

# On the Convergence Time of a Natural Dynamics for Linear Programming

Vincenzo Bonifaci

Institute for the Analysis of Systems and Informatics, National Research Council of Italy (IASI-CNR), Rome, Italy  
vincenzo.bonifaci@iasi.cnr.it

---

## Abstract

We consider a system of nonlinear ordinary differential equations for the solution of linear programming (LP) problems that was first proposed in the mathematical biology literature as a model for the foraging behavior of acellular slime mold *Physarum polycephalum*, and more recently considered as a method to solve LP instances. We study the convergence time of the continuous Physarum dynamics in the context of the linear programming problem, and derive a new time bound to approximate optimality that depends on the relative entropy between projected versions of the optimal point and of the initial point. The bound scales logarithmically with the LP cost coefficients and linearly with the inverse of the relative accuracy, establishing the efficiency of the dynamics for arbitrary LP instances with positive costs.

**1998 ACM Subject Classification** F.1.1 Models of Computation, G.1.6 Optimization

**Keywords and phrases** linear programming, natural algorithm, Physarum polycephalum, relative entropy, Mirror Descent

**Digital Object Identifier** 10.4230/LIPIcs.ISAAC.2017.17

## 1 Introduction

The theoretical analysis of natural systems has historically been the domain of mathematical biology, dynamical systems theory, and physics, but certain natural processes are capable of exhibiting remarkable information processing abilities which are often best understood from an optimization perspective. Indeed, the application of a “computational lens” to such processes has been advocated in different disciplines, and efforts are underway to identify, classify and analyze these so-called *natural algorithms* [9, 15].

One such example can be found in the slime mold *Physarum polycephalum*. *P. polycephalum* is an acellular, amoeboid slime mold in the *Mycetozoa* group. In controlled experiments, the slime mold’s capabilities have been leveraged to determine the shortest path between two locations in a network [14, 20] and, more generally, to adaptively form efficient transport networks [22]. In fact, a dynamical model proposed by the mathematical biologists to describe the time evolution of *P. polycephalum*’s network physiology [21] has been rigorously proved to be algorithmically efficient for problems such as the single-source single-sink shortest path problem [4, 8] and the minimum-cost transshipment problem [11, 18].

More recently, a variant of what we will call for short the *Physarum dynamics* has been proposed for solving linear programming (LP) problems [12]. Such dynamics is a direct mathematical extension of the one that has been studied for the shortest path and transshipment problems. It was shown that, under very mild assumptions on the linear program, the dynamics converges to an optimal LP solution [19]. However, the bound for the time of convergence of a discretization of the dynamics to an approximate solution has



© Vincenzo Bonifaci;

licensed under Creative Commons License CC-BY

28th International Symposium on Algorithms and Computation (ISAAC 2017).

Editors: Yoshio Okamoto and Takeshi Tokuyama; Article No. 17; pp. 17:1–17:12

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

only been proved to be polynomial when the LP cost coefficients are polynomially bounded and the constraint matrix has bounded maximum subdeterminant.

The contribution of this paper is to study the convergence time of the continuous Physarum dynamics in the context of the linear programming problem, and to derive a new time bound to approximate optimality that does not depend on the maximum subdeterminant of the constraint matrix, and depends only logarithmically on the LP costs, establishing efficiency for any LP instance with positive costs. The proof is based on convex duality and on a potential function that involves the relative entropy between the optimal and the current LP solution. The main technical novelty is represented by two ingredients: the use of a dimensionless potential function, and the explicit recognition of the crucial role played by the relative entropy function. We leave the study of a discretized version of the dynamics for future work, but it is natural to conjecture that some appropriate discretization should behave similarly to the continuous time dynamics. Indeed, several convex optimization methods can be interpreted as the discretization of an ordinary differential equation system, the solutions of which are guaranteed to converge to the set of minimizers; a well-known example is the interior point method [3, 13].

There is another compelling reason to study the convergence properties of the Physarum dynamics. It has been showed that, at least when started from a feasible point, this dynamics can be interpreted as a *natural gradient descent* algorithm in a space endowed with a non-Euclidean metric obtained from an entropy-like function [19]. This is also the case for certain incarnations of well-known meta-algorithms, such as Mirror Descent [5, 16], which are at the basis of very effective approximation algorithms for machine learning and convex optimization problems [2]. One can show that when the feasible LP region is the unit simplex, but independently of whether the initial point is feasible or not, the Physarum dynamics is identical to the continuous Mirror Descent dynamics of Nemirovski and Yudin [16] in the metric generated by the negative entropy function. Thus, when the feasible LP region is the simplex, the dynamics can be interpreted as a Mirror Descent method in a non-Euclidean metric [1, 10]. However, a similar connection may not hold more generally, suggesting that the Physarum dynamics is different from any known convex optimization method. A full characterization of the meta-algorithm behind the dynamics remains open; we believe it deserves to be investigated, due to its potential to suggest a novel iterative approach to linear optimization problems.

## 1.1 Linear programming and the Physarum dynamics

Let  $N$  and  $E$  be two finite index sets. Given a real matrix  $\mathbf{A} \in \mathbb{R}^{N \times E}$ , a positive vector  $\mathbf{c} \in \mathbb{R}_{>0}^E$ , and a vector  $\mathbf{b} \in \mathbb{R}^N$ , we consider the linear programming problem

$$\begin{aligned} \min \text{cost}(\mathbf{x}) & \\ \text{s.t. } \mathbf{Ax} = \mathbf{b} & \\ \mathbf{x} \geq \mathbf{0}, \mathbf{x} \in \mathbb{R}^E & \end{aligned} \tag{1}$$

where  $\text{cost}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{c}^\top \mathbf{x}$ . We assume that  $\mathbf{A}$  has full rank and that a nonzero optimal solution to (1) exists; uniqueness is not required. We denote by  $\mathbf{x}^*$  an arbitrary optimal solution to (1), and denote by  $\text{opt} \stackrel{\text{def}}{=} \text{cost}(\mathbf{x}^*)$  its value.

We describe the (*directed*) *Physarum dynamics* [4, 7, 8, 11, 18, 19, 21] that solves (1). Let  $\mathbf{x} \in \mathbb{R}_{>0}^E$  be a positive vector, and let  $\mathbf{C}$  be the diagonal matrix with entries  $x_j/c_j$ , for  $j \in E$ . Let  $\mathbf{L} \stackrel{\text{def}}{=} \mathbf{ACA}^\top$ ; the matrix  $\mathbf{L}$  is nonsingular and positive definite. Let  $\mathbf{p} \in \mathbb{R}^N$  be the

unique solution to  $\mathbf{L}\mathbf{p} = \mathbf{b}$ , and let  $\mathbf{q} \stackrel{\text{def}}{=} \mathbf{C}\mathbf{A}^\top \mathbf{p}$ . The *Physarum dynamics* for the linear program (1) is

$$\dot{x}_j(t) = q_j(t) - x_j(t) \quad \text{for all } j \in E \quad (2)$$

over the domain  $\Omega \stackrel{\text{def}}{=} \mathbb{R}_{>0}^E$ , where we used the notation  $\dot{x}_j(t) \stackrel{\text{def}}{=} (d/dt)x_j(t)$ . In vector notation, and omitting the implicit dependency on time, the Physarum dynamics can be written as

$$\dot{\mathbf{x}} = \mathbf{C}\mathbf{A}^\top \mathbf{L}^{-1} \mathbf{b} - \mathbf{x}. \quad (3)$$

The dynamical system has an initial condition of the form  $\mathbf{x}(0) = \mathbf{s}$  for some  $\mathbf{s} \in \mathbb{R}_{>0}^E$ . Existence of a solution  $\mathbf{x}(t)$  to (3) for  $t \in [0, \infty)$  has been proved by Straszak and Vishnoi [19, Theorem 1.1]. The system (3) is well-defined irrespective of whether the starting vector  $\mathbf{x}(0)$  satisfies  $\mathbf{A}\mathbf{x}(0) = \mathbf{b}$  or not; the case where it does is referred to as the *feasible start* case. In the special case where  $\mathbf{A}$  is derived from the signed incidence matrix of a graph, problem (1) is a *minimum-cost transshipment problem* and several of the quantities defined above have an intuitive interpretation; we refer to Section 2.2 for details.

## 1.2 Our contribution

From previous results, it is known that the solution to (3) exists, and that it converges to a feasible and optimal solution of the linear program (1). The known bound on the convergence time, however, depends on the largest absolute value of a subdeterminant of the constraint matrix  $\mathbf{A}$ . Our main contribution is to show that, in the case of feasible start, this dependence is unnecessary, and that one can obtain a bound that only depends logarithmically on the ratio between the starting cost and the optimal cost, and on the relative entropy of the optimal solution with respect to the starting solution. More precisely, we prove the following theorem.

► **Theorem 1.** *For a feasible initial condition  $\mathbf{s} \in \mathbb{R}_{>0}^E$ , consider the solution  $\mathbf{x} : [0, \infty) \rightarrow \Omega$  to the Physarum dynamics (3) with  $\mathbf{x}(0) = \mathbf{s}$ . Then  $\mathbf{x}(t)$  is a feasible solution to (1) for any  $t \geq 0$ , and for any  $\epsilon > 0$ , it holds that  $\text{cost}(\mathbf{x}(t)) \leq (1 + \epsilon)\text{opt}$  for all*

$$t \geq \frac{6}{\epsilon} \left( \ln \frac{\text{cost}(\mathbf{x}(0))}{\text{opt}} + \text{KL}(\boldsymbol{\xi}^*, \boldsymbol{\xi}(0)) \right),$$

where  $\text{KL}(\cdot, \cdot)$  denotes the relative entropy (Kullback-Leibler divergence) between distributions, and  $\xi_j(0) \stackrel{\text{def}}{=} c_j x_j(0) / \text{cost}(\mathbf{x}(0))$ ,  $\xi_j^* \stackrel{\text{def}}{=} c_j x_j^* / \text{opt}$  for  $j \in E$ . In particular,  $\text{cost}(\mathbf{x}(t)) \leq (1 + \epsilon)\text{opt}$  for all

$$t \geq \frac{6}{\epsilon} \left( 2 \ln \frac{\text{cost}(\mathbf{x}(0))}{\text{opt}} + \ln \mu \right),$$

where  $\mu \stackrel{\text{def}}{=} \max_{j \in E} x_j^* / x_j(0)$ .

We remark that our result applies to the continuous formulation of the dynamics, and not necessarily to its Euler discretization that has been considered, together with the continuous one, in previous papers. While we conjecture that *some* discretization may be similarly efficient as the bound in Theorem 1 suggests, it may also be the case that a simple Euler discretization is insufficient to obtain such a result and that a more accurate discretization technique, such as a Runge-Kutta method, would help in this sense.

The appearance of the relative entropy term in our potential function is not an accident: it can be shown that when the feasible LP region is the unit simplex, independently of whether the dynamics is initialized with a feasible point or not, its trajectories coincide with those of the continuous Mirror Descent method of Nemirovski and Yudin [16] in a metric with geometry dictated by the negative entropy function – also known as the *information geometry* metric [1].

### 1.3 Related work

An undirected variant of the Physarum dynamics has been first proposed in the mathematical biology literature by Tero, Kobayashi and Nakagaki [21] as a model for the foraging physiology of the true slime mold *Physarum polycephalum*, an acellular organism that has been proved capable of solving shortest path problems effectively in laboratory experiments [14]. The convergence to optimality of the continuous dynamics for the shortest path problem and for its close generalization – the minimum-cost transshipment problem – has been studied analytically by Bonifaci, Mehlhorn and Varma [8] and by Ito et al. [11]. An analysis of the convergence time of the Euler discretization of the dynamics was carried out by Becchetti et al. [4] for the shortest path problem, and by Straszak and Vishnoi [18] for the minimum-cost transshipment problem. In summary, these works proved that the Physarum dynamics yields a polynomial-time approximation scheme to the shortest path problem and to the transshipment problem, assuming that the costs associated to the edges of the network are polynomially bounded. Observe that, in the statement of Theorem 1, the costs are confined within logarithms, and thus a discrete version of the dynamics that achieved a similar convergence time as in Theorem 1 would *not* require the costs to be polynomially bounded to be efficient.

The generalization of the Physarum dynamics to linear programming problems that we consider here has been first suggested by Johansson and Zou [12]. Most relevant to the current paper is the work of Straszak and Vishnoi [19], who initiated the rigorous study of the Physarum dynamics for LP problems of the form (1). Straszak and Vishnoi proved that a solution to the dynamics exists over the entire time horizon  $[0, \infty)$ , and that a bound on the convergence time of the continuous dynamics can be expressed in terms of the parameter  $\mathcal{D}$ , the *largest absolute value of a subdeterminant* of the constraint matrix  $\mathbf{A}$ , as summarized by their theorem that we quote here for comparison.

► **Theorem 2.** [19, Theorem 6.3] *Suppose that  $\mathbf{x} : [0, \infty) \rightarrow \Omega$  is any solution to the Physarum dynamics. Then, for some  $R, \nu > 0$  depending only on  $\mathbf{A}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ ,  $\mathbf{x}(0)$ , we have*

$$|\text{cost}(\mathbf{x}(t)) - \text{opt}| \leq R \cdot \exp(-\nu t),$$

where one can take  $\nu = \mathcal{D}^{-3}$  and  $R = \exp(8\mathcal{D}^2 \cdot \|\mathbf{c}\|_1 \cdot \|\mathbf{b}\|_1) \cdot (|E| + M_x)^2$ . Here,

$$\mathcal{D} \stackrel{\text{def}}{=} \max\{|\det(\mathbf{A}')| : \mathbf{A}' \text{ a square submatrix of } \mathbf{A}\},$$

and

$$M_x \stackrel{\text{def}}{=} \max\left(\max_{j \in E} x_j(0), \max_{j \in E} x_j^{-1}(0)\right).$$

Compared to this result of Straszak and Vishnoi, our contribution is to derive a bound that avoids the dependency on  $\mathcal{D}$ , thus showing that the dynamics are efficient –to the extent made precise in the statement of Theorem 1– for all linear programs of the form (1), not just

for those with special constraint matrices. Note that, in general, the bounds of Theorem 1 and 2 are incomparable: for a fixed relative error  $\epsilon$ , the time for convergence guaranteed by Theorem 1 scales polynomially in the input encoding length, while Theorem 2 only yields an exponential dependence; on the other hand, for a fixed input, Theorem 2 achieves a polynomial dependence on  $\log(1/\epsilon)$  (by taking  $t \gg \mathcal{D}^3$ ), while this is  $O(1/\epsilon)$  in Theorem 1. It is known that a simultaneous polynomial dependence on  $\log(1/\epsilon)$  and on the input length cannot be achieved [18, Appendix B].

As mentioned in the introduction, several convex optimization methods can be interpreted as discretizations of ordinary differential equation systems: for example, the Interior Point method [3, 13] and the Mirror Descent method [16, Chapter 3]. Straszak and Vishnoi [19] proved that the Physarum dynamics with feasible start can be interpreted as natural gradient descent in an appropriate information metric. Amari [1] gives an overview of natural gradient methods in the context of information geometry; see also Raskutti and Mukherjee [17].

## 1.4 Organization of the paper

The remainder of the paper is organized as follows. In Section 2 we prove some basic facts about the Physarum dynamics, including an alternative characterization of the vector  $\mathbf{q} \in \mathbb{R}^E$  defined in Section 1.1. In Section 3 we discuss the time of convergence to the feasible region of the LP and prove that the set of feasible LP solutions is an invariant set for the dynamics. In Section 4 we consider the time evolution of the cost of a feasible solution and prove our main result, Theorem 1. We summarize and discuss our findings in Section 5.

## 2 Basic properties of the dynamics

### 2.1 Notation

In the paper we reserve boldface symbols for vectors or matrices and non-boldface symbols for scalars or sets. We use the standard norms: for example, for  $\mathbf{v} \in \mathbb{R}^n$ :  $\|\mathbf{v}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^n |v_i|$ ,  $\|\mathbf{v}\|_2 \stackrel{\text{def}}{=} (\sum_{i=1}^n v_i^2)^{1/2}$ . With the notation  $\mathbf{Diag}((d_i)_{i=1}^n)$  we mean the  $n \times n$  diagonal matrix with  $d_i$  as the  $i$ th term on the main diagonal.

For the whole paper, the linear program (1) is fixed, in other words the triple  $(\mathbf{A}, \mathbf{b}, \mathbf{c})$  is fixed. Whenever the matrices or vectors  $\mathbf{C} = \mathbf{Diag}((x_j/c_j)_{j \in E})$ ,  $\mathbf{L} = \mathbf{A}\mathbf{C}\mathbf{A}^\top$ ,  $\mathbf{p} = \mathbf{L}^{-1}\mathbf{b}$ , or  $\mathbf{q} = \mathbf{C}\mathbf{A}^\top \mathbf{p}$  appear, they should be understood as computed with respect to a point  $\mathbf{x} \in \mathbb{R}_{>0}^E$ . As  $\mathbf{x} = \mathbf{x}(t)$  evolves in time with the dynamics (3), the former quantities are time-varying as well. The quantity  $\mathcal{E} \stackrel{\text{def}}{=} \mathbf{q}^\top \mathbf{R}\mathbf{q}$  is called the *energy* of the vector  $\mathbf{q}$ .

For a strictly convex and differentiable function  $\psi : \Omega \rightarrow \mathbb{R}$ , the *Bregman divergence* under  $\psi$  is the function

$$D_\psi(\mathbf{x}', \mathbf{x}) \stackrel{\text{def}}{=} \psi(\mathbf{x}') - \psi(\mathbf{x}) - \nabla\psi(\mathbf{x})^\top \cdot (\mathbf{x}' - \mathbf{x}),$$

where  $\mathbf{x}', \mathbf{x} \in \Omega$ . The Bregman divergence is in general not symmetric, but it is nonnegative and satisfies  $D_\psi(\mathbf{x}', \mathbf{x}) = 0$  if  $\mathbf{x}' = \mathbf{x}$ . The *Legendre dual* of  $\psi$  is the function  $\psi^* : \mathbb{R}^E \rightarrow \mathbb{R}$  defined by

$$\psi^*(\mathbf{y}) \stackrel{\text{def}}{=} \sup_{\mathbf{x} \in \mathbb{R}^E} (\mathbf{x}^\top \mathbf{y} - \psi(\mathbf{x})),$$

Note that a vector  $\mathbf{x} \in \Omega$  is a maximizer of  $\mathbf{x}^\top \mathbf{y} - \psi(\mathbf{x})$  iff  $\mathbf{y} = \nabla\psi(\mathbf{x})$ .

## 2.2 Intuition: The network case

The interpretation of the dynamics defined in Section 1.1 in the case where the constraint matrix is a network matrix is particularly appealing, as most of the statements below have physical interpretations in that case. Indeed, when  $\mathbf{A}$  is derived from the (signed) node-edge incidence matrix of a graph with node set  $N$  and edge set  $E$ <sup>1</sup>, the dynamics (3) have a natural interpretation in terms of electrical networks: the vector  $\mathbf{b}$  prescribes the external in-flow of current at each node, the matrix  $\mathbf{L}$  is the (reduced) graph Laplacian, the vector  $\mathbf{p}$  defines the Kirchhoff node potentials, the vector  $\mathbf{q}$  is the electrical flow, and  $\mathcal{E}$  is the energy dissipation (per unit time) of the network. In this context, Lemma 3 below is nothing but the *principle of least action* for electrical networks, also known as Thomson’s principle, stating that the electrical flow is the feasible flow that minimizes energy dissipation [6, Theorem IX.2]. The duality relation (5) becomes Ohm’s law. Proposition 4 is the conservation of energy principle, stating that if one replaces a network with a current source  $s$  and a sink  $\bar{s}$  with a single wire whose resistance is the effective resistance of the network, then the total energy in the system does not change [6, Theorem IX.3]. Proposition 5 is known as Tellegen’s theorem. Of course, the difference with classical circuit theory is that the resistor values are dynamically adjusted in response to the flow: the Physarum dynamics adjusts the edges’ resistances  $c_j/x_j$ , by updating the  $x_j$  via (2). In the network case, the dynamics converges to the solution of a minimum-cost transshipment problem with cost function prescribed by  $\mathbf{c}$  and node demands/supplies prescribed by  $\mathbf{b}$  (see for example [18, Theorem 1.2]). However, we remark that in the following statements we never require  $\mathbf{A}$  to be derived from a network matrix: our results hold for any full-rank matrix.

## 2.3 Extremal properties

We start by giving an alternative characterization of the vector  $\mathbf{q} \stackrel{\text{def}}{=} \mathbf{CA}^\top \mathbf{L}^{-1} \mathbf{b}$ .

► **Lemma 3.** *The vector  $\mathbf{q} \in \mathbb{R}^E$  defined in Section 1.1 equals the unique optimal solution to the continuous quadratic optimization problem:*

$$\begin{aligned} \min \mathbf{f}^\top \mathbf{R} \mathbf{f} \\ \text{s.t. } \mathbf{A} \mathbf{f} = \mathbf{b}. \end{aligned} \quad (4)$$

where  $\mathbf{R} \stackrel{\text{def}}{=} \mathbf{C}^{-1} \in \mathbb{R}^{E \times E}$  is the diagonal matrix with value  $r_j \stackrel{\text{def}}{=} c_j/x_j$  for the  $j$ -th element of the main diagonal. Moreover,

$$\mathbf{R} \mathbf{q} = \mathbf{A}^\top \mathbf{p}. \quad (5)$$

**Proof.** To establish (5), simply left-multiply with  $\mathbf{R}$  the identity  $\mathbf{q} = \mathbf{CA}^\top \mathbf{p}$ . It remains to establish the first part of the claim. Since the objective function in (4) is strictly convex, the problem has a unique optimal point. Consider any feasible point  $\mathbf{f}$ , and define  $\mathbf{g} = \mathbf{f} - \mathbf{q}$ . Then  $\mathbf{A} \mathbf{g} = \mathbf{b} - \mathbf{b} = \mathbf{0}$  and hence

$$\mathbf{f}^\top \mathbf{R} \mathbf{f} = (\mathbf{q} + \mathbf{g})^\top \mathbf{R} (\mathbf{q} + \mathbf{g}) = \mathbf{q}^\top \mathbf{R} \mathbf{q} + 2\mathbf{g}^\top \mathbf{R} \mathbf{q} + \mathbf{g}^\top \mathbf{R} \mathbf{g} \geq \mathbf{q}^\top \mathbf{R} \mathbf{q},$$

since  $\mathbf{g}^\top \mathbf{R} \mathbf{g} \geq 0$  and  $\mathbf{g}^\top \mathbf{R} \mathbf{q} = \mathbf{g}^\top \mathbf{A}^\top \mathbf{p} = (\mathbf{A} \mathbf{g})^\top \mathbf{p} = \mathbf{0}^\top \mathbf{p} = 0$ . Therefore, the objective function value of any feasible point  $\mathbf{f}$  is at least as large as the objective function value of  $\mathbf{q}$ . ◀

<sup>1</sup> More precisely, since we stipulated that  $\mathbf{A}$  should be full rank, we omit from  $N$  one of the nodes and omit the corresponding row from  $\mathbf{A}$ ; this corresponds to “grounding” the potential value of this node to zero.

► **Proposition 4.**  $\mathcal{E} = \mathbf{b}^\top \mathbf{L}^{-1} \mathbf{b} = \mathbf{p}^\top \mathbf{L} \mathbf{p}$ .

**Proof.**  $\mathbf{q}^\top \mathbf{R} \mathbf{q} = (\mathbf{b}^\top \mathbf{L}^{-1} \mathbf{A} \mathbf{C}) \mathbf{R} (\mathbf{C} \mathbf{A}^\top \mathbf{L}^{-1} \mathbf{b}) = (\mathbf{b}^\top \mathbf{L}^{-1}) (\mathbf{A} \mathbf{C} \mathbf{A}^\top) (\mathbf{L}^{-1} \mathbf{b}) = \mathbf{p}^\top \mathbf{L} \mathbf{p}$ . ◀

► **Proposition 5.** Let  $\mathbf{f}$  satisfy  $\mathbf{A} \dot{\mathbf{f}} = \mathbf{b}$ . Then

$$\mathbf{f}^\top \mathbf{A}^\top \mathbf{p}(t) = \mathcal{E}(t). \quad (6)$$

**Proof.** Since  $\mathbf{A} \mathbf{f} = \mathbf{b}$ , we have  $\mathbf{p}^\top \mathbf{A} \mathbf{f} = \mathbf{p}^\top \mathbf{b} = \mathbf{p}^\top \mathbf{L} \mathbf{p} = \mathcal{E}$ . The last equality is due to Proposition 4. ◀

### 3 Convergence to the feasible region

In this section we discuss the time of convergence to the feasible region  $\mathbf{A} \mathbf{x} = \mathbf{b}$ . In particular, we aim to show that feasibility is invariant under the dynamics: a feasible starting point remains feasible at all times. It turns out that a stronger property holds: the Euclidean norm of the “infeasibility error”  $\mathbf{e} \stackrel{\text{def}}{=} \mathbf{A} \mathbf{x} - \mathbf{b}$  approaches zero exponentially fast (Lemma 7).

► **Proposition 6.**  $\mathbf{A} \dot{\mathbf{x}} = \mathbf{b} - \mathbf{A} \mathbf{x}$ .

**Proof.** Using the definition of the dynamics (3),  $\mathbf{A} \dot{\mathbf{x}} = \mathbf{A} \mathbf{C} \mathbf{A}^\top \mathbf{L}^{-1} \mathbf{b} - \mathbf{A} \mathbf{x} = \mathbf{L} \mathbf{L}^{-1} \mathbf{b} - \mathbf{A} \mathbf{x} = \mathbf{b} - \mathbf{A} \mathbf{x}$ . ◀

► **Lemma 7.** Let  $\mathbf{e}(t) \stackrel{\text{def}}{=} \mathbf{A} \mathbf{x}(t) - \mathbf{b}$ . Then  $\|\mathbf{e}(t)\|_2 = \|\mathbf{e}(0)\|_2 \exp(-t)$  for any  $t > 0$ . In particular, if  $\mathbf{A} \mathbf{x}(0) = \mathbf{b}$  then  $\mathbf{A} \mathbf{x}(t) = \mathbf{b}$  for all  $t > 0$ .

**Proof.** We have

$$\begin{aligned} \frac{d}{dt} \|\mathbf{e}\|_2^2 &= \frac{d}{dt} (\mathbf{A} \mathbf{x} - \mathbf{b})^\top (\mathbf{A} \mathbf{x} - \mathbf{b}) = \frac{d}{dt} (\mathbf{x}^\top \mathbf{A}^\top \mathbf{A} \mathbf{x} - 2\mathbf{b}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{b}) = \\ &= 2\mathbf{x}^\top \mathbf{A}^\top \mathbf{A} \dot{\mathbf{x}} - 2\mathbf{b}^\top \mathbf{A} \dot{\mathbf{x}} = 2(\mathbf{x}^\top \mathbf{A}^\top - \mathbf{b}^\top) \mathbf{A} \dot{\mathbf{x}} = -2\|\mathbf{e}\|_2^2, \end{aligned}$$

where we used Proposition 6. Solution of the differential equation above yields  $\|\mathbf{e}(t)\|_2^2 = \|\mathbf{e}(0)\|_2^2 \exp(-2t)$ . Taking square roots yields the claim. ◀

### 4 Convergence in cost value

To analyze the convergence in cost values, and eventually prove Theorem 1, it will be useful to consider normalized versions of the candidate solution  $\mathbf{x}(t)$  and of the optimal vector  $\mathbf{x}^*$ . For any  $j \in E$ , let  $\xi_j(t) \stackrel{\text{def}}{=} c_j x_j(t) / \text{cost}(\mathbf{x}(t))$ ,  $\xi_j^* \stackrel{\text{def}}{=} c_j x_j^* / \text{opt}$ . Then, by construction,  $\mathbf{1}^\top \boldsymbol{\xi}^* = \mathbf{1}^\top \boldsymbol{\xi}(t) = 1$ ,  $\boldsymbol{\xi}(t) > \mathbf{0}$  and  $\boldsymbol{\xi}^* \geq \mathbf{0}$ , so  $\boldsymbol{\xi}(t)$  and  $\boldsymbol{\xi}^*$  can be interpreted as probability distributions over  $E$ . The *relative entropy* of  $\boldsymbol{\xi}^*$  with respect to  $\boldsymbol{\xi}$ , or *Kullback-Leibler divergence*  $\text{KL}(\boldsymbol{\xi}^*, \boldsymbol{\xi}(t))$ , is defined as:

$$\text{KL}(\boldsymbol{\xi}^*, \boldsymbol{\xi}(t)) \stackrel{\text{def}}{=} \sum_{j \in E} \xi_j^* \ln \frac{\xi_j^*}{\xi_j(t)}.$$

The KL divergence is the Bregman divergence of the negative entropy function  $\mathbf{x} \mapsto \sum_j x_j \ln x_j$ ; it is always nonnegative, and it is zero iff  $\boldsymbol{\xi}^* = \boldsymbol{\xi}(t)$  (see for example [1, Chapter 1]).

## 17:8 Convergence Time of Natural LP Dynamics

We can now define the potential function that is central to our analysis. Let

$$\Phi(t) \stackrel{\text{def}}{=} \ln \frac{\text{cost}(\mathbf{x}(t))}{\text{opt}} + \text{KL}(\boldsymbol{\xi}^*, \boldsymbol{\xi}(t)). \quad (7)$$

Note that the first term is nonnegative whenever  $\mathbf{x}(t)$  is feasible for the LP, and the second term is always nonnegative. Similar to previous analysis of Physarum dynamics based on potential functions [4, 18, 19], the potential function  $\Phi$  contains a term that depends on the cost of the candidate solution  $\mathbf{x}$ , and an “entropic barrier” term that captures the geometry of the feasible region: in particular, the second term penalizes distributions that get too close to the boundary of the positive orthant whenever the corresponding coordinate of the optimal solution is not on the boundary (that is,  $\xi_j(t) \approx 0$  but  $\xi_j^* > 0$ ). A difference with respect to previous papers is that the potential function (7) is dimensionless, which is natural since our aim is to bound the relative, rather than absolute, approximation error.

To proceed further, we study the evolution of the potential function over time. We start by bounding the derivative of various terms that compose it.

► **Lemma 8.** For any  $\mathbf{x}(t) \in \mathbb{R}_{>0}^E$ ,

$$\frac{d}{dt} \ln \frac{\text{cost}(\mathbf{x}(t))}{\text{opt}} \leq \left( \frac{\mathcal{E}(t)}{\text{cost}(\mathbf{x}(t))} \right)^{1/2} - 1. \quad (8)$$

**Proof.**

$$\begin{aligned} \frac{d}{dt} \ln \frac{\text{cost}(\mathbf{x})}{\text{opt}} &= \frac{\frac{d}{dt} \text{cost}(\mathbf{x})}{\text{cost}(\mathbf{x})} = \frac{\mathbf{c}^\top \dot{\mathbf{x}}}{\text{cost}(\mathbf{x})} \\ &\stackrel{(2)}{=} \frac{\mathbf{c}^\top (\mathbf{q} - \mathbf{x})}{\text{cost}(\mathbf{x})} = \frac{\mathbf{c}^\top \mathbf{q}}{\text{cost}(\mathbf{x})} - 1 \\ &\stackrel{(*)}{=} \frac{\sum_{j \in E} r_j q_j x_j}{\text{cost}(\mathbf{x})} - 1 \\ &\stackrel{(**)}{\leq} \frac{\left( \sum_{j \in E} r_j q_j^2 \right)^{1/2} \left( \sum_{j \in E} r_j x_j^2 \right)^{1/2}}{\text{cost}(\mathbf{x})} - 1 \\ &= \frac{(\mathcal{E} \text{cost}(\mathbf{x}))^{1/2}}{\text{cost}(\mathbf{x})} - 1, \end{aligned}$$

where in the third equality we used the definition of the dynamics, in (\*) we used  $r_j = c_j/x_j$ , and in (\*\*) we used the Cauchy-Schwarz inequality. For the last equality, we used the definition of the energy  $\mathcal{E} = \mathbf{q}^\top \mathbf{R} \mathbf{q}$  and (once more) the fact  $r_j = c_j/x_j$ . ◀

The following lemma is instrumental in bounding the time derivative of the KL divergence term in (7).

► **Lemma 9.** For any  $\mathbf{x}(t) \in \mathbb{R}_{>0}^E$ ,

$$\frac{d}{dt} \sum_{j \in E} \frac{c_j x_j^*}{\text{opt}} \ln \frac{x_j^*}{x_j(t)} = 1 - \frac{\mathcal{E}(t)}{\text{opt}}. \quad (9)$$

**Proof.** We start by computing

$$\sum_j \frac{c_j}{\text{opt}} x_j^* \ln \frac{x_j^*}{x_j} = -\frac{1}{\text{opt}} \sum_j c_j x_j^* \ln x_j + \frac{1}{\text{opt}} \sum_j c_j x_j^* \ln x_j^*.$$

The second term above is constant, so

$$\begin{aligned}
\frac{d}{dt} \sum_j \frac{c_j}{\text{opt}} x_j^* \ln \frac{x_j^*}{x_j} &= -\frac{1}{\text{opt}} \sum_j c_j x_j^* \frac{\dot{x}_j}{x_j} = \\
&= -\frac{1}{\text{opt}} \sum_j c_j x_j^* \frac{q_j - x_j}{x_j} = \\
&= 1 - \frac{1}{\text{opt}} \sum_j r_j q_j x_j^* = \\
&= 1 - \frac{1}{\text{opt}} \mathbf{x}^{*\top} \mathbf{R} \mathbf{q} = \\
&\stackrel{(5)}{=} 1 - \frac{1}{\text{opt}} \mathbf{x}^{*\top} \mathbf{A}^\top \mathbf{p} = \\
&\stackrel{(6)}{=} 1 - \frac{\mathcal{E}}{\text{opt}}.
\end{aligned}$$

We used (5) (Lemma 3) and the alternative characterization of the energy (6) given by Proposition 5.  $\blacktriangleleft$

► **Lemma 10.** For any  $\mathbf{x}(t) \in \mathbb{R}_{>0}^E$ ,

$$\frac{d}{dt} \text{KL}(\boldsymbol{\xi}^*, \boldsymbol{\xi}(t)) \leq \left( \frac{\mathcal{E}(t)}{\text{cost}(\mathbf{x}(t))} \right)^{1/2} - \frac{\mathcal{E}(t)}{\text{opt}}.$$

**Proof.**

$$\begin{aligned}
\frac{d}{dt} \sum_j \frac{c_j x_j^*}{\text{opt}} \ln \frac{c_j x_j^* / \text{opt}}{c_j x_j / \text{cost}(\mathbf{x})} &= \frac{d}{dt} \sum_j \frac{c_j x_j^*}{\text{opt}} \ln \frac{x_j^*}{x_j} + \frac{d}{dt} \sum_j \frac{c_j x_j^*}{\text{opt}} \ln \frac{\text{cost}(\mathbf{x})}{\text{opt}} \\
&\stackrel{(9)}{=} 1 - \frac{\mathcal{E}}{\text{opt}} + \frac{d}{dt} \ln \frac{\text{cost}(\mathbf{x})}{\text{opt}} \cdot 1 \\
&\stackrel{(8)}{\leq} 1 - \frac{\mathcal{E}}{\text{opt}} + \left( \frac{\mathcal{E}}{\text{cost}(\mathbf{x})} \right)^{1/2} - 1 \\
&= \left( \frac{\mathcal{E}}{\text{cost}(\mathbf{x})} \right)^{1/2} - \frac{\mathcal{E}}{\text{opt}}.
\end{aligned}$$

In the second equality we used Lemma 9; for the inequality we applied Lemma 8.  $\blacktriangleleft$

We are ready to prove that the more expensive a solution is, the more the potential function has to decrease.

► **Lemma 11.** If  $\text{cost}(\mathbf{x}(t)) \geq (1 + \epsilon)^2 \text{opt}$  for some  $\epsilon \in (0, 1/2)$ , then  $(d/dt)\Phi(t) \leq -\epsilon/2$ .

**Proof.** Let  $\gamma \stackrel{\text{def}}{=} \mathcal{E}/\text{opt}$  and  $\delta \stackrel{\text{def}}{=} 1/(1 + \epsilon)$ . Combining Lemma 8 and Lemma 10 yields

$$\begin{aligned}
\frac{d}{dt} \Phi(t) &= 2 \left( \frac{\mathcal{E}}{\text{cost}(\mathbf{x})} \right)^{1/2} - 1 - \frac{\mathcal{E}}{\text{opt}} \\
&\stackrel{(*)}{\leq} 2\delta\gamma^{1/2} - 1 - \gamma = \\
&= -2(1 - \delta)\gamma^{1/2} - (1 - \gamma^{1/2})^2,
\end{aligned}$$

## 17:10 Convergence Time of Natural LP Dynamics

where in (\*) we used the assumption  $\text{cost}(\mathbf{x}) \geq (1 + \epsilon)^2 \text{opt}$ . Note that both summands in the last expression are negative. We distinguish two cases. If  $(1 - \gamma^{1/2})^2 \geq \epsilon/2$ , then by ignoring the first summand above we obtain

$$\frac{d}{dt}\Phi \leq -(1 - \gamma^{1/2})^2 \leq -\epsilon/2,$$

which proves the claim. Otherwise, if  $(1 - \gamma^{1/2})^2 < \epsilon/2$ , then  $\gamma^{1/2} > 1 - (\epsilon/2)^{1/2}$  and by ignoring the second summand we obtain

$$\frac{d}{dt}\Phi \leq -2(1 - \delta)(1 - (\epsilon/2)^{1/2}) \leq -2 \cdot \frac{1 - (\epsilon/2)^{1/2}}{1 + \epsilon} \epsilon \leq -2 \frac{1/2}{3/2} \epsilon < -\epsilon/2. \quad \blacktriangleleft$$

The next lemma ensures that, for feasible solutions, the energy is always a valid lower bound on the cost.

► **Lemma 12.** *Suppose  $\mathbf{x}(t) \geq 0$ ,  $\mathbf{Ax}(t) = \mathbf{b}$ . Then  $\mathcal{E}(t) \leq \text{cost}(\mathbf{x}(t))$ .*

**Proof.** By the assumption,  $\mathbf{x}(t)$  is a feasible LP solution. By Lemma 3,  $\mathbf{q}(t)$  is a minimizer of the quadratic form  $\mathbf{f}^\top \mathbf{R} \mathbf{f}$  among all vectors  $\mathbf{f}$  satisfying  $\mathbf{A} \mathbf{f} = \mathbf{b}$ . One possible such vector is  $\mathbf{x}$ . Thus,

$$\mathcal{E} = \mathbf{q}^\top \mathbf{R} \mathbf{q} \leq \mathbf{x}^\top \mathbf{R} \mathbf{x} = \sum_{j \in E} \frac{c_j}{x_j} x_j^2 = \text{cost}(\mathbf{x}). \quad (10)$$

As a corollary, by Lemma 8 the cost of a feasible solution does not increase over time.

► **Corollary 13.** *Suppose  $\mathbf{x}(t) \geq 0$ ,  $\mathbf{Ax}(t) = \mathbf{b}$ . Then  $(d/dt)\text{cost}(\mathbf{x}(t)) \leq 0$ .*

**Proof.** Combine Lemma 12 and Lemma 8. ◀

All ingredients are now into place to derive our main claim, from which Theorem 1 will directly follow.

► **Theorem 14.** *Suppose  $\mathbf{x}(0) > 0$ ,  $\mathbf{Ax}(0) = \mathbf{b}$ . Then*

- (a)  $\mathbf{x}(t)$  is feasible for LP (1) for any  $t \geq 0$ ;
- (b)  $\text{cost}(\mathbf{x}(t)) \leq (1 + \epsilon) \text{opt}$  for all

$$t \geq \frac{\Phi(0)}{\epsilon/6} = \frac{6}{\epsilon} \left( \ln \frac{\text{cost}(\mathbf{x}(0))}{\text{opt}} + \text{KL}(\boldsymbol{\xi}^*, \boldsymbol{\xi}(0)) \right).$$

**Proof.** By assumption, we start with a feasible initial solution  $\mathbf{x}(0)$ , thus by Lemma 7 the solution  $\mathbf{x}(t)$  stays feasible for all  $t \geq 0$ ; this proves point (a). By Corollary 13, the cost of  $\mathbf{x}(t)$  can only decrease as  $t$  increases. To prove point (b), assume, by contradiction, that  $\text{cost}(\mathbf{x}(t_0))$  is larger than  $(1 + \epsilon) \text{opt}$  for some  $t_0$  that is larger than  $\Phi(0)/(\epsilon/6)$ . By Lemma 11,  $(d/dt)\Phi(t) \leq -\epsilon/6$  for all  $t$  such that

$$\text{cost}(\mathbf{x}(t)) \geq (1 + \epsilon) \text{opt} = (1 + 3\epsilon') \text{opt} \geq (1 + \epsilon')^2 \text{opt},$$

where  $\epsilon' \stackrel{\text{def}}{=} \epsilon/3$ . In particular,  $(d/dt)\Phi(t) \leq -\epsilon/6$  would hold for all  $t \in [0, t_0]$ . This implies the desired contradiction, since  $\Phi(t_0) = \Phi(0) + \int_0^{t_0} \frac{d}{dt}\Phi(t) \leq \Phi(0) - (\epsilon/6)t_0$  would have to be negative. This is impossible since  $\mathbf{x}(t)$  is feasible at all times and thus  $\Phi(t)$  is nonnegative for all  $t$ . ◀

**Proof of Theorem 1.** Theorem 14 already proves the first part of Theorem 1. For the second part, observe that if

$$\mu \stackrel{\text{def}}{=} \max_{j \in E} \frac{x_j^*}{x_j(0)},$$

then

$$\begin{aligned} \text{KL}(\boldsymbol{\xi}^*, \boldsymbol{\xi}(t)) &= \sum_{j \in E} \frac{c_j x_j^*}{\text{opt}} \ln \frac{c_j x_j^* / \text{opt}}{c_j x_j / \text{cost}(\mathbf{x})} \\ &= \sum_j \frac{c_j x_j^*}{\text{opt}} \ln \frac{x_j^*}{x_j} + \sum_j \frac{c_j x_j^*}{\text{opt}} \ln \frac{\text{cost}(\mathbf{x})}{\text{opt}} \\ &\leq (\ln \mu) \cdot \sum_j \frac{c_j x_j^*}{\text{opt}} + \left( \ln \frac{\text{cost}(\mathbf{x})}{\text{opt}} \right) \cdot \sum_j \frac{c_j x_j^*}{\text{opt}} \\ &= \ln \mu + \ln \frac{\text{cost}(\mathbf{x})}{\text{opt}}. \end{aligned}$$

Substitution in the bound of Theorem 14 yields the claim. ◀

## 5 Discussion

We have shown that the Physarum dynamics converges fast for LP instances with positive costs when starting from a feasible point. More precisely, the convergence is inversely proportional in time and logarithmic on the ratio between the initial cost and the optimal one, and the ratio between coordinates of the initial vector and the optimal solution. This result avoids all dependence on the coefficients of the constraint matrix, as opposed to a previous bound which was polynomial in the maximum subdeterminant of this matrix.

We were able to study only the continuous variant of the dynamics and did not derive bounds for the discretized dynamics that could be deduced from it. However, the fact that the continuous dynamics has desirable properties and converges fast is often a solid indication that the resulting discrete algorithm might work well. Clearly, establishing this formally is a nontrivial task and improving the bounds of Straszak and Vishnoi [19] for the discrete variant remains an important open question in this setting. Moreover, our argument relied on the assumption of a feasible starting point, which is most likely not required by the result. The main obstacle, from this point of view, is to appropriately replace Lemma 12.

We also observe that the dependence of accuracy in time has been proved to be at most of order  $t^{-1}$  in our analysis, whereas gradient methods for linear programming typically have worse bounds, of the order of  $t^{-1/2}$ . It would be interesting to know if such an improved rate can be maintained when performing a time-discretization of the dynamics. From a broader perspective, as pointed out in the introduction, a full characterization of the meta-algorithm behind the Physarum dynamics remains open.

**Acknowledgments.** The author would like to thank Kurt Mehlhorn for suggesting a shorter proof of Lemma 3.

---

## References

- 1 S. Amari. *Information Geometry and Its Applications*. Springer, 2016.
- 2 S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

- 3 D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming, I. Affine and projective scaling trajectories. *Trans. of the American Mathematical Society*, 314:499–526, 1989.
- 4 L. Becchetti, V. Bonifaci, M. Dirnberger, A. Karrenbauer, and K. Mehlhorn. Physarum can compute shortest paths: Convergence proofs and complexity bounds. In *Proc. of the 40th Int. Colloquium on Automata, Languages and Programming*, volume 7966 of *Lecture Notes in Computer Science*, pages 472–483. Springer, 2013.
- 5 A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.*, 31(3):167–175, 2003.
- 6 B. Bollobás. *Modern Graph Theory*. Springer, New York, 1998.
- 7 V. Bonifaci. Physarum can compute shortest paths: A short proof. *Inf. Process. Lett.*, 113(1-2):4–7, 2013.
- 8 V. Bonifaci, K. Mehlhorn, and G. Varma. Physarum can compute shortest paths. In *Proc. of the 23rd ACM-SIAM Symposium on Discrete Algorithms*, pages 233–240. SIAM, 2012.
- 9 B. Chazelle. Natural algorithms and influence systems. *Communications of the ACM*, 55(12):101–110, 2012.
- 10 J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.
- 11 K. Ito, A. Johansson, T. Nakagaki, and A. Tero. Convergence properties for the Physarum solver. arXiv:1101.5249v1, Jan 2011.
- 12 A. Johansson and J. Y. Zou. A slime mold solver for linear programming problems. In *How the World Computes - Turing Centenary Conference and 8th Conference on Computability in Europe*, pages 344–354. Springer, 2012.
- 13 N. K. Karmarkar. Riemannian geometry underlying interior-point methods for linear programming. In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pages 51–75. American Mathematical Society, 1990.
- 14 T. Nakagaki, H. Yamada, and Á. Tóth. Maze-solving by an amoeboid organism. *Nature*, 407:470, 2000.
- 15 S. Navlakha and Z. Bar-Joseph. Algorithms in nature: the convergence of systems biology and computational thinking. *Molecular Systems Biology*, 7:546, 2011.
- 16 A. S. Nemirovski and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley, 1983.
- 17 G. Raskutti and S. Mukherjee. The information geometry of mirror descent. *IEEE Trans. Information Theory*, 61(3):1451–1457, 2015.
- 18 D. Straszak and N. K. Vishnoi. Natural algorithms for flow problems. In *Proc. of the 27th ACM-SIAM Symposium on Discrete Algorithms*, pages 1868–1883. SIAM, 2016.
- 19 D. Straszak and N. K. Vishnoi. On a natural dynamics for linear programming. In *Proc. of the 2016 ACM Conf. on Innovations in Theoretical Computer Science*, page 291. ACM, 2016.
- 20 A. Tero, R. Kobayashi, and T. Nakagaki. Physarum solver: A biologically inspired method of road-network navigation. *Physica A*, 363:115–119, 2006.
- 21 A. Tero, R. Kobayashi, and T. Nakagaki. A mathematical model for adaptive transport network in path finding by true slime mold. *Journal of Theoretical Biology*, 244:553–564, 2007.
- 22 A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki. Rules for biologically inspired adaptive network design. *Science*, 327:439–442, 2010.