

Article

Information Processing in the Brain as Optimal Entropy Transport: A Theoretical Approach

Carlos Islas ^{1,†}, Pablo Padilla ^{2,*,†} and Marco Antonio Prado ^{1,2,†}

¹ Universidad Autónoma de la Ciudad de México, Doctor García Diego núm. 168, Cuauhtémoc, Ciudad de México 06720, Mexico; carlos.islas@uacm.edu.mx (C.I.); marcoantonio.prado@uacm.edu.mx (M.A.P.)

² IIMAS, Universidad Nacional Autónoma de México Circuito Escolar, Cd. Universitaria, Ciudad de México 04510, Mexico

* Correspondence: pablo@uxmym1.iimas.unam.mx

† These authors contributed equally to this work.

Received: 9 September 2020; Accepted: 15 October 2020; Published: 29 October 2020



Abstract: We consider brain activity from an information theoretic perspective. We analyze the information processing in the brain, considering the optimality of Shannon entropy transport using the Monge–Kantorovich framework. It is proposed that some of these processes satisfy an optimal transport of informational entropy condition. This optimality condition allows us to derive an equation of the Monge–Ampère type for the information flow that accounts for the branching structure of neurons via the linearization of this equation. Based on this fact, we discuss a version of Murray’s law in this context.

Keywords: informational entropy; neuroscience; Monge–Ampère equation; optimal transport; variational calculus; Murray’s law; neuronal branching structures

1. Introduction

The brain as the organ responsible for processing information in the body has been subjected to evolutionary pressure. “The business of the brain is computation, and the task it faces is monumental. It must take sensory inputs from the external world, translate this information into a computationally accessible form, then use the results to make decisions about which course of action is appropriate given the most probable state of the world” [1]. It is therefore arguable that information processing has been optimized, at least to some extent, by natural selection [2,3]. This is a rather abstract claim that should ultimately be contrasted with experiments (in this respect, see [4,5]). In a subsequent work, we explore this question in more detail, but a plausible connection can be established with experimental and theoretical results via fMRI in which different measures of cost have been proposed. The brain as an informational system is the subject of active research (see for instance [6–8]). For example, “behavior analysis has adopted these tools [Information theory] as a novel means of measuring the interrelations between behavior, stimuli, and contingent outcomes” [1]. In this context, in [1], it was shown that informational measures of contingency appear to be a reasonable basis for predicting behavior. We also refer the reader to the issue devoted to information theory in neuroscience [9] for a recent account of this perspective. In the work just mentioned, several papers investigate optimization principles (for instance, the maximum entropy principle [10] or the free energy principle; see [11,12]) as tools for understanding inference, coding, and other brain functionalities. Information theory will allow filling the gaps between disciplines, for example psychology, neurobiology, physics, mathematics, and computer science. In the present paper, we adopt a related setting and mathematically formalize information processing in the brain within the framework of optimal transport. The rationale for

this is consistent with the view that some essential brain functionalities such as inference and the coordination of tasks (e.g., auditory and motor activities) involve the transportation of information and that such processes should be efficient, have been subjected to evolutionary pressure, and as a consequence, are (pseudo) optimal. As was already pointed out, our theoretical proposal should be contrasted with experimental results.

It is necessary to observe from the very beginning that information processing and transport are of an intrinsically spatiotemporal nature. Therefore, our proposal should include these two features. In doing so, we expect the spatial part of the optimization to give rise to spatial patterns, for instance network-like or branching hierarchical components, as well as the temporal structure, such as periodic or synchronized patterns. However, in this paper, we begin by considering only the spatial part, except for a few general remarks. Since our proposal is an attempt to establish a methodological framework to study informational entropy transport in the brain, we deal with spacial aspects first, leaving a study of the dynamical aspects for a subsequent work. Furthermore, in order to simplify the problem, we consider only the one-dimensional case in the mathematical formalism. However, it is certain that the geometry of the brain has to play an essential role in all the processes [13], and we extrapolate some of the results we obtain to two and three dimensions.

Now, we provide an overview of the paper. In Section 2, we present a general framework for information processing in the brain as an optimal transport of entropy, as well as some mathematical results in the context of the Monge–Kantorovich problem. The main idea is that instead of considering that some sort of physical mass is being transported, it is informational entropy. In order to provide the mathematical results and for the sake of completeness, we adapt the material on the existence of a solution in the optimal mass transportation case as presented by Bonnotte [14] to the informational case. We conclude this section with a derivation of the Monge–Ampère equation in the one-dimensional case. We begin Section 3 by recalling the linearization of the Monge–Ampère equation around the square of the distance function, which involves the Laplacian. Then, we argue that adding a nonlinear term is justified by the physiological nature of transmission along neurons. The resulting model is a semilinear elliptic equation. At the end of this section, we relate the qualitative features of the solutions with the branching structure of neural networks in the brain. In other words, we show that the optimal transport process of informational entropy is consistent with the geometric branching structure of neural branching. In Section 4, we elaborate on the relationship of the branching structure of neurons and Murray’s law [15], which provides the optimal branching ratio of the father to the daughter branch sections, as well as the optimal bifurcation angle. We propose a modified version of Murray’s law when the underlying transport network carries information instead of a fluid. The last section is devoted to concluding remarks, further research, and open questions.

2. The Monge–Kantorovich Problem

We present a general overview of the results on optimal transportation theory needed in this work. For a complete exposition, see [14,16,17] or [18], but for the sake of completeness, we include a general discussion without giving the proofs of the results, but including appropriate references. Our presentation follows closely [14] and some parts of [16].

The original Monge–Kantorovich problem was formulated in the context of mass transport (Monge) or budget allocation (Kantorovich). In a later section, we will adapt the setting to include the transport of informational entropy.

Monge’s problem: Given two probability measures μ and ν on \mathbb{R}^n and a cost function $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty]$, the problem of Monge can be stated as follows:

$$\text{Find } T : \mathbb{R}^n \rightarrow \mathbb{R}^n, \text{ such that } \nu = T\#\mu \text{ and } \int_{\mathbb{R}^n} c(x, T(x)) d\mu(x) \text{ is minimal.} \quad (1)$$

The condition $\nu = T\#\mu$ means that T transports μ onto ν ; that is, ν is the push-forward of μ by T : for any ζ , $\int_{\mathbb{R}^n} \zeta(y) d\nu(y) = \int_{\mathbb{R}^n} \zeta(T(x)) d\mu(x)$.

Monge–Kantorovich problem: Monge’s problem might have no solution; hence, it is better to take the following generalization proposed by Leonid Kantorovich: instead of looking for a map, find a measure:

$$\pi \in \Pi(\mu, \nu) \quad \text{such that} \quad \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\pi(x, y) \quad \text{is minimal,} \tag{2}$$

where $\Pi(\mu, \nu)$ stands for the set of all transport plans between μ and ν , i.e., the probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ and ν . This problem really extends Monge’s problem. For any transport map T , sending μ onto ν yields a measure $\pi \in \Pi(\mu, \nu)$, which is given by $\pi = (Id, T)\#\mu$, i.e., the only measure π on $\mathbb{R}^n \times \mathbb{R}^n$ such that:

$$\forall \zeta \in C_b, \int_{\mathbb{R}^n \times \mathbb{R}^n} \zeta(x, y) d\pi(x, y) = \int_{\mathbb{R}^n} \zeta(T(x)) d\mu(x),$$

and the associated costs of transportation are the same. In this version, it is not difficult to show that there is always a solution ([14] or [16]).

2.1. Dual Formulation

There is a duality between the Monge–Kantorovich problem (2) and the following problem:

Find $\psi \in L^1(\mu), \phi \in L^1(\nu)$, such that $\psi(x) + \phi(y) \leq c(x, y)$ and $\int_X \psi(x) d\mu + \int_Y \phi(y) d\nu$ is maximal.

It seems natural to look for a solution of this problem among the pairs $(\psi, \phi) \in \mathcal{P}(X \times Y)$ that satisfy:

$$\phi(y) = \inf_x \{c(x, y) - \psi(x)\} \quad \text{and} \quad \psi(x) = \inf_y \{c(x, y) - \phi(y)\}.$$

We will write $\phi(y) = \psi^c(y)$ and $\psi(x) = \phi^c(x)$.

Definition 1. A function ψ is said to be c -concave if $\psi = \phi^c$ for some function ϕ . In that case, ψ^c and ϕ^c are called the c -transform of ψ and ϕ , respectively. We also say that (ψ, ψ^c) is an admissible pair and ψ, ψ^c are admissible potentials.

Then, the problem becomes:

$$\text{Find } \psi \in L^1(\mu) \quad \text{such that} \quad \int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu \quad \text{is maximal.} \tag{3}$$

The function ψ is called a Kantorovich potential between μ and ν .

The following proposition explains how to relate the Monge–Kantorovich problem (2) with (3), known as the Kantorovich duality principle.

Proposition 1 (Kantorovich duality principle). Let μ and ν be Borel probability measures on X and $Y \in \mathbb{R}^n$, respectively. If the cost function $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty]$ is lower semi-continuous and:

$$\int_Y \int_X c(x, y) d\mu(x) d\nu(y) < \infty,$$

then there is a Borel map $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ that is c -concave and optimal for (3). Moreover, the resulting maximum is equal to the minimum of the Monge–Kantorovich problem (2); i.e.,

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\pi(x, y) = \max_{\phi \in L^1(\mu)} \left\{ \int_X \phi(x) d\mu(x) + \int_Y \phi^c(y) d\nu(y) \right\},$$

or:

$$\min_{\pi \in \Pi(\mu, \nu)} I[\pi] = \max_{(\phi, \phi^c) \in \Phi} J(\phi, \phi^c), \tag{4}$$

where,

$$\Phi = \left\{ (\phi, \psi) \in L^1(\mu) \times L^1(\nu) \mid \phi(x) + \psi(y) \leq c(x, y) \text{ for } \mu - a.e. x \in X \text{ and } \nu - a.e. y \in Y \right\}. \tag{5}$$

If $\pi \in \Pi(\mu, \nu)$ is optimal, then $\phi(x) + \psi^c(y) = c(x, y)$ almost everywhere for π .

Proof. A proof of this result can be found in [16]. □

2.2. Solution in the Real Line: Optimal Transportation Case

For the rest of this section, we only consider the one-dimensional case, as discussed in the Introduction.

Let X and Y be two bounded smooth open sets in \mathbb{R} and $\mu(dx), \nu(dy)$ the probability measures of X and Y , respectively, with $\mu(dx) = f dx, \nu(dy) = g dy, f = 0$ in $\mathbb{R} \setminus X$, and $g = 0$ in $\mathbb{R} \setminus Y$. Let $F : \mathbb{R} \rightarrow [0, 1]$ and $G : \mathbb{R} \rightarrow [0, 1]$ be the cumulative distributions of μ and ν , respectively, defined by $F(x) = \mu((-\infty, x])$ and $G(y) = \nu((-\infty, y])$.

Proposition 2. Let $h \in C^1(\mathbb{R})$ be a non-negative, strictly convex function. Let μ and ν be Borel probability measures on \mathbb{R} such that:

$$\int_Y \int_X h(x - y) d\mu(x) d\nu(y) < \infty. \tag{6}$$

If μ has no atom and F and G stand for the respective cumulative distribution of μ and ν , respectively, then:

$$T = G^{-1} \circ F$$

solves Monge’s problem (1) for the cost:

$$c(x, y) = h(x - y). \tag{7}$$

If π is the induced transform plan, that is:

$$\pi = (Id, T) \# \mu$$

defined as $T \# \mu(E) = \mu(T^{-1}(E))$ for $E \in Y$, then π is optimal for the Monge–Kantorovich problem (2).

Proof. A proof of Proposition (2) can be found in [14]. □

In order to get the previous result, one has to consider the functional:

$$\int_{X \times Y} c(x, y) d\pi(x, y),$$

where $c : X \times Y \rightarrow \mathbb{R}$ is some given cost function and $\Pi(x, y)$ stands for the set of all transport plans between μ and ν , meaning the probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν ; i.e.,

$$\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(X \times Y) \mid \int_Y d\pi(x, y) = d\mu(x), \int_X d\pi(x, y) = d\nu(y) \right\}, \tag{8}$$

or more rigorously:

$$\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(X \times Y) \mid \pi(A \times Y) = \mu(A), \pi(X, B) = \nu(B), A \subseteq X, B \subseteq Y \right\}.$$

Our goal is to get a similar result to Proposition (2) in the case when entropy transportation is considered instead of mass transportation. This is the content of the next section.

3. Solution in the Real Line: Optimal Entropy Transportation Case

As was pointed out in [1], “the fundamental measure in information theory is the entropy. It corresponds to the uncertainty associated with a signal. “Entropy” and “information” are used interchangeably, because the uncertainty about an outcome before it is observed, corresponds to the information gained by observing it” (for a review, see [1,19] or [20]). In that context, we will prove the existence of an optimal entropy transport for the cost function $c(x, y)$ satisfying (6) and (7), similarly to the optimal transportation case discussed in the last section. We will also find the Monge–Ampère equation for quadratic cost $|x - y|^2/2$ for this optimal entropy transport. A few words are in order regarding the choice of c . From the mathematical perspective, it considerably simplifies the analysis. As a matter of fact, the optimal transportation problem has not been solved for the general case of nonquadratic costs. On the other hand, from the physiological point of view, it is natural to assume that the energy required to send a signal from one point of the brain to another can be taken as a monotone function of the distance.

Let μ and ν be probability measures defined as above, then take $X = Y = \Omega \in \mathbb{R}$, and let the entropy be characterized by Shannon’s proposal:

$$-\rho(x) \ln(\rho(x)), \tag{9}$$

where $x \in \Omega$, ρ , and $\tilde{\rho}$ are the distribution densities with respect to μ and ν , respectively, in the same way as the formulation of the optimal transportation problem states; i.e.,

$$\mu = \rho dx, \quad \nu = \tilde{\rho} dy,$$

satisfying (8). We wish (9) to be related to the probability measure μ in Ω . It is natural to think that in passing from the state characterized by (9) to the one characterized by $-\tilde{\rho}(y) \ln(\tilde{\rho}(y))$, where $y \in \Omega$ and wish it to be related to the probability measure ν in Ω . Similarly, $-\rho(x) \ln(\rho(x))$ and $-\tilde{\rho}(y) \ln(\tilde{\rho}(y))$ will be the marginals of $-\rho(x, y) \ln(\rho(x, y))$.

As a first concrete proposal, we consider the following functional:

$$\int_{\Omega \times \Omega} c(x, y) \rho(x, y) \ln(\rho(x, y)) \, dx dy,$$

where c is a spatial cost function. Notice that we have dropped the minus sign, so looking for maximal entropy is equivalent to minimizing the previous expression. The problem is then to find the optimal entropy transport strategy between x and y (analogous to Monge’s problem). There is however a standard problem, which is the fact that in the continuous case, the entropy could be negative, whereas in the discrete case (i.e., for discrete probability functions), the entropy is always positive. We therefore consider the absolute value of the entropy defined in (9). More precisely, define the following:

$$|\rho \ln \rho(x)| = (\rho \ln \rho)^+(x) + (\rho \ln \rho)^-(x),$$

with:

$$(\rho \ln \rho)^+(x) = \max_{\Omega}(\rho \ln \rho, 0) \quad \text{and} \quad (\rho \ln \rho)^-(x) = -\min_{\Omega}(\rho \ln \rho, 0).$$

It is natural to assume that:

$$\int_{\Omega} |\rho \ln \rho(x)| \, dx = K, \tag{10}$$

for some constant $K \in \mathbb{R}^+ \setminus \{0\}$, and let:

$$d\mu = \frac{1}{K} |\rho(x) \ln \rho(x)| dx, \tag{11}$$

with:

$$d\mu^+ = \frac{1}{K} (\rho \ln \rho)^+(x) dx \quad \text{and} \quad d\mu^- = \frac{1}{K} (\rho \ln \rho)^-(x) dx.$$

Observation 1. If $\rho(x, y)$ is the distribution density of $\pi(x, y)$ in $\Omega \times \Omega$, then (11) is a well-defined Borel probability measure on Ω , and:

$$d\nu = \frac{1}{K} |\tilde{\rho}(y) \ln \tilde{\rho}(y)| dy \tag{12}$$

is also a well-defined Borel probability measure on Ω . Then, we can define:

$$\widehat{\Pi}(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\Omega \times \Omega) \mid \int_{Y=\Omega} d\pi(x, y) = d\mu(x), \int_{X=\Omega} d\pi(x, y) = d\nu(y) \right\}, \tag{13}$$

with μ and ν given by (11) and (12); or more rigorously:

$$\widehat{\Pi}(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\Omega \times \Omega) \mid \pi(A, \Omega) = \mu(A), \pi(\Omega, B) = \nu(B), A, B \subseteq \Omega \right\},$$

with μ and ν given by (11) and (12); we can take the cumulative distributions of μ and ν , respectively, by $F(x) = \mu((-\infty, x])$ and $G(y) = \nu((-\infty, y])$ with μ and ν given as before. By definition, they are non-decreasing.

Now, we are in the condition to establish the problem (1) analogous to that of Monge, namely:

Find a map $T : \mathbb{R} \rightarrow \mathbb{R}$, such that $\nu = T\#\mu$ with μ and ν given by (11) and (12) and

$$\int_{\Omega} c(x, T(x)) d\mu(x) \text{ is minimal,} \tag{14}$$

and analogous to the Monge–Kantorovich problem:

Find a measure $\pi \in \widehat{\Pi}(\mu, \nu)$ given by (13) with μ and ν given by (11) and (12) such that

$$\int_{\Omega \times \Omega} c(x, y) d\pi(x, y) \text{ is minimal.} \tag{15}$$

Next, we present the equivalent proposition of (2), for the measures given by (11) and (12), namely:

Proposition 3. Let $h \in C^1(\mathbb{R})$ be a non-negative, strictly convex function. Let μ and ν be Borel probability measures on \mathbb{R} given by (11) and (12), respectively. Suppose that:

$$\int_{\Omega} \int_{\Omega} h(x - y) d\mu(x) d\nu(y) < \infty. \tag{16}$$

for $\Omega \in \mathbb{R}$. If μ has no atom and F and G represent the corresponding cumulative distribution functions of μ and ν , respectively, then:

$$T = G^{-1} \circ F$$

solves Problem (14) for the cost:

$$c(x, y) = h(x - y).$$

If π is the induced entropy transform plan, that is:

$$\pi = (Id, T)\#\mu,$$

defined as $T\#\mu(E) = \mu(T^{-1}(E))$ for $E \in \Omega$, then π is optimal for Problem (15).

Proof. The proof of this result is adapted from [14] and for the sake of completeness is given in Appendix A. \square

Our next goal is to deduce the Monge–Ampère equation for this case. In order to do that, we will need an analog of the Kantorovich duality principle (Proposition (1)), for the measures given by (11) and (12), namely:

Proposition 4. Let μ and ν be the Borel probability measures on $\Omega \in \mathbb{R}$ given by (11) and (12), respectively. If the cost function:

$$c : \mathbb{R} \times \mathbb{R} \longrightarrow [0, \infty)$$

is lower semi-continuous and:

$$\int_{\Omega} \int_{\Omega} c(x, y) d\mu(x) d\nu(y) < \infty,$$

then there is a Borel map $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ that is c -concave and optimal for (3). Moreover, the resulting maximum is equal to the minimum of Problem (15); i.e.,

$$\min_{\pi \in \hat{\Pi}(\mu, \nu)} \int_{\mathbb{R} \times \mathbb{R}} c(x, y) d\pi(x, y) = \max_{\psi \in L^1(\mu)} \left\{ \int_{\Omega} \psi(x) d\mu(x) + \int_{\Omega} \psi^c(y) d\nu(y) \right\}$$

or:

$$\min_{\pi \in \hat{\Pi}(\mu, \nu)} I[\pi] = \max_{(\psi, \psi^c) \in \Phi} J(\psi, \psi^c), \tag{17}$$

where,

$$\Phi = \left\{ (\psi, \psi^c) \in L^1(\mu) \times L^1(\nu) \mid \psi(x) + \psi^c(y) \leq c(x, y) \text{ for } \mu - a.e. x \in \Omega \text{ and } \nu - a.e. y \in \Omega \right\}, \tag{18}$$

and if $\pi \in \hat{\Pi}(\mu, \nu)$ given by (13) is optimal, then $\psi(x) + \psi^c(y) = c(x, y)$ almost everywhere for π .

Proof. A proof of this result can be found in Appendix A. \square

If we propose $C(x, y) = |x - y|^2/2$, we wish T to be expressed as $T = \nabla\phi$ for some convex function ϕ and then to be able to find the corresponding Monge–Ampère equation related to the measures μ and ν given by (11) and (12). This fact is guaranteed by Brenier’s theorem. The details of its proof adapted for our case are important and can be found in Appendix A.

Theorem 1 (Brenier). Let μ and ν be the Borel probability measures on $\Omega \in \mathbb{R}$ given by (11) and (12), respectively, and with finite second-order moments; that is, such that:

$$\int_{\Omega} |x|^2 d\mu(x) < \infty \quad \text{and} \quad \int_{\Omega} |y|^2 d\nu(y) < \infty. \tag{19}$$

Then, if μ is absolutely continuous on Ω , there exists a unique $T : \mathbb{R} \longrightarrow \mathbb{R}$ such that $\nu = T\#\mu$ and:

$$\int_{\Omega} |x - T(x)|^2 d\mu(x) = \min_{\gamma \in \hat{\Pi}(\mu, \nu)} \int_{\Omega} |x - y|^2 d\gamma(x, y),$$

with $\widehat{\Pi}(\mu, \nu)$ given by (13). Moreover, there is only one optimal transport plan, γ , which is necessarily $(Id, T)\#\mu$, and T is the gradient of a convex function φ , which is therefore unique up to an additive constant. There is also a unique (up to an additive constant) Kantorovich potential, ψ , which is locally Lipschitz and linked to φ through the relation:

$$\varphi(x) = \frac{1}{2}|x|^2 - \psi(x).$$

Proof. See Appendix A. \square

Observation 2. Observe that Theorem 1 holds for the general case on \mathbb{R}^n .

The Monge–Ampère Equation

Let:

$$d\mu(x) = \frac{1}{K}|\rho(x) \ln[\rho(x)]|dx \quad \text{and} \quad d\nu(y) = \frac{1}{K}|\tilde{\rho}(y) \ln[\tilde{\rho}(y)]|dy$$

be two probability measures, absolutely continuous with respect to the Lebesgue measure. By Theorem 1, there exists a unique gradient of a convex function, φ , such that:

$$\int_{\Omega} \zeta(y)|\tilde{\rho}(y) \ln[\tilde{\rho}(y)]|dy = \int_{\Omega} \zeta(\nabla\varphi(x))|\rho(x) \ln[\rho(x)]|dx, \tag{20}$$

for all test functions $\zeta \in C_b(\mathbb{R})$. Since φ is strictly convex, then $\nabla\varphi$ is C^0 and one-to-one. Hence, taking $y = \nabla\varphi(x)$, we get:

$$\int_{\Omega} \zeta(y)|\tilde{\rho}(y) \ln[\tilde{\rho}(y)]|dy = \int_{\Omega} \zeta(\nabla\varphi(x))|\tilde{\rho}(\nabla\varphi(x)) \ln[\tilde{\rho}(\nabla\varphi(x))]| \det D^2\varphi(x)dx. \tag{21}$$

From (20) and (21), we get:

$$|\rho(x) \ln[\rho(x)]| = |\tilde{\rho}(\nabla\varphi(x)) \ln[\tilde{\rho}(\nabla\varphi(x))]| \det D^2\varphi(x), \tag{22}$$

and the Monge–Ampère equation:

$$\det D^2\varphi(x) = \frac{|\rho(x) \ln[\rho(x)]|}{|\tilde{\rho}(\nabla\varphi(x)) \ln[\tilde{\rho}(\nabla\varphi(x))]|}, \tag{23}$$

corresponding to this case.

4. Neural Branching Structure and the Linearization of the Monge–Ampère Equation

As was pointed out in the Introduction, the purpose of this section is to propose a model for the branching structure of the neurons, which is consistent with the process of information transport previously introduced. The basic idea is as follows. If we consider that information transport is optimized in some brain processes, we consequently have (as discussed in the previous section) an associated Monge–Ampère equation for the transport plan potential. Besides, it is natural to consider a cost that is close to some power of the distance function, since physiological cost can be taken to depend on the distance traveled by the corresponding signal, as is usually assumed in transport networks. For technical reasons, and in order to be able to adapt results well known in the literature, we take this cost function to be close to:

$$\frac{|x - y|^2}{2},$$

From a qualitative perspective, this choice should not change the results much, as long as the cost function remains convex (see [16]). If this is the case, we can then compute the linearization of the Monge–Ampère equation around this quadratic cost function. The resulting equation is a linear elliptic equation. We argue that a self-activating mechanism should be incorporated in the form of a nonlinear term (as a result of the excitable nature of the transport of electric impulses along axons). In this way, we end up with a semilinear equation that can be used to explain branching processes in biological networks (see [21]). More specifically, the solution to this equation could be associated with the concentration of a morphogen, e.g., a growth factor, and if such a concentration is above a certain threshold, a branching mechanism is triggered. It is then consistent to look for solutions that are close to the quadratic cost function and therefore to linearize around it.

In order to linearize the Monge–Ampère equation, we assume then that φ is very close to $|x|^2/2$, so $|\rho(x) \ln[\rho(x)]|$ is very close to $|\tilde{\rho}(\nabla\varphi(x)) \ln[\tilde{\rho}(\nabla\varphi(x))]|$. In that case, following [16,22,23], make:

$$\varphi(x) \equiv \varphi_\varepsilon(x) = \frac{|x|^2}{2} + \varepsilon\eta + O(\varepsilon^2) \quad (24)$$

and:

$$\left| \tilde{\rho}(\nabla\varphi(x)) \ln[\tilde{\rho}(\nabla\varphi(x))] \right| = \left| \tilde{\rho}(\nabla\varphi(x)) \ln[\tilde{\rho}(\nabla\varphi(x))] \right|_\varepsilon \equiv (1 + \varepsilon h + O(\varepsilon^2)) \left| \rho(x) \ln[\rho(x)] \right|, \quad (25)$$

with $\eta, h \in L^1(\mu)$ and $\varepsilon \ll 1$. We leave the details of this computation for Appendix A. Substituting (24) and (25) in the Monge–Ampère Equation (23), we get as the linearized operator:

$$L\eta = h, \quad (26)$$

with:

$$L = \left(-\Delta + \nabla \left(-\log \left| \rho(x) \ln \rho(x) \right| \right) \cdot \nabla \right).$$

Then, the Laplacian plus a transport term can be seen as the linearized version of the Monge–Ampère equation for our proposal. We notice that the main mechanism responsible for the flow of information along the axons is the propagation of electrical impulses. This is well known to be an excitable process that involves, among others, a self-activating component, as for instance in the standard Hodgkin–Huxley or FitzHugh–Nagumo models. If we include this into the linearization previously obtained (Equation (26)), we get:

$$L\phi = \left(-\Delta\phi + \nabla \left(-\log \left| \rho(x) \ln \rho(x) \right| \right) \cdot \nabla\phi \right) + F(\phi),$$

where F is a function describing the self-activating mechanism and can be taken typically as a power of ϕ :

$$F(\phi) = \phi^p,$$

with $p > 1$. Solutions of this type of equations have been studied by many authors since the pioneering work by Ni and Takagi ([24] and the references therein), since they appear in different contexts. These solutions typically exhibit concentrations that can be responsible for branching structures.

Indeed, if one assumes that the concentration of a solution to the previous equation is correlated with a growth factor morphogen, then a branch will stem out of the main branch. This or similar models have been proposed using reaction-diffusion models following Turing’s original proposal ([25] or [26]), for pattern formation; in particular for branching structures in plants ([27,28] or [29]), lungs ([21]), and other vascular systems ([30,31]). Figures 1 and 2 show numerical simulations for a particular case of the linearization of the Monge–Ampère equation. Growth is induced by the concentration of the solution, and it can be seen that the process gives rise to lateral branches.

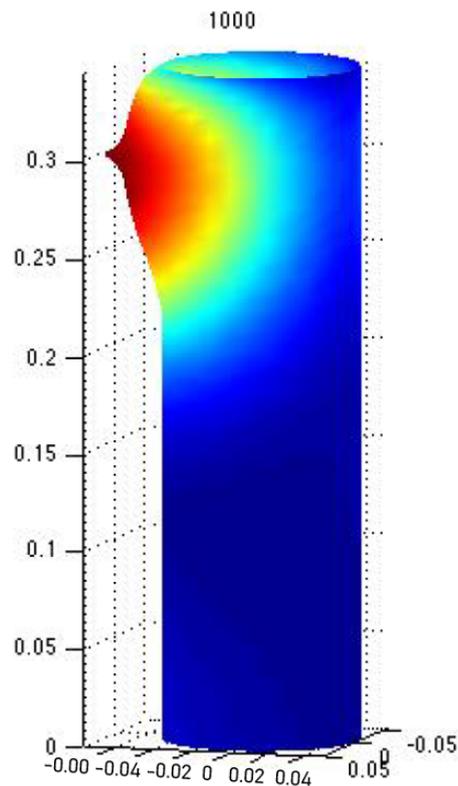


Figure 1. Numerical solution of the linearization of the Monge-Ampère equation including growth. Branching occurs when there is the concentration of the solution, the morphogen, above a certain threshold (the color code stands for standard heat maps: red, high; blue, low). This simulation was provided by Jorge Castillo-Medina and developed in COMSOL. For more details, the reader is referred to [21] and the references therein.

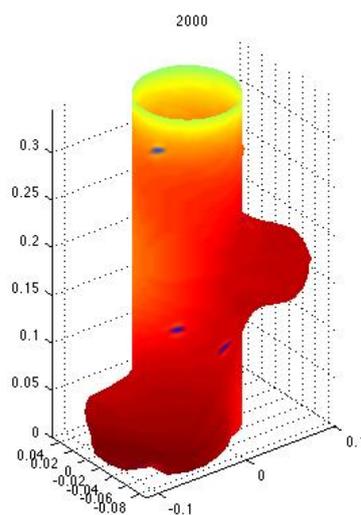


Figure 2. A similar simulation as in the previous figure with a different growth rate. Notice the different branching structure. Simulation performed by J. Castillo as well; see also [21].

5. Murray's Law and Neural Branching

In the previous section, we argued that neuronal branching is compatible with reaction diffusion processes. On the other hand, we deduced the corresponding equations by considering the

transport of information along neural networks. The question of whether there is some connection with Murray’s law arises naturally. Recall that Murray’s law refers to a transport network ([32,33]).

In his original paper [32], Murray obtained from optimizing considerations a relationship for the different parameters associated with a branching transport network that was later generalized in [15]. In what follows, we deduce it from scratch for the sake of completeness (we refer the reader to [15,34,35] for further details). The total power required for the flow to overcome the viscous drag is described by:

$$W_t = \frac{8\mu L f^2}{\pi r^4} + m\pi L r^2, \tag{27}$$

where μ is the dynamic viscosity of the fluid, L is the vessel length, f is the volumetric flow rate, m is an all-encompassing metabolic coefficient that includes the chemical cost of keeping the blood constituents fresh and functional and the general cost owing to the weight of the blood and the vessel, and r is the vessel radius. For our purposes, we modify Equation (27) as follows:

$$W_t = \frac{a f^2}{r^2} + b r^\alpha, \tag{28}$$

where the first term corresponds to the power required for an electrical impulse to propagate along the axon. Notice in particular that the factor of r^2 in the denominator follows from the fact that electrical resistance is inversely proportional to the area of the section of the conducting material. On the other hand, the second term is proportional to a power, α , of the radius and describes the fact that metabolic cost can vary depending on the type of neurons with which we are dealing. For instance, the degree of myelination of the axon could determine the effective cost associated with information transport. The minimum power is found by differentiating with respect to r and equating to zero:

$$\frac{dW_t}{dr} = \frac{-2a f^2}{r^3} + \alpha b r^{\alpha-1}. \tag{29}$$

With this, the optimal radius:

$$r^{2+\alpha} = \frac{2a}{\alpha b} f^2, \tag{30}$$

and the optimal relation between volumetric flow rate and vessel radius, such that the power requirement is minimized, is obtained with:

$$f = k r^{1+\frac{\alpha}{2}}, \tag{31}$$

where $k = \sqrt{\frac{\alpha b}{2a}}$.

Using the construction of [33], if the radius of the main branch (r_0), lateral branches (r_1, r_2), and x and y are the angles between the lateral branches and the main branch, we obtain a generalized version of Murray’s law:

$$f_0 = f_1 + f_2 = k r_0^{1+\frac{\alpha}{2}} = k \left(r_1^{1+\frac{\alpha}{2}} + r_2^{1+\frac{\alpha}{2}} \right);$$

thus, we get the general law:

$$r_0^{1+\frac{\alpha}{2}} = r_1^{1+\frac{\alpha}{2}} + r_2^{1+\frac{\alpha}{2}}.$$

for $\alpha \in \mathbb{R}$. Using these relations, we obtain three general equations associated with the branching angles (see Figure 3) x, y , and $x + y$:

$$\cos(x) = \frac{r_0^4 + r_1^4 - \left(r_0^{1+\frac{\alpha}{2}} - r_1^{1+\frac{\alpha}{2}} \right)^{\frac{8}{2+\alpha}}}{2r_0^2 r_1^2}, \tag{32}$$

$$\cos(y) = \frac{r_0^4 + r_2^4 - \left(r_0^{1+\frac{\alpha}{2}} - r_2^{1+\frac{\alpha}{2}}\right)^{\frac{8}{2+\alpha}}}{2r_0^2 r_2^2}, \tag{33}$$

$$\cos(x+y) = \frac{\left(r_1^{1+\frac{\alpha}{2}} + r_2^{1+\frac{\alpha}{2}}\right)^{\frac{8}{2+\alpha}} - r_1^4 - r_2^4}{2r_1^2 r_2^2}, \tag{34}$$

which correspond to a different generalized Murray’s law for different values of $\alpha \in \mathbb{R}$.

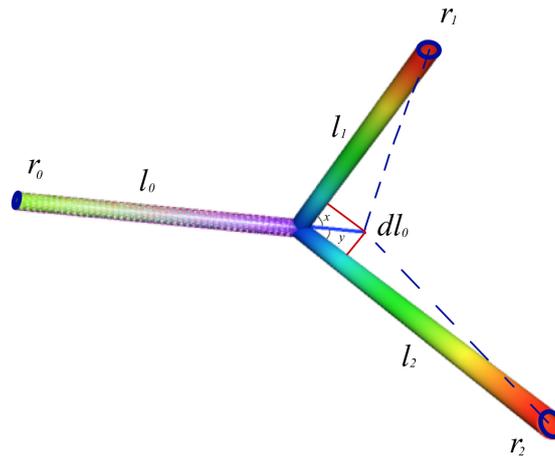


Figure 3. The figure shows schematically the geometric configuration when branching occurs on a plane. Illustration after [33] by the authors.

Observation 3. If $\alpha = 2$, then by (32)–(34), we get $\cos(x+y) = 1$ and then $x+y = 0$, which is consistent with the fact that $\cos x = 1$, then $x = 0$ and $\cos y = 1$, and then, $y = 0$. This case corresponds to no branching. In terms of the cost, this would imply that it is more efficient for the network not to bifurcate.

If $\alpha = 4$, then $\cos(x) = \frac{r_0^4 + r_1^4 - (r_0^3 - r_1^3)^{\frac{4}{3}}}{2r_0^2 r_1^2}$, $\cos(y) = \frac{r_0^4 + r_2^4 - (r_0^3 - r_2^3)^{\frac{4}{3}}}{2r_0^2 r_2^2}$, and $\cos(x+y) = \frac{(r_1^3 + r_2^3)^{\frac{4}{3}} - r_1^4 - r_2^4}{2r_1^2 r_2^2}$, which corresponds to Murray’s original proposal [33], which states that the angle in the bifurcation of an artery should not be less than 75° (74.9° to be more exact). This is consistent with the numerical and experimental results in [36].

If $\alpha = 6$, we get $\cos(x+y) = 0$ and then $x+y = \frac{\pi}{2}$. On the other hand, $\cos(x) = r_1^2/r_0^2$, then $\cos(x) > 0$ since $r_0, r_1 \neq 0$; this implies that $\cos(x) \neq 0$ and $x \neq \frac{\pi}{2}$. Similarly, $y \neq \frac{\pi}{2}$. We obtain that $x+y = \frac{\pi}{2}$ and $x, y \in (0, \frac{\pi}{2})$ for this case. In other words, for $\alpha = 6$, the angle between the bifurcated branches is $\pi/2$, but orthogonal branching is ruled out.

We conclude then that the relevant values for our purposes are for $\alpha \in [2, 6]$. It would be interesting to contrast these possible scenarios with experimental data for different kinds of nervous tissues. To our knowledge, no systematic experimental study of branching angles has been carried out.

6. Conclusions

We proposed that information flow in some brain processes can be analyzed in the framework of optimal transportation theory. A Monge–Ampère equation was obtained for the optimal transportation plan potential in the one-dimensional case. Extrapolating to higher dimensions, the corresponding linearization around a quadratic distance cost was derived and shown to be consistent with the branching structure of the nervous system. Finally, a generalized version of Murray’s law was derived assuming different cost functions, depending on a parameter related to the metabolic maintenance term. Future work includes a detailed comparison of the methodological proposal with experimental

data. In particular, it would be interesting to carry out the program here proposed in a concrete cognitive experiment. Possible concrete experiments to compare with could be found in [2,3,5]. Here, we outline a simple procedure with fMRI data in which a direct connection with optimal transport theory can be tested. Consider the brain activity map for the resting state given by standard fMRI. Once normalized, this map will provide the initial probability density and entropy to be transported. In fact, in [37], another possible methodology for measuring the entropy with fMRI can be found. Later on, the subject is asked to perform a simple motor task, for instance move the right hand. The corresponding density after the task is done can then be registered as before and will provide the final density in the optimal transport problem. Some intermediate densities should be determined as well. This information will provide a transport plan that can be compared with the mathematical solution of the problem. Correspondingly, the branching structures and their bifurcation angles and radii should be compared with experimental results as well. In principle, Murray's law should be consistent with the Monge–Ampère equation and its linearization, and it should be possible to derive it from them. A precise relationship between the maximum entropy principle and optimal entropy transport should be clarified.

Author Contributions: All authors contributed equally to the manuscript: read and approved the final version, conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing—original draft preparation, writing—review and editing, visualization, supervision and project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: M.A.P. was supported by Universidad Autónoma de la Ciudad de México, sabbatical approval UACM/CAS/010/19. This work was supported by the Departamento de Matemáticas y Mecánica (MyM) of the Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) of the Universidad Nacional Autónoma de la Ciudad de México (UNAM). M.A.P. would like to thank P.P. and IIMAS-UNAM for the support during his sabbatical leave. The authors would also like to thank Jorge Castillo-Medina at the Universidad Autónoma de Guerrero in Acapulco for his kind permission to use Figures 1 and 2.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Relevant Theorems and Some Proofs

Theorem A1 (The continuity and support theorem). *Let μ be a probability measure on \mathbb{R} with cumulative distribution function F . The following properties are equivalent:*

a. *The function F is strictly increasing in the interval:*

$$\{x \in \mathbb{R} \mid 0 < F(x) < 1\}.$$

b. *The inverse function F^{-1} is continuous.*

c. *The inverse measure μ^{-1} is non-atomic.*

d. *The support of μ , given by:*

$$\text{supp}(\mu) = \overline{\{A \in \Sigma \mid \mu(A) > 0\}}$$

with Σ a σ -algebra defined on \mathbb{R} , is a closed interval in the real line, finite or not.

Proof. The proof of this result can be found in [38], Appendix A. \square

Proposition A1. *Let $h \in C^1(\mathbb{R})$ be a non-negative, strictly convex function. Let μ and ν be Borel probability measures on \mathbb{R} given by (11) and (12), respectively. Suppose that:*

$$\int_{\Omega} \int_{\Omega} h(x-y) d\mu(x) d\nu(y) < \infty. \quad (\text{A1})$$

for $\Omega \in \mathbb{R}$. If μ has no atom and F and G represent the respective cumulative distribution functions of μ and ν , respectively, then:

$$T = G^{-1} \circ F$$

solves Problem (14) for the cost:

$$c(x, y) = h(x - y).$$

If π is the induced entropy transform plan, that is:

$$\pi = (Id, T)\#\mu,$$

defined as $T\#\mu(E) = \mu(T^{-1}(E))$ for $E \in \Omega$, then π is optimal for Problem (15).

Proof. 1. T is well defined:

The only problem we might have with the definition of $T = G^{-1} \circ F$ could be when $F(x) = 0$. However, if $F(x) = 0$, then:

$$\begin{aligned} T(x) &= G^{-1}(F(x)) \\ &= G^{-1}(0) \\ &= \min \{y \in [-\infty, \infty] \mid 0 \leq G(y)\} \\ &= -\infty, \end{aligned}$$

but if for some $a \in \mathbb{R}$, we have $F(a) = 0$, then $\mu((-\infty, a]) = 0$, which means that $a = -\infty$ and T is well defined, as desired.

2. $\nu = T\#\mu$:

Let F and G be defined as in observation (1). Then, $T = G^{-1} \circ F$ is non-decreasing, since F and G are non-decreasing. Then:

$$\begin{aligned} T^{-1}((-\infty, y]) &= \{x \in [-\infty, \infty] \mid T(x) = y\} \\ &= \{x \in [-\infty, \infty] \mid G^{-1}(F(x)) = y\} \\ &= \{x \in [-\infty, \infty] \mid F(x) = G(y)\}, \end{aligned}$$

Since T is non-decreasing, $T^{-1}((-\infty, y])$ is an interval.

Claim 1. Since μ has no atom, F is increasing and continuous, and then, $T^{-1}((-\infty, y])$ is a closed interval.

If μ has no atom, let $A \in \Omega$ and $\alpha \in \mathbb{R}^+$ such that $\mu(A) = \alpha$. Then, given $\beta \in \mathbb{R}^+$ such that $0 < \beta \leq \alpha$, there exists $B \subset A$ such that $\mu(B) = \beta$. Then, $\forall \gamma \in \mathbb{R}$ such that $0 < \gamma \leq \alpha$, there exists $x_\gamma \in A$ such that:

$$\mu((-\infty, x_\gamma]) = \gamma,$$

then $0 < F(x_\gamma) < F(x_\alpha) \leq 1$ for all γ, α such that $x_\gamma < x_\alpha$, and so, F is increasing and continuous; Theorem A1 implies that F^{-1} is continuous, and then:

$$\{x \in \mathbb{R} \mid 0 \leq F(x) \leq 1\} = \text{supp}(\mu),$$

is closed. In particular, $\{x \in \mathbb{R} \mid 0 \leq F(x) \leq G(y)\}$ is closed. Then, we conclude that $T^{-1}((-\infty, y])$ is a closed interval, as desired.

We have proven Claim 1.

Now, if $x = \sup\{x \in \mathbb{R} \mid F(x) \leq G(y)\}$, then $F(x) = G(y)$, and we have:

$$\begin{aligned} \mu(T^{-1}((-\infty, y])) &= \mu((-\infty, x]) \\ &= F(x) \\ &= G(y) \\ &= \nu((-\infty, y]), \end{aligned}$$

and then, $\nu = T\#\mu$, as desired.

3. $T = G^{-1} \circ F$ is optimal:

Observe that (μ, ν) given by (11) and (12) satisfy:

$$\int_{Y=\Omega} d\pi(x, y) = d\mu(x) \quad \text{and} \quad \int_{X=\Omega} d\pi(x, y) = d\nu(y)$$

where:

$$d\pi(x, y) = \frac{1}{K} |\rho(x, y) \ln(\rho(x, y))| dx dy,$$

and:

$$d\mu(x) = \frac{1}{K} |\rho(x) \ln(\rho(x))| dx,$$

then $\pi(x, y) \in \widehat{\Pi}(\mu, \nu)$ as given by (13).

Now, observe that:

$$\int_{\Omega} h'(x - T(x)) \frac{1}{K} |\rho(x) \ln(\rho(x))| dx = \int_{\Omega} h'(x - T(x)) d\mu;$$

since T and h' are non-decreasing,

$$h'(u - T(u)) \leq h'(u - T(x))$$

for $u \geq x$. Then:

$$\begin{aligned} \int_x^y h'(u - T(u)) du &\leq \int_x^y h'(u - T(x)) du \\ &\leq h(y - T(x)) - h(x - T(x)). \end{aligned}$$

On the other hand,

$$h'(u - T(u)) \geq h'(u - T(x))$$

for $u \leq x$, and then:

$$\begin{aligned} \int_x^y h'(u - T(u)) du &= - \int_y^x h'(u - T(u)) du \\ &\leq - \int_y^x h(u - T(x)) du; \\ &= -[h(x - T(x)) - h(y - T(x))] \\ &= h(y - T(x)) - h(x - T(x)). \end{aligned}$$

As a consequence, in any case, we have:

$$\int_x^y [h(u - T(u))du] \leq h(y - T(x)) - h(x - T(x)). \tag{A2}$$

Set:

$$\psi(y) = \int_0^y h'(u - T(u))du,$$

then (A2) becomes:

$$\psi(y) - \psi(x) \leq h(y - T(x)) - h(x - T(x)),$$

which implies:

$$\begin{aligned} \psi^c(T(x)) &= \inf_y \{h(y - T(x)) - \psi(y)\} \\ &= h(x - T(x)) - \psi(x); \end{aligned}$$

as a consequence, ψ is c -concave, according to Definition (1).

Hypothesis A1 implies the existence of $x_0, y_0 \in \Omega$ such that:

$$\int_{\Omega} h(x - y_0)d\mu(x) < \infty \quad \text{and} \quad \int_{\Omega} h(x_0 - y)d\nu(y) < \infty. \tag{A3}$$

Claim 2: $\psi \in L^1(\mu)$ and $\psi^c \in L^1(\nu)$: Observe that:

$$\begin{aligned} h(x - y_0) - \psi^c(y_0) &= h(x - y_0) - \inf_y \{h(y - y_0) - \psi(y)\} \\ &\geq h(x - y_0) - (h(x - y_0) - \psi(x)) \\ &= \psi(x), \end{aligned} \tag{A4}$$

and:

$$\begin{aligned} h(x_0 - T(x)) - \psi(x_0) &\geq \inf_y \{h(y - T(x)) - \psi(y)\} \\ &= h(x - T(x)) - \psi(x) \\ &= \psi^c(T(x)); \end{aligned}$$

hence:

$$\psi(x) \geq -\psi^c(T(x)) \geq -h(x_0 - T(x)) + \psi(x_0), \tag{A5}$$

then, by (A4) and (A5),

$$h(x - y_0) - \psi^c(y_0) \geq \psi(x) \geq -h(x_0 - T(x)) + \psi(x_0); \tag{A6}$$

now, by Hypothesis A3 and since $\nu = T\#\mu$:

$$\int_{\Omega} h(x_0 - T(x))d\mu = \int_{\Omega} h(x_0 - y)d\nu < \infty; \tag{A7}$$

therefore, integrating (A6) with respect to μ , we get:

$$\int_{\Omega} [h(x - y_0) - \psi^c(y_0)]d\mu(x) \geq \int_{\Omega} \psi(x)d\mu(x) \geq \int_{\Omega} [-h(x_0 - T(x)) + \psi(x_0)]d\mu(x);$$

as a consequence:

$$\int_{\Omega} |\psi(x)| d\mu(x) < \infty,$$

and we can conclude that $\psi \in L^1(\mu)$.

Similarly:

$$h(x_0 - T(x)) - \psi(x_0) \geq \psi^c(T(x)) \geq -\psi(x) \geq -h(x - y_0) + \psi^c(y_0)$$

by (A4) and (A5); using (A3) and (A7), if we integrate with respect to μ ,

$$\int_{\Omega} [h(x_0 - T(x)) - \psi(x_0)] d\mu(x) \geq \int_{\Omega} \psi^c(T(x)) d\mu(x) \geq \int_{\Omega} [-h(x - y_0) + \psi^c(y_0)] d\mu(x);$$

therefore:

$$\int_{\Omega} |\psi^c(T(x))| d\mu(x) = \int_{\Omega} |\psi^c(y)| d\nu(y) < \infty$$

Hence, $\psi^c \in L^1(\nu)$. We have proven Claim 2.

Integrating $\phi(x) + \phi^c(T(x)) = h(x - T(x)) = c(x, y)$ with respect to μ , we get:

$$\int_{\Omega} \psi(x) d\mu(x) + \int_{\Omega} \psi^c(T(x)) d\mu(x) = \int_{\Omega} h(x - T(x)) d\mu(x),$$

and then:

$$\int_{\Omega} \psi(x) d\mu(x) + \int_{\Omega} \psi^c(y) d\nu(y) = \int_{\Omega} c(x, T(x)) d\mu(x).$$

Finally, observe that $\psi(x) + \psi^c(y) \leq c(x, y)$ for every $(x, y) \in \Omega \times \Omega$; if $\pi_1 \in \widehat{\Pi}(\mu, \nu)$ is another entropy transport plan, the associated total entropy transport cost $c(x, y)$ is greater, by the definition of ψ and ψ^c ; then, the equality holds only for the optimal entropy transport plan π and $T = G^{-1} \circ F$; hence, it solves Problem (14), and $\pi = (Id, T)\# \mu$ solves Problem (15).

We have proven the proposition.

□

Proposition A2. Let μ and ν be the Borel probability measures on $\Omega \in \mathbb{R}$ given by (11) and (12), respectively. If the cost function:

$$c : \mathbb{R} \times \mathbb{R} \longrightarrow [0, \infty)$$

is lower semi-continuous and:

$$\int_{\Omega} \int_{\Omega} c(x, y) d\mu(x) d\nu(y) < \infty,$$

then there is a Borel map $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ that is c -concave and optimal for (3). Moreover, the resulting maximum is equal to the minimum of Problem (15); i.e.,

$$\min_{\pi \in \widehat{\Pi}(\mu, \nu)} \int_{\mathbb{R} \times \mathbb{R}} c(x, y) d\pi(x, y) = \max_{\psi \in L^1(\mu)} \left\{ \int_{\Omega} \psi(x) d\mu(x) + \int_{\Omega} \psi^c(y) d\nu(y) \right\}$$

or:

$$\min_{\pi \in \widehat{\Pi}(\mu, \nu)} I[\pi] = \max_{(\psi, \psi^c) \in \Phi} J(\psi, \psi^c), \tag{A8}$$

where,

$$\Phi = \left\{ (\phi, \psi) \in L^1(\mu) \times L^1(\nu) \mid \phi(x) + \psi(y) \leq c(x, y) \text{ for } \mu - a.e. x \in \Omega \text{ and } \nu - a.e. y \in \Omega \right\}, \tag{A9}$$

and if $\pi \in \hat{\Pi}(\mu, \nu)$ given by (13) is optimal, then $\psi(x) + \psi^c(y) = c(x, y)$ almost everywhere for π .

Proof. The existence of a maximizing pair and Relation (A8) have been proven in Proposition A1. Now, choosing an optimal $\pi \in \hat{\Pi}(\mu, \nu)$, we have:

$$\begin{aligned} \int_{\Omega \times \Omega} [c(x, y) - \psi(x) - \psi^c(y)] d\pi(x, y) &= \int_{\Omega \times \Omega} c(x, y) d\pi(x, y) - \int_{\Omega} \psi(x) d\mu(x) \\ &\quad - \int_{\Omega} \psi^c(y) d\nu(y) \\ &= 0, \end{aligned}$$

and since $c(x, y) - \psi(x) - \psi^c(y) \geq 0$ in $\Omega \times \Omega$, it must vanish for π -a.e. $(x, y) \in \Omega \times \Omega$. The proof is completed. \square

Theorem A2 (Brenier). Let μ and ν be the Borel probability measures on $\Omega \in \mathbb{R}$ given by (11) and (12), respectively, and with second-order moments; that is, such that:

$$\int_{\Omega} |x|^2 d\mu(x) < \infty \quad \text{and} \quad \int_{\Omega} |y|^2 d\nu(y) < \infty \tag{A10}$$

Then, if μ is absolutely continuous on Ω , there exists a unique $T : \mathbb{R} \rightarrow \mathbb{R}$ such that $\nu = T\#\mu$ and:

$$\int_{\Omega} |x - T(x)|^2 d\mu(x) = \min_{\gamma \in \hat{\Pi}(\mu, \nu)} \int_{\Omega} |x - y|^2 d\gamma(x, y),$$

with $\hat{\Pi}(\mu, \nu)$ given by (13). Moreover, there is only one optimal transport plan γ , which is necessarily $(Id, T)\#\mu$, and T is the gradient of a convex function ϕ , which is therefore unique up to an additive constant. There is also a unique (up to an additive constant) Kantorovich potential, ψ , which is locally Lipschitz and linked to ϕ through the relation:

$$\phi(x) = \frac{1}{2}|x|^2 - \psi(x).$$

Proof. Claim 1: $\pi = (Id, \nabla\phi)\#\mu$, and it is optimal:

Set $c(x, y) = \frac{|x - y|^2}{2}$. Proposition A1 guarantees the existence of the pair $(\psi, \psi^c) \in \Phi$ as given by (A9) such that $\psi(x) + \psi^c(y) \leq \frac{|x - y|^2}{2}$ for all $(x, y) \in \Omega \times \Omega$; then:

$$\psi(x) + \psi^c(y) \leq \frac{x^2 - 2xy + y^2}{2}$$

and then:

$$xy \leq \left[\frac{x^2}{2} - \psi(x) \right] + \left[\frac{y^2}{2} - \psi^c(y) \right].$$

Set:

$$\phi(x) = \left[\frac{x^2}{2} - \psi(x) \right] \quad \text{and} \quad \phi^c(y) = \left[\frac{y^2}{2} - \psi^c(y) \right]. \tag{A11}$$

Hypothesis A10 means:

$$M = \int_{\Omega} \frac{|x|^2}{2} d\mu(x) + \int_{\Omega} \frac{|y|^2}{2} d\nu(y) < \infty,$$

then:

$$I[\pi] = \int_{\Omega \times \Omega} \frac{|x-y|^2}{2} d\pi(x,y) \leq \int_{\Omega \times \Omega} [|x|^2 + |y|^2] d\pi(x,y) = 2M;$$

hence, $I[\pi]$ is always finite in $\widehat{\Pi}(\mu, \nu)$ given by (13). As a consequence,

$$\begin{aligned} \inf_{\widehat{\Pi}(\mu, \nu)} I[\pi] &= \frac{1}{2} \left[\int_{\Omega} |x|^2 d\mu(x) + \int_{\Omega} |y|^2 d\nu(y) \right] \\ &\quad - \sup \left\{ \int_{\Omega \times \Omega} xy d\pi(x,y) \mid \pi \in \widehat{\Pi}(\mu, \nu) \right\} \\ &= M - \inf_{\widehat{\Phi}} \left\{ \int_{\Omega} \varphi(x) d\mu(x) + \int_{\Omega} \varphi^c(y) d\nu(y) \right\} \\ &= \sup_{\widehat{\Phi}} \left\{ \int_{\Omega} \varphi(x) d\mu(x) + \int_{\Omega} \varphi^c(y) d\nu(y) \right\} \\ &= \sup_{\widehat{\Phi}} J(\varphi, \varphi^c), \end{aligned}$$

where $J(\varphi, \varphi^c) = \int_{\Omega} \varphi(x) d\mu(x) + \int_{\Omega} \varphi^c(y) d\nu(y)$, $\widehat{\Phi}$ given by (18), and:

$$\widehat{\Phi} = \left\{ (\varphi, \psi) \in L^1(\mu) \times L^1(\nu) \text{ with values in } \mathbb{R} \cup \{\infty\} \mid xy \leq \varphi(x) + \psi(y) \text{ for } \mu - \text{a.e. } x \in \Omega \right. \tag{A12}$$

$$\left. \text{and } \nu - \text{a.e. } y \in \Omega \right\}.$$

Then, the Kantorovich duality principle (4) becomes:

$$\sup \left\{ \int_{\Omega \times \Omega} xy d\pi(x,y) \mid \pi \in \widehat{\Pi}(\mu, \nu) \right\} = \inf_{\widehat{\Phi}} J(\varphi, \varphi^c). \tag{A13}$$

Now, $xy \leq \varphi(x) + \varphi^c(y)$ since $(\varphi, \varphi^c) \in \widehat{\Phi}$; hence:

$$\varphi^c \geq \sup_x \{xy - \varphi(x)\} \equiv \varphi^*(y).$$

$\varphi^*(y)$ is called the Legendre transform of $\varphi(x)$ and satisfies:

$$\varphi(x) + \varphi^*(y) \geq xy.$$

It can be proven that $\inf_{\widehat{\Phi}} J = J(\varphi, \varphi^*)$ (see [16]; Theorem 2.9); now, Proposition A1 guarantees the existence of an optimal transport plan π . As a consequence, Relation (A13) implies:

$$\begin{aligned} \int_{\Omega} \varphi(x) d\mu(x) + \int_{\Omega} \varphi^*(y) d\nu(y) &= \int_{\Omega \times \Omega} xy d\pi(x,y) \\ &= \int_{\Omega \times \Omega} [\varphi(x) + \varphi^c(y)] d\pi(x,y), \end{aligned}$$

and then:

$$\int_{\Omega \times \Omega} [\varphi(x) + \varphi^c(y) - xy] d\pi(x,y) = 0,$$

for all $(x, y) \in \Omega \times \Omega$. However, $\varphi(x) + \varphi^c(y) - xy \geq 0$ by the definition of the Legendre transform. Then, for π -a.e. $(x, y) \in \Omega \times \Omega$, we have:

$$\begin{aligned} \varphi(x) + \varphi^*(y) - xy = 0 &\iff xy \geq \varphi(x) + \varphi^*(y) \\ &\iff xy \geq \varphi(x) + (zy - \varphi(z)), \forall z \in \mathbb{R} \\ &\iff \varphi(z) \geq \varphi(x) + y(z - x) \\ &\iff y \in \partial\varphi(x) \end{aligned} \tag{A14}$$

(by symmetry, we can also conclude that $x \in \partial\varphi^*(y)$). Now, since φ is convex, it can be proven that it is locally Lipschitz and continuous; Rademacher’s theorem (see [39]; Section 3.1.2) implies then that φ is differentiable μ -a.e. on \mathbb{R} . As a consequence:

$$\partial\varphi(x) = \{ \nabla\varphi(x) \}. \tag{A15}$$

One can prove that $y = \nabla\varphi(x)$ (see [16]). Therefore, $\pi = (Id, T)\#\mu = (Id, \nabla\varphi)\#\mu$, and it is optimal. We have proven Claim 1.

Claim 2: $\pi = (Id, \nabla\varphi)\#\mu$ is the only transport plan:

Let $\hat{\varphi}$ be another convex function such that $\nu = \nabla\hat{\varphi}\#\mu$.

We want to prove $\nabla\varphi(x) = \nabla\hat{\varphi}(x)$ μ -a.e. $x \in \Omega$.

By Claim 1, $(Id, \nabla\hat{\varphi})\#\mu$ is an optimal transport plan, and the pair $(\hat{\varphi}, \hat{\varphi}^*)$ is optimal for the dual problem, just like (φ, φ^*) . Therefore:

$$\int_{\Omega} \hat{\varphi}(x)d\mu(x) + \int_{\Omega} \hat{\varphi}^*(y)d\nu(y) = \int_{\Omega} \varphi(x)d\mu(x) + \int_{\Omega} \varphi^*(y)d\nu(y).$$

Let π be the optimal transport plan associated with φ . Then:

$$\int_{\Omega \times \Omega} [\hat{\varphi}(x) + \hat{\varphi}^*(y)]d\pi(x, y) = \int_{\Omega \times \Omega} [\varphi(x) + \varphi^*(y)]d\pi(x, y) = \int_{\Omega \times \Omega} xy d\pi(x, y);$$

hence:

$$\int_{\Omega} [\hat{\varphi}(x) + \hat{\varphi}^*(\nabla\varphi(x))]d\mu(x) = \int_{\Omega} x\nabla\varphi(x)d\mu(x),$$

and:

$$\int_{\Omega} [\hat{\varphi}(x) + \hat{\varphi}^*(\nabla\varphi(x)) - x\nabla\varphi(x)]d\mu(x) = 0;$$

since $\hat{\varphi}(x) + \hat{\varphi}^*(\nabla\varphi(x)) - x\nabla\varphi(x) \geq 0$, we conclude:

$$\hat{\varphi}(x) + \hat{\varphi}^*(\nabla\varphi(x)) - x\nabla\varphi(x) = 0,$$

μ -a.e. on Ω . Hence, by (A14):

$$\nabla\varphi(x) \in \partial\hat{\varphi}(x), \quad \mu - \text{a.e. } x \in \Omega.$$

Now, since $\hat{\varphi}$ is differentiable μ -a.e. on Ω :

$$\nabla\varphi(x) = \nabla\hat{\varphi}(x).$$

Claim 2 has been proven.

Finally, we have proven not only the uniqueness of Problem (15) (analogous to the Monge–Kantorovich problem), but also the uniqueness of the gradient of a convex function $\nabla\phi$ such that $\nabla\phi\#\mu = \nu$ and:

$$\min_{\gamma \in \hat{\Pi}(\mu, \nu)} \int_{\Omega \times \Omega} |x - y|^2 d\gamma(x, y) = \int_{\Omega} |x - \nabla\phi(x)|^2 d\mu(x) = \int_{\Omega} |x - T(x)|^2 d\mu(x).$$

We have proven the theorem.

□

Appendix B. Linearization of the Monge–Ampère Equation

Assume that ϕ is very close to $\frac{|x|^2}{2}$, so $|\rho(x) \ln[\rho(x)]|$ is very close to $|\tilde{\rho}(\nabla\phi(x)) \ln[\tilde{\rho}(\nabla\phi(x))]|$. In that case, following [16,22,23], make:

$$\phi(x) \equiv \phi_\varepsilon(x) = \frac{|x|^2}{2} + \varepsilon\eta + O(\varepsilon^2),$$

and:

$$|\tilde{\rho}(\nabla\phi(x)) \ln[\tilde{\rho}(\nabla\phi(x))]| \equiv \left| \tilde{\rho}(\nabla\phi(x)) \ln[\tilde{\rho}(\nabla\phi(x))] \right|_\varepsilon = (1 + \varepsilon h + O(\varepsilon^2)) |\rho(x) \ln[\rho(x)]|.$$

with $\eta, h \in L^1(\mu)$. Then, by the Neumann series:

$$\begin{aligned} \frac{|\rho(x) \ln[\rho(x)]|}{|\tilde{\rho}(\nabla\phi(x)) \ln[\tilde{\rho}(\nabla\phi(x))]|} &= \frac{|\rho(x) \ln[\rho(x)]|}{(1 + \varepsilon h + O(\varepsilon^2)) |\rho(\nabla\phi(x)) \ln[\rho(\nabla\phi(x))]|} \\ &= \frac{(1 - \varepsilon h + O(\varepsilon^2)) |\rho(x) \ln[\rho(x)]|}{|\rho(\nabla\phi(x)) \ln[\rho(\nabla\phi(x))]|}. \end{aligned} \tag{A16}$$

On the other hand, it is easy to see that:

$$\begin{aligned} \det D^2\phi(x) &= \det D^2\left(\frac{|x|^2}{2} + \varepsilon\eta + O(\varepsilon^2)\right) \\ &= \det D^2\left(\frac{|x|^2}{2}\right) + \varepsilon \operatorname{trace}(HD^2\eta) + O(\varepsilon^2). \end{aligned}$$

Definition A1. The term **trace** ($HD^2\eta$) is called the linearization of the Monge–Ampère equation, where $H = (H)_{ij}$ is the matrix of co-factors of $D^2\left(\frac{|x|^2}{2}\right) = I$.

Then:

$$\det D^2(\phi(x)) = 1 + \varepsilon \operatorname{trace}(D^2\eta) + O(\varepsilon^2). \tag{A17}$$

Substituting (A16) and (A17) into the Monge–Ampère Equation (22) and omitting order superior terms, we get:

$$(1 + \varepsilon\Delta\eta) |\rho(\nabla\phi(x)) \ln[\rho(\nabla\phi(x))]| = (1 - \varepsilon h) |\rho(x) \ln[\rho(x)]|, \tag{A18}$$

and:

$$\begin{aligned} \left| \rho(\nabla\varphi(x)) \ln[\rho\nabla\varphi(x)] \right| &= \left| \rho \ln \rho \right| (\nabla\varphi(x)) \\ &= \left| \rho \ln \rho \right| \nabla \left(\frac{|x|^2}{2} + \varepsilon\eta \right) \\ &= \left| \rho \ln \rho \right| (x + \varepsilon\nabla\eta) \\ &= \left| \rho(x) \ln \rho(x) \right| + \varepsilon\nabla\eta \cdot \nabla \left(\left| \rho(x) \ln \rho(x) \right| \right), \end{aligned}$$

by using the order one Taylor's series. as a consequence, (A18) becomes:

$$(1 + \Delta\eta) \left[\left| \rho(x) \ln \rho(x) \right| + \varepsilon\nabla\eta \cdot \nabla \left(\left| \rho(x) \ln \rho(x) \right| \right) \right] = (1 - \varepsilon h) \left| \rho(x) \ln \rho(x) \right|,$$

then:

$$\begin{aligned} (1 - \varepsilon h) &= (1 + \Delta\eta) \left[1 + \varepsilon\nabla\eta \cdot \left(\frac{\nabla \left| \rho(x) \ln \rho(x) \right|}{\left| \rho(x) \ln \rho(x) \right|} \right) \right] \\ &= (1 + \Delta\eta) \left[1 + \varepsilon\nabla\eta \cdot \nabla \log \left| \rho(x) \ln \rho(x) \right| \right], \end{aligned}$$

and then:

$$(1 - \varepsilon h) = 1 + \varepsilon\nabla\eta \cdot \nabla \log \left| \rho(x) \ln \rho(x) \right| + \varepsilon\Delta\eta + \varepsilon^2\Delta\eta \left[\nabla\eta \cdot \nabla \log \left| \rho(x) \ln \rho(x) \right| \right];$$

omitting again order two terms, we get:

$$-h = \nabla\eta \cdot \nabla \log \left| \rho(x) \ln \rho(x) \right| + \Delta\eta,$$

or simply:

$$L\eta = h, \tag{A19}$$

with:

$$L = \left(-\Delta + \nabla \left(-\log \left| \rho(x) \ln \rho(x) \right| \right) \cdot \nabla \right).$$

We have proven that the Laplace equation can be seen as the linearized version of the Monge–Ampère equation for our proposal.

References

- Jensen, G.; Ward, R.D.; Balsam, P.D. Information: Theory, brain, and behavior. *J. Exp. Anal. Behav.* **2013**, *100*, 408–431. [[CrossRef](#)] [[PubMed](#)]
- Pregowska, A.; Szczepanski, J.; Wajnryb, E. Temporal code versus rate code for binary Information Sources. *Neurocomputing* **2016**, *216*, 756–762. [[CrossRef](#)]
- Pregowska, A.; Szczepanski, J.; Wajnryb, E. How Far can Neural Correlations Reduce Uncertainty? Comparison of Information Transmission Rates for Markov and Bernoulli Processes. *Int. J. Neural Syst.* **2019**, *29*, 1950003. [[CrossRef](#)] [[PubMed](#)]
- Harris, J.J.; Jolivet, R.; Engl, E.; Attwell, D. Energy-Efficient Information Transfer by Visual Pathway Synapses. *Curr. Biol.* **2015**, *25*, 3151–3160. [[CrossRef](#)] [[PubMed](#)]
- Harris, J.J.; Engl, E.; Attwell, D.; Jolivet, R.B. Energy-efficient information transfer at thalamocortical synapses. *PLoS Comput. Biol.* **2019**, *15*, 1–27. [[CrossRef](#)] [[PubMed](#)]
- Keshmiri, S. Entropy and the Brain: An Overview. *Entropy* **2020**, *22*, 917. [[CrossRef](#)]
- Salmasi, M.; Stemmler, M.; Glasauer, S.; Loebel, A. Synaptic Information Transmission in a Two-State Model of Short-Term Facilitation. *Entropy* **2019**, *21*, 756. [[CrossRef](#)]
- Crumiller, M.; Knight, B.; Kaplan, E. The Measurement of Information Transmitted by a Neural Population: Promises and Challenges. *Entropy* **2013**, *15*, 3507–3527. [[CrossRef](#)]
- Panzeri, S.; Piasini, E. Information Theory in Neuroscience. *Entropy* **2019**, *21*, 62. [[CrossRef](#)]
- Isomura, T., A Measure of Information Available for Inference. *Entropy* **2018**, *20*, 512. [[CrossRef](#)]

11. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138. [[CrossRef](#)]
12. Ramstead, M.J.D.; Badcock, P.B.; Friston, K. Answering Schrodinger's question: A free-energy formulation. *Phys. Life Rev.* **2016**, *24*, 1–16. [[CrossRef](#)] [[PubMed](#)]
13. Luczak, A. Measuring neuronal branching patterns using model-based approach. *Front. Comput. Neurosci.* **2010**, *4*, 135. [[CrossRef](#)] [[PubMed](#)]
14. Bonnotte, N. Unidimensional and Evolution Methods for Optimal Transportation. Ph.D. Thesis, Scuola Normale Superiore di Pisa and Université Paris-Sud XI, Orsay, France, 2013.
15. Stephenson, D.; Patronis, A.; Holland, D.M.; Lockerby, D.A. Generalizing Murray's Law: An optimization principle for fluidic networks of arbitrary shape and scale. *J. Appl. Phys.* **2015**, *118*, 174302. [[CrossRef](#)]
16. Villani, C. *Topics in Optimal Transportation*, 1st ed.; Graduate Studies in Mathematics; American Mathematical Society: Providence, RI, USA, 2003; Volume 58.
17. Villani, C. *Optimal Transport Old and New*, 1st ed.; Grundlehren der Mathematischen Wissenschaften: A Series of Comprehensive Studies in Mathematics; Springer: Berlin/Heidelberg, Germany, 2009; Volume 338. [[CrossRef](#)]
18. Evans, L.C. Partial Differential Equations and Monge–Kantorovich Mass Transfer. *Curr. Dev. Math.* **1997**, *1997*, 65–126. [[CrossRef](#)]
19. Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
20. Applebaum, D. *Probability and Information: An Integrated Approach*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2010. [[CrossRef](#)]
21. Alarcón, T.; Castillo, J.; García-Ponce, B.; Padilla, P. Growth rate and shape as possible control mechanisms for the selection of mode development in optimal biological branching processes. *Eur. Phys. J. Spec. Top.* **2016**, *225*, 2581–2589. [[CrossRef](#)]
22. Gutierrez, C.E.; Caffarelli, L.A. Properties of the solutions of the linearized Monge–Ampère equation. *Am. J. Math.* **1997**, *119*, 423–465. [[CrossRef](#)]
23. Gutiérrez, C.E. *The Monge–Ampère Equation*, 2nd ed.; Progress in Nonlinear Differential Equations and Their Applications; Birkhäuser: Basel, Switzerland, 2016; Volume 89. [[CrossRef](#)]
24. Ni, W.M. Diffusion, cross-diffusion, and their spike-layer steady states. *Not. AMS* **1998**, *45*, 9–18.
25. Turing, A.M. The Chemical Basis of Morphogenesis. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **1952**, *237*, 37–72.
26. Zhu, X.; Yang, H. Turing Instability-Driven Biofabrication of Branching Tissue Structures: A Dynamic Simulation and Analysis Based on the Reaction–Diffusion Mechanism. *Micromachines* **2018**, *9*, 109. [[CrossRef](#)] [[PubMed](#)]
27. Meinhardt, H.; Koch, A.J.; Bernasconi, G. Models of pattern formation applied to plant development. In *Symmetry in Plants*; Series in Mathematical Biology and Medicine: Volume 4; World Scientific: Singapore, 1998; pp. 723–758. [[CrossRef](#)]
28. Cortes-Poza, Y.; Padilla-Longoria, P.; Alvarez-Buylla, E. Spatial dynamics of floral organ formation. *J. Theor. Biol.* **2018**, *454*, 30–40. [[CrossRef](#)]
29. Barrio, R.A.; Romero-Arias, J.R.; Noguez, M.A.; Azpeitia, E.; Ortiz-Gutiérrez, E.; Hernández-Hernández, V.; Cortes-Poza, Y.; Álvarez-Buylla, E. Cell Patterns Emerge from Coupled Chemical and Physical Fields with Cell Proliferation Dynamics: The Arabidopsis thaliana Root as a Study System. *PLoS Comput. Biol.* **2013**, *9*, e1003026. [[CrossRef](#)]
30. Serini, G.; Ambrosi, D.; Giraudo, E.; Gamba, A.; Preziosi, L.; Bussolino, F. Modeling the early stages of vascular network assembly. *EMBO J.* **2003**, *22*, 1771–1779. [[CrossRef](#)] [[PubMed](#)]
31. Köhn, A.; de Back, W.; Starruß, J.; Mattiotti, A.; Deutsch, A.; Perez-Pomares, J.M.; Herrero, M.A. Early Embryonic Vascular Patterning by Matrix-Mediated Paracrine Signalling: A Mathematical Model Study. *PLoS ONE* **2011**, *6*, e24175.
32. Murray, C. The Physiological Principle of Minimum Work: I. The Vascular System and the Cost of Blood Volume. *Proc. Natl. Acad. Sci. USA* **1926**, *12*, 207–214. [[CrossRef](#)]
33. Murray, C. The Physiological Principle of Minimum Work Applied to the Angle of Branching of Arteries. *J. Gen. Physiol.* **1926**, *9*, 835–841. [[CrossRef](#)]
34. McCulloh, K.A.; Sperry, J.S.; Adler, F.R. Water transport in plants obeys Murray's law. *Nature* **2003**, *421*, 939–942. [[CrossRef](#)]

35. Zheng, X.; Shen, G.; Wang, C.; Li, Y.; Dunphy, D.; Tawfique Hasan, T.; Brinker, C.J.; Su, B.L. Bio-inspired Murray materials for mass transfer and activity. *Nat. Commun.* **2017**, *8*, 14921. [[CrossRef](#)]
36. Özdemir, H.I. The structural properties of carotid arteries in carotid artery diseases—A retrospective computed tomography angiography study. *Pol. J. Radiol.* **2020**, *85*, e82–e89. [[CrossRef](#)]
37. Wang, Z.; Li, Y.; Childress, A.R.; Detre, J.A. Brain Entropy Mapping Using fMRI. *PLoS ONE* **2014**, *9*, e89948. [[CrossRef](#)] [[PubMed](#)]
38. Bobkov, S.; Ledoux, M. *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*, 1st ed.; Memoirs of the American Mathematical Society; American Mathematical Society: Providence, RI, USA, 2016; Volume 261. [[CrossRef](#)]
39. Evans, L.C.; Gariepy, R.F. *Measure Theory and Fine Properties of Functions*, Revised ed.; Textbooks in Mathematics; CRC Press: Boca Raton, FL, USA, 2015.

Sample Availability: Samples of the compounds are available from the authors.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).