Continuous Predictive Coding

immediate

July 20, 2022

Abstract

1 Encoder/Decoder without feedforward error correction

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

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

with $\mathcal{E}_0^n = \mathcal{E}_0$ for each $n \geq 0$ acting as a source term. At $j = J$, one gets

$$\mathcal{E}_J^{n+1} = \beta \mathcal{W}_{J-1,J}^f \mathcal{E}_{J-1}^{n+1} + (1-\beta)\mathcal{E}_J^n.$$

Here, we assume that each $\mathcal{E}_j^n \in \mathbb{R}^d$ with $d \geq 1$. As a consequence for each j the matrices $\mathcal{W}_{j-1,j}^f$ and $\mathcal{W}_{j+1,j}^b$ are in $\mathscr{M}_d(\mathbb{R})$. We can equivalently rewrite the above recurrence equation in a condensed form as follows. First, we denote $\mathcal{E}^n := (\mathcal{E}_1, \cdots, \mathcal{E}_J)^{\mathbf{t}} \in \mathbb{R}^{dJ}$ and introduce two matrices $\mathbb{W}^f \in \mathscr{M}_{dJ}(\mathbb{R})$ and $\mathbb{W}^b \in \mathscr{M}_{dJ}(\mathbb{R})$ given by

$$\mathbb{W}^f = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \mathcal{W}_{1,2}^f & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \mathcal{W}_{2,3}^f & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & 0 & \mathcal{W}_{J-1,J}^f & 0 \end{pmatrix}, \quad \mathbb{W}^b = \begin{pmatrix} 0 & \mathcal{W}_{2,1}^b & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \mathcal{W}_{3,2}^b & 0 & \cdots & \vdots \\ \vdots & \ddots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & \mathcal{W}_{J,J-1}^b \\ 0 & \cdots & \cdots & \cdots & 0 & 0 \end{pmatrix}.$$

We thus obtain

$$\mathcal{E}^{n+1} = \beta \mathbb{W}^f \mathcal{E}^{n+1} + \lambda \mathbb{W}^b \mathcal{E}^n + \mathbb{D}\mathcal{E}^n + \mathcal{S},$$

where we also set

$$\mathbb{D} = \begin{pmatrix} (1-\beta-\lambda)I_d & 0 & 0 & \cdots & \cdots & 0 \\ 0 & (1-\beta-\lambda)I_d & 0 & 0 & \cdots & \vdots \\ \vdots & \ddots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & (1-\beta-\lambda)I_d & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & (1-\beta)I_d \end{pmatrix} \in \mathscr{M}_{dJ}(\mathbb{R}),$$

and $\mathcal{S} = \left(\beta \mathcal{W}^f_{0,1}\mathcal{E}_0, 0, \cdots, 0\right)^{\mathbf{t}} \in \mathbb{R}^{dJ}$. As a consequence, we can rewrite the recurrence equation as

$$\mathcal{E}^{n+1} = \underbrace{\left(I_{dJ} - \beta \mathbb{W}^f\right)^{-1}\left(\lambda \mathbb{W}^b + \mathbb{D}\right)}_{:=\mathbb{M}} \mathcal{E}^n + \left(I_{dJ} - \beta \mathbb{W}^f\right)^{-1} \mathcal{S}$$

whose solution is

$$\mathcal{E}^n = \mathbb{M}^n \mathcal{E}^0 + \left(\sum_{k=0}^{n-1} \mathbb{M}^k\right)\left(I_{dJ} - \beta \mathbb{W}^f\right)^{-1} \mathcal{S}, \quad n \geq 1.$$

So the important quantities are the eigenvalues $(\mu_\ell)_{1\leq \ell \leq dJ}$ of the matrix $\mathbb{M}$ and whether they are larger or smaller in modulus than one.

It is possible to get an expression for $\left(I_{dJ} - \beta \mathbb{W}^f\right)^{-1}$, indeed one has

$$\left(I_{dJ} - \beta \mathbb{W}^f\right)^{-1} = \begin{pmatrix} I_d & \cdots & \cdots & \cdots & \cdots & 0 \\ \beta \mathcal{W}^f_{1,2} & I_d & \cdots & \cdots & \cdots & 0 \\ \beta^2 \mathcal{W}^f_{2,3}\mathcal{W}^f_{1,2} & \beta \mathcal{W}^f_{2,3} & I_d & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & I_d & 0 \\ \beta^{d-1}\mathcal{W}^f_{J-1,J}\mathcal{W}^f_{J-2,J-1}\cdots \mathcal{W}^f_{1,2} & \cdots & \cdots & \beta^2 \mathcal{W}^f_{J-1,J}\mathcal{W}^f_{J-2,J-1} & \beta \mathcal{W}^f_{J-1,J} & I_d \end{pmatrix}.$$

One also has that $\det(\mathbb{M}) = (1-\beta-\lambda)^{d-1}(1-\beta) \neq 0$ if $\beta < 1$ and $\beta + \lambda < 1$.

2 Possible physical interpretation via transport equations

The idea is to interpret the recurrence equation for encoder/decoder as a *transport* equation. To start with, we assume at first that $\mathcal{W}^f_{j-1,j}$ and $\mathcal{W}^b_{j+1,j}$ are the identity such that we have

$$\mathcal{E}^{n+1}_j - \mathcal{E}^n_j = \beta\left(\mathcal{E}^{n+1}_{j-1} - \mathcal{E}^n_j\right) + \lambda\left(\mathcal{E}^n_{j+1} - \mathcal{E}^n_j\right).$$

Let introduce $\Delta t > 0$ and $\Delta x > 0$ and let rewrite the above equation as

$$(1-\beta)\frac{\mathcal{E}^{n+1}_j - \mathcal{E}^n_j}{\Delta t} = \beta \frac{\Delta x}{\Delta t}\frac{\mathcal{E}^{n+1}_{j-1} - \mathcal{E}^{n+1}_j}{\Delta x} + \lambda \frac{\Delta x}{\Delta t}\frac{\mathcal{E}^n_{j+1} - \mathcal{E}^n_j}{\Delta x}.$$

And if now $\mathcal{E}^n_j$ represents an approximation some continuous function $\mathcal{E}(t,x)$ evaluated at $t_n = n\Delta t$ and $x_j = j\Delta x$, that is $\mathcal{E}^n_j \sim \mathcal{E}(t_n, x_j)$, then in the limit $\Delta t \to 0$, $\Delta x \to 0$ with $\frac{\Delta x}{\Delta t} = \nu > 0$ fixed, one gets (when $\beta \neq \lambda$) the partial differential equation

$$\partial_t \mathcal{E}(t,x) + \frac{\nu(\beta-\lambda)}{1-\beta}\partial_x \mathcal{E}(t,x) = 0, \quad t > 0, \quad x > 0,$$

with boundary condition $\mathcal{E}(t, x=0) = \mathcal{E}_0$ and initial condition $\mathcal{E}(t=0,x) = \mathcal{E}^{in}_0(x)$ satisfying the compatibility condition $\mathcal{E}_0 = \mathcal{E}^{in}_0(0)$. A first information that one gets is that $\frac{\nu(\beta-\lambda)}{1-\beta} > 0$ in order for the information to be transported to the right (for example $0 < \lambda < \beta < 1$ ensures the positivity). Solutions are given by

$$\mathcal{E}(t,x) = \begin{cases} \mathcal{E}_0, & x \leq \frac{\nu(\beta-\lambda)}{1-\beta}t, \\ \mathcal{E}^{in}_0\left(x - \frac{\nu(\beta-\lambda)}{1-\beta}t\right), & x > \frac{\nu(\beta-\lambda)}{1-\beta\nu}t. \end{cases}$$

We now work with $\mathcal{W}^f_{j-1,j}$ and $\mathcal{W}^b_{j+1,j}$ and suppose that they can be read out from some continuous matrix-valued functions $\mathcal{W}^f(x)$ and $\mathcal{W}^b(x)$ at respectively x_{j-1} and x_{j+1} that is: $\mathcal{W}^f_{j-1,j} = \mathcal{W}^f(x_{j-1})$ and $\mathcal{W}^b_{j+1,j} = \mathcal{W}^b(x_{j+1})$. With this assumption, the recurrence equation now reads

$$\mathcal{E}_j^{n+1} = \beta \mathcal{W}^f(x_{j-1})\mathcal{E}_{j-1}^{n+1} + \lambda \mathcal{W}^b(x_{j+1})\mathcal{E}_{j+1}^{n} + (1-\beta-\lambda)\mathcal{E}_j^n.$$

And rearranging terms, we get exactly

$$\begin{aligned}\left(I-\beta\mathcal{W}^f(x_j)\right)\left(\mathcal{E}_j^{n+1}-\mathcal{E}_j^n\right) &= \beta\left(\mathcal{W}^f(x_{j-1})\mathcal{E}_{j-1}^{n+1}-\mathcal{W}^f(x_j)\mathcal{E}_j^{n+1}\right)+\beta\left(\mathcal{W}^f(x_j)-I\right)\mathcal{E}_j^n\\ &\quad+\lambda\left(\mathcal{W}^b(x_{j+1})\mathcal{E}_{j+1}^{n}-\mathcal{W}^b(x_j)\mathcal{E}_j^{n}\right)+\lambda\left(\mathcal{W}^b(x_j)-I\right)\mathcal{E}_j^n,\end{aligned}$$

which is equivalent to by dividing by $\Delta t > 0$

$$\begin{aligned}\left(I-\beta\mathcal{W}^f(x_j)\right)\frac{\left(\mathcal{E}_j^{n+1}-\mathcal{E}_j^n\right)}{\Delta t} &= \beta\frac{\Delta x}{\Delta t}\frac{\left(\mathcal{W}^f(x_{j-1})\mathcal{E}_{j-1}^{n+1}-\mathcal{W}^f(x_j)\mathcal{E}_j^{n+1}\right)}{\Delta x}+\beta\frac{\left(\mathcal{W}^f(x_j)-I\right)}{\Delta t}\mathcal{E}_j^n\\ &\quad+\lambda\frac{\Delta x}{\Delta t}\frac{\left(\mathcal{W}^b(x_{j+1})\mathcal{E}_{j+1}^{n}-\mathcal{W}^b(x_j)\mathcal{E}_j^{n}\right)}{\Delta x}+\lambda\frac{\left(\mathcal{W}^b(x_j)-I\right)}{\Delta t}\mathcal{E}_j^n.\end{aligned}$$

In the limit $\Delta t \to 0$, $\Delta x \to 0$ with $\frac{\Delta x}{\Delta t} = \nu > 0$, we obtain at first orders:

$$\left(I-\beta\mathcal{W}^f(x)\right)\partial_t\mathcal{E}(t,x)+\nu\partial_x\left(\left[\beta\mathcal{W}^f(x)-\lambda\mathcal{W}^b(x)\right]\mathcal{E}(t,x)\right)=\left(\beta\frac{\left(\mathcal{W}^f(x)-I\right)}{\Delta t}+\lambda\frac{\left(\mathcal{W}^b(x)-I\right)}{\Delta t}\right)\mathcal{E}(t,x).$$

We remark that:

- the righthand side is of order Δt^{-1} and this somehow indicates that one should have $\mathcal{W}^f(x) \sim I$ and $\mathcal{W}^b(x) \sim I$ in the fully continuous approximation;
- if one assumes that $\mathcal{W}^f(x) = I + \Delta t\widetilde{\mathcal{W}}^f(x)$ and $\mathcal{W}^b(x) = I + \Delta t\widetilde{\mathcal{W}}^b(x)$, then the equation reduces at leading order in Δt to

  $$(1-\beta)\partial_t\mathcal{E}(t,x)+\nu(\beta-\lambda)\partial_x\mathcal{E}(t,x)=\left(\beta\widetilde{\mathcal{W}}^f(x)+\lambda\widetilde{\mathcal{W}}^b(x)\right)\mathcal{E}(t,x)$$

  which close to the previous case but now with a source term given by $\beta\widetilde{\mathcal{W}}^f(x)+\lambda\widetilde{\mathcal{W}}^b(x)$. Solutions can be explicitly constructed as

  $$\mathcal{E}(t,x)=\begin{cases}\exp\left(\int_{t-\frac{1-\beta}{\nu(\beta-\lambda)}x}^{t}\widetilde{\mathcal{W}}^s\left(x-\frac{\nu(\beta-\lambda)}{1-\beta}(t-s)\right)\mathrm{d}s\right)\mathcal{E}_0, & x\leq\frac{\nu(\beta-\lambda)}{1-\beta}t,\\ \exp\left(\int_0^t\widetilde{\mathcal{W}}^s\left(x-\frac{\nu(\beta-\lambda)}{1-\beta}(t-s)\right)\mathrm{d}s\right)\mathcal{E}_0^{in}\left(x-\frac{\nu(\beta-\lambda)}{1-\beta}t\right), & x>\frac{\nu(\beta-\lambda)}{1-\beta\nu}t,\end{cases}$$

  where we have defined

  $$\widetilde{\mathcal{W}}^s(x) := \beta\widetilde{\mathcal{W}}^f(x)+\lambda\widetilde{\mathcal{W}}^b(x).$$

  Some *energy* estimates of the solutions can be derived without using the exact form of the solutions. Indeed, if $\langle\cdot,\cdot\rangle$ denotes the scalar product on $\mathbb{R}^{dJ}$, one gets

  $$\frac{(1-\beta)}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_0^{+\infty}\langle\mathcal{E}(t,x),\mathcal{E}(t,x)\rangle\mathrm{d}x\right)=\frac{\nu(\beta-\lambda)}{2}\langle\mathcal{E}_0,\mathcal{E}_0\rangle+\int_0^{+\infty}\left\langle\left(\beta\widetilde{\mathcal{W}}^f(x)+\lambda\widetilde{\mathcal{W}}^b(x)\right)\mathcal{E}(t,x),\mathcal{E}(t,x)\right\rangle\mathrm{d}x.$$

One can conclude if for example one assumes that there exists some constant $\omega > 0$ such that

$$\left\langle \left(\beta\widetilde{\mathcal{W}}^f(x) + \lambda\widetilde{\mathcal{W}}^b(x)\right)\mathcal{E}(t,x), \mathcal{E}(t,x)\right\rangle < -\omega\langle\mathcal{E}(t,x), \mathcal{E}(t,x)\rangle,$$

then

$$\int_0^{+\infty}\langle\mathcal{E}(t,x), \mathcal{E}(t,x)\rangle \mathrm{d}x < e^{-\frac{2\omega}{1-\beta}t}\langle\mathcal{E}_0^{in}(x), \mathcal{E}_0^{in}(x)\rangle + \frac{\nu(\beta-\lambda)}{2\omega}\langle\mathcal{E}_0, \mathcal{E}_0\rangle,$$

and the *energy* of the solution is preserved in time and remains bounded.

Key point: The *space* variable x here represents a *continuous* depth of a neural network.

What happens when $\beta = \lambda$? This time, we take the limit limit $\Delta t \to 0$, $\Delta x \to 0$ with $\frac{\Delta x^2}{\Delta t} = \delta > 0$, and in the homogeneous case $\mathcal{W}^f = \mathcal{W}^b = I$, one gets

$$\partial_t\mathcal{E}(t,x) = \frac{\beta\delta}{1-\beta}\partial_x^2\mathcal{E}(t,x), \quad t > 0, \quad x > 0,$$

and we get a heat equation instead of a transport equation.

3 Incorporating a feedforward error correction

First, we introduce the reconstruction error $\mathcal{R}_{j-1}^n$ at layer $j-1$ defined as the mean square error between the representation $\mathcal{E}_{j-1}^n$ and the predicted reconstruction $\mathcal{W}_{j,j-1}^b\mathcal{E}_j^n$, that is

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

Taking the gradient with respect to $\mathcal{E}_j^n$, one gets

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

Introducing such a term into the above recurrence equation gives

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

Once again, if we denote $\mathcal{E}^n := (\mathcal{E}_1, \cdots, \mathcal{E}_J)^{\mathbf{t}} \in \mathbb{R}^{dJ}$ the augmented vector, we obtain

$$\mathcal{E}^{n+1} = \beta\mathbb{W}^f\mathcal{E}^{n+1} + \lambda\mathbb{W}^b\mathcal{E}^n + \mathbb{D}\mathcal{E}^n + \frac{\alpha}{d}\mathbb{E}\mathcal{E}^n + \widetilde{\mathcal{S}},$$

where

$$\mathbb{E} = \begin{pmatrix} -(\mathcal{W}_{1,0}^b)^{\mathbf{t}}\mathcal{W}_{1,0}^b & \cdots & \cdots & \cdots & \cdots & 0 \\ (\mathcal{W}_{2,1}^b)^{\mathbf{t}} & -(\mathcal{W}_{2,1}^b)^{\mathbf{t}}\mathcal{W}_{2,1}^b & \cdots & \cdots & \cdots & 0 \\ 0 & (\mathcal{W}_{3,2}^b)^{\mathbf{t}} & \ddots & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & (\mathcal{W}_{J,J-1}^b)^{\mathbf{t}} & -(\mathcal{W}_{J,J-1}^b)^{\mathbf{t}}\mathcal{W}_{J,J-1}^b \end{pmatrix},$$

and $\widetilde{\mathcal{S}} = \left(\beta\mathcal{W}_{0,1}^f\mathcal{E}_0 + \frac{\alpha}{d}(\mathcal{W}_{1,0}^b)^{\mathbf{t}}\mathcal{E}_0, 0, \cdots, 0\right)^{\mathbf{t}} \in \mathbb{R}^{dJ}$. The solution is now given by

$$\mathcal{E}^n = \left[\left(I_{dJ}-\beta\mathbb{W}^f\right)^{-1}\left(\lambda\mathbb{W}^b+\mathbb{D}+\frac{\alpha}{d}\mathbb{E}\right)\right]^n\mathcal{E}^0+\left(\sum_{k=0}^{n-1}\left[\left(I_{dJ}-\beta\mathbb{W}^f\right)^{-1}\left(\lambda\mathbb{W}^b+\mathbb{D}+\frac{\alpha}{d}\mathbb{E}\right)\right]^k\right)\left(I_{dJ}-\beta\mathbb{W}^f\right)^{-1}\widetilde{\mathcal{S}},$$

for each $n \geq 1$. The matrix of interest is now

$$\mathbb{A} := \left(I_{dJ}-\beta\mathbb{W}^f\right)^{-1}\left(\lambda\mathbb{W}^b+\mathbb{D}+\frac{\alpha}{d}\mathbb{E}\right),$$

whose eigenvalues will determine the asymptotic behavior of $\mathcal{E}^n$ as n gets large.

Of course, we can repeat the arguments done previously to identify a continuous limiting model. Assuming first that $\mathcal{W}^f_{j-1,j}$ and $\mathcal{W}^b_{j+1,j}$ are the identity, we get

$$\mathcal{E}_j^{n+1}-\mathcal{E}_j^n = \beta\left(\mathcal{E}_{j-1}^{n+1}-\mathcal{E}_j^n\right)+\lambda\left(\mathcal{E}_{j+1}^n-\mathcal{E}_j^n\right)+\frac{\alpha}{d}\left(\mathcal{E}_{j-1}^n-\mathcal{E}_j^n\right),$$

which yields to the PDE

$$\partial_t\mathcal{E}(t,x)+\frac{\nu(\beta+\frac{\alpha}{d}-\lambda)}{1-\beta}\partial_x\mathcal{E}(t,x)=0, \quad t>0, \quad x>0.$$

Once again, a good regime of parameters is $\beta < 1$ and $\lambda < \beta + \frac{\alpha}{d}$. Relaxing the assumption that $\mathcal{W}^f_{j-1,j}$ and $\mathcal{W}^b_{j+1,j}$ are the identity, we get

$$\begin{aligned}\left(I-\beta\mathcal{W}^f(x)\right)\partial_t\mathcal{E}(t,x)+\nu\partial_x\left(\left[\beta\mathcal{W}^f(x)-\lambda\mathcal{W}^b(x)\right]\mathcal{E}(t,x)\right)+\nu\frac{\alpha}{d}\mathcal{W}^b(x)^{\mathbf{t}}\partial_x\left(\mathcal{W}^b(x)\mathcal{E}(t,x)\right)\\ =\left(\beta\frac{(\mathcal{W}^f(x)-I)}{\Delta t}+\lambda\frac{(\mathcal{W}^b(x)-I)}{\Delta t}-\frac{\alpha}{d}\mathcal{W}^b(x)^{\mathbf{t}}\frac{(\mathcal{W}^b(x)-I)}{\Delta t}\right)\mathcal{E}(t,x).\end{aligned}$$

If one assumes that $\mathcal{W}^f(x) = I+\Delta t\widetilde{\mathcal{W}}^f(x)$ and $\mathcal{W}^b(x) = I+\Delta t\widetilde{\mathcal{W}}^b(x)$, then the equation reduces at leading order in Δt to

$$(1-\beta)\partial_t\mathcal{E}(t,x)+\nu\left(\beta+\frac{\alpha}{d}-\lambda\right)\partial_x\mathcal{E}(t,x)=\left(\beta\widetilde{\mathcal{W}}^f(x)+\left(\lambda-\frac{\alpha}{d}\right)\widetilde{\mathcal{W}}^b(x)\right)\mathcal{E}(t,x).$$

4 One dimensional case

Here, we assume that $d = 1$ together with $\mathcal{W}^f_{j-1,j} = \mathcal{W}^b_{j+1,j} = 1$, we obtain a recurrence relation of the form

$$\mathcal{E}_j^{n+1}-\mathcal{E}_j^n = \beta\left(\mathcal{E}_{j-1}^{n+1}-\mathcal{E}_j^n\right)+\lambda\left(\mathcal{E}_{j+1}^n-\mathcal{E}_j^n\right)+\alpha\left(\mathcal{E}_{j-1}^n-\mathcal{E}_j^n\right)$$

that we prefer to write

$$\mathcal{E}_j^{n+1}-\beta\mathcal{E}_{j-1}^{n+1} = \alpha\mathcal{E}_{j-1}^n+(1-\beta-\lambda-\alpha)\mathcal{E}_j^n+\lambda\mathcal{E}_{j+1}^n.$$

To get a better understanding of the problem, we will first consider $j \in \mathbb{Z}$. Postulating an Ansatz of the form $\mathcal{E}_j^n = \rho^n e^{\mathbf{i}\theta j}$ for $\rho \in \mathbb{C}$ and $\theta \in [-\pi,\pi]$, we find a relation for ρ and θ given by

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

where we assumed that $0 \leq \beta < 1$ for the denominator to be well-defined. We first remark that we always have

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

Next we will check under which condition $|\rho(\theta)| \leq 1$ for all $\theta \in [-\pi, \pi]$ to guaranty stability of the recurrence equation. For that, 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 $\pm\theta$ 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

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

As a consequence, $|\rho(\theta)|^2 \leq 1$ if and only if $\lambda + \alpha \leq 1$.

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

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

and using the previous factorization we get

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

and 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 $\rho(\pm\pi) = -1$.