Appl. Comput. Math., V.xx, N.xx, 20xx, pp.xx-xx

ON THE STABILITY OF A NON-HYPERBOLIC NONLINEAR MAP WITH NON-BOUNDED SET OF NON-ISOLATED FIXED POINTS WITH APPLICATIONS TO MACHINE LEARNING

R. HANSEN[1], M. VERA[1,2], L. ESTIENNE[1,3], L. FERRER[3] AND P. PIANTANIDA[4]

Abstract. This paper deals with the convergence analysis of the SUCPA (Semi Unsupervised Calibration through Prior Adaptation) algorithm, defined from a first-order non-linear difference equations, first developed to correct the scores output by a supervised machine learning classifier. The convergence analysis is addressed as a dynamical system problem, by studying the local and global stability of the nonlinear map derived from the algorithm. This map, which is defined by a composition of exponential and rational functions, turns out to be non-hyperbolic with a non-bounded set of non-isolated fixed points. Hence, a non-standard method for solving the convergence analysis is used consisting of an ad-hoc geometrical approach. For a binary classification problem (two-dimensional map), we rigorously prove that the map is globally asymptotically stable. Numerical experiments on real-world application are performed to support the theoretical results by means of two different classification problems: Sentiment Polarity performed with a Large Language Model and Cat-Dog Image classification. For a greater number of classes, the numerical evidence shows the same behavior of the algorithm, and this is illustrated with a Natural Language Inference example. The experiment codes are publicly accessible online at the following repository: `https://github.com/LautaroEst/sucpa-convergence`.

Keywords: Discrete Dynamical Systems, Non-Hyperbolic Maps, Algorithm Convergence, Calibration

AMS Subject Classification: 37C75, 39A30, 68R01

1. Introduction

Difference equations have gained a lot of attention in the last decades due to their applicability in many branches of scientific knowledge. These equations model discrete natural and social phenomena and perform a fundamental role in many applications like population dynamics, control engineering, genetics, signal processing, health sciences and ecology [20]. In most cases, difference equations arise naturally as discretization of differential equations [8] or are defined by an iterative algorithm that describe a dynamic system. In the later case, the algorithm can be fully described by a map (i.e., a function) and the nature of this map defines the convergence properties of the algorithm. As consequence, studying the map and the difference equation associated to a dynamic system is usually considered a very important topic because they define the behaviour of the system's dynamic.

[1] Universidad de Buenos Aires, Facultad de Ingeniería, Buenos Aires, Argentina,

[2] CONICET, Centro de Simulación Computacional para Aplicaciones Tecnológicas (CSC), Buenos Aires, Argentina,

[3] CONICET, Instituto de investigación en Ciencias de la Computación (ICC), Buenos Aires, Argentina,

[4] International Laboratory on Learning Systems (ILLS) and Quebec AI Institute (MILA), CNRS and Centrale-Supélec - Université Paris-Saclay, Quebec, Canada,

e-mail: rhansen@fi.uba.ar, mvera@fi.uba.ar, lestienne@fi.uba.ar, lferrer@dc.uba.ar, pablo.piantanida@cnrs.fr

Manuscript received xx.

1.1. **Non-hyperbolic maps and fixed points.** Depending on the discipline in which they are applied, different methods were developed throughout the literature to demonstrate the convergence of iterative algorithms (e.g. [1, 22, 30]). In most cases, the map associated with the algorithm is hyperbolic and the solutions of its associated difference equation (or equivalently, the points to which the algorithm converges) form a finite set of points –and therefore isolated. This points are usually referred to as the *fixed points* of the system. Roughly speaking, a fixed point is called *isolated* if there exists a neighborhood of it that does not contain any other fixed points of the system, otherwise it is called non-isolated. Also, a fixed point of a map is called *hyperbolic*, if the Jacobian matrix of the map at this point has no eigenvalues of modulus one, otherwise, it is called non-hyperbolic [6, 9].

As it is known, the eigenvectors provide the contracting/expanding directions of the local dynamics according to whether their corresponding eigenvalues have modulus greater/lower than one. The behaviour of dynamical systems around a non-hyperbolic fixed point is much more subtle than for a hyperbolic one, because one cannot get a definite conclusion of the local stability from just the linearization of the map when it has eigenvalues in the unit circle. These are called *resonant* cases and the scenarios are very different depending on whether the eigenvalues are 1, -1 or a pair of complex conjugates, and on the amount of each of them [21]. Neither is there a general classification of them, so a case-by-case analysis is required [2]. The local stability near a non-hyperbolic fixed point is addressed by the theory of Central Manifold, which is an invariant set in a low dimensional space where the local dynamics can be reduced. This theory is relevant and mostly applied in the case of a system's parameter bifurcation. It consists of transforming the map into its normal form, first by setting the fixed point to the origin through a suitable change of coordinates, and then splitting up the algebraic expression of the new map into a linear and a nonlinear terms (the interested readers on this theory can see [4, 9, 21, 32]). Thus, this formulation is used mostly in theoretical problems, or successfully applied when the map under study posses a relative simple algebraic expression (polynomial, exponential, etc.) as in [25]; otherwise it is very cumbersome to be utilized.

In this work, we study the convergence of the recently proposed SUCPA algorithm [12] by means of the local and global stability analysis of the map derived from it. This K-dimensional nonlinear map ($K \geq 2$) is *non-hyperbolic* and its fixed points form an *unbounded* set of *non-isolated* points (indeed, they are a straight line). We prove the convergence for the two-classes case, and show different properties and conjectures in the general case.

The map associated with the SUCPA algorithm consists of equations that involve a composition of exponential and rational functions, and the non-hyperbolic fixed points have only one eigenvalue equal to one and the rest of them have modulus less than one. The particular study of –high order– rational difference equations like the ones present in this map has gained a lot of attention in the last time, because there is no an effective general method to deal with, so they currently are a class of challenging problems, and only specific type of them are treated and reported in the literature [10, 17, 29]. In addition, despite that the study of non-hyperbolic maps is addressed in several works, they do not usually bring together all the features presented here. In most cases they are about isolated fixed points [27], one-dimensional maps [5, 19, 31], flows [34], area-preserving maps [23] or a two dimensional map with double eigenvalue equal one [18], to name a few.

Another problem addressed in this work is the global stability of the map, which would account for the convergence of the SUCPA algorithm for any input. The standard tools for this purpose –as obtaining a suitable Lyapunov function– cannot be applied here either for the same reason that the fixed points are not isolated. This fact, together with the no so simple algebraic equations of the system, force us to prove the local and global stability by non-standard ways, defining auxiliary functions and solving it by an ad-hoc geometrical approach.

1.2. **SUCPA as a calibration algorithm.** The SUCPA algorithm was first developed in the context of supervised machine learning classifiers. A supervised machine learning classifier is a function $\hat{h}(\cdot)$ that usually maps an input $\mathbf{x}$ composed by real-world data (e.g., images, text or audio) to an output y that can be of the same type as the input or could be numerical. An example of these functions is an image classifier, which takes in an image in the form of a 3D tensor $\mathbf{x} \in \mathbb{R}^{C \times N \times M}$ and outputs the category y to which that image belongs. (For instance, $y = \{y_1, y_2\} = \{\text{cat}, \text{dog}\}$ in the case of binary classification of images of animals). In most cases, classifiers map the input to a category contained in an unordered finite set of discrete categories.

The goal of supervised learning is not to manually design the function $\hat{h}(\cdot)$, but to obtained it by exposing the system to a set $\mathcal{D} = \{(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(N)}, y^{(N)})\}$ of multiple input-output examples (called, the *training set*). More formally, the function $\hat{h}(\cdot)$ is obtained by minimizing a loss $\mathcal{L}_{\mathcal{D}}(h)$ function over the entire set of possible functions considered in the problem. Then, this function is used to predict an output $\hat{y} = \hat{h}(\mathbf{x})$ from a new input $\mathbf{x}$. Particularly in the case of probabilistic machine learning classifiers, the training set is used to obtain a so called *posterior probability distribution* $P(y|\mathbf{x})$ over the set $\mathcal{Y} = \{y_1, \ldots, y_K\}$ of possible categories, for each possible input $\mathbf{x}$. It is a well known result that once this distribution is known, it is trivial to map the input $\mathbf{x}$ to the label $\hat{y}$ by making Bayes decisions, which are those that minimize the Bayes risk [7], a standard criterion for measuring performance of a classifier. A well known instance of this risk (and the one used in this work) is the one obtained by assigning equal cost to each decision. In this case, it can be shown that the risk is minimized by choosing the class $\hat{y}$ that maximizes the posterior $P(y|\mathbf{x})$:

$$\hat{y} = \hat{h}(\mathbf{x}) = \underset{y \in \mathcal{Y}}{\arg\max}\, P(y|\mathbf{x}) \tag{1}$$

Concerning the quality of the estimation of the posterior $P(y|\mathbf{x})$, it has been studied in the last years that the posterior distribution produced by most modern classifiers is miscalibrated [15], which means that the Bayes decisions for that posteriors are not optimal for all cost functions [13][1]. This may happen for a few different reasons. A model that overfitted the training data tends to overestimate its certainty about the class, resulting in posteriors that take sub-optimally extreme values [15]. Further, when the data distribution of the training and the test data are different, the model may also be miscalibrated. In particular, when the training and test class probabilities (so-called *priors probability distributions*) are different, the posteriors will be sub-optimal since the priors are implicitly encoded in the posteriors [14].

Many approaches have been proposed in the literature to calibrate machine learning systems that are miscalibrated. One simple calibration method is based on logistic regression [3], taking as input features the log posteriors produced by the model (i.e. the logarithm of the estimated probability $P(k|\mathbf{x})$ of a class k, $k = 1, \ldots, K$, given an observed input $\mathbf{x}$) and applying an affine transformation of the following form:

$$\log \tilde{P}(k|\mathbf{x}) = \alpha_k \log P(k|\mathbf{x}) + \beta_k + \gamma(\boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{2}$$

where $\tilde{P}(k|\mathbf{x})$ is the calibrated posterior probability, $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1, \ldots, \alpha_K\end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix}\beta_1, \ldots, \beta_K\end{bmatrix}$ and $\gamma(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is determined so that $\sum_{k=1}^{K} \tilde{P}(k|\mathbf{x}) = 1$. Here, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are parameters of the affine transformation, and they are trained to minimize a Proper Scoring Rule (usually the Negative Log-Likelihood or *cross-entropy*) [13].

Different assumptions in (2) are usually made according to the specific problem that needs to be solved. For instance, temperature scaling [15], one of the most widely used calibration methods, corresponds to taking $\beta_k = 0$ for all k and $\alpha_k = \frac{1}{T}$ a single scalar. In recent work, we proposed the SUCPA algorithm to address the scenario of mismatched priors for text classification tasks solved with large language models [12]. This algorithm is derived for the case in which $\alpha_k = 1$ for all $k \in \{1, \ldots, K\}$, which is equivalent to assume that the source of miscalibration is

[1] A more formal definition of calibration is that a system is calibrated if the estimated probability of the predicted label matches with the true probability of that category. For more details, consult [13].

only due to a mismatch in class priors. As logistic regression, SUCPA is obtained by minimizing the cross-entropy loss on the training set $\mathcal{D}=\{(\mathbf{x}^{(1)},y^{(1)}),\dots,(\mathbf{x}^{(N)},y^{(N)})\}$:

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N} -\log \tilde{P}(y^{(i)}|\mathbf{x}^{(i)}) = -\frac{1}{N}\sum_{i=1}^{N}\left[\log P(y^{(i)}|\mathbf{x}^{(i)}) + \beta_{y^{(i)}} + \gamma_i(\mathbf{1},\boldsymbol{\beta})\right] \tag{3}$$

If we set the derivative of the cross-entropy to zero, we can derive the following expression for β_k:

$$\beta_k = \log\left(\frac{N_k}{N}\right) - \left(\frac{1}{N}\sum_{i=1}^{N}\frac{P(y_k|\mathbf{x}^{(i)})}{\sum_{j=1}^{K}P(y_j|\mathbf{x}^{(i)})e^{\beta_j}}\right) \tag{4}$$

where N_k is the number of samples with $y^{(i)}=k$. Mathematical details can be read in [12]. If the proportion $\frac{N_k}{N}$ is known as external knowledge from the nature of the task, then (4) can be used to iteratively estimate the value of β_k for every $k\in\{1,\dots,K\}$ using just the samples $\{\mathbf{x}^{(1)},\dots,\mathbf{x}^{(N)}\}$ (i.e., with no labels). It is important to highlight the need to have this prior information, which limits the spectrum of possible applications to tasks where this knowledge is precise.

The rest of the paper is organized as follows. In Section 2 we present the K-dimensional nonlinear map derived from the algorithm, and pose the main conjecture of the work. Section 3 summarizes some essential properties of the map. In Section 4 we prove the convergence for two classes ($K=2$). Numerical examples are shown in Section 5. Finally, in Section 6, some concluding remarks are discussed.

2. The SUCPA-map

In this section, we formally introduce the K-dimensional map, $\mathbf{f}$, derived from the SUCPA algorithm presented in [12], henceforth the *SUCPA-map*. We also recall some useful definitions and formulate the main conjecture of the work in terms of this map. The discrete time is denoted by $t\in\mathbb{N}$.

Definition 2.1. Let $\boldsymbol{\beta}^{[t]}=\left[\beta_1^{[t]},\dots,\beta_K^{[t]}\right]\in\mathbb{R}^K$ and $\boldsymbol{\beta}^{[0]}\in\mathbb{R}^K$ be an initial condition. Let also, $[N_1,\dots,N_K]\in\mathbb{N}^K$, with $\sum_{k=1}^K N_k=N$ and $\mathbf{P}\in\mathbb{R}^{N\times K}$ a matrix with coefficients $P_{i,k}>0$ and $\sum_{k=1}^K P_{i,k}=1$. We define the *SUCPA algorithm* as:

$$\beta_k^{[t+1]} = -\log\left(\frac{1}{N_k}\sum_{i=1}^{N}\frac{P_{i,k}}{\sum_{j=1}^{K}P_{i,j}e^{\beta_j^{[t]}}}\right),\quad 1\le k\le K. \tag{5}$$

The *SUCPA-map*, $\mathbf{f}=[f_1,\dots,f_K]:\mathbb{R}^K\to\mathbb{R}^K$, is the map defined by (5), as $\boldsymbol{\beta}^{[t+1]}=\mathbf{f}\big(\boldsymbol{\beta}^{[t]}\big)$. Then:

$$f_k(\boldsymbol{\beta}) = -\log\left(\frac{1}{N_k}\sum_{i=1}^{N}\frac{P_{i,k}}{\sum_{j=1}^{K}P_{i,j}e^{\beta_j}}\right),\quad 1\le k\le K. \tag{6}$$

As usual, $\mathbf{f}^{(t+1)}(\boldsymbol{\beta})=\mathbf{f}\big(\mathbf{f}^{(t)}(\boldsymbol{\beta})\big)$, $\forall\, t\in\mathbb{N}$, and $\mathbf{f}^{(0)}(\boldsymbol{\beta})=\boldsymbol{\beta}$. Note that Eq. (6) is a system of K first-order difference equations involving a composition of exponential and rational functions.

Definition 2.2. Let $\mathbf{f}$ be a map, then [6, 32]:

(1) $\boldsymbol{\beta}^*\in\mathbb{R}^K$ is a *fixed point of* $\mathbf{f}$ if it verifies $\mathbf{f}(\boldsymbol{\beta}^*)=\boldsymbol{\beta}^*$.
(2) The *ω-limit set* of $\boldsymbol{\beta}$ is the set:

$$\omega(\boldsymbol{\beta}) = \bigcap_{n=0}^{\infty}\overline{\{\mathbf{f}^{(t)}(\boldsymbol{\beta}),t\ge n\}} \tag{7}$$

where the over-bar accounts for the closure of a set.

The fixed points are those that kept invariant when undergoing iterations of the algorithm. The ω-limit is the set toward the algorithm, started in $\boldsymbol{\beta}$, converges with the successive iterations. Note that $\omega(\boldsymbol{\beta})$ is a limit set because it is the intersection of non-increasing sets. It is also a closed set (by definition), and it can be of different type –a single point, a periodic orbit, an infinite set, an unbounded set, an empty set, and it may have also complicated structures– depending on the $\mathbf{f}$ structure [32].

As said in the introduction, this work deals with the convergence of the SUCPA algorithm for any initial condition (i.c.), $\boldsymbol{\beta}^{[0]} \in \mathbb{R}^K$, which leads us to study the stability properties of the related SUCPA-map.

Conjecture 2.3. *For each i.c. $\boldsymbol{\beta}^{[0]} \in \mathbb{R}^K$, there exist a unique fixed point of the SUCPA-map, $\mathbf{f}$, $\boldsymbol{\beta}^*\in \mathbb{R}^K$ $(\boldsymbol{\beta}^*(\boldsymbol{\beta}^{[0]}))$, such that $\omega(\boldsymbol{\beta}^{[0]}) = \{\boldsymbol{\beta}^*\}$ (it is a single point set).*

This result is completely demonstrated in the case of two dimensions, i.e. $K=2$, corresponding to a binary classification problem for the original algorithm. The case $K>2$ of this conjecture is not yet proven, however many interesting properties can be shown, and this will be done in the next section.

3. General properties of the SUCPA-map

Within the literature of discrete dynamical systems, most problems deal with the stability of isolated fixed points, for which the standard tools to study both, local and global stability focus on the linearization of the system at those points, and/or obtaining the so-called Lyapunov functions [6,32]. Furthermore, the number of fixed points is usually finite –therefore they form a bounded set. The SUCPA-map defined in (6) actually presents a fairly simple behavior, in terms of its dynamics –for example, there is no presence of chaos despite being non-linear. However, as will be demonstrated, the fixed points of the system form a non-bounded set of non-isolated points. Indeed, the set of fixed points is a straight line. These two features together with the algebraic type of equations, make it necessary to study the stability of the system by alternative ways. We did this, for the case $K=2$, by defining properly auxiliary functions that allowed us to argue geometrically.

3.1. Fixed points analysis.

In terms of the SUCPA-map $\mathbf{f}$ of Eq. (6), and according to Def. 2.2(1), $\boldsymbol{\beta}^* = [\beta_1^*, \dots, \beta_K^*]$ is a fixed point of $\mathbf{f}$, if and only if, $\forall\, k=1, \cdots, K$:

$$e^{-\beta_k^*} = \frac{1}{N_k} \sum_{i=1}^{N} \frac{P_{i,k}}{\sum_{j=1}^{K} P_{i,j} e^{\beta_j^*}} \tag{8}$$

Equations (8) cannot be solved explicitly, so we must approach this point from another side.

The SUCPA-map has a particular behavior against constant vector sums. It is easy to see the following result.

Lemma 3.1. *Let $\boldsymbol{\lambda} = [\lambda, \dots, \lambda] \in \mathbb{R}^K$ a constant vector (all entries the same), then for all $\boldsymbol{\beta} \in \mathbb{R}^K$, the SUCPA-map satisfies:*

$$\mathbf{f}(\boldsymbol{\beta}+\boldsymbol{\lambda}) = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\lambda} \tag{9}$$

Proof. It is straightforward evaluating Eq. (6) in $\boldsymbol{\beta}+\boldsymbol{\lambda}$. □

Naturally, this property remains valid for successive iterations of $\mathbf{f}$ and it is the reason for the geometry structure of the fixed points set –previously subjected to the proof of its existence.

Corollary 3.2. *Let $\boldsymbol{\lambda} = [\lambda, \dots, \lambda] \in \mathbb{R}^K$. If $\boldsymbol{\beta}^*$ is a fixed point of $\mathbf{f}$, then the points in the straight line, $\mathcal{S}(\boldsymbol{\beta}^*)$, with $\boldsymbol{\lambda}$ direction and through $\boldsymbol{\beta}^*$:*

$$\mathcal{S}(\boldsymbol{\beta}^*) = \{\boldsymbol{\beta} = \boldsymbol{\lambda} + \boldsymbol{\beta}^*\} \tag{10}$$

are also fixed points of $\mathbf{f}$.

Proof. It is straightforward by induction in $t\in\mathbb{N}$. Lemma 3.1 is the case for $t=1$. Assuming $\mathbf{f}^{(t)}(\boldsymbol{\beta}+\boldsymbol{\lambda}) = \mathbf{f}^{(t)}(\boldsymbol{\beta}) + \boldsymbol{\lambda}$ is valid for $t>1$, then for $t+1$, Lemma 3.1 implies:

$$\mathbf{f}^{(t+1)}\big(\boldsymbol{\beta}+\boldsymbol{\lambda}\big) = \mathbf{f}\left(\mathbf{f}^{(t)}(\boldsymbol{\beta}+\boldsymbol{\lambda})\right) = \mathbf{f}\big(\mathbf{f}^{(t)}(\boldsymbol{\beta})+\boldsymbol{\lambda}\big) = \mathbf{f}^{(t+1)}(\boldsymbol{\beta}) + \boldsymbol{\lambda} \tag{11}$$

□

This result shows that, if for some i.c. $\boldsymbol{\beta}^{[0]}$ the algorithm converges to $\boldsymbol{\beta}^{*}$, then for a shifted i.c. $\boldsymbol{\beta}^{[0]}+\boldsymbol{\lambda}$, it converges to $\boldsymbol{\beta}^{*}+\boldsymbol{\lambda}$ (an example of this behaviour can be seen in Fig. 3). Therefore, if $\boldsymbol{\beta}^{*}$ is a fixed point of $\mathbf{f}$, the points of the straight line with $\mathbf{1}=[1,\dots,1]$ direction and through $\boldsymbol{\beta}^{*}$ are also fixed points of $\mathbf{f}$. In fact, there exists abundant numerical evidence to formulate the following conjecture, although only the case $K=2$ is formally proven in Sec. 4.

Conjecture 3.3. *The map* $\mathbf{f}$ *has a unique straight line of fixed points which is of the form* (10)*.*

An example for $K=3$ can be seen in Fig. 4.

3.2. **Jacobian Matrix of the SUCPA-map.** The Jacobian matrix of a system is an essential feature to study, for example, the local dynamics around the fixed points of a map [6, 32].

The Jacobian matrix of the map $\mathbf{f}$ evaluated at $\boldsymbol{\beta}$, $\mathbf{J}(\boldsymbol{\beta})\in\mathbb{R}^{K\times K}$, has elements, $J_{k,\ell}(\boldsymbol{\beta})$, which can be computed as:

$$J_{k,\ell}(\boldsymbol{\beta}) = \frac{\partial f_k}{\partial \beta_\ell}(\boldsymbol{\beta}) = \frac{\displaystyle\sum_{i=1}^{N} \frac{P_{i,k}P_{i,\ell}\, e^{\beta_\ell}}{\big(\sum_{j=1}^{K} P_{i,j}\, e^{\beta_j}\big)^2}}{\displaystyle\sum_{i=1}^{N} \frac{P_{i,k}}{\sum_{j=1}^{K} P_{i,j}\, e^{\beta_j}}}\,, \quad k,\ell=1,\dots,K \tag{12}$$

In particular, $\mathbf{J}(\boldsymbol{\beta})$ turns to be a *regular transition probability matrix* for all $\boldsymbol{\beta}\in\mathbb{R}^K$, and so, it has the following properties [24]:

Lemma 3.4. *The* $\mathbf{J}(\boldsymbol{\beta})$ *matrix defined by the elements of* (12) *verifies:*

(i) $J_{k,\ell}(\boldsymbol{\beta})>0,\ \forall\, k,\ell.$

(ii) *Each row of the matrix adds up to 1, i.e.:*

$$\sum_{\ell=1}^{K} J_{k,\ell}(\boldsymbol{\beta})=1,\ \forall\ 1\le k\le K \tag{13}$$

This condition necessarily means that $\mathbf{J}(\boldsymbol{\beta})$ *has an eigenvalue* $\mu=1$ *with* $\mathbf{1}=[1,\dots,1]$ *as an associated eigenvector.*

(iii) $\mu=1$ *has multiplicity one, and all others eigenvalues verify* $|\mu|<1$*.*

Proof. Items (i) and (ii) are straightforward to find out. Item (iii) is a consequence of Perron's Theorem [24, Chapter 8]. □

3.3. **Complementary definitions and properties.** To deepen into some details, we add other necessary concepts and results.

Definition 3.5. The *forward orbit of* $\boldsymbol{\beta}\in\mathbb{R}^K$ is the set:

$$O^{+}(\boldsymbol{\beta}) = \left\{\mathbf{f}^{(t)}(\boldsymbol{\beta}), t\ge 0\right\} \tag{14}$$

$O^{+}(\boldsymbol{\beta})$ is the set of all points that the algorithm will pass through when starting at $\boldsymbol{\beta}$.

Definition 3.6. $\mathcal{X}\subset\mathbb{R}^K$ is an *invariant set* if $\mathbf{f}(\mathcal{X})\subset\mathcal{X}$. If $\mathbf{f}(\mathcal{X})=\mathcal{X}$, $\mathcal{X}$ is called *strongly invariant* (or s-invariant for short).

This means that all forward orbits of points in $\mathcal{X}$ lie in $\mathcal{X}$. In particular, any set of fixed points is s-invariant. Also, for all $\boldsymbol{\beta}$, the set $\omega(\boldsymbol{\beta})$ is a s-invariant set.

Definition 3.7. A fixed point $\boldsymbol{\beta}^*$ of a map is called *unstable*, if at least one eigenvalue of $\mathbf{J}(\boldsymbol{\beta}^*)$ has modulus greater than one. Otherwise, it is called *stable*. If all eigenvalues have modulus less than one, it is called *asymptotically stable*.

Let $\Delta_k(t) = \beta_k^{[t+1]} - \beta_k^{[t]}$ be the increment of the k^{th} component of an orbit. The following interesting lemma is related to them.

Lemma 3.8. *For $\Delta_k(t)$ the following condition holds:*

$$\sum_{k=1}^{K} N_k e^{-\Delta_k(t)} = N \tag{15}$$

Proof. By Eq. (5), it is not hard to see that each $\Delta_k(t)$ can be written as:

$$\Delta_k(t) = -\log\left(\frac{1}{N_k}\sum_{i=1}^{N}\frac{P_{i,k}e^{\beta_k^{[t]}}}{\sum_{j=1}^{K}P_{i,j}e^{\beta_j^{[t]}}}\right) \tag{16}$$

taking exponential functions on both sides of the equation:

$$N_k e^{-\Delta_k(t)} = \sum_{i=1}^{N}\frac{P_{i,k}e^{\beta_k^{[t]}}}{\sum_{j=1}^{K}P_{i,j}e^{\beta_j^{[t]}}} \tag{17}$$

and finally adding all the terms over k, the proof is done. □

Lemma 3.8 shows that, in some sense, the results at every step of the algorithm values are "balanced", since, if for some k, β_k tends to increase with the iterations, then there must be, at least, another k', for which $\beta_{k'}$ tend to decrease (and vice versa).

4. The case of two classes

For $K=2$, $\boldsymbol{\beta}=[\beta_1,\beta_2]$, the SUCPA algorithm (5), and the map $\mathbf{f}=[f_1,f_2]\colon\mathbb{R}^2\to\mathbb{R}^2$, become into:

$$\begin{cases} f_1(\boldsymbol{\beta}^{[t]}) = \beta_1^{[t+1]} = -\log\left(\dfrac{1}{N_1}\displaystyle\sum_{i=1}^{N}\frac{P_{i,1}}{P_{i,1}\,e^{\beta_1^{[t]}}+P_{i,2}\,e^{\beta_2^{[t]}}}\right) \\ f_2(\boldsymbol{\beta}^{[t]}) = \beta_2^{[t+1]} = -\log\left(\dfrac{1}{N_2}\displaystyle\sum_{i=1}^{N}\frac{P_{i,2}}{P_{i,1}\,e^{\beta_1^{[t]}}+P_{i,2}\,e^{\beta_2^{[t]}}}\right) \end{cases} \tag{18}$$

where $N_1+N_2=N$, $P_{i,k}>0$, $k=1,2$ and $P_{i,1}+P_{i,2}=1$, for all $i=1,\ldots,N$.

Regarding the increments, Lemma 3.8 remains:

$$N_1\,e^{-\Delta_1(t)}+N_2\,e^{-\Delta_2(t)} = N \tag{19}$$

In addition, for $K=2$ it is possible to derive an additional property. Once it is known that the algorithm must be convergent, it is necessary that $\Delta_1(t)$ and $\Delta_2(t)$ go to zero when $t\to\infty$. From Eq. (19), and applying the L'Hopital's rule, it can be obtained the direction of convergence:

$$\lim_{t\to+\infty}\frac{\Delta_2(t)}{\Delta_1(t)} = \lim_{\Delta_1\to 0}\frac{-\log\left(\dfrac{N-N_1e^{-\Delta_1}}{N_2}\right)}{\Delta_1} = \lim_{\Delta_1\to 0}\frac{-N_2}{(N-N_1e^{-\Delta_1})}\,\frac{N_1}{N_2}\,e^{-\Delta_1} = -\frac{N_1}{N_2} \tag{20}$$

This means that $O^+(\boldsymbol{\beta}^{[0]})\subset\mathbb{R}^2$ becomes tangent to a straight line of slope $-N_1/N_2$, and this is true for every i.c. $\boldsymbol{\beta}^{[0]}$.

In order to obtain the fixed points of Eq. 18, for $x\in\mathbb{R}$, let $\mathcal{S}_x$ be the straight line with slope 1 and crossing the y-axis at x value:

$$\mathcal{S}_x = \{\boldsymbol{\beta}=[\lambda,\lambda+x],\ \lambda\in\mathbb{R}\} \tag{21}$$

As the straight line of fixed point to be found is of this type, it will be enough to determine only its intercept x.

4.1. **Main results.** The results will be splitted up into those related to convergence and those related to the Jacobian matrix.

4.1.1. *About the convergence.* As said before, Conjectures 2.3 and 3.3 can be proven when the number of classes $K=2$. The main result can be stated as follow: *There exists a unique straight line of slope 1 built up by fixed points, to which the algorithm converges, for all initialization.* In terms of dynamical systems, this is equivalent to state the following theorems.

Theorem 4.1. *Let* $\mathbf{f}$ *be the map defined in* (18). *Then, there exists a unique* $b \in \mathbb{R}$ *, such that the fixed point set of* $\mathbf{f}$ *is* $\mathcal{S}_b$.

Note that all fixed points are not-isolated, and $\mathcal{S}_b$ is an unbounded set.

Theorem 4.2. *Let* $\mathbf{f}$ *be the map defined in* (18). *For all i.c.* $\boldsymbol{\beta}^{[0]}$, *there exist a fixed point,* $\boldsymbol{\beta}^* \in \mathcal{S}_b$ *(*$\boldsymbol{\beta}^*(\boldsymbol{\beta}^{[0]})$*), such that* $\omega(\boldsymbol{\beta}^{[0]})=\{\boldsymbol{\beta}^*\}$.

Both proofs are developed in Section 4.2.

4.1.2. *About the Jacobian matrix.* As said before, the eigenvalues and eigenvectors of the Jacobian matrix of a map, evaluated at a fixed point, account for the local dynamics of the system around that point. A fixed point $\boldsymbol{\beta}^*$ of a map $\mathbf{f}$ is call *hyperbolic*, if the Jacobian matrix, $\mathbf{J_f}(\boldsymbol{\beta}^*)$, has no eigenvalues, μ in the unit circle. Otherwise, it is called *non-hyperbolic.* The invariant spaces, E^s, E^u and E^c, of $\mathbf{J_f}(\boldsymbol{\beta}^*)$, associated to $|\mu|>1$, $|\mu|<1$ and $|\mu|=1$, respectively, correspond to the *local* stable, the *local* unstable and the *local* "central" directions, respectively [6,32].

By Lemma 3.4, the two eigenvalues of $\mathbf{J_f}(\boldsymbol{\beta}^*)$ are, 1 with associated eigenvector $\mathbf{1}$, and the other, μ, must verify $|\mu|<1$, so, the fixed points are all of non-hyperbolic type. The eigenvalue 1 is in accordance with the obtaining of non-isolated fixed points [18], and the eigenvector $\mathbf{1}$ is in accordance with the direction of the fixed points straight line. Therefore, $E^u=\emptyset$, $E^c=\text{span}\{\mathbf{1}\}$ and $E^s=\text{span}\{\mathbf{v}\}$, where $\mathbf{v}$ is the eigenvector associated to μ to be determined. The following lemma shows some features for the eigen-pair of the Jacobian matrix for the fixed points.

Lemma 4.3. *Let* $\boldsymbol{\beta}^* \in \mathbb{R}^2$ *a fixed point of* $\mathbf{f}$. *The eigen-pairs of* $\mathbf{J_f}(\boldsymbol{\beta}^*)$ *are* $(1,\mathbf{1})$ *and* $(\mu,\mathbf{v})$ *where:*

(i) *The value of* μ *and* $\mathbf{v}$ *are the same for all fixed point* $\boldsymbol{\beta}^*$
(ii) $0 \le \mu < 1$
(iii) $\mathbf{v} = [N_2, -N_1]$

The proof can be seen in Section 4.2. Fig. 1 illustrate the phase portrait of the local dynamics near $\mathcal{S}_b$. The eigenvalues and eigenvectors of $\mathbf{J}(\boldsymbol{\beta}^*)$ account for the local dynamics, so every point $\boldsymbol{\beta}^* \in \mathcal{S}_b$ is stable, which means that there exists a neighborhood of $\boldsymbol{\beta}^*$, say $B(\boldsymbol{\beta}^*,\varepsilon)$, for some $\varepsilon>0$, such that every orbit that enters it, stays there. Besides, as the ε value is *the same* for all $\boldsymbol{\beta}^*$ (because the Jacobian matrix is), it turns out that the set:

$$\mathcal{S}_b(\varepsilon) = \bigcup_{\boldsymbol{\beta}^* \in \mathcal{S}_b} B(\boldsymbol{\beta}^*,\varepsilon) \tag{22}$$

is an invariant set for the map $\mathbf{f}$. The set $\mathcal{S}_b(\varepsilon)$ is nothing but the strip of 2ε-width around $\mathcal{S}_b$, and the dynamics within it will be characterize by the sign of μ (see Fig. 1).

In order to prove the main results, in the next section we introduce some auxiliary functions useful in the technical details of the proofs.

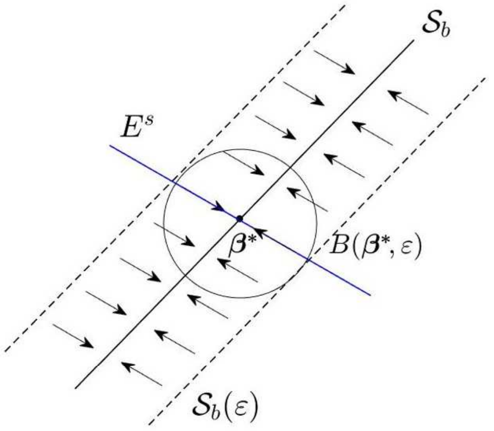

$\mathcal{S}_b$
E^s
$\boldsymbol{\beta}^*$
$B(\boldsymbol{\beta}^*,\varepsilon)$
$\mathcal{S}_b(\varepsilon)$

FIGURE 1. Schematic representation of local stability at a fixed point $\boldsymbol{\beta}^*\in\mathcal{S}_b$ within a ball of radius ε centered at $\boldsymbol{\beta}^*$, $B(\boldsymbol{\beta}^*,\varepsilon)$. The stability direction E^s (blue line). The direction of E^c matches that of the $\mathcal{S}_b$. The local dynamics is a replica for each $\boldsymbol{\beta}^*\in\mathcal{S}_b$ within the 2ε-width strip $\mathcal{S}_b(\varepsilon)$

4.1.3. *Auxiliary functions and its properties.* Let $\alpha_1,\alpha_2:\mathbb{R}\to\mathbb{R}$ be the following function:

$$\alpha_1(x)=\frac{1}{N_1}\sum_{i=1}^{N}\frac{P_{i,1}}{P_{i,1}+P_{i,2}\,e^x}\qquad \alpha_2(x)=\frac{1}{N_2}\sum_{i=1}^{N}\frac{P_{i,2}}{P_{i,1}+P_{i,2}\,e^x}\tag{23}$$

with $P_{i,k}>0,\ k=1,2,\ P_{i,1}+P_{i,2}=1,\ i=1,\dots,N$. Functions $\alpha_k(x)$ are clearly positive and continuous derivable. Also is easy to see they satisfy the following relation, for $x\in\mathbb{R}$:

$$N_1\alpha_1(x)+N_2\,e^x\alpha_2(x)=N\tag{24}$$

Lemma 4.4. *$\alpha_k(x)$ are monotone decreasing on x, $k=1,2$.*

Proof. Indeed, for all $x\in\mathbb{R}$:

$$\alpha_1'(x)=\frac{1}{N_1}\sum_{i=1}^{N}\frac{-P_{i,1}P_{i,2}\,e^x}{(P_{i,1}+P_{i,2}\,e^x)^2}<0\qquad \alpha_2'(x)=\frac{1}{N_2}\sum_{i=1}^{N}\frac{-P_{i,2}^2\,e^x}{(P_{i,1}+P_{i,2}\,e^x)^2}<0\tag{25}$$

□

4.2. **Proofs of the main results.** Bellow are the proofs of Theo. 4.1, Lemma 4.3 and Theo. 4.2.

The first theorem is about the existence and uniqueness of a straight line, $\mathcal{S}_x$, of fixed points of $\mathbf{f}$, for some x.

Proof. Theo. 4.1 The strategy is to prove that there exists a unique value $b\in\mathbb{R}$, such that $\boldsymbol{\beta}^*=[0,b]$ is a fixed point of $\mathbf{f}$. Thus, by virtue of Corollary 3.2, the straight line $\mathcal{S}_b$ is the desired line (see Eq. (21)). For $[0,b]$ to be a fixed point of $\mathbf{f}$, it has to verify, $\mathbf{f}\big([0,b]\big)=[0,b]$ in Eq. (18).

$$\begin{cases} f_1\big([0,b]\big)=-\log\left(\dfrac{1}{N_1}\displaystyle\sum_{i=1}^{N}\frac{P_{i,1}}{P_{i,1}\,e^0+P_{i,2}\,e^b}\right)=0\\ f_2\big([0,b]\big)=-\log\left(\dfrac{1}{N_2}\displaystyle\sum_{i=1}^{N}\frac{P_{i,2}}{P_{i,1}\,e^0+P_{i,2}\,e^b}\right)=b\end{cases}\tag{26}$$

Using the auxiliary functions α_k defined in Eq. (23), this is equivalent to prove that there exists a unique $b\in\mathbb{R}$, such that $\alpha_1(b)=1$ and $\alpha_2(b)=e^{-b}$.

Indeed, taking limits for $\alpha_1(x)$ when $x\to-\infty$ and $x\to+\infty$ respectively, it is obtained:

$$\lim_{x\to-\infty}\alpha_1(x)=\lim_{x\to-\infty}\frac{1}{N_1}\sum_{i=1}^{N}\frac{P_{i,1}}{P_{i,1}+P_{i,2}\,e^x}=\frac{N}{N_1}>1 \tag{27}$$

$$\lim_{x\to+\infty}\alpha_1(x)=\lim_{x\to+\infty}\frac{1}{N_2}\sum_{i=1}^{N}\frac{P_{i,1}}{P_{i,1}+P_{i,2}\,e^x}=0<1 \tag{28}$$

The existence follows from the mean value theorem and the uniqueness from the α_1 strict monotony. The proof of $\alpha_2(b)=e^{-b}$ comes straight out from (24). □

Note that b is obtained implicitly, and only can be found by numerical methods.

The second theorem is about the global stability of the map $\mathbf{f}$. The strategy is to prove that $\mathcal{S}_b$ is a *global attractor* set for $\mathbf{f}$.

4.2.1. *About the intercepts.* For every point $\boldsymbol{\beta}=[\beta_1,\beta_2]\in\mathbb{R}^2$ there is a single straight line, $\mathcal{S}_x$, of slope 1 and intercept, $x=\beta_2-\beta_1$, that passes through it. This section is devoted to show that every $\mathcal{S}_x$ monotonously approaches to $\mathcal{S}_b$, where b is the intercept of the fixed points line, previously obtained. Subtracting the equations in (18) and using the auxiliary functions α_k of 23, it is obtained:

$$\beta_2^{[t+1]}-\beta_1^{[t+1]}=\log\left(\frac{\alpha_1\big(\beta_2^{[t]}-\beta_1^{[t]}\big)}{\alpha_2\big(\beta_2^{[t]}-\beta_1^{[t]}\big)}\right) \tag{29}$$

Thus, for each $x\in\mathbb{R}$, the function:

$$\phi(x)=\log\left(\frac{\alpha_1(x)}{\alpha_2(x)}\right) \tag{30}$$

defines the update of the $\mathcal{S}_x$ intercepts, according to the iterations of $\mathbf{f}$. In particular, for the $\mathcal{S}_b$ line of fixed points, this relationship implies:

$$b=\phi(b)=\log\left(\frac{\alpha_1(b)}{\alpha_2(b)}\right) \tag{31}$$

The following lemma shows that the intercepts x monotonously approach to b.

Lemma 4.5.

(i) *For* $x>b$*:* $b\le\phi(x)<x$

(ii) *For* $x<b$*:* $x<\phi(x)\le b$

Proof. The right inequality in Lemma 4.5 (i) is fairly straightforward to prove. By Lemma 4.4 and (28), we have $\alpha_1(x)<1=\alpha_1(b)$, and then, (24) implies $e^x\alpha_2(x)>1$. Thus, $\alpha_1(x)<e^x\alpha_2(x)$, which is equivalent to the statement ($\alpha_2(x)>0$).

To obtain the left inequality, we must "split hair". It is equivalent to prove:

$$\frac{\alpha_2(x)}{\alpha_1(x)}\le e^{-b} \tag{32}$$

Let g be a function on x defined as:

$$g(x)=e^{\phi(x)}=\frac{\alpha_2(x)}{\alpha_1(x)} \tag{33}$$

Note that $g(b)=e^{-b}$, via (31). Thus, to obtain (32), all that remains is to prove that g is a monotone non-increasing function, i.e., $g'(x)\le 0$, for all $x>b$. This is equivalent to show that:

$$N_1N_2\big(\alpha_2'(x)\alpha_1(x)-\alpha_1'(x)\alpha_2(x)\big)\le 0 \tag{34}$$

By using the expressions of α_k and α'_k in 23 and 25:

$$N_1N_2\,\alpha'_2(x)\alpha_1(x) - N_1N_2\,\alpha'_1(x)\alpha_2(x) = \tag{35}$$

$$= \left[\sum_{i=1}^{N} \frac{-P_{i,2}^2\,e^x}{\left(P_{i,1}+P_{i,2}\,e^x\right)^2}\right]\left[\sum_{j=1}^{N} \frac{P_{j,1}}{P_{j,1}+P_{j,2}\,e^x}\right] - \left[\sum_{i=1}^{N} \frac{-P_{i,1}P_{i,2}\,e^x}{\left(P_{i,1}+P_{i,2}\,e^x\right)^2}\right]\left[\sum_{j=1}^{N} \frac{P_{j,2}}{P_{j,1}+P_{j,2}\,e^x}\right] \tag{36}$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{N} \frac{-P_{i,2}^2\,e^xP_{j,1} + P_{i,1}P_{i,2}\,e^xP_{j,2}}{\left(P_{i,1}+P_{i,2}\,e^x\right)^2\left(P_{j,1}+P_{j,2}\,e^x\right)} = e^x\sum_{j=1}^{N}\sum_{i=1}^{N} \frac{P_{i,2}\left(P_{i,1}P_{j,2}-P_{i,2}P_{j,1}\right)}{\left(P_{i,1}+P_{i,2}\,e^x\right)^2\left(P_{j,1}+P_{j,2}\,e^x\right)} \tag{37}$$

$$= e^x\sum_{j=1}^{N}\sum_{i=1}^{N} \frac{P_{i,2}\left(P_{i,1}P_{j,2}-P_{i,2}P_{j,1}\right)\left(P_{j,1}+P_{j,2}\,e^x\right)}{\left(P_{i,1}+P_{i,2}\,e^x\right)^2\left(P_{j,1}+P_{j,2}\,e^x\right)^2} \tag{38}$$

$$= e^x\sum_{j=1}^{N}\sum_{i=1}^{N} \frac{-\left(P_{i,1}P_{j,2}-P_{i,2}P_{j,1}\right)^2}{\left(P_{i,1}+P_{i,2}\,e^x\right)^2\left(P_{j,1}+P_{j,2}\,e^x\right)^2} \leq 0 \tag{39}$$

The last equality comes from the fact that in Eq. (38), the terms with same subscripts $(i=j)$, equal zero; and for $i\neq j$, by lumping together and adding two symmetric terms on i and j, the numerators become:

$$P_{i,2}\left(P_{i,1}P_{j,2}-P_{i,2}P_{j,1}\right)\left(P_{j,1}+P_{j,2}\,e^x\right) + P_{j,2}\left(P_{j,1}P_{i,2}-P_{j,2}P_{i,1}\right)\left(P_{i,1}+P_{i,2}\,e^x\right) = \tag{40}$$

$$= \left(P_{i,1}P_{j,2}-P_{i,2}P_{j,1}\right)\left[P_{i,2}\left(P_{j,1}+P_{j,2}\,e^x\right)-P_{j,2}\left(P_{i,1}+P_{i,2}\,e^x\right)\right] \tag{41}$$

$$= \left(P_{i,1}P_{j,2}-P_{i,2}P_{j,1}\right)\left[P_{i,2}P_{j,1} + P_{i,2}P_{j,2}\,e^x - P_{j,2}P_{i,1} - P_{j,2}P_{i,2}\,e^x\right] \tag{42}$$

$$= -\left(P_{i,1}P_{j,2}-P_{i,2}P_{j,1}\right)^2 \tag{43}$$

Then, (34) holds, $g'(x) \leq 0$ and $g(x) \leq e^{-b}$. In an analogous way, it can be proven the same inequalities corresponding to Lemma 4.5 (ii), for $x<b$. □

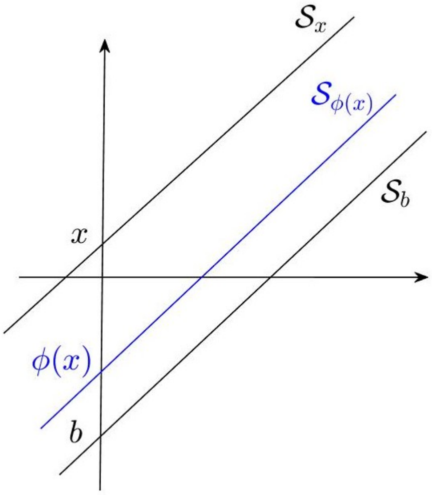


FIGURE 2. Any unitary slope straight line $\mathcal{S}_x$ above the line of fixed points, $\mathcal{S}_b$, is mapped onto another one, $\mathbf{f}(\mathcal{S}_x) = \mathcal{S}_{\phi(x)}$ (in blue), placed between them, but strictly above $\mathcal{S}_x$

The geometrical interpretation of this lemma is the following. Let $\mathcal{S}_b^{\uparrow}$ and $\mathcal{S}_b^{\downarrow}$ be the semi-planes located above and below $\mathcal{S}_b$, respectively, then they are invariant sets for $\mathbf{f}$.

Corollary 4.6. *If* $\boldsymbol{\beta}\in\mathcal{S}_b^{\uparrow}$*, then* $O^+(\boldsymbol{\beta})\in\mathcal{S}_b^{\uparrow}$*, and if* $\boldsymbol{\beta}\in\mathcal{S}_b^{\downarrow}$*, then* $O^+(\boldsymbol{\beta})\in\mathcal{S}_b^{\downarrow}$*.*

Moreover, if $x > b$ (i.e. $\mathcal{S}_x$ is placed above $\mathcal{S}_b$), then, Lemma 4.5 states that $\mathcal{S}_{\phi(x)}$ will be placed between $\mathcal{S}_x$ and $\mathcal{S}_b$, but *strictly* under the first one. This feature can be seen in the Fig. 2.

Proof. Lemma 4.3 Theorem 4.1 states that the fixed point of $\mathbf{f}$ are those $\boldsymbol{\beta}^* = [\lambda, \lambda+b]$, $\lambda \in \mathbb{R}$, with b the only value that verifies $\alpha_1(b)=1$ and $\alpha_2(b)=e^{-b}$. By Eq.(12) the Jacobian matrix of $\mathbf{f}$ at these points results:

$$\mathbf{J}_b = \mathbf{J_f}\big([\lambda, \lambda+b]\big) = \begin{bmatrix} \dfrac{1}{N_1}\displaystyle\sum_{i=1}^{N} \dfrac{P_{i,1}^2}{(P_{i,1}+P_{i,2}\,e^b)^2} & \dfrac{1}{N_1}\displaystyle\sum_{i=1}^{N} \dfrac{P_{i,1}P_{i,2}\,e^b}{(P_{i,1}+P_{i,2}\,e^b)^2} \\ \dfrac{1}{N_2}\displaystyle\sum_{i=1}^{N} \dfrac{P_{i,1}P_{i,2}\,e^b}{(P_{i,1}+P_{i,2}\,e^b)^2} & \dfrac{1}{N_2}\displaystyle\sum_{i=1}^{N} \dfrac{P_{i,2}^2(e^b)^2}{(P_{i,1}+P_{i,2}\,e^b)^2} \end{bmatrix} \tag{44}$$

Note that $\mathbf{J}_b$ does not depend on λ, i.e., it is the same for all $\boldsymbol{\beta}^* \in \mathcal{S}_b$, and so, the same the eigenvalues and eigenvectors. As a consequence, the local dynamics at one fixed point results in an exact replica at every $\boldsymbol{\beta}^*$. This fact proves item (i).

In order to obtain both, the other eigenvalue μ and its eigenvector $\mathbf{v}$, the matrix entries in Eq. (44) are written in simplified form by means of the auxiliary functions $\alpha_k, k=1,2$. Using Eqs. (13) and (25), the matrix in (44), becomes:

$$\mathbf{J}_b = \mathbf{J_f}\big([\lambda, \lambda+b]\big) = \begin{bmatrix} 1+\alpha_1'(b) & -\alpha_1'(b) \\ 1+e^b\alpha_2'(b) & -e^b\alpha_2'(b) \end{bmatrix} \tag{45}$$

The trace of $\mathbf{J}_b$ equals the sum of its eigenvalues, so:

$$\mu = \alpha_1'(b) - e^b\alpha_2'(b) \tag{46}$$

Let $h(x)$ be another auxiliary function on x defined as $h(x) = \alpha_2'(x)\alpha_1(x) - \alpha_1'(x)\alpha_2(x)$. By virtue of (34), we have that $h(x) \leq 0$ for all $x \in \mathbb{R}$. In particular, for $x = b$, and using that $\alpha_1(b)=1$ and $\alpha_2(b)=e^{-b}$, it is possible to obtain the sign of μ:

$$h(b) = \alpha_2'(b)\alpha_1(b) - \alpha_1'(b)\alpha_2(b) \leq 0 \tag{47}$$

$$\alpha_2'(b) - \alpha_1'(b)\,e^{-b} \leq 0 \tag{48}$$

$$\alpha_2'(b)\,e^b - \alpha_1'(b) \leq 0 \tag{49}$$

$$-\mu \leq 0 \tag{50}$$

Therefore, item (ii) is proven. As a consequence of μ being non-negative, any trajectory that approaches $\mathcal{S}_b$ from one side does not cross to the other side.

For item (iii), on the one hand, the anti-diagonal entries of $\mathbf{J}_b$ in (44) satisfy:

$$N_1 J_{1,2} = N_2 J_{2,1} \tag{51}$$

then, the entries of $\mathbf{J}_b$ in (45) meet:

$$-N_1\alpha_1'(b) = N_2\big(1+e^b\alpha_2'(b)\big) \tag{52}$$

On the other hand, using the expression of μ in (46) we have:

$$E^s = \mathrm{Nul}\big(\mathbf{J}_b - \mu\,\mathbf{I}\big) = \mathrm{Nul}\left(\begin{bmatrix} 1+e^b\alpha_2'(b) & -\alpha_1'(b) \\ 1+e^b\alpha_2'(b) & -\alpha_1'(b) \end{bmatrix}\right) \tag{53}$$

$$= \mathrm{span}\left\{[\alpha_1'(b), 1+e^b\alpha_2'(b)]\right\} \tag{54}$$

$$= \mathrm{span}\big\{[N_2, -N_1]\big\} \tag{55}$$

□

Then, $E^s = \text{span}\,\{\mathbf{v} = [N_2, -N_1]\}$. Note that, in case that $N_1 = N_2$, E^s and $E^c = \mathcal{S}_b$ become orthogonal.

The strategy for proving Theo. 4.2 depends largely on the geometrical interpretation of Lemma 3.1 which, in the case of two dimensions, means that any unitary straight line is mapped by $\mathbf{f}$ onto another unitary straight line. The idea is to prove that any $\mathcal{S}_x$ gets closer and closer to the fixed line $\mathcal{S}_b$ with successive iterations of the map. As they are parallel lines, it is enough to study only the sequence given by the intercepts x. First, the following auxiliary lemmas will be useful.

Lemma 4.7. *For all $x \in \mathbb{R}$, $\mathbf{f}^{(t)}(\mathcal{S}_x) \to \mathcal{S}_b$, as $t \to +\infty$*

Proof. Let $x \in \mathbb{R}$, and $\boldsymbol{\beta} = [0, x]$. Evaluating $\mathbf{f}$ in β in Eq. (18), and using functions α_k, we have that:

$$\mathbf{f}\big([0, x]\big) = [-\log \alpha_1(x), -\log \alpha_2(x)] \tag{56}$$

so, by Lemma 3.1:

$$\mathbf{f}\big([0, x] + \boldsymbol{\lambda}\big) = \boldsymbol{\lambda} + [-\log \alpha_1(x), -\log \alpha_2(x)] \tag{57}$$

which, in terms of Eq. (21), means that $\mathbf{f}(\mathcal{S}_x) = \mathcal{S}_{\phi(x)}$, where $\phi(x)$ is the function defined in Eq. (30), i.e., the unitary slope straight line with x intercept, is mapped to the unitary slope straight line with $\phi(x)$ intercept (see Fig. 2). Recursively, by applying Lemma 4.3 to $\phi(x) > b$, we have $\mathbf{f}^{(2)}(\mathcal{S}_x) = \mathcal{S}_{\phi^2(x)}$, where $\phi^2 = \phi \circ \phi$, and so on. In other words, for $x > b$, the sequence of intercepts, $\{\phi^t(x)\}_{t \geq 1}$, is monotone decreasing and:

$$\lim_{t \to +\infty} \phi^t(x) = b \tag{58}$$

The analogous way, if $x < b$, $\mathcal{S}_{\phi(x)}$ will be placed between $\mathcal{S}_b$ and $\mathcal{S}_x$, but *strictly* above the latter. Following the previous arguments, the succession $\{\phi^t(x)\}_{t \geq 1}$ is monotone increasing and so, it converges to b. □

As a natural consequence of this lemma, it is possible to say that the orbit of any initial condition approaches $\mathcal{S}_b$.

Corollary 4.8. *For any i.c. $\boldsymbol{\beta}^{[0]}$, $O^+(\boldsymbol{\beta}^{[0]}) \to \mathcal{S}_b$, for $t \to +\infty$*

Proof. Indeed, any point, $\boldsymbol{\beta}^{[0]} \in \mathbb{R}^2$, $\boldsymbol{\beta}^{[0]} = \big[\beta_1^{[0]}, \beta_2^{[0]}\big]$ belongs to $\mathcal{S}_x$, for $x = \beta_1^{[0]} - \beta_1^{[0]}$. □

Proof. Teo. 4.2

The latest results show that the line $\mathcal{S}_b$ is a global asymptotically stable set for $\mathbf{f}$. However, in order to demonstrate the convergence of SUCPA algorithm, it is still necessary to show that the orbit of any i.c. does not approach *asymptotically* to $\mathcal{S}_b$ but, in fact, reaches a point $\boldsymbol{\beta}^*$ on it. Although it is already evident that the asymptotic behavior is ruled out by the local dynamics illustrated in Fig. 1, we will also give rigorous argument on this point.

Let $\boldsymbol{\beta}^{[t]}$ and $\boldsymbol{\beta}^{[t+1]}$ be two consecutive points of $O^+(\boldsymbol{\beta}^{[0]})$ and $\Delta_k(t) = \beta_k^{[t+1]} - \beta_k^{[t]}$, $k = 1, 2$, their increments. Then, the slope of the straight line through $\boldsymbol{\beta}^{[t]}$ and $\boldsymbol{\beta}^{[t+1]}$, is $\frac{\Delta_2(t)}{\Delta_1(t)}$.

Proceeding by *reductio ad absurdum*, if $O^+(\boldsymbol{\beta}^{[0]})$ where asymptotic to $\mathcal{S}_b$, then the slopes should approach the unity:

$$\lim_{t \to +\infty} \frac{\Delta_2(t)}{\Delta_1(t)} = 1 \tag{59}$$

Indeed, $O^+(\boldsymbol{\beta}^{[0]})$ cannot cross over the straight line $\mathcal{S}_b$ due to Cor. 4.6, and so the "oscillating" asymptotic approach cannot take place. Finally, Eq. (59) is ruled out by Eq. (19), given that the possibility for the increments are two: they are both null ($\Delta_k(t) = 0$, and so, $\boldsymbol{\beta}^{[t]} = \boldsymbol{\beta}^{[t+1]}$), or they are of different sign, so $\dfrac{\Delta_2(t)}{\Delta_1(t)} < 0$, for all t, and this is in contradiction to (59).

Thus, having excluded the asymptotic behavior, there must be a fixed point, $\boldsymbol{\beta}^* \in \mathcal{S}_b$, to which the orbit converges, and then, the proof of Theo. 4.2, for $K = 2$, is completed. □

5. Numerical examples

In this section we present three real-world applications, showing the features proved for the $K=2$ cases and the conjectures made for the $K=3$ case. The first two examples correspond to language models and the third to an image classification application. Implementation details can be found in the repository of this work[2].

5.1. Application to Language models with K=2.

The first example corresponds to the calibration of the posterior probabilities of a language model when it is used as a zero-shot classifier. Specifically, a generative language model is a function that maps a string of characters x (the prompt) to a probability distribution $P_{LM}(t|x)$ over a set of tokens $V=\{t_1,\dots,t_{|V|}\}$ (the vocabulary), which represents the probability of the next token after the input string. Language models can be used as classifiers by computing the probability that the next token is the one that represents the class y to be predicted. For instance, we can compute the probability that the words *"positive"* or *"negative"* appear after the sentence *"Identify if the next review has a positive or negative connotation. Review: 'I love this movie'. Sentiment:"*. This is an example of zero-shot classification because the prompt does not contain examples of correctly classified phrases. In order to obtain a probability distribution $P(y_k|x)$ for each class y_k, we compute the score $s_k=P_{LM}(w_k|x)$ with a predefined set of tokens (in the example, the words $w_1 =$ *"positive"* and $w_2 =$ *"negative"*) and then we normalized over all the scores:

$$P(y_k|x) = \frac{s_k}{\sum_{k=1}^{K} s_k} \tag{60}$$

This is a particularly good use-case scenario because these probabilities produced by the language model are likely to be uncalibrated. For an extended explanation of this problem, see [12].

One of the families of classification problems are those known as *polarity classification* and consists in determining if a given sentence has a positive or negative connotation. For this experiment, a subset of the SST-2 dataset [28], which contains examples of annotated movie reviews was used. The number of positive and negative samples was $N_1=1729$ and $N_2=2271$, respectively. To obtain the class scores, the GPT-2 model [26] was used.

Using the above values of N_1 and N_2 and the probabilities output by the model, a line $\mathcal{S}_b$ of fixed points with $b=-1.39726$ is obtained. Fig. 3 shows this line and the orbits of five different i.c.: $\boldsymbol{\beta}_1^{[0]}=[0,2]$ (in red), $\boldsymbol{\beta}_2^{[0]}=[1.5,3.5]$ (blue), $\boldsymbol{\beta}_3^{[0]}=[3,0]$ (cyan), $\boldsymbol{\beta}_4^{[0]}=[4,-1]$ (green) and $\boldsymbol{\beta}_5^{[0]}=[5,1]$ (magenta). It can be appreciated that orbits starting at on one side of the line $\mathcal{S}_b$, converge to it on the same side, without crossing the semi-planes. Also, Lemma 3.1 and Corollary 3.2 are illustrated: The blue orbit initialized at $\boldsymbol{\beta}_2^{[0]}=\boldsymbol{\beta}_1^{[0]}+\boldsymbol{\lambda}$, i.e. shifted from the red one (in the straight line direction), converges with the same shift, to $\boldsymbol{\beta}_2^*=\boldsymbol{\beta}_1^*+\boldsymbol{\lambda}$. In addition, the convergence of the algorithm is really fast, managing to converge in at most 5 iterations.

5.2. Application to Language models with K=3.

For the second example, we repeat the same procedure for a classification task with $K=3$. We choose the MNLI dataset [33], which contains examples of sentence pairs with textual entailment annotations. Specifically, given a premise sentence and a hypothesis sentence, the task is to predict whether the premise entails the hypothesis (entailment), contradicts the hypothesis (contradiction), or neither (neutral). The prompt template used for this task was the following: *"Premise:* `[premise]`*. Hypothesis:* `[hypothesis]`*. Entailment, neutral or contradiction? Answer:"* and the model used to obtain the probabilities was the same as the first experiment.

In this case, the parameter values are $N_1=1407$, $N_2=1298$ and $N_3=1295$. Fig. 4 shows the line of fixed points, $\mathcal{S}(\boldsymbol{\beta}^*)$, with direction vector $[1,1,1]$; the orbits of five different i.c.: $\boldsymbol{\beta}^{[0]}=[1,-1,1]$ (red), $\boldsymbol{\beta}^{[0]}=[2,-2,2]$ (blue), $\boldsymbol{\beta}^{[0]}=[2.5,-2,1]$ (magenta), $\boldsymbol{\beta}^{[0]}=[2.5,-0.5,1]$ (green) and $\boldsymbol{\beta}^{[0]}=[1,-1.5,1.5]$ (cyan).

[2] `https://github.com/LautaroEst/sucpa-convergence`

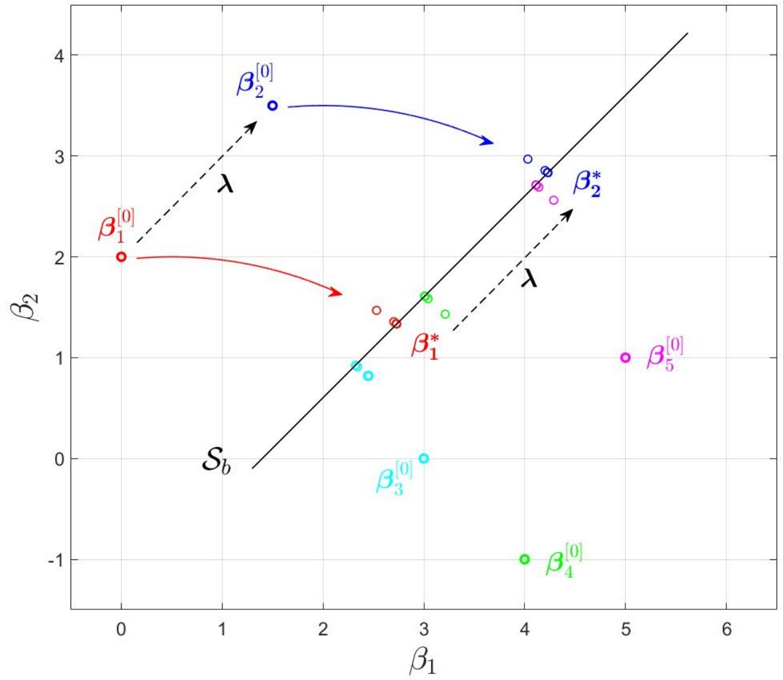


FIGURE 3. Example with $K = 2$. Orbits of five different i.c.: $\boldsymbol{\beta}_1^{[0]} = [0, 2]$ (red), $\boldsymbol{\beta}_2^{[0]} = [1.5, 3.5]$ (blue), $\boldsymbol{\beta}_3^{[0]} = [3, 0]$ (cyan), $\boldsymbol{\beta}_4^{[0]} = [4, -1]$ (green) and $\boldsymbol{\beta}_5^{[0]} = [5, 1]$ (magenta). Only five points of each orbit were plotted due to rapid convergence. Also the line of fixed points, $\mathcal{S}_b$, with $b = -1.39726$. The shift vector is $\boldsymbol{\lambda} = [1.5, 1.5]$

As we mention in Conj. 2.3 and 3.3, all i.c. converges to the unique straight line of unitary slope. This example would seem to indicate that the characteristics shown for $K = 2$ can be extended to a greater number of classes. In addition, the convergence of the algorithm is really fast, managing to converge in at most 5 iterations.

5.3. **Application to Image Classification.** The final example shows that the features shown in the previous examples occur in other completely different applications. In this case, a recognizer of images of dogs and cats was trained on the Asirra dataset [11] obtained from the Dogs-vs-Cats Kaggle competition website[3]. Here, the pretrained ResNet-18 [16] model was fine-tuned on 80% of the training data and validated on the rest. A final accuracy of 0.98 on the validation split was obtained after one epoch using Adam optimization with defaults parameters, a learning rate of 1e-4 and batch size of 64.

In this case, the parameter values are $N_1 = 10053$, $N_2 = 9947$, and $b = -0.03465$ is obtained. Fig. 5 shows the line of fixed points and the orbits of different i.c. The same conclusions as in the previous cases can be seen in this example with the difference that up to 200 steps were needed to achieve convergence.

6. CONCLUSIONS

In this work we proved several convergences properties of SUCPA, a calibration algorithm described by a non-hyperbolic map with a non-bounded set of non-isolated fixed points. Among them it is worth highlighting, on the one hand, the main results: Corollary 3.2 and Lemma 3.4.

[3] `https://www.kaggle.com/c/dogs-vs-cats`

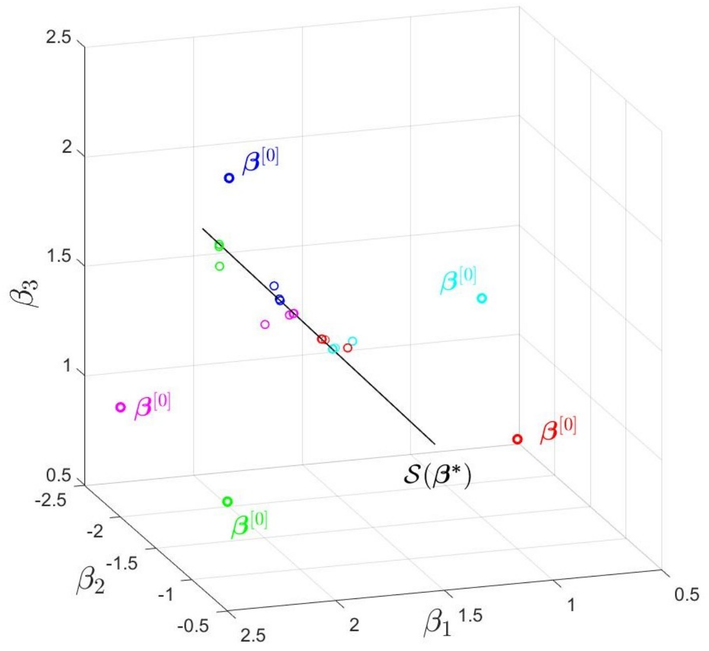


FIGURE 4. Example with $K = 3$. Orbits of five different i.c.: $\boldsymbol{\beta}^{[0]} = [1, -1, 1]$ (red), $\boldsymbol{\beta}^{[0]} = [2, -2, 2]$ (blue), $\boldsymbol{\beta}^{[0]} = [2.5, -2, 1]$ (magenta), $\boldsymbol{\beta}^{[0]} = [2.5, -0.5, 1]$ (green) and $\boldsymbol{\beta}^{[0]} = [1, -1.5, 1.5]$ (cyan). Five points of each orbit are plotted. Also the line of fixed points, $\mathcal{S}(\boldsymbol{\beta}^*)$

The first one establishes that if there exists a fixed point, then, all points in the straight line of direction $[1, \dots, 1]$ and passing through it must be also fixed points of the map (an unbounded set of non-isolated fixed points); while the second one shows that the Jacobian matrix of the system is a regular transition probability matrix (for all $\boldsymbol{\beta} \in \mathbb{R}^K$) which makes it a non-hyperbolic problem.

On the other hand, stronger conclusions were proved for the two–classes case by means of Theo. 4.1 which proves the existence and uniqueness of the fixed points straight line, and Theo. 4.2 which shows that every i.c. converges. These results seem to be valid for the general case, but it remains as future work to prove it formally. Additionally, interesting real-world application examples were presented in language models and image classification for which all the mentioned results are experimentally corroborated.

Another possible study to continue this work is the sensitivity of the priors. SUCPA algorithm greatly degrades its performance when the proportion of labels is poorly estimated. Perhaps this phenomenon can be explained from the point of view of convergence, studying how the fixed points change according to variations in the N_k values. This sensitivity analysis is under study and will be reported elsewhere.

7. Acknowledgments

This work was supported by the University of Buenos Aires (grant UBACYT 20020220200045BA, 20020220400162BA and 20020190100032BA), CONICET (grant PIP 11220200101826CO) and the ANPCyT (grant PICT 2020-01336).

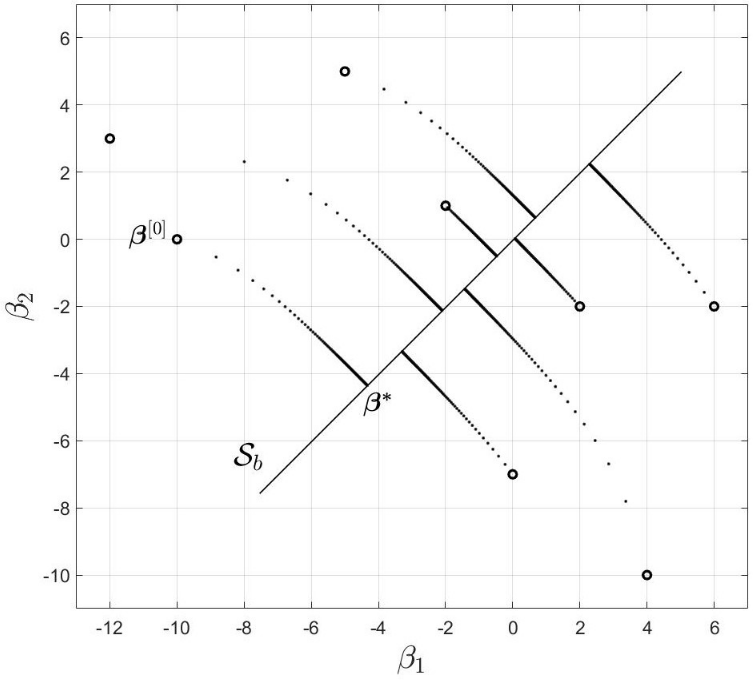


FIGURE 5. Example with $K = 2$ in an image classification task. Orbits of eight different i.c. are plotted. At least 200 points of each one were needed to reach the corresponding limit point $\boldsymbol{\beta}^*$

REFERENCES

[1] Akimoto Y., Auger A., Hansen N. An ODE method to prove the geometric convergence of adaptive stochastic algorithms. *Stochastic Processes and their Applications*, 145:269–307, 2022.

[2] Bauerschmidt R., Brydges D. C., Slade G. Structural stability of a dynamical system near a non-hyperbolic fixed point. *Ann. Henri Poincaré*, 16:1033–1065, 2015.

[3] Brümmer N. *Measuring, refining and calibrating speaker and language information extracted from speech.* PhD thesis, Stellenbosch University, 2010.

[4] Carr J. *Applications of Centre Manifold Theory.* Springer-Verlag, New York, Heidelberg, Berlin, 1982.

[5] Dannan F. M., Elaydi S. N., Ponomarenko V. Stability of hyperbolic and nonhyperbolic fixed points of one-dimensional maps. *Journal of Difference Equations and Applications*, 9(5):449–457, 2003.

[6] Devaney R. *An Introduction to Chaotic Dynamical Systems.* Addison-Wesley, 1987.

[7] Duda R., Hart P., Stork D. *Pattern Classification.* John Wiley, 2 edition, 2001.

[8] Elaydi S. *An Introduction to Difference Equations.* Springer, third edition, 2005.

[9] Elaydi S. *Discrete Chaos with Applications in Sciences and Engineering.* Taylor & Francis Group, LLC, second edition, 2007.

[10] Elsayed E. M. Dynamics and behavior of a higher order rational difference equation. *The Journal of Nonlinear Science and Applications*, 9(4):1463–1474, 2016.

[11] Elson J., Douceur J., Howell J., Saul J. Asirra: A captcha that exploits interest-aligned manual image categorization. In *Proceedings of the 14th ACM Conference on Computer and Communications Security*, CCS '07, page 366–374, New York, NY, USA, 2007. Association for Computing Machinery.

[12] Estienne L., Ferrer L., Vera M., Piantanida P. Unsupervised calibration through prior adaptation for text classification using large language models. *ArXiv*, abs/2307.06713, 2023.

[13] Filho T., Song H., Perello-Nieto M., Santos-Rodriguez R., Kull M., Peter F. Classifier calibration: a survey on how to assess and improve predicted class probabilities. *Machine Learning*, 112(9):3211–3260, 2023.

[14] Godau P., Kalinowski P., Christodoulou E., Reinke A., Tizabi M., Ferrer L., Jäger P., Maier-Hein L. Deployment of image analysis algorithms under prevalence shifts. In *Proc. of MICCAI*, 2023.

[15] Guo C., Pleiss G., Sun Y., Weinberger K. On calibration of modern neural networks. In *Proceedings of the International Conference on Machine Learning ICML*, Sydney, 2017.

[16] He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[17] Ibrahim T. F., Khan A. Q., Oğul B., Şimşek D. Closed-form solution of a rational difference equation. *Advances in Stability and Control of Dynamical Systems*, 2021(ID 3168671):12 pages, 2021.

[18] Jamieson W., Merino O. Local dynamics of planar maps with a non-isolated fixed point exhibiting 1-1 resonance. *Advances in Difference Equations, vol.142*, 2018.

[19] Kapçak S. *A Note on Non-hyperbolic Fixed Points of One-Dimensional Maps*, pages 257–267. Springer, 2021.

[20] Khan A. Q., Ibrahim T. F. Stability and bifurcations analysis of discrete dynamical systems. *Discrete Dynamics in Nature and Society*, 2019(ID 8474706), 2019.

[21] Kuznetsov Y. A. *Elements of Applied Bifurcation Theory.* Springer, New York, 2004.

[22] Li S., Cheah C. Transfer learning algorithm for image classification task and its convergence analysis. In *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, 2023.

[23] Liu Z., Chen G. On area-preserving non-hyperbolic chaotic maps: A case study. *Chaos, Solitons & Fractals*, 16(5):811–818, 2003.

[24] Meyer C. *Matrix Analysis and Applied Linear Algebra.* SIAM, 2000.

[25] Psarros N., Papaschinopoulos G., Schinas C. J. Study of the stability of a 3x3 system of difference equations using centre manifold theory. *Applied Mathematics Letters*, 64:185–192, 2017.

[26] Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever I. Language models are unsupervised multitask learners, 2019.

[27] Schaumann F. *Investigation of nonhyperbolic dynamical systems.* PhD thesis, Max-Planck-Institut für Meteorologie, 2019.

[28] Socher R., Perelygin A., Wu J., Chuang J., Manning C. D., Ng A., Potts C. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.

[29] Stević S., Diblík J., Iričanin B., Šmarda Z. On a solvable system of rational difference equations. *Journal of Difference Equations and Applications*, 20(5-6):811–825, 2014.

[30] Tarlowski D. On asymptotic convergence rate of evolutionary algorithms. Technical report, https://doi.org/10.36227/techrxiv.20103404.v2, 2022.

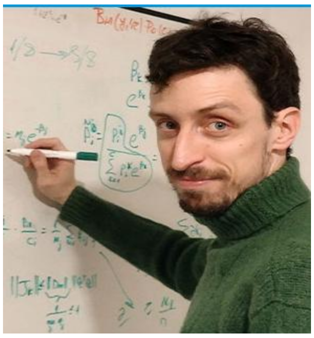
[31] Urbański M., Zdunik A. Geometry and ergodic theory of non-hyperbolic exponential maps. *Transactions of the American Mathematical Society*, 359(8):3973–3997, 2007.

[32] Wiggins S. *Introduction to Applied Nonlinear Dynamical Systems and Chaos, 2nd. Ed.* Springer-Verlag, 2003.

[33] Williams A., Nangia N., Bowman S. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[34] Zolfaghari-Nejad M., Charmi M., Hassanpoor H. A new chaotic system with only nonhyperbolic equilibrium points: Dynamics and its engineering application. *Complexity*, 2022:16 pages, 2022.

Roberta Hansen - graduated from Faculty of Exact and Natural Sciences of the University of Buenos Aires. She received her Ph.D. degree in Mathematics from the same university in 2009. At the present she is Associate Professor at the Department of Mathematics of the Faculty of Engineering, UBA. and a member of its Nonlinear Dynamics Group. Dr. Hansen's research interests also include fractal geometry, multifractality and complexity in time series.



Matias Vera - graduated from Engineering School of the University of Buenos Aires in 2014. He received his Ph.D. degree on Engineering (summa cum laude) from the University of Buenos Aires in 2020. Presently, he is an Associate Professor at the University of Buenos Aires and member of the National Scientific and Technical Research Council in Argentina. Dr. Vera's research interests include machine learning, information theory and statistics.

Lautaro Estienne - graduated from Engineering School of the University of Buenos Aires in 2022. He is currently a Ph.D. student at the Engineering School of the University of Buenos Aires and works in Probabilistic Machine Learning, Natural Language and Speech Processing and Calibration.

Luciana Ferrer - graduated from Engineering School of the University of Buenos Aires, Argentina, in 2001. She received her Ph.D. degree in Electronic Engineering from Stanford University, USA, in 2009. She is currently a researcher at the Computer Science Institute, affiliated to the University of Buenos Aires (UBA) and the National Scientific and Technical Research Council (CONICET), Argentina. Her primary research area is machine learning applied to speech, audio, and language processing tasks, with special focus on the problem of score calibration and robustness to domain mismatch.

Pablo Piantanida received the B.Sc. degree in electrical engineering and the M.Sc. degree from the University of Buenos Aires, Argentina, in 2003, and the Ph.D. degree from Université Paris-Sud, Orsay, France, in 2007. He is currently director of the International Laboratory and Learning Systems (ILLS), professor at CentraleSupélec (Université Paris-Saclay) with CNRS and Associate Academic Member Mila - Quebec AI Institute (Mila). From 2019 to 2021, he was also an associate member of Comète – Inria research team (Lix - Ecole Polytechnique). In 2018 and 2019, he was vising professor at the Université de Montréal and Laboratoire de Mathématiques d'Orsay (LMO). His research interests include information theory, machine learning, security of learning systems and the secure processing of information and applications to computer vision, health, natural language processing, among others. He has served as the General Co-Chair for the 2019 IEEE International Symposium on Information Theory (ISIT). He served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and Editorial Board of Section "Information Theory, Probability and Statistics" for Entropy. He is member of the IEEE Information Theory Society Conference Committee.