HAL
open science

Mathematical derivation of wave propagation properties in hierarchical neural networks with predictive coding feedback dynamics

Grégory Faye, Guilhem Fouilhé, Rufin VanRullen

To cite this version:

Grégory Faye, Guilhem Fouilhé, Rufin VanRullen. Mathematical derivation of wave propagation properties in hierarchical neural networks with predictive coding feedback dynamics. Bulletin of Mathematical Biology, In press, 10.1007/s11538-023-01186-9 . hal-04062705

HAL Id: hal-04062705

https://hal.science/hal-04062705

Submitted on 12 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mathematical derivation of wave propagation properties in hierarchical neural networks with predictive coding feedback dynamics

Grégory Faye*[1,3], Guilhem Fouilhé[1,2,3], and Rufin VanRullen[2,3]

[1]Institut de Mathématiques de Toulouse ; UMR5219, Université de Toulouse ; UPS IMT, F-31062 Toulouse Cedex 9 France
[2]CerCo, CNRS, 31052 Toulouse, France
[3]ANITI, Université de Toulouse, 31062, Toulouse, France

April 12, 2023

Abstract

Sensory perception (e.g. vision) relies on a hierarchy of cortical areas, in which neural activity propagates in both directions, to convey information not only about sensory inputs but also about cognitive states, expectations and predictions. At the macroscopic scale, neurophysiological experiments have described the corresponding neural signals as both forward and backward-travelling waves, sometimes with characteristic oscillatory signatures. It remains unclear, however, how such activity patterns relate to specific functional properties of the perceptual apparatus. Here, we present a mathematical framework, inspired by neural network models of predictive coding, to systematically investigate neural dynamics in a hierarchical perceptual system. We show that stability of the system can be systematically derived from the values of hyper-parameters controlling the different signals (related to bottom-up inputs, top-down prediction and error correction). Similarly, it is possible to determine in which direction, and at what speed neural activity propagates in the system. Different neural assemblies (reflecting distinct eigenvectors of the connectivity matrices) can simultaneously and independently display different properties in terms of stability, propagation speed or direction. We also derive continuous-limit versions of the system, both in time and in neural space. Finally, we analyze the possible influence of transmission delays between layers, and reveal the emergence of oscillations at biologically plausible frequencies.

*gregory.faye@math.univ-toulouse.fr

1 Introduction

The brain's anatomy is characterized by a strongly hierarchical architecture, with a succession of brain regions that process increasingly complex information. This functional strategy is mirrored by the succession of processing layers found in modern deep neural networks (and for this reason, we use the term "layer" in this work to denote one particular brain region in this hierarchy, rather than the laminar organization of cortex that is well-known to neuroscientists). The hierarchical structure is especially obvious in the organization of the visual system [15], starting from the retina through primary visual cortex (V1) and various extra-striate regions, and culminating in temporal lobe regions for object recognition and in parietal regions for motion and location processing.

In this hierarchy of brain regions, the flow of information is clearly bidirectional: there are comparable number of fibers sending neural signals down (from higher to lower levels of the hierachy) as there are going up [8]. While the bottom-up or "feed-forward" propagation of information is easily understood as integration of sensory input (and matches the functional structure found in artificial deep learning networks), the opposite feedback direction of propagation is more mysterious, and its functional role remains unknown.

Predictive coding is one dominant theory to explain the function of cortical feedback [27]. Briefly, the theory states that each layer in the cortical hierarchy generates predictions about what caused their own activity; these predictions are sent to the immediately preceding layer, where a prediction error can be computed, and carried forward to the original layer, which can then iteratively update its prediction. Over time (and as long as the sensory input does not change), the system settles into a state where top-down predictions agree with bottom-up inputs, and no prediction error is transmitted. Like any large-scale theory of brain function, the predictive coding theory is heavily debated [22]. But macroscopic (EEG) experiments have revealed characteristic propagation signatures that could be hallmarks of predictive coding. For instance, Alamia and VanRullen [1] showed evidence for alpha-band (7-15Hz) oscillatory travelling waves propagating in both directions (feed-forward and feedback); the oscillation frequency and dynamics were compatible with a simplistic hierarchical model that included a biologically plausible time delay for transmitting signals between layers, and was also confirmed by a rudimentary mathematical model. In another study, Bastos et al [4, 5] found that beta (15-30Hz) and gamma-frequency (30-100Hz) oscillations could reflect, respectively, the predictions and prediction errors signals carried by backward and forward connections.

More recently, predictive coding has been explored in the context of deep neural networks [9, 25, 32]. For instance, Choksi et al [9] augmented existing deep convolutional networks with feedback connections and a mechanism for computing and minimizing prediction errors, and found that the augmented system displayed more robust perception, better aligned with human abilities. In another study, Pang et al [25] used a similar system and reported the emergence of illusory contour perception comparable to what humans (but not standard deep neural networks) would typically perceive.

While the concept of predictive coding is potentially fundamental for understanding brain function, and its large-scale implementation in deep artificial neural networks provides empirical support for its potential functional relevance, there is a gap of theoretical knowledge about the type of brain activity that predictive coding could engender, and the potential conditions for its stability. Here, we propose a mathematical framework where a potentially infinite number of neuronal layers exchange signals in both directions according to predictive coding principles. The stable propagation of information in such a system can be explored analytically as a function of its initial state, its internal parameters (controlling the strength of

inputs, predictions, and error signals) and its connectivity (e.g. convolution kernels). Our approach considers both a discrete approximation of the system, as well as continuous abstractions. We demonstrate the practical relevance of our findings by applying them to a ring model of orientation processing. Finally, we extend our analytical framework to a more biologically plausible situation with communication delays between successive layers. This gives rise to oscillatory signals resembling those observed in the brain.

2 Model description

Our initial model is inspired by the generic formulation of predictive coding proposed in the context of deep learning models by Choksi et al. [9]. This formulation considers different update terms at each time step: feed-forward inputs, memory term, feedback- and feed-forward prediction error corrections. By modulating the hyper-parameters controlling each of these terms, the model can be reconciled with different formulations of predictive coding (for instance, the Rao and Ballard model [27] by setting the feed-forward input term to zero) or other models of hierarchical brain function (e.g. similar to Heeger's model [19] by setting the feed-forward error correction to zero). Indeed, our objective is precisely to characterize the propagation dynamics inside the network as a function of the relative value of these hyper-parameters, which in turn alters the model's functionality.

We consider the following recurrence equation where $\mathcal{E}_j^n \in \mathbb{R}^d$ represents an encoder at step n and layer j

$$\mathcal{E}_j^{n+1} = \beta \mathcal{W}^f \mathcal{E}_{j-1}^{n+1} + (1-\beta)\mathcal{E}_j^n - \alpha \mathcal{F}_{j-1}^n - \lambda \mathcal{B}_j^n, \quad j = 1, \cdots, J-1\,,$$

where $\mathcal{W}^f \in \mathscr{M}_d(\mathbb{R})$ is a $d \times d$ square matrix representing the weights of feedforward connections which we assume to be the same for each layer such that $\mathcal{W}^f \mathcal{E}_{j-1}^{n+1}$ models an instantaneous feedforward drive from layer $j-1$ to layer j, controlled by hyper-parameter β. The term $\mathcal{F}_{j-1}^n$ encodes a feedforward error correction process, controlled by hyper-parameter α, where the reconstruction error $\mathcal{R}_{j-1}^n$ at layer $j-1$, defined as the square error between the representation $\mathcal{E}_{j-1}^n$ and the predicted reconstruction $\mathcal{W}^b \mathcal{E}_j^n$, that is

$$\mathcal{R}_{j-1}^n := \frac{1}{2}\|\mathcal{E}_{j-1}^n - \mathcal{W}^b \mathcal{E}_j^n\|^2,$$

propagates to the layer j to update its representation. Here, $\mathcal{W}^b \in \mathscr{M}_d(\mathbb{R})$ is a $d \times d$ square matrix representing the weights of feedback connections which we assume to be the same for each layer. Following [1, 9, 27, 32], the contribution $\mathcal{F}_{j-1}^n$ is then taken to be the gradient of $\mathcal{R}_{j-1}^n$ with respect to $\mathcal{E}_j^n$, that is

$$\mathcal{F}_{j-1}^n = \nabla \mathcal{R}_{j-1}^n = -(\mathcal{W}^b)^{\mathbf{t}} \mathcal{E}_{j-1}^n + (\mathcal{W}^b)^{\mathbf{t}} \mathcal{W}^b \mathcal{E}_j^n.$$

On the other hand, $\mathcal{B}_j^n$ incorporates a top-down prediction to update the representation at layer j. This term thus reflects a feedback error correction process, controlled by hyper-parameter λ. Similar to the feedforward process, $\mathcal{B}_j^n$ is defined as the the gradient of $\mathcal{R}_j^n$ with respect to $\mathcal{E}_j^n$, that is

$$\mathcal{B}_j^n = \nabla \mathcal{R}_j^n = -\mathcal{W}^b \mathcal{E}_{j+1}^n + \mathcal{E}_j^n.$$

As a consequence, our model reads

$$\mathcal{E}_j^{n+1} = \beta \mathcal{W}^f \mathcal{E}_{j-1}^{n+1} + \alpha (\mathcal{W}^b)^{\mathbf{t}} \mathcal{E}_{j-1}^n + \left[(1-\beta-\lambda)\mathbf{I}_d - \alpha (\mathcal{W}^b)^{\mathbf{t}} \mathcal{W}^b\right] \mathcal{E}_j^n + \lambda \mathcal{W}^b \mathcal{E}_{j+1}^n, \tag{2.1}$$

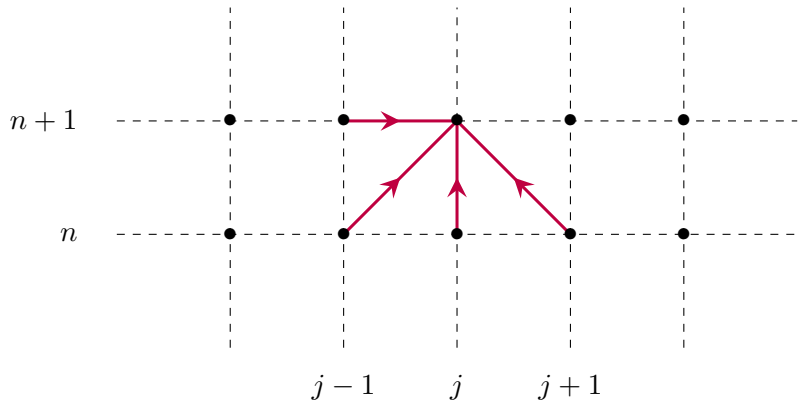


Figure 1: *Schematic illustration of the network structure of model (2.1) where each point represents a given neuronal layer index j (x-axis) at a particular time step n (y-axis), and the red arrows indicate the contributions leading to the update of $\mathcal{E}_j^{n+1}$.*

for each $j = 1, \cdots, J-1$ and $n \geq 0$ where we denoted $\mathbf{I}_d$ the identity matrix of $\mathscr{M}_d(\mathbb{R})$. We supplement the recurrence equation (2.1) with the following boundary conditions at layer $j = 0$ and layer $j = J$. First, at layer $j = 0$, we impose

$$\mathcal{E}_0^n = \mathcal{S}_0^n, \quad n \geq 0, \tag{2.2}$$

where $\mathcal{S}_0^n \in \mathbb{R}^d$ is a given source term, which can be understood as the network's constant visual input. At the final layer $j = J$, there is no possibility of incoming top-down signal, and thus one gets

$$\mathcal{E}_J^{n+1} = \beta \mathcal{W}^f \mathcal{E}_{J-1}^{n+1} + \alpha (\mathcal{W}^b)^{\mathbf{t}} \mathcal{E}_{J-1}^n + \left[(1-\beta)\mathbf{I}_d - \alpha (\mathcal{W}^b)^{\mathbf{t}} \mathcal{W}^b \right] \mathcal{E}_J^n, \quad n \geq 0. \tag{2.3}$$

Finally, at the initial step $n = 0$, we set

$$\mathcal{E}_j^0 = \mathcal{H}_j, \quad j = 0, \cdots, J, \tag{2.4}$$

for some given initial sequence $(\mathcal{H}_j)_{0,\cdots,J}$. For instance, in Choksi et al [9], $\mathcal{H}_j$ was initialized by a first feedforward pass through the system, i.e. $\beta > 0$ and $\alpha = \lambda = 0$. Throughout we assume the natural following compatibility condition between the source terms and the initial condition, namely

$$\mathcal{S}_0^0 = \mathcal{H}_0. \tag{2.5}$$

Regarding the hyper-parameters of the problem we assume that

$$0 \leq \beta < 1, \quad \text{with} \quad 0 \leq \alpha + \lambda \leq 1. \tag{2.6}$$

Our key objective is to characterize the behavior of the solutions of the above recurrence equation (2.1) as a function of the hyper-parameters and the feedforward and feedback connections matrices $\mathcal{W}^f$ and $\mathcal{W}^b$. We would like to stay as general as possible to encompass as many situations as possible, keeping in mind that we already made strong assumptions by imposing that the weight matrices of feedforward and feedback connections are identical from one layer to another. Motivated by concrete applications, we will mainly consider matrices $\mathcal{W}^f$ and $\mathcal{W}^b$ which act as convolutions on $\mathbb{R}^d$.

3 The identity case

It turns out that we will gain much information by first treating the simplified case where $\mathcal{W}^f$ and $\mathcal{W}^b$ are both identity. That is, from now on, and throughout this section we assume that

$$\mathcal{W}^f = \mathcal{W}^b = \mathbf{I}_d.$$

That is, each neuron in a layer is only connected to the corresponding neuron in the immediately preceding and following layer, with unit weight in each direction. Under such a setting, the recurrence equation (2.1) reduces to a scalar equation, that is

$$e_j^{n+1} = \beta e_{j-1}^{n+1} + \alpha e_{j-1}^n + (1-\beta-\lambda-\alpha)e_j^n + \lambda e_{j+1}^n, \quad j = 1, \cdots, J-1, \tag{3.1}$$

with this time the unknown $e_j^n \in \mathbb{R}$, together with

$$e_0^n = s_0^n, \quad n \geq 0, \tag{3.2}$$

and

$$e_J^{n+1} = \beta e_{J-1}^{n+1} + \alpha e_{J-1}^n + (1-\beta-\alpha)e_J^n, \quad n \geq 0. \tag{3.3}$$

with

$$e_j^0 = h_j, \quad j = 0, \cdots, J. \tag{3.4}$$

3.1 Wave propagation on an infinite depth network

It will be first useful to consider the above problem set on an infinite domain and look at

$$e_j^{n+1} = \beta e_{j-1}^{n+1} + \alpha e_{j-1}^n + (1-\beta-\lambda-\alpha)e_j^n + \lambda e_{j+1}^n, \quad j \in \mathbb{Z}, \tag{3.5}$$

given some initial sequence

$$e_j^0 = h_j, \quad j \in \mathbb{Z}.$$

This situation has no direct equivalent in the brain, where the number of hierarchically connected layers is necessarily finite; but it is a useful mathematical construct. Indeed, such recurrence equations set on the integers $\mathbb{Z}$ are relatively well understood from the mathematical numerical analysis community. The behavior of the solution sequence $(e_j^n)_{j\in\mathbb{Z}}$ can be read out from the so-called amplification factor function defined as

$$\rho(\theta) := \frac{\alpha\left(e^{-\mathbf{i}\theta}-1\right) + 1 - \beta + \lambda\left(e^{\mathbf{i}\theta}-1\right)}{1-\beta e^{-\mathbf{i}\theta}}, \quad \theta \in [-\pi, \pi], \tag{3.6}$$

and which relates spatial and temporal modes. Indeed, formally, the sequence $(\rho(\theta)^n e^{\mathbf{i}j\theta})_{j\in\mathbb{Z}}$ is an explicit solution to (3.5) for each $\theta \in [-\pi, \pi]$. Actually one can be much precise and almost explicit in the sense that one can relate the expression of the solutions to (3.5) starting from some initial sequence $(h_j)_{j\in\mathbb{Z}}$ to the properties of ρ in a systematic way that we now briefly explain.

Let us first denote by $\mathcal{G}^n = (\mathcal{G}_j^n)_{j\in\mathbb{Z}}$ the sequence which is the fundamental solution of (3.5) in the special case where $(\mathcal{H}_j)_{j\in\mathbb{Z}}$ is the Dirac delta sequence $\boldsymbol{\delta}$. The Dirac delta sequence $\boldsymbol{\delta}$ is defined as $\boldsymbol{\delta}_0 = 1$ and $\boldsymbol{\delta}_j = 0$ for all $j \in \mathbb{Z}\backslash\{0\}$. As a consequence, we have $\mathcal{G}^0 = \boldsymbol{\delta}$ and for each $n \geq 0$

$$\mathcal{G}_j^{n+1} - \beta\mathcal{G}_{j-1}^{n+1} = \alpha\mathcal{G}_{j-1}^n + (1-\beta-\lambda-\alpha)\mathcal{G}_j^n + \lambda\mathcal{G}_{j+1}^n, \quad j \in \mathbb{Z}.$$

The starting point of the analysis is the following representation formula, obtained via inverse Fourier transform, which reads

$$\mathcal{G}_j^n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\mathbf{i}j\theta}\rho(\theta)^n \mathrm{d}\theta, \quad n \geq 1, \quad j \in \mathbb{Z}. \tag{3.7}$$

Then, given any initial sequence $(h_j)_{j\in\mathbb{Z}}$, the solution $(e_j^n)_{j\in\mathbb{Z}}$ to (3.5) can be represented as the convolution product between the initial sequence and the fundamental solution, namely

$$e_j^n = \sum_{\ell\in\mathbb{Z}} \mathcal{G}_{j-\ell}^n h_\ell, \quad j \in \mathbb{Z}, \quad n \geq 1. \tag{3.8}$$

That is, having characterized the fundamental solution for a simple input pattern ($\boldsymbol{\delta}$), with a unitary impulse provided to a single layer, we can now easily generalize to any arbitrary input pattern, by applying the (translated) fundamental solution to each layer.

Our aim is to understand under which conditions on the hyper-parameters we can ensure that the solutions of (3.5) given through (3.8) remain bounded for all $n \geq 1$ independently of the choice of the initial sequence $(h_j)_{j\in\mathbb{Z}}$. More precisely, we introduce the following terminology. We say that the recurrence equation is *stable* if for each bounded initial sequence $(h_j)_{j\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$, the corresponding solution $(e_j^n)_{j\in\mathbb{Z}}$ given by (3.8) satisfies

$$\sup_{j\in\mathbb{Z}}|e_j^n| \underset{n\to\infty}{\longrightarrow} 0.$$

On the other hand, we say that the recurrence equation is *unstable* if one can find a bounded initial sequence $(h_j)_{j\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$ such that the corresponding solution $(e_j^n)_{j\in\mathbb{Z}}$ given by (3.8) satisfies

$$\sup_{j\in\mathbb{Z}}|e_j^n| \underset{n\to\infty}{\longrightarrow} +\infty.$$

Finally, we say that the recurrence equation is *marginally stable* if there exists a universal constant $C > 0$ such that for each bounded initial sequence $(h_j)_{j\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$, the corresponding solution $(e_j^n)_{j\in\mathbb{Z}}$ given by (3.8) satisfies

$$\sup_{j\in\mathbb{Z}}|e_j^n| \leq C \sup_{j\in\mathbb{Z}}|h_j|, \quad n \geq 1.$$

It turns out that one can determine the stability properties of the recurrence equation by solely looking at the amplification factor function. Indeed, from [29], we know that

$$\lim_{n\to\infty} \|\mathcal{G}^n\|_{\ell^1(\mathbb{Z})}^{1/n} = \max_{\theta\in[-\pi,\pi]}|\rho(\theta)|,$$

where we have set

$$\|\mathcal{G}^n\|_{\ell^1(\mathbb{Z})} := \sum_{j\in\mathbb{Z}} |\mathcal{G}_j^n|.$$

As a consequence, we directly deduce that the recurrence equation is stable when $|\rho(\theta)| < 1$ for all $\theta \in [-\pi, \pi]$, whereas it is unstable if there exists $\theta_0 \in [-\pi, \pi]$ such that $|\rho(\theta_0)| > 1$. The limiting case occurs precisely when $\max_{\theta\in[-\pi,\pi]}|\rho(\theta)| = 1$ and there is actually a long history of works [10, 12, 14, 26, 30] that have studied the marginal stability of the recurrence equation in that case. All such results rely on a very precise understanding of the amplification factor function and lead to the following statement.

Theorem 1 ([10, 12, 14, 26, 30]). *Suppose that there exist finitely many $\theta_1, \cdots, \theta_K \in [-\pi, \pi]$ such that for all $\theta \in [-\pi, \pi] \setminus \{\theta_1, \cdots, \theta_k\}$ one has $|\rho(\theta)| < 1$ and $|\rho(\theta_k)| = 1$ for each $k = 1, \cdots, K$. Furthermore, assume that there exist $c_k \in \mathbb{R}$, $\sigma_k \in \mathbb{C}$ with $\mathrm{Re}(\sigma_k) > 0$ and an integer $\mu_k \geq 1$ such that*

$$\frac{\rho(\theta_k + \theta)}{\rho(\theta_k)} = \exp\left(-\mathbf{i}c_k\theta - \sigma_k\theta^{2\mu_k} + \mathcal{O}(|\theta|^{2\mu_k+1})\right), \text{ as } \theta \to 0.$$

Then the recurrence equation is marginally stable.

Based on the above notions of stability/instability, we see that the only interesting situation is when the recurrence equation is marginally stable, and thus when the amplification function is contained in the unit disk with finitely many tangent points to the unit circle with prescribed asymptotic expansions. This is also the only interesting situation from a biological standpoint, as it ensures that the network remains active, yet without runaway activations.

3.1.1 Study of the amplification factor function

Since we assumed that $0 \leq \beta < 1$, the denominator in (3.6) never vanishes and is well-defined. Next, we crucially remark that we always have

$$\rho(0) = 1.$$

We will now check under which conditions $|\rho(\theta)| \leq 1$ for all $\theta \in [-\pi, \pi]$ to guarantee marginal stability of the recurrence equation.

To assess stability, we compute

$$\begin{aligned}
|\rho(\theta)|^2 &= \frac{((\lambda+\alpha)(\cos(\theta)-1)+1-\beta)^2 + (\lambda-\alpha)^2\sin(\theta)^2}{1-2\beta\cos(\theta)+\beta^2} \\
&= \frac{(\lambda+\alpha)^2(\cos(\theta)-1)^2 + 2(1-\beta)(\lambda+\alpha)(\cos(\theta)-1) + (1-\beta)^2 + (\lambda-\alpha)^2(1-\cos(\theta)^2)}{(1-\beta)^2 + 2\beta(1-\cos(\theta))} \\
&= \frac{(1-\cos(\theta))\left((\lambda+\alpha)^2(1-\cos(\theta)) - 2(1-\beta)(\lambda+\alpha) + (\lambda-\alpha)^2(1+\cos(\theta))\right) + (1-\beta)^2}{(1-\beta)^2 + 2\beta(1-\cos(\theta))} \\
&= \frac{(1-\cos(\theta))\left(-4\alpha\lambda\cos(\theta) - 2(1-\beta)(\lambda+\alpha) + 2(\lambda^2+\alpha^2)\right) + (1-\beta)^2}{(1-\beta)^2 + 2\beta(1-\cos(\theta))}
\end{aligned}$$

such that $|\rho(\theta)|^2 \leq 1$ is equivalent to

$$(1-\cos(\theta))\left(2\beta + 4\alpha\lambda\cos(\theta) + 2(1-\beta)(\lambda+\alpha) - 2(\lambda^2+\alpha^2)\right) \geq 0, \quad \theta \in [-\pi, \pi],$$

and since $1 - \cos(\theta) \geq 0$ we need to ensure

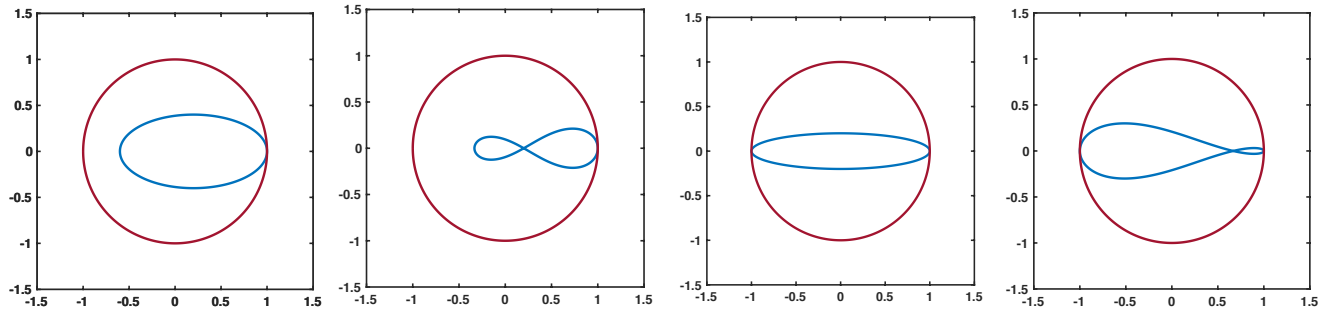
$$\beta + 2\alpha\lambda\cos(\theta) + (1-\beta)(\lambda+\alpha) - \lambda^2 - \alpha^2 \geq 0, \quad \theta \in [-\pi, \pi],$$

and evaluating at $\pm\pi$ the above inequality we get

$$\beta + (1-\beta)(\lambda+\alpha) - (\lambda+\alpha)^2 \geq 0.$$

But we remark that the above expression can be factored as

$$(\beta + \lambda + \alpha)(1 - \lambda - \alpha) \geq 0.$$



(a) $\beta = 0$ and $\alpha + \lambda < 1$. (b) $\beta > 0$ and $\alpha + \lambda < 1$. (c) $\beta = 0$ and $\alpha + \lambda = 1$. (d) $\beta > 0$ and $\alpha + \lambda = 1$.

Figure 2: *Several representative illustration of the curve $\theta \mapsto \rho(\theta)$ for $\theta \in [-\pi, \pi]$ in the case where $\beta = 0$ and $\beta \neq 0$. (a)-(b) Amplification factor function $\rho(\theta)$ (blue curve) with a unique tangency point on the unit circle at $z = 1$ corresponding $\theta = 0$. (c)-(f) When $\alpha + \lambda = 1$ the function $\rho(\theta)$ (blue curve) has two tangency points on the unit circle at $z = 1$ corresponding $\theta = 0$ and $z = -1$ corresponding to $\theta = \pm\pi$.*

As a consequence, $|\rho(\theta)|^2 \leq 1$ if and only if $\lambda + \alpha \leq 1$. This is precisely the condition that we made in (2.6). We can actually track cases of equality which are those values of $\theta \in [-\pi, \pi]$ for which we have

$$(1 - \cos(\theta)) \left(2\beta + 4\alpha\lambda \cos(\theta) + 2(1 - \beta)(\lambda + \alpha) - 2(\lambda^2 + \alpha^2)\right) = 0.$$

We readily recover that at $\theta = 0$ we have $|\rho(0)| = 1$. So, now assuming that $\theta \neq 0$, we need to solve

$$\beta + 2\alpha\lambda \cos(\theta) + (1 - \beta)(\lambda + \alpha) - \lambda^2 - \alpha^2 = 0,$$

which we write as

$$\beta - 2\alpha\lambda + (1 - \beta)(\lambda + \alpha) - \lambda^2 - \alpha^2 + 2\alpha\lambda\left(\cos(\theta) + 1\right)) = 0,$$

and using the previous factorization we get

$$(\beta + \lambda + \alpha)\left(1 - \lambda - \alpha\right) + 2\alpha\lambda\left(\cos(\theta) + 1\right)) = 0,$$

and we necessarily get that both $1 + \cos(\theta) = 0$ and $1 - \lambda - \alpha = 0$ must be satisfied. As consequence, $|\rho(\pm\pi)| = 1$ if and only if $1 = \lambda + \alpha$.

As a summary we have obtained that:

- if $0 \leq \lambda + \alpha < 1$ and $0 \leq \beta < 1$, then $|\rho(\theta)| < 1$ for all $\theta \in [-\pi, \pi] \backslash \{0\}$ with $\rho(0) = 1$;
- if $\lambda + \alpha = 1$ and $0 \leq \beta < 1$, then $|\rho(\theta)| < 1$ for all $\theta \in (-\pi, \pi) \backslash \{0\}$ with $\rho(0) = 1$ and $\rho(\pm\pi) = -1$.

We present in Figure 2 several representative illustrations of the spectral curves $\rho(\theta)$ for various values of the hyper-parameters recovering the results explained above.

Furthermore, near $\theta \sim 0$, we get that ρ admits the following asymptotic expansion

$$\rho(\theta) = \exp\left(-\mathbf{i}\frac{\beta + \alpha - \lambda}{1 - \beta}\theta - \frac{\beta(1 - \alpha - \lambda) + \alpha + \lambda - (\lambda - \alpha)^2}{2(1 - \beta)^2}\theta^2 + \mathcal{O}(|\theta|^3)\right), \text{ as } \theta \to 0,$$

provided that

$$\beta(1 - \alpha - \lambda) + \alpha + \lambda - (\lambda - \alpha)^2 \neq 0.$$

In fact, since $-(\lambda-\alpha)^2 \geq -(\lambda+\alpha)^2$ as both α and λ are positive, we remark that

$$\beta(1-\alpha-\lambda)+\alpha+\lambda-(\lambda-\alpha)^2 \geq \beta(1-\alpha-\lambda)+\alpha+\lambda-(\lambda+\alpha)^2 = (\beta+\lambda+\alpha)\,(1-\lambda-\alpha) \geq 0.$$

Finally, we remark that when $\alpha+\lambda=1$ we have

$$\rho(\theta+\pi) = -\exp\left(-\mathbf{i}\frac{-\beta+\alpha-\lambda}{1+\beta}\theta - \frac{1-(\alpha-\lambda)^2}{2(1+\beta)^2}\theta^2 + \mathcal{O}(|\theta|^3)\right), \text{ as } \theta \to 0.$$

From now on, we denote

$$c_0 := \frac{\beta+\alpha-\lambda}{1-\beta}, \quad \sigma_0 := \frac{\beta(1-\alpha-\lambda)+\alpha+\lambda-(\lambda-\alpha)^2}{2(1-\beta)^2},$$
$$c_\pi := \frac{-\beta+\alpha-\lambda}{1+\beta}, \quad \sigma_\pi := \frac{1-(\alpha-\lambda)^2}{2(1+\beta)^2},$$

and we always assume that

$$\sigma_0 > 0, \text{ and } \sigma_\pi > 0,$$

which is equivalent to assume that $0<\alpha<1$ and $0<\lambda<1$.

Here, (c_0,σ_0) and (c_π,σ_π) are derived, respectively, from the asymptotic expansions of the amplification factor function $\rho(\theta)$ near $\theta=0$ and $\theta=\pi$, as defined above. On the one hand c_0 reflects the propagation speed of the solution associated with $\rho(0)$, while σ_0 can be understood as its spatio-temporal spread (and similarly for the solution potentially associated with $\rho(\pi)$). In the following, we explore the fundamental solutions of this system for various values of its hyper-parameters.

3.1.2 Turning off the instantaneous feedforward connections: case $\beta = 0$

We first investigate the case where there is no instantaneous feedforward connections in the network, that is we set $\beta=0$. This case, although less generic, is compatible with the prominent Rao-Ballard formulation of predictive coding [27], in which feedforward connections—after contributing to setting the initial network activity—only convey prediction errors, as captured by the hyper-parameter α. In that case, the model is fully explicit: the update at time step $n+1$ only depends on the internal states at the previous step n since we simply have

$$e_j^{n+1} = \alpha e_{j-1}^n + (1-\lambda-\alpha)e_j^n + \lambda e_{j+1}^n, \quad j \in \mathbb{Z}.$$

As we assumed that $\alpha+\lambda \leq 1$, the right-hand side of the recurrence equation is a positive linear combination of elements of the sequence (e_j^n) such that we have positivity principle of the solution, namely

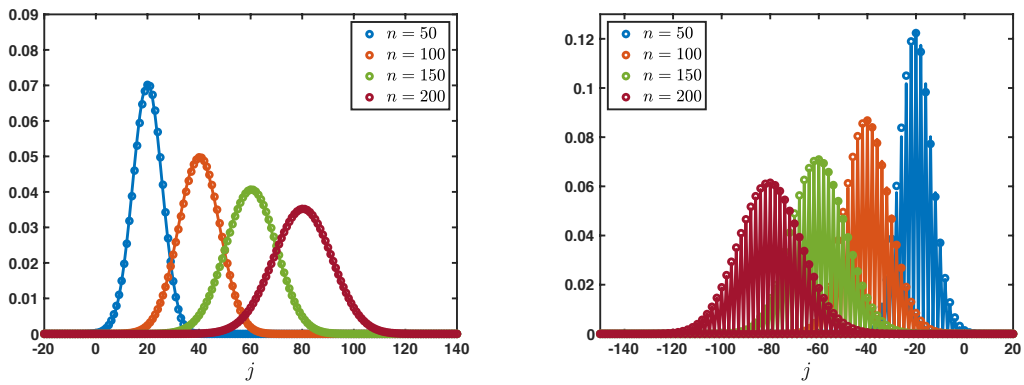
$$\forall j \in \mathbb{Z}, \quad 0 \leq h_j \quad \Longrightarrow \quad \forall j \in \mathbb{Z}, \quad n \geq 1, \quad 0 \leq e_j^n.$$

Furthermore, since the recurrence equation is explicit, we have finite speed propagation, in the following sense. Recall that when $\beta=0$, the fundamental solution $\mathcal{G}^n$ is solution to

$$\mathcal{G}_j^{n+1} = \alpha\mathcal{G}_{j-1}^n + (1-\lambda-\alpha)\mathcal{G}_j^n + \lambda\mathcal{G}_{j+1}^n, \quad n \geq 1, \qquad j \in \mathbb{Z},$$

starting from $\mathcal{G}^0 = \boldsymbol{\delta}$. Finite speed of propagation then refers to the property that

$$\mathcal{G}_j^n = 0, \quad |j| > n.$$



(a) $\beta = 0$, $\lambda < \alpha$ and $\alpha + \lambda < 1$.

(b) $\beta = 0$, $\alpha < \lambda$ and $\alpha + \lambda = 1$.

Figure 3: *Illustration of the evolution of the fundamental solution $\mathcal{G}^n$ starting from the Dirac delta sequence at $j = 0$ in the case $\beta = 0$ at several time iterations. The circles represent the numerically computed solution while the plain lines represent the Gaussian approximation. (a) When $\lambda < \alpha$ there is a rightward propagation along a Gaussian profile whose leading profile is given by $\frac{1}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|j-c_0 n|^2}{4\sigma_0 n}\right)$. (b) When $\alpha < \lambda$ there is a leftward propagation along a Gaussian profile whose leading profile is given by $\frac{1+(-1)^{n+j}}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|j-c_0 n|^2}{4\sigma_0 n}\right)$ which vanishes whenever $n + j$ is odd.*

This in turn implies that necessarily $c_0 \in (-1, 1)$ which is readily seen from the explicit formula $c_0 = \alpha - \lambda$ in that case. Actually, it is possible to be more precise and to give a general expression for the fundamental solution. Roughly speaking, each $\mathcal{G}_j^n$ ressembles a discrete Gaussian distribution centered at $j = c_0 n$ and we refer to the recent theoretical results of [10, 12, 14, 26] for a rigorous justification.

Essentially, the results can be divided into two cases depending on whether or not $\alpha + \lambda = 1$. As can be seen above, the special case $\alpha + \lambda = 1$ results in a cancellation of the "memory" term, such that a neuronal layer j's activity does not depend on its own activity at the previous time step, but only on the activity of its immediate neighbors $j - 1$ and $j + 1$. More precisely, we have the following:

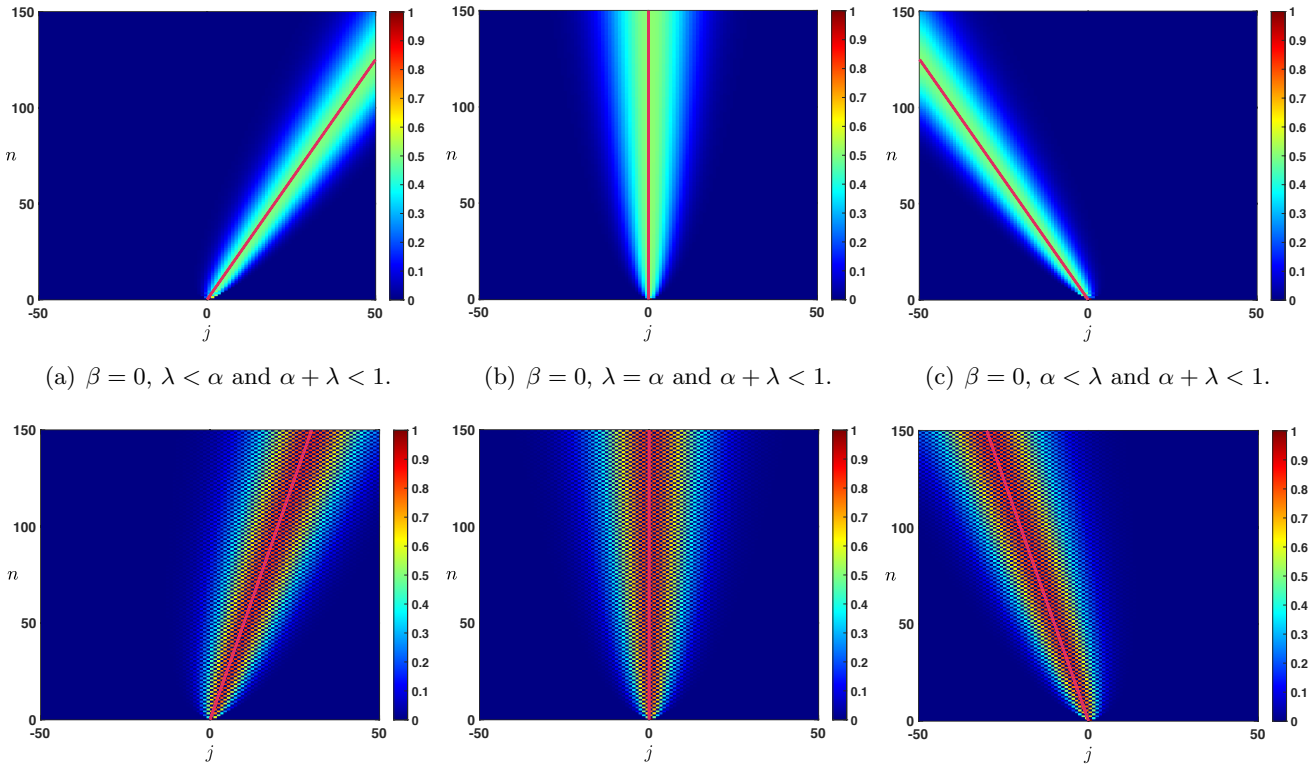
- Case: $0 \leq \lambda + \alpha < 1$. The fundamental solution can be decomposed as

$$\mathcal{G}_j^n = \frac{1}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|j - c_0 n|^2}{4\sigma_0 n}\right) + \mathcal{N}_j^n, \quad |j| \leq n,$$

  where the remainder term satisfies an estimate

$$\left|\mathcal{N}_j^n\right| \leq \frac{C}{n} \exp\left(-\kappa \frac{|j - c_0 n|^2}{n}\right), \quad |j| \leq n,$$

  for some universal constants $C, \kappa > 0$ which only depend on the hyper-parameters and not n and j. In Figure 3(a), we represented the fundamental solution $\mathcal{G}_j^n$ at different time iterations (circles) in the case $\lambda < \alpha$ where there is rightward propagation with $c_0 > 0$ and compared it with the leading order fixed Gaussian profile centered at $j = c_0 n$ (plain line). On the other hand, in Figure 4, panels (a)-(b)-(c), we illustrate the above results by presenting a space-time color plot of the fundamental solution rescaled by a factor $\sqrt{n}$. We observe rightward (respectively leftward) propagation with $c_0 > 0$ (respectively $c_0 < 0$) when $\lambda < \alpha$ (respectively $\alpha < \lambda$), while when $\alpha = \lambda$ we have $c_0 = 0$ and no propagation occurs.



(a) $\beta = 0$, $\lambda < \alpha$ and $\alpha + \lambda < 1$.

(b) $\beta = 0$, $\lambda = \alpha$ and $\alpha + \lambda < 1$.

(c) $\beta = 0$, $\alpha < \lambda$ and $\alpha + \lambda < 1$.

(d) $\beta = 0$, $\lambda < \alpha$ and $\alpha + \lambda = 1$.

(e) $\beta = 0$, $\lambda = \alpha$ and $\alpha + \lambda = 1$.

(f) $\beta = 0$, $\alpha < \lambda$ and $\alpha + \lambda = 1$.

Figure 4: *Illustration of the evolution of the rescaled solution sequence* $(\sqrt{n}\,\mathcal{G}_j^n)_{j\in\mathbb{Z}}$ *starting from the Dirac delta sequence at* $j = 0$ *in the case* $\beta = 0$. *First row:* $\alpha + \lambda < 1$ *and second row:* $\alpha + \lambda = 1$. *When* $\lambda \lesseqgtr \alpha$, *we observe a rightward/leftward propagation while when* $\alpha = \lambda$ *no propagation occurs. In all panels, the pink curve is given by* $j = nc_0$, *clearly illustrating the fact that* c_0 *measures the propagation speed of the solution. Note that in the case* $\beta = 0$ *and* $\alpha + \lambda = 1$, *we have* $c_0 = c_\pi$ *which results in the tiled patterns observed in panels (d)-(e)-(f).*

- Case: $\lambda + \alpha = 1$. In this case, we first note that we have $c_0 = c_\pi$ together with $\sigma_0 = \sigma_\pi$ and

$$\mathcal{G}_j^n = \frac{1 + (-1)^{n+j}}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|j - c_0 n|^2}{4\sigma_0 n}\right) + \mathcal{N}_j^n, \quad |j| \le n,$$

  where the remainder term satisfies an estimate

$$\left|\mathcal{N}_j^n\right| \le \frac{C}{n} \exp\left(-\kappa \frac{|j - c_0 n|^2}{n}\right), \quad |j| \le n,$$

  for some universal constants $C, \kappa > 0$. In Figure 3(b), we represented the fundamental solution $\mathcal{G}_j^n$ at different time iterations (circles) in the case $\alpha < \lambda$ where there is leftward propagation with $c_0 < 0$ and compared it with the leading order fixed Gaussian profile centered at $j = c_0 n$ (plain line). Similarly as in the previous case, in Figure 4, panels (d)-(e)-(f), we illustrate the above results by presenting a space-time color plot of the fundamental solution rescaled by a factor $\sqrt{n}$. The direction of propagation still depends on the sign of c_0 and whether or not $\lambda \lesseqgtr \alpha$. Unlike the case $\alpha + \lambda < 1$, we observe a tiled pattern where $\mathcal{G}_j^n = 0$ for even or odd integers alternatively for each time step.

As a partial intermediate summary, we note that the sign of c_0 (directly related to the sign of $\alpha - \lambda$) always indicates in which direction the associated Gaussian profile propagates. Namely if $\alpha > \lambda$ and $c_0 > 0$ (resp.

$\alpha < \lambda$ and $c_0 < 0$) there is rightward (resp. leftward) propagation. Intuitively, this behavior reflects the functional role of each hyper-parameter, with α and λ controlling feed-forward and feed-back prediction error correction, respectively. When $\alpha = \lambda$, the two terms are equally strong, and there is no dominant direction of propagation. In addition, when $\lambda + \alpha = 1$, the Gaussian profile is *oscillating* because of the presence of $(-1)^{n+j}$. As will be seen later when considering continuous versions of our model, this oscillatory pattern might not be truly related to neural oscillations observed in the brain, but could instead arise here as a consequence of discrete updating.

Finally, we note that the fundamental solution sequence $(\mathcal{G}_j^n)_{j\in\mathbb{Z}}$ is uniformly integrable for all values of the parameters, that is there exists some universal constant $C > 0$, depending only on the hyper-parameters such that

$$\|\mathcal{G}^n\|_{\ell^1(\mathbb{Z})} := \sum_{j\in\mathbb{Z}} |\mathcal{G}_j^n| \leq C, \quad n \geq 1.$$

As a consequence, since given any bounded initial sequence $(h_j)_{j\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$, the solution $(e_j^n)_{j\in\mathbb{Z}}$ to (3.5) can be represented as the convolution product between the initial sequence and the fundamental solution, namely

$$e_j^n = \sum_{\ell\in\mathbb{Z}} \mathcal{G}_{j-\ell}^n h_\ell, \quad j \in \mathbb{Z}, \quad n \geq 1,$$

we readily deduce that the solution $(e_j^n)_{j\in\mathbb{Z}}$ is uniformly bounded with respect to n, that is there exists some universal constant denoted $C > 0$, such that

$$\sup_{j\in\mathbb{Z}} \left|e_j^n\right| \leq C \sup_{j\in\mathbb{Z}} |h_j|, \quad n \geq 1.$$

This is exactly our definition of marginal stability.

3.1.3 Turning on the instantaneous feedforward connections: case $\beta > 0$

We now turn to the general case where $\beta > 0$. That is, the feed-forward connections continue to convey sensory inputs at each time step following the network initializing, and β controls the strength of these signals. In that case, the recurrence equation is no longer explicit but implicit and the positivity property together with the finite speed propagation no longer hold true in general. Indeed, upon introducing the shift operator

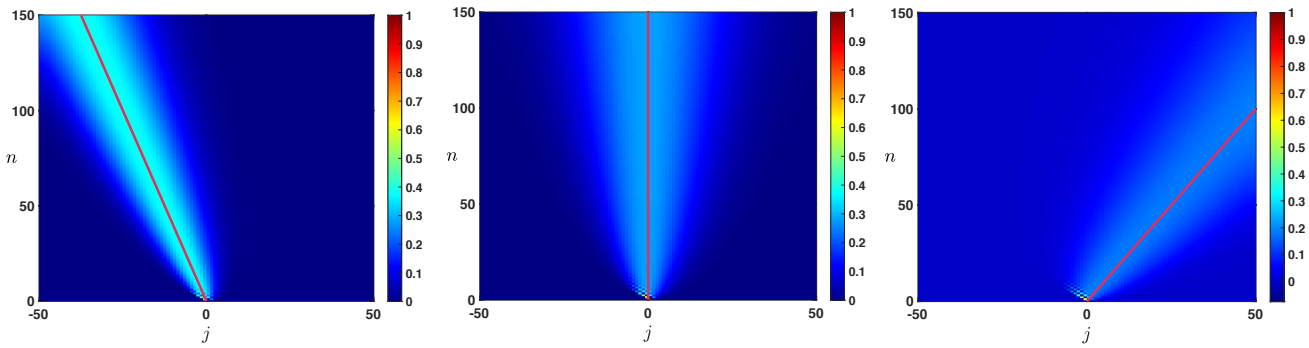
$$\mathbf{S} : (u_j)_{j\in\mathbb{Z}} \mapsto (u_{j+1})_{j\in\mathbb{Z}},$$

we remark that equation (3.5) can be written

$$\left(\mathrm{Id} - \beta\mathbf{S}^{-1}\right) e^{n+1} = \alpha\mathbf{S}^{-1}e^n + (1 - \beta - \lambda - \alpha)e^n + \lambda\mathbf{S}e^n, \quad n \geq 0,$$

with $e^n = (e_j^n)_{j\in\mathbb{Z}}$. Since $0 < \beta < 1$ and $|||\mathbf{S}^{-1}|||_{\ell^q(\mathbb{Z})\to\ell^q(\mathbb{Z})} = 1$ for any $q \in [1, +\infty]$, the operator $\mathrm{Id} - \beta\mathbf{S}^{-1}$ is invertible on $\ell^q(\mathbb{Z})$ for any $q \in [1, +\infty]$ with inverse

$$\left(\mathrm{Id} - \beta\mathbf{S}^{-1}\right)^{-1} = \sum_{\ell=0}^{\infty} \beta^\ell \mathbf{S}^{-\ell}.$$



(a) $\beta + \alpha < \lambda$. (b) $\beta + \alpha = \lambda$. (c) $\lambda < \beta + \alpha$.

Figure 5: *Effects of turning on $\beta > 0$ when $\alpha + \lambda < 1$. When $\beta + \alpha < \lambda$ we have $c_0 < 0$ and observe backward propagation in panel (a) while when $\lambda < \beta + \alpha$ we have $c_0 > 0$ and have forward propagation as seen in panel (c). At the transition $\beta + \alpha = \lambda$, the wave speed vanishes $c_0 = 0$ and there is no propagation as illustrated in panel (b).Intuitively, α and β, both propagating signals in the forward (rightward) direction, compete with λ carrying the feedback (leftward) prediction signals; this competition determines the main direction of propagation of neural activity in the system.*

As a consequence, the recurrence equation can be recast as a convolution operator across the network layers with infinite support, namely

$$e_j^{n+1} = \alpha \sum_{\ell=0}^{\infty} \beta^\ell e_{j-\ell-1}^n + (1 - \beta - \lambda - \alpha) \sum_{\ell=0}^{\infty} \beta^\ell e_{j-\ell}^n + \lambda \sum_{\ell=0}^{\infty} \beta^\ell e_{j-\ell+1}^n, \quad j \in \mathbb{Z}, \quad n \geq 0.$$

From the above expression, we readily deduce that the positivity of the solution is preserved whenever $0 < \beta < 1 - \lambda - \alpha$. Furthermore, for the fundamental solution starting from the Dirac delta solution which solves

$$\mathcal{G}_j^{n+1} - \beta \mathcal{G}_{j-1}^{n+1} = \alpha \mathcal{G}_{j-1}^n + (1 - \beta - \lambda - \alpha) \mathcal{G}_j^n + \lambda \mathcal{G}_{j+1}^n, \quad j \in \mathbb{Z}, \quad n \geq 0,$$

we only have that

$$\mathcal{G}_j^n = 0, \quad j < -n,$$

which implies that $-1 < c_0, c_\pi < +\infty$. Indeed, from the formula of c_0 we get that

$$c_0 \sim \frac{1 - \alpha - \beta}{1 - \beta} \longrightarrow +\infty, \qquad \text{as } \beta \to 1^-.$$

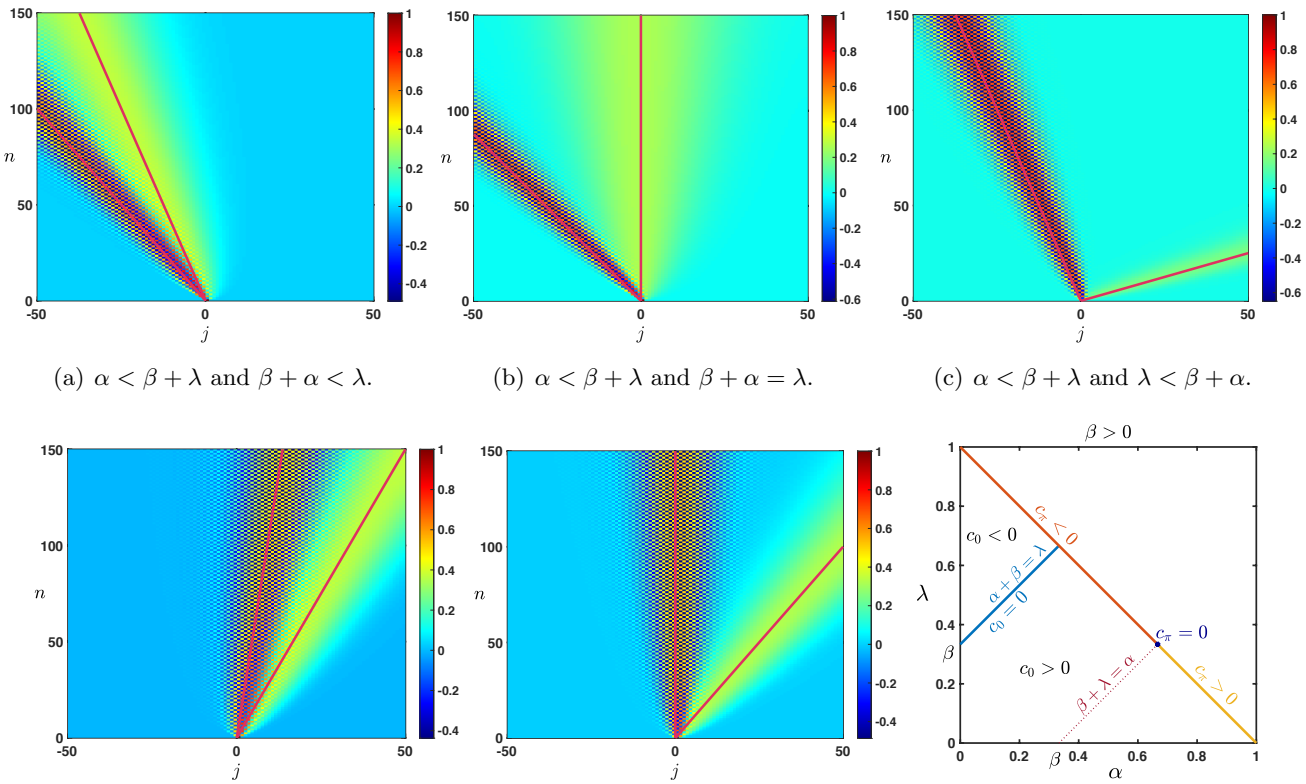
Once again, as in the case with $\beta = 0$, we can precise the behavior of the fundamental solution by using the combined results of [10, 12].

- Case: $0 \leq \lambda + \alpha < 1$. There exist some universal constants $C, \kappa > 0$ and $L > 0$ such that

$$\mathcal{G}_j^n = \frac{1}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|j - c_0 n|^2}{4\sigma_0 n}\right) + \mathcal{N}_j^n, \quad -n \leq j \leq Ln,$$

  where the remainder term satisfies a Gaussian estimate

$$|\mathcal{N}_j^n| \leq \frac{C}{n} \exp\left(-\kappa \frac{|j - c_0 n|^2}{n}\right), \quad -n \leq j \leq Ln.$$



(a) $\alpha < \beta + \lambda$ and $\beta + \alpha < \lambda$. (b) $\alpha < \beta + \lambda$ and $\beta + \alpha = \lambda$. (c) $\alpha < \beta + \lambda$ and $\lambda < \beta + \alpha$.

(d) $\beta + \lambda < \alpha$. (e) $\beta + \lambda = \alpha$. (f)

Figure 6: *Effects of turning on $\beta > 0$ when $\alpha + \lambda = 1$. We now observe a secondary wave with associated wave speed c_π whose sign depends on the competition between α and $\beta + \lambda$. (a)-(b)-(c) When $\alpha < \beta + \lambda$, the wave speed of the secondary wave always verifies $c_\pi < 0$, and the competition between λ and $\beta + \alpha$ gives the direction of the primary wave as previously reported in Figure 5. (d) When $\beta + \lambda < \alpha$ which always implies that $\alpha + \beta > \lambda$, we have $0 < c_\pi < c_0$ traducing forward propagation for both waves. We remark that the secondary wave is slower. (e) When $\beta + \lambda = \alpha$ which also implies that $\alpha + \beta > \lambda$, we get $0 = c_\pi < c_0$ such that the secondary wave is blocked. (f) Summary of the sign of the wave speeds c_0 and c_π when $\beta > 0$ when $\alpha + \lambda \leq 1$.*

While for $j > nL$ we simply get a pure exponential bound

$$\left|\mathcal{G}_j^n\right| \leq Ce^{-\kappa n - \kappa j}, \quad nL < j.$$

Inspecting the formula for c_0, we notice that when $\alpha + \beta \lesseqgtr \lambda$ we have $c_0 \lesseqgtr 0$ and the wave speed vanishes precisely when $\alpha + \beta = \lambda$. This is illustrated in Figure 5, where we see that α and β, both propagating signals in the forward (rightward) direction, compete with λ carrying the feedback (leftward) prediction signals; this competition determines the main direction of propagation of neural activity in the system.

- Case: $\lambda + \alpha = 1$. What changes in that case is the existence of a secondary wave with associated wave speed c_π whose sign depends on the competition between α and $\beta + \lambda$. When $\alpha < \beta + \lambda$ then we have $c_\pi < 0$, and the competition between λ and $\beta + \alpha$ will determine the sign of c_0, as illustrated in panels (a)-(b)-(c) of Figure 6. On the other hand, when $\beta + \lambda < \alpha$ implying that $c_\pi > 0$, we note that $\alpha + \beta > \lambda$ and thus $c_0 > 0$. In that case, the explicit formula for c_π and c_0 shows that

$0 < c_\pi < c_0$ and the secondary wave associated to c_π is slower to propagate into the network, see Figure 6(d). Finally, when $\beta + \lambda = \alpha$ we have $0 = c_\pi < c_0$ and the secondary wave is blocked, see Figure 6(e).

We have summarized in the diagram of Figure 6(f) all possible configurations for the sign of the wave speeds c_0 and c_π when $\beta \in (0, 1)$ when $\alpha + \lambda \leq 1$. We notably observe that when β is increased the region of parameter space where $c_0 < 0$ diminishes while the region of parameter space where $c_\pi < 0$ increases, indicating that for high values of β the primary wave is most likely to be forward while the secondary wave is most likely to be backward.

3.2 Wave propagation on a semi-infinite network with a forcing source term

Now that we have understood the intrinsic underlying mechanisms of wave propagation for our model (3.1) set on an infinite domain, we turn to the case where the network is semi-infinite. That is, the network admits an *input* layer that is only connected to the layer above. The problem now reads

$$\begin{cases} e_j^{n+1} - \beta e_{j-1}^{n+1} = \alpha e_{j-1}^n + (1 - \beta - \lambda - \alpha) e_j^n + \lambda e_{j+1}^n, \quad j \geq 1, \quad n \geq 0, \\ e_0^n = s_0^n, \quad n \geq 0, \\ e_j^0 = h_j, \quad j \geq 1. \end{cases} \tag{3.9}$$

We see that the system depends on the source term s_0^n applied to its input layer at each time step, also called a *boundary value*, and on the starting activation value (h_j) applied to each layer at the initial time point, also called the *initial value*. In fact, the linearity principle tells us that the solutions of the above problem can be obtained as the linear superposition of the solutions to the following two problems, the boundary value problem, where all layers except the input layer are initialized at zero:

$$\begin{cases} g_j^{n+1} - \beta g_{j-1}^{n+1} = \alpha g_{j-1}^n + (1 - \beta - \lambda - \alpha) g_j^n + \lambda g_{j+1}^n, \quad j \geq 1, \quad n \geq 0, \\ g_0^n = s_0^n, \quad n \geq 0, \\ g_j^0 = 0, \quad j \geq 1, \end{cases} \tag{3.10}$$

and the initial value problem, where the input layer source term is set to zero for all time steps:

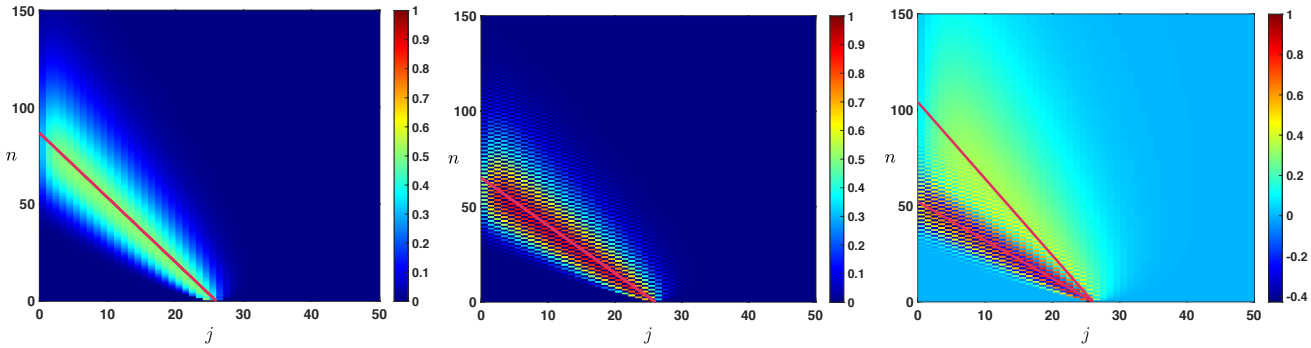
$$\begin{cases} f_j^{n+1} - \beta f_{j-1}^{n+1} = \alpha f_{j-1}^n + (1 - \beta - \lambda - \alpha) f_j^n + \lambda f_{j+1}^n, \quad j \geq 1, \quad n \geq 0, \\ f_0^n = 0, \quad n \geq 0, \\ f_j^0 = h_j, \quad j \geq 1. \end{cases} \tag{3.11}$$

Subsequently, the generic solution sequence $(e_j^n)_{j \geq 1}$ can be obtained as

$$e_j^n = f_j^n + g_j^n, \quad j \geq 1, \quad n \geq 1.$$

3.2.1 The initial value problem (3.11)

It is first natural to investigate the initial value problem (3.11) since it is really close to the infinite network case of the previous section. Here, we consider the effect of the initial value assigned to each layer $j > 0$



(a) (b) (c)

Figure 7: *Space-time evolution of the rescaled solution sequence* $(\sqrt{n}\,\mathcal{G}^n_{\rm ivp}(j,j_0))_{j\geq 1}$ *to* (3.11) *starting with a Dirac delta sequence at* $j_0 = 25$ *in different cases with leftward propagation. (a)* $\beta = 0$ *&* $\alpha+\lambda < 1$ *with* $c_0 < 0$*. (b)* $\beta = 0$ *&* $\alpha+\lambda = 1$ *with* $c_0 < 0$*. (c)* $0 < \beta < 1$ *&* $\alpha+\lambda = 1$ *with* $-1 < c_\pi < c_0 < 0$*.*

at the first time step ($n = 0$), except the input layer ($j = 0$) which is set to zero. The dynamics of (3.11) is still read out from the amplification factor function ρ defined in (3.6) and once again the solutions to (3.11) can be obtained as the convolution of the initial sequence with the fundamental solution associated to the problem. For $j_0 \geq 1$, we denote by $\boldsymbol{\delta}^{j_0}$ the Dirac delta sequence defined as $\boldsymbol{\delta}^{j_0}_{j_0} = 1$ and $\boldsymbol{\delta}^{j_0}_{j} = 0$ for all $j \geq 1$ and $j \neq j_0$. Correspondingly, we denote by $\mathcal{G}^n_{\rm ivp}(\cdot, j_0) = (\mathcal{G}^n_{\rm ivp}(j,j_0))_{j\geq 1}$ the solution to (3.11) starting from $\boldsymbol{\delta}^{j_0}$, and let us remark that the solutions to (3.11) starting from any initial condition $(h_j)_{j\geq 1}$ can be represented as

$$f^n_j = \sum_{j_0=1}^{+\infty} \mathcal{G}^n_{\rm ivp}(j,j_0) h_{j_0}, \quad j \geq 1, \quad n \geq 1.$$

Combining the results of [10, 12] together with those of [13, 16, 17] which precisely deal with recurrence equations with boundary conditions, one can obtain very similar results as in the previous case. The very first obvious remark that we can make is that for all $j, j_0 \geq 1$ and $1 \leq n < j_0$ we have

$$\mathcal{G}^n_{\rm ivp}(j,j_0) = \mathcal{G}^n_{j-j_0},$$

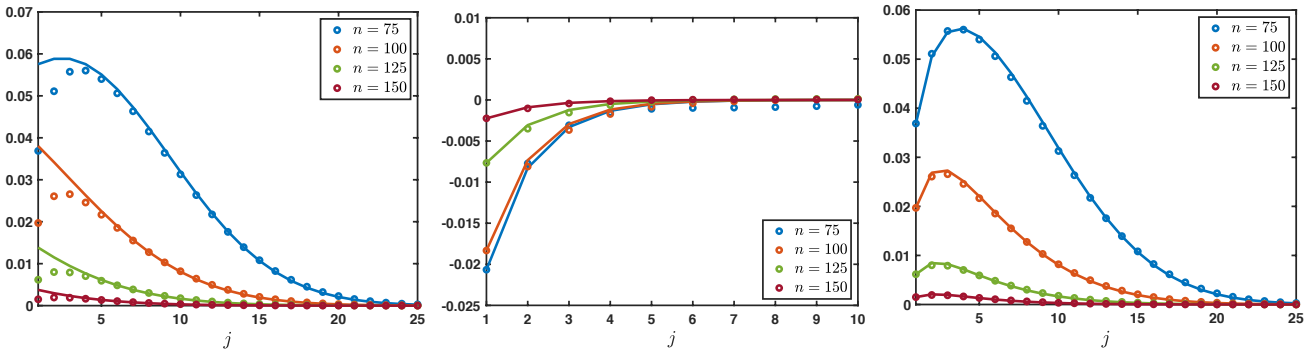
meaning that it takes $n = j_0$ iterations before the solution arrives at the boundary $j = 0$ and for $1 \leq n < j_0$ the problem is similar to the one set on the infinite network. This behavior is illustrated in Figure 7 for several values of the hyper-parameters where we represent the spatio-temporal evolution of the rescaled solution sequence $(\sqrt{n}\,\mathcal{G}^n_{\rm ivp}(j,j_0))_{j\geq 1}$. We clearly observe a Gaussian behavior before the solution reaches the boundary. And for all $n \geq j_0$, we can write

$$\mathcal{G}^n_{\rm ivp}(j,j_0) = \mathcal{G}^n_{j-j_0} + \mathcal{G}^n_{\rm bl}(j,j_0),$$

where $\mathcal{G}^n_{\rm bl}(j,j_0)$ is a remainder term generated by the boundary condition at $j = 0$. It is actually possible to bound $\mathcal{G}^n_{\rm bl}(j,j_0)$ in each of the cases treated above.

When $\beta = 0$ and $\alpha+\lambda < 1$ with $\alpha < \lambda$ such that $c_0 < 0$, then $\mathcal{G}^n_{\rm bl}(j,j_0)$ is well approximated by

$$\mathcal{G}^n_{\rm bl}(j,j_0) \approx \begin{cases} -\frac{1}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|-j_0-c_0 n|^2}{4\sigma_0 n}\right) \left(\frac{\alpha}{\lambda}\right)^j, & 1 \leq j \leq j_0, \\ e^{-\kappa n - \kappa(j-j_0)}, & j > j_0, \end{cases}$$



(a) Gaussian approximation. (b) Visualization of the boundary layer. (c) Gaussian & boundary layer approximation.

Figure 8: *Illustration of the solution* $\mathcal{G}^n_{\rm ivp}(j,j_0)$ *to* (3.11) *in the case where* $\beta = 0$ *and* $\alpha+\lambda<1$ *with* $\alpha<\beta$. *(a) Visualizations of the solution* $\mathcal{G}^n_{\rm ivp}(j,j_0)$ *(circles) at different time iterations. The plain lines correspond to the Gaussian approximation* $\frac{1}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|j-j_0-c_0n|^2}{4\sigma_0 n}\right)$ *and remark the presence of a boundary layer (seen as a mismatch between the circles and the Gaussian lines approximation). (b) We represent the boundary layer by plotting (circles)* $\mathcal{G}^n_{\rm ivp}(j,j_0) - \frac{1}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|j-j_0-c_0n|^2}{4\sigma_0 n}\right)$ *and we compare it to our boundary layer approximation* $-\frac{1}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|-j_0-c_0n|^2}{4\sigma_0 n}\right)\left(\frac{\alpha}{\lambda}\right)^j$ *(plain lines). (c) Finally we compare the solution* $\mathcal{G}^n_{\rm ivp}(j,j_0)$ *(circles) to its first order approximation* $\frac{1}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|j-j_0-c_0n|^2}{4\sigma_0 n}\right) - \frac{1}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|-j_0-c_0n|^2}{4\sigma_0 n}\right)\left(\frac{\alpha}{\lambda}\right)^j$ *(plain lines).*

while when $\lambda<\alpha$ with $c_0>0$, then $\mathcal{G}^n_{\rm bl}(j,j_0)$ is well approximated by

$$\mathcal{G}^n_{\rm bl}(j,j_0) \approx \begin{cases} e^{-\kappa n-\kappa(j_0-j)}, & 1\le j\le j_0,\\ -\frac{1}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|j-c_0n|^2}{4\sigma_0 n}\right)\left(\frac{\lambda}{\alpha}\right)^{j_0}, & j_0<j,\end{cases}$$

this is illustrated in Figure 8 in the case $c_0<0$.

On the other hand for $\alpha+\lambda=1$ with $\alpha<\lambda$ such that $c_0<0$, then $\mathcal{G}^n_{\rm bl}(j,j_0)$ is well approximated by

$$\mathcal{G}^n_{\rm bl}(j,j,_0) \approx \begin{cases} -\frac{1+(-1)^n}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|-j_0-c_0n|^2}{4\sigma_0 n}\right)\left(\frac{\alpha}{\lambda}\right)^{j}, & 1\le j\le j_0,\\ e^{-\kappa n-\kappa(j-j_0)}, & j>j_0,\end{cases}$$

while when $\lambda<\alpha$ with $c_0>0$, then $\mathcal{G}^n_{\rm bl}(j,j_0)$ is well approximated by

$$\mathcal{G}^n_{\rm bl}(j,j_0) \approx \begin{cases} e^{-\kappa n-\kappa(j_0-j)}, & 1\le j\le j_0,\\ -\frac{1+(-1)^n}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|j-c_0n|^2}{4\sigma_0 n}\right)\left(\frac{\lambda}{\alpha}\right)^{j_0}, & j_0<j.\end{cases}$$

When $0<\beta<1$ and $\alpha+\lambda<1$ the approximations are similar as for the case with $\beta=0$. We thus need to discuss three cases.

- Case $-1<c_\pi<c_0<0$. In that case, we have for $1\le j\le j_0$ that

$$\mathcal{G}^n_{\rm bl}(j,j_0) \approx -\frac{1}{\sqrt{4\pi\sigma_0 n}}\exp\left(-\frac{|-j_0-c_0n|^2}{4\sigma_0 n}\right)\left(\frac{\alpha+\beta}{\lambda}\right)^j - \frac{(-1)^n}{\sqrt{4\pi\sigma_\pi n}}\exp\left(-\frac{|-j_0-c_\pi n|^2}{4\sigma_\pi n}\right)\left(\frac{\alpha-\beta}{\lambda}\right)^j,$$

  with an exponential bound for $j>j_0$. This situation is presented in Figure 7(c)

- Case $-1 < c_\pi < 0 < c_0$. In this case we have

$$\mathcal{G}^n_{\mathrm{bl}}(j, j_0) \approx \begin{cases} -\frac{(-1)^n}{\sqrt{4\pi\sigma_\pi n}} \exp\left(-\frac{|-j_0 - c_\pi n|^2}{4\sigma_\pi n}\right) \left(\frac{\alpha-\beta}{\lambda}\right)^j, & 1 \le j \le j_0, \\ -\frac{1}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|j - c_0 n|^2}{4\sigma_0 n}\right) \left(\frac{\lambda}{\alpha+\beta}\right)^{j_0}, & j_0 < j < Ln. \end{cases}$$

- Case $-1 < 0 < c_\pi < c_0$. In this case we have

$$\mathcal{G}^n_{\mathrm{bl}}(j, j_0) \approx -\frac{1}{\sqrt{4\pi\sigma_0 n}} \exp\left(-\frac{|j - c_0 n|^2}{4\sigma_0 n}\right) \left(\frac{\lambda}{\alpha+\beta}\right)^{j_0} - \frac{(-1)^n}{\sqrt{4\pi\sigma_\pi n}} \exp\left(-\frac{|j - c_\pi n|^2}{4\sigma_\pi n}\right) \left(\frac{\lambda}{\alpha-\beta}\right)^{j_0}$$

  for $j_0 < j < Ln$.

3.2.2 The boundary value problem (3.10)

We now turn our attention to the boundary value problem (3.10) where the network is initialized with zero activity, for all layers except the input. Motivated by applications, we will only focus on the case where $s^n_0 = s_0 \in \mathbb{R}$ for all $n \ge 0$ (i.e., a constant sensory input) and thus study:

$$\begin{cases} g^{n+1}_j - \beta g^{n+1}_{j-1} = \alpha g^n_{j-1} + (1 - \beta - \lambda - \alpha) g^n_j + \lambda g^n_{j+1}, \quad j \ge 1, \quad n \ge 0, \\ g^n_0 = s_0, \quad n \ge 0, \\ g^0_j = 0, \quad j \ge 1. \end{cases} \tag{3.12}$$

Case $\beta = 0$. Here, the stimulus information s_o does not directly propagate through the network via its feedforward connections (since $\beta = 0$), but may still propagate towards higher layers $j > 0$ via the feedforward prediction error correction mechanism, governed by parameter α. When $\alpha + \lambda \le 1$, we distinguish between three cases. Here and throughout, we denote by erf the error function defined by

$$\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} \mathrm{d}z, \quad x \in \mathbb{R}.$$

- Case $\alpha < \lambda$. In this case we have

$$g^n_j = s_0 \left(\frac{\alpha}{\lambda}\right)^j \left(1 + \omega^n_j\right), \quad \text{with} \quad \left|\omega^n_j\right| \le C e^{-\kappa n - \kappa j}, \quad j \ge 1, \quad n \ge 1.$$

  It is interesting to note that the sequence $\left(s_0 \left(\frac{\alpha}{\lambda}\right)^j\right)_{j \le 1}$ is a stationary solution to (3.12) and we have uniform convergence at exponential rate toward this stationary solution, that is

$$\sup_{j \ge 1} \left| g^n_j - s_0 \left(\frac{\alpha}{\lambda}\right)^j \right| \le C e^{-\kappa n} \underset{n \to +\infty}{\longrightarrow} 0.$$

  We illustrate this uniform convergence in Figure 9 (a)-(d).

- Case $\alpha = \lambda$. We have

$$\left| g^n_j - s_0 \left(1 - \mathrm{erf}\left(\frac{j}{\sqrt{4\sigma_0 n}}\right)\right) \right| \le \frac{C}{n} \exp\left(-\kappa \frac{j^2}{n}\right), \quad j \ge 1, \quad n \ge 1.$$

In this case, we observe a slow convergence to the steady state s_0. Indeed, for each $\delta \in (0, 1/2)$ we have

$$\lim_{n\to+\infty} \sup_{1\le j\le n^\delta} \left|g_j^n - s_0\right| = 0,$$

while for any $\delta > 1/2$ we get

$$\lim_{n\to+\infty} \sup_{j\ge n^\delta} \left|g_j^n\right| = 0.$$

The propagation is thus diffusive along $j \sim \sqrt{n}$. This can be seen in Figure 9 (b)-(e).

- Case $\lambda < \alpha$. In this case we have

$$\left|g_j^n - \frac{s_0}{2}\left(1 - \operatorname{erf}\left(\frac{j - c_0 n}{\sqrt{4\sigma_0 n}}\right)\right)\right| \le \frac{C}{\sqrt{n}} \exp\left(-\kappa \frac{(j - c_0 n)^2}{n}\right), \quad j \ge 1, \quad n \ge 1.$$

In this case, we deduce that we have local uniform convergence towards the steady state s_0, actually we have spreading at speed c_0. More precisely, for any $c \in (0, c_0)$ we have

$$\lim_{n\to+\infty} \sup_{1\le j\le cn} \left|g_j^n - s_0\right| = 0,$$

while for any $c > c_0$, we get

$$\lim_{n\to+\infty} \sup_{j\ge cn} \left|g_j^n\right| = 0.$$

We refer to Figure 9 (c)-(f) for an illustration. The figure clearly shows the competition between hyperparameters α and λ, with forward propagation of the sensory input only when $\alpha \ge \lambda$.

Case $0 < \beta < 1$. Here, the stimulus information s_o propagates through the network not only via its feedforward connections (governed by $\beta > 0$) but also via the feedforward prediction error correction mechanism, governed by parameter α. In the case where $\alpha + \lambda \le 1$, the results from the case $\beta = 0$ remain valid, the only differences coming from the fact that the above approximations in the case $\lambda \le \alpha$ are only valid for $1 \le j \le Ln$ for some large constant $L > 0$ with exponential localized bounds for $j \ge Ln$ and that the steady state is now $\left(s_0 \left(\frac{\alpha+\beta}{\lambda}\right)^j\right)_{j\ge1}$ whenever $\alpha + \beta < \lambda$. This confirms that the feedforward propagation of the input s_0 is now dependent on both terms α and β, jointly competing against the feedback term λ.

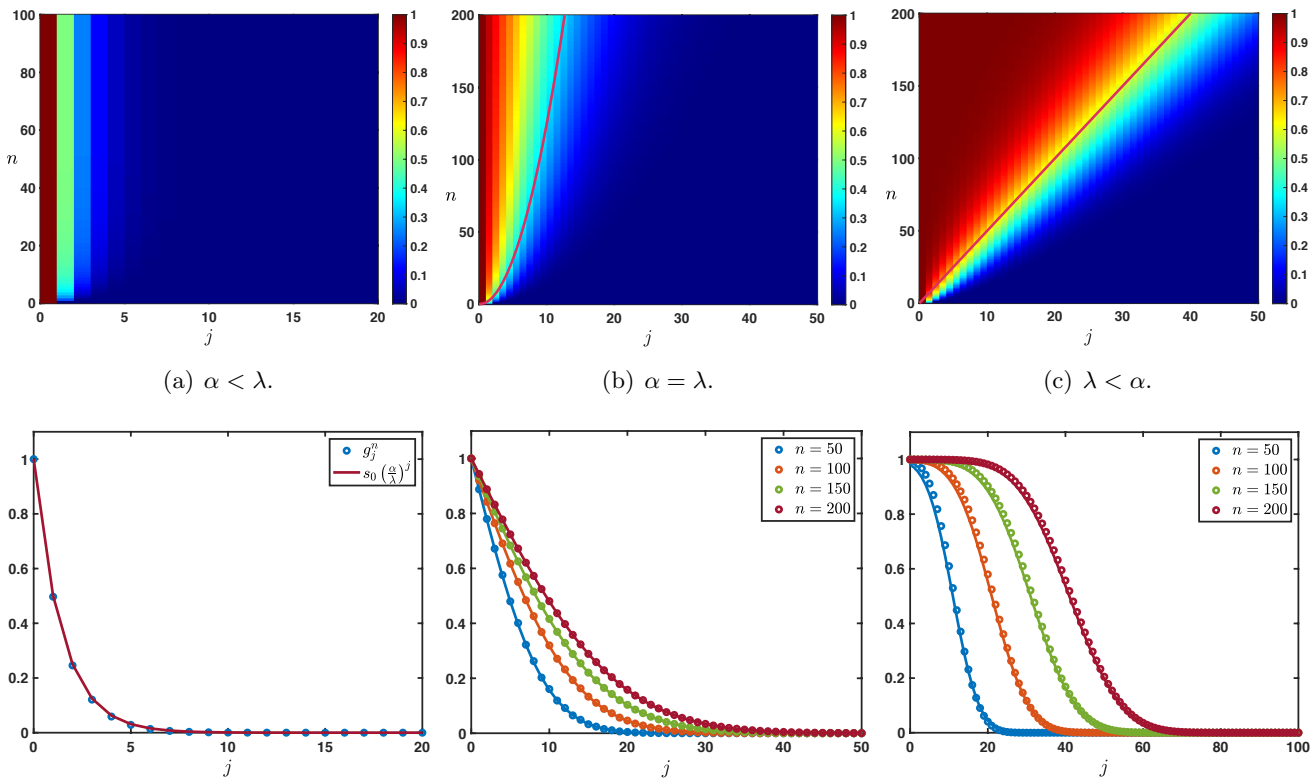
Let us remark that when $0 < \beta < \alpha - \lambda$ and in the special case $\alpha + \lambda = 1$, where a second stable point exists for the amplification factor function at $\rho(\pi)$, we can get a slightly more accurate description of the solution in the form

$$g_j^n = \frac{s_0}{2}\left(1 - \operatorname{erf}\left(\frac{j - c_0 n}{\sqrt{4\sigma_0 n}}\right)\right) - \frac{s_0}{2(1+\beta)} \frac{(-1)^j}{\sqrt{4\pi\sigma_\pi n}} \exp\left(-\frac{(j - c_\pi n)^2}{4\sigma_\pi n}\right) + r_j^n, \quad n \ge 1, \quad 1 \le j \le Ln,$$

where the remainder term satisfies an estimate of the form

$$\left|r_j^n\right| \le \frac{C}{\sqrt{n}} \exp\left(-\kappa \frac{(j - c_0 n)^2}{n}\right) + \frac{C}{n} \exp\left(-\kappa \frac{(j - c_\pi n)^2}{n}\right).$$

This is illustrated in Figure 10. It should be noted here that, while the main wavefront reflecting solution c_0 is a generic property of our network in the entire range of validity of parameters $0 \le \beta < 1$ and $\alpha + \lambda \le 1$, the second oscillatory pattern reflecting c_π only appears in the special case of $\beta \ne 0$ and $\alpha + \lambda = 1$. This oscillation is, in fact, an artifact from the discrete formulation of our problem, as will become evident in the next section, where we investigate continuous formulations of the problem.



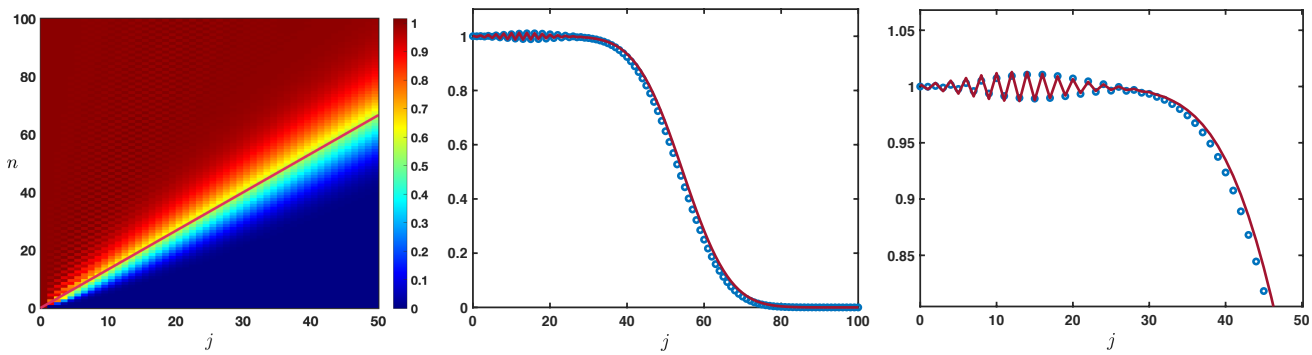
(a) $\alpha < \lambda$. (b) $\alpha = \lambda$. (c) $\lambda < \alpha$.

(d) $\alpha < \lambda$. (e) $\alpha = \lambda$. (f) $\lambda < \alpha$.

Figure 9: *Case $\beta = 0$ and $\alpha + \lambda \leq 1$. Visualization of the solution $(g_j^n)_{j\geq 1}$ of (3.12) when $s_0^n = 1$ for all $n \geq 0$. Top row: space-time plots of the solution depending on α and λ. Bottom row: solution profiles at different time iterations (circles) compared with the leading order approximations. In the case $\alpha < \lambda$, we observe uniform convergence (in space) at exponential rate (in time) toward the stationary solution $s_0\left(\left(\frac{\alpha}{\lambda}\right)^j\right)_{j\geq 1}$ while when $\lambda < \alpha$ we observe the propagation of a wavefront where the uniform steady state s_0 is propagating at speed $c_0 > 0$. In the intermediate case where $\alpha = \lambda$ we get a diffusive invasion as illustrated by the curve $j = \sqrt{4\sigma_0 n}$.*

3.3 Towards continuous predictive models

Starting from a discrete approximation of our system made sense, not only for mathematical convenience but also because artificial neural networks and deep learning systems implementing similar predictive coding principles are intrinsically discrete. Nonetheless, it can be useful to discard this discrete approximation and investigate our system in the continuous limit. Note that in the following, we will explore continuous extensions of our model in both time *and* space. Biological neural networks, like any physical system, operate in continuous time and thus it is more biologically accurate to relax the temporal discretization assumption. This is what we do in the first part of this section. In the spatial domain, however, the discretization of our system into successive processing layers was not just an approximation, but also a reflection of the hierarchical anatomy of the brain. Nonetheless, we can still represent neuronal network depth continuously, even if only as a mathematical abstraction. This is what we will do in the subsequent part of this section. Understanding such continuous limits can allow us to test the robustness of our framework, and to relate it to canonical models whose dynamics have been more exhaustively characterized.



(a) Space-time plot. (b) Profile of the solution at $n = 70$. (c) Zoom of the profile near the boundary.

Figure 10: *Case* $0 < \beta < 1$ *and* $\alpha + \lambda = 1$ *with* $0 < \beta < \alpha - \lambda$. *Visualization of the solution* $(g_j^n)_{j\geq 1}$ *of* (3.12) *when* $s_0^n = 1$ *for all* $n \geq 0$. *The solution is the super-position of a leading rightward front spreading at speed* $c_0 > 0$ *and an oscillatory Gaussian profile propagating at speed* $c_\pi > 0$ *with* $c_\pi < c_0$.

3.3.1 Continuous in time interpretation

As a first step, we present a continuous in time interpretation of the model (3.9). We let $\Delta t > 0$ be some parameter which will represent a time step and reformulate the recurrence equation as

$$(1-\beta)\frac{e_j^{n+1} - e_j^n}{\Delta t} = \frac{\beta}{\Delta t}\left(e_{j-1}^{n+1} - e_j^{n+1}\right) + \frac{\lambda}{\Delta t}\left(e_{j+1}^n - e_j^n\right) - \frac{\alpha}{\Delta t}\left(e_j^n - e_{j-1}^n\right).$$

We now interpret e_j^n as the approximation of some smooth function of time $\mathbf{e}_j(t)$ evaluated at $t_n := n\Delta t$, that is $e_j^n \sim \mathbf{e}_j(t_n)$. As a consequence, we get that

$$e_j^{n+1} \sim \mathbf{e}_j(t_{n+1}) = \mathbf{e}_j(t_n + \Delta t) = \mathbf{e}_j(t_n) + \Delta t \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{e}_j(t_n) + \mathcal{O}(|\Delta t|^2), \text{ as } \Delta t \to 0,$$

such that

$$\frac{e_j^{n+1} - e_j^n}{\Delta t} \sim \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{e}_j(t_n), \text{ as } \Delta t \to 0.$$

Now, introducing the scaled parameters

$$\widetilde{\beta} := \frac{\beta}{\Delta t}, \quad \widetilde{\lambda} := \frac{\lambda}{\Delta t}, \text{ and } \widetilde{\alpha} := \frac{\alpha}{\Delta t},$$

we get at the limit $\Delta t \to 0$ the following lattice ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{e}_j(t) = (\widetilde{\beta} + \widetilde{\alpha})\mathbf{e}_{j-1}(t) - (\widetilde{\beta} + \widetilde{\alpha} + \widetilde{\lambda})\mathbf{e}_j(t) + \widetilde{\lambda}\mathbf{e}_{j+1}(t), \quad t > 0. \tag{3.13}$$

When defined on the infinite lattice $\mathbb{Z}$, one can represent the solutions as

$$\mathbf{e}_j(t) = \sum_{k\in\mathbb{Z}} \mathbf{G}_{j-k}(t)h_k, \quad j \in \mathbb{Z}, \quad t > 0,$$

starting from the initial sequence $\mathbf{e}_j(t = 0) = (h_j)_{j\in\mathbb{Z}}$ where $(\mathbf{G}_j(t))_{j\in\mathbb{Z}}$ is the fundamental solution to (3.13) starting from the Dirac delta sequence $\boldsymbol{\delta}$. Once again, each $\mathbf{G}_j(t)$ can be represented by the inverse

Fourier transform and reads

$$\mathbf{G}_j(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\nu(\theta)t} e^{\mathbf{i}j\theta}\mathrm{d}\theta, \quad t > 0, \quad j \in \mathbb{Z},$$

where the function $\nu(\theta)$ is defined as

$$\nu(\theta) := (\widetilde{\beta} + \widetilde{\alpha})e^{-\mathbf{i}\theta} - (\widetilde{\beta} + \widetilde{\alpha} + \widetilde{\lambda}) + \widetilde{\lambda}e^{\mathbf{i}\theta}, \quad \theta \in [-\pi, \pi].$$

The function $\nu(\theta)$ serves as an amplification factor function for the time continuous equation (3.13). To ensure stability[1], one needs to impose that $\mathrm{Re}(\nu(\theta)) \leq 0$ for each $\theta \in [-\pi, \pi]$. From its formula, we obtain that

$$\mathrm{Re}(\nu(\theta)) = (\widetilde{\beta} + \widetilde{\alpha} + \widetilde{\lambda})(\cos(\theta) - 1), \quad \theta \in [-\pi, \pi],$$

such that we deduce that $\mathrm{Re}(\nu(0)) = 0$ and $\mathrm{Re}(\nu(\theta)) < 0$ for all $\theta \in [-\pi, \pi]\backslash\{0\}$. In particular, it is now evident that, contrary to the discrete case, $\nu(\pi)$ cannot be a stable solution for the continuous system (except in the trivial case where all hyperparameters $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\lambda}$ are zero). This confirms that the previously observed oscillations associated with $\rho(\pi)$ in specific cases were merely an artifact of the temporal discretization.

We note that, near the tangency point at $\theta = 0$, the function $\nu(\theta)$ has the following asymptotic expansion

$$\nu(\theta) = -(\widetilde{\beta} + \widetilde{\alpha} + \widetilde{\lambda})\mathbf{i}\theta - \frac{\widetilde{\beta} + \widetilde{\alpha} + \widetilde{\lambda}}{2}\theta^2 + \mathcal{O}(|\theta|^3), \text{ as } \theta \to 0.$$

It is also possible to prove a Gaussian approximation in that case, and following for example [6], we have

$$\mathbf{G}_j(t) = \frac{1}{\sqrt{4\pi\widetilde{\sigma_0}t}}\exp\left(-\frac{|j - \widetilde{c_0}t|^2}{4\widetilde{\sigma_0}t}\right) + \mathbf{R}_j(t), \quad j \in \mathbb{Z}, \quad t > 0,$$

with

$$|\mathbf{R}_j(t)| \leq \frac{C}{\sqrt{t}}\exp\left(-\kappa\frac{|j - \widetilde{c_0}t|^2}{t}\right), \quad j \in \mathbb{Z}, \quad t > 0,$$

for some universal constants $C > 0$ and $\kappa > 0$. Here, $\widetilde{c_0}$ and $\widetilde{\sigma_0}$ are given by

$$\widetilde{c_0} = \widetilde{\beta} + \widetilde{\alpha} - \widetilde{\lambda}, \quad \text{and} \quad \widetilde{\sigma_0} = \frac{\widetilde{\beta} + \widetilde{\alpha} + \widetilde{\lambda}}{2} > 0.$$

We remark that both $\widetilde{c_0}$ and $\widetilde{\sigma_0}$ are linked to c_0 and σ_0 (the propagation speed and spread of the solution in the case of the discrete model) in the following sense

$$\frac{c_0}{\Delta t} \to \widetilde{c_0}, \quad \text{and} \quad \frac{\sigma_0}{\Delta t} \to \widetilde{\sigma_0}, \quad \text{as } \Delta t \to 0.$$

We also note that the spatially homogeneous solutions of (3.13) are trivial in the sense that if we assume that $\mathbf{e}_j(t) = \mathbf{e}(t)$ for all $j \in \mathbb{Z}$ then the equation satisfied by $\mathbf{e}(t)$ is simply

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{e}(t) = 0.$$

Finally, we conclude by noticing that in this continuous in time regime, there is no possible oscillations either in space or time, in the sense that the fundamental solution always resembles a fixed Gaussian profile advected at wave speed $\widetilde{c_0}$. The formula for $\widetilde{c_0}$ highlights the intuitive functional relation between the propagation (or advection) direction and the "competition" between the feedforward influences $\widetilde{\alpha} + \widetilde{\beta}$ and the feedback influence $\widetilde{\lambda}$.

[1]Note that the notions of stability/unstability and marginal stability introduced in the fully discrete setting naturally extend to the semi-continuous setting.

3.3.2 Fully continuous interpretation: both in time and depth

In this section, we give a possible physical interpretation of the discrete model (3.9) via continuous transport equations, in which both time and space (i.e., neuronal network depth) are made continuous. Let us introduce $\Delta t > 0$, $\Delta x > 0$ and set $\nu := \frac{\Delta x}{\Delta t}$. As before, we can view Δt as a time step for our system; additionally, Δx can be viewed as a spatial step in the (continuous) neuronal depth dimension, and thus ν becomes akin to a neural propagation speed or a conduction velocity. We then rewrite the recurrence equation as

$$(1-\beta)\frac{e_j^{n+1}-e_j^n}{\Delta t} = \beta\nu\frac{e_{j-1}^{n+1}-e_j^{n+1}}{\Delta x} + \lambda\nu\frac{e_{j+1}^n-e_j^n}{\Delta x} - \alpha\nu\frac{e_j^n-e_{j-1}^n}{\Delta x}.$$

The key idea is to now assume that e_j^n represents an approximation of some smooth function $\mathbf{e}(t,x)$ evaluated at $t_n := n\Delta t$ and $x_j := j\Delta x$, that is $e_j^n \sim \mathbf{e}(t_n, x_j)$. Then passing to the limit $\Delta t \to 0$, $\Delta x \to 0$ with $\frac{\Delta x}{\Delta t} = \nu > 0$ fixed and assuming that $\beta + \alpha \neq \lambda$, one gets the partial differential equation

$$\partial_t \mathbf{e}(t,x) + \frac{\nu(\beta+\alpha-\lambda)}{1-\beta}\partial_x \mathbf{e}(t,x) = 0, \quad t > 0, \quad x > 0, \tag{3.14}$$

with boundary condition $\mathbf{e}(t, x = 0) = s_0(t)$ and initial condition $\mathbf{e}(t = 0, x) = h(x)$ satisfying the compatibility condition $s_0(0) = h(0)$ where $s_0(t)$ is a smooth function such that $s_0(t^n) = s_0^n$ and $h(x_j) = h_j$. The above partial differential equation is a transport equation with associated speed $\frac{\nu(\beta+\alpha-\lambda)}{1-\beta} = \nu c_0$. Depending on the sign of c_0, we have a different representation for the solutions of (3.14).

- Case $c_0 < 0$. Solution is given by

  $$\mathbf{e}(t,x) = h\left(x - \nu c_0 t\right), \quad t > 0, \quad x > 0.$$

  Let us remark that when $c_0 < 0$ the trace of the solution at $x = 0$ is entirely determined by the initial data $h(x)$ since

  $$\mathbf{e}(t, x = 0) = h\left(-\nu c_0 t\right), \quad t > 0.$$

  Intuitively, this reflects the dominance of backward (leftward) propagation in this network, with solutions determined entirely by the initial value $h(x)$, even for $x = 0$ (the source term, $s_0(t)$, having no influence in this case).

- Case $c_0 > 0$. Solution is given by

  $$\mathbf{e}(t,x) = \begin{cases} s_0\left(t - \frac{x}{\nu c_0}\right), & x \leq \nu c_0 t, \\ h\left(x - \nu c_0 t\right), & x > \nu c_0 t, \end{cases} \qquad t > 0, \quad x > 0.$$

  Intuitively, this reflects the dominance of forward (rightward) propagation in this network, with both the source term $s_0(t)$ and the initial values $h(x)$ transported at constant velocity νc_0.

Thanks to the explicit form of the solutions, we readily obtain many qualitative properties of the solution $\mathbf{e}(t,x)$. Boundedness and positivity of the solutions are inherited from the functions $s_0(t)$ and $h(x)$. In the case where $\beta + \alpha = \lambda$ (i.e., with balanced feed-forward and feedback influences), the limiting equation is slightly different. Indeed, in this case, introducing $\delta := \frac{\Delta x^2}{\Delta t}$ and letting $\Delta t \to 0$, $\Delta x \to 0$ with $\delta > 0$ fixed, on gets the partial differential equation

$$\partial_t \mathbf{e}(t,x) = \frac{\delta(\beta+\alpha+\lambda)}{2(1-\beta)}\partial_x^2 \mathbf{e}(t,x), \quad t > 0, \quad x > 0, \tag{3.15}$$

and we readily observe that when $\beta + \alpha = \lambda$, we have that

$$\frac{\beta + \alpha + \lambda}{2(1-\beta)} = \frac{\beta(1-\alpha-\lambda) + \alpha + \lambda - (\lambda - \alpha)^2}{2(1-\beta)^2} = \sigma_0 > 0.$$

We obtain a heat equation with a boundary condition $\mathbf{e}(t, x = 0) = s_0(t)$ and initial condition $\mathbf{e}(t = 0, x) = h(x)$. Upon denoting

$$\mathcal{S}(t, x, y) := \frac{1}{\sqrt{4\pi\delta\sigma_0 t}} \left(e^{-\frac{(x-y)^2}{4\delta\sigma_0 t}} - e^{-\frac{(x+y)^2}{4\delta\sigma_0 t}} \right),$$

the solution of the equation is given by

$$\mathbf{e}(t, x) = s_0(t) + \int_0^{+\infty} \mathcal{S}(t, x, y) \left(h(y) - s_0(0)\right) \mathrm{d}y - \int_0^t \int_0^{+\infty} \mathcal{S}(t-s, x, y) s_0'(s) \mathrm{d}y \mathrm{d}s, \quad t > 0, \quad x > 0,$$

Let us remark that when $s_0(t) = s_0 \in \mathbb{R}$ is constant for all $t \geq 0$, the above expression simplifies to

$$\mathbf{e}(t, x) = s_0 \left(1 - \mathrm{erf}\left(\frac{x}{\sqrt{4\delta\sigma_0 t}}\right)\right) + \int_0^{+\infty} \mathcal{S}(t, x, y) h(y) \mathrm{d}y, \quad t > 0, \quad x > 0.$$

In conclusion, this section extended our discrete model towards a continuous limit in both space and time. In the temporal domain, it allowed us to understand our stable solution as an advection behavior, and alerted us that the other apparently oscillatory solutions previously observed in specific cases were mainly due to our discretization approximation. In the spatio-temporal domain, the continuous limit (3.14) allowed us to realize that our main equation (3.1) was merely a discrete version of a transport equation.
In the following sections, we will systematically return to discrete implementations (with gradually increasing functionality), before considering, again, their continuous formulations.

4 Beyond the identity case

In the previous section we have studied in depth the case where $\mathcal{W}^f$ and $\mathcal{W}^b$ are both the identity matrix: each neuron in any given layer directly conveys its activation value to a single corresponding neuron in the next layer, and to a single neuron in the previous layer. Motivated by concrete implementations of the model in deep neural networks [9, 32], we aim to investigate more realistic situations with more complex connectivity matrices. While the generic unconstrained case (i.e. two unrelated and dense connection matrices $\mathcal{W}^f$ and $\mathcal{W}^b$) does not easily lend itself to analytical study, we will consider here two situations of practical interest: in the first one, the forward and backward connection matrices are symmetric and identical; in the second case, each matrix is symmetric, but the two are not necessarily identical.

4.1 The symmetric Rao & Ballard case

Following the pioneering work of Rao & Ballard [27], we will assume in this section that $\mathcal{W}^f = (\mathcal{W}^b)^{\mathbf{t}}$ and $\mathcal{W}^f$ is symmetric, which implies that

$$\mathcal{W}^f = \mathcal{W}^b \in \mathscr{S}_d(\mathbb{R}),$$

where we denoted $\mathscr{S}_d(\mathbb{R})$ the set of symmetric matrices on $\mathbb{R}^d$.

The underlying interpretation is that, if a strong synaptic connection exists from neuron a to neuron b, then there is also a strong connection from b to a. This assumption, which follows from Hebbian plasticity rules ("neurons that fire together wire together") does not capture all of the diversity of brain connectivity patterns, but can be considered a good first approximation [27].

4.1.1 Neural basis change and neural assemblies

The spectral theorem ensures that $\mathcal{W}^f$ (and thus $\mathcal{W}^b$) is diagonalizable in an orthonormal basis. Namely, there exists an orthogonal invertible matrix $P \in \mathscr{M}_d(\mathbb{R})$ such that $PP^{\mathbf{t}} = P^{\mathbf{t}}P = \mathbf{I}_d$ and there exists a diagonal matrix denoted $\mathcal{D} \in \mathscr{M}_d(\mathbb{R})$ such that

$$\mathcal{W}^f = \mathcal{W}^b = P\mathcal{D}P^{\mathbf{t}}.$$

We denote by $\gamma_p \in \mathbb{R}^d$, $p = 1, \cdots, d$ the diagonal elements of $\mathcal{D}$ and without loss of generality we may assume that

$$\gamma_1 \leq \cdots \leq \gamma_d.$$

Thanks to this diagonalization, we can now perform a change of basis for our neuronal space. We set $\mathcal{U}_j^n := P^t\mathcal{E}_j^n$ as the new basis, with $P\mathcal{U}_j^n := \mathcal{E}_j^n$. Each $\mathcal{U}_j^n$ can now be understood as a neural *assembly*, reflecting one of the principal components of the weight matrix $\mathcal{W}^f = \mathcal{W}^b$. Importantly, although assemblies may overlap, activity updates induced by feedforward or feedback connections to one given assembly do not affect the other assemblies, since the matrix P is orthogonal. Therefore, our problem is much simplified when considering activity update equations at the level of these neural assemblies $\mathcal{U}_j^n$ rather than across individual neurons $\mathcal{E}_j^n$. Our model (2.1) becomes

$$\mathcal{U}_j^{n+1} = \beta\mathcal{D}\mathcal{U}_{j-1}^{n+1} + \alpha\mathcal{D}\mathcal{U}_{j-1}^n + \left[(1-\beta-\lambda)\mathbf{I}_d - \alpha\mathcal{D}^2\right]\mathcal{U}_j^n + \lambda\mathcal{D}\mathcal{U}_{j+1}^n.$$

Note that, because all matrices in the above equation are diagonal, we have totally decoupled the d components of the vector $\mathcal{U}_j^n$. More precisely, by denoting $u_{j,p}^n$ the pth component of $\mathcal{U}_j^n$, that is $\mathcal{U}_j^n = (u_{j,1}^n, \cdots, u_{j,d}^n)^{\mathbf{t}}$, we obtain

$$u_{j,p}^{n+1} - \beta\gamma_p u_{j-1,p}^{n+1} = \alpha\gamma_p u_{j-1,p}^n + (1-\beta-\lambda-\alpha\gamma_p^2)u_{j,p}^n + \lambda\gamma_p u_{j+1,p}^n, \quad p = 1, \cdots, d.$$

This indicates that one needs to study

$$u_j^{n+1} - \beta\gamma u_{j-1}^{n+1} = \alpha\gamma u_{j-1}^n + (1-\beta-\lambda-\alpha\gamma^2)u_j^n + \lambda\gamma u_{j+1}^n, \tag{4.1}$$

where $\gamma \in \mathbb{R}$ is a given parameter. Here, γ can be thought of as the connection strength across layers (both feedforward and feedback, since we assumed here symmetric connectivity) of the neural assembly under consideration. By construction, each assembly in a given layer is only connected to the corresponding assembly in the layer above, and similarly in the layer below. Note that when $\gamma = 1$, we encounter again the exact situation that we studied in the previous section (3.9), but now with neural assemblies in lieu of individual neurons.

4.1.2 Study of the amplification factor function

Based on our previous analysis, the behavior of the solutions to (4.1) are intrinsically linked to the properties of the amplification factor function:

$$\rho_\gamma(\theta) = \frac{\alpha\gamma\left(e^{-\mathbf{i}\theta} - \gamma\right) + 1 - \beta + \lambda\left(\gamma e^{\mathbf{i}\theta} - 1\right)}{1 - \beta\gamma e^{-\mathbf{i}\theta}}, \quad \theta \in [-\pi, \pi],$$

where one needs to ensure that $|\rho_\gamma(\theta)| \leq 1$ for all $\theta \in [-\pi, \pi]$. The very first condition is to ensure that $1 \neq \beta|\gamma|$ (to avoid division by zero). Next, we investigate the behavior of $\rho_\gamma(\theta)$ at $\theta = 0$ and check under which condition on γ we can ensure that $-1 \leq \rho_\gamma(0) \leq 1$. We have

$$\rho_\gamma(0) = 1 + \frac{(1-\gamma)(\alpha\gamma - \lambda - \beta)}{1 - \beta\gamma},$$

which readily tells us that $\rho_\gamma(0) = 1$ if and only if $\gamma = 1$ or $\gamma = \frac{\lambda+\beta}{\alpha}$. And on the other hand $\rho_\gamma(0) = -1$ if and only if $\gamma = \gamma^0_\pm$ with

$$\gamma^0_\pm = \frac{\lambda + \alpha - \beta \pm \sqrt{(\lambda + \alpha - \beta)^2 + 4\alpha(2 - \lambda - \beta)}}{2\alpha},$$

with $\gamma^0_- < 0 < \gamma^0_+$ since $\lambda + \beta < 2$ by assumption. One also notices that

$$(\lambda + \alpha - \beta)^2 + 4\alpha(2 - \lambda - \beta) = (\lambda + 3\alpha - \beta)^2 + 8\alpha(1 - \alpha - \lambda),$$

such that either $\alpha + \lambda = 1$ and in that case $\gamma^0_- = -1$ and $\gamma^0_+ = 1 + \frac{1-\beta}{\alpha} > 1$, or $\alpha + \lambda < 1$ and in that case $\gamma^0_- < -1$ and $\gamma^0_+ > 1 + \frac{1-\beta}{\alpha} > 1$. Next, we remark that

$$\rho_\gamma(\pi) = 1 - \frac{(1+\gamma)(\alpha\gamma + \lambda + \beta)}{1 + \beta\gamma} = \rho_{-\gamma}(0),$$

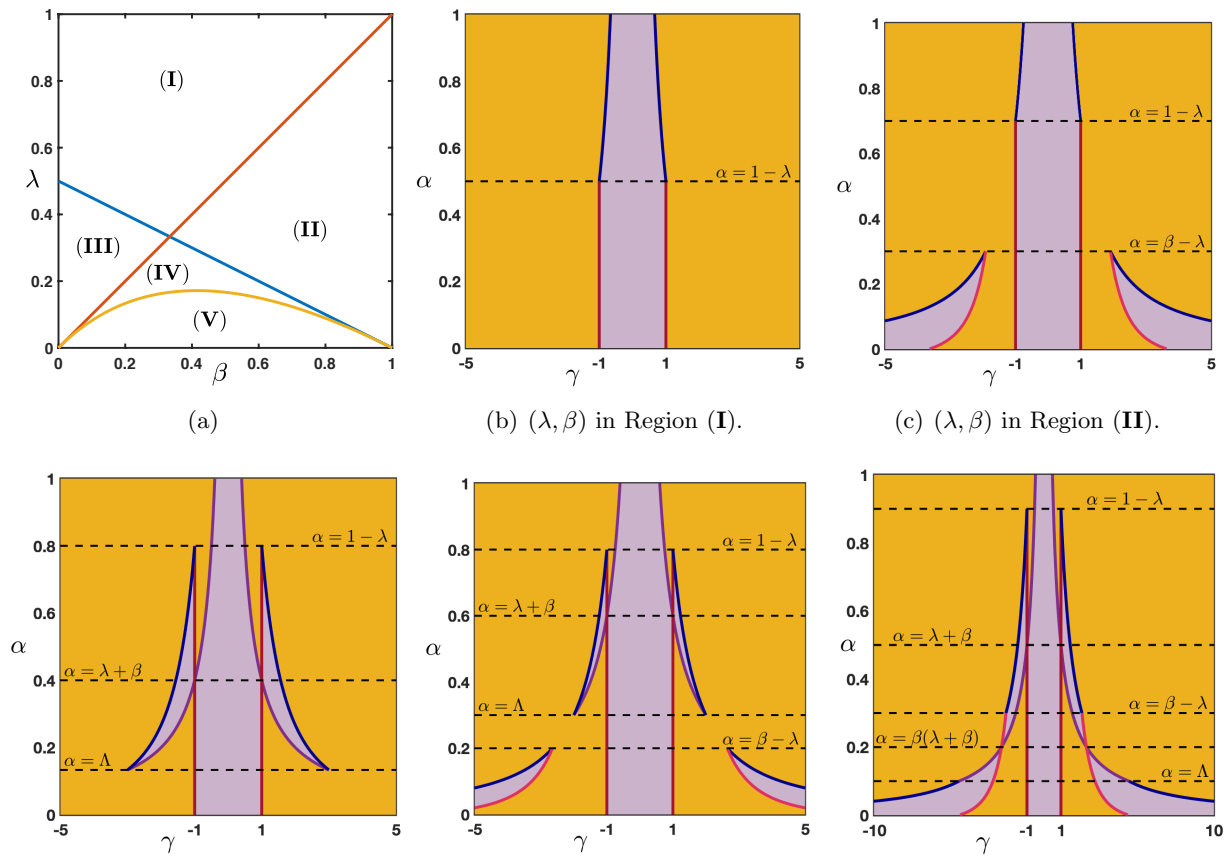
which then implies that $\rho_\gamma(\pi) = 1$ if and only if $\gamma = -1$ or $\gamma = -\frac{\lambda+\beta}{\alpha}$ and $\rho_\gamma(\pi) = -1$ if and only if $\gamma = -\gamma^0_\pm$.

As explained in the beginning, our aim is to completely characterize under which conditions on $\gamma \in \mathbb{R}$, $0 \leq \beta < 1$ and $0 < \alpha, \lambda < 1$ with $\alpha + \lambda \leq 1$, one can ensure that $|\rho_\gamma(\theta)| \leq 1$ for all $\theta \in [-\pi, \pi]$.

Potential regions of marginal stability are thus given by those values of the parameters satisfying $\gamma = \pm 1$, $\gamma = \pm\frac{\lambda+\beta}{\alpha}$, $\gamma = \pm\gamma^0_+$ and $\gamma = \pm\gamma^0_-$, and it is important to determine the intersections among the above regions. We have already proved that $\gamma^0_- = -1$ whenever $\alpha = 1 - \lambda$. Next, we compute that $\gamma^0_- = -\frac{\lambda+\beta}{\alpha}$ whenever $\alpha = \frac{\lambda(\lambda+\beta)}{1-\lambda-\beta} =: \Lambda$, while $\gamma^0_- = -\gamma^0_+$ whenever $\alpha = \beta - \lambda$ and $\gamma^0_+ = \frac{\lambda+\beta}{\alpha}$ when $\alpha = \beta(\lambda + \beta)$. Let us already point out that Λ is only defined if $\lambda + \beta < 1$ and in that case $\Lambda > 0$.

We now introduce five regions in the quadrant $(\beta, \lambda) \in [0, 1) \times (0, 1)$ which are depicted in Figure 11(a). First, since $1 - \lambda = \Lambda = \lambda + \beta$ if and only if $2\lambda + \beta = 1$ (which corresponds to the blue line in Figure 11(a)), we deduce that when $2\lambda + \beta \geq 1$ we have $1 - \lambda \leq \min(\Lambda, \lambda + \beta)$ which leads us to define the following two regions:

$$(\mathbf{I}) := \{(\beta, \lambda) \in [0, 1) \times (0, 1) \mid 2\lambda + \beta \geq 1 \text{ and } \beta \leq \lambda\},$$
$$(\mathbf{II}) := \{(\beta, \lambda) \in [0, 1) \times (0, 1) \mid 2\lambda + \beta \geq 1 \text{ and } \beta > \lambda\}.$$



(a)

(b) (λ, β) in Region (**I**).

(c) (λ, β) in Region (**II**).

(d) (λ, β) in Region (**III**).

(e) (λ, β) in Region (**IV**).

(f) (λ, β) in Region (**V**).

Figure 11: *Stability/instability regions and their boundaries as a function of* (α, γ) *for* (4.1) *while* (λ, β) *being fixed in one the five regions given in panel (a). Shaded orange regions correspond to an instability for* (4.1) *while purple regions correspond to a stability for* (4.1)*. The boundaries of the stability/instability regions are given by the intersections of the parametrized curves* $\gamma = \pm 1$ *(dark red curves),* $\gamma = \pm\gamma_-^0$ *(dark blue curves),* $\gamma = \pm\gamma_+^0$ *(pink curves) and* $\gamma = \pm\frac{\lambda+\beta}{\alpha}$ *(magenta curves). Note that the region of interest is* $0 < \alpha \leq 1 - \lambda$*. Along each such parametrized curves equation* (4.1) *is marginally stable.*

Now, when $2\lambda + \beta < 1$, we have the strict ordering $0 < \Lambda < \lambda + \beta < 1 - \lambda$ and when $\beta > \lambda$ it is thus necessary to compare Λ to $\beta - \lambda$. We remark that $\Lambda = \beta - \lambda = \beta(\lambda + \beta)$ if and only if $\beta^2 - \beta(1 - \lambda) + \lambda = 0$, which corresponds to the yellow parabola in Figure 11(a). We thus define the following three regions:

$$
\begin{aligned}
(\mathbf{III}) &:= \{(\beta, \lambda) \in [0, 1) \times (0, 1) \mid 2\lambda + \beta < 1 \text{ and } \beta \leq \lambda\}, \\
(\mathbf{IV}) &:= \left\{(\beta, \lambda) \in [0, 1) \times (0, 1) \mid 2\lambda + \beta < 1, \beta > \lambda \text{ and } \beta^2 - \beta(1 - \lambda) + \lambda \geq 0\right\}, \\
(\mathbf{V}) &:= \left\{(\beta, \lambda) \in [0, 1) \times (0, 1) \mid 2\lambda + \beta < 1, \beta > \lambda \text{ and } \beta^2 - \beta(1 - \lambda) + \lambda < 0\right\}.
\end{aligned}
$$

Note that when (β, λ) is in region (**IV**), we have the ordering

$$0 < \beta - \alpha \leq \Lambda < \lambda + \beta < 1 - \lambda,$$

while for (β, λ) in region (**V**), we have

$$0 < \Lambda < \beta(\lambda + \beta) < \beta - \alpha < \lambda + \beta < 1 - \lambda.$$

	$\gamma=1$	$\gamma=-1$	$\gamma=\frac{\lambda+\beta}{\alpha}$	$\gamma=-\frac{\lambda+\beta}{\alpha}$	$\gamma=\gamma^0_\pm$	$\gamma=-\gamma^0_\pm$
c_0^γ	$\frac{\beta+\alpha-\lambda}{1-\beta}$	$\frac{\alpha-\lambda-\beta}{1+\beta}$	$\frac{(\beta+\lambda)(\beta+\alpha-\lambda)}{\alpha-\beta(\lambda+\beta)}$		$\frac{\gamma^0_\pm(\beta+\lambda-\alpha)}{1-\beta\gamma^0_\pm}$	
c_π^γ	$\frac{\alpha-\lambda-\beta}{1+\beta}$	$\frac{\beta+\alpha-\lambda}{1-\beta}$		$\frac{(\beta+\lambda)(\beta+\alpha-\lambda)}{\alpha-\beta(\lambda+\beta)}$		$\frac{\gamma^0_\pm(\beta+\lambda-\alpha)}{1-\beta\gamma^0_\pm}$

Table 1: *Expressions of the wave speed c_0^γ and c_π^γ for values of γ corresponding the boundaries of the stability regions from Figure 11. Let us note that c_π^1 and c_0^{-1} only exist when $\alpha+\lambda=1$.*

We can characterize the stability of our equation separately for each of the five regions defined in Figure 11(a). Since the region already determines the value of the parameters β and λ, the stability will be expressed as a function of the two remaining parameters α and γ (Figure 11(b-f)). We refer to Figures 11(b-f) for a comprehensive representation of the stability regions. Note that the boundaries of the stability/instability regions are precisely given by the intersections of the parametrized curves $\gamma=\pm1$ (dark red curves), $\gamma=\pm\gamma^0_-$ (dark blue curves), $\gamma=\pm\gamma^0_+$ (pink curves) and $\gamma=\pm\frac{\lambda+\beta}{\alpha}$ (magenta curves). Along each such parametrized curves equation (4.1) is marginally stable. We comment below the case (λ,β) in Region (**III**). The other cases can be described in the same way, but we leave this out for conciseness.

Suppose that (λ,β) belongs to Region (**III**). We present the results of Figure 11(d) by letting α vary between 0 and $1-\lambda$ and $\gamma\in\mathbb{R}$. More precisely, for each fixed $\alpha\in(0,1-\lambda)$ we investigate the stability properties as a function of γ. We have to distinguish between several subcases.

(i) If $0<\alpha<\Lambda$. Then, equation (4.1) is stable for each $\gamma\in(-1,1)$, unstable for $|\gamma|>1$ and marginally stable at $\gamma=\pm1$ with $\rho_1(0)=1$ and $\rho_{-1}(\pm\pi)=1$.

(ii) If $\alpha=\Lambda$. Then $\gamma^0_-=-\frac{\lambda+\beta}{\alpha}$ and equation (4.1) is stable for each $\gamma\in(-1,1)$, unstable for $|\gamma|>|\gamma^0_-|$ and $|\gamma^0_-|>|\gamma|>1$, whereas it is marginally stable at $\gamma=\pm1$ and at $\gamma=\pm\gamma^0_-$ with $\rho_1(0)=1$, $\rho_{-1}(\pm\pi)=1$, $\rho_{\gamma^0_-}(0)=-1$ and $\rho_{-\gamma^0_-}(\pm\pi)=-1$ together with $\rho_{-\gamma^0_-}(0)=1$, $\rho_{\gamma^0_-}(\pm\pi)=1$.

(iii) If $\Lambda<\alpha<\lambda+\beta$. Then, equation (4.1) is stable for each $\gamma\in(-1,1)$ and $\frac{\lambda+\beta}{\alpha}<|\gamma|<|\gamma^0_-|$, unstable for $|\gamma|>|\gamma^0_-|$ and $\frac{\lambda+\beta}{\alpha}>|\gamma|>1$ and marginally stable at $\gamma\in\left\{\pm1,\pm\frac{\lambda+\beta}{\alpha},\pm\gamma^0_-\right\}$ with $\rho_1(0)=1$, $\rho_{-1}(\pm\pi)=1$, $\rho_{\frac{\lambda+\beta}{\alpha}}(0)=1$, $\rho_{-\frac{\lambda+\beta}{\alpha}}(\pm\pi)=1$, $\rho_{\gamma^0_-}(0)=-1$ and $\rho_{-\gamma^0_-}(\pm\pi)=-1$.

(iv) If $\alpha=\lambda+\beta$. Then, equation (4.1) is stable for each $\gamma\in(\gamma^0_-,-\gamma^0_-)\backslash\{\pm1\}$, unstable for $|\gamma|>|\gamma^0_-|$ and marginally stable at $\gamma\in\left\{\pm1,\pm\gamma^0_-\right\}$ with $\rho_1(0)=1$, $\rho_{-1}(\pm\pi)=1$, $\rho_{\gamma^0_-}(0)=-1$ and $\rho_{-\gamma^0_-}(\pm\pi)=-1$. Remark that in this case we have $\frac{\lambda+\beta}{\alpha}=1$.

(v) If $\lambda+\beta<\alpha<1-\lambda$. Then, equation (4.1) is stable for each $\gamma\in\left(-\frac{\lambda+\beta}{\alpha},\frac{\lambda+\beta}{\alpha}\right)$ and $1<|\gamma|<|\gamma^0_-|$, unstable for $|\gamma|>|\gamma^0_-|$ and $1>|\gamma|>\frac{\lambda+\beta}{\alpha}$ and marginally stable at $\gamma\in\left\{\pm1,\pm\frac{\lambda+\beta}{\alpha},\pm\gamma^0_-\right\}$ with $\rho_1(0)=1$, $\rho_{-1}(\pm\pi)=1$, $\rho_{\frac{\lambda+\beta}{\alpha}}(0)=1$, $\rho_{-\frac{\lambda+\beta}{\alpha}}(\pm\pi)=1$, $\rho_{\gamma^0_-}(0)=-1$ and $\rho_{-\gamma^0_-}(\pm\pi)=-1$.

(vi) If $\alpha=1-\lambda$. Then, equation (4.1) is stable for each $\gamma\in\left(-\frac{\lambda+\beta}{\alpha},\frac{\lambda+\beta}{\alpha}\right)$, unstable for $|\gamma|>|\gamma^0_-|=1$ and $1>|\gamma|>\frac{\lambda+\beta}{\alpha}$ and marginally stable at $\gamma\in\left\{\pm1,\pm\frac{\lambda+\beta}{\alpha}\right\}$ with $\rho_1(0)=1$ and $\rho_1(\pm\pi)=-1$, $\rho_{-1}(\pm\pi)=1$ and $\rho_{-1}(0)=-1$, with $\rho_{\frac{\lambda+\beta}{\alpha}}(0)=1$, $\rho_{-\frac{\lambda+\beta}{\alpha}}(\pm\pi)=1$. Remark that in this case we have $\gamma^0_-=-1$.

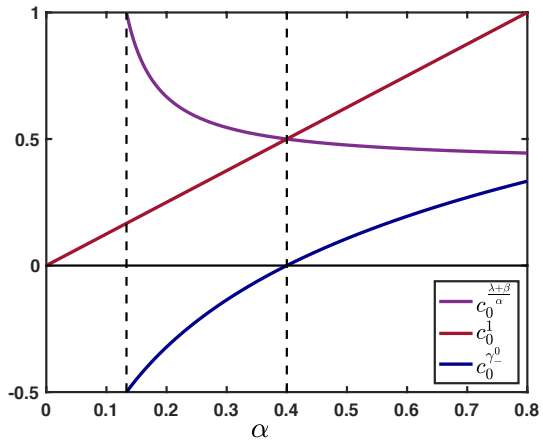


Figure 12: *Plot of the wave speeds* c_0^1, $c_0^{\frac{\lambda+\beta}{\alpha}}$ *and* $c_0^{\gamma_-^0}$ *as a function of* α *in the case* (λ, β) *in Region* **III** *associated to Figure 11(d).*

Summary. In summary, we see that stability is nearly guaranteed whenever $-1 < \gamma < 1$, regardless of the values of other parameters (as long as $0 < \alpha < 1 - \lambda$). This makes intuitive sense, as γ represents the connection strength across layers of a particular neural assembly, a connection weight $|\gamma| < 1$ implies that activity of this assembly will remain bounded across layers. Additionally, and perhaps more interestingly, in some but not all regions (e.g. Regions II and V) stability can be obtained for much larger values of $|\gamma|$; this, however, appears to coincide with low values of the α parameter. In other words, for high connection strengths $|\gamma|$, the feedforward error correction term α makes the system unstable.

4.1.3 Wave speed characterization

In the previous section (The Identity Case), we have proved that the direction of propagation was given by the sign of c_0 and c_π whenever they exist which could be read off from the behavior of $\rho_\gamma(\theta)$ near $\theta = 0$ or $\theta = \pm\pi$. We have reported the values of c_0^γ and c_π^γ for different values of γ in Table 1. For example, in Figure 12 we illustrate the changes in propagation speed and direction for c_0^γ for the case (λ, β) in Region (**III**) (as defined in Figure 11a), but the calculations remain valid for the other regions.

It is worth emphasizing that for fixed values of the hyper-parameters α, β and λ, we see here that varying γ can give rise to different propagation speeds or even different directions. As each neuronal assembly $u_{j,p}$ in a given layer j is associated with its own connection strength γ_p, it follows that different speeds and even different directions of propagation can concurrently be obtained in a single network, one for each assembly. For instance, in a given network with hyperparameters $\alpha = 0.2$, $\beta = 0.2$ and $\lambda = 0.3$ (region **III**), a neural assembly with a connection strength of $\gamma = 1$ would propagate forward at a relatively slow speed, while another with $\gamma = 2.5$ would propagate in the same direction at a much faster speed, and yet another assembly with $\gamma = \gamma_-^0 \approx -2.09$ would simultaneously propagate in the opposite backward direction.

4.1.4 Continuous in time interpretation

We can repeat the "continuous system" analysis conducted in the previous section (The Identity Case), which has lead to (3.13), but this time with Rao-Ballard connection matrices between layers. With the

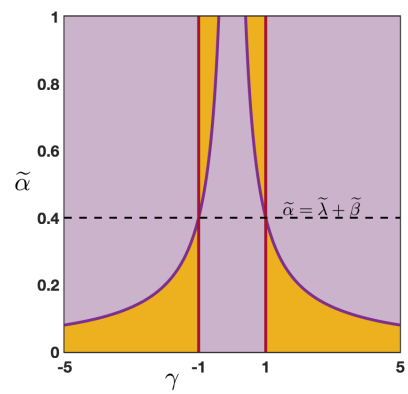


Figure 13: *Stability/instability regions and their boundaries as a function of* $(\widetilde{\alpha}, \gamma)$ *for* (4.2) *for any* $(\widetilde{\lambda}, \widetilde{\beta})$ *fixed. The shaded orange region corresponds to an instability for* (4.2) *while the purple region corresponds to a stability for* (4.2)*. The boundaries of the stability/instability regions are given by the intersections of the parametrized curves* $\gamma = \pm 1$ *(dark red curves) and* $\gamma = \pm\frac{\lambda+\beta}{\alpha}$ *(magenta curves) where equation* (4.2) *is marginally stable.*

same scaling on the hyperparameters

$$\widetilde{\beta} := \frac{\beta}{\Delta t}, \quad \widetilde{\lambda} := \frac{\lambda}{\Delta t}, \text{ and } \widetilde{\alpha} := \frac{\alpha}{\Delta t},$$

we get that, at the limit $\Delta t \to 0$, the equation (4.1) becomes the following lattice ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}_j(t) = (\widetilde{\beta} + \widetilde{\alpha})\gamma \mathbf{u}_{j-1}(t) - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma^2)\mathbf{u}_j(t) + \widetilde{\lambda}\gamma \mathbf{u}_{j+1}(t), \quad t > 0. \tag{4.2}$$

Note that the neuronal layer activity is now expressed in terms of neural assemblies $\mathbf{u}_j$ rather than individual neurons $\mathbf{e}_j$.

The amplification factor function in this case is given by

$$\nu_\gamma(\theta) = (\widetilde{\beta} + \widetilde{\alpha})\gamma e^{-\mathbf{i}\theta} - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma^2) + \widetilde{\lambda}\gamma e^{\mathbf{i}\theta}, \quad \theta \in [-\pi, \pi],$$

whose real part is given by

$$\mathrm{Re}(\nu_\gamma(\theta)) = (\widetilde{\beta} + \widetilde{\alpha} + \widetilde{\lambda})\gamma \cos(\theta) - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma^2), \quad \theta \in [-\pi, \pi].$$

When $\gamma > 0$, we observe that

$$\max_{\theta \in [-\pi,\pi]} \mathrm{Re}(\nu_\gamma(\theta)) = \mathrm{Re}(\nu_\gamma(0)) = (\widetilde{\lambda} + \widetilde{\beta} - \widetilde{\alpha}\gamma)(\gamma - 1),$$

whereas when $\gamma < 0$, we have

$$\max_{\theta \in [-\pi,\pi]} \mathrm{Re}(\nu_\gamma(\theta)) = \mathrm{Re}(\nu_\gamma(\pm\pi)) = -(\widetilde{\lambda} + \widetilde{\beta} + \widetilde{\alpha}\gamma)(\gamma + 1).$$

As a consequence, the stability analysis in this case is very simple and depends only on the relative position of γ with respect to ± 1 and $\pm\frac{\widetilde{\lambda}+\widetilde{\beta}}{\widetilde{\alpha}}$. It is summarized in Figure 13.

The simple behavior illustrated in Figure 13 for our continuous system contrasts with the number and diversity of behaviors obtained for the discrete version of the same system (Figure 11). A number of

points are worth highlighting. For instance, although the values of β and λ were critical for the discrete system (to define the region (I) to (V)), they do not affect the qualitative behavior of the continuous system. Furthermore, some observations in the continuous system appear to contradict the conclusions made previously in the discrete case. We see that stability can still be obtained with high values of the connection weight $\gamma >> 1$, but this time the stable regions coincide with high α values, whereas it was the opposite in Figure 11 panels (b),(f). This qualitative difference in behavior can be taken as a point of caution, to remind us that a discrete approximation of the system can be associated with important errors in interpretation.

Finally we note that, while stability regions are qualitatively different in the continuous case compared to the discrete approximation, the speed and direction of propagation of neural signals (reflected in the variables c_0 and c_π when they exist) remains comparable.

4.1.5 A class of examples

In this section, we provide a class of examples of $\mathcal{W}^f$ amenable to a complete analysis. Namely we consider $\mathcal{W}^f$ as the following linear combination

$$\mathcal{W}^f = \zeta \mathbf{I}_d + \xi \mathbf{A}, \tag{4.3}$$

for some $\zeta, \xi \in \mathbb{R}$ where $\mathbf{A} \in \mathscr{M}_d(\mathbb{R})$ is given by

$$\mathbf{A} = \begin{pmatrix} -2 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & -2 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & -2 & 1 \\ 0 & \cdots & \cdots & 0 & 1 & -2 \end{pmatrix}.$$

The matrix $\mathbf{A}$ is nothing but the discrete laplacian and $\mathcal{W}^f$ acts as a convolution operator on $\mathbb{R}^d$. More precisely, $\mathcal{W}^f$ combines a convolution term with a residual connection term, as in the well-known ResNet architecture [18]. Let us also note that the spectrum of $\mathbf{A}$ is well known and given by

$$\mathrm{Spec}(\mathbf{A}) = \left\{ -4\sin^2\left(\frac{p\pi}{2(d+1)}\right), \quad p = 1, \cdots, d \right\}.$$

As a consequence, the spectrum of $\mathcal{W}^f$ is simply given by

$$\mathrm{Spec}(\mathcal{W}^{\mathbf{f}}) = \left\{ \zeta - 4\xi\sin^2\left(\frac{p\pi}{2(d+1)}\right), \quad p = 1, \cdots, d \right\}.$$

One can for example set

$$\zeta = \frac{\sin^2\left(\frac{d\pi}{2(d+1)}\right) + \sin^2\left(\frac{\pi}{2(d+1)}\right)}{\sin^2\left(\frac{d\pi}{2(d+1)}\right) - \sin^2\left(\frac{\pi}{2(d+1)}\right)} \quad \text{and} \quad \xi = \frac{1}{2\left(\sin^2\left(\frac{d\pi}{2(d+1)}\right) - \sin^2\left(\frac{\pi}{2(d+1)}\right)\right)},$$

such that

$$\zeta - 4\xi\sin^2\left(\frac{d\pi}{2(d+1)}\right) = -1, \quad \zeta - 4\xi\sin^2\left(\frac{\pi}{2(d+1)}\right) = 1,$$

and for all $p = 2, \cdots, d-1$

$$\zeta - 4\xi \sin^2\left(\frac{p\pi}{2(d+1)}\right) \in (-1, 1).$$

Next, for any $p = 1, \cdots, d$ the eigenvector corresponding to the eigenvalue $-4\sin^2\left(\frac{p\pi}{2(d+1)}\right)$ is

$$U_p = \left(\cos\left(\frac{p\pi}{d+1}\right), \cdots, \cos\left(\frac{pk\pi}{d+1}\right), \cdots, \cos\left(\frac{pd\pi}{d+1}\right)\right)^{\mathbf{t}} \in \mathbb{R}^d.$$

U_p is the projection vector that corresponds to the pth neural assembly $u_{j,p}$ as defined above.
Along U_1, the recurrence equation reduces to (4.1) with $\gamma = 1$, while along U_d, the recurrence equation reduces to (4.1) with $\gamma = -1$, and we can apply the results of the previous section (the Identity case). In between (for all $1 \leq p \leq d$) we see that the eigenvalues of our connection matrix $\mathcal{W}^{\mathbf{f}}$ span the entire range between -1 and 1, that they can be explicitly computed, and thus that the stability, propagation speed and direction of activity in the corresponding neural assembly can be determined.

4.1.6 Fully continuous interpretation in time, depth and width.

For the same class of example (connection matrix composed of a convolution and residual terms), we now wish to provide a fully continuous interpretation for model (2.1) in the special case $\zeta = 1$ and ξ adjusted as follows. By fully continuous, we mean that we explore the limit of our model when not only time t, but also network depth x *and* neuronal layer width y are considered as continuous variables. Although we already presented a model that was continuous in both *time* and *depth* in subsection 3.3.2, the layers in that model only comprised a single neuron, and had no intrinsic spatial dimension. We now introduce this third continuous dimension. The starting point is to see $\mathcal{E}^n_{j,k}$, the kth element of $\mathcal{E}^n_j$, as an approximations of some continuous function $\mathcal{E}(t, x, y)$ evaluated at $t_n = n\Delta t$, $x_j = j\Delta x$ and $y_k = k\Delta y$ for some $\Delta t > 0$, $\Delta x > 0$ and $\Delta y > 0$. Let us first remark that the action of $\mathbf{A}$ on $\mathcal{E}^n_j$ is given by

$$(\mathbf{A}\mathcal{E}^n_j)_k = \mathcal{E}^n_{j,k-1} - 2\mathcal{E}^n_{j,k} + \mathcal{E}^n_{j,k+1},$$

which can be seen at a discrete approximation of $\partial^2_y \mathcal{E}(t_n, x_j, y_k)$ up to a scaling factor of order Δy^2. Once again, setting $\nu = \frac{\Delta x}{\Delta t}$ and introducing $\kappa = \frac{\Delta y^2}{\Delta t}$, we may rewrite (2.1) with $\mathcal{W}^f = \mathcal{W}^b = \mathbf{I}_d + \xi\mathbf{A}$ as

$$\begin{aligned}(1-\beta)\frac{\mathcal{E}^{n+1}_j - \mathcal{E}^n_j}{\Delta t} &= \beta\nu\frac{\mathcal{E}^{n+1}_{j-1} - \mathcal{E}^{n+1}_j}{\Delta x} + \lambda\nu\frac{\mathcal{E}^n_{j+1} - \mathcal{E}^n_j}{\Delta x} - \alpha\nu\frac{\mathcal{E}^n_j - \mathcal{E}^n_{j-1}}{\Delta x} \\ &\quad + \beta\xi\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}^{n+1}_{j-1} + \alpha\xi\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}^n_{j-1} - 2\alpha\xi\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}^n_j - \alpha\xi^2\kappa\Delta y^2\frac{\mathbf{A}}{\Delta y^2}\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}^n_j \\ &\quad + \lambda\xi\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}^n_{j+1}.\end{aligned}$$

Now letting $\Delta t \to 0$, $\Delta x \to 0$ and $\Delta y \to 0$ with ν and κ fixed, we obtain the following partial differential equation

$$\partial_t \mathcal{E}(t, x, y) + \frac{\nu(\beta + \alpha - \lambda)}{1-\beta}\partial_x \mathcal{E}(t, x, y) = \frac{\xi\kappa(\beta + \lambda - \alpha)}{1-\beta}\partial^2_y \mathcal{E}(t, x, y).$$

This is a diffusion equation along the y dimension while being a transport in the x direction. As such, it is only well defined (or stable) when the sign of the diffusion coefficient in front of $\partial^2_y \mathcal{E}(t, x, y)$ is positive. This depends on the sign of ξ and $\beta + \lambda - \alpha$, which need to verify $\xi(\beta + \lambda - \alpha) > 0$. In that case, the system diffuses neural activity along the dimension y such that the entire neuronal layer converges to a single, uniform activation value when $t \to \infty$.

4.2 The general symmetric case

Finally, we now wish to relax some of the assumptions made in the previous Rao-Ballard case. Thus, the last case that we present is one where we assume that

(i) $\mathcal{W}^f$ and $\mathcal{W}^b$ are symmetric matrices, that is $\mathcal{W}^f, \mathcal{W}^b \in \mathscr{S}_d(\mathbb{R})$,

(ii) $\mathcal{W}^f$ and $\mathcal{W}^b$ commute, that is $\mathcal{W}^f\mathcal{W}^b = \mathcal{W}^b\mathcal{W}^f$.

But we do not necessarily impose that $\mathcal{W}^f = (\mathcal{W}^b)^{\mathbf{t}}$ as in the Rao & Ballard's previous case. Let us already note that examples of matrices verifying the above conditions are residual convolution matrices introduced in (4.3), that is $\mathcal{W}^f = \zeta_f \mathbf{I}_d + \xi_f \mathbf{A}$ and $\mathcal{W}^b = \zeta_b \mathbf{I}_d + \xi_b \mathbf{A}$ for some $\zeta_{b,f}, \xi_{b,f} \in \mathbb{R}$. Under assumptions (i) and (ii), $\mathcal{W}^f$ and $\mathcal{W}^b$ can be diagonalized in the same orthonormal basis, meaning that there exist an invertible orthogonal matrix $P \in \mathscr{M}_d(\mathbb{R})$ such that $PP^{\mathbf{t}} = P^{\mathbf{t}}P = I_d$, and two diagonal matrices $\mathcal{D}^f \in \mathscr{M}_d(\mathbb{R})$ and $\mathcal{D}^b \in \mathscr{M}_d(\mathbb{R})$ with the properties that

$$P^{\mathbf{t}}\mathcal{W}^f P = \mathcal{D}^f, \quad \text{and} \quad P^{\mathbf{t}}\mathcal{W}^b P = \mathcal{D}^b.$$

For future reference, we denote by $\gamma_p^{f,b}$ for each $1 \leq p \leq d$ the diagonal elements of $\mathcal{D}^{f,b}$. Once again, we can use the matrix P to apply an orthonormal basis change and create neural *asssemblies* $\mathcal{U}_j^n := P^t\mathcal{E}_j^n$. With $P\mathcal{U}_j^n := \mathcal{E}_j^n$, the recurrence equation becomes

$$\mathcal{U}_j^{n+1} - \beta\mathcal{D}^f\mathcal{U}_{j-1}^{n+1} = \alpha\mathcal{D}^b\mathcal{U}_{j-1}^n + \Big[(1-\beta-\lambda)\mathbf{I}_d - \alpha\mathcal{D}^b\mathcal{D}^b\Big]\mathcal{U}_j^n + \lambda\mathcal{D}^b\mathcal{U}_{j+1}^n.$$

Note that, because all matrices in the above equation are diagonal, we have also totally decoupled the d components of the vector $\mathcal{U}_j^n$. More precisely, by denoting $u_{j,p}^n$ the pth component of $\mathcal{U}_j^n$, that is $\mathcal{U}_j^n = (u_{j,1}^n, \cdots, u_{j,d}^n)^{\mathbf{t}}$, we obtain

$$u_{j,p}^{n+1} - \beta\gamma_p^f u_{j-1,p}^{n+1} = \alpha\gamma_p^b u_{j-1,p}^n + (1-\beta-\lambda-\alpha\left(\gamma_p^b\right)^2)u_{j,p}^n + \lambda\gamma_p^b u_{j+1,p}^n, \quad p = 1, \cdots, d.$$

This indicates that one needs to study

$$u_j^{n+1} - \beta\gamma_1 u_{j-1}^{n+1} = \alpha\gamma_2 u_{j-1}^n + (1-\beta-\lambda-\alpha\gamma_2^2)u_j^n + \lambda\gamma_2 u_{j+1}^n, \tag{4.4}$$

where $\gamma_{1,2} \in \mathbb{R}$ are now two given parameters. As before, $\gamma_{1,2}$ can be thought of as the connection strength across layers of the neural assembly under consideration. By construction, each assembly in a given layer is only connected to the corresponding assembly in the layer above, and similarly in the layer below, with γ_1 for the feedforward direction and γ_2 for the feedback direction. Note that $\gamma_1 = \gamma_2$ would then correspond to the Rao-Ballard situation studied previously.

4.2.1 Study of the amplification factor function

Repeating the previous analysis, one needs to understand the amplification factor

$$\rho_{\gamma_1,\gamma_2}(\theta) = \frac{\alpha\gamma_2\left(e^{-\mathbf{i}\theta} - \gamma_2\right) + 1 - \beta + \lambda\left(\gamma_2 e^{\mathbf{i}\theta} - 1\right)}{1 - \beta\gamma_1 e^{-\mathbf{i}\theta}}, \quad \theta \in [-\pi, \pi].$$

We already note a symmetry property of the amplification factor function which reads

$$\rho_{\gamma_1,\gamma_2}(\theta) = \rho_{-\gamma_1,-\gamma_2}(\theta \pm \pi), \quad \theta \in [-\pi, \pi].$$

As a consequence, whenever $\rho_{\gamma_1,\gamma_2}(0) = \pm 1$ one has $\rho_{-\gamma_1,-\gamma_2}(\pm\pi) = \pm 1$ for the same values of the parameters. Then, we note that

$$\rho_{\gamma_1,\gamma_2}(0) = 1 \Longleftrightarrow \gamma_1 = \chi(\gamma_2),$$

where the function $\chi(x)$, depending only on the hyper-parameters, is given by

$$\chi(x) := \frac{\alpha x^2 - (\alpha+\lambda)x + \lambda + \beta}{\beta}, x \in \mathbb{R}. \tag{4.5}$$

Thus, using the above symmetry, we readily deduce that

$$\rho_{\gamma_1,\gamma_2}(\pm\pi) = 1 \Longleftrightarrow \gamma_1 = -\chi(-\gamma_2).$$

Finally, we compute that

$$\rho_{\gamma_1,\gamma_2}(0) = -1 \Longleftrightarrow \gamma_1 = \zeta(\gamma_2),$$

where the function $\zeta(x)$, depending only on the hyper-parameters, is given by

$$\zeta(x) := \frac{-\alpha x^2 + (\alpha+\lambda)x + 2 - \lambda - \beta}{\beta}, x \in \mathbb{R}. \tag{4.6}$$

Using the above symmetry, we readily deduce that

$$\rho_{\gamma_1,\gamma_2}(\pm\pi) = -1 \Longleftrightarrow \gamma_1 = -\zeta(-\gamma_2).$$

A complete and exhaustive characterization of all possible cases as a function of $\gamma_{1,2}$ and the hyper-parameters is beyond the scope of this paper. Nevertheless, we can make some few further remarks. The four above curves $\gamma_1 = \chi(\gamma_2)$, $\gamma_1 = -\chi(-\gamma_2)$, $\gamma_1 = \zeta(\gamma_2)$ and $\gamma_1 = -\zeta(-\gamma_2)$ form parabolas in the plane (γ_1, γ_2) that can intersect and provide the boundaries of the stability regions. For example, we can notice that $\gamma_1 = \zeta(\gamma_2)$ and $\gamma_1 = -\zeta(-\gamma_2)$ intersect if and only if $\gamma_2 = \pm\sqrt{\frac{2-\lambda-\beta}{\alpha}}$ whereas $\gamma_1 = \chi(\gamma_2)$ and $\gamma_1 = -\chi(-\gamma_2)$ can never intersect. We refer to Figure 14 for an illustration of the stability regions and their boundaries in the case $(\alpha, \beta, \lambda) = (0.4, 0.2, 0.3)$. Here, we see that stability can be obtained with large values of the feedforward connection strength γ_1, but this requires the feedback connections strength γ_2 to remain low. Of course, different qualitative behaviors and stability regions may be obtained for different choices of the hyperparameters (α, β, λ); while it is beyond the scope of the present study to characterize them all, it is important to point out that such a characterization is feasible using the present method, for any choice of the hyperparameters.

More interestingly, we can investigate the dependence of the wave speed as a function of the parameters γ_1 and γ_2. For example, when $\gamma_1 = \chi(\gamma_2)$, we have that

$$\rho_{\chi(\gamma_2),\gamma_2}(\theta) = \exp\left(-\mathbf{i}\frac{(\alpha-\lambda)\gamma_2 + \beta\chi(\gamma_2)}{1-\beta\chi(\gamma_2)}\theta + \mathcal{O}(|\theta|^2)\right), \text{ as } \theta \to 0,$$

such that the associated wave speed is given by

$$c_0^\chi = \frac{(\alpha-\lambda)\gamma_2 + \beta\chi(\gamma_2)}{1-\beta\chi(\gamma_2)},$$

whose sign may vary as γ_2 is varied. We refer to the forthcoming section 4.2.4 below for a practical example (see Figure 18).

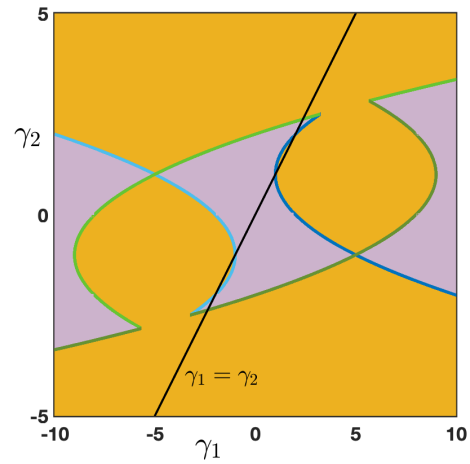


Figure 14: *Stability/instability regions and their boundaries as a function of* (γ_1, γ_2) *for* (4.4) *for fixed values of the hyperparameters* (α, β, λ)*. The shaded orange region corresponds to an instability for* (4.4) *while the purple region corresponds to a stability for* (4.4)*. The boundaries of the stability/instability regions are given by the intersections of the parametrized curves* $\gamma_1 = \chi(\gamma_2)$ *(blue curve),* $\gamma_1 = -\chi(-\gamma_2)$ *(light blue curve),* $\gamma_1 = \zeta(\gamma_2)$ *(dark green curves) and* $\gamma_1 = -\zeta(-\gamma_2)$ *(light green curves) where equation* (4.4) *is marginally stable. We represented the line* $\gamma_1 = \gamma_2$ *(black curve) which corresponds to the case studied in Figure 11(d) with* (β, λ) *in Region* (**III**) *and* α *fixed in* $(\Lambda, \lambda + \beta)$*.*

4.2.2 Continuous in time interpretation

As done in previous sections, we now perform a continuous in time limit of the model (4.4). With the same scaling on the hyperparameters

$$\widetilde{\beta} := \frac{\beta}{\Delta t}, \quad \widetilde{\lambda} := \frac{\lambda}{\Delta t}, \text{ and } \widetilde{\alpha} := \frac{\alpha}{\Delta t},$$

we get that, at the limit $\Delta t \to 0$, the equation (4.4) becomes the following lattice ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}_j(t) = (\widetilde{\beta}\gamma_1 + \widetilde{\alpha}\gamma_2)\mathbf{u}_{j-1}(t) - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma_2^2)\mathbf{u}_j(t) + \widetilde{\lambda}\gamma_2\mathbf{u}_{j+1}(t), \quad t > 0. \tag{4.7}$$

The amplification factor function in this case is given by

$$\nu_{\gamma_1,\gamma_2}(\theta) = (\widetilde{\beta}\gamma_1 + \widetilde{\alpha}\gamma_2)e^{-\mathbf{i}\theta} - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma_2^2) + \widetilde{\lambda}\gamma_2 e^{\mathbf{i}\theta}, \quad \theta \in [-\pi, \pi],$$

whose real part is given by

$$\mathrm{Re}(\nu_{\gamma_1,\gamma_2}(\theta)) = (\widetilde{\beta}\gamma_1 + (\widetilde{\alpha} + \widetilde{\lambda})\gamma_2)\cos(\theta) - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma_2^2), \quad \theta \in [-\pi, \pi].$$

When $\widetilde{\beta}\gamma_1 + (\widetilde{\alpha} + \widetilde{\lambda})\gamma_2 > 0$, we observe that

$$\max_{\theta\in[-\pi,\pi]} \mathrm{Re}(\nu_{\gamma_1,\gamma_2}(\theta)) = \mathrm{Re}(\nu_{\gamma_1,\gamma_2}(0)) = \widetilde{\beta}\gamma_1 + (\widetilde{\alpha} + \widetilde{\lambda})\gamma_2 - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma_2^2),$$

such that

$$\max_{\theta\in[-\pi,\pi]} \mathrm{Re}(\nu_{\gamma_1,\gamma_2}(\theta)) = 0 \Longleftrightarrow \gamma_1 = \frac{\widetilde{\alpha}\gamma_2^2 - (\widetilde{\alpha} + \widetilde{\lambda})\gamma_2 + \widetilde{\beta} + \widetilde{\lambda}}{\widetilde{\beta}}.$$

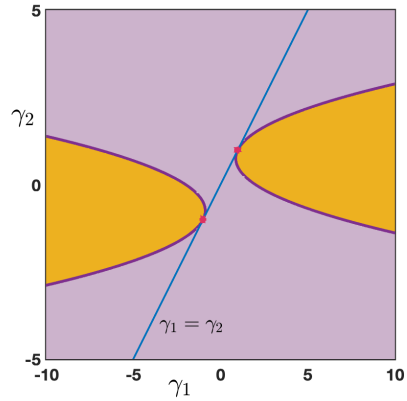


Figure 15: *Stability/instability regions and their boundaries as a function of* (γ_1, γ_2) *for* (4.7) *for any* $(\widetilde{\alpha}, \widetilde{\lambda}, \widetilde{\beta})$ *fixed with* $\widetilde{\alpha} = \widetilde{\lambda} + \widetilde{\beta}$. *The shaded orange region corresponds to an instability for* (4.7) *while the purple region corresponds to a stability for* (4.7). *The boundaries of the stability/instability regions are given by the intersections of the parabolas* $\gamma_1 = \frac{\widetilde{\alpha}\gamma_2^2 - (\widetilde{\alpha}+\widetilde{\lambda})\gamma_2 + \widetilde{\beta} + \widetilde{\lambda}}{\widetilde{\beta}}$ *and* $\gamma_1 = \frac{-\widetilde{\alpha}\gamma_2^2 - (\widetilde{\alpha}+\widetilde{\lambda})\gamma_2 - \widetilde{\beta} - \widetilde{\lambda}}{\widetilde{\beta}}$ *(magenta curves) where equation* (4.7) *is marginally stable. We represented the line* $\gamma_1 = \gamma_2$ *(blue curve) which corresponds to the case studied in Figure 13.*

Whereas, when $\widetilde{\beta}\gamma_1 + (\widetilde{\alpha} + \widetilde{\lambda})\gamma_2 < 0$, we observe that

$$\max_{\theta \in [-\pi, \pi]} \mathrm{Re}(\nu_{\gamma_1,\gamma_2}(\theta)) = \mathrm{Re}(\nu_{\gamma_1,\gamma_2}(\pm\pi)) = -\widetilde{\beta}\gamma_1 - (\widetilde{\alpha} + \widetilde{\lambda})\gamma_2 - (\widetilde{\beta} + \widetilde{\lambda} + \widetilde{\alpha}\gamma_2^2),$$

such that

$$\max_{\theta \in [-\pi, \pi]} \mathrm{Re}(\nu_{\gamma_1,\gamma_2}(\theta)) = 0 \Longleftrightarrow \gamma_1 = \frac{-\widetilde{\alpha}\gamma_2^2 - (\widetilde{\alpha} + \widetilde{\lambda})\gamma_2 - \widetilde{\beta} - \widetilde{\lambda}}{\widetilde{\beta}}.$$

As a consequence, the stability regions are determined by the locations of the parabolas $\gamma_2 \mapsto \frac{\widetilde{\alpha}\gamma_2^2 - (\widetilde{\alpha}+\widetilde{\lambda})\gamma_2 + \widetilde{\beta} + \widetilde{\lambda}}{\widetilde{\beta}}$ and $\gamma_2 \mapsto \frac{-\widetilde{\alpha}\gamma_2^2 - (\widetilde{\alpha}+\widetilde{\lambda})\gamma_2 - \widetilde{\beta} - \widetilde{\lambda}}{\widetilde{\beta}}$ in the plane (γ_1, γ_2). We observe that they never intersect and are oriented in the opposite directions and refer to Figure 15 for a typical configuration. Here, we see that the system is stable for a very large range of values of both γ_1 and γ_2. In particular, for large enough values of the feedback connection weight (e.g. $|\gamma_2| > 3$), stability is guaranteed regardless of the value of the feedforward connection weight γ_1 (within a reasonable range, e.g. $\gamma_1 \in (-10, 10)$). This is the opposite behavior as that obtained for the discrete system in Figure 14, where stability was impossible under the same conditions for $\gamma_{1,2}$. This highlights again the errors of interpretation that can potentially be caused by discrete approximation of a continuous system.

4.2.3 Fully continuous interpretation when $\mathcal{W}^f = \mathbf{I}_d + \xi_f \mathbf{A}$ and $\mathcal{W}^b = \mathbf{I}_d + \xi_b \mathbf{A}$.

When $\mathcal{W}^f = \mathbf{I}_d + \xi_f \mathbf{A}$ and $\mathcal{W}^b = \mathbf{I}_d + \xi_b \mathbf{A}$, one can once again identify $\mathcal{E}^t_{j,k}$ as the approximation of some smooth function $\mathcal{E}(t, x, y)$ at $t_n = n\Delta t$, $x_j = j\Delta x$ and $y_k = k\Delta y$, along the three dimensions of time,

network depth and neuronal layer width. We may rewrite (2.1) in this case as

$$
\begin{aligned}
(1-\beta)\frac{\mathcal{E}_j^{n+1}-\mathcal{E}_j^n}{\Delta t} &= \beta\nu\frac{\mathcal{E}_{j-1}^{n+1}-\mathcal{E}_j^{n+1}}{\Delta x}+\lambda\nu\frac{\mathcal{E}_{j+1}^n-\mathcal{E}_j^n}{\Delta x}-\alpha\nu\frac{\mathcal{E}_j^n-\mathcal{E}_{j-1}^n}{\Delta x}\\
&\quad+\beta\xi_f\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}_{j-1}^{n+1}+\alpha\xi_b\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}_{j-1}^n-2\alpha\xi_b\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}_j^n-\alpha\xi_b^2\kappa\Delta y^2\frac{\mathbf{A}}{\Delta y^2}\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}_j^n\\
&\quad+\lambda\xi_b\kappa\frac{\mathbf{A}}{\Delta y^2}\mathcal{E}_{j+1}^n,
\end{aligned}
$$

such that in the limit $\Delta t\to 0$, $\Delta x\to 0$ and $\Delta y\to 0$ with ν and κ fixed, we obtain the following partial differential equation

$$
\partial_t\mathcal{E}(t,x,y)+\frac{\nu(\beta+\alpha-\lambda)}{1-\beta}\partial_x\mathcal{E}(t,x,y)=\kappa\frac{\beta\xi_f+(\lambda-\alpha)\xi_b}{1-\beta}\partial_y^2\mathcal{E}(t,x,y).
$$

As before, this is a diffusion equation along the y dimension, whose stability depends on the positivity of the diffusion coefficient, i.e. $\beta\xi_f+(\lambda-\alpha)\xi_b\geq 0$.

4.2.4 Application to a ring model of orientations

Going back to our discrete system, in this section we consider the case where neurons within each layer encode for a given orientation in $[0,\pi]$. Here, we have in mind visual stimuli which are made of a fixed elongated black bar on a white background with a prescribed orientation. We introduce the following matrix $\mathbf{A}_{\text{per}}\in\mathscr{M}_d(\mathbb{R})$ given by

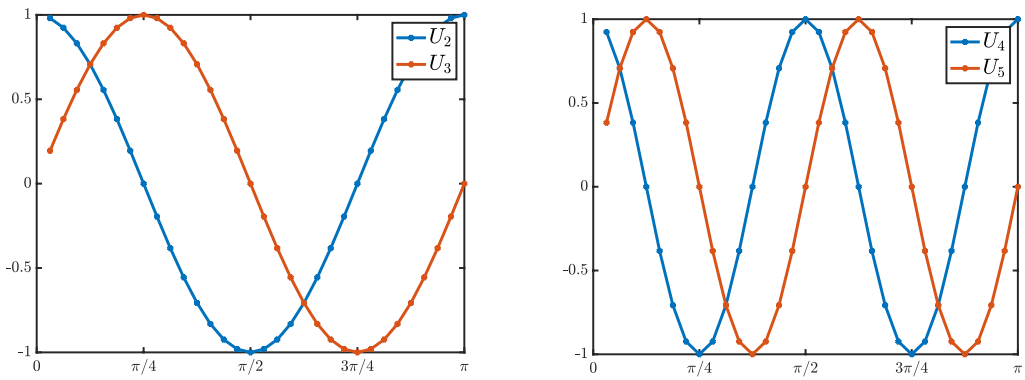
$$
\mathbf{A}_{\text{per}}=\begin{pmatrix}
-2 & 1 & 0 & \cdots & 0 & 1\\
1 & -2 & 1 & \ddots & \ddots & 0\\
0 & \ddots & \ddots & \ddots & \ddots & \vdots\\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \ddots & \ddots & 1 & -2 & 1\\
1 & 0 & \cdots & 0 & 1 & -2
\end{pmatrix},
$$

which is nothing but the discretizing of the Laplacian with boundary condition. Indeed, for each $\mathcal{E}_j^n\in\mathbb{R}^d$, we assume that neuron $\mathcal{E}_{j,k}^n$ encodes for orientation $\frac{k}{d}\pi$ for $k=1,\cdots,d$. We readily remark that $0\in\text{Spec}(\mathbf{A}_{\text{per}})$ with corresponding eigenvector $U_1=(1,\cdots,1)^{\mathbf{t}}\in\mathbb{R}^d$. Furthermore, we have:

- if $d=2m+1$ is odd, then $\lambda_p=-4\sin\left(\frac{p\pi}{d}\right)^2$ with $p=1,\cdots,m$ is an eigenvalue of $\mathbf{A}_{\text{per}}$ of multiplicity 2 with associated eigenvectors

$$
\begin{aligned}
U_{2p}&=\left(\cos\left(\frac{2p\pi}{d}\right),\cdots,\cos\left(\frac{2kp\pi}{d}\right),\cdots,1\right)^{\mathbf{t}}\in\mathbb{R}^d,\\
U_{2p+1}&=\left(\sin\left(\frac{2p\pi}{d}\right),\cdots,\sin\left(\frac{2kp\pi}{d}\right),\cdots,0\right)^{\mathbf{t}}\in\mathbb{R}^d;
\end{aligned}
$$

- if $d=2m$ is even, then $\lambda_p=-4\sin\left(\frac{p\pi}{d}\right)^2$ with $p=1,\cdots,m-1$ is an eigenvalue of $\mathbf{A}_{\text{per}}$ of multiplicity 2 with associated eigenvectors U_{2p} and U_{2p+1} as above. And $\lambda=-4$ is a simple eigenvalue of $\mathbf{A}_{\text{per}}$ with associated eigenvector $U_d=(-1,1,-1,1,\cdots,-1,1)\in\mathbb{R}^d$.



(a) Eigenvectors U_2 and U_3.

(b) Eigenvectors U_4 and U_5.

Figure 16: *We plot the eigenvectors U_{2p} and U_{2p+1} for $p = 1$ and $p = 2$ as a function of $\frac{k}{d}\pi$ for $k = 1, \cdots, d$. We note that U_{2p} and U_{2p+1} encode the first Fourier modes. Here we have set $d = 2^5$.*

It may be interesting to note that any linear combinations of U_{2p} and U_{2p+1} can always be written in the form

$$aU_{2p} + bU_{2p+1} = A\left(\cos\left(\frac{2p\pi}{d} + \varphi\right), \cdots, \cos\left(\frac{2kp\pi}{d} + \varphi\right), \cdots, \cos(\varphi)\right)^{\mathbf{t}} \in \mathbb{R}^d,$$

where $A = \sqrt{a^2 + b^2} > 0$ and $\varphi = -\arctan\left(\frac{b}{a}\right) \in (-\pi/2, \pi/2)$ whenever $a \neq 0$ and $b \neq 0$. This means that U_{2p} and U_{2p+1} span all possible translations modulo $[0, \pi]$ of a fixed profile. We refer to Figure 16 for a visualization of the first eigenvectors. In short, these eigenvectors U_i implement a Fourier transform of the matrix $\mathbf{A}_{\text{per}}$.

We now set $\mathcal{W}^b$ to be

$$\mathcal{W}^b = \frac{1}{2}\mathbf{I}_d - \frac{1}{4}\mathbf{A}_{\text{per}} = \begin{pmatrix} 1 & -\frac{1}{4} & 0 & \cdots & 0 & -\frac{1}{4} \\ -\frac{1}{4} & 1 & -\frac{1}{4} & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -\frac{1}{4} & 1 & -\frac{1}{4} \\ -\frac{1}{4} & 0 & \cdots & 0 & -\frac{1}{4} & 1 \end{pmatrix},$$

which means that $\mathcal{W}^b$ acts as a convolution with local excitation and lateral inhibition. From now on, to fix ideas, we will assume that $d = 2m$ is even. We define the following matrix

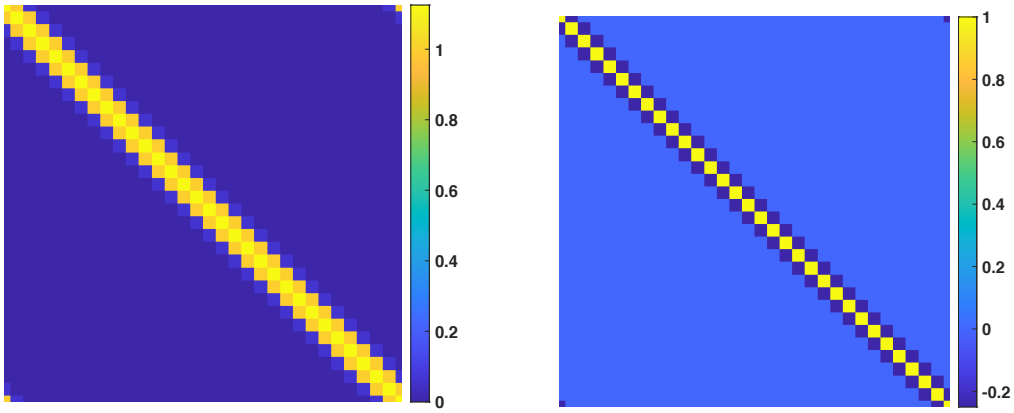
$$P = (U_1, U_2, \cdots, U_d) \in \mathscr{M}_d(\mathbb{R}).$$

As a consequence, we have the decomposition

$$P^{\mathbf{t}}\mathcal{W}^b P = \mathcal{D}^b,$$

with $\mathcal{D}^b = \text{diag}\left(\frac{1}{2}, \frac{1}{2} - \frac{1}{4}\lambda_1, \frac{1}{2} - \frac{1}{4}\lambda_1, \cdots, \frac{1}{2} - \frac{1}{4}\lambda_{m-1}, \frac{1}{2} - \frac{1}{4}\lambda_{m-1}, \frac{3}{2}\right) \in \mathscr{M}_d(\mathbb{R})$. Now, for given values of the hyper-parameters (α, β, λ) with $\beta > 0$, we set $\mathcal{D}^f := \chi(\mathcal{D}^b)$ where the map χ, defined in (4.5), is applied to the diagonal elements of $\mathcal{D}^b$, that is

$$\mathcal{D}^f = \text{diag}\left(\chi\left(\frac{1}{2}\right), \chi\left(\frac{1}{2} - \frac{1}{4}\lambda_1\right), \cdots, \chi\left(\frac{1}{2} - \frac{1}{4}\lambda_{m-1}\right), \chi\left(\frac{3}{2}\right)\right) \in \mathscr{M}_d(\mathbb{R}).$$



(a) $\mathcal{W}^f$.

(b) $\mathcal{W}^b$.

Figure 17: *Illustration of the matrices $\mathcal{W}^f$ and $\mathcal{W}^b$ for $d = 2^5$ neurons and values of the hyper-parameters fixed to $(\alpha, \beta, \lambda) = (0.1, 0.1, 0.5)$. Note the band structure of $\mathcal{W}^f$ with local excitation.*

And then we set $\mathcal{W}^f := P\mathcal{D}^f P^{\mathbf{t}}$. We refer to Figure 17 for an illustration of the structures of matrices $\mathcal{W}^f$ and $\mathcal{W}^b$. For large set of values of the hyper-parameters, $\mathcal{W}^f$ still present a band structure with positive elements on the diagonals indicating that $\mathcal{W}^f$ can also be interpreted as a convolution with local excitation. For the values of the hyper-parameters fixed in Figure 17, the feedforward matrix $\mathcal{W}^f$ is purely excitatory.

Reproducing the analysis developed in the previous Subsection 4.2, we perform a change of orthonormal basis to express neural activities in terms of the relevant *assemblies* $\mathcal{U}_j^n := P^t \mathcal{E}_j^n$. With $P\mathcal{U}_j^n := \mathcal{E}_j^n$, the recurrence equation becomes

$$\mathcal{U}_j^{n+1} - \beta \mathcal{D}^f \mathcal{U}_{j-1}^{n+1} = \alpha \mathcal{D}^b \mathcal{U}_{j-1}^n + \left[(1-\beta-\lambda)\mathbf{I}_d - \alpha \mathcal{D}^b \mathcal{D}^b\right] \mathcal{U}_j^n + \lambda \mathcal{D}^b \mathcal{U}_{j+1}^n.$$

Then, if we denote by γ_p the pth diagonal element of $\mathcal{D}^b$, then for each $p = 1, \cdots, d$ the above recurrence writes

$$u_{j,p}^{n+1} - \beta\chi(\gamma_p) u_{j-1,p}^{n+1} = \alpha\gamma_p u_{j-1,p}^n + (1-\beta-\lambda-\alpha\gamma_p^2) u_{j,p}^n + \lambda\gamma_p u_{j+1,p}^n,$$

where $u_{j,p}^n$ is the pth component (or neural assembly) of $\mathcal{U}_j^n$. For each $p = 1, \cdots, d$, the associated amplification factor function reads

$$\rho_p(\theta) = \frac{\alpha\gamma_p\left(e^{-\mathbf{i}\theta} - \gamma_p\right) + 1 - \beta + \lambda\left(\gamma_p e^{\mathbf{i}\theta} - 1\right)}{1 - \beta\chi(\gamma_p) e^{-\mathbf{i}\theta}}, \quad \theta \in [-\pi, \pi],$$

and with our specific choice of function χ, we have that $\rho_p(0) = 1$ with

$$\rho_p(\theta) = \exp\left(-\mathbf{i}\frac{(\alpha-\lambda)\gamma_p + \beta\chi(\gamma_p)}{1-\beta\chi(\gamma_p)}\theta - \sigma_0^p\theta^2 + \mathcal{O}(|\theta|^3)\right), \text{ as } \theta \to 0,$$

such that the associated wave speed is given by

$$c_0^p = \frac{(\alpha-\lambda)\gamma_p + \beta\chi(\gamma_p)}{1-\beta\chi(\gamma_p)},$$

and where we have set

$$\sigma_0^p = \frac{\alpha(1+4\lambda)\gamma_p^2 + \beta + \lambda - (\alpha+\lambda)\gamma_p(\alpha\gamma_p^2 + \beta + \lambda)}{2(1-\beta\chi(\gamma_p))^2}.$$

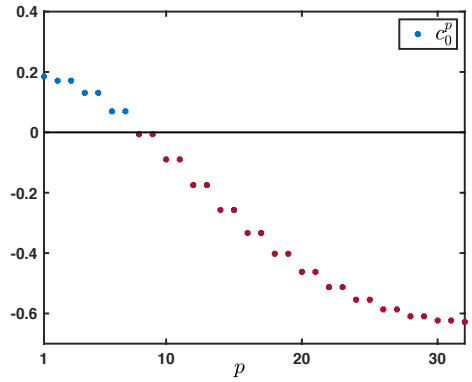


Figure 18: *Plot of the wave speed c_0^p for $p = 1, \cdots, d$ (colored dots). The color code (blue/red) refers to the sign of c_0^p: blue when positive and dark red when negative. Note that only the elements associated to the eigenvalues $\frac{1}{2}$, $\frac{1}{2} - \frac{1}{4}\lambda_1$, $\frac{1}{2} - \frac{1}{4}\lambda_2$ and $\frac{1}{2} - \frac{1}{4}\lambda_3$ are positive. We also remark that c_0^p is a monotone decreasing function. Here $d = 2^5$ and values of the hyper-parameters are fixed to $(\alpha, \beta, \lambda) = (0.1, 0.1, 0.5)$.*

From now on, we assume that we have tuned the hyper-parameters such that $|\rho_p(\theta)| < 1$ for all $\theta \in [-\pi, \pi] \setminus \{0\}$ and each $p = 1, \cdots, d$. This can in fact be systematically checked numerically for a given set of hyper-parameters. We report in Figure 18 the shape of $p \mapsto c_0^p$ for the same values of the hyper-parameters as the ones in Figure 17 and $d = 2^5$. We first remark that $p \mapsto c_0^p$ is a monotone decreasing map, and in our specific case we have

$$c_0^d < c_0^{d-1} = c_0^{d-2} < \cdots < c_0^9 = c_0^8 < 0 < c_0^7 = c_0^6 < c_0^5 = c_0^4 < c_0^3 = c_0^2 < c_0^1.$$

Given a fixed input entry $\mathcal{E}_0 \in \mathbb{R}^d$ presented at $j = 0$ to the network continually at each time step, we can deduce which components of $\mathcal{E}_0 \in \mathbb{R}^d$ will be able to propagate forward through the network. More precisely, we can decompose $\mathcal{E}_0$ along the basis $(U_1, \cdots U_d)$ of eigenvectors, that is

$$\mathcal{E}_0 = \sum_{p=1}^{d} a_p U_p,$$

for some real coefficients a_p for $p = 1, \cdots, d$. Assuming that the network was at rest initially, we get that the dynamics along each eigenvector (or neural assembly) is given by

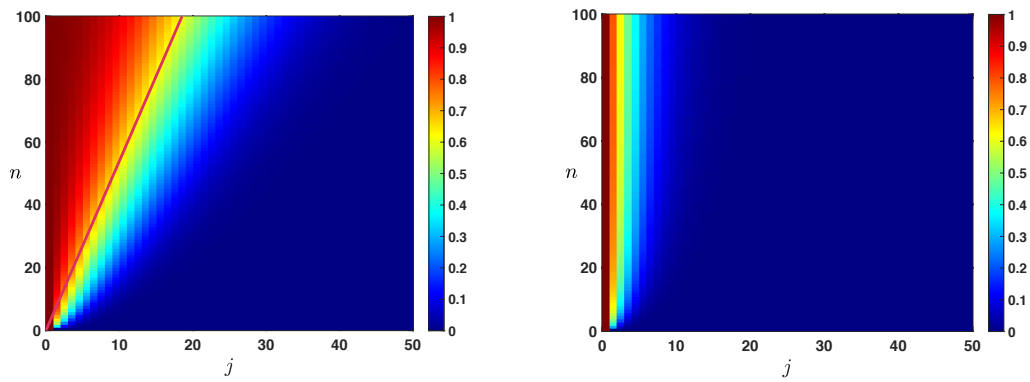
$$\begin{cases} u_{j,p}^{n+1} - \beta\chi(\gamma_p) u_{j-1,p}^{n+1} = \alpha\gamma_p u_{j-1,p}^n + (1 - \beta - \lambda - \alpha\gamma_p^2) u_{j,p}^n + \lambda\gamma_p u_{j+1,p}^n, \quad j \geq 1, \quad n \geq 0, \\ u_{0,p}^n = a_p, \quad n \geq 0, \\ u_{j,p}^0 = 0, \quad j \geq 1. \end{cases} \tag{4.8}$$

Thus, we readily obtain that

$$\mathcal{E}_j^n = \sum_{p=1}^{d} u_{j,p}^n U_p, \quad j \geq 1, \quad n \geq 1,$$

where $u_{j,p}^n$ is a solution to (4.8).

As a consequence, the monotonicity property of the map $p \mapsto c_0^p$ indicates that the homogeneous constant mode U_1 is the fastest to propagate forward into the network with associated spreading speed c_0^1, it is then followed by the modes (U_2, U_3) propagating at speed $c_0^2 = c_0^3$. In our numerics, we have set the parameters



(a) Case $p = 2$ with $\gamma_2 = \frac{1}{2} - \frac{1}{4}\lambda_1 \sim 0.5096$.

(b) Case $p = 10$ with $\gamma_{10} = \frac{1}{2} - \frac{1}{4}\lambda_5 \sim 0.7222$.

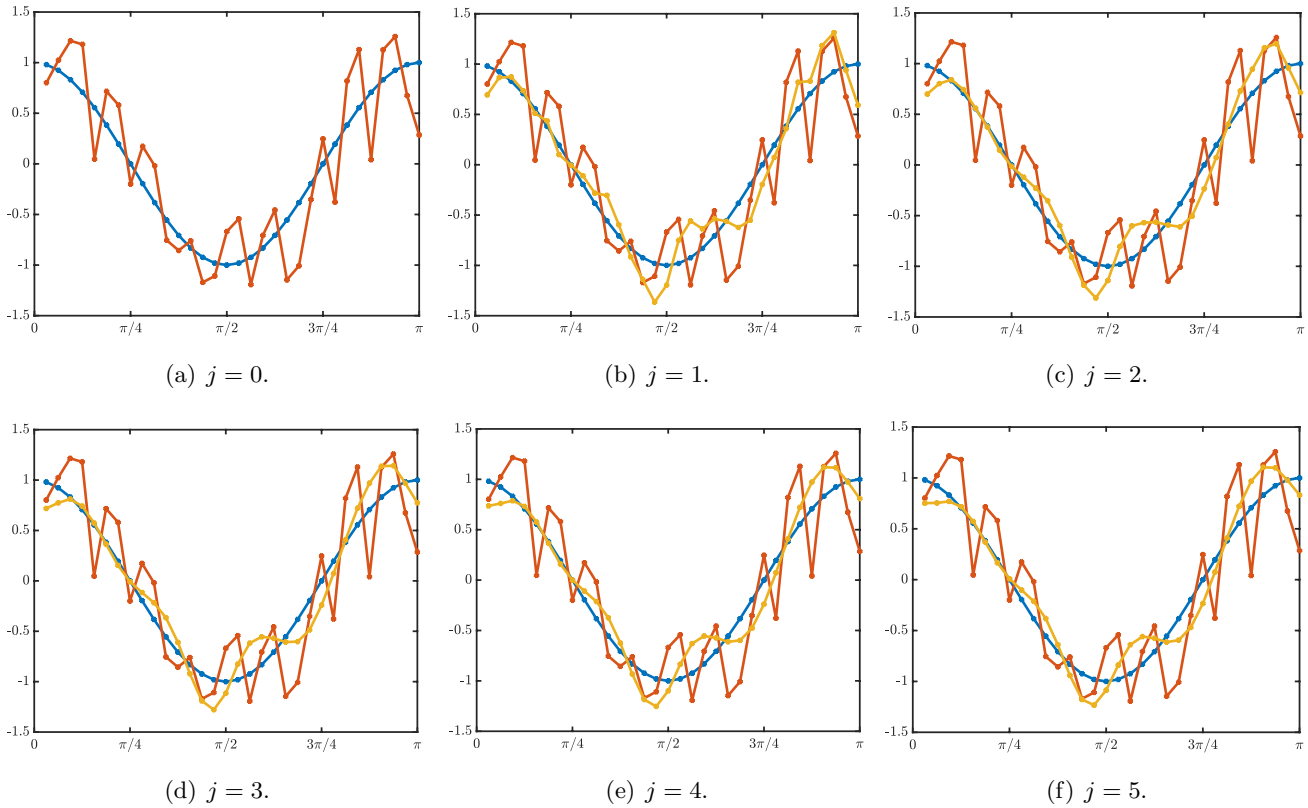
Figure 19: *Space-time plot of the solution of the recurrence equation for* (4.8) *for* $p = 2$ *and* $p = 10$ *associated to respectively positive wave speed* $c_0^2 > 0$ *and negative wave speed* $c_0^{10} < 0$*. The neural assembly associated with the 2nd eigenvector of the connectivity matrix propagates its input signal into the network at constant speed; but the neural assembly associated with the 10th eigenvector does not propagate the signals it receives on the input layer.*

such that $c_0^1 \approx c_0^2 = c_0^3$ with a significant gap with the other wave speeds. Lets us remark, that all modes U_p with $p \geq 8$ are not able to propagate into the network (see Figure 19). Thus our architecture acts as a mode filter.

Even more precisely, let us remark that the sequence $\left(a_p \left(\frac{\alpha\gamma_p + \beta\chi(\gamma_p)}{\lambda\gamma_p}\right)^j\right)_{j\geq 0}$ is a stationary solutions of (4.8) which remains bounded whenever p is such that the associated wave speed is negative, that is $c_0^p < 0$, since in that case, one has $\alpha\gamma_p + \beta\chi(\gamma_p) < \lambda\gamma_p$. The solution $\mathcal{E}_j^n$ can then be approximated as

$$\mathcal{E}_j^n \simeq \sum_{p\,:\,c_0^p>0} \frac{a_p}{2}\left(1 - \mathrm{erf}\left(\frac{j - c_0^p n}{\sqrt{4\sigma_0^p n}}\right)\right) U_p + \sum_{p\,:\,c_0^p<0} a_p \left(\frac{\alpha\gamma_p + \beta\chi(\gamma_p)}{\lambda\gamma_p}\right)^j U_p, \quad j \geq 1, \quad n \geq 1,$$

This is illustrated by a first example simulation in Figures 20 and 21. We present at $j = 0$ a fixed input $\mathcal{E}_0$ which is generated as the superposition of a tuned curve at $\vartheta = 0$ (blue) with some fixed random noise: namely we select $a_1 = 0$, $a_2 = 1$, $a_3 = 0$ and all other coefficients a_p for $p = 4, \cdots, d$ are drawn from a normal law with an amplitude pre-factor of magnitude ε set to $\varepsilon = 0.1$. The shape of the input $\mathcal{E}_0$ is shown in Figure 20(a). The profile of $\mathcal{E}_j^n$ at time iteration $n = 200$ along the first layers of the network $j \in \{1, 2, 3, 4, 5\}$ is given in Figure 20(b)-(c)-(d)-(e)-(f) respectively. We first observe that the network indeed acts as a filter since across the layers of the network the solution profile $\mathcal{E}_j^n$ has been denoised and get closer to the tuned curve at $\vartheta = 0$. Let us also remark that the filtering is more efficient for layers away from the boundary and is less efficient for those layers near the boundary. This is rather natural since the impact of the input $\mathcal{E}_0$ is stronger on the first layers. We see that already at layer $j = 5$, we have almost fully recovered the tuned curve at $\vartheta = 0$ (see Figure 20(f)). On the other hand, in Figure 21, we show the time evolution of $\mathcal{E}_j^n$ at a fixed layer far away from the boundary, here $j = 10$. Initially, at $n = 0$, the layer is inactivated (see Figure 21(a)), and we see that after several time iterations that the solution profile $\mathcal{E}_j^n$ start to be activated. It is first weakly tuned (see Figures 21(b)-(c)-(d)) and then it becomes progressively fully tuned and converges to the tuned curve at $\vartheta = 0$ (see Figures 21(e)-(f)).



(a) $j = 0$. (b) $j = 1$. (c) $j = 2$.

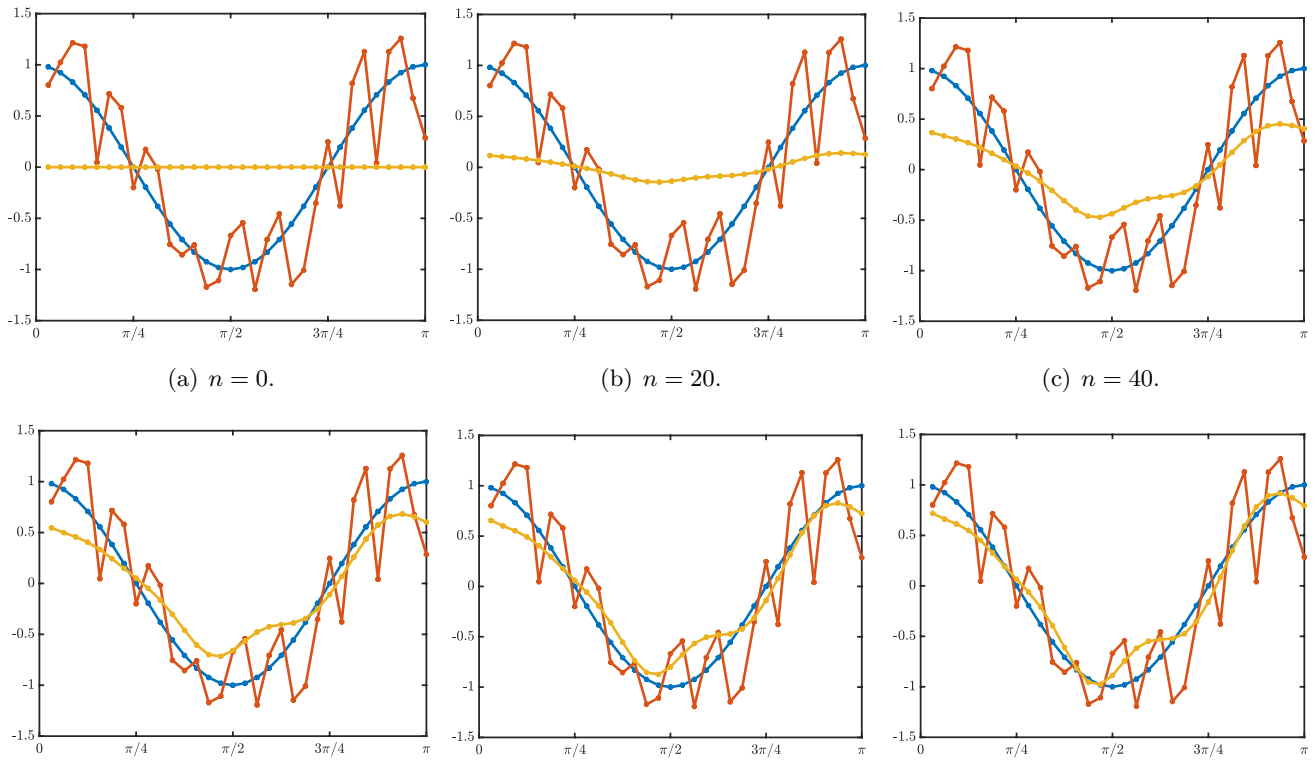
(d) $j = 3$. (e) $j = 4$. (f) $j = 5$.

Figure 20: *A fixed input $\mathcal{E}_0$ (red) which is the superposition of a tuned curve at $\vartheta = 0$ (blue) with some fixed random noise is presented at layer $j = 0$. Profile (yellow) of $\mathcal{E}_j^n$ at time iteration $n = 200$ along the first layers of the network $j \in \{1, 2, 3, 4, 5\}$.*

In a second example simulation (Figure 22), we highlight the dynamics of the different modes in a situation where the input is a narrow Gaussian profile (close to a Dirac function), with a superposition of various Fourier modes. As expected from the different values of the propagation speed c_0 (Figure 18), we see that the mode associated with the first Fourier component is the first to reach layer $j = 10$, later followed by successive modes associated with later Fourier components. In other words, this hierarchically higher layer $j = 10$ first receives information about the coarse spatial structure of the input signal, and then gradually about finer and finer spatial details.

4.3 Summary

In this section, we saw that the results obtained initially (The Identity Case) with the amplification function can be extended to more realistic situations with forward and backward connection matrices, for instance implementing (residual) convolutions or orientation processing. When we consider neural *assemblies* capturing the principal components of the connection matrices, we see that each assembly can be treated independently in terms of stability and signal propagation speed and direction. The exact behavior of the system will depend on the actual connection matrices (and thus on the function that they implement in the neural network), but the important point is that our generic framework can always be applied in practice. In some example cases (ring model of orientations), we saw that only a few assemblies support



(a) $n = 0$. (b) $n = 20$. (c) $n = 40$.

(d) $n = 60$. (e) $n = 80$. (f) $n = 100$.

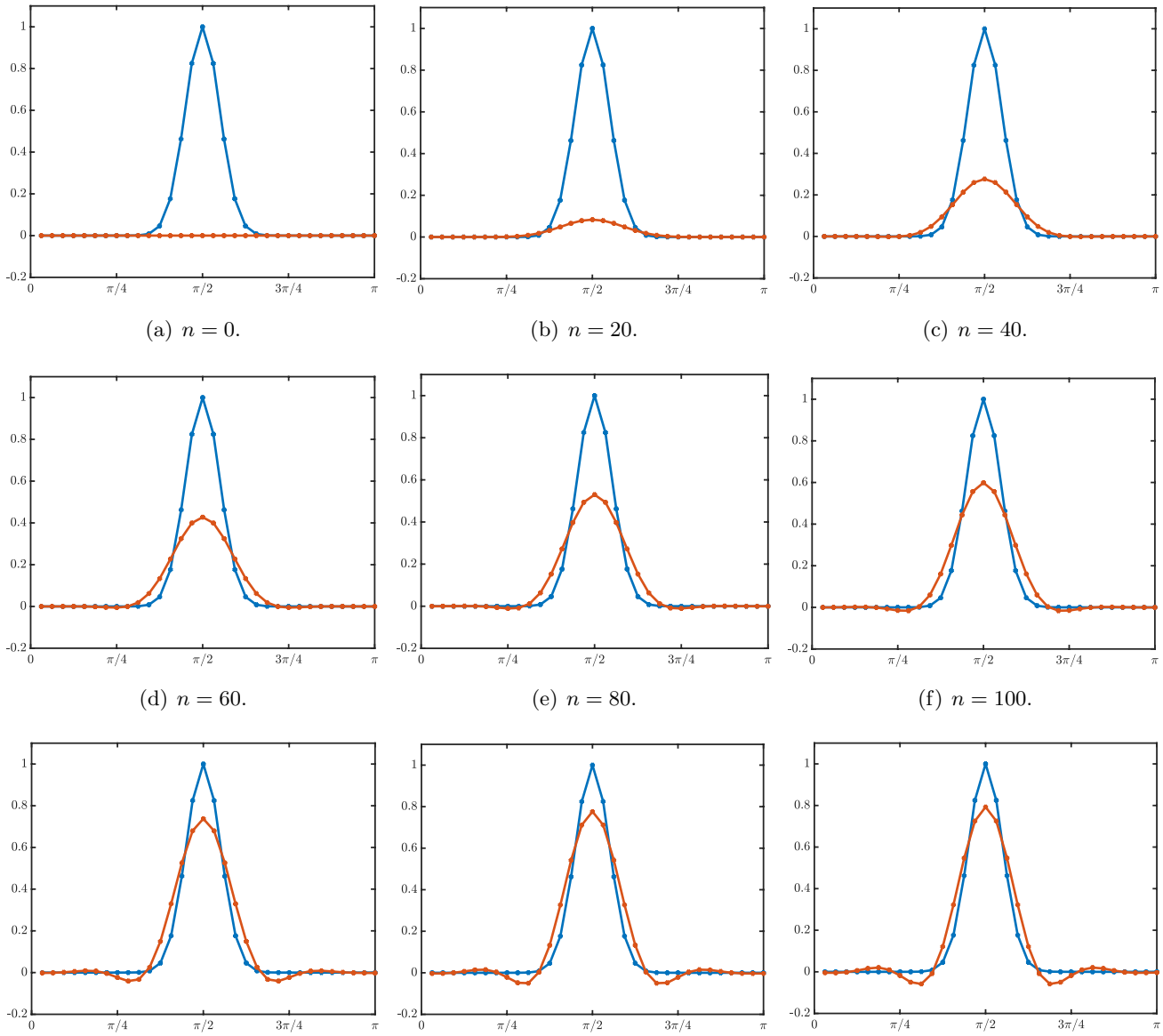
Figure 21: *Time evolution of $\mathcal{E}_j^n$ at layer $j = 10$ (yellow) for $n \in \{0, 20, 40, 60, 80, 100\}$ with fixed input $\mathcal{E}_0$ at layer $j = 0$ (red). The input $\mathcal{E}_0$ is the superposition of a tuned curve at $\vartheta = 0$ (blue) with some fixed random noise.*

signal propagation (implying that the system acts a filter on its inputs), and these assemblies propagate information at different speeds (implementing a coarse-to-fine analysis). In other cases (e.g. Figure 12), we have even seen that distinct assemblies can simultaneously propagate information in opposite directions, with one assembly supporting feedforward propagation while another entails feedback propagation.

We have extended our equations to the continuous limit in time, and found that the amplification factor function can give rise to qualitatively different stability regions compared to the discrete model. This served as a cautionary note for situations where the discrete implementation must be chosen; in that case, using smaller time steps will be preferable, because it makes such discrepancies less likely.

Finally, we also showed that it is possible to consider fully continuous versions of our dynamic system, where not only time but also network depth and neural layer width are treated as continuous variables. This gives rise to diffusion equations, whose stability can also be characterized as a function of hyperparameter values.

In the following, we address possible extensions of the model to more sophisticated and more biologically plausible neural architectures, taking into account the significant communication delays between layers.



(a) $n = 0$. (b) $n = 20$. (c) $n = 40$.

(d) $n = 60$. (e) $n = 80$. (f) $n = 100$.

(g) $n = 200$. (h) $n = 300$. (i) $n = 400$.

Figure 22: *Time evolution of* $\mathcal{E}_j^n$ *at layer* $j = 10$ *(orange) for* $n \in \{0, 20, 40, 60, 80, 100, 200, 300, 400\}$ *with fixed input* $\mathcal{E}_0$ *at layer* $j = 0$ *(blue). The input* $\mathcal{E}_0$ *is a Gaussian centered at* $\vartheta = \pi/2$.

5 Extension of the model: taking into account transmission delays

Deep feedforward neural networks typically implement *instantaneous* updates, as we did in Eq (2.1) with our feedforward term $\mathcal{E}_j^{n+1} = \beta \mathcal{W}^f \mathcal{E}_{j-1}^{n+1} + ...$ Similarly, artificial recurrent neural networks sequentially update their activity from *one time step to the next*, as we did with the other terms in our equation (2.1) (memory term, feedforward and feedback prediction error correction terms): $\mathcal{E}_j^{n+1} = ... + \alpha(\mathcal{W}^b)^{\mathbf{t}} \mathcal{E}_{j-1}^n + (1 - \beta - \lambda)\mathcal{E}_j^n - \alpha(\mathcal{W}^b)^{\mathbf{t}} \mathcal{W}^b \mathcal{E}_j^n + \lambda \mathcal{W}^b \mathcal{E}_{j+1}^n$. However, in the brain there are significant transmission delays whenever neural signals travel from one area to another. These delays could modify the system's dynamics

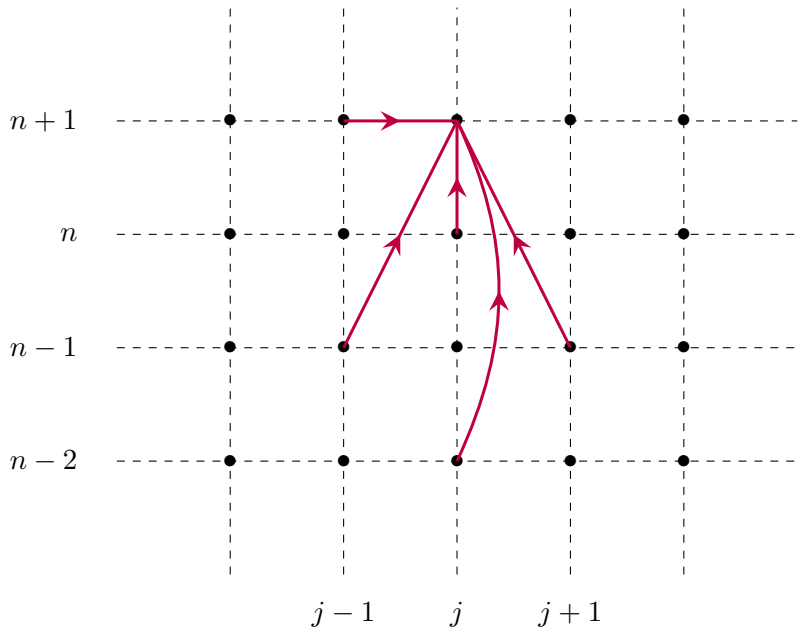


Figure 23: *Illustration of the network structure of model* (5.1) *for* $k = 1$ *where the red arrows indicate the contributions leading to the update of* $\mathcal{E}_j^{n+1}$.

and its stability properties. Therefore, in this section we modify model (2.1) by assuming that it takes k time steps to receive information from a neighboring site in the feedback/feedforward dynamics, namely we consider the following recurrence equation

$$\mathcal{E}_j^{n+1} - \beta \mathcal{W}^f \mathcal{E}_{j-1}^{n+1} = \alpha (\mathcal{W}^b)^{\mathbf{t}} \mathcal{E}_{j-1}^{n-k} + (1-\beta-\lambda)\mathcal{E}_j^n - \alpha (\mathcal{W}^b)^{\mathbf{t}} \mathcal{W}^b \mathcal{E}_j^{n-2k} + \lambda \mathcal{W}^b \mathcal{E}_{j+1}^{n-k}, \tag{5.1}$$

where $k \geq 1$ is some given fixed integer (see Figure 23 for an illustration with $k = 1$), and we refer to [23] for the justification of the derivation of the model. (Note in particular that we did not modify the *instantaneous* nature of our feedforward updating term $\mathcal{E}_j^{n+1} = \beta \mathcal{W}^f \mathcal{E}_{j-1}^{n+1} + \ldots$. This is because, as motivated in [9, 23], we aim for the feedforward part of the system to be compatible with state-of-the-art deep convolutional neural networks, and merely wish to investigate how adding recurrent dynamics can modify its properties.) We may already notice that when $k = 0$, we recover our initial model (2.1). In what follows, for the mathematical analysis, we restrict ourselves to the identity case $\mathcal{W}^f = \mathcal{W}^n = \mathbf{I}_d$ and when the model is set on $\mathbb{Z}$. Indeed, our intention is to briefly explain what could be the main new propagation properties that would emerge by including transmission delays. Thus, we consider

$$e_j^{n+1} - \beta e_{j-1}^{n+1} = \alpha e_{j-1}^{n-k} + (1-\beta-\lambda) e_j^n - \alpha e_j^{n-2k} + \lambda e_{j+1}^{n-k}, \quad j \in \mathbb{Z}. \tag{5.2}$$

Let us also note that the system (5.2) depends on a "history" of $2k+1$ time steps; thus one needs to impose $2k+1$ initial conditions:

$$e_j^m = h_j^m, \quad m = 0, \cdots, 2k, \quad j \in \mathbb{Z},$$

for $2k+1$ given sequences $(h_j^m)_{j \in \mathbb{Z}}$ with $m = 0, \cdots, 2k$.

To proceed in the analysis, we first introduce a new vector unknown capturing each layer's recent history:

$$\mathbf{E}_j^n := \begin{pmatrix} e_j^{n-2k} \\ \vdots \\ e_j^{n-2} \\ e_j^{n-1} \\ e_j^n \end{pmatrix} \in \mathbb{R}^{2k+1}, \quad n \geq 1, \quad j \in \mathbb{Z},$$

such that the above recurrence (5.2) can then be rewritten as

$$\mathbf{E}_j^{n+1} - \beta Q_{-1}\mathbf{E}_{j-1}^{n+1} = \alpha Q_1 \mathbf{E}_{j-1}^n + Q_0 \mathbf{E}_j^n + \lambda Q_1 \mathbf{E}_{j+1}^n, \quad n \geq 1, \quad j \in \mathbb{Z}, \tag{5.3}$$

where the matrices $Q_1, Q_0, Q_{-1} \in \mathscr{M}_{2k+1}(\mathbb{R})$ are defined as follows

$$Q_0 = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 1 \\ -\alpha & 0 & \cdots & \cdots & 0 & 1-\beta-\lambda \end{pmatrix},$$

and $Q_{\pm 1}$ have a single nonzero element on their last row:

$$(Q_{-1})_{2k+1,2k+1} = 1, \quad (Q_1)_{2k+1,k+1} = 1.$$

5.1 Mathematical study of the recurrence equation (5.3)

We now postulate an Ansatz of the form $\rho^n e^{\mathbf{i}\theta j}\mathbf{E}$ for some non zero vector $\mathbf{E} \in \mathbb{C}^{2k+1}$, and obtain

$$\underbrace{\left(\rho\left[\mathbf{I}_{2k+1} - \beta e^{-\mathbf{i}\theta} Q_{-1}\right] - Q_0 - (\alpha e^{-\mathbf{i}\theta} + \lambda e^{\mathbf{i}\theta}) Q_1\right)}_{:=\mathcal{A}_k(\rho,\theta)} \mathbf{E} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

which is equivalent to

$$\det\left(\rho\left[\mathbf{I}_{2k+1} - \beta e^{-\mathbf{i}\theta} Q_{-1}\right] - Q_0 - (\alpha e^{-\mathbf{i}\theta} + \lambda e^{\mathbf{i}\theta}) Q_1\right) = 0,$$

that is

$$(1-\beta e^{-\mathbf{i}\theta})\rho^{2k+1} - \rho^{2k}(1-\beta-\lambda) - \rho^k\left(\alpha e^{-\mathbf{i}\theta} + \lambda e^{\mathbf{i}\theta}\right) + \alpha = 0. \tag{5.4}$$

The above system has $2k+1$ roots in the complex plane that we denote $\rho_m(\theta)$ for $m = 1, \cdots 2k+1$. We remark at $\theta = 0$, $\rho = 1$ is always a root of the equation since in this case (5.4) reduces to

$$(1-\beta)\rho^{2k+1} - \rho^{2k}(1-\beta-\lambda) - \rho^k(\alpha+\lambda) + \alpha = 0. \tag{5.5}$$

By convention, we assume that $\rho_1(0) = 1$. We further note that $\mathbf{E}_1 = (1, \cdots, 1)^{\mathbf{t}}$ is the associated eigenvector. As usual, we can perform a Taylor expansion of ρ_1 near $\theta = 0$ and we obtain that

$$\rho_1(\theta) = \exp\left(-\mathbf{i}\frac{\alpha+\beta-\lambda}{1-\beta+k(\lambda-\alpha)}\theta + \mathcal{O}(|\theta|^2)\right), \text{ as } \theta \to 0,$$

so that the associated wave speed is this time given by

$$c_0^k = \frac{\alpha+\beta-\lambda}{1-\beta+k(\lambda-\alpha)},$$

and depends explicitly on the delay k. We readily conclude that:

- When $\alpha < \lambda$, then c_0^k is well defined for all values of k. Furthermore, the amplitude of the wave speed $k \mapsto |c_0^k|$ decreases as k increases with $|c_0^k| \to 0$ as $k \to +\infty$. That is, the activity waves may go forward or backward (depending on the hyperparameter values), but the transmission delay always slows down their propagation.
- When $\alpha = \lambda$, then $c_0^k = \frac{\beta}{1-\beta} > 0$ is independent of the delay k. This is compatible with our implementation choice, where the initial feedforward propagation term (controlled by β) is not affected by transmission delays.
- When $\lambda < \alpha$, then c_0^k is well defined whenever $k \neq \frac{1-\beta}{\alpha-\lambda} > 0$. Furthermore, the wave speed $c_0^k > 0$ for $1 \leq k < \frac{1-\beta}{\alpha-\lambda}$ and increases with the delay k on that interval. That is, in this parameter range neural activity waves propagate forward and, perhaps counterintuively, accelerate when the transmission delay increases. On the other hand $c_0^k < 0$ for $k > \frac{1-\beta}{\alpha-\lambda}$ and $k \mapsto |c_0^k|$ decreases as k increases on that domain with $|c_0^k| \to 0$ as $k \to +\infty$. In this parameter range, waves propagate backward, and decelerate when the transmission delay increases.

Coming back to (5.5), we can look for other potential roots lying on the unit disk, i.e., marginally stable solutions. That is we look for $\omega \in (0, 2\pi)$ such that $\rho = e^{\mathbf{i}\omega}$. We obtain a system of two equations

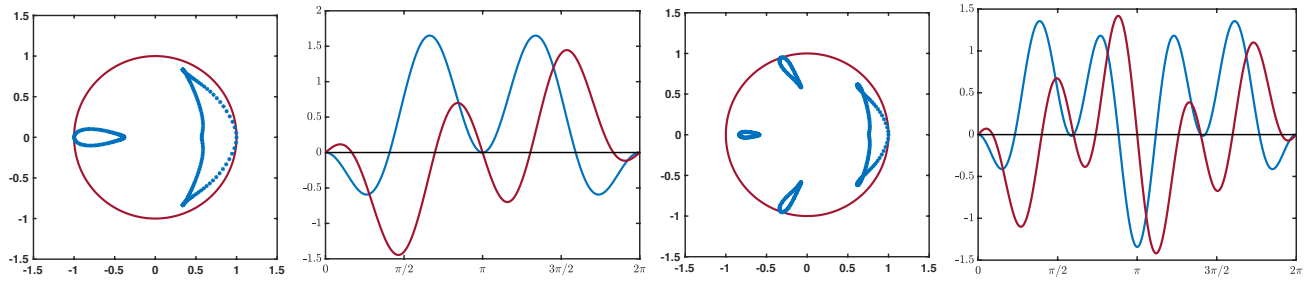
$$\begin{cases} (1-\beta)\cos((2k+1)\omega) - (1-\beta-\lambda)\cos(2k\omega) - (\alpha+\lambda)\cos(k\omega) + \alpha = 0, \\ (1-\beta)\sin((2k+1)\omega) - (1-\beta-\lambda)\sin(2k\omega) - (\alpha+\lambda)\sin(k\omega) = 0. \end{cases} \tag{5.6}$$

Case $k = 1$. When $k = 1$, coming back to (5.5), we see that the two other roots are real and given by $-\frac{\lambda}{2(1-\beta)} \pm \frac{\sqrt{\lambda^2+4\alpha(1-\beta)}}{2(1-\beta)}$, such that when $\alpha+\beta+\lambda = 1$ the negative root is precisely -1 such that $\omega = \pi$ is a solution which we assume, without loss of generality, to be the second root, that is $\rho_2(0) = -1$ whenever $\alpha+\beta+\lambda = 1$. In this specific case, the associated eigenvector is $\mathbf{E}_{-1} = (1, -1, 1)^{\mathbf{t}}$. Recall that $\mathbf{E}$ reflects the *history* of activity across the $2k+1 = 3$ preceding time steps. In this case, the eigenvector $\mathbf{E}_{-1}$ is a rapid alternation of activity, i.e. an oscillation. We refer to Figure 24(a) for an illustration of the spectral configuration in that case. We can perform a Taylor expansion of ρ_2 near $\theta = 0$ and we obtain that

$$\rho_2(\theta) = -\exp\left(-\mathbf{i}\frac{\alpha+\beta-\lambda}{5-5\beta-\alpha-3\lambda}\theta + \mathcal{O}(|\theta|^2)\right), \text{ as } \theta \to 0,$$

which provides an associated wave speed $\widetilde{c}_0$ given by

$$\widetilde{c}_0 = \frac{\alpha+\beta-\lambda}{5-5\beta-\alpha-3\lambda}.$$



(a) (b) (c) (d)

Figure 24: *Spectral configurations in the case $k = 1$ (a) and $k = 2$ (c) with tangency points associated to $\theta = 0$ in (5.5). In (b)-(d), we plot the left-hand side of the equations defining system (5.6) in the case $k = 1$ and $k = 2$ respectively where the first component is in blue and the second component in dark red. For $k = 1$, we have a solution at $\omega = 0$ and $\omega = \pi$ which can be seen in panel (a) with the tangency points at ± 1. For $k = 2$, we have three solutions $\omega = 0$ and $\omega \sim 1.885$ and $\omega \sim 2\pi - 1.885$ which can be seen in panel (c) with the tangency points at 1 and $e^{\pm \mathbf{i} 1.885}$. Parameters are set to $(\alpha, \beta, \lambda) = (0.4, 0.3, 0.3)$ in (a)-(b) and $(\alpha, \beta, \lambda) = (0.3, 0.3292, 0.3)$.*

As a consequence of the above analysis, if $\mathbf{G}_j^n$ denotes the fundamental solution of (5.3) starting from a Dirac delta mass centered at $j = 0$ along the direction $\mathbf{E} \in \mathbb{R}^3$, then we have the following representation for $\mathbf{G}_j^n$:

- If $\alpha + \beta + \lambda \neq 1$, then

$$\mathbf{G}_j^n \approx \frac{1}{\sqrt{4\pi\sigma_0^k n}} \exp\left(-\frac{|j - c_0^k n|^2}{4\sigma_0^k n}\right) \left\langle (0,0,1)^{\mathbf{t}}, \pi_1(\mathbf{E}) \right\rangle_{\mathbb{R}^3},$$

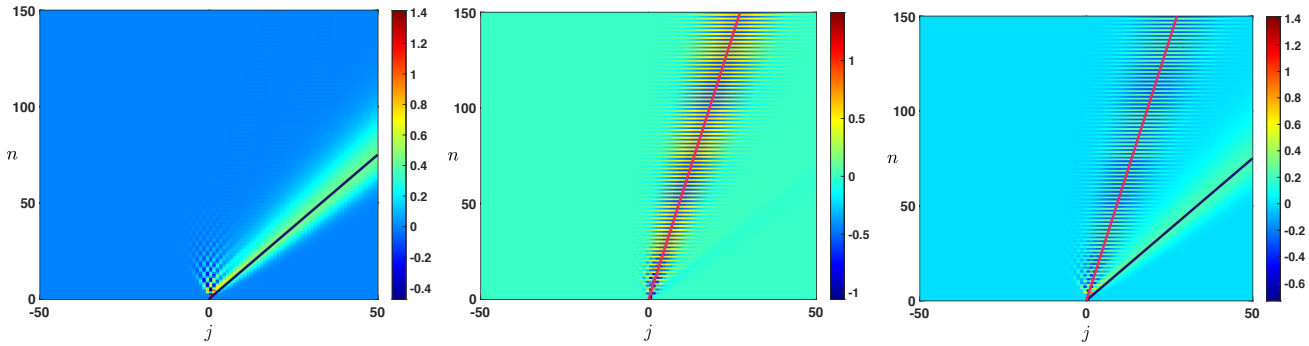
  where π_1 is the spectral projection of $\mathbb{R}^3$ along the direction $\mathbf{E}_1$ and $\langle \cdot, \cdot \rangle_{\mathbb{R}^3}$ is the usual scalar product. Here σ_0^k is some positive constant that can be computed explicitly by getting the higher order expansion of $\rho_1(\theta)$.

- If $\alpha + \beta + \lambda = 1$, then

$$\mathbf{G}_j^n \approx \frac{1}{\sqrt{4\pi\sigma_0^k n}} \exp\left(-\frac{|j - c_0^k n|^2}{4\sigma_0^k n}\right) \left\langle (0,0,1)^{\mathbf{t}}, \pi_1(\mathbf{E}) \right\rangle_{\mathbb{R}^3} + \frac{(-1)^n}{\sqrt{4\pi\widetilde{\sigma}_0 n}} \exp\left(-\frac{|j - \widetilde{c}_0 n|^2}{4\widetilde{\sigma}_0 n}\right) \left\langle (0,0,1)^{\mathbf{t}}, \pi_{-1}(\mathbf{E}) \right\rangle_{\mathbb{R}^3},$$

  where π_{-1} is the spectral projection of $\mathbb{R}^3$ along the direction $\mathbf{E}_{-1}$. Here $\widetilde{\sigma}_0$ is some positive constant that can be computed explicitly by getting the higher order expansion of $\rho_2(\theta)$.

In Figure 25, we illustrate the previous results in the case where $\alpha + \beta + \lambda = 1$. In panel (a), we have set $\mathbf{E} = \mathbf{E}_1$ (a constant history of activity over the previous 3 time steps), such that $\pi_1(\mathbf{E}_1) = \mathbf{E}_1$ and $\pi_{-1}(\mathbf{E}_1) = 0_{\mathbb{R}^3}$ so that we only observe a Gaussian profile propagating at speed c_0^k. On the other hand in panel (b), we have set $\mathbf{E} = \mathbf{E}_{-1}$ (an oscillating history of activity over the previous 3 time steps), such that $\pi_1(\mathbf{E}_{-1}) = 0_{\mathbb{R}^3}$ and $\pi_{-1}(\mathbf{E}_{-1}) = \mathbf{E}_{-1}$ so that we only observe an oscillating (in time) Gaussian wave profile propagating at speed $\widetilde{c}_0$. Note that in this case, the period of the oscillation is necessarily equal to $2k$, i.e. twice the transmission delay between layers. Finally in panel (c), we observe a super-position of the two Gaussian profiles propagating at speed c_0^1 and $\widetilde{c}_0$.



(a) $\mathbf{E} = \mathbf{E}_1 = (1,1,1)^{\mathbf{t}}$. (b) $\mathbf{E} = \mathbf{E}_{-1} = (1,-1,1)^{\mathbf{t}}$. (c) $\mathbf{E} = \frac{1}{2}\mathbf{E}_1 + \frac{1}{2}\mathbf{E}_{-1} = (1,0,1)^{\mathbf{t}}$.

Figure 25: *Space-time plots of the last component of the rescaled fundamental solution* $\mathbf{E}_j^n$ *of* (5.3) *starting from a Dirac delta mass centered at* $j = 0$ *along different directions* $\mathbf{E}$ *when* $k = 1$ *and* $\alpha + \beta + \lambda = 1$. *(a) When* $\mathbf{E} = \mathbf{E}_1$ *(constant history of activity) is the eigenvector associated to* $\mathcal{A}_1(1,0)$ *we observe propagation at wave speed* c_0^1. *(b) When* $\mathbf{E} = \mathbf{E}_{-1}$ *(oscillating history of activity) is the eigenvector associated to* $\mathcal{A}_1(-1,0)$ *we observe propagation of an oscillatory wave at wave speed* $\widetilde{c}_0$. *(c) When* $\mathbf{E}$ *is a linear combination of* $\mathbf{E}_1$ *and* $\mathbf{E}_{-1}$, *we observe two propagating waves (one of them oscillating) at wave speed* c_0^1 *and* $\widetilde{c}_0$ *respectively. Parameter values are set to* $(\alpha, \beta, \lambda) = (0.4, 0.3, 0.3)$.

Case $k \geq 2$**.** Studying the above system (5.6) in full generality is a very difficult task. We refer to Figure 24(c)-(d) for an illustration in the case $k = 2$ with three tangency points associated to $\theta = 0$ lying on the unit circle. Increasing the delay k while keeping fixed the other hyper-parameters (α, β, λ) will generically tend to destabilize the spectrum (as shown in Figure 26).

5.2 Continuous in time interpretation

As done before, we now re-examine our model (with transmission delays) in the time-continuous limit. First, we recall our notations for the scaled parameters

$$\widetilde{\beta} := \frac{\beta}{\Delta t}, \quad \widetilde{\lambda} := \frac{\lambda}{\Delta t}, \text{ and } \widetilde{\alpha} := \frac{\alpha}{\Delta t},$$

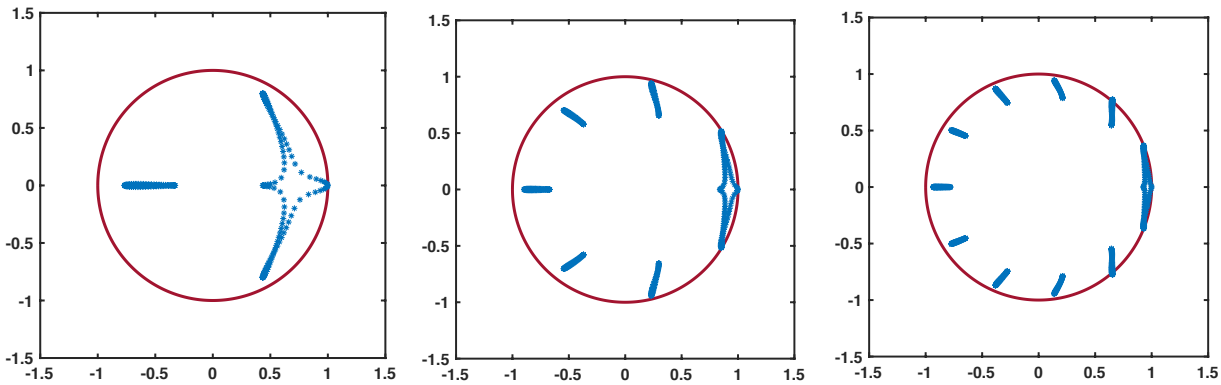
where $\Delta t > 0$ is some time step. Next we introduce the following rescaled time delay (representing the transmission time for neural signals between adjacent areas)

$$\tau := k\Delta t.$$

Identifying e_j^n as the approximation of some continuous fonction $\mathbf{e}_j(t_n)$ at $t_n = n\Delta t$, we readily derive a delayed version of (3.13), namely

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{e}_j(t) = \widetilde{\beta}\mathbf{e}_{j-1}(t) - (\widetilde{\beta} + \widetilde{\lambda})\mathbf{e}_j(t) + \widetilde{\alpha}\mathbf{e}_{j-1}(t-\tau) + \widetilde{\lambda}\mathbf{e}_{j+1}(t-\tau) - \widetilde{\alpha}\mathbf{e}_j(t-2\tau), \quad t > 0, \quad j \in \mathbb{Z}.$$

In what follows, we first investigate the case of homogeneous oscillations, which are now possible because of the presence of time delays into the equation. Then, we turn our attention to oscillatory traveling waves.



(a) $k = 1$. (b) $k = 3$. (c) $k = 5$.

Figure 26: *Destabilization of the spectrum by increasing the delay k while keeping fixed the hyper-parameters to* $(\alpha, \beta, \lambda) = (0.3, 0.1, 0.3)$.

5.2.1 Homogeneous oscillations

One key difference of the above delayed equation compared to (3.13) is that spatially homogeneous solutions (i.e., solutions $\mathbf{e}_j(t)$ that are independent of the layer j) may now have a non trivial dynamics, such as a broadly synchronized oscillation resembling brain rhythmic activity. Indeed, looking for solutions which are independent of j, we get the delayed ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{e}(t) = -\widetilde{\lambda}\mathbf{e}(t) + (\widetilde{\alpha} + \widetilde{\lambda})\mathbf{e}(t-\tau) - \widetilde{\alpha}\mathbf{e}(t-2\tau), \quad t > 0.$$

Looking for pure oscillatory exponential solutions $\mathbf{e}(t) = e^{\mathbf{i}\omega t}$ for some $\omega \in \mathbb{R}$ we obtain

$$\mathbf{i}\omega = -\widetilde{\lambda} + (\widetilde{\alpha} + \widetilde{\lambda})e^{-\mathbf{i}\tau\omega} - \widetilde{\alpha}e^{-2\mathbf{i}\tau\omega}.$$

This leads to the system of equations

$$\begin{cases} 0 = -\widetilde{\lambda} + (\widetilde{\alpha} + \widetilde{\lambda})\cos(\tau\omega) - \widetilde{\alpha}\cos(2\tau\omega), \\ \omega = -(\widetilde{\alpha} + \widetilde{\lambda})\sin(\tau\omega) + \widetilde{\alpha}\sin(2\tau\omega). \end{cases}$$

Introducing $\varrho = \widetilde{\lambda}/\widetilde{\alpha} > 0$, we observe that the above system writes instead

$$\begin{cases} 0 = -\varrho + (1+\varrho)\cos(\tau\omega) - \cos(2\tau\omega), \\ \omega = \widetilde{\alpha}\left(-(1+\varrho)\sin(\tau\omega) + \sin(2\tau\omega)\right). \end{cases} \tag{5.7}$$

Using trigonometry identities, the first equation can be factorized as

$$0 = (1 - \cos(\tau\omega))(\varrho - 1 - 2\cos(\tau\omega)).$$

We distinguish several cases. If $\varrho > 3$, then the above equation has solutions if and only if $\tau\omega = 2k\pi$ for $k \in \mathbb{Z}$. Inspecting the second equation, we see that necessarily $k = 0$ and $\omega = 0$ is the only possible solution. When $\varrho = 3$, we notice that the equation reduces to $0 = (1 - \cos(\tau\omega))^2$, and the solutions are

again given by $\tau\omega = 2k\pi$ for $k \in \mathbb{Z}$, which yields $\omega = 0$ because of the second equation. Now, if $\varrho \in (0,3)$, we deduce that either $\tau\omega = 2k\pi$ for $k \in \mathbb{Z}$ or

$$\cos(\tau\omega) = \frac{\varrho - 1}{2}.$$

In the first case, we recover that $\omega = 0$. Assuming now that $\omega \neq 0$, i.e., a true oscillation with non-zero frequency, we derive that

$$\tau\omega = \pm\arccos\left(\frac{\varrho - 1}{2}\right) + 2k\pi, \quad k \in \mathbb{Z}.$$

Injecting the above relation into the right-hand side of the second equation yields that

$$\omega = \widetilde{\alpha}\left(-(1+\varrho)\sin(\tau\omega) + \sin(2\tau\omega)\right) = \mp\widetilde{\alpha}\sqrt{(1+\varrho)(3-\varrho)},$$

and thus necessarily

$$(\tau, \omega) = \left(\frac{-\arccos\left(\frac{\varrho-1}{2}\right) + 2k\pi}{\widetilde{\alpha}\sqrt{(1+\varrho)(3-\varrho)}}, \pm\widetilde{\alpha}\sqrt{(1+\varrho)(3-\varrho)}\right), \quad k \in \mathbb{Z}.$$

We recover the fact that the system (5.7) is invariant by $\omega \mapsto -\omega$. Since $\arccos\left(\frac{\varrho-1}{2}\right) \in [0,\pi]$, we deduce that the smallest positive τ is always achieved at $k = 1$. We computed for several values of $\widetilde{\alpha}$ the corresponding values of τ and ω (for $k = 1$) as a function of ϱ, which are presented in Figure 27(a)-(b). We observe that for values of ϱ in the range $(1/2, 1)$ the corresponding time delay τ takes values between $12ms$ to $23ms$ for values of $1/\widetilde{\alpha}$ ranging from $5ms$ to $10ms$. Correspondingly, in the same range of values for ϱ, the frequency $\omega/2\pi$ takes values between $30Hz$ to $60Hz$.

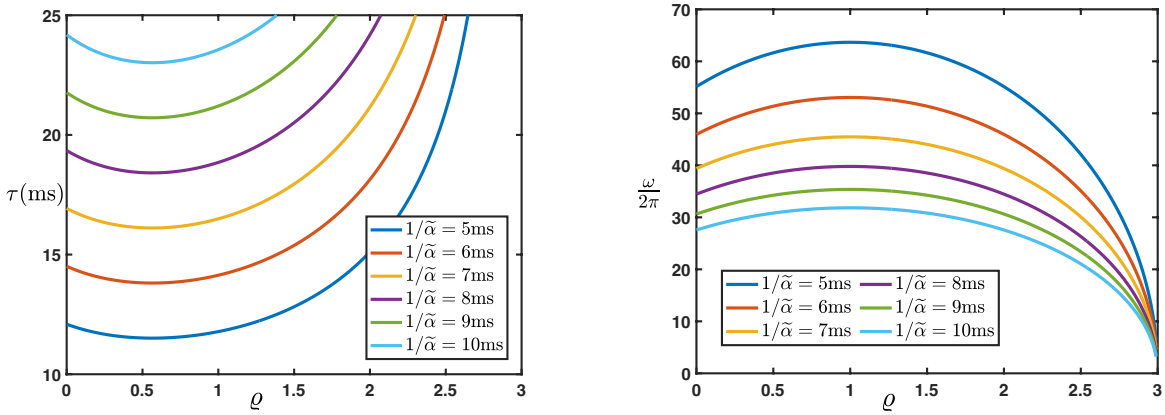
This tells us that, when the time delay τ is chosen to be around $10 - 20ms$, compatible with biologically plausible values for communication delays between adjacent cortical areas, and when hyperparameters $\widetilde{\alpha}$ and $\widetilde{\lambda}$ are suitably chosen ($\widetilde{\alpha}$ in particular must be strong enough to allow rapid feed-forward error correction updates, i.e. around $1/\widetilde{\alpha} < 8ms$, while $\widetilde{\lambda}$ can be chosen more liberally, as long as it stays $< 3\widetilde{\alpha}$), then the network produces globally synchronized oscillations, comparable to experimentally observed brain rhythms in the γ-band regime (30-60Hz). In this context, it is interesting to note that theoretical and neuroscientific considerations have suggested that error correction in predictive coding systems is likely to be accompanied by oscillatory neural activity around this same γ-frequency regime [4].

5.2.2 Oscillatory traveling waves

However, experimental and computational studies have also suggested that oscillatory signatures of predictive coding could be found at lower frequencies, in the so-called α-band regime, around 7-15Hz. Furthermore, these oscillations are typically not homogeneous over space, as assumed in the previous section, but behave as forward- or backward-travelling waves with systematic phase shifts between layers [1]. To explore this idea further, we now investigate the possibility of having traveling wave solutions of the form

$$\mathbf{e}_j(t) = e^{\mathbf{i}(\omega t + j\theta)}, \quad t > 0, \quad j \in \mathbb{Z},$$

for some $\omega \in \mathbb{R}$ (representing the wave's temporal frequency) and $\theta \in [0, 2\pi)$ (representing the wave's spatial frequency, i.e. its phase shift across layers), and we are especially interested in deriving conditions



(a) (b)

Figure 27: *(a) Representation of the (minimal) time delay τ expressed in milliseconds as a function of the parameter ϱ for various values of $1/\widetilde{\alpha}$ ranging from $5ms$ to $10ms$. We observe that for values of ϱ in the range $(1/2, 1)$ the corresponding time delay τ takes values between $12ms$ to $23ms$. (b) Representation of the frequency $\omega/2\pi$ (in Hertz) as a function of the parameter ϱ for various values of $1/\widetilde{\alpha}$ ranging from $5ms$ to $10ms$. We observe that for values of ϱ in the range $(1/2, 1)$ the corresponding frequency $\omega/2\pi$ takes values between $30Hz$ to $60Hz$.*

under which one can ensure that $\theta \neq 0$ (since otherwise, we would be again facing the homogeneous oscillation case). We only focus on the case $\widetilde{\beta} = 0$ (as postulated, e.g. in Rao and Ballard's work [27]) and leave the case $\widetilde{\beta} > 0$ for future investigations. As a consequence, the equation reduces to

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{e}_j(t) = -\widetilde{\lambda}\mathbf{e}_j(t) + \widetilde{\alpha}\mathbf{e}_{j-1}(t-\tau) + \widetilde{\lambda}\mathbf{e}_{j+1}(t-\tau) - \widetilde{\alpha}\mathbf{e}_j(t-2\tau), \quad t > 0, \quad j \in \mathbb{Z}.$$

Plugging the ansatz $\mathbf{e}_j(t) = e^{\mathbf{i}(\omega t + j\theta)}$, we obtain:

$$\mathbf{i}\omega = \widetilde{\alpha}\left(e^{-\mathbf{i}(\omega\tau+\theta)} - e^{-2\mathbf{i}\omega\tau}\right) + \widetilde{\lambda}\left(e^{-\mathbf{i}(\omega\tau-\theta)} - 1\right).$$

Taking real and imaginary parts, we obtain the system

$$\begin{cases} 0 = \widetilde{\alpha}\left(\cos(\omega\tau+\theta) - \cos(2\omega\tau)\right) + \widetilde{\lambda}\left(\cos(\omega\tau-\theta) - 1\right), \\ \omega = -\widetilde{\alpha}\left(\sin(\omega\tau+\theta) - \sin(2\omega\tau)\right) - \widetilde{\lambda}\sin(\omega\tau-\theta). \end{cases}$$

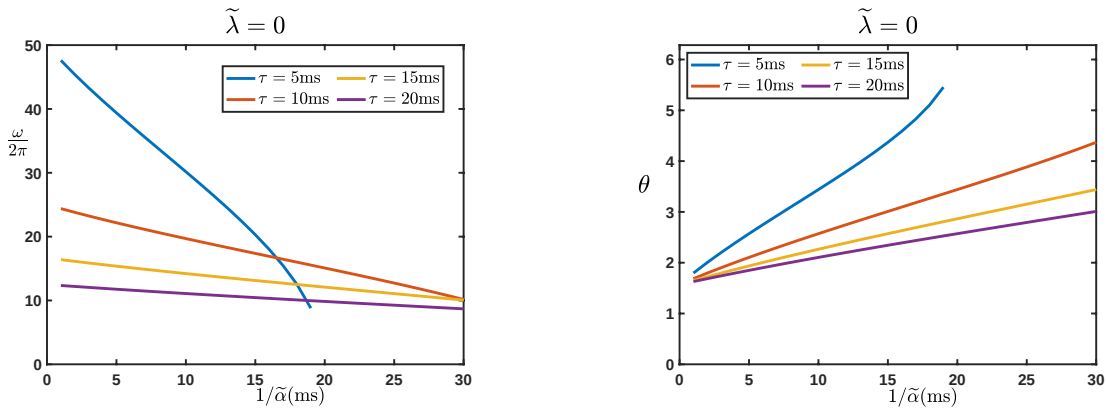
Once again, we introduce $\varrho := \frac{\widetilde{\lambda}}{\widetilde{\alpha}} \geq 0$ where we implicitly assumed that we always work in the regime $\widetilde{\alpha} > 0$. Then, we note that the right-hand side of the first equation of the above system can be factored as

$$\widetilde{\alpha}\left(\cos(\omega\tau+\theta) - \cos(2\omega\tau)\right) + \widetilde{\lambda}\left(\cos(\omega\tau-\theta) - 1\right) = -2\widetilde{\alpha}\sin\left(\frac{\theta-\omega\tau}{2}\right)\left(\varrho\sin\left(\frac{\theta-\omega\tau}{2}\right) + \sin\left(\frac{\theta+3\omega\tau}{2}\right)\right).$$

As a consequence, either $\sin\left(\frac{\theta-\omega\tau}{2}\right) = 0$, that is $\omega\tau = \theta + 2k\pi$ for $k \in \mathbb{Z}$, which then leads, from the second equation, to $\omega = 0$ and $\theta = 0$ since we restrict $\theta \in [0, 2\pi)$, or $\sin\left(\frac{\theta-\omega\tau}{2}\right) \neq 0$. In the latter case, assuming that $\omega\tau \neq \theta + 2k\pi$ for $k \in \mathbb{Z}$, we get that

$$0 = \varrho\sin\left(\frac{\theta-\omega\tau}{2}\right) + \sin\left(\frac{\theta+3\omega\tau}{2}\right).$$

We will now study several cases.



(a) (b)

Figure 28: *Representation of the temporal frequency $\omega/2\pi$ (in Hz) and the spatial frequency $\theta \in [0, 2\pi)$ (panel (a) and (b) respectively) in the case $\widetilde{\lambda} = 0$ as a function of $1/\widetilde{\alpha}$ (in ms) for several values of the time delay τ.*

Case $\varrho = 0$. (In other words, this case implies $\widetilde{\lambda} = 0$, that is, a system with no feedback error correction.) From $\sin\left(\frac{\theta+3\omega\tau}{2}\right) = 0$, we deduce that $\theta = -3\omega\tau + 2k\pi$ for some $k \in \mathbb{Z}$, and reporting into the second equation of the system, we end up with

$$\omega = 2\widetilde{\alpha}\sin(2\omega\tau).$$

We always have the trivial solution $\omega = 0$ with $\theta = 0$. In fact, when $4\widetilde{\alpha}\tau \leq 1$, $\omega = 0$ is the only solution of the above equation. On the other hand, when $4\widetilde{\alpha}\tau > 1$, there can be multiple non trivial solutions. At least, for each $(\widetilde{\alpha}, \tau)$ such that $4\widetilde{\alpha}\tau > 1$ there always exist a unique $\omega_c(\widetilde{\alpha}, \tau) \in \left(0, \frac{\pi}{2\tau}\right)$ solution of the above equation. This gives a corresponding $\theta_c^k = -3\omega_c(\widetilde{\alpha}, \tau)\tau + 2k\pi$ with $k \in \mathbb{Z}$, and retaining the corresponding value of θ in the interval $[0, 2\pi)$, we have $\theta_c = -3\omega_c(\widetilde{\alpha}, \tau)\tau + 2\pi$. We refer to Figure 28 for an illustration of the solutions (ω, θ) for several values of the parameters.

Interestingly, we see that for biologically plausible values of the time delay τ between $10ms$ and $20ms$, the observed oscillation frequency is lower than in the previous case, and now compatible with the α-frequency regime (between $10Hz$ and $20Hz$). Furthermore, the phase shift between layers θ varies roughly between 2 and 4 radians. As phase shifts below π or above π radians indicate respectively backward- or forward-travelling waves, we see that the exact value of the parameters τ and $\widetilde{\alpha}$ critically determines the propagation direction of the travelling waves: stronger feedforward error correction (lower values of $1/\widetilde{\alpha}$) and longer communication delays τ will tend to favor backward-travelling waves; and vice-versa, weaker feedforward error correction (higher values of $1/\widetilde{\alpha}$) and shorter communication delays τ will favor forward-travelling waves.

Case $\varrho = 1$. Now we assume that $\widetilde{\lambda} \neq 0$, that is, the system now includes feedback error correction. At first, we consider the simpler case when $\varrho = 1$, that is when $\widetilde{\alpha} = \widetilde{\lambda}$, where the equation can also be solved easily. Indeed, we have either

$$\frac{\omega\tau - \theta}{2} = \frac{\theta + 3\omega\tau}{2} + 2k\pi, \quad k \in \mathbb{Z},$$

or

$$\frac{\omega\tau - \theta}{2} = \pi - \frac{\theta + 3\omega\tau}{2} + 2k\pi, \quad k \in \mathbb{Z}.$$

This equivalent to

$$\theta = -\omega\tau + 2k\pi, \quad k \in \mathbb{Z},$$

or

$$\omega\tau = \frac{\pi}{2} + k\pi, \quad k \in \mathbb{Z}.$$

Let assume first that $\theta = -\omega\tau + 2k\pi$ for some $k \in \mathbb{Z}$, then the second equation of the system gives $\omega = 0$ since $\widetilde{\alpha} = \widetilde{\lambda}$ when $\varrho = 1$, and thus we end up with $\theta = 0$. Now, if $\omega\tau = \frac{\pi}{2} + k\pi$ for some $k \in \mathbb{Z}$, the second equation leads to

$$\omega = -2\widetilde{\alpha}\cos(\theta + k\pi),$$

from which we deduce that necessarily we must have

$$\frac{(2k+1)\pi}{-4\widetilde{\alpha}\tau} = \cos(\theta + k\pi), \quad k \in \mathbb{Z}.$$

We first remark that if $4\widetilde{\alpha}\tau < \pi$, then the above equation has no solution. On the other hand if $4\widetilde{\alpha}\tau \geq \pi$, one can obtain solutions to the above equation. Indeed, let us denote

$$p := \left\lfloor \frac{4\widetilde{\alpha}\tau}{\pi} \right\rfloor \geq 1$$

the integer part of $\frac{4\widetilde{\alpha}\tau}{\pi}$. Then, for each $k \in \mathbb{Z}$ such that $|2k+1| \leq p$, we have

$$\theta = \pm\text{arcos}\left((-1)^{k+1}\frac{(2k+1)\pi}{4\widetilde{\alpha}\tau}\right) + 2m\pi, \quad m \in \mathbb{Z},$$

with corresponding ω given by

$$\omega = \frac{\pi}{2\tau} + \frac{k\pi}{\tau}.$$

That is, the set of solutions is given by

$$(\omega, \theta) = \left(\frac{\pi}{2\tau} + \frac{k\pi}{\tau}, \pm\text{arcos}\left((-1)^{k+1}\frac{(2k+1)\pi}{4\widetilde{\alpha}\tau}\right) + 2m\pi\right),$$

for each $k \in \mathbb{Z}$ such that $|2k+1| \leq \left\lfloor \frac{4\widetilde{\alpha}\tau}{\pi} \right\rfloor$ and $m \in \mathbb{Z}$. If we only retain the smallest ω and the corresponding value of $\theta \in [0, 2\pi)$, we must take $k = 0$, and we have two solutions

$$(\omega, \theta) = \left(\frac{\pi}{2\tau}, \text{arcos}\left(-\frac{\pi}{4\widetilde{\alpha}\tau}\right)\right), \text{ and } (\omega, \theta) = \left(\frac{\pi}{2\tau}, -\text{arcos}\left(-\frac{\pi}{4\widetilde{\alpha}\tau}\right) + 2\pi\right) \quad \text{if } 4\widetilde{\alpha}\tau \geq \pi.$$

We note that in this case the temporal frequency only depends on the time delay τ since it is given by $\frac{\omega}{2\pi} = \frac{1}{4\tau}$ and ranges from 12.5Hz (α-band regime) to 25Hz (β-band regime) for values of τ between 10ms to 20ms (as long as $\widetilde{\alpha}$ is fixed such that $4\widetilde{\alpha}\tau \geq \pi$ is verified). The corresponding spatial frequencies are $\text{arcos}\left(-\frac{\pi}{4\widetilde{\alpha}\tau}\right) \in (\pi/2, \pi)$ and $-\text{arcos}\left(-\frac{\pi}{4\widetilde{\alpha}\tau}\right) + 2\pi \in (3\pi/2, 2\pi)$.

In summary, when the feedforward and feedback error correction strengths are matched (that is when $\widetilde{\alpha} = \widetilde{\lambda}$) and sufficiently high (such that $4\widetilde{\alpha}\tau \geq \pi$), then the system will show two simultaneous travelling waves at the same frequency in the α-band or β-band regime, but travelling in opposite directions, one as a feedforward wave and the other as a feedback wave.

Case $\varrho > 1$. Here, the feedback error correction λ is stronger than the feedforward α. In this case, we remark that

$$\varrho \sin\left(\frac{\theta - \omega\tau}{2}\right) + \sin\left(\frac{\theta + 3\omega\tau}{2}\right) = (\varrho + \cos(2\omega\tau)) \sin\left(\frac{\theta - \omega\tau}{2}\right) + \cos\left(\frac{\theta - \omega\tau}{2}\right) \sin(2\omega\tau).$$

Since $\varrho > 1$ and $\omega\tau \neq \theta + 2k\pi$ for $k \in \mathbb{Z}$, we have $(\varrho + \cos(2\omega\tau)) \sin\left(\frac{\theta-\omega\tau}{2}\right) \neq 0$, and thus $\cos\left(\frac{\theta-\omega\tau}{2}\right) \sin(2\omega\tau) \neq 0$ otherwise we would reach a contradiction since we try to solve

$$\varrho \sin\left(\frac{\theta - \omega\tau}{2}\right) + \sin\left(\frac{\theta + 3\omega\tau}{2}\right) = 0.$$

As a consequence, $\cos\left(\frac{\theta-\omega\tau}{2}\right) \neq 0$ and we can rewrite the above equation as

$$\tan\left(\frac{\theta - \omega\tau}{2}\right) = -\frac{\sin(2\omega\tau)}{\varrho + \cos(2\omega\tau)},$$

so that

$$\theta = \omega\tau - 2\arctan\left(\frac{\sin(2\omega\tau)}{\varrho + \cos(2\omega\tau)}\right) + 2k\pi, \quad k \in \mathbb{Z}.$$

Injecting this expression for θ into the second equation, we find, after simplification, that

$$\omega = \widetilde{\alpha} \frac{2\sin(2\omega\tau)\left(1 - \varrho^2\right)}{2\varrho\cos(2\omega\tau) + \varrho^2 + 1}.$$

We first remark that $\omega = 0$ is always a solution, giving $\theta = 0$. Now, inspecting the right-hand of the above expression, we get that

$$\frac{2\widetilde{\alpha}\left(1 - \varrho^2\right)}{2\varrho\cos(2\omega\tau) + \varrho^2 + 1} < 0, \text{ for all } \omega \in \mathbb{R}.$$

As a consequence, we look for the negative minima of the function $\omega \mapsto \frac{\sin(2\omega\tau)}{2\varrho\cos(2\omega\tau)+\varrho^2+1}$ which are given by $\omega_0 = \frac{\pi}{2\tau} + \frac{1}{2\tau}\text{arcos}\left(\frac{2\varrho}{1+\varrho^2}\right) + \frac{k\pi}{\tau}$ for $k \in \mathbb{Z}$, at such minima, one gets that

$$\frac{\sin(2\omega_0\tau)}{2\varrho\cos(2\omega_0\tau) + \varrho^2 + 1} = \frac{1}{1 - \varrho^2}.$$

This implies that if $4\widetilde{\alpha}\tau < \pi + \text{arcos}\left(\frac{2\varrho}{1+\varrho^2}\right)$, then there is no other solution than $(\omega, \theta) = (0, 0)$. As a consequence, one needs to assume $4\widetilde{\alpha}\tau \geq \pi + \text{arcos}\left(\frac{2\varrho}{1+\varrho^2}\right)$ to ensure the existence of at least one non trivial solution. We remark that this condition is consistent with our condition $4\widetilde{\alpha}\tau \geq \pi$ derived in the case $\varrho = 1$.

Case $0 < \varrho < 1$. We start once again from the equation

$$0 = (\varrho + \cos(2\omega\tau)) \sin\left(\frac{\theta - \omega\tau}{2}\right) + \cos\left(\frac{\theta - \omega\tau}{2}\right) \sin(2\omega\tau).$$

This time, it is possible that $\varrho + \cos(2\omega\tau) = 0$, which gives necessarily that

$$\omega\tau = \pm\frac{1}{2}\text{arcos}(-\varrho) + k\pi, \quad k \in \mathbb{Z}.$$

But if $\varrho + \cos(2\omega\tau) = 0$, then one has $0 = \cos\left(\frac{\theta - \omega\tau}{2}\right)\sin(2\omega\tau)$.

Let us first assume that it is $0 = \cos\left(\frac{\theta-\omega\tau}{2}\right)$, such that $\theta = \omega\tau + (2k+1)\pi$ for $k \in \mathbb{Z}$. Now, looking at the second equation, we find that

$$\omega = 2\widetilde{\alpha}\sin(2\omega\tau) = \pm 2\widetilde{\alpha}\sqrt{1-\varrho^2},$$

which implies that it is possible only if

$$\tau = \frac{\operatorname{arcos}(-\varrho) + k\pi}{4\widetilde{\alpha}\sqrt{1-\varrho^2}}, \quad k \geq 0.$$

As a conclusion, if τ and $0 < \varrho < 1$ satisfy $\tau = \frac{\operatorname{arcos}(-\varrho)+k\pi}{4\widetilde{\alpha}\sqrt{1-\varrho^2}}$ for some positive integer $k \geq 0$, then

$$(\omega, \theta) = \left(2\widetilde{\alpha}\sqrt{1-\varrho^2}, \frac{1}{2}\operatorname{arcos}(-\varrho) - \pi\right), \quad \text{and} \quad (\omega, \theta) = \left(-2\widetilde{\alpha}\sqrt{1-\varrho^2}, -\frac{1}{2}\operatorname{arcos}(-\varrho) + \pi\right),$$

are corresponding solutions of the problem.

Next, let us assume that it is $\sin(2\omega\tau) = 0$, implying that $2\omega\tau = k\pi$ for $k \in \mathbb{Z}$. Now we readily remark that since $0 < \varrho < 1$, we have $\operatorname{arcos}(-\varrho) \in (\pi/2, \pi)$. As a consequence, we should have

$$\pm\operatorname{arcos}(-\varrho) + 2k\pi = p\pi, \quad k, p \in \mathbb{Z},$$

this is impossible and thus $\sin(2\omega\tau) \neq 0$ and we are back to the case treated before.

We now assume that $\varrho + \cos(2\omega\tau) \neq 0$. In that case, we can proceed as in the case $\varrho > 1$ and obtain that

$$\theta = \omega\tau - 2\arctan\left(\frac{\sin(2\omega\tau)}{\varrho + \cos(2\omega\tau)}\right) + 2k\pi, \quad k \in \mathbb{Z},$$

which gives

$$\omega = \widetilde{\alpha}\frac{2\sin(2\omega\tau)\left(1-\varrho^2\right)}{2\varrho\cos(2\omega\tau) + \varrho^2 + 1}.$$

Once again, $(\omega, \theta) = (0, 0)$ is always a solution. What changes in this case is that now

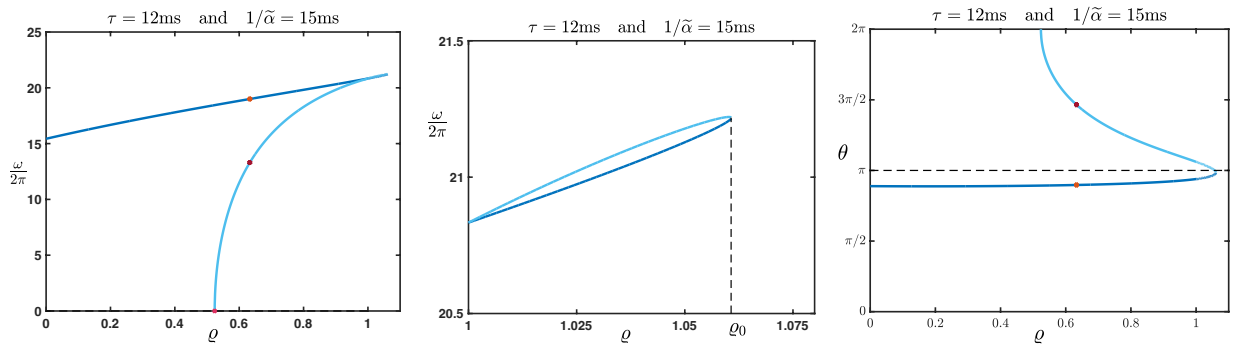
$$\frac{2\widetilde{\alpha}\left(1-\varrho^2\right)}{2\varrho\cos(2\omega\tau) + \varrho^2 + 1} > 0, \text{ for all } \omega \in \mathbb{R}.$$

This time, one needs to look at the positive maxima of the map $\omega \mapsto \frac{\sin(2\omega\tau)}{2\varrho\cos(2\omega\tau)+\varrho^2+1}$ which are given by $\omega_0 = \frac{\pi}{2\tau} - \frac{1}{2\tau}\operatorname{arcos}\left(\frac{2\varrho}{1+\varrho^2}\right) + \frac{k\pi}{\tau}$ for $k \in \mathbb{Z}$, at such maxima, one gets that

$$\frac{\sin(2\omega_0\tau)}{2\varrho\cos(2\omega_0\tau) + \varrho^2 + 1} = \frac{1}{1-\varrho^2}.$$

As a consequence, if $4\widetilde{\alpha}\tau < \pi - \operatorname{arcos}\left(\frac{2\varrho}{1+\varrho^2}\right)$, then there is no other solution than $(\omega, \theta) = (0, 0)$. To obtain at least one non trivial positive solution, one needs to impose that $4\widetilde{\alpha}\tau \geq \pi - \operatorname{arcos}\left(\frac{2\varrho}{1+\varrho^2}\right)$. Once again, this condition is consistent with the condition $4\widetilde{\alpha}\tau \geq \pi$ derived in the case $\varrho = 1$. We can also derive a second simple condition which ensures the existence of a non trivial solution by looking at the behavior near $\omega \sim 0$ where we have

$$\widetilde{\alpha}\frac{2\sin(2\omega\tau)\left(1-\varrho^2\right)}{2\varrho\cos(2\omega\tau) + \varrho^2 + 1} \sim \frac{4\widetilde{\alpha}\tau\left(1-\varrho\right)}{1+\varrho}\omega.$$



(a) $\omega/2\pi$ as a function of ϱ. (b) Zoom of panel (a) near $\varrho \sim 1$. (c) θ as a function of ϱ.

Figure 29: *Representation of the temporal frequency $\omega/2\pi$ (in Hz) and the spatial frequency $\theta \in [0, 2\pi)$ (panel (a) and (c) respectively) as a function of ϱ for fixed values of the time delay $\tau = 12ms$ and $1/\widetilde{\alpha} = 15ms$. Panel (b) represents a zoom of panel (a) near $\varrho \sim 1$ where the two branches terminate.*

Thus if

$$4\widetilde{\alpha}\tau > \frac{1+\varrho}{1-\varrho},$$

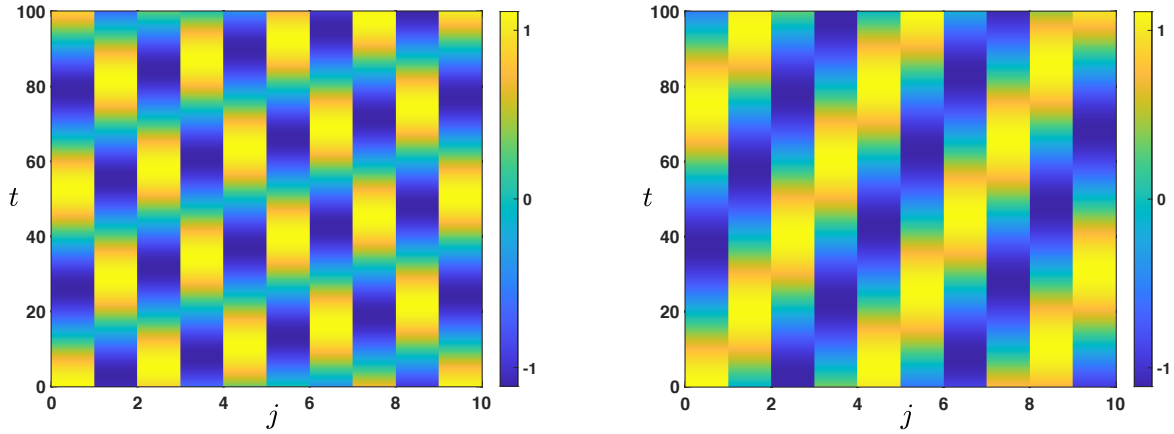
then there exists at least one positive solution $\omega \in (0, \pi/2\tau)$ to the above equation (and also one negative solution in $(-\pi/2\tau, 0)$ by symmetry). Note that the condition $4\widetilde{\alpha}\tau > \frac{1+\varrho}{1-\varrho}$ is consistent with the condition $4\widetilde{\alpha}\tau > 1$ derived in the case $\varrho = 0$.

Examples. To illustrate the different scenarios and their possible interpretations in terms of brain oscillations, we take here two distinct examples corresponding to the situations described above. We report in Figure 29 the non trivial branches of solutions corresponding to positive values of ω, as a function of ϱ for fixed values of the time delay $\tau = 12$ms and $1/\widetilde{\alpha} = 15$ms. These values are biologically plausible and correspond to the values used in [1]. Upon denoting

$$\varrho_c := \frac{1 - 4\widetilde{\alpha}\tau}{1 + 4\widetilde{\alpha}\tau} \in (0, 1),$$

for all $\varrho \in [0, \varrho_c)$, we get the existence of a unique branch of solution (blue curve) for the temporal frequency $\omega/2\pi$. A second branch (light blue curve) of solutions emerges precisely at $\varrho = \varrho_c$. These two branches cross at $\varrho = 1$ where $\omega/2\pi = \frac{1}{4\tau}$ and terminate at a value of $\varrho = \varrho_0 \sim 1.06$ (see Figure 29(b)). The branch of solutions which exists for all values of $\varrho \in [0, \varrho_0]$ has an associated spatial frequency which is almost constant and whose value is around $\sim 2.82 \in (0, \pi)$. On the other hand, the branch of solutions which only exists for values of $\varrho \in (\varrho_c, \varrho_0]$ has an associated spatial frequency which lies in $(\pi, 2\pi)$. Let us remark that at $\varrho = 1$, the spatial frequencies of the two solutions are different and symmetric with respect to π. Furthermore, at $\varrho = \varrho_0 \sim 1.06$ where the two branches collide the associated spatial frequency is $\theta \sim \pi$. Let us finally note that for $\varrho \in [1, \varrho_0]$, the spatial frequencies of the two branches are almost identical, although the secondary branch is slightly above the primary one.

Correspondingly we illustrate in Figure 30, the space-time plot of two points along the two different branches which correspond to the orange and dark red points in Figure 29. The corresponding values are $(\omega, \theta) \sim (0.12, 2.82)$ and $(\omega, \theta) \sim (0.08, 4.60)$ and associated to the same value of $\varrho \sim 0.633$. In



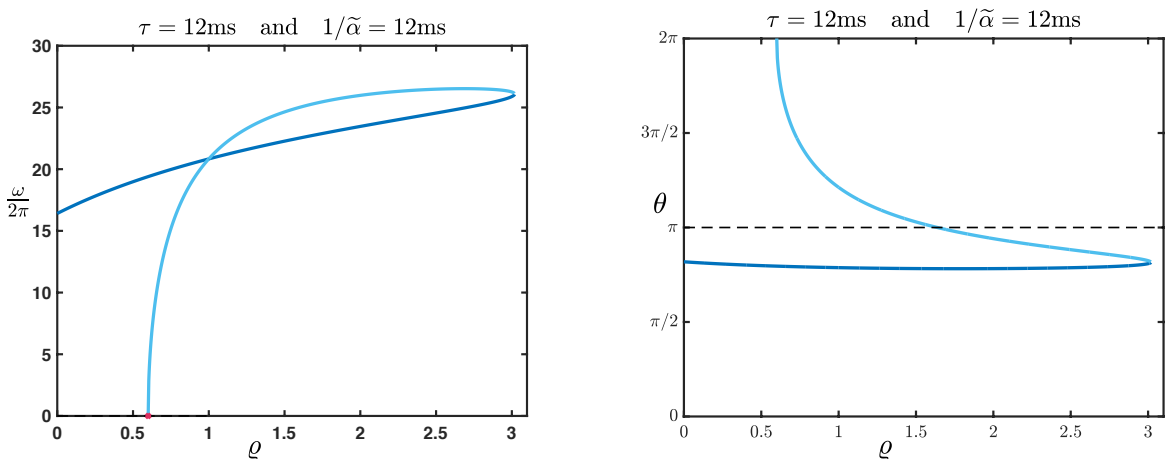
(a) $(\omega, \theta) \sim (0.12, 2.82)$ (b) $(\omega, \theta) \sim (0.08, 4.60)$

Figure 30: *Space-time plot of* $\cos(\omega t + \theta j)$ *for values of* (ω, θ) *which correspond to the orange and dark red points of Figure 29 lying respectively on the blue and light blue curves, time* t *is in ms. In (a), the temporal frequency is* $\omega/2\pi \sim 19Hz$ *while in (b) it is* $\omega/2\pi \sim 13.3Hz$*. In (a), since* $\theta \in (0, \pi)$*, we observe a backward propagation of the wave while in (b) we have a forward propagation since* $\theta \in (\pi, 2\pi)$*.*

the first panel of Figure 30(a), which corresponds to the point on the branch of solution defined for all $\varrho \in [0, \varrho_0]$, since the corresponding value of the spatial frequency is $\theta \in (0, \pi)$, we observe an apparent backward propagation, while in the second panel of Figure 30(b), we observe a forward propagation. This corresponds to the point on the lower branch of the solutions defined for values of $\varrho \in (\varrho_c, \varrho_0]$ with associated spatial frequency $\theta \in (\pi, 2\pi)$. From a biological point of view, this indicates that the more interesting range of the parameters is the one with $\varrho \in (\varrho_c, \varrho_0]$ and the corresponding branch of solutions which emerges at $\varrho = \varrho_c$ from the trivial solution $(\omega, \theta) \sim (0, 0)$ since in this case we obtain an oscillatory traveling wave with forward propagation into the network.

In Figure 31, we show the global structure of the branches for a second example, with fixed values of the time delay $\tau = 12$ms and $1/\widetilde{\alpha} = 12$ms, which are still biologically relevant values. We observe that the two branches terminate at a value of $\varrho = \varrho_0 \sim 3.03$ with a crossing at $\varrho = 1$. For $\varrho \in [1, \varrho_0]$, the primary branch (blue curve) has a temporal frequency below the secondary branch (light blue curve), the difference in frequencies is almost 5Hz for values of $\varrho \sim 2$. Even more interestingly, we see that the corresponding spatial frequencies along the secondary branch are decreasing from 2π to a final value below π at ϱ_0 indicating that by increasing the value of ϱ we can reverse the direction of propagation from forward to backward oscillatory traveling waves. The transition occurs for $\varrho \sim 1.65$, that is for values of $1/\widetilde{\lambda} \sim 7-8$ms. It is further noticed that the associated temporal frequencies in the backward regime are around 25Hz (β-frequency regime) much higher than for forward traveling waves whose temporal frequencies range from 0Hz to 20Hz (and include the α-frequency regime).

Summary In this section, we saw that including temporal delays in the time-continuous version of the system produces non-trivial dynamics that can be characterized analytically. Contrary to the discrete version of the system, which can only be analyzed for a few discrete time delays $k = 1, 2, ...$, the continuous version is informative for a wide range of biologically plausible delays, and the resulting frequencies are very diverse. In particular, we observed homogenous synchronized oscillations in the gamma band $(30 - 60Hz)$



(a) (b)

Figure 31: *Representation of the temporal frequency $\omega/2\pi$ (in Hz) and the spatial frequency $\theta \in [0, 2\pi)$ (panel (a) and (b) respectively) as a function of $\varrho \in [0, 1]$ for fixed values of the time delay $\tau = 12ms$ and $1/\widetilde{\alpha} = 12ms$.*

that emerged when the feed-forward error correction term $\widetilde{\alpha}$ was strong enough (roughly, with $1/\widetilde{\alpha} < 8ms$). But we also found situations in which the oscillatory activity was not homogenous, but propagated as a travelling wave through the network. With biologically plausible values for the various parameters, the waves could propagate forward in the alpha-band ($7-15Hz$) frequency range, and when the feedback error correction term $\widetilde{\lambda}$ was strong enough (e.g. $1/\widetilde{\lambda} < 8ms$ while $1/\widetilde{\alpha} = 12ms$), they started moving backward at a faster frequency in the beta-band ($15 - 30Hz$). Altogether, this pattern of results is compatible with various (sometimes conflicting) observations from the Neuroscience literature [1, 4], and informs us about the conditions in which the corresponding dynamic behaviors might emerge in the brain.

6 Discussion

6.1 Contributions

We proposed a mathematical framework to explore the properties and stability of neural network models of the visual system comprising a hierarchy of visual processing areas (or "layers"), mutually connected according to the principles of predictive coding. Using a discrete model, as is typically done in the recent deep learning literature, we introduced the amplification factor function, which serves to characterize the interesting (i.e., "marginally stable") regions as a function of the model hyperparameters. When considered on an infinite domain, we showed that the response of our linear neural network to a Dirac delta initialization presents a universal behavior given by a Gaussian profile with fixed variance and which spreads at a given speed. Both speed and variance could be explicitly characterized in terms of the model hyperparameters. This universal Gaussian profile was then the key to understand the long-time dynamics of the linear neural network set on a semi-infinite domain with a fixed constant source term at the left boundary of the network.

At first, we ignored the influence of neuronal selectivity and used feed-forward and feedback connection

matrices set to the identity matrix. When $\beta = 0$ (no feedforward update after the network initialization), we observed that hyperparameters α and λ compete for forward and backward propagation, respectively. When $\beta > 0$, the constant feedforward input makes things more complex, with λ (feedback error correction) now competing with $\beta + \alpha$ (feedforward drive and feedforward error correction). In the special case when $\alpha + \lambda = 1$, a second (but spurious) mode of propagation with rapidly alternating activity can emerge, whose direction is determined by the competition between α and $\beta + \lambda$.

Next, to evaluate the influence of a more complex and functionally relevant connectivity matrix, we defined *neural assemblies* reflecting the eigenvectors of the matrix. Each of these neural assemblies can be analyzed separately, and its behavior depends on the corresponding eigenvalue (in addition to the hyperparameters α, β and λ, as explained above). Different assemblies can simultaneously support different dynamics, so that some may propagate information forward, others may not propagate at all (acting as a filter on the inputs), while yet others might propagate backward (e.g. carrying "priors" set by preceding activations). We again saw a number of cases where "fringe" or spurious behavior arose, e.g. rapid alternations in activity, and understood that this could be caused by the discrete nature of our model, when the time steps defining the model's temporal resolution are too coarse.

The time-continuous version of the model helped us overcome this issue, and characterize dynamics in the limit of infinitely small time steps. The amplification factor function is still crucial in this situation, but it produces more robust results, without fringe behavior or spurious oscillations. In particular, the analysis of stability and propagation direction/speed was greatly simplified in this continuous case.

The same time-continuous model also allowed us to investigate the inclusion of biologically plausible communication delays between layers. In this case, we demonstrated the emergence of genuine oscillatory dynamics and travelling waves in various frequency bands compatible with neuroscientific observations (alpha-band from 7 to 15Hz, beta-band from 15 to 30Hz and gamma-band from 30 to 60Hz).

Finally, we considered fully continuous versions of the model, not only in time but also in space, both across network depth (across neuronal layers) and width (across neurons in the same layer). This mathematical abstraction revealed that our model could be understood as a transport equation, and that it produced diffusion dynamics.

6.2 Biological interpretations

The mathematical framework that we proposed naturally lends itself to interpretation in biological terms. The model's hyperparameters reflect the strength of feedforward and feedback signalling in the brain. These are determined not only by axonal density and synaptic strength (that vary slowly throughout development and learning), but can also be gated by other brain regions and control systems, e.g. through the influence of neurotransmitters, and thus vary much more dynamically. For instance, the feedforward drive β could be more active to capture sensory information immediately after each eye movement, and decrease over time until the next eye movement [21]; similarly, feedback error correction λ could dominate over the feedforward error correction α for one given second (e.g. because top-down attention drives expectation signals) and decrease in the next second (e.g. because unexpected sensory inputs have been detected) [31]. In this dynamic context, it is fundamental to be able to characterize the dependence of the system's behavior on the exact hyperparameter values. Fortunately, our framework reveals that when the hyperparameters vary, the stability of the system, and its ability to propagate signals and maintain activity, change in predictable

ways. Some hyperparameter combinations would not support signal propagation at all; others would render the system unstable, e.g. because of runaway excitation. Under the (reasonable) assumption that the brain behaves as a predictive coding system, our equations inform us about the plausible parameter regimes for the brain.

Using our time-continuous model, we found that predictive coding dynamics associated with inter-areal communication delays result in oscillatory activity. This finding resonates with both experimental observations and neuroscientific theories [1, 4].

Bastos and colleagues [4, 5] suggested that feedforward error correction could be accompanied by gamma-band oscillations; this suggestion was verified in our model, with synchronized gamma rhythms appearing when the corresponding hyperparameter $\widetilde{\alpha}$ was strong enough (and with a frequency that monotonically increased from 30 to 60Hz when the value of $1/\widetilde{\alpha}$ decreased from 10ms to 5ms). However, considering that the communication delay τ between two adjacent brain regions is a fixed property of the system (a reasonable first approximation), our analysis shows that this oscillatory mode will only happen for a narrow range and a very precise combination of hyperparameter values $\widetilde{\alpha}$ and $\widetilde{\lambda}$ (see Figure 27). This could explain why gamma-band oscillations are not always found during electrophysiological recording experiments [28].

By relaxing the phase delay between brain areas, our equations also revealed the potential emergence of oscillatory travelling waves across brain regions, similar to those observed in human EEG experiments [1–3, 24]. Again, for a fixed communication delay τ, these waves may only happen for specific values and combinations of the hyperparameters $\widetilde{\alpha}$ and $\widetilde{\lambda}$. In certain regimes (see e.g. Figure 31 with $1/\widetilde{\alpha} = 1/\widetilde{\lambda} = 12ms$), two waves might coexist at the same frequency, but going in opposite directions. This matches experimental reports of co-occurring feedforward and feedback waves in the brain [2, 3]. Upon increasing the feedback strength $\widetilde{\lambda}$, we saw that an initial alpha-band (7-15Hz) feed-forward wave could accelerate (towards the beta-band, 15-30Hz) and eventually reverse its direction, producing a feedback wave. Similar reversal phenomena have also been reported for oscillatory waves in the human brain [2, 3, 24].

6.3 Limitations and future extensions

"All models are wrong, but some are useful" [7]. Our model, like all mathematical models, is based on simplifications, approximations and assumptions, and can only be valid under those assumptions. Some (if not all) of these assumptions are questionable, and future work will need to determine the robustness of the model, or its potential modifications, when relaxing these assumptions.

Even tough we assumed that the brain follows the general principles of predictive coding [27], our system's hyperparameters can in fact be modulated to accommodate many variants of this framework [9, 19, 31, 32]. One other important assumption that we made was to simplify the connectivity matrices between neuronal layers—which determines the selectivity of each neuron, and thus the functionality of the entire system. Even when we moved past the "identity" assumption, the connection matrices that we adopted were constrained to be symmetric, and most importantly, were assumed to be similar from one layer to the next. This made our equations tractable, but it constitutes a clear restriction, and a departure from biological plausibility that will need to be addressed in future extensions. Another important limitation that we wish to relax in future works is the fact that we have considered a linear model although real biological networks or deep neural networks are intrinsically nonlinear. Going beyond the linear analysis that we have presented here would need the development of new theoretical techniques which constitutes

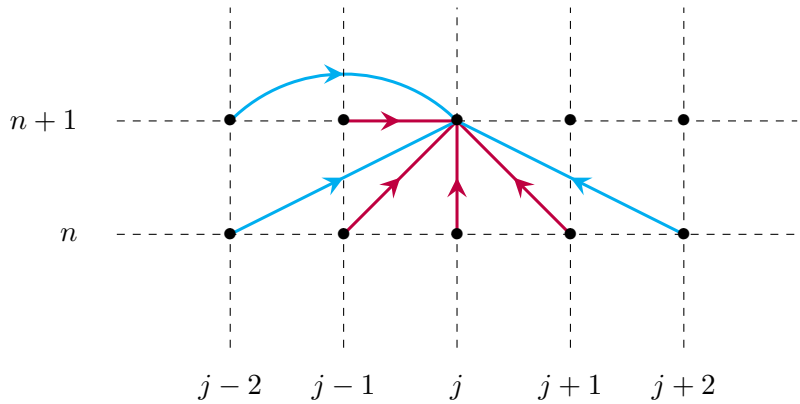


Figure 32: *Illustration of the network structure of model* (6.1) *where blue arrows indicate the new long-range interactions coming from layer* $j \pm 2$.

a major open problem to be addressed in forthcoming works.

Aside from exploring richer patterns of connectivity between adjacent layers, another natural extension of the model could be to incorporate long-range interactions, beyond the immediately adjacent layers. For instance, one could explore a simple second-order layer model (illustrated in Figure 32), whose scalar version reads:

$$e_j^{n+1} - \beta_1 e_{j-1}^{n+1} - \beta_2 e_{j-2}^{n+1} = \alpha_1 e_{j-1}^n + \alpha_2 e_{j-2}^n + (1 - \beta^* - \lambda^* - \alpha^*) e_j^n + \lambda_1 e_{j+1}^n + \lambda_2 e_{j+2}^n, \quad j \in \mathbb{Z}, \tag{6.1}$$

where we have set $\beta^* := \beta_1 + \beta_2$, $\alpha^* := \alpha_1 + \alpha_2$ and $\lambda^* := \lambda_1 + \lambda_2$. Once again the fate of such a system (6.1) would be dictated by the amplification factor function

$$\rho(\theta) = \frac{\alpha_1 e^{-\mathbf{i}\theta} + \alpha_2 e^{-2\mathbf{i}\theta} + 1 - \beta^* - \lambda^* - \alpha^* + \lambda_1 e^{\mathbf{i}\theta} + \lambda_2 e^{2\mathbf{i}\theta}}{1 - \beta_1 e^{-\mathbf{i}\theta} - \beta_2 e^{-2\mathbf{i}\theta}}, \quad \theta \in [-\pi, \pi].$$

This as well as higher-order interaction models, possibly including "hub" regions like the thalamus that would be mutually interconnected with all layers in the hierarchy [20], are promising directions for follow-up studies.

6.4 Conclusion

The mathematical framework proposed here, guided by both computational considerations and neuroscientific inspiration, can be of use to both fields. In machine learning, the framework may serve to provide guarantees about the stability of a predictive coding system given its chosen hyperparameters, or to choose a valid range for these hyperparameters. For neuroscientists, our equations can be used directly to understand biological vision and to make predictions about biological behavior in various situations compatible with predictive coding. But this general mathematical framework (a number of hierarchically connected layers with source terms, boundary conditions, feedforward and feedback connectivity matrices, analyzed via its amplification factor function) may also be adapted to fit other models of biological perception and cognition beyond predictive coding. We hope that the various derivations made in the present work can serve as a template for future applications in this direction. And more generally, that this study may be helpful to the larger computational neuroscience community.

Acknowledgements

G.F. acknowledges support from the ANR via the projects: Indyana under grant agreement ANR-21-CE40-0008, ChaMaNe under grant agreement ANR-19-CE40-0024 and an ANITI (Artificial and Natural Intelligence Toulouse Institute) Research Chair. R.V. is supported by ANR OSCI-DEEP (ANR-19-NEUC-004) and an ANITI (Artificial and Natural Intelligence Toulouse Institute) Research Chair (ANR-19-PI3A-0004).

References

[1] A. Alamia and R. VanRullen. Alpha oscillations and traveling waves: Signatures of predictive coding? *PLoS Biology*, 17(10):e3000487, 2019.

[2] A. Alamia, C. Timmermann, D.J. Nutt, R. VanRullen, and R.L. and Carhart-Harris. DMT alters cortical travelling waves. *Elife* (2020), 9, e59784.

[3] A. Alamia, L. Terral, M.R. d'Ambra and R. VanRullen. Distinct roles of forward and backward alpha-band waves in spatial visual attention. *Elife*, (2023), 12, e85035.

[4] A. Bastos, W. Usrey, R. Adams, G.R. Mangun, P. Fries and K. Friston. Canonical microcircuits for predictive coding *Neuron*, (2012), vol 76 (4), pp 695-711

[5] A. Bastos, J. Vezoli, C.A. Bosman, J.M. Schoffelen, R. Oostenveld, J.R. Dowdall and P. Fries. Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* (2015), 85(2), pp 390-401

[6] C. Besse, G. Faye, J.-M. Roquejoffre and M. Zhang. The logarithmic Bramson correction for Fisher-KPP equations on the lattice $\mathbb{Z}$. *arXiv:2206.04358*, 2022.

[7] G. E. P Box. Robustness in the strategy of scientific model building, in Launer, R. L.; Wilkinson, G. N. (eds.), *Robustness in Statistics*, (1979), Academic Press, pp. 201?236

[8] J. Bullier. Feedback connections and conscious vision. *Trends in cognitive sciences* (2001), vol 5 (9), pp 369-370.

[9] B. Choksi, M. Mozafari, C. Biggs O'May, B. Ador, A. Alamia, and R. VanRullen. Predify: Augmenting deep neural networks with brain-inspired predictive coding dynamics. *Advances in Neural Information Processing Systems*, (2021), 34, 14069-14083.

[10] L. Coeuret. Local limit theorem for complex valued sequences. *arXiv*, 2201.01514, 2022.

[11] L. Coeuret. Tamed instability for finite difference approximations of hyperbolic equations with boundary conditions. *arXiv:2304.02612*, 2023.

[12] J-F. Coulombel and G. Faye. Generalized gaussian bounds for discrete convolution powers. *Rev. Mat. Iberoam.* 38 (2022), no. 5, pp. 1553–1604.

[13] J-F. Coulombel and G. Faye. Sharp stability for finite difference approximations of hyperbolic equations with boundary conditions. IMA Journal of Numerical Analysis, (2021) , pp. 1-34.

[14] P. Diaconis and L. Saloff-Coste. Convolution powers of complex functions on $\mathbb{Z}$. *Math. Nachr.*, 287(10):1106–1130, 2014.

[15] D.J. Felleman D.C. and Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, (1991), vol 1 (1), pp 1-47.

[16] M. Goldberg and E. Tadmor. Convenient stability criteria for difference approximations of hyperbolic initial-boundary value problems. *Mathematics of computation* 44.170 (1985): 361-377.

[17] M. Goldberg and E. Tadmor. Convenient stability criteria for difference approximations of hyperbolic initial-boundary value problems. II. *Mathematics of computation* 48.178 (1987): 503-520.

[18] K. He, X. Zhang, S. Ren and J. Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770-778

[19] D.J. Heeger. Theory of cortical function. *Proceedings of the National Academy of Sciences* (2017), vol 114 (8), pp 1773-1782.

[20] K. Hwang, M.A. Bertolero, W.B. Liu and M. D'Esposito. The human thalamus is an integrative hub for functional brain networks. *Journal of Neuroscience*, (2017), vol 37 (23), pp 5594-5607.

[21] J. Knoell, P. Binda, M.C. Morrone and F. Bremmer. Spatiotemporal profile of peri-saccadic contrast sensitivity. *Journal of vision* (2011), vol 11 (14), pp 15-15.

[22] B. Millidge, A. Seth, and C.L. Buckley. Predictive coding: a theoretical and experimental review. *arXiv* (2021) preprint arXiv:2107.12979.

[23] Z. Pang. Predictive coding in the brain and deep neural networks. *Dissertation thesis*, Université Paul Sabatier-Toulouse III, 2022.

[24] Z. Pang, A. Alamia and R. VanRullen. Turning the stimulus on and off changes the direction of α traveling waves. *Eneuro*, (2020), vol 7 (6).

[25] Z. Pang, C.B. O'May, B. Choksi, and R. VanRullen. Predictive coding feedback results in perceived illusory contours in a recurrent neural network. *Neural Networks*, (2021) 144, 164-175.

[26] E. Randles and L. Saloff-Coste. On the convolution powers of complex functions on $\mathbb{Z}$. *J. Fourier Anal. Appl.*, 21(4):754–798, 2015.

[27] R.P. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* (1999) vol 2 (1), pp 79-87.

[28] S. Ray and J.H. Maunsell. Do gamma oscillations play a role in cerebral cortex? *Trends Cogn Sci.* (2015) vol 19 (2), pp 78-85

[29] F. Riesz and B. Sz.-Nagy. Functional Analysis. Ungarn,New York, 1955.

[30] V. Thomée. Stability of difference schemes in the maximum-norm. *J. Differential Equations* 1 (1965), 273–292.

[31] A. Tschantz, B. Millidge, A.K. Seth and C.L. Buckley. Hybrid Predictive Coding: Inferring, Fast and Slow. *arXiv* (2022). https://doi.org/https://arxiv.org/abs/2204.02169v2

[32] H. Wen, K. Han, J. Shi, Y. Zhang, E. Culurciello and Z. Liu. Deep predictive coding network for object recognition. *International Conference on Machine Learning*, (2018), pp. 5266-5275.