Optimal Control of Semilinear Graphon Systems*

Martin T. Köhler, Artemi Makarow, and Christian Kirches

***Abstract*— Controlling the dynamics of large-scale networks is essential for a macroscopic reduction of overall consumption and losses in the context of energy supply, finance, logistics, and mobility. We investigate the optimal control of semilinear dynamical systems on asymptotically infinite networks, using the notion of graphons. Graphons represent a limit object of a converging graph sequence and serve as a generalization of graphs. This enables the systematic analysis of converging graph sequences as the number of vertices tends to infinity. Based on the theory of graphons and optimal control, we derive novel convergence results for the states, adjoints, and controls for dynamical systems on converging graph sequences. In other words, we can approximate the optimal system behavior on the infinite-dimensional limit object up to arbitrary precision by solving an optimal control problem subject to dynamics on a dense but finite graph, which is sampled from the limit object of the converging graph sequence. Numerical experiments support our theoretical results and verify our derived convergence rates.**

I. Introduction

In our daily lives, we encounter networks in the field of energy supply, logistics, mobility, telecommunication, but also in the interconnection of neurons in our brain, social relationships, food webs, power grids, the internet of things and lately neural networks powering AI. Oftentimes such networks are very large and dense, hindering efficient computation for dynamical systems and optimization.

In order to study large networks from a theoretical standpoint, the notion of graphons has been introduced as a generalization of graphs [1]–[6]. The theory allows the adjacency matrix of a graph to be mapped to a piecewise constant function on a continuous space, specifically the unit square $[0, 1] \times [0, 1]$. A distance measure called the *cut distance* is constructed to enable the comparison of graphs with different sizes. As a result, using the cut distance, we can analyze converging sequences of dense graphs and more specifically its limit object: A graphon. This generalization strongly relies on functional analysis and therefore enables the use of continuous optimization of dynamical systems in Banach spaces.

In [7], [8], the authors investigate the evolution of dynamical systems on graphons, without implementing optimal control. They show that the coupled dynamics on the elements of a converging graph sequence converge to those of the limit system described by a graphon. This convergence has a strong impact on the evolution of stable and unstable regions in the state domain [7]. However, the discussion is restricted to $\{0, 1\}$ valued graphons.

In [8], the authors highlight the challenges of rigorously defining symmetries compared to finite graphs and present symmetry groups of several graphons. Most interestingly, the limit object can have more symmetries than the finite graph of its converging sequence.

Gao and Caines investigate the connection between dynamical systems on a graphon and optimal control in [9], [10]. They establish the exact and approximate controllability as well as convergence properties for linear time-invariant graphon systems. Furthermore, they devise a method for controlling large network systems and integrate their results with the minimum energy state-to-state and linear quadratic regulator problem.

All these research directions in the context of graphons find applications in graphon mean field games, which is the study of decision-making in dynamic games of infinite agents on an asymptotically infinite network [11], [12]. This problem formulation is especially useful in modeling the interactions and controls in social, financial and communication networks of large scale.

In this paper, we consider a graphon dynamical system in an optimal control context, where we aim to minimize a cost function subject to state constraints given by a semilinear differential equation on a graphon. We extend the theory from the control of linear to semilinear graphon systems and present novel convergence results. The key result is that we can approximate the optimal evolution of the control trajectory of the dynamical limit system by using a finite graph system with local nonlinear dynamics, stemming from a converging graph sequence.

Contributions: Section II introduces the preliminaries required to build on the graphon theory.

In Section III, we first examine convergence properties in the absence of optimal control. Here, we first prove that the states of the semilinear graphon system converge in the L^2-norm under mild assumptions as we refine the graph. Additionally, we are able to provide a convergence rate in relation to the size of the graph.

In Section IV, we introduce the corresponding optimal control problem and show that the adjoint variables also converge in the L^2 norm. The Hamiltonian function of our optimal control problem bridges the relationship between the state, adjoint, and control variables. If we combine both results, we notice that the control variables must also

*Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This work is supported by ERC grant SCARCE, 101087662.

Martin T. Köhler, Artemi Makarow, and Christian Kirches are with Institute for Mathematical Optimization, Technische Universität Braunschweig, 38106 Braunschweig, Germany.
{martin.koehler1,artemi.makarow,c.kirches}@tu-bs.de

converge. In other words, if we solve the optimal control problem on an approximate finite graph with local nonlinear dynamics, we approximate the optimal controls of the limit system on the asymptotically infinite network. Finally, we substantiate our claims in Section V with numerical examples and viszualize convergence rates.

Notation: In this paper, we use y_* or A_* to denote the limit object of a sequence (y_N) or (A_N), where N denotes the number of vertices in a finite graph. In contrast, y^* refers to the optimal value of the associated optimization problem and A^* to the adjoint of the linear operator A. Additionally, boldface letters, such as $\mathbf{a}$, $\mathbf{A}$ can signify vectors, matrices or their elements. Furthermore, we write $\dot{y}$ for a derivative with respect to the variable t, meaning time, and D_y for the Frechét derivative with respect to the argument y. The form $\langle \cdot, \cdot \rangle$ denotes the L^2 inner product. Lastly, *a.e.* means *almost everywhere*, indicating that a property holds for all elements in a set X except on a subset of X with measure zero.

II. Preliminaries on Graphons

We consider the following graphon differential equation:

$$\dot{y}(t) = Ay(t) + Bu(t) + f(t, y(t), u(t)), \tag{1}$$

where the semilinear dynamical system consists of linear graphon operators $A, B : L^2([0,1]) \to L^2([0,1])$ and a nonlinear term $f : \mathbb{R} \times L^2([0,1]) \times L^2([0,1]) \to L^2([0,1])$. The state y and control variables u are elements of $L^2([0,1] \times [0, t_\mathrm{f}])$, but for simplicity and focus on the time evolution, we omit the spatial variable and write $y(t)$ and $u(t)$ instead of $y(\cdot, t)$ and $u(\cdot, t)$.

A. From Graphs to Graphons

Our analysis focuses on graphons, which are limits of graph sequences. This section provides an overview of the theory of graphons.

Let $G = (V, E)$ be an undirected graph with a set V of N vertices and a set of edges $E \subset V \times V$, representing a network. We restrict our analysis to dense networks, meaning the number of edges is in the order of $\mathcal{O}(N^2)$.
Let $\mathcal{W}$ denote the space of all bounded symmetric Lebesgue measurable functions $W : [0,1]^2 \to \mathbb{R}$. Elements of $\mathcal{W}$ are called kernels.

Definition 2.1 (Graphon, see [3]): Let $\mathcal{W}_0$ denote the set of all kernels such that $W : [0,1]^2 \to [0,1]$. Additionally, let the set of all $W : [0,1]^2 \to [-1,1]$ be denoted by $\mathcal{W}_1$. An element $W \in \mathcal{W}_1$ is called a graphon.

Depending on the literature, either sets $\mathcal{W}_0$ and $\mathcal{W}_1$ may be called graphons and in rare cases even $\mathcal{W}$. Unless stated otherwise, we refer to the elements of $\mathcal{W}_1$ as graphons.

Graphons generalize the notion of a symmetric graph in the following sense.

Definition 2.2 (Step function, see [9]): Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be an adjacency matrix representing a graph $G = (V, E)$ with N vertices and let there be a uniform partition $P = \{P_1, \dots, P_N\}$ of $[0,1]$ such that $P_k = \left(\frac{k-1}{N}, \frac{k}{N}\right)$ for $k = \{1, \dots, N\}$. A step function $W_G \in \mathcal{W}$ for a graph G is given by

$$W_G(x, y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{1}_{P_j}(x) \mathbb{1}_{P_i}(y) \mathbf{a}_{ij}, \tag{2}$$

where $x, y \in [0,1]$. The indicator function $\mathbb{1}_X(x) = 1$ if $x \in X$, otherwise $\mathbb{1}_X(x) = 0$. Here, $\mathbf{a}_{ij}$ denotes the weight of the edge $(i,j) \in E$ within the adjacency matrix $\mathbf{A}$.

The step function W_G allows for any undirected graph G of arbitrary size, weighted or unweighted, to be represented as a piecewise constant function on the unit square.

B. Convergence of Graph Sequences

The distance between two graphons is based on a norm on the space $\mathcal{W}$.

Definition 2.3 (Cut Norm, see [3]): Let $W \in \mathcal{W}$ be a kernel. The *cut norm* is defined by

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y) \, \mathrm{d}x \, \mathrm{d}y \right| \tag{3}$$

taken over all measurable subsets $S, T \subseteq [0,1]$.
Computing the cut norm is a difficult task. Therefore, the authors in [13] devise an approximation algorithm for its discretized counterpart.

For elements $W \in \mathcal{W}_1$ we can show the following relation:

$$\|W\|_\square \leq \|W\|_1 \leq \|W\|_2 \leq \|W\|_\infty \leq 1. \tag{4}$$

This relation is practical for bounding terms that include a graphon.

For the distance between two graphons, we consider all possible relabelings. An *unlabeled graphon* is a graphon $W^\phi(x,y) := W(\phi(x), \phi(y))$ up to all measure preserving bijections $\phi \in S_{[0,1]}$ on $[0,1]$. We rely on the following distance measure.

Definition 2.4 (Cut Distance, see [3]): For $W, V \in \mathcal{W}$, the *cut distance* is defined as

$$\delta_\square(W, V) = \inf_{\phi \in S_{[0,1]}} \left\| W^\phi - V \right\|_\square . \tag{5}$$

With the cut distance, we minimize the maximum discrepancy between two graphons over all possible relabelings. It is important to mention that for two distinct kernels $W, V \in \mathcal{W}$, the cut distance (5) can be zero. We therefore identify W and V, denoted by $W \sim V$, if and only if $\delta_\square(W, V) = 0$. Under this equivalence relation, the set $\widetilde{\mathcal{W}}$, endowed with the cut distance (5), forms a metric space.

We say that a sequence of graphs (G_n) is convergent if it is a Cauchy sequence under the distance measure $\delta_\square$. If such a sequence is Cauchy, then it always converges to some $W \in \mathcal{W}$, i.e.,

$$\delta_\square(W_{G_n}, W) \to 0 \text{ for } n \to \infty. \tag{6}$$

Figure 1 visualizes the convergence of a graph sequence in $\mathcal{W}_0$. Lovász and Szegedy [2] prove one of the most important results in the theory of graphons.

Theorem 2.1: The space $(\widetilde{\mathcal{W}}, \delta_\square)$ is compact. See [2], [3].

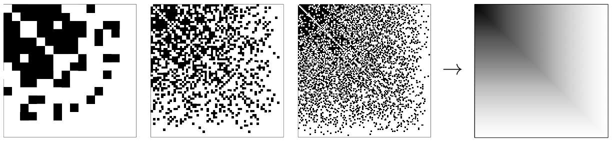

Fig. 1. A sequence of graphs represented as step functions converging to the uniform attachment graphon $W(x,y) = 1 - \max(x,y)$.

This means that the space of graphons $\widetilde{\mathcal{W}}$ is a completion of the space of finite graphs with the cut norm. One can also argue in the weak* topology that sequences of graphons converge in a weak* sense [14], [15], although the convergence in the cut distance has favorable properties such as being stronger compared to weak* and also conserving the subgraph density, which is useful for combinatorial analysis [15].

As graphs and graphons are in the same space $\widetilde{\mathcal{W}}$, shown in Theorem 2.1, the elements of converging graph sequences inherit certain properties of the limit object such as densities, degree distributions, clusterings or, as we show later, network dynamics.

C. Graphon Operator

In order to model the evolution in a network, we need an operator that governs the change in time. A powerful interpretation arises when we view a graphon as a linear operator that maps $L^2([0,1])$ functions to the space of $L^2([0,1])$.

Definition 2.5 (Kernel Operator, see [3]): Let $W \in \mathcal{W}$ be a kernel. For any $f \in L^2([0,1])$, the *kernel operator* is defined by

$$(Wf)(x) = \int_0^1 W(x,y) f(y) \, \mathrm{d}y, \quad x \in [0,1]. \tag{7}$$

For simplicity of notation, we write Wf.

Between two nodes $x, y \in [0,1]$, a graphon encodes the weight of the relationship $W(x,y)$. This means for $x \in [0,1]$ the operation Wf accumulates contributions $f(y)$ of all nodes $y \in [0,1]$ to x based on the weight of the relationship $W(x,y)$.

Any $W \in \mathcal{W}_1$ in (7) is a linear self-adjoint operator, that is $W = W^*$. This property will be a key element in one of our proofs later on. We define the operator norm to be

$$\|W\|_{\text{op}} := \sup_{0 \neq f \in L^2([0,1])} \frac{\|Wf\|_2}{\|f\|_2}, \tag{8}$$

which is shown to be equivalent to the cut norm [16]:

$$\|W\|_{\square} \leq \|W\|_{\text{op}} \leq 4 \|W\|_{\square}. \tag{9}$$

Therefore, the convergence of graph sequences in the cut distance (5) is equivalent to the convergence in the operator norm (8).

Now, in our dynamical system (1), we assume the linear operators A and B to be defined according to (7). The graphons A, B can be the limit object of the graph sequence or step functions according to Definition 2.2, meaning we can represent the finite graphs of the graph sequence or its limit graphon. In this case, the notion of graphons allows for a generic notation of a network dynamical systems, since all objects exist in the same space and use the same mathematical foundation.

D. Integration of Semigroup Theory

Since the operators A, B in our dynamical system (1) are bounded linear operators, we rely on semigroup theory [17].

Definition 2.6 (Strongly Continuous Semigroup, see [17]): Let X be a Banach space. If a family of bounded linear operators $T(t)$, $t \in [0, \infty)$, where $T(t) : X \to X$, fulfill

- $T(0) = I$, where I is the identity operator on X,
- $T(t)T(s) = T(t+s)$ for all $s, t \geq 0$,
- $\lim_{t \to 0} T(t)x = x$ for all $x \in X$,

it is called a *strongly continuous semigroup*.

An identity element I for graphons can be defined, but it is not considered to be a graphon itself [9]. Let $A \in \mathcal{W}_1$ be a graphon. We know from semigroup theory [17] that A is the infinitesimal generator of the strongly continuous semigroup

$$T_A(t) := e^{At} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}. \tag{10}$$

For the following Cauchy problem

$$\begin{cases} \dot{y}(t) = Ay(t), & t \in (0, t_{\mathrm{f}}], \\ y(0) = y_0, \end{cases} \tag{11}$$

where $y_0 \in L^2([0,1])$ is the initial value, the solution satisfies

$$y(t) = e^{At} y_0, \quad t \in [0, t_{\mathrm{f}}], \tag{12}$$

and is unique and continuously differentiable in time [17]. Similarly, for our dynamical system (1), the (mild) solution must satisfy

$$y(t) = e^{At} y_0 + \int_0^t e^{A(t-s)} \big(Bu(s) + f(s, y(s), u(s)) \big) \, \mathrm{d}s, \tag{13}$$

for $t \in [0, t_{\mathrm{f}}]$, given the functions u and f are continuous in t and uniformly Lipschitz-continuous on $L^2([0,1])$ [17]. The use of semigroup theory gives us an explicit form for the solution of the network dynamical system, thus it greatly simplifies its analysis.

III. State Convergence of Semilinear Systems

We first want to analyze the convergence behavior of the graphons system with respect to the state $y(t)$ and in the absence of optimal control. To do so, we impose mild assumptions under which we will consider our analysis.

Let $x_a(t) := (y_a(t), u_a(t))$ and let the following assumptions hold for all $y_a(t) \in L^2([0,1])$, $u_a(t) \in U_{\text{ad}}$, $t \in [0, t_{\mathrm{f}}]$.

(A1) The nonlinear function f is Lipschitz continuous with $L_f > 0$ and satisfies:

$$\|f(t, x_1(t)) - f(t, x_2(t))\|_2 \leq L_f \|x_1(t) - x_2(t)\|_2. \tag{14}$$

(A2) For a solution tuple $x(t) := (y(t), u(t))$, there exists a constant M such that

$$\sup_{t\in[0,t_{\mathrm{f}}]} \|f(t, x(t))\|_2 \le M \tag{15}$$

holds for all $y(t) \in L^2([0,1])$, $u(t) \in U_{\mathrm{ad}}$, $t \in [0, t_{\mathrm{f}}]$.

Moreover, for the graphons governing the dynamical system (1), we assume the following.

(A3) There exist sequences of finite graphs with N vertices represented as step functions denoted by (A_N), $(B_N) \subset \mathcal{W}_1$ that converge to their respective limit graphons A_*, $B_* \in \mathcal{W}_1$ in the operator norm as N grows to infinity, i.e.,

$$\begin{aligned}\lim_{N\to\infty} \|A_N - A_*\|_{\mathrm{op}} = 0,\\ \lim_{N\to\infty} \|B_N - B_*\|_{\mathrm{op}} = 0.\end{aligned} \tag{16}$$

A. Convergence of the States Variables

We can approximate a graphon $A_* \in \mathcal{W}_1$ up to arbitrary precision using a step function $A_N \in \mathcal{W}_1$ that describes a graph with N vertices. We now demonstrate that the difference of the state variables y_* and y_N of the dynamical systems

$$\dot{y}_*(t) = A_*\, y_*(t) + B_* u(t) + f(t, y_*(t), u(t)), \tag{17a}$$
$$\dot{y}_N(t) = A_N\, y_N(t) + B_N u(t) + f(t, y_N(t), u(t)), \tag{17b}$$
$$y_*(0) = y_0 \in L^2([0,1]), \tag{17c}$$
$$y_N(0) = y_{0,N} \in L^2([0,1]), \tag{17d}$$

converge in the L^2 norm for a given control function $u \in L^2([0,1] \times [0, t_{\mathrm{f}}])$. The initial value $y_{0,N}$ is constructed such that for $N \to \infty$, the piecewise constant function $y_{0,N}$ converges to y_0 in $L^2([0,1])$. A related statement and proof for the nonlinear heat equation on graph limits has been shown in [7], where only $\{0,1\}$ valued graphons are taken into consideration. However, we take a different approach using semigroup theory, that does not restrict the set of graphons.

The convergence statement means we can approximate the graphon dynamical system formulation using a finite graph description. We now show that the state variables of the finite system converge to that of the graphon dynamical system.

Theorem 3.1 (State Convergence): Let $u \in L^2([0,1] \times [0, t_{\mathrm{f}}])$ with $U := \sup_{t\in[0,t_{\mathrm{f}}]} \|u(t)\|_2$. Suppose Assumptions (A1), (A2), and (A3) hold for the dynamical system (17) with

$$\lim_{N\to\infty} \|y_{0,N} - y_0\|_2 = 0, \tag{18}$$

then

$$\lim_{N\to\infty} \sup_{t\in[0,t_{\mathrm{f}}]} \|y_N(t) - y_*(t)\|_2 = 0. \tag{19}$$

Before we prove this result, we consider the linear differential equations governed by their respective graphon operators.

Lemma 3.1 (Gao and Caines [9]): Let $A_N, A_* \in \mathcal{W}_1$ be two graphons. For any $y \in L^2([0,1])$, the following holds:

$$\left\|e^{A_N t} y - e^{A_* t} y\right\|_2 \le t e^t \|A_N - A_*\|_{\mathrm{op}} \|y\|_2 . \tag{20}$$

Furthermore, if a sequence $(A_N) \subset \mathcal{W}_1$ converges to A_* in the operator norm, i.e.,

$$\lim_{N\to\infty} \|A_N - A_*\|_{\mathrm{op}} = 0, \tag{21}$$

then

$$\lim_{N\to\infty} \sup_{t\in[0,t_{\mathrm{f}}]} \left\|e^{A_N t} y - e^{A_* t} y\right\|_2 = 0 \tag{22}$$

for any $y \in L^2([0,1])$ and any $t_{\mathrm{f}} > 0$.

Proof: The proof can be found in Gao [9, Thm. 4]. ■

For our purposes, the linear system governed by A_N has a piecewise constant initial value $y_{0,N}$, therefore we must consider the following corollary.

Corollary 3.1: Let $(A_N) \subset \mathcal{W}_1$ be a sequence converging to $A_* \in \mathcal{W}_1$ in the operator norm, i.e.,

$$\lim_{N\to\infty} \|A_N - A_*\|_{\mathrm{op}} = 0 \tag{23}$$

and let $y_0, y_{0,N} \in L^2([0,1])$ with

$$\lim_{N\to\infty} \|y_{0,N} - y_0\|_2 = 0, \tag{24}$$

then

$$\lim_{N\to\infty} \sup_{t\in[0,t_{\mathrm{f}}]} \left\|e^{A_N t} y_{0,N} - e^{A_* t} y_0\right\|_2 = 0 \tag{25}$$

for any $t_{\mathrm{f}} > 0$.

Proof: By adding $0 = e^{A_* t} y_{0,N} - e^{A_* t} y_{0,N}$ to (25) and through reformulation, we obtain

$$\begin{aligned} E_1 &:= \left\|e^{A_N t} y_{0,N} - e^{A_* t} y_0 + e^{A_* t} y_{0,N} - e^{A_* t} y_{0,N}\right\|_2 \\ &\le \underbrace{\left\|e^{A_N t} y_{0,N} - e^{A_* t} y_{0,N}\right\|_2}_{=:e_1} + \underbrace{\left\|e^{A_* t} (y_{0,N} - y_0)\right\|_2}_{=:e_2} . \end{aligned} \tag{26}$$

The term e_1 is equivalent to (20) in Lemma 3.1 and e_2 tends to zero as $N \to \infty$ because of (24) and $\left\|e^{A_* t}\right\|_{\mathrm{op}} \le e^{t\|A_*\|_{\mathrm{op}}} \le e^{t_{\mathrm{f}}}$ on a finite time interval $[0, t_{\mathrm{f}}]$. ■

The convergence of the state variables for linear graphon systems has been established [9]. We now aim to prove the convergence of the state variables in semilinear systems, thereby extending the existing theory of graphon dynamical systems control.

The inequality (20) and the Grönwall-Bellman inequality presented in [18], [19] are essential to the proof of Theorem 3.1.

For a shorter notation, let

$$\begin{aligned} F(y_N(t)) &:= f(t, y_N(t), u(t)), \quad t \in [0, t_{\mathrm{f}}], \\ F(y_*(t)) &:= f(t, y_*(t), u(t)), \quad t \in [0, t_{\mathrm{f}}], \\ L(y_N(t)) &:= \ell(t, y_N(t), u(t)), \quad t \in [0, t_{\mathrm{f}}], \\ L(y_*(t)) &:= \ell(t, y_*(t), u(t)), \quad t \in [0, t_{\mathrm{f}}]. \end{aligned}$$

Proof: [Proof of Theorem 3.1] We show that $\|y_N - y_*\|_2$ is smaller than some finite bound and converges to zero as N tends to infinity. As $y_N(t)$ and $y_*(t)$ are solutions to the semilinear dynamical system (1) and satisfy (13) for $t \in [0, t_\mathrm{f}]$, reordering and using the triangle inequality leads to

$$\|y_N(t) - y_*(t)\|_2 \leq \underbrace{\left\|e^{A_N t} y_{0,N} - e^{A_* t} y_0\right\|_2}_{=:E_1} + \underbrace{\left(\left\|\int_0^t e^{A_N(t-s)}\big(B_N u(s) + F(y_N(s))\big) - e^{A_*(t-s)}\big(B_* u(s) + F(y_*(s))\big)\,\mathrm{d}s\right\|_2\right)}_{=:E_2}.$$

The first term E_1 is equivalent to (26), which rests upon Corollary 3.1, and with $\left\|e^{A_* t}\right\|_\mathrm{op} \leq e^t$ it follows that the term is bounded by

$$E_1 \leq te^t \|A_N - A_*\|_\mathrm{op} \|y_0\|_2 + e^t \|y_{0,N} - y_0\|_2 .$$

We now focus on the second term E_2. Temporarily, we define

$$G_N(s) := B_N u(s) + F(y_N(s)),$$
$$G_*(s) := B_* u(s) + F(y_*(s)).$$

By adding $0 = e^{A_*(t-s)} G_N(s) - e^{A_*(t-s)} G_N(s)$ inside the integral, subsequently we get

$$\begin{aligned} E_2 &= \left\|\int_0^t e^{A_N(t-s)} G_N(s) - e^{A_*(t-s)} G_*(s) + e^{A_*(t-s)} G_N(s) - e^{A_*(t-s)} G_N(s)\,\mathrm{d}s\right\|_2 \\ &\leq \underbrace{\left\|\int_0^t \left(e^{A_N(t-s)} - e^{A_*(t-s)}\right) G_N(s)\,\mathrm{d}s\right\|_2}_{=:D_1} + \underbrace{\left\|\int_0^t e^{A_*(t-s)}\left(G_N(s) - G_*(s)\right)\,\mathrm{d}s\right\|_2}_{=:D_2}. \end{aligned}$$

The estimate consists of two terms, which we will analyze separately. Starting with the first term D_1, the structure inside the integral is in essence of the same to (20), where instead of a fixed $L^2([0,1])$ function, we have G_N dependent on the variable of integration s. Looking at G_N more closely, we observe that B_N is bounded in the operator norm by the value 1, u is bounded by U, and $F(y_N(s))$ is bounded by M by Assumption (A2), hence we have that $\|G_N(s)\|_2 \leq U + M$ for all $s \in [0, t]$. In total, we obtain:

$$\begin{aligned} D_1 &\leq \left|\int_0^t (t-s)e^{t-s} \|A_N - A_*\|_\mathrm{op} \|G_N(s)\|_2 \,\mathrm{d}s\right| \\ &\leq \int_0^t (t-s)e^{t-s}\,\mathrm{d}s\, \|A_N - A_*\|_\mathrm{op} (U+M) \\ &\leq te^t \|A_N - A_*\|_\mathrm{op} (U+M). \end{aligned} \tag{27}$$

Here, we overestimate the integral $\int_0^t (t-s)e^{t-s}\,\mathrm{d}s$ by a simpler expression te^t for $t \geq 0$.

Lastly, we derive an estimate for the second term D_2. For this step, we expand the abbreviation and use Assumption (A1), where F is locally Lipschitz continuous. We thus have that D_2 is smaller or equal to

$$\begin{aligned} &\left\|\int_0^t e^{A_* t}\big(\,(B_N - B_*)\,u(s) + F(y_N(s)) - F(y_*(s))\big)\,\mathrm{d}s\right\|_2 \\ &\leq \left\|e^t \int_0^t (B_N - B_*)u(s) + L_f \|y_N(s) - y_*(s)\|_2 \,\mathrm{d}s\right\|_2 \\ &\leq te^t \|B_N - B_*\|_\mathrm{op} U + e^t L_f \int_0^t \|y_N(s) - y_*(s)\|_2 \,\mathrm{d}s. \end{aligned} \tag{28}$$

Here, the structure of the integral suggests the use of the Grönwall-Bellman inequality to bound $\|y_N(s) - y_*(s)\|_2$. For simplicity, we define $\alpha_N(t)$ to contain all the estimates independent of y_N and y_*:

$$\begin{aligned} \alpha_N(t) := {} & te^t \|A_N - A_*\|_\mathrm{op} \|y_0\|_2 \\ & + e^t \|y_{0,N} - y_0\|_2 \\ & + te^t \|A_N - A_*\|_\mathrm{op} (U+M) \\ & + te^t \|B_N - B_*\|_\mathrm{op} U. \end{aligned} \tag{29}$$

Furthermore, we define $E := e^{t_\mathrm{f}}$. We can now summarize the previous steps to

$$\|y_N(t) - y_*(t)\|_2 \leq \alpha_N(t) + EL_f \int_0^t \|y_N(t) - y_*(t)\|_2 \,\mathrm{d}s. \tag{30}$$

Note that $\alpha_N(t)$ is a positive non-decreasing function for $t \in [0, t_\mathrm{f}]$ and for all N, hence applying the Grönwall-Bellman inequality according to [18], [19] leads us to

$$\|y_N(t) - y_*(t)\|_2 \leq \alpha_N(t)\, e^{tEL_f} \quad t \in [0, t_\mathrm{f}]. \tag{31}$$

This inequality shows that the upper bound is no longer dependent on $y_N(t)$ and $y_*(t)$. The term $\sup_{t\in[0,t_\mathrm{f}]} \alpha_N(t)$ tends to zero for $N \to \infty$ by Assumption (A3) and (18), thus we have proven the original statement. ∎

Corollary 3.2 (State Convergence Rate): Let A_N and B_N converge to A_* and B_* respectively in the operator norm with the convergence rate of $\mathcal{O}(N^{-1/2})$ with high probability [3]. Then

$$\sup_{t\in[0,t_\mathrm{f}]} \|y_N(t) - y_*(t)\|_2 \in \mathcal{O}(N^{-1/2}). \tag{32}$$

Proof: The term α_N is in large parts dependent on the differences $\|A_N - A_*\|_\mathrm{op}$ and $\|B_N - B_*\|_\mathrm{op}$. Therefore, $\alpha_N(t) \in \mathcal{O}(N^{-1/2})$ and subsequently $\|y_N(t) - y_*(t)\| \in \mathcal{O}(N^{-1/2})$ for all $t \in [0, t_\mathrm{f}]$. ∎

IV. Optimal Control with Graphons

While we assume an arbitrary control function $u \in L^2([0,1] \times [0, t_\mathrm{f}])$ in (17) for proving state convergence of semilinear graphon systems, we now want to use the optimal control functions. We aim to explore whether and under which conditions the optimal control function on a finite graph system converges to the optimal control function of the limit system. Therefore, we embed the semilinear graphon system in (1) into the following optimal control problem:

$$\min_{y,\,u} \quad J(y,u) = \int_0^{t_\mathrm{f}} \ell(t, y(t), u(t))\,\mathrm{d}t \tag{33a}$$
$$\text{s.t.} \quad \dot{y}(t) = Ay(t) + Bu(t) + f(t, y(t), u(t)), \tag{33b}$$
$$y(0) = y_0 \in L^2([0,1]), \tag{33c}$$
$$u(t) \in U_\mathrm{ad} \subset L^2([0,1]), \tag{33d}$$
$$t \in [0, t_\mathrm{f}] \text{ a.e.}. \tag{33e}$$

The running cost and nonlinear term are continuous mappings $\ell : \mathbb{R} \times L^2([0,1]) \times L^2([0,1]) \to \mathbb{R}$ and $f : \mathbb{R} \times L^2([0,1]) \times L^2([0,1]) \to L^2([0,1])$, respectively. To prove convergence of the adjoint and control variables in the following, we require further mild assumptions about the nonlinear functions.

(A4) The mapping $x(t) \mapsto \ell(t, x(t))$ is Lipschitz continuous with $x(t) = (y(t), u(t))$ and $L_\ell > 0$ and satisfies:

$$|\ell(t, x_1(t)) - \ell(t, x_2(t))| \leq L_\ell \left\| x_1(t) - x_2(t) \right\|_2 . \tag{34}$$

(A5) The Frechét derivatives of the running cost function ℓ and the nonlinear function f with respect to y are continuous and satisfy

$$\begin{aligned} &\left\| D_y \ell(t, y_1(t), u(t)) - D_y \ell(t, y_2(t), u(t)) \right\|_2 \\ &\quad \leq L_{D\ell} \left\| y_1(t) - y_2(t) \right\|_2 , \\ &\left\| D_y f(t, y_1(t), u(t)) - D_y f(t, y_2(t), u(t)) \right\|_\mathrm{op} \\ &\quad \leq L_{Df} \left\| y_1(t) - y_2(t) \right\|_2 , \end{aligned} \tag{35}$$

with $L_{D\ell}, L_{Df} > 0$.

A. Convergence of the Adjoint Variables

We now first want to analyze the convergence behavior of the adjoint variables in order to derive convergence claims for the control variables in a following step. The relationship of the state, control, and adjoint variables is stated in the Hamiltonian function

$$\begin{aligned} H(t, y(t), u(t), \lambda(t)) &:= \ell(t, y(t), u(t)) \\ &+ \left\langle \lambda(t), Ay(t) + Bu(t) + f(t, y(t), u(t)) \right\rangle. \end{aligned} \tag{36}$$

Similar to the state variables, the adjoint variables are in the space $L^2([0,1] \times [0, t_\mathrm{f}])$, but we omit writing the spacial variable and focus on its time component. Therefore, we represent it as $\lambda(t)$.

The well-known Pontryagin maximum principle is used to solve the Hamiltonian system and states the following.

Lemma 4.1 (See, e.g., [20]): If the tuple (y^*, u^*) is an optimal solution to (33), then there exists a $\lambda \in C(0, t_\mathrm{f}; L^2([0,1]))$ such that

$$\begin{aligned} -\dot{\lambda}(t) &= D_y \big(Ay^*(t) + Bu^*(t) + f(t, y^*(t), u^*(t)) \big)^* \lambda(t) \\ &\quad + D_y \ell(t, y^*(t), u^*(t)), \\ \lambda(t_\mathrm{f}) &= 0 \end{aligned} \tag{37}$$

hold and

$$H(t, y^*(t), u^*(t), \lambda(t)) = \max_{u \in U} H(t, y^*(t), u(t), \lambda(t)) \tag{38}$$

for $t \in [0, t_\mathrm{f}]$ a.e. with $U \subset L^2([0,1] \times [0, t_\mathrm{f}])$.

Proof: For the proof of the infinite-dimensional case, refer to [20] and the references therein. ■

We further show that also the difference of the adjoint variables λ_* and λ_N converge in the L^2 norm. With λ_N we denote the adjoint equation with step functions A_N and B_N. Additionally, with λ_* we denote the adjoint equation on the limit graphons A_* and B_*.

Theorem 4.1 (Adjoint Convergence): Suppose Assumptions (A1), (A3), (A4), and (A5) hold. Let $u \in L^2([0,1] \times [0, t_\mathrm{f}])$ with $U := \sup_{t \in [0, t_f]} \left\| u(t) \right\|_2$, then

$$\lim_{N \to \infty} \sup_{t \in [0, t_\mathrm{f}]} \left\| \lambda_N(t) - \lambda_*(t) \right\|_2 = 0, \tag{39}$$

with λ_N and λ_* denoting the solutions to (37).

Proof: The differential equations for the adjoint variables follow from Lemma 4.1:

$$\begin{aligned} -\dot{\lambda}_*(t) &= D_y \left(A_* y_*(t) + B_* u(t) + F(y_*(t)) \right)^* \lambda_*(t) \\ &\quad + D_y L(y_*(t)), \\ -\dot{\lambda}_N(t) &= D_y \left(A_N y_N(t) + B_N u(t) + F(y_N(t)) \right)^* \lambda_N(t) \\ &\quad + D_y L(y_N(t)). \end{aligned} \tag{40}$$

We now analyze the difference of the adjoint variables:

$$\begin{aligned} \dot{\lambda}_N(t) - \dot{\lambda}_*(t) &= A_* \lambda_*(t) - A_N \lambda_N(t) \\ &\quad + D_y \big(F(y_*(t)) \big)^* \lambda_*(t) - D_y \big(F(y_N(t)) \big)^* \lambda_N(t) \\ &\quad + D_y \big(L(y_*(t)) - L(y_N(t)) \big). \end{aligned} \tag{41}$$

In this step, we have used the property of graphon operators being self-adjoint and $D_y(B_a u(t)) \lambda_a(t) = 0$. To continue, we add $0 = A_* \lambda_N(t) - A_* \lambda_N(t)$ and $0 = D_y F(y_*(t))^* \lambda_N(t) - D_y F(y_*(t))^* \lambda_N(t)$ and thus by rearranging, we obtain:

$$\begin{aligned} \dot{\lambda}_N(t) - \dot{\lambda}_*(t) &= A_* (\lambda_*(t) - \lambda_N(t)) + (A_* - A_N) \lambda_N(t) \\ &\quad + D_y \big(F(y_*(t)) \big)^* (\lambda_*(t) - \lambda_N(t)) \\ &\quad + D_y \big(F(y_*(t)) - F(y_N(t)) \big)^* \lambda_N(t) \\ &\quad + D_y \big(L(y_*(t)) - L(y_N(t)) \big). \end{aligned} \tag{42}$$

Combining terms leads to the equation

$$\begin{aligned} \dot{\lambda}_N(t) - \dot{\lambda}_*(t) &= \big(A_* + D_y (F(y_*(t)))^* \big) (\lambda_*(t) - \lambda_N(t)) \\ &\quad + \big((A_* - A_N) + D_y (F(y_*(t)) - F(y_N(t)))^* \big) \lambda_N(t) \\ &\quad + D_y \big(L(y_*(t)) - L(y_N(t)) \big), \end{aligned} \tag{43}$$

which we integrate on both sides from t to t_f:

$$\begin{aligned} \lambda_N(t) - \lambda_*(t) &= \lambda_*(t_\mathrm{f}) - \lambda_N(t_\mathrm{f}) \\ &\quad + \int_t^{t_\mathrm{f}} \big(A_* + D_y F(y_*(s))^* \big) (\lambda_*(s) - \lambda_N(s)) \\ &\quad + \big((A_* - A_N) \\ &\quad + D_y (F(y_*(s)) - F(y_N(s)))^* \big) \lambda_N(s) \\ &\quad + D_y \big(L(y_*(s)) - L(y_N(s)) \big) \,\mathrm{d}s. \end{aligned} \tag{44}$$

Recall that the terminal values $\lambda_*(t_\mathrm{f})$ and $\lambda_N(t_\mathrm{f})$ are zero by Lemma 4.1. We now apply the triangle inequality:

$$\begin{aligned}\|\lambda_N(t)-\lambda_*(t)\|_2 \le \int_t^{t_\mathrm{f}} & \|A_* + D_y(F(y_*(s))^*\|_{\mathrm{op}} \\ & \cdot \|\lambda_*(s)-\lambda_N(s)\|_2 \\ & + \big(\|A_*-A_N\|_{\mathrm{op}} \\ & + \left\|D_y\left(F(y_*(s))-F(y_N(s))\right)^*\right\|_{\mathrm{op}}\big)\|\lambda_N(s)\|_2 \\ & + \|D_y\left(L(y_*(s))-L(y_N(s))\right)\|_2 \,\mathrm{d}s. \end{aligned} \tag{45}$$

Note that $\|A_*\|_{\mathrm{op}} \le 1$ and $\|D_yF\|_{\mathrm{op}}$ is bounded by the constant L_f by Assumption (A1). Additionally, by Assumption (A5), L and F are locally Lipschitz continuous with constants $L_{D\ell}$ and L_{Df}, respectively. Therefore, we have that:

$$\begin{aligned}\|\lambda_N(t)-\lambda_*(t)\|_2 &\le (1+L_f)\int_t^{t_\mathrm{f}} \|\lambda_N(s)-\lambda_*(s)\|_2 \,\mathrm{d}s \\ &+ \int_t^{t_\mathrm{f}} \big(\|A_*-A_N\|_{\mathrm{op}} \\ &\quad + L_{Df}\|y_*(s)-y_N(s)\|_2\big)\|\lambda_N(s)\|_2 \\ &\quad + L_{D\ell}\|y_*(s)-y_N(s)\|_2 \,\mathrm{d}s. \end{aligned} \tag{46}$$

For a short notation, let us define:

$$\begin{aligned}\beta_N(t) := \int_t^{t_\mathrm{f}} & \big(\|A_*-A_N\|_{\mathrm{op}} \\ & + L_{Df}\|y_*(s)-y_N(s)\|_2\big)\|\lambda_N(s)\|_2 \\ & + L_{D\ell}\|y_*(s)-y_N(s)\|_2 \,\mathrm{d}s. \end{aligned} \tag{47}$$

By using a backward-in-time variant of the Grönwall-Bellman inequality, where now the function β_N must be non-increasing, see, e.g., [18], [19], we obtain for all $t \in [0, t_\mathrm{f}]$:

$$\|\lambda_N(t)-\lambda_*(t)\|_2 \le \beta_N(t)e^{(1+L_f)(t_\mathrm{f}-t)}. \tag{48}$$

Because of Assumption (A3), subsequently

$$\lim_{N\to\infty}\sup_{t\in[0,t_\mathrm{f}]}\|y_N(t)-y_*(t)\|_2 = 0, \tag{49}$$

as shown in the proof of Theorem 3.1. This implies that

$$\lim_{N\to\infty}\sup_{t\in[0,t_\mathrm{f}]}\beta_N(t) = 0. \tag{50}$$

Finally, we obtain that:

$$\lim_{N\to\infty}\sup_{t\in[0,t_\mathrm{f}]}\|\lambda_N(t)-\lambda_*(t)\|_2 = 0. \tag{51}$$

■

B. Convergence of the Control Variables

So far we have shown that the state variables converge under Assumption (A3) according to Theorem 3.1. Moreover, if we obtain state convergence, we also ensure by Theorem 4.1 that the adjoint variables converge in L^2 for all $t \in [0, t_\mathrm{f}]$. We are now able to present the main statement in the form of the following corollary.

Corollary 4.1 (Control Convergence): Suppose Assumptions (A1), (A2), (A3), (A4), and (A5) hold for the optimal control problem (33) subject to state constraints (17a), (17b). Then

$$\lim_{N\to\infty}\sup_{t\in[0,t_\mathrm{f}]}\|u_N^*(t)-u_*^*(t)\|_2 = 0, \tag{52}$$

with optimal controls $u_N^*(t), u_*^*(t) \in L^2([0,1])$ for $t \in [0, t_\mathrm{f}]$.

Proof: From the Assumptions (A1) and (A4), we know that the Hamiltonian functions

$$\begin{aligned} &H_*(t,y(t),u(t),\lambda(t)) := \ell(t,y(t),u(t)) \\ &\quad + \big\langle \lambda(t), A_*y(t)+B_*u(t)+f(t,y(t),u(t))\big\rangle, \\ &H_N(t,y(t),u(t),\lambda(t)) := \ell(t,y(t),u(t)) \\ &\quad + \big\langle \lambda(t), A_Ny(t)+B_Nu(t)+f(t,y(t),u(t))\big\rangle, \end{aligned} \tag{53}$$

are continuous with respect to $t, y(t), u(t)$, and $\lambda(t)$. With Theorem 3.1, we ensure that as $A_N \to A_*$ in the operator norm, we have state convergence $y_N(t) \to y_*(t)$ in L^2 for all $t \in [0, t_\mathrm{f}]$. Moreover, in Theorem 4.1, we show that similarly as $A_N \to A_*$ and $B_N \to B_*$ in the operator norm, it follows that the adjoint variables converge, i.e., $\lambda_N(t) \to \lambda_*(t)$ in L^2 for all $t \in [0, t_\mathrm{f}]$. Therefore, as the state and adjoint variables are continuously dependent on the controls since, we can deduce that for the respective optimal solutions u_N^* and u_*^* it follows that:

$$\lim_{N\to\infty}\sup_{t\in[0,t_\mathrm{f}]}\|u_N^*(t)-u_*^*(t)\|_2 = 0. \tag{54}$$

■

Thus, we conclude that approximating the dynamical system with a sequence of step functions converging to the limit graphon ensures that the optimal controls also converge to those of the limit system.

C. Network Dynamical System

In this section, we show that our semilinear graphon system (1) covers the following finite dimensional network dynamical system

$$\dot{\mathbf{y}}_i(t) = \frac{1}{N}\left(\sum_{j=1}^N \mathbf{a}_{ij}\mathbf{y}_j(t) + \mathbf{b}_{ij}\mathbf{u}_j(t)\right) + \bar{f}(t,\mathbf{y}_i(t),\mathbf{u}_i(t)) \tag{55}$$

with $\bar{f}(t,y(x,t),u(x,t)) = f(t,y(t),u(t))(x)$, for all $i \in \{1,\dots,N\}$ with $\mathbf{y}(t), \mathbf{u}(t) \in \mathbb{R}^N$, $(\mathbf{a}_{ij})_{i,j=1}^N, (\mathbf{b}_{ij})_{i,j=1}^N \in \mathbb{R}^{N\times N}$, where N denotes the number of nodes in this network. We follow [9] and use step functions $A_N, B_N \in \mathcal{W}_1$ to transform the graphon system into a network system and vice versa.

Let $P = \{P_1,\dots,P_N\}$ be the uniform partition belonging to A_N and B_N, where each subset P_k has measure $\mu(P_k) = \frac{1}{N}$ for $k \in \{1,\dots,N\}$. In the following, the spatial component is of interest, thus we recall the graphon dynamical system (1):

$$\begin{aligned}\dot{y}_N(x,t) = &\int_0^1 A_N(x,\eta)y_N(\eta,t)\,\mathrm{d}\eta \\ &+ \int_0^1 B_N(x,\eta)u_N(\eta,t)\,\mathrm{d}\eta \\ &+ f(t,y_N(x,t),u_N(x,t)). \end{aligned} \tag{56}$$

Given that the graphons A_N, B_N are piecewise constant, we can split up the integral into a sum of integrals over each partition. Consequently for any $x \in P_i$, $i \in \{1, \ldots, N\}$, we can take the constant value $\mathbf{a}_{ij}$ of A_N out of the integral, thus we obtain:

$$\begin{aligned}\int_0^1 A_N(x,\eta) y_N(\eta,t)\,\mathrm{d}\eta &= \sum_{j=1}^{N} \int_{P_j} A_N(x,\eta) y_N(\eta,t)\,\mathrm{d}\eta \\ &= \sum_{j=1}^{N} \mathbf{a}_{ij} \int_{P_j} y_N(\eta,t)\,\mathrm{d}\eta\end{aligned} \tag{57}$$

and similarly:

$$\int_0^1 B_N(x,\eta) u_N(\eta,t)\,\mathrm{d}\eta = \sum_{j=1}^{N} \mathbf{b}_{ij} \int_{P_j} u_N(\eta,t)\,\mathrm{d}\eta. \tag{58}$$

Since we integrate over piecewise constant functions y_N and u_N, we define $\mathbf{y}_j(t) := y_N(\eta,t) \in \mathbb{R}$, $\mathbf{u}_j(t) := u_N(\eta,t) \in \mathbb{R}$ for all $\eta \in P_j$, $j \in \{1, \ldots, N\}$ and $t \in [0, t_\mathrm{f}]$. Then the integrals become $\int_{P_j} u_N(\eta,t)\,\mathrm{d}\eta = \frac{1}{N}\mathbf{u}_j(t)$ and $\int_{P_j} y_N(\eta,t)\,\mathrm{d}\eta = \frac{1}{N}\mathbf{y}_j(t)$, therefore we represent each piecewise constant value as an entry in a N-dimensional vector. In total, the piecewise constant dynamical system (56) with state and controls in $L^2([0,1])$ for given $t \in [0, t_\mathrm{f}]$ can be represented as (55) with $\mathbf{y}(t), \mathbf{u}(t) \in \mathbb{R}^N$ for $t \in [0, t_\mathrm{f}]$ and $(\mathbf{a}_{ij})_{i,j=1}^N =: \mathbf{A}_N, (\mathbf{b}_{ij})_{i,j=1}^N =: \mathbf{B}_N \in \mathbb{R}^{N \times N}$. Therefore, our previously infinite dimensional system with piecewise constant graphons A_N, B_N and functions y_N, u_N is represented as a system on a finite graph using adjacency matrices $\mathbf{A}_N$, $\mathbf{B}_N$.

V. Numerical Experiments

We now verify the derived convergence properties using two numerical examples. In the first step, we address Theorem 3.1 and only focus on state convergence. The second example exemplifies the state and control convergence in the optimal control context, see Corollary 4.1.

A. Experimental Setup

To generate an adjacency matrix that represents the approximating step function of the limit graphon $W \in \mathcal{W}_0$, we take each subset P_i of a uniform partition $P = \{P_1, \ldots, P_N\}$ and define a midpoint $x_i = \frac{i-0.5}{N}$. Then for all $i,j = 1, \ldots, N$ we have the relation

$$\mathbf{a}_{ij} = \mathbf{a}_{ji} \sim \mathrm{Ber}(W(x_i, x_j)), \tag{59}$$

with $\mathrm{Ber}(p)$ denoting the Bernoulli distribution with parameter $p \in [0,1]$. Using the Hoeffding inequality [21], the convergence rate of $\|A_N - W\|_\square$ can be shown to be in $\mathcal{O}\left(N^{-1/2}\right)$ with high probability [3], where A_N is a step function constructed according to (2).

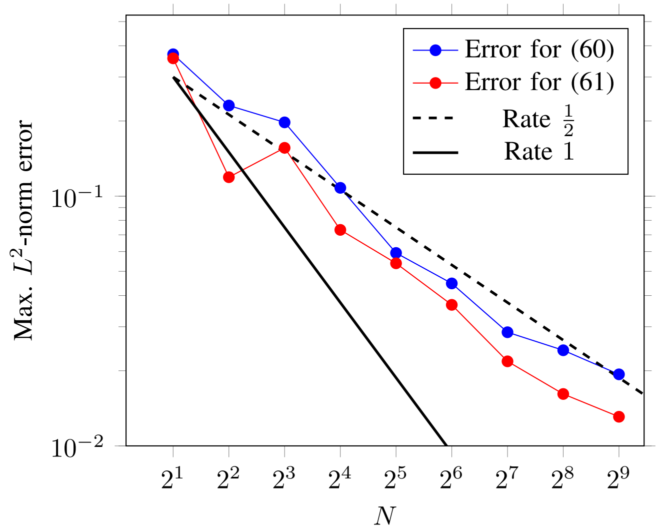

Fig. 2. Convergence in the state variables of graphon dynamical systems

B. Numercial Results

Our first example addresses the state convergence we have proven in Theorem 3.1. Let us consider the following dynamical systems:

$$\begin{aligned}\dot{y}(t) &= Ay(t) + f(t, y(t)), \\ y(x,0) &= \frac{1}{2}, \\ f(t,y(t))(x) &= y(x,t)(1 - y(x,t)), \\ (x,t) &\in [0,1] \times [0,3] \text{ a.e.,}\end{aligned} \tag{60}$$

and

$$\begin{aligned}\dot{y}(t) &= By(t) + g(t, y(t)), \\ y(x,0) &= \sin(\pi x), \\ g(t,y(t))(x) &= y^2(x,t) - y^3(x,t), \\ (x,t) &\in [0,1] \times [0,3] \text{ a.e.}\end{aligned} \tag{61}$$

Here, the control variables are assumed to be zero and to have no impact on the evolution of the systems. Therefore, we do not need to include $u(t)$. For the nonlinear terms in the dynamical systems (60) and (61) we have taken inspiration from reaction-diffusion equations in [22]. In (60), the nonlinear term is adopted from the Fisher or also called logistic equation. The second system (61) incorporates the nonlinear term from the Zeldovich equation.

The graphon operator $A \in \mathcal{W}_0$ is based on the kernel $W_A(x,y) = 1 - \max(x,y)$, which represents a uniform attachment graphon. The graphon $B \in \mathcal{W}_0$ is based on $W_B(x,y) = 0.5$, which represents an Erdős-Rényi graphon [3]. The evolution of this system is computed using a second order implicit Runge-Kutta method, more specifically the trapezoid rule. Our reference solution is computed on an approximate graph A_N of size $N = 1024$ and will be denoted by y_*. Using the adaptive high-order quadrature rule provided by [23], we evaluate the maximum L^2 error $\|y_*(t_i) - y_N(t_i)\|_2$ over all $i \in \{0, \ldots, 127\}$ for $N = 2, 4, \ldots, 512$.

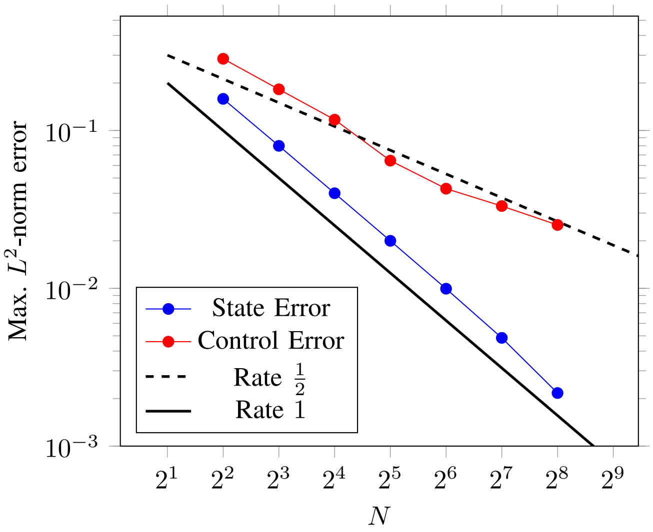

Fig. 3. Convergence of the state and control variables of a graphon system embedded into an optimal control problem.

Figure 2 shows the expected state convergence to our reference as N increases. Additionally, we can empirically deduce the convergence rate to be roughly $1/2$, which exactly coincides with the rate we have shown before, see Corollary 3.2.

The optimal control example aims to minimize the square deviation of the state variables to the origin in the L^2 norm while also minimizing the square of the controls u over time $t \in [0, 3]$:

$$\min_{y,u} \quad \int_0^3 \|y(x,t)\|_2^2 + \|u(x,t)\|_2^2 \, \mathrm{d}t \tag{62a}$$

$$\text{s.t.} \quad y(x,0) = \sin(\pi x), \tag{62b}$$

$$\dot{y}(t) = Ay(t) + Bu(t), + h(t, y(t), u(t)), \tag{62c}$$

$$u(t) \in L^2([0,1]), \tag{62d}$$

$$h(t, y(t), u(t))(x) = y(x,t)(1 - y(x,t)), + u(x,t), \tag{62e}$$

$$(x,t) \in [0,1] \times [0,3] \text{ a.e.} \tag{62f}$$

Here, the graphon operators $A, B \in \mathcal{W}_0$ are represented by a uniform attachment graphon $W_A(x,y) = 1 - \max(x,y)$ (Figure 1) and a k-nearest-neighbor graphon $W_B(x,y) = 1$ if $|x-y| \le 0.1$ or $|x-y+1| \le 0.1$ else 0.
We then discretize in time using a direct collocation ansatz with the trapezoid rule, see, e.g., [24] and model the nonlinear program using JuMP.jl [25]. The underlying solver is Ipopt.jl (wrapper for Ipopt [26]), which is based on an interior point method and the linear sparse solver MUMPS [27], [28].

Our reference is an optimal control solution with approximate graphs A_N, B_N of size $N = 512$ and a time discretization of 128 grid points. The solution will be denoted by u_* with its resulting state y_*. We then solve the optimal control problem for $N = 4, 8, 16, \ldots, 256$ on the same time discretization and compute $\max_{i\in\{0,\ldots,127\}} \|u_*(t_i) - u_N(t_i)\|_2$ and $\max_{i\in\{0,\ldots,127\}} \|y_*(t_i) - y_N(t_i)\|_2$ with an adaptive high-order quadrature scheme [23].

TABLE I
STATE AND CONTROL ERRORS FOR OPTIMAL CONTROL PROBLEM

N	State Error	Control Error	Objective
4	0.15868	0.28484	0.88314
8	0.07994	0.18240	0.96684
16	0.04003	0.11712	0.81315
32	0.02000	0.06442	0.74881
64	0.00994	0.04279	0.77620
128	0.00485	0.03322	0.76397
256	0.00217	0.02520	0.76081
Ref.	-	-	0.75668

The results shown in Figure 3 do comply with our expectation in that the state and control variables converge as N becomes larger. Additionally, we see that the control variables converge with a rate of 0.5 and the state variables with a rate of 1. However, this higher convergence rate in the state variables does not invalidate our proof showing the rate of 0.5. Possible reasons might include missing regularity properties of our reference object or superconvergence, observed in the field of optimal control of partial differential equations [29]–[31].

Most importantly, our finite dynamical systems approach the reference objective value of 0.75668 as the number of vertices tends to our reference size of $N = 512$, see Table I.

VI. CONCLUSION

We have formally derived state, adjoint, and control convergence of semilinear dynamical network systems towards their limit objects defined on a graphon system in the context of optimal control. Thereby, we have extended existing theory on linear dynamical graphon systems due to [9] and adopted the main assumption of converging graph sequences, namely that the piecewise constant graphons $A_N \to A_*$ and $B_N \to B_*$ in the operator norm as the number of vertices N tends to infinity.

The proof for the state convergence relies on semigroup theory and the Grönwall-Bellman equation, while assuming a single but arbitrary L^2-type control function for all dynamical graphon systems. We show the convergence of the adjoint variables of approaching graphons systems by investigating the infinite-dimensional Hamiltonian function and a backward-in-time variant of the Grönwall-Bellman inequality. Since our Hamiltonian function is continuous in the remaining control variable, we finally proof convergence of the control functions.

In other words, we have shown that we can approximate the optimal control performance of an infinite-dimensional semilinear network system by a dense but finite-dimensional semilinear network system up to arbitrary precision. These results enable the systematic design and evaluation of semilinear network systems and its control performance. The numerical examples in this paper verify convergence rates above 0.5 for the state and control variables.

References

[1] L. Lovász and B. Szegedy, "Limits of dense graph sequences," *Journal of Combinatorial Theory, Series B*, vol. 96, no. 6, pp. 933–957, 2006.

[2] ——, "Szemerédi's lemma for the analyst," *Geometric and Functional Analysis*, vol. 17, pp. 252–270, 04 2007.

[3] L. Lovász, *Large Networks and Graph Limits*, ser. American Mathematical Society colloquium publications. American Mathematical Society, 2012.

[4] C. Borgs, J. Chayes, L. Lovász, V. Sós, and K. Vesztergombi, "Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing," *Advances in Mathematics*, vol. 219, no. 6, pp. 1801–1851, 2008.

[5] C. Borgs, J. Chayes, L. Lovász, V. Sós, and K. Vesztergombi, "Convergent sequences of dense graphs ii. multiway cuts and statistical physics," *Annals of Mathematics*, vol. 176, pp. 151–219, 2012.

[6] C. Borgs, J. Chayes, L. Lovász, V. Sós, and K. Vesztergombi, "Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing," *Advances in Mathematics*, vol. 219, no. 6, pp. 1801–1851, 2008.

[7] G. S. Medvedev, "The nonlinear heat equation on dense graphs and graph limits," *SIAM Journal on Mathematical Analysis*, vol. 46, no. 4, pp. 2743–2766, 2014.

[8] C. Bick and D. Sclosa, "Dynamical systems on graph limits and their symmetries," *Journal of Dynamics and Differential Equations*, 2024.

[9] S. Gao and P. E. Caines, "Graphon control of large-scale networks of linear systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 4090–4105, 2020.

[10] ——, "Graphon linear quadratic regulation of large-scale networks of linear systems," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 5892–5897.

[11] P. E. Caines and M. Huang, "Graphon mean field games and their equations," *SIAM Journal on Control and Optimization*, vol. 59, no. 6, pp. 4373–4399, 2021.

[12] ——, "Graphon mean field games and the gmfg equations," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 4129–4134.

[13] N. Alon and A. Naor, "Approximating the cut-norm via grothendieck's inequality." *SIAM J. Comput.*, vol. 35, pp. 787–803, 01 2006.

[14] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň, "Relating the cut distance and the weak* topology for graphons," *Journal of Combinatorial Theory, Series B*, vol. 147, pp. 252–298, 2021.

[15] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň, "Cut distance identifying graphon parameters over weak* limits," *Journal of Combinatorial Theory, Series A*, vol. 189, p. 105615, 2022.

[16] S. Janson, "Graphons, cut norm and distance, couplings and rearrangements," *NYJM Monographs*, vol. 4, 2013.

[17] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, 1st ed., ser. Applied Mathematical Sciences. Springer New York, 1983, vol. 44.

[18] R. Bellman, "Asymptotic series for the solutions of linear differential-difference equations," *Rendiconti del Circolo Matematico di Palermo*, vol. 7, no. 3, pp. 261–269, 9 1958.

[19] B. Pachpatte, "Chapter one - linear integral inequalities," in *Inequalities for Differential and Integral Equations*, ser. Mathematics in Science and Engineering, B. Pachpatte, Ed. Elsevier, 1998, vol. 197, pp. 9–97.

[20] M. I. Krastanov, N. K. Ribarska, and T. Y. Tsachev, "A pontryagin maximum principle for infinite-dimensional problems," *SIAM Journal on Control and Optimization*, vol. 49, no. 5, pp. 2155–2182, 2011.

[21] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

[22] B. Gilding and R. Kersner, *Travelling Waves in Nonlinear Diffusion-Convection Reaction*. Birkhäuser Basel, 2004.

[23] S. G. Johnson, "QuadGK.jl: Gauss–Kronrod integration in Julia," https://github.com/JuliaMath/QuadGK.jl, 2013.

[24] T. Binder, L. Blank, H. Bock, R. Bulirsch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J. Schlöder, and O. Von Stryk, *Introduction to Model Based Optimization of Chemical Processes on Moving Horizons*. Springer, Berlin, Heidelberg, 2001, pp. 295–340.

[25] M. Lubin, O. Dowson, J. Dias Garcia, J. Huchette, B. Legat, and J. P. Vielma, "JuMP 1.0: Recent improvements to a modeling language for mathematical optimization," *Mathematical Programming Computation*, 2023.

[26] A. Wächter and L. T. Biegler, "On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 2006.

[27] P. Amestoy, I. S. Duff, J. Koster, and J.-Y. L'Excellent, "A fully asynchronous multifrontal solver using distributed dynamic scheduling," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, no. 1, pp. 15–41, 2001.

[28] P. Amestoy, A. Buttari, J.-Y. L'Excellent, and T. Mary, "Performance and Scalability of the Block Low-Rank Multifrontal Factorization on Multicore Architectures," *ACM Transactions on Mathematical Software*, vol. 45, pp. 2:1–2:26, 2019.

[29] C. Meyer and A. Rösch, "Superconvergence properties of optimal control problems," *SIAM Journal on Control and Optimization*, vol. 43, no. 3, pp. 970–985, 2004.

[30] H. Guan and D. Shi, "Superconvergence analysis of a nonconforming finite element method for monotone semilinear elliptic optimal control problems," *Numerical Methods for Partial Differential Equations*, vol. 36, no. 6, pp. 1405–1417, 2020.

[31] Y. Dai and Y. Chen, "Superconvergence for general convex optimal control problems governed by semilinear parabolic equations," *International Scholarly Research Notices*, vol. 2014, no. 1, p. 579047, 2014.