INTERACTION SPACES: TOWARDS A MATHEMATICAL THEORY OF COMPLEX SYSTEMS

P. GIORDANO AND J.D. FARMER

Abstract. We present the first steps of *interaction spaces theory*, a universal mathematical theory of complex systems which is able to include cellular automata, agent based models, master equation based models, networked dynamical models, neural networks and evolutionary algorithms in a single notion. Therefore, interaction spaces theory represents a common precise mathematical language that can be used to describe several complex systems modeling frameworks. We consider Markovian interaction spaces and we prove for them a master equation, but also non-Markovian interaction spaces, which are described by a system of mean derivative equations. We define the notion of complex adaptive system by following the original intuition of G.K. Zipf, and we prove for them a power law. We finally illustrate the definition of complex adaptive system using a generalization of Von Thünen's model, which can be easily generalized to other economic systems. Every notion is introduced both using an intuitive description and listing a lot of examples, and using a modern mathematical language.

1. Introduction: why do we need a mathematical theory of complex systems?

Do we need to update the following citations?

Throughout the history of science, several disciplines have considerably gained by a sound mathematical foundation: quantum mechanics, continuum mechanics, thermodynamics, medicine, biology, information science, economics, social sciences, and urban studies, to name but a few. Indeed, the contribution of mathematics to many disciplines can be considered a general process that occurs when the solution of problems requires the strongest notion of rational truth corroborated by a meaningful validation.

At present, different theories are used for the modeling of complex systems (CS): we can cite, e.g., cellular automata, [35], agent based models (ABM), [59], master equation based models, [45], networked dynamical systems, [37], neural networks, [34], and evolutionary algorithms, [2]. However, there is no universal mathematical theory of CS, i.e. a theory sufficiently powerful to range from ABM to systems described by some type of differential equations (see e.g. [27, 8, 12, 46, 20, 5, 10] for the discussion in the literature concerning this problem).

In this article, we introduce a new mathematical structure, called *interaction space* (IS), having the property to include in a single notion all the previously listed modeling frameworks. Clearly, such a founding mathematical theory could provide great impact from the perspectives of precision, common language and general results which would hence be applicable in all these settings.

Other aims we have in mind are the following:

- With our definition of IS, we do not aim to provide an explicit definition of CS but rather to show a common structure that every CS appears to possess.
- A common mathematical language can be useful to precisely formulate problems like phase transitions and critical phenomena, pattern formation theory, ergodic theory, study of ABM as dynamical systems, etc. (see e.g. [12] for similar problems).
- IS represents a proposal for a good mathematical definition of ABM. This definition would open the possibility to start a mathematical study of a large family of these models (see e.g. [32] for a mathematical approach to the dynamics of some types of ABM; in [8] it is also stated: "*At present [1997], there exists no decent mathematical theory of such processes*").

In other words, a mathematical theory of CS aims to link phenomenological studies (e.g. estimates of power laws) to a modern mathematical theory, so that to make a step further obtaining more general, clear and widely applicable results.

1.1. **What this article is not for.** IS theory is a useful mathematical language to model systems having an arbitrary number of interacting entities: from a small finite number to a continuum. Here, we are distinguishing a "language" from a fundamental theory: obviously we are not going to introduce the analogous of Newton's laws of dynamics satisfied by every CS. We can say that IS theory is more fundamental and basic than dynamical laws, because it is a language which can be used to describe every CS. Indeed, in all CS we must necessarily have a set of interacting entities, they must have a state space, they are involved in interactions, these interactions occur at (stochastic) times and produce something that depends on the information collected in some subset of the system and can modifies the state of other entities. IS theory is all these necessary conditions stated using a precise mathematical language.

1.2. **Schema of the paper.** We start the paper by giving an intuitive description of IS (Sec. 2.1) and of their dynamics (Sec. 2.2). Only after, we start the mathematical theory by giving some mathematical definitions faithfully corresponding to this intuition (Sec. 2.3). In Sec. 4, we consider the case of Markovian IS and we prove for them a master equation. More generally, IS theory includes also non-Markovian dynamics, which can be described using a system of mean derivative equations (Sec. 5). In Sec. 6, we introduce the notion of *complex adaptive systems* (CAS) by sizing the original intuition of G.K. Zipf, [63] (Sec. 6) in what we call *generalized evolution principle* (GEP). Following the classical idea of B. Mandelbrot, we prove in a very general context that a large class of systems that follow the GEP, satisfy a power law. In Sec. 7, we finally illustrate the GEP using a generalization of Von Thünen's model, which can be easily generalized to other systems.

2. Interaction spaces: a universal mathematical theory for complex systems

We firstly describe a generic IS by using only an intuitive approach and giving several examples, exactly like agent based models (ABM) are frequently presented. Secondly, we give a more mathematical approach, which is a necessary step for modern mathematics.

IS can be seen as a good interpolation between artificial intelligence (AI) based methods, more typical of ABM (see e.g. [59] and references therein), and physics' methods more typical of synergetics and econophysics (see e.g. [23, 24, 45, 44, 51, 58]).

2.1. **Informal description of interaction spaces.** IS theory aims at modeling complex systems enclosed in the following general frame:

Interacting entities: The system is made by *interacting entities* described by dynamical *state variables*. An interacting entity is everything able to send or receive signals to interact with something else.

Examples of interacting entities are: cells of a CA and agents of a ABM, a vehicle, a traffic light or the piece of road between two following cars, advertisement in a street, goods exchanged in a market, a whole population of individuals sharing common features, etc.

Interactions: Each interaction i of type α can be described as a causally directed process in which a set of *agent* entities $a_1, \ldots, a_n$, modify the state of a *patient* entity p through a *propagator* entity r. We distinguish between the type α of the interaction, which is usually a label useful to classify different interactions, and the interaction $i = (a_1, \ldots, a_n, r, \alpha, p)$ that includes all the information. The propagator r can be thought of as a signal-entity activated by agents, and carrying the cause-effect relation sent by agents $a_1, \ldots, a_n$ to the patient p. We also say that the state space of the propagator r works as a *resource space* for the changing of the state of the patient p, see below.
The general form of an interaction i is hence:

$$a_1, \ldots, a_n \text{ have an interaction } \alpha \text{ with } p \text{ through } r \tag{2.1}$$

which will be also indicated with the notation

$$a_1, \ldots, a_n \xrightarrow{r,\alpha} p \tag{2.2}$$

or with a diagram as in Fig. 2.1.

Examples: a physical interaction between one particle p_1 that sends a signal s to another particle p_2, $i = (p_1, s, \texttt{sendSignal}, p_2)$; or a firm (agent) sending an advertisement (propagator) and hence changing the state of several people (patients); a suitable set of goods in a market (agents) sending a signal (propagator) that carries information useful for buyers (patients); a biological entity (agents) sending a chemical signal (propagator) to another entity (patients) having receptors able to recognize that signal; an object in an object oriented program sending a message to another object. In urban models, agents can be individuals acting in the urban space, patients can be lots of terrain, signals can be volumes and surfaces produced for different uses so that the state space of propagators is linked to the available surface and volume at disposal, depending on the master plan (resources; see [54, 55]).

Activation: An interaction can occur only if at least one of the agents $a_1, \ldots, a_n$ involved in a given interaction i is *active for that interaction*. Indeed, in the state of each interacting entity there is always a time dependent Boolean variable $\mathrm{ac}_e^i(t) \in \{0, 1\}$ indicating if, with respect to the given interaction, the entity is active or not. Intuitively, an agent a_k is active at time t_1, and we write $\mathrm{ac}_{a_k}^i(t_1) = 1$, with respect to an interaction i if it can start that interaction at $t = t_1$. At the same time t_1, agents activate the propagator r: $\mathrm{ac}_r^i(t_1) = 1$. The propagator r will take a certain time $t_2 - t_1$ to arrive at the patient. If no other entity and interaction stops r, it

arrives and activates the patient at time t_2: $\mathrm{ac}_p^i(t_2) = 1$. If the interaction i lasts the time $t_3 - t_1$, then the patient p will remain active, for the interaction i, for the whole time interval $[t_2, t_3]$ after which it will be deactivated: $\mathrm{ac}_p^i(t_3) = 0$ together with all the other entities involved in i.
Active agents can also be interpreted in biological terms as entities sending some kind of chemical signal to patients entities having suitable receptors to recognize it; in this description, propagators are entities carrying the signals. Therefore, agents which are not already active in the initial condition of the system, can pass to an active state as a consequence of an interaction (endogenous or exogenous).

Examples: in the above mentioned example about firm's advertisement, only people activated, in some way, for the advertised products will have a state modification; only the biological entities having suitable receptors are active for the corresponding interactions; only hungry predators are active for hunting preys; only software objects with a suitable public state variable can receive a message to change that variable.

Neighborhood of an interaction: The occurrence of an interaction and its effects depends only on the history of the state of a set of entities called the *neighborhood* of the interaction. The neighborhood of the interaction i is intuitively defined by all the entities from which i takes the information it needs to operate. The neighborhood of an interaction always includes agent, patient and propagator entities whenever they are active for that interaction.

Examples: if an agent is searching for a new house, only the information collected in some order in its memory will affect its future decisions; only the state of the cells belonging to the neighborhood can influence the future state of a given cell in a cellular automata; only the objects in the visual field of a pedestrian may influence its goal-oriented path; the information collected in a graphical user interface may influence the possible starting of a given computer program.

Goods and resources: When and interaction i starts, its agents probabilistically extract a quantity π (called *good*) from the state space of the propagator r of i (i.e. the space of resources) and send the signal (r, π) to the patient p. Goods' probability distribution depends only on the history of interaction's neighborhood. A situation where the resources are exhausted before the finishing of the interaction, is an example of situation where the propagator is deactivated before the ending of the interaction, and hence the interaction itself is deactivated.

Examples: an excited electron (agent) produces a photon (propagator) that changes the state of another electron (patient) in a scattering interaction. The input currents of a neuron are the signals that will be integrated to produce a suitable changing of the output synapses. A developer decides to build a new house and produces as signal the house's project, hold in the state of a suitable abstract propagator entity. Starting from this project, the state of the building's plot will change in a suitable amount of time, unless the municipal administration blocks the project.

Occurrence times: Generally speaking, interactions occur at random times, whose distribution depends on the history of interaction's neighborhood only. Using the previous notations, we can say that t_1, t_2, t_3 are random times following suitable model-depending distributions. Of course, deterministic times can always be included as particular cases using Dirac delta distributions.

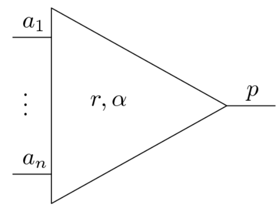

FIGURE 2.1. Representation of an interaction using a diagram.

Examples: an interaction where an agent chooses a shop on the basis of its information about quality, prices, and goods availability, occurs at random times with a suitable distribution depending both on objective and subjective characteristics; an interaction describing a house leasing occurs at random times depending on several factors, e.g. the rate of birth, of marriage, of immigration, etc; the infection of an organism by a virus depends randomly on the hosts encountered; if this virus is considered as the propagator of the infection interaction, then it will arrive to the possible next organism after a random time depending on its aging; an excited electron (agent) produces a photon (propagator) that, in a deterministic time depending on the media, changes the state of another electron (patient) in a scattering interaction; the starting of a program randomly depends on the interaction of the user with the program's interface.

Transition functions: When an interaction i starts to change the state variables of a patient p, this change is determined by a suitable *transition function* f_i depending on i and on the extracted goods π. In other words, when a subsequent interaction j s $t_1' \leq t_2' \leq t_3'$ (where $t_1' \geq t_3$) on the same patient p, then p must be deactivated for i, i.e. $\mathrm{ac}_p^i(t_1') = 0$, we will have that $\mathrm{ac}_p^j(t_2') = 1$, and the transition function will change from f_i to f_j. We can also say that interacting entities are modeled like dynamical systems driven by input currents produced by other interacting entities. At each time instant, every patient can be active for only one interaction.

Examples due to an interaction and before the occurring of a subsequent interaction: a bouncing billiard ball; a pedestrian between two subsequent interactions with other pedestrians or obstacles; the process of building a house after its starting time and before its end, and the internal evolution of a box in a flow chart representing a computer program.

The intuitive description above can be summarized by saying: in an interaction, agents activate and send the propagator and the goods as a signal to modify the state of the patient; the modification depends on information collected from the neighborhood of that interaction; the starting time, the speed of the signal and the duration of the interaction can be stochastic. We hence note that interactions are local in the sense that they are affected only by entities in the neighborhood; secondly their occurrence is causally constrained by logical conditions expressed by the activation of the entities.

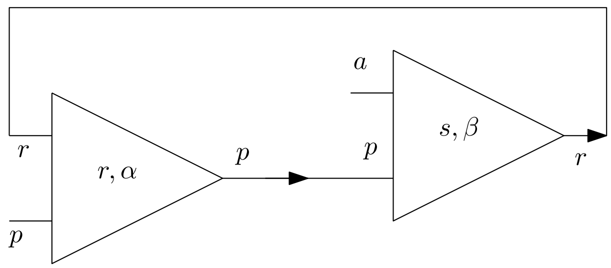

FIGURE 2.2. Graphical representation of two interactions where the propagator is at the same time agent and propagator of the first interaction and patient of the second interaction. Using the agent a, we can change the status of r and hence the resources of the first interaction.

If we represent an interaction by means of a graph, like in the following figure 2.2, and connect two graphs when they share an entity, we obtain a network representing the mentioned causal flows in the system.

2.2. **Dynamics of an interaction space.** The dynamics of a generic IS, is determined by answering the following questions:

(a) *When do interactions occur?* For each interaction i that is active at time t (this means that at least one of its agents is already active, i.e. that at least one cause of i have already occurred before), and for a given past history x of the state of the neighborhood of i, we have to provide two random variables, which specify the *starting time* S_i *of* i and the *arrival time* $A_i \geq S_i$ *of the propagator* r *of* i. As an idealized case, we can take $A_i = S_i$ and choose a suitable $\Delta t > 0$ such that every active interaction will occur at discrete deterministic time $k \cdot \Delta t$, $k \in \mathbb{N}$ (synchronous occurrences). As another example, if we take a suitable (problem depending) small $\Delta t > 0$, we can assume that events occurring at a time distance less than Δt are stochastically independent (asynchronous occurrences with an exponential distribution). We always assume that, at each time, the number of active interactions is finite.

(b) *What does each interaction produce when it occurs?* Let us assume that $i = \left(a_1, \ldots, a_n \xrightarrow{r,\alpha} p\right)$ is an active interaction starting at time t_1. The goods π to change the state x_p of the patient p are extracted using a given probability G_i on the state space X_r of the propagator r (the space of resources of i).

(c) *How do we change patients' states?* The propagator arrives and activates the patient p at time $t_2 \geq t_1$. Using the produced goods π, the state of p can start to change by following a function of the form $x_p(t) = f_i(t, \omega; t_1, t_2, t_3, \pi)$, called the *transition function of* p *rel.* $i = \left(a_1, \ldots, a_n \xrightarrow{r,\alpha} p\right)$. Here, ω represents the possibility of having a stochastic dynamics. At time $t_3 \geq t_2$ the interaction i stops, and all agents, patient and propagator are deactivated for i.

(d) *What happens if two interactions occur at the same time on the same patient?* The modeler has to devise a criterion to answer this question, either using a deterministic or a random choice, depending on our problem. In fact, we recall that at each time instant, every patient can be active for only one interaction.

The algorithm to simulate an IS is now clear:

(I) Because an interaction can start only if at least one of its agents is active, we necessarily need to start any simulation with at least one entity that is already (exogenously or endogenously) active for some interaction.

(II) If no interaction has at least one active agent, the simulation stops.

(III) For every active interaction i and every time t, using S_i, we extract the starting time $t_{1i} > t$ of all the interactions and compute $t_1 := \min\{t_{1i} \mid i \text{ is active at time } t\}$.

(IV) Let i be an interaction that occurs at the minimum time t_1. We can hence set the time to $t = t_{1i}$. If the time t is greater than the final time t_{end}, then the simulation stops. Otherwise, using the random variables G_i and A_i, we extract the goods π produced by the agents of i and the arrival time t_{2i} of the propagator. If two or more interactions act on the same patient at time $t = t_{2i}$, using the criterion fixed in (d), we decide the unique one that actually occurs.

(V) If the arrival time t_{2i} represents the next event (i.e., before the starting times t_{1j} of all the other interactions j), we change the state of the patient p of i using the corresponding transition function. We set the time at $t = t_{2i}$. If $t > t_{\text{end}}$, the simulation stops. Because we have changed the state of an entity, we restart from step (II). Otherwise, we repeat from step (IV) for the next occurring interaction.

The terms agent, patient, propagator and members of a neighborhood are collectively named *roles* of entities in an interaction. Of course, interacting entities can play different roles in different interactions and more than a role in the same interaction, e.g. a propagator of i can also be at the same time an agent of the same interaction and a patient of another interaction j which triggers the resources of i, see Fig. 2.2. Note that this informal description already allows for a practical implementation of simulated IS (see e.g. [53]).

2.3. **Mathematical definition of IS.**

2.3.1. *Interacting entities and interactions.* A mathematical definition of IS is a necessary step to start a mathematical theory, and hence to prove general theorems in a clear way.

Already in the informal description of IS, it is clear that many components are needed to define an IS: a set of entities, a set of types of interactions, state maps and state spaces, probabilities for times and goods, etc. For this reason, using a nested approach, we introduce several structures that will define the notion of IS.

Definition 1. A *system of entities and interactions* $\mathcal{EI} = (E, t_{\text{st}}, t_{\text{end}}, \mathcal{T}, I)$ is given by the following data which satisfy the following conditions:

(i) A set E, called *the set of interacting entities.*

(ii) A *time interval* $[t_{\text{st}}, t_{\text{end}}]$, with $0 \le t_{\text{st}} < t_{\text{end}} \le +\infty$.

(iii) A finite set $\mathcal{T}$ called *the set of types of interactions.*

(iv) A set I called *the set of interactions* verifying the following condition: every interaction $i \in I$ can be written as $i = (a_1, \dots, a_n, r, \alpha, p)$ for some type of interaction $\alpha \in \mathcal{T}$, some entities $a_1, \dots a_n$, r, $p \in E$, where $n \ge 0$ depends on i;

Remark 2.

(a) We set $E_i := \{a_1, \dots, a_n, r, p\}$, $\mathrm{ag}(i) := \{a_1, \dots, a_n\}$, $\mathrm{pa}(i) := \{p\}$ and $\mathrm{pr}(i) := \{r\}$ to denote all the interacting entities involved in the interaction i, agents, patient and propagator of i, resp.
(b) There is no a priori limitation on the cardinality of the set E of interacting entities, even though in the majority of cases it is usually finite.
(c) The system is studied in the time interval $[t_{\mathrm{st}}, t_{\mathrm{end}}]$; if $t_{\mathrm{end}} = +\infty$, we will use the notation $[t_{\mathrm{st}}, t_{\mathrm{end}}] = [t_{\mathrm{st}}, +\infty]$ to mean $[t_{\mathrm{st}}, +\infty)$.
(d) Generally speaking, the interactions are non Newtonian: they involve more than one agent and they are, in general, not reversible, i.e. there is not an action-reaction principle. For example, it does not seem useful to think as Newtonian the interaction of a pedestrian with an obstacle or of a builder with a house or of an object in an object oriented programming language with another object: even if frequently to these interactions correspond another interaction as answer, in general there is no useful way to say that the intensity (force) of the cause interaction is the opposite of the intensity (force) of the reaction interaction.

2.3.2. *State spaces and activation.*

Definition 3. Let $\mathcal{EI} = (E, t_{\mathrm{st}}, t_{\mathrm{end}}, \mathcal{T}, I)$ be a system of entities and interactions. A *system of state spaces and activation maps* $\mathcal{SA} = (\mathrm{ac}, S, \mathfrak{S}, x)$ *for* $\mathcal{EI}$ is given by the following data which satisfy the following conditions:

(i) An *activation map* ac that satisfies
  (a) $\forall i \in I \, \forall e \in E \, \forall t \in [t_{\mathrm{st}}, t_{\mathrm{end}}] : \ \mathrm{ac}^i_e(t) \in \{0, 1\}$, i.e. the activation map is a Boolean function.
  (b) For all $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$, the set $E_t := \left\{ e \in E \mid \exists i \in I : \ \mathrm{ac}^i_e(t) = 1 \right\}$ of all the entities which are active for some interaction at time t, is non empty.

  For every interaction $i \in I$, we say that *i is active at time t*, and we write $\mathrm{ac}^i(t) = 1$, if at least one of the entities involved in i is active at t, i.e.
$$\exists e \in E_i : \ \mathrm{ac}^i_e(t) = 1, \tag{2.3}$$
  Otherwise, we set $\mathrm{ac}^i(t) := 0$. Moreover, we denote by
$$I_t := \left\{ i \in I \mid \mathrm{ac}^i(t) = 1 \right\}$$
  the set of all the interactions which are active at time t. We finally request, by definition, the following condition:
  (c) At each time instant, for all entity $p \in E$ there is one and only one active interaction that has p as active patient, i.e.
$$\forall t \in [t_{\mathrm{st}}, t_{\mathrm{end}}] \, \forall p \in E \, \exists! i \in I_t : \ p = \mathrm{pa}(i) \text{ and } \mathrm{ac}^i_p(t) = 1 \tag{2.4}$$
(ii) For every interacting entity $e \in E$, a Borel space $(S_e, \mathfrak{S}_e)$, i.e. a measurable and topological space whose measurable sets are generated by open sets, called *the state space of the interacting entity* e.
(iii) A *state map* x that satisfies
$$\forall t \in [t_{\mathrm{st}}, t_{\mathrm{end}}] \, \forall e \in E_t : \ x_t(e) \in S_e \tag{2.5}$$

Remark 4.

(a) The request on the state space $(S_e, \mathfrak{S}_e)$ to be a measurable space is the least possible one, from the mathematical point of view, even if usually on a state space there is a richer structure, e.g. $S_e = \mathbb{R}^d$ for some $d > 0$ depending on $e \in E$. Let us note explicitly that the state space is not time dependent.

(b) Condition (2.5) implies that we can talk of the state $x_t(e)$ of the entity e at time t only if the entity is already active for some interaction, i.e. only if $e \in E_t$.

(c) Depending by the progress of the interaction i, we can have: I) at least one agent is active when the interaction starts; II) the propagator is active and it is reaching the patient; c) the patient is active and its state space is changing due to i. Once an entity finishes its task, it can be deactivated. This justifies the existence quantifier in (2.3). Since E_t is finite, also I_t is finite.

(d) Condition (2.4) implies that we cannot have two interactions acting on the same patient at the same time. Mathematically, if $i,\ j \in I_t$, $\mathrm{ac}^i_{\mathrm{pa}(i)}(t) = 1$, $\mathrm{ac}^j_{\mathrm{pa}(j)}(t) = 1$, then $\mathrm{pa}(i) \neq \mathrm{pa}(j)$. It is therefore implicit in this condition that the modeler has to chose a way to decide among interactions that occur at the same time on the same patient. This choice depends on the model and it can be deterministic or stochastic.

2.3.3. *Information to start an interaction.* If we think at the state function x as the output of a simulation, then neighborhood, start, arrival and stop times of an interaction i, and the probability to extract goods from the space of resources $S_{\mathrm{pr}(i)}$, are all the information that we need to run i before applying its transition function to its patient.

Definition 5. Let $\mathcal{EI} = (E, t_{\mathrm{st}}, t_{\mathrm{end}}, \mathcal{T}, I)$ be a system of entities and interactions, and let $\mathcal{SA} = (\mathrm{ac}, S, \mathfrak{S}, x)$ be a system of state spaces and activation map for $\mathcal{EI}$. A *system of information* $\mathcal{I} = (\mathcal{U}, T, G)$ *for the interactions of* $\mathcal{EI}$ *and* $\mathcal{SA}$ is given by the following data which satisfy the following conditions:

(i) For all $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$ and all active interactions $i \in I_t$, a subset $\mathcal{U}_i(t) \subseteq E_t$ called the *neighborhood of i at time t.*

(ii) The neighborhood must contain all the active entities of the interaction i, i.e.

$$\forall e \in E_i: \ e \in E_t \ \Rightarrow \ e \in \mathcal{U}_i(t).$$

For all entities $e \in \mathcal{U}_i(t)$ in the neighborhood of i, and all times $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$, we define the *story of the state $x_{(-)}(e)$ up to time t* as the map

$$\mathrm{n}_i\, x_t(e): \ \tau \in \{t_1 \in [t_{\mathrm{st}}, t] \mid e \in E_{t_1}\} \mapsto x_\tau(e) \in S_e.$$

We denote by $\mathrm{n}_i\, x$ the function $(e, t) \mapsto \mathrm{n}_i\, x_t(e)$, which is defined only for $e \in \mathcal{U}_i(t)$.

(iii) If $i \in I_t$ is an active interaction at time $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$, then $T_i(-; \mathrm{n}_i\, x, t)$ is a probability on the set

$$\left\{(t_1, t_2, t_3) \in [t, t_{\mathrm{end}}]^3 \mid t_1 \le t_2 \le t_3\right\}.$$

If (S, A, E) are random variables with joint probability $T_i(-; \mathrm{n}_i\, x, t)$, S is called the *starting time of* i, A is called the *arrival time of the propagator of* i, and E is called *the ending time of* i. The difference $E - S$ is called the *duration of* i.

(iv) If $i \in I_t$ is an active interaction at time $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$, then $G_i(-; \mathrm{n}_i\, x, t)$ is a probability on $\left(S_{\mathrm{pr}(i)}, \mathfrak{S}_{\mathrm{pr}(i)}\right)$. $G_i(-; \mathrm{n}_i\, x, t)$ is called the *probability to extract goods from the space of resources* $S_{\mathrm{pr}(i)}$ of the interaction i.

Note that both the probabilities $T_i(-; \mathrm{n}_i\, x, t)$ and $G_i(-; \mathrm{n}_i\, x, t)$ depend on the story $\mathrm{n}_i\, x$ of the state of the entities in the neighborhood of i. So, if we consider another state map y such that $\mathrm{n}_i\, x = \mathrm{n}_i\, y$, then $T_i(-; \mathrm{n}_i\, x, t) = T_i(-; \mathrm{n}_i\, y, t)$ and $G_i(-; \mathrm{n}_i\, x, t) = G_i(-; \mathrm{n}_i\, y, t)$.

2.3.4. *The transition functions.* We can finally introduce the notion of transition function of an interaction:

Definition 6. Let $\mathcal{I} = (\mathcal{U}, T, G)$ be a system of information for the interactions of $\mathcal{EI} = (E, t_{\mathrm{st}}, t_{\mathrm{end}}, \mathcal{T}, I)$ *and* $\mathcal{SA} = (\mathrm{ac}, S, \mathfrak{S}, x)$. A *system of transition functions* $\mathcal{TF} = (f, \Omega, \mathcal{F}, P)$ *for* $\mathcal{EI}$, $\mathcal{SA}$ *and* $\mathcal{I}$ satisfies for all $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$ and $i \in I_t$ the following conditions

(i) $(\Omega_i, \mathcal{F}_i, P_i)$ is a probability space.

(ii) If t_1, t_2, $t_3 \in [t, t_{\mathrm{end}}]$, with $t_1 \le t_2 \le t_3$, and $\pi \in S_{\mathrm{pr}(i)}$, then

$$f_i(t, -; \mathrm{n}_i\, x, t_1, t_2, t_3, \pi) : (\Omega_i, \mathcal{F}_i) \longrightarrow \left(S_{\mathrm{pa}(i)}, \mathfrak{S}_{\mathrm{pa}(i)}\right) \text{ is measurable.}$$

(iii) If $\mathrm{ac}^i_{\mathrm{pa}(i)}(t) = 1$, then there exist $\omega \in \Omega_i$, t_1, t_2, $t_3 \in [t, t_{\mathrm{end}}]$, with $t_1 \le t_2 \le t_3$, and $\pi \in S_{\mathrm{pr}(i)}$ such that $t_2 \le t < t_3$

$$x_t(\mathrm{pa}(i)) = f_i(t, \omega; \mathrm{n}_i\, x, t_1, t_2, t_3, \pi). \tag{2.6}$$

Moreover

$$\begin{aligned} &\mathrm{ac}^i_{\mathrm{pr}(i)}(t_1) = 1 \text{ and } \exists a \in \mathrm{ag}(i) : \ \mathrm{ac}^i_a(t_1) = 1 \\ &\mathrm{ac}^i_{\mathrm{pa}(i)}(t_2) = 1 \\ &\forall e \in E_i : \ \mathrm{ac}^i_e(t_3) = 0. \end{aligned} \tag{2.7}$$

Remark 7.

(a) Condition (2.6) clarifies mathematically the intuition about the state map x, which results as a possible path of the state variables of our system. In other words: by running a simulation of the system which follows the algorithm presented in Sec. 2.2, we obtain an outcome of the state variables which is analog to x; again: we could say that the state maps x are the unknown of our CS that we need to compute and study. Therefore, in (2.6) we can think that (t_1, t_2, t_3) have been extracted using the distribution $T_i(-; \mathrm{n}_i\, x, t)$, and π has been extracted using the distribution $G_i(-; \mathrm{n}_i\, x, t)$.

(b) Conditions (2.7) state formally the dynamics of the activation function: at the starting time t_1 at least one agent is active and the propagator $\mathrm{pr}(i)$ is activated; at the arrival time t_2 of the propagator, the patient $\mathrm{pa}(i)$ is activated. Since in (iii) we started by assuming that $\mathrm{ac}^i_{\mathrm{pa}(i)}(t) = 1$ this means that $t_2 \le t < t_3$. When the interaction stops, we are at time t_3 and every entity involved with the interaction i is deactivated.

To illustrate this concept, we can consider the following simple examples:

1. a person a is throwing a stone: the propagator $r = a$ is the same agent carrying in its state the information of the initial velocity $\vec{v}_0$ and position x_0 (hence in this example we have $\pi = (x_0, \vec{v}_0)$); we have $t_1 = t_2$ and, e.g., t_3 when the stone reaches the ground; in the time interval $[t_1, t_3)$ the transition function is of the form $f_i(t; x, t_1, t_1, t_3, \pi) = f_i(t; t_1, x_0, \vec{v}_0)$ and gives the deterministic dynamics of the stone (t_3 is uniquely determined by the other variables). In this case we have a trivial space for stochastic internal evolution, i.e. $|\Omega_i| = 1$.
2. In a more realistic modeling, we can consider the initial condition $(x_0, \vec{v}_0)$ distributed as a 6-dimensional normal distribution, so that we have $S_r = \mathbb{R}^6$ and $G_i(-; x, t) : \mathcal{B}(\mathbb{R}^6) \longrightarrow [0, 1]$ is this normal distribution ($\mathcal{B}(\mathbb{R}^6)$ is the sigma algebra generated by open sets of $\mathbb{R}^6$). The transition function $f_i(t; t_1, x_0, \vec{v}_0) = x_0 + \vec{v}_0(t - t_1) + \frac{1}{2}\vec{g}(t - t_1)^2$ becomes a random variable in the probability space

$(S_r, \mathcal{B}(\mathbb{R}^6), G_i(-; x, t))$ if we thought it as a function of $\pi = (x_0, \vec{v}_0) \in S_r$. Once again the space Ω_i is trivial.

3. Let us consider a pedestrian p that receives at time t_2 a signal r from a source a, and starts to move in the direction $\vec{\pi} \in \mathbb{R}^3$, $|\vec{\pi}| = 1$, with a certain stochastic deviation, both in the direction and in the magnitude of the velocity. We will have $f_i(t, \omega; t_1, t_2, t_3, \vec{\pi}) = x_0 + \vec{v}(\omega) \cdot (t - t_2)$, where x_0 is the position of p at time t_2 and where the expected value of $\vec{v}$ in the space $(\Omega_i, \mathcal{F}_i, P_i)$ is $E(\vec{v}) = v_0 \cdot \vec{\pi}$; both v_0 and x_0 are taken from the state space S_p of the patient p.

2.3.5. *Interaction spaces and classical models for complex systems.*

Definition 8. An *interaction space* $\mathfrak{I} = (\mathcal{EI}, \mathcal{SA}, \mathcal{I}, \mathcal{TF})$ is given by considering all the previously defined systems:

(i) A system of entities and interactions $\mathcal{EI} = (E, t_{\mathrm{st}}, t_{\mathrm{end}}, \mathcal{T}, I)$.
(ii) A system of state spaces and activation maps $\mathcal{SA} = (\mathrm{ac}, S, \mathfrak{S}, x)$ *for* $\mathcal{EI}$.
(iii) A system of information $\mathcal{I} = (\mathcal{U}, T, G)$ for the interactions of $\mathcal{EI}$ and $\mathcal{SA}$.
(iv) A system of transition functions $\mathcal{TF} = (f, \Omega, \mathcal{F}, P)$ for $\mathcal{EI}$, $\mathcal{SA}$ and $\mathcal{I}$.

We first study the case of Markovian IS, and subsequently show that classical models of CS are all included as particular cases of IS.

3. Classical models for complex systems as interaction spaces

We can now explain how classical models for complex system can be embedded as interaction spaces. Of course, IS are not trivial mathematical structures, i.e. these embeddings are injective: e.g. if two cellular automata are equal when viewed as IS, then they are necessarily equal as cellular automata.

Cellular automata: Cells with their state space are the interacting entities; depending on the type of cellular automaton, we can have either local or global interactions. Every local interaction has the same type of neighborhood, which corresponds to that of the cell. In every interaction, there is only one agent and one patient corresponding to the cell on which the interaction acts. The dynamics is synchronous, and hence propagator entities are useless, exactly like starting times, speed and duration, which are trivially determined by the synchronous dynamics. Transition functions correspond to the mathematical functions that define the state change of the cellular automaton. We can either consider every cell as always active, or as activated every time an interaction changes its state. Interactions may also depend on a suitable space of resources and goods (see e.g. [54]).

Agent based models: For ABM, we refer to the mathematical definition given by [59]. Although agents naturally correspond to interacting entities, we have to consider that frequently ABM are identified with the corresponding implementation in an object oriented programming language, and the corresponding mathematical formalization is not always considered. In that case, the state space of an agent can also include its behavioral rules or methods. Since IS are a mathematical theory, such methods have to be associated to a corresponding mathematical function, but this is clearly always possible. In the language of IS, these are exactly the transition functions. The environment itself has to be considered as an interacting entity. Neighborhoods are defined by all the entities (agents and environment) from which the methods (interactions) take the information they need to operate. In case another agent is not contained in this neighborhood, it is completely hidden for the interactions of the considered agent. Since an interaction is of the form $i = \left(a_1, \ldots, a_n \xrightarrow{r,\alpha} p\right)$, in general interactions are only of local nature, depending on the agents $a_1, \ldots, a_n, p$, on r, α, and on the cause-effect relations with other agents. The dynamics is naturally asynchronous, depending on the signals r sent between agents: these corresponds to propagator entities whose speed has to be modeled only in particular cases.

Master equation based models: Under the assumptions (I), (II), (III), (IV) we have seen in the previous section, the dynamics of a Markovian IS is described by a master equation. This case includes several models used in synergetics. Usually, additional structures such as propagators, starting time, speed, duration, neighborhoods, etc. are not used in these descriptions.

Networked dynamical systems: In this case, interacting entities are the nodes of the network with their state space. The network can be easily formalized considering $\{0, 1\}$-valued interactions between nodes and corresponding to the adjacency matrix (their transition function can be trivialized considering e.g. constant functions equal to the entry of the adjacency matrix). Each node interacts only with adjacent nodes (which is hence the neighborhood) and the transition functions correspond to the functions that update the state of each node, and can hence come, e.g., from the solution of suitable differential equations. As it is well known, the update algorithm can be synchronous or asynchronous, and as such it has to be implemented as an IS.

Neural networks: As in the previous case, interacting entities are the neurons of the network. The structure of the network and the neighborhoods can be formalized, in the language of IS, using the adjacency matrix as above. The state of each neuron includes the values of the input variables, the bias for each one of these inputs, the activation function, the property of being a start or an end node. State space of propagators include the weights of the links. State space of neurons could also include the property of being active with respect to a change of the inputs or a change of the weights associated to its input links. The most important interactions depend on the type of learning and, in general, have propagators (and hence weights) as patient entities. In case of supervised or reinforcement learning, the pairs of examples can be seen as stochastic resources of suitable propagators. We can also have neurons as patients if the learning algorithm changes their activation functions. The dynamics is in general synchronous. If one is interested in computation times of activation functions, propagators speed and times can also be considered.

Evolutionary algorithms: The population of candidate solutions (phenotypes) with their state space (genotype) are the interacting entities. Stochastic interactions are clearly the core of these models. Mutation, crossover, inversion and selection operators can be easily implemented as fitness depending interactions of an IS. The algorithm is synchronous, but asynchronous versions can also be implemented, e.g. by considering more fitted populations as single interacting entities that spread out their genetic code over the entire set of interacting entities. In this case, propagators can be considered, with their speeds and times. In this generalization, the introduction of suitable neighborhoods is also relevant.

It is also clear that these ways to include these classical models as IS are sufficiently detailed to allow a reconstruction of the initial model from the corresponding IS. In other words, they are embedded as IS.

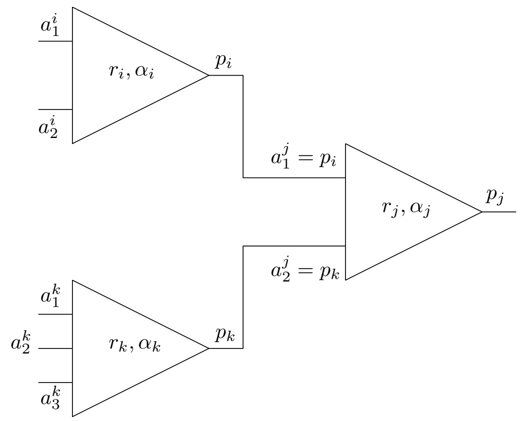

FIGURE 4.1. An example of interactions in cause-effect cascade

4. The Markovian case: master equation

There are only six (possible) sources of randomness acting in an IS $\mathfrak{I}$. The first one can derive from the stochastic evolution of (2.6). We have this type of stochastic behavior because we do not want to model some details of the evolution of the interaction i, or because it is not possible at all (it can happens, e.g., that a pedestrian randomly chooses between two exists which are equally located with respect to its present position). Three sources of randomness are times t_1, t_2, t_3. The fifth source of randomness is due to extraction of resources. The sixth and last possible stochasticity can happen when two or more interactions occur at the same time and involve the same patient: this source is hidden in condition (2.4). It is therefore possible to define a new probability space that combines all these random sources. The elementary events of this space are of the form $(t_1, t_2, t_3, \pi, \omega, \gamma)$, where $(t_1, t_2, t_3) \in \left\{(t_1, t_2, t_3) \in [t, t_{\mathrm{end}}]^3 \mid t_1 \le t_2 \le t_3\right\}$ are distributed as $T_i(-; \mathrm{n}_i\, x, t)$, $\pi \in S_{\mathrm{pr}(i)}$ is distributed as $G_i(-; \mathrm{n}_i\, x, t)$, $\omega \in \Omega_i$ is distributed as P_i, and $\gamma \in I$ is the output of the choice between interactions that occur at the same time on the same patient. All these quantities have to be considered for all interactions that can be active along a possible dynamical state path x of the system $\mathfrak{I}$. The joint probability that we have to settle on this space of elementary events is certainly not easy. On the other hand, the causal graph (see e.g. Fig. 2.2) defined by the activation function, can be of great help in finding this probability: we can say that the interaction i is a cause of the interaction j if along any possible path x the interaction i always s previous to those of j, and if i activates an agent of j (see e.g. Fig. 4.1).

If the resulting directed cause-effect graph is acyclic, we can interpret it as a beliefs network and apply the methods of Bayesian networks to define the joint probability, see e.g. [16].

For the interaction space $\mathfrak{I}$, we use the same notations introduced above. The *global state space* of $\mathfrak{I}$ is the set

$$M := \prod_{\substack{t \in [t_{\mathrm{st}}, t_{\mathrm{end}}] \\ e \in E_t}} S_e,$$

i.e. the product of all the (measurable) space states S_e of all the active entities $e \in E_t$ at some time $t \in [t_{\rm st}, t_{\rm end}]$. Note that an element of M is a map of the form $(e,t) \mapsto x_t(e) \in S_e$ for all $t \in [t_{\rm st} t_{\rm end}]$ and all $e \in E_t$. We also set

$$M_t := \prod_{e \in E_t} S_e,$$

which is the state of the system $\mathfrak{I}$ at time t. Any $\sigma \in M_t$ is a map of the form $e \in E_t \mapsto \sigma_e \in S_e$.

On the basis of what we said above, we can assume to have a probability space $(\Omega^{\rm g}, \mathcal{F}^{\rm g}, P^{\rm g})$ (the superscript "g" stands for *global*) and a random variable $X : \Omega^{\rm g} \longrightarrow M$ representing the possible state maps of $\mathfrak{I}$. Mathematically, this means that for every elementary event $w \in \Omega^{\rm g}$, the map $(e,t) \mapsto X(w)_t(e) \in S_e$ is a possible state map of the IS $\mathfrak{I}$. In other words, for any fixed $w \in \Omega^{\rm g}$, if we replace everywhere the state map $(e,t) \mapsto x_t(e) \in S_e$ with the map $(e,t) \mapsto X(w)_t(e) \in S_e$, all the conditions in the definition of the IS $\mathfrak{I}$ are still satisfied.

We now assume that:

(I) All the entity $e \in E$ are always active

$$\forall t \in [t_{\rm st}, t_{\rm end}]\, \exists i \in I : \ \mathrm{ac}_e^i(t) = 1, \tag{4.1}$$

i.e. $E = E_t$ for all time t. By (4.1), it follows that $M \simeq M_t = \prod_{e \in E} S_e$. We can hence think $M = \prod_{e \in E} S_e$.

(II) The IS $\mathfrak{I}$ is Markovian, i.e. its future history is completely determined by its present state. In our case, the latter assumption corresponds to assume that

$$\forall t \in [t_{\rm st}, t_{\rm end}]\, \forall \sigma \in M : \ P^{\rm g}\left[\forall e \in E : \ X_t(e) = \sigma_e\right] > 0. \tag{4.2}$$

(III) The global state space M is at most countable.

Condition (4.2) yields the possibility to define

Definition 9. Let t, $\tau \in [t_{\rm st}, t_{\rm end}]$, $\sigma \in M$, $\mu \in M$.

(i) We use the following notation for the event

$$\{\omega \in \Omega^{\rm g} \mid \forall e \in E_t : \ X(\omega)_t(e) = \sigma_e\} = [\forall e \in E_t : \ X_t(e) = \sigma_e] =: [X_t = \sigma] \in \mathcal{F}^{\rm g}.$$

(ii) $p(\sigma, t) := P^{\rm g}[X_t = \sigma]$ is the probability that the system $\mathfrak{I}$ is in the global state σ at time t.

(iii) The quantity

$$p(\mu, \tau \mid \sigma, t) := \frac{P^{\rm g}[X_\tau = \mu, \ X_t = \sigma]}{P^{\rm g}[X_t = \sigma]}$$

is the conditional probability that the system $\mathfrak{I}$ is in the global state μ at time τ given that the system was in the global state σ at time t (Markovian assumption).

Lemma 10. *Let t, $\tau \in [t_{\rm st}, t_{\rm end}]$ and $\mu \in M$. Then*

$$p(\mu, \tau) = \sum_{\sigma \in M} p(\mu, \tau \mid \sigma, t) \cdot p(\sigma, t).$$

Proof. Def. 9.(iii) yields $p(\mu,\tau \mid \sigma,t)\cdot p(\sigma,t) = P^{\mathrm{g}}\left[X_\tau=\mu,\ X_t=\sigma\right]$, and hence (note that M is at most countable)

$$\sum_{\sigma\in M} p(\mu,\tau \mid \sigma,t)\cdot p(\sigma,t) = \sum_{\sigma\in M} P^{\mathrm{g}}\left[X_\tau=\mu,\ X_t=\sigma\right] = \\ = \sum_{\sigma\in M} P^{\mathrm{g}}\left(\left[X_\tau=\mu\right]\cap\left[X_t=\sigma\right]\right).$$

The family of events $\sigma\in M\mapsto [X_\tau=\mu]\cap[X_t=\sigma]$ is pairwise disjoint, so

$$\sum_{\sigma\in M} p(\mu,\tau \mid \sigma,t)\cdot p(\sigma,t) = P^{\mathrm{g}}\left(\left[X_\tau=\mu\right]\cap\bigcup_{\sigma\in M}\left[X_t=\sigma\right]\right). \tag{4.3}$$

But for all $w\in\Omega^{\mathrm{g}}$ we have $X(w)\in M$, so $w\in[X_t=X(w)_t(-)]$ and hence $\Omega^{\mathrm{g}}=\bigcup_{\sigma\in M}[X_t=\sigma]$. This and (4.3) yield the conclusion. Note that if M is countable, the convergence of the series follows from (4.3) and the sigma-additivity of P^{g}. □

We now assume that

(IV) For all σ, $\mu\in M$ and all $t\in[t_{\mathrm{st}},t_{\mathrm{end}}]$, the function

$$\tau\in[t_{\mathrm{st}},t_{\mathrm{end}}]\mapsto p(\mu,\tau\mid\sigma,t)\in[0,1]$$

is differentiable at $\tau=t$.

This allows the following

Definition 11. Let σ, $\mu\in M$ and $t\in[t_{\mathrm{st}},t_{\mathrm{end}}]$, then the derivative

$$w_t(\mu,\sigma) := \frac{\partial p(\mu,\tau\mid\sigma,t)}{\partial\tau}(t)$$

is called *rate of transition from σ to μ at time t.*

Theorem 12 (Master equation)**.** *If (I), (II), (III), (IV) hold, and if*

$$\sum_{\mu\in M}\left|w_t(\sigma,\mu)\right| < +\infty, \tag{4.4}$$

then for all $t\in[t_{\mathrm{st}},t_{\mathrm{end}}]$ and all $\sigma\in M$

$$\frac{\partial p}{\partial t}(\sigma,t) = \sum_{\mu\in M} w_t(\sigma,\mu)\cdot p(\mu,t) = \sum_{\substack{\mu\in M\\ \mu\neq\sigma}}\left[w_t(\sigma,\mu)\cdot p(\mu,t) - w_t(\mu,\sigma)\cdot p(\sigma,t)\right]. \tag{4.5}$$

Proof. Using Lem. 10, for all $\tau\in[t_{\mathrm{st}},t_{\mathrm{end}}]$, we have

$$\sum_{\mu\in M} p(\mu,\tau\mid\sigma,t) = \frac{1}{P^{\mathrm{g}}\left[X_t=\sigma\right]}\cdot\sum_{\mu\in M} P^{\mathrm{g}}\left[X_\tau=\mu,\ X_t=\sigma\right] = \\ = \frac{1}{P^{\mathrm{g}}\left[X_t=\sigma\right]}\cdot\sum_{\mu\in M} p(\sigma,t\mid\mu,\tau)\cdot p(\mu,\tau) = \\ = \frac{1}{P^{\mathrm{g}}\left[X_t=\sigma\right]}\cdot p(\sigma,t) = 1. \tag{4.6}$$

This result implies that

$$\sum_{\mu\in M}\frac{p(\mu,t+h\mid\sigma,t)-p(\mu,t\mid\sigma,t)}{h}\cdot p(\sigma,t)=$$

$$=\frac{p(\sigma,t)}{h}\left[\sum_{\mu\in M}p(\mu,t+h\mid\sigma,t)-\sum_{\mu\in M}p(\mu,t\mid\sigma,t)\right]=0. \quad (4.7)$$

Now, using Lem. 10, we have

$$\sum_{\mu\in M}\frac{p(\sigma,t+h\mid\mu,t)-p(\sigma,t\mid\mu,t)}{h}\cdot p(\mu,t)=\frac{1}{h}\sum_{\mu\in M}p(\sigma,t+h\mid\mu,t)\cdot p(\mu,t)-$$

$$-\frac{1}{h}\sum_{\mu\in M}p(\sigma,t\mid\mu,t)\cdot p(\mu,t)=\frac{1}{h}\left[p(\sigma,t+h)-p(\sigma,t)\right],$$

Taking the limit for $h\to 0$ in this equality, we obtain

$$\frac{\partial p}{\partial t}(\sigma,t)=\lim_{h\to 0}\sum_{\mu\in M}\frac{p(\sigma,t+h\mid\mu,t)-p(\sigma,t\mid\mu,t)}{h}\cdot p(\mu,t). \quad (4.8)$$

If M is countable, assumption (4.4) implies that we can exchange limits and series in this equality. In fact, if $\{\mu_n\mid n\in\mathbb{N}\}=M$ is an enumeration of M, then e.g. $\left|\frac{p(\sigma,t+h|\mu_n,t)-p(\sigma,t|\mu_n,t)}{h}\right|\le|w_t(\sigma,\mu_n)|+\frac{1}{n^2}$ for all h sufficiently small. We can hence apply Lebesgue's dominated convergence theorem. Therefore, (4.8) yields the first part of (4.5). Considering (4.7), we also get

$$\frac{\partial p}{\partial t}(\sigma,t)=\sum_{\mu\in M}\left[w_t(\sigma,\mu)\cdot p(\mu,t)-w_t(\mu,\sigma)\cdot p(\sigma,t)\right],$$

which gives the second part of (4.5) because the summand is zero for $\mu=\sigma$. □

We can already derive an immediate consequence of the inclusion of classical model for CS into IS: if anyone of these modeling systems satisfies assumptions (I), (II), (III), (IV), then a master equation (4.5) holds.

5. The non-Markovian case: mean derivative equations

to adapt and generalize from: `http://www.mat.univie.ac.at/~giordap7/MDE.pdf`

6. Complex adaptive systems following Zipf's idea

It is widely recognized that complex adaptive systems (CAS) are among the most interesting types of CS (see, e.g., [36]). There are several attempts to provide a precise definition of CAS, [38, 3, 24, 61]. Due to its unifying properties, IS theory can provide the appropriate context for a general mathematical definition of CAS, and the understanding of general dynamical laws governing these systems. In particular, we want to try a formalization of the original ideas that G.K. Zipf expressed in [63]. In fact, these ideas seem intuitively clear, meaningfully and applicable to a large class of CS. In [63], the appearing of a power law is connected to what we think is an adapting behavior of the system. Zipf's *principle of least effort* intuitively explains this adaptation as a result of two opposing processes: *unification* and *diversification*. We will try to explain this principle through several examples listed below. Clearly both these terms are sufficiently imprecise and could hence be misinterpreted. The subsequent mathematical formalization we will present also aims to gain a more clear understanding of these processes. In our insight, unification processes are related to a decreasing in convenient *costs* (possibly meant in an abstract way: e.g. loss of profits, loss of common goods, probability to get hungry, probability to loose reproductive possibilities, etc.), whereas diversification processes are linked to *long term changes* of suitable interaction, i.e. to the increasing of the *information entropy of the goods* generated by these interactions. It is the implementation of these interactions and the most diversified exchange of fluxes of goods that enable the population to be resilient and keep a low value of costs.

To start a first understanding and a validation of the subsequent mathematical definition, we can keep in mind the following examples:

1. In a natural language, unification processes drift to shorten most frequently used words (or better: frequently used sounds, see [6]); diversification ones make evolve the language towards longer and specialized words, [13, 63].
2. In cities and their markets development, unification tends to bring near people so as to decrease suitable costs of living; diversification tends to use all the possible living locations so as to approach the appropriate rent costs, [21].
3. In natural selection, unification forces push giraffes to search for eatable trees (see e.g. [7] and references therein); diversification selects all the best genetic codes that allow for a longer neck, [15]. We recall that a costs decreasing process (unification) can cause an evolution into different phenotypes (diversification), see [26]. More generally, natural selection seems to result by the dynamics of two processes: representation of biological information as chemical properties and control flow (diversification), and the energetic constraints limiting the maintenance of that information (unification), [48, 49, 50]. See [47] for an alternative evolutionary explanation of long neck in giraffes.
4. Determining the direction to navigate to a safe place, such as a home or nest, is a fundamental behavior for all complex animals and a crucial first step in navigation. We can say that unification processes evaluate costs related to the achievement of these goals, whereas diversification ones are related to judgments based on a previously learned or planned behavior, [9].
5. Companies with a longer life span are able not only to decrease costs and increase profits (unification), but are also able to adapt to their complex environment by implementing long-term robustness. The latter are often realized through diversification processes such as: maintain heterogeneity of people, ideas, and endeavors, and preserve redundancy among components, [43].
6. In autism (but also in schizophrenia), a necessary decreasing in high costs related to social interactions (unification) is frequently compensated by higher abilities (diversification) in very specialized or creative activities, [40, 11].
7. The ability to manage costs related to large varieties of goods (unification) is related to the ability to implement the same stable economic decisions (interactions) applicable to different goods (diversification), [60].
8. Phyllotaxis, the regular arrangement of leaves or flowers around a plant stem, is an example of developmental pattern formation. Phyllotaxis is characterized by the divergence angles between the organs, the most common angle being 137.5°, the golden angle. Different approaches and hypotheses has been used to model

this formation mechanism, see e.g. [31]. In this process, we can see unification forces related to energy exploitation by each primordium, and diversification forces that tend to uniformly distribute these energy sources between old and new primordia.

9. An example of **non**-adaptive but still complex system is traders payments of Wall Street employees. It is well known that this payment has a base salary and a bonus, which is usually a percentage of trader's profit. If traders lose, they still get their base, and only if their loss is great enough, they are fired. However, they never have to return the money lost by the company due to their wrong trading. This is a clear financial incentive to be reckless because it rewards short-term gains (costs which can be identified as unification forces) without regard to long-term consequences (diversification forces), [1].
10. The efficiency of a parliament (unification processes interpreted as decreasing of suitable costs) can be improved by inserting randomly selected legislators (increasing of diversification among legislators), see [41].
11. Whereas classical economic theories prescribe specialization of countries industrial production, inspection of the country databases of exported products shows that this is not the case: successful countries are extremely diversified, in analogy with biosystems evolving in a competitive dynamical environment. In fact, together with classical and necessary costs reduction (unification), diversification represents the hidden potential for development and growth, [52].
12. The evolutionary emergence of an egalitarian attitude in a population can be explained by using an evolutionary model of group-living individuals competing for resources and reproductive success (i.e. unifications as costs to decrease), see [22]. Although the differences in fighting abilities lead to the emergence of hierarchies where stronger individuals take away resources from weaker individuals, the logic of within-group competition implies that each individual benefits if the transfer of resources from a weaker group member to a stronger one is prevented. This model shows that this effect can result in the evolution of a particular behavior causing individuals to interfere in a bully–victim conflict on the side of the victim. A necessary condition in this process is a high efficiency of coalitions in conflicts against the bullies. The egalitarian drive leads to a dramatic reduction in within-group inequality. Simultaneously it creates the conditions for the emergence of inequity aversion via the internalization of these adapted behavioral rules. All these interactions are general because they can be applied to different situations. They hence represent a long-term and diversified improving of the society.
13. The present network of financial exposures among institutions (e.g. banks, companies) shows that this system is **not** well adapted. The centrality of certain institutions does not allow the system to be resilient to financial fails of few institutions, see [4]. As we will see later, even if in an idealized system such as that of Von Thünen's market, this is the same mechanism showed by monopolies and the corresponding lack of diversity in the system. We could say that the system is flawed because it lacks in interactions that act when a node of the financial network fail: a more diversified network would show a greater resilience because of a lower centrality of these nodes. It is also clear that a globally directed taxation system can prevent the formation of such a centralized nodes. Even if these global taxes do not allow a maximization of profits (which could be clearly

higher without this tax), they could be rightly justified as a measure to prevent global irreparable problems to the entire system, and hence to the institutions themselves. See also [56].

To underline that we are going to give our interpretation of Zipf's principle of least effort, we prefer to name our mathematical version *generalized evolution principle* (GEP). This name has also the merit to link this adapting dynamics to evolutionary theories that, in our opinion, are well inscribed into it.

As above, we start with an intuitive description of the GEP:

1. In an IS we can have several populations of interacting entities. Since only some of these populations have to be described as adapting, we have to talk of the *adaptation of a given* subset (=: *population*) *of interacting entities*. In other words, it is not necessarily correct to talk of an entire "complex adaptive IS".
2. Therefore, in an adaptation process, we need to identify an adapted population $\mathcal{P}$ of interacting entities, $\mathcal{P} \subseteq E$, and an interaction $i \in I$ performed by the population $\mathcal{P}$ and representing how $\mathcal{P}$ is going to *stabilize the adaptation process through a suitable diversification of its resources*. An interaction performed by $\mathcal{P}$ is defined as follows:

Definition 13. Let $\mathcal{P} \subseteq E$, we say that $i \in I$ is *an interaction of (the population)* $\mathcal{P}$ if we can find interaction $i_1, \dots, i_h$ such that

(i) At least one agents of each interaction i_h is a member of the population $\mathcal{P}$: $\forall k = 1, \dots, h \, \exists a \in \mathrm{ag}(i_k) \cap \mathcal{P}$

(ii) The resource space of i is the product of the resources of all the i_k

$$(S_{\mathrm{pr}(i)}, \mathfrak{S}_{\mathrm{pr}(i)}) = \prod_{k=1}^{h} (S_{\mathrm{pr}(i_k)}, \mathfrak{S}_{\mathrm{pr}(i_k)}).$$

(iii) For all $x \in M$ and $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$, if $\Pi(-; \mathrm{n}_{i_k} x, t)$ is distributed as $G_{i_k}(-; \mathrm{n}_{i_k} x, t)$, then $\Pi(-; \mathrm{n}_i x, t)$ is distributed as the join probability of $(\Pi(-; \mathrm{n}_{i_k} x, t))_{k=1}^{h}$.

3. Together with the interaction i, we have to identify a state functions $C_i : \prod_{e \in \mathcal{P}} S_e \longrightarrow \mathbb{R}_{\geq 0}$ called *cost associated to* i: *a CAS reacts to an increase in this costs* C_i *by changing state and trying to decrease* C_i. The interactions that allow a decreasing of the cost function represent the unification processes. Therefore, we are going to define when a CS is adaptive *with respect to the given cost functions* C_i.
4. In most cases C_i is computed by averaging costs of the same type undergone by single interacting entities, see Sec. 6.2, 7. For this reason, to verify an adaptation process, we also need to *identify a probability* P_{C_i} *on the global state space* $\left(\prod_{e \in \mathcal{P}} S_e, \prod_{e \in \mathcal{P}} \mathfrak{S}_e\right) =: (S_{\mathcal{P}}, \mathfrak{S}_{\mathcal{P}})$ *of* $\mathcal{P}$ *to evaluate the expected value of the cost function* C_i.
5. We precisely define when it happen that *at state* y *and time* t *the population* $\mathcal{P}$ *is better adapted than in the state* x *at time* s. Therefore, the population $\mathcal{P}$ must be active both at time t and s: $\mathcal{P} \subseteq E_t \cap E_s$. A particular case will be when y at time t is one of the *best* possible states, which usually corresponds to a steady or an equilibrium state.
6. The first important condition, which intuitively is a consequence of the unification interactions in passing from the state x at time s to y at time t, is that t*he expected values of the cost function* C_i *is decreasing by passing from* x *at time* s

to y at time t, i.e.

$$E\left(C_i(y_t|_{\mathcal{P}})\right) \le E\left(C_i(x_s|_{\mathcal{P}})\right).$$

We note that $x \in M = \prod_{\substack{t\in[t_{\text{st}},t_{\text{end}}]\\ e\in E_t}} S_e$ and hence the evaluation $x_s : e \in E_s \mapsto x_s(e) \in S_e$ at time s of the state function x can be restricted to act only on $\mathcal{P} \subseteq E_s$: $x_s|_{\mathcal{P}} : e \in \mathcal{P} \mapsto x_s(e) \in S_e$. Analogously for y.

7. The other important element of the adaptation process is the interaction $i \in I_t \cap I_s$ associated to the cost function C_i. The interaction i is active at both times and realizes the diversification to a possible increasing of C_i, and hence to stabilize this decreasing of costs. We measure the diversification "forces" with the *information entropy of the fluxes of goods extracted by population interaction i from its resource spaces*.
   (i) We therefore need to finally identify suitable probabilities Q_i to evaluate the corresponding forces of diversification D_i as: if $\Pi_i(-;\mathrm{n}_i\, y,t)$ is a random variable distributed as $G_i(-;\mathrm{n}_i\, y,t)$, then

$$D_i(\mathrm{n}_i\, y,t) := \text{Entropy}\left(\Pi_i(-;\mathrm{n}_i\, y,t)\right) := E\left(-\ln\left(P_i\left(\Pi_i(-;\mathrm{n}_i\, y,t)\right)\right)\right). \tag{6.1}$$

   Analogously for the state x at time s. Therefore, the map Q_i must be a probability on the measurable state space $(S_{\text{pr}(i)}, \mathfrak{S}_{\text{pr}(i)})$ and the expected value in (6.1) is computed in the probability space $(Q_i, S_{\text{pr}(i)}, \mathfrak{S}_{\text{pr}(i)})$. Finally, in (6.1), the symbol $P_i(X)$ denotes the probability mass function of the random variable $X : S_{\text{pr}(i)} \longrightarrow \mathbb{R}$ with respect to the probability Q_i.
8. The second important condition states that the *diversification forces are greater* in y at time t with respect to x at time s:

$$D_i(\mathrm{n}_i\, y,t) \ge D_i(\mathrm{n}_i\, x,s) \quad \forall i \in \mathcal{I}.$$

Definition 14 (Generalized evolution principle, CAS)**.** Let $\mathfrak{I}$ be an IS, and let s, $t \in [t_{\text{st}}, t_{\text{end}}]$, $\mathcal{P} \subseteq E_s \cap E_t$, x, $y \in M$, $i \in I_s \cap I_t$. Then we say that *at y, t the population $\mathcal{P}$ is better adapted than at x, s with respect to C_i, P_{C_i}, Q_i (briefly: $\mathcal{P}$ is a CAS) if*

(i) $C_i : S_{\mathcal{P}} \longrightarrow \mathbb{R}_{\ge 0}$ is a random variable.
(ii) P_{C_i} is a probability on the state space $(S_{\mathcal{P}}, \mathfrak{S}_{\mathcal{P}})$ of the population $\mathcal{P}$.
(iii) i is an interaction of $\mathcal{P}$.
(iv) Q_i is a probability on the resource space $(S_{\text{pr}(i)}, \mathfrak{S}_{\text{pr}(i)})$.
(v) We have

$$E\left(C_i(y_t|_{\mathcal{P}})\right) \le E\left(C_i(x_s|_{\mathcal{P}})\right),$$
$$D_i(\mathrm{n}_i\, y,t) \ge D_i(\mathrm{n}_i\, x,s),$$

where D_i are defined by (6.1) and the expected value $E(-)$ is computed using P_{C_i}.
(vi) We finally say that *y is a best possible state at time t for $\mathcal{P}$* if for all state x and time s, at y, t the population $\mathcal{P}$ is better adapted than in x, s (with respect to C_i, P_{C_i}, Q_i).

In other words, by definition, we say that the population $\mathcal{P}$ in the IS $\mathfrak{I}$ adapts to a possible increase in cost C_i if the entities of $\mathcal{P}$ are able to change state x and to *decrease* the *average* value of the state function C_i; moreover, this adapting reaction also aims to *increase* the information entropy (called *force of diversification D_i*) of the goods produced by the interaction $i \in \mathcal{I}$. If the cost function C_i is constant and y is a best possible state, we hence get the classical entropy maximization principle for the probabilities Q_i on the space of resources of the interaction $i \in \mathcal{I}$. On the other hand, if we trivialize the interaction i by choosing constant resources (deterministic extraction of goods), then the population is adapting if it minimizes the cost function C_i. Even if this means that the GEP includes both the entropy maximization principle and minimization algorithms, more interesting possibilities necessarily originate from meaningful modeling choices, where the relations between cost C_i and the interaction of the population $\mathcal{P}$ is clear and it is not trivial. This further clarifies that both the cost function C_i and the population interaction i are modeling choices, and hence they need to be validated.

Because the expected costs $E(C_i(y_t|_{\mathcal{P}}))$ are calculated with a probability distribution supported over the global state of the population $\mathcal{P}$, we can interpret the quantity $E(C_i(y_t|_{\mathcal{P}}))$ as inversely proportional to a quantification of suitable *common goods for the population* $\mathcal{P}$. As a consequence of the maximization properties of Shannon's entropy, the forces of diversification D_i can be interpreted as a *gauge of long-term changes* (with respect to the given cost function C_i).

Clearly, we may have a step from a state (x, s) to a better (y, t) even if there does not exists a best possible state. An intuitive example that seems to satisfy this property is Darwinian evolution. Think at giraffes and their elongation of neck: the cost are related to the probability of finding leaves to eat; we have at least two interacting entities: giraffes and trees, but only one is adapting with respect to given costs; the population interaction related to the force of diversification is to have a genetic code that causes a neck elongation. It seems that there is no a maximum length of neck minimizing this cost, even if such a maximum is reached due to the increasing of other costs. Assuming costs related to the probability to win in a sexual competition, we have a model for the hypothesis expressed in [47].

We now prove, in a more abstract setting and under mild conditions, that in every CAS $\mathcal{P}$ the probabilities Q_i used to compute the forces of diversification satisfy a power law when the population is at a best possible state, i.e. when costs are minimum and diversification is maximum. We then give some examples of cost functions in Sec. 6.2, in particular in a generalized Von Thünen's model in Sec.7.

topics to write:

- first connections with other definitions of CAS? Is the GEP a different concept?

6.1. **Power law distribution following Mandelbrot's idea.**

- Topics to write:
  - power laws in linguistics, economics, biology, information sciences, etc.
  - other explanations of power laws

This section follows the classical ideas of B. Mandelbrot presented in [33]. We imagine to have a population $\mathcal{P}$ whose state is described by a vector $x \in S_{\mathcal{P}} \subseteq \mathbb{R}^n$. The systems has to be thought of as a CAS that changes its state so as to decrease a suitable cost function $C : S_{\mathcal{P}} \longrightarrow \mathbb{R}_{>0}$ and, at the same time, to increase a corresponding information entropy (force of diversification):

$$D(x) = -\sum_{j=1}^{d} q_j(x_j) \cdot \log_2 q_j(x_j) > 0 \qquad \forall x \in S_{\mathcal{P}}.$$

More precisely, the system adapts (e.g. it evolves) so as to minimize the ratio $\frac{C}{D}$ at the interior point $y \in S_{\mathcal{P}}$:

$$\forall x \in S_{\mathcal{P}} : \ 0 < \frac{C(y)}{D(y)} \leq \frac{C(x)}{D(x)}.$$

By assumption, the probabilities always enter into the global state of the system

$$q_j(x_j) = x_j \qquad \forall j = 1, \ldots, d,$$

where $0 < d \leq n$. This implies that by changing these probabilities, we are describing a different system having a different state. Using the language of IS theory, we can say that the space of resources of the interaction $i \in \mathcal{I}$ is given by $(x_1, \ldots, x_d)$ which are the first d components of the global state $S_{\mathcal{P}} \subseteq \mathbb{R}^n$ of the population $\mathcal{P}$. Using these notations, both the cost C and the force of diversification D depend, in general, by the considered interaction i.
The cost function C must satisfies the inequality

$$\partial_k C(y) \leq \alpha_k(y) \cdot \log_2 k \qquad \forall k = 2, \ldots, d \tag{6.2}$$

for some $\alpha_k : S_{\mathcal{P}} \longrightarrow \mathbb{R}$. Note that (6.2) is not required to hold for $k = 1$. Finally, we assume that

$$\sum_{k=1}^{d} k^{-\alpha_k(y) \cdot \frac{D(y)}{C(y)}} =: N(y) \geq \frac{1}{q_1(y)} \geq e. \tag{6.3}$$

Note that this implies $q_1(y) \leq e^{-1} \simeq 0.368$. This is another restriction on the value $\alpha_k(y) \in \mathbb{R}$: whereas condition (6.2) states that we can take $\alpha_k(y)$ as large as we want, the inequality (6.3) yields that the larger is $\alpha_k(y)$, the more difficult will be to arrive at a value $N(y) \geq e$. Moreover, note that the normalization factor $N(y)$ depends on $\alpha_k(y)$.

We have the following

Theorem 15. *Let $y \in S_{\mathcal{P}} \subseteq \mathbb{R}^n$ be an open set and let $q_j \in \mathcal{C}^1(S_{\mathcal{P}}, \mathbb{R}_{\geq 0})$, for all $j = 1, \ldots, d \leq n$, be such that*

$$\forall x \in S_{\mathcal{P}} : \ \left(q_j(x)\right)_{j=1,\ldots,d} \ \text{is a probability.}$$

Set

$$D(x) := -\sum_{j=1}^{d} q_j(x_j) \cdot \log_2 q_j(x_j) \qquad \forall x \in S_{\mathcal{P}}.$$

Let $C \in \mathcal{C}^1(S_{\mathcal{P}}, \mathbb{R}_{>0})$ be such that

$$\forall x \in S_{\mathcal{P}} : \ 0 < \frac{C(y)}{D(y)} \leq \frac{C(x)}{D(x)}. \tag{6.4}$$

Finally assume that

$$\begin{aligned} & q_j(x_j) = x_j \qquad \forall j = 1, \ldots, d \ \forall x \in S_{\mathcal{P}} \\ & \partial_k C(y) \leq \alpha_k(y) \cdot \log_2 k \qquad \forall k = 2, \ldots, d \\ & \sum_{k=1}^{d} k^{-\alpha_k(y) \cdot \frac{D_i(y)}{C(y)}} =: N(y) \geq \frac{1}{q_1(y)} \geq e, \end{aligned} \tag{6.5}$$

where $\alpha_k : S_{\mathcal{P}} \longrightarrow \mathbb{R}$. Then we have

(i) $q_k(y) = q_1(y) \cdot k^{-\alpha_k(y) \cdot \frac{D(y)}{C(y)}}$ *for all $k = 1, \ldots, d$.*
(ii) $q_1(y) = \frac{1}{N(y)}$.

Proof. Since $S_{\mathcal{P}}$ is an open set, $y \in S_{\mathcal{P}}$ and $C,\ D \in \mathcal{C}^1(S_{\mathcal{P}}, \mathbb{R}_{>0})$, by (6.4) we get $\partial_k \left(\frac{C}{D}\right)(x) = 0$. For simplicity, all the functions that will appear in the following are evaluated at the point y. We have

$$\partial_k C \cdot D + C \cdot \sum_j \left(\partial_k y_j \cdot \log_2 y_j + y_j \frac{1}{y_j} \log_2 e \cdot \partial_k y_j \right) = 0,$$

Where we used $q_j(y) = y_j$ and hence $\partial_k y_j = \frac{\partial}{\partial x_k}(x_j)|_{x_j = y_j} = \delta_{kj}$, so

$$D \cdot \partial_k C + C\left(\log_2 y_k + \log_2 e\right) = 0.$$

By (6.5), for $k \geq 2$ we obtain

$$\log_2 y_k = -\frac{D}{C} \partial_k C - \log_2 e \geq -\frac{D}{C} \alpha_k \log_2 k - \log_2 e,$$

and hence

$$y_k = q_k \geq 2^{\left(-\frac{D}{C}\alpha_k \log_2 k - \log_2 e\right)} = k^{-\alpha_k \frac{D}{C}} e^{-1}.$$

We assumed that $N \geq e$, so

$$q_k \geq \frac{1}{N} \cdot k^{-\alpha_k \frac{D}{C}} \qquad \forall k = 1, \ldots, d.$$

Note that this inequality holds also for $k = 1$ because we assumed that $q_k(y) \geq \frac{1}{N(y)}$.

Finally, we note that we cannot have $q_h > \frac{1}{N} \cdot h^{-\alpha_k \frac{D}{C}}$ for some $h = 1, \ldots, d$ because otherwise we would have

$$\sum_{j=1}^{d} q_j = 1 > \sum_{j=1}^{d} \frac{1}{N} k^{-\alpha_k \frac{D}{C}} = 1.$$

Therefore, we must have $q_k = \frac{1}{N} k^{-\alpha_k \frac{D}{C}}$ for all $k = 1, \ldots, d$. Finally, for $k = 1$ we get $q_1 = \frac{1}{N}$ which proves both our conclusions. □

Examples of cost functions that satisfy (6.2) are given by the average value $C(x) = \sum_{k=1}^{n} c_k(x) \cdot p_k(x)$, where $(p_1(x), \ldots, p_n(x))$ (which is the analogous of P_{C_i} in the notations of the GEP) is a probability, and the k-th component of the cost c_k can be any one of the following examples.

6.2. **Examples of costs functions.**

(i) In this first case, $p_k(x) = q_k(x) = x_k$ and $c_k(x) = \frac{a}{q_k(x)^s} \log_2 k = \frac{a}{x_k^s} \log_2 k$, where $a,\ s \in \mathbb{R}_{>0}$. This example represents costs that are decreasing with an increasing of the probabilities $q_k(x) = x_k$ and they are increasing with an increasing of the number of bits $\log_2 k$ which are necessary to transmit the rank k. In this case we have $\partial_k C(y) = \partial_k c_k(y) \cdot y_k + c_k(y) = \frac{a}{y_k^s}(1-s) \log_2 k$. Therefore, it suffices to take $\alpha_k(y) = \frac{a}{y_k^s}(1-s)$.

(ii) For $k \geq 2$, assume that $c_k(x) \leq a \cdot \log_b(k + k_0) + j_0$, where $a,\ j_0 \in \mathbb{R}_{>0}$, $b > 2$, $k_0 \leq b - 2$ and also that $\partial_k C(x) \leq c_k(x)$. This example is considered, with the equality signs, in [33]. Usually, b is the number of letters in an alphabet we are considering. The second inequality e.g. holds if the costs c_k do not depend on the probabilities y_k, so that $\partial_k C(x) = c_k(x)$. We have $\log_b(k + k_0) \leq \log_2 k$ if and only if

$$k + k_0 \leq b^{\log_2 k} = \left(b^{\log_b k}\right)^{\frac{1}{\log_b 2}} = k^{\log_2 b}. \tag{6.6}$$

Since $b > 2$, we have $\log_2 b > 1$ and the function $k^{\log_2 b} - k$ is increasing in k. Therefore if $k_0 \le 2^{\log_2 b} - 2 = b - 2$ the inequality (6.6) always holds, and hence $c_k(x) \le a \log_2 k + j_0$. If we take $\alpha_k(x) \equiv a + j_0$, we get $\alpha_k \log_2 k = a \log_2 k + j_0 \log_2 k \ge a \log_2 k + j_0 \ge a \cdot \log_b(k + k_0) + j_0 \ge c_{ik}(x) \ge \partial_k C_i(x)$ for $k \ge 2$.

(iii) We can choose a better estimate of α_k if, for $k \ge 2$, $c_k(y) \le j_0 + a \cdot \log_b k$ and $\partial_k C(x) \le c_k(x)$, i.e. if $k_0 = 0$ in the previous example. This case is considered in [13]. In fact, for $k \ge 2$ we have

$$\begin{aligned}\partial_k C(y) \le c_k(y) \le j_0 + a \log_b k = j_0 \log_2 2 + a \log_b 2 \cdot \log_2 k \le \\ \le j_0 \log_2 k + a \log_b 2 \cdot \log_2 k = (a \log_b 2 + j_0) \cdot \log_2 k.\end{aligned}$$

We can hence set $\alpha_k \equiv a \log_b 2 + j_0$.

(iv) $c_k(y) = \gamma_k > 0$. This is the case of constant costs depending only on the rank k (e.g. we can have that γ_k is proportional to the length of the words of rank k). We therefore have to take α so that $\alpha_k(y) \ge \frac{\gamma_k}{\log_2 k}$.

(v) other examples of costs functions?

Note that in the first example (i) it is not reasonable to assume that this formula holds also for $k = 1$ because this would yield a null cost $c_1(y) = 0$. This, and the calculations we realized in the second example, motivate that (6.2) holds only for $k \ge 2$.

7. Generalizing Von Thünen's model

topics to write:

- the necessity of complex systems modeling in economics
- complex adaptive economic systems as stable economic systems

Von Thünen's model tries to answer the basic questions of location and land use theory: "where should a certain activity be located?" and "which activity should be chosen at a certain location?". Both questions address the principles underlying the spatial layout of an economy. Several original assumptions of Von Thünen's model can be weakened and generalized by simply using an appropriate mathematical notations and modeling; later, we will see them in details. Most important for us is that in this model costs and forces of diversification have a clear economic meaning and hence, this model helps to explain how to define costs and force of diversification in other economic systems.

The explicit use of unit of measurement greatly helps the understanding of the different economic quantities that we are going to introduce. We will use notations such as $\left[\frac{€}{\mathrm{m}^2}\right]$ to denote the 1-dimensional (totally ordered real vector) space of quantities whose unit of measurement is € per square meter. Mathematically it can be identified with the space of polynomials in the unknowns $€ \cdot \mathrm{m}^{-2}$.

7.1. Von Thünen's impedance zones.

We need to introduce several quantities and notations:

1. $B \in \mathbb{N}_{>0}$ the *number of commodities* produced by the considered economy. There are no a priory assumptions on the type of commodities (e.g. not necessarily of agricultural type).
2. For all $b = 1, \dots, B$, a unit of measurement u_b for the commodity b. For example u_b can be ton, $n \in \mathbb{N}_{>0}$ so that $[u_b] = \mathbb{R}$ (dimensionless), box, etc.
3. $x_m \in \mathbb{R}^2$ is the *location of the market* m.

4. $A \subseteq \mathbb{R}^2$ possible *locations for the companies* that produce some commodity $b = 1, \ldots, B$.
5. $y_b : A \longrightarrow \left[\frac{u_b}{\mathrm{m}^2}\right]_{>0}$, $y_b(x) > 0$ is the *yield of the commodity* b if the company is located at $x \in A$. The model is not stochastic, so quantities like these can denote average values. The use of a spatial unit of measurement like m^2 is only useful so as to not use too heavy notations. More appropriate units of production v_b, depending on the commodity b, could be introduced instead of m^2 (e.g. it could be $v_b = \mathrm{kwh}$, or $v_b = \frac{\mathrm{hour}}{\mathrm{man}}$).
6. $p_b(x_m) \in \left[\frac{€}{u_b}\right]_{>0}$ *price of the commodity* b in the market x_m per units of b.
7. $c_b : A \longrightarrow \left[\frac{€}{u_b}\right]_{>0}$, $c_b(x) > 0$ is the *production cost* of b at the location $x \in A$ per units of b.
8. $j(-, x_m) : A \longrightarrow [i]_{\geq 0}$, $j(x, x_m)$ is the *impedance* between the location $x \in A \subseteq \mathbb{R}^2$ and the market $x_m \in \mathbb{R}^2$; the impedance has unit of measurement i (e.g. i can be $i = \mathrm{hour}$, $i = \frac{\mathrm{hour}}{\mathrm{man}}$, $i = \mathrm{km}$, $i = €$, etc.). We do not need any assumption about existence or non existence of possible routes, neither on the nature of the transportation, nor on the linearity of j with respect to the distance between x and x_m along the shortest route.
9. $F_b : [i]_{\geq 0} \longrightarrow \left[\frac{€}{u_b}\right]_{>0}$, $F_b(d)$ is the *transportation cost* of the commodity b per units of b and for any pair of points, $(x, x_m) \in \mathbb{R}^2$ having impedance $d \in [i]_{>0}$. For example, $F_b(d)$ can be lower if we assume the possibility to use refrigerators for the transportation of the dairy product b. This modeling assumption includes the particular case where $F_b(d) = \bar{F}_b \cdot d$, where $\bar{F}_b \in \left[\frac{€}{u_b \cdot i}\right]_{>0}$, i.e. the case where the transportation cost is proportional to the impedance d. We always assume that the transportation cost is increasing with the impedance d:
$$\forall d_1, d_2 \in [i]_{\geq 0} : \; d_1 \leq d_2 \Rightarrow F_b(d_1) \leq F_b(d_2). \tag{7.1}$$
10. $k_b(-, x_m) : A \longrightarrow \left[\frac{€}{\mathrm{m}^2}\right]_{>0}$, $k_b(x, x_m)$ is the average *life cost* (everything but the rent of the company's location), per m^2, of the owner producing b at x.

We can now introduce the basic quantities of Von Thünen's model

Definition 16. The *locational rent* or *land value* at $x \in A$ with respect to the production of the commodity $b = 1, \ldots, B$ and the market x_m is
$$L_b(x, x_m) := y_b(x) \cdot [p_c(x_m) - c_b(x) - F_b(j(x, x_m))] \in \left[\frac{€}{\mathrm{m}^2}\right].$$
The *ideal rent* for a company located *at x with respect to the market located at x_m* is
$$R(x, x_m) := \max_{b=1,\ldots,B} [L_b(x, x_m) - k_b(x, x_m)] \in \left[\frac{€}{\mathrm{m}^2}\right]. \tag{7.2}$$
From this, we deduce that the life cost must satisfy the constraint $k_b(x, x_m) < L_b(x, x_m)$ for all $x \in A$.

Moreover, we say that x *is a good (location) for (the production of)* b if
$$R(x, x_m) = L_b(x, x_m) - k_b(x, x_m),$$

i.e. if the ideal rent R equals the land value L_b minus the life cost k_b. Finally, the *impedance boundaries around the market* x_m are given by

$$\underline{r}_b(x_m) := \inf \{j(x, x_m) \mid x \in A,\ x \text{ is good for } b\} \in [i]_{\geq 0} \tag{7.3}$$
$$\overline{r}_b(x_m) := \sup \{j(x, x_m) \mid x \in A,\ x \text{ is good for } b\} \in [i]_{\geq 0}. \tag{7.4}$$

Therefore, if a company is located at x, which is a good location for the production of b, we trivially have that $\underline{r}_b(x_m) \leq j(x, x_m) \leq \overline{r}_b(x_m)$, that is the company is located in the corresponding *impedance zone* bounded by $\underline{r}_b(x_m) \leq \overline{r}_b(x_m)$. Note that if at least two locations are good for b and have different impedance, the zone is non trivial, i.e. $\underline{r}_b(x_m) < \overline{r}_b(x_m)$.

7.2. **Disjoint impedance zones.** We now want to see under what assumptions the impedance zones are disjoint, i.e. when $\underline{r}_\beta(x_m) \leq \overline{r}_\beta(x_m) \leq \underline{r}_b(x_m) \leq \overline{r}_b(x_m)$ if $b \neq \beta$ are two different commodities. Let us start to see when the land value L_b decrease for an increase of the impedance j. It is therefore natural to use the assumption (7.1). We also need to hypothesize that the yield y_b of b, the production cost c_b and the cost of life k_b do not depend on the location x. We will use e.g. the notation $y_b \equiv y_b(-) \in \left[\frac{u_b}{\mathrm{m}^2}\right]$. This is clearly an assumption which holds only if $A \subseteq \mathbb{R}^2$ is not too large; e.g. it surely does not hold for locations situated in different countries, with different cost of labor, different climate conditions and different life costs. In that case, (7.1) simply implies

$$x, y \in A,\ j(x, x_m) > j(y, x_m) \Rightarrow L_b(x, x_m) \leq L_b(y, x_m).$$

Let $b,\ \beta = 1, \ldots, B,\ b \neq \beta$, be two commodities. For simplicity, we omit the dependence by the market's location x_m. By contradiction, assume that

$$\overline{r}_\beta > \underline{r}_b \tag{7.5}$$

Definition (7.4) yields the existence of a location $x \in A$ which is good for β and such that $\underline{r}_b < j(x) \leq \overline{r}_\beta$. Since x is good for β, we have

$$R(x) = L_\beta(x) - k_\beta = \max_b \left[L_b(x) - k_b\right] \geq L_b(x) - k_b. \tag{7.6}$$

Analogously, from (7.3) and $\underline{r}_b < j(x)$ we get the existence of $y \in A$ which is good for b and such that

$$\underline{r}_b \leq j(y) < j(x) \tag{7.7}$$
$$R(y) = L_b(y) - k_b \geq L_\beta(y) - k_\beta. \tag{7.8}$$

How can it happen that x is a good location for β and y is not a good location for it even if $j(y) < j(x)$? To understand this point, we compare the *gain of net land value passing from x to y*, for the commodities β and b:

$$\begin{aligned} \Delta\eta_\beta(x, y) :&= [L_\beta(y) - k_\beta] - [L_\beta(x) - k_\beta] = L_\beta(y) - L_\beta(x) = \\ &= y_\beta \cdot [F_\beta(j(x)) - F_\beta(j(y))] =: -y_\beta \cdot \Delta F_\beta(y, x). \end{aligned} \tag{7.9}$$

Note that $y_\beta \cdot \Delta F_\beta(y, x) \in \left[\frac{€}{\mathrm{m}^2}\right]$ represents the variation of transportation cost, per square meter, passing from x to y. We therefore have that $\Delta\eta_\beta(x, y) \geq \Delta\eta_b(x, y)$ if and only if $y_\beta \cdot \Delta F_\beta(y, x) \leq y_b \cdot \Delta F_b(y, x)$. If we assume these inequalities, considering (7.7), (7.8), (7.6), (7.9), we get $L_\beta(y) = L_\beta(x) + \Delta\eta_\beta(x, y) \geq L_\beta(y) + \Delta\eta_b(x, y) = L_\beta(y) + L_b(y) - L_b(x) \geq L_\beta(y) + L_b(y) - L_b(x) - k_\beta + k_\beta \geq L_b(x) - k_b + L_b(y) - L_b(x) + k_\beta$ and hence $L_\beta(y) - k_\beta \geq L_b(y) - k_b$ so that $L_\beta(y) - k_\beta = L_b(y) - k_b$.

Therefore, y is a good location both for β and b. We can summarize this arguments with the following

Theorem 17 (J.H. von Thünen)**.** *Let us assume that y_b, k_b and c_b do not depend on the location $x \in A$. Let b, β be two commodities such that for all x, $y \in A$*

$$y_\beta \cdot \Delta F_\beta(y,x) \le y_b \cdot \Delta F_b(y,x) \tag{7.10}$$

$$L_\beta(y) - k_\beta \ne L_b(y) - k_b. \tag{7.11}$$

Then impedance zones of b and β (around the market x_m) are disjoint, i.e. $\underline{r}_\beta \le \overline{r}_\beta \le \underline{r}_b \le \overline{r}_b$. Therefore, if any pair of different commodities always have different land values (7.11) *and different variations of transportation costs* (7.10) *(for all x, $y \in A$), then we can order the commodities so that*

$$\underline{r}_{b_1} \le \overline{r}_{b_1} \le \underline{r}_{b_2} \le \overline{r}_{b_2} \le \ldots \le \underline{r}_{b_B} \le \overline{r}_{b_B}.$$

All this underscores that we can have disjoint impedance zones even if we do not have an optimized economy. Indeed, we proved that if x is good for β, y is good for b but $\underline{r}_b \le j(x) < j(y) \le \overline{r}_\beta$ and $y_\beta \cdot \Delta F_\beta(y,x) \le y_b \cdot \Delta F_b(y,x)$, then y is also good for β, i.e. the land values of the two commodities at the location y are equal: $L_\beta(y) - k_\beta = L_b(y) - k_b$. We cannot thus say that producing β at y instead of producing b gives us a better system, e.g. with decreased costs for some agent in the system. In the next section, we will consider what happen when the subpopulation of companies producing the same commodity adapts following the GEP, i.e. evolves decreasing costs in the most diversified (long-term) way.

7.3. **Von Thünen's model and the GEP.**

7.3.1. *Cost minimization.* In a non adapted economy, we do not necessarily have that rents coincide with their ideal values (7.2), e.g. because of ignorance of some agent with respect to the entire market configuration. For simplicity, in this section we assume to consider only one market x_m, so that all the prices, rents and values always refer to x_m; e.g. $L_b(x)$ is the simplified notation for $L_b(x, x_m)$. Assume to have $n_b \in \mathbb{N}_{>0}$ companies producing the commodity $b = 1, \ldots, B$. We can easily think at these companies, and the related renters, as agents and patients of an IS. Then, the configuration space of a generic economy located around the market x_m is given by:

(i) A location $x_b^a \in A$ for each company $a = 1, \ldots, n_b$ producing the commodity $b = 1, \ldots, B$.
(ii) In each one of these locations, we have a rent $r_b^a \in \left[\frac{€}{\mathrm{m}^2}\right]$ really applied to companies that wish to produce b at x_b^a.

Using these notations, we can model configurations representing a positive cost for some agent acting as location renter or tenant owning a company. For example the inequalities

$$r_b^a < L_b(x_b^i) - k_b(x_b^a) < R(x_b^a)$$

imply

(i) $r_b^a < L_b(x_b^a) - k_b(x_b^a)$: the location x_b^a is rented at a price r_b^a which is less than the land value (it could be rented at a higher price).
(ii) $r_b^a < R(x_b^a)$: the rent r_b^a is less than the maximum rent that it would be possible to ask in the location x_b^a (at another company producing a different commodity).

Considering $<$, $=$ or $>$, all the possible inequalities are $3^3 = 27$, but several of them are mathematically impossible, repetitions, or impossible relations due to the definition of $R(x)$ (see (7.2)). There remain the following possible inequalities

Definition 18. The configuration space M of the economies centered around the market x_m is defined by

$$(x_b^1, r_b^1, \dots, x_b^{n_b}, r_b^{n_b})_{b=1,\dots,B} \in M$$

if and only if $x_b^1 \in A$, $r_b^1 \in \left[\frac{€}{\mathrm{m}^2}\right]$ and at least one of the following condition is satisfied

$$\begin{aligned} r_b^a &= L_b(x_b^a) - k_b(x_b^a) = R(x_b^a) \\ r_b^a &= L_b(x_b^a) - k_b(x_b^a) < R(x_b^a) \\ r_b^a &< L_b(x_b^a) - k_b(x_b^a) = R(x_b^a) \\ r_b^a &< L_b(x_b^a) - k_b(x_b^a) < R(x_b^a) \\ r_b^a &> L_b(x_b^a) - k_b(x_b^a) = R(x_b^a) \end{aligned} \tag{7.12}$$

$$\begin{aligned} r_b^a &= R(x_b^a) > L_b(x_b^a) - k_b(x_b^a) \\ r_b^a &> R(x_b^a) > L_b(x_b^a) - k_b(x_b^a) \\ R(x_b^a) &> r_b^a > L_b(x_b^a) - k_b(x_b^a). \end{aligned} \tag{7.13}$$

We also use the simplified notation $(x, r) = (x_b^1, r_b^1, \dots, x_b^{n_b}, r_b^{n_b})_{b=1,\dots,B}$ to denote a configuration. We will think at (x_b^a, r_b^a) as two components of the state space of the interacting entity $a = 1, \dots, n_b$ (company) that produces the commodity $b = 1, \dots B$.

The first one (7.12) of these conditions is called *von Thünen configuration*. Each one of these, but the von Thünen one, corresponds to a possible configuration of an economy where at least one of its agents is paying a cost or is loosing a profit:

Definition 19. Let $b = 1, \dots, B$, $a = 1, \dots, n_b$ and $(x_b^1, r_b^1, \dots, x_b^{n_b}, r_b^{n_b})_{b=1,\dots,B} \in M$, then

(i) The cost paid by the tenant/company a, which produces the commodity b, is located in x_b^a and paying the rent r_b^a is

$$c_\mathrm{t}(x_b^a, r_b^a) := \begin{cases} r_b^a - [L_b(x_b^a) - k_b(x_b^a)] & \text{if } r_b^a > L_b(x_b^a) - k_b(x_b^a) \\ c_{0\mathrm{t}} & \text{otherwise.} \end{cases}$$

The quantity $c_{0\mathrm{t}} > 0$ represents a minimum non avoidable cost related to the rent of this location (e.g. administrative cost) and paid by the tenant. We assume $c_{0\mathrm{t}}$ sufficiently small:

$$\forall b: \ r_b^a > L_b(x_b^a) - k_b(x_b^a) \ \Rightarrow \ 0 < c_{0\mathrm{t}} \le r_b^a - [L_b(x_b^a) - k_b(x_b^a)]\,. \tag{7.14}$$

(ii) The loss of profit of the renter of the location x_b^a at rent r_b^a for the production of the commodity b is

$$l_\mathrm{r}^1(x_b^a, r_b^a) := \begin{cases} [L_b(x_b^a) - k_b(x_b^a)] - r_b^a & \text{if } r_b^a < L_b(x_b^a) - k_b(x_b^a) \\ l_{1\mathrm{r}} > 0 & \text{otherwise.} \end{cases}$$

The quantity $l_{1r} > 0$ represents a minimum non avoidable cost related to the rent of this location (e.g. some tax or the cost to know the land value $L_b(x_b^a)$

and the cost of life $k_b(x_b^a)$) and paid by the renter. We assume $l_{1\mathrm{r}}$ sufficiently small

$$\forall b: \ r_b^a < L_b(x_b^a) - k_b(x_b^a) \ \Rightarrow \ 0 < l_{1\mathrm{r}} \le [L_b(x_b^a) - k_b(x_b^a)] - r_b^a. \tag{7.15}$$

(iii) The loss of profit of the renter of the location x_b^a at rent r_b^a with respect to all the possible commodities is

$$l_{\mathrm{r}}^2(x_b^a, r_b^a) := \begin{cases} R(x_b^a) - r_b^a & \text{if } r_b^a < R(x_b^a) \\ l_{2\mathrm{r}} > 0 & \text{otherwise} \end{cases}$$

The quantity $l_{2r} > 0$ represents a minimum non avoidable cost related to the rent of this location (e.g. some tax or the cost to know the ideal rent $R(x_b^b)$) and paid by the renter. We assume $l_{2\mathrm{r}}$ sufficiently small

$$\forall b: \ r_b^a < R(x_b^a) \ \Rightarrow \ l_{2\mathrm{r}} \le R(x_b^a) - r_b^a. \tag{7.16}$$

To these costs and losses, we associate the following probabilities and hence expected costs and losses

Definition 20. Let $j = 1, 2$ and $(x, r) \in M$, then

(i) The cost of the tenat C_{t} can be computed as

$$\begin{aligned} N_{\mathrm{t}}(x,r) &:= \sum_{a,b} e^{c_{\mathrm{t}}(x_b^a, r_b^a)} \\ p_{\mathrm{t}}(x_b^a, r_b^a) &:= \frac{1}{N_{\mathrm{t}}(x,r)} e^{c_{\mathrm{t}}(x_b^a, r_b^a)} \\ C_{\mathrm{t}}(x,r) &:= \sum_{b=1}^{B} \sum_{a=1}^{n_b} p_{\mathrm{t}}(x_b^a, r_b^a) \cdot c_{\mathrm{t}}(x_b^a, r_b^a) \ge c_{0\mathrm{t}}. \end{aligned}$$

(ii) The losses of profits of the renter can be computed as

$$\begin{aligned} N_{\mathrm{r}}^j(x,r) &:= \sum_{a,b} e^{c_{\mathrm{r}}^j(x_b^a, r_b^a)} \\ p_{\mathrm{r}}^j(x_b^a, r_b^a) &:= \frac{1}{N_{\mathrm{r}}^j(x,r)} e^{c_{\mathrm{r}}^j(x_b^a, r_b^a)} \\ L_{\mathrm{r}}^j(x,r) &:= \sum_{b=1}^{B} \sum_{a=1}^{n_b} p_{\mathrm{r}}^j(x_b^a, r_b^a) \cdot l_{\mathrm{r}}^j(x_b^a, r_b^a) \ge l_{j\mathrm{r}}. \end{aligned}$$

These expected costs and losses are minimum at a von Thünen configuration: $r_b^a = L_b(x_b^a) - k_b(x_b^a) = R(x_b^a)$ for all a, b. Vice versa, when a given configuration $(x_b^1, r_b^1, \ldots, x_b^{n_b}, r_b^{n_b})_{b=1,\ldots,B} \in M$ minimize costs and losses, is it a von Thünen configuration? In general this is false: let $A = \{y_1, y_2\}$, $B = 2$, with

$$\begin{aligned} L_1(y_j) - k_1(y_j) &= 1\frac{€}{\mathrm{m}^2} \\ L_2(y_j) - k_2(y_j) &= 2\frac{€}{\mathrm{m}^2}. \end{aligned}$$

That is the commodity $b_1 = 1$ has a net land value equal to $1\frac{€}{\mathrm{m}^2}$ in both location y_1, y_2, whereas the commodity $b_2 = 2$ has a double net land value in both locations. From the point of view of renter, in both locations it would be better to produce commodity $b_2 = 2$, but this is not possible because in any configuration which situates both commodities in different localizations. Assume, by contradiction, that $(x_1^1, r_1^1, x_2^2, r_2^2) \in M$ minimizes costs and losses. We have

$$\begin{aligned} c_{\mathrm{t}}(x_1^1, r_1^1) = c_{0\mathrm{t}} &\iff r_1^1 \le L_1(x_1^1) - k_1(x_1^1) = 1\frac{€}{\mathrm{m}^2} \\ l_r^1(x_1^1, r_1^1) = l_{1\mathrm{r}} &\iff r_1^1 \ge L_1(x_1^1) - k_1(x_1^1) = 1\frac{€}{\mathrm{m}^2} \\ l_r^2(x_1^1, r_1^1) = l_{2\mathrm{r}} &\iff r_1^1 \ge R(x_1^1) = 2\frac{€}{\mathrm{m}^2}. \end{aligned} \tag{7.17}$$

These are not satisfied by any possible value of r_1^1 (and of $x_1^1 \in A$), and this proves that, in this system, no configuration minimizes costs and losses to the values c_{ot}, $l_{1\mathrm{r}}$, $l_{2\mathrm{r}}$. In other words, this system does not allow for a von Thünen configuration. The problem here is that the system is lacking in configurations and the loss of profits (7.17). This justifies the main assumption of the following

Theorem 21. *If any company can be situated in a good location:*

$$\forall b = 1, \ldots, B \,\forall a = 1, \ldots, n_b \,\exists y \in A : \; y \text{ is good for } b, \tag{7.18}$$

then a given configuration $(x, r) \in M$ minimizes the average costs and losses

$$\begin{aligned} &C_t(x, r) \le C_t(y, s) \\ &L_r^1(x, r) \le L_r^1(y, s) \quad \forall (y, s) \in M \\ &L_r^2(x, r) \le L_r^2(y, s) \end{aligned} \tag{7.19}$$

if and only if (x, r) is a von Thünen configuration, i.e.

$$r_b^a = L_b(x_b^a) - k_b(x_b^a) = R(x_b^a) \quad \forall a, b.$$

Proof. If (x, r) is a von Thünen configuration, then $C_{\mathrm{t}}(x, r) = c_{0\mathrm{t}} \le C_t(y, s)$, $L_{\mathrm{r}}^1(x, r) = l_{1\mathrm{r}} \le L_{\mathrm{r}}^1(y, s)$, $L_{\mathrm{r}}^2(x, r) = l_{2\mathrm{r}} \le L_{\mathrm{r}}^2(y, s)$ by Def. 19 and because of (7.14), (7.15), (7.16). Vice versa, using assumption (7.18), we can construct a von Thünen configuration (y, s) by choosing a good location for every commodity:

$$\forall b, a \,\exists y_b^a \in A : \; R(y_b^a) = L_b(y_b^a) - k_b(y_b^a) =: s_b^a.$$

But (7.19) yields

$$\begin{aligned} c_{0\mathrm{t}} &\le C_{\mathrm{t}}(x, r) \le C_{\mathrm{t}}(y, s) = c_{0\mathrm{t}} \\ l_{1\mathrm{r}} &\le L_{\mathrm{r}}^1(x, r) \le L_{\mathrm{r}}^1(y, s) = l_{1\mathrm{r}} \\ l_{2\mathrm{r}} &\le L_{\mathrm{r}}^2(x, r) \le L_{\mathrm{r}}^2(y, s) = l_{2\mathrm{r}}. \end{aligned}$$

Therefore, also (x, r) is necessarily a von Thünen configuration. □

Clearly, assumption (7.18) is not realistic. Once again, this underscores that what cost functions to consider in this kind if economic models, is a modeling/-philosophical/political choice. Depending on our political choices, other types of costs can be considered, such as: environmental costs, energy consumption, loss of profits, loss of job places because of transfer of branches, company's stock price, etc. Let us consider, e.g., the difference between the cost C_{t} and the losses of profits L_{r}^j: if the first one is paid by a company for too much time, this could cause its failure (i.e. its deactivation, using the language of IS). This is not the case for the losses of profits L_{r}^j. One can also argue that also the maximization of profit is unrealistic because it implies the full knowledge of the entire system and unrealistic tendencies, such as that of trying the renting of centered locations to jeweleries as soon as the demand of jewels increases.

In another political approach, we can denote by $\mathrm{tax}(x_b^a)$ the real estate tax that must be paid by the owner (renter) of the real estate property located (and rented) at x_b^a; then, we can consider the previous cost of the tenant C_t and, instead of the losses of profits L_r^j, and examine the cost

$$c_\mathrm{r}(x_b^a, r_b^a) := \begin{cases} \mathrm{tax}(x_b^a) - r_b^a & \text{if } r_b^a < \mathrm{tax}(x_b^a) \\ 0 & \text{otherwise,} \end{cases} \tag{7.20}$$

and the corresponding average cost C_r. In this model, any configuration where $r_b^a \le L_b(x_b^a) - k_b(x_b^a)$ and $r_b^a \ge \mathrm{tax}(x_b^a)$ minimizes the costs C_t and C_r. We have hence a more realistic model without any assumption of profits maximization or of full information about the market.

7.3.2. *Maximization of diversification forces.* A very natural flux of goods and resources is the amount of commodity $b = 1, \dots, B$ that each company $a = 1, \dots, n_b$ is selling to the market x_m. We therefore want to see what the maximization of the diversification forces yields if we consider this interaction. The population $\mathcal{P}_b$ is hence the collection of the aforementioned companies, so that its cardinality is $|\mathcal{P}_b| = n_b$. We can therefore think at an interaction $i_b^a : a \xrightarrow{g_b^a} x_m$ having the company a as agent and the market x_m as patient. The good is a quantity $\varphi_b^a \in [0, Q_b] =: S_{g_b^a}$ belonging to the state space of the propagator $\mathrm{pr}(i_b^a) = g_b^a$. To understand, in a simpler deterministic model of constant production, if s_b^a is the surface of the company a used for the production of 1year of b (more generally, the amount of units of production v_b used in 1year, see Sec. 7.1, definition 5.), then we have

$$\varphi_b^a = y_b(x_b^a) \cdot s_b^a \in \left[\frac{u_b}{\mathrm{year}}\right]$$
$$Q_b = \sum_{a=1}^{n_b} \varphi_b^a.$$

We can thus think at Q_b as the amount of commodity b demanded by x_m in 1year. We always assume $Q_b = \sum_{a=1}^{n_b} \varphi_b^a > 0$ and the probability of extraction of goods equals to a Bernoulli process with parameter $p = \frac{\varphi_b^a}{Q_b} =: p_b^a$. In other words, the probability to extract a unit of commodity b produced by the company a, among all those flowing to the market x_m in one year, equals the fraction $\frac{\varphi_b^a}{Q_b}$ of goods φ_b^a produced by a over the total Q_b.

The interaction i_b of the population $\mathcal{P}_b$ has hence the following state space of the propagator

$$\left(S_{\mathrm{pr}(i_b)}, \mathfrak{S}_{\mathrm{pr}(i_b)}\right) = \prod_{a=1}^{n_b} \left(S_{\mathrm{pr}(i_b^a)}, \mathfrak{S}_{\mathrm{pr}(i_b^a)}\right) = [0, Q_b]^{n_b}.$$

The probability to extract a good from this product space of resources of b is given by the joint discrete probability $\left(p_b^1, \dots, p_b^{n_b}\right)$. Assume now that the system is in a best possible state y (at any time t) for the population $\mathcal{P}_b$ of companies producing the commodity b. We have that

$$D_{i_b}(\mathrm{n}_{i_b}\, y, t) = D_{i_b}(\varphi_b^1, \dots, \varphi_b^{n_b}) = -\sum_{a=1}^{n_b} p_b^a \cdot \log_2 p_b^a \geq D_{i_b}(\mathrm{n}_{i_b}\, x, s) \quad \forall (x, s). \tag{7.21}$$

The necessary result yielded by this maximization is well known, and it can be briefly repeated here: using the Lagrange's multiplier method, we get

$$\exists \lambda \in \mathbb{R} : \left. \frac{\partial D_{i_b}(x^1, \dots, x^{n_b})}{\partial x^a} \right|_{x^{(-)}=\varphi_b^{(-)}} =$$

$$= \frac{\partial}{\partial x^a} \left(-\sum_{a=1}^{n_b} \frac{x^a}{Q_b} \cdot \log_2 \frac{x^a}{Q_b} \right) \Bigg|_{x^{(-)}=\varphi_b^{(-)}} =$$

$$= \lambda \cdot \frac{\partial}{\partial x^a} \left(\sum_{a=1}^{n_b} \frac{x^a}{Q_b} - 1 \right) \Bigg|_{x^{(-)}=\varphi_b^{(-)}} .$$

This gives $\varphi_b^a = \frac{2^{-\lambda}}{eQ_b}$, so that $Q_b = \sum_{a=1}^{n_b} \varphi_b^a = n_b \cdot \frac{2^{-\lambda}}{eQ_b}$ and

$$\varphi_b^a = \frac{Q_b}{n_b}. \tag{7.22}$$

Fluxes of commodities b are hence equally divided among all the companies. We can interpret this property as a resilience characteristic of the adapted population $\mathcal{P}_b$, due to the absence of monopolies, maximal distribution of work, etc. We can hence say that in this case the GEP coincides with a well studied property of suitable economies. Equation (7.22) can also be obtained by Thm. 15 and considering costs such as those of Def. (19) or (7.20) (in general any cost function C such that $\frac{\partial C}{\partial x^k} = 0$. In fact, assuming (7.21) and setting $\alpha_k = 0$, we have $\sum_{k=1}^{n_b} k^{-\alpha_k \cdot \frac{D_{i_b}}{C}} = n_b$. Therefore, if at least $n_b \geq 3$, Thm. 15 yields $q_k(y) = p_b^k(y) = p_b^1(y) = \frac{1}{n_b} = \frac{\varphi_b^a}{Q_b}$.

It is clear that the result (7.22) does not correspond to a realistic behavior: However, it gives elements to start thinking at the GEP as a way to measure how far our economy is from a stable, adaptive, resilient system. We recall that these notions strictly depend on our choice of the cost function C_i and of the interaction i of the population $\mathcal{P}$.

Do we want to include other examples in the present definition of CAS?

8. Conclusions and future developments

Possible conclusions and future developments can be listed as follows.

We can firstly consider topics related to a further general development of IS theory:

- The validation of the definitions of IS and CAS consists in the verification that a sufficient number of examples, which are considered of interest by the related scientific community, satisfy these definitions. A first step could be considering the long list of examples of Sec. 6.
- Other important validations of a unifying mathematical theory of CS are both its providing efficient methods for real world problems, and its potential in solving important problems. What precise benefits and methods IS theory can bring to classical open problems concerning CS? In other words: can IS theory convince a wide audience that it is a good framework for considering models of real CS?

All this would show the potential impact of IS theory on applied disciplines and eventually on the real world.

- For these reasons, we plan several applications of IS theory to classical models such as ECHO [28], Kauffman NK Model [30, 29], game of life, Sugarscape [18, 19], Yakovenko's model [14, 62, 17], Page's model, [39].
- The notion of cause-effect graph of interactions, briefly introduced in Sec. 4 is strictly tied to the notion of morphism of IS. These concepts seem to represent an innovative approach to hierarchy of CS, learning systems and even software verification.

We can also consider topics more near to economics:

- The notion of best possible state following the GEP seems to have strong relations with other notions of economic equilibrium, such as Pareto efficiency and Nash equilibrium. The investigation of the reciprocal relations is therefore mandatory.
- In Sec. 6, examples 9. and 13., we mentioned at famous non-adapted economic systems. IS theory can be used to simulate sufficiently idealized models of systems derived by these, but satisfying the GEP with respect to several types of costs. The idea is to transform them into simulated adaptive system, e.g. using a suitable global taxation method.

Finally, the presented von Thünen's model can be improved from several point of views:

- How far suitable generalizations of the present notations can meaningfully include more than one market?
- We can transform the present von Thünen's model into a more realistic, dynamic, stochastic IS so as to numerically study its dynamics.
- something else?

References

[1] Arnade C., Why It's Smart to Be Reckless on Wall Street, Scientific American Guest blog 2013, https://blogs.scientificamerican.com/guest-blog/why-its-smart-to-be-reckless-on-wall-street/

[2] Bäck T., *Evolutionary algorithms in theory and practice*, Oxford University Press, 1996.

[3] Bak P., Creutz M., *Fractals and Self Organized Criticality*. Springer Verlag, Berlin, 1994.

[4] Battiston S., Puliga M., Kaushik R., Tasca P., Caldarelli G., DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk, Scientific Reports, 2:541, 2012, doi:10.1038/srep00541.

[5] Bedau M., McCaskill J., Packard P., Rasmussen S., Green D., Ikegami T., Kaneko K., Ray T., Open problems in artificial life. Artificial Life 2000, 6, 363-376.

[6] Bouchard-Côté A., Hall D., Griffiths T.L., Klein D., Automated reconstruction of ancient languages using probabilistic models of sound change, PNAS vol. 110, no. 11, 2012, 4224–4229, doi: 10.1073/pnas.1204678110.

[7] Cameron E.Z., du Toit J.T., Winning by a Neck: Tall Giraffes Avoid Competing with Shorter Browsers, The American Naturalist 169, no. 1, 2007, 130-135, doi: 10.1086/509940

[8] Casti J., Would-be worlds: toward a theory of complex systems. Artificial Life and Robotics, 1(1):11–13, 1997.

[9] Chadwick M.J., Jolly A.E.J., Amos D.P., Hassabis D., Spiers H.J., A Goal Direction Signal in the Human Entorhinal/Subicular Region, Current Biology 25, 87–92, January 5, 2015, http://dx.doi.org/10.1016/j.cub.2014.11.001

[10] Chu D., Strand R., Fjelland R., Theories of complexity. Complexity, 8(3), 19-30, 2003.

[11] Crespi B., Stead P., Elliot M., Comparative genomics of autism and schizophrenia, PNAS, vol. 107, suppl. 1, 2010, doi:10.1073/pnas.0906080106.

[12] Crutchfield J.T., Complex systems theory? Santa Fe Institute Retreat, 2003. See http://cse.ucdavis.edu/ chaos/chaos/talks/CSTheorySFIRetreat.pdf.

[13] B. Conrad, M. Mitzenmacher, Power Laws for Monkeys Typing Randomly: The Case of Unequal Probabilities, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 50, NO. 7, JULY 2004.

[14] Cottrell A.F., Cockshott P., Michaelson G.J., Wright I.P., Yakovenko V.M., Classical Econophysics, Series: Routledge Advances in Experimental and Computable Economics, Routledge, 2009.

[15] Danowitz M., Vasilyev A., Kortlandt V., Solounias N., Fossil evidence and stages of elongation of the *Giraffa camelopardalis* neck, *R. Soc. opensci.* **2**:150393, 2015. http://dx.doi.org/10.1098/rsos.150393.

[16] Darwiche A., *Modeling and Reasoning with Bayesian Networks*, Cambridge University Press, 2009.

[17] Drăgulescu A., Yakovenko V.M., Statistical mechanics of money, Eur. Phys. J. B 17, pp. 723-729, 2000.

[18] Epstein J., Agent-based computational models and generative social science. Complexity 1999, 4(5), 41-60.

[19] Epstein J., Axtell R., Growing Artificial Societies: Social Science from the bottom-up, The MIT Press, Cambridge, 1998.

[20] Fontana W., Ballati S., Complexity - why the sudden fuss. Complexity 1999, 4(3), 14-15.

[21] Gabaix X., Zipf 's law for cities: An explanation. Quarterly Journal of Economics, CXIV:739–767, 1999.

[22] Gavrilets S., On the evolutionary origins of the egalitarian syndrome, PNAS vol. 109, no. 35, 2012, 14069–14074, doi: 10.1073/pnas.1201718109

[23] Haag G., *Dynamic Decision Theory: Application to Urban and Regional Topics,* Kluwer Academic, 1989.

[24] Haken H., *Synergetics. Introduction and Advanced Topics*, Springer Berlin, 2004.

[25] Haken H., *Information and self-organization: A macroscopic Approach to Complex Systems.* Springer Series in Synergetics. Springer-Verlag, Berlin and Heidelberg, 2000. 2nd edition.

[26] Herron M.D., Doebeli M., Parallel Evolutionary Dynamics of Adaptive Diversification in Escherichia coli. PLoS Biol 11(2): e1001490, 2013. doi:10.1371/ journal.pbio.1001490.

[27] Holland J., Can there be a unified theory of complex adaptive systems? In: The Mind and the Brain, and Complex Adaptive Systems, volume XXII of Studies in the Sciences of Complexity, Morowitz, H., Singer, J., Eds.; Santa Fe Institute, 1995, pp 45-50.

[28] Holland J., Echoing emergence: objectives, rough definitions, and speculations for ECHO-class mod- els. In: Complexity—Metaphores, Models, and Reality, volume XIX of Studies in the Sciences of Complexity; Cowan, G., Pines, D., Meltzer, D., Eds.; Santa Fe Institute, 1994, pp 310-334.

[29] Kauffman S., The Origins of Order , Oxford University Press, Oxford, 1993.

[30] Kauffman S. A., Levin S., Towards a general theory of adaptive walks on rugged landscapes. Journal of Theorectical Biology, Vol 128, pp11 - 45, 1987.

[31] Kuhlemeier C., Phyllotaxis, Trends in Plant Science, Vol.12 No.4, 2007, doi:10.1016/j.tplants.2007.03.004.

[32] Lerman K., Galstyan A., A general methodology for mathematical analysis of multi-agent systems. Technical Report ISI-TR-529, USC Information Sciences, 2001.

[33] Mandelbrot B., An informational theory of the statistical structure of language, in W. Jackson (eds.), Communication Theory, Butterworths, London, 1953.

[34] Mandic D.P., Chambers J.A., *Recurrent Neural Networks for Prediction*, John Wiley & Sons, 2001.

[35] Margolus N., Toffoli T., *Cellular automata machines. A new environment for modeling.* MIT Press, MA, USA, 1987.

[36] Miller J.H., Page S.E., *Complex Adaptive Systems: An Introduction to Computational Models of Social Life.* Princeton University Press, 2007.

[37] Newman M.E.J., *Networks: An Introduction.* Oxford University Press, 2010.

[38] Nicolis G., Prigogine I., *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations.* Wiley, New York, 1977.

[39] Page S.E., Type interaction models and the rule of six, Economic Theory (2007) 30, pp. 223–241.

[40] Power R.A. et al., Polygenic risk scores for schizophrenia and bipolar disorder predict creativity, Nature neuroscience, Brief communications, 2015, doi:10.1038/nn.4040.

[41] Pluchino A., Garofalo C., Rapisarda A., Spagano S., Caserta M., Accidental politicians: How randomly selected legislators can improve parliament efficiency, Physica A: Statistical Mechanics and its Applications, Volume 390, Issues 21–22, 2011, 3944-3954, doi:10.1016/j.physa.2011.06.028.
[42] Puu T., *Mathematical Location and Land Use Theory. An Introduction*, Springer-Verlag Berlin Heidelberg 1997.
[43] Reeves M., Levin S., Ueda D., The Biology of Corporate Survival. Natural ecosystems hold surprising lessons for business, Harvard Business Review, January-February 2016.
[44] Schweitzer F., *Brownian Agents and Active Particles*, Springer, Berlin, 2003.
[45] Schweitzer F., *modeling Complexity in Economic and Social Systems*. World Scientific, Singapore, 2002.
[46] Simon H.A., Can there be a science of complex systems? In Y. Bar-Yam (Ed.), Unifying themes in complex systems: Proceedings from the International Conference on Complex Systems 1997 (pp. 4-14). Cambridge, MA: Perseus Press.
[47] Simmons R.E., Scheepers L., Winning by a Neck: Sexual Selection in the Evolution of Giraffe, The American Naturalist 148, no. 5, 1996, 771-786,doi:10.1086/285955.
[48] Smith E., Thermodynamics of natural selection I: Energy flow and the limits on organization, Journal of Theoretical Biology 252 (2008) 185–197.
[49] Smith E., Thermodynamics of natural selection II: Chemical Carnot cycles, Journal of Theoretical Biology 252 (2008) 198–212.
[50] Smith E., Thermodynamics of natural selection III: Landauer's principle in computation and chemistry, Journal of Theoretical Biology 252 (2008) 213–220.
[51] Sornette D., *Critical Phenomena in Natural Sciences*, Springer Series in Synergetics, Springer-Verlag Berlin Heidelberg, 2006.
[52] Tacchella A., Cristelli M., Caldarelli G., Gabrielli A., Pietronero, L. A New Metrics for Countries' Fitness and Products' Complexity. Sci. Rep. 2, 723, 2012; doi:10.1038/srep00723.
[53] Vancheri A., Giordano P., Andrey D., Fuzzy logic based modeling of traffic flows induced by regional shopping malls. Advances in Complex Systems Vol. 17, N. 3 & 4, 2014, (39 pages).
[54] Vancheri A., Giordano P., Andrey D., Albeverio S., A model for urban growth processes with continuous state cellular automata, multi-agents and related differential equation. Part 1: Theory. Environment and Planning B: Planning and Design 2008, volume 35, issue 4, pages 723-739.
[55] Vancheri A., Andrey D., Giordano P., Albeverio S., A model for urban growth processes with continuous state cellular automata, multi-agents and related differential equation. Part 2: Computer Simulations. Environment and Planning B: Planning and Design 2008, volume 35, pages 863-880.
[56] Victor P., Questioning economic growth, Nature 468, 370–371, 2010, doi:10.1038/468370a.
[57] von Thünen J.H., *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalokonomie*, Perthes, Hamburg, 1826.
[58] Weidlich W., *Sociodynamics: A Systematic Approach to Mathematical modeling in the Social Sciences*, The Harwood Academic Publisher, 2000.
[59] Wooldridge M.J., *An introduction to multiagent systems*, John wiley & Sons, 2002.
[60] Xie J., Padoa-Schioppa C., Neuronal remapping and circuit persistence in economic decisions, Nature neuroscience 2016, doi:10.1038/nn.4300.
[61] Yukalov V.I., Sornette D., Self-organization in complex systems as decision making, Advances in Complex Systems Vol. 17, No. 03n04, 2014.
[62] Yakovenko V.M., Barkley Rosser J., Statistical mechanics of money, wealth, and income, Rev. Mod. Phys. 81, 1703 2009.
[63] Zipf G.K., *Human Behavior and the Principle of Least Effort*, Addison Wesley, Cambridge, MA., 1949.

P. Giordano, Wolfgang Pauli Institute, Vienna, AT.

J.D. Farmer, The Institute for New Economic Thinking at the Oxford Martin School.