their utility over all points in time. Though already rather challenging from an economic point of view due to this endogenous character of the model, the equilibrium is given as solution of a nonlinear optimization problem with equality constraints when time is considered in discrete periods, or as an optimal control problem under an ordinary differential equation, if time is continuous.
The Ramsey model is quite abstract and universal which gives rise to several more specific applications. In Chapter 2.3, we summarize several modifications of the Ramsey model, such as including technological progress in the model, or to consider a set of heterogeneous agents to motivate the spatial extension.
In Chapter 3, we give an introduction to the *Geographic Economics* which deal with spatial extensions of economic growth models. After providing all important mathematical background information to analyze the spatial models according to well-posedness in Chapter 3.2, we follow the derivation of the common spatial Ramsey model according to Brito (2001) in Chapter 3.3. In this *local spatial Ramsey model*, as we will call it throughout this thesis, capital mobility across space is modeled as Laplace operator. This (local) diffusion operator implies an infinite adjustment speed of the production factor and denies any impact of the welfare of areas 'far away' on the capital stock in a respective location.
To our knowledge, we are the first who impeach the validity of this assumption. In Chapter 3.4, we derive a new extension of the spatial Ramsey model, includ-

ing nonlocal diffusion effects and endogenous productivity growth. Such nonlocal diffusion effects are modeled as integral operators which describe slower adjustment speed in the diffusion process and also jumps. We assume that technological progress, which we identify with an increase of productivity, rises in time and moreover depends on the capital stock in a respective location and the welfare of the surrounding. The equilibrium in this nonlocal spatial Ramsey model is finally given as the solution of an optimal control problem with a quite general, convex objective function under a semilinear parabolic partial integro-differential equation.

Although Ramsey himself considers an infinite time horizon in his model which is convenient in economic terms, we do have to restrict the time horizon to a finite terminal date in order to derive some existence results. However, we introduce a terminal condition, that captures the infinite time horizon character of the Ramsey model, if chosen properly.

In the second part of this thesis, which is mainly the fourth chapter, we analyze the nonlocal spatial Ramsey model with endogenous productivity growth over an unbounded spatial domain. Unbounded spatial domains are of special interest in economic growth theory as they can be interpreted as one closed economy that does not interact with any other economy. There is no need to introduce any boundary conditions which also affect the optimal capital and labor distribution.

In Chapter 4.1, we derive an abstract existence result of a weak solution of a linear but homogeneous partial integro-differential equation with a local-nonlocal diffusion operator. Then, we apply this result to show existence of a solution in weak sense of the capital accumulation equation in the nonlocal Ramsey model with endogenous productivity growth. Here, we exploit the Lipschitz continuity assumption on the production function and the uniform boundedness of the productivity operator to show that the solution operator, that maps a right-hand side to the solution of such a PIDE, is a contraction on small time intervals. Since we only consider a finite time horizon in our model, we can construct the weak solution on the whole time-space cylinder after finitely many steps.

After we stated the existence result, we study the weak solution with respect to some stronger regularities. In the one-dimensional, spatially unbounded setting, we are indeed able to prove the overall boundedness and continuity of the weak solution in time and space. Here, we apply a result that guarantees the continuity of a weak solution of a semilinear parabolic differential equation under some regularity assumptions on the initial condition on compact subsets of the domain of interest and extend it to the whole unbounded space domain.

Following this result on the regularity of the weak solution, we prove the existence of the equilibrium in the nonlocal spatial Ramsey model with endogenous productivity growth in Chapter 4.2. We follow a common technique to prove the existence of an optimal control under a semilinear parabolic differential equation on bounded spatial domains and exploit the property of the kernel function in the nonlocal integral operator to vanish towards the edges. In this way, this kernel induces very naturally a weight and allows us to work in ‘ordinary’ Sobolev spaces instead of weighted function spaces that are usually considered in the context of unbounded space domains. The existence of an optimal control in our growth model over unbounded spatial domains is one of the main results in this thesis.

The last part of this monograph is provided in Chapters 5 and 6. Here, we analyze the nonlocal spatial Ramsey model with endogenous productivity growth over a bounded spatial domain with respect to the existence of a weak solution and implement the capital accumulation equation and the optimal control problem numerically. In this setting, we have to define some boundary conditions which are given as volume constraints acting on a set with nonzero volume in contrary to ordinary boundary conditions that are only defined on the surface of the set of interest. These volume constraints describe the interaction of the considered economy with the surrounding. Since the economy in this setting is still bounded but not closed any more, the transitional dynamics of the solution are important application results in these chapters. Moreover, we study how pure nonlocal diffusion affects the capital accumulation in contrast to the previous chapter, where we considered a combined local-nonlocal diffusion. Analyzing the model with respect to the existence of a weak solution of the capital accumulation equation, we need a different theory in order to derive some existence results, since the considered capital equation, we consider in this chapter, is a pure integral equation with differentiation only in time direction. Hence, we begin Chapter 5 with an introduction to the nonlocal vector calculus by Du et al. (2012a) in Section 5.1. This theory provides some important results such as the nonlocal version of Green’s identities and allows us to apply an existence result for a weak solution of a linear nonlocal differential equation of parabolic type to our semilinear case.
Modeling the capital accumulation over space as a pure nonlocal diffusion process is convenient for the application. However, the lack of derivatives in space direction in the capital accumulation equation leads to weak regularities of the weak solution. We point out that we can increase the regularity of the weak solution in the interior of the set of interest when we consider higher regular initial conditions and inhomogeneities, however we do not gain any smoothness across the boundary. This makes it very difficult to prove the existence of a Ramsey equilibrium also in the spatially bounded case, or an optimal control respectively. We will further discuss this challenge in Chapter 5.3.

Nevertheless, we solve the nonlocal spatial Ramsey model with endogenous productivity growth on bounded spatial domains numerically and construe the results. We introduce two approaches to implement an optimal control problem, namely the first discretize, then optimize and the first optimize, then discretize approach. But first of all, we illustrate the impact of the nonlocal diffusion and the productivity operator on the capital accumulation process in time and space by solving the integro-differential equation in Chapter 6.2.
In Chapter 6.3, we consider the numerical solution of the optimal control problem.

We implement the first discretize approach using a Crank-Nicolson scheme and a Gaussian quadrature rule and interpret the solutions in Section 6.3.1.
We do not implement the first optimize approach, but analyze the Ramsey model with respect to Fréchet differentiability and derive the necessary first order conditions in Section 6.3.2. These are given as a system of a semilinear and a linear inhomogeneous and nonlocal integro-differential equations.
We conclude this monograph with a comparison of the numerical solutions of the common local spatial model with the ones of our new nonlocal modification in Chapter 6.4.

Acknowledgments

During my time at the Department of Mathematics at Trier University, I was pleased about a steady support and help, without I would not have been able to write this thesis.

First of all, I want to express my thanks to my supervisor Prof. Dr. Ekkehard Sachs, who not only gave me the opportunity to write this thesis but also ongoing support. He always shared his wealth of experience, concerning not only mathematical knowledge, from which I have profited a lot. Moreover, he always encouraged me to present my work at conferences and workshops and introduced me to many interesting and important mathematicians.
At one occasion, I got to know Prof. Dr. Fredi Tröltzsch, whom I want to thank for much constructive advice, his interest in my work, and his willingness to review my thesis.
As this thesis is highly interdisciplinary, I want to express my gratitude to Prof. Dr. Georg Müller-Fürstenberger for interesting insights in economics, creative and fruitful discussions, and his participation as a referee.
Further, I want to thank Prof. Dr. Leonhard Frerick for his constant support, his encouraging words, and enthusiasm for my research.
This thesis has been developed within the Research Training Group *Algrothmic Optimization*. At this point I want to thank all principal investigators for many interesting and informative courses and constant support, independent of their responsibilities. In addition, I want to express my gratitude to the *German Research Foundation* (Deutsche Forschungsgemeinschaft DFG) for the financial support within the research training group.
During my time in ALOP, I had the opportunity to stay two months at the *Statistical and Applied Mathematical Science Institute* SAMSI in Raleigh. I want to thank SAMSI for financial support, the warm welcome, and a great experience.
At this point, I also want to thank Dr. Martha D'Elia for her interest and good advice.

I want to thank all my colleagues at the Trier University for a cheerful and amicable atmosphere. A special thanks goes to Dr. Christina Müller and Julian Wagner for uncountable constructive discussions. Moreover, I want to thank Dennis Kreber, Gennadij Heidel, Dr. Christina Müller, Bernd Perscheid, Christian Vollmann, Julian Wagner and Dr. Kathrin Welker for proofreading my monograph. I want to mention Yvonne Scheid, whom I also owe my gratitude for proofreading, although she finished with maths right after school.

And last, but in no case least, I want to thank my family.
Meine lieben Eltern Ute und Georg, meine beste Schwester Carolyn und auch meine lieben Großeltern, ihr habt mich mein gesamtes Leben lang mit Liebe, Vertrauen und Unterstützung begleitet. Ihr habt an mich geglaubt und gabt mir die Kraft und die Zuversicht, diese Arbeit zu beginnen und auch zu beenden. Auch dir, lieber Markus, möchte ich von Herzen danken. Du hast mir von Anfang an mit unermüdlicher Geduld zur Seite gestanden. Euer aller Rückhalt bedeutet mir unvorstellbar viel.

CHAPTER 1

Introduction and Outline

'Individual wealth [...] varies greatly within countries. Indeed at the sub-national level, individual wealth can vary across provinces, regions, cities and urban and rural classifications.' (Morrissey et al., 2011, p.80)

The *Brexit* may be one of the most revolutionizing decisions in the European Union since its foundation. Once founded as a customs union, the EU of today is a political and economic union of currently 28 member states. This economic and political coalition has a huge impact on the economy and the welfare of the participating countries. Considering the results of the Brexit vote, as illustrated in Figure 1.1, the almost strict disjunction between Brexit supporters located in the south and Brexit opponents in the north seems rather exceptional. Actually, this strict line exists in Britain not only with respect to the political attitude. Indeed, the United Kingdom is divided into north and south in many different fields which is a unique feature among the north-western European countries, at least according to Gordon Brown, the former Prime Minister of the UK (c. Brown, 2016). Although the discussion on the gap between incomes, employment rates, education standards, and the health system in northern and southern Britain is not new (cf. Armstrong and Riley, 1987, Green, 1988, and Martin, 1988), the recent British paper press reveals that this inequality is still an up-to-date topic (cf. *'Britain's North-South Divide: How it affects education, economy and*

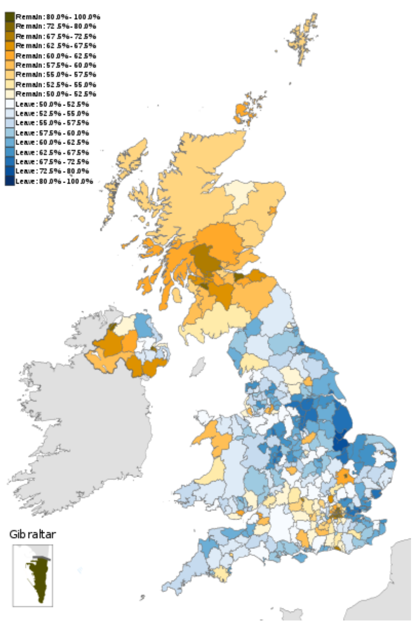


Figure 1.1: Geographical Differences in Brexit Vote, Mirrorme22 et al. (2016)

gender pay-gap?', Barnett, 2016).

This *North-South Divide* is not in line with the neoclassical competitive equilibrium theory (cf. Blackaby and Manning, 1990, p.1), where capital and labor flows should delete such disparities over time. Such neoclassical competitive equilibrium models appear in the economic growth theory, which tries to understand how economic growth, which has to be understood as the increase of the gross domestic product (GDP) or GDP per capita, is generated and how it can last. The neoclassical equilibrium models have become a working horse in this macroeconomic field, due to their abstract, universal structure.
One neoclassical growth model that is widely used is the Ramsey-Cass-Koopmans model, for short Ramsey model, which can be traced back to Ramsey (1928) and was further developed by Cass (1965) and Koopmans (1965), independently from each other. In contrast to others, the Ramsey model suggests not only an optimization intension in the producing sector of an economy, but considers a lifetime utility maximization approach in the consuming sector as well. The *optimal saving decisions* that generate the maximal welfare of an economy are defined endogenously in the so called *Ramsey equilibrium*, the state in which both optimization approaches are satisfied and all markets are cleared.
Originally, the flows of capital, labor, and production goods in the Ramsey economy were modeled only in time. But with the development of the so called *New Economic Geography*, spatial extensions of this model have been introduced. In these spatialized versions, production factors and goods move likewise through time and space. In the common spatial Ramsey models, going back to Brito (2001), the mobility of production factors through space is described as a (local) diffusion. This leads to an even capital and labor (or per capita capital) distribution across space, independent of the initial capital distribution. Moreover, these spatial models ignore technological progress which is also a driving factor of economic growth. Aldashev et al. (2014, p.14) point out, that the capital equation in the spatial Ramsey model by Brito (2001) is not appropriate. In addition to that, most neoclassical growth models do either not consider any heterogeneity at all or enforce any initial heterogeneity to disappear over time. Neither the Ramsey model nor its spatial extension are an exception. However, the example of the UK shows that *the Heterogeneity* matters for the progress of economic growth and political decisions, and thus should be taken into account.

In this thesis, we develop a new capital accumulation equation of the spatial Ramsey model. We consider a nonlocal diffusion effect that describes the capital mobility across space. In that way, we are able to model jumps of capital stocks and to preserve heterogeneities in initial capital distributions. Moreover, we introduce a new productivity-production operator that describes the technological progress of the considered economy. We claim that the increase of an initial productivity distribution is exponential in time and depends on the capital distribution in the surrounding of a respective location. Both adjustments of the capital equation in this neoclassical model aim to complete the accumulation process of capital in time and space, which is mainly driven by the interaction of diffusion and agglomeration effects (cf. Camacho et al., 2008), and to take any heterogeneity in an initial spatial welfare distribution into account.

The outline of this thesis is as follows:

In **Chapter 2**, we give a short introduction to *neoclassical growth models*. These economic models are *competitive equilibrium models* that determine the market clearing prices and the optimal allocation of production factors and goods as the result of the interaction of supply and demand in an economy. In general, these models consider an economy that is described by two sectors, the producing and the consuming sector. Neoclassical growth models are characterized by the special form of the production function that models the producing side of the economy and of the utility function which describes the consuming sector. These functions exhibit constant returns to scale and diminishing returns.

We introduce the *Ramsey model in discrete time*. The Ramsey equilibrium in this setting is given as the solution of a nonlinear optimization problem,

$$\begin{aligned} &\max_{c,k} \sum_{0}^{\infty} \beta^t u(c_t) \\ s.t.\ &k_{t+1} + \delta k_t - p(k_t) = -c_t,\ t = 0, 1, 2... \\ &k_{t+1}, c_t \geq 0,\ t = 0, 1, 2... \\ &k_0 > 0 \text{ given.} \end{aligned} \tag{1.1}$$

Based on the economic *welfare theorems*, the solution of this optimization problem is indeed the market equilibrium in this economy. We end this chapter with a short overview on existing applications, where we focus on borrowing constraints, heterogeneities in the consuming sector, and the modeling of technological change.

We continue the introduction of the economic background of this thesis with a short presentation of the *Geographical Economics* in **Chapter 3**, a field of economic growth theory that considers mobility of production factors and goods also through space and not only through time. We focus on the spatial extensions of the Ramsey model and discuss the capital accumulation equations in these models.
After giving the basic background information on partial differential equations, we derive the common spatial Ramsey model following Brito (2001). In this spatial expansion, the Ramsey equilibrium is defined as the solution of an optimal control problem under a parabolic partial differential equation,

$$\max_{k,c}\ \mathcal{J}(k, c) := \int_0^\infty \int_{\mathbb{R}} U(c(y, t)) e^{-\tau t - \gamma |x|} dt dy,$$

subject to

$$\begin{aligned} \frac{\partial k}{\partial t} - \frac{\partial^2 k}{\partial x^2} + \delta k - Ap(k) &= -c && \text{on } \mathbb{R} \times \mathbb{R}_+ \\ \lim_{x \to \pm\infty} \frac{\partial k}{\partial x} &= 0 && \text{in } \mathbb{R}_+ \\ k(\cdot, 0) = k_0(\cdot) &> 0 && \text{in } \mathbb{R} \\ k, c &\geq 0 && \text{on } \mathbb{R} \times \mathbb{R}_+ \end{aligned}$$

for discount factors $\tau, \gamma > 0$, a productivity growth factor A which may either be a constant in $\mathbb{R}_+$, or a nonnegative, real valued function in x and/or t, and a depreciation rate $\delta > 0$. Here, the function k denotes the (per capita) capital stock of an agent located in x at time t and c stands for the consumption. By $\mathbb{R}_+$, we mean all nonnegative real numbers. We discuss the choice of the infinite time horizon and an unbounded spatial domain.
The capital accumulation equation in this (local) spatial Ramsey model exhibits some weaknesses which motivate the definition of a new, nonlocal capital constraint in the spatial Ramsey model. We give a detailed derivation of our expanded version of this equation, introducing a nonlocal diffusion operator and a nonlinear, nonlocal operator that describes the technological progress in the economy. In contrast to other models, the productivity operator in our version determines the technological change endogenously.

In **Chapter 4**, we analyze the well-posedness of our nonlocal spatial Ramsey model with endogenous productivity growth for a finite time horizon and on unbounded spatial domains. The choice of such an unbounded domain is convenient for the application, since the absence of any boundary conditions allows the analysis of the economy without any interaction with other economies, or some abstract surrounding.

We show that, under appropriate assumptions, the nonlocal capital accumulation equation,

$$\begin{aligned} k_t - \mathcal{L}(k) + \delta k - \mathcal{P}(k) &= -c && \text{on } \mathbb{R}^n \times (0, T) \\ k(\cdot, 0) = k_0(\cdot) &> 0 && \text{in } \mathbb{R}^n, \end{aligned}$$

admits a unique weak solution $k \in W(0,T)$ for initial data $c \in L^2(0,T; H^{-1}(\mathbb{R}^n))$ and $k_0 \in L^2(\mathbb{R}^n)$.
In the setting of an unbounded spatial domain, we define the local-nonlocal diffusion operator $\mathcal{L}$ as

$$\mathcal{L}(k)(x,t) := \alpha\ \Delta k(x,t) + \beta \int_{\mathbb{R}^n} (k(y,t) - k(x,t))\Gamma_\varepsilon(x,y)\ dy$$

for coefficients $\alpha, \beta > 0$ and $\varepsilon > 0$, and the kernel as the density function of the multivariate normal distribution,

$$\Gamma_\nu(x,y) := \frac{1}{\sqrt{(2\pi\nu^2)^n}} \exp\left(-\frac{1}{2}(x-y)^T\Sigma_\nu^{-1}(x-y)\right)$$

for given constants $\nu \in \{\varepsilon, \mu\}$ with $0 < \mu < \varepsilon$ and a covariance matrix $\Sigma_\nu \in \mathbb{R}^{n\times n}$ with $\det \Sigma_\nu = \nu^{2n}$. The nonlocal operator $\mathcal{P}$ on the left-hand side describes the production of the economy and is given as

$$\begin{aligned}\mathcal{P}(k)(x,t) :&= P(k)(x,t)\ p(k(x,t))\\ &= A_0(x) \exp\left(\frac{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\varepsilon(x,y)dy + \xi}\ t\right)\ p(k(x,t)),\end{aligned}$$

where $A_0 : \mathbb{R}^n \to \mathbb{R}_+$ denotes the initial productivity distribution over space, $\xi > 0$ is a constant, and $\phi : \mathbb{R} \to \mathbb{R}_+$ is a continuous function such that the integrals exist. We will refer to ϕ as the nominal function.

Moreover, we prove the a priori estimates

$$\|k\|_{L^2(0,T;H^1(\mathbb{R}^n))}\ +\ \mathrm{C}_1\|k\|_{L^\infty(0,T;L^2(\mathbb{R}^n))} \le \mathrm{C}_2\left(\|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + \|k_0\|_{L^2(\mathbb{R}^n)} + 1\right)$$

and

$$\|k\|_{W(0,T)} \le \mathrm{C}_3(\|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + \|k_0\|_{L^2(\mathbb{R}^n)} + 1),$$

for some constants $\mathrm{C}_1,\ \mathrm{C}_2,\ \mathrm{C}_3 > 0$.

One of the main results in this chapter is the proof of the continuity and the essential boundedness of the weak solution. This seminal result enables us to finally show the existence of an optimal control of the nonlocal spatial Ramsey model on unbounded spatial domains, which yields the existence of a competitive market equilibrium.

Since most economies are not completely autarkic, it is also of interest to study the nonlocal Ramsey model with endogenous productivity growth on a bounded spatial domain Ω. This is done in **Chapter 5**. In this setting, we need to define so called volume constraints which model the interaction of the considered economy with its surrounding. We consider a pure nonlocal diffusion model, hence do not include any local diffusion operator in the capital accumulation equation. We truncate the action of the kernel function Γ_ν to a ν-surrounding of a respective location of interest, which enables us to embed our model in the nonlocal vector calculus of Du et al. (2012a). For initial data $c \in L^2(0,T;V_c'(\Omega))$ and $k_0 \in V_c(\Omega)$, we derive an existence result of a weak solution $k \in \mathcal{C}(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}})) \cap H^1(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}}))$ of the PIDE constraint. Moreover, for a constant $C_\infty > 0$, we show the a priori estimate

$$\|k\|_{H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))} \leq C_\infty \left(\|c\|_{L^2(0,T;L^2(\Omega))} + \|k_0\|_{L^2(\Omega)} + 1\right).$$

In the last chapter of this thesis, **Chapter 6**, we implement the nonlocal spatial Ramsey model with endogenous productivity growth as introduced and analyzed in Chapter 5. We introduce a quadrature-based finite difference method to discretize the PIDE and compare it to a finite element solution. We point out the agglomerative effect of the productivity-production operator and the ability of the nonlocal diffusion operator for preserving heterogeneities and discontinuities in the initial data.
In addition to that, we implement the *first discretize, then optimize* approach to solve the whole optimal control problem and discuss the Fréchet differentiability of the nonlinearity and the state-solution operator in order to derive the necessary first order conditions.
We complete this chapter and this monograph with a detailed comparison of the numerical solution of the new, nonlocal spatial Ramsey model with endogenous productivity growth and of the local model by Brito (2001).

CHAPTER 2

Fundamentals of Neoclassical Growth Theory

'The process of economic growth and the sources of differences in economic performance across nations are some of the most interesting, important, and challenging areas in modern social science.' (Acemoglu, 2009, p.XV)

The Ramsey model is a widely used neoclassical growth model, which aims in analyzing how economic growth is generated and under what circumstances it can be persistent. In this chapter, we provide a short introduction to the *(neoclassical) economic growth theory* in Section 2.1. Afterwards, we derive the equilibrium problem of the Ramsey model in discrete time in Section 2.2 in order to explain the dynamic and endogenous character of the model. We end this introduction with a summary of some interesting applications and modifications of the original model in Section 2.3.

2.1 Economic Growth and Equilibrium Models

Economic growth is a dynamic process where output, capital, consumption, and population patterns change over time. Of course, there is a great interest in understanding how growth, which is to be understood as the increase of economic quantities such as income per capita, is generated and under which conditions it is persistent (cf. Acemoglu, 2009). The analysis of differences and interdependencies across countries relates to this field of interest (cf. Barro and Sala-i Martin, 1995, Acemoglu, 2009). Theoretical models have been developed that explain economic growth and in this way, give not only insight, but also policy advice to fields like development economics.

The origin of the modern growth theory may go back to the late 1770th, when Adam Smith published his outstanding work about *'The Wealth of Nations'* (Smith, 1776). In the following two centuries, many other impressive economists like Ricardo (1891), Young (1928), Schumpeter (1939), and Knight (1944) further developed Smiths' theories and expanded the so called *classical growth theory*, which can be characterized mainly by the ground-breaking assumptions of *'competitive behavior and equilibrium dynamics, the role of diminishing returns and its relation to the accumulation of physical and human capital, the interplay between per capita income and the growth rate of population, the effects of technological progress in the form of increased specialization of labor and discoveries of new goods and methods of production, and the role of monopoly power as an incentive for technological advance'* (Barro and Sala-i Martin, 1995, p.9).

The turning point from classical to *neoclassical growth theory* was reached in the 1950th, when Solow (1956) and Swan (1956) independently developed a relatively simple growth model, an equilibrium model that also took productivity growth and savings as economic growth driving factors into account. *'The key aspect of the Solow-Swan model is the neoclassical form of the production function, a specification that assumes constant returns to scale, diminishing returns to each input, and some [...] elasticity assumptions'* (Barro and Sala-i Martin, 1995, p.10). In this model, economic growth is driven by four factors: labor, capital, savings, and technology (or productivity). The saving rate in the consuming sector is assumed to be given exogenously. The purpose of the Solow-Swan model is to detemine the appropriate factor constellation that generates a so called *(economic) equilibrium*.

Equilibrium models are literally workhorses in macroeconomics. *'Based on the Walrasian tradition, [...] equilibrium models describe the allocation of resources in a market economy as the result of the interaction of supply and demand, leading to equilibrium prices. The building blocks of these models are equations representing the behavior of the relevant economic agents'* (Borges, 1986, p.8). According to Magill and Quinzii (1998, pp.29-30), five ingredients are needed to built such an equilibrium model:

1. The *time-uncertainty setting* describes the commodity space and how time and uncertainty is modeled.

2. The *real side of economy* describes goods, agents, resources, preferences, and technology.

3. The *market structure* describes all trading arrangements between the agents.

4. The *behavior of agents* describes how agents make their decisions.

and finally

5. The *concept of equilibrium*, *'which describes the conditions under which the agents' decisions are mutually consistent'*.

The most common conditions that are stated in the context of neoclassical growth models are *market clearing conditions* and the *compatibility of expectations* of all agents (Magill and Quinzii, 1998, p.36). Hence, the underlying consideration in these models is the assumption that market forces will lead to an equilibrium between demand and supply. These equilibrium models enable the calculation of equilibrium prices under which all markets clear (Borges, 1986, p.8).
The study of market equilibria is a central aspect of economic analysis, as well as the exploration of optimality. The notion of market equilibrium and the so called *Pareto-optimality* are closely inter-related. In economics, optimality usually has to be understood in the sense of Pareto. In mathematical and economic terms, a Pareto optimal state is defined as the state, or allocation of resources, where no agent or preference criterion can be made better off, without making any other agent or preference criterion worse off. Adam Smith already recommended that any market equilibrium is also 'socially optimal', but he did neither give a precise definition, nor did he study this fact analytically. The so called *welfare theorems*, proven by Arrow (1951) and Debreu (1951), state that market equilibria and Pareto optima are indeed equivalent under certain assumptions. For a formal treatment of this equivalence due to Lerner and Lange, see for example Samuelson (1947). Moreover, a good summary of all necessary assumptions under which either a market equilibrium is Pareto optimal, or under which any Pareto-optimal state of an economy can be achieved via a market clearing choice of prices and a reallocation of income, can be found in the book of Acemoglu (2009, pp.164-176).

A neoclassical equilibrium growth model, which has been quite unnoticed until the 1960s, is the model introduced by Frank P. Ramsey in 1928. His model was *'mathematically quite demanding and surely in advance of his time'* (Heinemann, 2015, p.57). The combination of Ramsey's idea of an endogenous saving rate with the simpler Solow growth model, as introduced in two path-breaking works by Cass (1965) and Koopmans (1965), made the so called *Ramsey-Cass-Koopmans model* become a cornerstone in neoclassical growth theory.
The essential difference between the Solow model and the Ramsey model is that the latter also takes the households' utility optimization into account. Instead of exogenously committing a fixed saving rate, households seek to maximize their lifetimes' utility. In this way, the saving rate is endogenously determined, whereas it is exogenously given in the Solow model. The optimization approach in the households' sector is the reason why the Ramsey model is also known as the *'model of optimal growth'* (Heinemann, 2015, p.57).

2.2 The Ramsey Model

The Ramsey model is one of the standard models in the neoclassical growth theory. It has to reduce the complexity in a real economy to a level where the dynamics in such an economy can be caught by mathematical equations. Nevertheless, *'the theoretical superiority of the general equilibrium approach, [that is used in the*

Ramsey model], has always been accepted' (Borges, 1986, p.9).

Considering a general market economy, the simplifications in the Ramsey model concern the number and heterogeneity of agents, the supply of available and tradable goods, and multiple social interactions. Like the Solow model mentioned above, the Ramsey model considers a simple one-good and closed economy (cf. Acemoglu, 2009, p.27). Time can be measured in discrete periods, $t = 1, 2, ...,$ or as a continuous interval and the time horizon can either be infinite ($t \in \mathbb{R}_+$), or finite ($t \in [0, T],\ T \in \mathbb{N}$). In the original model, Ramsey considers an infinite-horizon economy in continuous time. Although individuals have finite lifetimes, the assumption of immortal households is appropriate if these correspond *'to finite-lived individuals who are connected via pattern of operative intergenerational transfers that are based on altruism'* (Barro and Sala-i Martin, 1995, p.60). Since this thesis deals with finite time horizon economies only, the introduction in this chapter will be limited to finite time modeling. We will argue below that we can force a solution of our spatial Ramsey model to have infinite time horizon-character when we introduce a sustainable terminal condition on the capital stock.

Many economic growth models have the same underlying general equilibrium structure. An economy is described as a two-sector model with one consuming and one producing sector. The consuming side can either be described by a single so called *representative household* (cf. Acemoglu, 2009 or Barro and Sala-i Martin, 1995) or by a set of *heterogeneous households* that differ according to their time and utility preferences (see for example Becker, 1980 or Becker and Foias, 1987). Representing the production side in the economy by one representative firm does not require as stringent assumptions on the structure of the production side as it is the case in the consuming sector. Hence, *'the entire production side of the economy [is] represented by an aggregated production possibility set'* without any loss of generality (Acemoglu, 2009, p.158).

As we only aim to give a broad description of the Ramsey model in this section, the heterogeneity in the individuals' preferences is excluded in this context. Instead, the existence of one *normative representative household* is assumed, whose saving and consumption decisions reflect the decisions of a set of heterogeneous households. In the case that all households have so called *Gorman preferences*, which means that preferences *'can be represented by [some] special linear indirect utility functions'* (Acemoglu, 2009, p.151), the existence of such a representative household is not as implausible as it may appear at first glance. For further information see Acemoglu (2009, pp.147–155).

In the following, the time horizon is assumed to be finite, hence $T \in \mathbb{N}$ and time is measured in discrete periods, $t \in \mathcal{T}_0 := \{0, 1, ..., T\}$.

The Producing Sector

The production side in the Ramsey economy is modeled via a so called *neoclassical production function*. The form of the production function gives its name to the whole theory that tries to analyze economic growth in such settings. As already mentioned, the firms, that constitute the production side, use aggregated quantities, i.e. aggregated capital and aggregated labor, which they demand from the households. Throughout this chapter, capital letters will refer to such aggregated variables, whereas small letters will be used for variables of single agents. In this frame, let the aggregated capital stock in a period $t \in \mathcal{T}_0$ be defined by

$$K_t \in \mathbb{R}_+$$

and the aggregated labor by

$$L_t \in \mathbb{R}_+.$$

The production side is described by a neoclassical production function,

$$F : \mathbb{R}_+^2 \to \mathbb{R},\ (K, L) \mapsto F(K, L),$$

which satisfies the following assumptions:

Assumption 2.1 (Neoclassical Production)**:**
The production function F is continuous on $\mathbb{R}_+$ and twice continuously differentiable on $\mathbb{R}_+\backslash\{0\}$ with

$$F_K(K,L) = \frac{\partial F(K,L)}{\partial K} > 0, \qquad F_L(K,L) = \frac{\partial F(K,L)}{\partial L} > 0,$$

$$F_{KK}(K,L) = \frac{\partial^2 F(K,L)}{\partial K^2} < 0, \quad F_{LL}(K,L) = \frac{\partial^2 F(K,L)}{\partial L^2} < 0,$$

for all $K, L \in \mathbb{R}_+\backslash\{0\}$ and it satisfies the Inada conditions

$$\lim_{K\to 0} F_K(K,L) = \infty \text{ and } \lim_{K\to\infty} F_K(K,L) = 0 \text{ for all } L > 0,$$

$$\lim_{L\to 0} F_L(K,L) = \infty \text{ and } \lim_{L\to\infty} F_L(K,L) = 0 \text{ for all } K > 0.$$

Moreover, $F(0, L) = 0 = F(K, 0)$ for all $K, L \geq 0$ and it exhibits constant returns to scale in K and L (cf. Acemoglu, 2009, p.29).

All components of Assumption 2.1 are important: The requirement that F is strictly concave in every variable means that the marginal products of both capital and labor are diminishing, hence *'more capital, holding everything else constant, increases output less and less. And the same applies to labor. This property is [...] referred to as "diminishing returns" to capital and labor'* (Acemoglu, 2009, p.29). The property of diminishing returns is what characterizes this production function as neoclassical. The *constant returns to scale* assumption is equivalent to F being *linearly homogeneous*, which means homogeneous of degree one in both variables.

Definition 2.2:
Let $n \in \mathbb{N}$ be a given dimension. A function $g : \mathbb{R}^n \to \mathbb{R}$ is homogeneous of degree $m \in \mathbb{R}_+$ in $x \in \mathbb{R}^n$ if

$$g(\lambda x) = \lambda^m g(x) \text{ for all } \lambda \in \mathbb{R}.$$

The following theorem is important while rewriting the output function in terms per capita:

Theorem 2.3 (Euler's Theorem for Homogeneous Functions)**:**
Let $g : \mathbb{R}^{n+2} \to \mathbb{R}$ be differentiable in the first two arguments $x, y \in \mathbb{R}$ with partial derivatives denoted by g_x and g_y and homogeneous of degree $m \in \mathbb{R}_+$ in x, y. Then

$$mg(x,y,z) = g_x(x,y,z)x + g_y(x,y,z)y \text{ for all } x,y \in \mathbb{R} \text{ and } z \in \mathbb{R}^n.$$

Moreover, $g_x(x,y,z)$ and $g_y(x,y,z)$ are homogeneous of degree $m-1$ in x and y.

For the proof, see for example Acemoglu (2009, p.30).

Given an aggregated production function, meaning a production function whose input variables are aggregated labor and capital, and assuming that the representative firm can buy labor and capital on some factor markets, the firm's optimization problem can be stated as follows:

For given factor prices $R_t = (1 + r_t)$ and ω_t, the profit maximization problem of the representative firm in period $t \in \mathcal{T}_0$ is given by the following static problem

$$\max_{K \geq 0, L \geq 0} F(K,L) - R_t K - \omega_t L$$

(cf. Acemoglu, 2009, p.32).
This optimization problem does not have a well-defined solution due to Assumption 2.1. Since F has constant returns to scale, there may not exist any solution $(K,L) \in \mathbb{R}_+^2$ that maximizes the firm's profit, or the solution may not be unique (Acemoglu, 2009, p.32). Remark that this representation does not need any prices for the final good since the price of the final good has been normalized to one, what is without any loss of generality. Moreover, the firm is taking the factor prices R_t and ω_t, which are to be understood as in terms of the final good, as given. This implies that there exist some *competitive markets* that will be described further in the following.
The constant returns to scale feature of the production function enables to write the optimization problem in terms per capita, which makes notation easier in the equilibrium case. The output in period $t \in \mathcal{T}_0$, defined as $Y_t = F(K_t, L_t)$, can be

rewritten as

$$y_t = f(k_t), \tag{2.1}$$

where $y_t := \frac{Y_t}{L_t}$, $k_t = \frac{K_t}{L_t}$ and $f(k_t) = F(\frac{K_t}{L_t}, 1)$, where the function $f : \mathbb{R}_+ \to \mathbb{R}$ still satisfies the Inada conditions and the constant returns to scale requirement.

Finally, the optimization problem in the producing sector in terms per capita is defined as

$$\max_{k_t \in \mathbb{R}_+} f(k_t) - R_t k_t - \omega_t. \tag{2.2}$$

The Consuming Sector

The main aspect in the Ramsey-Cass-Koopmans model is the endogenous saving rate that is determined via the households' utility optimization. The consuming sector in neoclassical growth models can either be characterized as *'a unit measure of households, [which means as an] uncountable number of households with total measure normalized to one'* (Acemoglu, 2009, p.147), i.e. the unit interval, or as an infinite but countable set of households, i.e. the set of natural numbers $\mathbb{N}$, or as a finite set $\mathcal{H} := \{1, ...H\},\ H \in \mathbb{N}$ (cf. Becker and Foias (1987)).
Like in the basic general equilibrium theory, the preference ordering of every household $h \in \mathcal{H}$ is represented by a so called *felicity function* or *utility function* denoted by $U^h : \mathbb{R}_+^{T+1} \to \mathbb{R}$ that captures the utility which an individual (or household) derives from consumption in all points of time. It is commonly assumed that this felicity function is *time-separable and stationary*, which means that the utility at any date is independent from consumption in all former and future dates and that the instantaneous utility function is the same in all periods. Moreover, it is assumed that the households discount future utility exponentially. Hence, the lifetime utility of a household h is given as

$$U^h(c_0^{(h)}, c_1^{(h)}, ..., c_T^{(h)}) = \sum_{t=0}^{T} (\beta^{(h)})^t u^h(c_t^{(h)}), \tag{2.3}$$

where $\beta^{(h)} \in (0,1)$ is the *time discount rate* that discounts future utility and $u^h : \mathbb{R}_+ \to \mathbb{R}$ denotes the *instantaneous utility function* of the h^{th} household (Acemoglu, 2009, p.148). The variable $c_t^{(h)} \in \mathbb{R}_+$ denotes the consumption of household h at time t. The notation $\boldsymbol{c}^{(h)}$ will be used in the following in order to refer to the whole vector $(c_0^{(h)}, ..., c_T^{(h)}) \in \mathbb{R}_+^{T+1}$.
Ramsey (1928) was the first who hypothesizes an economic behavior in the consuming sector. It was his seminal idea of a lifetime utility maximization in the households' sector that led to a new comprehension of economic growth.

In the Ramsey-Cass-Koopman model (as in many other growth models) the existence of *competitive markets* is assumed. *'In competitive markets, households and firms act in a price-taking manner and pursue their own objectives, and prices* *clear markets'* (Acemoglu, 2009, p.30). Moreover, it is supposed that all production factors, i.e. labor and capital, are owned by the households. It is a common assumption that labor is supplied inelastically, which means that the whole endowment of labor in an economy is supplied independently of the factor price, which is the wage rate in this case, as long as it is nonnegative. This *labor market clearing condition* can be written as a complementary slackness condition

$$L_t \leq \overline{L}_t,\ \omega_t \geq 0 \text{ and } (L_t - \overline{L}_t)\omega_t = 0,$$

where $\omega_t \geq 0$ denotes the wage rate, $\overline{L}_t$ the endowment of labor in the economy, and L_t the demand for labor at time t, or as equation

$$L_t = \overline{L}_t.$$

The households hold all capital in this economy and lend it to the firms. In return, they get interest payments. The *capital market clearing condition* is analog to the labor market clearing condition, hence

$$K_t = \overline{K}_t,$$

where $\overline{K}_t$ denotes the whole capital stock of the economy and K_t stands for the capital that the firms demand in the respective period. The factor price for the capital is the interest rate which is denoted by $R_t \geq 1$.

As already mentioned above, we assume the existence of one *representative household*. This means that the consuming side of the economy is structured such that all aggregated consuming and saving decisions, as well as all aggregated labor supply decisions, can be represented by the decisions of one single (fictive) household. An elementary example for an economy with one representative household is an economy where all households are identical, which means they have the same time and utility preferences and equal labor endowments.

For a given (representative) utility function $U : \mathbb{R}_+^{T+1} \to \mathbb{R}$ of the form (2.3), that is increasing and concave, the households' side of the economy is represented by the following maximization problem:

$$\max_{\boldsymbol{c}} U(\boldsymbol{c}), \tag{2.4}$$

where $\boldsymbol{c} \in \mathbb{R}_+^{T+1}$ denotes the consumption stream of the representative household (Acemoglu, 2009, p.150).
Households gain utility by consuming. They finance consumption with income, which they receive from the firms for offering labor and capital. Hence, they get an income stream and have to decide how much of their income they spend on consumption and how much they save in order to transfer capital to future dates. These saving decisions link the capital assets of one period t to the asset of the

following period $t+1$ as described, for example by Acemoglu (2009, p.220), in the following way: Assuming that the representative household offer one unit of labor in every period and that an initial capital endowment $k_0 > 0$ is given, its capital stock in period $t+1$ is recursively defined as

$$k_{t+1} = R_t k_t + \omega_t - c_t \quad \text{for } t \in \mathcal{T}_0, \tag{2.5}$$

where k_t denotes the capital asset or capital stock of the representative household in time t. In order to ensure that k_t does not reach negative infinity, the capital assets are often bounded from below. A common budget or borrowing constraint states

$$k_{t+1} \geq 0 \quad \text{for } t \in \mathcal{T}_0 \tag{2.6}$$

(cf. Becker (1980), Cass (1965), or Lucas and Stokey (1984)). This constraint means that the household is not allowed to hold negative assets at any point in time. Alternative constraints consider nonnegative lifetime budgets (cf. Acemoglu, 2009, pp.175), or allow to incur debt as long it can be repayed in a fixed number of future periods (cf. Becker et al., 2015, pp.5).

Not only the production function, but also the utility function has to satisfy some conditions:

Assumption 2.4 (Neoclassical Preferences)**:**
$u : \mathbb{R}_+ \to \mathbb{R}$ is a strictly increasing function, it is concave and twice continuously differentiable with $u'(c) > 0$ and $u''(c) < 0$ for all $c \in \mathbb{R}_+ \backslash \{0\}$. Moreover, it satisfies the Inada-conditions:

$$u(0) = 0, \ \lim_{c \to \infty} u'(c) = 0 \quad \textit{and} \quad \lim_{c \to 0} u'(c) = \infty$$

(cf. Acemoglu, 2009, p.287).

Then, the optimization intention in the consuming sector yields the following maximization problem: The representative household tends to maximize its lifetime utility (2.4) making those consumption and saving decisions that satisfy (2.5) and (2.6).

The Equilibrium

The main aspect in the Ramsey model is the (competitive) equilibrium, which is in neoclassical growth theory defined as the allocation of labor and capital and the time paths of factor prices such that both optimization approaches in the households' and firms' sector are satisfied and all markets are cleared. The following definition makes clear, that the households and firms both behave price-taking:

Definition 2.5 (Competitive Equilibrium)**:**
A competitive equilibrium *[...] consists of paths of consumption, capital stock, wage rates, and rental rates of capital,* $\{C_t, K_{t+1}, \omega_t, R_t\}_{t=0}^{T}$*, such that the representative household maximizes its utility given an initial capital stock* K_0 *and taking the time path of prices* $\{\omega_t, R_t\}$ *as given; firms maximize profits taking the time path of factor prices* $\{\omega, R_t\}_{t=1}^{T+1}$ *as given; and factor prices are such that all markets clear (Acemoglu, 2009, p.293).*

In the equilibrium, the factor prices are determined via the firms' optimization as

$$R_t = f'(k_t)$$

and

$$\omega_t = f(k_t) - k_t f'(k_t),$$

where f is the neoclassical production function defined in (2.1). Together with the market clearing conditions, the competitive equilibrium problem in the Ramsey model with one representative household and firm and the no-debt borrowing constraint is given by the following nonlinear optimization problem

$$\begin{aligned} &\max_{\boldsymbol{c},\boldsymbol{k}} \sum_{t=0}^{T} \beta^t u(c_t) \\ &s.t.\ k_{t+1} = f(k_t) - c_t, \\ &\qquad k_{t+1}, c_t \geq 0 \text{ for } t \in \mathcal{T}_0,\ k_0 > 0 \text{ given.} \end{aligned} \tag{2.7}$$

Defining the equilibrium in the Ramsey model by determining the factor prices via the equilibrium relationship is mathematically more convenient. However, Definition 2.5 gives an conceptual insight.

As already mentioned, there is a close relation between the competitive equilibrium, which is defined as the solution of the optimization problem (3.14), and the Pareto optimal factor distribution in the considered economy. This relation is based on the *first and second welfare theorems of economics*. The first theorem states that every market equilibrium is Pareto optimal whenever all households are nonsatiated and the aggregated price for the production factors is finite. Or vice versa, the second theorem says that every Pareto optimal factor distribution is a market equilibrium whenever every production set is convex and every preference relation is convex and nonsatiated (a more detailed version and the proofs of both welfare theorems can be found for example in Acemoglu, 2009, pp.163-171).

Ramsey (1928) introduces this model originally as continuous in time, but the discrete time-points make the whole derivation of the model, with focus on the separability assumptions on u and the homogeneity assumption on f, easier. Moreover, the discrete version is widely used in economic applications. When we consider time to be continuous, the nonlinear problem in (2.7) becomes an optimal control problem with ordinary differential equation (ODE) and state constraint, as described amongst others by Acemoglu (2009, p.299). In that case, the variables are functions, the production function in (2.1) and the utility function in Assumption 2.4 become superposition operators (see Definition 6.6), and the resulting optimal control problem is given as

$$\begin{aligned} &\max_{c,k} \int_0^T e^{\beta t} u(c(t))dt \\ &s.t.\ k_t = f(k(t)) - \delta k(t) - c(t), \\ &\quad\ k, c \geq 0 \end{aligned} \tag{2.8}$$

with $k(0) > 0$ given. Here, a new parameter, $\delta > 0$, is introduced that describes the depreciation of capital through time.

2.3 Applications of the Ramsey Model

The neoclassical growth model developed by Ramsey, Cass, and Koopmans in the last century has been modified and analysed a lot. A vast literature, dealing with different types of equilibrium, with finite or infinite time horizons in a continuous or a discrete setting, exists. The alpplications of this model vary from the consideration of taxation (cf. Sorger, 2002) over the impact of tax-funded public spending (amongst others in Barro, 1990) to a growing population (cf. Acemoglu (2009)). Some models consider a central planning authority (Cass, 1965) whereas some others study the existence of Pareto optima or even Nash equilibria in models with a set of households with different preferences (as mainly done by Becker, 1980, or Van and Vailakis, 2003). We do not consider to give an overview of all existing modifications of the Ramsey model, however we want to discuss the flexibility of the model and list some seminal works.

A Ramsey economy is usually described as a two-sector model with one consuming and one producing sector. As already mentioned, the consuming side can either be described by a single so called *representative household* (like in Acemoglu, 2009; Barro and Sala-i Martin, 1995; Brock et al., 2014; Heinemann, 2015; Ramsey, 1928; Weil, 2013, and Young, 1928) or by a set of *heterogeneous households* that differ according to their time and utility preferences (cf. Becker, 1980; Becker and Foias, 1987; Borissov and Dubey, 2015; Lucas and Stokey, 1984; Sorger, 2002, and Van and Vailakis, 2003). To our knowledge, Becker (1980) was the first who considered a set of distinct households with heterogeneous preferences in the Ramsey model. In his model, the consuming sector is described by a finite, or at least countable, set of different utility functions, each indexed by a respective household. This can either be interpreted as an economy with only finitely many households or as an economy where households with homogeneous preferences form finitely many heterogeneous unions. He shows that the Ramsey equilibrium in his model, which

is given as the solution of a nonlinear multicriteria optimization problem, is indeed Pareto optimal.
A direct implication of Becker's model is that not only the utility preferences, but also the optimal capital and consumption paths are heterogeneous. In that way, it can be seen as a first attempt to *spatialize* the Ramsey model.

The most interesting part in the Ramsey model is the endogenous saving rate that is determined by the welfare optimization of the households' sector. Since the saving rate is directly related to the capital stock via the capital constraint, any budget or borrowing constraint affects the equilibrium. Ramsey himself does not restrict the capital stock held by one of the individuals in the respective economy. He distinguishes between several possible constellations how, and under which circumstances the *maximum obtainable rate of enjoyment or utility* - he calls it *Bliss* - can be reached, if it is reached at all (Ramsey, 1928, p.545). However, in order to ensure the existence of a solution of the households' optimization problem (and its uniqueness) and to simplify, the capital stock of the households is often restricted. So called *borrowing constrains* are introduced, which are huge market imperfections in the context of economics. The most common way to restrict the households' capital stock is to bound it from below by zero,

$$k_t \geq 0, \quad \text{for all } t,$$

respectively in a setting of heterogeneous households,

$$k_t^h \geq 0, \ \text{for all } h, t,$$

see amongst others Alt (2002), Becker (1980), Becker and Foias (1987) and Sorger (2002). Becker et al. (1991) weaken this condition by restricting only the aggregated capital stock,

$$\sum_h k_t^h \geq 0, \quad \text{for all } t.$$

In the context of an infinite time horizon this market imperfection can be replaced by the so called *no-Ponzi-game condition*, which states that *'the present value of lifetime consumption must not exceed the present value of lifetime earnings'* (Sorger, 2002, p.229),

$$\lim_{t\to\infty} k(t) \prod_{s=1}^{t-1} \frac{1}{R(t)} \geq 0,$$

(cf Acemoglu, 2009, p.305). This *inter-temporal budget constraint* ensures that *'the individual does not asymptotically tend to negative wealth'* (Acemoglu, 2009, p.291) and prevent imperfection (Sorger, 2002, p.229).
Other budget constraints in the Ramsey model try to weaken the market imperfection which is a direct consequence of any borrowing constraints. Becker et al. (2015) for example introduce a liberal borrowing constraint to the Ramsey model in discrete time where single households are *'able to borrow against their future*

wage incomes for finitely many time periods' (Becker et al., 2015, p.3). This version is an extension of the model by Borissov and Dubey (2015), who relax the no-debt condition by allowing the households to borrow against their next period wage income.

Besides some modifications on the budget constraints, the literature provides also many modifications of the capital equation of the Ramsey model, including taxation or public spendings. A summary is given by Acemoglu et al. (2011). In this paper, the authors extend the common Ramsey model by a government budget constraint, which is linked to the budget constraint of the households. They study the dynamic taxation of capital and labor with linear taxation rates τ_k and τ_l, and derive a competitive equilibrium. The latter is given as the solution of the coupled optimization problem consisting of the optimization problems of the households,

$$\max_{c_h,k_h,l_h,b_h} \sum_{t=0}^{\infty} \beta^t (u(c_{h,t}) - v(l_{h,t}))$$

$$\begin{aligned} s.t.\ \ c_{h,t} + k_{h,t} + q_{t+1} b_{h,t+1} &\leq (1-\tau_{l,t})\omega_t l_{h,t} + (1-\tau_{k,t}) r_t k_{h,t} + \iota_t b_{h,t} \\ c_{h,t} \geq 0,\ k_{h,t+1} &\geq 0,\ l_{h,t+1} \geq 0,\ b_{h,t+1} \geq 0,\ \forall\ t,\ h \in \mathcal{H}, \end{aligned}$$

and the decision problem of the elected politician,

$$\max_{\tau,q,\iota,x} \sum_{t=0}^{\infty} \delta^t \nu(x_t)$$

$$\begin{aligned} s.t.\ x_t + \iota_t b_t &\leq \tau_{k,t} r_t K_t + \tau_{l,t} \omega_t L_t + q_{t+1} b_{t+1}, \\ x_t &\geq 0,\ \forall\ t. \end{aligned}$$

Here, the authors consider a finite continuum of households, $h \in \mathcal{H}$, and model labor supply separately from the capital. The function v describes the preferences with respect to free time of the household and is assumed to satisfy the Inada conditions and to be strictly convex and continuous. As usual, $k_{h,t}$ denotes the capital stock of a household h at time t, $c_{h,t}$ the consumption demand, and $l_{h,t}$ the labor supply. The variable b_t denotes the government debt, $\iota \in \{0,1\}$ the debt default decision, and $b_{h,t}$ the government bond holding of the respective household at time t.
In contrast to that, Barro (1990) considers *'public services as an input to private production'* (Barro, 1990, p.106), hence he adapts the production function itself. His model is continuous in time, and given as

$$\max_{c,k} \int_0^{\infty} u(c) e^{-\beta t} dt$$

$$\begin{aligned} s.t.\ k_t &= f(k(t), g(t)) - c(t), \\ g(t) &= \tau f(k(t), g(t)), \end{aligned}$$

where the production function f depends on the capital stock k and the public service g. The additional constraint states that this governmental spending is financed via taxations with taxation rate τ.

Amongst others, Brida and Accinelli (2007) consider a population growth in the Ramsey economy. They extend the ODE constraint in the continuous Ramsey model by an additional ODE, describing a logistic population growth law.

In economic growth theory, it was recognized at an early stage that not only savings, population development, or taxation, but also *technological change* over time drives economic growth. Especially in the Ramsey model, where agents try to maximize their welfare or profit, it would be convenient to link technological progress to intentional, endogenous determined investment decisions. However, most models insinuate an exogenous, even constant technology growth rate (see AK models in Acemoglu, 2009, Chapter 8). Romer (1986) *'started the endogenous growth literature'* (Acemoglu, 2009, p.387). In his model, he equalizes technological progress with knowledge accumulation, where the latter is defined as a by-product of capital accumulation. He models so called *spillover effects*, which work through human capital. The production function in this model is given as

$$Y(t) = F(K(t); A(t)L(t)),$$

with

$$A(t) = BK(t),$$

which means that the stock of technology is proportional to the capital stock. Acemoglu (2009, p.399) interprets this assumption as 'learning-by-doing', since investments increase productivity in this framework.

All in all, there exists a vast literature on modified Ramsey models. We picked out some few applications to make clear that the spatialized Ramsey models, or at least models that consider a continuum of agents in the consuming sector and models with dynamic technological progress, have been studied for a long time. Thus, our research on a new capital accumulation equation in the spatialized Ramsey model, that takes endogenous technological progress into account, is in line with former and on-going research in economic growth theory.

CHAPTER 3

The Spatial Economics and Optimal Control Theory

'Why are some countries so rich and others so poor? Does it have to be this way?' (Weil, 2013, p.21)

Various neoclassical growth models, based on the fundamental works of Ramsey (1928) or Solow (1956), are spatially homogeneous, which means they do not consider any motion of capital or labor across space, but only through time. However, observations of differences regarding population development, income distribution, and even education or technological deployment across space signalize that there is a gap between the economic growth theory and these observations. This could explain, why we observe a growing interest in the so called *Geographical Economics* over the last few years. Starting with the monograph *Geography and Trade* by Krugman (1991), this branch of economics has developed rapidly in the last three decades. Here, the models do not only consider changes in production factors over time, but also across space. These *new economic geography models* are general equilibrium models that tend to explain consumption, price formation, and production in the whole economy and they are based on assumptions considering market structure and mobility of production factors (Camacho et al., 2008, p.1). The approach of Krugman is based mainly on agglomeration effects which rely on increasing returns and transportation costs (c. Krugman, 2011, p.3). We do not consider such an approach in this thesis but stay in the neoclassical context. However, we follow the new geographical economy by modeling the spatial agglomeration effects by spillover effects and the autonomous mobility of production factors through space.

In this chapter, an introduction to the spatial economic growth theory is provided. We begin in Section 3.1 with a broad literature overview of spatially heterogeneous

Ramsey models. Since the resulting equilibrium problems in continuous time are optimal control problems with PDE constraints, we give some basic information on (parabolic) PDEs in Section 3.2. Afterwards, we introduce the local spatial Ramsey model by Brito (2001) in Section 3.3. The main part of this chapter deals with the derivation of the nonlocal spatial Ramsey model with endogenous productivity growth in Section 3.4.

3.1 The Spatial Growth Models

Since the Geographical Economics have become increasingly important in recent times, also the neoclassical growth theory has been spatialized. There are two hegemonial approaches on how to include the spatial component in a neoclassical growth model: The first one is a discrete extension of the model as seen in Bala and Sorger (2001), or Fujita et al. (1999). The other one is to embed a continuous space dependence into the considered model, as mainly done by Brito (2001, 2004), Boucekkine et al. (2009), and Camacho et al. (2008) in the earlier past.

The model introduced by Bala and Sorger (2001) is based on a discrete set of agents indexed by time and by family. Here, the direct neighborhood of agents is essential for the growth rate of the economy. Stationary equilibria are studied in a framework where productivity and investment are driven by local spillover effects of human capital and global market participation.
Pavilos and Wang (1996) model the space dimension via transportation costs for production factors and knowledge. Hence, their model stays only time dependent, however, a dependence of economic development on spatial distances is no longer denied.
We also note the modifications of the Ramsey model where a continuum of heterogenous households is considered. As already mentioned in Chapter 2.3, such models are for example described by Becker (1980), Becker and Foias (1987), Becker et al. (1991), and Sorger (2002). The heterogeneity in the preference in the consuming sector can be interpreted as a first step towards spatial extensions of the growth model.

To our knowledge, Brito (2001) was the first who introduced a continuous spatial extension of the established Ramsey-Cass-Koopmans model of optimal capital accumulation. In this framework, the spatial capital dynamics are driven by differences in access to productivity factors and economic variables. Camacho et al. (2008, p.1) point out that the *'alternative of a continuous space structure fits better modern economics, since this structure implies that all locations have access to goods'*. The process of optimal time-space accumulation of capital in the Ramsey model with a central planner is described by an optimal control problem with a semilinear parabolic partial differential equation over a bounded spatial domain and an infinite time horizon. As shown in Boucekkine et al. (2009), applying Pontryargin's maximum principle may lead to a dynamic system of PDE's that

is ill-posed in the sense of Hadamard under certain conditions. Boucekkine et al. (2009) study a Benthamian Ramsey model with a spatial discounting and linear utility functions of the households to circumvent this problem, whereas Camacho et al. (2008) consider an analogue optimal growth model but with a finite time horizon. Desmet and Rossi-Hansberg (2009) randomize the spatial growth model. They assume that realizations of local innovations are random and in this way extend the spatial model a bit further.
A more general mathematical approach to spatial growth theory can be found in the paper of Brock et al. (2014). They make use of operator theory in order to derive local and global equilibria in a Ramsey type capital accumulation problem with geographical spillovers. They are - to our knowledge - the first who take nonlocal spillover effects in technology into account, but still deny any capital agglomeration across space independent of these spillover effects.
Boucekkine et al. (2013), as well as Aldashev et al. (2014), develop the spatial optimal growth model with continuous space and time for an AK growth model in which the production technology is linear in capital, returns on capital are constant, and where technology changes by a constant and given factor over time. For further information on AK-models, see amongst others Acemoglu (2009, pp.388-408). As stated by Brito (2001), the resulting partial differential equations are (parabolic) diffusion equations. Aldashev et al. (2014) are able to generalize the model in Boucekkine et al. (2013) concerning the objective function of the social planner.

Although there already exists an extensive collection of spatial Ramsey models, taking many different parameters of economic growth into account, a rigorous mathematical analysis of the existence of solutions of the capital accumulation equations or the equilibrium problem is missing in most of the cases, or the models are restricted to quite simple modifications in order to fit the theory of classical solutions of differential equations. We will analyze the nonlocal spatial Ramsey model, that will be introduced below, in a more general mathematical context, namely in the weak solution theory. We first give a short introduction to weak solutions of partial differential equations in the next section.

3.2 On Partial Differential Equations, Sobolev Spaces and Embedding Theorems

This section aims to provide a compact overview of the theory of parabolic differential equations, the concept of weak solutions, and the corresponding function spaces. Essential is the notion of weak derivatives and embeddings between the considered function spaces, which will be of great importance when it comes to existence results of weak solutions of parabolic differential equations, or optimal controls.

Throughout this thesis, $\Omega \subseteq \mathbb{R}^n$ denotes an open and connected set, a so called *domain*. Whether it is also bounded or not, is stated explicitly or made clear in the context. We always consider a finite time horizon $T \in \mathbb{N}$, although Ramsey himself defined his model with an infinite time horizon. This infinite time horizon may be meaningful in economic terms, however it makes a mathematical analysis of the economic model even more complicated. We will later on introduce a suitable terminal condition in order to be on one hand able to capture the infinite time horizon character of the original model and, on the other hand, to make the model tractable.
The partial differential equations, which we consider here, are semilinear, parabolic, and nonlocal. Most practical applications lead to semilinear, or even nonlinear differential equations, however, a theoretical treatment is quite complicated. Thus, we have to make some rather strong assumptions on the appearing nonlinearities, such as Lipschitz continuity, or uniformly boundedness in order to prove existence or to state some regularity results of the weak solutions. Nevertheless, we are able to keep the Ramsey model, that we will introduce in this thesis, as universal as possible.

In general, a partial differential equation (PDE) is an equation involving an unknown function and some of its partial derivatives. An expression of the form

$$F(D^m u(x), D^{m-1}u(x), ..., Du(x), u(x), x) = 0 \qquad (x \in \Omega), \tag{3.1}$$

is called a m^{th}-order partial differential equation, where

$$F : \mathbb{R}^{n^m} \times \mathbb{R}^{n^{m-1}} \times \ldots \times \mathbb{R}^n \times \mathbb{R} \times \Omega \to \mathbb{R}$$

is a given functional and $u : \Omega \to \mathbb{R}$ is the unknown function (compare for example Evans (1997, p.1)). Here, $D^m u(x)$ denotes the m^{th} total derivative of u, $m \in \mathbb{N}$.

There are several types of differential equations, depending on how the derivatives appear in it. As already said, we will only focus on *semilinear* PDE, where all but the highest order derivative may appear nonlinearly. For a more detailed definition see for example Evans (1997, p.2). Whenever a differential equation can be written as

$$u_t = F(x, t, u, u_{x_i}, u_{x_i x_j}) \text{ on } \Omega \times (t_0, T),$$

with F elliptic and $0 \leq t_0 < T$, this PDE is called *parabolic* (Jost, 2013, p.5).

A function u *solves* the PDE, if it satisfies the equation (3.1). However, when it comes to the solution of partial differential equations, it is not that clear how a ‘solution’ of such a PDE should look like. For sure, it should fulfill some smoothness assumptions such that the equation (3.1) makes sense at all. A solution of (3.1), which is at least m times continuously differentiable, is called a *classical* solution. But, in most of the cases, the PDEs that are considered in applications cannot be solved in the classical sense. That is why there is a need for a new, *general-*

ized or *weak* notion of solutions in the context of PDEs (cf. Evans, 1997, p.8-9). Appropriate function spaces, that may contain such weak solutions of PDEs, are the so called *Sobolev spaces*, which contain less smooth functions and moreover, do have nice structure. Before we can define a Sobolev space, it is necessary to define the so called *weak derivatives* as below. The function spaces, which are needed in the following, are the spaces of continuous and continuously differentiable functions and Lebesgue spaces. According to Wloka (1982, p.12), we denote the space of continuous and bounded functions $\varphi : \Omega \to \mathbb{R}$ on Ω by $\mathcal{C}(\Omega)$. The space of functions which have bounded derivatives up to order $m \in \mathbb{N}$ on Ω is denoted by $\mathcal{C}^m(\Omega)$. Endowed with the norm

$$\|\varphi\|_{\mathcal{C}} := \sup_{x\in\Omega} |\varphi(x)|,$$

the space $(\mathcal{C}(\Omega), \|\cdot\|_{\mathcal{C}})$ is a Banach space. If $\mathcal{C}^m(\Omega)$ is endowed with the norm

$$\|\varphi\|_{\mathcal{C}^m} = \sup_{\substack{|\alpha|\leq m \\ x\in\Omega}} |D^\alpha \varphi(x)|,$$

it is a Banach space as well. Here, the notation $D^\alpha\varphi$ denotes the partial derivative of φ with respect to the multiindex α. We define according to Wloka (1982, p.12)

$$|\alpha| := \sum_{i=1}^{n} \alpha_i \leq m.$$

Although the notation $\mathcal{C}_b^m$ is common for bounded and continuous functions, we will follow the notation in Wloka throughout this monograph and suppress the b. The term $\mathcal{C}_0^\infty(\Omega)$ denotes the set of infinitely often differentiable functions $\varphi : \Omega \to \mathbb{R}$, with compact support in Ω. Endowed with the appropriate topology, this space will be referred to as the space of *test functions* later on.
The space of functions $\varphi \in \mathcal{C}^m(\Omega)$ for which $D^\alpha\varphi$ is not only bounded, but also uniformly continuous on Ω for $0 \leq |\alpha| \leq m$, is denoted according to Adams and Fournier (2003, p.10) as $\mathcal{C}^m(\overline{\Omega})$. Note, that this notation may be misleading in the case where Ω is unbounded. For example it is $\mathcal{C}^m(\overline{\mathbb{R}^n}) \neq \mathcal{C}^m(\mathbb{R}^n)$ although $\overline{\mathbb{R}^n} = \mathbb{R}^n$ (Adams and Fournier, 2003, p.10). However, especially for $m = 0$ and on bounded domains, the notation is uncomplicated (cf. Adams and Fournier, 2003, 1.30, p.11). Since $\mathcal{C}^m(\overline{\Omega})$ is a closed subspace of $\mathcal{C}^m(\Omega)$, it is a Banach space endowed with the same norm as $\mathcal{C}^m(\Omega)$.
Another important function space is the space of Hölder continuous functions with exponent λ, $\mathcal{C}^{m,\lambda}(\Omega)$. This space consists of all functions $\varphi \in \mathcal{C}^m(\overline{\Omega})$ whose partial derivatives of order m satisfy the Hölder condition of exponent λ in Ω, hence for which there exists a constant K such that

$$|D^\alpha\varphi(x) - D^\alpha\varphi(y)| \leq K|x-y|^\lambda, \quad x, y \in \Omega.$$

Provided the norm

$$\|\varphi\|_{\mathcal{C}^{m,\lambda}(\Omega)} := \|\varphi\|_{\mathcal{C}^m(\Omega)} + \max_{0\leq|\alpha|\leq m} [D^\alpha \varphi]_{\mathcal{C}^{0,\lambda}(\Omega)},$$

where $[\cdot]_{\mathcal{C}^{0,\lambda}(\Omega)}$ denotes the λ-th Hölder seminorm

$$[D^\alpha \varphi]_{\mathcal{C}^{0,\lambda}(\Omega)} := \sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{|D^\alpha\varphi(x) - D^\alpha\varphi(y)|}{|x-y|^\lambda},$$

$(\mathcal{C}^{m,\lambda}(\Omega), \|\cdot\|_{\mathcal{C}^{m,\lambda}(\Omega)})$ is a Banach space (Evans, 1997, pp.240).

These spaces of well-behaved functions are usually no suitable function spaces for PDE theory due to their strong requirements on differentiability. Moreover, they are only Banach spaces and do not have any Hilbert space property. Adams and Fournier (2003, p.23-24) give the following definition of Lebesgue spaces spaces that contain less smooth functions, but do have the required structural properties:

Definition 3.1 (Lebesgue Spaces)**:**
Let Ω be a domain in $\mathbb{R}^n$. For $1 \leq p < \infty$, let

$$\mathcal{L}^p(\Omega) := \{u : \Omega \to \mathbb{R} \text{ measurable} : \int_\Omega |u(x)|^p dy \ < \infty\}$$

denote the space of all p-Lebesgue measurable functions on Ω. Moreover, define

$$\mathcal{N}(\Omega) := \{g : \Omega \to \mathbb{R} : g = 0 \text{ a.e.}\} \subset \mathcal{L}^p.$$

Then, the Lebesgue space $L^p(\Omega)$ consists of all (equivalence classes of) p-Lebesgue measurable functions,

$$L^p(\Omega) := \mathcal{L}^p(\Omega)/\mathcal{N}(\Omega).$$

In the case $p = \infty$, $\mathcal{L}^\infty(\Omega)$ captures all essentially bounded functions u with

$$ess \sup_{x\in\Omega} |u(x)| < \infty.$$

The space $L^\infty(\Omega)$ is defined as $\mathcal{L}^\infty(\Omega)/\mathcal{N}(\Omega)$.
The space $L^p_{loc}(\Omega)$ denotes the space of all p-locally integrable functions on Ω, hence

$$L^p_{loc}(\Omega) = \{u : \Omega \to \mathbb{R} \text{ measurable} : \ u|_K \in L^p(K) \ \forall \ K \Subset \Omega \text{ compact}\}.$$

In the following, we lways identify a function $u \in L^p(\Omega)$ as the whole equivalence class. A very important property of the Lebesgue spaces is their completeness:

Corollary 3.2:
Endowed with the norm

$$\|u\|_{L^p(\Omega)} := \left(\int_\Omega |u(x)|^p dx \right)^{\frac{1}{p}}, \ 1 \le p < \infty,$$

respectively

$$\|u\|_{L^\infty(\Omega)} := ess \sup_{x \in \Omega} |u(x)|, \ p = \infty,$$

$L^p(\Omega)$ *is a Banach space for* $1 \le p \le \infty$.

The proof can be found for example in Adams and Fournier (2003, p.29). Especially for $p = 2$, $L^2(\Omega)$ is a separable Hilbert space with inner product

$$\langle u, v \rangle = \int_\Omega u(x)v(x)dx,$$

which is relevant in the following.

We define the space of the so called *test functions*, which is of great importance when it comes to the weak formulation of a PDE, according to Adams and Fournier (2003, p.20) in distributional sense:

Definition 3.3 (Space of test functions)**:**
Let Ω *be a domain. A sequence* $\{\varphi_j\}$ *of functions in* $\mathcal{C}_0^\infty(\Omega)$ *is said to* converge in the sense of $\mathcal{D}(\Omega)$ *to a function* $\varphi \in \mathcal{C}_0^\infty(\Omega)$ *provided the following conditions are satisfied,*

- *There exists* $K \Subset \Omega$ *such that* $supp(\varphi_j - \varphi) \subset K$ *for all* j.
- $lim_{j \to \infty} D^\alpha \varphi_j(x) = D^\alpha \varphi(x)$ *uniformly on* K *for each multi-index* α.

There exists a finest locally convex topology $\mathcal{T}$ *on* $\mathcal{C}_0^\infty(\Omega)$ *with respect to which a linear functional* T *is continuous if*

$$\varphi_j \to \varphi \text{ (in the sense of } \mathcal{D}(\Omega)) \ \Rightarrow T(\varphi_j) \to \ T(\varphi) \text{ in } \mathbb{R}.$$

The locally convex topological vector space $(\mathcal{C}_0^\infty(\Omega), \mathcal{T})$ *defines the space of test functions, which will be denoted by* $\mathcal{D}(\Omega)$.

Remark 3.4:
Whenever we use the notion $\mathcal{C}_0^\infty$, *this has to be understood in distributional sense. This is convenient since we did not further define the topology* $\mathcal{T}$.

Definition 3.5 (Distributions)**:**
A linear map $T : \mathcal{D}(\Omega) \to \mathbb{R}$ *is called a distribution on* Ω, *if for all sequences* $\varphi_i \to 0$ *(in the sense of* $\mathcal{D}(\Omega)$*) it holds* $T(\varphi_i) \to 0$ *(in* $\mathbb{R}$*). The space of distributions is denoted by* $\mathcal{D}'(\Omega)$.

An analog definition is stated as Theorem 1.4 in Wloka (1982, p.21). An example of a distribution is the following integral operator: Corresponding to every function $u \in L^1_{loc}(\Omega)$ there is a distribution $T_u \in \mathcal{D}'(\Omega)$ defined by

$$T_u(\varphi) = \int_\Omega u(x)\varphi(x)dx, \quad \varphi \in \mathcal{D}(\Omega).$$

The mapping $u \mapsto T_u$ is injective (Wloka, 1982, p.21), so we can identify the distribution T_u with the function u. With this example in mind, we can state the definition of a weak derivative.

Definition 3.6:
Let Ω be a domain in $\mathbb{R}^n$ and suppose that $u,\ v_\alpha \in L^1_{loc}(\Omega)$. The function v_α is called the α^{th}- weak partial derivative of u, *in notation $D^\alpha u = v_\alpha$, if $T_{v_\alpha} = D^\alpha T_u$ in $\mathcal{D}'(\Omega)$, or*

$$\int_\Omega u D^\alpha \varphi dx = (-1)^{|\alpha|} \int_\Omega v_\alpha \varphi dx, \tag{3.2}$$

for all test functions $\varphi \in \mathcal{D}(\Omega)$.

It should be noticed that if a weak derivative of u is considered, u does not have to be an element in $\mathcal{C}^m(\Omega)$ with $m = |\alpha|$. There only has to exist a locally summable function v_α for which the formula (3.2) is valid (Evans, 1997, p.242). If such a function v_α exists, it is uniquely defined up to a set of measure zero (Evans, 1997, p.243). Since the m-th weak derivative is equal to the classical derivative whenever u is m-times differentiable in the classical sense, it is convenient to use the same symbol for both kinds of derivatives. For further details see for example Theorem 1.7. by Wloka (1982, p.26).

In the following, the parameters $1 \leq p \leq \infty$ and $m \in \mathbb{N}_0$ are fixed. Now, we finally are able to define the function spaces named after Sergei Sobolev. These *'Sobolev spaces are vector spaces whose elements are functions defined on domains in n-dimensional Euclidean space $\mathbb{R}^n$ and whose partial derivatives satisfy certain integrability conditions'* (Adams and Fournier, 2003, p.1). In mathematical terms, we define the following:

Definition 3.7:
Let $\Omega \subseteq \mathbb{R}^n$ be a domain. The Sobolev space *$W^{m,p}(\Omega)$ consists of all (locally) summable functions $u : \Omega \to \mathbb{R}$ such that for each multiindex α with $|\alpha| \leq m$, $D^\alpha u$ exists in the weak sense and belongs to the Lebesgue space $L^p(\Omega)$, hence*

$$W^{m,p}(\Omega) := \{u \in L^p(\Omega) : \ D^\alpha u \in L^p(\Omega),\ 0 \leq |\alpha| \leq m\}.$$

The following theorem states the most important properties of Sobolev spaces, which make the mathematical treatment of weak solutions possible:

Theorem 3.8:
For $m \in \mathbb{N}$ and $1 \leq p \leq \infty$, the Sobolev space $W^{m,p}(\Omega)$, endowed with the norm

$$\|u\|_{W^{m,p}(\Omega)} := \begin{cases} \left(\Sigma_{|\alpha|\leq m} \int_\Omega |D^\alpha u|^p dx\right)^{\frac{1}{p}} & (1 \leq p < \infty) \\ \Sigma_{|\alpha|\leq m} ess\sup_\Omega |D^\alpha u| & (p = \infty) \end{cases}, \tag{3.3}$$

is a Banach space. If $p = 2$, $W^{m,2}(\Omega)$ is a separable Hilbert space, hence it is often denoted by $H^m(\Omega)$.

The proof is amongst others given by Evans (1997, p.249) or Wloka (1982, p.69).

Remark 3.9:
The Sobolev spaces are closely related to the space of test functions. The space $W_0^{m,p}(\Omega)$ is defined as the closure of $\mathcal{C}_0^\infty(\Omega)$ in $W^{m,p}(\Omega)$. In the context of PDE, also the dual spaces of the considered function spaces are of great importance. The dual space of $H^1(\Omega)$ is denoted by $H^{-1}(\Omega)$.

In the following part of this section, we will introduce some important embedding theorems, that will play an important role for the existence theory of weak solutions and optimal controls of the spatial Ramsey model the following chapters. We will mainly follow the books of Adams and Fournier (2003) and Wloka (1982). We start with the following definition as taken from Adams and Fournier (2003, p.68).

Definition 3.10:
A domain Ω satisfies the segment condition *if every $x \in \partial\Omega$ has a neighborhood U_x and a nonzero vector y_x such that if $z \in \overline{\Omega} \cap U_x$, then $z + \lambda y_x \in \Omega$ for $0 < \lambda < 1$.*

Theorem 3.11:
If Ω satisfies the segment condition, then the set of restrictions to Ω of functions in $C_0^\infty(\mathbb{R}^n)$ is dense in $W^{m,p}(\Omega)$ for $1 \leq p < \infty$.

The result is taken from Adams and Fournier (2003, p.68), and the proof can be found on the following page of this book. A very useful connection exists between the Sobolev spaces and Lebesgue spaces as defined above. Before we can state the respective theorem, we have to give further properties of a domain in $\mathbb{R}^n$.

Definition 3.12:
$\Omega \subseteq \mathbb{R}^n$ satisfies the so called cone condition*, if there exists a finite cone C, such that each $x \in \Omega$ is the vertex of a finite cone C_x contained in Ω, and congruent to C.*

The following *Sobolev embedding theorem* enables us to state some fundamental results on the existence of weak solutions and optimal controls:

Theorem 3.13 (The Sobolev Embedding Theorem)**:**
Suppose that $\Omega \subseteq \mathbb{R}^n$ *satisfies the cone condition, let* $1 \leq p \leq q \leq \infty$ *and* $1 \leq k \leq n$*. If either* $mp \geq n$ *or* $m = n$ *and* $p = 1$*, then*

$$W^{m,p}(\Omega) \hookrightarrow L^q(\Omega).$$

If $mp < n$ *and either* $n - mp < k \leq n$ *or* $p = 1$ *and* $n - m \leq k \leq n$*, then*

$$W^{m,p}(\Omega) \hookrightarrow L^q(\Omega)$$

for $p \leq q \leq p^* = np/(n - mp)$ *(Adams and Fournier, 2003, p. 85).*
Here, $\hookrightarrow$ *denotes a continuous embedding,* $\hookrightarrow^c$ *will stand for a compact embedding.*

Other important connections which exist between the spaces of continuous functions are shown by Adams and Fournier (2003, pp.11-12).

Theorem 3.14:
Let m *be a nonnegative integer and let* $0 < \nu < \lambda \leq 1$*. Then the following embeddings exist*

$$\begin{aligned} \mathcal{C}^{m+1}(\overline{\Omega}) &\hookrightarrow \mathcal{C}^{m}(\overline{\Omega}), \\ \mathcal{C}^{m,\nu}(\Omega) &\hookrightarrow \mathcal{C}^{m}(\overline{\Omega}), \\ \mathcal{C}^{m,\lambda}(\Omega) &\hookrightarrow \mathcal{C}^{m,\nu}(\Omega). \end{aligned}$$

If Ω *is bounded, the last two embeddings are compact.*

Remark 3.15:
For a bounded Ω *and the case that* $m = 0$*, a function* $\varphi \in \mathcal{C}^0(\overline{\Omega})$ *has to be understood as the unique continuous extension of* φ *to the closure of* Ω*. In particular that means, that the space of Hölder continuous functions with exponent* ν *is compactly embedded into the space of continuous functions (see amongst others 1.30 in Adams and Fournier (2003, p.11)).*

A term, that describes the interaction between the function spaces involved in PDE theory, and which is of great importance when it comes to the existence of weak solutions, is a so called *Gelfand triple*. According to Wloka (1982, p.253), a reflexive Banach space V and a Hilbert space H such that V is continuously, injectively, and densely embedded in H form a Gelfand triple

$$V \hookrightarrow H \hookrightarrow V'.$$

Lemma 3.16:
Let $\Omega \subseteq \mathbb{R}^n$ *satisfy the cone condition. The Sobolev space* $H^1(\Omega)$ *and the Lebesgue*

space $L^2(\Omega)$ *form a Gelfand triple, hence the embeddings*

$$H^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$$

are dense and continuous.

Proof. The proof can directly be derived from Theorem 3.13 and the fact that every Hilbert space is reflexive. □

The dynamics of the spatial Ramsey model do not only depend on space, but also on time. Hence, the capital accumulation equation has to be solved in the space-time cylinder. Thus, when it comes to parabolic differential equations, an appropriate function space has to be defined. We follow the general definition of Wloka (1982, pp.378-381).

Definition 3.17:
Let V, H *be two separable Hilbert spaces such that* $V \hookrightarrow H \hookrightarrow V'$ *built a Gelfand triple. Let* $0 < T < \infty$ *be given. For* $p \in \mathbb{N}$, *we define the space*

$$W^{1,p}(0,T;V) := \{f \in L^p(0,T;V) : \frac{\partial f}{\partial t} \in L^p(0,T;V')\} \tag{3.4}$$

where for $p \geq 1$ *the Lebesgue space of functions with values in a Hilbert space is defined as* $L^p(K;V) := \{f : K \to V : f \text{ weakly measurable and } \int_K \|f(s)\|_V^p ds < \infty\}$.

Remark that $L^p(K;V)$ is a Banach space for all $p \geq 1$. Moreover, in the case of $p = 2$, $L^2(K;V)$ is a Hilbert space endowed with the scalar product

$$\langle x, y \rangle_{L^2(K;V)} := \int_K \langle x(s), y(s) \rangle_V ds.$$

In the following, we will shorten the expression for $W^{1,2}(0,T;V)$ and will denote it by $W(0,T)$, where V is always made clear in the context. The next theorem points out the nice structure of the considered function space.

Theorem 3.18:
Let V, H *be two separable Hilbert spaces such that* $V \hookrightarrow H \hookrightarrow V'$ *built a Gelfand triple. Let* $0 < T < \infty$ *be given.*

(a) Endowed with the scalar product

$$\langle f, g \rangle_{L^2(0,T;V)} := \int_0^T \langle f(t), g(t) \rangle_V dt,$$

$L^2(0,T;V)$ *is a Hilbert space.*

(b) *Endowed with the scalar product*

$$\langle f,g\rangle_W := \int_0^T \langle f(t),g(t)\rangle_V dt + \int_0^T \langle \frac{\partial f(t)}{\partial t}, \frac{\partial g(t)}{\partial t}\rangle_{V'} dt$$

$W(0,T)$ *is a Hilbert space.*

(c) *All functions in* $W(0,T)$ *are continuous with values in* H*, hence* $f \in W(0,T)$ *implies that*

$$f : [0,T] \to H \text{ is continuous}$$

(cf. Wloka, 1982, pp.380-382). In particular, (c) implies that there exists a constant c_E such that

$$\|u\|_{\mathcal{C}([0,T];H)} \leq c_E\|u\|_{W(0,T)}.$$

Throughout this thesis, we will only consider domains $\Omega \subseteq \mathbb{R}^n$ which satisfy the segment and cone condition. Hence, we can revert to the Gelfand triple and embedding theorems.

3.3 The Local Ramsey Model - Optimal Control with PDE Constraint

In this section, we introduce the spatial Ramsey-Cass-Koopmans model as first modeled by Brito (2001) and analysed, respectively modified in the following years mainly by Brito (2004, 2012) himself, Boucekkine et al. (2009, 2013), and Camacho et al. (2008). As the name hypothesizes, the main difference between the common Ramsey model and its spatial version is that in the latter, the capital accumulation is a process depending not only on time but also on space. In the spatial Ramsey models, the households do not only have the possibility to shift capital towards future dates in time by saving, but there are also capital flows among different locations or regions allowed. The spatial domain, which we denote by $\Omega \subseteq \mathbb{R}^n$ in the following, may either be interpreted as heterogeneity in the continuum of households if n is equal to one, or as geographical space in case n is two. Ramsey himself considered an infinite time horizon. However, especially when the spatial domain is unbounded, a mathematical analysis of the spatial model with respect to existence and uniqueness of solutions is difficult. Nevertheless, most spatial Ramsey models also consider an infinite time horizon (see amongst others Brito, 2001, 2004, 2012, Brock and Xepapadeas, 2006, Brock et al., 2014, 2012, 2013, or Boucekkine et al., 2009, 2013). An exception is the model introduced by Camacho et al. (2008), who introduce a finite time horizon $T \in \mathbb{N}$ and define a terminal condition on the capital stock. We will point out the advantage of such a terminal condition later and only consider the infinite time horizon models in this section, hence $t \in \mathbb{R}_+$.

In the spatial Ramsey models, it is assumed that there exists a continuum of *'potentially heterogeneous and interacting households'* whose support is identified with the spatial domain (Brito, 2004, p.6). Moreover, these households should be evenly distributed over space. Every household is naturally endowed with labor and capital such that these production factors, or the capital-labor ratio, become space dependent. Moreover, also consumption and production depend on the respective location or the household. The common assumption of one homogeneous production good is transferred from the only time depending model. The capital-labor ratio and consumption are described by functions

$$\begin{aligned} k &: \Omega \times \mathbb{R}_+ \to \mathbb{R},\ (x,t) \mapsto k(x,t), \\ c &: \Omega \times \mathbb{R}_+ \to \mathbb{R},\ (x,t) \mapsto c(x,t). \end{aligned} \tag{3.5}$$

Here, $c(x,t)$ stands for the consumption and $k(x,t)$ for the respective capital stock of the households located in $x \in \Omega$ at time $t \in \mathbb{R}_+$.
The assumption of a homogeneous production leads to the definition of a production function p, which is a neoclassical production function and, as in the original only time depending model, depends on the capital-labor ratio only

$$p : \mathbb{R} \to \mathbb{R},\ k \mapsto p(k).$$

All variables are in terms per capita, which means that labor force is already normalized to one.

In the recent literature, some variants considering technology in the spatial Ramsey model can be found. In the model of Brito (2001, 2004), technology is assumed to be space independent. He defines a productivity factor $A \in \mathbb{R}_+$, which is multiplied with the production function and is given exogenously. Boucekkine et al. (2009) generalize this model. Here, the total factor productivity may be heterogeneous in space and time, but still is pre-determined as a function $A : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$. In contrast to these models listed above, where the production is described as the product of a productivity factor and the production function, Brock and Xepapadeas (2006) and Brock et al. (2012, 2013, 2014) consider spillover effects to endogenize productivity and to drive capital agglomeration. However, they do not consider any diffusion effects in the capital equation and therefore can be assigned to a different type of spatial Ramsey models.

A common assumption in spatial economic growth models is that *'households increase the scale of production by accumulating physical capital'* (Brito, 2001, p.2). The gross investment in location x at time t is modeled as the derivative of the capital function with respect to time, which should satisfy the following budget constraint

$$\frac{\partial k}{\partial t}(x,t) = Ap(k(x,t)) - c(x,t) - \tau(x,t), \qquad (x,t) \in \Omega \times \mathbb{R}_+. \tag{3.6}$$

The function $\tau : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ describes the trade balance which *'is equal to the net lending capacity'* (Brito, 2001, p.2). According to Brito (2001), assuming the aggregate economy to be closed yields

$$\int_\Omega \left(\frac{\partial k(x,t)}{\partial t} + c(x,t) - Af(k(x,t)) + \tau(x,t) \right) dx = 0, \quad \forall t. \tag{3.7}$$

Due to (3.6), the equation (3.7) is also true for subsets $\mathcal{O}$ of the spatial domain Ω

$$\int_\mathcal{O} \left(\frac{\partial k(x,t)}{\partial t} + c(x,t) - Af(k(x,t)) + \tau(x,t) \right) dx = 0, \quad \forall t.$$

The term

$$\int_\mathcal{O} \tau(x,t) dx$$

describes the net trade balance in a region $\mathcal{O}$ at time t. The central assumption in this local Ramsey model, as it will be called in the following, is that *'capital flows from regions with lower marginal productivity of capital to higher ones'* (Boucekkine et al., 2009, p.6). According to Boucekkine et al. (2009) and Brito (2004), this is equivalent to *'assuming that capital flows from regions which are abundant in capital towards regions which are relatively scarce'* (Boucekkine et al., 2009, p.4). Neglecting any institutional barriers to capital flows and adjustment speed, the trade balance is equal to the capital flow through $\mathcal{O}$, hence

$$\int_\mathcal{O} \tau(x,t) dx = - \int_\mathcal{O} \frac{\partial^2 k(x,t)}{\partial x^2} dx.$$

Brito (2001) also assumes that capital movements will eliminate all inter-regional arbitrage opportunities, which motivates

$$- \int_\mathcal{O} \frac{\partial^2 k(x,t)}{\partial x^2} dx = 0.$$

The aggregated budget constraint in the local spatial Ramsey model is then described by

$$\int_\mathcal{O} \left(\frac{\partial k(x,t)}{\partial t} - \frac{\partial^2 k(x,t)}{\partial x^2} - Af(k(x,t)) + c(x,t) \right) dx = 0, \quad \forall t.$$

For small regions $\mathcal{O}$, the accumulation process of capital in time and space is then given by the following semilinear parabolic differential equation

$$\frac{\partial k(x,t)}{\partial t} = \frac{\partial^2 k(x,t)}{\partial x^2} + Af(k(x,t)) - c(x,t) \quad \text{on } \Omega \times \mathbb{R}_+.$$

Moreover, the capital stock should satisfy an initial value condition, hence

$$k(x,0) = k_0(x) > 0 \quad \forall x \in \Omega.$$

Analog to the original only time depending model, the existence of a social central planner, who tries to maximize the households' lifetime utility benevolently, is also assumed in the spatial Ramsey models. In this ethical framework, the best action the central planner can choose is the one that maximizes the common welfare. For example, Brito (2001) assumes a Benthamian utility function, which is given as unweighted sum of the individual intertemporal utility functions for every household located in every $x \in \Omega$:

$$\int_{\mathbb{R}_+} \int_{\Omega} U(c(x,t)) e^{-\tau t} dx dt.$$

Such an objective is the most common functional to model a social central planner. The households discount future utility exponentially with a time discount factor $\tau \in (0,1)$. Although time is discounted, the integral over the space domain does not necessarily exist if Ω is not bounded. Brito (2004) introduces some potential remedies: For example, he considers not only a time discounting, but a space discounting as well, which leads to a *'symmetry between time and space, by penalizing dates and locations far away from the origin'* (Brito, 2004, p.13). However, according to Brito (2004) and Camacho et al. (2008), there may be no meaningful evidence for punishing distances from the origin in economic interpretation. Therefore, Brito (2004) considers an alternative in order to guarantee the existence of the objective integral. He introduces a so called *spatial averaging* where all locations $x \in \Omega$ are weighted by the inverse of their relative distance to the origin. Considering a so called *Millian intertemporal utility function*,

$$V(x,t) := \lim_{x \to \infty} \frac{1}{2x} \int_{-x}^{x} \int_{0}^{\infty} u(c(y,t)) e^{-\beta t} dt dy,$$

he can make sure that the integral will be bounded for the *'steady state spatially symmetric distributions of consumption'* (Brito, 2004, 14). However, this kind of objective functions is not of Benthamian type.
Boucekkine et al. (2009) point out that they are able to stay in the social central planner setting, even without any spatial discounting in the objective. However, they have to assume that the utility function is linear and introduce some free boundary conditions.

Solving the partial differential equation on bounded spatial domains requires the definition of such boundary conditions. Also in the spatially unbounded case $\Omega = \mathbb{R}$, Brito (2004), Boucekkine et al. (2009), and Camacho et al. (2008) claim that the capital flow vanishes far away from the origin, hence

$$\lim_{x \to \pm\infty} \frac{\partial k(x,t)}{\partial x} = 0, \qquad \forall t \in \mathbb{R}_+.$$

Combining the social central planner's objective, the budget constraint, and the capital flow constraint yields the following optimal control problem, which defines the local spatial Ramsey model:

$$\max_{k,c} \ \mathcal{J}(k,c) := \int_0^\infty \int_{\mathbb{R}} U(c(t,y)) e^{-\tau t - \gamma |x|} dt dy, \tag{3.8}$$

subject to

$$\begin{aligned} \frac{\partial k(x,t)}{\partial t} - \frac{\partial^2 k(x,t)}{\partial x^2} - Af(k(x,t)) + \delta k(x,t) = -c(x,t) \qquad & \text{on } \mathbb{R} \times \mathbb{R}_+ \\ \lim_{x \to \pm\infty} \frac{\partial k(x,t)}{\partial x} = 0, \qquad & \text{in } \mathbb{R}_+ \\ k(0,\cdot) = k_0(\cdot) > 0, \qquad & \text{in } \mathbb{R} \\ k, c \geq 0 \qquad & \text{on } \mathbb{R} \times \mathbb{R}_+ \end{aligned} \tag{3.9}$$

for discount factors $\tau, \gamma > 0$, a productivity factor A which may either be a constant in $\mathbb{R}_+$ or a nonnegative, real valued function and a depreciation rate $\delta > 0$.

As already mentioned, studying existence and uniqueness of the solution of the PDE constraint (3.9) and the optimal control problem (3.8)-(3.9) is a rather challenging task mostly due to the unbounded space and the inifinite time horizon. Indeed, most of the available existence results are stated for spatial models on bounded spatial domains. Aldashev et al. (2014) and Boucekkine et al. (2013) define the spatial domain as unit circle in $\mathbb{R}^n$. According to Boucekkine et al. (2013), *'the choice of the unit circle to represent space is not innocuous'*, however it is a traditional modeling of space in economics and it *'allows to avoid the specification of boundary conditions'* (Boucekkine et al., 2013, p.2, p.5).
Camacho et al. (2008, p.8) state an existence result for a classical solution of the PDE constraint on unbounded space but with finite time horizon. They derive the necessary first order conditions applying Pontryagin's maximum principle. However, the well-posedness of the resulting system of PDE is only guaranteed under some strong assumptions such as the boundedness of the consumption path and that the capital stock is positive on the whole time-space cylinder, including the terminal point in time.
Boucekkine et al. (2009, p.13) admit that their optimal control model in infinite time and unbounded space is ill-posed. However, they are able to derive an existence result for a linear utility function

$$\mathcal{J}(k,c) := \int_0^\infty \int_{\mathbb{R}} c(x,t)\psi(x)e^{-\tau t}dxdt$$

for a function $\psi : \mathbb{R} \to \mathbb{R}_+\backslash\{0\}$ with

$$\int_{\mathbb{R}} \psi(x)dx = 1.$$

It it remarkable, that most analysis considering the existence of a solution of the optimal control problem (3.8)-(3.9) are based on the theory of classical solutions. Especially for semilinear differential equations, this may not be appropriate. In the following, we consider weak solutions of the underlying PDE constraints. In that way, we are much freer in the choice of our nonlinearities and the shape of the objective function. Moreover, we can stay in the setting of a social central planner, which is one of the highlighted advantages stated in the paper of Camacho et al. (2008).

3.4 The Nonlocal Spatial Ramsey Model with Endogenous Productivity Growth

In the recent literature dealing with continuous spatial Ramsey models, the capital accumulation process via time and space is modeled as a parabolic differential equation. In the models introduced by Brito (2001, 2004, 2012), Camacho et al. (2008), and Boucekkine et al. (2009), the mobility of capital over space is described by a Laplace operator. In order to compute the derivatives at a point (x,t), the function k, that describes the capital distribution in time and space, has to be known only in the respective point and in an arbitrary small neighborhood. Such (local) diffusion equations play a mayor role in mathematical modeling, describing phenomena in physics, chemistry, or finance. A prime example is the heat equation, which models the motion of heat in a homogeneous and isotropic medium Ω,

$$u_t - \Delta u = 0 \qquad \text{on } \Omega,$$

where the function u describes the temperature.
Considering the mobility of capital and labor force across space, local diffusion effects are no longer sufficient to model the process of accumulation and diffusion of capital and labor close to reality. Coming back to the example of the heat equation, this becomes quite obvious: Temperature, as well as odor or color molecules in the air or in liquids, always need physical contact to the direct surrounding to move from one location to another. If we think about labor force, agents endowed with labor may move from one location to another without undertaking work in every single location they pass through on their way. The same behavior is observable for the dispersion of capital. Investments arise only on some separate locations, they do of course affect the surrounding but do not spread evenly from

one point to another. Capital, as well as labor force, can literally 'jump' through space.

So called *jump diffusion models* have mainly been studied in financial mathematics because they are able to capture sudden, discontinuous changes in asset prices. Other fields of application are physics (modeling heat diffusion over a crack in the isotropic medium), or biology (population and swarm simulation models). The nonlocal diffusion equation we will consider here arises naturally from a probabilistic process in which capital moves randomly in space, subject to a probability that allows long jumps. However, the model is not stochastic, but deterministic as we consider an economy with a central planner, who observes any spatial consumption distribution in all points of time, and can determine the capital distribution according to the capital accumulation process.
Another motivation for introducing those jump diffusion effects to the Ramsey model is that the heat equation insinuates an infinite adjustment speed of the molecules. The accumulation of capital or labor is much slower in real world observations. Moreover, the local diffusion operator enforces an even distribution of molecules or heat in the medium and does neither allow slow adjustment speed, nor the conservation of heterogeneity, or even discontinuities. Especially when modeling initial capital distributions that exhibit gaps or cracks, the models as introduced by Brito (2001) and others lead to even capital distributions and smooth out every disparities.
To our knowledge, we are the first who introduce such *nonlocal diffusion effects* in the spatial Ramsey model. We add an additional nonlocal diffusion operator, complementing the (local) Laplace operator. Thus, in our version of this economic growth model, capital mobility in a location does not only depend on the respective one but also on 'far away' locations.

We consider a domain of interest as a bouned or unbounded subset $\Omega \subseteq \mathbb{R}^n$. According to Brock et al. (2012), we can interpret Ω either geographically, which means as physical space and would motivate to choose $n = 2$, or as economic space, where the location has to be understood as a set of *'attributes related to economic quantities of interest'* (Brock et al., 2012, p.3). Analog to Brito (2004), we assume that the population or attributes are evenly distributed across space, which legitimates the assumption on Ω to be a connected set.
We consider a function $k : \Omega \times [0, T] \to \mathbb{R}$, which describes the capital stock distribution in time and space. Although the capital distribution is in this way heterogeneous in time and space, we only consider one single capital distribution function. This is in line with our procedure to consider only one single agent who makes decisions. We are aware of the criticism against this representative agent approach, see for example Kirman (1992) or Stiglitz (2018). However, our main interest in the context of this monograph lies in questions of existence and computability of economic growth in time and space and how nonlocal diffusion effects in the capital accumulation process and endogenous productivity growth influence

it. We recommend a research on decentralization and distributional issues on our nonlocal Ramsey model for future research.

For an only space dependent function $k : \mathbb{R}^n \to \mathbb{R}$, the nonlocal diffusion operator is given as an integral operator,

$$\mathcal{NL}(k)(x) := \int_{\mathbb{R}^n} (k(y) - k(x))\gamma(x, y)dy$$

(cf. Burch and Lehoucq, 2011; Du et al., 2014; Gunzburger and Lehoucq, 2010; D'Elia and Gunzburger, 2014). The function γ is a so called *kernel function* and throughout this thesis it is assumed to be a nonnegative and symmetric function. If we can rewrite γ as $\hat{\gamma} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},\ x, y \mapsto \hat{\gamma}(x - y)$, and if moreover, $\hat{\gamma}$ is a probability density such that

$$\int_{\mathbb{R}^n} \hat{\gamma}(s)ds = 1,$$

the rate of diffusion can be interpreted as *'the difference in the rate at which k enters x* [from all other locations], $\int_{\mathbb{R}^n} k(y,t)\gamma(x,y)dy$*, and the rate at which k departs x* [to all other locations], $k(x)$*'* (Burch and Lehoucq, 2011, p.32). Here, the value $\hat{\gamma}(x-y)$ is thought of as the probability of k jumping from y to x (Chasseigne et al., 2006, p.1). It is obvious by the definition, that this operator does have smoothing effects. Briani et al. (2004) indeed prove that, for an appropriate kernel function γ, the integral operator behaves like a weak Laplace operator, which is one of the reasons why discontinuities can be preserved. Moreover, this operator reduces the adjustment speed of capital in space, as compared to the local model. The kernel function, which we will consider in our nonlocal Ramsey model, is the Gaussian probability density function with a given variance or covariance matrix and a mean value x. We will analyze the Ramsey model on bounded and unbounded spatial domains. We define the kernel function, depending on Ω, as the multivariate Gaussian probability density function,

$$\Gamma_\nu(x, y) := \begin{cases} \frac{1}{\sqrt{(2\pi\nu^2)^n}} \exp\left(-\frac{1}{2}(x - y)^T \Sigma_\nu^{-1}(x - y)\right) & \Omega = \mathbb{R}^n, \\ \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{1}{2}(x - y)^T \Sigma_\sigma^{-1}(x - y)\right) \mathbb{1}_{\mathcal{B}_\nu(x)}(y) & \mathbb{R}^n \backslash \Omega \neq \emptyset, \end{cases} \tag{3.10}$$

where we choose $\nu \in \{\mu, \varepsilon\}$ for two constants $0 < \mu < \varepsilon$ and $\sigma > 0$. $\mathcal{B}_\nu(x)$ denotes the multidimensional ball with radius ν and center x.
When we consider a bounded spatial domain, it is appropriate to consider different parameters σ, which describes the covariances of the spatial directions, and ν, which denotes the interaction radius. Both parameters affect the diffusive effect of the nonlocal operator $\mathcal{NL}$, but contrariwise. We will point out in Chapter 6, that an increasing interaction radius ν increases the diffusive effect, whereas a smaller σ drives diffusion.

The matrices $\Sigma_\nu \in \mathbb{R}^{n\times n}$ and $\Sigma_\sigma \in \mathbb{R}^{n\times n}$ are covariance matrices, hence positive semidefinite and symmetric, with determinants $\det(\Sigma_\nu) = \nu^{2n}$ and $\det(\Sigma_\sigma) = \sigma^{2n}$. We can control the interdependency of the spatial directions with the entries next to the diagonals. Whenever the covariance matrix has diagonal structure, this means that capital can move through space completely uncorrelated. Moreover, the entries on the diagonal reflect the central planners priorities with respect to the space directions.
The choice of this density function is application driven. First of all, we are in this case in the setting of Briani et al. (2004) such that we can motivate the assumption of weak diffusion effects to drive capital accumulation in space. Second, the Gaussian density function has a special shape, weighting points near the expected value x higher than points far away, and moreover is rapidly decreasing towards the edges. In economic terms, this means that the probability of capital and labor force jumping to 'near by' areas is higher than moving suddenly to far away locations. By varying the value of the parameter ν, we can make the area where capital or labor movements are more likely, bigger or smaller. This characteristic is in line with the assumption of a central planner, who can decide in what areas jumps of production factors are more likely or appropriate.

A groundbreaking innovation of the Ramsey model is the endogenous saving rate, which means that the optimal saving rate, that maximizes the welfare of the economy, is determined via the households' lifetime maximization intention during the optimization process within the model itself. In that point, the Ramsey model differs from many other neoclassical growth models. As already mentioned, economic growth is also driven by technological progress, or the increase of productivity, which can both be modeled by so called *spillover effects*. In the common (local) Ramsey model, this productivity growth is assumed to be growing at a constant rate A (compare equation (3.6)). In our opinion, this exogenously pre-defined productivity growth rate sets the endogenous character of the Ramsey model aside. We introduce a new, nonlocal productivity operator P, that aims to endogenize the process of productivity growth, and in that way, preserves the self-contained character of the Ramsey model. We assume that there is a correlation between the development (meaning an increase) of productivity and the state of the system, namely the capital stock in a surrounding of a respective location. Moreover, we assume that productivity naturally increases over time. We model the productivity growth as integral term as well. Inspired by a paper of Olson Jr. et al. (2000), we assume that the productivity growth is exponentially in time in all locations, depending on an initial level of productivity in the respective location and on the distribution of capital in space. Combining all intentions, we define the *nonlocal productivity growth operator* at a location x and at time t as

$$P(k)(x,t) := A_0(x) \exp\left(\frac{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\varepsilon(x,y)dy + \xi} t \right), \tag{3.11}$$

with a continuous and nonnegative function $\phi : \mathbb{R} \to \mathbb{R}_+$ and $\xi > 0$. Examples for

ϕ we have in mind are the absolute value function or a differentiable approximation. We will refert to ϕ as the nominal function and assume throughout that the integrals in (3.11) are welldefined. Following Olson Jr. et al. (2000), A_0 denotes a given initial productivity distribution that is compounded over time. In contrast to a simple AK-model as studied in Boucekkine et al. (2013), we consider a nonlinear production function that is multiplied with the productivity growth operator. Hence, the production side in the nonlocal spatial Ramsey model is modeled as nonlinear and nonlocal productivity-production operator

$$\mathcal{P}(k)(x,t) := P(k)(x,t)p(k(x,t)), \tag{3.12}$$

where p denotes the neoclassical production function defined in (2.1).
The idea of such spillover effects that drive productivity growth was already studied in Brock et al. (2012). In their paper, the authors distinguish two cases: a deterministic, exogenous defined spillover effect and a spillover effect that is endogenously determined by the state of the system. The latter case is within our framework. However, in contrast to our model, Brock et al. (2012) insert the nonlocal spillover effect in the production function p and do not consider it as a rate to compound time and capital to an initial productivity distribution. Moreover, and this is crucial, the authors model the capital agglomeration in space to be only driven by the geographical spillovers. Thus, although depending on space, the capital accumulation equation in Brock et al. (2012) is an ordinary differential equation.

Based on the second welfare theorem, the representative agent approach allows us to consider the competitive equilibrium in the spatial Ramsey model with nonlocal capital diffusion and endogenous productivity growth as a solution of the following optimal control problem. Thus, the social optimum of the economy in our setting is defined as the minimum of

$$\mathcal{J}(k,c) := -\int_0^T \int_\Omega U(x,t,c(x,t))dxdt + \frac{1}{\rho}\|k(T) - k_T\|^2_{L^2(\Omega)}, \tag{3.13}$$

where $k_T : \Omega \to \mathbb{R}$ is a given terminal condition on the capital stock k and $\rho > 0$ denotes the penalty parameter, such that the capital accumulation equation

$$\frac{\partial k}{\partial t} - \alpha\Delta k - \beta\mathcal{NL}(k) + \delta k = \mathcal{P}(k) - c \tag{3.14}$$

holds on $\Omega \times [0,T]$, with constants $\alpha, \beta, \delta \geq 0$, where δ is as usual a given depreciation rate.
The function $U : \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}$ is a neoclassical utility function as in Assumption 2.4. We assume a central planner, hence both preferences and production are homogeneous in space. This is analog to the setting in Brito (2001). Nevertheless, our model differs according to the economic structure of accumulation and heterogeneity. Brito considers *'the simplest economic structure in which the only*

difference as far as space is concerned is related to the level of the (local) economic variables' (Brito, 2001, p.2). By introducing the nonlocal diffusion operator $\mathcal{NL}$ and the productivity growth operator P, the cross-sectional heterogeneity in our spatial Ramsey model is driven by endogenous and nonlocal effects that change across space.
Consistent with many common economic models, we will introduce a time discounting for a discount rate $\tau > 0$. This time discounting means, that the central planner values a contemporary gain of utility higher than future consumption. Whenever we consider an unbounded spatial domain, we also consider a space discounting analog to Camacho et al. (2008). This is again within the framework of a central planner and models a population density, or a political decision of the Benthamain planner. Moreover, the preferences of the decision maker impose a pre-ordering of the set of consumption bundles, which differ with respect to time and space.
We assume a finite time horizon T and introduce a terminal capital function $k_T \geq 0$. In that way, we are able to consider a larger (also infinite) time horizon, as long as we are able to meaningfully determine k_T. This terminal condition could be understood as a sustainability condition.

The capital accumulation equation (3.14) is to be interpreted like in the local model. We assume that the households can decide on their own, how much of their salary they spent for consumption and how much they save or invest. Analog to Camacho et al. (2008), the households can invest in all firms in space. These firms are assumed to be represented by one representative production function p, hence are equal in every point $x \in \Omega$. This is in line with the homogeneous production assumption. Nevertheless, due to the (non-constant) productivity operator, we generate a heterogeneous distribution of investment returns. The households are allowed to choose the most profitable location for investment. The investment in any location x at time t is given as

$$i(x,t) = \mathcal{P}(k)(x,t) - \delta k(x,t) - c(x,t),$$

and can be positive or negative.
Analog to the local model, we denote by $c : \Omega \times [0,T] \to \mathbb{R}$ the consumption distribution in time and space. In this monograph, we consider one single good, which is produced and consumed.
We introduce the initial value constraint

$$k(x,0) = k_0(x), \tag{3.15}$$

which should be satisfied in Ω. We will consider weak solutions of the PIDE constraint (3.14), hence the highest regularity we should assume on k_0 will be L^∞. This is especially interesting for the economic application, since we are allowed to start with a discontinuous initial capital distribution across space. We will point out in the numerical results, that the nonlocal diffusion operator preserves any heterogeneity and even discontinuities over time much longer than the pure local model.

Depending on whether we consider a bounded or unbounded spatial domain, we have to introduce boundary or volume constraints in order to make the problem well-posed. Common boundary conditions describe the function only on the surface $\partial\Omega$ of the bounded set Ω. In the context of nonlocal diffusion effects such surface constraints are not sufficient any more. Natural extensions of such boundary conditions are so called *volume constraints*, that act on a non-zero volume domain. Analogously to the local case, several types of these constraints are distinguished. Here, we consider Dirichlet- and Neumann-type volume constraints which both have an economic meaning in the Ramsey economy with a central planner.

In order to avoid that every household runs into debts, economists often introduce a so called *no Ponzi game condition* (cf. Brito, 2001, p.7). Under that constraint, the agents are not allowed to get into debt under multilevel selling. Since we are only considering a finite time horizon, we can reduce this condition with out any loss of generality to the state constraint

$$k(x,t) \geq 0, \ \forall (x,t) \in \Omega \times [0,T]. \tag{3.16}$$

Moreover, it is commonly assumed that also the consumption is bounded from below by zero since negative consumption means that the agent is starving, hence we additionally assume

$$c(x,t) \geq 0, \ \forall (x,t) \in \Omega \times [0,T]. \tag{3.17}$$

Boucekkine et al. (2009, p.3) confess that their spatial Ramsey model with local diffusion effects in the capital agglomeration equation is ill-posed in the sense of Hadamard, which means they cannot prove neither existence nor uniqueness of a solution. Although the model of Boucekkine et al. (2009), like the one of Brito (2001), can be seen as a special case of our nonlocal spatial Ramsey model, we are able to overcome this ill-posedness. In the following chapters, we will analyze the nonlocal spatial Ramsey model (3.13)-(3.17) according to well-posedness and solve it numerically. We will distinguish between two settings and consider the model on a bounded and an unbounded spatial domain. Both settings have meaningful interpretations and give interesting insights from application and mathematical points of view.

CHAPTER 4

The Nonlocal Spatial Ramsey Model on Unbounded Spatial Domains

'In some instances boundaries can be classified as significant borders, that is, as places where the economic conditions change abruptly because of some change, for example in the tax system, or in transport costs. In other instances we can speak of irrelevant borders, where nothing actually happens from an economic standpoint.' (Arbia, 2001, p.415)

In the originally space independent model, Ramsey (1928) himself considered an infinite time horizon. This assumption is appropriate from an economic point of view. Although no agent lives forever, this non-terminated time naturally introduces a sustainability condition. In some discrete models, as for example introduced by Acemoglu (2009, Chapter 6), an immortal agent is explicitly interpreted as a dynasty, where single individuals have the incentive to pass a non-zero capital stock to future generations. Whenever a space dimension is introduced to the Ramsey model, it is necessary to decide whether the spatial domain should be bounded or not. The combination of an infinite time horizon and an unbounded spatial domain holds some difficulties concerning the well-posedness of the spatial model (cf. Boucekkine et al., 2009, p.3). As already pointed out in Chapter 3.4, we circumvent these difficulties by introducing a terminal capital distribution k_T that should not be undercut. In this way, we mimic an infinite time horizon, but do only have to deal with a finite terminal time. Moreover, we introduce a spatial discounting in the objective function, which is convenient in the setting of a central planner. These additional constraints on the state variable and the special structure of the objective function allow us to consider an unbounded spatial domain in the nonlocal spatial Ramsey model. Such infinite space domains are of interest because they can be interpreted as one single and closed economy, where no flows of production factors to, or interactions with any other economies take place. Moreover, due to the spatial discounting, we do not need to define any boundary conditions in order to guarantee well-posedness of the model. Several types of boundary conditions, such as Neumann, Dirichlet, or Cauchy conditions and their economic meaning are for example discussed by Brito (2004, p.14). Here, it becomes obvious that the choice of the appropriate type of boundary conditions is not an easy task and that it heavily influences the solution of the underlying partial differential equation. Camacho et al. (2008) also consider a finite time horizon and unbounded spatial domain, but they disclaim any spatial discounting. This is the reason why they have to introduce free boundary conditions that enforce the capital distribution to become flat towards infinity, what restricts the set of possible solutions of the partial differential equation too much.

In this chapter, we provide an existence result of a weak solution of the capital equation over unbounded spatial domains and derive some regularity statements in Section 4.1. Moreover, we prove the existence of an optimal control in the nonlocal spatial Ramsey model. The latter is the main result of Section 4.2.

The capital accumulation equation, which we consider in the following, is a mixed local-nonlocal diffusion equation, i.e. the weights α, β in equation (3.14) are both positive. We see later that we can indeed choose α, which is the weight of the local diffusion term, very small, but that we cannot neglect it. Due to the unbounded spatial domain, we do not have to introduce any boundary, or volume constraints. Moreover, we do not have to truncate the kernel function in the nonlocal diffusion operator in this setting, but are able to analyze the dynamics of the Ramsey model on the whole, unbounded, and untruncated spatial domain. We fix the finite time horizon $T \in \mathbb{N}$ and the unbounded and open domain $\Omega \subseteq \mathbb{R}^n$. The capital accumulation equation, which we consider in the spatially unbounded case, is hence defined according to (3.14) as

$$\begin{aligned} k_t - \mathcal{L}(k) + \delta k - \mathcal{P}(k) &= -c \qquad \text{on } \Omega \times (0,T), \\ k(\cdot,0) &= k_0(\cdot) > 0 \qquad \text{in } \Omega, \end{aligned} \tag{4.1}$$

where the local-nonlocal diffusion operator $\mathcal{L}$ is defined as

$$\mathcal{L}(k)(x,t) := \alpha\ \Delta k(x,t) + \beta \int_\Omega (k(y,t) - k(x,t))\Gamma_\varepsilon(x,y)dy, \tag{4.2}$$

for coefficients $\alpha, \beta > 0$ and $\varepsilon > 0$. The kernel function is given according to equation (3.10) as the density function of the multivariate normal distribution,

$$\Gamma_\varepsilon(x,y) := \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{1}{2}(x-y)^T\Sigma_\varepsilon^{-1}(x-y)\right), \tag{4.3}$$

for a given covariance matrix Σ_ε with $\det(\Sigma_\varepsilon) = \varepsilon^{2n}$, $\varepsilon > 0$. In the following, we assume that the matrix Σ_ε is a diagonal matrix with constant entries,

$$\Sigma_\varepsilon = \begin{bmatrix} \varepsilon^2 & & \\ & \ddots & \\ & & \varepsilon^2 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

This assumption is application driven. We assume that capital can move through space without any barriers, or transition costs, thus capital flows are absolutely free in space. Moreover, the central planner does not prioritize any space direction, but weights them all equally. Hence, the spatial directions in the spatialized Ramsey model are completely uncorrelated and the variances are equal.
Given this special form of the covariance matrices, we can rewrite the kernel function as

$$\Gamma_\varepsilon(x, y) = \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|x - y\|_2^2}{2\varepsilon^2}\right), \tag{4.4}$$

where $\|\cdot\|_2$ denotes the Euclidean norm.

The nonlocal operator $\mathcal{P}$ on the left-hand side describes the production of the economy and is given according to equation (3.12) as

$$\begin{aligned} \mathcal{P}(k)(x,t) &:= P(k)(x,t)\ p(k(x,t)) \\ &= A_0(x) \exp\left(\frac{\int_\Omega \phi(k(y,t))\Gamma_\mu(x,y)dy}{\int_\Omega \phi(k(y,t))\Gamma_\varepsilon(x,y)dy + \xi}\ t\right)\ p(k(x,t)), \end{aligned} \tag{4.5}$$

where $A_0 : \mathbb{R}^n \to \mathbb{R}$ denotes the initial productivity distribution over space, $\phi : \mathbb{R} \to \mathbb{R}_+$ is the continuous nominal function, and $p : \mathbb{R} \to \mathbb{R}_+$ denotes the productivity function. The kernel function Γ_μ is defined analogously to (4.4) for a parameter $0 < \mu < \varepsilon$. The boundedness of the fraction in the exponential function is an important property, that we will exploit very often in this chapter. We state this property in the next lemma.

Lemma 4.1:
Let $\Omega \subseteq \mathbb{R}^n$ be an unbounded domain, let $\phi : \mathbb{R} \to \mathbb{R}_+$ generate a nonnegative superposition operator, $\xi > 0$, and let the kernel functions Γ_μ and Γ_ε for parameters $0 < \mu < \varepsilon$ be defined according to equation (4.4). Then the estimate

$$\frac{\int_\Omega \phi(k(y,t))\Gamma_\mu(x,y)dy}{\int_\Omega \phi(k(y,t))\Gamma_\varepsilon(x,y)dy + \xi} \leq \left(\frac{\varepsilon}{\mu}\right)^n \tag{4.6}$$

holds for all $x \in \Omega$.

Proof. Without any loss of generality, we choose $x = 0$. As x is by definition the expected value of Γ_ν, $\nu \in \{\mu, \varepsilon\}$, the proof will be analog for every other x, but with translational displaced Γ_ν. We define $\Gamma_\nu(0, y) =: \Gamma_\nu(y)$. The inequality (4.6) can be rewritten as

$$\int_\Omega \phi(k(y,t))\Gamma_\mu(y)dy \leq \left(\frac{\varepsilon}{\mu}\right)^n \left(\int_\Omega \phi(k(y,t))\Gamma_\varepsilon(y)dy + \xi\right).$$

Subtracting the left term, we get

$$0 \leq \int_\Omega \phi(k(y,t)) \left(\left(\frac{\varepsilon}{\mu}\right)^n \Gamma_\varepsilon(y) - \Gamma_\mu(y)\right) dy + \left(\frac{\varepsilon}{\mu}\right)^n \xi,$$

which is in particular true whenever

$$\left(\frac{\varepsilon}{\mu}\right)^n \Gamma_\varepsilon(y) - \Gamma_\mu(y) \geq 0$$

for all $y \in \Omega$, since we assume ϕ to be nonnegative. But this inequality follows with the monotonicity of the exponential function. Let therefore $y \in \Omega$ be arbitrary, then it holds

$$\begin{aligned}\left(\frac{\varepsilon}{\mu}\right)^n \Gamma_\varepsilon(y) - \Gamma_\mu(y) &= \left(\frac{\varepsilon}{\mu}\right)^n \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|y\|_2^2}{2\varepsilon^2}\right) \\ &\quad - \frac{1}{\sqrt{(2\pi\mu^2)^n}} \exp\left(-\frac{\|y\|_2^2}{2\mu^2}\right) \\ &= \frac{1}{\sqrt{(2\pi\mu^2)^n}} \left(\exp\left(-\frac{\|y\|_2^2}{2\varepsilon^2}\right) - \exp\left(-\frac{\|y\|_2^2}{2\mu^2}\right)\right) \geq 0,\end{aligned}$$

whenever

$$-\frac{\|y\|_2^2}{2\varepsilon^2} \geq -\frac{\|y\|_2^2}{2\mu^2},$$

which completes the proof. □

4.1 The Weak Solution over Unbounded Spatial Domains

In this section, we give an existence and uniqueness result of a weak solution of the capital accumulation equation of the nonlocal spatial Ramsey model with endogenous productivity growth over unbounded spatial domains, as defined in equation (4.1). Before we consider the economic application, we introduce an abstract existence result of a weak solution of a linear, but inhomogeneous nonlocal parabolic differential equation applying an argument from Wloka (1982). We will then extend this result to the semilinear case as the application to our nonlocal spatial Ramsey model with productivity growth.

We consider the following abstract, nonlocal, and linear initial value problem

$$
\begin{aligned}
k_t - a\ \Delta k - b\ \int_\Omega (k(y,\cdot) - k(\cdot,\cdot))\gamma(\cdot,y)dy + c\ k \ &= \ f \qquad\qquad \text{on } \Omega \times (0,T),\\
k(\cdot,0) &= k_0(\cdot) \qquad\quad \text{in } \Omega,
\end{aligned}
\tag{4.7}
$$

where a, b, c are either time depending functions, mapping between $(0,T)$ and $\mathbb{R}_+$, or fixed coefficients in $\mathbb{R}_+$. The function $\gamma : \Omega \times \Omega \to \mathbb{R}$ denotes a kernel function.

Definition 4.2:
Consider the Gelfand triple $H^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$. A function $k \in W(0,T)$ is called a weak solution of (4.7) if

- *for all $v \in H^1(\Omega)$ yields*

$$
\begin{aligned}
&\langle k_t(\cdot), v\rangle_{L^2(\Omega)} + \int_\Omega a(\cdot)\ \nabla_x k(x,\cdot)^T \nabla_x v(x) + c(\cdot)\ k(x,\cdot)v(x)\ dx\\
&- b(\cdot)\int_\Omega (k(y,\cdot) - k(x,\cdot))\gamma(x,y)dy\ v(x)\ dx = \int_\Omega f(x,\cdot)v(x)\ dx,
\end{aligned}
\tag{4.8}
$$

  in the sense of $\mathcal{D}'(]0,T[)$,

- *$k(\cdot,0) = k_0(\cdot)$ almost everywhere on Ω.*

The weak formulation (4.8) motivates the definition of the bilinear form $\mathbf{l} : H^1(\Omega) \times H^1(\Omega) \times [0,T] \to \mathbb{R}$,

$$
\begin{aligned}
\mathbf{l}(k,v)(t) = a(t)\ &\int_\Omega \nabla_x k(x)^T\ \nabla_x v(x)\ dx + c(t)\int_\Omega k(x)v(x)\ dx\\
&- b(t)\int_\Omega\int_\Omega (k(y) - k(x))\gamma(x,y)dy\ v(x)\ dx.
\end{aligned}
\tag{4.9}
$$

Assumption 4.3:
Let the following assumptions hold for $\mathbf{l}$:

(a) The bilinear form $\mathbf{l}(u,v)(\cdot)$ is measurable on $[0,T]$ (for fixed u, v).

(b) $\mathbf{l}(\cdot,\cdot)(t)$ is continuous, hence there exists a constant $c_1 > 0$ (independent of t), such that

$$|\mathbf{l}(u,v)(t)| \le c_1 \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)},$$

for all $t \in [0,T]$ and $u, v\ \in L^2(\Omega)$.

(c) There exist some constants $c_2 \ge 0$ and $c_3 > 0$ (independent of t), such that $\mathbf{l}$ satisfies the Gårding inequality

$$\mathbf{l}(u,u)(t) + c_2\|u\|^2_{L^2(\Omega)} \ge c_3\|u\|_{H^1(\Omega)},$$

for all $t \in [0,T]$ and $u \in L^2(\Omega)$.

We hide the nonlocality inside the bilinear form. Since the result of Wloka (1982) is quite abstract and the bilinear form $\mathbf{l}$ is assumed to satisfy all necessary conditions, we get the following quite abstract existence result for a linear nonlocal partial differential equation.

Theorem 4.4:
Let $f \in L^2(0,T;H^{-1}(\Omega))$ and $k_0 \in L^2(\Omega)$ be given functions, and suppose that the bilinear form $\mathbf{l}$ satisfies Assumption 4.3. Then the problem (4.7) has a unique weak solution $k \in W(0,T)$.

The proof by Wloka (1982, pp.384-389) can be adapted one to one to the bilinear form defined in (4.9).

We apply the result for linear PIDEs to the semilinear case, exploiting the Lipschitz continuity of the nonlinearity p and the boundedness of the exponential term. The spatial domain, which we consider in the following, is the untruncated $\mathbb{R}^n$, $n \in \mathbb{N}$. First, we derive the weak formulation of the PIDE (4.1), multiplying the equation with a function $v \in H^1(\mathbb{R}^n)$ and integrating over $\mathbb{R}^n$. Integrating by parts, we get

$$\int_{\mathbb{R}^n} k_t(x,\cdot)\; v(x)\; dx \;+\; \int_{\mathbb{R}^n} (\alpha\; \nabla_x k(x,\cdot)^T \nabla_x v(x) \;+ \delta k(x,\cdot)\; v(x))\; dx$$
$$- \beta \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (k(y,\cdot) - k(x,\cdot))\Gamma_\varepsilon(x,y) dy\; v(x)\; dx \;=\; \int_{\mathbb{R}^n} (\mathcal{P}(k)(x,\cdot) \;-\; c(x))\, v(x)\; dx,$$

where the equality has to be understood in distributional sense with respect to t.

This weak formulation motivates the following definition of a bilinear form.

Definition 4.5:
We define the bilinear form $\mathbf{a} : H^1(\mathbb{R}^n) \times H^1(\mathbb{R}^n) \to \mathbb{R}$ as

$$\begin{aligned}\mathbf{a}(u,v) :=\; &\alpha \int_{\mathbb{R}^n} \nabla_x u^T \nabla_x v\; dx \;+\; \delta \int_{\mathbb{R}^n} u\; v\; dx \\ &- \beta \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (u(y) - u(x))\Gamma_\varepsilon(x,y)\; dy\; v(x)\; dx.\end{aligned} \tag{4.10}$$

In order to apply Theorem 4.4, we have to show that $\mathbf{a}$ is continuous and weakly coercive. Note that this bilinear form is independent of time, since we have chosen α and β to be constants.

Lemma 4.6:
There exist some constants $c_1, c_3 > 0$, and $c_2 \geq 0$ such that the bilinear form $\mathbf{a}$ *as defined in (4.10) satisfies the following properties for all functions $u, v \in H^1(\mathbb{R}^n)$:*

$$\begin{aligned} &(i)\,\textit{Continuity: } |\mathbf{a}(u,v)| \leq c_1 \|u\|_{H^1(\mathbb{R}^n)} \|v\|_{H^1(\mathbb{R}^n)}, \\ &(ii)\,\textit{Gårding Inequality: } \mathbf{a}(u,u) + c_2 \|u\|^2_{L^2(\mathbb{R}^n)} \geq c_3 \|u\|^2_{H^1(\mathbb{R}^n)}. \end{aligned} \tag{4.11}$$

Proof.
(i) For the first and second term of the bilinear form defined in (4.10), it is true that

$$\left| \int_{\mathbb{R}^n} \big(\alpha \; \nabla_x u^T \nabla_x v + \delta \; uv\big) \; dx \right| \leq (\alpha + \delta) \|u\|_{H^1(\mathbb{R}^n)} \|v\|_{H^1(\mathbb{R}^n)} \qquad \forall \; u, v \in H^1(\mathbb{R}^n),$$

using the Hölder inequality two times and the definition of the $H^1(\mathbb{R}^n)$ norm.
In order to estimate the nonlocal term, a little more work has to be done. We rewrite the term for $y := x - z$ and apply the fundamental theorem of calculus. This yields

$$\begin{aligned}
&\left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (u(x-z) - u(x)) \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz \; v(x) \; dx \right| \\
=& \left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \int_0^1 \nabla_x u(x - \xi z)^T z d\xi \; \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz \; v(x) \; dx \right| \\
=& \left| \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \nabla_x u(x - \xi z) v(x) \; dx \; z \; \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz d\xi \right| \\
\leq& \left| \int_0^1 \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |\nabla_x u(x - \xi z)|^2 dx \right)^{1/2} \|v\|_{L^2(\mathbb{R}^n)} \; z \; \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz d\xi \right| \\
=& \left| \int_0^1 \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |\nabla_x u(y)|^2 \; dy \right)^{1/2} \|v\|_{L^2(\mathbb{R})} \; z \; \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz d\xi \right| \\
\leq& \left| \int_{\mathbb{R}^n} \|\nabla_x u\|_{L^2(\mathbb{R}^n)} \|v\|_{L^2(\mathbb{R}^n)} \; z \; \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz \right| \\
\leq& \|\nabla_x u\|_{L^2(\mathbb{R}^n)} \|v\|_{L^2(\mathbb{R}^n)} \left| \int_{\mathbb{R}^n} \; z \; \frac{1}{(\sqrt{2\pi\varepsilon^2})^n} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz \right| \\
\leq& \kappa \|\nabla_x u\|_{L^2(\mathbb{R}^n)} \|v\|_{L^2(\mathbb{R}^n)} \\
\leq& \kappa \|u\|_{H^1(\mathbb{R}^n)} \|v\|_{H^1(\mathbb{R}^n)},
\end{aligned}$$

with

$$0 \leq \kappa := \int_{\mathbb{R}^n} |z| \; \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(-\frac{\|z\|_2^2}{2\varepsilon^2}\right) dz.$$

Note that this integral is finite, since $\int_{\mathbb{R}^n} |x| \exp(-a\|x\|_2^2)\, dx$ is bounded whenever a is positive.

Combining all estimates, the continuity of $\mathbf{a}$,

$$|\mathbf{a}(u,v)| \leq c_1 \|u\|_{H^1(\mathbb{R}^n)} \|v\|_{H^1(\mathbb{R}^n)} \qquad \forall\, u, v \in H^1(\mathbb{R}^n),$$

where $0 \leq c_1 := \alpha + \delta + \beta\kappa$, is proven.

(ii) To prove the weak coercivity of $\mathbf{a}$, the procedure is the same as in (i), hence every term is estimated separately. For the first term, we have

$$\int_{\mathbb{R}^n} (\alpha\, |\nabla_x u|^2 + \delta\, u^2)\, dx = \alpha \|\nabla_x u\|^2_{L^2(\mathbb{R}^n)} + \delta \|u\|^2_{L^2(\mathbb{R}^n)} = \alpha \|u\|^2_{H^1(\mathbb{R}^n)} + (\delta - \alpha)\|u\|^2_{L^2(\mathbb{R}^n)}.$$

We use an estimation from (i), which leads to the following estimation for the last term:

$$-\beta \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (u(x-z) - u(x))\Gamma_\varepsilon(z) dz\; u(x)\; dx \geq -\beta\kappa \|\nabla_x u\|_{L^2(\mathbb{R}^n)} \|u\|_{L^2(\mathbb{R}^n)}.$$

Using Young's inequality for an arbitrary $c > 0$, we get

$$-\beta\kappa \|\nabla_x u\|_{L^2(\mathbb{R}^n)} \|u\|_{L^2(\mathbb{R}^n)} \geq -\frac{c}{2} \|\nabla_x u\|^2_{L^2(\mathbb{R}^n)} - \frac{(\beta\kappa)^2}{2c} \|u\|^2_{L^2(\mathbb{R}^n)}.$$

Combining both estimates then completes the proof,

$$\mathbf{a}(u,u) \geq \alpha \|u\|^2_{H^1(\mathbb{R}^n)} - \left(\alpha + \frac{(\beta\kappa)^2}{2c} - \delta\right) \|u\|^2_{L^2(\mathbb{R}^n)} - \frac{c}{2} \|\nabla_x u\|^2_{L^2(\mathbb{R}^n)}$$

which is equivalent to

$$\mathbf{a}(u,u) + \left(\alpha + \frac{(\beta\kappa)^2}{2c} - \delta\right) \|u\|^2_{L^2(\mathbb{R}^n)} \geq \left(\alpha - \frac{c}{2}\right) \|u\|^2_{H^1(\mathbb{R}^n)}.$$

□

Remark 4.7:
At the end of the proof of Lemma 4.6 (ii), we can choose $c > 0$ small enough such that

$$c_3 := \left(\alpha - \frac{c}{2}\right) > 0 \quad \textit{and} \quad c_2 := \left(\alpha + \frac{(\beta\kappa)^2}{2c} - \delta\right) \geq 0.$$

In that case, we get the Gårding inequality

$$\mathbf{a}(u,u) + c_2 \|u\|^2_{L^2} \geq c_3 \|u\|^2_{H^1}.$$

Instead of considering $k \in W(0,T)$ satisfying the capital accumulation equation, we can consider a function

$$z(t) = k(t)\exp(-c_2 t) \ \in W(0,T),$$

which has to satisfy the same equation (with a slightly modified right-hand side) and whose corresponding bilinear form is strictly coercive. Hence, we can interpret the bilinear forms in the proofs below as the one of z and assume the coercivity of **a** *without any loss of generality (cf. Wloka, 1982, p.384). We point that out later.*

Remark 4.8:
Note that at the end of the proof, we need the parameter α, which is the weighting parameter of the local diffusion operator, to be positive such that $\alpha - \frac{c}{2}$ is positive. The constant c, which comes from Young's inequality, is positive, so we cannot choose $\alpha = 0 = c$. Hence, at this point it becomes obvious why we need the local diffusion term in the Ramsey model on unbounded spatial domains. In the case of bounded spatial domains and volume constraints, we are able to apply Green's identity as introduced in the nonlocal vector calculus developed in Du et al. (2012a) and Du et al. (2012b). Thus we change the sign in front of the nonlocal diffusion operator. Here, no Green formula is available. Nevertheless, we can choose c to be very small and so are able to minimize the local diffusion effect in the spatial Ramsey model over unbounded spatial domains.

In order to prove the existence of a weak solution of the semilinear PIDE, we have to make some assumptions on the nonlinearities which are stated below. Although these may seem quite restrictive, we justify later that the assumptions are appropriate in the context of the Ramsey model.

Assumption 4.9:
The nonlinear functions in the nonlocal spatial Ramsey model with endogenous productivity growth are assumed to satisfy the following properties:

- *The production function $p : \mathbb{R} \to \mathbb{R}$ is concave and Lipschitz continuous, hence there exists a constant $L_p > 0$, such that*

$$|p(x) - p(y)| \leq L_p |x - y|, \ \forall\ x, y \in \mathbb{R}.$$

- *The production function p is bounded, hence there exists a constant $M_p > 0$, such that*

$$|p(x)| \leq M_p, \ \forall\ x \in \mathbb{R}.$$

- *The production function satisfies*

$$p(0) = 0.$$

- *The initial productivity distribution satisfies $A_0 \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$.*

- *The nominal function $\phi : \mathbb{R} \to \mathbb{R}_+$ is Lipschitz continuous with Lipschitz constant $L_\phi > 0$, hence*

$$|\phi(x) - \phi(y)| \leq L_\phi |u - v|, \ \forall \ x, y \in \mathbb{R}$$

and satisfies for some $a_0, \ a_1 \geq 0$

$$\phi(x) \leq a_0 + a_1 \ |x| \text{ for all } x \in \mathbb{R}.$$

We want to apply the existence result stated in Theorem 4.4 for linear differential equations. Therefore, we need to show the following regularity of the productivity-production operator $\mathcal{P}$.

Lemma 4.10:
Let Assumption 4.9 be valid. The functional $\mathcal{P}$ is bounded between $L^2(0,T;L^2(\mathbb{R}^n))$ and $L^2(0,T;L^2(\mathbb{R}^n))$.

Proof. Let k be a function in $L^2(0,T;L^2(\mathbb{R}))$. Then, we estimate

$$\begin{aligned}
\|\mathcal{P}(k)\|^2_{L^2(0,T;L^2(\mathbb{R}^n))} &= \int_0^T \int_{\mathbb{R}^n} |\mathcal{P}(k)(x,t)|^2 \ \ dxdt \\
= \int_0^T \int_{\mathbb{R}^n} &\left| A_0(x) \exp\left(\frac{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\mu(x,y) \ dy}{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\varepsilon(x,y) \ dy + \xi} \ t \right) \ p(k(x,t)) \right|^2 \ dxdt \\
\leq \|A_0\|^2_{L^\infty(\mathbb{R}^n)} &\int_0^T \int_{\mathbb{R}^n} \left| \exp\left(2t \frac{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\mu(x,y) \ dy}{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\varepsilon(x,y) \ dy + \xi} \right) \ p^2(k(x,t)) \right| \ dxdt \\
\leq \|A_0\|^2_{L^\infty(\mathbb{R}^n)} &\int_0^T \int_{\mathbb{R}^n} \left| \exp\left(\frac{2t\varepsilon^n}{\mu^n} \right) p^2(k(x,t)) \right| \ dxdt \\
\leq \|A_0\|^2_{L^\infty(\mathbb{R}^n)} &e^{\frac{2T\varepsilon^n}{\mu^n}} \int_0^T \int_{\mathbb{R}^n} |p(k(x,t))|^2 \ dxdt \\
= \|A_0\|^2_{L^\infty(\mathbb{R}^n)} &e^{\frac{2T\varepsilon^n}{\mu^n}} \int_0^T \int_{\mathbb{R}^n} |p(k(x,t)) - p(0)|^2 \ dxdt \\
\leq \|A_0\|^2_{L^\infty(\mathbb{R}^n)} &e^{\frac{2T\varepsilon^n}{\mu^n}} L_p^2 \int_0^T \int_{\mathbb{R}^n} |k(x,t)|^2 \ dxdt \\
= \|A_0\|^2_{L^\infty(\mathbb{R}^n)} &e^{\frac{2T\varepsilon^n}{\mu^n}} L_p^2 \|k\|^2_{L^2(0,T;L^2(\mathbb{R}^n))} < \infty.
\end{aligned}$$

□

In particular, we have proven that $\mathcal{P}(k)$ is an element of $L^2(0,T;H^{-1}(\mathbb{R}^n))$, since $L^2(0,T;L^2(\mathbb{R}^n)) \hookrightarrow L^2(0,T;H^{-1}(\mathbb{R}^n))$. Now we have all at hand to proof the existence and uniqueness of a weak solution of the PIDE constraint in our nonlocal Ramsey model, as stated in the next theorem.

Theorem 4.11:
Let $k_0 \in L^2(\mathbb{R}^n),\ c \in L^2(0,T;H^{-1}(\mathbb{R}^n))$ and let the functions p, ϕ, and A_0 satisfy Assumption 4.9. Then the capital accumulation equation in the nonlocal spatial Ramsey model with endogenous productivity growth (4.1) admits a unique weak solution $k \in W(0,T)$.

Proof. We give the proof to Theorem 4.11, following a common technique, which is based on Banach's fixed point theorem and the Lipschitz continuity of the non-linearity. We exploit the boundedness of the productivity growth operator

$$P(k)(x,t) = A_0(x)\exp\left(\frac{\int_{\mathbb{R}^n}\phi(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\mathbb{R}^n}\phi(k(y,t))\Gamma_\varepsilon(x,y)dy+\xi}\ t\right)$$

in $L^\infty(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$, the Lipschitz continuity of the functions p, ϕ, and the local Lipschitz continuity of the exponential function to derive the Lipschitz continuity of the whole productivity-production operator.
First, we fix $T^* \in (0,T)$. We show that the solution mapping, which maps a right-hand side to the solution of the linearized differential equation, is a contraction for T^* sufficiently small.
Let $v \in \mathcal{C}([0,T^*];L^2(\mathbb{R}^n))$, for short $\mathcal{C}(0,T^*;L^2(\mathbb{R}^n))$. As proven in Lemma 4.10, $\mathcal{P}(v) \in L^2(0,T^*;L^2(\mathbb{R}^n))$, where we have used the inequality

$$\|v\|_{L^2(0,T^*;L^2(\mathbb{R}^n))} \leq \|v\|_{L^\infty(0,T^*;L^2(\mathbb{R}^n))}$$

for all finite T^*. According to Theorem 4.4, there exists a unique weak solution $u \in W(0,T^*)$ of

$$u_t - \mathcal{L}(u) + \delta u = \mathcal{P}(v) - f \qquad \text{in } \mathbb{R}^n \times (0,T^*),$$

with $u(\cdot,0) = k_0(\cdot)$ on $\mathbb{R}^n$. Theorem 3.18 (c) then guarantees that u is an element of $\mathcal{C}(0,T^*;L^2(\mathbb{R}^n))$. This defines the operator

$$S : \mathcal{C}(0,T^*;L^2(\mathbb{R}^n)) \to \mathcal{C}(0,T^*;L^2(\mathbb{R}^n)),\ S(v) = u.$$

In the following, we prove that S is a contraction. Consider the difference $S(v_1) - S(v_2)$ for two arbitrary functions $v_1, v_2 \in \mathcal{C}(0,T^*;L^2(\mathbb{R}^n))$ with $S(v_1) = u_1$ and $S(v_2) = u_2$. We choose the function $u := u_1 - u_2 \in W(0,T^*)$ and deduce the weak formulation as

$$\int_0^t \int_{\mathbb{R}^n} u_t(x,s)u(x,s)\ dx + \mathbf{a}(u,u)(s)\ ds =$$
$$\int_0^t \int_{\mathbb{R}^n} (\mathcal{P}(v_1)(x,s) - \mathcal{P}(v_2)(x,s))u(x,s)\ dxds,$$

for all $t \in [0,T^*]$.

We can estimate the left-hand side (LHS) using a calculation from Lemma 4.6 (i):

$$\begin{aligned} LHS \geq &\int_0^t \int_{\mathbb{R}^n} u_t(x,s)u(x,s)\ dxds + \alpha \int_0^t \int_{\mathbb{R}} |\nabla_x u|^2(x,s)\ \ dxds \\ &+ \delta \int_0^t \int_{\mathbb{R}^n} u^2(x,s)\ dxds - \kappa_1 \int_0^t \|\nabla_x u(s)\|_{L^2(\mathbb{R}^n)} \|u(s)\|_{L^2(\mathbb{R}^n)} ds, \end{aligned}$$

where

$$0 \leq \kappa_1 := \beta \int_{\mathbb{R}^n} |z| \frac{1}{\sqrt{(2\pi\varepsilon^2)^n}} \exp\left(\frac{-\|z\|_2^2}{2\varepsilon^2}\right)\ dz.$$

Applying Young's inequality for an arbitrary $\eta > 0$, we get

$$\begin{aligned} LHS \geq &\int_0^t \int_{\mathbb{R}^n} u_t(x,s)u(x,s)\ dxds + \alpha \int_0^t \int_{\mathbb{R}^n} |\nabla_x u|^2(x,s)\ \ dxds \\ &+ \delta \int_0^t \int_{\mathbb{R}^n} u^2(x,s)\ \ dxds - \int_0^t \frac{\eta}{2} \|\nabla_x u(s)\|^2_{L^2(\mathbb{R}^n)} + \frac{\kappa_1^2}{2\eta} \|u(s)\|^2_{L^2(\mathbb{R}^n)}\ ds. \end{aligned}$$

We choose $\eta \leq 2\alpha$, which yields together with the identity

$$\int_0^t \int_{\mathbb{R}^n} u_t(x,s)u(x,s)\ dxds = \frac{1}{2}\|u(\cdot,t)\|^2_{L^2(\mathbb{R}^n)}$$

the following estimate for the left-hand side

$$LHS \geq \frac{1}{2}\|u(\cdot,t)\|^2_{L^2(\mathbb{R}^n)} + \int_0^t \delta \|u(\cdot,s)\|^2_{L^2(\mathbb{R}^n)} + \frac{\kappa_1^2}{2\eta}\|u(\cdot,s)\|^2_{L^2(\mathbb{R}^n)}\ ds.$$

For the right-hand side (RHS), we get

$$\begin{aligned} &\int_0^t \int_{\mathbb{R}^n} (\mathcal{P}(v_1)(x,s) - \mathcal{P}(v_2)(x,s))\ u(x,s)\ \ dxds \\ \leq &\int_0^t \|\mathcal{P}(v_1)(\cdot,s) - \mathcal{P}(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)} \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds \\ = &\int_0^t \|\mathcal{P}(v_1)(\cdot,s) - P(v_1)(\cdot,s)p(v_2(\cdot,s)) + P(v_1)(\cdot,s)p(v_2(\cdot,s)) - \mathcal{P}(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)} \\ &\quad \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds \\ \leq &\int_0^t \|\mathcal{P}(v_1)(\cdot,s) - P(v_1)(\cdot,s)p(v_2(\cdot,s))\|_{L^2(\mathbb{R}^n)}\ \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds \\ &+ \int_0^t \|P(v_1)(\cdot,s)p(v_2(\cdot,s)) - \mathcal{P}(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)}\ \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds. \end{aligned}$$

To estimate the first term, we exploit the Lipschitz continuity of the production function p and the boundedness of the productivity growth operator P and get

$$\begin{aligned}
&\|\mathcal{P}(v_1)(\cdot,s) - P(v_1)(\cdot,s)p(v_2(\cdot,s))\|_{L^2(\mathbb{R}^n)} \\
&= \|P(v_1)(\cdot,s)(p(v_1(\cdot,s)) - p(v_2(\cdot,s)))\|_{L^2(\mathbb{R}^n)} \\
&\leq \|P(v_1)(\cdot,s)\|_{L^\infty(\mathbb{R}^n)}\|p(v_1(\cdot,s)) - p(v_2(\cdot,s))\|_{L^2(\mathbb{R}^n)} \\
&\leq L_p \; \|P(v_1)(\cdot,s)\|_{L^\infty(\mathbb{R}^n)}\|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)},
\end{aligned}$$

where we deduce from (4.6)

$$\begin{aligned}
\|P(v)(\cdot,s)\|_{L^\infty(\mathbb{R}^n)} &:= ess \sup_{x\in\mathbb{R}^n}\left| A_0(x)\exp\left(\frac{\int_{\mathbb{R}^n}\phi(v(y,s))\Gamma_\mu(x,y)dy}{\int_{\mathbb{R}^n}\phi(v(y,s))\Gamma_\varepsilon(x,y)dy+\xi}s\right)\right| \\
&= \|A_0\|_{L^\infty(\mathbb{R}^n)}e^{\frac{s\varepsilon^n}{\mu^n}} \\
&=: \frac{\kappa_2}{L_p}e^{\frac{s\varepsilon^n}{\mu^n}} < \infty,
\end{aligned}$$

for all $s \in [0,t]$. Estimating the second term yields

$$\begin{aligned}
&\|P(v_1)(\cdot,s)p(v_2(\cdot,s)) - \mathcal{P}(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)} \\
&= \|p(v_2(\cdot,s))(P(v_1)(\cdot,s) - P(v_2)(\cdot,s))\|_{L^2(\mathbb{R}^n)} \\
&\leq \|p(v_2(\cdot,s))\|_{L^\infty(\mathbb{R}^n)}\|P(v_1)(\cdot,s) - P(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)} \\
&\leq M_p\|P(v_1)(\cdot,s) - P(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)}.
\end{aligned}$$

Here, it is

$$\begin{aligned}
&\|P(v_1)(\cdot,s) - P(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)} = \\
&\Bigg\| A_0(\cdot)\Bigg[\exp\left(\frac{\int_{\mathbb{R}^n}\phi(v_1(y,s))\Gamma_\mu(\cdot,y)dy}{\int_{\mathbb{R}^n}\phi(v_1(y,s))\Gamma_\varepsilon(\cdot,y)dy+\xi}s\right) \\
&\qquad\qquad - \exp\left(\frac{\int_{\mathbb{R}^n}\phi(v_2(y,s))\Gamma_\mu(\cdot,y)dy}{\int_{\mathbb{R}^n}\phi(v_2(y,s))\Gamma_\varepsilon(\cdot,y)dy+\xi}s\right)\Bigg]\Bigg\|_{L^2(\mathbb{R}^n)}.
\end{aligned}$$

We shorten the expression in the following and define for $\nu \in \{\varepsilon,\mu\}$ the operator

$$\Phi_\nu(v)(x,s) := \int_{\mathbb{R}^n}\phi(v(y,s))\Gamma_\nu(x,y)dy.$$

We exploit the boundedness of the fraction according to the inequality (4.6),

$$\left|\frac{\Phi_\mu(v_1)(x,s)}{\Phi_\varepsilon(v_1)(x,s)+\xi}s - \frac{\Phi_\mu(v_2)(x,s)}{\Phi_\varepsilon(v_2)(x,s)+\xi}s\right| \leq 2\frac{\varepsilon^n}{\mu^n}$$

for all v_1, v_2, so we can use the local Lipschitz continuity of the exponential function with Lipschitz constant $L_{\exp} > 0$ and get

$$
\begin{aligned}
&\left\| A_0(\cdot) \left[\exp\left(\frac{\Phi_\mu(v_1)(\cdot,s)}{\Phi_\varepsilon(v_1)(\cdot,s)+\xi} s \right) - \exp\left(\frac{\Phi_\mu(v_2)(\cdot,s)}{\Phi_\varepsilon(v_2)(\cdot,s)+\xi} s \right) \right] \right\|_{L^2(\mathbb{R}^n)} \\
&\leq s L_{exp} \left\| A_0(\cdot) \left[\frac{\Phi_\mu(v_1)(\cdot,s)}{\Phi_\varepsilon(v_1)(\cdot,s)+\xi} - \frac{\Phi_\mu(v_2)(\cdot,s)}{\Phi_\varepsilon(v_2)(\cdot,s)+\xi} \right] \right\|_{L^2(\mathbb{R}^n)} .
\end{aligned}
$$

The calculation

$$
\begin{aligned}
&\left| \frac{\Phi_\mu(v_1)(\Phi_\varepsilon(v_2)+\xi) - \Phi_\mu(v_2)(\Phi_\varepsilon(v_1)+\xi)}{(\Phi_\varepsilon(v_1)+\xi)(\Phi_\varepsilon(v_2)+\xi)} \right| \\
&= \left| \frac{\Phi_\mu(v_1)\Phi_\varepsilon(v_2) + \Phi_\mu(v_1)\xi - \Phi_\mu(v_2)\Phi_\varepsilon(v_1) - \Phi_\mu(v_2)\xi}{(\Phi_\varepsilon(v_1)+\xi)(\Phi_\varepsilon(v_2)+\xi)} \right| \\
&= \left| \frac{\Phi_\mu(v_1)\Phi_\varepsilon(v_2) - \Phi_\mu(v_1)\Phi_\varepsilon(v_1) + \Phi_\mu(v_1)\Phi_\varepsilon(v_1) - \Phi_\mu(v_2)\Phi_\varepsilon(v_1) + \Phi_\mu(v_1)\xi - \Phi_\mu(v_2)\xi}{(\Phi_\varepsilon(v_1)+\xi)(\Phi_\varepsilon(v_2)+\xi)} \right| \\
&\leq \left| \frac{\Phi_\mu(v_1)}{(\Phi_\varepsilon(v_1)+\xi)(\Phi_\varepsilon(v_2)+\xi)} \right| |(\Phi_\varepsilon(v_1) - \Phi_\varepsilon(v_2))| \\
&\quad + \left(\left| \frac{\Phi_\varepsilon(v_1)}{(\Phi_\varepsilon(v_1)+\xi)(\Phi_\varepsilon(v_2)+\xi)} \right| + \left| \frac{\xi}{(\Phi_\varepsilon(v_1)+\xi)(\Phi_\varepsilon(v_2)+\xi)} \right| \right) |(\Phi_\mu(v_1) - \Phi_\mu(v_2))| \\
&\leq \frac{\varepsilon^n}{\mu^n \xi} (\Phi_\varepsilon(v_1) - \Phi_\varepsilon(v_2)) + \frac{2}{\xi} (\Phi_\mu(v_1) - \Phi_\mu(v_2)),
\end{aligned}
$$

yields

$$
\begin{aligned}
&\left\| A_0(\cdot) \left[\frac{\Phi_\mu(v_1)(\cdot,s)(\Phi_\varepsilon(v_2)(\cdot,s)+\xi)}{(\Phi_\varepsilon(v_1)(\cdot,s)+\xi)(\Phi_\varepsilon(v_2)(\cdot,s)+\xi)} - \frac{\Phi_\mu(v_2)(\cdot,s)(\Phi_\varepsilon(v_1)(\cdot,s)+\xi)}{(\Phi_\varepsilon(v_1)(\cdot,s)+\xi)(\Phi_\varepsilon(v_2)(\cdot,s)+\xi)} \right] \right\|_{L^2(\mathbb{R}^n)} \\
&\leq \frac{2}{\xi} \left\| A_0(\cdot) \left[\Phi_\mu(v_1)(\cdot,s) - \Phi_\mu(v_2)(\cdot,s) \right] \right\|_{L^2(\mathbb{R}^n)} + \frac{\varepsilon^n}{\xi \mu^n} \left\| A_0(\cdot) \left[\Phi_\varepsilon(v_1)(\cdot,s) - \Phi_\varepsilon(v_2)(\cdot,s) \right] \right\|_{L^2(\mathbb{R}^n)} .
\end{aligned}
$$

Here, we have estimated the term

$$
\left| \frac{\Phi_\mu(v_1)}{(\Phi_\varepsilon(v_1)+\xi)(\Phi_\varepsilon(v_2)+\xi)} \right| \leq \frac{1}{\xi} \left| \frac{\Phi_\mu(v_1)}{(\Phi_\varepsilon(v_1)+\xi)} \right| \leq \frac{1}{\xi} \left(\frac{\varepsilon}{\mu} \right)^n ,
$$

applying the inequality (4.6) and $(\Phi_\varepsilon(v_1)+\xi)\Phi_\varepsilon(v_2) \geq 0$ by assumption. The other terms are estimated in a similar way.
We now apply the Lipschitz continuity of the function ϕ and get

$$
\begin{aligned}
&\|A_0(\cdot)\left[\Phi_\nu(v_1)(\cdot,s)-\Phi_\nu(v_2)(\cdot,s)\right]\|_{L^2(\mathbb{R}^n)}\\
&=\left\|A_0(\cdot)\left[\int_{\mathbb{R}^n}\phi(v_1(y,s))\Gamma_\nu(\cdot,y)dy-\int_{\mathbb{R}^n}\phi(v_2(y,s))\Gamma_\nu(\cdot,y)dy\right]\right\|_{L^2(\mathbb{R}^n)}\\
&=\left(\int_{\mathbb{R}^n}\left|A_0(x)\int_{R^n}(\phi(v_1(y,s))-\phi(v_2(y,s)))\Gamma_\nu(x,y)dy\right|^2dx\right)^{\frac{1}{2}}\\
&\leq\left(\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}(\phi(v_1(y,s))-\phi(v_2(y,s)))^2dy\right)\left(\int_{\mathbb{R}^n}A_0(x)^2\Gamma_\nu^2(x,y)dy\right)dx\right)^{\frac{1}{2}}\\
&\leq L_\phi\|v_1-v_2\|_{L^2(\mathbb{R}^n)}\left(\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}A_0^2(x)\Gamma_\nu^2(x,y)dydx\right)^{\frac{1}{2}},
\end{aligned}
$$

for $\nu\in\{\varepsilon,\mu\}$.

Note that, since the kernel function Γ_ν is a multivariate Gaussian probability density function, it holds

$$
ess\sup_{x\in\mathbb{R}^n}\int_{\mathbb{R}^n}\frac{1}{(2\pi\nu^2)^n}\exp\left(-\frac{\|x-y\|_2^2}{\nu^2}\right)dy=\frac{1}{(2\nu\sqrt{\pi})^n}<\infty.
$$

Since we have assumed

$$
\|A_0\|_{L^2(\mathbb{R}^n)}<\infty,
$$

we can finally deduce

$$
\|P(v_1)(\cdot,s)-P(v_2)(\cdot,s)\|_{L^2(\mathbb{R}^n)}\leq s\kappa_3\|v_1(\cdot,s)-v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)},
$$

with a positive constant

$$
\begin{aligned}
\kappa_3:&=2M_pL_{exp}L_\phi\|A_0\|_{L^2(\mathbb{R}^n)}\max\left\{\frac{2}{\xi}\|\Gamma_\mu\|_{L^\infty(\mathbb{R}^n)\times L^2(\mathbb{R}^n)},\frac{\varepsilon^n}{\xi\mu^n}\|\Gamma_\varepsilon\|_{L^\infty(\mathbb{R}^n)\times L^2(\mathbb{R}^n)}\right\}\\
&=2M_pL_{exp}L_\phi\|A_0\|_{L^2(\mathbb{R}^n)}\max\left\{\frac{2}{\xi\sqrt{(2\sqrt{\pi}\mu)^n}},\frac{\varepsilon^n}{\xi\mu^n\sqrt{(2\sqrt{\pi}\varepsilon)^n}}\right\}<\infty.
\end{aligned}
$$

Combining both estimates, we have

$$
\begin{aligned}
&\int_0^t \int_{\mathbb{R}^n} (\mathcal{P}(v_1)(x,s) - \mathcal{P}(v_2)(x,s))(u(x,s))\, dxds \\
\leq &\int_0^t e^{\frac{\varepsilon^n s}{\mu^n}} \kappa_2 \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)} \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds \\
&+ \int_0^t s\kappa_3 \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)} \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds \\
\leq &\max\{\kappa_2,\kappa_3\} \int_0^t (s + e^{\frac{\varepsilon^n s}{\mu^n}}) \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)} \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds.
\end{aligned}
$$

Applying Young's inequality for a $\varsigma > 0$ and denoting by $\kappa_\infty := \max\{\kappa_2,\kappa_3\}$ yields

$$
\begin{aligned}
&\kappa_\infty \int_0^t (s + e^{\frac{\varepsilon^n s}{\mu^n}}) \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)} \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)} ds \\
\leq &\frac{\kappa_\infty^2}{2\varsigma} \int_0^t (s + e^{\frac{\varepsilon^n s}{\mu^n}})^2 \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)}^2 ds + \frac{\varsigma}{2} \int_0^t \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)}^2 ds.
\end{aligned}
$$

All in all, we have

$$
\begin{aligned}
&\frac{1}{2}\|u(\cdot,t)\|_{L^2(\mathbb{R}^n)}^2 + \int_0^t \delta \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)}^2 + \frac{\kappa_1^2}{2\eta} \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)}^2\, ds \\
\leq &\frac{\kappa_\infty^2}{2\varsigma} \int_0^t (s + e^{\frac{\varepsilon^n s}{\mu^n}})^2 \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)}^2 ds + \frac{\varsigma}{2} \int_0^t \|u(\cdot,s)\|_{L^2(\mathbb{R}^n)}^2 ds.
\end{aligned}
$$

Sorting the inequality leads to

$$
\begin{aligned}
\frac{1}{2}\|u(\cdot,t)\|_{L^2(\mathbb{R}^n)}^2 \leq &\left(\frac{\varsigma}{2} - \frac{\kappa_1^2}{2\eta} - \delta\right) \int_0^t \|u(\cdot,s)\|_{L(\mathbb{R}^n)}^2 ds \\
&+ \frac{\kappa_\infty^2}{2\varsigma} \int_0^t (s + e^{\frac{\varepsilon^n s}{\mu^n}})^2 \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\mathbb{R}^n)}^2 ds,
\end{aligned}
$$

where we can choose the parameters ς and η, such that $(\varsigma/2 - \kappa_1^2/2\eta - \delta) \geq 0$. Taking the maximum for all $t \in [0,T^*]$, we end up with

$$
\begin{aligned}
&\frac{1}{2}\|u_1 - u_2\|_{L^\infty(0,T^*;L^2(\mathbb{R}^n))}^2 \\
&\leq T^* \left(\frac{\varsigma}{2} - \frac{\kappa_1^2}{2\eta} - \delta\right) \|u_1 - u_2\|_{L^\infty(0,T^*;L^2(\mathbb{R}^n))}^2 + \frac{\kappa_\infty^2}{2\varsigma} \int_0^{T^*} (s + e^{\frac{\varepsilon^n s}{\mu^n}})^2 ds \|v_1 - v_2\|_{L^\infty(0,T^*;L^2(\mathbb{R}^n))}^2
\end{aligned}
$$

as $\delta, \kappa^2 \geq 0$. Hence, we have

$$
\|u_1 - u_2\|_{L^\infty(0,T^*;L^2(\mathbb{R}^n))}^2 \leq C(T^*)\ \|v_1 - v_2\|_{L^\infty(0,T^*;L^2(\mathbb{R}^n))}^2, \qquad (\#)
$$

with

$$C(T^*) := \frac{\hat{\kappa}_\infty(\frac{{T^*}^3}{3} + \frac{\mu^n}{2\varepsilon^n} e^{\frac{2T^*\varepsilon^n}{\mu^n}} - \frac{\mu^n}{2\varepsilon^n} + \frac{\mu^n}{\varepsilon^n} e^{\frac{\varepsilon^n T^*}{\mu^n}} T^* - \frac{\mu^{2n}}{\varepsilon^{2n}} e^{\frac{\varepsilon^n T^*}{\mu^n}} + \frac{\mu^{2n}}{\varepsilon^{2n}})}{\frac{1}{2} - T^* \left(\frac{\varsigma}{2} - \frac{\kappa_1^2}{2\eta} - \delta \right)}.$$

Taking the limit $T^* \to 0$ yields

$$C(T^*) \to 0.$$

Especially, the exists a $T^* \in \mathbb{R}_+$, such that $C(T^*) < 1$.

Note that we can divide by $\frac{1}{2} - T^* \left(\frac{\varsigma}{2} - \frac{\kappa_1^2}{2\eta} - \delta \right)$, because we can choose ς, η appropriately, such that the term is positive, at least for small T^*. So, all in all, we have shown that S is a contraction for T^* small enough and we can apply the fixed point theorem of Banach which yields the existence of a unique fixed point $S(u) = u$ on $W(0, T^*)$. Now, we have to construct a solution k on the whole time-space-cylinder, but since the local solution u is independent of the time horizon T^*, we can proceed on the interval $[T^*, 2T^*]$ using the same arguments as above with a new initial condition $u(\cdot, T^*)$. After finitely many steps, we can construct a weak solution $k \in W(0, T)$ of (4.1). Moreover, this solution is unique, which follows from the inequality (#). □

Lemma 4.12:
Let $c \in L^2(0, T; L^2(\mathbb{R}^n))$. If the bilinear form $\mathbf{a}$ is coercive, then there exist two constants $\mathrm{C}_1, \mathrm{C}_2 > 0$ such that the solution $k \in W(0, T)$ of (4.1) satisfies the following a priori estimate:

$$\begin{aligned} \|k\|_{L^2(0,T;H^1(\mathbb{R}^n))} \; + \; \mathrm{C}_1 \|k\|_{L^\infty(0,T;L^2(\mathbb{R}^n))} & \\ \leq \mathrm{C}_2 \left(\|k_0\|_{L^2(\mathbb{R}^n)} + \|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + 1 \right). & \end{aligned} \tag{4.12}$$

Proof. The coercivity assumption on $\mathbf{a}$ yields a constant $c_{coer} > 0$ such that $\mathbf{a}(k, k) \geq c_{coer} \|k\|^2_{H^1(\mathbb{R}^n)}$ for all $k \in H^1(\mathbb{R}^n)$. Now, we fix a $t \in [0, T]$ and derive the weak formulation of (4.1) for the test function $k \in W(0, T)$. We then have

$$\int_0^t \int_{\mathbb{R}^n} \frac{\partial k}{\partial s} k \;\; dxds + \int_0^t \mathbf{a}(k(s), k(s)) \; ds = \int_0^t \int_{\mathbb{R}^n} \mathcal{P}(k) k \; dxds - \int_0^t \int_{\mathbb{R}^n} c \; k \; dxds.$$

To estimate the right-hand side, we exploit the Lipschitz continuity of p, $p(0) = 0$, and the boundedness of the fractional term in the productivity growth operator

(4.6). Altogether, this yields

$$
\begin{aligned}
RHS &\leq \int_0^t \|\mathcal{P}(k)\|_{L^2(\mathbb{R}^n)}\|k\|_{L^2(\mathbb{R}^n)}ds + \int_0^t \|c\|_{L^2(\mathbb{R}^n)}\|k\|_{L^2(\mathbb{R}^n)}ds \\
&\leq \|A_0\|_{L^\infty(\mathbb{R}^n)} \int_0^t \left(\int_{\mathbb{R}^n} (e^{\frac{s\varepsilon^n}{\mu^n}} p(k))^2 dx\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} |k|^2 dx\right)^{\frac{1}{2}} ds + \int_0^t \|c\|_{L^2(\mathbb{R}^n)}\|k\|_{L^2(\mathbb{R}^n)}ds \\
&\leq \|A_0\|_{L^\infty(\mathbb{R}^n)} L_p \int_0^t e^{\frac{s\varepsilon^n}{\mu^n}} \|k\|^2_{L^2(\mathbb{R}^n)}ds + \int_0^t \|c\|_{L^2(\mathbb{R}^n)}\|k\|_{L^2(\mathbb{R}^n)}ds.
\end{aligned}
$$

In order to estimate the left-hand side, we use the coercivity assumption for $\mathbf{a}$,

$$
\begin{aligned}
LHS &= \int_0^t \int_{\mathbb{R}^n} \frac{\partial k}{\partial s} k \; dxds + \int_0^t \mathbf{a}(k,k)ds \\
&\geq \int_0^t \int_{\mathbb{R}^n} \frac{\partial k}{\partial s} k \; dxds + c_{coer} \int_0^t \|k\|^2_{H^1(\mathbb{R}^n)}ds \\
&= \frac{1}{2}\|k(t)\|^2_{L^2(\mathbb{R}^n)} - \frac{1}{2}\|k_0\|^2_{L^2(\mathbb{R}^n)} + c_{coer} \int_0^t \|k\|^2_{H^1(\mathbb{R}^n)}ds.
\end{aligned}
$$

Combining both estimates and applying Young's inequality with $\eta_1, \eta_2 > 0$, we get

$$
\begin{aligned}
&\frac{1}{2}\|k(t)\|^2_{L^2(\mathbb{R}^n)} + c_{coer} \int_0^t \|k\|^2_{H^1(\mathbb{R}^n)}ds \\
\leq &\|A_0\|_{L^\infty(\mathbb{R}^n)} L_p \left(\frac{\eta_1 \mu^n}{4\varepsilon^n} \left(e^{\frac{2\varepsilon^n t}{\mu^n}} - 1\right) + \frac{1}{2\eta_1} \int_0^t \|k\|^2_{L^2(\mathbb{R}^n)}ds \right) \\
&+ \frac{\eta_2}{2} \int_0^t \|c\|^2_{L^2(\mathbb{R}^n)}ds + \frac{1}{2\eta_2} \int_0^t \|k\|^2_{L^2(\mathbb{R}^n)}ds + \frac{1}{2}\|k_0\|^2_{L^2(\mathbb{R}^n)}.
\end{aligned}
$$

Taking the maximum of all $t \in [0,T]$, we have

$$
\begin{aligned}
&\frac{1}{2}\|k\|^2_{L^\infty(0,T;L^2(\mathbb{R}^n))} + c_{coer}\|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} \\
\leq &\|A_0\|_{L^\infty(\mathbb{R}^n)} L_p \left(\frac{\eta_1 \mu^n}{4\varepsilon^n} \left(e^{\frac{2T\varepsilon^n}{\mu^n}} - 1\right) + \frac{1}{2\eta_1} \|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} \right) \\
&+ \frac{\eta_2}{2}\|c\|^2_{L^2(0,T;L^2(\mathbb{R}^n))} + \frac{1}{2\eta_2}\|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} + \frac{1}{2}\|k_0\|^2_{L^2(\mathbb{R}^n)}.
\end{aligned}
$$

We multiply with 2, sort all terms, and end up with

$$
\begin{aligned}
&\|k\|^2_{L^\infty(0,T;L^2(\mathbb{R}^n))} + \left(2c_{coer} - \frac{\|A_0\|_{L^\infty(\mathbb{R}^n)} L_p}{\eta_1} - \frac{1}{\eta_2}\right) \|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} \\
&\leq \eta_1 \; \|A_0\|_{L^\infty(\mathbb{R}^n)} L_p \frac{\mu^n}{2\varepsilon^n} \left(e^{\frac{2\varepsilon^n \; T}{\mu^n}} - 1\right) + \eta_2\|c\|^2_{L^2(0,T;L^2(\mathbb{R}^n))} + \|k_0\|^2_{L^2(\mathbb{R}^n)}.
\end{aligned}
$$

Since $\eta_1, \eta_2 > 0$ were arbitrary, we choose both constants large enough such that

$$\left(2c_{coer} - \frac{\|A_0\|_{L^\infty(\mathbb{R}^n)} L_p}{\eta_1} - \frac{1}{\eta_2}\right) > 0,$$

which completes the proof.

□

As already mentioned, the coercivity assumption on the bilinear form $\mathbf{a}$ is not as restrictive as it may seem. We have already shown in Lemma 4.6 that the corresponding bilinear form of the capital accumulation equation in the Ramsey model is continuous in $H^1(\mathbb{R}^n)$. According to Wloka (1982, Theorem 17.9, p.264 and (28), p.265), this yields the existence of a unique, bijective, linear, and continuous operator $L : H^1(\mathbb{R}^n) \to H^{-1}(\mathbb{R}^n)$ with

$$\mathbf{a}(u, v) = \langle L(u), v \rangle_{L^2(\mathbb{R}^n)}.$$

Hence, we can rewrite the nonlocal PIDE problem as follows.
For $k_0 \in L^2(\mathbb{R}^n)$ and $c \in L^2(0, T; H^{-1}(\mathbb{R}^n))$, find a function $k \in W(0, T)$ such that $k(\cdot, 0) = k_0(\cdot)$ on $\mathbb{R}^n$ and

$$L(k) + \frac{\partial k}{\partial t} = \mathcal{P}(k) - c.$$

We now assume that this equation has a solution $k \in W(0, T)$. We define

$$z(t) = k(t) \exp(-c_2 t),$$

where c_2 is the constant of the Gårding inequality in Lemma 4.6. Differentiating z after t yields

$$\frac{\partial z}{\partial t} = \frac{\partial k}{\partial t} \exp(-c_2 t) - c_2 k(t) \exp(-c_2 t) = \frac{\partial k}{\partial t} \exp(-c_2 t) - c_2 z(t).$$

Substituting $k(t) = z(t) \exp(c_2 t)$ in the PIDE, we get

$$\frac{\partial z}{\partial t} + (L(z) + c_2 E) z(t) = \exp(-c_2 t) \mathcal{P}(z(t) \exp(c_2 t)) - \exp(-c_2 t) c.$$

We define a new linear and continuous operator $\hat{L} := L + c_2 E$, which yields a new bilinear form $\hat{\mathbf{a}}$, derived analogously to (4.10). This bilinear form is coercive, since

$$\begin{aligned}
\hat{\mathbf{a}}(u, u) &= \langle \hat{L} u, u \rangle_{L^2(\mathbb{R}^n)} = \langle L u + c_2 E u, u \rangle_{L^2(\mathbb{R}^n)} \\
&= \langle L u, u \rangle_{L^2(\mathbb{R}^n)} + \langle c_2 E u, u \rangle_{L^2(\mathbb{R}^n)} \\
&= \langle L u, u \rangle_{L^2(\mathbb{R}^n)} + c_2 \|u\|^2_{L^2(\mathbb{R}^n)} \\
&\geq c_3 \|u\|^2_{H^1(\mathbb{R}^n)},
\end{aligned}$$

with $c_3 > 0$ as in Lemma 4.6. Hence, we can interpret any solution k as a solution

z with $c_2 = 0$.

The following a priori estimate of the weak solution is crucial for the proof of existence of an optimal control.

Lemma 4.13:
There exists a constant $\tilde{C} > 0$ *such that weak solution of (4.1) satisfies*

$$\|k\|_{W(0,T)} \leq \tilde{C}(\|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + \|k_0\|_{L^2(\mathbb{R}^n)} + 1). \tag{4.13}$$

Proof. We follow the proof of Theorem 3.13 by Tröltzsch (2005, p.121). First note that

$$\|k\|^2_{W(0,T)} = \|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} + \|k_t\|^2_{L^2(0,T;H^{-1}(\mathbb{R}^n))}.$$

For the first term, we have already proven in Lemma 4.12 that there exists a constant C > 0 such that the inequality

$$\|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} \leq \mathrm{C}\left(\|k_0\|_{L^2(\mathbb{R}^n)} + \|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + 1\right)^2$$

holds true. In order to estimate the second term, a bit more work has to be done. First, we define the linear functionals $F_i(t) : H^1(\mathbb{R}^n) \to \mathbb{R}$, $i = 1, ..., 5$ as

$$\begin{aligned}
F_1(t) :\ & v \mapsto \langle \alpha \nabla_x k(t), \nabla_x v \rangle_{L^2(\mathbb{R}^n)}, \\
F_2(t) :\ & v \mapsto \langle \beta \int_{\mathbb{R}^n} (k(y,t) - k(\cdot,t))\Gamma_\varepsilon(\cdot,y)dy, v \rangle_{L^2(\mathbb{R}^n)}, \\
F_3(t) :\ & v \mapsto \langle \delta k(t), v \rangle_{L^2(\mathbb{R}^n)}, \\
F_4(t) :\ & v \mapsto \langle \mathcal{P}(k)(t), v \rangle_{L^2(\mathbb{R}^n)}, \\
F_5(t) :\ & v \mapsto \langle c(t), v \rangle_{L^2(\mathbb{R}^n)}.
\end{aligned}$$

The weak formulation of the PIDE then yields

$$\|k_t\|_{L^2(0,T;H^{-1}(\mathbb{R}^n))} \leq \sum_{i=1}^{5} \|F_i\|_{L^2(0,T;H^{-1}(\mathbb{R}^n))}.$$

We estimate all summands separately. It holds

$$\begin{aligned}
|F_1(t)v| = |\alpha \langle \nabla_x k(t), \nabla_x v \rangle_{L^2(\mathbb{R}^n)}| &\leq \alpha \|\nabla_x k(t)\|_{L^2(\mathbb{R}^n)} \|\nabla_x v\|_{L^2(\mathbb{R}^n)} \\
&\leq \alpha \|k(t)\|_{H^1(\mathbb{R}^n)} \|v\|_{H^1(\mathbb{R}^n)},
\end{aligned}$$

hence

$$\|F_1\|^2_{L^2(0,T;H^{-1}(\mathbb{R}^n))} \leq \int_0^T \|F_1(t)\|^2_{H^{-1}(\mathbb{R}^n)} dt \leq \int_0^T \hat{c}\|k(t)\|^2_{H^1(\mathbb{R}^n)} dt = \hat{c}\|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))}.$$

for a constant $\hat{c} \geq 0$. Finally, from Lemma 4.12 we know that

$$\|F_1\|^2_{L^2(0,T;H^{-1}(\mathbb{R}^n))} \leq \hat{c}\|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} \leq \mathrm{C}\left(\|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + \|k_0\|_{L^2(\mathbb{R}^n)} + 1\right)^2.$$

with C > 0. For the other summands we proceed analogously and deduce

$$|F_2(t)v| \leq \beta\kappa\|k(t)\|_{H^1(\mathbb{R}^n)}\|v\|_{H^1(\mathbb{R}^n)},$$

$$|F_3(t)v| \leq \delta\|k(t)\|_{H^1(\mathbb{R}^n)}\|v\|_{H^1(\mathbb{R}^n)},$$

$$|F_4(t)v| \leq \|\mathcal{P}(k)(t)\|_{L^2(\mathbb{R}^n)}\|v\|_{L^2(\mathbb{R}^n)} \leq \hat{\mathrm{c}}(t)\|k(t)\|_{H^1(\mathbb{R}^n)}\|v\|_{H^1(\mathbb{R}^n)},$$

for contants which we have already derived in the proof of Lemma 4.12. For the last term, it holds

$$|F_5(t)v| \leq \|c(t)\|_{L^2(\mathbb{R}^n)}\|v\|_{H^1(\mathbb{R}^n)}.$$

Taking the maximum of all $t \in [0, T]$ and combining all estimates for $F_i,\ i = 1, ..., 5$ we get

$$\|k_t\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} \leq \hat{\mathrm{C}}\left(\|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + \|k_0\|_{L^2(\mathbb{R}^n)} + 1\right)^2,$$

and together with the a priori estimate in Lemma 4.12, we finally have

$$\begin{aligned}\|k\|^2_{W(0,T)} &= \|k\|^2_{L^2(0,T;H^1(\mathbb{R}^n))} + \|k_t\|^2_{L^2(0,T;H^{-1}(\mathbb{R}^n))} \\ &\leq \tilde{\mathrm{C}}\left(\|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} + \|k_0\|_{L^2(\mathbb{R}^n)} + 1\right)^2.\end{aligned}$$

□

With this inequality, we have amongst others shown the uniform boundedness of all solutions k in $W(0, T)$ for any $c \in \mathcal{U}_{ad}$, if $\mathcal{U}_{ad}$ is bounded and closed. This will be of importance in the proof of Theorem 4.19, when we derive the weak convergence of a sequence of states in the proof of the existence of an optimal control.

4.2 Existence of an Optimal Control

In this section, we finally have all at hand to prove the existence of an optimal control in the spatial Ramsey model with endogenous productivity growth. This optimal control and the corresponding state variable define the market equilibrium according to the second welfare theorem of economics.

Consider the following optimal control problem

$$\begin{aligned}\min_{k,c} \quad &\int_0^T \int_{\mathbb{R}^n} -U(c(x,t))e^{-\tau t - \gamma\|x\|_2^2}\, dxdt \\ &+ \frac{1}{2\rho_1}\|k(\cdot, T) - k_T(\cdot)\|^2_{L^2(\mathbb{R}^n)} + \frac{1}{2\rho_2}\|\min\{0, k\}\|^2_{L^2(0,T;L^2(\mathbb{R}^n))},\end{aligned} \tag{4.14}$$

$$
\begin{aligned}
\text{s.t.} \frac{\partial k}{\partial t} - \mathcal{L}(k) + \delta k - \mathcal{P}(k) &= -c \qquad \text{on } \mathbb{R}^n \times (0,T), \\
k(x,0) &= k_0(x) \quad \text{on } \mathbb{R}^n, \\
c &\in \mathcal{U}_{ad}.
\end{aligned}
\tag{4.15}
$$

Here, we have replaced the box constraint on the state variable in (3.16), $k \geq 0$ in $\mathbb{R}^n \times (0,T)$, by a Moreau-Yosida penalty function as in Pearson et al. (2011) and introduced the penalty parameters $\rho_1, \rho_2 > 0$.
Before we begin, we have to state some assumptions on the utility function U and the set of feasible controls $c \in \mathcal{U}_{ad}$. First note that we need the set of feasible controls to be bounded. Hence, we have to introduce a *maximal aggregated consumption level* $\overline{C}$. Although this might seem quite restrictive considering the economic interpretation, this assumption is not as prohibitive as one would think. Since the production function p is, according to Assumption 4.9, assumed to be bounded in order to guarantee the existence of a weak solution of the capital accumulation equation, an infinite consumption is never possible. We can justify the boundedness assumption of the production function economically, for example with the boundedness of space on earth. Thus, any utopian but finite upper bound on the consumption will do.
The assumption on p to be zero for a zero input is also economically driven. It states that the producing sector can only generate positive output, if it can devote any production factors. The assumption on the Lipschitz continuity is not in line with the neoclassical theory, since Lipschitz continuous functions do not satisfy the Inada condition, as $p'(x) \nrightarrow \infty, x \to 0$. However, we can choose the Lipschitz constant L_p arbitrarily large and so approximate the Inada condition in 0. Note that the boundedness assumption on p still is in line with the neoclassical theory.

We denote the set of feasible controls as $\mathcal{U}_{ad}$. The first assumption, which we need to make in order to prove the existence of an optimal control in the nonlocal spatial Ramsey model, is:

(1) *$\mathcal{U}_{ad}$ is a bounded, closed, and convex subset of* $L^2(0,T;L^2(\mathbb{R}^n))$.

For example, for a given maximal consumption function $c_{max} \in L^2(\mathbb{R}^n \times (0,T)) \cap L^\infty(\mathbb{R}^n \times (0,T))$ and a maximum aggregated consumption level $\overline{C} \in \mathbb{R}_+$, we consider

$$\mathcal{U}_{ad} := \{c \in L^2(0,T;L^2(\mathbb{R}^n)) : \; 0 \leq c \leq c_{max} \; \wedge \; \|c\|_{L^2(0,T;L^2(\mathbb{R}^n))} \leq \overline{C}\}.$$

The nonlocal diffusion operator $\mathcal{L}$ and the productivity-production operator $\mathcal{P}$ are defined as in Chapter 3.4. The assumptions concerning the Lipschitz continuity and the uniform boundedness of the production function p and the essential boundedness of the initial productivity distribution A_0 remain.

(2) *The production function p and the initial productivity distribution A_0 satisfy Assumption 4.9. As in (4.2), the diffusion operator is defined as*

$$\mathcal{L}(k)(x,t) := \alpha \Delta k(x,t) + \beta \int_{\mathbb{R}^n} (k(y,t) - k(x,t)) \Gamma_\varepsilon(x,y) dy,$$

for some weights $\alpha, \beta > 0$ and $\varepsilon > 0$ and the productivity-production operator is given as in (4.4) by

$$\mathcal{P}(k)(x,t) := A_0(x) \exp\left(\frac{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\mathbb{R}^n} \phi(k(y,t))\Gamma_\varepsilon(x,y)dy} \, t \right) \, p(k(x,t)).$$

Here, the necessary properties of ϕ motivate the next assumption:

(3) *The nominal function ϕ is (Lipschitz) continuous. We assume that $\phi(k) > 0$ for all k.*

In the following we assume that ϕ denotes a continuous approximation of the absolute value function. The example that we have in mind is

$$\phi(k) = \sqrt{k^2 + \eta},$$

which depends on a parameter η. We assume this parameter to be a priori defined and fixed, so it is convenient to omit the dependence of $\phi = \phi(\eta)$ on this parameter. However, for the proof of the existence of an optimal control, any continuous and positive function ϕ is sufficient.

Considering the objective function, we assume the following.

(4) *The utility function $U : \mathbb{R} \to \mathbb{R}$ is bounded and locally Lipschitz continuous, hence there exists a constant K, such that*

$$|U(0)| \leq K,$$

and a constant $L(M)$, such that for all $c_1, c_2 \in [-M, M]$

$$|U(c_1) - U(c_2)| \leq L(M)|c_1 - c_2|.$$

Moreover, we assume that U is concave.

It is worth to mention once more, that U is the utility function which describes the consuming sector in the Ramsey economy. Hence, the concavity is essential for the economic interpretation. The assumptions on U, together with the measurability of the function $(x,t) \mapsto e^{-\tau t - \gamma \|x\|_2^2}$ for $\tau, \gamma > 0$, are necessary in order to guarantee that the objective function $\mathcal{J}$ is convex, continuous, and bounded from below on $\mathcal{U}_{ad}$.

The following Lemma is crucial for the proof of existence of an optimal control. In contrast to the spatially bounded case, that we study in the following chapters, we are able to derive much stronger regularity statements on the weak solution of the nonlocal capital accumulation equation for controls in L^2:

Lemma 4.14:
We consider the case $n = 1$. Let L_p be the Lipschitz constant of the production function p and let the assumptions $(1) - (4)$ hold. Assume that the initial value function $k_0 \in L^2(\mathbb{R})$ is also Hölder continuous of exponent $\lambda > 0$ on $\mathbb{R}$. Let $\mathcal{U}_{ad}$ and k_0 be chosen such that there exists a constant $\theta > 0$ satisfying

$$\begin{aligned} &(i)\;\; 4L_p^2 \leq \theta \text{ and} \\ &(ii)\; |k_0(x)| + \int_0^T |c(x,s)|^2 ds < \frac{1}{16 L_p T e^{\theta T/2}} \quad \forall c \in \mathcal{U}_{ad}. \end{aligned} \tag{4.16}$$

Then, the weak solution of the spatial nonlocal Ramsey model (4.15) is bounded and continuous on $\mathbb{R} \times [0,T]$ for every $c \in \mathcal{U}_{ad}$.

Proof. The boundedness of the weak solution is a direct application of Theorem 3.1 by Ran and Zhang (2010, p.956). We can interpret the nonlocal diffusion term as compact perturbation of the right-hand side and use the boundedness of the productivity operator $\mathcal{P}$. By assumption, k_0 and $c \in \mathcal{U}_{ad}$ are chosen such that (4.16) is satisfied. Hence, the theorem mentioned above yields that there exists a constant $0 < M(c) := M < \infty$, such that $\|k\|_{L^\infty(\mathbb{R}\times[0,T])} \leq M$.

In order to prove the continuity of the weak solution on $\mathbb{R} \times [0,T]$, a bit more work has to be done. We want to apply a result from Ladyženskaya et al. (1968), where the authors prove the local Hölder continuity of any essentially bounded weak solution of a quasi-linear differential equation of parabolic type on $\Omega \times [0,T]$. Although Ladyženskaya et al. (1968) assumed a bounded and open domain $\Omega \subset \mathbb{R}^n$, we can adapt the statement to our case. Since the result does not require any boundary conditions on $\partial\Omega$, we can apply Theorem 1.1 in Ladyženskaya et al. (1968, p.419) also to the spatially unbounded case and get the Hölder continuity of exponent $\lambda_1 > 0$ of the weak solution on all compact subsets $K \times [t_1, t_2] \Subset \mathbb{R} \times (0,T)$ with $K \subset \mathbb{R}$ compact and $0 < t_1 < t_2 < T$. The Hölder constant λ_1 depends only on M and the coercivity constant of the bilinear form (4.10), $c_{coer} > 0$.
Since the initial value function k_0 is assumed to be Hölder continuous of exponent λ, we can apply the extended result of Theorem 1.1 in Ladyženskaya et al. (1968, p.419) and get the local Hölder continuity of exponent $\lambda_2 > 0$ depending on M, c_{coer} and λ of the weak solution also in $t = 0$. Remark that we can extend our left-hand sides c and $\mathcal{P}$ to $[0, T + \varepsilon]$ for any $\varepsilon > 0$, so we can assume the local Hölder continuity of the weak solution also in T without any loss of generality. Note that we have all assumptions in Ladyženskaya et al. (1968, p.418) fulfilled since the solution k is bounded and the constants α and β in (4.2) are positive.

Hence, we have

$$k \in \mathcal{C}^{\lambda',\lambda'/2}(K \times [0,T]) \quad \forall K \Subset \mathbb{R},$$

with some positive λ' depending only on M, T, c_{coer}, and λ. Every Hölder continuous function is also continuous, hence we have

$$k \in \mathcal{C}(K \times [0,T]) \quad \forall K \Subset \mathbb{R}.$$

Considering the exhaustion by compact sets of $\mathbb{R}$, meaning a sequence of compact sets $\{K_m\}_{m\in\mathbb{N}}$ with $K_m \subset \mathring{K}_{m+1}$ and $\bigcup_{m\in\mathbb{N}} K_m = \mathbb{R}$, we can extend the local result to the whole unbounded $\mathbb{R}$ due to the locality of continuity. This means, for every $x_0 \in \mathbb{R}$, there exists a m large enough such that $(x_0,t) \in \mathring{K}_m \times [0,T]$, where $\mathring{K}$ denotes the interior of a set K.

□

Remark 4.15:
The fundamental result from Ladyženskaya et al. (1968) can also be found in a slightly different version in DiBenedetto (1993, p.41). Moreover, in Remark 1.1, p.17, DiBenedetto (1993) points out that, in the L^2 setting, even a locally bounded weak solution of a non-degenerate parabolic differential equation is locally Hölder continuous. Hence, the assumptions on k_0 and $\mathcal{U}_{ad}$, that guarantee the global (essential) boundedness of the weak solution in the spatial Ramsey model with endogenous productivity growth, may already be too strong. However, we do not have any restrictions on the choice of our feasible controls or the initial capital distribution for application reasons. Thus, we do not study how we could weaken the assumptions on c or k_0 in this thesis, but leave that question to further research.

Before we can state the most important result in this chapter, namely the existence result of an optimal control, we show that the nonlocal diffusion operator is continuous from $L^2(0,T;L^2(\mathbb{R}))$ to $L^2(0,T;L^2(\mathbb{R}))$. Therefore, we need the following proposition:

Proposition 4.16 (Young's Inequality for Convolution)**:**
Let $p,q,r \geq 1$. If $(1/p) + (1/q) = 1 + (1/r)$ and if $u \in L^p(\mathbb{R}^n)$ and $v \in L^q(\mathbb{R}^n)$, then $u * v \in L^r(\mathbb{R}^n)$ and

$$\|u * v\|_{L^r(\mathbb{R}^n)} \leq \|u\|_{L^p(\mathbb{R}^n)} \|v\|_{L^q(\mathbb{R}^n)}.$$

The result and proof can for example be found in Adams and Fournier (2003, p.34).

Remark 4.17:
Note that, as special case, we have for $u \in L^1(\mathbb{R}^n)$ and $v \in L^2(\mathbb{R}^n)$

$$\|u * v\|_{L^2(\mathbb{R}^n)} \leq \|u\|_{L^1(\mathbb{R}^n)} \|v\|_{L^2(\mathbb{R}^n)}.$$

Lemma 4.18:
The nonlocal diffusion operator

$$\mathcal{NL}(\cdot)(x,t): \; k \mapsto \int_{\mathbb{R}} (k(y,t) - k(x,t))\Gamma_\varepsilon(x,y)dy$$

is continuous from $L^2(0,T;L^2(\mathbb{R}))$ *to* $L^2(0,T;L^2(\mathbb{R}))$.

Proof. We proceed as in Lemma 4.10. Since $L^2(0,T;L^2(\mathbb{R}))$ is a vector space, we can estimate the two terms separately. To estimate the first term, we apply Young's inequality for convolution and recall that the kernel function Γ_ε indeed has the form $\Gamma_\varepsilon(x,y) = \Gamma_\varepsilon(x-y)$. Since

$$\int_{\mathbb{R}} \Gamma_\varepsilon(x)dx = 1,$$

we can deduce

$$\begin{aligned}
\left\| \int_{\mathbb{R}} k(y,t)\Gamma_\varepsilon(x-y)dy \right\|^2_{L^2(0,T;L^2(\mathbb{R}))} &= \int_0^T \|k(\cdot,t) * \Gamma_\varepsilon(\cdot - y)\|^2_{L^2(\mathbb{R})}dt \\
&\leq \int_0^T \|k(\cdot,t)\|^2_{L^2(\mathbb{R})}\|\Gamma_\varepsilon\|^2_{L^1(\mathbb{R})}dt \\
&= \|k\|^2_{L^2(0,T;L^2(\mathbb{R}))},
\end{aligned}$$

The estimation of the second term yields

$$\begin{aligned}
\left\| \int_{\mathbb{R}} k(x,t)\Gamma_\varepsilon(x,y)dy \right\|^2_{L^2(0,T;L^2(\mathbb{R}))} &= \int_0^T \int_{\mathbb{R}} \left(\int_{\mathbb{R}} k(x,t)\Gamma_\varepsilon(x,y)dy \right)^2 dxdt \\
&= \int_0^T \int_{\mathbb{R}} k(x,t)^2 \left(\int_{\mathbb{R}} \Gamma_\varepsilon(x,y)dy \right)^2 dxdt \\
&\leq \|k\|^2_{L^2(0,T;L^2(\mathbb{R}))},
\end{aligned}$$

again exploiting $\int_{\mathbb{R}} \Gamma_\varepsilon(x,y)dy = 1$ for all $x \in \mathbb{R}$ by definition.

□

In particular, this yields $\mathcal{NL}(k) \in L^2(0,T;H^{-1}(\mathbb{R}))$ for every $k \in L^2(0,T;L^2(\mathbb{R}))$ with the same arguments as in Lemma 4.10.

Now, we finally have all at hand to state the following theorem:

Theorem 4.19:
Consider the case $n = 1$. *Let the assumptions* $(1)-(4)$ *hold. Moreover, assume that* $k_0 \in L^2(\mathbb{R}) \cap \mathcal{C}^\lambda(\mathbb{R})$ *for some* $\lambda > 0$ *and* $\mathcal{U}_{ad}$ *fulfill the assumptions of Lemma 4.14, hence are given such that the weak solution of the capital accumulation equation is bounded and continuous. Then, there exist an optimal control* $\overline{c} \in \mathcal{U}_{ad}$ *and a*

corresponding optimal state $\overline{k} \in W(0,T)$ *of the spatial nonlocal Ramsey model (4.14) and (4.15).*

Proof. As already shown in Theorem 4.11, the state equation has a unique weak solution $k := k(c) \in W(0,T)$ for every control $c \in \mathcal{U}_{ad}$ and $k_0 \in L^2(\mathbb{R})$. The additional assumptions on the initial value function k_0 then yield, according to Lemma 4.14, the uniform boundedness of this weak solution, i.e. the existence of a constant $0 < M := M(c) < \infty$ such that

$$\|k\|_{L^\infty(\mathbb{R}\times[0,T])} \leq M,$$

for all states corresponding to a control $c \in \mathcal{U}_{ad}$ and the continuity of k, i.e. $k \in C(\mathbb{R} \times [0,T])$.

Due to the boundedness of $\mathcal{U}_{ad}$, the uniform boundedness of k in $W(0,T)$ according to Lemma 4.13, and the assumption on the objective, there exists a finite infimum J_{inf} of $\mathcal{J}$. Since $L^2(0,T;L^2(\mathbb{R}))$ is reflexive, we can choose a minimizing sequence $(c_m)_{m\in\mathbb{N}}$ in $\mathcal{U}_{ad}$ that has a weak convergent subsequence $(c_{m_j})_{j\in\mathbb{N}}$ with limit $\overline{c} \in L^2(0,T;L^2(\mathbb{R}))$. Without any loss of generality, we can identify this subsequence with $(c_m)_{m\in\mathbb{N}}$. Assumption (1) states that $\mathcal{U}_{ad}$ is closed and convex, thus weakly sequentially closed, which guaranties that $\overline{c} \in \mathcal{U}_{ad}$. Hence, we get

$$c_m \rightharpoonup \overline{c} \in \mathcal{U}_{ad},\ m \to \infty.$$

This sequence of controls defines a sequence of corresponding states $(k_m)_{m\in\mathbb{N}} := (k(c_m))_{m\in\mathbb{N}}$. We define

$$\rho_m := \mathcal{P}(k_m) \quad \text{and} \quad \kappa_m := \int_{\mathbb{R}} (k_m(y,t) - k_m(x,t))\Gamma_\varepsilon(x,y)dy.$$

As already shown in the Lemmas 4.10 and 4.18, ρ_m and κ_m are elements of $L^2(0,T;L^2(\mathbb{R}))$ for $k_m \in L^2(0,T;L^2(\mathbb{R}))$. Moreover, the continuity of $\mathcal{P}$ and $\mathcal{NL}$ in $L^2(0,T;L^2(\mathbb{R}))$ yields the uniform boundedness of the sequences due to the uniform boundedness of $(k_m)_{m\in\mathbb{N}}$ in $W(0,T)$. Here, we have applied Lemma 4.13 and recalled that $\mathcal{U}_{ad}$ is assumed to be bounded in $L^2(0,T;L^2(\mathbb{R}))$. Hence, we can assume that there exist some subsequences, again denoted by $(\rho_m)_{m\in\mathbb{N}}$ and $(\kappa_m)_{m\in\mathbb{N}}$ without any loss of generality, that converge weakly to some $\overline{\rho}$ and $\overline{\kappa}$ in $L^2(0,T;L^2(\mathbb{R}))$.

Now, we consider the linear parabolic initial value problems given by

$$\begin{aligned} \frac{\partial k_m}{\partial t} - \alpha\Delta k_m + \delta k_m &= p_m + \kappa_m - c_m \quad && \text{on } \mathbb{R}\times(0,T] \\ k_m(\cdot,0) &= k_0 && \text{on } \mathbb{R} \end{aligned}$$

for $m \in \mathbb{N}$. We know that the right-hand side converges weakly towards $\overline{p} +$

$\overline{\kappa} - \overline{c}$ in $L^2(0,T;L^2(\mathbb{R}))$, hence also in $L^2(0,T;H^{-1}(\mathbb{R}))$ since the embedding $L^2(0,T;L^2(\mathbb{R})) \hookrightarrow L^2(0,T;H^{-1}(\mathbb{R}))$ is continuous. Due to the continuity of the solution mapping, that maps a right-hand side and an initial value function to the solution of a linear parabolic differential equation (c. Wloka, 1982, p.382), this mapping is also weakly continuous from $L^2(0,T;H^{-1}(\mathbb{R})) \times L^2(\mathbb{R})$ to $W(0,T)$. Thus we get the weak convergence of the left-hand side as well. We have

$$k_m \rightharpoonup \overline{k} \text{ in } W(0,T),\ m \to \infty.$$

Moreover, the continuity of the solution mapping guarantees that $\overline{k} \in W(0,T)$. With the same arguments as used in the proof of Lemma 4.14, we get

$$k_m \rightharpoonup \overline{k} \text{ in } C^{\lambda',\lambda'/2}(K \times [0,T]),\ m \to \infty,$$

for all compact subsets K of $\mathbb{R}$ and an $\lambda' > 0$, depending on $M,\ T,\ coer$, and λ. It is true that

$$C^{\lambda',\lambda'/2}(K \times [0,T]) \hookrightarrow^c C(K \times [0,T]),$$

(cf. Adams and Fournier, 2003, p.12), thus we get the strong convergence of the sequence of states in the space of continuous functions on all compact subsets $K \times [0,T]$ of $\mathbb{R} \times [0,T]$.

Now, we need to show the convergence of the integrals in the weak formulation. So far, we have derived

$$\begin{aligned} &(i) \quad c_m \rightharpoonup \overline{c} \text{ in } L^2(0,T;L^2(\mathbb{R})), \\ &(ii) \ \ k_m \rightharpoonup \overline{k} \text{ in } W(0,T), \\ &(iii)\ k_m \to \overline{k} \text{ in } C(K \times [0,T]) \text{ for all } K \Subset \mathbb{R}. \end{aligned} \tag{4.17}$$

Due to the a priori estimates in (4.12) and (4.13) and the weak (or weak star) compactness of unit balls in the spaces $L^2(0,T;H^1(\mathbb{R}))$, $L^\infty(0,T;L^2(\mathbb{R}))$, and $L^2(0,T;H^{-1}(\mathbb{R}))$, we can adapt the arguments by Dautray and Lions (1992, p.515) and extract a subsequence $(k_{\overline{m}})_{\overline{m}\in\mathbb{N}}$ with

$$\begin{aligned} &(i) \ \ k_{\overline{m}} \rightharpoonup \overline{k} \text{ in } L^2(0,T;H^1(\mathbb{R})), \\ &(ii)\ k_{\overline{m}} \rightharpoonup^* \overline{k} \text{ in } L^\infty(0,T;L^2(\mathbb{R})). \end{aligned} \tag{4.18}$$

Note that the embedding $W(0,T) \hookrightarrow L^2(0,T;H^1(\mathbb{R}))$ is continuous, hence weakly continuous, which guarantees that the subsequence $(k_{\overline{m}})_{\overline{m}\in\mathbb{N}}$ converges to the same limit in $L^2(0,T;H^1(\mathbb{R}))$ as $(k_m)_{m\in\mathbb{N}}$ in $W(0,T)$.

For now, we choose the test functions $\psi := \varphi \otimes v$ for $\varphi \in \mathcal{C}_0^\infty([0,T[)$ with $\varphi(0) \neq 0$ and $v \in \mathcal{C}_0^\infty(\mathbb{R})$. We derive the weak formulation of the capital accumulation equation as

$$
\begin{aligned}
&-\int_0^T \langle k_{\overline{m}}(t), \psi_t(t)\rangle_{L^2(\mathbb{R})} dt + \int_0^T \mathbf{a}(k_{\overline{m}}(t), \psi(t))\ dt - \int_0^T \langle \mathcal{P}(k_{\overline{m}})(t), \psi(t)\rangle_{L^2(\mathbb{R})} dt \\
&= -\int_0^T \langle c_{\overline{m}}(t), \psi(t)\rangle_{L^2(\mathbb{R})} dt + \langle k_0, \psi(0)\rangle_{L^2(\mathbb{R})}, \\
&\psi = \varphi \otimes v,\ \forall \varphi \in \mathcal{C}_0^\infty([0,T[),\ \varphi(0) \neq 0,\ v \in \mathcal{C}_0^\infty(\mathbb{R}),
\end{aligned}
$$

with the definition of the bilinear form in 4.10.
From (4.17)(i), we can deduce

$$
\int_0^T \langle c_{\overline{m}}(t), \psi(t)\rangle_{L^2(\mathbb{R})} dt \to \int_0^T \langle \overline{c}(t), v\rangle_{L^2(\mathbb{R})} \varphi(t)\ dt,\ m \to \infty
$$

and from (4.18)(i), we get

$$
\int_0^T \langle k_{\overline{m}}(t), \psi'_{\overline{m}}(t)\rangle_{L^2(\mathbb{R})} dt \to \int_0^T \langle \overline{k}(t), \psi'(t)\rangle_{L^2(\mathbb{R})} dt,\ m \to \infty.
$$

We can rewrite the bilinear form $\mathbf{a}(k,\psi)$ in verctorial form as $\langle Ak, \psi\rangle_{H^1(\mathbb{R})}$ with $Ak(\cdot) \in L^2(0,T;H^{-1}(\mathbb{R}))$ (cf. Dautray and Lions, 1992, p.515), hence (4.18)(i) implies

$$
\int_0^T \mathbf{a}(k_{\overline{m}}(t), \psi(t)) dt \to \int_0^T \mathbf{a}(\overline{k}(t), \psi(t)) dt \quad \text{for } \overline{m} \to \infty.
$$

So far, we were able to use the same arguments as Dautray and Lions (1992, p.515). The convergence of the nonlinear productivity term needs some further analysis. We exploit the strong convergence of the sequence of states on compact sets, (4.17)(iii) and the properties of the kernel function Γ_ε, respectively Γ_μ. We start with the exponential term. Due to the boundedness of every $k_{\overline{m}}$ in $L^\infty(\mathbb{R} \times [0,T])$, the continuity of ϕ, and the property of Γ_ε to be decreasing for large absolute values of input variables, we can choose a radius $R > 0$ large enough such that

$$
\int_{\mathbb{R}\setminus\mathcal{B}_R(0)} \phi(k_{\overline{m}}(y,t)) \Gamma_\varepsilon(x,y) dy \leq \tilde{\varepsilon}/4
$$

for all $\overline{m} \in \mathbb{N}$ and an $\tilde{\varepsilon} > 0$. l Then it is also

$$
\int_{\mathbb{R}\setminus\mathcal{B}_R(0)} \phi(k_{\overline{m}}(y,t)) \Gamma_\mu(x,y) dy \leq \tilde{\varepsilon}/4
$$

for $\mu < \varepsilon$. With this, we get

$$
\begin{aligned}
&\left|\int_{\mathbb{R}} \phi(k_{\overline{m}}(y,t))\Gamma_\varepsilon(x,y)dy - \int_{\mathbb{R}} \phi(\overline{k}(y,t))\Gamma_\varepsilon(x,y)dy\right| \\
\leq &\left|\int_{\mathcal{B}_R(0)} (\phi(k_{\overline{m}}(y,t)) - \phi(\overline{k}(y,t)))\Gamma_\varepsilon(x,y)dy\right| \\
&+ \left|\int_{\mathbb{R}\setminus\mathcal{B}_R(0)} \phi(k_{\overline{m}}(y,t))\Gamma_\varepsilon(x,y)dy\right| + \left|\int_{\mathbb{R}\setminus\mathcal{B}_R(0)} \phi(\overline{k}(y,t))\Gamma_\varepsilon(x,y)dy\right| \\
\leq &\left|\int_{\mathcal{B}_R(0)} (\phi(k_{\overline{m}}(y,t)) - \phi(\overline{k}(y,t)))\Gamma_\varepsilon(x,y)dy\right| + \tilde{\varepsilon}/2.
\end{aligned}
$$

By assumption, ϕ is continuous, hence there exists a $N \in \mathbb{N}$ such that if $|k_{\overline{m}} - \overline{k}| \leq \delta$ for $\overline{m} \geq N$, it is $|\phi(k_{\overline{m}}(y,t)) - \phi(\overline{k}(y,t))| \leq \tilde{\varepsilon}(N)$. This yields

$$
\int_{\mathcal{B}_R(0)} |(\phi(k_{\overline{m}}(y,t)) - \phi(\overline{k}(y,t)))|\Gamma_\varepsilon(x,y)dy \leq \tilde{\varepsilon}(N) \int_{\mathcal{B}_R(0)} \Gamma_\varepsilon(x,y)dy.
$$

We can choose N large enough such that $\tilde{\varepsilon}(N) \leq \tilde{\varepsilon}/2$ and end up with

$$
\left|\int_{\mathbb{R}} \phi(k_{\overline{m}}(y,t))\Gamma_\varepsilon(x,y)dy - \int_{\mathbb{R}} \phi(\overline{k}(y,t))\Gamma_\varepsilon(x,y)dy\right| \leq \tilde{\varepsilon},
$$

since $\int_{\mathcal{B}_R(0)} \Gamma_\varepsilon(x,y)dy \leq 1$.
According to (4.6), the exponential term is bounded by

$$
\exp\left(\frac{\int_{\mathbb{R}} \phi(k_{\overline{m}}(y,t))\Gamma_\mu(x,y)dy}{\int_{\mathbb{R}} \phi(k_{\overline{m}}(y,t))\Gamma_\varepsilon(x,y)dy} t\right) \leq e^{\frac{T\varepsilon}{\mu}}
$$

for all $k_{\overline{m}}$. Hence, we can exploit the property of the chosen test function v having compact support on $\mathbb{R}$. We can finally show the convergence

$$
\int_0^T \langle \mathcal{P}(k_{\overline{m}})(t), \psi(t)\rangle_{L^2(\mathbb{R})} dt \to \int_0^T \langle \mathcal{P}(\overline{k})(t), \psi(t)\rangle_{L^2(\mathbb{R})} dt, \ \overline{m} \to \infty,
$$

since $k_{\overline{m}} \to \overline{k}$ strongly on all compact sets $K \times [0,T]$. Combining all limits for $\overline{m} \to \infty$ in the weak formulation,we obtain

$$
\begin{aligned}
&- \int_0^T \langle \overline{k}(t), v\rangle_{L^2(\mathbb{R})} \varphi'(t)dt + \int_0^T \mathbf{a}(\overline{k}(t), v)\varphi(t)dt - \int_0^T \langle \mathcal{P}(\overline{k})(t), v\rangle_{L^2(\mathbb{R})}\varphi(t)dt \\
&= \langle k_0, \psi(0)\rangle_{L^2(\mathbb{R})} - \langle \overline{c}(t), v\rangle_{L^2(\mathbb{R})}\varphi(t)dt.
\end{aligned}
$$

Since this equality has to hold for every $\varphi \in \mathcal{C}_0^\infty([0,T[)$ with $\varphi(0) \neq 0$ and $v \in \mathcal{C}_0^\infty(\mathbb{R})$, we have finally shown that $\overline{k} \in W(0,T)$ is indeed a weak solution of the capital accumulation equation in the nonlocal spatial Ramsey model with endogenous productivity growth.

The only thing left to show in order to finish this proof of existence of an optimal control is the optimality of $(\overline{c}, \overline{k})$. But this follows immediately from the convexity and continuity of the objective function: Recall that every continuous and convex function is lower semicontinous. Hence, for

$$\begin{aligned} J(c,k) &= \int_0^T \int_{\mathbb{R}} -U(c(x,t))e^{-\tau t - \gamma x^2}\ dxdt \\ &+ \frac{1}{2\rho_1}\|k(\cdot,T) - k_T(\cdot)\|^2_{L^2(\mathbb{R})} + \frac{1}{2\rho_2}\|\min\{0,k\}\|^2_{L^2(0,T;L^2(\mathbb{R}))} \\ &:= F(c) + Q(k) \end{aligned}$$

it follows

$$J_{inf} = \lim_{\overline{m}\to\infty} J(c_{\overline{m}}, k_{\overline{m}}) = \lim_{\overline{m}\to\infty} F(c_{\overline{m}}) + \lim_{\overline{m}\to\infty} Q(k_{\overline{m}}) \geq F(\overline{c}) + Q(\overline{k}) = J(\overline{c}, \overline{k}).$$

Since J_{inf} was the infimum of $\mathcal{J}$, we get the equality. □

The proof of existence of an optimal control is crucial, not only for the mathematical study. From an economic point of view, this means that there exists a competitive market equilibrium in the closed spatialized Ramsey economy, where households may be heterogeneous in their initial capital distribution, productivity is heterogeneous in space and time, and no interaction with the surrounding takes place. Due to its complexness, the nonlocal spatial Ramsey model with endogenous productivity growth is quite general. Brito (2001, 2004, 2012), Boucekkine et al. (2009, 2013), and Camacho et al. (2008) admit that their spatial versions of the Ramsey model are not well-posed in the sense of Hadamard, at least if they consider a quite general, convex utility function and no further restrictions on the set of interest. As already mentioned, all approaches analyzing the (local) model with respect to existence of an optimal control are based on the theory of classical solutions. We considered a weaker notion of solution and were able to proof not only the existence of a weak solution of the capital accumulation equation but also the existence of an optimal control. Since our model is very general, we can capture the dynamics of the local model by Brito as a special case (for example by setting $\beta = 0$, $\mu = \varepsilon$). Hence, we have enhanced the economic theory on the spatial Ramsey model and provided not only a proof of existence for our model, but also for the common spatial Ramsey models.

CHAPTER 5

The Nonlocal Spatial Ramsey Model on Bounded Spatial Domains

'Freedom of movement is a fundamental characteristic of human beings and human values [...]. Barriers of movement [defined as discontinuities in the interaction between two countries] may concern people, goods, capital but also ideas, cultural standards, regulations, or intangible items.' (Topaloglou and Petrakos, 2008, p.3)

Economic geography depends not only on space, but also on the spreading behavior of production factors between several disjoint economies. In times of globalization and international trade agreements, it is also important in geographic economics to consider cross-border dependencies of production factors and economic welfare. By restricting the spatial domain of interest to a bounded domain $\Omega \subset \mathbb{R}^n$, we have naturally defined a border of an economy. Such a bounded economy could be seen as a country or a trade association. In contrast to the previous chapter, where we considered an unbounded spatial domain, we now have to introduce some boundary conditions in order to make the Ramsey problem well defined. As already mentioned in the introduction to the nonlocal spatial Ramsey model in Section 3.4, these boundary constraints do not only act on the surface of the domain Ω, but on a non-zero volume, the so called interaction, domain $\Omega_{\mathcal{I}}$. We refer to these constraints acting on the interaction domain as *volume constraints*. Whenever we understand the domain Ω as a bounded economy, $\Omega_{\mathcal{I}}$ can be interpreted either as a border area or region, where the central planner controls the production factor distribution, or as a trade-off set of production factors, where the exchange of production factors takes place naturally and independently. The meaning of the interaction domain $\Omega_{\mathcal{I}}$ hereby depends on the type of the volume constraints. A Dirichlet-type volume constraint, where the state variable is fixed on the interaction domain, corresponds to the central planner, who forces the production factor distribution to some value. By considering Neumann-type volume constraints, we assume that the exchange of production factors between Ω and $\Omega_{\mathcal{I}}$ follows some balancing law. Here, we only allow capital and labor to leave the economy Ω and be replaced by the production factors in $\Omega_{\mathcal{I}}$. Moreover, we assume that neither production goods leave the economy nor consumption goods can be traded in the interaction set and that there is no production in $\Omega_{\mathcal{I}}$, hence $A_0 = 0$ on $\Omega_{\mathcal{I}}$.

We are not the first who restrict the spatial domain to a bounded set. For example, Boucekkine et al. (2013) define the spatial domain as the unit ball and Aldashev et al. (2014), following Boucekkine et al. (2013), consider the parameterized circle as the interval $[0, 2\pi]$. They do not define any boundary conditions but interpret the unit circle as the global economy. Brock et al. (2014) and Brock et al. (2013) consider a nonlocal model on an (arbitrary) compact interval. In their model, they do not consider any diffusion effects of the state variable but only time dependent spillover effects and thus do not need to define any boundary conditions. A model close to our setting is described in Aniţa et al. (2013). Here, the authors consider a bounded space domain and introduce homogeneous Neumann boundary conditions. In the optimal control problem, they restrict the time line to a finite time horizon. Although their model is nonlocal as well, the quality of the nonlocality is different. Instead of considering nonlocal diffusion effects to model the mobility of capital, they include a pollution function in the capital accumulation equation. This pollution function is modeled as a partial differential equation with an integral term depending only on the capital function as right-hand side. However, the works listed above show that considering a bounded spatial domain and a finite time horizon is convenient for the economic application.

This chapter is organized as follows: We start with a short introduction to the nonlocal vector calculus developed by Du et al. (2012a) in Section 5.1. This theory provides all tools to analyze the nonlocal capital accumulation equation with pure nonlocal diffusion and Dirichlet-type volume constraints with respect to the existence of weak solutions and the regularity of these solutions. The results are stated and proved in Section 5.2. We conclude this chapter with an overview of the difficulties which arise when we consider the existence of an optimal control in this setting in Section 5.3.

5.1 A Nonlocal Vector Calculus

The fundamental theorem of calculus combines the concepts of the differential calculus and the integral calculus (cf. Elstrodt, 2005, p.304). Du, Gunzburger, Lehoucq, and Zhou introduce a new, nonlocal vector calculus. This theory aims at defining an analogon to the well known vector calculus for differential equations. In their papers, Du et al. (2012b,a, 2014) and Gunzburger and Lehoucq (2010) derive a notion of nonlocal divergence and gradient operators and some fundamental relationships between the nonlocal operators and their derivatives. They are able to mimic the classical differential calculus to the framework of nonlocal operators and they prove identities like the Gaussian theorem or Green's identities for nonlocal diffusion equations. In that way, they make it possible to use the techniques of the analysis of common differential equations in the context of partial integro-differential equations.

In this section, we introduce the nonlocal vector calculus developed by Du et al. (2012a,b) and Gunzburger and Lehoucq (2010). We follow the notation in Du et al. (2012a,b) and D'Elia et al. (2014), and denote by $\Omega \subset \mathbb{R}^n$ an open and bounded domain with sufficiently smooth boundary. Throughout this chapter, we assume Ω to be a Lipschitz domain.
The nonlocal vector calculus exploits the form of the nonlocal divergence and gradient operators. It is crucial to understand the *nonlocal diffusion operator*, commonly defined for a function $u : \mathbb{R}^n \to \mathbb{R}$ as

$$\mathcal{NL}u(x) := \int_{\mathbb{R}^n} (u(y) - u(x))\gamma(x, y)dy \qquad \text{for } x \in \Omega, \tag{5.1}$$

where Ω has nonzero volume and γ denotes a nonnegative and symmetric kernel function, as a composition of divergence and gradient operators, like in the local case. In the following, we consider the two vector mappings $\nu,\ \alpha :\ \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$, with α antisymmetric. The *nonlocal divergence operator* $\mathcal{D}$ on ν, $\mathcal{D}(\nu) : \mathbb{R}^n \to \mathbb{R}$, is then defined as

$$\mathcal{D}(\nu)(x) := \int_{\mathbb{R}^n} (\nu(x, y) + \nu(y, x))^T \alpha(x, y)dy \qquad \text{for } x \in \mathbb{R}^n \tag{5.2}$$

(c. Du et al., 2012b, p.10).
For a given mapping $u : \mathbb{R}^n \to \mathbb{R}$, Du et al. (2012b) derive the the adjoint operator $\mathcal{D}^*$ corresponding to $\mathcal{D}$ with respect to the standard L^2 duality pairing as

$$\mathcal{D}^*(u)(x, y) = -(u(y) - u(x))\alpha(x, y) \qquad \text{for } x, y \in \mathbb{R}^n. \tag{5.3}$$

The function $\mathcal{D}^*(u)$ maps from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^m$. The nonlocal adjoint operator $-\mathcal{D}^*$ can be interpreted as *nonlocal gradient operator*. Now, if the kernel function γ in (5.1) is given as

$$\gamma = \alpha^T(\Theta\alpha),$$

for a symmetric second order tensor Θ, the nonlocal diffusion operator in (5.1) can be represented as

$$\mathcal{NL}(u) = -\frac{1}{2}\mathcal{D}(\Theta\mathcal{D}^* u).$$

'*Thus, the operator $\mathcal{NL}$ is a composition of nonlocal divergence and gradient operators so that if Θ is the identity tensor, $\mathcal{NL}$ can be interpreted as a nonlocal Laplacian operator.[...] if Θ is also positive definite, the operator $-\mathcal{NL}$ is nonnegative*' (Du et al., 2012b, p.676).

Below, Θ is always assumed to be the identity tensor thus we can skip it in the notation.

The operator $\mathcal{NL}$ is called *'nonlocal because the value of $\mathcal{NL}(u)$ at a point x requires information about u at points $y \neq x$; this should be contrasted with local operators, e.g. the value of Δu at a point x requires information about u only on x'* (Du et al., 2012a, p.2). Due to this nonlocal character of $\mathcal{NL}$, it is not sufficient to consider boundary conditions that only act on the boundary $\partial\Omega$ of the set of interest, which is only a surface in $\mathbb{R}^n$. Instead and as already mentioned before, we have to introduce so called *volume constraints* which act on an *interaction domain* with nonzero volume. This interaction domain, denoted by $\Omega_{\mathcal{I}} \subset \mathbb{R}^n$, is the natural nonlocal extension of the surface-boundary of Ω. Throughout this thesis, we require that $\Omega_{\mathcal{I}} \cap \Omega = \emptyset$. Du et al. (2012b) define the interaction domain as

$$\Omega_{\mathcal{I}} := \{y \in \mathbb{R}^n \backslash \Omega : \ \alpha(x, y) \neq 0 \text{ for some } x \in \Omega\},$$

'so that $\Omega_{\mathcal{I}}$ consists of those points outside of Ω that interact with points in Ω' (Du et al., 2012b, p.16). Note that there is no assumption made about the geometric relation between the two sets Ω and $\Omega_{\mathcal{I}}$. The figure below is taken from Du et al. (2012b, p.16) and illustrates some possible constructions of Ω and $\Omega_{\mathcal{I}}$.

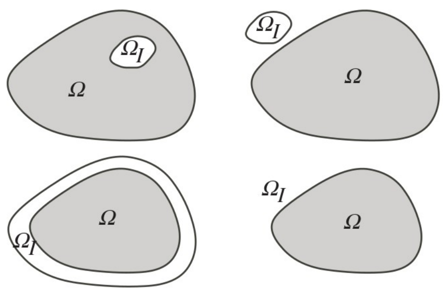


Figure 5.1: Configurations for Ω and $\Omega_{\mathcal{I}}$

The interaction of points in the domain of interest Ω with points in the interaction domain $\Omega_{\mathcal{I}}$ is modeled by a so called *nonlocal interaction operator* $\mathcal{V}$, an analogon to the local flux operator $\partial u / \partial \overrightarrow{n}$. For a function $\nu : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$, it is defined as

$$\mathcal{V}(\nu)(x) := -\int_{\Omega \cup \Omega_{\mathcal{I}}} (\nu(x, y) + \nu(y, x))^T \alpha(x, y) dy,$$

where x is an element in $\Omega_{\mathcal{I}}$. This operator can be interpreted as nonlocal flux from Ω into $\Omega_{\mathcal{I}}$.

For two functions $u, v : \mathbb{R}^n \to \mathbb{R}$ and $\nu : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$, Du et al. (2012a) are able to prove the *nonlocal Gauss' theorem*

$$\int_{\Omega} \mathcal{D}(\nu)\ dx = \int_{\Omega_{\mathcal{I}}} \mathcal{V}(\nu)\ dx,$$

the *nonlocal integration by parts formula*,

$$\int_{\Omega} u\mathcal{D}(\nu)\ dx - \int_{\Omega\cup\Omega_{\mathcal{I}}}\int_{\Omega\cup\Omega_{\mathcal{I}}} \mathcal{D}^*(u)^T\nu\ dydx = \int_{\Omega_{\mathcal{I}}} u\mathcal{V}(\nu)\ dx,$$

the *nonlocal Green's first identity*

$$\int_{\Omega} v\mathcal{D}(\mathcal{D}^*(u))\ dx - \int_{\Omega\cup\Omega_{\mathcal{I}}}\int_{\Omega\cup\Omega_{\mathcal{I}}} \mathcal{D}^*(v)^T(\mathcal{D}^*(u))\ dydx = \int_{\Omega_{\mathcal{I}}} v\mathcal{V}(\mathcal{D}^*(u))\ dx, \quad (5.4)$$

and the *nonlocal Green's second identity*

$$\int_{\Omega} u\mathcal{D}(\mathcal{D}^*(v))\ dx - \int_{\Omega} v\mathcal{D}(\mathcal{D}^*(u))\ dx = \int_{\Omega_{\mathcal{I}}} u\mathcal{V}(\mathcal{D}^*(v))\ dx - \int_{\Omega_{\mathcal{I}}} v\mathcal{V}(\mathcal{D}^*(u))\ dx.$$

Not only does the nonlocal vector calculus theory provides some tools to analyze differential equations with nonlocal diffusion, the construction of the nonlocal diffusion operator as the composition of the nonlocal gradient and divergence operator leads to the definition of a function space that is - under some circumstances - equivalent to the volume-constraint space of quadratic Lebesgue integrable functions

$$L^2_c(\Omega\cup\Omega_{\mathcal{I}}) := \{u \in L^2(\Omega\cup\Omega_{\mathcal{I}}):\ E_c(u;0) = 0\},$$

where the constraint functional E_c is defined below in (5.7). As appropriate function space for the (weak) solutions of the nonlocal (differential) equations, we consider the so called *nonlocal energy space*

$$V(\Omega\cup\Omega_{\mathcal{I}}) := \{u \in L^2(\Omega\cup\Omega_{\mathcal{I}}) : |||u||| < \infty\}$$

as defined by Du et al. (2012a), endowed with the *nonlocal energy norm*

$$|||u||| := \left(\frac{1}{2}\int_{\Omega\cup\Omega_{\mathcal{I}}}\int_{\Omega\cup\Omega_{\mathcal{I}}} \mathcal{D}^*(u)(x,y)^T(\mathcal{D}^*(u)(x,y))\ dydx\right)^{\frac{1}{2}}.$$

Dealing with volume constraints, the *nonlocal volume-constrained energy space* is then defined as

$$V_c(\Omega\cup\Omega_{\mathcal{I}}) := \{u \in V(\Omega\cup\Omega_{\mathcal{I}}):\ E_c(u;0) = 0\}, \quad (5.5)$$

where E_c denotes the constraint functional in (5.7).
Du et al. (2012a) introduce two types of volume constraints, Dirichlet- and Neumann-

type. Let $\Omega_{\mathcal{I}_d} \subseteq \Omega_{\mathcal{I}}$ denote the Dirichlet interaction set. The Dirichlet volume constraints are defined analogously to the local case as

$$u = g_d \text{ on } \Omega_{\mathcal{I}_d}$$

for a given data function $g_d : \Omega_{\mathcal{I}_d} \to \mathbb{R}$. Let the Neumann interaction set be denoted by $\Omega_{\mathcal{I}_n}$, then the Neumann-type volume constraints are given as

$$-\mathcal{V}(\mathcal{D}^* u) = g_n \text{ on } \Omega_{\mathcal{I}_n}$$

with $g_n : \Omega_{\mathcal{I}_n} \to \mathbb{R}$. Calculating $\mathcal{V}(\mathcal{D}^* u)$ yields

$$-\int_{\Omega \cup \Omega_{\mathcal{I}}} (u(y) - u(x))\gamma(x, y)\ dy = g_n \text{ on } \Omega_{\mathcal{I}_n}. \tag{5.6}$$

The type of the volume constraints determine the constraint functional E_c. In the case of Dirichlet volume constraints, that is $\emptyset \neq \Omega_{\mathcal{I}_d} \subseteq \Omega_{\mathcal{I}}$, Du et al. (2012a, p.680) set

$$E_c(u; g) = E_c(u; g_d) = \int_{\Omega_{\mathcal{I}_d}} (g_d - u)^2 dx. \tag{5.7}$$

If $\Omega_{\mathcal{I}} = \Omega_{\mathcal{I}_n}$, the constraint functional E_c is given as

$$E_c(u; g) = E_c(u; g_n) = \left(g_n - \int_{\Omega \cup \Omega_{\mathcal{I}}} u\ dx \right)^2. \tag{5.8}$$

These constraint functionals characterize the solution space and ensure the uniqueness of the solution (Du et al., 2012a, pp.679).

We have a closer look at the circumstances under which the nonlocal constrained energy space is equivalent to the L^2_c. According to D'Elia et al. (2014), the kernel function γ in (5.1) or (5.6) has to have the following properties:

Proposition 5.1:
For $x \in \Omega$, let $B_\varepsilon(x) := \{y \in \mathbb{R}^n : \ \|y - x\|_2 \leq \varepsilon\}$ be the n-dimensional ball with a given radius $\varepsilon > 0$. Let the kernel function γ satisfy the following properties:

1. *$\gamma(x, y) \geq 0$ for all $y \in \mathcal{B}_\varepsilon(x)$.*
2. *$\gamma(x, y) \geq \gamma_0 > 0$ for all $y \in \mathcal{B}_{\varepsilon/2}(x)$.*
3. *$\gamma(x, y) = 0$ for all $y \in (\Omega \cup \Omega_{\mathcal{I}}) \backslash \mathcal{B}_\varepsilon(x)$.*
4. *There exists a constant $\gamma_1 > 0$ such that*

$$\gamma_1 \leq \int_{(\Omega \cup \Omega_{\mathcal{I}}) \cap \mathcal{B}_\varepsilon(x)} \gamma(x, y)\ dy \qquad \forall x \in \Omega.$$

5. There exists a constant $\gamma_2 > 0$ such that

$$\int_{\Omega\cup\Omega_{\mathcal{I}}} \gamma^2(x,y)\ dy \leq \gamma_2^2 \qquad \forall x \in \Omega.$$

Then, the nonlocal volume-constrained energy space $V_c(\Omega\cup\Omega_{\mathcal{I}})$ is equivalent to the volume constrained Lebesgue space,

$$L_c^2(\Omega\cup\Omega_{\mathcal{I}}) := \{u \in L^2(\Omega\cup\Omega_{\mathcal{I}}):\ \ E_c(u;0) = 0\}.$$

Hence, there exist some constants C_1 and C_2, both positive, such that

$$C_1\|u\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \leq |||u||| \leq C_2\|u\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \qquad \forall u \in V_c(\Omega\cup\Omega_{\mathcal{I}}). \tag{5.9}$$

Moreover, $V_c(\Omega\cup\Omega_{\mathcal{I}})$ endowed with the norm $|||\cdot|||$ is a Hilbert space.

The proof is given by Du et al. (2012a, p.684).

The dual space of $V_c(\Omega\cup\Omega_{\mathcal{I}})$ with respect to the standard $L^2(\Omega\cup\Omega_{\mathcal{I}})$ pairing is denoted by $V_c'(\Omega\cup\Omega_{\mathcal{I}})$. If γ satisfies all properties of Proposition 5.1, this dual is equivalent to $L_c^2(\Omega\cup\Omega_{\mathcal{I}})$ as well. The norm on $V_c'(\Omega\cup\Omega_{\mathcal{I}})$ can naturally be defined as

$$\|f\|_{V_c'(\Omega\cup\Omega_{\mathcal{I}})} := \sup_{u\in V_c(\Omega\cup\Omega_{\mathcal{I}}),\ u\neq 0} \frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} fu\ dx}{|||u|||}.$$

Especially for the kernel function considered in this context, it is true that $V_c'(\Omega\cup\Omega_{\mathcal{I}})$ is equivalent to $L_c^2(\Omega\cup\Omega_{\mathcal{I}})$ such that $\hat{C}_1\|f\|_{V_c'(\Omega\cup\Omega_{\mathcal{I}})} \leq \|f\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})}$ (c. D'Elia and Gunzburger, 2014, p.248).

The spatial Ramsey model is defined over a space-time cylinder, hence we have to consider the time dependent spaces

$$L^2(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})) := \{u(\cdot,t) \in V_c(\Omega\cup\Omega_{\mathcal{I}}):\ |||u(\cdot,\cdot)||| \in L^2(0,T)\},$$

and

$$L^2(0,T;V_c'(\Omega\cup\Omega_{\mathcal{I}})) := \{u(\cdot,t) \in V_c'(\Omega\cup\Omega_{\mathcal{I}}):\ \|u(\cdot,\cdot)\|_{V_c'} \in L^2(0,T)\},$$

for $T > 0$ respectively (D'Elia et al., 2014, p.10).
For functions that are weakly differentiable according to time, we define the space

$$H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})) := \{u \in L^2(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})):\ \frac{\partial u}{\partial t} \in L^2(0,T;V_c'(\Omega\cup\Omega_{\mathcal{I}}))\}.$$

We can conclude later that the weak solution of our problem is not only weakly differentiable, but also continuous in the time variable. Hence, the function space where we expect our weak solution to live in, is

$$\mathcal{C}(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))\cap H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})).$$

This intersection has to be understood as subspace of $\mathcal{C}(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))$.

5.2 The Weak Solution over Bounded Spatial Domains

We now transfer the nonlocal spatial Ramsey model on bounded spatial domains as introduced in Chapter 3.4 to the nonlocal vector calculus. We base our assumptions on the papers of D'Elia and Gunzburger (2014) and D'Elia et al. (2014). First, we discuss the nonlocal vector calculus with respect to applicability to our model. Afterwards, we derive an existence and several regularity results of the weak solution of the nonlocal spatial Ramsey model on bounded domains.

5.2.1 Embedding the Nonlocal Spatial Ramsey Model in the Nonlocal Vector Calculus

We assume that the domain $\Omega\subset\mathbb{R}^n$ has a (at least piecewise) smooth boundary and satisfies the cone condition as stated in Definition 3.12. The interaction domain $\Omega_{\mathcal{I}}$ and the nonlocal closure $\Omega\cup\Omega_{\mathcal{I}}$ are assumed to have the same properties. We choose the nonlocal interaction domain as

$$\Omega_{\mathcal{I}}:=\{y\in\mathbb{R}^n\backslash\Omega:\ \|y-x\|_2<\varepsilon \text{ for } x\in\Omega\}.$$

The parameter ε will be referred to as the interaction radius.
We set $\alpha:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$,

$$\alpha_\varepsilon(x,y):=\operatorname{sign}(\|x\|_2-\|y\|_2)\left(\frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left(-\frac{1}{2}(x-y)^T\Sigma_\sigma^{-1}(x-y)\right)\mathbb{1}_{B_\varepsilon(x)}(y)\right)^{\frac{1}{2}},$$

for $\varepsilon,\ \sigma>0$, and a covariance matrix $\Sigma_\sigma\in\mathbb{R}^{n\times n}$. Note, that we do explicitly allow to choose $\sigma\neq\varepsilon$. But since ε is the only parameter that is important for estimates and the calculations below, we only keep the dependence of α_ε on σ in mind and do not use it in the notation.
The function α_ε is obviously antisymmetric and we can easily calculate that the kernel function in the nonlocal Ramsey model indeed has the form

$$\Gamma_\varepsilon(x,y)=\alpha_\varepsilon^2(x,y)=\frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left(-\frac{1}{2}(x-y)^T\Sigma_\sigma^{-1}(x-y)\right)\mathbb{1}_{B_\varepsilon(x)}(y).$$

We assume that the covariance matrix is a diagonal matrix with equal entries,

$$\Sigma_\sigma := \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix},$$

such that $\det(\Sigma_\sigma) = \sigma^{2n}$. Again, this means that capital can move through space without any barriers or transition costs and that the central planner does not prioritize any space direction, but weights them all equally. Then, the kernel function satisfies all properties required in order to fit the nonlocal vector calculus.

Lemma 5.2:
The kernel function Γ_ε satisfies all properties of Proposition 5.1.

Proof. The kernel function Γ_ε is given by

$$\Gamma_\varepsilon(x,y) := \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{\|x-y\|_2^2}{2\sigma^2}\right) \mathbb{1}_{B_\varepsilon(x)}(y),$$

which is obviously symmetric. We go on checking all properties as in Proposition 5.1.

(1) ,(2) Let $0 < \eta \leq \varepsilon$. For all $y \in B_\eta(x)$ it is true that

$$\Gamma_\varepsilon(x,y) \geq \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{\eta^2}{2\sigma^2}\right) > 0$$

(3) The third property follows with the definition of the indicator function.

(4) For $x \in \Omega$, we calculate

$$\begin{aligned} \int_{(\Omega\cup\Omega_{\mathcal{I}})\cap B_\varepsilon(x)} \Gamma_\varepsilon(x,y)\, dy &= \int_{B_\varepsilon(x)} \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{\|x-y\|_2^2}{2\sigma^2}\right) dy \\ &\geq \int_{B_\varepsilon(x)} \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) dy \\ &= c_n\varepsilon^n \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) > 0 \end{aligned}$$

where c_n denotes the volume of the unit sphere in $\mathbb{R}^n$.

(5) For the last property, we calculate for $x \in \Omega$,

$$\begin{aligned}\int_{\Omega \cup \Omega_{\mathcal{I}}} \Gamma_\varepsilon^2(x,y)\, dy &= \int_{\Omega \cup \Omega_{\mathcal{I}}} \frac{1}{(2\pi\sigma^2)^n} \exp\left(-\frac{\|x-y\|_2^2}{\sigma^2}\right) \mathbb{1}_{B_\varepsilon(x)}(y)\, dy \\ &= \int_{B_\varepsilon(x)} \frac{1}{(2\pi\sigma^2)^n} \exp\left(-\frac{\|x-y\|_2^2}{\sigma^2}\right) dy \\ &\leq \int_{B_\varepsilon(x)} \frac{1}{(2\pi\sigma^2)^n} \exp\left(-\frac{0}{\sigma^2}\right) dy \\ &= \frac{c_n \varepsilon^n}{(2\pi\sigma^2)^n} < \infty,\end{aligned}$$

which completes the proof.

□

We now have all at hand to define the nonlocal spatial Ramsey model with endogenous productivity growth under a finite time horizon and an open, but bounded spatial domain. We restrict the consideration to homogeneous Dirichlet volume constraints, which means that the central planner in the considered economy enforces the value of the capital stock in the interaction domain to zero. Hence, for a given initial condition $k_0 \in V_c(\Omega)$, the central planner faces the problem to find an optimal control $c^* \in \mathcal{U}_{ad} \subset L^2(0,T;V_c'(\Omega))$ and an optimal state $k^* \in \mathcal{C}(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}})) \cap H^1(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}}))$, such that

$$\mathcal{J}(k,c) := \int_0^T \int_\Omega -U(c(x,t)) e^{-\tau t - \gamma \|x\|_2^2}\, dxdt + \frac{1}{2\rho} \int_\Omega (k(x,T) - k_T(x))^2\, dx \tag{5.10}$$

is minimized subject to k and c satisfying

$$\begin{aligned}\frac{\partial k}{\partial t} - \beta \mathcal{NL}_\varepsilon(k) + \delta k &= \mathcal{P}(k) - c && \text{on } \Omega \times (0,T), \\ k &= 0 && \text{on } \Omega_{\mathcal{I}} \times (0,T), \\ c \in \mathcal{U}_{ad},\ k &\geq 0 && \text{on } \Omega \times (0,T), \\ k(\cdot,0) = k_0 &> 0 && \text{in } \Omega,\end{aligned} \tag{5.11}$$

where $\mathcal{NL}_\varepsilon$ is given as in (5.1) with the kernel function $\gamma := \Gamma_\varepsilon$ and the nonlocal productivity-production operator $\mathcal{P}$ is defined analogously to (3.12) as

$$\mathcal{P}(k)(x,t) := A_0(x) \exp\left(\frac{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi(k(y,t)) \Gamma_\mu(x,y)\, dy}{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi(k(y,t)) \Gamma_\varepsilon(x,y)\, dy + \xi} t \right) p(k(x,t))$$

for a non-negative, real valued function ϕ, $\xi > 0$, and $0 < \mu < \varepsilon$. The set $\mathcal{U}_{ad}$ denotes the set of feasible controls, which will be described in more detail later.

Remark 5.3:
Note that the definition of the kernel function as a truncated Gaussian density function, depending on an indicator function which is determined by the parameters μ and ε, guarantees that

$$\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(k(y,t))\Gamma_\mu(x,y)\ dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(k(y,t))\Gamma_\varepsilon(x,y)\ dy + \xi} \leq 1.$$

5.2.2 Existence of a Weak Solution

The main part of this section will be the proof of the existence of a weak solution of the problem (5.11). We apply the fixed point theorem of Banach. To do so, we exploit the Lipschitz continuity of the productivity-production operator $\mathcal{P}$ and refer to a result by Du et al. (2012a, p.686) which states that the linear nonlocal diffusion problem with homogeneous Dirichlet boundary constraints has a unique weak solution.
Throughout this section, we assume that the depreciation rate $\delta > 0$ and that the initial productivity distribution function A_0, the nominal function ϕ, and the nonlinear production function p satisfy the following assumptions:

Assumption 5.4:
Consider the functions $p : \mathbb{R} \to \mathbb{R}_+$, $\phi : \mathbb{R} \to \mathbb{R}_+$, and $A_0 : \mathbb{R}^n \to \mathbb{R}_+$. Then, we assume

- *The production function p is continuous and satisfies the Inada conditions.*
- *The production function satisfies $p(0) = 0$.*
- *The production function p is Lipschitz continuous with Lipschitz constant $L_p > 0$.*
- *The production function p is bounded from above by a constant $M_p > 0$.*
- *The nominal function ϕ is Lipschitz continuous with constant $L_\phi > 0$.*
- *The initial productivity distribution function A_0 is in $L^\infty(\Omega)$.*

We derive the weak formulation of the system (5.11), i.e. we multiply the state equation with a test function $\varphi \in C(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))$ and integrate over $\Omega\times(0,T)$ which yields

$$\begin{aligned}\int_0^T \int_\Omega k_t \varphi \, dxdt - \int_0^T \int_\Omega \mathcal{NL}_\varepsilon(k) \, \varphi \, dxdt + \delta \int_0^T \int_\Omega k \varphi \, dxdt \\ = \int_0^T \int_\Omega (\mathcal{P}(k) - c) \varphi \, dxdt.\end{aligned} \tag{5.12}$$

Applying the nonlocal Green's first identity (5.4) and the homogeneous volume constraint then give us

$$\begin{aligned}\int_0^T \int_\Omega k_t \varphi \, dxdt + \frac{1}{2} \int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \int_{\Omega\cup\Omega_\mathcal{I}} \mathcal{D}^*(k)^T \mathcal{D}^*(\varphi) \;\; dydxdt + \delta \int_0^T \int_\Omega k \varphi \, dxdt \\ = \int_0^T \int_\Omega (\mathcal{P}(k) - c) \varphi \, dxdt.\end{aligned} \tag{5.13}$$

Analogously to the spatially unbounded case, this weak formulation of the capital accumulation equation gives rise to the following definition of a bilinear form

$$\mathbf{a} : V_c(\Omega \cup \Omega_\mathcal{I}) \times V_c(\Omega \cup \Omega_\mathcal{I}) \to \mathbb{R},$$

$$\mathbf{a}(u, v) := \frac{1}{2} \int_{\Omega\cup\Omega_\mathcal{I}} \int_{\Omega\cup\Omega_\mathcal{I}} \mathcal{D}^*(u)^T \mathcal{D}^*(v) \; dydx + \delta \int_\Omega u \; v \; dx. \tag{5.14}$$

We prove the coercivity and continuity of $\mathbf{a}$ in the following Lemma.

Lemma 5.5:
The bilinear form $\mathbf{a}$ *is coercive and continuous, hence there exist constants* $c_1 > 0$ *such that*

(i) $|\mathbf{a}(u, v)| \leq c_1 |||u||| \; |||v|||,$

(ii) $\mathbf{a}(u, u) \geq |||u|||^2.$

Proof. (i) Choose $u, v \in V_c(\Omega \cup \Omega_\mathcal{I})$. Then,

$$\begin{aligned}|\mathbf{a}(u, v)| &= \left| \frac{1}{2} \int_{\Omega\cup\Omega_\mathcal{I}} \int_{\Omega\cup\Omega_\mathcal{I}} \mathcal{D}^*(u)(x, y)^T \mathcal{D}^*(v)(x, y) \; dydx + \delta \int_\Omega u \; v \; dx \right| \\ &\leq \frac{1}{2} \int_{\Omega\cup\Omega_\mathcal{I}} \int_{\Omega\cup\Omega_\mathcal{I}} |\mathcal{D}^*(u)(x, y)^T \mathcal{D}^*(v)(x, y)| \; dydx + \delta \int_\Omega |uv| \; dx.\end{aligned}$$

We use the Cauchy Schwartz inequality which yields together with the norm equivalence (5.9),

$$\begin{aligned}|\mathbf{a}(u,v)| &\leq |||u|||\ |||v||| + \delta \int_{\Omega} |u\ v|\ dx \\ &\leq |||u|||\ |||v||| + \delta \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\ &\leq \left(1 + \frac{\delta}{C_1^2}\right) |||u|||\ |||v|||\end{aligned}$$

for the constant $C_1 > 0$ from (5.9).

(ii) Applying the Poincare inequality, we have

$$\begin{aligned}\mathbf{a}(u,u) &= \frac{1}{2} \int_{\Omega\cup\Omega_{\mathcal{I}}} \int_{\Omega\cup\Omega_{\mathcal{I}}} (\mathcal{D}^*(u)(x,y))^2\ dydx + \delta \int_{\Omega} u^2(x)\ dx \\ &\geq |||u|||^2 + \delta \|u\|^2_{L^2(\Omega)} \\ &\geq |||u|||^2,\end{aligned}$$

exploiting $\delta > 0$.

□

Proceeding as in Chapter 4.1, we can now prove the existence of a weak solution of the capital accumulation equation. We again apply Banach's fixed point theorem and an existence result for a linear but inhomogeneous nonlocal diffusion equation given by Du et al. (2012a, Theorem 5.1, p.686).

Theorem 5.6:
For a given $c \in L^2(0,T;V_c'(\Omega))$ and $k_0 \in V_c(\Omega)$, the problem 5.13 with $k(x,0) = k_0(x)$ on Ω and $k = 0$ on $\Omega_{\mathcal{I}} \times (0,T)$, has a unique weak solution $k^* \in C(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})) \cap H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))$.

Proof. Let $\mathcal{S} : C(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})) \cap H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})) \to C(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}})) \cap H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))$ be the operator that maps a function v to the unique function k that satisfies $k(x,0) = k_0(x)$ on Ω, $k = 0$ on $\Omega_{\mathcal{I}} \times (0,T)$ and that solves the weak formulation of the linear equation

$$\int_0^T \int_{\Omega} k_t(x,t)\varphi(x,t)\ dxdt + \int_0^T \mathbf{a}(k(\cdot,t),\varphi(\cdot,t))dt = \int_0^T \int_{\Omega} (\mathcal{P}(v)(x,t) - c(x,t))\varphi(x,t)\ dxdt,$$

for all $\varphi \in \mathcal{C}(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))$. We fix $T^* \in (0,T)$ sufficiently small, and consider the difference $S(v_1) - S(v_2)$ for two arbitrary functions $v_1,\ v_2 \in C(0,T^*;V_c(\Omega\cup\Omega_{\mathcal{I}}))$ with $S(v_1) = k_1$ and $S(v_2) = k_2$. We choose the test function $k_1 - k_2 \in$

$\mathcal{C}(0, T^*; V_c(\Omega \cup \Omega_{\mathcal{I}})) \cap H^1(0, T^*; V_c(\Omega \cup \Omega_{\mathcal{I}}))$. Then, $k_1 - k_2$ solves

$$\int_0^t \int_\Omega (k_1 - k_2)_t(x,s)(k_1 - k_2)(x,s)\ dx + \mathbf{a}(k_1 - k_1, k_1 - k_2)(s)ds =$$
$$\int_0^t \int_\Omega \left(\mathcal{P}(v_1)(s,x) - \mathcal{P}(v_2)(x,s)\right)(k_1 - k_2)(x,s)\ dxds,$$

for all $t \in [0, T^*]$ according to Du et al. (2012a, p.686).
We estimate the left-hand side using the coercivity property of the bilinear form $\mathbf{a}$ according to Lemma 5.5:

$$\begin{aligned}
&\int_0^t \int_\Omega (k_1 - k_2)_t(x,s)(k_1 - k_2)(x,s)\ dx + \mathbf{a}(k_1 - k_1, k_1 - k_2)(s)ds \\
&\geq \int_0^t \int_\Omega (k_1 - k_2)_t(x,s)(k_1 - k_2)(x,s)\ dx + |||k_1 - k_2(s)|||^2 ds \\
&= \frac{1}{2}\|k_1 - k_2(t)\|^2_{L^2(\Omega)} + \int_0^t |||k_1 - k_2(s)|||^2 ds.
\end{aligned}$$

For the right-hand side, we exploit the Lipschitz property of $\mathcal{P}$ on bounded spatial domains as follows: First, we apply the Hölder inequality,

$$\begin{aligned}
&\int_0^t \int_\Omega \left(\mathcal{P}(v_1)(x,s) - \mathcal{P}(v_2)(x,s)\right)(k_1(x,s) - k_2(x,s))\ dxds \\
&\leq \int_0^t \|\mathcal{P}(v_1)(\cdot,s) - \mathcal{P}(v_2)(\cdot,s)\|_{L^2(\Omega)} \|k_1(\cdot,s) - k_2(\cdot,s)\|_{L^2(\Omega)} ds := (\#).
\end{aligned}$$

Now, we add a 'clever zero' and calculate

$$\begin{aligned}
(\#) = &\int_0^t \|\mathcal{P}(v_1)(\cdot,s) - P(v_1)(\cdot,s)p(v_2(\cdot,s)) + P(v_1)(\cdot,s)p(v_2(\cdot,s)) - \mathcal{P}(v_2)(\cdot,s)\|_{L^2(\Omega)} \\
&\quad \cdot \|k_1(\cdot,s) - k_2(\cdot,s)\|_{L^2(\Omega)} ds \\
\leq &\int_0^t \|\mathcal{P}(v_1)(\cdot,s) - P(v_1)(\cdot,s)p(v_2(\cdot,s))\|_{L^2(\Omega)} \|k_1(\cdot,s) - k_2(\cdot,s)\|_{L^2(\Omega)} ds \\
+ &\int_0^t \|P(v_1)(\cdot,s)p(v_2(\cdot,s)) - \mathcal{P}(v_2)(\cdot,s)\|_{L^2(\Omega)} \|k_1(\cdot,s) - k_2(\cdot,s)\|_{L^2(\Omega)} ds \\
\leq &\int_0^t \|P(v_1)(\cdot,s)\|_{L^\infty(\Omega)} \|p(v_1(\cdot,s) - p(v_2(\cdot,s))\|_{L^2(\Omega)} \|k_1 - k_2(s)\|_{L^2(\Omega)} ds \\
+ &\int_0^t \|p(v_2(\cdot,s))\|_{L^\infty(\Omega)} \|P(v_1)(\cdot,s) - P(v_2)(\cdot,s)\|_{L^2(\Omega)} \|k_1(\cdot,s) - k_2(\cdot,s)\|_{L^2(\Omega)} ds
\end{aligned}$$

$$\leq \int_0^t L_p \|P(v_1)(\cdot,s)\|_{L^\infty(\Omega)} \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\Omega)} \|k_1(\cdot,s) - k_2(\cdot,s)\|_{L^2(\Omega)} ds$$
$$+ \int_0^t M \; \|P(v_1)(\cdot,s) - P(v_2)(\cdot,s)\|_{L^2(\Omega)} \|k_1 - k_2(s)\|_{L^2(\Omega)} ds.$$

We have a closer look at the terms $\|P(v_1)(\cdot,s)\|_{L^\infty(\Omega)}$ and $\|P(v_1)(\cdot,s)-P(v_2)(\cdot,s)\|_{L^2(\Omega)}$. Again, we exploit the boundedness of the fraction in the exponential term of $\mathcal{P}$, which is bounded by

$$\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v(y,s))\Gamma_\mu(x,y)\; dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v(y,s))\Gamma_\varepsilon(x,y)\; dy + \xi} \leq 1$$

by the definition of the indicator function, and the monotonicity of the integral. Hence, we can estimate

$$\begin{aligned}
\|P(v)(\cdot,s)\|_{L^\infty(\Omega)} &= ess\ \sup_{x\in\Omega} |P(v)(x,s)| \\
&= ess\ \sup_{x\in\Omega} \left| A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v(y,s))\Gamma_\mu(x,y)\; dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v(y,s))\Gamma_\varepsilon(x,y)\; dy + \xi}\; s \right) \right| \\
&\leq \|A_0\|_{L^\infty(\Omega)} \exp(s),
\end{aligned}$$

and

$$\begin{aligned}
\|P(v_1)(\cdot,s) - P(v_2)(\cdot,s)\|_{L^2(\Omega)} \leq \|A_0\|_{L^\infty(\Omega)} \; &\left\| \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_1(y,s))\Gamma_\mu(\cdot,y)\; dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_1(y,s))\Gamma_\varepsilon(\cdot,y)\; dy + \xi}\; s \right) \right. \\
&\left. - \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_2(y,s))\Gamma_\mu(\cdot,y)\; dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_2(y,s))\Gamma_\varepsilon(\cdot,y)\; dy + \xi}\; s \right) \right\|_{L^2(\Omega)}
\end{aligned}$$

In order to keep a compact representation, we define the operator

$$\Phi_\nu(v)(x,s) := \int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v(y,s))\Gamma_\nu(x,y)\; dy,$$

for $\nu \in \{\mu, \varepsilon\}$. The exponential function is Lipschitz continuous on compact sets and due to the boundedness of the fractions occurring in the nonlocal productivity-production operator, we can estimate

$$\left\|\exp\left(\frac{\Phi_\mu(v_1)(\cdot,s)}{\Phi_\varepsilon(v_1)(\cdot,s)+\xi}\ s\right)-\exp\left(\frac{\Phi_\mu(v_2)(\cdot,s)}{\Phi_\varepsilon(v_2)(\cdot,s)+\xi}\ s\right)\right\|_{L^2(\Omega)}$$

$$\leq L_{exp}\ s\left\|\frac{\Phi_\mu(v_1)(\cdot,s)}{\Phi_\varepsilon(v_1)(\cdot,s)+\xi}-\frac{\Phi_\mu(v_2)(\cdot,s)}{\Phi_\varepsilon(v_2)(\cdot,s)+\xi}\ \right\|_{L^2(\Omega)}$$

$$= L_{exp}\ s\left\|\frac{\Phi_\mu(v_1)(\cdot,s)\left(\Phi_\varepsilon(v_2)(\cdot,s)+\xi\right)}{\left(\Phi_\varepsilon(v_1)(\cdot,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(\cdot,s)+\xi\right)}-\frac{\left(\Phi_\varepsilon(v_1)(\cdot,s)+\xi\right)\Phi_\mu(v_2)(\cdot,s)}{\left(\Phi_\varepsilon(v_1)(\cdot,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(\cdot,s)+\xi\right)}\ \right\|_{L^2(\Omega)}$$

$$\leq L_{exp}\ s\Bigg(\int_\Omega\Bigg[\Bigg|\frac{\Phi_\mu(v_1)(x,s)\Phi_\varepsilon(v_2)(x,s)}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}-\frac{\Phi_\mu(v_1)(x,s)\Phi_\varepsilon(v_1)(x,s)}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}\Bigg|$$

$$+\left|\frac{\Phi_\mu(v_1)(x,s)\Phi_\varepsilon(v_1)(x,s)}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}-\frac{\Phi_\mu(v_2)(x,s)\Phi_\varepsilon(v_1)(x,s)}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}\right|$$

$$+\left|\frac{\xi\left(\Phi_\mu(v_1)(x,s)-\Phi_\mu(v_2)(x,s)\right)}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}\right|\Bigg]^2\ dx\Bigg)^{\frac{1}{2}}$$

$$\leq L_{exp}\ s\Bigg(\int_\Omega\left(\frac{1}{\xi}\left|\Phi_\varepsilon(v_1)(x,s)-\Phi_\varepsilon(v_2)(x,s)\right|+\frac{2}{\xi}\left|\Phi_\mu(v_1)(x,s)-\Phi_\mu(v_2)(x,s)\right|\right)^2\ dx\Bigg)^{\frac{1}{2}},$$

since

$$\left|\frac{\Phi_\mu(v_1)(x,s)}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}\right|\leq\left|\frac{\Phi_\varepsilon(v_1)(x,s)}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}\right|\leq\frac{1}{\xi},$$

and

$$\left|\frac{\xi}{\left(\Phi_\varepsilon(v_1)(x,s)+\xi\right)\left(\Phi_\varepsilon(v_2)(x,s)+\xi\right)}\right|\leq\frac{1}{\xi},$$

for all $v_1, v_2 \in \mathcal{C}(0,T^*;V_c(\Omega\cup\Omega_{\mathcal{I}}))$ and $(x,s)\in\Omega\times[0,T]$.

Applying the Minkowski and the Hölder inequalities and exploiting the Lipschitz continuity of ϕ, we end up with

$$\left\| \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_1(y,s))\Gamma_\mu(\cdot,y)\ dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_1(y,s))\Gamma_\varepsilon(\cdot,y)\ dy + \xi}\ s\right) - \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_2(y,s))\Gamma_\mu(\cdot,y)\ dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(v_2(y,s))\Gamma_\varepsilon(\cdot,y)\ dy + \xi}\ s\right)\right\|_{L^2(\Omega)}$$

$$\leq sK\ \|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})}$$

with

$$K := \frac{1}{\xi}\left(L_{exp}L_\phi\|\Gamma_\varepsilon\|_{L^2(\Omega\times(\Omega\cup\Omega_{\mathcal{I}}))} + 2L_{exp}L\phi\|\Gamma_\mu\|_{L^2(\Omega\times(\Omega\cup\Omega_{\mathcal{I}}))}\right) < \infty.$$

Note that on the bounded domain $\Omega\cup\Omega_{\mathcal{I}}$, we have

$$\|\Gamma_\nu\|_{L^2(\Omega\times(\Omega\cup\Omega_{\mathcal{I}}))} < \infty$$

for all $\nu > 0$. Thus, we can estimate

$$\|P(v_1)(\cdot,s) - P(v_2)(\cdot,s)\|_{L^2(\Omega)} \leq sK\|A_0\|_{L^\infty(\Omega)}\|v_1(\cdot,s) - v_2(\cdot,s)\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})}.$$

Combining both estimates for the left- and right-side of the PIDE and applying Young's inequality for a constant $\beta > 0$, we get

$$\frac{1}{2}\|k_1 - k_2(t)\|^2_{L^2(\Omega)} + \int_0^t |||k_1 - k_2(s)|||^2 ds \leq \int_0^t \frac{L(s)^2}{2\beta}\|v_1 - v_2(s)\|^2_{L^2(\Omega\cup\Omega_{\mathcal{I}})} + \frac{\beta}{2}\|k_1 - k_2\|^2_{L^2(\Omega)} ds$$

with $L(s) := \|A_0\|_{L^\infty(\Omega)}(L_p\exp(s) + MKs)$.

We choose $2C_1 < \beta < 2C_1\left(\frac{1}{2C_2} + 1\right)$, where C_1 and C_2 are the constants from (5.9). Then, again with (5.9), we can interpret the inequality in terms of the $V_c(\Omega\cup\Omega_{\mathcal{I}})$-norm as follows

$$\frac{1}{2C_2}|||k_1 - k_2(t)|||^2 \leq \int_0^t \left(\frac{L(s)^2}{2\beta C_1}|||v_1 - v_2(s)|||^2 + \left(\frac{\beta}{2C_1} - 1\right)|||k_1 - k_2|||^2\right)\ ds.$$

Note that we have once more exploited the Dirichlet volume constraints in order to rewrite

$$\|k_1 - k_2(t)\|_{L^2(\Omega)} = \|k_1 - k_2(t)\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})}$$

Taking the maximum over all $t \in [0, T^*]$ and sorting the terms, we have

$$\frac{1}{2C_2}\|k_1 - k_2\|^2_{L^\infty(0,T^*;V_c(\Omega\cup\Omega_{\mathcal{I}}))} + (1 - \frac{\beta}{2C_1})T^*\|k_1 - k_2\|^2_{L^\infty(0,T^*;V_c(\Omega\cup\Omega_{\mathcal{I}}))} \leq C(T^*)\|v_1 - v_2\|^2_{L^\infty(0,T^*;V_c(\Omega\cup\Omega_{\mathcal{I}}))},$$

where

$$C(T^*) := \frac{1}{2\beta C_1} ess \sup_{t\in[0,T^*]} \int_0^t L(s)^2 ds.$$

Taking the limit $T^* \to 0$, we obtain $C(T^*) \to 0$ since

$$\int_0^t L(s)^2 ds \leq \int_0^{T^*} L(s)^2 ds \leq \tilde{C} \int_0^{T^*} (\exp(s) + s)^2 ds = \tilde{C}\left(\frac{1}{2}\exp(2T^*) - \frac{1}{2} + \frac{1}{3}{T^*}^2 + \exp(T^*)(T^* - 1) + 1\right) \to 0.$$

Thus, we conclude that there exists a T^* small enough such that

$$\frac{C(T^*)}{\left(\frac{1}{2C_2} + (1 - \frac{\beta}{2C_1})T^*\right)} < 1.$$

Note that in particular

$$\left(\frac{1}{2C_2} + (1 - \frac{\beta}{2C_1})T^*\right) > 0$$

for $T^* \leq 1$ by the choice of β. Hence, we have shown that $\mathcal{S}$ is a contraction on a sufficiently small time interval. According to Banach's fixed point theorem, $\mathcal{S}$ has a unique fixed point on every bounded set. Since the local solution k is independent of the time horizon T^*, we can proceed on the interval $[T^*, 2T^*]$ using the same arguments as above but with a new initial condition $k(\cdot, T^*)$. After finitely many steps, we can construct a weak solution of (4.1) on the whole time space cylinder after finitely many steps. Moreover, this solution is unique. □

Now, we have a closer look at the regularity of the weak solution k. We start calculating an a priori estimate, which depends only on the initial value condition and the inhomogeneity.

Corollary 5.7:
There exists a constant $C_\infty > 0$ independent of the data c and k_0 such that the weak solution of (5.11) satisfies the following a priori estimate

$$\|k\|_{H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))} \leq C_\infty \left(\|c\|_{L^2(0,T;L^2(\Omega))} + \|k_0\|_{L^2(\Omega)} + 1\right).$$

In particular, this estimate gives us the continuity of the solution operator

$$G : L^2(0,T;L^2(\Omega)) \times L^2(\Omega) \to H^1(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}}))$$

that maps any inhomogeneity c and initial condition k_0 to the solution of (5.11) (cf. Tröltzsch, 2005, p.112).

Proof. First, we recall that

$$\|k\|^2_{H^1(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))} = \|k\|^2_{L^2(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))} + \|k_t\|^2_{L^2(0,T;V_c'(\Omega\cup\Omega_{\mathcal{I}}))}.$$

We estimate the first term exploiting the coercivity of the bilinear form $\mathbf{a}$. We choose a $t \in [0,T]$ and derive the weak formulation of the capital equation for the test function $k \in H^1(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}})) \cap \mathcal{C}(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}}))$ which yields

$$\int_0^t \int_\Omega \frac{\partial k}{\partial t} k \; dxds + \int_0^t \mathbf{a}(k,k)ds = \int_0^t \int_\Omega (\mathcal{P}(k) - c)k \; dxds.$$

As already proven in Lemma 5.5, $\mathbf{a}(k,k) \geq |||k|||^2$. Hence, we can estimate the left-hand side as

$$\begin{aligned} LHS &= \int_0^t \int_\Omega \frac{\partial k}{\partial t} k \; dxds + \int_0^t \mathbf{a}(k,k)ds \\ &\geq \int_0^t \int_\Omega \frac{\partial k}{\partial t} \; dxds + \int_0^t |||k(s)|||^2 ds \\ &= \frac{1}{2}\|k(t)\|^2_{L^2(\Omega)} - \frac{1}{2}\|k_0\|^2_{L^2(\Omega)} + \int_0^t |||k(s)|||^2 ds \end{aligned}$$

for all $t \in [0,T]$. In order to derive an upper bound for the right-hand side, we exploit the Lipschitz continuity and the boundedness of the production function p, $p(0) = 0$, and the boundedness of the fraction

$$\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(k(y,t))\Gamma_\mu(x,y) \; dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi(k(y,t))\Gamma_\varepsilon(x,y) \; dy + \xi} \leq 1$$

for all $x \in \Omega$ and $t \in [0,T]$. With these properties, we get

$$\begin{aligned} RHS &= \int_0^t \int_\Omega (\mathcal{P}(k) - c)k \; dxds \\ &\leq \int_0^t \|\mathcal{P}(k)(s)\|_{L^2(\Omega)} \|k(s)\|_{L^2(\Omega)} ds + \int_0^t \|c(s)\|_{L^2(\Omega)} \|k(s)\|_{L^2(\Omega)} ds \\ &\leq \|A_0\|_{L^\infty(\Omega)} \int_0^t \left(\int_\Omega |e^s p(k(s))|^2 \; dx\right)^{\frac{1}{2}} \left(\int_\Omega |k(s)|^2 \; dx\right)^{\frac{1}{2}} ds \end{aligned}$$

$$
\begin{aligned}
&+ \int_0^t \|c(s)\|_{L^2(\Omega)} \|k(s)\|_{L^2(\Omega)} ds \\
&\leq \|A_0\|_{L^\infty(\Omega)} M_p |\Omega| \left(\int_0^t e^{2s} ds \right)^{\frac{1}{2}} \left(\int_0^t \|k(s)\|^2_{L^2(\Omega)} ds \right)^{\frac{1}{2}} \\
&+ \int_0^t \|c(s)\|_{L^2(\Omega)} \|k(s)\|_{L^2(\Omega)} ds \\
&\leq \|A_0\|_{L^\infty(\Omega)} M_p |\Omega| \left(\frac{e^{2t}}{2} - \frac{1}{2} \right)^{\frac{1}{2}} \|k\|_{L^2(0,t;L^2(\Omega))} + \int_0^t \|c(s)\|_{L^2(\Omega)} \|k(s)\|_{L^2(\Omega)} ds.
\end{aligned}
$$

Using Young's inequality for two constants $\eta_1, \eta_2 > 0$, we have

$$
\begin{aligned}
RHS &\leq \|A_0\|_{L^\infty(\Omega)} M_p |\Omega| \left(\frac{\eta_1}{2} \left(\frac{e^{2t}}{2} - \frac{1}{2} \right) + \frac{1}{2\eta_1} \|k\|^2_{L^2(0,t;L^2(\Omega))} \right) \\
&\quad + \frac{\eta_2}{2} \|c\|^2_{L^2(0,t;L^2(\Omega))} + \frac{2}{\eta_2} \|k\|^2_{L^2(0,t;L^2(\Omega))} \\
&\leq \|A_0\|_{L^\infty(\Omega)} M_p |\Omega| \left(\frac{\eta_1}{2} \left(\frac{e^{2t}}{2} - \frac{1}{2} \right) + \frac{1}{2\eta_1} \|k\|^2_{L^2(0,t;V_c(\Omega \cup \Omega_{\mathcal{I}}))} \right) \\
&\quad + \frac{\eta_2}{2} \|c\|^2_{L^2(0,t;L^2(\Omega))} + \frac{2}{\eta_2} \|k\|^2_{L^2(0,t;V_c(\Omega \cup \Omega_{\mathcal{I}}))}.
\end{aligned}
$$

Combining both estimates yields

$$
\begin{aligned}
&\frac{1}{2} \|k(t)\|^2_{L^2(\Omega)} + \int_0^t |||k(s)|||^2 ds \\
&\leq \frac{1}{2} \|k_0\|^2_{L^2(\Omega)} + \|A_0\|_{L^\infty(\Omega)} M_p |\Omega| \left(\frac{\eta_1}{2} \left(\frac{e^{2t}}{2} - \frac{1}{2} \right) + \frac{1}{2\eta_1} \|k\|^2_{L^2(0,t;V_c(\Omega \cup \Omega_{\mathcal{I}}))} \right) \\
&\quad + \frac{\eta_2}{2} \|c\|^2_{L^2(0,t;L^2(\Omega))} + \frac{2}{\eta_2} \|k\|^2_{L^2(0,t;V_c(\Omega \cup \Omega_{\mathcal{I}}))}.
\end{aligned}
$$

Taking the maximum over all $t \in [0, T]$ finally gives us

$$
\frac{1}{2} \|k\|^2_{L^\infty(0,T;L^2(\Omega))} + \hat{c} \|k\|^2_{L^2(0,T;V_c(\Omega \cup \Omega_{\mathcal{I}}))} \leq C \left(\|k_0\|_{L^2(\Omega)} + \|c\|_{L^2(0,T;L^2(\Omega))} + 1 \right)^2
$$

with

$$
\hat{c} := 1 - \frac{\|A_0\|_{L^\infty(\Omega)} M_p |\Omega|}{2\eta_1} - \frac{2}{\eta_2} > 0
$$

for $\eta_1, \eta_2 > 0$ sufficiently large.
In order to estimate the second term, we define some linear functionals analogously to Tröltzsch (2005, pp.119), namely

$$F_1(t) : v \mapsto \langle k(t), v \rangle_{V(\Omega\cup\Omega_{\mathcal{I}})} \qquad F_2(t) : v \mapsto \langle \delta k(t), v \rangle_{L^2(\Omega)}$$
$$F_3(t) : v \mapsto \langle \mathcal{P}(k)(t), v \rangle_{L^2(\Omega)} \qquad F_4(t) : v \mapsto \langle c(t), v \rangle_{L^2(\Omega)}$$

These functionals are continuous since

$$|F_1(t)v| \leq |||k(t)|||\ |||v||| \quad \text{and} \quad |F_2(t)v| \leq \delta|||k(t)|||\ |||v|||$$

using the Cauchy-Schwartz inequality. For the third and fourth functional we get

$$|F_3(t)v| \leq \text{const}(t)|||k(t)|||\ |||v||| \quad \text{and} \quad |F_4(t)v| \leq \|c(t)\|_{L^2(\Omega)}|||v|||$$

using the estimates of the proof of Lemma 5.5. Here we denote by const(t) a constant depending only on t. For fixed k and c, we can interpret the values $|||k(t)|||$ and $\|c(t)\|_{L^2(\Omega)}$ as constants of the definition of the continuity of $F_i,\ i = 1, ..., 4$. According to Tröltzsch (2005, p.120), we can find a constant $\hat{c}$ such that

$$\|F_i(t)\|_{V_c(\Omega\cup\Omega_{\mathcal{I}})'} \leq \hat{c}|||k(t)|||,\ i = 1, 2, 3 \quad \text{and } \|F_4(t)\|_{V_c'(\Omega\cup\Omega_{\mathcal{I}})} \leq \hat{c}\|c(t)\|_{L^2(\Omega)}.$$

From the weak formulation, we know that

$$\|k_t\|^2_{L^2(0,T;V_c'(\Omega\cup\Omega_{\mathcal{I}}))} \leq \sum_{i=1}^{4} \|F_i\|_{L^2(0,T;V_c'(\Omega\cup\Omega_{\mathcal{I}}))}.$$

Using the estimation for k,

$$\|k\|^2_{L^2(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))} \leq C \left(\|k_0\|_{L^2(\Omega)} + \|c\|_{L^2(0,T;L^2(\Omega))} + 1\right)^2,$$

we have

$$\|k_t\|^2_{L^2(0,T;V_c'(\Omega\cup\Omega_{\mathcal{I}}))} \leq \tilde{C} \left(\|k_0\|_{L^2(\Omega)} + \|c\|_{L^2(0,T;L^2(\Omega))} + 1\right)^2.$$

Summing up both estimates, we finally achieve

$$\|k\|^2_{L^2(0,T;V_c(\Omega\cup\Omega_{\mathcal{I}}))} + \|k_t\|^2_{L^2(0,T;V_c'(\Omega\cup\Omega_{\mathcal{I}}))} \leq C_\infty^2 \left(\|k_0\|_{L^2(\Omega)} + \|c\|_{L^2(0,T;L^2(\Omega))} + 1\right)^2$$

which completes the proof. □

So far, we have only considered the initial data and the right-hand side of the PIDE to be $L^2(\Omega \times [0, T])$ functions. The highest regularity, we can achieve in that case, is $\mathcal{C}(0, T; L^2(\Omega \cup \Omega_{\mathcal{I}}))$. We cannot expect a higher regularity in the space direction, since there is no operator, such as the differential operator, that drives regularity. Nevertheless, we would expect a higher regularity of the weak solution, whenever we choose a higher regularity for the data. The following theorem shows, that the weak solution of the nonlocal capital accumulation has indeed the same regularity as the data.

Theorem 5.8:
Let all assumptions of Theorem 5.6 hold, and let $k_0 \in L^\infty(\Omega)$ *and* $c \in L^\infty(\Omega \times [0,T])$*. Then, the weak solution of the capital accumulation equation (5.11) is* $\mathcal{C}(0,T; V_c(\Omega \cup \Omega_{\mathcal{I}})) \cap H^1(0,T; L^\infty(\Omega))$*.*

Remark 5.9:
By the intersection $\mathcal{C}(0,T; V_c(\Omega \cup \Omega_{\mathcal{I}})) \cap H^1(0,T; L^\infty(\Omega))$*, we mean a subspace of the* $\mathcal{C}([0,T]; V_c(\Omega \cup \Omega_{\mathcal{I}}))$ *space. We define the Banach space*

$$\mathcal{V}^\infty := \{u \in \mathcal{C}([0,T]; V_c(\Omega \cup \Omega_{\mathcal{I}})) : \ ess \sup_{(x,t) \in \Omega \times (0,T)} |u(x,t)| < \infty\}$$

endowed with the norm

$$\|u\|_{\mathcal{V}^\infty} := \|u\|_{\mathcal{C}([0,T]; V_c(\Omega \cup \Omega_{\mathcal{I}}))} + \|u\|_{L^\infty(\Omega \times (0,T))}$$

and refer to $\mathcal{V}^\infty$ *whenever we consider the intersection space.*

Proof. Consider the solution $k^* \in \mathcal{C}([0,T]; V_c(\Omega \cup \Omega_{\mathcal{I}}))$ of the capital accumulation equation. For such k^*, the production-productivity operator $\mathcal{P}$ maps to L^∞ since we have assumed A_0 to be a $L^\infty(\Omega)$ function and the production function p to be bounded. We can calculate

$$\|\mathcal{P}(k^*)\|_{L^\infty(\Omega \times [0,T])} = \|A_0\|_{L^\infty(\Omega)} e^T \ M_p,$$

where M_p denotes the uniform upper bound of p. Moreover, we know that

$$\int_{\Omega \cup \Omega_{\mathcal{I}}} \Gamma_\varepsilon(x,y) \ dy =: \hat{\Gamma}_\varepsilon(x) \leq 1$$

and using Hölder's inequality, it follows that

$$\int_{\Omega \cup \Omega_{\mathcal{I}}} k^*(y,t) \Gamma_\varepsilon(x,y) \ dy < \infty$$

for all $x \in \Omega$. For a fixed $x \in \Omega$, we consider the capital accumulation equation

$$\begin{aligned} \frac{\partial k}{\partial t}(x,t) - \mathcal{NL}(k)(x,t) + \delta k(x,t) - \mathcal{P}(k)(x,t) = -c(x,t) \qquad &\text{on } (0,T) \\ k(x,0) = k_0(x). \end{aligned}$$

We rewrite the equation as

$$\frac{\partial k}{\partial t}(x,t) + (\hat{\Gamma}_\varepsilon(x) + \delta) k(x,t) = \int_{\Omega \cup \Omega_{\mathcal{I}}} k(y,t) \Gamma_\varepsilon(x,y) \ dy + \mathcal{P}(k)(x,t) - c(x,t) \text{ on } (0,T).$$

We neglect the dependence of k of the right-hand side, since it maps every k to

L^∞, and define

$$g_x(t) := \int_{\Omega \cup \Omega_{\mathcal{I}}} k(y,t)\Gamma_\varepsilon(x,y)\ dy + \mathcal{P}(k)(x,t) - c(x,t) \in L^\infty(\Omega \times [0,T]).$$

Note, that the regularity of g_x is determined by the regularity of c. Now, we consider the inhomogeneous linear ordinary differential equation

$$\frac{\partial k_x}{\partial t} + (\hat{\Gamma}_{x,\varepsilon} + \delta)\ k = g_x$$

depending on the parameter x. We know that g_x is continuous in t. Hence, the equation has a solution $\overline{k}$ given as

$$\overline{k}_x(t) = e^{-t(\hat{\Gamma}_{x,\varepsilon}+\delta)} \left(k_{0,x} + \int_0^t g_x(s) e^{s(\hat{\Gamma}_{x,\varepsilon}+\delta)} ds \right)$$

which is bounded for every $x \in \Omega$. Thus, we conclude

$$\overline{k} \in L^\infty(\Omega \times [0,T]).$$

Since the solution of the capital accumulation equation is unique we get $k^* = \overline{k}$, which ends the proof. □

5.3 Discussion of the Existence of an Optimal Control

So far, we have derived the existence of a weak solution of the capital accumulation equation in the nonlocal spatial Ramsey model with endogenous productivity growth on a bounded spatial domain. The next, natural step would be to analyze the model (5.10) and (5.11) with respect to the existence of an optimal control c^* and the corresponding optimal state k^*. However, this task is quite difficult - at least to the best of our knowledge and judgment.
In contrast to Chapters 3.4 and 4, where we introduced a very general version of the nonlocal spatial Ramsey model, we decided to not consider any local diffusion effects in the model analyzed above. The main motivation was to analyze how pure nonlocal diffusion affects the accumulation of capital over space. This was an application-driven decision since any movement of capital and labor across space is in fact a nonlocal process. Another consideration was simplicity. Combining both effects -local and nonlocal diffusion- on bounded spatial domains, we would have had to introduce some boundary conditions, acting on the surface $\partial\Omega$ as well as the volume constraints, living on a set of non-zero volume $\Omega_{\mathcal{I}}$. But the combination of surface and volume constraints is a topic that is not well studied yet. Even defining the appropriate nonlocal volume constraints is a task that is not easy to do in practice (D'Elia et al., 2016, p.2).

Moreover, the literature on such coupled boundary (surface or volume type) conditions for local-nonlocal models is quite rare. Though, a field called *Local-to-Nonlocal Coupling Strategies* has focused on the combination of local and nonlocal diffusion effects, whereupon the motivation for this topic is different to our case. The idea is to exploit the computational efficiency of PDEs, also in fields, where nonlocal models are needed. One paper which we want to mention in this context and where the basic idea of those coupling strategies becomes quite obvious is *'A coupling strategy for nonlocal and local diffusion models with mixed volume constraints and boundary conditions'* by D'Elia et al. (2016). Here, the authors consider an application from physics. They are in a setting where they need to model processes like heat flow in a medium with cracks, where local-nonlocal diffusion models turned out to be the most appropriate. Nevertheless, the situation is different to our case. D'Elia et al. (2016) are able to confine two separated spatial domains which only share the boundary. On those domains, they consider a local and a nonlocal diffusion problem separately, such that they end up with two (not completely independent) problems, one with integral operator and volume constraints, and the other with Laplace operator and common boundary constraints. In that way, they circumvent the combination of both types of boundary constraints in one single problem.
Another field where models with both, Laplace and space-integral, operators occur rather often is Financial Mathematics. When considering stochastic jump diffusion models with Lévy processes for option pricing problems, the transformation of stochastic differential equations into deterministic partial differential equations via Itō's lemma and the Feynman-Kac formula leads to partial integro-differential equations that have the same form as the capital accumulation equation in (3.14). However, a characteristic of these types of models is the unbounded spatial domain. The space variable in these models describes the price of an underlying asset and is usually not restricted a priori. A paper we want to mention here as an example is *'Convergence of numerical schemes for viscosity solutions in integro-differential degenerate parabolic problems arising in financial theory'* by Briani et al. (2004) which we have mentioned before. Here, the authors consider the $\mathbb{R}^n$ as spatial domain, which makes the definition of boundary conditions unnecessary. Especially in the application considered in that paper, they show that the error, which they produce when truncating the spatial domain in order to solve the equation numerically, is small and hence negligible. Other models with both, local and nonlocal diffusion, and unbounded spatial domain, hence without any boundary conditions, are written for example by Chandra and Mukherjee (2016), Cont and Voltchkova (2005), Matache et al. (2004), or Sachs and Strauss (2008).

From an economic point of view, it is justifiable (or even necessary) to neglect any local diffusive effects in the capital accumulation process across space. However, this decision now takes its toll when it comes to proving existence of an optimal control. Since we do not need any derivative of the state function in space direction, the regularity of the solution of the PIDE in the spatial Ramsey model is much weaker than if we had considered local diffusion as well. The highest regularity we can expect in the space direction is L^∞, at least for data k_0 and c in L^∞, and this regularity is only valid in the interior of the set of interest. We do not gain any information on the regularity inside the interaction domain or between Ω and $\Omega_{\mathcal{I}}$. For data with less regularity, also the regularity of the weak solution is weaker. Most techniques which are used to derive the existence of an optimal control under semilinear differential equations require much higher regularity on k, for example continuity (cf. Casas, 2006; Tröltzsch, 2005; Lions, 1971). However, the weak solutions in our model are not necessarily continuous. On the contrary, we will point out in the next chapter, that one advantage of our model is to conserve any discontinuity of a given initial capital distribution.

For optimal control problems subject to semilinear parabolic differential equations, Lions (1971), Tröltzsch (2005), and Casas (2006) have shown existence results. The approach is the same in all three works and as already mentioned, the regularity of the weak solution of the state equation is crucial for this theory. The proof uses some convergence arguments of sequences of controls and states, and embedding theorems of the considered function spaces. The basic idea, as already described in detail in Chapter 4.2, is to consider a minimizing sequence of controls that exists under appropriate assumptions on the objective and the set of feasible controls. Then, the semilinear state equations are linearized in order to derive a convergent sequence of states in an appropriate norm and it is shown that the limit of this sequence solves the PDE constraint. After showing that the limit of the control and state sequence minimizes the cost functional, the proof is concluded.

Although solutions of such nonlocal equations as considered in the nonlocal vector calculus are much less regular, compared to the weak solutions of local PDEs with derivatives in space direction, it is not impossible to prove the existence of optimal controls in this setting. In the paper *'Optimal Distributed Control of Nonlocal Steady Diffusion Problems'* by D'Elia and Gunzburger (2014), the authors consider a model for which they can show the existence of an optimal control straightforwardly. They consider an elliptic, inhomogeneous, though linear nonlocal differential constraint equation and a quadratic matching functional as objective. The authors are able to exploit the special structure of the objective functional by rewriting the reduced cost functional as a sum of a symmetric, continuous, and coercive bilinear form and a continuous and linear functional. Applying the Lax-Milgram theorem then completes the proof, although the regularity of the weak solution is only L^2.
However, our cost functional does not have any matching type structure, but is a quite general, convex function which we cannot rewrite in that way. Moreover, our state equation is parabolic and semilinear, thus we cannot proceed similarly to D'Elia and Gunzburger (2014).

We expect that the weak solution of our nonlocal spatial Ramsey model is continuous in the interior of the set of interest Ω for appropriate (i.e. continuous) data, since the integral operator does not weaken any initial regularity over time and space. The transition between Ω and $\Omega_{\mathcal{I}}$ has to be analyzed further, also the impact of the regularity of the weak solution in the interaction domain on the existence of an optimal control. However, we recommend this to further research.

CHAPTER 6

Numerical Results and Economic Evaluation

'The disparities between cities and regions are generally more stable [...], even at times of growth and structural change. But it is also true that disparities can be substantial and persistent, lasting many decades.' (Breinlich et al., 2014, p.1)

This chapter is dedicated to the numerical solution of the nonlocal spatial Ramsey model with endogenous productivity growth. We start the numerical treatment illustrating the impact of the kernel function in the nonlocal diffusion operator on the quality of the diffusion in Section 6.1. After that we discuss the implementation of the capital accumulation equation (5.11) in Section 6.2.
In contrast to Chapter 5, we consider a Neumann-type volume constraint for the numerical implementation, hence we consider the case $\Omega_{\mathcal{I}} = \Omega_{\mathcal{I}_n}$. Throughout this numerical realization, we ignore the constraint on the accumulated capital stock

$$\int_{\Omega \cup \Omega_{\mathcal{I}}} k(x,t)\; dx = g(t) \text{ on } (0,T),$$

which defines the constraint functional in (5.8) and was only introduced to ensure the uniqueness of the weak solution. This is convenient as we are only interested in one solution, not a particular one. However, this constraint would lead to an additional linear equality constraint, which could be implemented quite forward using a quadrature rule. As already mentioned, a Neumann-type volume constraint refers to a natural exchange of production factors between the considered (bounded) economy and its surrounding, independent of the preferences of a central planner, which is an other interesting scenario for the application.
The numerical solution of partial integro-differential equations is often a rather challenging task. Thereby, the most problematic term is the integral part of the PIDE. Depending on the kernel function, many spatial discretization schemes like finite elements or finite differences combined with quadrature rules lead to dense

matrices. Hence, a fully implicit time discretization is very expensive in terms of memory capacity and computation time. There exist some techniques to reduce computational cost. Anderson and Anderson (2000) consider an alternating directions implicit method (ADI) combined with a fast Fourier transformation. In the ADI, every time step in the discretization is split in two. In the first half-step, the dense part of the discretized problem is treated with an explicit Euler scheme, whereas the sparse part is discretized with an implicit Euler scheme. In the second half-step, the time discretization methods are swapped. Other approaches are splitting schemes (see amongst others Briani et al., 2004; Cont and Voltchkova, 2005), where the discretization matrix of the PIDE is split into a dense and a sparse part. Then, the dense part is treated explicitly with respect to the time discretization, whereas the sparse part can be discretized with a higher order implicit scheme.

In the context of the nonlocal vector calculus, the kernel function enables us to reduce the computational cost as well, at least if the parameter in the kernel is chosen appropriately. The kernel function, which we introduced in the previous chapter, is a truncated Gaussian probability density function. If the set where it is truncated to, is smaller than the set of interest, the discretization matrices are not dense but do have band structure and are symmetric. Moreover, depending on the discretization scheme, the coefficient matrices have Toeplitz structure, which reduces the memory capacity of an $M \times N$ matrix from order $\mathcal{O}(MN)$ to $\mathcal{O}(M+N)$. We can exploit this structure in our numerical implementation of the optimal control problem.

For the numerical solution of the state equation, we implement a finite element method and study another quite direct discretization, based only on quadrature rules and finite differences as introduced by Lin and Tait (1993) and further developed in Tian and Du (2013). We will discuss both approaches with respect to applicability to our model in Section 6.2.

In Section 6.3, we introduce two approaches for solving the optimal control problem. The first, more direct approach is the *first discretize, then optimize* (FDTO) approach. Here, the optimization problem is discretized using a product-quadrature rule to approximate the involved two-dimensional integrals and the finite differences to approximate the PIDE constraint. We solve the resulting discrete nonlinear optimization problem with a MATLAB solver based on a SQP method.

The second procedure to solve an optimal control problem is the *first optimize, then discretize* (FOTD) approach. In contrast to the FDTO approach, the FOTD approach requires the continuous solution of the minimization problem and afterwards the discretization of the resulting optimality system. In that way, the FOTD approach yields a coupled system of semilinear nonlocal partial differential equations. In this context, we show the Fréchet differentiability of the control-solution operator which is crucial for this approach. We heuristically derive the necessary first order conditions in order to give a broad insight into the structure of our Ramsey model, however we do not implement it. We recommend to solve this system using a gradient projection method as in Sachs and Strauss (2008) or

Tröltzsch (2005, pp.233) and leave this to further studies.
We conclude this chapter in Section 6.4 with a detailed comparison of the local model as introduced in Chapter 3.3 and our nonlocal model with endogenous productivity growth.

It is worth to mention, that we restrict our numerical implementation to the spatially one dimensional case. This is convenient for the application. We can interpret any space interval as a continuum of households, which are sorted according to their initial income, and not according to their actual geographic position.

In order to guarantee the convergence of our numerical schemes, we have to adjust the nonlocal model analyzed in the previous chapter. We use a Moreau-Yosida penalty function with a penalty parameter $\rho_1 > 0$ to substitute the state constraint $k \geq 0$. Moreover, we replace the quite general, continuous function ϕ in the productivity growth operator $\mathcal{P}$ by a differentiable approximation of the absolute value function. In that way, we make sure that $\mathcal{P}$ is Frechet-differentiable with respect to the state variable. The resulting optimal control problem, that we will consider here, is then given as follows:

For an open domain $\Omega \subset \mathbb{R}$ and $T \in \mathbb{N}$, find the optimal control function $c \in \mathcal{U}_{ad}$ and the corresponding optimal state function k which satisfy

$$\begin{aligned}\min_{c,k}\ \mathcal{J}(k,c) := \ & -\int_0^T \int_\Omega U(c(x,t))e^{-\tau t-\gamma x^2}\ dxdt\ + \frac{1}{2\rho_1}\|\min\{0,k\}\|^2_{L^2(0,T;L^2(\Omega))} \\ & + \frac{1}{2\rho_2}\|k(\cdot,T)-k_T(\cdot)\|^2_{L^2(\Omega)}\end{aligned} \tag{6.1}$$

such that the state equation

$$\frac{\partial k}{\partial t} - \beta\int_{\Omega\cup\Omega_\mathcal{I}}(k(y,t)-k(x,t))\Gamma_\varepsilon(x,y)dy +\ \delta k - \mathcal{P}(k) = -c, \tag{6.2}$$

holds on $\Omega \times (0,T)$, the Neumann-type volume constraints

$$-\beta\int_{\Omega\cup\Omega_\mathcal{I}}(k(y,t)-k(x,t))\Gamma_\varepsilon(x,y)dy = 0 \tag{6.3}$$

are valid on the interaction domain-time cylinder $\Omega_\mathcal{I} \times (0,T)$, and such that the initial value constraint

$$k(x,0) = k_0(x) > 0 \tag{6.4}$$

is satisfied on Ω.

For our numerical realization, we choose $\Omega = (1,3)$ and $T = 2$. We define the interaction domain $\Omega_{\mathcal{I}}$ as

$$\Omega_{\mathcal{I}} := (1-\varepsilon, 1] \cup [3, 3+\varepsilon)$$

for an interaction radius $\varepsilon > 0$. Whenever the value of this interaction radius is not clearly defined by the context, we fixed $\varepsilon := 0.5$ and the interaction radius in the productivity-operator as $\mu := 0.3$. Considering the set of feasible controls $\mathcal{U}_{ad}$, which we assumed throughout to be convex and bounded, we requested that the discretized variable is pointwise bounded by 0 and an utopian upper bound. We choose the variance of the kernel function,

$$\Gamma_\nu(x,y) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \mathbb{1}_{B_\nu(x)}(y),\ \nu \in \{\mu, \varepsilon\},$$

relatively small to achieve fairly high diffusive effects. Whenever not stated explicitly, we chose $\sigma = 0.2$ in the following examples.
We consider the nonlinearity defined in (3.12),

$$\mathcal{P}(k)(x,t) = A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy + \xi} t\right) p(k(x,t)), \tag{6.5}$$

where the function $\phi_\eta : \mathbb{R} \to \mathbb{R}_+ \backslash \{0\}$ is defined as

$$\phi_\eta(k) := (\sqrt{k^2+\eta}),$$

depending on a parameter $\eta > 0$. Since η is a fixed parameter, we do not further mention the dependence of $\mathcal{P}$ in the notation. Moreover, for any $k \in \mathbb{R}$ we have

$$||k| - \phi_\eta(k)| \leq \sqrt{\eta},$$

but in contrast to the absolute value function, ϕ_η is differentiable. Note that in this case, we can drop the additional parameter $\xi > 0$ in the denominator. We had to introduce this parameter to make sure that we do not divide by zero. Since ϕ_η is always bounded from below by $\sqrt{\eta} > 0$, it is not necessary any more.
We consider the production function

$$p(k) := \begin{cases} k^\rho, & x > 0.01, \\ 0.01^{\rho-1}k, & x \leq 0.01. \end{cases} \tag{6.6}$$

This function is a Lipschitz continuous approximation to the *Cobb-Douglas production function* for $\rho \in (0,1)$, which is a standard production function in economics. The Cobb-Douglas function satisfies the Inada conditions and is the prime example of a neoclassical production function with decreasing returns to scale.

As an objective function U, we consider a *constant relative risk aversion* (CRRA) utility function. This type of function is commonly used in economic applications. These CRRA functions are characterized by a constant Arrow-Pratt coefficient, which is defined for a twice differentiable and strictly concave function u as

$$APC := -\frac{u''(c)c}{u'(c)}.$$

Integrating gives the family of CRRA utility functions as

$$u(c) = \begin{cases} \frac{c^{1-\theta}-1}{1-\theta}, & 0 \leq \theta < 1, \\ \log(c), & \theta = 1, \end{cases}$$

(cf. Acemoglu, 2009, p.308). Since we stated in assumption 2.4 that the utility function is neoclassical, we consider the slightly adapted function

$$U(c) := \log(c+1).$$

Before we start the numerical implementation of the nonlocal Ramsey model, we point out the diffusive effect of the nonlocal diffusion operator. The following section aims in giving an insight, which parameters are responsible for the strength of diffusion and how the dispersion of the capital stock in the considered economy can be controlled.

6.1 Nonlocal Diffusion - The Role of the Kernel Function

The development of the capital distribution across space over time should be driven by a combination of agglomerative and dispersive effects (cf. Aldashev et al., 2014, p.11). In our model, the agglomerative effects are given by the nonlocal, nonlinear production operator which is heterogeneous in space whenever the initial capital or productivity distribution is heterogeneous in space and which, moreover, depends on time. In contrast to the common spatial Ramsey models, we describe the dispersive effects on the capital distribution over space as integral operator

$$\mathcal{NL}(k)(x,t) = \int_{\Omega \cup \Omega_{\mathcal{I}}} (k(y,t) - k(x,t))\Gamma_{\varepsilon}(x,y)dy, \ (x,t) \in \Omega \times (0,T).$$

In this section, we want to illustrate the diffusive effect of this operator which is highly dependent on the kernel function of the nonlocal integral operator $\mathcal{NL}$,

$$\Gamma_{\varepsilon}(x,y) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \mathbb{1}_{\mathcal{B}_{\varepsilon}(x)}(y).$$

The shape of this kernel function, and hence the strength of the diffusion, depends on the variance σ and the interaction radius ε. Thus, by varying these two parameters, we are able to picture a much larger set of scenarios than the stringent Laplace operator, which is used in the common spatial Ramsey model.

We get a first expectation on how the diffusive impact of the integral operator on the state varibale depends on the variance σ from a result of Briani et al. (2004). In their paper, the authors compare two differential equations,

$$u_t + au_x - bu_{xx} + cu = \lambda \int_{\mathbb{R}} (u(x+z,t) - u(x,t))\gamma_\sigma(z)dz \tag{6.7}$$

and

$$v_t + av_x - bv_{xx} + cv = \frac{\lambda\sigma^2}{2} v_{xx}, \tag{6.8}$$

with the same initial condition

$$u(x,0) = v(x,0) = f_0(x), \; x \in \mathbb{R}.$$

Briani et al. (2004) define the kernel function γ_σ as the Gaussian probability density

$$\gamma_\sigma(z)dz = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{z^2}{2\sigma^2}\right).$$

In this framework, they prove that the difference of the solutions of both equations is of order $\mathcal{O}(T\sigma^3)$, i.e. that the solutions are close for a small variance σ.

Lemma 6.1 (Briani et al., 2004)**:**
Let u be the solution of (6.7) and v the solution of (6.8) with the same initial condition $f_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. Then, if $\sigma \ll 1$,

$$\|u - v\|_{L^\infty(0,T;L^1(\mathbb{R}))} \leq \mathcal{O}(T\sigma^3).$$

In other words, the smaller the variance in the kernel function gets, the closer are the solutions of the two differential equations (6.8) and (6.7). Note that we can in particular choose the parameters $a = 0 = b$, and that the Laplace operator on the right-hand side in (6.8) is weighted with the variance σ. Whenever we choose a nonzero variance, we expect any heterogeneity of the initial value function to be conserved longer in our nonlocal Ramsey model as compared to the local one. The bigger σ, the stronger is the preservation of the inequalities in the initial data. Moreover, we have already pointed out in section 5.3 that the solution of the pure nonlocal equation does not even need to be continuous. Thus, the bigger we choose the variance, the sharper should these discontinuities appear over time. However, the variance should not be chosen too small, since the diffusive effect vanishes for $\sigma \to 0$.

When we consider the impact of the nonlocal diffusion operator with respect to the interaction radius ε, we make an initial guess based on the following result of Du et al. (2012a, p.686). In order to apply it directly to our one-dimensional examples, we adapt it to the one-dimensional case.

Lemma 6.2:
Let $\Omega \subset \mathbb{R}$ denote a bounded domain, independent of the interaction radius ε of the kernel function γ_ε. Denote the second moment of γ_ε as $C := \lim_{\varepsilon \to 0} C_\varepsilon$, where

$$C_\varepsilon = \int_{B_\varepsilon(0)} z^2 \gamma_\varepsilon(|z|) dz.$$

Then, if the kernel function γ_ε satisfies the properties 1. and 2. of Proposition 5.1 such that $\gamma_\varepsilon(|z|) = 0$ for $|z| > \varepsilon$,

$$\lim_{\varepsilon \to 0} \int_{\Omega \cup \Omega_{\mathcal{I}}} \int_{\Omega \cup \Omega_{\mathcal{I}}} \mathcal{D}^* v^T \mathcal{D}^* u dy dx = C \int_\Omega v_x \; u_x \; dx,$$

holds for any $u, v \in H^1(\Omega \cup \Omega_{\mathcal{I}})$ with support in Ω.

Note that C_ε lies in the interval $[0, \sigma^2]$ by definition of the kernel function as Gauss probability density function and that in our case $C \to 0$ as $\varepsilon \to 0$. Hence, we can conclude that the smaller ε gets, the smaller is the diffusive effect of $\mathcal{NL}_\varepsilon$.

We want to make both conjectures on how the parameters σ and ε affect the action of the nonlocal diffusion operator clear in some numerical examples, that should illustrate the impact of both parameters and give a hint, how to choose them in applications. We compare the terminal capital distributions of our nonlocal model with the ones of the local diffusion model by Brito (2001) under varied parameters σ and ε. We consider a homogeneous equation, hence $A_0, c \equiv 0$ on Ω, respectively $\Omega \times (0,T)$, and $\delta = 0$. This yields the capital equation

$$\frac{\partial k(x,t)}{\partial t} = \frac{\partial^2 k(x,t)}{\partial x^2} \text{ on } \Omega \times (0,T)$$

with homogeneous Neumann boundary conditions for the local model. The non-local capital equation is then given as

$$\frac{\partial k(x,t)}{\partial t} = \int_{\Omega \cup \Omega_{\mathcal{I}}} (k(y,t) - k(x,t)) \Gamma_\varepsilon(x,y) dy \quad \text{on } \Omega \times (0,T),$$

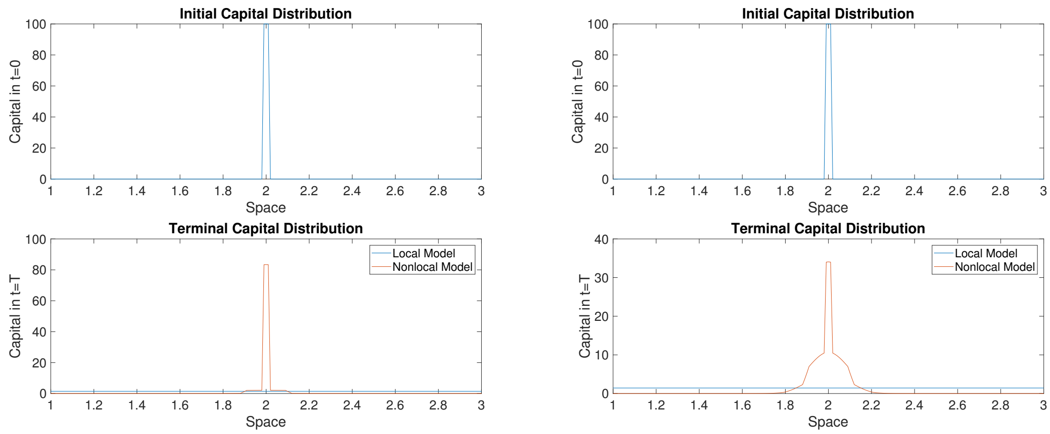
under homogeneous Neumann-type volume constraints,

$$-\int_{\Omega \cup \Omega_{\mathcal{I}}} (k(y,t) - k(x,t)) \Gamma_\varepsilon(x,y) dy = 0 \quad \text{on } \Omega_{\mathcal{I}} \times (0,T).$$

The two Lemmas 6.1 and 6.2 explain the impact of the variance parameter in the kernel function on the strength of the diffusive effect in the nonlocal model. An additional effect of the interaction radius ε becomes clear by definition: The bigger ε is, the wider is the area where smoothing appears. The combination of both parameters generates the distributive effect of the nonlocal diffusion operator. Therefore, the choice of both parameters is crucial for the solution of the nonlocal capital accumulation equation.
For illustration, we consider an extreme initial capital distribution,

$$k_0(x) := \begin{cases} 100, & x \in \mathcal{B}_{1/100}(2) \\ 0, & \text{otherwise.} \end{cases}$$

As illustrated in the figures below, the terminal capital distribution in the local model is evenly distributed across space, thus any heterogeneity of the initial distribution has been smoothed out completely over time.

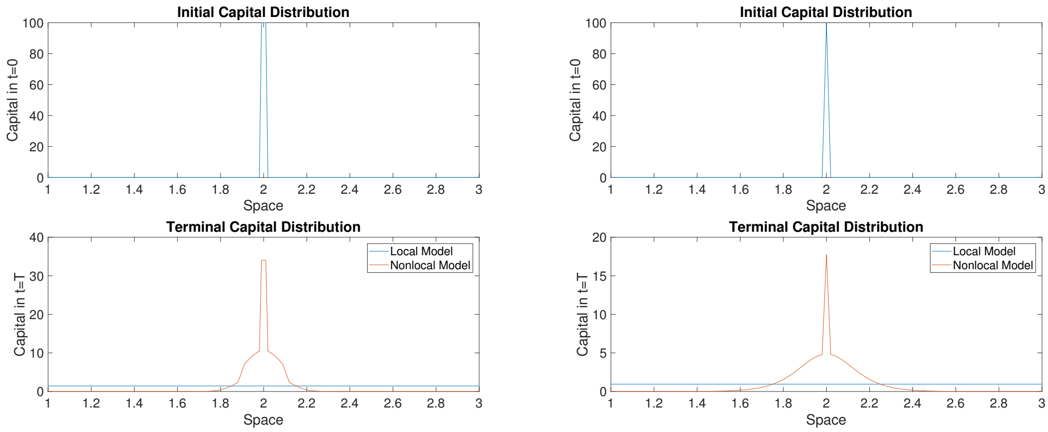


a) Local vs nonlocal diffusion with variance $\sigma = 1$ and interaction radius $\varepsilon = 0.1$

b) Local vs nonlocal diffusion with variance $\sigma = 0.1$ and interaction radius $\varepsilon = 0.1$

Figure 6.1: Impact of the Variance σ

In contrast to that, the smoothing effects in the nonlocal model are much weaker but increase in the length of the interaction radius and decrease in the variance.
Especially when we choose a high value for the variance, e.g. $\sigma = 1$, combined with a small interaction radius like $\varepsilon = 0.1$, there is almost no diffusive effect observable (see Figure 6.1a).
Compared to that, a smaller variance causes a much higher diffusive effect for a fixed ε as illustrated in Figure 6.1b. Thus, we do not only observe a spreading of capital towards the edges of Ω, such that a kind of basis is generated underneath the peak of the initial capital distribution, but the absolute value of the capital stock in the center of Ω is reduced as well.

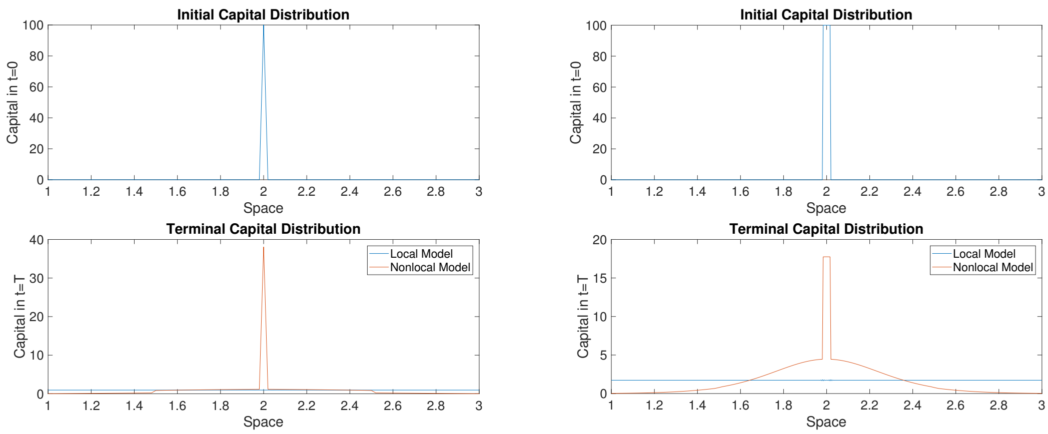


a) Local vs nonlocal diffusion with variance $\sigma = 0.1$ and interaction radius $\varepsilon = 0.1$

b) Local vs nonlocal diffusion with variance $\sigma = 0.1$ and interaction radius $\varepsilon = 1$

Figure 6.2: Impact of the Interaction Radius ε

When we keep the small value of the variance of Γ_ε and increase the interaction radius further, the smoothing effect of the integral operator increases again. The absolute value of the capital stock in the center is reduced and the basis is much wider in Figure 6.2b, as compared to Figure 6.2a.



a) Local vs nonlocal diffusion with variance $\sigma = 0.8$ and interaction radius $\varepsilon = 0.5$

b) Local vs nonlocal diffusion with variance $\sigma = 0.2$ and interaction radius $\varepsilon = 0.5$

Figure 6.3: Two Moderate Scenarios

For a moderate choice of the two diffusion parameters, we obtain two interesting cases. In Figure 6.3a, the center is still much richer as compared to the rest in Ω, but the terminal capital distribution has increased in a quite wide interval around $x = 2$ and has reached the level of the solution of the local model almost everywhere. The case considered in Figure 6.3b appears as the most applicable parameter constellation for a general study of the nonlocal model. The diffusive effect of the nonlocal operator is quite high, since the width of the basis is quite big, and the absolute value of the capital stock in the center point is much smaller as at the initial point in time. Thus, if not stated otherwise, we always choose the parameters

$$\varepsilon = 0.5 \quad \text{and} \quad \sigma = 0.2,$$

for the kernel function in the following examples. Although the solution of the capital accumulation equation in our model is highly sensitive to the choice of these parameters, we leave a better calibration of the kernel function to real life applications to further research.

6.2 Solving the Partial Integro-Differential Equation

Considering the discretization of a parabolic initial-boundary problem with finite elements or finite differences, it can be useful to first discretize either after time or after space, not after both variables at once. Such an approach is called *semi-discretization* or *method of lines*. In general, the *vertical* and the *horizontal method of lines* are distinguished. In the first case, the parabolic differential equation is discretized in space. The resulting system of ordinary differential equations is then discretized in time with appropriate schemes like Runge-Kutta or multi-step methods (cf. Grossmann and Roos, 2005, p.317). In contrast to that, in the horizontal method of lines the parabolic differential equation is first discretized in time. In these, so called Rothe methods, the PDE is approximated by a sequence of elliptic differential equations, which are solved afterwards (cf. Grossmann and Roos, 2005, p.337).
In this section, we introduce two different approaches to solve the seminlinear PIDE in the nonlocal spatial Ramsey model. We first consider a horizontal method of lines, discretizing the time dimension with a *semi-implicit* scheme as introduced by Tröltzsch (2005, p.234) and solving the resulting elliptic equations with a (standard) finite element method. The second alternative that we introduce here, is a vertical method of lines. We discretize the PIDE in space direction, using a quadrature-based finite difference scheme and consider a fully implicit time discretization to approximate the time derivative.

6.2.1 Finite Element Method

The capital accumulation equation in this nonlocal spatial Ramsey model is semilinear. In contrast to semilinear elliptic differential equations, the numerical realization of semilinear parabolic PDE is rather easy. Instead of using a complete implicit scheme for the time discretization, we apply a *semi-implicit* scheme (cf. Tröltzsch, 2005, p.234). This means, that we will approximate the nonlinearity $\mathcal{P}$ at time step t_{i+1} by the evaluation in time step t_i. By doing so, we get a system of linear elliptic differential equations, which then can be solved applying a standard finite element scheme.

The finite element method makes use of the weak formulation of a PDE. Hence, there is no requirement on the strong formulation of a PDE to be satisfied pointwise. On the contrary, *'the finite element method is a particular case of the variational approximation'* and therefore fits the theory of weak solutions best (cf. Dautray and Lions, 1985, p.160-166). This is the reason, why we start our numerical realization with a finite element method.
We use a Galerkin approach to discretize the PIDE in the nonlocal spatial Ramsey model. We consider the standard piecewise linear hat functions

$$\phi_i(x) = \begin{cases} (x - x_{i-1})/h, & \text{for } x \in (x_{i-1}, x_i), \\ (x_{i+1} - x)/h, & \text{for } x \in [x_i, x_{i+1}), \\ 0, & \text{otherwise} \end{cases}$$

for $i \in \mathcal{I}_{\Omega \cup \Omega_{\mathcal{I}}}$, which denotes the index set of the spatial discretization.
To approximate the integrals over the kernel functions Γ_ε and Γ_μ, we use a Gaus-Legendre quadrature. Note, that this quadrature rule does not have equidistant steps in space direction.
Since the initial capital distribution k_0 is only given in Ω, we have to start any routine to solve the capital equation with the determination of $k(x, 0)$ on $\Omega_{\mathcal{I}}$, exploiting the Neumann volume constraint. We can extend the initial capital distribution on the whole nonlocal closure by solving the linear system

$$\mathbb{B}\mathbf{k}^0_{\Omega_{\mathcal{I}}} = b_\Omega,$$

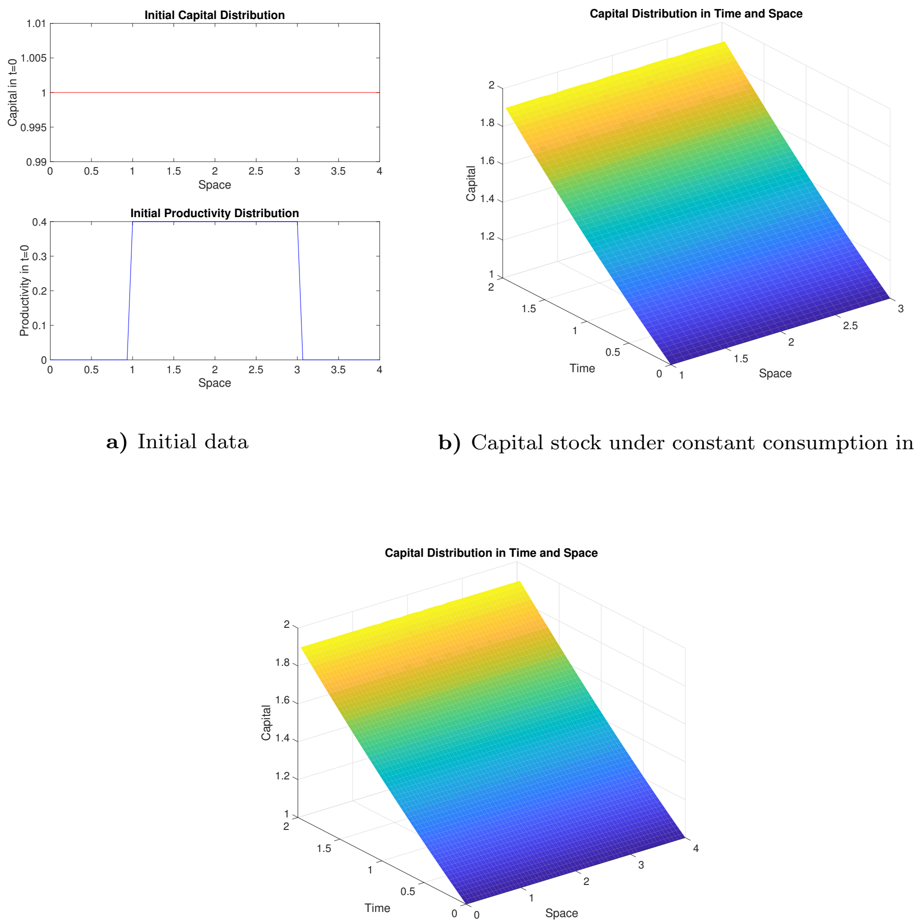
where $\mathbb{B} \in \mathbb{R}^{2N \times 2N}$ is the sum of a diagonal matrix and a block matrix of the form

$$\begin{bmatrix} B_1 & \mathbf{0}_N \\ \mathbf{0}_N & B_2 \end{bmatrix}$$

for two dense matrices $B_1, B_2 \in \mathbb{R}^{N \times N}$ and $\mathbf{k}^0$ denote the discrete approximation on k_0. The integer N denotes the number of steps in the interaction domain.

The Figures 6.4-6.6 below illustrate the distribution of the capital stock across space and over time under a given and constant consumption c. In the following examples, we fix the depreciation rate $\delta = 0.01$, the parameter of the production function $\rho = 0.6$, and $c \equiv 0$. If the value of c is chosen nonzero, it should be less than the initial value of k. We compare the solutions of the PIDE for varying initial capital and productivity distributions k_0 and A_0.

For a constant initial capital distribution $k_0 \equiv 1$ and a constant initial productivity distribution $A_0 \equiv 0.4$ (here, 'constant' means homogeneous in space), we can observe an even growth of capital across space over time (see Figure 6.4). The intersection between Ω and $\Omega_{\mathcal{I}}$ is very smooth and there is no difference between the capital development in the domain of interest or the one in the interaction domain.



a) Initial data

b) Capital stock under constant consumption in Ω

c) Capital stock under constant consumption in $\Omega \cup \Omega_{\mathcal{I}}$

Figure 6.4: Finite Element Solution: Constant Initial Productivity, Constant Initial Capital Distribution

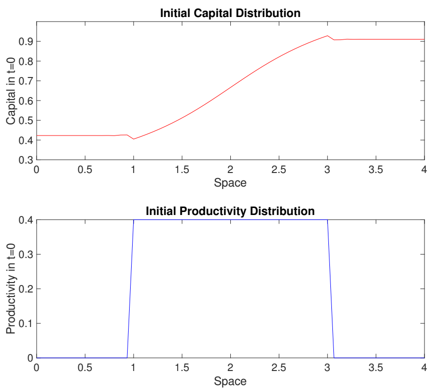
Although the comparison with the local model is the topic of Chapter 6.4, it is worth to mention at this point that for constant initial data the solution of the nonlocal and local capital accumulation equations are very similar. The explanation is quite forward: Since there is no heterogeneity at all, neither the preserving, nor the smoothing effect of the integral, respectively the differential operator arise. Moreover, since the productivity is homogeneous in space, the agglomerative effect of the nonlinear term is insignificant as well.

The situation is different if we choose the same initial productivity distribution, but start with a continuous, non-constant initial capital distribution. In all examples below, we always choose

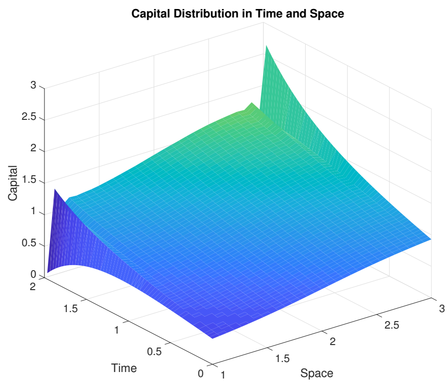
$$k_0(x) = \frac{1}{3}(2 + \arctan(x - 2)) \tag{6.9}$$

whenever we refer to a continuous but heterogeneous initial condition.

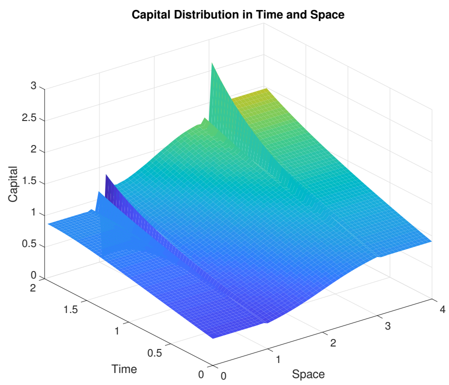
As it becomes obvious in Figure 6.5a, the Neumann volume constraints leads to an uneven transition between Ω and $\Omega_{\mathcal{I}}$. The oscillations appear du to the choice of the standard continuous piecewise linear hat basis functions. These functions are continuous over the boundaries of single elements and hence cannot capture discontinuous solutions. At least in $t = 0$, the oscillations are within the limits of the discretization error.



a) Initial data



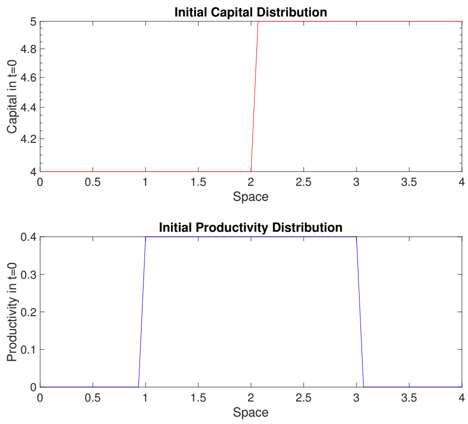
b) Capital stock under constant consumption in Ω



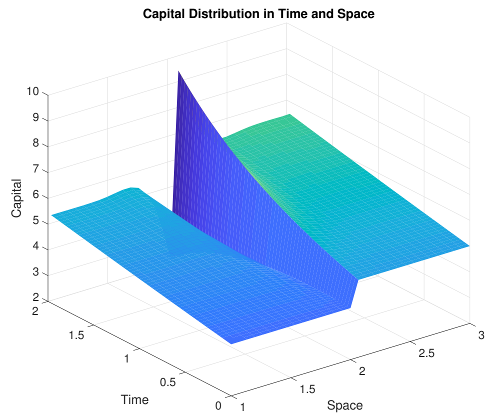
c) Capital stock under constant consumption in $\Omega \cup \Omega_{\mathcal{I}}$

Figure 6.5: Finite Element Solution: Constant Initial Productivity, Heterogeneous but Continuous Initial Capital Distribution

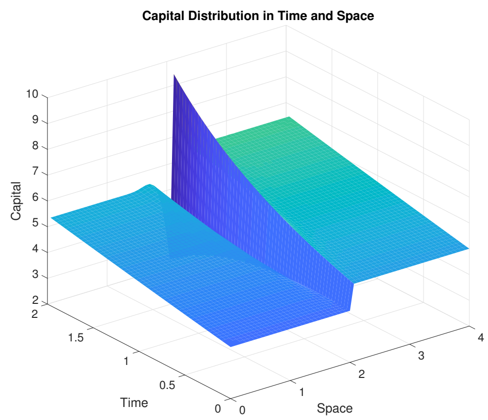
An advantage of the nonlocal model is that any heterogeneities and even discontinuities, that appear over time in the spatial capital distribution, are preserved. However, this property is at the same time a disadvantage, since also numerical inaccuracies are preserved and even increased over time. This leads to numerical results which distort the solution of the model and thus impede any economic interpretation. Nevertheless, Figure 6.5 illustrates that the initial heterogeneity in the capital distribution is perfectly conserved over time. Moreover, we observe a small increase of the heterogeneity in the spatial distribution of the state variable, meaning that the curvature of the terminal capital distribution is stronger than the one of k_0. This increase may be driven by the nonlinearity.



a) Initial data



b) Capital stock under constant consumption in Ω



c) Capital stock under constant consumption in $\Omega \cup \Omega_{\mathcal{I}}$

Figure 6.6: Finite Element Solution: Constant Initial Productivity, Piecewise Constant but Discontinuous Initial Capital Distribution

As already mentioned, one advantage of our model is the possibility to preserve not only heterogeneities, but also discontinuities in the initial data. This conjecture is illustrated in Figure 6.6 for a discontinuous initial capital distribution,

$$k_0(x) := \begin{cases} 4, & x \in (1,2] \\ 5, & x \in (2,3) \end{cases}. \tag{6.10}$$

The previous examples lead to the hypothesis that the pure nonlocal diffusion, weighted with $\beta = 1$, has very little smoothing effects. Indeed, this is confirmed by the result of Briani et al. (2004) which we stated in Lemma 6.1.

As already recommended by Aldashev et al. (2014) or Barro (1990), these few examples already illustrated that any increase of technology, that is heterogeneous over space and depends on time, has agglomerative effects on the capital distribution. We will discuss the interplay of the accumulation effect, driven by the nonlocal productivity-production operator, and the diffusion effects, which are modeled by the nonlocal diffusion operator, in the nonlocal spatial Ramsey model in the following sections.

6.2.2 Quadrature-Based Finite Difference Method

Especially the example illustrated in Figure 6.6 and the boundaries between Ω and $\Omega_{\mathcal{I}}$ in Figure 6.5 indicate that the simple FEM discretization we used is not appropriate since it cannot handle discontinuities. The discretization errors at the discontinuities are even increased due to the productivity-production operator over time. Since the integral operator has only little smoothing effects, the oscillations do not vanish, but become huge over time. This makes an economic analysis and interpretation of the numerical results rather complicated.
The discontinuities at the boundaries between Ω and $\Omega_{\mathcal{I}}$ are per se expectable. Since we chose the volume of the interaction domain $\Omega_{\mathcal{I}}$ much smaller than the volume of the set of interest Ω (in relation 1:3), the homogeneous Neumann-type volume constraint causes jumps of the initial capital distribution between $\Omega_{\mathcal{I}}$ and Ω. This is in line with the economic interpretation of the Neumann volume constraint since the capital flows from Ω to $\Omega_{\mathcal{I}}$ and vice versa must be balanced. Thus, the smaller the interaction domain is, the more capital must be available in $\Omega_{\mathcal{I}}$ in order to counterbalance the flows from Ω.
However, the FEM scheme we introduced above is not the best choice to solve our state constraint. We want to transpire the property of our model to keep and even create discontinuities without any agglomeration of oscillations that are caused by discretization errors. One possibility would be to use a Galerkin approach with discontinuous basis functions in the FEM discretization. A nice introduction is for example given by Chen and Gunzburger (2011).

An alternative method to the FEM is to solve the PIDE with a finite differences method and a quadrature rule. This may be considered as a direct approach, since we discretize the PIDE itself and do not use the variational formulation. We discretize the integral operator with a Riemann quadrature following the approach of Tian and Du (2013). We use a θ-scheme to approximate the derivative in time.

The idea of the quadrature-based finite differences method (QBFDM) is quite simple. Since we have truncated the kernel function in the nonlocal diffusion operator to an ε-interval, we can exploit the additivity of the integral and the symmetry of the kernel function in the following manner:
Let $\Omega := (x_0, x_f)$ and $\Omega_{\mathcal{I}} := (x_0 - \varepsilon, x_0] \cup [x_f, x_f + \varepsilon)$. Given a parameter $\alpha \in [0, 2]$, we consider the nonlocal diffusion operator with truncated support and calculate

$$\begin{aligned}\mathcal{NL}_\varepsilon(k)(x) &= \int_{x-\varepsilon}^{x+\varepsilon} (k(y) - k(x))\Gamma_\varepsilon(x-y)dy \\ &= \int_{-\varepsilon}^{\varepsilon} (k(x+s) - k(x))\Gamma_\varepsilon(s)ds \\ &= \int_{0}^{\varepsilon} \frac{k(x-s) - 2k(x) + k(x+s)}{s^\alpha} s^\alpha \Gamma_\varepsilon(s)ds.\end{aligned}$$

This reformulation makes clear that the nonlocal operator $\mathcal{NL}$ can be viewed as a *'weighted average of second order difference operators'* (Tian and Du, 2013, p.3462).

We consider $N+1$ discretization points in Ω and define $h = (x_f - x_0)/N$. For the time interval $[0, T]$, we choose $M+1$ discretization points and define the step size $\Delta t = T/M$. We set the parameter $\varepsilon = rh$ for a nonnegative integer $r < N/2$, which guarantees that the volume of the interaction set is smaller than the one of the set of interest. Moreover, we define $\mu = qh$ for a nonnegative integer $q < r$. We denote the set of grid points in $\Omega \cup \Omega_{\mathcal{I}}$ by

$$\{x_n := x_0 - \varepsilon + nh\}_{n=-r+1}^{N+r+1}.$$

According to Tian and Du (2013, p.3461), we define $I_j = ((j-1)h, jh)$ for $1 \leq j \leq r$.

Using a simple Riemann sum to approximate the integrals yields the following approximation of the nonlocal diffusion operator in $k_n := k(x_n)$:

$$\mathcal{NL}^h(k_n) = \sum_{j=1}^{r} \frac{k_{n-j} - 2k_n + k_{n+j}}{(jh)^\alpha} \int_{I_j} s^\alpha \Gamma_\varepsilon(s)ds, \ n = 1, ..., N+1, \ \alpha \in [0, 2].$$

This is only well defined if $s^\alpha \Gamma_\varepsilon(s)$ is integrable for $\alpha \in [0, 2]$. In our case, Γ_ε is the truncated density function of the Gaussian probability, hence the second moment is bounded by the variance $\sigma^2 < \infty$. The quadrature order is $\mathcal{O}(h)$ for a fixed ε. As mentioned in Tian and Du (2013), higher order quadrature rules lead to better accuracy, however the choice is limited to quadrature rules with equidistant step

sizes.
We calculate the weight

$$\int_{I_j} \left(\frac{s}{j}\right)^{\alpha} \Gamma_{\varepsilon}(s)ds$$

for every j using a Gauss-Legendre quadrature to improve accuracy. To discretize the nonlinearity, we use a trapezoidal rule to approximate the integrals.
The complete discretization of the nonlocal diffusion operator yields a matrix

$$A^{\varepsilon} \in \mathbb{R}^{(N+1)\times(N+2r+1)},$$

whose entries are given as

$$A^{\varepsilon}_{[i,n]} = \begin{cases} -\frac{2}{h^{\alpha}} \sum_{l=1}^{r} \int_{I_l} \frac{s^{\alpha}}{l^{\alpha}} \Gamma_{\varepsilon}(s)ds, & i = n \\ \frac{1}{h^{\alpha}} \int_{I_m} \frac{s^{\alpha}}{m^{\alpha}} \Gamma_{\varepsilon}(s)ds, & 1 \leq |i-n| \leq r \\ 0, & \text{otherwise} \end{cases}$$

for $i = 1, ..., N+1$ and $n = 1-r, ..., N+1+r$. Here, we denote the entry in the i-th row and n-th column of the matrix by $A^{\varepsilon}_{[i,n]}$. For $r < N/2$, this matrix is a Toeplitz matrix with band structure of the form

$$A^{\varepsilon} = \begin{bmatrix} t_0 & t_1 & t_2 & \dots & t_{2r} & 0 & \dots & 0 \\ 0 & t_0 & t_1 & t_2 & \dots & t_{2r} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots & \ddots & 0 \\ 0 & \dots & 0 & t_0 & t_1 & t_2 & \dots & t_{2r} \end{bmatrix}.$$

Though the matrix may not be sparse, especially for large r, the Toeplitz structure reduces the computational cost as well. The memory capacity of such a matrix is of order $\mathcal{O}((N+1)+(N+2r+1))$ since it is uniquely defined by its first row and column.

To discretize the remaining linear part of the PIDE, we define the matrices M^k and M^c as

$$M^k := \begin{bmatrix} \mathbf{0}_{(N+1)\times r} & (1-\delta)\mathrm{I}_{N+1} & \mathbf{0}_{(N+1)\times r} \end{bmatrix} \in \mathbb{R}^{(N+1)\times(N+2r+1)},$$

and

$$M^c := \begin{bmatrix} \mathbf{0}_{(N+1)\times r} & \mathrm{I}_{N+1} & \mathbf{0}_{(N+1)\times r} \end{bmatrix} \in \mathbb{R}^{(N+1)\times(N+2r+1)}.$$

Here, $\mathbf{0}_{n\times m}$ denotes the zero matrix in $\mathbb{R}^{n\times m}$ and I_n the unit matrix in $\mathbb{R}^{n\times n}$.

To disretize the nonlinearity, we define the vector

$$\Phi^{\mathbf{k}} := (\phi(k^{1-r}), ..., \phi(k^{N+r+1}))' \in \mathbb{R}^{N+2r+1},$$

and the matrix

$$B^{\mu} := \begin{bmatrix} \mathbf{0}_{(N+1)\times(r-q)} & \mathbf{\Gamma}^{\mu} & \mathbf{0}_{(N+1)\times(r-q)} \end{bmatrix} \in \mathbb{R}^{(N+1)\times(N+2r+1)},$$

where $\mathbf{\Gamma}^{\mu} \in \mathbb{R}^{(N+1)\times(N+2q+1)}$ is the Toeplitz matrix with first column

$$\mathbf{\Gamma}^{\mu}_{[:,1]} := \left(\frac{h}{2}\Gamma_{\mu}(\mu), 0, \dots, 0\right)'$$

and first row

$$\mathbf{\Gamma}^{\mu}_{[1,:]} := \left(\frac{h}{2}\Gamma_{\mu}(\mu), h\Gamma_{\mu}(\mu - h), ..., h\Gamma_{\mu}(-\mu + h), \frac{h}{2}\Gamma_{\mu}(-\mu), 0, \dots, 0\right).$$

Moreover, we define the matrix $B^{\varepsilon} := \mathbf{\Gamma}^{\varepsilon} \in \mathbb{R}^{(N+1)\times(N+2r+1)}$, where $\mathbf{\Gamma}^{\varepsilon}$ is constructed analogously to $\mathbf{\Gamma}^{\mu}$, and the vector

$$\mathbf{A}_0 = (\mathbf{0}_r, A_0(x_1), ..., A_0(x_{N+1}), \mathbf{0}_r)' \in \mathbb{R}^{N+2r+1},$$

where $\mathbf{0}_r$ denotes the zero vector in $\mathbb{R}^r$. Note that the vector product

$$B^{\nu}_{[n,:]}\Phi^{\mathbf{k}^m},$$

with

$$\mathbf{k}^m = (k^m_{1-r}, k^m_{2-r}, ..., k^m_1, ..., k^m_{N+1}, ..., k^m_{N+r+1})',$$

where as usual k^m_n denotes the approximation of $k(x_n, t_m)$, is a positive real value for all n and $\nu \in \{\mu, \varepsilon\}$. Combining the discretizations of all components of the operator $\mathcal{P}$, the discretization of the nonlinearity yields a time dependent vector

$$P^{(t_m,\mathbf{k}^m)} \in \mathbb{R}^{N+2r+1}$$

with entries

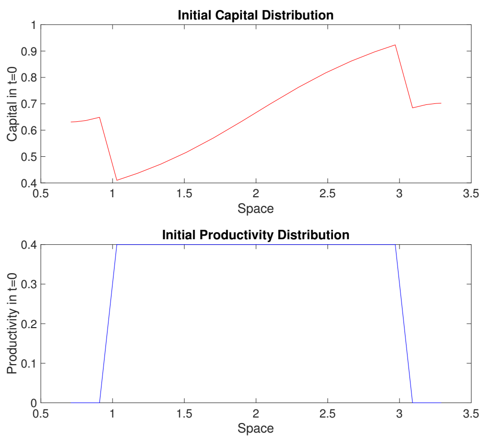
$$P^{(t_m,\mathbf{k}^m)}_{[n]} := \mathbf{A}_{0,[n]} \exp\left((B^{\mu}_{[n,:]}\Phi^{\mathbf{k}^m})(B^{\varepsilon}_{[n,:]}\Phi^{\mathbf{k}^m})^{-1} t_m\right) p(\mathbf{k}^m_{[n]}),$$

for $n = 1-r, ..., N+r+1,\ m = 0, ..., M$. Since $P^{(t_m,\mathbf{k}^m)}$ depends not only on time, but also on the current iterate $\mathbf{k}^m$, it has to be calculated in every step in time during the iteration process.

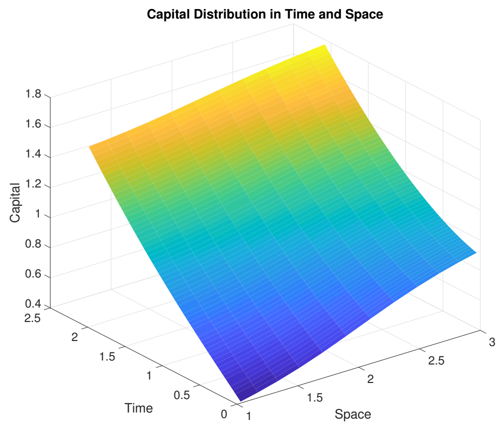
Finally, for given $\theta \in [0,1]$, $t_m := m\Delta t$, and vectors $\mathbf{k}^m, \mathbf{c}^m \in \mathbb{R}^{N+2r+1}$, $(m = 0, ..., M)$, the fully discretized system is given for the matrices M^k, M^c, $A^{\varepsilon} \in \mathbb{R}^{(N+1)\times(N+2r+1)}$ and the vectors $P^{(t_m,\mathbf{k}^m)} \in \mathbb{R}^{N+2r+1}$ $(m = 0, ..., M)$ by

$$
\begin{aligned}
(M^k - \Delta t\theta A^\varepsilon)\mathbf{k}^{m+1} - \Delta t\theta P^{(t_{m+1},\mathbf{k}^{m+1})} - \Delta t\theta M^c\mathbf{c}^{m+1} =& \\
(M^k + \Delta t(1-\theta)A^\varepsilon)\mathbf{k}^m + \Delta t(1-\theta)P^{(t_m,\mathbf{k}^m)} - \Delta t(1-\theta)M^c\mathbf{c}^m& \\
A^\varepsilon\mathbf{k}^0 =& 0.
\end{aligned}
\tag{6.11}
$$

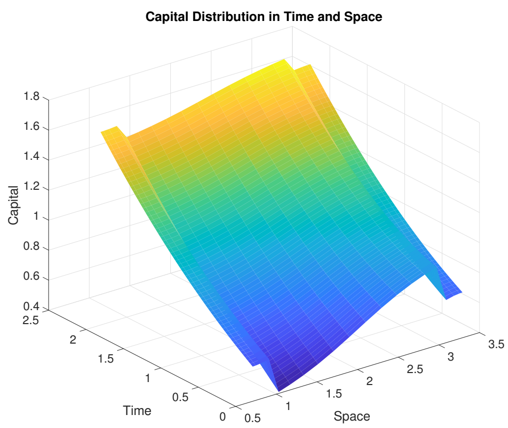
In the examples below, we want to highlight that, if we use this discretization scheme, the oscillations near the discontinuities do not appear. Moreover, we discuss the economic meaning of the numerical solutions.



a) Initial data



b) Capital stock under constant consumption in Ω



c) Capital stock under constant consumption in $\Omega \cup \Omega_{\mathcal{I}}$

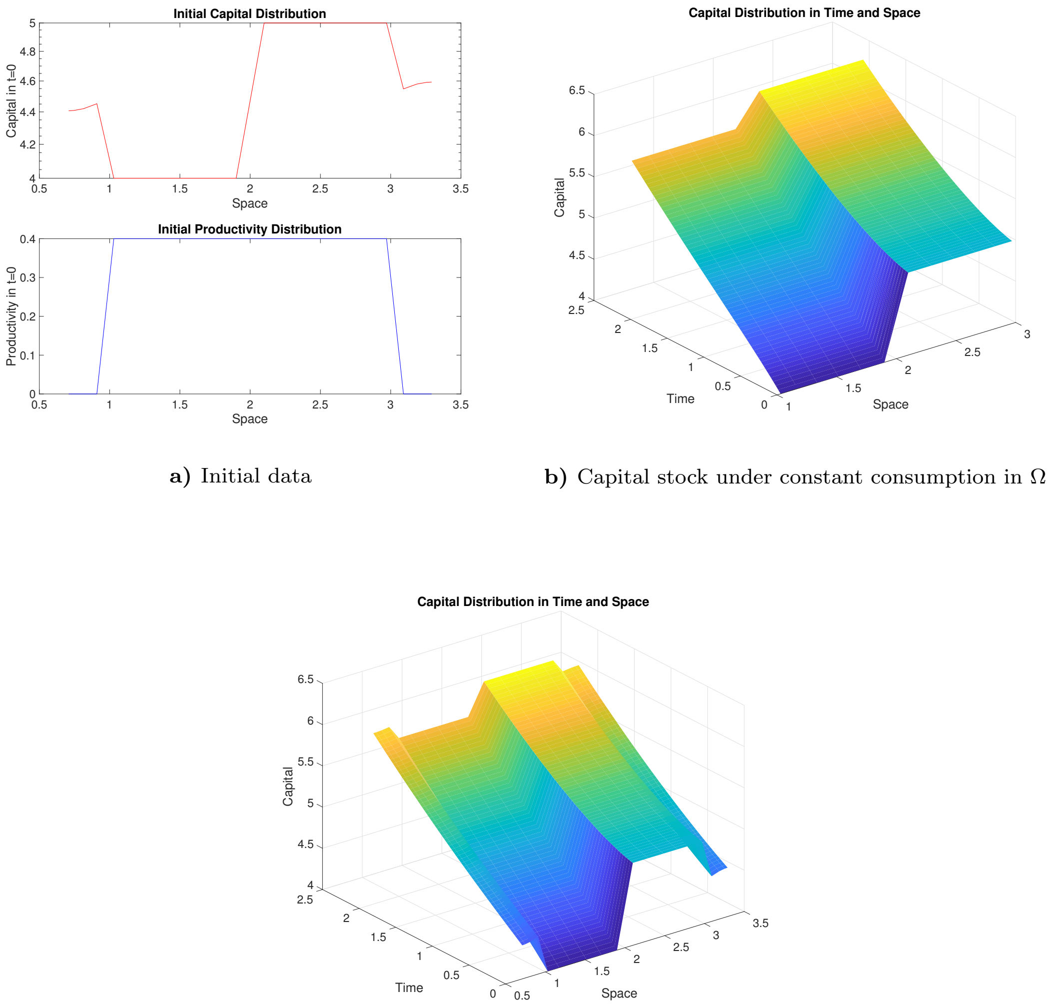
Figure 6.7: Finite Differences Solution: Constant Initial Productivity, Heterogeneous but Continuous Initial Capital Distribution

Figure 6.7 illustrates the time-space development of the capital distribution for zero consumption under the constant initial productivity distribution

$$A_0(x) := \begin{cases} 0.4, & x \in (1,3), \\ 0, & x \in (1-\varepsilon, 1] \cup [3, 3+\varepsilon) \end{cases}$$

and the heterogeneous initial capital distribution as defined in (6.9). The discontinuities between Ω and $\Omega_{\mathcal{I}}$ are sustained over time and oscillations do not appear.

Economic Interpretation of Figure 6.7:
The households in this economy are ordered from poor to rich. In economic terms, the monotone initial capital distribution means that the heterogeneity in the initial income distribution is evenly distributed over space. The constant productivity distribution A_0 implies that all agents in the economy are endowed with the same level of initial knowledge or production skills. Since we assume a zero consumption level, which corresponds to a saving rate of 100%, and a small capital depreciation rate ($\delta = 0.01$) the capital stock naturally increases in every location over time.
The discontinuities between the set of interest Ω and the interaction domain $\Omega_{\mathcal{I}}$ are important for the economic meaning. The capital stock in $\Omega_{\mathcal{I}}$ is almost equally high on the left-hand and the right-hand side of Ω. However, the impact of the interaction of agents located at the left-hand side of Ω with the interaction domain is positive, whereas it is negative on the right-hand side. This means, that the value of $k(\cdot, t)$ in $\Omega_{\mathcal{I}}^- := (0.5, 1]$ is higher than the value of $k(1, t)$ for all points in time $t \in (0, T)$. Thus, interactions with the outside increase the value of capital on the left-hand side of the economy. In contrast to that, the initial capital distribution $k(3, t)$ is much higher than the capital in $\Omega_{\mathcal{I}}^+ := [3, 3, 5)$ for all points in time. This implies that the value of the capital stock decreases in the richer part of the economy whenever any interaction with $\Omega_{\mathcal{I}}^+$ takes place. Thus, the interaction of the poorer households with the outside of Ω is of a different quality than the exchange of production factors on the richer side of the economy. In other words, the poorer agents benefit more from the interaction with the outside of the economy than the richer agents do.
The impact of the interaction domain combined with the weak smoothing effect of the nonlocal diffusion operator leads to a slightly vanishing heterogeneity over time, but the initial disparities do not vanish completely. The slightly catch up of the poorer households can be explained by the dispersive effects of the capital accumulation process over time and the positive impact of the interaction with $\Omega_{\mathcal{I}}$, and not by the agglomerative effect of the production operator, since productivity is equally distributed over space.
In the next example illustrated in Figure 6.8, we observe a similar effect.



a) Initial data

b) Capital stock under constant consumption in Ω

c) Capital stock under constant consumption in $\Omega \cup \Omega_{\mathcal{I}}$

Figure 6.8: Finite Differences Solution: Constant Initial Productivity, Piecewise Constant but Discontinuous Initial Capital Distribution

Here, we again start with a constant initial productivity distribution, but consider the discontinuous initial capital distribution defined in 6.10.

Economic Interpretation of Figure 6.8:
In this economy, there are two distinct, but equally large unions of homogeneous agents. One unit is endowed with a much higher initial capital distribution, but all agents are equally productive in $t = 0$. At first glance, it is obvious that the gap between rich and poor does not vanish over time. It is not evenly smoothed out, but it becomes smaller after time since the poorer agents gain capital faster than the richer ones. Again, the capital stock in the interaction domain is almost equally high in the left part $\Omega_{\mathcal{I}}^-$ and the right part $\Omega_{\mathcal{I}}^+$. The positive impact of the interaction domain on the left-hand side of the economy is once more one reason

for the decreasing disparity between rich and poor.

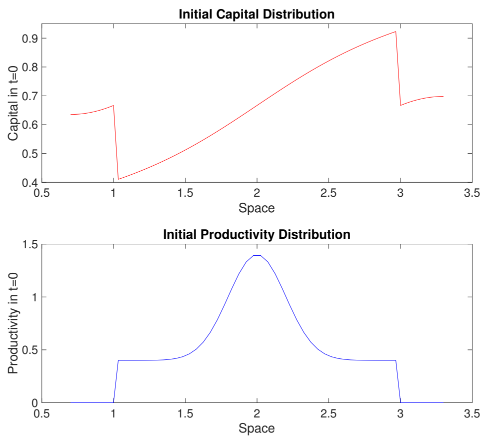
If we consider a heterogeneous initial capital distribution and a non-constant, bell-shaped initial productivity distribution such as

$$A_0(x) := 0.4 \left(1 + \exp\left(\frac{-(x-2)^2}{0.08}\right)\right), \tag{6.12}$$

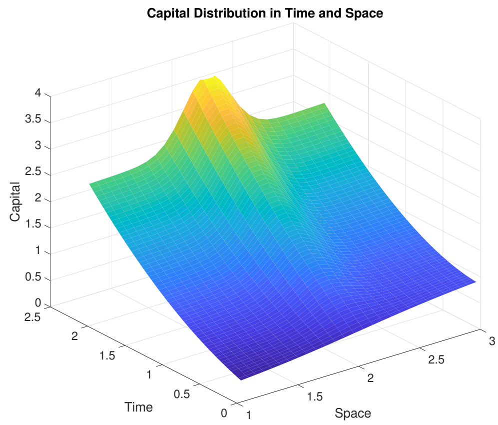
we can observe the severe agglomerate impact of the nonlinear inhomogeneity on the capital stock in Figure 6.9a. Combined with the preservative character of the nonlocal diffusion operator, this leads to a huge increase of capital in locations with higher initial productivity.

Economic Interpretation of Figure 6.9:
This figure illustrates that, under some certain parameter constellations, the agglomerative effect of the technology progress is much stronger than the diffusive effect of the nonlocal operator. The initial productivity distribution A_0 (see Figure 6.9a) describes an economy with one urban center, such as a city or huge industrial area, and its rural surrounding. The households are again ordered from poor to rich. Although the integral operator conserves the heterogeneity in the initial capital distribution, the shape of A_0 has the most significant impact on the space-time development of the capital distribution.
From an economic point of view, this means for a given depreciation and diffusion rate that the increase of capital is higher in more productive areas. Or in other words, for the appearance of wealth, the initial distribution of capital is not as important as the efficiency in production.
This is in line with the observations of Camacho et al. (2008, p.15). The authors also consider a heterogeneous initial capital distribution and a bell-shaped initial technology distribution, similar to our example. Following the model of Brito, Camacho et al. (2008) model the capital mobility across space as a local diffusion process. This implies that the initial disparity in capital does not have any meaning over time. However, they see that the time depending technology preserves the heterogeneity and that its impact on the capital distribution is much stronger than the diffusive effect of the Laplace operator.

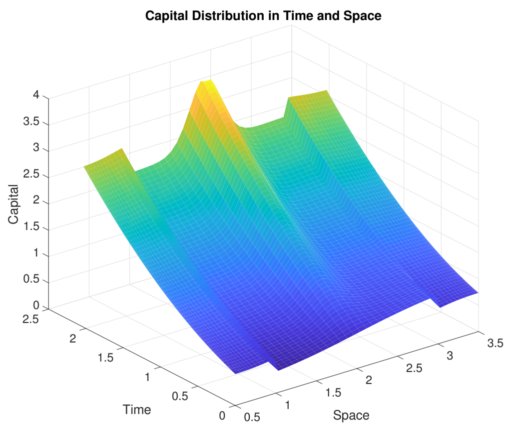
It is remarkable in Figure 6.9, that not only the initial heterogeneity in capital, but also the associated differences in the interaction with $\Omega_\mathcal{I}$ appears to vanish over time. Whereas the capital stock in $\Omega_\mathcal{I}^-$ is much higher than the capital at the left boundary of Ω over the whole time period, the relation of the capital located at the right boundary of Ω and the capital in $\Omega_\mathcal{I}^+$ is reversed in the end. In $t = T$, the impact of the interaction with the outside is the same, on the left- and the right-hand side of the economy.



a) Initial data



b) Capital stock under constant consumption in Ω



c) Capital stock under constant consumption in $\Omega \cup \Omega_{\mathcal{I}}$

Figure 6.9: Finite Differences Solution: Bell-shaped Initial Productivity, Heterogeneous but Continuous Initial Capital Distribution

6.3 Solving the Optimal Control Problem

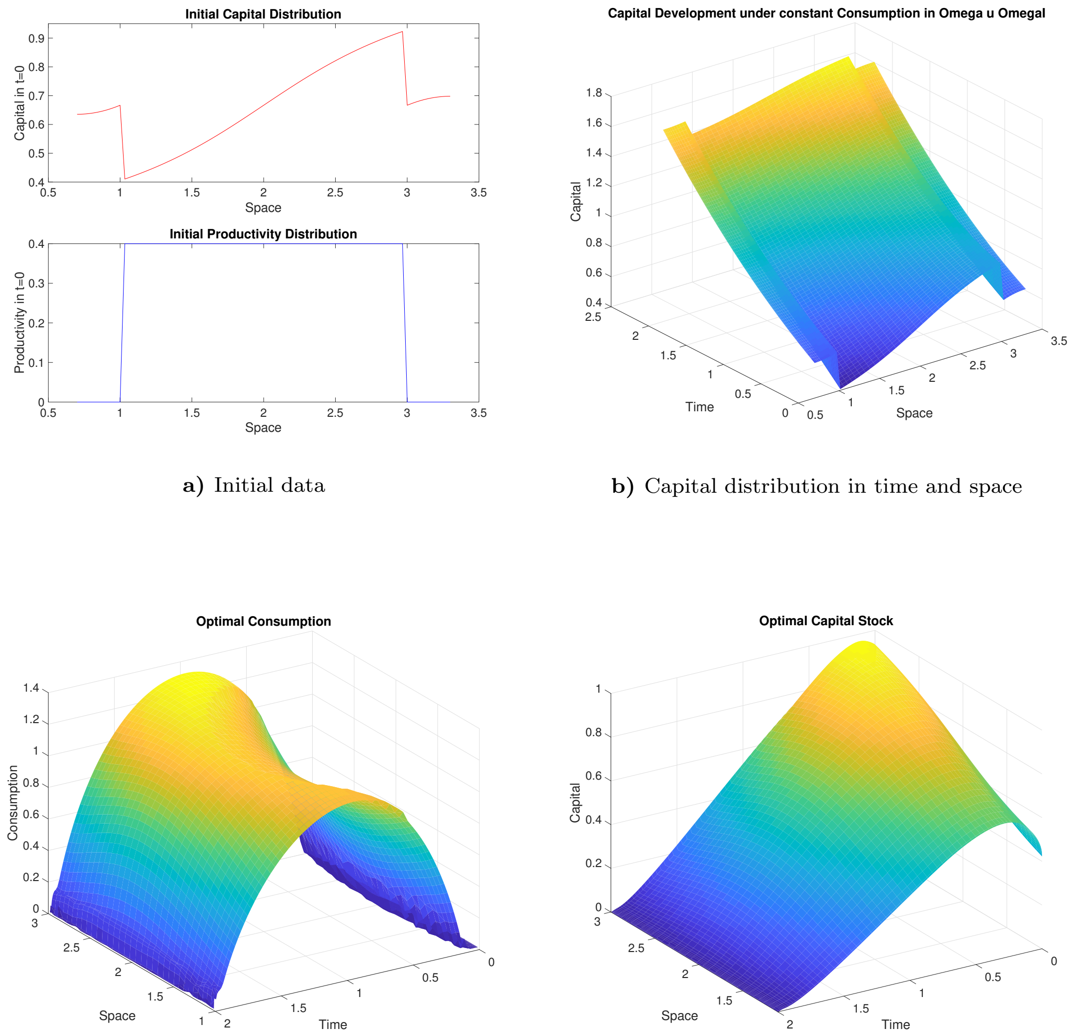
When it comes to the numerical solution of an optimal control problem, there is the choice between two essentially different approaches, the *first discretize, then optimize* (FDTO) and the *first optimize, then discretize* (FOTD) approach. The FDTO approach is also known as *direct approach*. The control problem, which is an optimization problem in a function space, is discretized *'by means of parametric functions with local support, and then the resulting nonlinear program in finitely many optimization variables is solved'* (Sager, 2009, p.6). This approach is straight forward and widely used in practical applications. However, it produces less accurate solutions than the indirect methods. In the *indirect approach*, the FOTD, the necessary first order conditions (NFOC) are derived which are given as a coupled boundary value problem. These NFOC are then solved applying an appropriate discretization. In this approach it is notably possible to exploit the structure of the optimal control problem. The state and adjoint equation may have special structures, which can be exploited to reach higher accuracies or to save computational cost. However, the latter approach demands the differentiability of the optimal control problem, which is often hard to prove.

We start this section with a detailed economic analysis of the numerical results which we achieved with the FDTO approach in Section 6.3.1. Since the main focus of the thesis is the development of a new capital constraint equation for the spatial Ramsey model, we do not pay attention on efficient numerical implementation or advanced optimization techniques.
In Section 6.3.2, we analyze the optimal control problem with respect to Fréchet differentiability of the control-solution operator. We derive the necessary first order conditions, but do not regard any numerical solution of the resulting system of coupled partial integro-differential equations.

6.3.1 First Discretize, Then Optimize

The biggest advantage of the FDTO approach is that there is no need to study the structure of the problem, but the numerical implementation can be started right away. We apply the product rule based on univariate trapezoidal quadrature rules to discretize the double integrals in the objective function. We do not define any terminal condition at k_T. However, we expect that the optimal value of k_T is equal to 0, because the agents do not gain any utility by holding capital over the finite time horizon, but by spending capital on consumption goods. We solve the discretized nonlinear optimization problem with nonlinear equality and inequality constraints using the MATLAB solver *fmincon*. The box constraints on the control variable are consigned to the solver as well.

The following Figure 6.10 illustrates the optimal time-space capital and consumption distributions in an economy with equally distributed poor and rich households. The initial capital distribution is continuous, but strictly monotone and the agents are equally productive over space. Note that the time lines in the Figures 6.10c and 6.10d are shifted from 2 to 0, and space and time axis are interchanged.



a) Initial data

b) Capital distribution in time and space

c) Optimal consumption distribution in time and space

d) Optimal capital distribution in time and space

Figure 6.10: Market Equilibrium: Constant Initial Productivity and Heterogeneous but Continuous Initial Capital Distribution

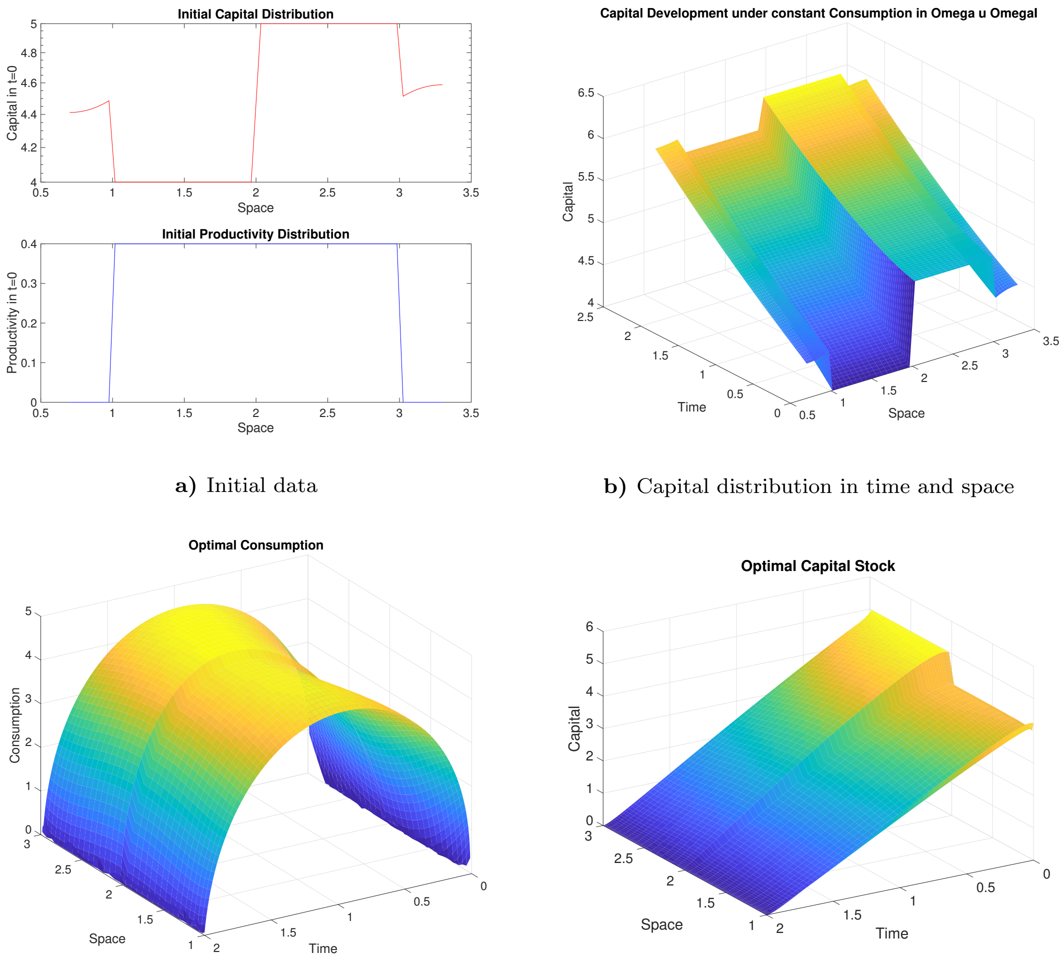
Economic Interpretation of Figure 6.10:

As expected, we observe that the optimal capital path falls down to zero at the end of time. As it can be seen in Figure 6.10d, the capital stock of the poorer households located near $x = 1$ increases at the beginning and starts sinking down to zero after a short time (almost at $t = 0.5$), whereas the capital stock of the richer agents is decreasing over the whole time period.
The consumption distribution over space increases everywhere, until it reaches its zenith after almost half of the time. Though, the consumption level depends on the capital distribution: The richer the agent is, the higher is the consumption. As capital decreases, the consumption level falls down to zero as well, since it has to be financed and households are not allowed to incur debt, implied by the state constraint $k \geq 0$ in every point of time and location in space.

The most important observation in this example is that the heterogeneity in the initial capital distribution is not only preserved in the pure capital accumulation process, as already discussed in the previous section, but is also visible in the optimal capital and consumption distributions. Although the optimal capital distribution becomes more and more homogeneous across space, until it is equally small, the saving behavior of the agents is as heterogeneous as the initial capital distribution. The poorer the agent is, the more capital is saved. This leads to an almost equally distributed consumption level towards the end of the time horizon. However, the absolute consumption is much higher in the locations, where the households are endowed with a higher initial capital stock.
The productivity does not have a big impact on the heterogeneity of the capital and consumption distributions, since it is constant at the beginning.
The optimal solutions depend on the choice of the time and space discount rate. The smaller we choose the time discount rate, the more patient are the agents, and vice versa. We chose the time discount rate in this example as $\tau = 0.01$, which is quite small. This means that the agents are patient enough to save some capital and spend it on consumption after almost the half of time. However, consumption in $t = 1$ is of higher interest than towards the end of time. Thus, the time discounting is essential for the tunnel shaped surface of the consumption distribution in time direction.

When we consider a discontinuous initial capital distribution, we do not expect the discontinuity to disappear. In contrast to the local model, where the mobility of capital over space is modeled as a common diffusion effect, we have shown in Section 5.2 that the regularity of the solution of the capital accumulation equation is of the same quality as the initial data. However, we have already pointed out that the nonlocal diffusion effect does have some smoothing, or spreading impact on the state variable. Thus, we expect that on one hand any discontinuity of the initial capital distribution is preserved, but on the other hand that the disparity between richer and poorer households vanishes slightly.

Economic Interpretation of Figure 6.11:
Indeed, in Figure 6.11b it is shown that the gap between the poorer households at the left-hand side of the economy ($x \in (1, 2]$), and the richer ones at the right-hand side ($x \in (2, 3)$), decreases in time, but persists. We will point out below, that the constant initial productivity distribution is crucial for this result.
For the economic application, this example illustrates that disparities in the initial endowment of production factors, given an equally distributed productivity, are weakened over time, but do not vanish completely. At least in a short time horizon, the lead of the richer households is impossible to be overhauled by the poorer ones, when all other external circumstances are equal.



a) Initial data

b) Capital distribution in time and space

c) Optimal consumption distribution in time and space

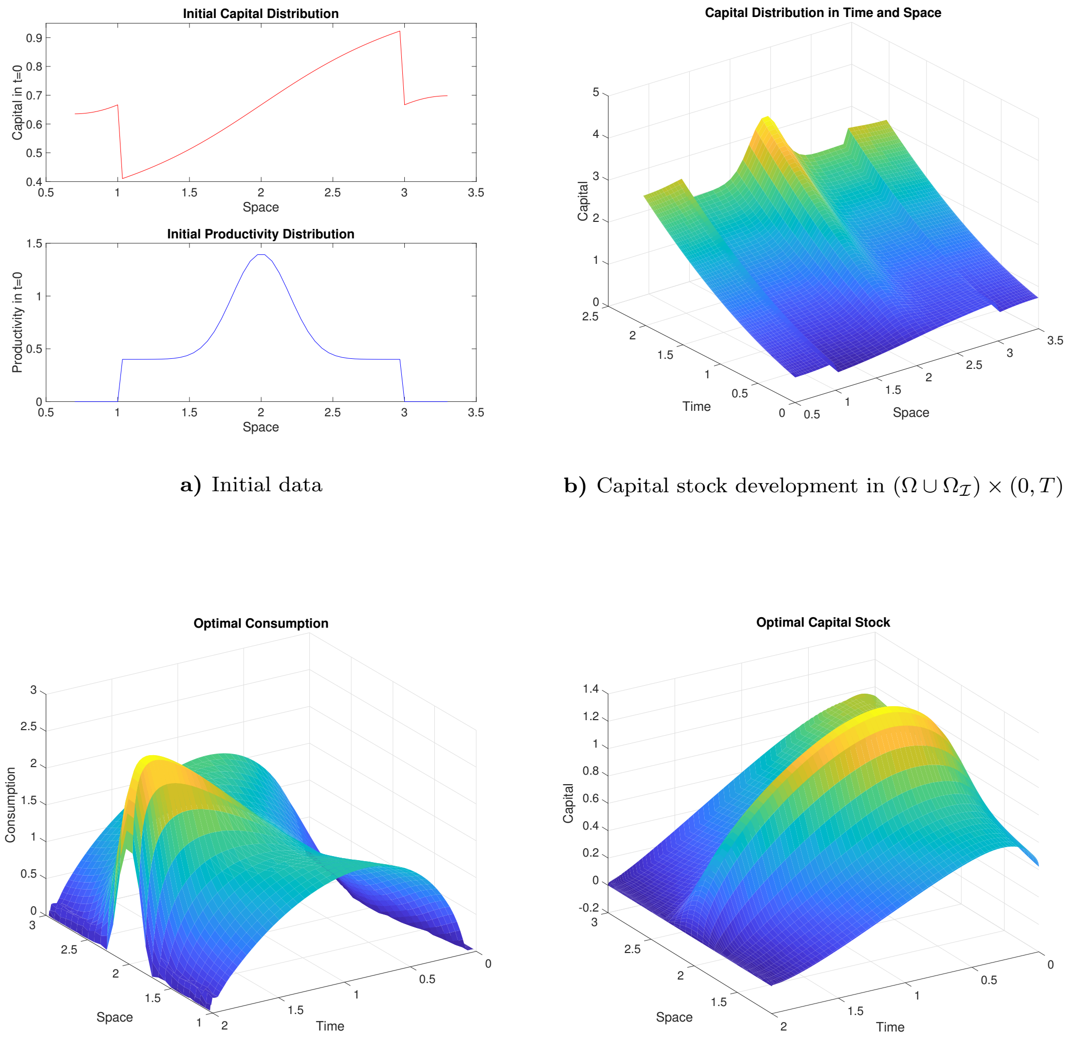
d) Optimal capital distribution in time and space

Figure 6.11: Market Equilibrium: Constant Initial Productivity and Piecewise Constant but Discontinuous Initial Capital Distribution

The last example, that we will consider here, illustrates the impact of the initial productivity distribution and the time depending productivity-production operator. We consider the same initial distributions as in the example illustrated in Figure 6.9. It is remarkable, that the characteristic bell-shape of the initial productivity distribution is carried over to the optimal capital and consumption distributions.

Economic Interpretation of Figure 6.12:
The agents are again evenly distributed from poor to rich across space and there exists an urban center in location 2 with rural surroundings. The benefiting area of the highly productive center is relatively narrow, it reaches from the center in $x = 2$ to $x = 1.5$ to the left, and to $x = 2.5$ to the right.
The productivity-production operator depends on time and on the capital stock in a μ-surrounding of a location x compared to the capital stock in the respective

ε surrounding. The time discount rate in this example is again chosen as $\tau = 0.01$.



a) Initial data

b) Capital stock development in $(\Omega \cup \Omega_{\mathcal{I}}) \times (0, T)$

c) Optimal consumption distribution in time and space

d) Optimal capital distribution in time and space

Figure 6.12: Market Equilibrium: Bell-shaped Initial Productivity and Heterogeneous but Continuous Initial Capital Distribution

In contrast to the example with constant initial productivity, the capital- and time-depending production induces all agents, not only the poorest, to save money at the beginning. Moreover, the increase of capital in the regions with higher initial productivity is the greatest. The most productive agents do not only save the most capital, but they also consume the most towards the end of time. Even at time $t = T$, the consumption level of the agents located in the urban center is strictly positive. The delayed consumption can be much higher, since the capital stock can increase more over time due to the increasing productivity. Due to the time depending increase of productivity, the agents do have an incentive to save money, and to shift consumption into the future. The social planner is absolutely aware of the initial productivity distribution and how it develops over time, and can determine the optimal saving rates of the agents straight from the beginning.

The numerical examples, which we have considered in this section and also in the previous one, illustrate that the market equilibrium in the nonlocal spatial Ramsey model with endogenous productivity growth depends on many parameters and on the interplay of several components of the model. The complexity of the model makes a rigorous calibration with respect to the initial data, the interaction radii ε and μ, and the variance of the kernel function σ inevitable for every application.

6.3.2 First Optimize, Then Discretize

The FOTD approach is based on the calculus of variations or the maximum principle. The optimal state and control trajectories are given as the solution of the necessary first order conditions, a two-point (or even multiple-point) boundary value problem. This so called indirect method provides solutions with high accuracy, as it can exploit the structure of the problem better than direct methods, and is of less computational effort. An additional advantage of the FOTD approach for solving an optimal control problem is that these methods *'provide a better insight into the core of the optimization process in the theory of economic growth'* (c. Ratković, 2016, p.43). However they are sometimes very difficult to solve, especially for complex problems, and require a more elaborate analysis of the optimal control problem.

In this subsection, we want to give an insight into the structure of the nonlocal spatial Ramsey model with endogenous productivity growth. In order to derive the necessary first order conditions of the optimal control problem (6.1)-(6.4), we have to analyze whether the control-state operator, which maps a right-hand side c to the solution of the capital accumulation equation (6.2)-(6.4), is differentiable with respect to the state variable. We discuss here the Fréchet differentiability of this operator. Afterwards, we heuristically derive the necessary first order conditions. We do neither consider the well-posedness of the adjoint system, nor do we implement it. Moreover, we do not pay attention to the appropriate space of the adjoint variable. We follow the Lagrangian approach in Lebherz et al. (2018) to derive a representation of the derivative of the unconstrained optimization problem which we derive from (6.1)-(6.4). This is convenient in this context, since we only want to provide a short outlook to future research.

On the Fréchet Differentiability of the Operators in the Nonlocal Spatial Ramsey Model

We start this analysis with a short introduction to Fréchet differentiability and superposition operators, which play an important role in our optimal control problem. According to Rudin (1973, p.248), we define the Fréchet differential in Banach spaces as follows:

Definition 6.3 (Fréchet Derivative)**:**
Let X, Y be two Banach spaces, $\mathcal{A}$ be an open subset of X and $F : \mathcal{A} \to Y$. Let $a \in \mathcal{A}$. If there exists $\Lambda \in \mathcal{B}(X,Y)$, where $\mathcal{B}(X,Y)$ denotes the Banach space of all bounded and linear functionals on X into Y, such that

$$\lim_{\|x\|_X \to 0} \frac{\|F(a+x) - F(a) - \Lambda x\|_Y}{\|x\|_X} = 0,$$

then Λ is called a Fréchet derivative of F at a. We will use $\partial F(a)$ for the notation. If $\partial F(a)$ exists for every $a \in \mathcal{A}$, and if

$$a \to \partial F(a)$$

is a continuous mapping of $\mathcal{A}$ into $\mathcal{B}(X,Y)$, then F is said to be continuously differentiable.

The following lemma by makes clear that the Fréchet derivative reflects the idea of a linear approximation of F.

Lemma 6.4 (Werner (2007), p.113)**:**
Let X, Y and $\mathcal{A}$ be defined as in Definition 6.3. $F : \mathcal{A} \to Y$ is Fréchet differentiable in $a \in \mathcal{A}$ with derivative $\partial F(a)$ if and only if

$$F(a+x) = F(a) + \partial F(a)(x) + r(x) \quad \text{with} \lim_{\|x\|_X \to 0} \frac{r(x)}{\|x\|_X} = 0.$$

Remark that the Fréchet derivative is unique, if it exists. The next corollary states some fundamental properties and calculation rules of the Fréchet derivative.

Corollary 6.5 (Werner (2007), p.120)**:**
Let X, Y and Z be Banach spaces and $\mathcal{A} \subset X$, $\mathcal{B} \subset Y$ be two open sets.

(a) If F, $G : \mathcal{A} \to Y$ are Fréchet differentiable in $a \in \mathcal{A}$, then $F + G$ and λF ($\lambda \in \mathbb{R}$) are Fréchet differentiable in $a \in \mathcal{A}$ with derivatives given by

$$\partial(F+G)(a) = \partial F(a) + \partial G(a), \quad \partial(\lambda F)(a) = \lambda \partial F(a).$$

(b) Let $F : \mathcal{A} \to Y$ and $G : \mathcal{B} \to Z$ with $F(\mathcal{A}) \subset \mathcal{B}$ be Fréchet differentiable in $a \in \mathcal{A}$ and $F(a) \in \mathcal{B}$. Then, $G \circ F$ is Fréchet differentiable in $a \in \mathcal{A}$ with derivative

$$\partial(G \circ F)(a) = \partial G(F(a)) \circ \partial F(a).$$

The PIDE (6.2) is semilinear. Hence, we have to deal with the nonlinearities when studying (Fréchet) differentiability. The nonlinearity $\mathcal{P}$ can be understood as a composition of integral operators and oprators which are generated by real valued functions. For the latter operators an elaborate theory on their differentiability exists. We will shortly state some fundamental properties of these so called *superposition* or *Nemyzki operators* in Lebesgue spaces as stated in Appell and P.P.Zabrejko (1990), which will be crucial for the analysis of the differentiability of $\mathcal{P}$ and the control-state operator later on.

Definition 6.6 (Superposition Operator)**:**
Let Ω be an arbitrary set and let $f = f(s,u)$ be a function defined on $\Omega \times \mathbb{R}$ mapping to $\mathbb{R}$. Applying f on a given a function $x : \Omega \to \mathbb{R}$ defines the so called superposition *or* Nemyzki *operator*

$$Fx(s) = f(s, x(s)).$$

Theorem 6.7 (Fundamental Theorem on Superposition Operators in Lebesgue Spaces)**:**
The superposition operator F generated by f maps $L^p(\Omega)$ into $L^q(\Omega)$ if and only if there exists a function $\alpha \in L^q(\Omega)$ and a constant $\beta \geq 0$, such that

$$|f(s,u)| \leq \alpha(s) + \beta |u|^{\frac{p}{q}} \tag{6.13}$$

for all $(s,q) \in \Omega \times \mathbb{R}$.

Following Appell and P.P.Zabrejko (1990, p.67), we will refer to the equation (6.13) as *acting condition.*

Theorem 6.8:
Let f be a measurable function and suppose that the superposition operator F generated by f acts from $L^p(\Omega)$ to $L^q(\Omega)$. Then, F is continuous if and only if f is a Carathéodory function.

Appell and P.P.Zabrejko (1990) give a more general version of this theorem, involving functions f that are superpositionally equivalent to some Carathéodory functions, but since we only work with Carathéodory functions, we will keep the theorems and definitions more specified.

Theorem 6.9 (Appell and P.P.Zabrejko (1990), p.75)**:**
Let f be a Carathéodory function and suppose that the superposition operator F generated by f acts from $L^p(\Omega)$ to $L^q(\Omega)$ with $p \geq q$. Then, the following conditions are equivalent:

(i) The operator F satisfies a (local) Lipschitz condition

$$\|Fx_1 - Fx_2\|_{L^q(\Omega)} \leq h(r)\|x_1 - x_2\|_{L^p(\Omega)}, \; x_1, x_2 \in \mathcal{B}_r(L^p(\Omega)).$$

(ii) The function f satisfies a (local) Lipschitz condition

$$|f(s,u_1) - f(s,u_2)| \leq g(s,v)|u_1 - u_2|, \ (|u_i| \leq w, \ i = 1,2),$$

where the function g generates a superposition operator that maps the ball $\mathcal{B}_r(L^p)$ into the ball $\mathcal{B}_r(L^q/L^p)$. Here, L^q/L^p has to be understood as the multiplicator space

$$L^q/L^p = L_{pq/(p-q)}.$$

In the case $p = q$, the function g is essentially bounded.

A nice property of superposition operators is the fact that differentiability of the generating function can be carried over under some circumstances:

Theorem 6.10 (Appell and P.P.Zabrejko (1990), p.78)**:**
Let f be a Carathéodory function and $1 \leq q \leq p < \infty$. Suppose that the superposition operator F generated by f acts from $L^p(\Omega)$ to $L^q(\Omega)$ and let $a \in L^p(\Omega)$. If the superposition operator G generated by

$$g(s,u) = \begin{cases} \frac{1}{u}(f(s,x(s)+u) - f(s,x(s)) & , \ \text{if } u \neq 0 \\ a(s) & , \ \text{if } u = 0, \end{cases}$$

acts from $L^p(\Omega)$ into L^q/L^p and is continuous at the zero-operator, then F is differentiable at x with

$$\partial F(x)h(s) = a(s)h(s).$$

Remark 6.11:
A more intuitive version of it is given by Tröltzsch (2005, p.153). The latter version states that, if the function f is Carathéodory, if the generated superposition operator F maps from $L^p(\Omega)$ to $L^q(\Omega)$ with $0 \leq q < p < \infty$, if the function f is partially differentiable after x for almost every $s \in \Omega$, and if moreover the by f_u generated superposition operator acts from $L^p(\Omega)$ to L^q/L^p, then F is differentiable from $L^p(\Omega)$ to $L^q(\Omega)$ with

$$\partial F(x)h(s) = f_x(s,x(s))h(s). \tag{6.14}$$

This version gives us an explicit formula for the calculation of ∂F.

Especially for $p = \infty$, the differentiability of superposition operators can be concluded straight forward according to the next lemma:

Lemma 6.12 (Tröltzsch (2005), p. 151)**:**
Let f be a Carathéodory function and partially differentiable with respect to x. If f and $\partial_x f$ are moreover bounded in $x = 0$ and locally Lipschitz continuous, then the generated superposition operator F is Fréchet differentiable in $L^\infty(\Omega)$ with

derivative according to equation 6.14).

Now, we have all at hand to start the analysis of differentiability of the control-state operator of the semilinear PIDE (6.2), presupposed the problem has a solution. We start with the analysis of the nominal function, defined as the approximated absolute value function $\phi_\eta : \mathbb{R} \to \mathbb{R}_+ \backslash \{0\},\ k \mapsto \sqrt{k^2+\eta}$, $\eta > 0$, which generates the superposition operator

$$\Phi_\eta(k)(x,t) := \sqrt{k(x,t)+\eta},\ (x,t) \in \mathbb{R} \times [0,T].$$

Lemma 6.13:
The superposition operator Φ_η generated by ϕ_η maps from $L^2((\Omega \cup \Omega_\mathcal{I}) \times (0,T))$ to $L^2((\Omega \cup \Omega_\mathcal{I}) \times (0,T))$ and is Fréchet differentiable from $L^2((\Omega \cup \Omega_\mathcal{I}) \times (0,T))$ in $L^1((\Omega \cup \Omega_\mathcal{I}) \times (0,T))$ with

$$(\partial\Phi_\eta(k)\lambda)(x,t) = \frac{k(x,t)}{\sqrt{k(x,t)^2+\eta}}\lambda(x,t).$$

In addition to that, Φ_η and $\partial\Phi_\eta(k)$ are also locally Lipschitz continuous.

Proof. First, we prove that Φ_η satisfies the acting conditions for $L^2((\Omega \cup \Omega_\mathcal{I}) \times [0,T])$. This is quite forward since for all $\eta > 0$ it is true that

$$|\phi_\eta(k)| = |\sqrt{k^2+\eta}| \leq \sqrt{k^2} + \sqrt{\eta} = \alpha + |k|,$$

for $\alpha \equiv \sqrt{\eta}$. Here, we exploited the concavity of the root function.
The function $\phi_\eta : k \to \sqrt{k^2+\eta}$ is a Carathéodory function which implies together with the acting condition and Theorem 6.8 the continuity of Φ_η.
If the operator Φ_η has a Fréchet derivative, we can calculate this as the Gâteaux derivative of the operator Φ_η as follows:

$$\begin{aligned}
&(\partial\Phi_\eta(k)\lambda)(y) := \lim_{s\to 0} \frac{1}{s} \left(\phi_\eta(k(y)) + s\lambda(y)) - \phi_\eta(k(y))\right) \\
&= \lim_{s\to 0} \frac{1}{s}\left(\sqrt{(k(y)+s\lambda(y))^2+\eta} - \sqrt{k^2(y)+\eta}\right) \\
&= \lim_{s\to 0} \frac{1}{s}\left(\sqrt{(k(y)+s\lambda(y))^2+\eta} - \sqrt{k^2(y)+\eta}\right)\left(\frac{\sqrt{(k(y)+s\lambda(y))^2+\eta} + \sqrt{k^2(y)+\eta}}{\sqrt{(k(y)+s\lambda(y))^2+\eta} + \sqrt{k^2(y)+\eta}}\right) \\
&= \lim_{s\to 0} \frac{2k(y)\lambda(y) + s\lambda(y)^2}{\sqrt{(k(y)+s\lambda(y))^2+\eta} + \sqrt{k^2(y)+\eta}} = \frac{k(y)}{\sqrt{k^2(y)+\eta}}\lambda(y), \quad y \in (\Omega \cup \Omega_\mathcal{I}) \times [0,T].
\end{aligned}$$

This operator is continuous, since it is a linear operator in λ and the fraction in k is always bounded by one. Hence,

$$(\partial\Phi_\eta(k)\lambda)(x,t) = \frac{k(x,t)}{\sqrt{k(x,t)^2+\eta}}\lambda(x,t)$$

defines the Gâteaux derivative of Φ_η. In a next step, we show that $\partial\Phi_\eta(\cdot)$ is continuous as a map between $L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))$ and $\mathcal{B}(L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)), L^1((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)))$ which then implies the Fréchet differentiability. Therefore, we consider a sequence

$$k_n \to k_0 \text{ in } L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))$$

and estimate the operator norm in $\mathcal{B}(L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)), L^1((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)))$ as follows

$$\begin{aligned}
&\|\partial\Phi_\eta(k_n)-\partial\Phi_\eta(k_0)\|_{\mathcal{B}(L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)),L^1((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)))}\\
=&\sup_{\|\lambda\|_{L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))}\le 1}\|\partial\Phi_\eta(k_n)\lambda-\partial\Phi_\eta(k_0)\lambda\|_{L^1((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))}\\
\le&\sup_{\|\lambda\|_{L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))}\le 1}\|\partial\Phi_\eta(k_n)-\partial\Phi_\eta(k_0)\|_{L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))}\|\lambda\|_{L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))}\\
=&\left(\int_0^T\int_{\Omega\cup\Omega_{\mathcal{I}}}\left(\frac{k_n(x,t)}{\sqrt{k_n^2(x,t)+\eta}}-\frac{k_0(x,t)}{\sqrt{k_0^2(x,t)+\eta}}\right)^2 dxdt\right)^{\frac{1}{2}}\\
\overset{!}{\longrightarrow}&\ 0.
\end{aligned}$$

We show the convergence with a contradiction, exploiting the almost everywhere convergence of a subsequence of k_n to k_0. Therefore, we assume that there exists a sequence $k_n \to k_0$ in $L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))$ such that

$$\|\partial\Phi_\eta(k_n)-\partial\Phi_\eta(k_0)\|_{\mathcal{B}(L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)),L^1((\Omega\cup\Omega_{\mathcal{I}})\times(0,T)))}\nrightarrow 0.$$

Then, we can choose a subsequence (w.l.o.g) $k_n \to k_0$ in $L^2((\Omega\cup\Omega_{\mathcal{I}})\times(0,T))$ and $k_n \to k_0$ almost everywhere. For this subsequence, we estimate

$$\begin{aligned}
&\int_0^T\int_{\Omega\cup\Omega_{\mathcal{I}}}\left(\frac{k_n(x,t)}{\sqrt{k_n^2(x,t)+\eta}}-\frac{k_0(x,t)}{\sqrt{k_0^2(x,t)+\eta}}\right)^2 dxdt\\
&\le\int_0^T\int_{\Omega\cup\Omega_{\mathcal{I}}}\left(\frac{k_n(x,t)}{\sqrt{k_n^2(x,t)+\eta}}-\frac{k_0(x,t)}{\sqrt{k_n^2(x,t)+\eta}}\right)^2 dxdt\\
&\quad+\int_0^T\int_{\Omega\cup\Omega_{\mathcal{I}}}\left(\frac{k_0(x,t)}{\sqrt{k_n^2(x,t)+\eta}}-\frac{k_0(x,t)}{\sqrt{k_0^2(x,t)+\eta}}\right)^2 dxdt.
\end{aligned}$$

For the second term it is true that

$$\begin{aligned}
&\int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \left(\frac{k_0(x,t)}{\sqrt{k_n^2(x,t)+\eta}} - \frac{k_0(x,t)}{\sqrt{k_0^2(x,t)+\eta}} \right)^2 dxdt \\
&= \int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} |k_0(x,t)|^2 \left| \frac{1}{\sqrt{k_n^2(x,t)+\eta}} - \frac{1}{\sqrt{k_0^2(x,t)+\eta}} \right|^2 dxdt \\
&\leq \int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} k_0^2(x,t) \left| \frac{1}{\sqrt{\eta}} - \frac{1}{\sqrt{k_0^2(x,t)+\eta}} \right|^2 dxdt \\
&\leq \int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \frac{k_0^2(x,t)}{\eta} dxdt
\end{aligned}$$

and we can follow the convergence of

$$\int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \left(\frac{k_0(x,t)}{\sqrt{k_n^2(x,t)+\eta}} - \frac{k_0(x,t)}{\sqrt{k_0^2(x,t)+\eta}} \right)^2 dxdt \to 0 \ (n \to \infty)$$

with Lebesgue's dominated convergence theorem and the point-wise convergence of $k_n \to k_0$. The first term also yields

$$\begin{aligned}
&\int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \left| \frac{k_n(x,t)}{\sqrt{k_n^2(x,t)+\eta}} - \frac{k_0(x,t)}{\sqrt{k_n^2(x,t)+\eta}} \right|^2 dxdt \\
&= \int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \left| \frac{1}{\sqrt{k_n^2(x,t)+\eta}} \right|^2 |k_n(x,t) - k_0(x,t)|^2 dxdt \\
&\leq \int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \frac{1}{\eta} |k_n(x,t) - k_0(x,t)|^2 dxdt,
\end{aligned}$$

which converges to 0 for $n \to \infty$ as $k_n \to k_0$ in $L^2((\Omega \cup \Omega_\mathcal{I}) \times (0,T))$ for $n \to \infty$. Hence, we have

$$\int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \left(\frac{k_n(x,t)}{\sqrt{k_n^2(x,t)+\eta}} - \frac{k_0(x,t)}{\sqrt{k_0^2(x,t)+\eta}} \right)^2 dxdt \to 0 \quad (n \to \infty),$$

which is a contradiction to

$$\|\partial\Phi_\eta(k_n) - \partial\Phi_\eta(k_0)\|_{\mathcal{B}(L^2((\Omega\cup\Omega_\mathcal{I})\times(0,T)),L^1((\Omega\cup\Omega_\mathcal{I})\times(0,T)))} \nrightarrow 0.$$

Thus, we can conclude that $\partial\Phi_\eta(k)$ is continuous between $L^2((\Omega \cup \Omega_\mathcal{I}) \times (0,T))$ and $\mathcal{B}(L^2((\Omega \cup \Omega_\mathcal{I}) \times (0,T)), L^1((\Omega \cup \Omega_\mathcal{I}) \times (0,T)))$, which yields that $\partial\Phi_\eta(k)$ is not only the Gâteaux but also the Fréchet derivative of Φ_η.
In order to prove the Lipschitz continuity, we exploit the Lipschitz continuity of

the root function on $[\eta, \infty)$ for every $\eta > 0$ and calculate for $k_1, k_2 \in [-M, M]$,

$$\begin{aligned}|\phi_\eta(k_1) - \phi_\eta(k_2)| &= |\sqrt{k_1^2 + \eta} - \sqrt{k_2^2 + \eta}| \\ &\leq L|k_1^2 + \eta - (k_2^2 + \eta)| \\ &= L|k_1^2 - k_2^2| = L|(k_1 - k_2)(k_1 + k_2)| \\ &\leq L_1|k_1 - k_2|\end{aligned}$$

where L denotes the Lipschitz constant of the root function and $L_1 := 2ML$. In order to prove the Lipschitz continuity of the derivative, we estimate for every $\eta > 0$ and $k_1, k_2 \in [-M, M]$,

$$\begin{aligned}&\left|\phi_\eta'(k_1) - \phi_\eta'(k_2)\right| \\ &= \left|\frac{k_1}{\sqrt{k_1^2 + \eta}} - \frac{k_2}{\sqrt{k_2^2 + \eta}}\right| = \left|\frac{\sqrt{k_2^2 + \eta}\ k_1\ - \sqrt{k_1^2 + \eta}\ k_2}{\sqrt{k_1^2 + \eta}\sqrt{k_2^2 + \eta}}\right| \\ &\leq \frac{1}{\eta}\left|\sqrt{k_2^2 + \eta}k_1 - \sqrt{k_1^2 + \eta}k_2\right| \\ &= \frac{1}{\eta}\left|(\sqrt{k_2^2 + \eta} - \sqrt{k_1^2 + \eta})(k_1 - k_2) - k_1\sqrt{k_1^2 + \eta} + k_2\sqrt{k_2^2 + \eta}\right| \\ &\leq \frac{2\sqrt{M^2 + \eta}}{\eta}\ |k_1 - k_2| + \left|k_1\sqrt{k_1^2 + \eta} - k_2\sqrt{k_1^2 + \eta} + k_2\sqrt{k_1^2 + \eta} - k_2\sqrt{k_2^2 + \eta}\right| \\ &\leq \frac{2\sqrt{M^2 + \eta}}{\eta}\ |k_1 - k_2| + \left|(k_1 - k_2)\sqrt{k_1^2 + \eta}\right| + \left|\left(\sqrt{k_1^2 + \eta} - \sqrt{k_2^2 + \eta}\right) k_2\right| \\ &\leq \frac{2\sqrt{M^2 + \eta}}{\eta}\ |k_1 - k_2| + \sqrt{M^2 + \eta}|k_1 - k_2| + ML_1|k_1 - k_2| \\ &\leq L_2|k_1 - k_2|\end{aligned}$$

for $L_2 \geq \frac{2\sqrt{M^2+\eta}}{\eta} + \sqrt{M^2 + \eta} + ML_1$. Applying Theorem 6.9 concludes the proof.
□

For initial data $c \in L^2(\Omega \times (0, T))$, the highest regularity of the weak solution of our nonlocal capital accumulation equation that we can expect on the union of Ω and $\Omega_\mathcal{I}$ is $\mathcal{C}([0, T]; V_c(\Omega \cup \Omega_\mathcal{I}))$, where V_c is equivalent to the L^2, since the nonlocal volume constraints are in integral form and do not drive the regularity of the solution. Analyzing the productivity-production operator $\mathcal{P}$ according to Fréchet differentiability would hence be meaningful in $L^2((\Omega \cup \Omega_\mathcal{I}) \times (0, T))$. According to Appell and P.P.Zabrejko (1990, p.64), the operator $\mathcal{P}$ must be affine if it was Fréchet differentiable from L^2 to L^2. This is not the case. We could try to show the differentiability of $\mathcal{P}$ from $L^2((\Omega \cup \Omega_{\mathcal{I}}) \times (0,T))$ to $L^1((\Omega \cup \Omega_{\mathcal{I}}) \times (0,T))$, which would not be appropriate with respect to the differentiability of the control-state operator. However, we have shown in Theorem 5.8 that in the case of Dirichlet-type volume constraints, the weak solution of the capital accumulation equation is L^∞ on Ω and L^2-regular only on $\Omega_{\mathcal{I}}$ for $c \in L^\infty(\Omega \times (0,T))$. For the problem with Neumann-type volume constraints we expect the same regularity, since the integral operator on $\Omega_{\mathcal{I}}$ does neither drive nor reduce the regularity of the solution. Thus, we prove the Fréchet differentiability between the solution space $\mathcal{V}^\infty$ and $L^\infty(\Omega \times (0,T))$, where we have defined the Banach space

$$\mathcal{V}^\infty := \{u \in \mathcal{C}([0,T]; V_c(\Omega \cup \Omega_{\mathcal{I}})) : \ ess \sup_{(x,t)\in\Omega\times(0,T)} |u(x,t)| < \infty\},$$

endowed with the norm

$$\|u\|_{\mathcal{V}^\infty} := \|u\|_{\mathcal{C}([0,T];V_c(\Omega\cup\Omega_{\mathcal{I}}))} + \|u\|_{L^\infty(\Omega\times(0,T))}$$

in Remark 5.9.

Lemma 6.14:
If the (Carathéodory) production function $p = p(k) : \mathbb{R} \to \mathbb{R}$ is differentiable with respect to k, locally Lipschitz continuous, and bounded, then it is true that the productivity operator $\mathcal{P}$ is Fréchet differentiable from $\mathcal{V}^\infty$ to $L^\infty(\Omega \times (0,T))$.

Proof. The proof is straight forward. In a first step, we show that the composition

$$k(\cdot) \mapsto \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,\cdot))\Gamma_\mu(\cdot,y)dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,\cdot))\Gamma_\varepsilon(\cdot,y)dy} \cdot \right)$$

is Fréchet differentiable between $\mathcal{V}^\infty$ to $L^\infty(\Omega \times (0,T))$. First note that the integral operator

$$k \mapsto \int_{\Omega\cup\Omega_{\mathcal{I}}} k(y,\cdot)\Gamma_\nu(x,y)dy, \ x \in \Omega$$

is linear and continuous from $L^1((\Omega \cup \Omega_{\mathcal{I}}) \times (0,T)) \to L^\infty(\Omega \times (0,T))$ for $\nu \in \{\mu, \varepsilon\}$ and hence Fréchet differentiable.
We have already shown that $\Phi_\eta(k) = \phi_\eta(k(\cdot,\cdot))$ defines a Fréchet differentiable superposition operator that maps $\mathcal{V}^\infty$ to $L^2((\Omega \cup \Omega_{\mathcal{I}}) \times (0,T))$ and which is Fréchet differentiable from $\mathcal{V}^\infty$ to $L^1((\Omega \cup \Omega_{\mathcal{I}}) \times (0,T))$. Since compositions of Fréchet differentiable operators are Fréchet differentiable,

$$k \mapsto \int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,\cdot))\Gamma_\nu(x,y)dy, \ x \in \Omega, \ \nu \in \{\mu, \varepsilon\},$$

is Fréchet differentiable from $\mathcal{V}^\infty$ to $L^\infty(\Omega \times (0,T))$ with derivative

$$\hat{k} \mapsto \int_{\Omega \cup \Omega_{\mathcal{I}}} \phi'_\eta(k(y,\cdot))(\hat{k}(y,\cdot))\Gamma_\nu(x,y)dy,\ k \in \mathcal{V}^\infty,\ x \in \Omega,\ \nu \in \{\mu, \varepsilon\}.$$

The superposition operator

$$k \mapsto \frac{1}{k+\epsilon}$$

is Fréchet differentiable in $L^\infty(\Omega \times (0,T))$ to $L^\infty(\Omega \times (0,T))$ for every $\epsilon > 0$ according to Lemma 6.12. The exponential function also generates a superposition operator between $L^\infty(\Omega \times (0,T))$ and $L^\infty(\Omega \times (0,T))$. Since we have chosen μ and ε such that the fraction

$$\frac{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}$$

is always bounded by 1, we can exploit the boundedness and local Lipschitz continuity of the exponential function on compact sets. This yields the differentiability of the generated superposition operator according to Lemma 6.12. Thus, we can conclude the Fréchet differntiability of the composition

$$k(x,t) \mapsto \exp\left(\frac{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} t\right)$$

in k with derivative

$$\hat{k}(x,t) \mapsto \exp\left(\frac{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} t\right) \cdot \left[\frac{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi'_\eta(k(y,t))\hat{k}(y,t)\Gamma_\mu(x,y)dy}{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} - \frac{\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy\ \int_{\Omega \cup \Omega_{\mathcal{I}}} \phi'_\eta(k(y,t))\hat{k}(y,t)\Gamma_\varepsilon(x,y)dy}{\left(\int_{\Omega \cup \Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy\right)^2}\right].$$

Since we assumed p to be a differentiable and Lipschitz continuous Carathéodory function, the generated superposition operator $p(k)$ is also Fréchet differentiable between $L^\infty(\Omega \times (0,T))$ and $L^\infty(\Omega \times (0,T))$, especially between $\mathcal{V}^\infty$ and $L^\infty(\Omega \times (0,T))$. Since every composition and product of Fréchet differentiable operators is Fréchet differentiable according to Corollary 6.5, we can conclude the proof.

□

The derivatives derived in the proof allow to construct the Fréchet derivative of the productivity-production operator straightforwardly.

Remark 6.15:
According to Corollary 6.5(b), we can calculate the Fréchet derivative of $\mathcal{P}$ *with*

respect to k *in direction* $\hat{k} \in \mathcal{V}^\infty$ *applying the chain rule. Note that* $\partial\mathcal{P}(k)$ *is indeed linear in* $\hat{k}$*, though nonlocal:*

$$\partial\mathcal{P}(k)(\hat{k})[x,t] = A_0(x)\exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}}\phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}}\phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}t\right)p'(k(x,t))\hat{k}(x,t)$$

$$+\,t\;A_0(x)\exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}}\phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}}\phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}t\right)p(k(x,t))$$

$$\cdot\left[\frac{\int_{\Omega\cup\Omega_\mathcal{I}}\phi'_\eta(k(y,t))\hat{k}(y,t)\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}}\phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}\right.$$

$$\left.-\frac{\int_{\Omega\cup\Omega_\mathcal{I}}\phi_\eta(k(y,t))\Gamma_\mu(x,y)dy\;\int_{\Omega\cup\Omega_\mathcal{I}}\phi'_\eta(k(y,t))\hat{k}(y,t)\Gamma_\varepsilon(x,y)dy}{\left(\int_{\Omega\cup\Omega_\mathcal{I}}\phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy\right)^2}\right],$$

$(x,t) \in \Omega \times (0,T)$.

In order to derive the necessary first order conditions, we have to show that the control-state operator, that maps a control to the solution of the capital accumulation equation in the nonlocal spatial Ramsey model on bounded domains, is Féchet differentiable. A fundamental theorem, which we use in the proof, is the theorem on implicit functions in Banach spaces.

Theorem 6.16 (Implicit Function Theorem in Banach Spaces)**:**
Let X, Y*, and* W *be three Banach spaces,* $k \geq 1$, $A \subset X \times Y$ *an open set, let* $(x_0, y_0) \in A$ *and* $f : A \to W$ *be a* $\mathcal{C}^k$ *map such that* $f(x_0, y_0) = 0$*. Assume that* $D_y f(x_0, y_0) : Y \to W$ *is a bounded invertible linear transformation. Then there is an open neighborhood* U_0 *of* x_0 *in* X *such that for all connected open neighborhoods* U *of* x_0 *contained in* U_0*, there is a unique and continuous* $u : U \to Y$ *such that* $u(x_0) = y_0$, $(x, u(x)) \in A$ *and* $f(x, u(x)) = 0$ *for all* $x \in U$*. Moreover* u *is necessarily* $\mathcal{C}^k$ *and*

$$Du(x) = -D_y f(x,u(x))^{-1} D_x f(x,u(x)) \text{ for all } x \in U$$

(cf. Driver, 2003, p.436).

Remark 6.17:
Tröltzsch (2005) outline two different approaches how to derive the Fréchet differentiability of the control-state operator

$$G : L^\infty(\Omega \times (0,T)) \to \mathcal{V}^\infty,\; c \mapsto k,$$

that maps a control c *to the solution of the nonlocal capital accumulation equation (6.2)-(6.4). The first approach exploits the Lipschitz continuity of the operator* G*.*

Crucial for this is the boundedness and monotonicity of the nonlinearity. Considering the formula for $\partial\mathcal{P}(k)$ given in Remark 6.15, we cannot state if the derivative is nonnegative or not.
The second approach uses the implicit function theorem. We will not be able to apply the Fredholm alternative in the setting of the L^∞ spaces in order to confirm that all assumptions of Theorem 6.16 are satisfied. However, we can apply the Lemma of the Neumann series (cf. Werner, 2007, Theorem II.1.11, p.56):

Lemma 6.18 (Neumann Series)**:**
Let X be a normed space and $T : X \to X$ linear and continuous. If the Neumann series converges in the operator norm, then $\mathrm{Id} - T$ is invertible. Especially when X is a Banach space, this holds true whenever $\|T\| < 1$.

We recall that we have denoted the Lipschitz constant and the uniform upper bound of the production function p by L_p and M_p. The parameter $\eta > 0$ characterizes the function $\phi_\eta : k \to \sqrt{k^2 + \eta}$. By G_Ω, we denote the control-state operator that maps the control $c \in L^\infty(\Omega \times (0, T))$ to the solution $k \in \mathcal{V}^\infty$ of the homogeneous nonlocal Neumann volume constrained problem with zero initial condition. The norm $\|\cdot\|_{op}$ defines the operator norm on $\mathcal{V}^\infty$. With these definitions, we can finally state the following theorem:

Theorem 6.19:
Let the assumptions of Lemma 6.14 hold and let the initial productivity distribution A_0 satisfy

$$\|A_0\|_{L^\infty(\Omega)} < \Theta(T, L_p, M_p, \eta, \Gamma_\mu, \Gamma_\varepsilon, G_\Omega),$$

where $\Theta \in \mathbb{R}$ denotes a sufficiently small constant depending on the parameters T, L_p, M_p, η, the kernel functions Γ_μ and Γ_ε, and the norm of the operator G_Ω. Then, the control-state operator G, that maps a control c to the state $k = k(c)$ as the solution of the capital accumulation equation in the nonlocal spatial Ramsey model (6.2) - (6.4), is Fréchet differentiable from $\mathcal{V}^\infty$ to $\mathcal{V}^\infty$.

Proof. For given data $u \in L^\infty(\Omega \times (0, T))$, $v \in \mathcal{C}(0, T)$, and initial condition $w \in L^\infty(\Omega)$, we consider the linear nonlocal problem with homogeneous Neumann volume constraints

$$\begin{aligned} k_t - \mathcal{NL}(k) + \delta k &= u && \text{on } \Omega \times (0, T) \\ -\mathcal{NL}(k) &= 0 && \text{on } \Omega_{\mathcal{I}} \times (0, T) \\ \int_{\Omega \cup \Omega_{\mathcal{I}}} k(y, t) dy &= v && \text{on } (0, T) \\ k(x, 0) &= w && \text{on } \Omega. \end{aligned} \tag{6.15}$$

Assuming that the Neumann-constrained spatial Ramsey problem has a solution, we define the linear and continuous solution operators of (6.15) as

$$G_\Omega : L^\infty(\Omega \times (0,T)) \to \mathcal{V}^\infty$$

corresponding to the linear problem with $w = 0$ and

$$G_0 : L^\infty(\Omega) \to \mathcal{V}^\infty$$

corresponding to the linear problem with $u = 0$.

Now, we rewrite the PIDE (6.2) - (6.4) as

$$\begin{aligned} k_t - \mathcal{NL}(k) + \delta k &= \mathcal{P}(k) - c && \text{on } \Omega \times (0,T) \\ -\mathcal{NL}(k) &= 0 && \text{on } \Omega_\mathcal{I} \times (0,T) \\ k(x,0) &= k_0 && \text{on } \Omega. \end{aligned}$$

Note that we have dropped the constraint on the aggregated capital stock, since this information is already put in the considered solution space V_c. The solution of (6.2)-(6.4) is then of the form

$$k = G_\Omega(\mathcal{P}(k) - c) + G_0(k_0).$$

We rewrite this equation and define the operator $F : \mathcal{V}^\infty \times L^\infty(\Omega \times (0,T)) \to \mathcal{V}^\infty$ as

$$0 = k - G_\Omega(\mathcal{P}(k) - c) - G_0(k_0) =: F(k,c).$$

As already shown in Lemma 6.14, the nonlinear productivity-production operator $\mathcal{P}$ is Fréchet differentiable from $\mathcal{V}^\infty$ to $L^\infty(\Omega \times (0,T))$. Moreover, the operators G_Ω and G_0 are linear and continuous, hence Fréchet differentiable. This yields the Fréchet differentiability of the operator F as a composition of Fréchet differentiable operators.
In order to derive the differentiability of G, we have to prove that $\partial_k F$ is invertible. The derivative of F with respect to k is exactly of the form

$$\partial_k F(k,c) = \mathrm{Id} - T(k,c),$$

where $T(k,c) := G_\Omega(\partial_k \mathcal{P}(k))$ is a linear and continuous operator. According to the Lemma on Neumann series 6.18, we have to prove that

$$\|T(k,c)\|_{op} < 1,$$

which is equivalent to show

$$\|\partial_k \mathcal{P}(k)\|_{op} < \|G_\Omega\|_{op}^{-1}.$$

For $\partial_k \mathcal{P}(k) : \mathcal{V}^\infty \to L^\infty(\Omega \times (0,T))$ given in Remark 6.15, we estimate for $\hat{k} \in \mathcal{V}^\infty$ with $\|\hat{k}\|_{\mathcal{V}^\infty} \leq 1$:

$$
\begin{aligned}
&\left\| \partial_k \mathcal{P}(k)(\hat{k}) \right\|_{L^\infty(\Omega\times(0,T))} \\
&= ess \sup_{(x,t)\in\Omega\times(0,T)} \left| A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} t \right) p'(k(x,t))\hat{k}(x,t) \right. \\
&\quad + t\ A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} t \right) p(k(x,t)) \\
&\quad \cdot \left[\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta'(k(y,t))\hat{k}(y,t)\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} \right. \\
&\quad \left. \left. - \frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy\ \int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta'(k(y,t))\hat{k}(y,t)\Gamma_\varepsilon(x,y)dy}{\left(\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy\right)^2} \right] \right|
\end{aligned}
$$

We analyze the three terms separately. For the first term it holds

$$
\begin{aligned}
ess \sup_{(x,t)\in\Omega\times(0,T)} &\left| A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} t \right) p'(k(x,t))\hat{k}(x,t) \right| \\
&\leq \|A_0\|_{L^\infty(\Omega)} e^T L_p \|\hat{k}\|_{L^\infty(\Omega\times(0,T))} \leq \|A_0\|_{L^\infty(\Omega)} e^T L_p \|\hat{k}\|_{\mathcal{V}^\infty}
\end{aligned}
$$

by the definition of the $\mathcal{V}^\infty$-norm. The estimation of the second term yields

$$
\begin{aligned}
ess \sup_{(x,t)\in\Omega\times(0,T)} &\left| t\ A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} t \right) p(k(x,t)) \right. \\
&\left. \cdot \frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta'(k(y,t))\hat{k}(y,t)\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} \right| \\
&\leq \mathrm{C}T \|A_0\|_{L^\infty(\Omega)} e^T M_p \frac{1}{\sqrt{\eta}} \|\hat{k}\|_{L^\infty(0,T;L^2(\Omega\cup\Omega_\mathcal{I}))} \\
&\leq \hat{\mathrm{C}}T \|A_0\|_{L^\infty(\Omega)} e^T M_p \frac{1}{\sqrt{\eta}} \|\hat{k}\|_{\mathcal{V}^\infty},
\end{aligned}
$$

where we have used Hölder's inequality and exploited $k^2, \eta > 0$ in order to estimate

$$
\begin{aligned}
& ess \sup_{(x,t)\in\Omega\times(0,T)} \left| \frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi'_\eta(k(y,t))\hat{k}(y,t)\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} \right| \\
\leq\ & ess \sup_{(x,t)\in\Omega\times(0,T)} \frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} |\phi'_\eta(k(y,t))\Gamma_\mu(x,y)|\ |\hat{k}(y,t)|dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} \\
\leq\ & ess \sup_{(x,t)\in\Omega\times(0,T)} \frac{\left(\int_{\Omega\cup\Omega_{\mathcal{I}}} |\phi'_\eta(k(y,t))\Gamma_\mu(x,y)|^2 dy\right)^{\frac{1}{2}}}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} \ \|\hat{k}(\cdot,t)\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \\
=\ & ess \sup_{(x,t)\in\Omega\times(0,T)} \underbrace{\frac{\left(\int_{\Omega\cup\Omega_{\mathcal{I}}} \frac{|k(y,t)|^2}{k(y,t)^2+\eta}\Gamma^2_\mu(x,y)dy\right)^{\frac{1}{2}}}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \sqrt{k(y,t)^2+\eta}\ \Gamma_\varepsilon(x,y)dy}}_{=:\#} \ \|\hat{k}(\cdot,t)\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})}.
\end{aligned}
$$

For the term # it holds true that

$$
\# \leq \frac{\left(\int_{\Omega\cup\Omega_{\mathcal{I}}} \Gamma^2_\mu(x,y)dy\right)^{\frac{1}{2}}}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \sqrt{k(y,t)^2+\eta}\ \Gamma_\varepsilon(x,y)dy} \leq \frac{1}{\sqrt{\eta}} \frac{\left(\int_{\Omega\cup\Omega_{\mathcal{I}}} \Gamma^2_\mu(x,y)dy\right)^{\frac{1}{2}}}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \Gamma_\varepsilon(x,y)dy}.
$$

We define

$$
0 < \mathrm{C}(x) := \frac{\left(\int_{\Omega\cup\Omega_{\mathcal{I}}} \Gamma^2_\mu(x,y)dy\right)^{\frac{1}{2}}}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \Gamma_\varepsilon(x,y)dy}.
$$

Note that we have already shown in Lemma 5.2(5), that

$$
\mathrm{C} := ess \sup_{x\in\Omega} \mathrm{C}(x) < \infty.
$$

We define $\hat{\mathrm{C}} := C_1\mathrm{C}$, where C_1 is the constant in 5.9. An analogous estimate yields for the third term

$$
\begin{aligned}
ess \sup_{(x,t)\in\Omega\times(0,T)} & \Bigg| t\ A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy} t\right) p(k(x,t)) \\
& \cdot \frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy\ \int_{\Omega\cup\Omega_{\mathcal{I}}} \phi'_\eta(k(y,t))\hat{k}(y,t)\Gamma_\varepsilon(x,y)dy}{\left(\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy\right)^2} \Bigg| \\
& \leq \tilde{\mathrm{C}} T \|A_0\|_{L^\infty(\Omega)} e^T M_p \frac{1}{\sqrt{\eta}} \|\hat{k}\|_{\mathcal{V}^\infty},
\end{aligned}
$$

for a constant $\infty > \tilde{\mathrm{C}} \geq \hat{\mathrm{C}}$.
Combining all estimates leads to

$$\begin{aligned}
&\sup_{\|\hat{k}\|_{\mathcal{V}^\infty} \leq 1} \|\partial_k \mathcal{P}(k)(\hat{k})\|_{L^\infty(\Omega \times (0,T))} \\
\leq &\sup_{\|\hat{k}\|_{\mathcal{V}^\infty} \leq 1} \|A_0\|_{L^\infty(\Omega)} \max\{L_p, M_p\} e^T \left(\|\hat{k}\|_{\mathcal{V}^\infty} + \frac{2\tilde{\mathrm{C}}T}{\sqrt{\eta}} \|\hat{k}\|_{\mathcal{V}^\infty} \right) \\
\leq &\|A_0\|_{L^\infty(\Omega)} \max\{L_p, M_p\} e^T \left(1 + \frac{2\tilde{\mathrm{C}}T}{\sqrt{\eta}} \right).
\end{aligned}$$

We define the constant θ such that

$$\|A_0\|_{L^\infty(\Omega)} \max\{L_p, M_p\} e^T \left(1 + \frac{2\tilde{\mathrm{C}}T}{\sqrt{\eta}} \right) < \|G_\Omega\|_{op}^{-1},$$

which yields

$$\|T(k,c)\|_{op} < 1.$$

According to theorem 6.16, the operator F is Fréchet differentiable. Thus, we can conclude that also the control-solution operator G is differentiable as one representation of a solution of (6.2)-(6.4).

□

Remark 6.20:
An analogous procedure to prove the Fréchet differentiability of the control-state operator can be found in the paper of Casas et al. (2013, pp.10–12). In this paper, the considered PDE constraint is transformed with a weight $e^{\lambda t}$. This transformation leads to a substitution of the original state variable y by a new, weighted variable $v = e^{-\lambda t} y$. The idea behind this transformation is that for sufficiently large λ, the right-hand side of the seminlinear equation becomes small enough. We could follow the same approach instead of restricting the initial productivity distribution A_0. However, our derivation allows a better insight in the structure of the optimal control problem.

The Necessary First Order Conditions

We have finally all at hand to heuristically derive the necessary first order conditions of the nonlocal spatial Ramsey model with endogenous productivity growth. We assume there exists a (locally) optimal pair $(\overline{k}, \overline{c})$, where $\overline{c}$ denotes the optimal control belonging to a convex set of feasible controls $\mathcal{U}_{ad}$ and $\overline{k} = k(\overline{c})$ is the corresponding optimal state. We have already shown that the control-state operator

$$G : L^\infty(\Omega \times (0,T)) \to \mathcal{V}^\infty, \; c \mapsto G(c) = k(c)$$

is continuous and differentiable. The assumptions on the objective $\mathcal{J}$ defined in (6.1) (U Inada and bounded) make sure that $\mathcal{J} : \mathcal{V}^\infty \times L^\infty(\Omega \times (0,T)) \to \mathbb{R}$ is Fréchet differentiable. Thus also the composition

$$f(c) := \mathcal{J}(G(c), c)$$

is Fréchet differentiable as well. Since the set of feasible controls $\mathcal{U}_{ad}$ is assumed to be convex, this yields the necessary first order conditions in form of the variational inequality

$$f'(\overline{c})(c - \overline{c}) \geq 0 \ \forall \ c \in \mathcal{U}_{ad} \tag{6.16}$$

(cf. Tröltzsch, 2005, p.211).

In order to derive a representation of the derivative $f'(c)$, we follow the *adjoint approach* as outlined by Lebherz et al. (2018, Theorem 2.5, p.5). Therefore, we define the Lagrange function of the model for an appropriate Lagrangian parameter $\lambda \in Z^*$ (where we do not pay attention to the function space Z^* of the variable λ, since we only aim to give a broad insight in the structure of the optimal control problem) as

$$\begin{aligned}\mathcal{L}(k,c,\lambda)(x,t) := & -\int_0^T \int_\Omega U(c(x,t)) e^{-\tau t - \gamma \|x\|_2^2} \ dxdt \\ & + \frac{1}{2\rho_1} \|k(\cdot,T) - k_T\|^2_{L^2(\Omega)} + \frac{1}{2\rho_2} \|\min\{0,k\}\|^2_{L^2(0,T;L^2(\Omega))} \\ & + \int_0^T \int_\Omega \left(k_t - \int_{\Omega \cup \Omega_\mathcal{I}} (k(y,t) - k(x,t)) \Gamma_\varepsilon(x,y) dy + \delta k + c - \mathcal{P}(k)(x,t) \right) \lambda(x,t) \ dxdt \\ & - \int_0^T \int_{\Omega_\mathcal{I}} \int_{\Omega \cup \Omega_\mathcal{I}} (k(y,t) - k(x,t)) \Gamma_\varepsilon(x,y) dy \ \lambda(x,t) \ dxdt + \int_\Omega (k(x,0) - k_0(x)) \lambda(x,0) \ dx.\end{aligned}$$

and get

$$\begin{aligned}\mathcal{L}(k,c,\lambda)(x,t) = & -\int_0^T \int_\Omega U(c(x,t)) e^{-\tau t - \gamma \|x\|_2^2} \ dxdt \\ & + \frac{1}{2\rho_1} \int_\Omega (k(x,T) - k_T(x))^2 \ dx + \frac{1}{2\rho_2} \int_0^T \int_\Omega \min_{x,t}\{0, k(x,t)\}^2 \ dxdt \\ & - \int_0^T \int_\Omega k(x,t) \left(\lambda_t(x,t) - \delta \lambda(x,t) \right) - c\lambda(x,t) \ dxdt \\ & - \int_0^T \int_{\Omega \cup \Omega_\mathcal{I}} k(x,t) \int_{\Omega \cup \Omega_\mathcal{I}} (\lambda(y,t) - \lambda(x,t)) \Gamma_\varepsilon(x,y) dy \ dxdt \\ & - \int_0^T \int_\Omega \mathcal{P}(k)(x,t) \lambda(x,t) \ dxdt + \int_\Omega k(x,T) \lambda(x,T) dx - \int_\Omega k_0(x) \lambda(x,0) \ dx.\end{aligned}$$

We have integrated by parts and used a variable shift in the linear nonlocal term, to shift the operation to the Lagrange variable λ. Theorem 2.5 in Lebherz et al. (2018) states that for a control $c \in L^\infty(\Omega \times (0,T))$, the derivative of the operator f can be represented by

$$f'(c) = \mathcal{L}_c(G(c), c, \lambda),$$

where the variable λ is defined as the solution of the adjoint equation, which can be expressed by the identity

$$\mathcal{L}_k(k, c, \lambda) = 0.$$

We heuristically differentiate the Lagrangian with respect to c in direction $\hat{c}$ and get

$$\mathcal{L}_c(k,c,\lambda)(\hat{c})(x,t) = \int_0^T \int_\Omega (U'(c(x,t))e^{-\tau t - \gamma \|x\|_2^2} + \lambda(x,t))\ \hat{c}(x,t)\ dxdt$$

for $(x,t) \in \Omega \times (0,T)$.

We calculate the directional derivative $\mathcal{L}_k$ in direction $\hat{k}$ using the formula of the derivative of $\mathcal{P}$. We get

$$\begin{aligned}
&\mathcal{L}_k(k,c,\lambda)(\hat{k})(x,t) = \\
&\quad \frac{1}{\rho_1}\int_\Omega (k(x,T) - k_T(x))\hat{k}(x,T)\ dx + \frac{1}{\rho_2}\int_0^T \int_\Omega \min\{0, k(x,t)\}\hat{k}(x,t)\ dxdt \\
&\quad - \int_0^T \int_\Omega \hat{k}(x,t)\left(\lambda_t(x,t) - \delta\lambda(x,t)\right)\ dxdt + \int_\Omega \hat{k}(x,T)\lambda(x,T)\ dx \\
&\quad - \int_0^T \int_{\Omega\cup\Omega_\mathcal{I}} \hat{k}(x,t) \int_{\Omega\cup\Omega_\mathcal{I}} (\lambda(y,t) - \lambda(x,t))\Gamma_\varepsilon(x,y)dy\ dxdt \\
&\quad - \int_0^T \int_\Omega A_0(x)\exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}\ t\right) p'(k(x,t))\lambda(x,t)\hat{k}(x,t)\ dxdt \\
&\quad - \int_0^T \int_\Omega t\ A_0(x)\exp\left(\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}\ t\right) p(k(x,t))\lambda(x,t) \\
&\quad \cdot \left[\frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta'(k(y,t))\Gamma_\mu(x,y)\hat{k}(y,t)dy}{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}\right. \\
&\quad \left. - \frac{\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy\ \int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta'(k(y,t))\Gamma_\varepsilon(x,y)\hat{k}(y,t)dy}{(\int_{\Omega\cup\Omega_\mathcal{I}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy)^2}\right]\ dxdt,
\end{aligned}$$

for $(x,t) \in \Omega \times (0,T)$. Exploiting the equality of

$$\int_{\Omega\cup\Omega_{\mathcal{I}}} k(x)\ dx = \int_{\Omega} k(x)\ dx + \int_{\Omega_{\mathcal{I}}} k(x)\ dx$$

since $\Omega \cup \Omega_{\mathcal{I}} = \emptyset$ and subtly choosing the variable λ, we obtain the linear but nonlocal adjoint equation on $\Omega \times (0,T)$ as

$$\begin{aligned}
-\lambda_t = &\int_{\Omega\cup\Omega_{\mathcal{I}}} (\lambda(y,t) - \lambda(x,t))\Gamma_{\varepsilon}(x,y)dy - \delta\lambda \\
&+ A_0(x) \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(x,y)dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(x,y)dy}\ t \right) p'(k(x,t))\lambda(x,t) \\
&+ \phi_\eta'(k(x,t)) \int_\Omega \mathcal{R}_1(k)(z,t)\Gamma_\mu(x,z)\lambda(z,t)dz \\
&- \phi_\eta'(k(x,t)) \int_\Omega \mathcal{R}_2(k)(z,t)\Gamma_\varepsilon(x,z)\lambda(z,t)dz,
\end{aligned} \tag{6.17}$$

with volume constraints

$$\begin{aligned}
-\int_{\Omega\cup\Omega_{\mathcal{I}}} (\lambda(y,t) - \lambda(x,t))\Gamma_{\varepsilon}(x,y)dy = &\ \phi_\eta'(k(x,t)) \int_\Omega \mathcal{R}_1(k)(z,t)\Gamma_\mu(x,z)\lambda(z,t)dz \\
&- \phi_\eta'(k(x,t)) \int_\Omega \mathcal{R}_2(k)(z,t)\Gamma_\varepsilon(x,z)\lambda(z,t)dz,
\end{aligned} \tag{6.18}$$

acting on $\Omega_{\mathcal{I}} \times (0,T)$, and the terminal condition

$$\lambda(x,T) = \frac{1}{\rho_1}(k_T(x) - k(x,T)) \tag{6.19}$$

on Ω. We shortened the expressions, introducing some nonlocal operators $\mathcal{R}_1$ and $\mathcal{R}_2$ as

$$\mathcal{R}_1(k)(z,t) := t\ A_0(z) \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(z,y)dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(z,y)dy}\ t \right) \frac{p(k(z,t))}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(z,y)dy}$$

and

$$\begin{aligned}
\mathcal{R}_2(k)(z,t) := &\ t\ A_0(z) \exp\left(\frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(z,y)dy}{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(z,y)dy}\ t \right) \\
&\cdot p(k(z,t))\ \frac{\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\mu(z,y)dy}{(\int_{\Omega\cup\Omega_{\mathcal{I}}} \phi_\eta(k(y,t))\Gamma_\varepsilon(z,y)dy)^2}
\end{aligned}$$

Thus, we can finally state the (heuristic) necessary first order conditions as follows:

Theorem 6.21 (Necessary First Order Conditions)**:**
Let there exist a (locally) optimal control $\overline{c} \in \mathcal{U}_{ad}$ *with corresponding optimal state* $\overline{k}$ *of (6.1)-(6.4) and assume that the adjoint system (6.17)-(6.19) has a unique solution* $\overline{\lambda}$*. Under the assumptions of Theorem 6.19, every locally optimal control satisfies the variational inequality*

$$\int_0^T \int_\Omega \Big(U'(\overline{c}(x,t)) e^{-\tau t - \gamma \|x\|_2^2} + \overline{\lambda}(x,t) \Big) \ (c(x,t) - \overline{c}(x,t)) \ dxdt \geq 0 \quad \forall \ c \in \mathcal{U}_{ad}. \tag{6.20}$$

When it comes to the numerical implementation of the FOTD approach, we have to solve the coupled system of nonlocal partial differential equations, the state equation (6.2)-(6.4), the adjoint equation (6.17)-(6.19), and the design inequality (6.20). As already mentioned, we do not implement this coupled system in the context of this thesis, but recommend a rigorous numerical treatment of this approach to future research.
Considering the adjoint equation, we suggest to solve the problem (6.17)-(6.19) sequentially time. The terminal condition (6.19) defines the adjoint variable in the set of interest Ω. Analogously to the initial condition of the state variable, we can determine the adjoint on the interaction domain making use of the volume constraint (6.18). Note that in order to calculate the right hand side of the equation (6.18), we only have to know the adjoint variable in Ω. Then, we solve the equation (6.17), where we need to know the values of λ on the nonlocal closure $\Omega \cup \Omega_{\mathcal{I}}$. Going backwards in time, we can determine the adjoint on the whole space-time cylinder.
The adjoint equation has a very similar structure as the state equation. We do have a nonlocal diffusion term, Neumann volume constraints, and a terminal conditions that is only defined in the set of interest. Moreover, the adjoint equation is linear, hence less complex. Except of the time shift in this equation, the second disparity between the state system and the adjoint system lies in the volume constraints. We do not have homogeneous Neumann constraints anymore, but the interaction with $\Omega_{\mathcal{I}}$ depends also on the aggregated value of λ in Ω. This makes an analysis regarding the existence of a solution of (6.17)-(6.19) quite complicated. However, if there exists a solution, we suggest that $\lambda \in \mathcal{V}^\infty$ as well.

6.4 Comparison with the Local Model

6.4.1 Comparison of the Capital Accumulation Equations

The major result in the papers of Brito (2001, 2004, 2012), Boucekkine et al. (2009), and Boucekkine et al. (2013) is the *'asymptotic disappearance of spatial inequality, i.e. convergence of the capital stock over time to the same level in all regions, despite the initially heterogeneous spatial distribution of capital'* (Aldashev et al., 2014, p.2). On one hand, this is expectable due to the shape of the neoclassical production function in the Ramsey model, which enforces diminishing returns on capital, and due to the local diffusion effect that smoothes out heterogeneity very quickly. On the other hand, this is in line with the economic intuition that capital flows from capital-abundant to poorer regions. This behavior conjoins the assumption of a social central planner, who maximizes the welfare of the economy as a whole and does not consider any individual preferences. However, as Breinlich et al. (2014) points out, the smoothing effects in the capital distributions between rich and poorer regions is much slower in reality. It is crucial, also for economic growth models, to describe reality as good as possible in order to give meaningful policy advices. As Aldashev et al. (2014, p.11) state, a pure local diffusion effect, which models the mobility of capital across space, may not be sufficient to capture the dynamics of capital flows.
The local diffusion operator, which describes the capital mobility over space in the common spatial Ramsey model, is not the only driving factor, that generates the homogeneous optimal capital and consumption distribution over space. Furthermore, every technological progress is assumed to be constant in all the spatial models listed above. Not only Aldashev et al. (2014, p.11), but also Camacho et al. (2008, p.3) propose that differences in technology induce an agglomerational effect on the capital accumulation process, which should not be denied and which may cause heterogeneities in capital and consumption behavior in time.
In our nonlocal spatial Ramsey model with endogenous productivity growth, we combine both suggestions, the diffusive and agglomerative effect, to enrich the model of the capital accumulation across time and space. And indeed, as the first numerical results may hypothesize, the combination of nonlocal diffusion and space-time-depending technological progress enables us to conserve and create heterogeneities and disparities between different locations. Also the optimal consumption distribution reflects this property of our capital accumulation equation.

In this section, we finally compare the numerical solution of the capital accumulation equation and the optimal saving decisions of the households in the local model by Brito (2001) with the ones in our nonlocal spatial Ramsey model. In a first approach, we want to illustrate, how the nonlocal diffusion operator changes the capital accumulation in space and time compared to the pure diffusion model. In order to keep the models comparable, we extend also the local model by the nonlocal productivity-production operator. Thus, we compare the capital accumulation equation of our nonlocal model

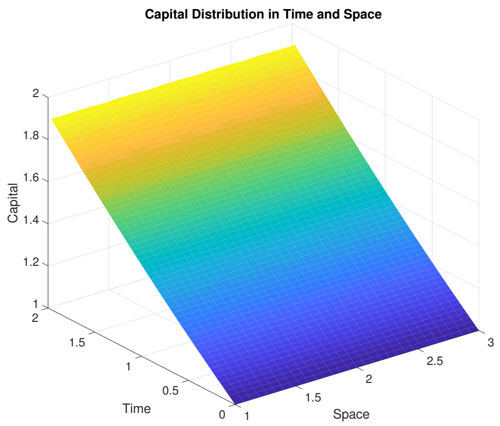
$$
\begin{aligned}
k_t - \int_{\Omega \cup \Omega_{\mathcal{I}}} (k(y,t) - k(x,t))\Gamma_\varepsilon(x,y)dy + \delta k - \mathcal{P}(k) &= 0 && \text{on } \Omega \times (0,T), \\
- \int_{\Omega \cup \Omega_{\mathcal{I}}} (k(y,t) - k(x,t))\Gamma_\varepsilon(x,y)dy &= 0 && \text{in } \Omega_{\mathcal{I}}, \\
k(x,0) &= k_0(x) && \text{in } \Omega,
\end{aligned}
$$

with the capital accumulation equation with local diffusion

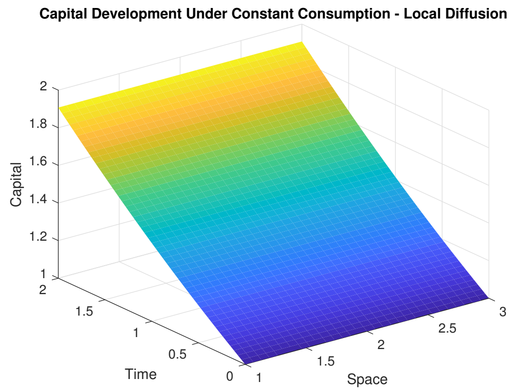
$$
\begin{aligned}
k_t - \Delta k + \delta k - \mathcal{P}(k) &= 0 && \text{on } \Omega \times (0,T), \\
- \int_{\Omega \cup \Omega_{\mathcal{I}}} (k(y,t) - k(x,t))\Gamma_\varepsilon(x,y)dy &= 0 && \text{in } \Omega_{\mathcal{I}}, \\
k(x,0) &= k_0(x) && \text{in } \Omega.
\end{aligned}
$$

We consider the volume constraints in the local diffusion model as well, since the nonlocal production-productivity operator requires information on the state variable k also in $\Omega_{\mathcal{I}}$.

We start our comparison with the spatially homogeneous case, as illustrated in Figure 6.13. Here, we choose the constant initial capital and productivity distributions $k_0 \equiv 1$ and $A_0 \equiv 0.4$ in Ω.



a) Nonlocal spatial Ramsey model

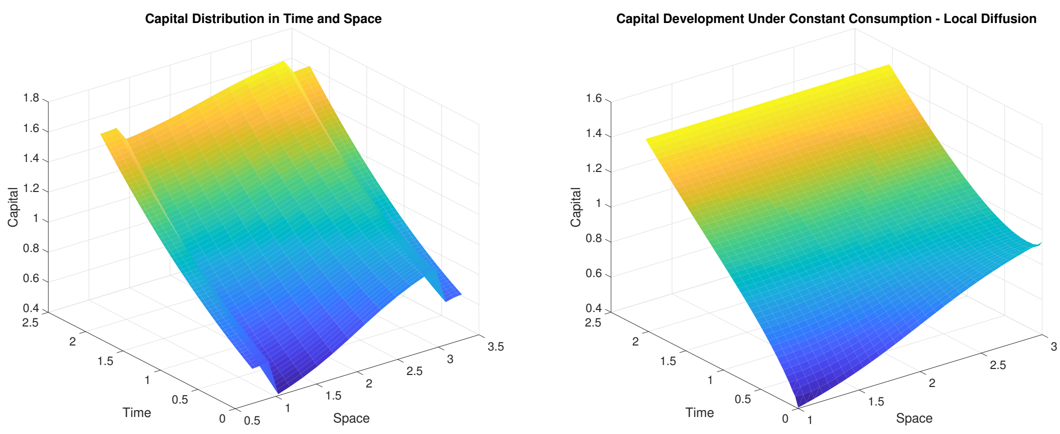


b) Local spatial Ramsey model

Figure 6.13: Comparison: Capital Accumulation across Space and Time under Constant Initial Capital and Productivity Distribution

As we expected, there is no difference in the solutions of the two capital equations. The reason is that there is neither any heterogeneity which could be smoothed out in the initial income distribution, nor any uneven agglomeration of capital due to production. Since capital and productivity are evenly distributed in space in all points in time, no diffusive effect appears.

The difference between the two models becomes clearer, whenever we face heterogeneities in the initial capital or productivity distribution. .

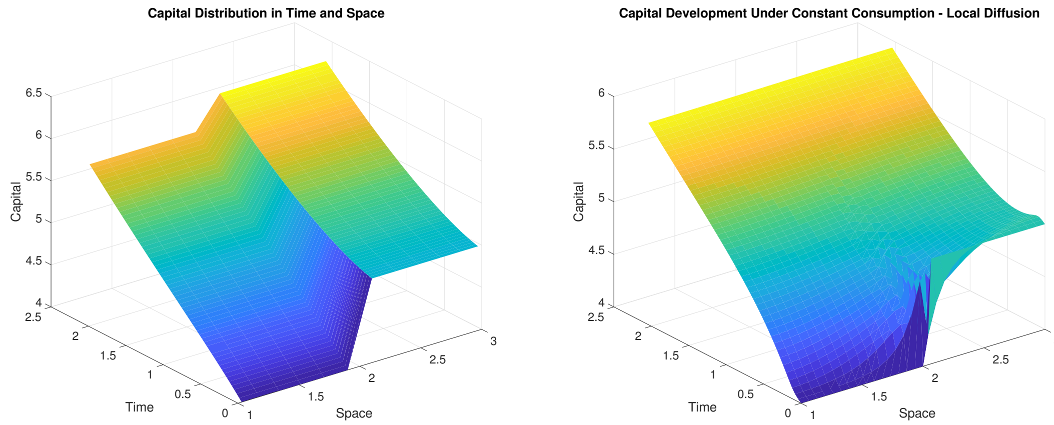


a) Nonlocal spatial Ramsey model **b)** Local spatial Ramsey model

Figure 6.14: Comparison: Capital Accumulation across Space and Time under Constant Initial Productivity and Heterogeneous but Continuous Initial Capital Distribution

If we consider the continuous, but not constant initial capital distribution k_0 defined in (6.9) and a constant initial productivity $A_0 \equiv 0.4$, we see in Figure 6.14 that the heterogeneity vanishes in short time in the local model, whereas it is well preserved in our nonlocal setting.
If we even start with a discontinuous initial capital distribution as shown in Figure 6.15, it becomes obvious that the solution of the capital equation in the nonlocal model can be discontinuous whereas the local diffusion effect smoothes out this discontinuity very quickly.

As already seen in Section 6.2, not only inhomogeneous initial capital distributions, but also non-constant productivity distributions lead to space-heterogeneous solutions of the PIDE. We illustrate this conjecture in Figure 6.16.

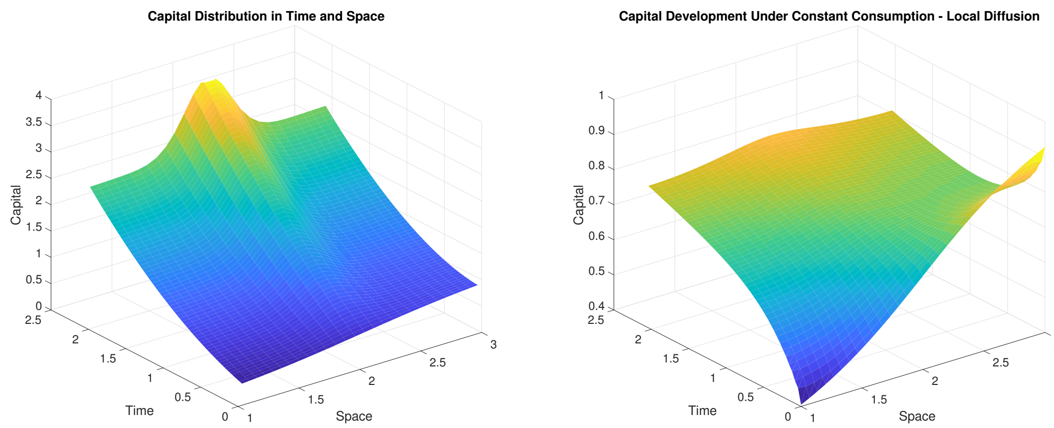


a) Nonlocal spatial Ramsey model

b) Local spatial Ramsey model

Figure 6.15: Comparison: Capital Accumulation across Space and Time under Constant Initial Productivity and Discontinuous Initial Capital Distribution

Also in the local model, we would expect an increasing heterogeneity, driven by the agglomerative effects of an increasing technological progress over time. Since we added the nonlocal productivity-production operator to the local model by Brito (2001), the figures in this section really illustrate the impact of the different types of diffusion operators. We consider the bell-shaped initial productivity distribution in Figure 6.16.



a) Nonlocal spatial Ramsey model

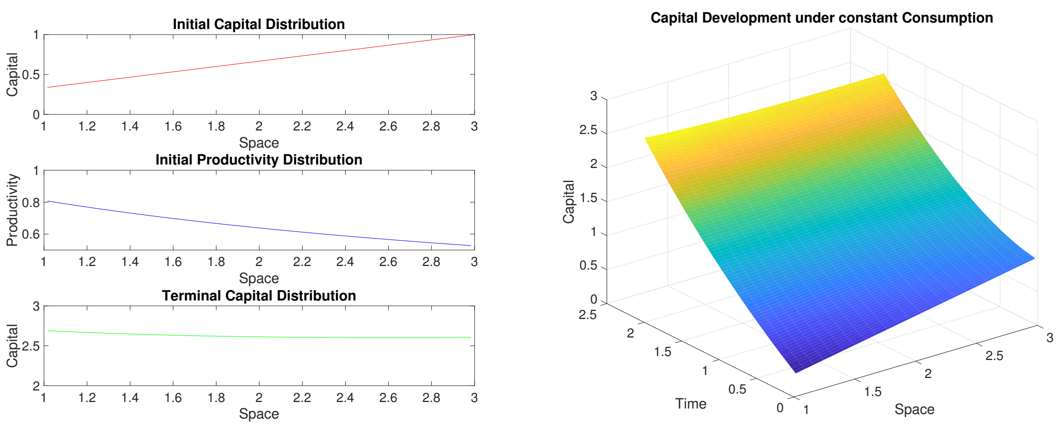
b) Local spatial Ramsey model

Figure 6.16: Comparison: Capital Accumulation across Space and Time under Bell-shaped Initial Productivity and Heterogeneous but Continuous Initial Capital Distribution

In line with a result of Camacho et al. (2008), we observe that the capital distribution shapes analogously to A_0, also in the local model (see Figure 6.16b). However, due to the fast spreading effects of the local diffusion operator, the capital distribution k is more flat than the initial productivity distribution A_0 and the overall agglomeration in the local is much weaker than compared to our pure nonlocal model.

6.4.2 Compensating Effects of Initial Capital and Productivity Distributions

From an economic point of view, a heterogeneous terminal capital distribution may not be in line with the assumption of a benevolent planner. The Figures 6.17 and 6.18 below illustrate, that our nonlocal model with nonlinear productivity is general enough, to describe not only economies with heterogeneous capital and consumption distributions over time and space, but that we can also model a capital distribution which becomes homogeneous in space over time, without any action of the central planner. The figures on the left-hand side illustrate the initial capital and productivity distributions and the terminal capital distribution.
It becomes obvious by means of the terminal capital distribution and of the time-space capital distribution on the right-hand side that A_0 and k_0 have compensating effects. Thus, an appropriate choice of the data may hence enable us to mimic the dynamics of the local model and mimic an benevolent planner who homogenizes wealth across space.



a) Initial data and terminal capital distribution **b)** Capital distribution in time and space

Figure 6.17: Compensating Effects in the Nonlocal Model: Linear Initial Capital and Nonlinear Initial Productivity Distribution

In Figure 6.17, the initial capital and productivity distributions are given as

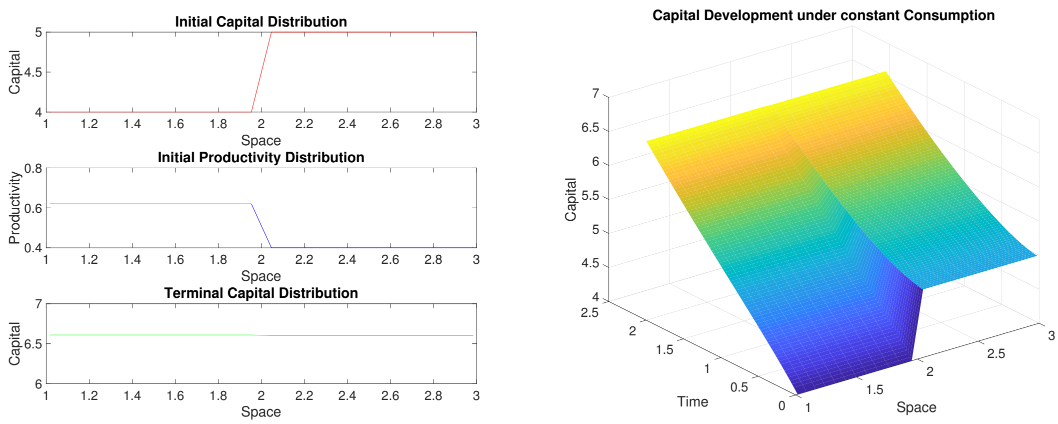
$$k_0(x) := \frac{1}{3}x, \qquad \text{and} \qquad A_0(x) := \frac{3}{2.7 + x}, \qquad x \in \Omega.$$

The function A_0 is only a guess, how the productivity should be distributed in order to compensate the initial capital distribution. To increase the compensating effects such that the terminal capital distribution is truly constant, we suggest to calibrate the data with an additional optimal control problem. Nevertheless, this initial guess is appropriate for our purpose. We only intend to point out that we can choose the initial data such that the terminal condition k_T is almost constant in space.

We can observe the same compensating effects of the initial capital and productivity distribution also for discontinuous initial data,

$$k_0(x) := \begin{cases} 4, & x \in (1,2] \\ 5, & x \in (2,3) \end{cases} \quad \text{and} \quad A_0(x) := \begin{cases} 0.61, & x \in (1,2] \\ 0.4, & x \in (2,3) \end{cases},$$

as illustrated in Figure 6.18.



a) Initial data and terminal capital distribution **b)** Capital distribution in time and space

Figure 6.18: Compensating Effects in the Nonlocal Model: Discontinuous Initial Capital and Reversed Initial Productivity Distribution

Economic Interpretation of Figures 6.17 and 6.18:

The initial capital distributions describe an economy with richer and poorer agents. The poorer agents are more productive, which means that they generate welfare more efficiently with the available production factors than the richer ones. Thus, after the finite time horizon, they have overcome any disparity in the initial welfare. If time goes on, the agents with higher efficiency will become even richer and, at some point, overtake the initially richer households such that the capital distribution is reversed.

For the economic application, this means that the productivity has a bigger impact on the welfare of an economy than the inital capital distribution. This is once more in line with the result in Camacho et al. (2008) and also expectable, since the production is a time depending process, which affects the capital stock of the economy in every point in time and not only at the beginning.

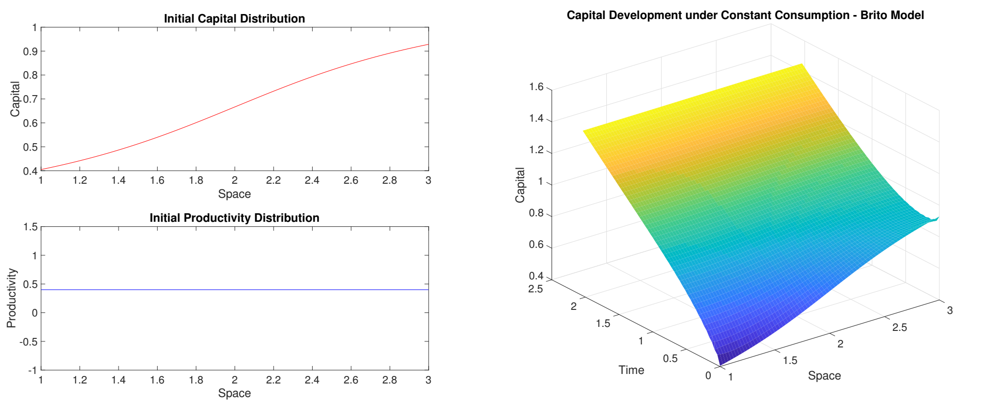
6.4.3 Comparison of the Ramsey Equilibria

In a last study, we compare the solutions of the optimal control problems, on one hand with an underlying local PDE constraint and on the other hand with the nonlocal diffusion equation. We implement the local model according to Brito (2001, 2004), hence we assume a constant productivity growth rate which is independent of time. We consider the nonlinear production function p instead of the simple AK-model which is studied in Brito (2012), such that the production side of the economy is modeled as

$$y(x,t) = A_0(x)p(k(x,t)) \quad (x,t) \in \Omega \times (0,T),$$

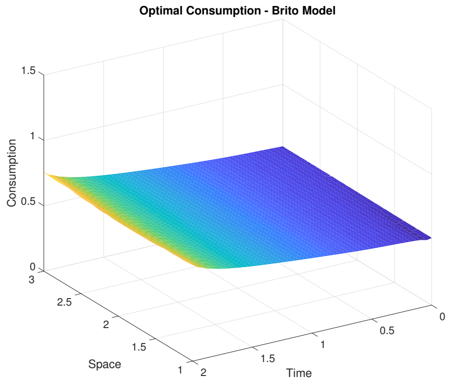
where as usual k denotes the capital stock in a point (x,t). We have to admit that the model by Brito and our model with time depending productivity growth are barely comparable. However, we want to point out with the following illustrations, how much more flexible our model is compared to the common spatial Ramsey model. We consider the same initial data for the numerical realization of Brito's model, as we have used in the examples illustrated in Figures 6.10, 6.11, and 6.12.

We start our study with the continuous, but heterogeneous initial capital distribution defined in (6.9) and a constant productivity growth rate A_0. We illustrate the optimal consumption path and the corresponding optimal capital stock distribution of Brito's model in Figure (6.19). The shape of the optimal consumption distribution in this local model is highly dependent on the discount rates in time and space. In the following examples, we assume that the central planner does not have any spatial priorities, thus we set the space discount rate $\gamma = 0$. The time discount rate is fixed as $\tau = 0.1$, which means that the central planner prefers consumption today over future consumption, or in other words, an equal gain in utility at a future date, compared to the gain of utility today, requires a higher consumption level.

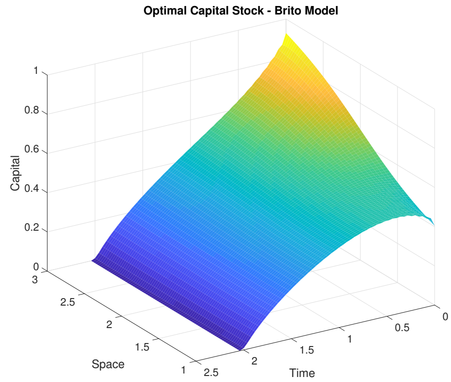


a) Initial data

b) Capital stock under constant consumption in Brito's model



c) Optimal consumption path in Brito's model



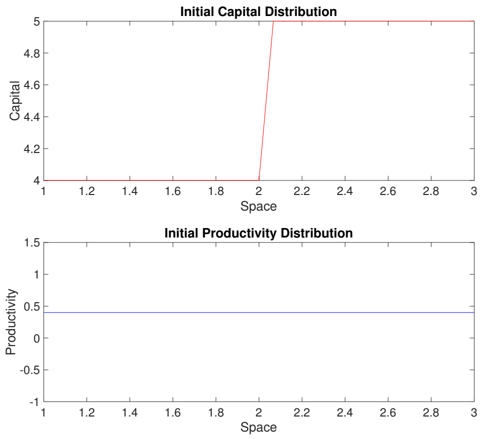
d) Optimal capital stocks in Brito's model

Figure 6.19: The Dynamics of Brito's Ramsey Model: Continuous but Heterogenous Initial Capital Distribution and Constant Initial Productivity

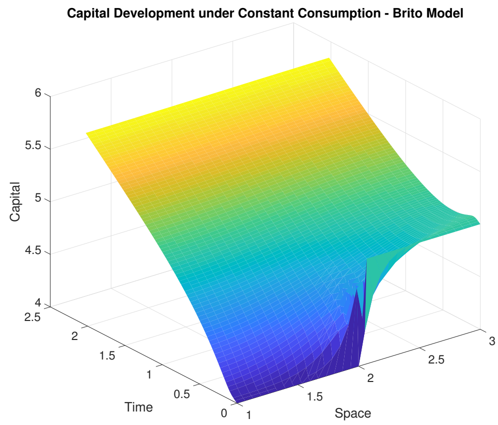
Not only that the capital distribution becomes homogeneous in space over time, when we consider a constant consumption level of the agent (Figure 6.19b, or 6.20b), it is obvious at first glance that the optimal consumption distribution in Brito's model in time and space differs a lot from the optimal solution in our non-local model. Whereas in Figures 6.10, 6.11, and 6.12, the level of consumption is the highest in locations where the initial capital distribution or productivity is the highest, the central planner in the local model tends to homogenize the consumption distribution over space in all examples illustrated in the Figures 6.19, 6.20, and 6.22.
The explanation for equally high consumption levels across space is the choice of a zero space discount rate. This means that the central planner weights every location equally. As we have expected, the consumption increases a bit towards the end of time, which we can explain by the time discounting of the central planner.

In contrast to the consumption distribution, the shape of the capital stock's surface depends on the initial data, however the impact of the initial capital distribution seems to be stronger than the impact of the given productivity distribution $A(x)$, $x \in \Omega$, which is in this model only dependent on space and not on time. Like in our nonlocal model, the optimal capital distribution falls down to zero towards the terminal date in all considered examples, since the agents do not gain any utility by holding capital.

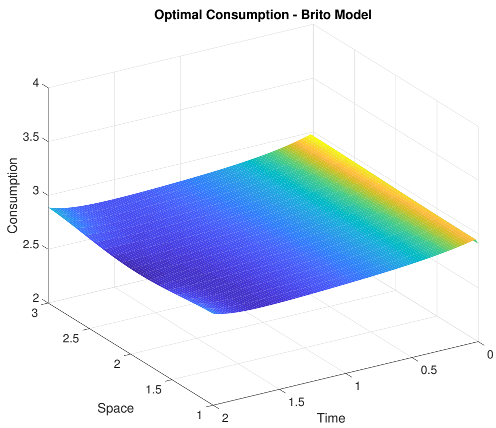
Especially in Figure 6.20, it becomes clear that the optimal capital distribution is influenced by the initial capital distribution, although the discontinuity in k_0 is smoothed out very quickly in the first time steps.



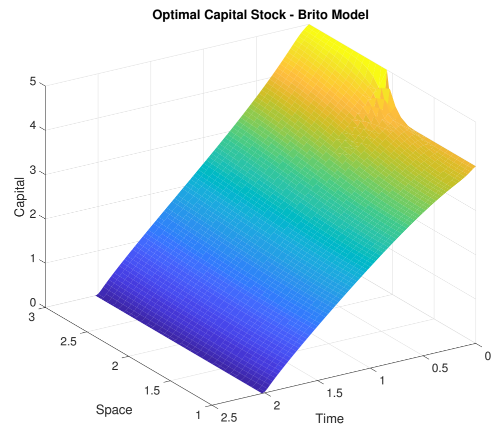
a) Initial data



b) Capital stock under constant consumption in Brito's model



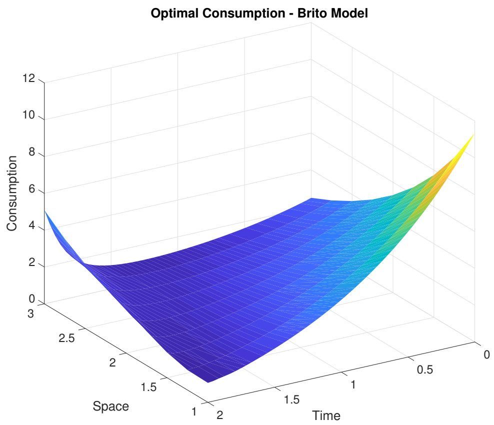
c) Optimal consumption path in Brito's model



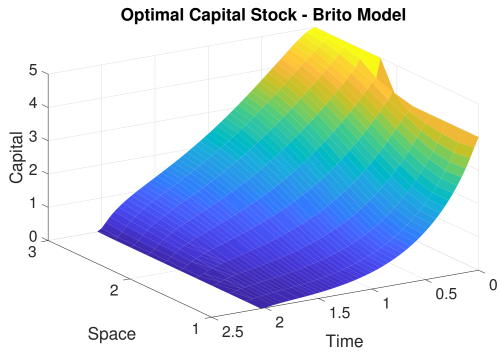
d) Optimal capital stocks in Brito's model

Figure 6.20: The Dynamics in Brito's Ramsey Model: Constant Initial Productivity and Piecewise Constant but Discontinuous Initial Capital Distribution

It is worth to mention that, whenever we consider a nonzero space discount rate $\gamma = 0.1$, the over-time-aggregated consumption is homogeneous as well. By the over-time-aggregated consumption level, we mean the integral over time of the consumption function $c(x, \cdot)$ for a fixed $x \in \Omega$, multiplied with the respective time and space discount rates. We show an example in Figure 6.21, where we solved the model with the same initial data as shown in Figure 6.19a.



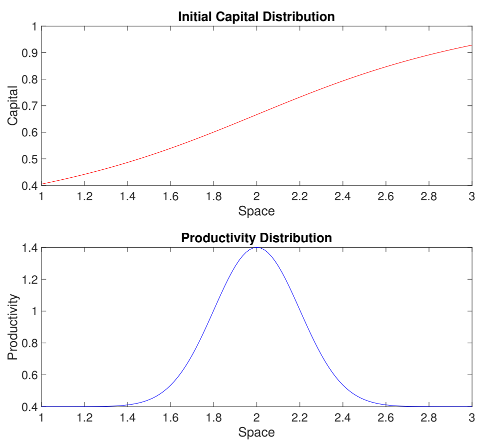
a) Optimal consumption path in Brito's model



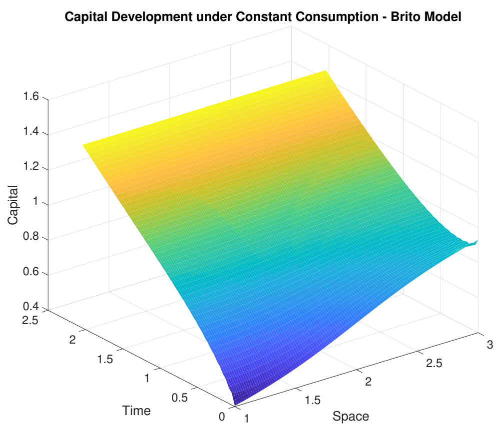
b) Optimal capital stocks in Brito's model

Figure 6.21: The Impact of Spatial Discounting in the Local Ramsey Model

This Figure 6.21 illustrates once more that the preference ordering of the central planner has bigger impact on the optimal consumption distribution than the initial data. Furthermore, the shape of the consumption paths seem to be independent of the initial capital distribution, or the constant productivity growth rate in all examples. Only the absolute values depend on the values of k_0.

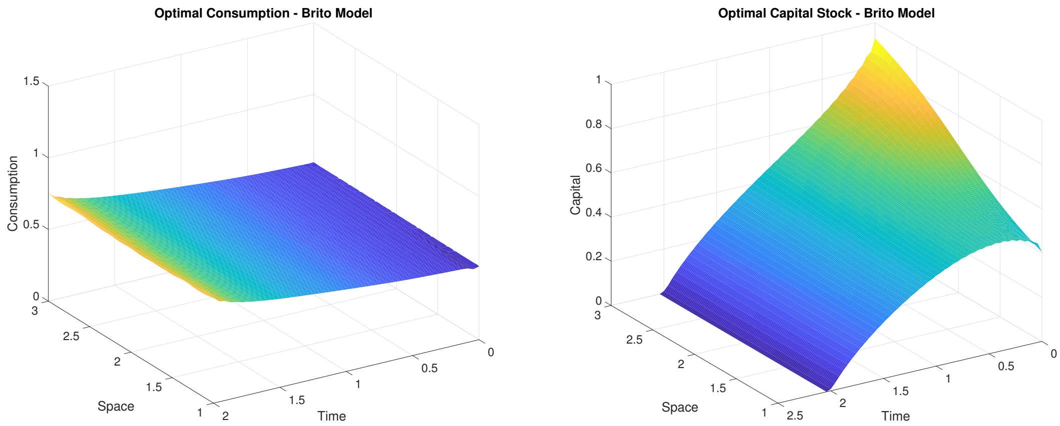


a) Initial data



b) Capital stock under constant consumption in Brito's model

Figure 6.22 makes clear, that the heterogeneity of the time-constant productivity growth rate has almost no impact, neither on the capital stock development under constant consumption, nor on the optimal solution of Brito's model.



c) Optimal consumption path in Brito's model

d) Optimal capital stocks in Brito's model

Figure 6.22: The Dynamics in Brito's Ramsey Model: Bell-shaped Initial Productivity and Continuous but Heterogenous Initial Capital Distribution

When we compare the market equilibria, by which we mean the solution of the optimal control problems, as illustrated in Figures 6.19, 6.20, and 6.22, to the solutions of our nonlocal spatial Ramsey model with endogenous productivity growth pictured in the Figures 6.10, 6.11 and 6.12, we see at first glance, that our model is able to capture the heterogeneity in the initial data. In that way, our model considers *the Heterogeneity* inside a single economy or between two distinct countries as an important factor for the establishment of policy advises.

CHAPTER 7

Summary and Conclusion

In this monograph, we have developed and studied a nonlocal extension of the spatial Ramsey model, a neoclassical growth model from economics, whose spatial extension was first studied by Brito (2001). We introduced a nonlocal diffusion operator to describe the mobility of the production factors across space. Moreover, we endogenized the productivity growth in order to preserve the self-complete character of the Ramsey model. We were able, based on the second welfare theorem, to describe the competitive equilibrium in the resulted economy as the solution of an optimal control problem under a semilinear partial integro-differential equation.
We analyzed the model with respect to well-posedness in two different settings. In Chapter 4, we derived an existence and uniqueness result of a weak solution of the capital accumulation equation in the nonlocal spatial Ramsey model with endogenous productivity growth over an unbounded spatial domain. We derived some strong regularity results, based on the fundamental works of DiBenedetto (1993) and Ladyženskaya et al. (1968), and were able to prove the existence of a market equilibrium. In this setting, the structure of our model allowed us to overcome the difficulties, which appear when considering unbounded spatial domains in PDE optimization, without the need of weighted Sobolev spaces.
In Chapter 5, we considered a pure nonlocal diffusion operator to model the capital mobility across space. We embedded the nonlocal spatial Ramsey model in the nonlocal vector calculus by Du et al. (2012a) and showed the existence and uniqueness of a weak solution of the capital accumulation equation under homogeneous Dirichlet-type volume constraints. Furthermore, we derived some a priori estimates and regularities of the weak solution.
Finally, we have implemented the spatial growth model for the scenario of an interacting economy. In a first analysis, we illustrated the impact of the nonlocal diffusion operator, considering the numerical solution of the (uncontrolled) capital accumulation equation. We implemented the optimal control problem with a first

optimize, then discretize approach. We concluded the numerical study with a comparison of the common local spatial Ramsey model and our nonlocal, endogenous version. Here, we have seen that the dynamics of the common spatial Ramsey model are almost independent of the initial data. The central planner tends to homogenize the overall consumption distribution, or to distribute it according to his preferences, taking neither the initial welfare of an agent nor his productivity into account. Moreover, the stringent local diffusion homogenized the capital stock across space, even with no intervention of the central planner. Thus, as already pointed out by Aldashev et al. (2014), the capital accumulation equation may be too poor to study the economic growth dynamics of a heterogeneous economy.

With our model, we have overcome this weakness. In our nonlocal version, the market equilibrium can be heterogeneous in space and time, depending on the initial state of the economy. We are able to control the dependence on the initial data and the quality of the heterogeneity in the capital and consumption distributions by the choice of several parameters, such as the interaction radius in the kernel function of the nonlocal diffusion operator, or the proportion of the areas, whose public welfare has impact on the productivity growth. Moreover, considering weak solutions in the analytical study, we were able to show that the spatial Ramsey model is well-posed (in weak sense). Thus, we were not only able to enrich the dynamics of production factor mobility in the Ramsey economy, but we also closed the gap in the literature, considering the question of existence of a market equilibrium in the spatial Ramsey model in a very general setting.

When studying a single economy, which is quite heterogeneous with respect to the income distribution, or infrastructure and productivity, such as the United Kingdom, which we referred to in the introduction, a more complex model may capture real life observations better than the common spatial Ramsey model. Depending on this intention, our nonlocal model may be a better choice to give policy advises. Especially as mentioned in Section 6.4.2, a rigorous calibration of the initial productivity distribution may give an insight, how policy makers can compensate an existing, or even growing disparity in the spatial welfare distribution of an economy.

With this monograph, we have provided a cornerstone for a rigorous study of heterogeneities inside single closed and interacting economies, and of cross-border disparities. We recommend a more extensive numerical study of our nonlocal model. The efficient numerical solution of partial integro-differential equations is a rather challenging task, with respect to computational storage costs and computing time. Moreover, we only heuristically derived the necessary first order conditions. Although the broad numerical examples, which we introduced in this context, give an insight on how the initial data, or the central planner's preferences, influence the optimal capital and consumption distributions in time and space, we suggest a rigorous sensitivity analysis and a numerical treatment of the adjoint system to future research.

Nomenclature

Functions

A_0	$: \mathbb{R}^n \to \mathbb{R}$, initial productivity distribution
c	$: \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$, consumption per capita
F	$: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, aggregated neoclassical production function, see Assumption 2.1
Γ_ν	Gaussian probability density function, see equation (3.10)
$\mathcal{J}$	real valued objective function, see equation (3.13)
k	$: \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$, capital per capita
k_0	$: \mathbb{R}^n \to \mathbb{R}_+$, initial capital distribution
k_T	$: \mathbb{R}^n \to \mathbb{R}_+$, terminal capital distribution, sustainability condition
$\mathcal{L}$	Lagrange Function
p	$: \mathbb{R} \to \mathbb{R}$, neoclassical production function, see (2.1)
ϕ	$: \mathbb{R} \to \mathbb{R}_+$, Lipschitz continuous nominal function
ϕ_η	$: \mathbb{R} \to \mathbb{R}_+$ continuous and differentiable function depending on $\eta > 0$, i.e. $k \mapsto \sqrt{k^2 + \eta}$
U	$: \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$, neoclassical instantaneous utility function, see Assumption 2.4

Function Sets/Spaces

$\mathcal{C}(\Omega)$	space of bounded and continuous functions $\varphi : \Omega \to \mathbb{R}$
$\mathcal{C}^m(\Omega)$	space of m times continuously differentiable functions $\varphi : \Omega \to \mathbb{R}$ with bounded derivatives up to order $m \in \mathbb{N}$
$\mathcal{C}^m(\overline{\Omega})$	space of functions with bounded and uniformly continuous derivatives up to order m on Ω
$\mathcal{C}^{m,\lambda}(\Omega)$	space of Hölder continuous functions up to order $m \in \mathbb{N}$ of exponent $\lambda \in (0, 1]$
$\mathcal{C}_0^\infty(\Omega)$	set of test functions, set of infinitely often continuously differentiable functions with compact support on Ω
$\mathcal{D}(\Omega)$	topological space of test functions $(\mathcal{C}_0^\infty, \mathcal{T})$, see Definition 3.3
$\mathcal{D}'(\Omega)$	set of distributions on Ω, see Definition 3.5
$L^p(\Omega)$	space of equivalence classes of p-Lebesgue integrable functions on Ω, $1 \le p < \infty$, see Definition 3.1
$L^p_{loc}(\Omega)$	space of $p-$locally integrable functions on Ω, see Definition 3.1
$L^\infty(\Omega)$	space (of equivalence classes of) functions that are essentially bounded on Ω, see Definition 3.1
$W^{m,p}(\Omega)$	Sobolev space of up to order $m \in \mathbb{N}$ weak differentiable, $p-$Lebesgue integrable functions on Ω whose partial derivatives up to order m are p-Lebesgue integrable on Ω, see Definition 3.7
$H^m(\Omega)$	$W^{m,2}(\Omega)$
$V_c(\Omega \cup \Omega_\mathcal{I})$	nonlocal volume-constrained energy space, see equation (5.5)
$\mathcal{V}^\infty$	high regularity solution space, see Remark 5.9
$W(0,T)$	shortcut for $W^{1,2}(0,T;V)$, see equation (3.4)

Mathematical Symbols

$\mathbb{N}$	set of natural numbers
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\mathbb{R}$	set of real numbers
$\mathbb{R}^n$	$n-$dimensional space of vectors with n real entries
$\mathbb{R}_+$	set of nonnegative real numbers
$\mathbb{R}^n_+$	set of all vectors with n nonnegative real entries
Ω	set of interest, $\Omega \subseteq \mathbb{R}^n$ open and connected
$\Omega_\mathcal{I}$	interaction domain, $\Omega_\mathcal{I} \subseteq \Omega^c$

Norms

$\lVert \cdot \rVert_2$	Euclidean norm
$\lVert \cdot \rVert_{\mathcal{C}_b^m(\Omega)}$	$\lVert\varphi\rVert_{\mathcal{C}(\Omega)} := \sup_{\substack{\lvert\alpha\rvert\leq m \\ x\in\Omega}} \lvert D^\alpha\varphi(x)\rvert$, $m \in \mathbb{N}_0$
$\lVert \cdot \rVert_{\mathcal{C}^{m,\lambda}(\Omega)}$	$\lVert\varphi\rVert_{\mathcal{C}^{m,\lambda}(\Omega)} := \lVert\varphi\rVert_{\mathcal{C}^m(\Omega)} + \max_{0\leq\lvert\alpha\rvert\leq m}[D^\alpha\varphi]_{\mathcal{C}^{0,\lambda}(\Omega)}$
$[\cdot]_{\mathcal{C}^{0,\lambda}(\Omega)}$	$[D^\alpha\varphi]_{\mathcal{C}^{0,\lambda}(\Omega)} := \sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{\lvert D^\alpha\varphi(x)-D^\alpha\varphi(y)\rvert}{\lvert x-y\rvert^\lambda}$, λ-th Hölder seminorm
$\lVert \cdot \rVert_{L^p(\Omega)}$	$\lVert u\rVert_{L^p(\Omega)} := \left(\int_\Omega \lvert u(x)\rvert^p dx\right)^{\frac{1}{p}},\ 1 \leq p < \infty$
$\lVert \cdot \rVert_{L^\infty(\Omega)}$	$\lVert u\rVert_{L^\infty(\Omega)} := ess\sup_{x\in\Omega} \lvert u(x)\rvert$
$\lVert \cdot \rVert_{W^{m,p}(\Omega)}$	$\lVert u\rVert_{W^{m,p}(\Omega)} := \left(\Sigma_{\lvert\alpha\rvert\leq m} \int_\Omega \lvert D^\alpha u\rvert^p dx\right)^{\frac{1}{p}} \quad (1 \leq p < \infty)$
$\lVert \cdot \rVert_{W^{m,\infty}(\Omega)}$	$\lVert u\rVert_{W^{m,\infty}(\Omega)} := \Sigma_{\lvert\alpha\rvert\leq m} ess\sup_\Omega \lvert D^\alpha u\rvert$
$\lVert \cdot \rVert_{W(0,T)}$	$\lVert u\rVert^2_{W(0,T)} := \lVert u\rVert^2_{L^2(0,T;H^1(\Omega))} + \lVert u_t\rVert^2_{L^2(0,T;H^{-1}(\Omega))}$
$\lvert\lvert\lvert \cdot \rvert\rvert\rvert$	$:= \lVert u\rVert_{V_c(\Omega\cup\Omega_\mathcal{I})} = \left(\frac{1}{2}\int_{\Omega\cup\Omega_\mathcal{I}}\int_{\Omega\cup\Omega_\mathcal{I}} \mathcal{D}^*(u)(x,y)^T(\mathcal{D}^*(u)(x,y))\ dydx\right)^{\frac{1}{2}}$
$\lVert \cdot \rVert_{\mathcal{V}^\infty}$	$\lVert u\rVert_{\mathcal{V}^\infty} := \lVert u\rVert_{\mathcal{C}([0,T];V_c(\Omega\cup\Omega_\mathcal{I}))} + \lVert u\rVert_{L^\infty(\Omega\times(0,T))}$

Operators

Δ	Laplace operator, $\Delta f = div(grad(f))$
$\mathcal{D}$	nonlocal divergence operator, see equation 5.2
$-\mathcal{D}^*$	nonlocal gradient operator, see equation 5.3
$\mathcal{L}$	local-nonlocal diffusion operator, see equation (4.2)
$\mathcal{NL}$	nonlocal diffusion operator, see equation (5.1)
P	productivity operator, see equation (3.11)
$\mathcal{P}$	productivity-production operator, see equation (3.12)

Parameters

α	> 0, local diffusion weight
β	≥ 0, nonlocal diffusion weight
γ	> 0, space discount rate
δ	> 0, capital depreciation rate
η	> 0, parameter of the function ϕ_η
ε	> 0, radius of interest
μ	$\in (0, \varepsilon)$, productivity radius
T	$\in \mathbb{N}$, finite time horizon
τ	> 0, time discount rate
ξ	> 0

List of Figures

Bibliography

Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. Princton University Press.

Acemoglu, D., Golosov, M., and Tsyvinski, A. (2011). Political economy of Ramsey taxation. *Journal of Public Economics*, 95(7-8): 467–475.

Adams, R. and Fournier, J. (2003). *Sobolev Spaces*. Elsevier, 2 edition.

Aldashev, G., Aldashev, S., and Carletti, T. (2014). On convergence in the spatial AK growth models. *Working Paper, arXiv preprint arXiv:1401.4887*.

Alt, W. (2002). *Nichtlineare Optimierung: Eine Einführung Theorie, Verfahren und Anwendungen*. Springer Verlag Heidelberg.

Anderson, L. and Anderson, J. (2000). Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing. *Review of Derivatives Research*, 4(3): 231–262.

Aniţa, S., Capasso, V., Kunze, H., and La Torre, D. (2013). Optimal control and long-run dynamics for a spatial economic growth model with physical capital accumulation and pollution diffusion. *Applied Mathematics Letters*, 26(8): 908–912.

Appell, J. and P.P.Zabrejko (1990). Nonlinear superposition operators. *Cambridge University Press*.

Arbia, G. (2001). Modelling the geography of economic activities on a continuous space. *Papers in Regional Science*, 80(4): 411–424.

Armstrong, H. and Riley, D. (1987). The north-south controversy and Britain's regional problems. *Local Economy*, 2(2): 93–105.

Arrow, K. J. (1951). An extension of the basic theorems of classical welfare economics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 507–532, Berkeley, Calif. University of California Press.

Bala, V. and Sorger, G. (2001). A spatial-temporal model of human capital accumulation. *Journal of Economic Theory*, 96: 153–179.

Barnett, D. (2016). Britain's north-south divide: How it affects education, economy and gender pay-gap?

Barro, R. J. (1990). Government spending in a simple model of endogeneous growth. *Journal of political economy*, 98(5, Part 2): S103–S125.

Barro, R. J. and Sala-i Martin, X. (1995). *Economic growth*. McGraw-Hill, Inc., New York.

Becker, R. (1980). On the long-run steady state in a simple dynamic model of equilibrium with heterogeneous households. *The Quarterly Journal of Economics*, 95: 375–382.

Becker, R., Borisso, K., and Dubey, R. (2015). Ramsey equilibrium in a model with liberal borrowing. *Caepr Working Papers 2015-003*. Center for Applied Economics and Policy Research, Economics Department, Indiana University Bloomington.

Becker, R., Boyd, J., and Foias, C. (1991). The existence of Ramsey equilibrium. *Econometrica*, 59(2): 441–460.

Becker, R. and Foias, C. (1987). A characteriszation of Ramsey equilibrium. *Journal of Economic Theory*, 41: 173–184.

Blackaby, D. H. and Manning, D. N. (1990). The north-south divide: questions of existence and stability. *The Economic Journal*, 100(401): 510–527.

Borges, A. M. (1986). Applied general equilibrium models: An assessment of their usefulness for policy analysis. *OECD journal: economic studies*, 7.

Borissov, K. and Dubey, R. S. (2015). A characterization of Ramsey equilibrium in a model with limited borrowing. *Journal of Mathematical Economics*, 56: 67–78.

Boucekkine, R., Camacho, C., and Fabbri, G. (2013). Spatial dynamics and convergence: The spatial AK model. *Journal of Economic Theory*, 148: 2719–2736.

Boucekkine, R., Camacho, C., and Zou, B. (2009). Bridging the gap between growth theory and the new economic geography: The spatial Ramsey model. *Macroeconomic Dynamics*, 13: 20–45.

Breinlich, H., Ottaviano, G. I., and Temple, J. R. (2014). Regional growth and regional decline. In *Handbook of Economic Growth*, volume 2, 683–779. Elsevier.

Briani, M., Chioma, C. L., and Natalini, R. (2004). Convergence of numerical schemes for viscosity solutions in integro-differential degenerate parabolic problems arising in financial theory. *Numerische Mathematik*. Springer Verlag.

Brida, J. and Accinelli, E. (2007). The Ramsey model with logistic population growth. *Economics Bulletin*, 3(15).

Brito, P. (2001). A Bentham-Ramsey model for spatially heterogeneous growth. *Working Papers Depeartment of Economics*. ISEG, University of Lisboa.

Brito, P. (2004). The dynamics of growth and distribution in a spatially heterogeneous world. *Working Papers Departement of Economics*. ISEG, University of Lisboa.

Brito, P. (2012). Global endogeneous growth and distributional dynamics. *Munich Personal RePEc Archive*.

Brock, W. and Xepapadeas, A. (2006). Diffusion-induced instability and pattern formation in infinite horizon recursive optimal control. 32: 2745–2787.

Brock, W., Xepapadeas, A., and Yannacopoulos, A. N. (2012). Optimal agglomerations in dynamic economics. *FEEM Working Paper*. Available at SSRN: https://ssrn.com/abstract=2161785 or http://dx.doi.org/10.2139/ssrn.2161785.

Brock, W., Xepapadeas, A., and Yannacopoulos, A. N. (2014). Optimal agglomerations in dynamic economics. *Journal of Mathematical Economics*, 53(C): 1–15.

Brock, W. A., Xepapadeas, A., and Yannacopoulos, A. (2013). Adjustment costs and long run spatial agglomerations. *FEEM Working Paper*. Available at SSRN: https://www.econstor.eu/bitstream/10419/89610/1/NDL2013-068.pdf.

Brown, G. (2016). We need a brexit deal that heals the north-south divide. *The Guardian - Online Journal International Edition*. Available at https://www.theguardian.com/commentisfree/2016/nov/08/gordon-brown-brexit-deal-north-south-divide-power-regions-britain; checked on March, 17, 2018.

Burch, N. and Lehoucq, R. (2011). Classical, nonlocal, and fractional diffusion equations on bounded domains. *International Journal for Multiscale Computational Engineering*, 9(6).

Camacho, C., Zou, B., and Briani, M. (2008). On the dynamics of capital accumulatin across space. *European Journall of Operational Research*, 186(2): 451–465.

Casas, E. (2006). Optimal control of PDE theory and numerical analysis. *HAL archives-ouvert, Castro Urdiales (Espagne)*, 3: pp. 136. https://cel.archives-ouvertes.fr/cel-00392187.

Casas, E., Ryll, C., and Tröltzsch, F. (2013). Sparse optimal control of the Schlögl and FitzHugh–Nagumo systems. *Computational Methods in Applied Mathematics*, 13(4): 415–442.

Cass, D. (1965). Optimum growth in an aggregative model of capital accumulation. *The Review of Economic Studies*, 32(3): 233–240.

Chandra, S. R. and Mukherjee, D. (2016). Barrier option under Lévy model: A PIDE and Mellin transform approach. *Mathematics*, 4(1): 2.

Chasseigne, E., Chaves, M., and Rossi, J. D. (2006). Asymptotic behavior for nonlocal diffusion equations. *Journal de Mathmatiques Pures et Appliques*, 86(3): 271 – 291.

Chen, X. and Gunzburger, M. (2011). Continuous and discontinuous finite element methods for a peridynamics model of mechanics. *Computer Methods in Applied Mechanics and Engineering*, 200(9-12): 1237–1250.

Cont, R. and Voltchkova, E. (2005). A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. *SIAM Journal on Numerical Analysis*, 43(4): 1596–1626.

Dautray, R. and Lions, J.-L. (1985). *Mathematical Analysis and Numerical Methods for Science and Technology: Volume 4 Integral Equations and Numerical Methods*. Springer-Verlag Berlin Heidelberg.

Dautray, R. and Lions, J.-L. (1992). *Mathematical Analysis and Numerical Methods for Science and Technology: Volume 5 Evolution Problems I*. Springer-Verlag Berlin Heidelberg.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3): 273–292.

D'Elia, M., Du, Q., Gunzburger, M., and Lehoucq, R. (2014). Finite range jump processes and volume-constrained diffusion problems. *Sandia National Labs SAND 2014-2584J*.

D'Elia, M. and Gunzburger, M. (2014). Optimal distributed control of nonlocal steady diffusion problems. *SIAM Journal on Control and Optimization*, 52(1): 243–273.

D'Elia, M., Perego, M., Bochev, P., and Littlewood, D. (2016). A coupling strategy for nonlocal and local diffusion models with mixed volume constraints and boundary conditions. *Computers & Mathematics with Applications*, 71(11): 2218–2230.

Desmet, K. and Rossi-Hansberg, E. (2009). Spatial development. *NBER Working Paper Series.* National Bureau of Economic Research, Cambridge.

DiBenedetto, E. (1993). *Degenerate Parabolic Equations.* Springer Verlag, New York.

Driver, B. K. (2003). Analysis tools with applications. *Lecture Notes.*

Du, Q., Gunzburger, M., Lehoucq, R., and Zhou, K. (2012a). Analysis and approximation of nonlocal diffusion problems with volume constraints. *SIAM Review*, 54(4): 667–696.

Du, Q., Gunzburger, M., Lehoucq, R., and Zhou, K. (2012b). A nonlocal vector calculus, nonlocal volume-constrained problems, and nonlocal balance laws. *Sandia National Labs SAND 2010-8353J.*

Du, Q., Huang, Z., and Lehoucq, R. (2014). Nonlocal convection-diffusion volume-constrained problems and jump processes. *Discrete and Continuous Dynamical Systems Series B*, 19(4): 961–977.

Elstrodt, J. (2005). *Maß- und Integrationstheorie*, volume 4. Spinger.

Evans, L. C. (1997). *Partial Differential Equations*, volume 19. Graduate Studies in Mathematics, American Mathematical Society.

Fujita, M., Krugman, P., and Venables, A. (1999). *The Spatial Economy.* MIT Press, Cambridge.

Green, A. E. (1988). The north-south divide in Great Britain: An examination of the evidence. *Transactions of the Institute of British Geographers*, 13(2): 179–198.

Grossmann, C. and Roos, H. (2005). *Numerische Behandlung partieller Differentialgleichungen.* Teubner.

Gunzburger, M. and Lehoucq, R. B. (2010). A nonlocal vector calculus with application to nonlocal boundary value problems. *Multiscale Modeling & Simulation*, 8(5): 1581–1598.

Heinemann, M. (2015). *Dynamische Makroökonomik*, volume 1. Gabler Verlag, Springer-Verlag Heidelberg Berlin.

Jost, J. (2013). *Partial Differential Equations.* Springer. Graduate Texts in Mathematics.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of economic perspectives*, 6(2): 117–136.

Knight, F. H. (1944). Diminishing returns from investment. *The Journal of Political Economy*, 26–47.

Koopmans, T. (1965). On the concept of optimal economic growth.

Krugman, P. (1991). *Geography and Trade.* MIT Press, Cambridge.

Krugman, P. (2011). The new economic geography, now middle-aged. *Regional Studies*, 45(1): 1–7.

Ladyženskaya, O., Solonnikov, V., and Uralćeva, N. (1968). *Linear and Quasi-linear Equations of Parabolic Type*, volume 23. american mathematical society. Translation of Mathematical Monographs.

Lebherz, J., Loerx, A., Petra, N., and Sachs, E. (2018). Second order adjoints in optimization. Unpublished manuscript.

Lin, Y. and Tait, R. (1993). Finite-difference approximations for a class of non-local parabolic boundary value problems. *Journal of computational and applied mathematics*, 47(3): 335–350.

Lions, J. (1971). *Optimal control of systems governed by partial differential equations.* Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer-Verlag.

Lucas, R. and Stokey, N. (1984). Optimal growth with many consumers. *Journal of Economic Theory.*

Magill, M. and Quinzii, M. (1998). *Theory of incomplete markets*, volume 2. MIT Press, Cambridge, Massechusetts.

Martin, R. (1988). The political economy of Britain's north-south divide. *Transactions of the Institute of British Geographers*, 13(4): 389–418.

Matache, A.-M., Von Petersdorff, T., and Schwab, C. (2004). Fast deterministic pricing of options on Lévy driven assets. *ESAIM: Mathematical Modelling and Numerical Analysis*, 38(1): 37–71.

Mirrorme22, Brythones, Nilfanion: English and Scottish Council Areas, TUBS: Welsh council areas, and Sting: Gibraltar (2016). United Kingdom EU referendum 2016 area results.svg. `https://commons.wikimedia.org/wiki/File/3AUnited_Kingdom_EU_referendum_2016_area_results.svg`, Creative Commons Attribution-Share Alike 3.0 Unported license `https://creativecommons.org/licenses/by-sa/3.0/deed.en`. Downloaded on March, 22, 2018.

Morrissey, K., ODonoghue, C., et al. (2011). The spatial distribution of labour force participation and market earnings at the sub-national level in Ireland. *Review of Economic Analysis*, 3(1): 80–101.

Olson Jr., M., Sarna, N., and Swamy, A. (2000). Governance and growth: A simple hypothesis explaining cross-country differences in productivity growth. *Public Choice*, 341–364. Kluwer Academic Printers.

Pavilos, T. and Wang, P. (1996). Spatial agglomeration and endogeneous growth. *Regional Science and Uraban Economics*, 26: 645–669. Elsevier Science B.V.

Pearson, J., Stoll, M., and Wathen, A. (2011). Preconditioners for state constrained optimal control problems with Moreau-Yosida penalty function. *Numerical Linear Algebra with Applications*, 1–16.

Ramsey, F. (1928). A mathematical theory of saving. *The Economic Journal*, 38(152): 543–559.

Ran, Q. and Zhang, T. (2010). Existence and uniqueness of bounded weak solutions of a semilinear parabolic PDE. *Journal of Theoretical Probability*, 23(4): 951–971.

Ratković, K. (2016). Limitations in direct and indirect methods for solving optimal control problems in growth theory. *Industrija*, 44(4): 19–46.

Ricardo, D. (1891). *Principles of political economy and taxation (1817)*. G. Bell and sons.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5): 1002–1037.

Rudin, W. (1973). *Functional Analysis*. McGraw-Hill Book Company.

Sachs, E. W. and Strauss, A. K. (2008). Efficient solution of a partial integro-differential equation in finance. *Appl. Numer. Math.*, 58(11): 1687–1703.

Sager, S. (2009). Reformulations and algorithms for the optimization of switching decisions in nonlinear optimal control. *Journal of Process Control*, 19(8): 1238–1247.

Samuelson, P. (1947). *Foundations of Economic Analysis*. Harvard University Press.

Schumpeter, J. A. (1939). *Business cycles*, volume 1. Cambridge University Press.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*.

Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 65–94.

Sorger, G. (2002). On the long-run distribution of capital in the Ramsey model. *Journal of Economic Theory*, 105: 226–243.

Stiglitz, J. E. (2018). Where modern macroeconomics went wrong. *Oxford Review of Economic Policy*, 34(1-2): 70–106.

Swan, T. W. (1956). Economic growth and capital accumulation. *Economic record*, 32(2): 334–361.

Tian, X. and Du, Q. (2013). Analysis and comparison of different approximations to nonlocal diffusion and linear peridynamic equations. *SIAM Journal on Numerical Analysis*, 51(6): 3458–3482.

Topaloglou, L. and Petrakos, G. (2008). Identifying barriers related to economic geography across the external borders of Europe. Technical report, Discussion Paper Series.

Tröltzsch, F. (2005). *Optimale Steuerung partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen*. Vieweg+Teubner Verlag.

Van, C. L. and Vailakis, Y. (2003). Existence of a competitive equilibrium in a one sector growth model with heterogeneous agents and irreversible investment. *Economic Theory*, 22.

Weil, D. (2013). *Economic Growth*, volume 3. Pearson Addison Wesley.

Werner, D. (2007). *Funktionalanalysis*. Springer-Lehrbuch. Springer Berlin Heidelberg.

Wloka, J. (1982). *Partielle Differentialgleichungen: Sobolevräume und Randwertaufgaben*. Mathematische Leitfäden. Teubner.

Young, A. A. (1928). Increasing returns and economic progress. *The Economic Journal*, 38(152): 527–542.