

Vector flows that compute the capacity of discrete memoryless channels

Original

Vector flows that compute the capacity of discrete memoryless channels / Beretta, Guglielmo; Pelillo, Marcello. - In: ENTROPY. - ISSN 1099-4300. - ELETTRONICO. - 27:4(2025). [10.3390/e27040362]

Availability:

This version is available at: 11583/2998761 since: 2025-04-02T23:05:57Z

Publisher:

MDPI

Published

DOI:10.3390/e27040362

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Article

Vector Flows That Compute the Capacity of Discrete Memoryless Channels [†]

Guglielmo Beretta ^{1,2,*}  and Marcello Pelillo ^{1,3,4,*} ¹ DAIS, Università Ca' Foscari di Venezia, Via Torino 155, 30170 Venezia, Italy² DAUIN, Politecnico di Torino, Corso Castelfidardo 34/d, 10138 Torino, Italy³ College of Mathematical Medicine, Zhejiang Normal University, Jinhua 321004, China⁴ European Centre for Living Technology, Ca' Bottacin, Dorsoduro 3911, Calle Crosera, 30123 Venezia, Italy* Correspondence: guglielmo.beretta@unive.it (G.B.); pelillo@unive.it (M.P.)[†] This paper is an extended version of our paper published in 7th International Conference, DIS 2024, Kalamata, Greece, 2–7 June 2024.

Abstract: One of the fundamental problems of information theory, since its foundation by C. Shannon, has been the computation of the capacity of a discrete memoryless channel, a quantity expressing the maximum rate at which information can travel through the channel. In this paper, we investigate the properties of a novel approach to computing the capacity, based on a continuous-time dynamical system. Interestingly, the proposed dynamical system can be regarded as a continuous-time version of the classical Blahut–Arimoto algorithm, and we can prove that the former shares with the latter an exponential rate of convergence if certain conditions are met. Moreover, a circuit design is presented to implement the dynamics, hence enabling analog computation to estimate the capacity.

Keywords: analog computation; capacity; convex optimization; discrete memoryless channel; dynamical systems; mutual information; ODE; optimal input distribution; vector flow



Academic Editor: Jun Chen

Received: 8 February 2025

Revised: 25 March 2025

Accepted: 25 March 2025

Published: 29 March 2025

Citation: Beretta, G.; Pelillo, M.Vector Flows That Compute the Capacity of Discrete Memoryless Channels. *Entropy* **2025**, *27*, 362.<https://doi.org/10.3390/e27040362>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Estimating the capacity of a discrete memoryless channel (DMC) is a well-known problem related to the reliability of point-to-point communication systems, as a consequence of C. Shannon's noisy-channel coding theorem [1–3]. A fundamental algorithm to address this problem is the classical Blahut–Arimoto algorithm (BAA), an iterative algorithm based on an alternating maximization procedure [4]. Named after S. Arimoto [5] and R. Blahut [6], who discovered it independently, the BAA requires only some mild conditions on the zero elements of the transition matrix and, unlike its antecedents, it is also applicable when the input alphabet and the output alphabet of the channel have different cardinalities. Notably, the BAA is still a subject of active research (see, e.g., Refs. [7–10]).

In contrast to the BAA, this paper describes a continuous-time dynamical system used to compute the capacity of a DMC. This dynamical system is obtained via the (forward) flow of a suitable ODE, and it can evolve a given distribution towards an optimal input distribution of the channel, hence enabling capacity computation. In studying this unconventional way to address capacity computation, we were inspired by the work by R. W. Brockett, who, as reported, e.g., in [11], provides a novel way, grounded in calculus, to approach problems traditionally addressed via algorithms [12]. Interestingly, the proposed capacity-computing ODE (denoted by CC-ODE in the sequel) has a connection with the BAA. Indeed, it can be regarded, in a sense, as a continuous-time version of the BAA. To

support this claim, we leverage the notion of the multiplicative-weight-update (MWU) rule [13] and its connection to both the BAA (see Refs. [8,14]) and to some discretization techniques used, e.g., in evolutionary game theory [15,16].

The link between the BAA and the CC-ODE flow can be further extended when studying the convergence rates. By construction, the BAA generates a sequence of input distributions, and understanding the convergence rate of this sequence to an optimal input distribution has been central in the study of the BAA ever since its origin [5]. Remarkably, estimating an optimal input distribution for a generic DMC involves some issues affecting not only the BAA, but also any iterative algorithm running on a Turing machine, as recently shown in [10] via computability theory arguments. Despite this, some technical conditions on DMCs ensure exponential convergence of this sequence to an optimal input distribution [5,9]. We prove that under these conditions, in the formulation given in [9], the CC-ODE flow converges exponentially to an optimal input distribution, which can be shown thanks to some tools of Lyapunov's stability theory [17]. The convergence rate can be further refined for a trivial family of DMCs, namely, the noiseless symmetric channels—see, e.g., Ref. [18] (p. 77). Even though a known formula exists for their capacity, we found it interesting that these channels are associated with a CC-ODE for which an explicit analytic solution is available, and we leverage this to produce a more precise asymptotic estimate of the flow.

Lastly, we propose a circuit design to implement the CC-ODE flow, thereby enabling analog computation of the capacity. Analog computation is an important alternative to digital computation [19], and is still a topic of active research—see, e.g., Refs. [20–27]. We speculate about how the proposed circuit could preserve its effectiveness even in the presence of noise, and we comment on its usage in association with the unavailability of some channel input symbols. See also Ref. [28] for a similar circuit design dealing with a labeling task.

This paper is a follow-up of [29], where the empirical usage of numerical methods applied to the CC-ODE was studied to compute the capacity. However, no mathematical proofs were given in [29], and except for a simplified version of Theorem 2, the results presented in this paper do not appear in [29]. To the best of our knowledge, no previous work has studied continuous-time dynamical systems to compute the capacity of DMCs.

Despite this, the CC-ODE can be regarded, under some technical conditions, as an instance of a class of ODEs discussed in [30], where smooth functions are optimized over polyhedra using some notions of Riemannian geometry. In addition to that, similar ODEs appear in the literature on convex optimization (see, e.g., Ref. [31]), as well as in some models pertaining to evolutionary game theory [16,32]. However, in contrast to [16,30,31], the objective function associated with the CC-ODE may be not differentiable on part of the boundary of the feasible set, and we provide the technical arguments that are required to adapt the existing results to the problem under discussion.

The subsequent sections of this paper are organized as follows. Notation conventions are established in Section 2, and Section 3 reviews the aforementioned class of ODEs that have been used in the literature to tackle optimization programs on a standard simplex. We also mention how relaxed hypotheses on the objective function may negatively affect trajectory convergence to stationary points, and in Lemma 1, we discuss some alternative conditions to overcome this issue. In Section 4, the reader is introduced to the problem of computing the channel capacity for DMCs and its formulation as a concave optimization program on a standard simplex. The main contributions of this paper are in Section 5, which deals with the CC-ODE, its properties, and its link with the BAA, and Section 6, where the convergence rates are discussed. The circuit designed to implement the CC-ODE flow is examined in Section 7. Some simulations are described in Section 8 so as to give more

insight into how the CC-ODE flow behaves in comparison with the BAA. Section 9 contains some remarks on our results, and final considerations and future research directions are reported in Section 10.

2. Notations

In this paper, the information content is measured in *nats*, since this choice simplifies the theoretical computations involving mutual information, as mentioned, e.g., in [6]. For a conversion to bits, we recall that 1 nat equals $\ln 2$ bits [2]. In the sequel, we consider the expression $\alpha \ln \alpha$ as well defined and equal to 0 in case $\alpha = 0$. For every integer $n > 0$, the *standard simplex* in \mathbb{R}^n is the set

$$\Delta_n = \left\{ z \in \mathbb{R}^n \mid z \geq \mathbf{0} \text{ and } \sum_{i=1}^n z_i = 1 \right\},$$

its (relative) *interior* is the set

$$\text{int}(\Delta_n) = \left\{ z \in \mathbb{R}^n \mid z > \mathbf{0} \text{ and } \sum_{i=1}^n z_i = 1 \right\},$$

and its (relative) *boundary* is the set $\partial\Delta_n = \Delta_n \setminus \text{int}(\Delta_n)$. For $z = (z_1, \dots, z_n)^\top \in \mathbb{R}^n$, the *support* of x is the set $\text{supp}(z) = \{ i \in [n] \mid z_i \neq 0 \}$. The canonical basis of \mathbb{R}^n is denoted by e_1, \dots, e_n , and $\mathbf{1}_n = \sum_{i=1}^n e_i$. Given $x, y \in \mathbb{R}^n$, we write $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ for the usual dot product between x and y , whereas we write $\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^n x_i^2}$ for the Euclidean norm of x , so that $\|x - y\|$ denotes the Euclidean distance between x and y . In this paper, gradients are column vectors. Given x_1, x_2, \dots , and $x_n \in \mathbb{R}$, we define

$$\text{MSQ}(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \left(x_i - n^{-1} \sum_{k=1}^n x_k \right)^2.$$

3. Optimizing Continuous-Time Dynamics

3.1. Preliminaries

We recall that a function $v : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ is (*globally*) *Lipschitz continuous* on Ω if there exists a constant $L > 0$ such that $\|v(x_1) - v(x_2)\| \leq L\|x_1 - x_2\|$ for every $x_1, x_2 \in \Omega$, whereas v is *locally Lipschitz continuous* on Ω if for every $x \in \Omega$, there exists an open neighborhood $O_x \subseteq \mathbb{R}^n$ of x such that the restriction of v to $\Omega \cap O_x$ is Lipschitz continuous on $\Omega \cap O_x$. Global and local Lipschitz continuity play a fundamental role in the existence and uniqueness results concerning solutions of an ODE, as in Picard–Lindelöf theorem—see, e.g., Refs. [33,34]. Given an open $\Omega \subseteq \mathbb{R}^n$, a function $v : \Omega \rightarrow \mathbb{R}^n$ that is locally Lipschitz continuous on Ω , consider the ODE

$$\dot{z} = v(z). \tag{1}$$

We will say that a non-empty set $S \subseteq \Omega$ is *invariant* under (1) if for every $y \in S$, there exists $z :]-\infty, +\infty[\rightarrow S$ that solves (1) and satisfies $z(0) = y$. Similarly, we will say that S is *forward invariant* under (1) if for every $y \in S$, there exists $z : [0, +\infty[\rightarrow S$ that solves (1) and satisfies $z(0) = y$. By the assumptions made on v , note that for every $y \in S$ there exists *at most one* solution $z : [0, +\infty[\rightarrow \Omega$ of (1) such that $z(0) = y$. This is a trivial consequence of Picard–Lindelöf theorem. However, note that these assumptions *do not guarantee the existence* of such a solution—which is related to the problem of extending local solutions (see also Ref. [35] (Ch. 17.4))—and we remark that stronger assumptions are generally required to

prove the existence, such as those reported in Appendix A. We also recall that a point $y \in \Omega$ is stationary for (1) if $v(y) = 0$, i.e., if the constant function $z(t) \equiv y$ solves (1).

3.2. Optimizing Differential Equation on the Standard Simplex

A program of the form

$$\max_{z \in \Delta_n} f(z) \tag{2}$$

is related, under suitable conditions on its objective function f , to the ODE

$$\dot{z}_i = z_i \left[\partial_i f(z) - \sum_{k=1}^n z_k \partial_k f(z) \right], \quad i \in [n],$$

which can also be written in vector notation as

$$\dot{z} = [\text{diag}(z) - zz^\top] \nabla f(z). \tag{3}$$

In particular, for $z \in \text{int}(\Delta_n)$, the ODE (3) is known as the *Shahshahani gradient system* associated with f —see, e.g., Refs. [16,32]. Many properties of (3) are discussed in the literature under the assumption that f admits a globally Lipschitz continuous gradient on Δ_n , which guarantees the invariance of Δ_n under (3) and that the function f increases strictly along any non-constant solution of (3) evolving on Δ_n ; see, e.g., Refs. [16,30–32,36], and also Appendix B. Indeed, strictly speaking, the condition that is often assumed is that the gradient of f is locally Lipschitz continuous on some open superset of Δ_n (see, e.g., Ref. [15]), which is a stronger assumption, as shown, e.g., in [35] (p. 400).

However, as we shall see in the following sections, we are interested in objective functions that can exhibit singularities on $\partial\Delta_n$ and so we cannot rely on the usual assumptions made in the literature. Crucially, if the function f is differentiable in $\text{int}(\Delta_n)$ but not on some points of $\partial\Delta_n$, then forward trajectories of (3) may converge to one of these points, where (3) is undefined, and this may happen even in case additional hypotheses such as concavity are met, as in the example shown in Appendix C. By contrast, this cannot happen under the circumstances described in the next lemma, which we will apply in the sequel:

Lemma 1. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ admit a locally Lipschitz continuous gradient on some open $\Omega \subseteq \mathbb{R}^n$. Suppose f is concave on some non-empty compact convex $K \subseteq \Omega \cap \Delta_n$. Let $z : [0, +\infty[\rightarrow K$ be a solution of (3).*

- (i) *There exists $z^* = \lim_{t \rightarrow +\infty} z(t) \in K$ and z^* is a stationary point for (3);*
- (ii) *If $z(0) \in \text{int}(\Delta_n)$, then z^* is a KKT point for (2);*
- (iii) *If $z(0) \in \text{int}(\Delta_n)$ and f is concave and continuous on Δ_n , then z^* is a global solution of (2).*

Proof. (i): We now make the following claim, whose technical proof is deferred to Appendix D:

Claim 1 (Convergence). *There exists $z^* = \lim_{t \rightarrow +\infty} z(t) \in K$.*

Note that $z^* \in K$ entails $z^* \in \Omega$; hence, ∇f is well defined and continuous in a neighborhood of z^* . A standard argument can now be applied to show that z^* is stationary for (3)—namely, since the ω -limit here does not escape Ω , then the singleton $\{z^*\}$ is forward invariant under (3) as a consequence of Gronwall’s lemma—see, e.g., Ref. [16].

(ii): We will use the following claim, which provides a known alternative description, for maximization programs over the standard simplex, of the KKT conditions:

Claim 2 (KKT points). *A point $\mathbf{x}^* \in \Delta_n$ in which f is differentiable is a KKT point for (2) if and only if there exists some $\lambda \in \mathbb{R}$ such that*

$$\partial_i f(\mathbf{x}^*) \leq \lambda \text{ for every } i \in [n], \text{ with equality for every } i \in \text{supp}(\mathbf{x}^*).$$

The interested reader may find a proof for Claim 2 in Appendix E. Assume $\mathbf{z}(0) \in \text{int}(\Delta_n)$, which entails $\mathbf{z}(t) > \mathbf{0}$ for every $t > 0$ by uniqueness of local solutions—see, e.g., Proposition A1.(iii). By (i), the point $\mathbf{z}^* = \lim_{t \rightarrow +\infty} \mathbf{z}(t)$ exists in Ω . Define $\mu(\mathbf{z}) = \sum_{k=1}^n z_k \partial_k f(\mathbf{z})$ and set $\lambda = \mu(\mathbf{z}^*)$. Since \mathbf{z}^* is a stationary point for (3), it follows that $\partial_i f(\mathbf{z}^*) = \lambda$ for every $i \in \text{supp}(\mathbf{z}^*)$. The proof is completed by *reductio ad absurdum* as in the proof of [15] (Prop. 3.5). In fact, let $i \notin \text{supp}(\mathbf{z}^*)$ —i.e., let $z_i^* = 0$ —and suppose also that $\partial_i f(\mathbf{z}^*) > \lambda = \mu(\mathbf{z}^*)$. Then, by continuity, also $\partial_i f(\mathbf{y}) > \mu(\mathbf{y})$ for every $\mathbf{y} \in \Omega$ sufficiently close to \mathbf{z}^* ; hence, $\dot{z}_i(t) > 0$ for every $t > 0$ sufficiently large, which contradicts $z_i(t) \rightarrow 0^+$. It follows that \mathbf{z}^* is a KKT point for (2) by Claim 2.

(iii): If f is concave on Δ_n , then every KKT point for (2) is a global solution for (2)—see, e.g., Ref. [37]; hence, (iii) follows by (ii). □

4. Problem Formulation

4.1. Discrete Memoryless Channel and Capacity

A discrete memoryless channel (DMC) is a communication system that can be described by a triplet $\mathcal{C} = (\mathcal{X}, \mathcal{Y}, \mathbf{P})$, in which $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ and $\mathcal{Y} = \{y_1, y_2, \dots, y_m\}$ are finite alphabets called the *input alphabet* and *output alphabet*, respectively, whereas $\mathbf{P} = [p(j|i)]_{i \in [n], j \in [m]} \in \mathbb{R}^{n \times m}$ is a stochastic matrix, called the *transition matrix*, where $p(j|i)$ expresses the probability that the symbol $y_j \in \mathcal{Y}$ is observed as output of the system whenever the symbol $x_i \in \mathcal{X}$ is sent to the system as input [3]. Without loss of generality, we will work under the following assumption:

For every $j \in [m]$, there exists at least one $i \in [n]$ such that $p(j|i) > 0$.

In other words, we are assuming that \mathcal{Y} represents the minimal output alphabet required for a description of the DMC. In fact, this assumption ensures that for any selected symbol $y \in \mathcal{Y}$ there exists a corresponding input distribution for which y occurs as output with positive probability.

Let X be the *input variable* of \mathcal{C} , i.e., the random variable with range in \mathcal{X} modeling the input of the channel, and let Y be the *output variable* of \mathcal{C} , i.e., the random variable modeling the output of the channel. Following [2], the *mutual information* between X and Y , denoted by $I[X; Y]$, is a non-negative quantity measuring the reduction in uncertainty about X that results from learning the value of Y , and its formulation involves the notion of *entropy*, which for a generic discrete random variable D with range in a set \mathcal{D} is given by $H[D] = -\sum_{d \in \mathcal{D}} p_D(d) \ln[p_D(d)]$. The random variable Y and all the random variables $Y|X = x$ obtained as $x \in \mathcal{X}$ appear, together with the *input distribution* p_X , in the following expression for $I[X; Y]$ (see, e.g., Ref. [2]):

$$I[X; Y] = H[Y] - \sum_{x \in \mathcal{X}} p_X(x) H[Y|X = x].$$

The *capacity* of \mathcal{C} is the maximum value C of $I[X; Y]$ over all possible choices for p_X :

$$C = \max_{p_X} I[X; Y].$$

The capacity C provides a theoretical bound for the information content that can be transmitted through the channel [1–3].

4.2. Optimization Program for the Capacity

Consider a channel $\mathcal{C} = (\mathcal{X}, \mathcal{Y}, \mathbf{P})$ and set $n = |\mathcal{X}|$ and $m = |\mathcal{Y}|$. The maximization problem associated with the capacity admits a well-known formulation as a constrained optimization program over a standard simplex—see, e.g., Ref. [10]. We now show how to derive this program and, in so doing, we take the chance to define some auxiliary functions and parameters that we will extensively use in this paper.

Define the function $\mathbf{q} = (q_1, \dots, q_m)^\top : \mathbb{R}^n \rightarrow \mathbb{R}^m$, where q_j is the linear function

$$\mathbf{z} = (z_1, \dots, z_n)^\top \mapsto q_j(\mathbf{z}) = \sum_{i=1}^n p(j|i)z_i.$$

If $z_i = p_X(x_i)$ for every $i \in [n]$, then $q_j(\mathbf{z}) = p_Y(y_j)$ for every $j \in [m]$ by the theorem of total probability; thus, \mathbf{z} and $\mathbf{q} = \mathbf{q}(\mathbf{z})$ can be identified with the distributions of X and Y respectively, and it is possible to write $I[X; Y]$ as a function of \mathbf{z} . Indeed, it is sufficient to set

$$c_i = \sum_{j=1}^m p(j|i) \ln[p(j|i)], \quad i \in [n], \tag{4}$$

define the function

$$\mathcal{I}(\mathbf{z}) = \sum_{i=1}^n c_i z_i - \sum_{j=1}^m q_j \ln q_j, \tag{5}$$

and observe that, for $\mathbf{z} \in \Delta_n$ corresponding to p_X , the equality $I[X; Y] = \mathcal{I}(\mathbf{z})$ holds. As a result, the capacity C of the channel \mathcal{C} satisfies

$$C = \max_{\mathbf{z} \in \Delta_n} \mathcal{I}(\mathbf{z}). \tag{6}$$

Following [10], we call *optimal input distribution* every global solution to (6), i.e., every $\mathbf{z}^* \in \Delta_n$ such that $\mathcal{I}(\mathbf{z}^*) = C$.

5. Vector Flow for Capacity Computation

5.1. Flow Definition and Its Properties

The function \mathcal{I} , which is the objective function of (6), is continuous and concave on Δ_n —see, e.g., Ref. [3]. Define the open set

$$\Omega = \{ \mathbf{z} \in \mathbb{R}^n \mid q_j(\mathbf{z}) > 0 \text{ for all } j \in [m] \},$$

and note that $\text{int}(\Delta_n) \subset \Omega$.

Proposition 1. *Let $\mathbf{z} \in \Omega$. Then,*

- (a) $\partial_k \mathcal{I}(\mathbf{z}) = c_k - 1 - \sum_{j=1}^m p(j|k) \ln q_j(\mathbf{z});$
- (b) $\sum_{k=1}^n z_k \partial_k \mathcal{I}(\mathbf{z}) = \mathcal{I}(\mathbf{z}) - \sum_{k=1}^n z_k;$
- (c) $\partial_{i,k}^2 \mathcal{I}(\mathbf{z}) = - \sum_{j=1}^m p(j|i) p(j|k) [q_j(\mathbf{z})]^{-1}.$

Proof. See Appendix F. \square

In particular, $\nabla \mathcal{I}$ is well defined and continuous on Ω , and so $\nabla \mathcal{I}$ is locally Lipschitz continuous on Ω . However, note that \mathcal{I} is singular on $\partial \Omega$.

Now, consider the ODE

$$\dot{z}_i = z_i \left[\partial_i \mathcal{I}(\mathbf{z}) - \sum_{k=1}^n z_k \partial_k \mathcal{I}(\mathbf{z}) \right], \quad i \in [n]. \tag{7}$$

We will also refer to Equation (7) as the capacity computing ODE (CC-ODE for short).

Theorem 1 (Forward invariance). *The set $\Omega \cap \Delta_n$ is forward invariant under (7).*

Proof. Let

$$\mathcal{Y}_0 = \{j \in [m] \mid p(j|i) = 0 \text{ for some } i \in [n]\}.$$

It is easy to see that for every $j \in [m]$ and every $\mathbf{z} \in \Delta_n$:

- If $j \notin \mathcal{Y}_0$, then $q_j(\mathbf{z}) \geq \min_{i \in [n]} p(j|i) > 0$;
- If $j \in \mathcal{Y}_0$, then $q_j(\mathbf{z}) = 0$ if and only if $\text{supp}(\mathbf{z}) \cap \mathcal{S}_j = \emptyset$, where the set $\mathcal{S}_j \subseteq [n]$ is defined by $\mathcal{S}_j = \{i \in [n] \mid p(j|i) \neq 0\}$.

For every $j \in \mathcal{Y}_0$, we define some the following objects:

- The vector $\mathbf{b}^{(j)} = \sum_{i=1}^n p(j|i)\mathbf{e}_i$;
- The function $\beta_j : \Omega \rightarrow \mathbb{R}$ given by

$$\beta_j(\mathbf{z}) = \langle \mathbf{b}^{(j)}, [\text{diag}(\mathbf{z}) - \mathbf{z}\mathbf{z}^\top] \nabla \mathcal{I}(\mathbf{z}) \rangle; \tag{8}$$

- The real number ε_j as a positive solution to the system of inequalities

$$c_i - p(j|i) \ln \varepsilon_j - C > 0 \quad \text{for all } i \in \mathcal{S}_j, \tag{9}$$

where c_i and C are defined as in (4), (6) (note that such an ε_j exists, since $-\ln u \rightarrow +\infty$ as $u \rightarrow 0^+$ and $p(j|i) \neq 0$ for $i \in \mathcal{S}_j$).

We make the following claim:

Claim 3. *Let $\mathbf{z} \in \Omega \cap \Delta_n$ and let $j \in \mathcal{Y}_0$. Then,*

$$0 < q_j(\mathbf{z}) \leq \varepsilon_j \implies \beta_j(\mathbf{z}) \geq 0. \tag{10}$$

Proof of Claim 3. By Proposition 1, the definition of \mathcal{S}_j , and using that $\sum_{i=1}^n z_i = 1$,

$$\begin{aligned} \beta_j(\mathbf{z}) &= \sum_{i=1}^n p(j|i)z_i \left[\partial_i \mathcal{I}(\mathbf{z}) - \sum_{k=1}^n z_k \partial_k \mathcal{I}(\mathbf{z}) \right] \\ &= \sum_{i=1}^n p(j|i)z_i \left\{ c_i - \sum_{J=1}^m p(J|i) \ln[q_J(\mathbf{z})] - \mathcal{I}(\mathbf{z}) \right\} \\ &= \sum_{i \in \mathcal{S}_j} p(j|i)z_i \left\{ c_i - \sum_{J=1}^m p(J|i) \ln[q_J(\mathbf{z})] - \mathcal{I}(\mathbf{z}) \right\}, \end{aligned}$$

and so

$$\beta_j(\mathbf{z}) \geq \sum_{i \in \mathcal{S}_j} p(j|i)z_i \{ c_i - p(j|i) \ln[q_j(\mathbf{z})] - C \}, \tag{11}$$

since $\mathbf{z} \in \Delta_n$ entails both $\mathcal{I}(\mathbf{z}) \leq C$ and $\ln[q_J(\mathbf{z})] \leq 0$ for every J . If $q_j(\mathbf{z}) \leq \varepsilon_j$, then $\ln q_j(\mathbf{z}) \leq \ln \varepsilon_j$, and so, by (9) and (11), it follows that $\beta_j(\mathbf{z}) \geq 0$. \square

Now, let $\mathbf{y} \in \Omega \cap \Delta_n$. We have to prove that a solution $\mathbf{z} : [0, +\infty[\rightarrow \Omega \cap \Delta_n$ of (7) exists satisfying $\mathbf{z}(0) = \mathbf{y}$. We first make the following claim:

Claim 4. *There exists a convex compact $K \subseteq \Omega \cap \Delta_n$ such that $\mathbf{y} \in K$, and K is forward invariant under (7).*

Proof of Claim 4. For every $j \in \mathcal{Y}_0$, set $\alpha_j = \min \{ \varepsilon_j, q_j(\mathbf{y}) \}$ and define

$$K = \{ \mathbf{z} \in \Delta_n \mid q_j(\mathbf{z}) \geq \alpha_j \text{ for every } j \in \mathcal{Y}_0 \}.$$

By construction, $\mathbf{y} \in K \subseteq \Omega \cap \Delta_n$. Note that K is a compact and convex polytope containing the elements $\mathbf{z} \in \mathbb{R}^n$ that satisfy the following constraints:

- $\langle \mathbf{e}_i, \mathbf{z} \rangle \geq 0$ for every $i \in [n]$;
- $\langle \mathbf{1}_n, \mathbf{z} \rangle = 1$;
- $\langle \mathbf{b}^{(j)}, \mathbf{z} \rangle \geq \alpha_j$ for every $j \in \mathcal{Y}_0$.

Therefore, for every $\mathbf{z} \in K$, the *tangent cone* $\text{TK}(\mathbf{z})$ —see Appendix A or, e.g., Ref. [16]—is the set of $\mathbf{u} \in \mathbb{R}^n$ such that

- (a) $u_i \geq 0$ for every $i \notin \text{supp}(\mathbf{z})$;
- (b) $\sum_{i=1}^n u_i = 0$;
- (c) For every $j \in \mathcal{Y}_0$, if $\langle \mathbf{b}^{(j)}, \mathbf{z} \rangle = \alpha_j$ then $\langle \mathbf{b}^{(j)}, \mathbf{u} \rangle \geq 0$.

Note now that $[\text{diag}(\mathbf{z}) - \mathbf{z}\mathbf{z}^\top] \nabla \mathcal{I}(\mathbf{z}) \in \text{TK}(\mathbf{z})$ for every $\mathbf{z} \in K$. In fact, (a) and (b) are immediate to check, whereas (c) holds by Claim 3. The thesis follows by Theorem A1 in Appendix A. \square

By Claim 4, there exists a compact $K \subseteq \Omega \cap \Delta_n$ and a solution $\mathbf{z} : [0, +\infty[\rightarrow K \subseteq \Omega \cap \Delta_n$ of (7) satisfying $\mathbf{z}(0) = \mathbf{y}$. \square

By Theorem 1, we can define a continuous-time dynamical system on $\Omega \cap \Delta_n$ via the (forward) *flow* generated on $\Omega \cap \Delta_n$ by the CC-ODE, i.e., via the function

$$\varphi_{\mathcal{I}} : \Omega \cap \Delta_n \times [0, +\infty[\rightarrow \Omega \cap \Delta_n \tag{12}$$

such that for every $\mathbf{y} \in \Omega \cap \Delta_n$, the function $\mathbf{z} = \varphi_{\mathcal{I}}(\mathbf{y}, \cdot)$ is the unique solution of

$$\begin{cases} \dot{\mathbf{z}} &= [\text{diag}(\mathbf{z}) - \mathbf{z}\mathbf{z}^\top] \nabla \mathcal{I}(\mathbf{z}), \quad t \geq 0 \\ \mathbf{z}(0) &= \mathbf{y}. \end{cases} \tag{13}$$

Proposition 2 (CC-ODE flow properties). *Let $\mathbf{y} \in \Omega \cap \Delta_n$ and set $\mathbf{z} = \varphi_{\mathcal{I}}(\mathbf{y}, \cdot)$.*

- (a) *The equality $\text{supp}(\mathbf{z}(t)) = \text{supp}(\mathbf{y})$ holds for every $t \geq 0$;*
- (b) *Either \mathbf{y} is a stationary point for (7) or $\mathcal{I} \circ \mathbf{z}$ is strictly increasing;*
- (c) *There exists $\mathbf{z}^* = \lim_{t \rightarrow +\infty} \mathbf{z}(t) \in \Omega \cap \Delta_n$, and \mathbf{z}^* is a stationary point for (7);*
- (d) *The restriction of \mathcal{I} to the set $\Gamma = \{ \mathbf{x} \in \Delta_n \mid \text{supp}(\mathbf{x}) \subseteq \text{supp}(\mathbf{y}) \}$ attains its maximum in \mathbf{z}^* :*

$$\mathbf{z}^* \in \arg \max_{\mathbf{z} \in \Gamma} \mathcal{I}(\mathbf{z}); \tag{14}$$

Proof. (a), (b): These properties hold for Shahshahani gradient systems trajectories, and similar proofs are valid in our setting by the uniqueness of local solutions of (7). For more details, see Appendix B.

(c): This follows from Claim 4 and Lemma 1.(i).

(d): Assume without loss of generality that only the first \tilde{n} coordinates of \mathbf{y} are positive, where $0 < \tilde{n} \leq n$. Consider the injection $\iota : \mathbb{R}^{\tilde{n}} \rightarrow \mathbb{R}^n$ given by $\iota(\mathbf{w})^\top = (\mathbf{w}^\top, \mathbf{0}^\top)$, and define the function $\mathcal{I}' = \mathcal{I} \circ \iota$. Observe that $\mathbf{z} = \iota \circ \mathbf{x}$ by uniqueness of solutions, where $\mathbf{x} : [0, +\infty[\rightarrow \Delta_{\tilde{n}}$ solves

$$\begin{cases} \dot{\mathbf{x}} &= [\text{diag}(\mathbf{x}) - \mathbf{x}\mathbf{x}^\top] \nabla \mathcal{I}'(\mathbf{x}), \quad t \geq 0 \\ \mathbf{x}(0) &= \iota^{-1}(\mathbf{y}). \end{cases}$$

Since $\Gamma = \iota(\Delta_{\bar{n}})$ and $\iota^{-1}(\mathbf{y}) \in \text{int}(\Delta_{\bar{n}})$ by construction, the result follows by Lemma 1 applied to $f = \mathcal{I}'$. \square

In particular, Proposition 2 gives the following fundamental theorem:

Theorem 2 (CC-ODE flow-attaining capacity). *Let $\mathbf{y} \in \text{int}(\Delta_n)$. Then $\lim_{t \rightarrow +\infty} \varphi_{\mathcal{I}}(\mathbf{y}, t)$ exists and is an optimal input distribution. Either $\varphi_{\mathcal{I}}(\mathbf{y}, \cdot) \equiv \mathbf{y}$ is a constant function, or $\mathcal{I}(\varphi_{\mathcal{I}}(\mathbf{y}, t))$ is strictly increasing in t and converges to C as $t \rightarrow +\infty$.*

Proof. The proof follows directly from Proposition 2. \square

5.2. Connection with Blahut–Arimoto Algorithm

The BAA is an iterative algorithm that can be described via a map $\Phi_{\text{BA}} : \Delta_n \rightarrow \Delta_n$, here called the *Blahut–Arimoto map*. Given an initial input distribution $\mathbf{z}^{(0)} \in \text{int}(\Delta_n)$, Blahut [6] and Arimoto [5] proved that the sequence $\{\mathbf{z}^{(k)}\}_k$ defined by $\mathbf{z}^{(k+1)} = \Phi_{\text{BA}}(\mathbf{z}^{(k)})$ satisfies $\mathcal{I}(\mathbf{z}^{(k)}) \rightarrow C$ as $k \rightarrow \infty$. It has been observed [8] that Φ_{BA} acts according to the MWU rule [13]:

$$z_i^{(k+1)} = z_i^{(k)} \frac{\exp[w_i(\mathbf{z}^{(k)})]}{\sum_{\ell=1}^n z_{\ell}^{(k)} \exp[w_{\ell}(\mathbf{z}^{(k)})]}, \tag{15}$$

where the weights $w_1(\mathbf{z}), \dots, w_n(\mathbf{z})$ satisfy

$$w_i(\mathbf{z}) = \sum_{j \in [m]: p(j|i) \neq 0} p(j|i) \ln \left(\frac{p(j|i)}{q_j(\mathbf{z})} \right). \tag{16}$$

Interestingly, there exists a numerical scheme used for approximating Shahshahani gradient systems that is also based on MWU, which is the following:

$$z_i^{(k+1)} = z_i^{(k)} \frac{\exp[\tau \partial_i f(\mathbf{z}^{(k)})]}{\sum_{\ell=1}^n z_{\ell}^{(k)} \exp[\tau \partial_{\ell} f(\mathbf{z}^{(k)})]}, \tag{17}$$

where $\tau > 0$ is the stepsize and f must be sufficiently smooth. The recurrence described in (17) is well known in evolutionary game theory—see, e.g., the deduction of the discrete replicator-dynamics model [15]. One fundamental property of (17) is that if $\mathbf{z}^{(k)}$ is an element of Δ_n and the gradient of f is defined in $\mathbf{z}^{(k)}$, then (17) defines $\mathbf{z}^{(k+1)}$ as an element of Δ_n having the same support of $\mathbf{z}^{(k)}$, in agreement with the support invariance of the continuous-time dynamics—see also Appendix B. In particular, if $\mathbf{z}^{(0)} \in \text{int}(\Delta_n)$ and f is differentiable in $\text{int}(\Delta_n)$, then the recurrence given by (17) can be used to produce a sequence $\{\mathbf{z}^{(k)}\}_k$ with arbitrary length (even though, in numerical implementations, the recurrence initialized in $\text{int}(\Delta_n)$ could generate points on $\partial\Delta_n$ due to floating-point arithmetic, see also Ref. [29]).

We can now explain in the following theorem how the CC-ODE flow can be regarded as a continuous-time version of the BAA:

Theorem 3. *The Blahut–Arimoto map coincides with the MWU rule (17) applied to $f = \mathcal{I}$ defined as in (5) with stepsize $\tau = 1$.*

Proof. Observe that

$$\partial_i \mathcal{I}(\mathbf{z}) = -1 + w_i(\mathbf{z}), \tag{18}$$

with $w_i(z)$ defined as in (16); hence, $z^{(k+1)} = \Phi_{BA}(z^{(k)})$ satisfies the following:

$$\begin{aligned} z_i^{(k+1)} &= z_i^{(k)} \frac{\exp[\partial_i \mathcal{I}(z^{(k)}) + 1]}{\sum_{\ell=1}^n z_\ell^{(k)} \exp[\partial_\ell \mathcal{I}(z^{(k)}) + 1]} \\ &= z_i^{(k)} \frac{\exp[\partial_i \mathcal{I}(z^{(k)})]}{\sum_{\ell=1}^n z_\ell^{(k)} \exp[\partial_\ell \mathcal{I}(z^{(k)})]}. \end{aligned}$$

□

6. Convergence Rate

6.1. Conditions for Exponential Convergence

Consider an optimal input distribution z^* . We now present some conditions that ensure that the flow converges exponentially to z^* . To this end, we will study a first-order approximation of

$$v(z) = [\text{diag}(z) - zz^\top] \nabla \mathcal{I}(z) \tag{19}$$

in the vicinity of z^* . In our continuous-time scenario, this corresponds to what is examined in [9], where a truncated Taylor expansion of $\Phi_{BA}(z)$ in z^* is studied to investigate the rate of convergence of the sequence given by $z^{(k+1)} = \Phi_{BA}(z^{(k)})$ to z^* , where $z^{(0)} \in \text{int}(\Delta_n)$. To this end, we consider a classification over $[n]$ introduced in [9] that involves the coordinates of z^* and $\nabla \mathcal{I}(z^*)$. Specifically, an index $i \in [n]$ is classified as follows (note that the case $\partial_i \mathcal{I}(z^*) > C - 1$ is excluded by the KKT conditions):

- type I if $i \in \text{supp}(z^*)$ and $\partial_i \mathcal{I}(z^*) = C - 1$;
- type II if $i \notin \text{supp}(z^*)$ and $\partial_i \mathcal{I}(z^*) = C - 1$;
- type III if $i \notin \text{supp}(z^*)$ and $\partial_i \mathcal{I}(z^*) < C - 1$.

In particular, we remark that the vector $\mu \in \mathbb{R}^n$ given by

$$\mu = (\mu_1, \dots, \mu_n)^T = \sum_{i=1}^n [C - 1 - \partial_i \mathcal{I}(z^*)] e_i \tag{20}$$

satisfies $\mu_i > 0$ if i is of type III, and $\mu_i = 0$ otherwise. As shown in the following proposition, this classification helps in the study of the first-order approximation of (19) near z^* :

Proposition 3. *Let $i \in [n]$. Then,*

$$\nabla v_i(z^*) = \begin{cases} z_i^* [\nabla \partial_i \mathcal{I}(z^*) - \nabla \mathcal{I}(z^*) + \mathbf{1}_n] & \text{if } i \text{ is of type I,} \\ \mathbf{0} & \text{if } i \text{ is of type II,} \\ [\partial_i \mathcal{I}(z) - \mathcal{I}(z) + 1] e_i & \text{if } i \text{ is of type III.} \end{cases}$$

Proof. By Proposition 1,

$$\begin{aligned} \nabla v_i(z) &= \nabla \left\{ z_i \left[\partial_i \mathcal{I}(z) - \mathcal{I}(z) + \sum_{k=1}^n z_k \right] \right\} \\ &= \left[\partial_i \mathcal{I}(z) - \mathcal{I}(z) + \sum_{k=1}^n z_k \right] e_i + z_i [\nabla \partial_i \mathcal{I}(z) - \nabla \mathcal{I}(z) + \mathbf{1}_n], \end{aligned}$$

and note that $\mathcal{I}(z^*) = C$. □

By Proposition 3, if i is of type II, then v_i has a null gradient, and higher-order terms would be required to study v_i in the vicinity of \mathbf{z}^* . Indeed, in order to obtain a non-singular linearization of (7) in a neighborhood of \mathbf{z}^* , we will make the following assumption:

(N1) For some positive $n' \leq n$, the indices $i = 1, \dots, n'$ are of type I and the remaining $n''' = n - n'$ indices are of type III.

Similarly to what is done in [9], it is useful to introduce some additional notation to distinguish between indices of different types. To this end, for every $\mathbf{x} \in \mathbb{R}^n$, define $\mathcal{P}_I \mathbf{x} = (x_1, \dots, x_{n'})^\top \in \mathbb{R}^{n'}$ and $\mathcal{P}_{III} \mathbf{x} = (x_{n'+1}, \dots, x_n)^\top \in \mathbb{R}^{n'''}$ so that $\mathbf{x}^\top = (\mathcal{P}_I \mathbf{x}^\top, \mathcal{P}_{III} \mathbf{x}^\top)$. Note that $(\mathbf{z}^*)^\top = ((\mathcal{P}_I \mathbf{z}^*)^\top, \mathbf{0}^\top)$ and $\boldsymbol{\mu}^\top = (\mathbf{0}^\top, (\mathcal{P}_{III} \boldsymbol{\mu})^\top)$. Moreover, let \mathbf{H} be the Hessian of \mathcal{I} in \mathbf{z}^* , and write the submatrix of \mathbf{H} containing its first n' rows as $[\mathbf{H}_{I,I} | \mathbf{H}_{I,III}]$, where $\mathbf{H}_{I,I} \in \mathbb{R}^{n' \times n'}$ and $\mathbf{H}_{I,III} \in \mathbb{R}^{n' \times n'''}$.

Theorem 4. Assume (N1). Define $\mathbf{Z} = \text{diag}(\mathcal{P}_I \mathbf{z}^*) \in \mathbb{R}^{n' \times n'}$, let $\mathbf{D} = \text{diag}(\mathcal{P}_{III} \boldsymbol{\mu}) \in \mathbb{R}^{n''' \times n'''}$, and let $\mathbf{J} \in \mathbb{R}^{n \times n}$ be the Jacobian matrix of \mathbf{v} in \mathbf{z}^* . Then $\mathbf{J}(\mathbf{z} - \mathbf{z}^*) = \mathbf{M}(\mathbf{z} - \mathbf{z}^*)$ for every $\mathbf{z} \in \Delta_n$, where $\mathbf{M} \in \mathbb{R}^{n \times n}$ is the upper-triangular block matrix

$$\mathbf{M} = \left[\begin{array}{c|c} \mathbf{Z}\mathbf{H}_{I,I} & \mathbf{Z}\mathbf{H}_{I,III} + (\mathcal{P}_I \mathbf{z}^*)(\mathcal{P}_{III} \boldsymbol{\mu})^\top \\ \hline \mathbf{0} & -\mathbf{D} \end{array} \right]. \tag{21}$$

Proof. For every $i \in [n]$, the i -th row of \mathbf{H} is precisely $\nabla v_i(\mathbf{z})^\top$. Apply Proposition 3, using the fact that $\sum_{i=1}^n z_i - z_i^* = 1 - 1 = 0$. \square

Before proving our main result on the convergence rate of the CC-ODE flow, we need the following additional assumption:

(N2) The first n' rows of the transition matrix \mathbf{P} are independent.

Theorem 5 (Exponential convergence rate). Suppose (N1) and (N2) hold. Then the maximum eigenvalue λ_{\max} of \mathbf{M} is negative. Moreover, \mathbf{z}^* is the only optimal input distribution for the channel \mathcal{C} , and for every $0 < \alpha < |\lambda_{\max}|$, there exists $\delta, K > 0$ such that for every $\mathbf{y} \in \Omega \cap \Delta_n$, if $\|\mathbf{y} - \mathbf{z}^*\| < \delta$, then

$$\|\varphi_{\mathcal{I}}(\mathbf{y}, t) - \mathbf{z}^*\| \leq K \|\mathbf{y} - \mathbf{z}^*\| \exp(-\alpha t) \tag{22}$$

for every $t \geq 0$.

Proof. Consider first the matrix \mathbf{M} defined in (21), whose eigenvalues are the union, counting multiplicity, of the eigenvalues of $\mathbf{Z}\mathbf{H}_{I,I}$ and those of $-\mathbf{D}$. The eigenvalues of $-\mathbf{D}$ are $-\mu_{n'+1}, \dots, -\mu_n$, whereas those of $\mathbf{Z}\mathbf{H}_{I,I}$ are described in the following lemma, proved in [9] (Section III.C):

Lemma 2 (Nakagawa et al. [9]). The matrix $\mathbf{Z}\mathbf{H}_{I,I}$ is diagonalizable, has eigenvalues $-1 = \lambda_1 \leq \dots \leq \lambda_{n'} \leq 0$, and $\lambda_{n'} < 0$ if and only if the first n' rows of the transition matrix \mathbf{P} are independent.

By Lemma 2 and Assumption (N2), it follows that $\lambda_{\max} < 0$.

The remaining part of the proof relies on a standard application of Lyapunov’s stability theory (see, e.g., Refs. [17,38]), combined with the application of Theorem 4. Given $0 < \alpha < |\lambda_{\max}|$, the matrix $\tilde{\mathbf{M}} = \mathbf{M} + \alpha \mathbf{I}$ has only real negative eigenvalues by construction. Consequently, the Lyapunov’s equation [17]

$$\tilde{\mathbf{M}}^\top \mathbf{B} + \mathbf{B} \tilde{\mathbf{M}} = -\mathbf{I} \tag{23}$$

is solved by a positive definite symmetric matrix $B \in \mathbb{R}^{n \times n}$. Consider the quadratic form $V(x) = \langle x, Bx \rangle$. Calling the minimum and the maximum eigenvalue of B respectively a and A , it follows that $0 < a < A$ and

$$a\|x\|^2 \leq V(x) \leq A\|x\|^2. \tag{24}$$

Recall that

$$v(z) = J(z - z^*) + r(z - z^*)$$

where J is the Jacobian matrix of v in z^* and

$$\frac{\|r(z - z^*)\|}{\|z - z^*\|} \rightarrow 0 \quad \text{as} \quad z \rightarrow z^*. \tag{25}$$

Therefore, setting $z(t) = \varphi_{\mathcal{I}}(y, t)$ and performing the substitution $\eta(t) = z(t) - z^*$ yields

$$\begin{aligned} \frac{d}{dt}V(\eta) &= 2\langle v(z), B\eta \rangle \\ &= 2\langle \eta, BJ\eta \rangle + 2\langle r(\eta), B\eta \rangle \\ &= 2\langle \eta, BM\eta \rangle + 2\langle r(\eta), B\eta \rangle, \end{aligned}$$

where the last equality is a consequence of Theorem 4. By definition of \tilde{M} , by (23), and using that $B = B^\top$,

$$\begin{aligned} 2\langle \eta, BM\eta \rangle &= 2\langle \eta, B\tilde{M}\eta \rangle - 2\alpha\langle \eta, B\eta \rangle \\ &= \langle \eta, (\tilde{M}^\top B + B\tilde{M})\eta \rangle - 2\alpha V(\eta) \\ &= -\|\eta\|^2 - 2\alpha V(\eta). \end{aligned}$$

Hence,

$$\begin{aligned} \frac{d}{dt}V(\eta) &= -\|\eta\|^2 - 2\alpha V(\eta) + 2\langle r(\eta), B\eta \rangle \\ &= -\left(1 - 2\left\langle \frac{r(\eta)}{\|\eta\|}, \frac{B\eta}{\|\eta\|} \right\rangle\right)\|\eta\|^2 - 2\alpha V(\eta). \end{aligned}$$

Suppose now that for some $T > 0$, the expression

$$1 - 2\left\langle \frac{r(\eta)}{\|\eta\|}, \frac{B\eta}{\|\eta\|} \right\rangle \tag{26}$$

is positive for every $t \in [0, T[$. Then, the inequality $d/dt[V(\eta)] \leq -2\alpha V(\eta)$ holds on $[0, T[$. By Gronwall's lemma [17], this implies that $V(\eta(t)) \leq V(\eta(0)) \exp(-2\alpha t)$, and so by (24)

$$\|\eta(t)\| \leq \sqrt{\frac{A}{a}}\|\eta(0)\| \exp(-\alpha t) \tag{27}$$

for every $t \in [0, T[$. What is left is proving that for $\delta > 0$ sufficiently small, if $\|y - z^*\| = \|\eta(0)\| < \delta$, then (26) is positive for every $t \geq 0$. This follows easily from (25) and (27).

Finally, to prove that z^* is the unique optimal input distribution, suppose by absurd that there exists an optimal input distribution $y^* \neq z^*$. Note that, by convexity of \mathcal{I} , the set of optimal input distributions is a convex subset of Δ_n . Then every convex combination of y^* and z^* is an optimal input distribution. In particular, infinitely many stationary points exist whose distance from z^* does not exceed δ , and this contradicts the definition of δ . \square

Theorem 5 is a continuous-time counterpart of the following result reported in [9]:

Theorem 6 (Nakagawa et al. [9]). *Suppose (N1) and (N2) hold, and that there exists a unique optimal input distribution \mathbf{z}^* for the channel \mathcal{C} . Define ϑ_{\max} as the maximum of the set $\{1 + \lambda_i \mid \lambda_i \in \sigma(\mathbf{ZH})\} \cup \{e^{-\mu_i} \mid i = n' + 1, \dots, n\}$. Then for every $\vartheta_{\max} < \vartheta < 1$, there exists $\delta, K > 0$ such that for every $\mathbf{y} \in \text{int}(\Delta_n)$, if $\|\mathbf{y} - \mathbf{z}^*\| < \delta$, then*

$$\|\Phi_{BA}^{\circ N}(\mathbf{y}) - \mathbf{z}^*\| \leq K\vartheta^N \|\mathbf{y} - \mathbf{z}^*\|. \tag{28}$$

6.2. Noiseless Symmetric Channels

In this section, we refine the result given in Theorem 5 under the additional hypothesis that \mathcal{C} is a noiseless symmetric channel. Following [18] (pp. 77–78), we recall that the channel $\mathcal{C} = (\mathcal{X}, \mathcal{Y}, \mathbf{P})$ is called

- *Deterministic* if the output Y is a deterministic function of the input X .
- *Lossless* if the input X is completely determined by the output Y .
- *Noiseless* if it is both deterministic and lossless.
- *Symmetric* if in the transition matrix \mathbf{P} every row is a permutation of every other row, and every column is a permutation of every other column.

We now assume that \mathcal{C} is noiseless and symmetric. Then $|\mathcal{Y}| = |\mathcal{X}| = n$ and, up to a suitable reordering of the output alphabet, we may assume that the transition matrix is the $n \times n$ identity matrix. Note that $\mathbf{q} = \mathbf{z}$ in this case, and so $\Omega =]0, +\infty[^n$, the objective function is

$$\mathcal{I}(\mathbf{z}) = - \sum_{i=1}^n z_i \ln z_i,$$

and the CC-ODE is

$$\dot{z}_i = -z_i \left(\ln z_i - \mathcal{I}(\mathbf{z}) + 1 - \sum_{k=1}^n z_k \right), \quad i \in [n]. \tag{29}$$

The channel admits a unique optimal input distribution, which is $\mathbf{b} = n^{-1}\mathbf{1}_n$, the barycenter of Δ_n ; thus $C = \mathcal{I}(\mathbf{b}) = \ln n$ —see, e.g., Ref. [18] (pp. 84–85). In addition to that, it is easy to verify an interesting property of noiseless symmetric channels, namely that, for these channels, the BAA requires at most one iteration to find the optimal input distribution. By Theorem 2, we deduce that, for $\mathbf{y} \in \text{int}(\Delta_n) \setminus \{\mathbf{b}\}$, we have $\varphi_{\mathcal{I}}(\mathbf{y}, t) \rightarrow \mathbf{b}$ as $t \rightarrow +\infty$, and also $\varphi_{\mathcal{I}}(\mathbf{b}, t) \equiv \mathbf{b}$. We will see that a more precise asymptotic estimate on $\varphi_{\mathcal{I}}$ is available by leveraging the explicit solution of (29) on $\text{int}(\Delta_n)$.

Theorem 7. *If \mathcal{C} is the noiseless channel, then, for every $\mathbf{y} \in \text{int}(\Delta_n) \setminus \{\mathbf{b}\}$,*

$$\|\varphi_{\mathcal{I}}(\mathbf{y}, t) - \mathbf{b}\| \sim e^{-t} \left[\frac{\text{MSQ}(\ln y_1, \dots, \ln y_n)}{n} \right]^{1/2} \tag{30}$$

as $t \rightarrow +\infty$.

Proof. Given $\mathbf{y} \in \text{int}(\Delta_n)$, the function $\mathbf{z}(t) = \varphi_{\mathcal{I}}(\mathbf{y}, t)$ is given by the following explicit analytical expression:

$$z_i(t) = \frac{\exp(e^{-t} \ln y_i)}{\sum_{k=1}^n \exp(e^{-t} \ln y_k)}. \tag{31}$$

Note that $e^{-t} \rightarrow 0$ as $t \rightarrow +\infty$; hence, using the McLaurin expansion of the exponential function,

$$\begin{aligned} z_i - n^{-1} &= \frac{n \exp(e^{-t} \ln y_i) - \sum_{k=1}^n \exp(e^{-t} \ln y_k)}{n \sum_{k=1}^n \exp(e^{-t} \ln y_k)} \\ &= \frac{n(1 + e^{-t} \ln y_i) - \sum_{k=1}^n (1 + e^{-t} \ln y_k) + o(e^{-t})}{n[\sum_{k=1}^n 1 + o(1)]} \\ &= \frac{n \ln y_i - \sum_{k=1}^n \ln y_k + o(1)}{n^2 + o(1)} e^{-t}. \end{aligned}$$

Therefore,

$$\begin{aligned} n^2 e^{2t} \|z(t) - \mathbf{b}\|^2 &= \sum_{i=1}^n [n e^t (z_i - n^{-1})]^2 \\ &\rightarrow \sum_{i=1}^n \left(\ln y_i - n^{-1} \sum_{k=1}^n \ln y_k \right)^2 = n \text{MSQ}(\ln y_1, \dots, \ln y_n), \end{aligned}$$

which gives (30). \square

7. Analog Implementation

An attractive feature of continuous-time methods is their amenability to be mapped onto hardware circuits [39,40]. Figure 1 depicts a circuit implementing the CC-ODE flow to compute analogically the capacity of a DMC.

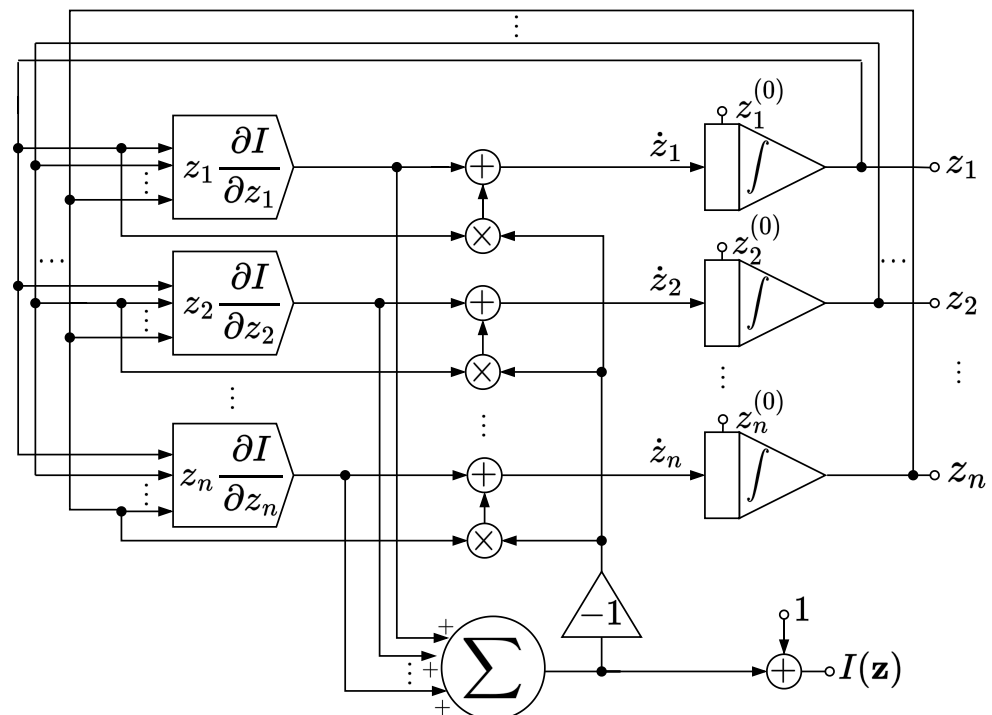


Figure 1. Ideal circuit.

The circuit requires $z_1(0), \dots, z_n(0)$ as input, which represent the coordinates of a point $\mathbf{z}(0) \in \text{int}(\Delta_n)$ in which a trajectory of (7) is initialized. Moreover, the circuit requires, for every $k \in [n]$, a module to implement $\mathbf{z} \mapsto z_k \partial_k \mathcal{I}(\mathbf{z})$, which we here consider as a

black-box. These modules encode the dependence of the circuit on the transition matrix P of the DMC. We remark on the following feature of these modules:

Proposition 4. For every $k \in [n]$, the map $\Omega \cap \Delta_n \ni \mathbf{z} \mapsto z_k \partial_k \mathcal{I}(\mathbf{z})$ can be extended to a continuous function defined on Δ_n .

Proof. Consider the function $L : [0, 1] \rightarrow \mathbb{R}$ given by $L(u) = u \ln u$ for $0 < u \leq 1$ and $L(0) = 0$, which is a continuous function on $[0, 1]$. Then, by Proposition 1, for every $\mathbf{z} \in \Omega \cap \Delta_n$,

$$\begin{aligned} z_k \partial_k \mathcal{I}(\mathbf{z}) &= z_k \left\{ c_k - 1 - \sum_{j=1}^m p(j|k) \ln [q_j(\mathbf{z})] \right\} \\ &= z_k (c_k - 1) - \sum_{j=1}^m \frac{z_k p(j|k)}{q_j(\mathbf{z})} L(q_j(\mathbf{z})), \end{aligned}$$

and since $q_j(\mathbf{z}) = \sum_{i=1}^n z_i p(j|i) \geq z_k p(j|k) \geq 0$, it follows that $\mathbf{z} \mapsto z_k \partial_k \mathcal{I}(\mathbf{z})$ is bounded and admits a continuous extension to Δ_n , which is given by

$$\mathbf{z} \mapsto z_k \left\{ c_k - 1 - \sum_{j \in [m] : q_j(\mathbf{z}) \neq 0} p(j|k) \ln [q_j(\mathbf{z})] \right\}.$$

□

The circuit outputs $\mathbf{z} = (z_1, \dots, z_n)^\top$, which at time $t = 0$ coincides with $\mathbf{z}(0)$, and thanks to the integrator elements appearing in the circuit [40], the recurrent design of the circuit evolves \mathbf{z} according to Volterra’s equation [33]:

$$z_i(t) = z_i(0) + \int_0^t \left[z_i(s) \partial_i \mathcal{I}(\mathbf{z}(s)) - z_i(s) \sum_{k=1}^n z_k(s) \partial_k \mathcal{I}(\mathbf{z}(s)) \right] ds.$$

Hence, $\mathbf{z}(t) = \varphi_{\mathcal{I}}(\mathbf{z}(0), t)$ for every $t \geq 0$. By Theorem 2, it follows that $\mathbf{z}(t)$ converges to an optimal input distribution as $t \rightarrow +\infty$, which is also a stationary point for the system. Moreover, if the channel admits a unique optimal input distribution, then Theorem 2 ensures that this optimal input distribution is also an asymptotically stable stationary point of the system. Besides that, the circuit computes $\sum_{k=1}^n z_k \partial_k \mathcal{I}(\mathbf{z}(t)) + 1$, which equals $\mathcal{I}(\mathbf{z})$ by Proposition 1 and converges to the capacity by Theorem 2.

We stress that the circuit drawn in Figure 1 is supposed to work in an ideal setting, where no noise affects \mathbf{z} . Indeed, the circuit relies on the property that $\Omega \cap \Delta_n$ is invariant under $\varphi_{\mathcal{I}}$, as shown in Theorem 1. By contrast, Figure 2 shows a variation in the circuit design that may mitigate the effect of small perturbations on \mathbf{z} thanks to an additional normalization module. The normalization module sets every negative signal received as input to zero, and then normalizes the resulting vector of non-negative signals with respect to the ℓ^1 -norm.

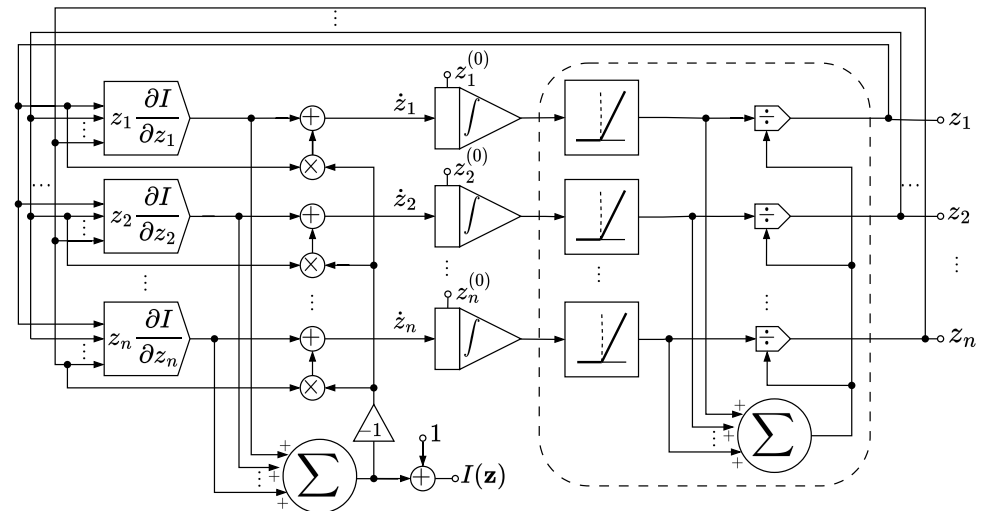


Figure 2. Circuit with normalizing module.

So long as z stays in $\Omega \cap \Delta_n$, as should theoretically happen if $z(0)$ also is in $\Omega \cap \Delta_n$, then the additional normalization module has no effect on the system, and both circuits in Figures 1 and 2 behave in the same way. However, in the presence of a perturbation that affects the system, then the normalization module prevents z from exiting Δ_n . In addition to this, if the modules $z \mapsto z_k \partial_k \mathcal{I}(z)$ are devised to implement the corresponding continuous extension described in Proposition 4, then by Peano’s existence theorem [33], the actual ODE solved by the circuit is well defined on Δ_n , and not just on $\Omega \cap \Delta_n$, even though some trajectories may lead to a suboptimal input distribution, as explained in Section 9.

8. Some Illustrative Examples

To give a graphical demonstration showing the qualitative behavior of the CC-ODE flow, we simulated the CC-ODE for some specific channels. This was accomplished by applying the MWU described in (17) as an integration scheme to discretize the flow. To run the simulations, we considered two channels, and for each channel, we selected a different starting point to initialize the dynamics. For each of these configurations, we computed the (unique) optimal input distribution by running the BAA for 10 000 iterations. In particular, we considered the DMC with transition matrix

$$P^{(1)} = \begin{bmatrix} 0.70 & 0.20 & 0.10 \\ 0.10 & 0.80 & 0.10 \\ 0.25 & 0.25 & 0.50 \end{bmatrix},$$

which admits the unique optimal input distribution $\approx (0.3645, 0.4169, 0.2186)^\top$, and we considered the corresponding CC-ODE flow initialized in $y \approx (0.333, 0.333, 0.333)^\top$. Similarly, we considered the noiseless and symmetric DMC with transition matrix

$$P^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

for which the optimal input distribution is approximately $(0.333, 0.333, 0.333)^\top$, and we considered the corresponding flow initialized in $(0.700, 0.200, 0.100)^\top$. To make a comparison between the CC-ODE vector flow and the BAA, we applied the MWU rule given by (17) with $\tau = 1$ and with $\tau = 0.01$, since the choice $\tau = 1$ yields exactly the Blahut–Arimoto map, whereas smaller values of τ lead to a better approximation of the CC-ODE flow.

Figure 3 shows how the components of the input distribution vary as t runs. Note that, for both channels, the dynamics drive the system towards the optimal input distribution of the corresponding channel. Moreover, for the noiseless symmetric channel, note in Figure 3b that the BAA converges in one step. Figure 4 illustrates how the mutual information of input and output variable evolves. Also in this case, the figure displays that the CC-ODE can be used to attain the capacity, similarly to what happens for the BAA. Clearly, for the noiseless symmetric channel, the capacity is attained via the BAA with just one step—see Figure 4b. Finally, in Figure 5, we depict some graphs plotted in semilog scale to provide some estimates about the speed of convergence of the CC-ODE. For every channel considered and for every integer value of t , we have marked with diamonds in Figure 5 the Euclidean distance between the input distributions obtained via the CC-ODE at time t and the corresponding input distribution computed with the BAA. As shown in the figure, this quantity exhibits exponential decay after just a few iterations of the BAA. Moreover, we reported with a solid red line the Euclidean distance between the limit optimal input distribution and the points on the orbit produced with $\tau = 0.01$. The channels considered have a non-singular transition matrix, and their optimal input distribution does not have null entries. Consequently, Theorem 5 is applicable for both channels. As displayed in Figure 5, the solid red line decays roughly as the dotted grey line, which is the graph of a function of the form $C \exp(\lambda_{\max} t)$, with λ_{\max} being the maximum eigenvalue discussed in Theorem 5, which equals -1 for the noiseless symmetric channel and is ≈ -0.1778 for the other. For the noiseless symmetric channel, we also report in Figure 5b the theoretical asymptotic behavior described in Theorem 7 for the error decay. It is possible to see that in Figure 5b, the computed error decay deviates from the expected behavior for $t \geq 32$. However, the plot also shows that the error stabilizes when it reaches a value close to 10^{-14} . Considering that the simulations were carried out using double precision, we interpret this outcome as a consequence of approaching the machine precision.

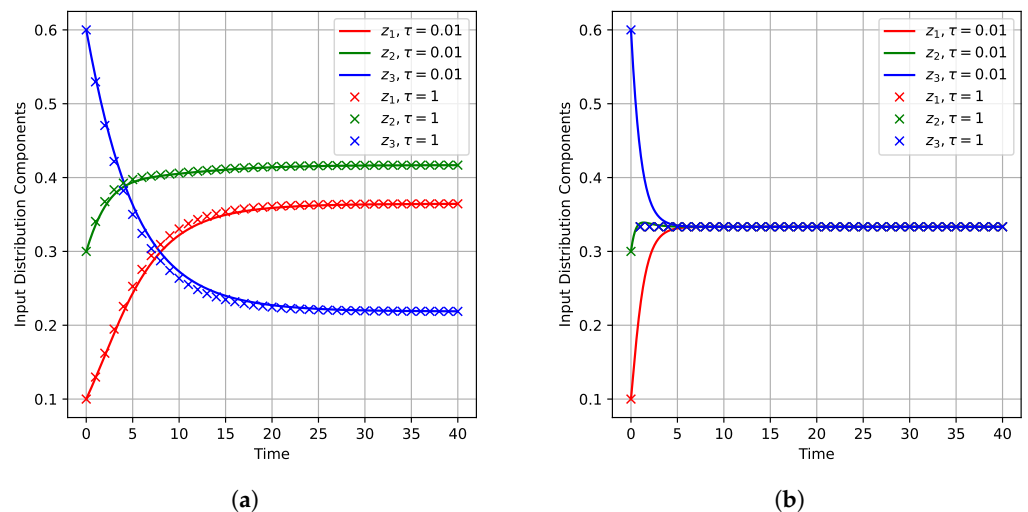


Figure 3. Evolution of CC-ODE and BAA orbits. (a) Channel with transition matrix $P^{(1)}$, systems initialized in $\mathbf{y} = (0.333, 0.333, 0.333)^\top$. (b) Channel with transition matrix $P^{(2)}$, systems initialized in $\mathbf{y} = (0.700, 0.200, 0.100)^\top$.

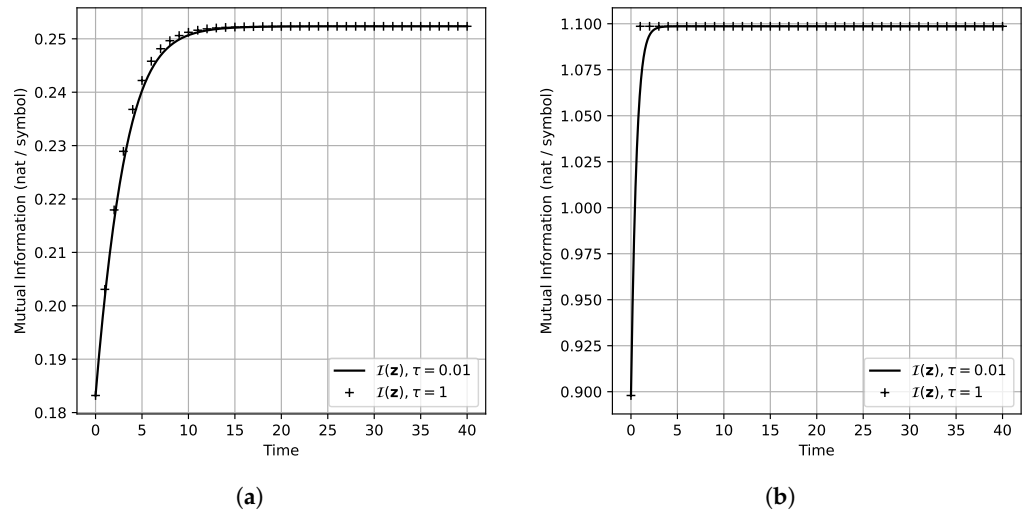


Figure 4. Evolution of $\mathcal{I}(z)$ computed along CC-ODE and BAA orbits. **(a)** Channel with transition matrix $P^{(1)}$, systems initialized in $\mathbf{y} = (0.333, 0.333, 0.333)^\top$. **(b)** Channel with transition matrix $P^{(2)}$, systems initialized in $\mathbf{y} = (0.700, 0.200, 0.100)^\top$.

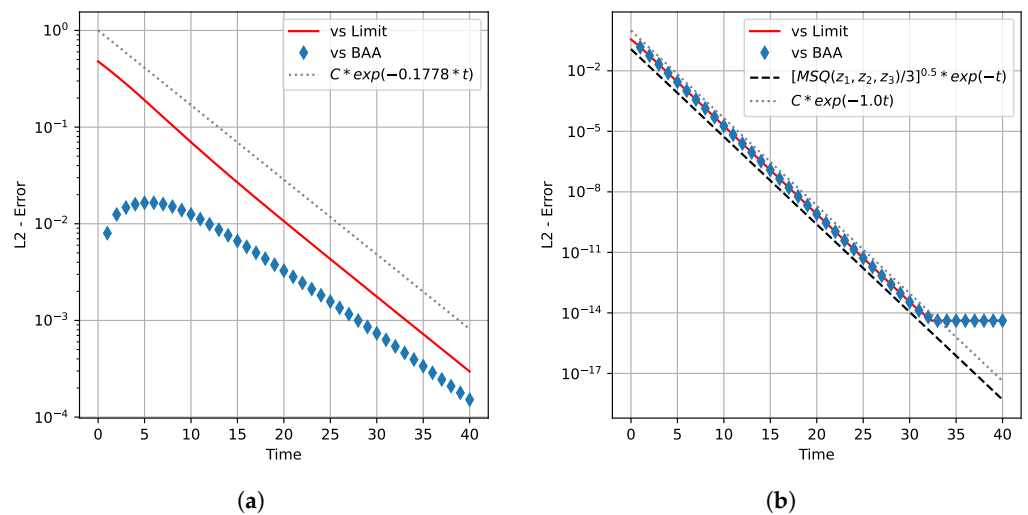


Figure 5. Speed of convergence of the CC-ODE: **(a)** Channel with transition matrix $P^{(1)}$, systems initialized in $\mathbf{y} = (0.333, 0.333, 0.333)^\top$. **(b)** Channel with transition matrix $P^{(2)}$, systems initialized in $\mathbf{y} = (0.700, 0.200, 0.100)^\top$. Diamonds are used for comparison versus the BAA and the solid red line reports the distance of z with the limit point of the dynamics. The dotted grey lines provide estimates on the minimal limiting steepness of the red line, as a consequence of Theorem 5. The dashed black line in **(b)** gives the theoretical asymptotic behavior described in Theorem 7.

The interested reader may also see Ref. [29], where other integration schemes were tested in conjunction with Algorithm 1.

Algorithm 1 Discretizing the CC-ODE to compute the capacity [29].

Input: τ , stepsize; $Niter$, maximum number of iterations; ε , tolerance; \mathbf{y} , initialization point.

Output: \hat{C} , estimated channel capacity; $\hat{\mathbf{z}}$, estimated optimal input distribution; k , number of integration steps performed.

```

1:  $k \leftarrow 0$ 
2:  $\mathbf{z} \leftarrow \mathbf{y}$  ▷ Initializing  $\mathbf{z}$ .
3:  $err \leftarrow +\infty$ 
4: while  $k < Niter \wedge err \geq \varepsilon$  do
5:    $\mathbf{z}_{new} \leftarrow SolverStep(\mathbf{z}, \tau)$ 
6:    $err \leftarrow \|\mathbf{z} - \mathbf{z}_{new}\|_1$  ▷ Note:  $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$  is the  $\ell^1$ -norm of  $\mathbf{x}$ .
7:    $\mathbf{z} \leftarrow \mathbf{z}_{new}$  ▷ Updating  $\mathbf{z}$ .
8:    $k \leftarrow k + 1$ 
9: end while
10:  $\hat{\mathbf{z}} \leftarrow \mathbf{z}$ 
11:  $\hat{C} \leftarrow \mathcal{I}(\hat{\mathbf{z}})$ 
12: return  $\hat{C}, \hat{\mathbf{z}}, k$ 

```

9. Discussion

We report in this section some additional remarks related to the results treated so far:

- In formulating Lemma 1, we tried to require only the essential hypotheses we used in its proof. Consequently, even though we apply Lemma 1 only in Proposition 2, it is applicable also in other, more general settings.
- In the proof of Theorem 1, note that if $\mathcal{Y}_0 = \emptyset$, then $\Delta_n \subset \Omega$, and the thesis is just a mere application of classical results—see, e.g., Refs. [16,31]. By contrast, the proof deviates from a standard setting in case $\mathcal{Y}_0 \neq \emptyset$, which implies $\partial\Delta_n \not\subset \Omega$.
- As far as the proposed circuits are concerned, note that $z_k \partial_k \mathcal{I}(\mathbf{z})$ could be produced by considering a module that outputs $\partial_k \mathcal{I}(\mathbf{z})$ and whose output is then multiplied by z_k . However, note that $\mathbf{z} \mapsto z_k \partial_k \mathcal{I}(\mathbf{z})$ is a bounded function of the input by Proposition 4, which in general is not true for $\mathbf{z} \mapsto \partial_k \mathcal{I}(\mathbf{z})$.
- We also remark on the following feature of the circuit presented in Figure 1. Assume that, for whatever reason, there are u input symbols $x_{i_1}, x_{i_2}, \dots, x_{i_u}$ that become unavailable, in the sense that any input distribution is now constrained to assign null probability to $x_{i_1}, x_{i_2}, \dots, x_{i_u}$. Indeed, this constraint on the channel $\mathcal{C} = (\mathcal{X}, \mathcal{Y}, \mathbf{P})$ corresponds to replacing it with $\mathcal{C}' = (\mathcal{X}', \mathcal{Y}, \mathbf{P}')$, where $\mathcal{X}' = \mathcal{X} \setminus \{x_{i_k} : k \in [u]\}$ and \mathbf{P}' is obtained from \mathbf{P} by removing the rows i_1, i_2, \dots, i_u . In case every symbol of the output alphabet \mathcal{Y} can still be obtained with positive probability, then the capacity of \mathcal{C}' may be computed without altering the circuit design displayed for \mathcal{C} , but by simply modifying the initialization rules on \mathbf{z} . Specifically, by Proposition 2.(d), it is sufficient to initialize $\mathbf{z}(0) \in \Delta_n$ so that $z_i(0) = 0$ if and only if $i = i_1, i_2, \dots, i_u$, and the dynamics evolves \mathbf{z} so as to maximize $\mathcal{I}(\mathbf{z})$ under this constraint;
- Note that noise could negatively affect the capacity computation, despite the normalization module of Figure 2. Indeed, suppose that a perturbation causes \mathbf{z} to change its support at time $t = t_1$, and that for $t > t_1$, no additional perturbations affect the dynamics. Then it is possible that \mathbf{z} converges towards an input distribution that is still optimal, but for the wrong channel, as in the case discussed in the previous remark. To check if this has occurred, a good idea could be to consider some “perturb and restart” strategies. For instance, when the capacity of \mathcal{C} is sought in the presence of noise, assume that a trajectory initialized in some $\mathbf{z}(0) \in \text{int}(\Delta_n)$ converges to some $\mathbf{z}^* \in \partial\Delta_n$. We may then restart the dynamics in some $\mathbf{z} \in \text{int}(\Delta_n)$ that is obtained by perturbing slightly \mathbf{z}^* , and then check whether $\mathcal{I}(\mathbf{z})$ converges once again to $\mathcal{I}(\mathbf{z}^*)$, or if \mathbf{z} converges to \mathbf{z}^* in case we know in advance that there exists a unique optimal input distribution.

- By Proposition 4 and Peano's theorem, the ODE (7) extended by continuity admits a solution for any initialization $z(0) \in \Delta_n$, regardless of whether $z(0)$ lies in Ω or not. Indeed, a solution for (7) can always be found by considering the "subchannel" of \mathcal{C} induced by the support of $z(0)$, similarly to the approach discussed before in Proposition 2.(d). However, in case $z(0) \notin \Omega$, note that the Picard–Lindelöf theorem cannot be applied, and we did not manage to prove that such a solution is unique. Despite this, for symmetric noiseless channels, note that the general integral (31) can be extended by continuity also for $z(0) \in \partial\Delta_n$, and arguing by induction it is not hard to see that, for these trivial channels, solutions for the extended ODE are unique.

10. Conclusions

We performed a theoretical analysis on the flow of the CC-ODE, which rules a continuous-time dynamical system that enables us to compute the capacity, as well as an optimal input distribution, of a DMC. We showed that the proposed dynamical system can be regarded as a continuous-time version of the BAA and that, under some technical conditions, the flow rate of convergence to an optimal input distribution is exponential, with constants that correspond to those arising from iterating the Blahut–Arimoto map. We described possible implementations of the CC-ODE flow in a circuit that may compute analogically the capacity of a DMC, and we discussed how the circuit may be still useful in case the channel changes due to the unavailability of some input symbols.

Possible future work stemming from this paper might be the exploration of continuous-time dynamics to maximize mutual information under some additional constraints, as in [6], or for quantum channels [41]. Moreover, an interesting aspect of Theorem 3 is that it shows that the CC-ODE flow is related to some accelerated versions of the BAA [8] that correspond to an application of (17) with $f = \mathcal{I}$ but where the stepsize is adjusted at every iteration. This link suggests investigating strategies to speed up analog computation by accelerating the CC-ODE flow. Lastly, a theoretical question that we have not addressed is whether solutions of the ODE obtained by extending (7) by continuity on Δ_n are unique, even though the hypotheses of the Picard–Lindelöf theorem are not met.

As mentioned by M. T. Chu in [42], continuous methods may help understand the corresponding discrete methods, and we hope that this paper may enrich our collective knowledge of the BAA, which is still a subject of active research.

Author Contributions: Conceptualization, G.B. and M.P.; methodology, G.B.; formal analysis, G.B.; investigation, G.B.; resources, M.P.; writing—original draft preparation, G.B.; writing—review and editing, G.B. and M.P.; visualization, G.B.; supervision, M.P., funding acquisition, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the support and the funding received from Politecnico di Torino and Università Ca' Foscari di Venezia.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors thank Giacomo Chiarot and Antonio Emanuele Cinà for the support given during the preliminary stage of this work, and the anonymous reviewers for their helpful suggestions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. ODEs on Compact Convex Sets

Let $C \subset \mathbb{R}^n$ be a closed convex set. Following Ref. [16], we recall that, for every $z \in C$, the *tangent cone* of C at z is the closed convex cone

$$TC(z) = \text{cl}(\{ \alpha(x - z) \mid x \in C, \alpha \in \mathbb{R}_+ \})$$

where $\text{cl}(\cdot)$ denotes the closure in the usual Euclidean topology of \mathbb{R}^n . We report a technical result that deals with the solutions of ODEs over C in case it is also compact.

Theorem A1 (ODEs on compact convex sets [16] (Theorem 4.A.5.(i))). *Let $C \subset \mathbb{R}^n$ be a compact convex set, and let $v : C \rightarrow \mathbb{R}^n$ be Lipschitz continuous. Suppose that $v(\hat{x}) \in TC(\hat{x})$ for all $\hat{x} \in C$. Then, for each $\zeta \in C$, there exists a unique $x : [0, +\infty[\rightarrow C$ satisfying $\dot{x} = v(x)$ and $x(0) = \zeta$.*

Appendix B. Support Invariance and Lyapunov Function

The following proposition reports some useful properties that are known for f admitting a Lipschitz continuous gradient on Δ_n . For completeness, we provide the details that extend the known proof to our more general setting.

Proposition A1. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ admit a locally Lipschitz continuous gradient on some open $\Omega \subseteq \mathbb{R}^n$. Let $a, b \in \mathbb{R}$, with $a \leq b$, and let the function $z : [a, b] \rightarrow \Omega$ satisfy $z(t_0) \in \Delta_n$ for some $t_0 \in [a, b]$ and*

$$\dot{z} = [\text{diag}(z) - zz^\top] \nabla f(z), \quad a \leq t \leq b.$$

- (i) *We have $z(t) \in \Delta_n$ for $a \leq t \leq b$;*
- (ii) *The support of $z(t)$ is invariant for $a \leq t \leq b$;*
- (iii) *Either z is a constant function or $f \circ z$ is strictly increasing.*

Proof. (i), (ii): It is sufficient to show that the function $\sigma(t) = \sum_{i=1}^n z_i(t)$ satisfies $\sigma(t) = 1$ for every $a \leq t \leq b$, and that for every $i \in [n]$ either $z_i \equiv 0$ or $z_i(t) > 0$ for every $a \leq t \leq b$. Consider first the function

$$v(t, s) = (1 - s) \left[\sum_{i=1}^n z_i(t) \partial_i f(z(t)) \right],$$

which is defined for $(t, s) \in [a, b] \times \mathbb{R}$ and Lipschitz continuous in s uniformly with respect to t —see, e.g., Ref. [34]. The ODE $\dot{s} = v(t, s)$ is solved by σ and admits the constant solution $s \equiv 1$. Since by hypothesis $\sigma(t_0) = 1 = s(1)$, then $\sigma(t) = s(t) = 1$ for every $a \leq t \leq b$ by the Picard-Lindelöf theorem. Similarly, for every $i \in [n]$, consider the function

$$u_i(t, x) = x \left[\partial_i f(z(t)) - \sum_{k=1}^n z_k(t) \partial_k f(z(t)) \right],$$

which is defined for $(t, x) \in [a, b] \times \mathbb{R}$ and Lipschitz continuous in x uniformly with respect to t . The ODE $\dot{x} = u_i(t, x)$ is solved by $z_i(t)$ and admits the constant solution $x \equiv 0$. Consequently, since by hypothesis $z(t_0) \geq 0$, by the Picard-Lindelöf theorem either $x_i \equiv 0$ or $x_i(t) > 0$ for every $a \leq t \leq b$.

(iii): Set $\mu(x) = \sum_{i=1}^n x_i \partial_i f(x)$ for every $x \in \Omega$. Note that

$$\frac{d}{dt} (f \circ z) = \sum_{i=1}^n \partial_i f(z) \dot{z}_i = \sum_{i=1}^n z_i [\partial_i f(z) - \mu(z)]^2,$$

and so

- The function $f \circ z$ has non-negative derivative, and is thereby non-decreasing.
- If for some t_1 the equality $d/dt(f \circ z)(t_1) = 0$ holds, then $\partial_i f(z(t_1)) - \mu(z(t_1)) = 0$ for every $i \notin \text{supp}(z(t_1))$, and it is easy to check that in this case $\dot{z}(t_1) = \mathbf{0}$ —thus, $z \equiv z(t_1)$ by the Picard-Lindelöf theorem.

□

Appendix C. An Example

Set $z^* = (1/2, 1/2, 0)^\top$ and choose some $y \in \text{int}(\Delta_3)$. Consider first the concave function $f_1 \in C^\infty(\mathbb{R}^3)$ defined by $f_0(z) = -\|z - z^*\|^2$. It is immediate to compute the Shahshahani gradient system associated with f_0 :

$$\dot{z} = -2[\text{diag}(z) - zz^\top](z - z^*). \tag{A1}$$

Let $z_0 : [0, +\infty[\rightarrow \Delta_3$ solve (A1) and suppose $z_0(0) = y$. Then $z_0 \rightarrow z^*$ as $t \rightarrow +\infty$, and z^* is a stationary point for (A1), as well as the global maximum of f over Δ_3 . Consider now the concave function f_1 defined on \mathbb{R}^3 by $f_1(z) = -\|z - z^*\|$, and the associated Shahshahani gradient system:

$$\dot{z} = -|f_1(z)|^{-1}[\text{diag}(z) - zz^\top](z - z^*). \tag{A2}$$

Let $z_1 : [0, +\infty[\rightarrow \Delta_3$ solve (A2) and suppose $z_1(0) = y$. By uniqueness of solutions, $z_1(t) = (z_0 \circ \tau)(t)$, with $\tau(t) = \int_0^t |2f_1(z_0(r))|^{-1} dr$. Consequently, $z^* = \lim_{t \rightarrow +\infty} z_1(t)$. However, ∇f_1 is undefined in z^* , and so (A2) is undefined in $z = z^*$; thus, z^* is *not* a stationary point for (A2).

Appendix D. Proof of Claim 1

The proof of Claim 1 here given adapts an argument used in [16] (Theorem 7.2.4), where a similar result is proved for a more stringent setting, namely, for f strictly convex and smooth on Δ_n .

Proof. Assume that $z(t)$ is not a constant function, since otherwise the proof is trivial. Call Λ the set of limit points of $z(t)$ as $t \rightarrow +\infty$; then note, by compactness of K , that $\emptyset \neq \Lambda \subseteq K$, and fix some $z^* \in \Lambda$. On $\mathcal{D}_{z^*} = \{x \in \mathbb{R}^n \mid \text{supp}(x) \supseteq \text{supp}(z^*)\}$, consider the function

$$h(x) = \sum_{i \in \text{supp}(z^*)} z_i^* \ln \frac{z_i^*}{x_i}. \tag{A3}$$

Additional properties are known for $h(x)$ if we constrain x in $\mathcal{D}_{z^*} \cap \Delta_n$, since in this case $h(x) = D(z^*||x)$, where the right-hand side denotes the relative entropy between z^* and x —see, e.g., Ref. [3]. Specifically, we are interested in the following:

- $h(x) \geq 0$, with equality only for $x = z^*$.
- $x \rightarrow z^*$ if and only if $h(x) \rightarrow 0$.

Note also that $z(t) \in \mathcal{D}_{z^*} \cap \Delta_n$ for every t by Proposition A1. Therefore,

- We have $h \circ z \geq 0$, and by definition of z^* ,

$$\liminf_{t \rightarrow +\infty} (h \circ z) = 0; \tag{A4}$$

- The proof is concluded if we also show that

$$\lim_{t \rightarrow +\infty} (h \circ z) = 0. \tag{A5}$$

By Proposition A1, the function $f \circ z$ is strictly increasing; hence, by continuity of f ,

$$f(z^*) = \sup_{t \geq 0} (f \circ z)(t) = \lim_{t \rightarrow +\infty} (f \circ z)(t). \tag{A6}$$

Moreover, we can now prove that

$$\frac{d}{dt}(h \circ z) \leq 0. \tag{A7}$$

In fact, using that $\sum_i z_i^* = 1$, we obtain

$$\begin{aligned} \frac{d}{dt}(h \circ z) &= - \sum_{i \in \text{supp}(z^*)}^n \frac{z_i^* \dot{z}_i}{z_i} \\ &= - \sum_{i=1}^n z_i^* \left[\partial_i f(z) - \sum_{k=1}^n z_k \partial_k f(z) \right] \\ &= \sum_{k=1}^n z_k \partial_k f(z) - \sum_{i=1}^n z_i^* \partial_i f(z) \\ &= \sum_{i=1}^n (z_i - z_i^*) \partial_i f(z) \\ &= - \langle \nabla f(z), z^* - z \rangle \\ &\leq f(z) - f(z^*). \end{aligned}$$

Then (A7) follows using that

$$f(z^*) \leq f(z) + \langle \nabla f(z), z^* - z \rangle, \tag{A8}$$

which is a consequence of the concavity of f on K —see, e.g., Ref. [37]. By (A7), the non-negative function $h \circ z$ is non-increasing, and since we know that (A4) holds, then this yields (A5). \square

Appendix E. Proof of Claim 2

Proof. The Lagrangian associated with (2) is the function $\mathcal{L} : \Delta_n \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$\mathcal{L}(x, \mu_0, \mu) = f(x) + \mu_0 (\langle \mathbf{1}_n, x \rangle - 1) + \langle \mu, x \rangle,$$

and the KKT conditions for x^* amount to the following system (see, e.g., Ref. [43]):

$$\begin{cases} (\nabla_x \mathcal{L})(x^*, \mu_0, \mu) = \nabla f(x^*) + \mu_0 \mathbf{1}_n + \mu = \mathbf{0}, & \text{(A9)} \\ \mu_i x_i^* = 0, \quad i \in [n], & \text{(A10)} \\ \mu \geq \mathbf{0}. & \text{(A11)} \end{cases}$$

Note that (A9) is equivalent to $\mu = -(\nabla f(x^*) + \mu_0 \mathbf{1}_n)$, which can be used to eliminate μ in (A10) and (A11), yielding

$$\begin{cases} (\partial_i f(x^*) + \mu_0) x_i^* = 0, \quad i \in [n], \\ \nabla f(x^*) + \mu_0 \mathbf{1}_n \leq \mathbf{0}, \end{cases}$$

i.e.,

$$\begin{cases} \partial_i f(x^*) = -\mu_0, \quad i \in \text{supp}(x^*), \\ \partial_i f(x^*) \leq -\mu_0, \quad i \in [n]. \end{cases}$$

The thesis then follows by considering $\lambda = -\mu_0$. \square

Appendix F. Proof of Proposition 1

Proof. (a): Let $k \in [n]$ and observe that

$$\begin{aligned}\partial_k \mathcal{I}(\mathbf{z}) &= c_k - \sum_{j=1}^m \partial_k (q_j \ln q_j)(\mathbf{z}) \\ &= c_k - \sum_{j=1}^m \{1 + \ln[q_j(\mathbf{z})]\} \partial_k q_j(\mathbf{z}) \\ &= c_k - \sum_{j=1}^m p(j|k) - \sum_{j=1}^m p(j|k) \ln q_j(\mathbf{z}) \\ &= c_k - 1 - \sum_{j=1}^m p(j|k) \ln q_j(\mathbf{z}).\end{aligned}$$

(b): Apply (a) to obtain

$$\begin{aligned}\sum_{k=1}^n z_k \partial_k \mathcal{I}(\mathbf{z}) &= \sum_{k=1}^n z_k \left[c_k - 1 - \sum_{j=1}^m p(j|k) \ln q_j \right] \\ &= \sum_{k=1}^n c_k z_k - \sum_{k=1}^n z_k - \sum_{j=1}^m q_j \ln q_j \\ &= \mathcal{I}(\mathbf{z}) - \sum_{k=1}^n z_k.\end{aligned}$$

(c): By (a),

$$\partial_{i,k}^2 \mathcal{I}(\mathbf{z}) = - \sum_{j=1}^m \frac{p(j|k)}{q_j(\mathbf{z})} \partial_i q_j(\mathbf{z}) = - \sum_{j=1}^m \frac{p(j|i)p(j|k)}{q_j(\mathbf{z})}.$$

\square

References

- Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
- MacKay, D.J.C. *Information Theory, Inference, and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
- Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
- Csiszár, I.; Tusnády, G. Information geometry and alternating minimization procedures. *Stat. Decis.* **1984**, *1*, 205–237.
- Arimoto, S. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inf. Theory* **1972**, *18*, 14–20. [\[CrossRef\]](#)
- Blahut, R. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theory* **1972**, *18*, 460–473. [\[CrossRef\]](#)
- Yu, Y. Squeezing the Arimoto-Blahut algorithm for faster convergence. *IEEE Trans. Inf. Theory* **2010**, *56*, 3149–3157. [\[CrossRef\]](#)
- Matz, G.; Duhamel, P. Information geometric formulation and interpretation of accelerated Blahut-Arimoto-type algorithms. In *2004 IEEE Information Theory Workshop*; IEEE Press: Piscataway, NJ, USA, 2004; pp. 66–70. [\[CrossRef\]](#)
- Nakagawa, K.; Takei, Y.; Hara, S.I.; Watabe, K. Analysis of the convergence speed of the Arimoto-Blahut algorithm by the second-order recurrence formula. *IEEE Trans. Inf. Theory* **2021**, *67*, 6810–6831. [\[CrossRef\]](#)
- Boche, H.; Schaefer, R.F.; Poor, H.V. Algorithmic computability and approximability of capacity-achieving input distributions. *IEEE Trans. Inf. Theory* **2023**, *69*, 5449–5462. [\[CrossRef\]](#)
- Helmke, U.; Moore, J.B. *Optimization and Dynamical Systems*; Springer: London, UK, 1994.
- Brockett, R.W. Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems. *Linear Algebra Its Appl.* **1991**, *146*, 79–91. [\[CrossRef\]](#)
- Arora, S.; Hazan, E.; Kale, S. The multiplicative weights update method. *Theory Comput.* **2012**, *8*, 121–164. [\[CrossRef\]](#)

14. Naja, Z.; Alberge, F.; Duhamel, P. Geometrical interpretation and improvements of the Blahut-Arimoto's algorithm. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 2505–2508. [[CrossRef](#)]
15. Weibull, J. *Evolutionary Game Theory*; MIT Press: Cambridge, MA, USA, 1995.
16. Sandholm, W.H. *Population Games and Evolutionary Dynamics*; MIT Press: Cambridge, MA, USA, 2010.
17. Khalil, H.K. *Nonlinear Systems*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002.
18. Roman, S. *Coding and Information Theory*; Graduate texts in mathematics; Springer: New York, NY, USA, 1992; Volume 134.
19. MacLennan, B.J. Analog Computation. In *Encyclopedia of Complexity and Systems Science*; Meyers, R.A. Ed.; Springer: New York, NY, USA, 2009; pp. 271–294. [[CrossRef](#)]
20. Taherkhani, A.; Belatreche, A.; Li, Y.; Cosma, G.; Maguire, L.P.; McGinnity, T.M. A review of learning in biologically plausible spiking neural networks. *Neural Netw.* **2020**, *122*, 253–272. [[CrossRef](#)]
21. Wang, Q.; Meng, C.; Wang, C. Analog continuous-time filter designing for Morlet wavelet transform using constrained L2-norm approximation. *IEEE Access* **2020**, *8*, 121955–121968. [[CrossRef](#)]
22. Schuman, C.D.; Kulkarni, S.R.; Parsa, M.; Mitchell, J.P.; Date, P.; Kay, B. Opportunities for neuromorphic computing algorithms and applications. *Nat. Comput. Sci.* **2022**, *2*, 10–19. [[CrossRef](#)] [[PubMed](#)]
23. Edwards, J.; Parker, L.; Cardwell, S.G.; Chance, F.S.; Koziol, S. Neural-inspired dendritic multiplication using a reconfigurable analog integrated circuit. In Proceedings of the 2024 IEEE International Symposium on Circuits and Systems (ISCAS), Singapore, 19–22 May 2024. [[CrossRef](#)]
24. Hasler, J. Energy-efficient programable analog computing. *IEEE Solid-State Circuits Mag.* **2024**, *16*, 32–40. [[CrossRef](#)]
25. Hasler, J.; Hao, C. Programmable analog system benchmarks leading to efficient analog computation synthesis. *ACM Trans. Reconfig. Technol. Syst.* **2024**, *17*, 1–25. [[CrossRef](#)]
26. Parker, L.; Cardwell, S.G.; Chance, F.S.; Koziol, S. Bio-inspired active silicon dendrite for direction selectivity. In Proceedings of the 2024 International Conference on Neuromorphic Systems (ICONS), Arlington, VA, USA, 30 July–2 August 2024; pp. 343–349. [[CrossRef](#)]
27. Ulmann, B. Beyond zeros and ones— analog computing in the twenty-first century. *Int. J. Parallel Emergent Distrib. Syst.* **2024**, *39*, 139–151. [[CrossRef](#)]
28. Torsello, A.; Pelillo, M. Continuous-time relaxation labeling processes. *Pattern Recognit.* **2000**, *33*, 1897–1908. [[CrossRef](#)]
29. Beretta, G.; Chiarot, G.; Cinà, A.E.; Pelillo, M. Computing the capacity of discrete channels using vector flows. In *Dynamics of Information Systems: 7th International Conference, DIS 2024, Kalamata, Greece, 2–7 June 2024, Revised Selected Papers*; Moosaei, H., Kotsireas, I., Pardalos, P.M., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 14661, pp. 119–128. [[CrossRef](#)]
30. Faybusovich, L. Dynamical systems which solve optimization problems with linear constraints. *IMA J. Math. Control Inf.* **1991**, *8*, 135–149. [[CrossRef](#)]
31. Alvarez, F.; Bolte, J.; Brahic, O. Hessian Riemannian gradient flows in convex programming. *SIAM J. Control Optim.* **2004**, *43*, 477–501. [[CrossRef](#)]
32. Hofbauer, J.; Sigmund, K. *Evolutionary Games and Population Dynamics*; Cambridge University Press: Cambridge, UK, 1998.
33. Birkhoff, G.; Rota, G.C. *Ordinary Differential Equations*; Wiley: Hoboken, NJ, USA, 1989.
34. Teschl, G. *Ordinary Differential Equations and Dynamical Systems*; American Mathematical Society: Providence, RI, USA, 2024.
35. Hirsch, M.W.; Smale, S.; Devaney, R.L. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, 3rd ed.; Academic Press: Waltham, MA, USA, 2013.
36. Mertikopoulos, P.; Sandholm, W.H. Riemannian game dynamics. *J. Econ. Theory* **2018**, *177*, 315–364. [[CrossRef](#)]
37. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.
38. Boyd, S.P.; El Ghaoui, L.; Feron, E.; Balakrishnan, V. *Linear Matrix Inequalities in System and Control Theory*; SIAM Studies in Applied Mathematics; SIAM: Philadelphia, PA, USA, 1994; Volume 15. [[CrossRef](#)]
39. Hopfield, J.J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 3088–3092. [[CrossRef](#)]
40. Cichocki, A.; Unbehauen, R. *Neural Networks for Optimization and Signal Processing*; Wiley: Hoboken, NJ, USA, 1993.
41. He, K.; Saunderson, J.; Fawzi, H. A Bregman proximal perspective on classical and quantum Blahut-Arimoto algorithms. *IEEE Trans. Inf. Theory* **2024**, *70*, 5710–5730. [[CrossRef](#)]
42. Chu, M.T. On the continuous realization of iterative processes. *SIAM Rev.* **1988**, *30*, 375–387. [[CrossRef](#)]
43. Luenberger, D.G.; Ye, Y. *Linear and Nonlinear Programming*; Springer: Cham, Switzerland, 2016.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.