Growth and agglomeration in the heterogeneous space: a generalized AK approach

GROWTH AND AGGLOMERATION IN THE HETEROGENEOUS SPACE: A GENERALIZED AK APPROACH

ABSTRACT. We provide an optimal growth spatio-temporal setting with capital accumulation and diffusion across space in order to study the link between economic growth triggered by capital spatio-temporal dynamics and agglomeration across space. The technology is AK, K being broad capital. The social welfare function is Benthamite. In sharp contrast to the related literature, which considers homogeneous space, we derive optimal location outcomes for any given space distributions for technology and population. Both the transitional spatio-temporal dynamics and the asymptotic spatial distributions are computed in closed form. Concerning the latter, we find, among other results, that: (i) due to inequality aversion, the consumption per capital distribution is much flatter than the distribution of capital per capita; (ii) endogenous spillovers inherent in capital spatio-temporal dynamics occur as capital distribution is much less concentrated than the (pre-specified) technological distribution; (iii) the distance to the center (or to the core) is an essential determinant of the shapes of the asymptotic distributions, that is relative location matters.

Key words: Growth, agglomeration, heterogeneous and continuous space, capital mobility, infinite dimensional optimal control problems

Journal of Economic Literature Classification: R1; O4; C61.

Date: March 31, 2021.

1. Introduction

Economic growth models with a spatial dimension have been already formulated in the context of the New Economic Geography stream, but, as observed by Desmet et Rossi-Hansberg (2010) in an illuminating survey (see also Nijkamp and Poot, 1998), they use to disregard intertemporal optimization, individual behaviors, and even capital accumulation. A paradigmatic example of such a growth modeling strategy can be seen in Fujita and Thisse (2002), Chapter 11. In this chapter, endogenous growth is driven by the manufacturing sector through horizontal differentiation *à la* Grossman-Helpman while skilled labor is the unique mobile factor.[1] Consumers do not save nor do they decide about schooling (no human capital accumulation). Indeed, with some notable exceptions (see for example the infrastructure location model developed by Martin and Rogers, 1995), the New Economic Geography has roughly left in the dark not only capital accumulation (over time) but also capital mobility through space.

This paper is concerned with the relationship between agglomeration and economic growth. As outlined by Fujita and Thisse (2002), "...in a world of globalization, agglomeration may well be the territorial counterpart of economic growth much in the same way as growth seems to foster inequality among individuals." (page 19). We shall provide a spatio-temporal setting with capital accumulation and diffusion across space showing the link between economic growth triggered by capital spatio-temporal dynamics and agglomeration across space.[2] In

[1]A more elaborate modeling of labor mobility and migrations can be found in Mossay (2003).

[2]The spatio-temporal setting is analogous to Brito's (2004) framework, which is itself an optimal control reformulation of the work of Isard and Liossatos (1979). In the latter, production uses a neoclassical production function at any location, output is used for *in situ* consumption and investment, while the net trade flow depends on the differentials of the spatially distributed capital stock, consistently with recent empirical results by Comin et al. (2012). Only a limited characterization of optimal solutions is possible in this case, see also Boucekkine et al. (2009).

line with Boucekkine et al. (2013), we choose the simplest production function generating growth endogenously, the AK technology. This is essential to get the analytical results gathered in this framework. It is worth pointing out here that, consistently with the growth literature (see for example, Barro and Sala-i-Martin, 1995, Chapter 4), the AK production function only makes sense if we have a broad view of capital: capital is not only physical, it also embodies human capital and knowledge. Capital diffusion across space makes therefore perfect sense. More importantly, our setting is a sharp generalization of Boucekkine et al. (2013): while in the latter space is homogeneous (same production function and one individual per location), we derive here optimal location outcomes for **any given space distributions for technology and population**. Technology space heterogeneity amounts to discrepancy on parameter A of the AK technology across locations, that is, roughly speaking, spatial differences in productivity, which can be itself due to a wide variety of pure technological or institutional factors.

In such a framework, we shall prove that capital accumulation and diffusion, and subsequent growth in the spatially heterogeneous economy, do come with agglomeration along the optimal spatio-temporal paths. Notice that here agglomeration occurs for different reasons than those usually invoked in the New Economic Geography. First, and trivially, capital accumulation and mobility is the dynamic engine of agglomeration in our story, and it is little doubtful that in real economies capital is more mobile than labor (see Aslund and Dabrowski, 2008, for a series of studies on this issue, especially in the European case). Of course, demand mobility, which is the main focus of the New Economic Geography literature, is of utmost relevance in regional dynamics, but it is also unquestionable that capital mobility is being a massive phenomenon, in particular in Europe. As such, the development of spatio-temporal models deriving the agglomeration implications of the latter sounds as a legitimate and necessary task. Second, we do not have increasing returns in our setting (the production function is linear) nor do we impose monopolistic competition (optimal growth setting). Third, using Krugman's terminology (1993), we do look for **first nature causes** for agglomeration as the technology and demographic distributions are exogenously given, and not for the **second nature causes** typically invoked in the New Economic Geography (like economies of scale or knowledge spillovers).

More precisely, we consider a planner problem whose objective is to maximize an intertemporal utilitarian social welfare function by identifying the optimal capital spatio-temporal paths for any given technological and demographic spatial (time-independent) distributions. The planner chooses the optimal path for consumption per location (and therefore also the investment path per location), and consequently drives the optimal capital flows through space at any time. There is a single consumption good and all the individuals (consumers) have the same (strictly concave) utility function whatever their location. The social welfare function is Benthamite (see discussion below), the most common specification in growth theory (see Barro and Sala-i-Martin, 1995, Chapter 2). Taking into account the two latter specifications, the social welfare function may be also interpreted in terms of an inequality-averse social planner with risk-neutral consumers. The main contribution of this paper is to characterize the optimal short-term spatio-temporal dynamics and the resulting optimal asymptotic distributions for the relevant variables together with the identification of the main economic mechanisms, resulting from technological and demographic spatial heterogeneity, at work. Incidentally, we address the research questions opened by Isard and Liossatos (1979) at the highest level of generality so far.

On the technical side, generalizing Boucekkine et al. (2013) approach to heterogeneous space is a daunting task. We have been able however to find a way to undertake it. More precisely, we are able to explicitly identify the maximal welfare (value function) and the optimal consumption profile in terms of technology and population spatial distributions and the initial spatial distribution of capital (Theorem 3.2). We also single out the partial differential equation which delivers the optimal spatio-temporal capital dynamics and study the asymptotic convergence properties associated. Ultimately, we are able to describe the long-run profile of the capital distribution in an explicit way by a suitable series of spatial functions (Theorem 3.4).[3] As a particular case, considering uniform distributions for both technology and population leads exactly to Boucekkine et al.'s uniform convergence results. We can therefore study the robustness of the asymptotic convergence to uniform spatial distributions to population and technology space dependence.

Indeed, we shall explore the properties of optimal spatio-temporal dynamics along many more dimensions. We proceed as follows. Mimicking the so-called Alonso-Mills-Muth monocentric city model (see Thisse and Fujita, 2002, Chapter 3), we consider three different types of pre-specified centers.[4] In the first case, we study the implications of a given technological center, *i.e* productivity showing a single-peaked spatial shape, while population distribution is uniform. In the second case, the demographic center configuration is analyzed, *i.e* population

[3]The results are obtained employing dynamic programming methods in infinite dimensions and the main methodological novelty of the present work with respect to the existing literature in spatial growth models: the use of the spectrum and the eigenfunctions of an appropriate Sturm-Liouville operator $\mathcal{L}$, the one associated to the (linear) zero consumption problem. A precise description of the techniques together with a complete proof of all the analytical results is given in Appendix A.
[4]Beside the obvious differences between the monocentric city models and ours (for example no land in our model and no capital in the former), it is worth pointing that the center corresponds to a point in the Alonso-Mills-Muth model, whereas it is a non-zero measure arc of circle in our setting.

density showing a single-peaked spatial shape, while productivity is identical across locations. Last but not least, we examine the case where productivity is related to population density, a larger population density being a driver of technological development. In particular, we use the specification of Allen and Arkolakis (2014).[5] For each predetermined center, we uncover the optimal spatio-temporal capital, consumption and output dynamics, and the corresponding asymptotic spatial distributions.

Our analysis produces several interesting results both for the short and long-run. As to the role of initial conditions, that is the initial distribution of capital, we show analytically that the asymptotic distribution is independent of the initial one, though the latter does matter in the short run spatio-temporal capital dynamics.[6] In contrast, the asymptotic distributions are closely related to the spatial distributions of population and technology. Though we analytically single out this link, it is shown to be remarkably complex. Numerical exercises using the three pre-specified spatial centers cases listed above are therefore needed to dig deeper into this relationship. These exercises allow to identify on an adequately calibrated version of the model two main effects at work when space distributions of technology and population are heterogeneous.

On the one hand, we have a technological spatial discrepancy effect or a **productivity effect**: the planner has the incentive to favor the concentration of the capital in the areas where it is more productive so that she will tend to promote (relatively more) investment in areas where

[5]Such a specification is not specific to economic geography, it is also common in unified growth theory, see Galor and Weil (2000). Boucekkine et al. (2007) provide with micro-foundations: larger population densities foster school creations, which in turn speeds up human capital accumulation, and therefore technological progress.

[6]While this property is consistent with the non-spatial AK model where the long-run capital level is proportional to its the initial one, it's not at all granted in our spatio-temporal model where capital flows across locations.

technology is better and to (relatively) privilege consumption in technologically lagged regions. On the other hand, we have a demographic spatial discrepancy effect or a **population effect**: the Benthamite form of the functional considered, entailing inequality-aversion, induces the planner to guarantee an adequate level of per capita consumption across space, so that areas with higher population get also a higher aggregate consumption and therefore a lower investment. Consumption and capital asymptotic distributions are characterized in several essential ways. At first, because of aversion to inequality, it is shown that the former is much flatter. Second, we identify a kind of endogenous spillover inherent in capital spatio-temporal dynamics: this shows up for instance in the fact that capital distribution is much less concentrated than the (pre-specified) technological distribution. Spatial spillovers do arise as the combination of capital (exogenous) diffusivity and the endogenous investment and consumption decisions taken by the planner. Third, we observe that the distance to the center (or to the core) is an essential determinant of the shapes of the asymptotic distribution, that is relative location matters.[7] Last but not least, the exogenous technological distribution does affect the shape of the asymptotic distribution of the per-capita consumption, while the demographic distribution only affects its level.

Incidentally, the two effects disentangled give a clear idea of why and how the "unequal treatment of equals" works in our spatio-temporal social optimum framework. Because the asymptotic spatial distributions do not depend on initial capital distributions, the two effects described above also apply to rigorously equal individuals, namely with same preferences and same initial capital endowment, as in the original related urban economics works due to Mills and McKinnon (1973) and Levhari et al. (1978). They are also equal from the intertemporal point

[7]This feature is shared with the Alonso-Mills-Muth model.

of view: they are infinitely-lived and population per location is constant over time, so for example, time discounting applies in the same way to all individuals at any time and place. While the resulting "unequal treatment of equals" outcome may be due to the Benthamite form of the social welfare function as in the static urban economics literature (see Mirrlees, 1972, and more recently, Wildasin, 1986),[8] we aim here at characterizing the optimal short-term and asymptotic spatial distributions due to capital to spatio-temporal dynamics in the standard social optimum set-up in growth theory, which is already a daunting analytical task. Within this set-up, we show how the population and technological effects work separately and then how they interact, in particular under the (nonlinear) specification previously implemented by Allen and Arkolakis (2014). Sensitivity exercises with respect to inequality parameters are also provided.

The paper proceeds as follows. Section 2 is devoted to description of the model. Section 3 presents the main analytical results. Section 4 concerns numerical simulations and associated remarks. Section 5 concludes. Appendix A provides the proofs of the analytical results.

2. The model

We study a spatial economy developing on the unit circle S^1 in the plane[9]:

$$S^1 := \{(\sin\theta, \cos\theta) \in \mathbb{R}^2 \,:\, \theta \in [0, 2\pi)\}.$$

[8]Boucekkine et al (2014) prove in a non spatial setting that when age structure matters, typically when lifetime is finite, and when the social planner chooses the optimal population size, the Benthamite social welfare function does ensure egalitarianism in consumption per capita across generations!

[9]The functions over S^1 can be clearly identified with 2π-periodic functions over $\mathbb{R}$. We shall then identify these functions, as well as the point $\theta \in [0, 2\pi)$ with the corresponding point $(\sin\theta, \cos\theta) \in S^1$. Hence, given a function $f : S^1 \to \mathbb{R}$, the derivatives with respect to $\theta \in S^1$ will be intended through the identification of functions defined on S^1 with 2π-periodic functions defined on $\mathbb{R}$.

We suppose that, for all time $t \geq 0$ and any point in the space $\theta \in [0, 2\pi)$, the production is a linear function of the employed capital:

$$Y(t,\theta) = A(\theta)K(t,\theta),$$

where $K(t,\theta)$ and $Y(t,\theta)$ represent, respectively, the aggregate capital and output at the location θ at time t, while $A(\theta)$ is the exogenous location-dependent technological level. In the model there is no state intervention and then, at any time, the local production is split into investment in local capital and local consumption, so that, once we include a location-dependent depreciation rate $\delta(\theta)$ and the net trade balance $\tau(t,\theta)$, we get the following accumulation law of capital:

$$\begin{aligned}\frac{\partial K}{\partial t}(t,\theta) &= I(t,\theta) - \delta(\theta)K(t,\theta) - \tau(t,\theta)\\ &= Y(t,\theta) - C(t,\theta) - \delta(\theta)K(t,\theta) - \tau(t,\theta)\\ &= (A(\theta) - \delta(\theta))K(t,\theta) - C(t,\theta) - \tau(t,\theta).\end{aligned}$$

We can always include the depreciation rate $\delta(\theta)$ in the coefficient $A(\theta)$ so the previous equation simply becomes

$$\frac{\partial K}{\partial t}(t,\theta) = A(\theta)K(t,\theta) - C(t,\theta) - \tau(t,\theta).$$

Now we model the term $\tau(t,\theta)$ in the above equation. Following the idea of Brito (2004) and then used in all the papers of the related stream of literature (see, for instance: Brock and Xepapadeas, 2008, Boucekkine et al., 2013, Fabbri, 2016, and the references therein), we assume that the left-to-right flow rate of capital across a point equals the opposite of the derivative of the capital level at such point. Imposing that the net trade balance of the region $(\theta_1, \theta_2) \subset [0, 2\pi)$ equals the outflow of capital at the boundaries θ_1 and θ_2 yields

$$\int_{\theta_1}^{\theta_2} \tau(t,\theta)d\theta = \frac{\partial K}{\partial \theta}(t,\theta_1) - \frac{\partial K}{\partial \theta}(t,\theta_2).$$

Since

$$\frac{\partial K}{\partial \theta}(t,\theta_1) - \frac{\partial K}{\partial \theta}(t,\theta_2) = -\int_{\theta_1}^{\theta_2} \frac{\partial^2 K}{\partial \theta^2}(t,\theta)d\theta,$$

we get

$$\int_{\theta_1}^{\theta_2} \left[\tau(t,\theta) + \frac{\partial^2 K}{\partial \theta^2}(t,\theta)\right] d\theta = 0,$$

hence, by arbitrariness of (θ_1, θ_2),

$$\tau(t,\theta) = -\frac{\partial^2 K}{\partial \theta^2}(t,\theta). \tag{1}$$

The capital evolution law reads then as

$$\frac{\partial K}{\partial t}(t,\theta) = \frac{\partial^2 K}{\partial \theta^2}(t,\theta) + A(\theta)K(t,\theta) - C(t,\theta).$$

Then, if for each (t,θ) we express the total consumption $C(t,\theta)$ as the product of the per-capita consumption[10] $c(t,\theta)$ and the time-independent exogenous (density of) population $N(\theta)$, we get the state equation[11]

$$\begin{cases} \frac{\partial K}{\partial t}(t,\theta) = \frac{\partial^2 K}{\partial \theta^2}(t,\theta) + A(\theta)K(t,\theta) - c(t,\theta)N(\theta), \ t > 0, \, \theta \in S^1, \\ K(0,\theta) = K_0(\theta), \quad \theta \in S^1, \end{cases} \tag{2}$$

where $K_0 : S^1 \to [0,\infty)$ is the function denoting the initial distribution of capital over the space S^1. Throughout the rest of the paper, we assume that

$$K_0 \text{ is square integrable, i.e. } \int_0^{2\pi} |K_0(\theta)|^2 d\theta < \infty. \tag{3}$$

[10] We suppose resources and consumption are equally distributed among the population of a certain location.

[11] Clearly the above derivation is only informal and the assumptions on the involved functions are at the moment not specified. The formal treatment of the capital evolution equation can be found in Appendix A.

We suppose that the policy maker operates to maximize the following intertemporal constant relative risk aversion functional:

$$\int_0^\infty e^{-\rho t}\left(\int_0^{2\pi}\frac{c(t,\theta)^{1-\sigma}}{1-\sigma}N(\theta)d\theta\right)dt, \tag{4}$$

where $\rho > 0$ and $\sigma \in (0,1)\cup(1,\infty)$ are given constants and the constraints

$$c(t,\theta) \geq 0, \qquad \text{and} \qquad K(t,\theta) \geq 0 \tag{5}$$

are imposed[12]. We note that, as the integrands keeps the sign, Tonelli's Theorem applies to get

$$\begin{aligned}&\int_0^\infty e^{-\rho t}\left(\int_0^{2\pi}\frac{c(t,\theta)^{1-\sigma}}{1-\sigma}N(\theta)d\theta\right)dt\\ =&\int_0^{2\pi} N(\theta)\left(\int_0^\infty e^{-\rho t}\frac{c(t,\theta)^{1-\sigma}}{1-\sigma}dt\right)d\theta.\end{aligned}$$

The latter is indeed a Benthamite functional in the following sense: at any time t, the planner linearly weights the per-capita utility at any location using the population density. In other terms, the consumption/utility of all the people in the economy matters in the same way in the target. This fact will have a certain importance in the following. It is also very important to notice that our functional can be interpreted as the social welfare function of an inequality-averse social planner with risk-neutral consumers. More generally, parameter σ could be interpreted as a mix of individual risk-aversion and societal inequality-aversion. Inequality aversion is indeed a fundamental ingredient of the problem, as we will see along the way.

The described model is a strict generalization of that considered by Boucekkine et al. (2013), because we consider here a technological level $A(\theta)$ and a population density $N(\theta)$ depending on the location θ. In

[12]More precisely, the planner chooses the function $c : [0,+\infty)\times S^1 \to [0,+\infty)$ with the goal of maximizing (4); see Appendix A for the rigorous formulation of the problem.

other words, here A and N are functions $A, N\colon S^1 \to \mathbb{R}$ instead of just two space-independent constants.

3. Main analytical results

The model presented in the previous section is, mathematically speaking, an optimal control problem with state equation (2), objective functional (4) and pointwise constraints (5). In this section we present the two main analytical results of this paper. The first characterizes the optimal strategies of the optimal control problem (2)-(4)-(5), while the second studies the long run behavior of the optimal capital path. As our results will be expressed in terms of the eigenvalues and the eigenfunctions of a suitable Sturm-Liouville problem, we begin our exposition by recalling the definitions of these concepts and some related results. In what follows we will avoid all mathematical difficulties which are unnecessary at this stage, hence many concepts will be expressed in an informal way: the reader interested in the complete mathematical setting can find precise definitions, statements and proofs in the technical Appendix A.

We will work under the following standing assumption on the (functional) parameters A, N:

(6)

$$A, N\colon S^1 \to [0, \infty) \text{ are measurable, bounded, not identically zero.}$$

We consider the differential operator associated to the zero-consumption diffusion dynamics of (2), namely

$$\mathcal{L}u(\theta) := \frac{\partial^2}{\partial \theta^2} u(\theta) + A(\theta) u(\theta). \tag{7}$$

The operator $\mathcal{L}$ is well defined on regular enough functions $\phi\colon S^1 \to \mathbb{R}$. A non identically zero regular function $\phi\colon S^1 \to \mathbb{R}$ is called *eigenfunction* of $\mathcal{L}$ if there exists a real number (*eigenvalue*) λ such that $\mathcal{L}\phi = \lambda\phi$.

The following claims are consequence of an application to our context of (regular) Sturm-Liouville theory with periodic boundary conditions[13] (see Chapter 7 and Chapter 8, Section 3, in Coddington and Levinson, 1955). There is a countable discrete set of eigenvalues $\{\lambda_n\}_{n\geq 0}$, which can be ordered in decreasing way; $\lambda_n \to -\infty$ as $n \to \infty$; the algebraic and geometric multiplicities of each eigenvalue coincide and are either 1 or 2; the highest eigenvalue, λ_0, is simple, i.e. its (algebraic/geometric) multiplicity is 1. Moreover, considering that eigenfunctions are clearly defined up to a multiplicative factor, we consider a normalized sequence of eigenfunctions $\{\mathbf{e}_n\}_{n\geq 0}$, associated to the sequence of eigenvalues[14] $\{\lambda_n\}_{n\geq 0}$, such that $\int_0^{2\pi} \mathbf{e}_n^2(\theta)d\theta = 1$; this sequence of eigenfunctions is an orthonormal basis of $L^2(S^1)$ (see (17) for the definition of this space); the eigenfunction $\mathbf{e}_0$ in this sequence is the only one without zeros and, without loss of generality, we assume that $\mathbf{e}_0(\theta) > 0$ for each $\theta \in S^1$.

We note that clearly $\{\lambda_n\}_{n\geq 0}$ and $\{\mathbf{e}_n\}_{n\geq 0}$ only depend on the distibution $A(\cdot)$. The eigenvalues are increasing in $A(\cdot)$ (see Theorem 2.9.1. of Brown et al.) in the following sense: if $\tilde{A}(\cdot)$ is another technological distribution and $\{\tilde{\lambda}_n\}_{n\geq 0}$ is the associated sequence of eigenvalues, then

$$\tilde{A}(\theta) \geq A(\theta) \quad \forall \theta \in S^1 \implies \tilde{\lambda}_n \geq \lambda_n \quad \forall n \geq 0.$$

It is easily seen that the strict inequality $\tilde{\lambda}_0 > \lambda_0$ holds above if and only if the set of S^1 where $\tilde{A}(\cdot) > A(\cdot)$ has positive measure. Concerning

[13]Indeed, the periodic boundary conditions $u(0) = u(2\pi)$ and $u'(0) = u'(2\pi)$ associated to the differential operator $\mathcal{L}$ acting on functions $u : [0, 2\pi] \to \mathbb{R}$ clearly correspond to considering, as we do here, the operator $\mathcal{L}$ on functions $u : S^1 \to \mathbb{R}$.

[14]In the sequence $\{\lambda_n\}_{n\geq 0}$ we consider that a certain value appears once, respectively twice, if its multiplicity is 1, respectively 2.

to this, it is worth noticing here that our model "does not see single points" in the sense that a change of the functional parameters $A(\cdot)$ and $N(\cdot)$ over isolated points (or even over a null measure set of points) does not affect the results: in order to have a change in the outputs which follow it is needed to change the functional parameters on "thick" sets, i.e. on sets of positive measure. A particular explicit comparison can be performed by increasing $A(\cdot)$ by a constant $\delta > 0$, i.e. by considering $\tilde{A}(\cdot) = A(\cdot) + \delta$. In this case it is easily seen that $\tilde{\lambda}_n = \lambda_n + \delta$ for each $n \geq 0$, whereas $\tilde{\mathbf{e}}_n = \mathbf{e}_n$ for each $n \geq 0$, where $\tilde{\mathbf{e}}_n$ is the eigenvector associated to $\tilde{\lambda}_n$ for the distribution $\tilde{A}(\cdot)$.

We have now collected the elements we need to describe the solution of the model and we can proceed by presenting it. We will work under the following spatial counterpart of the usual assumption needed in the standard one-dimensional AK model to ensure the finiteness of the intertemporal utility[15].

Hypothesis 3.1. *The discount rate* ρ *satisfies*

$$\rho > \lambda_0(1-\sigma). \tag{8}$$

We can now state the first important result on optimal spatio-temporal capital dynamics together with the optimal consumption strategy across time and space.

Theorem 3.2. *Denote by* α_0 *the value*[16]

$$\alpha_0 := \left(\frac{\sigma}{\rho - \lambda_0(1-\sigma)} \int_0^{2\pi} \mathbf{e}_0(\theta)^{-\frac{1-\sigma}{\sigma}} \mathbf{N}(\theta) d\theta \right)^{\frac{\sigma}{1-\sigma}}, \tag{9}$$

[15]The assumption that we will make on A will imply that λ_0 is positive (see Remark A.8). Hence, the condition (8) is obviously verified when $\sigma > 1$ (that is the case for reasonable calibrations of the model, see Section 4).

[16]This number is well defined and strictly positive thanks to (8).

and by $\boldsymbol{\beta}$ *the function* $\alpha_0 \mathbf{e}_0$. *Let* K^* *be the unique solution to the linear integro-PDE*

$$\begin{cases} \frac{\partial K}{\partial t}(t,\theta) = \mathcal{L}K(t,\theta) - \left(\int_0^{2\pi} \boldsymbol{\beta}(\eta)K(\eta)d\eta\right)\boldsymbol{\beta}(\theta)^{-\frac{1}{\sigma}}N(\theta), \\ K(0,\theta) = K_0(\theta), \ \theta \in S^1. \end{cases} \tag{10}$$

and assume that it is nonnegative. Then $K^*(t,\cdot)$ *is the optimal capital distribution at time* $t \geq 0$. *Moreover, the optimal consumption strategy* c^* *is given, as a feedback function of the current optimal state trajectory, as:*

$$c^*(t,\theta) = \left(\int_0^{2\pi} \boldsymbol{\beta}(\eta)K^*(t,\eta)d\eta\right)\left(\boldsymbol{\beta}(\theta)\right)^{-1/\sigma}, \qquad t \geq 0, \ \theta \in S^1. \tag{11}$$

Finally $c^*(t,\theta)$ *can also be expressed explicitly in terms of the initial capital density* $K_0(\theta)$ *as*

$$c^*(t,\theta) = \left(\int_0^{2\pi} \boldsymbol{\beta}(\eta)K_0(\eta)d\eta\right) e^{gt}\left(\boldsymbol{\beta}(\theta)\right)^{-1/\sigma}, \tag{12}$$

where g *is the optimal growth rate of the economy, given by*

$$g := \frac{\lambda_0 - \rho}{\sigma}. \tag{13}$$

Proof. See Appendix A and, in particular, Corollary A.5. □

Note that, on the one hand, as a consequence of the monotonicity of λ_0 with respect to $A(\cdot)$, in the sense described before Hypothesis 3.1, the growth rate g defined by (13) increases when the technological level increases over a set of positive measure of points of space, as expected. On the other hand, the population distribution does not play any role in the value of g (exactly as in the standard one-dimensional AK model the size of the population does not count for growth rate).

Once we compare the optimal consumption profile described in the previous theorem with the counterpart under space homogeneity (Boucekkine et al., 2013), we can immediately figure out the crucial role of a location-dependent technology (via coefficient A). Indeed under homogeneous space, the (per-capita and aggregate) optimal consumption level is always equal across locations, while here the expression of the optimal consumption is given by the space-independent term $\left(\int_0^{2\pi} \boldsymbol{\beta}(\eta) K_0(\eta) d\eta\right) e^{gt}$ and by the space-dependent term $(\boldsymbol{\beta}(\theta))^{-1/\sigma} = (\alpha_0 \mathbf{e}_0(\theta))^{-1/\sigma}$. The latter depends on $A(\cdot)$ both *via* α_0 and $\mathbf{e}_0$ and on $N(\cdot)$ *via* α_0. This fact is interesting from a theoretical point of view, since *a priori* one might guess that the egalitarian nature of the Benthamite functional could be enough to guarantee equalization of individual utility across space. As mentioned in the introduction, the "unequal treatment of equals" with a Benthamite social planner is not an odd result in urban economics, we simply show that it also hold in our spatio-temporal model with exogenous technology and demography. In our setting, the structural conditions of the economy can lead the planner to diversify per-capita consumption across locations (first nature causes).[17] As we will see in Section 4 the differentiation does not always go in the expected way.

As shown by (12), the "shape" of $c^*(t, \cdot)$ (i.e. the relative size of the consumption at different locations) only depends on the distribution of $A(\cdot)$, while the distribution of $N(\cdot)$ only impact its level (via the value of α_0). In particular, if we imagine to move some population from a low per-capita-consumption (i.e. $\mathbf{e}_0$ is high) location to a high (i.e. $\mathbf{e}_0$ low) per-capita-consumption location, nothing happens at the level of the shape of the per-capita distribution, but we have a level effect. Its

[17]Wildasin (1986) proposes to switch to Rawlsian planners to get rid of "unequal treatment of equals" (see also Fujita and Thisse, 2002, Chapter 3). However, as shown by Boucekkine et al. (2014) in a non-spatial dynamic model, the Benthamite planner can be egalitarian if she is allowed to choose population size.

sign depends on the value of σ: if $\sigma \in (0,1)$ it is negative, if $\sigma > 1$ it is positive. Observe that, nevertheless, the whole distribution of N takes action directly in the optimal evolution of the capital described by (10) via a dilution effect, so that, the higher the population in a certain region, the lower the long run capital. The direct effect of N on the capital distribution will be even clearer in Theorem 3.4, where the coefficients β_n of the series and then the shape of the long-run (detrended) distribution of the capital will explicitly depend on the shape of $N(\cdot)$.

Notice for now that, by the expression of the optimal consumption, we get the following expression for optimal social welfare:

$$V(K_0(\cdot)) = \frac{\alpha_0^{1-\sigma}\left(\int_0^{2\pi} K_0(\theta)\mathbf{e}_0(\theta)d\theta\right)^{1-\sigma}}{1-\sigma}. \tag{14}$$

Differently from the homogeneous space case, where maximal welfare only depends on aggregate capital, here the stock of capital in different locations enter the optimal welfare expression with different weights. Roughly speaking (see Section 4 for numerical examples), the spatial function $\mathbf{e}_0$ tends to be larger in the regions where A is bigger. So, for a given amount of initial aggregate capital, welfare will be higher if capital is more accumulated in the more productive locations. Finally, observe that this property holds true irrespectively of the population distribution, as one can realize by rewriting the expression of $V(K_0)$ above and disentangling the contributions of population and capital initial densities:

$$\left(\frac{\sigma}{\rho-\lambda_0(1-\sigma)}\int_0^{2\pi}\mathbf{e}_0(\theta)^{-\frac{1-\sigma}{\sigma}}\mathbf{N}(\theta)d\theta\right)^{\sigma}\frac{\left(\int_0^{2\pi} K_0(\theta)\mathbf{e}_0(\theta)d\theta\right)^{1-\sigma}}{1-\sigma}.$$

Heterogeneous technology and population distributions are also essential in our second result describing the long-run profile of the detrended optimal capital: while in case of space-constant A and N the space-distribution of the wealth always converges (under the hypotheses of Theorem 3.4) to a uniform profile, here an articulated expression, depending on the whole technological and human population distributions, arises. We need the following.

Hypothesis 3.3. *The optimal growth rate g defined in* (13) *satisfies*

$$g > \lambda_1, \tag{15}$$

where λ_1 is the second eigenvalue of the problem $\mathcal{L}\phi = \lambda\phi$.

Theorem 3.4. *Let the hypotheses of Theorem 3.2 hold and let Hypothesis 3.3 hold too. Define the detrended optimal path $K_g^*(t,\theta) := e^{-gt}K^*(t,\theta)$, for $t \geq 0$. Then*

$$K_g^*(t,\theta) \overset{t\to\infty}{\longrightarrow} \overline{K}_g^{K_0}(\theta) \quad \textit{uniformly in } \theta \in S^1,$$

where

$$\overline{K}_g^{K_0}(\theta) := \int_0^{2\pi} K_0(\eta)\boldsymbol{\beta}(\eta)d\eta \left(\frac{\mathbf{e}_0(\theta)}{\alpha_0} + \sum_{n\geq 1} \frac{\beta_n}{\lambda_n - g}\mathbf{e}_n(\theta) \right) \qquad \forall \theta \in S^1,$$

where

$$\beta_n := \int_0^{2\pi} (\boldsymbol{\beta}(\eta))^{-1/\sigma} N(\eta)\mathbf{e}_n(\eta)d\eta, \qquad \forall n \geq 1.$$

Proof. See Appendix A and, in particular, Proposition A.7. □

A natural question which arises here is the effect of the initial distribution $K_0(\theta)$ on the long-run optimal detrended distribution $\overline{K}_g^{K_0}$, i.e., roughly speaking, the ergodicity of the process of economic growth arising in the above Theorem 3.4. Indeed, the convergence result contained in Theorem 3.4 can be already seen as an ergodicity-type result. To show that, we first observe that, as the optimal capital path $K^*(t,\theta)$ is the unique solution to (10), the optimal detrended capital path $K_g^*(t,\theta)$

must be the unique solution to

$$(16)\quad \begin{cases} \frac{\partial K}{\partial t}(t,\theta) = (\mathcal{L}-g)K(t,\theta) - \left(\int_0^{2\pi} \boldsymbol{\beta}(\eta)K(\eta)d\eta\right)\boldsymbol{\beta}(\theta)^{-\frac{1}{\sigma}}N(\theta), \\ K(0,\theta) = K_0(\theta),\ \theta \in S^1. \end{cases}$$

This integro-PDE can be seen, as done in Appendix A, as a linear dynamical system in a Hilbert space $\mathcal{H}$ therein defined. In such a framework, well known results on ergodicity are the so called *Mean Ergodic Theorem* (see, e.g., for the discrete time case, Theorem 1.2, Chapter 2, in Petersen, 1983) and the stronger *Pointwise Ergodic Theorem* (see, e.g., again for the discrete time case, Theorem 2.3, Chapter 2, in Petersen, 1983): they concern the mean and pointwise, respectively, convergence of the time average of the solution to an equilibrium point. Stronger results in this area are on the so-called *"strongly mixing"* (see, e.g., for the discrete time case, Section 2.5 in Petersen, 1983), concerning the convergence of the solution to an equilibrium distribution. This is exactly what Theorem 3.4 says, as it states the uniform (in $\theta \in S^1$) convergence of $K_g(t,\cdot)$ to $\overline{K}_g^{K_0}$.

Concerning the effect of the initial distribution K_0 on the long-run distribution $\overline{K}_g^{K_0}$, here we can then say that, as it happens for all AK type models:

- the set of all long run distributions is a one-dimensional linear subspace of $\mathcal{H}$;[18]
- the shape of the long-run distribution $\overline{K}_g^{K_0}$ does not depend on K_0, as the influence of K_0 is only in the multiplicative constant $\int_0^{2\pi} K_0(\eta)\beta(\eta)d\eta$.

We conclude with some comments about the dependence of the outputs with respect to the model parameters.

[18]This fact has a straightforward proof based on the basic theory of ODEs and we omit it for brevity.

- The first term of the series defining the limit distribution of the detrended optimal capital $\overline{K}_g^{K_0}$ is
$$\left(\int_0^{2\pi} K_0(\eta)\mathbf{e}_0(\eta)d\eta\right)\mathbf{e}_0;$$
the latter expression only depends on A.
- The optimal consumption path c^* and the optimal social welfare V depend on both the technological and population distributions A and N *via* α_0. The shape of c^* only depend on A.
- The following monotonicity of the optimal growth rate g with respect to A holds as a consequence of the monotonicity of λ_0 with respect to A (in the sense discussed and precised before Hypothesis 3.1): if $A(\theta)$ increases for every $\theta \in S^1$, then g increases.[19]
- The following monotonicity with respect to N holds depending on σ: if $N(\theta)$ increases for every $\theta \in S^1$ then
  - α_0 and, consequently, V, increase if $\sigma \in (0,1)$ and decrease if $\sigma > 1$;
  - c^* decreases if $\sigma \in (0,1)$ and increases if $\sigma > 1$.

Remark 3.5. *We outline that the method and the results presented here are based on the fact that an explicit solution of the HJB equation is available. This nice feature strongly relies on the structure of the problem, i.e. on the linearity of the production function and the homogeneity of the utility function. Without such a structure explicit solutions are in general not available. However, some results can be still be proved and a qualitative analysis of the optimal paths is in principle possible. A first attempt in this direction for the same family of models has been done in Brito (2004), but to get stronger, and more*

[19]The monotonicity of the optimal social welfare V with respect to A is difficult to see from (14). However, such monotonicity can be proved using the fact that V is the supremum of the functional (4).

interesting, results one should make a deeper use of infinite dimensional control techniques like the ones described in the books Li and Yong (1995), Fabbri, Gozzi and Swiech (2017), and Fattorini (1999).

4. Numerical exercises

The explicit representation of the long-run configuration of the economy given in Theorem 3.4 can be used to undertake a numerical analysis of the system in some specific cases of interest.[20]

First we calibrate the model. In all the simulations we choose the discounting parameter ρ equal to 3% (consistent e.g. with the data of Lopez, 2008) and, except in Subsection 4.1.4, the parameter σ equal to 5 (its value is coherent with those found e.g. by Barsky et al., 1997). In the simulations we use uniform and non-uniform technological spatial distributions of $A(\cdot)$ whose values are in a range of values $(0.2 \div 0.25)$ compatible with the values of the ratio output-over-capital Y/K found by Piketty and Zucman (2014).

In the various situations, computing the first eigenvalue of the operator $\mathcal{L}$ defined in (7) and using (13), we get the reasonable values of the global growth rate equal closed to 3%. As a further check, we also observe that the (spatial-heterogeneous) saving rates we obtain are in line, for instance, with the World Bank data (see e.g. World Bank Group, 2016).

Hereafter, we start with the analysis of the asymptotic spatial distributions of the relevant variables in three different pre-specified technological and demographic spatial settings. Then, we briefly present some examples of transitional spatio-temporal dynamics for uniform versus non-uniform initial distributions of capital.

[20]To numerically compute the eigenfunctions e_n we use the package *Chebfun* written for MATLAB. See Birkisson and Driscoll (2011) and Driscoll and Hale (2016) for details on the implementation of the routines on linear differential operators and in particular on eigenfunctions of Sturm-Liouville operators in Chebfun.

4.1. **Numerical exploration of the asymptotic spatial distributions.** Partly imitating the monocentric city model *à la Alonso-Mills-Muth* (see Thisse and Fujita, 2002, Chapter 3), we investigate, in the following, the long-run spatial distributions of capital, consumption and output in the case of three different types of pre-specified centers. More precisely: (i) we study in Subsection 4.1.1 the situation where productivity is peaked in some *core* region, while population density is constant in space; (ii) we analyze in Subsection 4.1.2 the reverse circumstance of single-peaked spatial population and uniform productivity; (iii) we devote Subsection 4.1.3 to the case where, as widely recognized by empirical studies (see for instance Ciccone et Hall, 1996), the regions with higher population density are also the more productive ones; following the findings of Allen and Arkolakis (2014), we will suppose that they are linked by a power law.

4.1.1. *The case of the technological center.* The effect of a peaked spatial productivity distribution, whenever the population density is constant (with density everywhere equal to 1), is represented in Figure 1. We use the technological distribution $A(\cdot)$ on $[0, 2\pi]$ having a peak at the point π and attaining lower values in the further locations represented in the first picture of Figure 1. The long-run spatial distributions of detrended capital, output, consumption and investment are represented. In gray (and dotted) we visualize the benchmark, where homogeneous technological and demographic distributions are considered[21]. It is indeed the case of Boucekkine at al. (2013). We can promptly see the effect of the spatial polarization of the capital marginal (and average) productivity on capital accumulation in the first picture of the second line of Figure 1. In fact, the capital tends to

[21]The spatially constant value of A in the benchmark is a mean of the values of the single-peaked case and ensures the same asymptotic growth rate of the economy (equal to 3.17%).

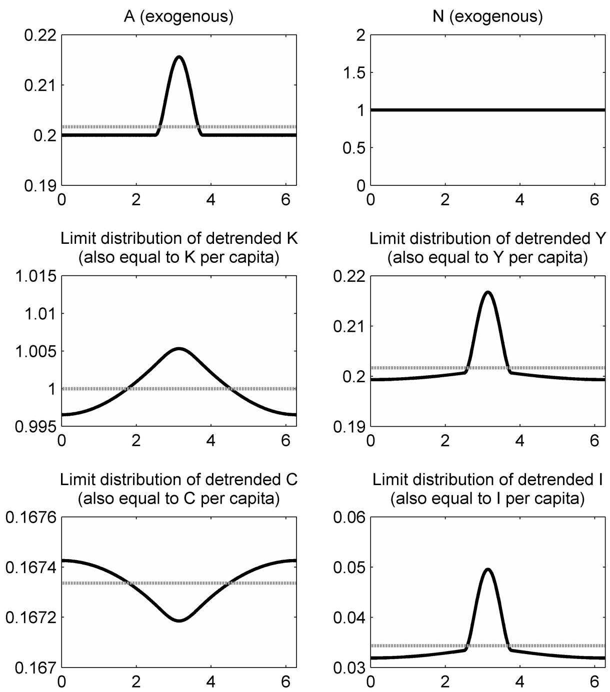

FIGURE 1. The productivity effect at work: long run distributions of the main variables for a homogeneous population distribution and a peaked total factor productivity profile

accumulate in the more productive areas, while those with lower productivity remain behind: the higher productivity of capital pushes the planner to increase investments and thus savings relatively more in the more productive regions, as shown in the second picture of the third line of Figure 1. As a byproduct the planner privileges consumption in peripheral regions, but this is a second-order effect of small magnitude, as one can see in the first picture of the third line of Figure 1. These are the outcomes of the productivity effect announced in the introduction.

Looking at the (spatial) relative magnitudes in the distributions of A compared to those of the long-run detrended K, we can easily realize that the capital distribution is much less concentrated than the technological level[22]. We can also observe that points with the same level of total factor productivity have, in the long run, different level of (detrended) capital (it is the case of the points in the line segment with extremes 4 and 6 for instance) due to the different distance with respect to the core. We have indeed an endogenous spatial spillover effect that is the combined result of both the capital exogenous diffusivity and the endogenous investment and consumption decisions by the planner.

The difference with respect to the results of Boucekkine at al. (2013) is crystal clear: once we introduce the spatial heterogeneity in capital productivity, the optimal detrended capital does not converge anymore to a spatial-homogeneous distribution. Indeed, the homogeneous space case, where all the detrended variables (capital, output, consumption, investment) converge to the spatial-homogeneous configuration, arises as a special case, only if A is constant over the locations.

Using the same parametrization, we can see, in Figure 2(a), a more substantial difference with respect to the results of Boucekkine et al. (2013). While in their case the long-run detrended net trade balance is zero everywhere, here the long-run value of τ is a non-trivial function of space (in black in the picture), having value equal to zero when integrated all along the circle. In particular the simulation reproduces a long-run flow of capital from the center (where we have a positive net-trade balance) to the periphery (where we have a negative net-trade balance) of the economy. The distribution of $A(\cdot)$ is the same as in Figure 1, it is in gray in the picture (the values of A are in the

[22]Conversely the concentration of the long-run detrended output is a little more peaked than A because the output has the form $Y = AK$.

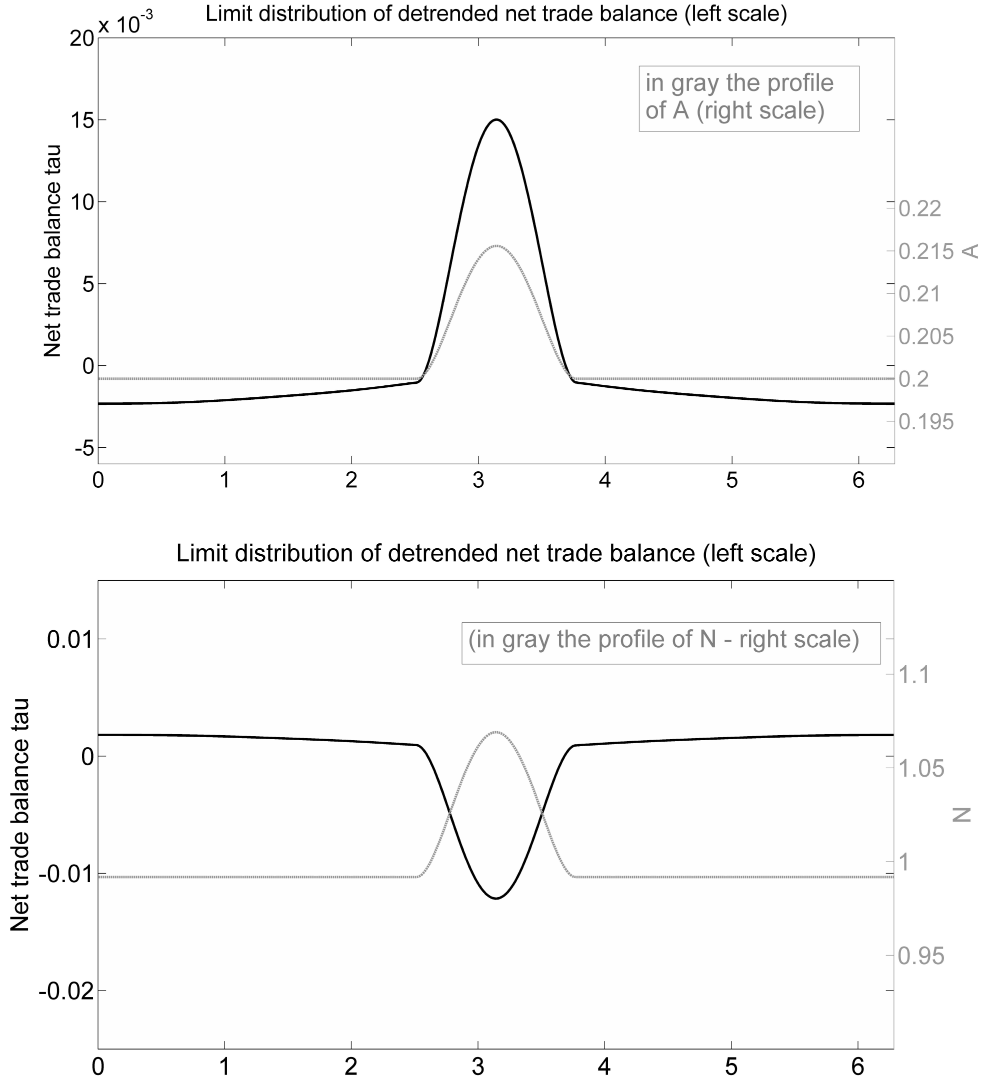

FIGURE 2. The detrended net trade balance at the long-run equilibrium

right side gray scale). The opposite dynamics arise in the case of a pre-specified demographic center (see next subsection) as shown in Figure 2(b).[23]

4.1.2. *The case of the demographic center.* Figure 3 is a first look at the population effects in the model. There we consider the same technological distribution as in Figure 1 and we vary uniformly the population

[23]The population distribution $N(\cdot)$, that is the same we will use in Figure 4, is gray and its values are represented in the right side gray scale.

density, more precisely we double the previous constant population density (in the picture the previous benchmark situation is in gray, with dotted line, while the new profile is in black, continuous line). The effect, in terms of aggregate optimal behavior is zero while per-capita variables are mechanically halved. This effect could be predicted directly from expression (11) taking into account the effect of population distribution on α_0 given by (9). Observe that this behavior is not due to the homogeneous distribution of the population we use: whatever the initial population distribution, a uniform increase of the population of $n\%$ in the whole space induces a spatial uniform proportional reduction (by a factor $\frac{1}{1+n/100}$) of per capita variables.

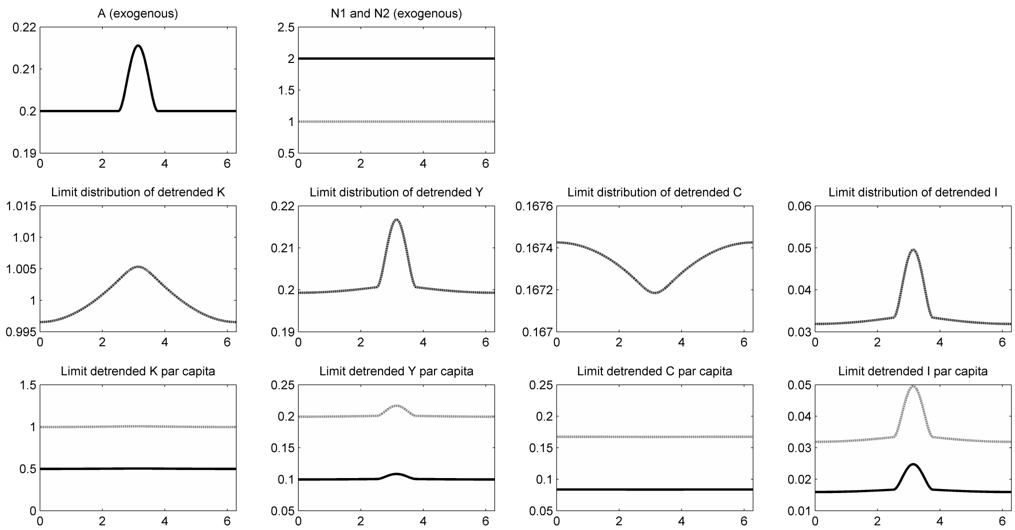

FIGURE 3. The population effect: long run distributions of the main variables for two homogeneous population distributions and a peaked total factor productivity profile

In Figure 4 we consider the case specular to the one considered in Figure 1. We consider here the case of a homogeneously distributed productivity and of a single-peaked population density distribution (and in gray, dotted line, the benchmark with homogeneous population and technological distribution, i.e. the case of Boucekkine et al., 2013). Total and per-capita capital, production and investments are, in the long run, lower in the more populated areas. The explanation of this population effect goes through the inequality-aversion of the Benthamite planner (see Subsection 4.1.4 for more on this point): to guarantee a reasonable level of consumption to everybody, the planner needs to maintain an higher level of aggregate consumption in more densely populated areas and therefore to lower investment at the same locations. This effect is completely transparent in the situation described in Figure 4 because, in absence of spatial productivity heterogeneities, the per capita consumption chosen by the planner is constant in space: all people leaving in the economy are entitled, at a fixed time, to the same consumption (see Theorem 3.2 and the subsequent discussion).

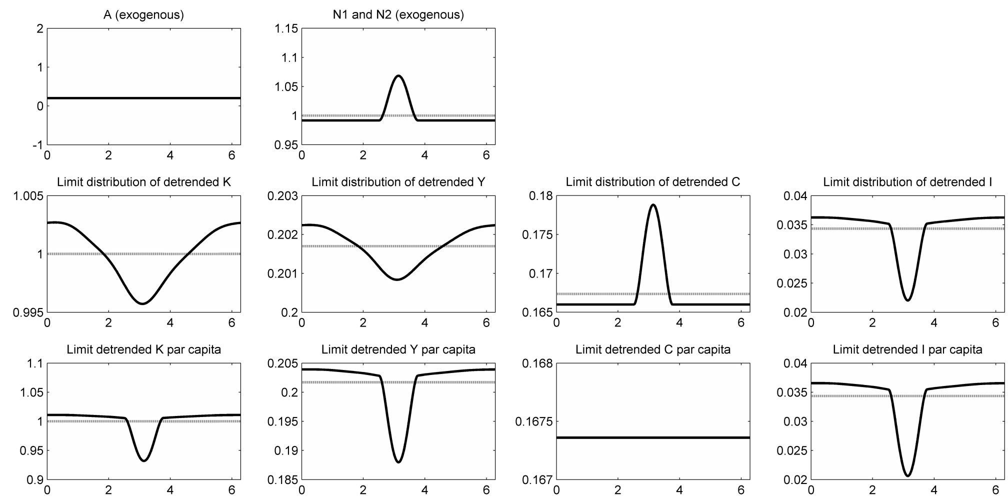

Figure 4. The population effect at work: long run distributions of the main variables for a peaked population distribution and a constant total factor productivity profile

4.1.3. *The case of productivity increasing in population density.* In Figure 5 we consider a concentration of capital productivity and population density in the same areas (a quite frequent configuration) showing how the productivity and population effects combine and can partially offset each other. In the first simulation we consider (in black and continuous line in the pictures) the same technology distribution of Figure 1 and the same population distributions of Figure 4 and we see how the productivity and the population effects combine. Observe that the choice we made was not arbitrary and indeed the (peaked) population of Figure 4 is proportional to the (peaked) technological distribution of Figure 1 so that, in the case of Figure 5, $A(\theta) = A_0 N(\theta)$ for all $\theta \in [0, 2\pi]$ for some positive real constant A_0. We also represent a benchmark (gray and dotted line in the picture) given by the homogeneous population and single-peaked technology situation of Figure 1.

In the new situation two distinct motivations drive the planner: on the one hand, she will tend to invest more in the more productive areas (productivity effect), but on the other, she wants to assign a reasonable per capita level of consumption in each region, increasing the consumption in the more populated areas (population effect). The total effect is depicted in the various pictures of Figure 5: the aggregate investment in more productive areas for the second population profile remains relatively higher[24] but the effect is mitigated because aggregate consumption is higher in these areas as well. This effects can be quantified with respect to the benchmark: the standard deviation of the capital distribution is 5 times greater in the case of uniform population than when the distribution of the population is more concentrated. Conversely, how we already observed, in the model the distribution of the population does not change the shape of the per-capita consumption distribution, so a higher population at a certain location corresponds, mechanically, to a higher aggregate consumption at the same location. For this reason, given the rather flat distribution of the

[24]This outcome depends on the chosen distribution of the population, a bigger concentration of the population would of course accentuate the population effect.

per-capita consumption, the higher concentration of the population of our second population distribution translates into a more concentrated aggregate consumption (the standard deviation increases by a factor 42 with respect to the benchmark). Since the distribution of long-run detrended capital is more uniform in the single-peaked population case, the per-capita capital accumulates more in less productive areas. For this reason the change in the population distribution translates into a (small) case of efficiency loss in the economic system and, even if almost no appreciable in the picture, per-capita consumption in the new configuration is always smaller that in the original one at any location.

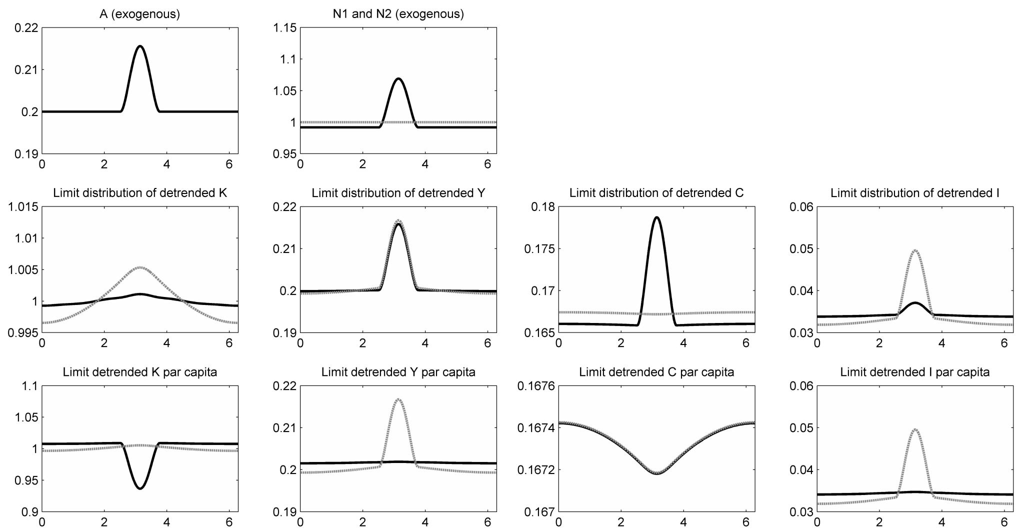

FIGURE 5. Balance between productivity and population effects: the case of proportional (single-peaked) technological and population profiles

In Figure 6 we consider a second situation where productivity and population density are higher in the same area. Here we use the specification of Allen and Arkolakis (2014), so that

$$A(\theta) = A_0 N(\theta)^\gamma$$

for some positive constants A_0 and γ. A_0 is here mostly a scaling parameter (one can see this fact by rewriting the relation as $N = (A_0)^{1/\gamma} A^{1/\gamma}$ and recalling what happens in Figure 3) and it is calibrated in order to have normalized population[25] and the same productivity of Figures 1 and 5 in the peripheral areas (this allows in particular to calibrate the model with a reasonable growth rate). Conversely, γ influences the ratio between population and productivity which is spatial dependent (apart from homogeneous distributions) as soon as $\gamma \neq 1$. Following the parametrization of Allen and Arkolakis (2014) we choose $\gamma = 0.1$. As in Figure 5 the two effects (productivity and population) are at work but a different equilibrium arises and the (aggregate and per capita) long run capital distribution is here more concentrated in less populated regions.

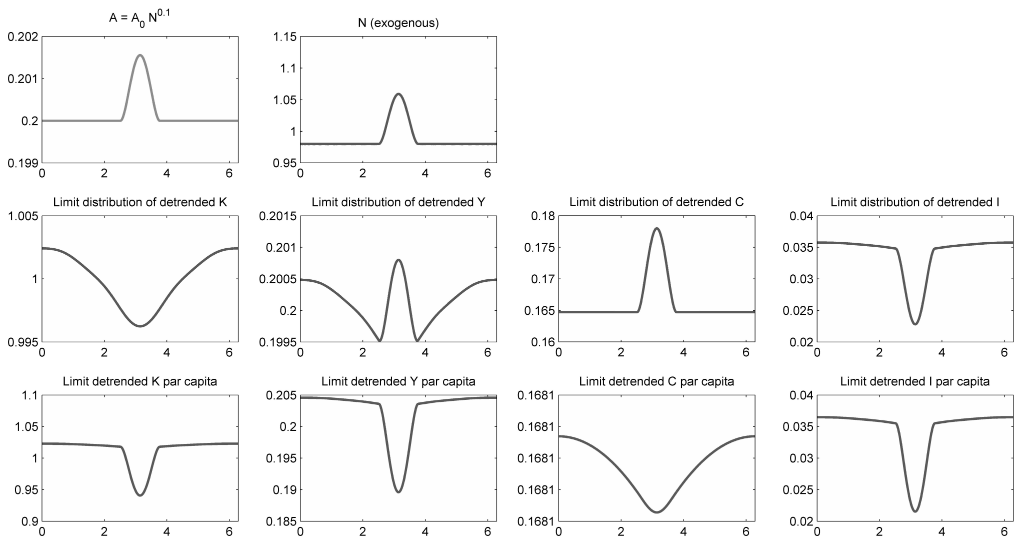

FIGURE 6. Productivity and population effects at work under the Allen and Arkolakis (2014)' specification: $A(\theta) = A_0 N(\theta)^{0.1}$

[25]All the population distributions in all the simulations have total mass equal to 2π.

Standard deviation ratio	σ_c/σ_k	σ_c/σ_y
Technological center	$0,028$	$0,017$
Demographic center	$1.2 \cdot 10^{-10}$	$6.0 \cdot 10^{-10}$
Proportional Demography and Technology	$4.2 \cdot 10^{-3}$	$0,68$
Arkolakis and Allen (2014)' case	$3.7 \cdot 10^{-4}$	$2.0 \cdot 10^{-3}$

Table 1. Ratio of standard deviation of long run distribution of per-capita consumption (σ_c) and standard deviation of long run distribution of per-capita capital (σ_k) and output (σ_y) in various scenarios.

Standard deviation ratio	σ_c/σ_k	σ_c/σ_y
$\sigma = 3$	$5.3 \cdot 10^{-4}$	$2.9 \cdot 10^{-3}$
$\sigma = 5$	$3.7 \cdot 10^{-4}$	$2.0 \cdot 10^{-3}$
$\sigma = 7$	$2.7 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$

Table 2. σ as inequality aversion parameter: shift of standard deviation ratios varying the value of σ (Arkolakis and Allen (2014)' specification)

.

4.1.4. *Aversion to inequality.* As already pointed out, the choice of a Bentahmite functional with per-capita concave utilities brings automatically some degree of the planner's inequality aversion and this fact is essential in driving the population effect as we described in the previous subsections. Indeed, since the planner's utility depends on the per-capita consumption (rather than per-capita distributions of other variables) we should expect that the dispersion of other relevant (per capita) variables is bigger than the the one of consumption. This is exactly what we have in all the cases we considered so far - Technological center, Demographic center, Proportional Demography and Technology, Arkolakis and Allen (2014)' case (corresponding to Figures 1, 4, 5 and 6). In Table 1 we show how, in each of them, the ratio between the standard deviation of the per-capita consumption σ_c and the standard deviation of the per-capita capital and production (respectively σ_k and σ_y) is always much smaller than 1.

Indeed, as already observed in Section 2, the parameter σ appearing in (4) can be interpreted as a mix of individual preferences parameter and a planner's inequality aversion parameter. Consistently with this latter interpretation we can see in Table 2 how the consumption distribution is more and more equal[26] when σ increases. The simulation is done using the Arkolakis and Allen (2014)' specification (Figure 6) and varying σ, similar results arise for other specifications of Table 1.

4.2. **Transitional spatio-temporal dynamics.** The analytical results we get, and in particular the expansion of the optimal spatio-temporal dynamics in terms of a (temporarily weighted) series of eigenfunctions (45) allow us to simulate the evolution of the spatial distribution of various variables in the economy. In Figure 7 we show what happens to the spatial capital distribution starting from two different initial configurations. We choose the parametrization of Figure 1 (technological center); the situation, *mutatis mutandis*, is similar in other cases. In Figure 7(a) the initial capital profile is homogeneous (namely $K_0(\theta) = 1$ for any $\theta \in S^1$); despite this, we see that progressively the agglomeration process we described takes place and the system converges toward the core-periphery configuration we have in Figure 1. In Figure 7(b) we show the dynamics when the initial capital distribution is peaked in some point (different from the central core point): the initial concentration of the capital smoothens and a new core emerges. In the two cases we get, in the long-run, the same detrended capital distribution. Observe that, to better show the two images, they are rotated in different ways, but in each of them the variable going from 0 to 1 is the time, while the one in the interval $[-\pi, \pi]$ is the space coordinate.

[26]In the table we show the dispersion of the per-capita consumption in terms of dispersion of other per-capita variables, but the same result arises if we only consider the concentration of the per-capita consumption itself.

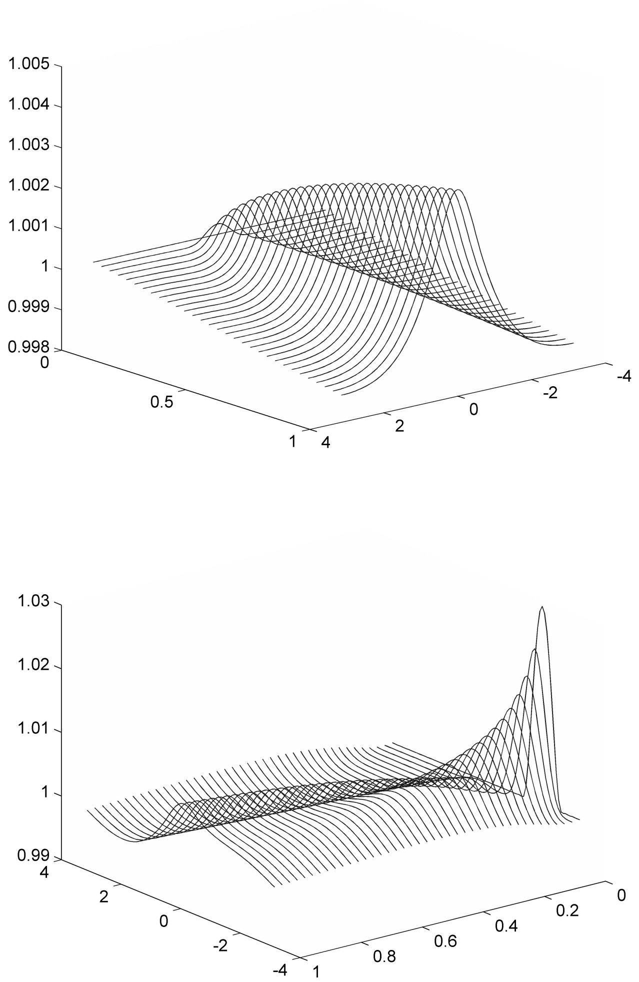

FIGURE 7. Detrended capital time evolution starting from two different initial capital distributions in space (technological center case)

Figure 8 represents the evolution of the instantaneous growth rate $\frac{\partial K_g^*(t,\theta)}{\partial t}$ of the detrended capital distribution at each point $\theta \in S^1$. As clearly reflected in the two pictures, short-run adjustments are necessary in the growth rate spatial dynamics to originate, in the long run, the core-periphery distribution of capital (and, in Figure 8(b), to overcome the initial peak of the capital distribution in the "wrong " position). Of course, in the long run, since detrended capital converges towards a (spatially heterogeneous) limit distribution, the growth rate converges, in each spatial point, to zero.

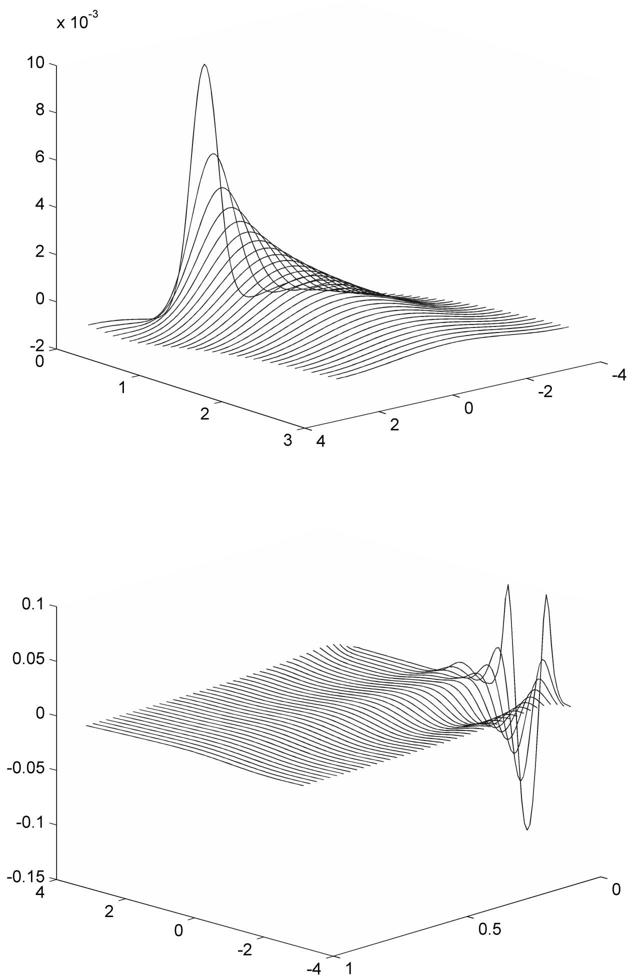

FIGURE 8. Evolution of the growth rate of the detrended capital starting from two different initial capital distributions in space (technological center case)

5. Conclusions

In this paper we introduce and study a general spatial model of economic growth. With respect to previous related contributions, our model is more general both for the possibility of studying heterogeneous spatial distributions of technology and for allowing for non-homogeneous spatially distributed population. We are able to solve it analytically by employing dynamic programming methods in infinite dimensions. This is made possible thanks to the use of the eigenfunctions of the linear Sturm-Liouville problem related to the consumption-free dynamics of the model. The numerical exercises allow to identify two opposing effects: productivity effect versus population effect. We show that the shape of agglomeration triggered by growth depends pretty much on the relative strengths of the two latter effects. Our setting delivers an agglomeration theory entirely based on optimal spatio-temporal capital dynamics for any given technology and population space distributions (first nature causes), which sharply departs from the agglomeration theories put forward in the New Economic Geography literature, which mostly disregards capital accumulation and focuses on second nature causes. We identify a form of endogenous spillover inherent in capital spatio-temporal dynamics and we observe how the distance to the core is an essential determinant of the shapes of the asymptotic distribution. The effect of the aversion to inequality on spatial distributions of the relevant variables is characterized as well.

References

Allen, T., and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics*, 129(3), 1085-1139.

Aslund, A., and Dabrowski, M. (2008). *Challenges of globalization: Imbalances and growth*. Peter G. Peterson for International Economics, Washington DC.

Barro, R., and Sala-i-Martin, X. (1995). *Economic Growth*. McGraw-Hill.

Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. *The Quarterly Journal of Economics*, 112(2), 537-579.

Birkisson, A., and Driscoll, T. A. (2012). Automatic Fréchet differentiation for the numerical solution of boundary-value problems, *ACM Transactions on Mathematical Software*, 38 1-26.

Boucekkine, R., Fabbri, G., and Gozzi, F. (2014). Egalitarianism under population change: Age structure does matter. *Journal of Mathematical Economics*, 55, 86-100.

Boucekkine, R., Camacho, C., and Fabbri, G. (2013). Spatial dynamics and convergence: The spatial AK model. *Journal of Economic Theory*, 148(6), 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(1), 20-45.

Boucekkine, R., de la Croix, D. and Peeters, D. (2007). Early litteracy achievements, population density, and the transition to modern growth. *The Journal of the European Economic Association*, 5, 183-226.

Brezis H. (2011). *Functional Analysis, Sobolev Sapces and Partial Differential Equations*, Springer Universitext, New York Dordrecht Heidelberg London.

Brito, P. (2004). The dynamics of growth and distribution in a spatially heterogeneous world. Working Papers Department of Economics 2004/14. ISEG, University of Lisbon.

Brock, W. and Xepapadeas, A. (2008). Diffusion-induced instability and pattern formation in infinite horizon recursive optimal control. *Journal of Economic Dynamics and Control*, 32(9), 2745-2787.

Brown B.M., Eastham M.S.P., Shmidt K.M. (2013). *Periodic Differential Operators*, Operator Theory: Advances and Applications Vol. 230, Birkhäuser, Basel.

Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. *The American Economic Review* 86(1), 54-70.

Coddington E.A. and N. Levinson (1955). *Theory of Ordinary Differential Equations*, McGraw-Hill, New York.

D. A. Comin, M. Dmitriev, and E. Rossi-Hansberg (2012). The spatial diffusion of technology. Technical Report w18534, National Bureau of Economic Research.

Desmet, K. and Rossi-Hansberg, E. (2010). On spatial dynamics. *Journal of Regional Science*, 50(1), 43-63.

Driscoll, T. A., and Hale, N. (2016). Rectangular spectral collocation. *IMA Journal of Numerical Analysis*, 36(1), 108-132.

Fabbri, G. (2016). Geographical structure and convergence: A note on geometry in spatial growth models. *Journal of Economic Theory*, 162(1), 114-136.

Fabbri G., Gozzi F., and A. Swiech (2017), *Stochastic Optimal control in Infinite Dimension. Dynamic Programming and HJB Equations.* Springer Series in Probability Theory and Stochastic Modeling, Vol. 82.

Fattorini, H.O. (1999), *Infinite Dimensional Optimization and Control Theory.* Cambridge University Press.

Fujita, M., and Thisse, J. (2002). *Economics of agglomeration: Cities, industrial location, and regional growth.* Cambridge University Press, Cambridge.

Galor, O., and Weil D. (2000). Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond. *American Economic Review*, 90, 806-828.

Isard, W. and Liossatos, P. (1979). *Spatial Dynamics and Optimal Space-time Development.* North-Holland.

Krugman, P. (1993). First nature, second nature, and the metropolitan location. *Journal of Regional Science*, 33(2), 129-144.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3), 483-499.

Levhari, D., Oron, Y., and Pibes, D. (1978). A note on unequal treatment of equals in an urban setting. *Journal of Urban Economics*, 5, 278-284.

Li X. and J. Yong (1995), *Optimal Control Theory for Infinite Dimensional Systems.* Birkhäuser, Boston.

Lopez, H. (2008). The social discount rate : estimates for nine Latin American countries. The World Bank Policy Research Working Paper Series 4639.

Martin, P., and Rogers, C.A. (1995). Industrial location and public infrastructure. *Journal of International Economics*, 39(3), 335-351.

Mills, E., and MacKinnon (1973). Notes on the new urban economics. *Bell Journal of Economics and Management Science*, 4, 593-601.

Mirrlees, J. (1972). The optimum town. *Swedish Journal of Economics*, 74, 115-135.

Mossay, P. (2003). Increasing returns and heterogeneity in a spatial economy. *Regional Science and Urban Economics*, 33(4), 419-444.

Nijkamp, P., and Poot, J. (1998). Spatial perspectives on new theories of economic growth. *The annals of regional science*, 32(1), 7-37.

Petersen K. (1983). *Ergodic Theory.* Cambridge University Press.

Piketty, T., and Zucman, G. (2014). Capital is back: wealth-income ratios in rich countries 1700-2010. *The Quarterly Journal of Economics*, 129(3), 1255-1310.

Wildasin, D. (1986). Spatial variation of marginal utility of income and unequal treatment of equals. *Journal of Urban Economics*, 19, 125-129.

World Bank Group (2016). *World Development Indicators 2016.* World Bank Publications.

Appendix A. Proofs of the analytical results

In order to use the infinite dimensional dynamic programming to prove Theorems 3.2 and 3.4 we first need to recall some preliminary concepts and results.

A.1. **The infinite dimensional setting.** We can represent (2) as an abstract dynamical system in infinite-dimension. Some steps are needed to describe this construction. Consider the space[27]

$$\mathcal{H} := L^2(S^1) := \left\{ f : S^1 \to \mathbb{R} \text{ measurable} \mid \int_0^{2\pi} |f(\theta)|^2 d\theta < \infty \right\}. \tag{17}$$

This is a Hilbert space when endowed with the inner product $\langle f, g\rangle := \int_0^{2\pi} f(\theta)g(\theta)d\theta$, inducing the norm $\|f\| = \int_0^{2\pi} |f(\theta)|^2 d\theta$. We will also use the following spaces of real functions defined on S^1:

$$L^\infty(S^1) := \{f \in \mathcal{H} \mid |f| \le C \text{ for some } C > 0\},$$
$$H^1(S^1) := \{f \in \mathcal{H} \mid \exists\, f' \text{ in weak sense and belongs to } \mathcal{H}\},$$
$$H^2(S^1) := \{f \in \mathcal{H} \mid \exists\, f' \text{ in weak sense and belong to } H^1(S^1)\}.$$

The differential operator

$$\mathcal{L}u := \frac{\partial^2 u}{\partial \theta^2} + A(\cdot)u, \qquad u \in H^2(S^1)$$

is well defined and $\mathcal{H}$-valued. It is also self-adjoint, i.e.

$$\mathcal{L}^* = \mathcal{L}. \tag{18}$$

The operator $\mathcal{L}$ is the sum of the Laplacian operator on S^1 with the bounded operator $\mathbf{A} : \mathcal{H} \to \mathcal{H},\ u \mapsto A(\cdot)u$. The Laplacian operator is closed on the domain

[27]To be precise, the definition of $L^2(S^1)$, as well as the definitions of the other spaces we introduce here, should involve a quotient with respect to the relation of *equality almost everywhere*. We omit these technical issues and refer to standard monographies on Lebesgue and Sobolev spaces, e.g. Brezis (2011).

$H^2(S^1)$ and generates a C_0-semigroup on the space $\mathcal{H}$. Hence, as $\mathbf{A}$ is bounded, we deduce that also $\mathcal{L}$ is closed on the domain

$$D(\mathcal{L}) := H^2(S^1)$$

and generates a C_0-semigroup on the space $\mathcal{H}$. From now on, in order to avoid confusion, we will denote the elements of $\mathcal{H}$ by bold letters. With this convention, we can formally rewrite (2) as an abstract dynamical system in the space $\mathcal{H}$:

$$\begin{cases} \mathbf{K}'(t) = \mathcal{L}\mathbf{K}(t) - \mathbf{c}(t)\mathbf{N}, & t \in \mathbb{R}^+, \\ \mathbf{K}(0) = \mathbf{K}_0 \in \mathcal{H}, \end{cases} \tag{19}$$

with the formal equalities $\mathbf{K}(t)(\theta) = K(t,\theta)$, $[\mathbf{c}(t)\mathbf{N}](\theta) = c(t,\theta)N(\theta)$ and we will read the original system as (19).[28]

By general theory of semigroups (see Proposition 3.1 and 3.2, Section II-1, of Bensoussan et al., 2007, also considering (18)), given $\mathbf{c} \in L^1_{loc}(\mathbb{R}^+;\mathcal{H})$, there exists a unique (weak) solution $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}} \in L^1_{loc}(\mathbb{R}^+;\mathcal{H})$ to (19) in the following sense: for each $\boldsymbol{\varphi} \in D(\mathcal{L})$ the function $t \mapsto \langle \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t), \boldsymbol{\varphi}\rangle$ is locally absolutely continuous and

$$\begin{cases} \frac{d}{dt}\langle \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t), \boldsymbol{\varphi}\rangle = \langle \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t), \mathcal{L}\boldsymbol{\varphi}\rangle - \langle \mathbf{c}(t)\mathbf{N}, \boldsymbol{\varphi}\rangle, & a.e.\ t \in \mathbb{R}^+, \\ \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(0) = \mathbf{K}_0 \in \mathcal{H}. \end{cases} \tag{20}$$

Consider the positive cone in $\mathcal{H}$, i.e. the set

$$\mathcal{H}^+ := \{\mathbf{K} \in \mathcal{H} \mid \mathbf{K}(\cdot) \geq 0\},$$

the positive cone in $\mathcal{H}$ without the zero function, i.e. the set

$$\mathcal{H}^+_0 := \{\mathbf{K} \in \mathcal{H} \mid \mathbf{K}(\cdot) \geq 0 \text{ and } \mathbf{K}(\cdot) \not\equiv 0\},$$

and define the set of admissible strategies as[29]

$$\mathcal{A}(\mathbf{K}_0) := \{\mathbf{c} \in L^1_{loc}(\mathbb{R}^+;\mathcal{H}^+) \mid \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t) \in \mathcal{H}^+_0 \quad \forall t \geq 0\}.$$

Then we can rewrite the original optimization problem as the one of maximizing the objective functional

$$J(\mathbf{K}_0;\mathbf{c}) := \int_0^\infty e^{-\rho t}\mathcal{U}(\mathbf{c}(t))dt, \tag{21}$$

over all $\mathbf{c} \in \mathcal{A}(\mathbf{K}_0)$ where

$$\mathcal{U} : \mathcal{H}^+ \to \mathbb{R}^+, \ \ \mathcal{U}(\mathbf{c}) := \int_0^{2\pi} \frac{\mathbf{c}(\theta)^{1-\sigma}}{1-\sigma}\mathbf{N}(\theta)d\theta.$$

In the following we call $(\mathbf{P})$ this problem and we define the associated value function as

$$V(\mathbf{K}_0) := \sup_{\mathbf{c} \in \mathcal{A}(\mathbf{K}_0)} J(\mathbf{K}_0;\mathbf{c}). \tag{22}$$

[28]The correspondence between the concept of solution to the abstract dynamical system in $\mathcal{H}$ that we introduce below (weak solution) and the solution of (2) can be argued as in Proposition 3.2, page 131, of Bensoussan et al. (2007).

[29]In this formulation we require the slightly sharper state constraint $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t) \in \mathcal{H}^+_0$ in place of the wider (original) one $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t)(\cdot) \geq 0$ almost everywhere. This is without loss of generality: indeed, if $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t) \equiv 0$ at some $t \geq 0$, the unique admissible (hence the optimal) control from t on is the trivial one $\mathbf{c}(\cdot) \equiv 0$, so we know how to solve the problem once we fall into this state and there is no need to define the Hamilton-Jacobi-Bellman equation at this point. The reason to exclude the zero function from the set $\mathcal{H}^+$ and considering the set $\mathcal{H}^+_0$ is that in this set our value function is well defined and solves the Hamilton-Jacobi-Bellman equation, while this does not happen in $\mathcal{H}^+$.

A.2. **HJB equation.** Through the dynamic programming approach we associate to the problem (**P**) the following Hamilton-Jacobi-Bellman (HJB) equation in $\mathcal{H}$ (which "should be" satisfied by the value function):

$$\rho v(\mathbf{K}) = \langle \mathbf{K}, \mathcal{L}\nabla v(\mathbf{K})\rangle + \sup_{\mathbf{c}\in\mathcal{H}^+} \{\mathcal{U}(\mathbf{c}) - \langle \mathbf{c}\mathbf{N}, \nabla v(\mathbf{K})\rangle\}. \tag{23}$$

An explicit solution of this equation can be given in a suitable half-space of $\mathcal{H}$ as shown by the following proposition.

Proposition A.1. *Let (8) hold. The function*

$$v(\mathbf{K}) = \frac{\langle \mathbf{K}, \alpha_0 \mathbf{e}_0\rangle^{1-\sigma}}{1-\sigma}, \quad \mathbf{K} \in \mathcal{H}^+_{\mathbf{e}_0}, \tag{24}$$

where

$$\mathcal{H}^+_{\mathbf{e}_0} := \{\mathbf{K} \in \mathcal{H} \mid \langle \mathbf{K}, \mathbf{e}_0\rangle > 0\}. \tag{25}$$

and

$$\alpha_0 := \left(\frac{\sigma}{\rho - \lambda_0(1-\sigma)} \int_0^{2\pi} \mathbf{e}_0(\theta)^{-\frac{1-\sigma}{\sigma}} \mathbf{N}(\theta) d\theta\right)^{\frac{\sigma}{1-\sigma}}, \tag{26}$$

is a classical solution[30] *of* (23) *over* $\mathcal{H}^+_{\mathbf{e}_0}$.

Proof. Let $\mathbb{R}^{++} := (0, \infty)$ and define the strictly positive cone of $\mathcal{H}$, i.e.

$$\mathcal{H}^{++} := \left\{ f : S^1 \to \mathbb{R}^{++} \mid \int_0^{2\pi} |f(\theta)|^2 d\theta < \infty \right\},$$

Setting

$$\mathcal{U}^*(\boldsymbol{\alpha}) := \sup_{\mathbf{c}\in\mathcal{H}^+} \{\mathcal{U}(\mathbf{c}) - \langle \mathbf{c}\mathbf{N}, \boldsymbol{\alpha}\rangle\}, \quad \boldsymbol{\alpha} \in \mathcal{H}^{++},$$

we have

$$\mathcal{U}^*(\boldsymbol{\alpha}) := \sup_{\mathbf{c}\in\mathcal{H}^+} \int_0^{2\pi} \left(\frac{\mathbf{c}(\theta)^{1-\sigma}}{1-\sigma}\mathbf{N}(\theta) - \mathbf{c}(\theta)\mathbf{N}(\theta)\boldsymbol{\alpha}(\theta)\right) d\theta = \int_0^{2\pi} u^*(\mathbf{N}(\theta), \boldsymbol{\alpha}(\theta)) d\theta,$$

where

$$u^*(N, q) := \sup_{c\geq 0} \left\{\frac{c^{1-\sigma}}{1-\sigma} N - qcN\right\} = \frac{\sigma}{1-\sigma} N q^{-\frac{1-\sigma}{\sigma}}, \quad q > 0, \ N \geq 0,$$

with optimizer

$$c^*(q) = q^{-\frac{1}{\sigma}}, \quad q > 0. \tag{27}$$

Plugging (24) into (23), and using that

$$\nabla v(\mathbf{K}) = \langle \mathbf{K}, \alpha_0\mathbf{e}_0\rangle^{-\sigma} \alpha_0 \mathbf{e}_0, \quad \mathbf{K} \in \mathcal{H}^+_{\mathbf{e}_0}, \tag{28}$$

we need to check the equality

$$\frac{\rho}{1-\sigma} \langle \mathbf{K}, \alpha_0 \mathbf{e}_0\rangle^{1-\sigma} = \langle \mathbf{K}, \alpha_0 \mathcal{L}\mathbf{e}_0\rangle \langle \mathbf{K}, \alpha_0\mathbf{e}_0\rangle^{-\sigma} + \frac{\sigma}{1-\sigma} \alpha_0^{-\frac{1-\sigma}{\sigma}} \left(\int_0^{2\pi} \mathbf{e}_0(\theta)^{-\frac{1-\sigma}{\sigma}} \mathbf{N}(\theta) d\theta\right) \langle \mathbf{K}, \alpha_0 \mathbf{e}_0\rangle^{1-\sigma}. \tag{29}$$

By definition of λ_0 and $\mathbf{e}_0$, we have $\mathcal{L}\mathbf{e}_0 = \lambda_0 \mathbf{e}_0$. So (29) holds by (26). □

For notational reasons we set

$$\boldsymbol{\beta} := \alpha_0 \mathbf{e}_0,$$

[30] By a classical solution of (23) in an open subset $\mathcal{H}_1$ of $\mathcal{H}$ we mean a function $\psi : \mathcal{H}_1 \to \mathbb{R}$ which is C^1 in its domain and which verifies (23) at every point $\mathbf{K} \in \mathcal{H}_1$.

so we can rewrite (24) as

$$v(\mathbf{K}) = \frac{\langle \mathbf{K}, \boldsymbol{\beta}\rangle^{1-\sigma}}{1-\sigma}, \quad \mathbf{K} \in \mathcal{H}^+_{\mathbf{e}_0}. \tag{30}$$

Moreover, by definition of $\boldsymbol{\beta}$ and by (26), we get the following identity that will be useful in the next subsection

$$\left(\int_0^{2\pi} \boldsymbol{\beta}(\theta)^{-\frac{1-\sigma}{\sigma}} \mathbf{N}(\theta) d\theta\right) = \frac{\rho - \lambda_0(1-\sigma)}{\sigma}. \tag{31}$$

A.3. **Solution of the optimal control problem via dynamic programming in infinite dimensions.** Proposition A.1 suggests to consider a different set of admissible controls, i.e.

$$\mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0) := \{\mathbf{c} \in L^1_{loc}(\mathbb{R}^+; \mathcal{H}^+) \mid \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t) \in \mathcal{H}^+_{\mathbf{e}_0} \quad \forall t \geq 0\}.$$

Since $\mathcal{H}^+_0 \subseteq \mathcal{H}^+_{\mathbf{e}_0}$, we have also $\mathcal{A}(\mathbf{K}_0) \subseteq \mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)$. We define an auxiliary problem associated to this new relaxed constraint, i.e.

$$(\tilde{\mathbf{P}}) \qquad \text{Maximize } J(\mathbf{K}_0; \mathbf{c}) \text{ over } \mathbf{c} \in \mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0) \tag{32}$$

The value function of the problem $(\tilde{\mathbf{P}})$ is

$$\tilde{V}(\mathbf{K}_0) := \sup_{\mathbf{c} \in \mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)} J(\mathbf{K}_0; \mathbf{c}). \tag{33}$$

Clearly we have the inequality

$$\tilde{V} \geq V \quad \text{over } \mathcal{H}^+_0. \tag{34}$$

The reason to consider the relaxed state constraint $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(\cdot) \in \mathcal{H}^+_{\mathbf{e}_0}$, in place of the stricter original one $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(\cdot) \in \mathcal{H}^+_0$, is that the former is somehow the "natural" one from the mathematical point of view and admits an explicit solution. On the other hand, the true constraint is still $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(\cdot) \in \mathcal{H}^+$, so we need to establish a relationship between the solutions of the two problems $(\mathbf{P})$ and $(\tilde{\mathbf{P}})$. Our approach relies on the following obvious result.

Lemma A.2. *If $\mathbf{c}^*$ is an optimal control for $(\tilde{\mathbf{P}})$ and $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(\cdot) \in \mathcal{H}^+_0$ (i.e. the solution of the optimization problem with relaxed state constraint actually satisfies the stricter one), then $\mathbf{c}^*$ is optimal also for $(\mathbf{P})$.*

We focus on the solution to $(\tilde{\mathbf{P}})$. Considering (27), the feedback map associated to the function v defined in (30) results in

$$\mathcal{H}^+_{\mathbf{e}_0} \to \mathcal{H}^+_0, \quad \mathbf{K} \mapsto \langle \boldsymbol{\beta}, \mathbf{K}\rangle \boldsymbol{\beta}^{-\frac{1}{\sigma}}, \tag{35}$$

where $\boldsymbol{\beta}^{-\frac{1}{\sigma}}(\theta) := (\boldsymbol{\beta}(\theta))^{-\frac{1}{\sigma}}$. By using the same results invoked for equation (19) above we find that the associated closed loop equation

$$\begin{cases} \mathbf{K}'(t) = \mathcal{L}\mathbf{K}(t) - \langle \boldsymbol{\beta}, \mathbf{K}(t)\rangle \boldsymbol{\beta}^{-\frac{1}{\sigma}} \mathbf{N}, \\ \mathbf{K}(0) = \mathbf{K}_0 \in \mathcal{H}^+_0, \end{cases} \tag{36}$$

admits a unique weak solution, in the sense that there exists a unique function $\mathbf{K}^{\mathbf{K}_0,*} \in L^1_{loc}(\mathbb{R}^+; \mathcal{H})$ such that the function $t \mapsto \langle \mathbf{K}^{\mathbf{K}_0,*}(t), \boldsymbol{\varphi}\rangle$ is absolutely continuous for every $\boldsymbol{\varphi} \in D(\mathcal{L})$ and

(37)

$$\begin{cases} \frac{d}{dt}\langle \mathbf{K}^{\mathbf{K}_0,*}(t), \boldsymbol{\varphi}\rangle = \langle \mathbf{K}^{\mathbf{K}_0,*}(t), \mathcal{L}\boldsymbol{\varphi}\rangle - \langle \boldsymbol{\beta}, \mathbf{K}^{\mathbf{K}_0,*}(t)\rangle \langle \boldsymbol{\varphi}, \boldsymbol{\beta}^{-\frac{1}{\sigma}} \mathbf{N}\rangle, \quad a.e.\ t \in \mathbb{R}^+, \\ \mathbf{K}^{\mathbf{K}_0,*}(0) = \mathbf{K}_0 \in \mathcal{H}^+_0. \end{cases}$$

Consider (31) and set

$$g := \lambda_0 - \int_0^{2\pi} \mathbf{N}(\theta) \boldsymbol{\beta}(\theta)^{-\frac{1-\sigma}{\sigma}} d\theta = -\frac{\rho - \lambda_0}{\sigma}. \tag{38}$$

Taking $\boldsymbol{\varphi} = \boldsymbol{\beta}$ in (37), we get

$$\langle \mathbf{K}^{\mathbf{K}_0,*}(t), \boldsymbol{\beta} \rangle = \langle \boldsymbol{\beta}, \mathbf{K}_0 \rangle e^{gt}, \quad t \geq 0, \tag{39}$$

Hence

$$\mathbf{K}_0 \in \mathcal{H}^+_{\mathbf{e}_0} \;\Rightarrow\; \mathbf{K}^{\mathbf{K}_0,*}(t) \in \mathcal{H}^+_{\mathbf{e}_0}.$$

So the control

$$\mathbf{c}^*(t) := \langle \boldsymbol{\beta}, \mathbf{K}(t) \rangle \boldsymbol{\beta}^{-\frac{1}{\sigma}} = \langle \boldsymbol{\beta}, \mathbf{K}_0 \rangle \boldsymbol{\beta}^{-\frac{1}{\sigma}} e^{gt}, \quad t \geq 0, \tag{40}$$

belongs to $\mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)$.

Lemma A.3. *For each* $\mathbf{c} \in \mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)$ *we have*

$$\langle \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t), \boldsymbol{\beta} \rangle \leq \langle \boldsymbol{\beta}, \mathbf{K}_0 \rangle e^{\lambda_0 t}, \quad \forall t \geq 0.$$

Proof. Denote by $\mathbf{0}$ the null control, i.e. the control $\mathbf{c}(t)(\theta) = 0$ for each $(t,\theta) \in \mathbb{R}^+ \times S^1$. Then (20) yields $\langle \mathbf{K}^{\mathbf{K}_0,\mathbf{0}}(t), \boldsymbol{\beta} \rangle = \langle \boldsymbol{\beta}, \mathbf{K}_0 \rangle e^{\lambda_0 t}$ for every $t \geq 0$. On the other hand, as $\boldsymbol{\beta}(\theta) > 0$ for each $\theta \in S^1$, standard comparison applied to the ODE (20) yields

$$\langle \mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(\cdot), \boldsymbol{\beta} \rangle \leq \langle \mathbf{K}^{\mathbf{K}_0,\mathbf{0}}(\cdot), \boldsymbol{\beta} \rangle, \tag{41}$$

and the claim follows. □

Theorem A.4. *Let* (8) *hold. Let* $\mathbf{K}_0 \in \mathcal{H}^+_{\mathbf{e}_0}$ *and let* $v : \mathcal{H}^+_{\mathbf{e}_0} \to \mathbb{R}$ *be the function defined in* (30). *Then* $v(\mathbf{K}_0) = \tilde{V}(\mathbf{K}_0)$ *and the control* $\mathbf{c}^*$ *defined in* (40) *is optimal for* $(\tilde{\mathbf{P}})$ *starting from the initial state* $\mathbf{K}_0$*; i.e.* $J(\mathbf{K}_0; \mathbf{c}^*) = \tilde{V}(\mathbf{K}_0)$.

Proof. The fact that $\mathbf{c}^* \in \mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)$ has been already observed in the discussion preceding Lemma A.3. We prove now the optimality. By the usual arguments employed to prove Verification Theorems within the Dynamic Programming approach, using the fact that v is a solution to (23) on $\mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)$ one gets, for every $\mathbf{c} \in \mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)$,

$$\begin{aligned} e^{-\rho t} v(\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(t)) - v(\mathbf{K}_0) &= -\int_0^t e^{-\rho s} \mathcal{U}(\mathbf{c}(s)) ds \\ &+ \int_0^t e^{-\rho s} \{ \mathcal{U}(\mathbf{c}(s)) - \langle \mathbf{c}(s)\mathbf{N}, \nabla v(\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(s)) \rangle - \mathcal{U}^*(\nabla v(\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(s)) \} ds \end{aligned} \tag{42}$$

We pass (42) to the limit for $t \to \infty$.

- We use (8) and Lemma A.3 in the left hand side;
- we use monotone convergence in the right hand side, as, by definition of $\mathcal{U}^*$, the integrand is nonpositive.

Hence, we get the so called *fundamental identity*, valid for each $\mathbf{c} \in \mathcal{A}^+_{\mathbf{e}_0}(\mathbf{K}_0)$:

$$\begin{aligned} v(\mathbf{K}_0) &= J(\mathbf{K}_0; \mathbf{c}) \\ &+ \int_0^\infty e^{-\rho s} \big\{ \mathcal{U}^*(\nabla v(\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(s)) - \big(\mathcal{U}(\mathbf{c}(s)) - \langle \mathbf{c}(s)\mathbf{N}, \nabla v(\mathbf{K}^{\mathbf{K}_0,\mathbf{c}}(s)) \rangle \big) \big\} ds. \end{aligned} \tag{43}$$

From (43), by definition of $\mathcal{U}^*$ we first get $v(\mathbf{K}_0) \geq \tilde{V}(\mathbf{K}_0)$. Then, observing that the integrand in (43) vanishes when $\mathbf{c} = \mathbf{c}^*$, we obtain $v(\mathbf{K}_0) = J(\mathbf{K}_0; \mathbf{c}^*)$. The claim follows. □

From Theorem A.4 and Lemma A.2, we get our first main result corresponding to Theorem 3.2.

Corollary A.5. *Let* (8) *hold. Let* $\mathbf{K}_0 \in \mathcal{H}^+_0$*, let* $\mathbf{c}^*$ *be the control defined in* (40) *and assume that* $\mathbf{c}^* \in \mathcal{A}(\mathbf{K}_0)$. *Then* $v(\mathbf{K}_0) = V(\mathbf{K}_0)$ *and* $\mathbf{c}^*$ *is optimal for* $(\mathbf{P})$.

Lemma A.6. *Let $\bar{A} := \sup_{\theta\in S^1} |A(\theta)|$ (the latter is finite due to* (6)*). There exists $C_0 > 0$ such that*

$$|\mathbf{e}_n|_\infty := \sup_{\theta\in S^1} |\mathbf{e}_n(\theta)| \le C_0(1+\lambda_n^2) \quad \forall n \ge 1.$$

Proof. Fix $n \ge 1$. Since $\mathbf{e}_n$ solves $\mathcal{L}\mathbf{e}_n = \lambda_n \mathbf{e}_n$, we have

$$|\mathbf{e}_n''(\theta)|^2 \le 2(\bar{A}^2 + \lambda_n^2 |\mathbf{e}_n(\theta)|^2) \quad \forall \theta \in S^1.$$

Integrating over S^1 and taking into account that $\|\mathbf{e}_n\| = 1$, we get the estimate

$$\int_0^{2\pi} |\mathbf{e}_n''(\theta)|^2 d\theta \le 4\pi\bar{A}^2 + 2\lambda_n^2,$$

form which we get, taking into account that $|x| \le x^2 + 1$ for each $x \in \mathbb{R}$,

$$\int_0^{2\pi} |\mathbf{e}_n''(\theta)| d\theta \le 4\pi\bar{A}^2 + 2\lambda_n^2 + 2\pi.$$

Weierestrass' and Fermat's Theorems ($\mathbf{e}_n$ is continuous and differentiable) ensure the existence of $\theta_0 \in S^1$ such that $\mathbf{e}_n'(\theta_0) = 0$. Hence we get the estimate for $|\mathbf{e}_n'(\theta)|$

$$|\mathbf{e}_n'(\theta)| \le \left| \int_{[\theta_0,\theta)} \mathbf{e}_n''(\theta) d\theta \right| \le \int_0^{2\pi} |\mathbf{e}_n''(\theta)| d\theta \le 4\pi\bar{A}^2 + 2\lambda_n^2 + 2\pi \quad \forall \theta \in S^1.$$

Considering that $\mathbf{e}_n(\theta_1) = 0$ for some $\theta_1 \in S^1$ (eigenfunctions for $n \ge 1$ have zeros), the latter provides

$$|\mathbf{e}_n(\theta)| \le \left| \int_{[\theta_0,\theta)} \mathbf{e}_n'(\theta) d\theta \right| \le \int_0^{2\pi} |\mathbf{e}_n'(\theta)| d\theta \le 8\pi^2\bar{A}^2 + 4\pi\lambda_n^2 + 4\pi^2 \quad \forall \theta \in S^1.$$

By arbitrariness of $n \ge 1$, we get the claim. □

The study of the convergence of the transitional dynamics to a stationary state gives the following claim corresponding to Theorem 3.4.

Proposition A.7. *Let* (8) *and* (15) *hold. Let $\mathbf{c}^*$ be the control defined in* (40) *and assume that $\mathbf{c}^* \in \mathcal{A}(\mathbf{K}_0)$. Define the detrended optimal path*

$$\mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(t) := e^{-gt}\mathbf{K}^{\mathbf{K}_0,\mathbf{c}^*}(t) \quad \forall t \ge 0,$$

and set $\beta_n := \langle \mathbf{e}_n, \boldsymbol{\beta}^{-\frac{1}{\sigma}}\mathbf{N}\rangle$ for each $n \ge 1$. Then

$$\mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(t)(\theta) \overset{t\to\infty}{\longrightarrow} \langle \mathbf{K}_0, \boldsymbol{\beta}\rangle \left(\alpha_0^{-1}\mathbf{e}_0(\theta) + \sum_{n\ge1} \frac{\beta_n}{\lambda_n - g}\mathbf{e}_n(\theta) \right) \quad \textit{uniformly in } \theta \in S^1.$$

Proof. As $\mathbf{K}^{\mathbf{K}_0,\mathbf{c}^*}(\cdot)$ is a weak solution of (36), $\mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(\cdot)$ is a weak solution of

$$\begin{cases} \mathbf{K}'(t) = \mathcal{L}\mathbf{K}(t) - g\mathbf{K}(t) - \langle \boldsymbol{\beta}, \mathbf{K}(t)\rangle \boldsymbol{\beta}^{-\frac{1}{\sigma}}\mathbf{N} \\ \mathbf{K}(0) = \mathbf{K}_0 \in \mathcal{H}_0^+, \end{cases}$$

i.e., for every $\boldsymbol{\varphi} \in D(\mathcal{L})$,

$$\begin{cases} \frac{d}{dt}\langle \mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(t), \boldsymbol{\varphi}\rangle = \langle \mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(t), (\mathcal{L}-g)\boldsymbol{\varphi}\rangle - \langle \boldsymbol{\beta}, \mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(t)\rangle \langle \boldsymbol{\varphi}, \boldsymbol{\beta}^{-\frac{1}{\sigma}}\mathbf{N}\rangle \\ \mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(0) = \mathbf{K}_0 \in \mathcal{H}_0^+. \end{cases} \tag{44}$$

As already recalled in Section 3, the sequence of eigenfunctions $\{\mathbf{e}_n\}_{n\ge0}$ is an orthonormal basis of $L^2(S^1)$, so we have the Fourier series expansion

$$\mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(t) = \sum_{n\ge0} K_{g,n}(t)\mathbf{e}_n, \tag{45}$$

where

$$K_{g,n}(t) := \langle \mathbf{K}_g^{\mathbf{K}_0,\mathbf{c}^*}(t), \mathbf{e}_n\rangle \quad \forall n \ge 0.$$

We compute now the Fourier coefficients $K_{g,n}(t)$.

- When $n = 0$, we already know from (39)

$$K_{g,0}(\cdot) \equiv \langle \mathbf{K}_0, \mathbf{e}_0 \rangle = \alpha_0^{-1} \langle \mathbf{K}_0, \boldsymbol{\beta} \rangle.$$

- When $n \geq 1$, we have, taking $\boldsymbol{\varphi} = \mathbf{e}_n$ in (44) and considering (39),

$$K'_{g,n}(t) = (\lambda_n - g) K_{g,n}(t) - \langle \mathbf{K}_0, \boldsymbol{\beta} \rangle \beta_n.$$

Hence we can explicitly express the Fourier coefficients for $n \geq 1$ as:

$$K_{g,n}(t) = \langle \mathbf{K}_0, \mathbf{e}_n \rangle e^{(\lambda_n - g)t} + \langle \mathbf{K}_0, \boldsymbol{\beta} \rangle \frac{\beta_n}{\lambda_n - g} (1 - e^{(\lambda_n - g)t}).$$

We write, taking into account that $\|\mathbf{e}_n\| = 1$ for each $n \geq 0$ and Lemma A.6,

$$\sup_{\theta \in \mathcal{S}^1} \left| \mathbf{K}_g^{\mathbf{K}_0, \mathbf{c}^*}(t)(\theta) - \langle \mathbf{K}_0, \boldsymbol{\beta} \rangle \left(\alpha_0^{-1} \mathbf{e}_0(\theta) + \sum_{n \geq 1} \frac{\beta_n}{\lambda_n - g} \mathbf{e}_n(\theta) \right) \right|$$

$$\leq \sup_{\theta \in \mathcal{S}^1} \sum_{n \geq 1} \left(|\langle \mathbf{K}_0, \mathbf{e}_n \rangle| \, e^{(\lambda_n - g)t} + |\langle \mathbf{K}_0, \boldsymbol{\beta} \rangle| \, \frac{\beta_n}{|\lambda_n - g|} e^{(\lambda_n - g)t} \right) |\mathbf{e}_n(\theta)|$$

$$\leq C_0 \|\mathbf{K}_0\| \sum_{n \geq 1} e^{(\lambda_n - g)t} (1 + \lambda_n^2) + C_0 \langle \mathbf{K}_0, \boldsymbol{\beta} \rangle \|\boldsymbol{\beta}^{-\frac{1}{\sigma}} \mathbf{N}\| \sum_{n \geq 1} \frac{1 + \lambda_n^2}{|\lambda_n - g|} e^{(\lambda_n - g)t}.$$

Note that, by (15), we have $\lambda_n \leq \lambda_1 < g$ for every $n \geq 1$. Then, the coefficients of the series converge in decreasing way to 0 as $t \to \infty$. Then one can conclude by dominated convergence of series, as the coefficients of the series above are nonnegative and decreasing in t, and the series above taken with $t = 1$ are convergent. □

Remark A.8. *The following estimates on λ_0 can be obtained from its representation provided in Section 2.10 of Brown et al. (2013):*

$$\frac{1}{2\pi} \int_0^{2\pi} A(\theta) d\theta \ \leq \ \lambda_0 \ \leq \ \sup_{S^1} |A|. \tag{46}$$

The lower bound in particular assures, given the positivity of $A(\cdot)$, the positivity of λ_0. The upper bound is useful to check (8),

Theorem 2.9.3 of Brown et al. (2013) also gives the following estimates for the second eigenvalue:

$$\lambda_1 \ \leq \sup_{S^1} A - 1,$$

useful to check (15).