Mathematical Programming
https://doi.org/10.1007/s10107-021-01644-z

FULL LENGTH PAPER

Series A

Iteratively reweighted least squares and slime mold dynamics: connection and convergence

Damian Straszak[1] · Nisheeth K. Vishnoi[2]

Received: 5 December 2019 / Accepted: 12 March 2021

Abstract

We present a connection between two dynamical systems arising in entirely different contexts: the Iteratively Reweighted Least Squares (IRLS) algorithm used in compressed sensing and sparse recovery to find a minimum ℓ_1-norm solution in an affine space, and the dynamics of a slime mold (*Physarum polycephalum*) that finds the shortest path in a maze. We elucidate this connection by presenting a new dynamical system – Meta-Algorithm – and showing that the IRLS algorithms and the slime mold dynamics can both be obtained by specializing it to disjoint sets of variables. Subsequently, and building on work on slime mold dynamics for finding shortest paths, we prove convergence and obtain complexity bounds for the Meta-Algorithm that can be viewed as a "damped" version of the IRLS algorithm. A consequence of this latter result is a slime mold dynamics to solve the undirected transshipment problem that computes a $(1+\varepsilon)-$approximate solution in time polynomial in the size of the input graph, maximum edge cost, and $\frac{1}{\varepsilon}$ – a problem that was left open by the work of (Bonifaci V et al. [10] Physarum can compute shortest paths. Kyoto, Japan, pp. 233–240).

Keywords Physarum · Iteratively Reweighted Least Squares · Dynamical Systems · Network Flows

Mathematics Subject Classification 37M99 Approximation methods and numerical treatment of dynamical systems · 92F99 Biology and other natural sciences · 68W40 Analysis of algorithms · 90C27 Combinatorial optimization · 90C25 Convex programming

---

Based on *IRLS and Slime Mold: Equivalence and Convergence* [59] by the same set of authors.

✉ Damian Straszak
damian.straszak@gmail.com

Nisheeth K. Vishnoi
nisheeth.vishnoi@yale.edu

1 Aleph Zero Foundation, Zürich, Switzerland

2 Yale University, New Haven, USA

Published online: 05 April 2021

1 Introduction

In the last few years, various connections between algorithms, optimization, and nature have been discovered. Algorithmic techniques have been used to understand biological phenomena such as how birds flock [20], how slime molds can solve a maze [10] and, when evolution is efficient [63,66]. Sometimes, such endeavors have resulted in development of genetic or bio-inspired algorithmic paradigms, one of the most famous being deep learning [3,17,28,35,43]. Even more surprisingly, these pursuits have often revealed similarities between natural and artificial algorithms, a striking example is a result of [19] that shows that the mathematical description of sexual evolution is equivalent to the multiplicative weight updates algorithm.

Here we present another such similarity – the Iteratively Reweighted Least Squares (IRLS) algorithm for the problem of finding the sparsest solution to an underdetermined system of linear equations is intimately related to the dynamics of the slime mold (Theorem 1). To establish this, we present a *Meta-Algorithm* (Fig. 1) and show that both the IRLS algorithm and the slime mold dynamics can be recovered from it by setting its "step-size" parameter $h \in (0, 1]$ and restricting it to a subset of variables. In particular, setting $h = 1$ gives the IRLS algorithm and letting $h \to 0$ results in the slime mold dynamics. Subsequently, we prove convergence and obtain complexity bounds for the Meta-Algorithm for some $h > 0$ (Theorem 2). A corollary of our result is a polynomial bound (with respect to the graph size, maximum edge cost and $\frac{1}{\varepsilon}$, with ε being the desired precision) on the convergence of the discretization of the slime mold dynamics for the transshipment problem (Theorem 3); a problem that was open since [10].

1.1 The basis pursuit problem

A classical problem in signal processing is to recover a sparse signal from a small number of linear measurements. Mathematically, this can be formulated as finding a solution to a linear system $Ax = b$ where $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$ are given and A has fewer rows than columns (i.e. $n \ll m$). Among all the solutions, one would like to recover one with the fewest non-zero entries.[1]

This problem, known as sparse recovery, is NP-hard and we cannot hope to find an efficient algorithm in general. However, it has been observed that, when dealing with real-world data, a solution to the following ℓ_1-minimization problem, also known as *basis pursuit*:

$$\min_{x\in\mathbb{R}^m} \sum_{i=1}^{m} |x_i| \quad \text{s.t. } Ax = b \tag{1}$$

is typically quite sparse, if not of optimal sparsity. It was first shown in [26,27] that the ℓ_1-norm objective is in fact equivalent to sparsity for a specific family of matrices

[1] We assume that the matrix A has rank n, i.e., all of its rows are linearly independent. Every linear system can be efficiently brought into this form by removing certain equations.

and, later, the same was argued for a class of random matrices [15]. Finally, the notion of Restricted Isometry Property (RIP) was formulated in [16] and shown to guarantee sparse recovery via (1). Consequently, optimization problems of the form (1) became important building blocks for applications in signal processing and statistics. Thus, fast algorithms for solving such problems are desired. Note that (1) can be cast as a linear program of size linear in n and m and, hence, any linear programming algorithm can be used to solve it. However, because of the special structure of the problem, many algorithms were developed which outperform standard LP solvers in terms of efficiency on real-world instances.

1.2 The transshipment problem

A fundamental problem in combinatorial optimization called the *transshipment problem* [22] is a special case of basis pursuit. In this problem one is given on input an undirected graph $G = (V, E)$, a demand $b_v \in \mathbb{Z}$ for every vertex $v \in V$ and a positive cost $c_e \in \mathbb{Z}_{>0}$ for every edge $e \in E$. The goal is to find a flow vector $f \in \mathbb{R}^E$ that minimizes the cost $\langle |f|, c\rangle = \sum_{e\in E} |f_e|c_e$. Since the constraint of f being a flow w.r.t. demands b can be written as $Bf = b$, where $B \in \mathbb{R}^{V\times E}$ is the signed incidence matrix of G, the transshipment problem indeed matches the form of (1) with the small difference that the objective is a weighted ℓ_1-norm. We note that such a weighted basis pursuit problem, with objective min $\sum_i c_i|x_i|$, is not more general than the basic variant (1) since one can simply substitute $y_i \leftarrow x_i c_i$ and rewrite the program in terms of $y \in \mathbb{R}^m$ by incorporating the weights in the matrix A. The transshipment problem generalizes several classical problems in combinatorial optimization, such as perfect matchings in bipartite graphs, maximum flow or minimum cost flow (see e.g. [22]). The history of work on these problems is especially rich and begins with the discovery of such fundamental combinatorial algorithms as the Ford-Fulkerson method [33], leads via its highly efficient refinements [30,34,38] and finally reaches the recent work on methods based on continuous optimization that currently achieve the best known asymptotic complexity bounds [44,54,56].

1.3 The two algorithms

We now describe the two approaches for solving the basis pursuit problem: the IRLS algorithm and the slime mold or Physarum dynamics. These two methods also operate in different spaces and the Physarum dynamics is a continuous-time dynamical system while the IRLS algorithm is a discrete-time dynamical system. However, they are both defined in terms of weighted ℓ_2-minimization, a problem where for a matrix $A \in \mathbb{R}^{n\times m}$, a vector $b \in \mathbb{R}^n$ and weights $w \in \mathbb{R}^m_{>0}$, one would like to find

$$\min_{x\in\mathbb{R}^m} sum_{i=1}^m w_i^{-1} x_i^2 \qquad \text{s.t.} \quad Ax = b. \tag{2}$$

It is well-known (see Fact 2) that a solution to the above always exists (unless the linear system $Ax = b$ is unsatisfiable), is unique and can be written explicitly as $WA^\top(AWA^\top)^{-1}b$ (see Fact 2), and from now on we denote it by $q(w) \in \mathbb{R}^m$.

The IRLS algorithm is defined as a discrete dynamical system for solving 1. It is initialized at any solution $z^{(0)}$ of $Ax = b$. Subsequently, given $z^{(k)}$, $z^{(k+1)}$ is determined[2] as follows:

$$z^{(k+1)} := \operatorname{argmin}_{x \in \mathbb{R}^m} \sum_{i=1}^{m} \frac{x_i^2}{\left|z_i^{(k)}\right|} \quad \text{s.t. } Ax = b. \tag{3}$$

In other words, $z^{(k+1)} = q\left(\left|z^{(k)}\right|\right)$. This gives rise to a sequence of vectors $\left(z^{(k)}\right)_{k \in \mathbb{N}}$ that we denote by IRLS $\left[A, b, z^{(0)}\right]$. The resulting algorithm does not require any preprocessing of the data or any special rules for choosing a starting point. These properties make the algorithm particularly attractive for practical use and, indeed, the IRLS algorithm is quite popular; see for instance [18,37]. However, from a theoretical viewpoint, the algorithm is still far from being understood.

It is known that no general global convergence analysis is possible since there are well-known examples to show that there are starting points for which the IRLS algorithm does not converge (see also Appendix B). The book [51] presents a local convergence result for the IRLS algorithm that assumes the starting point is sufficiently close to the optimum and no zero-entries appear in the iterates. To bypass this "zero" problem, the following *regularized* variant of the IRLS algorithm has been considered in the literature: fix a parameter $\eta > 0$ and replace $|z_i^{(k)}|$ in the denominator of (3) by $\sqrt{(z_i^{(k)})^2 + \eta^2}$. [24] gave a non-constructive local convergence result for this scheme when the matrix A satisfies a variant of RIP [7] proved that the sequence of points produced by this variant optimizes $\sum_{i=1}^{n}\left(x_i^2 + \eta^2\right)^{1/2}$ in the affine space $Ax = b$; a problem related to, but not the same as the basis pursuit problem.

Historically, the Physarum dynamics was introduced in [62] as a mathematical model of the behavior of a slime mold [47]. In this work we propose an extension of the Physarum dynamics for solving the basis pursuit problem. The Physarum dynamics is defined by the following system of differential equations, with $\sigma(t) \in \mathbb{R}^m_{\geq 0}$

$$\frac{d}{dt}\sigma(t) = |\varphi(t)| - \sigma(t), \tag{4}$$

where $\varphi(t) := \Sigma(t)A^\top(A\Sigma(t)A^\top)^{-1}b$ and $\Sigma(t)$ denotes a diagonal matrix with $\sigma(t)$ on the diagonal. (We note that $\varphi(t) = q(\sigma(t))$, i.e., it is the solution to a weighted ℓ_2-minimization problem.)

The above is a generalization of the Physarum dynamics for the shortest $s - t$ path problem in an undirected graph [62], where the ϕ and σ vectors have physical interpretations as flux and tube diameters in a network of tubular elements. For the shortest $s - t$ path case, it was shown by [10] that $\sigma(t)$ converges to the characteristic vector of the shortest $s - t$ path in G. Prior to this work, however, it was not known whether a *discretization* of this dynamics can also yield a finite algorithm, even for special case of the undirected transshipment problem.

[2] Note that if for some i and some $k \in \mathbb{N}$ it happens that $z_i^{(k)} = 0$ then the transition to $z^{(k+1)}$ is not well defined. In such a case the ith coordinate is ignored and $z_i^{(k+1)}$ is set to 0. See Remark 1 for more details.

1.4 Organization of the paper

We begin with a short Sect. 1.5 that lists basic notation used in this paper. In Sect. 2 we introduce the necessary background and we state the main results of this paper. Subsequently, we discuss related work in Sect. 3. Sect. 4 is devoted to the proofs of our main results (some proofs were also moved to the Appendix). We conclude and state open problems in Sect. 5. Finally, in Appendix B we provide an example for which the IRLS algorithm fails to converge.

1.5 Notation

Throughout the paper A denotes a real n by m matrix, i.e., $A \in \mathbb{R}^{n\times m}$. In our bounds the following quantity occasionally shows up:

$$\mathcal{D}(A) := \max \left\{ |\det(A')| : A' \text{ is a square submatrix of } A \right\}. \tag{5}$$

For a vector $x \in \mathbb{R}^m$ we denote by $|x|$ the entry-wise absolute value of x, i.e., a vector whose ith entry is $|x_i|$ for each $i \in \{1, 2, \ldots, m\}$. A capitalized version of a vector is used to denote the diagonal matrix with this particular vector on the diagonal. For instance for X (capitalized $x \in \mathbb{R}^m$) denotes a matrix in $\mathbb{R}^{m\times m}$ such that $X_{i,i} = x_i$ and $X_{i,j} = 0$ for $i \neq j$. Occasionally for a finite set A we use $\mathbb{R}^A$ to denote the set of vectors of dimension $|A|$ indexed by elements of A. Thus an element $x \in \mathbb{R}^A$ has an entry x_a for each $a \in A$. This notation naturally extends to matrices.

2 Our results

The main conceptual contribution of our paper is the discovery and formalization of a connection between the continuous Physarum dynamics and the IRLS algorithm. It is best explained via what we introduce as the *Meta-Algorithm* for solving the basis pursuit problem (1); see also Fig. 1.

2.1 Meta-Algorithm

For a given step size $h \in (0, 1]$ the Meta-Algorithm is initialized with a candidate solution $y^{(0)} \in \mathbb{R}^m$ that satisfies $Ay^{(0)} = b$ and a vector of weights $w^{(0)} \in \mathbb{R}^m_{>0}$, and proceeds according to the update rule:

$$(y^{(k+1)}, w^{(k+1)}) := (1-h)(y^{(k)}, w^{(k)}) + h(q^{(k)}, |q^{(k)}|), \tag{6}$$

where $q^{(k)} = q(w^{(k)})$ (as defined in (2)). We denote the sequence $\left((y^{(k)}, w^{(k)})\right)_{k\in\mathbb{N}}$ by

$$\mathrm{MA}\left[A, b, h, y^{(0)}, w^{(0)}\right].$$

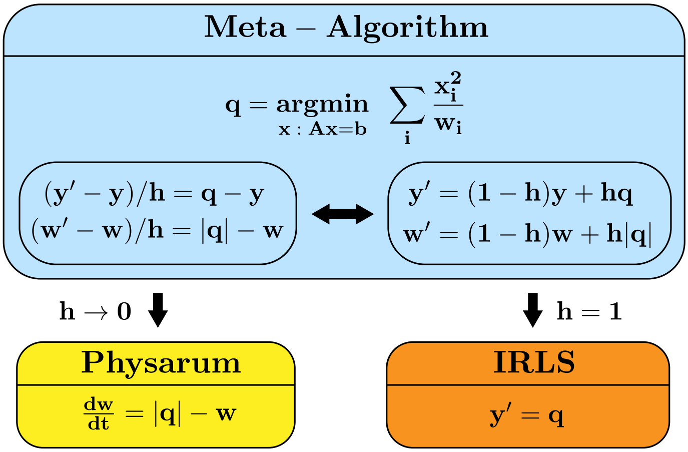

Fig. 1 An illustration of how the *Physarum dynamics* and the IRLS algorithm are derived as special cases of the Meta-Algorithm. The vectors y and w denote the current choices and y' and w' denote the vectors at the next time step. The update equations for y and w are written in two equivalent ways to make the connection apparent. The IRLS algorithm is obtained by taking the step size $h = 1$; for such an h the weight vector w is always equal to $|q| = |y|$. The Physarum algorithm is obtained in the limit as h tends to 0. The vector of weights corresponds to tube diameters σ and the weighted ℓ_2-minimizer q is equivalent to the flux φ

Note that $\left(w^{(k)}\right)_{k\in\mathbb{N}}$ depends only on h and $w^{(0)}$ and not on $y^{(0)}$.

Therefore, we also use $\mathrm{MA}\left[A, b, h, w^{(0)}\right]$ to denote the resulting sequence of weights $\left(w^{(k)}\right)_{k\in\mathbb{N}}$.

One can immediately see that for $h = 1$ the sequence $\{y^{(k)}\}_{k\in\mathbb{N}}$ coincides with the IRLS algorithm.

Remark 1 Note that when one sets $h = 1$ in the Meta-Algorithm, it might happen that the weight vector $w^{(k)}$ ends up having a zero coordinate, i.e., $w_i^{(k)} = 0$ for some $i \in \{1, 2, \ldots, m\}$ and $k \in \mathbb{N}$. In such a case, the corresponding weight in the ℓ_2-minimization problem should be treated as $+\infty$ and, hence, forces $q_i^{(k)} = 0$. In other words, in the presence of zeros in $w^{(k)}$, we set every zero-coordinate to 0 in $q^{(k)}$ and solve a weighted ℓ_2-minimization problem over the remaining coordinates. In fact, this is how the algorithm is formally defined.

Our first technical result asserts that both the IRLS and the Physarum dynamics can be seen as special cases of the Meta-Algorithm; this establishes the claimed connection between them. The Euler discretization of the Physarum dynamics that shows up in part 2 of this result has been studied in the past for the special case of the shortest path problem and flow problems [5,10].

Theorem 1 (Connection) *Let (A, b) be any instance of the basis pursuit problem.*

1. *Let $y^{(0)}$ be any solution to the linear system $Ax = b$ and let*

$$\left(y^{(k)}, w^{(k)}\right)_{k\in\mathbb{N}} := \mathrm{MA}\left[A, b, 1, y^{(0)}, \left|y^{(0)}\right|\right].$$

Then $\mathrm{IRLS}\left[A, b, y^{(0)}\right] = \left(y^{(k)}\right)_{k \in \mathbb{N}}$.

2. *Let* $\sigma(0) \in \mathbb{R}^m_{>0}$ *be an arbitrary positive vector.*
   (a) *For any* $h \in (0, 1]$*, the Meta-Algorithm* $\mathrm{MA}[A, b, h, \sigma(0)]$ *coincides with the Euler discretization of the Physarum dynamics* (4) *initialized at* $\sigma(0)$ *for step size* h.
   (b) *Let* $\sigma : [0, \infty) \to \mathbb{R}^m_{>0}$ *be a solution to the Physarum dynamics* (4) *starting at* $\sigma(0)$*. For any* $t > 0$ *we have*

$$\lim_{h \to 0^+} w_h^{(k(h))} = \sigma(t),$$

where $k(h) := \lfloor \frac{t}{h} \rfloor$ *and* $(w_h^{(k)})_{k \in \mathbb{N}} := \mathrm{MA}[A, b, h, \sigma(0)]$*. In words, for any fixed* t*, if the Meta-Algorithm and the Physarum dynamics are initiated at the same initial point* $\sigma(0)$ *then, as* $h \to 0$*, the value of the Physarum dynamics at time* t *is the same as that of the Meta-Algorithm after "infinitely many" steps.*

Once the suitable Meta-Algorithm has been defined, the proof of Theorem 1 is not hard. The first part follows from the definition of the Meta-Algorithm. In the second part, for (a) one just needs to write down the Euler discretization for the Physarum dynamics. Part 2 (b) requires showing that this discretization indeed closely tracks the continuous trajectory. While this is a standard task in the field of differential equations, this instance does not directly follow from classical convergence results, as the dynamics is not well behaved near the boundary of the domain. We refer to Appendix A for a detailed proof.
Our second technical result gives a quantitative convergence bound for the Meta-Algorithm.

Remark 2 In the quantitative convergence result below we assume that all the entries of A and b are integer. This allows us to state a clear and concrete bound depending on the maximum subdeterminant of A. In the general case, when A is not-necessarily integer or rational, the bound depends on the quantity

$$\mu_A := \max \left\{ w_i \left| a_i^\top \left(A W A^\top \right)^{-1} a_j \right| : w \in \mathbb{R}^m_{>0}, \quad i, j \in \{1, 2, \ldots, m\} \right\},$$

where a_i denotes the ith column of A; μ_A is finite by Lemma 3.

Theorem 2 (Convergence and complexity) *Let* (A, b) *be any integral instance to the basis pursuit problem and let* $x^\star \in \mathbb{R}^m$ *be any optimal solution to this instance. Suppose that* $y^{(0)}$ *and* $w^{(0)}$ *are chosen so as to satisfy* $Ay^{(0)} = b$ *and* $w^{(0)} \geq \left| y^{(0)} \right|$*. Furthermore, assume* $w_i^{(0)} \geq 1$ *for every* $i \in \{1, 2, \ldots, m\}$ *and* $\left\| w^{(0)} \right\|_1 \leq M \left\| x^\star \right\|_1$ *for some* $M \in \mathbb{R}$*. Let* $\varepsilon \in (0, 1/2)$ *and* $h \leq \frac{\varepsilon}{20 m \mathcal{D}(A)}$*. Then, for every* $\ell \in \mathbb{N}$ *we have that* $Ay^{(\ell)} = b$ *and for* $k = O\left(\frac{\ln M + \ln \|x^\star\|_1}{h \varepsilon^2}\right)$ *we have that* $\left\| y^{(k)} \right\|_1 \leq (1 + \varepsilon) \left\| x^\star \right\|_1$*, where* $\left((y^{(k)}, w^{(k)})\right)_{k \in \mathbb{N}} := \mathrm{MA}\left[A, b, h, y^{(0)}, w^{(0)}\right]$.

In the above statement $\mathcal{D}(A)$ stands for the maximum subdeterminant of A (see Eq. (5)). A few remarks are in order. The assumptions on the starting point $(y^{(0)}, w^{(0)})$ that we make in the statement are not crucial. However, they allow us to state the bounds in a simple form and make the proof much cleaner. Note that by taking *any* feasible solution $y^{(0)}$ (i.e., $Ay^{(0)} = b$) and defining $w_i^{(0)} := \max(1, |y_i^{(0)}|)$ for $i = 1, 2, \ldots, m$ we obtain an initial solution which trivially satisfies the condition of the theorem. For instance, one can obtain $y^{(0)}$ by solving the least squares problem over the subspace $Ax = b$, i.e., minimize $\|x\|_2$ instead of $\|x\|_1$ as in basis pursuit. Since the norms $\|\cdot\|_1$ and $\|\cdot\|_2$ differ only by at most a $\sqrt{m}$ multiplicative factor, one can take $M = O\left(\sqrt{m}\right)$ in the statement of Theorem 2. We remark that to guarantee convergence we require the step size to be $O(\varepsilon(m\mathcal{D}(A))^{-1})$ which might be exponentially small even when A has entries from a bounded interval. Thus in general, it does not follow that the Meta-Algorithm solves the basis pursuit problem in polynomial time. This can be deduced only for certain special cases, such as when A is a totally unimodular matrix (every subdeterminant of A is either 1 or -1).

We note that the Meta-Algorithm can be viewed as an *h-dampening* of the IRLS algorithm and, thus, Theorem 2 contributes to the mathematical understanding of the IRLS algorithm. The choice of h in Theorem 2 follows directly from our analysis and is not likely to be optimal. Our preliminary numerical simulations to test convergence of the Meta-Algorithm with random A, b show that even for constant $h > 0$ the algorithm converges and the number of iterations is much smaller than the theoretical upper bounds would suggest.

Our results on the Meta-Algorithm can be applied to obtain an iteration bound for the discrete Physarum dynamics for the transshipment problem, and thus derive Theorem 1.1 of [61]. For this result we use the Weighted Meta-Algorithm, i.e., the $q^{(k)}$ vector is now computed as

$$q^{(k)} := q\left(\frac{w_1^{(k)}}{c_1}, \ldots, \frac{w_m^{(k)}}{c_m}\right) = q(C^{-1}w^{(k)}).$$

For an instance (G, b, c) of the transshipment problem (where G is an n-node undirected graph) we denote by $b_P = \|b\|_1$ the total demand and by $c_{\max} = \max_{e \in E} c_e$ the maximum cost. Let also $f^\star \in \mathbb{R}^E$ be any optimal solution to this instance.

Theorem 3 (Iteration bound for the undirected transshipment problem) *Let (G, b, c) be an instance of the undirected transshipment problem and let $f^\star$ be an optimal solution to this instance. Choose the initial weight vector to be $w_e^{(0)} = b_P$ for every $e \in E$. Then*

1. *There exists an initial solution $y^{(0)}$ that satisfies $Ay^{(0)} = b$ and $\left|y^{(0)}\right| \leq w^{(0)}$ and it can be found in nearly linear time with respect to $|E|$. Every subsequent iterate $y^{(k)}$, for $k \in \mathbb{N}$, satisfies $Ay^{(k)} = b$.*

2. *For every $\varepsilon > 0$, by picking the step size $h := \frac{\varepsilon}{10 n c_{\max}}$ after $k = O\left(\frac{n c_{\max} \ln(n C b_P)}{\varepsilon^3}\right)$ iterations of the Weighted Meta-Algorithm we have*

$$\sum_{e \in E} c_e \left|f_e^\star\right| \leq \sum_{e \in E} c_e \left|y_e^{(k)}\right| \leq (1+\varepsilon) \sum_{e \in E} c_e \left|f_e^\star\right|.$$

Remark 3 One might observe that the iteration bound is not scale invariant with respect to c, even though the problem is (multiplying the cost by a constant does not change the instance, only multiplies the result by the same constant). This is a consequence of the fact that we require all the entries of c to be integer, and in particular the choice of the step size relies on this assumption.

By comparing the above to Theorem 2 for the case when $c \equiv 1$, one can see that up to logarithmic factors, the number of iterations is linear in n, as opposed to linear in m, which is a significant speed-up for dense graphs (since n is the number of vertices and m is the number of edges). This improvement is obtained by taking advantage of the fact that the weighted ℓ_2-minimization problem underlying the Physarum dynamics is intimately related to electrical flows (see e.g. [65]). Then, using properties of electrical flows we manage to obtain better bounds on the so-called potential differences on edges and in consequence arrive at an improved iteration bound. Also importantly the Spielman-Teng solver [56] allows us to execute each iteration[3] in $\tilde{O}(m)$ time giving yet another instance of the "Laplacian paradigm" [55,61,65].

Remark 4 This paper makes no claim (or attempt) to improve on the state-of-the-art algorithms for the basis pursuit and transshipment problem. The novelty of the paper lies in making the connection between IRLS algorithms and the slime mold dynamics and analyzing these natural dynamical systems. The fact that such simple and natural dynamics can solve fundamental problems such as the basis pursuit problem and the transshipment problem is surprising, and, prior to this work, was only known for the shortest path problem [5,10].

3 Related work

Basis pursuit and ℓ_p-regression. The basis pursuit and more generally the ℓ_p-regression problem, where one is asked to solve

$$\min_{x \in \mathbb{R}^m} \|x\|_p \qquad \text{s.t.} \quad Ax = b, \tag{7}$$

are among the most studied optimization problems. The cases $p = 1$ and $p = \infty$ can be seen as linear programs and thus can be solved using the fastest known algorithm for LP, i.e., roughly (ignoring the logarithmic dependency on the precision) in time

[3] We remark that Laplacian solvers output only approximate solutions, hence one would need to prove an equivalent of our results where the q vector is computed only approximately. The details of such a proof are long and not very enlightening. The reader is referred to the paper of Daitch and Spielman [23], where such an analysis has been carried out (see also [6,50]).

proportional to solving $\sqrt{m}$ linear systems of size $m \times m$ (here for simplicity we assume that the number of constraints in A is $n \approx m$). The remaining cases $p \in (1, \infty)$ are also convex programs hence can be solved using the Interior-point method framework developed by [48] also in time proportional to solving $\sqrt{m}$ linear systems. Interestingly, the Euclidean case $p = 2$ is special as it is equivalent to solving just a single linear system. The recent result of [12] improved the above running times for all cases $p \notin \{1, 2, \infty\}$ and showed that in fact only $\approx m^{|1/2-1/p|}$ linear systems are required. This was further improved in [1] to the current state-of-the-art: just $\approx m^{\frac{|p-2|}{2p+|p-2|}} \leq m^{1/3}$ linear systems. Even more recently, a novel adaptation of the IRLS methods proposed in [2] has been shown to yield a very practical algorithm and achieve only a slightly worse theoretical bound of $m^{\frac{|p-2|}{2p-2}}$.

3.1 Transshipment problem

This problem is equivalent to the well known problem in combinatorial optimization: minimum cost flow. In particular, it generalizes several classical tasks in combinatorial optimization such as maximum bipartite matching or maximum flow. Starting from the influential work of Ford and Fulkerson [33] a host of combinatorial algorithms for these tasks have been developed [34], for a detailed history of these developments we refer to the book [22]. The construction of a nearly-linear time solver for Laplacian linear systems by Spielman and Teng [56] led to the state-of-the-art exact algorithm for minimum cost flow [44] (using interior point methods) with complexity of $\widetilde{O}(m\sqrt{n})$ and a nearly linear time algorithm for approximate maximum flow [54] (using a non-Euclidean gradient descent). We refer to [44] for the history of algorithms based on convex optimization and Laplacian solvers. A different family of methods based on convex optimization for more general flow problems, based on rescaling, has been proposed and studied by [49,64].

3.2 IRLS

A number of different algorithms based on IRLS have been proposed for solving a variety of optimization problems. The book [51] presents (among others) the IRLS method for ℓ_1-minimization and proves a local convergence result (assuming the starting point is sufficiently close to the optimum and no zero-entries appear in the iterates). The paper [36] discusses a number of different IRLS schemes for finding sparse solutions to underdetermined linear systems. A general IRLS scheme for ℓ_p-regression (matching the setting of this paper for $p = 1$) has been proposed by [41]. There has been a significant amount of study on this scheme, see e.g., the survey [13] and [53], and a number of practical adjustments have been proposed [14,40,57]. Still, besides empirical evaluations suggesting that this method reliably converges in the regime $p \in (1.5, 3)$ not much is known about its global convergence.

We now discuss another line of work, for which rigorous convergence results are known. To circumvent mathematical difficulties related to zero-entries appearing in

IRLS iterates one can choose a small positive constant $\eta > 0$ and define a modified version of the IRLS update:

$$x^{(k+1)} = \operatorname{argmin}\left\{\sum_{i=1}^{n} \frac{x_i^2}{\sqrt{\left|x_i^{(k)}\right|^2 + \eta^2}} : x \in \mathbb{R}^n, Ax = b\right\}. \tag{8}$$

Note that the above minimization problem makes perfect sense even when $x_i^{(k)} = 0$ for some i. Consequently, it has a unique solution, for every choice of $x^{(k)}$. It was proved in [7] that the sequence of points produced by scheme (8) converges to the optimal solution of

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} \left(x_i^2 + \eta^2\right)^{1/2} \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{9}$$

The number of iterations required to get ε-close to the optimal solution is bounded by $O\left(m\left(\frac{\|y^{(0)}\|_1}{\|y^\star\|_1}\right)^2 \varepsilon^{-1}\eta^{-1}\right)$, where $y^{(0)}$ is the initial solution and $y^\star$ is the optimal solution.

The function $\sum_{i=1}^{n}\left(x_i^2 + \eta^2\right)^{1/2}$ approximates the ℓ_1-norm in the following sense:

$$\forall x \in \mathbb{R}^n \qquad \|x\|_1 \leq \sum_{i=1}^{n} \left(x_i^2 + \eta^2\right)^{1/2} \leq \|x\|_1 + n \cdot \eta.$$

In the case when the matrix A satisfies a variant of RIP (Restricted Isometry Property), [24] showed that a scheme similar to (9) (with $\eta_k \to 0$ in place of constant η) converges to the ℓ_1-optimizer. The proof relies on non-constructive arguments (compactness is repeatedly used to obtain certain accumulation points) hence no quantitative bounds on the global convergence rate follow from this analysis.

Subsequent to this work, a different regularization of the IRLS algorithm has been derived in [29] in order to come up with fast algorithms for basis pursuit (i.e., ℓ_1-minimization) and ℓ_∞-minimization. The authors show that their regularized IRLS solves both these problems up to multiplicative error $(1+\varepsilon)$ in $\widetilde{O}\left(m^{1/3}\varepsilon^{-2/3} + \varepsilon^{-2}\right)$ iterations.

The IRLS paradigm is also a popular choice in the design of effective heuristics for solving robust regression problems (see e.g. [4,9,11,21,25]). The recent result of [46] gives the first rigorous analysis of IRLS-based regression in adversarial settings.

3.3 Physarum dynamics

The discrete Physarum dynamics we propose for the basis pursuit problem can be seen as an analogue of the similarly looking, but technically very different, dynamics for linear programming studied in [39,58]. Its special case for the shortest path problem and for the transshipment problem have been studied in [5,10,45] respectively (yet the latter only studies the continuous-time dynamics and not the discretization). A generalization of Physarum dynamics from the discrete to continuous space and its connections to the Monge-Kantorovich problem were studied in [31].

Several follow-up works on Physarum dynamics appeared after publishing an initial version of this manuscript. In the work [8] the continuous-time Physarum dynamics for weighted basis pursuit is studied. The results of [8] generalize existence of solutions and convergence to the case when the weight vector is not strictly positive, but just non-negative (as long as a certain condition on the kernel of A is satisfied). The works [42] and [32] present convergence results for the "non-uniform" variant (roughly, different coordinates are updated at different rates) of the continuous Physarum dynamics.

4 Proofs

4.1 Structural results

In this section some structural results regarding the weighted ℓ_2-minimization and the behavior of the Meta Algorithm are presented. The first fact establishes a certain useful geometric condition that is maintained by the Meta-Algorithm throughout all steps of its execution, whenever it is initialized to satisfy it.

Fact 1 (Positivity and boundedness) *Suppose*

$$\left((y^{(k)}, w^{(k)})\right)_{k\in\mathbb{N}} := \mathrm{MA}\left[A, b, h, y^{(0)}, w^{(0)}\right]$$

is a sequence produced by the Meta-Algorithm for some $h \in (0, 1)$. *If* $w^{(0)} > 0$ *and* $\left|y^{(0)}\right| \leq w^{(0)}$ *then* $w^{(k)} > 0$ *and* $\left|y^{(k)}\right| \leq w^{(k)}$ *for every* $k \in \mathbb{N}$.

Proof The proof proceeds by induction. When $k = 0$, the claim is valid by the assumption on $y^{(0)}$ and $w^{(0)}$. For $k \geq 0$, we have

$$w_i^{(k+1)} = (1-h)w_i^{(k)} + h\left|q_i^{(k)}\right| > 0$$

because $h \in (0, 1)$. Similarly

$$\begin{aligned}\left|y_i^{(k+1)}\right| &= \left|(1-h)y_i^{(k)} + hq_i^{(k)}\right| \\ &\leq (1-h)\left|y_i^{(k)}\right| + h\left|q_i^{(k)}\right| \\ &\leq (1-h)w_i^{(k)} + h\left|q_i^{(k)}\right|\end{aligned}$$

$$= w_i^{(k+1)}.$$

□

The next fact summarizes some well known and useful properties of the weighted ℓ_2-minimization problem.

Fact 2 (Unique solution and its norm) *Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank n, $b \in \mathbb{R}^n$ and $w \in \mathbb{R}^m_{>0}$. Then, there exists a unique solution $q = q(w) \in \mathbb{R}^m$ to the weighted ℓ_2-minimization problem (2) and it is given by*

$$q = WA^\top (AWA^\top)^{-1} b.$$

Moreover, its weighted ℓ_2-norm satisfies

$$\sum_{i=1}^m \frac{q_i^2}{w_i} = b^\top L^{-1} b,$$

where $L := AWA^\top \in \mathbb{R}^{n\times n}$ is invertible.

The following lemma plays an important role in the proof of Theorem 2. Below, the columns of $A \in \mathbb{R}^{n\times m}$ are denoted by $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$.

Lemma 3 (Uniform upper bound) *Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank n and $w \in \mathbb{R}^m_{>0}$ be a weight vector. Then*

$$\forall i, j \in \{1, 2, \ldots, m\} \quad w_i \cdot |a_i^\top L^{-1} a_j| \le \alpha$$

where $L := AWA^\top$ and $\alpha \in \mathbb{R}$ is a constant that depends only on A. In case when A has integer entries, one can choose $\alpha := \mathcal{D}(A)$.

The fact that the upper bound is a constant (α) and does not depend on the weight vector w is instrumental in proving the convergence of the Meta-Algorithm for a *fixed* step size h (which, in turn, depends on α).

Proof of Lemma 3 Fix i and denote by p the solution to the system $Lp = a_i$. We can assume that $p^\top a_j \ge 0$ for every $j \in [m]$ by replacing the column a_j by $-a_j$ if necessary. One can easily see that such a change does not alter the problem, because L remains the same.

Let us first show that $a_i^\top L^{-1} a_i \le \frac{1}{w_i}$. Note that

$$w_i a_i a_i^\top \preceq \sum_{j=1}^m w_j a_j a_j^\top = L,$$

where $\preceq$ is the PSD order. This means that $w_i u^\top a_i a_i^\top u \le u^\top L u$, for every $u \in \mathbb{R}^n$. Let us pick $u = L^{-1} a_i$. We get

$$w_i u^\top a_i a_i^\top u \le u^\top L u$$

$$
\begin{aligned}
w_i a_i^\top L^{-1} a_i a_i^\top L^{-1} a_i &\leq a_i^\top L^{-1} L L^{-1} a_i \\
w_i (a_i^\top L^{-1} a_i)^2 &\leq a_i^\top L^{-1} a_i \\
a_i^\top L^{-1} a_i &\leq \frac{1}{w_i}.
\end{aligned}
$$

It remains to argue that for some α and for every $k \in [n]$,

$$a_k^\top p \leq \alpha a_i^\top p.$$

Fix $k \in [n]$, assume $k \neq i$. If $a_k^\top p = 0$, then we are done, assume $a_k^\top p > 0$. From $Lp = a_i$ we get

$$\sum_{j=1}^{m} w_j a_j (a_j^\top p) = a_i.$$

Hence the set $S_k \stackrel{\text{def}}{=} \{s \in \mathbb{R}^m_{\geq 0} : As = a_i \wedge s_k > 0\}$ is nonempty ($WA^\top p$ belongs to it).

To give some intuition about the set S_k note that it represents the set of all non-negative linear combinations of columns of A that give a_i as a result, with the restriction that the coefficient of a_k is positive. In the case when the considered basis pursuit instance corresponds to a shortest path problem in a graph, a_i is simply an edge (v_0, v_1) in the graph and the set S_k, roughly speaking, corresponds to all paths (and convex combinations of them) that start at v_0 and end in v_1 and traverse the edge number "k".

Take $s \in S_k$ with s_k maximum possible (it can be seen that s_k is bounded over all $s \in S_k$). Then $\sum_{j=1}^m s_j a_j = a_i$, hence $\sum_{j=1}^m s_j a_j^\top p = a_i^\top p$. Since $s_j a_j^\top p \geq 0$ for all j, we can deduce that $s_k a_k^\top p \leq a_i^\top p$ and hence $a_k^\top p \leq \frac{a_i^\top p}{s_k} = \alpha_k a_i^\top p$. It is enough to choose $\alpha = \max_k \alpha_k$.

For the quantitative bound one needs to note that α is chosen according to the following values: $\varepsilon_k = \max\{s_k : As = a_i, s \geq 0\}$. In fact $\alpha = \max_k \frac{1}{\varepsilon_k}$. Because linear programs attain optimal values in vertices, one can argue that $s^\star$ – the optimal solution to $\max\{s_k : As = a_i, s \geq 0\}$ (for some fixed k) can be chosen to be a vertex of the polyhedron $\{s : As = a_i, s \geq 0\}$. By the Cramer's rule, every positive entry of $s^\star$ is lower-bounded by $\mathcal{D}(A)^{-1}$. □

The following corollary of the above lemma is used multiple times in our proofs.

Corollary 1 *Let $A \in \mathbb{Z}^{n\times m}$ be a matrix of rank n, let $w \in \mathbb{R}^m_{>0}$ be a weight vector, and let $q = q(w)$ be the corresponding weighted ℓ_2-minimizer. Let $y \in \mathbb{R}^m$ be any solution to $Ay = b$ such that $\frac{|y_i|}{w_i} \leq K$ for every $i = 1, 2, \ldots, m$. Then*

$$\forall i \in \{1, 2, \ldots, m\} \qquad \frac{|q_i|}{w_i} \leq K \cdot \mathcal{D}(A) \cdot m.$$

Proof Let $L = AWA^\top$ (note that both L and L^{-1} are symmetric matrices), then $q = WA^\top L^{-1}b$ and hence

$$\frac{|q_i|}{w_i} = \left|a_i^\top L^{-1}b\right|.$$

Since $Ay = b$, we obtain

$$\begin{aligned}\frac{|q_i|}{w_i} &= \left|a_i^\top L^{-1}Ay\right| \\ &\leq \sum_{j=1}^{m}\left|y_j \cdot a_i^\top L^{-1}a_j\right| \\ &= \sum_{j=1}^{m}\left|y_j\right| \cdot \left|a_j^\top L^{-1}a_i\right| \\ &\overset{\text{Lemma 3}}{\leq} \sum_{j=1}^{m}\left|y_j\right| \cdot \frac{\alpha}{w_i^{(k)}} \\ &\leq K \cdot \alpha \cdot m.\end{aligned}$$

Since A is assumed to have integer entries, we can replace α by $\mathcal{D}(A)$. □

4.2 Convergence and complexity of the meta-algorithm

In this section we present a proof of Theorem 2. In the rest of this section, we assume that the starting vectors $(y^{(0)}, w^{(0)})$ satisfy the condition stated in Theorem 2. It then follows from Fact 1 that $w^{(k)} > 0$ and $\left|y^{(k)}\right| \leq w^{(k)}$ for every $k \in \mathbb{N}$. Our goal is to prove that $y^{(k)}$ approaches an optimal solution to the problem (1). Since $Ay^{(k)} = b$ for every k, the proof reduces to showing that $\left\|y^{(k)}\right\|_1 \to \|x^\star\|_1$. Towards this goal, we first introduce the potential functions that are used in Sect. 4.2.1. Subsequently, we show how these potential functions can be used to explain how the vector $y^{(k)}$ moves towards optimality. The analysis has two cases: one when the energy is significantly smaller than the cost (see Sect. 4.4.1), and one where the energy is higher than the optimal value (see Sect. 4.4.2).

4.2.1 Potential functions

Cost. We call $\left\|w^{(k)}\right\|_1$ the *cost* of the current solution. It follows from Fact 1 that $\left|y^{(k)}\right| \leq w^{(k)}$. Hence, proving that $\left\|w^{(k)}\right\|_1 \to \|x^\star\|_1$ implies the same for $y^{(k)}$. We show, in particular, that the cost decreases with k. To reason about the rate at which it decreases two additional potential functions are required.

4.3 Energy

The *energy* of the current solution is defined to be

$$E(k) := \sum_{i=1}^{m} \frac{\left(q_i^{(k)}\right)^2}{w_i^{(k)}}.$$

It corresponds to the optimal value of the ℓ_2-minimization problem that is solved at step k.

4.4 Entropy

The *relative entropy* of the current solution with respect to the optimal one, or the generalized Kullback-Leibler divergence between $x^\star$ and $w^{(k)}$, is denoted by

$$I(k) := D_{KL}\left(\left|x^\star\right|, w^{(k)}\right) = \sum_{i=1}^{m} \left|x_i^\star\right| \ln \frac{\left|x_i^\star\right|}{w_i^{(k)}} - \sum_{i=1}^{m} \left|x_i^\star\right| + \sum_{i=1}^{m} w_i^{(k)}.$$

While the proof idea and its intuitive meaning is best understood in terms of $I(k)$, it is more convenient to use a simplified variant of $I(k)$. Dropping the constant terms and the $\left\|w^{(k)}\right\|_1$ term, we arrive at the potential

$$\mathcal{B}(k) := \sum_{i=1}^{m} \left|x_i^\star\right| \ln w_i^{(k)} \tag{10}$$

which we call the *barrier*, as it resembles the logarithmic barrier function used in interior point methods; see [67].

Figure 2 depicts the evolution of the various potentials during an example run of the algorithm and illustrates the two possible cases that are discussed in the subsequent subsections.

4.4.1 Case 1: Cost is far from energy

Lemma 4 *For every $k \in \mathbb{N}$ it holds that $\left\|w^{(k+1)}\right\|_1 \le \left\|w^{(k)}\right\|_1$. Furthermore, if for some $\varepsilon \in \left(0, \frac{1}{2}\right)$, $\left\|w^{(k)}\right\|_1 > \left(1 + \frac{\varepsilon}{3}\right) E(k)$, then $\left\|w^{(k+1)}\right\|_1 \le \left(1 - \frac{h\varepsilon}{8}\right) \left\|w^{(k)}\right\|_1$.*

Proof Start by noting that

$$\begin{aligned} \left\|w^{(k)}\right\|_1 - \left\|w^{(k+1)}\right\|_1 &= h \sum_{i=1}^{m} \left(w_i^{(k)} - \left|q_i^{(k)}\right|\right) \\ &= h \left(\left\|w^{(k)}\right\|_1 - \left\|q^{(k)}\right\|_1\right). \end{aligned}$$

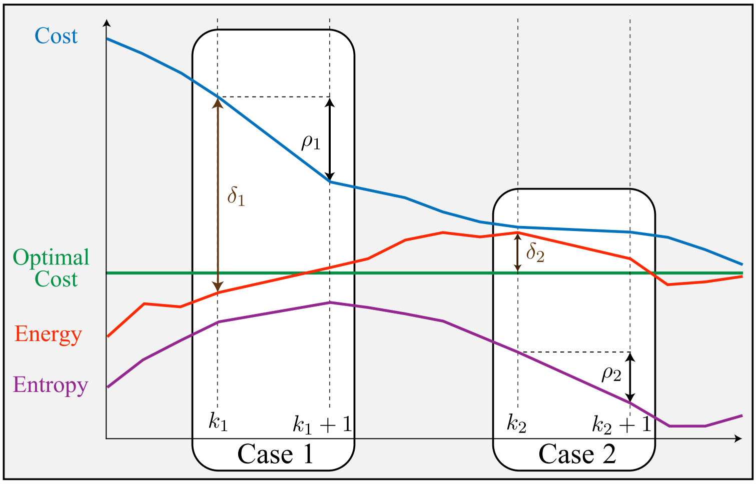

Fig. 2 A diagram showing a possible evolution of the cost, energy and entropy throughout the execution of the Meta-Algorithm. In Case 1, δ_1 denotes the difference between the cost and energy; δ_1 is large, which can be used to prove that the drop in cost ρ_1 in the next iteration will be large as well. In Case 2, the difference between cost and energy is very small, in such a case δ_2, the difference between energy and the optimal cost, is large. This can be shown to imply a large drop of entropy ρ_2 in the following iteration

Furthermore,

$$\left\| q^{(k)} \right\|_1 = \sum_{i=1}^{m} \left| q_i^{(k)} \right| = \sum_{i=1}^{m} \sqrt{w_i^{(k)}} \frac{|q_i^{(k)}|}{\sqrt{w_i^{(k)}}}.$$

Thus, by applying the Cauchy-Schwarz inequality, we obtain

$$\sum_{i=1}^{m} \sqrt{w_i^{(k)}} \frac{|q_i^{(k)}|}{\sqrt{w_i^{(k)}}} \leq \left\| w^{(k)} \right\|_1^{1/2} \cdot E(k)^{1/2}.$$

Consequently, the following holds:

$$h \left(\left\| w^{(k)} \right\|_1 - \left\| q^{(k)} \right\|_1 \right) \geq h \left\| w^{(k)} \right\|_1^{1/2} \left(\left\| w^{(k)} \right\|_1^{1/2} - E(k)^{1/2} \right).$$

Since $q^{(k)}$ minimizes the weighted ℓ_2-norm over the subspace $Ax = b$, it follows that:

$$E(k) = \sum_{i=1}^{m} \frac{\left(q_i^{(k)}\right)^2}{w_i^{(k)}} \leq \sum_{i=1}^{m} \frac{\left(y_i^{(k)}\right)^2}{w_i^{(k)}} \leq \sum_{i=1}^{m} \frac{\left(w_i^{(k)}\right)^2}{w_i^{(k)}} = \left\| w^{(k)} \right\|_1.$$

This establishes the first part of the lemma. Assume now that $\left\|w^{(k)}\right\|_1 > \left(1+\frac{\varepsilon}{3}\right)E(k)$. As a consequence,

$$\begin{aligned}\left\|w^{(k)}\right\|_1 - \left\|w^{(k+1)}\right\|_1 &\geq h\left\|w^{(k)}\right\|_1^{1/2}\left(\left\|w^{(k)}\right\|_1^{1/2} - E(k)^{1/2}\right)\\ &\geq h\left(1-\left(1+\frac{\varepsilon}{3}\right)^{-1/2}\right)\left\|w^{(k)}\right\|_1.\end{aligned}$$

To complete the proof of the lemma, it remains to note that $1-\left(1+\frac{\varepsilon}{3}\right)^{-1/2} \geq \frac{\varepsilon}{8}$. $\square$

4.4.2 Case 2: Energy is far from the optimal value

To track the convergence process for steps when the cost is close to energy we use the barrier potential $\mathcal{B}(k)$. The following lemma characterizes its behavior.

Lemma 5 *Suppose that $h \leq \frac{\varepsilon}{20\cdot m\cdot\mathcal{D}(A)}$, then for every k it holds that*

$$\mathcal{B}(k+1) \geq \mathcal{B}(k) + h\left(\left(1-\frac{\varepsilon}{10}\right)E(k) - \left(1+\frac{\varepsilon}{10}\right)\left\|x^\star\right\|_1\right).$$

Here $\mathcal{D}(A)$ is as defined in Equation (5).

The proof uses the following simple inequality

$$\forall x \in [-1/2, 1/2] \quad x - x^2 \leq \ln(1+x) \leq x. \tag{11}$$

Proof Consider the change in the barrier potential:

$$\begin{aligned}\mathcal{B}(k+1) - \mathcal{B}(k) &= \sum_{i=1}^m \left|x_i^\star\right| \ln\frac{w_i^{(k+1)}}{w_i^{(k)}}\\ &= \sum_{i=1}^m \left|x_i^\star\right| \ln\left(1+h\left(\frac{\left|q_i^{(k)}\right|}{w_i^{(k)}}-1\right)\right).\end{aligned}$$

We apply the left hand side of (11) to every summand. For simplicity let $z_i := \left(\frac{\left|q_i^{(k)}\right|}{w_i^{(k)}}-1\right)$. Thus, we obtain that

$$\begin{aligned}\mathcal{B}(k+1) - \mathcal{B}(k) &\geq \sum_{i=1}^m \left|x_i^\star\right|(hz_i - h^2z_i^2)\\ &= h\sum_{i=1}^m \left|x_i^\star\right| z_i - h^2\sum_{i=1}^m \left|x_i^\star\right| z_i^2.\end{aligned} \tag{12}$$

The linear and the quadratic terms are analyzed separately. For the linear term, note that

$$h\sum_{i=1}^{m}\left|x_i^\star\right| z_i = h\sum_{i=1}^{m}\left|x_i^\star\right|\left(\frac{\left|q_i^{(k)}\right|}{w_i^{(k)}}\right) - h\left\|x^\star\right\|_1 .$$

Henceforth, for brevity, denote the sum $\sum_{i=1}^{m}\left|x_i^\star\right|\left(\frac{\left|q_i^{(k)}\right|}{w_i^{(k)}}\right)$ by $\widetilde{E}(k)$. Then the above linear term becomes $h\cdot\widetilde{E}(k) - h\cdot\|x^\star\|_1$. To analyze the quadratic term in (12), we apply Corollary 1 to obtain:

$$\begin{aligned}
\sum_{i=1}^{m} h^2\left|x_i^\star\right| z_i^2 &\le h^2\cdot|m\mathcal{D}(A)+1|\cdot\sum_{i=1}^{m}\left|x_i^\star\right||z_i| \\
&\le h\cdot\frac{\varepsilon}{10}\cdot\sum_{i=1}^{m}\left|x_i^\star\right|\left(\frac{\left|q_i^{(k)}\right|}{w_i^{(k)}}+1\right) \\
&= h\cdot\frac{\varepsilon}{10}\widetilde{E}(k) + h\cdot\frac{\varepsilon}{10}\left\|x^\star\right\|_1 .
\end{aligned}$$

Combining the linear and quadratic order bounds, we obtain:

$$\mathcal{B}(k+1)-\mathcal{B}(k) \ge h\left(1-\frac{\varepsilon}{10}\right)\widetilde{E}(k) - h\left(1+\frac{\varepsilon}{10}\right)\left\|x^\star\right\|_1 .$$

To complete the proof, it suffices to show that $\widetilde{E}(k)\ge E(k)$. Towards this, note that

$$\begin{aligned}
\sum_{i=1}^{m}\left|x_i^\star\right|\left(\frac{\left|q_i^{(k)}\right|}{w_i^{(k)}}\right) &\ge \sum_{i=1}^{m} x_i^\star\frac{q_i^{(k)}}{w_i^{(k)}} \\
&= (x^\star)^\top\left(W^{(k)}\right)^{-1}q^{(k)} = (x^\star)^\top\left(W^{(k)}\right)^{-1}W^{(k)}A^\top L^{-1}b \\
&= (x^\star)^\top A^\top L^{-1}b = b^\top L^{-1}b
\end{aligned}$$

where $L := AW^{(k)}A^\top$. The above, together with Fact 2, gives

$$\widetilde{E}(k)\ge b^\top L^{-1}b = E(k),$$

which concludes the proof of the lemma. □

We note that the above lemma constrains the step size h to be small, of the order $\varepsilon(m\mathcal{D}(A))^{-1}$. Technically, this bound is required to apply a second order lower bound on the ln function, i.e., $x-x^2\le\ln(1+x)$, which is used to show that $\mathcal{B}(k)$ cannot drop too quickly with k. It seems that each step-by-step analysis that treats the $\mathcal{B}(k)$

and $E(k)$ separately, and does not assume any kind of "centering" of $w^{(k)}$ (as is the case in the analysis of Interior Point Methods), will necessarily require some such worst-case lower bound on $\mathcal{B}(k+1) - \mathcal{B}(k)$. Thus a way to improve the iteration bound should likely exploit more properties of the trajectories of the Meta-Algorithm.

4.4.3 Proof of Theorem 2

We would like to upper bound the number of steps until the first time when $\left\|w^{(k)}\right\|_1 \leq (1+\varepsilon)\left\|x^\star\right\|_1$. From Lemma 4, the ℓ_1-norm of $w^{(k)}$ is non-increasing with k and whenever $\left\|w^{(k)}\right\|_1 > \left(1+\frac{\varepsilon}{3}\right)E(k)$ (i.e., Case 1 occurs), $\left\|w^{(k)}\right\|_1$ decreases by a multiplicative factor of $(1-\frac{h\varepsilon}{8})$. This means that there can be at most

$$\frac{\ln\left(\frac{M}{1+\varepsilon}\right)}{\ln(1-h\varepsilon)^{-1}} = O\left(\frac{\ln M}{h\varepsilon}\right)$$

such steps. Consider a step for which $\left\|w^{(k)}\right\|_1 \leq \left(1+\frac{\varepsilon}{3}\right)E(k)$. We obtain:

$$(1+\varepsilon)\left\|x^\star\right\|_1 \leq \left\|w^{(k)}\right\|_1 \leq \left(1+\frac{\varepsilon}{3}\right)E(k).$$

This implies in particular that

$$E(k) \geq \left(1+\frac{\varepsilon}{2}\right)\left\|x^\star\right\|_1,$$

i.e., Case 2 occurs. We apply Lemma 5 to conclude that in this case

$$\mathcal{B}(k+1) \geq \mathcal{B}(k) + \frac{h\varepsilon}{5}\left\|x^\star\right\|_1.$$

We now analyze how $\mathcal{B}(k)$ changes. Start by observing that $\mathcal{B}(0) \geq 0$ (since $w_i^{(0)} \geq 1$ for every $i \in \{1, 2, \ldots, m\}$) and $\mathcal{B}(k)$ is upper bounded by $\|x^\star\|_1 \cdot (\ln M + \ln\|x^\star\|_1)$ (this holds because $\left\|w^{(k)}\right\|_1 \leq \left\|w^{(0)}\right\|_1 \leq M\|x^\star\|_1$). At every step k for which $\left\|w^{(k)}\right\|_1 > \left(1+\frac{\varepsilon}{3}\right)E(k)$, $\mathcal{B}(k)$ drops by at most

$$h\left(1+\frac{\varepsilon}{10}\right)\left\|x^\star\right\|_1 \leq 2h\left\|x^\star\right\|_1$$

by Lemma 5. Note that by the reasoning above there are at most $O\left(\frac{\ln M}{h\varepsilon}\right)$ such steps. On the other hand, if $\left\|w^{(k)}\right\|_1 \leq \left(1+\frac{\varepsilon}{3}\right)E(k)$ then $\mathcal{B}(k)$ increases by at least $\frac{h\varepsilon}{5}\|x^\star\|_1$. This means that the total decrease of $\mathcal{B}(k)$ is at most

$$O\left(\frac{\ln M}{\varepsilon}\left\|x^\star\right\|_1\right).$$

Therefore, the number of steps in which $\left\|w^{(k)}\right\|_1 \leq \left(1+\frac{\varepsilon}{3}\right) E(k)$ is at most

$$O\left(\frac{\frac{\ln M}{\varepsilon}\left\|x^{\star}\right\|_1+\left\|x^{\star}\right\|_1 \cdot\left(\ln M+\ln \left\|x^{\star}\right\|_1\right)}{\frac{h \varepsilon}{5}\left\|x^{\star}\right\|_1}\right)$$

which is bounded by $O\left(\frac{\ln M+\ln\left\|x^{\star}\right\|_1}{h \varepsilon^2}\right)$. This completes the proof of Theorem 2.

4.5 Iteration bound for undirected transshipment problem

The main component of our improved iteration bound for the transshipment problem is the following strengthening of Corollary 1 for the case where the matrix A is an incidence matrix of a graph.

Lemma 6 *Let $B \in \mathbb{R}^{V \times E}$ be an incidence matrix of an undirected graph $G=(V, E)$, $b \in \mathbb{Z}^E$ be a demand vector and $c \in \mathbb{Z}_{>0}^E$ a cost vector. Let $w \in \mathbb{R}_{>0}^E$ be any weight vector, and let $q=q(C^{-1} w)$ be the corresponding weighted ℓ_2-minimizer. Let $y \in \mathbb{R}^E$ be any solution to $B y=b$ such that $\frac{|y_e|}{w_e} \leq K$ for every $e \in E$. Then*

$$\forall e \in E \qquad \frac{|q_e|}{w_e} \leq K \cdot c_{\max} \cdot n,$$

where $c_{\max}:=\max_{e \in E} c_e$.

Proof By adapting the electrical network interpretation (see [65]) of the weighted ℓ_2-minimization problem, observe that the vector $p \in \mathbb{R}^V$ defined as

$$p:=\left(B C^{-1} W B^{\top}\right)^{-1} b$$

defines node potentials for the electrical flow on the graph G, where the resistance of an edge e is $\frac{c_e}{w_e}$ for all $e \in E$. This vector is not unique but any two such vectors differ by a multiple of the all-one vector, hence in particular the potential differences on edges: p_u-p_v for $u v \in E$ are well defined.

The quantity of our interest is $\frac{|q_e|}{w_e}$ for an edge $e=u v \in E$ which is exactly

$$\frac{|q_e|}{w_e}=\left|p_u-p_v\right| .$$

Thus it is enough to prove that for any pair of vertices $u, v \in V$ (not necessarily connected by an edge) we have

$$\left|p_u-p_v\right| \leq c_{\max} \cdot K \cdot n .$$

For this, sort all the potentials in nondecreasing order and pick two neighbouring ones p_u, p_v. In other words, take u, v, such that $p_u \leq p_v$ and for all $w \in V$ either $p_w \leq p_u$ or $p_w \geq p_v$. We show that $p_v-p_u \leq K \cdot c_{\max}$.

Assume the contrary: $p_v - p_u > K \cdot c_{\max}$. Recall that $c_{\max}$ is the maximum cost c_e over all $e \in E$. We define a partition of V into two sets $S, \bar{S}$:

$$S = \{w \in V : p_w \leq p_u\}, \qquad \bar{S} = \{w \in V : p_w \geq p_v\}.$$

Let $E_S \subseteq E$ be the set of edges going between S and $\bar{S}$. Since the graph is connected we know that $E_S \neq \emptyset$. Since the potentials in $\bar{S}$ are higher than those in S, it is clear that q sends some non-zero flow from $\bar{S}$ to S, in other words $b_S = \sum_{v \in S} b_v > 0$. Moreover, no flow is going back from S to $\bar{S}$ (as it would violate the potentials) hence

$$\sum_{e \in E_S} |q_e| = b_S.$$

On the other hand:

$$\sum_{e = u_1 v_1 \in E_S} |q_e| = \sum_{e \in E_S} \left| \frac{(p_{u_1} - p_{v_1}) w_e}{c_e} \right| > K \sum_{e \in E_S} \left| \frac{c_{\max} w_e}{c_e} \right|,$$

since for every $u_1 \in S$ and every $v_1 \in \bar{S}$ we have $|p_{u_1} - p_{v_1}| > K \cdot c_{\max}$. Note now that by our assumption there exists a solution $y \in \mathbb{R}^E$, i.e., $Ay = b$ such that for every $e \in E$ we have $|y_e| \leq K w_e$. Thus further we obtain:

$$b_S > K \sum_{e \in E_S} \left| \frac{c_{\max} w_e}{c_e} \right| \geq \sum_{e \in E_S} |y_e| \geq b_S,$$

since y being a valid flow implies that y sends at least b_S units of flow between S and $\bar{S}$. This contradiction concludes the proof. □

We are now ready to give a proof of Theorem 3.

Proof of Theorem 3 The argument follows closely that in the proof of Theorem 2 presented in Sect. 4.2. In particular the same set of potential functions and the same two cases are considered in the convergence analysis. Below we highlight where these arguments differ, and hence where the improvement comes from.

Note first that by initializing the algorithm with $w^{(0)} \equiv b_P$ we make sure that there is a feasible solution $y^{(0)}$ such that $\left|y^{(0)}\right| \leq w^{(0)}$ coordinatewise – such a solution is easy to construct by repeatedly saturating vertices using arbitrary flow paths that go from a positive demand to a negative demand. Such a solution can be efficiently constructed by finding a spanning tree of G and adding flow on paths in the tree using an efficient data structure that can support such queries in $O(\log |E|)$ time.

The only significant change in the argument occurs in Case 2: when energy is far from the optimal value. The quantity of interest there is

$$\sum_{e \in E} \ln \left(1 + h \left(\frac{\left|q_e^{(k)}\right|}{w_e^{(k)}} - 1 \right) \right)$$

and in particular a bound on $\frac{\left|q_e^{(k)}\right|}{w_e^{(k)}}$ is required in order to use the approximation $\ln(1+x) \approx x$. Given the bound from Lemma 6 one can pick $h \approx \frac{1}{n \cdot c_{\max}}$ for this to hold. Indeed from our above observation it follows that the constant K in Lemma 6 can be taken to be 1. Finally, in a later stage of this argument (for Case 2) one has to argue that

$$h \cdot \left(\frac{\left|q_e^{(k)}\right|}{w_e^{(k)}} - 1 \right) \leq \frac{\varepsilon}{10}.$$

This follows by choosing $h = \Theta\left(\frac{\varepsilon}{n \cdot c_{\max}}\right)$. □

5 Conclusion and future work

In this paper we have established convergence of a damped version of the IRLS algorithm. It is then natural to ask whether the standard version of this algorithm converges and what is its rate of convergence. The example presented in Appendix B shows that it does not converge to optimal solutions for all instances. The following two variants of the convergence question give a way to bypass this negative example:

(1) Does the IRLS algorithm converge from *almost* every starting point? Formally: is the set of "bad" starting points of measure zero?
(2) Does a stochastic variant of IRLS converge? By a stochastic version we mean one which perturbs the point in every iteration by a small amount of Gaussian noise.

A connection between physarum dynamics and IRLS

In this section we present a proof of Theorem 1. The proof has two parts that are established in the two subsequent subsections. While the first is rather trivial, the second takes some effort as it requires establishing several technical facts about the Physarum dynamics.

A.1 IRLS as the Meta-Algorithm for $h = 1$

Proof of Theorem 1, Part 1 When $h = 1$, the Meta-Algorithm proceeds as follows:

$$\left(y^{(k+1)}, w^{(k+1)}\right) = \left(q^{(k)}, \left|q^{(k)}\right|\right),$$

where $q^{(k)} = q(w^{(k)})$. Therefore, at each step, $w^{(k)} = \left|y^{(k)}\right|$. In particular, the dynamics is guided only by the y variables and is given by $y^{(k+1)} = q(|y^{(k)}|)$. This is exactly the same update rule as IRLS whose iterations correspond to $z^{(k+1)} = q(|z^{(k)}|)$. □

A.2 Physarum Dynamics as the limiting case ($h \to 0$) of the Meta-Algorithm

Fix an instance of the basis pursuit problem: $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$. Note that by Fact 1, the vector $\varphi(t)$ defined in (4) can be equivalently characterized as $q(\sigma(t))$. Let $G(\sigma) = |q(\sigma)| - \sigma$ so that the Physarum dynamics can then be written compactly as

$$\frac{d}{dt}\sigma(t) = G(\sigma(t)). \tag{13}$$

In this section, we use the ℓ_2-norm to measure lengths of vectors and, hence, $\|\sigma\|$ should be understood as $\sqrt{\sum_{i=1}^m \sigma_i^2}$. We now state some technical lemmas whose proofs appear in A.3.

Lemma 7 *(Properties of Physarum trajectories) Let $T \in \mathbb{R}_{>0}$ and $\sigma : [0, T) \to \mathbb{R}^m_{>0}$ be any solution to the Physarum dynamics (13).*

1. *For every $t \in [0, T)$ and for all $i \in \{1, 2, \ldots, m\}$, $\sigma_i(t) \geq \sigma_i(0)e^{-t}$.*
2. *The solution stays in a bounded region, i.e., $\sup_{t\in[0,T)} \|\sigma(t)\| < \infty$.*
3. *The limit $\lim_{t\to T^-} \sigma(t)$ exists and is a point in $\mathbb{R}^m_{>0}$.*

The next lemma implies the existence of a global solution of Physarum trajectories for all valid starting points.

Lemma 8 *(Existence of global solution) For every initial condition $\sigma(0) \in \mathbb{R}^m_{>0}$ there exists a global solution $\sigma : [0, +\infty) \to \mathbb{R}^m_{>0}$ to the Physarum dynamics (13).*

The final lemma, whose proof relies on the previous lemmas, implies the proof Theorem 1 part 2) b), trivially.

Lemma 9 *(Error analysis) Given $\sigma(0) \in \mathbb{R}^m_{>0}$ and $T \in \mathbb{R}_{\geq 0}$, let $\sigma : [0, T] \to \mathbb{R}^m_{>0}$ be the solution to (13) and $\sigma_{\min}(0) = \min_i \sigma_i(0)$. Then, there exists a constant $K > 0$, which depends on $\sigma(0)$ and T, such that for every step size $0 < h \leq \frac{\sigma_{\min}(0)}{2} \cdot (eK)^{-T}$, it holds for the sequence of weights $\left(w^{(k)}\right)_{k\in\mathbb{N}}$ produced by the Meta-Algorithm initialized at $w^{(0)} = \sigma^{(0)}$ with step size h that*

$$\left\| w^{(\ell)} - \sigma(h\ell) \right\| \leq hK^{h\ell}, \quad \textit{for every } \ell \in \{0, 1, \ldots, \lfloor T/h \rfloor\}.$$

Proof Fix any solution $\sigma : [0, T] \to \mathbb{R}^m_{>0}$ and let

$$\varepsilon := \frac{\sigma_{\min}}{2} e^{-T}.$$

Take F to be the closed ε-neighborhood of $\{\sigma(t) : t \in [0, T]\}$, i.e., the set of all points of distance at most ε from any point on the solution curve. Note that by Lemma 7, F is a compact subset of $\mathbb{R}^m_{>0}$. Let $L_1, L_2 > 0$ be constants such that

$$\|G(x) - G(y)\| \leq L_1 \|x - y\| \qquad \text{for all } x, y \in F,$$

$$\|\sigma(t) - \sigma(s)\| \leq L_2 \, |t - s| \qquad \text{for all } t, s \in [0, T].$$

Such an L_1 exists because G is locally Lipschitz (see Lemma 10). To see why L_2 exists one can use the fact that G is bounded on F (as a continuous function on a compact domain) together with the formula

$$\sigma(t) - \sigma(s) = \int_s^t G(\sigma(\tau)) d\tau.$$

We claim that for every $h \in (0, 1)$ and $\ell \in \mathbb{N}$ such that $h \cdot \ell \leq T$ and $hL_2 e^{L_1 T} \leq \varepsilon$,

$$\left\| w^{(\ell)} - \sigma(h\ell) \right\| \leq hL_2 e^{L_1 T}.$$

The above claim is enough to conclude the proof. Towards its proof, first define

$$d_h(\ell) := \left\| w^{(\ell)} - \sigma(h\ell) \right\|$$

for any $h \in (0, 1)$ and $\ell \in \mathbb{N}$. Recall that

$$w^{(\ell+1)} = w^{(\ell)} + hG(w^{(\ell)}).$$

We start by applying the triangle inequality to extract an error term

$$\begin{aligned} d_h(\ell+1) &= \left\| \sigma((\ell+1)h) - w^{(\ell+1)} \right\| \\ &= \left\| \sigma(\ell h) - w^{(\ell)} + \sigma((\ell+1)h) - \sigma(\ell h) - hG(w^{(\ell)}) \right\| \\ &\leq d_h(\ell) + \left\| \sigma((\ell+1)h) - \sigma(\ell h) - hG(w^{(\ell)}) \right\|. \end{aligned}$$

Next, we analyze the error term:

$$\begin{aligned} \left\| \sigma((\ell+1)h) - \sigma(\ell h) - hG(w^{(\ell)}) \right\| &= \left\| \int_{\ell h}^{(\ell+1)h} G(\sigma(\tau)) d\tau - hG(w^{(\ell)}) \right\| \\ &= \left\| \int_{\ell h}^{(\ell+1)h} \left[G(\sigma(\tau)) - G(w^{(\ell)}) \right] d\tau \right\| \\ &\leq \int_{\ell h}^{(\ell+1)h} \left\| G(\sigma(\tau)) - G(w^{(\ell)}) \right\| d\tau \\ &\leq \int_{\ell h}^{(\ell+1)h} L_1 \left\| \sigma(\tau) - w^{(\ell)} \right\| d\tau. \end{aligned}$$

Where in the last inequality we used the fact that $w^{(\ell)} \in F$ (which will be justified later). To bound the distance $\left\|\sigma(\tau) - w^{(\ell)}\right\|$ for any $\tau \in [\ell h, (\ell+1)h]$, note that

$$\left\|\sigma(\tau) - w^{(\ell)}\right\| \leq \|\sigma(\tau) - \sigma(\ell h)\| + \left\|\sigma(\ell h) - w^{(\ell)}\right\| \leq hL_2 + d_h(\ell).$$

Altogether, we obtain the following recursive bound on $d_h(\ell+1)$

$$d_h(\ell+1) \leq d_h(\ell) + hL_1(hL_2 + d_h(\ell)) = d_h(\ell)(1 + hL_1) + h^2 L_1 L_2.$$

By expanding the above expression, one can show that

$$d_h(\ell) \leq h^2 L_1 L_2 \sum_{i=0}^{\ell-1} (1 + hL_1)^i \leq hL_2(1 + hL_1)^\ell \leq hL_2 e^{h\ell L_1}.$$

In particular, whenever $h \cdot \ell \leq T$, we obtain that $d_h(\ell) \leq hL_2 e^{tL_1}$. Note that the above derivation is correct under the assumption that all the points $w^{(0)}, w^{(1)}, \ldots, w^{(\ell)}$ belong to F, however this is implied by the assumption that h is small: $hL_2 e^{L_1 T} \leq \varepsilon$, hence the upper bound on h in the statement. □

A.3 Technical lemmas and their proofs

Lemma 10 *(Local Lipschitzness) The function $G : \mathbb{R}^m_{>0} \to \mathbb{R}^m$ is locally Lipschitz, i.e., for every compact subset F of $\mathbb{R}^m_{>0}$, the restriction $G\restriction_F$ is Lipschitz.*

Proof Fix any compact subset $F \subseteq \mathbb{R}^m_{>0}$. Since $G(\sigma) = |q(\sigma)| - \sigma$ and the identity function is Lipschitz, it is enough to prove that $\sigma \mapsto |q(\sigma)|$ is Lipschitz on F. It follows from Fact 2 that

$$q(\sigma) = \Sigma A^\top (A \Sigma A^\top)^{-1} b$$

and, hence, Cramer's rule, applied to the linear system $(A\Sigma A^\top)\xi = b$ (with variables $\xi \in \mathbb{R}^n$), implies that $q_i(\sigma)$ is a rational function of the form $\frac{Q_i(\sigma)}{\det(A\Sigma A^\top)}$ where $Q_i(\sigma)$ is a polynomial. Since $\det(A\Sigma A^\top)$ is positive for $\sigma \in \mathbb{R}^m_{>0}$, $q_i(\sigma)$ is a continuously differentiable function on $\mathbb{R}^m_{>0}$. Further, since F is a compact set, the magnitude of the derivative of q is upper bounded by a finite quantity, i.e.,

$$\sup_{\sigma \in F} \|\nabla q_i(\sigma)\| \leq C \qquad \text{for all } i = 1, 2, \ldots, m$$

for some $C \in \mathbb{R}$. Now, for any $x, y \in F$ and any $i \in \{1, 2, \ldots, m\}$:

$$q_i(x) - q_i(y) = \int_0^1 \langle \nabla q_i(y + t(x-y)), x - y\rangle \, dt$$

and, thus, using the Cauchy-Schwarz inequality it follows that

$$|q_i(x) - q_i(y)| \leq \|x - y\| \cdot \sup_{\sigma \in [x,y]} \|\nabla q_i(\sigma)\| \leq C \|x - y\| .$$

Thus, the Lipschitz constant of q on F is at most $\sqrt{m} \cdot C$. ◻

Proof of Lemma 7 The first claim follows directly from Gronwall's inequality (see Sect. 2.3 in [52]) since

$$\frac{d}{dt}\sigma_i(t) = |q_i(\sigma(t))| - \sigma_i(t) \geq -\sigma_i(t).$$

For the second claim, it is enough to show that there exists a constant $C_1 > 0$ such that

$$\frac{|q_i(\sigma(t))|}{\sigma_i(t)} \leq C_1 \tag{14}$$

for all $t \in [0, T]$ and $i = 1, 2, \ldots, m$. Indeed, Gronwall's inequality then implies that $\sigma_i(t) \leq \sigma_i(0)e^{C_1 t}$, for all $t \in [0, T]$. Towards the proof of (14), let $y \in \mathbb{R}^m$ be any fixed solution to $Ay = b$. From the first claim of Lemma 7, for every $t \in [0, T]$ and $i = 1, 2, \ldots, m$ we obtain

$$\frac{|y_i|}{\sigma_i(t)} \leq |y_i|\, \sigma_i(0)^{-1}e^t \leq |y_i|\sigma_i(0)^{-1}e^T.$$

Hence, by Corollary 1, $\frac{|q_i(\sigma(t))|}{\sigma_i(t)}$ is upper bounded by a quantity C_1 that depends on $T, \sigma(0), A, b$ only; proving the second claim.

The last claim follows from the previous two and the continuity of G. Indeed, one can deduce that the solution curve $\{\sigma(t) : t \in [0, T]\}$ is contained in a compact set $F \subseteq \mathbb{R}^m_{>0}$. Denote $C_2 := \max\{\|G(\sigma)\| : \sigma \in F\}$. For any $s, t \in [0, T]$

$$\|\sigma(t) - \sigma(s)\| = \left\| \int_s^t G(\sigma(\tau))d\tau \right\| \leq C_2 \,|s - t| .$$

The above readily implies that $\lim_{t \to T^-} \sigma(t)$ exists. Since F is a compact set, the limit $\sigma(T)$ belongs to F and thus $\sigma(T) \in \mathbb{R}^m_{>0}$. ◻

Proof of Lemma 8 By Lemma 10, the function $G(\sigma)$ is locally Lipschitz and, hence, there is a maximal interval of existence $[0, T)$ of a solution $\sigma(t)$ to (13) where $0 < T \leq +\infty$ (see Theorem 1, Sect. 2.4 in [52]). Suppose that $x : [0, T) \to \mathbb{R}^m$ is a solution with $T \in \mathbb{R}_{>0}$. We show that it can be extended to a strictly larger interval $[0, T+\varepsilon)$ for some $\varepsilon > 0$. Let $\sigma(T) := \lim_{t \to T^-} \sigma(t)$; the limit exists and $\sigma(T) \in \mathbb{R}^m_{>0}$ by Lemma 7. Since $G(\sigma)$ is locally Lipschitz, one can apply the Fundamental Existence-Uniqueness theorem (see Sect. 2.2 in [52]) to obtain a solution $\tau : (T - \varepsilon, T + \varepsilon) \to \mathbb{R}^m_{>0}$ with $\tau(T) = \sigma(T)$ for some $\varepsilon > 0$. Because of uniqueness, τ and σ agree on $(T - \varepsilon, T]$

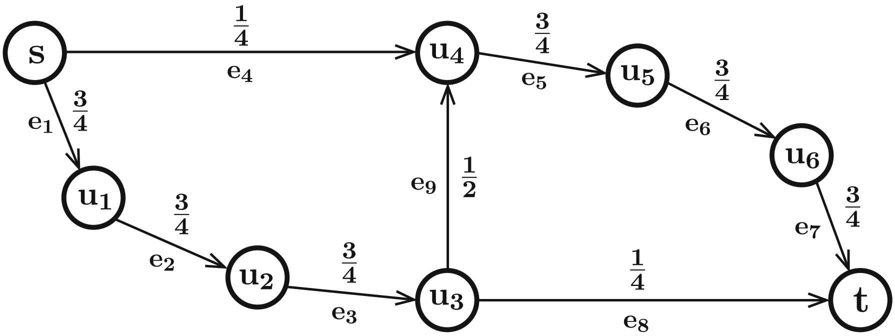

Fig. 3 The graph G together with a feasible solution $y^{(0)} \in \mathbb{R}^V$

and, hence, can be combined to yield a solution on a larger interval: $[0, T + \varepsilon)$. This concludes the proof of the lemma. □

B Example of non-convergence of IRLS

In this section, we present an example instance of the basis pursuit problem for which IRLS fails to converge to the optimal solution.

Theorem 4 *There exists an instance (A, b) of the basis pursuit problem (1) and a strictly positive point $y^{(0)} \in \mathbb{R}^m_{>0}$ (with $Ay^{(0)} = b$) such that if IRLS is initialized at $y^{(0)}$ (and $\{y^{(k)}\}_{k\in\mathbb{N}}$ is the sequence produced by IRLS) then $\|y^{(k)}\|_1$ does not converge to the optimal value.*

The proof is based on the simple observation that if IRLS reaches a point $y^{(k)}$ with $y_i^{(k)} = 0$ for some $k \in \mathbb{N}$, $i \in \{1, 2, \ldots, m\}$ then $y_i^{(l)} = 0$ for all $l > k$.

Consider an undirected graph $G = (V, E)$ with $V = \{u_0, u_1, ..., u_6, u_7\}$, $E = \{e_1, e_2, \ldots, e_9\}$, and also let $s = u_0$, $t = u_7$. G is depicted in Fig. 3.
We define $A \in \mathbb{R}^{8\times 9}$ to be the signed incidence matrix of G with edges directed according to increasing indices and let $b := (-1, 0, 0, 0, 0, 0, 0, 1)^\top$. Then the following problem

$$\min\ \|x\|_1 \qquad \text{s.t.}\ \ Ax = b$$

is equivalent to the shortest $s - t$ path problem in G. The linear system $Ax = b$ is stated explicitly in Fig. 4. The unique optimal solution is the path $s - u_4 - u_3 - t$, i.e., $y^\star = (1, 0, 0, 0, 0, 0, 0, 1, -1)^\top$ (note that we work with undirected graphs here). In particular, the edge (u_3, u_4) is in the support of the optimal vector (it corresponds to the last coordinate of $y^\star$).

Consider an initial solution $y^{(0)}$ given below

$$y^{(0)} = (3/4, 3/4, 3/4, 1/4, 3/4, 3/4, 3/4, 1/4, 1/2)^\top . \tag{15}$$

$$\begin{array}{rcl}
x_1 + \qquad\qquad\qquad +x_4 & = & 1 \\
-x_1 + x_2 & = & 0 \\
-x_2 + x_3 & = & 0 \\
-x_3 + \qquad\qquad x_8 + x_9 & = & 0 \\
-x_4 + x_5 \qquad - x_9 & = & 0 \\
-x_5 + x_6 & = & 0 \\
-x_6 + x_7 & = & 0 \\
-x_7 - x_8 & = & -1
\end{array}$$

Fig. 4 The linear system that encodes the shortest path problem in Fig. 3

Claim IRLS initialized at $y^{(0)}$ produces in one step a point $y^{(1)}$ with $y_9^{(1)} = 0$.
The above claim implies that IRLS initialized at $y^{(0)}$ (which has full support) does not converge to the optimal solution, which has 1 in the last coordinate. Thus to prove Theorem 4 it suffices to show Claim B.

Proof of Claim B IRLS chooses the next point $y^{(1)} \in \mathbb{R}^9$ according to the rule:

$$y^{(1)} = \underset{y \in \mathbb{R}^9}{\operatorname{argmin}} \sum_{i=1}^{9} \frac{x_i^2}{y_i} \qquad \text{s.t.} \quad Ax = b$$

which is the same as the unit electrical $s-t$ flow in G corresponding to edge resistances $\frac{1}{y_e}$. (This is due to the fact that electrical flows minimize energy, see [65].) In such an electrical flow the potentials of u_4 and u_3 are equal (the paths $s-u_4$ and $s-u_1-u_2-u_3$ have equal resistances), hence the flow through (u_3, u_4) is zero. □

C Remaining Proofs

Proof Of Fact 2 To prove the first claim, consider the strictly convex function $f : \mathbb{R}^m \to \mathbb{R}$ given by $f(x) := \sum_{i=1}^{m} \frac{x_i^2}{w_i}$. The optimality conditions for the convex program $\min\{f(x) : Ax = b\}$ are then given by

$$\begin{cases} AWA^\top \lambda = b \\ x = WA^\top \lambda \end{cases},$$

where $\lambda \in \mathbb{R}^n$ are Lagrangian multipliers for the linear constraints $Ax = b$. We claim that $L = AWA^\top$ is a full rank matrix. Indeed, suppose that $u \in \mathbb{R}^n$ is such that $Lu = 0$, then

$$0 = u^\top Lu = u^\top AWA^\top u = \left\| W^{1/2} A^\top u \right\|^2,$$

hence $u = 0$, since A, and consequently $W^{1/2}A$, have rank n. For this reason L is invertible and the optimizer is explicitly given by the expression $WA^\top(AWA^\top)^{-1}b$.

The proof of the second claim starts with the formula $q = WA^\top L^{-1}b$ established in the first claim. Thus, $\sum_{i=1}^m \frac{q_i^2}{w_i} = q^\top W^{-1}q$ and, hence:

$$\begin{aligned} q^\top W^{-1}q &= b^\top L^{-1}AWW^{-1}WA^\top L^{-1}b \\ &= b^\top L^{-1}(AWA^\top)L^{-1}b \\ &= b^\top L^{-1}b. \end{aligned}$$

□

References

1. Adil, D., Kyng, R., Peng, R., Sachdeva, S.: Iterative refinement for ℓ_p-norm regression. In: Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019, pp. 1405–1424 (2019). https://doi.org/10.1137/1.9781611975482.86
2. Adil, D., Peng, R., Sachdeva, S.: Fast, provably convergent IRLS algorithm for p-norm linear regression. In: H.M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E.B. Fox, R. Garnett (eds.) Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada, pp. 14166–14177 (2019). URL http://papers.nips.cc/paper/9565-fast-provably-convergent-irls-algorithm-for-p-norm-linear-regression
3. Afek, Y., Alon, N., Barad, O., Hornstein, E., Barkai, N., Bar-Joseph, Z.: A biological solution to a fundamental distributed computing problem. Science **331**(6014), 183–185 (2011). https://doi.org/10.1126/science.1193210. URL http://science.sciencemag.org/content/331/6014/183
4. Ba, D.E., Babadi, B., Purdon, P.L., Brown, E.N.: Convergence and stability of iteratively re-weighted least squares algorithms. IEEE Trans. Signal Process. **62**(1), 183–195 (2014). https://doi.org/10.1109/TSP.2013.2287685
5. Becchetti, L., Bonifaci, V., Dirnberger, M., Karrenbauer, A., Mehlhorn, K.: Physarum can compute shortest paths: Convergence proofs and complexity bounds. In: Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part II, pp. 472–483 (2013)
6. Becchetti, L., Bonifaci, V., Dirnberger, M., Karrenbauer, A., Mehlhorn, K.: Physarum can compute shortest paths: Convergence proofs and complexity bounds. In: Full version (2014)
7. Beck, A.: On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes. SIAM J. Opt. **25**(1), 185–209 (2015)
8. Becker, R., Bonifaci, V., Karrenbauer, A., Kolev, P., Mehlhorn, K.: Two results on slime mold computations. Theor. Comput. Sci. **773**, 79–106 (2019)
9. Bissantz, N., Dümbgen, L., Munk, A., Stratmann, B.: Convergence analysis of generalized iteratively reweighted least squares algorithms on convex function spaces. SIAM J. Opt. **19**(4), 1828–1845 (2009). https://doi.org/10.1137/050639132
10. Bonifaci, V., Mehlhorn, K., Varma, G.: Physarum can compute shortest paths. In: Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012, pp. 233–240 (2012)
11. Brimberg, J., Love, R.F.: Global convergence of a generalized iterative procedure for the minisum location problem with lp distances. Oper. Res. **41**(6), 1153–1163 (1993). https://doi.org/10.1287/opre.41.6.1153
12. Bubeck, S., Cohen, M.B., Lee, Y.T., Li, Y.: An homotopy method for ℓ_p regression provably beyond self-concordance and in input-sparsity time. In: Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, pp. 1130–1137 (2018). https://doi.org/10.1145/3188745.3188776

13. Burrus, C.: Iterative reweighted least squares (2012). URL https://cnx.org/contents/krkDdys0@12/Iterative-Reweighted-Least-Squares
14. Burrus, C.S., Barreto, J.A., Selesnick, I.W.: Iterative reweighted least-squares design of FIR filters. IEEE Trans. Signal Process. **42**(11), 2926–2936 (1994). https://doi.org/10.1109/78.330353
15. Candes, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theor. **52**(2), 489–509 (2006). https://doi.org/10.1109/TIT.2005.862083
16. Candès, E., Tao, T.: Decoding by linear programming. Inf. Theor., IEEE Trans. **51**(12), 4203–4215 (2005)
17. Cardelli, L., Csikász-Nagy, A.: The cell cycle switch computes approximate majority. Sci. Rep. **2**, 656 (2012). https://doi.org/10.1038/srep00656
18. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pp. 3869–3872 (2008)
19. Chastain, E., Livnat, A., Papadimitriou, C., Vazirani, U.: Algorithms, games, and evolution. Proceedings of the National Academy of Sciences **111**(29), 10620–10623 (2014). https://doi.org/10.1073/pnas.1406556111. URL http://www.pnas.org/content/111/29/10620.abstract
20. Chazelle, B.: Natural algorithms and influence systems. Commun. ACM **55**(12), 101–110 (2012). https://doi.org/10.1145/2380656.2380679
21. Chen, C., He, L., Li, H., Huang, J.: Fast iteratively reweighted least squares algorithms for analysis-based sparse reconstruction. Med. Image Analyt. **49**, 141–152 (2018). https://doi.org/10.1016/j.media.2018.08.002
22. Cook, W., Cunningham, W., Pulleyblank, W., Schrijver, A.: Comb. opt. wiley, New York (1998)
23. Daitch, S.I., Spielman, D.A.: Faster approximate lossy generalized flow via interior point algorithms. In: C. Dwork (ed.) Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008, pp. 451–460. ACM (2008). https://doi.org/10.1145/1374376.1374441
24. Daubechies, I., DeVore, R., Fornasier, M., Güntürk, C.S.: Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math. **63**(1), 1–38 (2010)
25. Dong, H., Yang, L.: Iteratively reweighted least squares for robust regression via SVM and ELM. CoRR abs/1903.11202 (2019). URL http://arxiv.org/abs/1903.11202
26. Donoho, D.L., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. Proceedings of the National Academy of Sciences **100**(5), 2197–2202 (2003). https://doi.org/10.1073/pnas.0437847100. URL http://www.pnas.org/content/100/5/2197.abstract
27. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inf. Theor. **47**(7), 2845–2862 (2001). https://doi.org/10.1109/18.959265
28. Eiben, A.E., Smith, J.: From evolutionary computation to the evolution of things. Nature **521**(7553), 476–482 (2015)
29. Ene, A., Vladu, A.: Improved convergence for ℓ_1 and ℓ_∞ regression via iteratively reweighted least squares. In: Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA, pp. 1794–1801 (2019). URL http://proceedings.mlr.press/v97/ene19a.html
30. Even, S., Tarjan, R.E.: Network flow and testing graph connectivity. SIAM J. Comput. **4**(4), 507–518 (1975). https://doi.org/10.1137/0204043
31. Facca, E., Cardin, F., Putti, M.: Towards a stationary monge-kantorovich dynamics: The physarum polycephalum experience. SIAM J. Appl. Math. **78**(2), 651–676 (2018)
32. Facca, E., Karrenbauer, A., Kolev, P., Mehlhorn, K.: Convergence of the non-uniform directed physarum model. Theor. Comput. Sci. **816**, 184–194 (2020). https://doi.org/10.1016/j.tcs.2020.01.034
33. Ford, L., Fulkerson, D.: Maximal flow through a network. Canad. J. Math. **8**, 399–404 (1956)
34. Goldberg, A.V., Rao, S.: Beyond the flow decomposition barrier. J. ACM **45**(5), 783–797 (1998). https://doi.org/10.1145/290179.290181
35. Gordon, D.M.: Ant Encounters: Interaction Networks and Colony Behavior. Primers in Complex Systems. Princeton University Press (2010). URL https://books.google.ch/books?id=MabwdXLZ9YMC
36. Gorodnitsky, I., Rao, B.: Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm. Trans. Signal Proc. **45**(3), 600–616 (1997). https://doi.org/10.1109/78.558475

37. Green, P.: Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion). J. Royal Statist. Soc., Series B: Methodol. **46**, 149–192 (1984)
38. Hopcroft, J.E., Karp, R.M.: An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. SIAM J. comput. **2**(4), 225–231 (1973)
39. Johannson, A., Zou, J.: A slime mold solver for linear programming problems. In: How the World Computes. Lecture Notes in Computer Science, vol. 7318, pp. 344–354. Springer, Berlin Heidelberg (2012)
40. Karam, L.J., McClellan, J.H.: Complex chebyshev approximation for fir filter design. IEEE Trans. Circuits Syst. II: Anal. Digit. Signal Process. **42**(3), 207–216 (1995)
41. Karlovitz, L.: Construction of nearest points in the l^p, p even, and l^∞ norms. i. J. Approx. Theor. **3**(2), 123–127 (1970)
42. Karrenbauer, A., Kolev, P., Mehlhorn, K.: Convergence of the non-uniform physarum dynamics. Theor. Comput. Sci. **816**, 260–269 (2020). https://doi.org/10.1016/j.tcs.2020.02.032
43. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
44. Lee, Y.T., Sidford, A.: Path finding methods for linear programming: Solving linear programs in $O(\sqrt{rank})$ iterations and faster algorithms for maximum flow. In: 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014, pp. 424–433 (2014). https://doi.org/10.1109/FOCS.2014.52
45. Miyaji, T., Ohnishi, I.: Physarum can solve the shortest path problem on riemannian surface mathematically rigourously. Int. J. Pure Appl. Matt. **47**(3), 353–369 (2008)
46. Mukhoty, B., Gopakumar, G., Jain, P., Kar, P.: Globally-convergent iteratively reweighted least squares for robust regression problems. In: K. Chaudhuri, M. Sugiyama (eds.) Proceedings of Machine Learning Research, *Proceedings of Machine Learning Research*, vol. 89, pp. 313–322. PMLR (2019). URL http://proceedings.mlr.press/v89/mukhoty19a.html
47. Nakagaki, T., Yamada, H., Toth, A.: Maze-solving by an amoeboid organism. Nature **407**(6803), 470 (2000)
48. Nesterov, Y., Nemirovskii, A.: Interior-point polynomial algorithms in convex programming, vol. 13. Society for Industrial and Applied Mathematics, (1994)
49. Olver, N., Végh, L.A.: A simpler and faster strongly polynomial algorithm for generalized flow maximization. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017, pp. 100–111 (2017). https://doi.org/10.1145/3055399.3055439
50. Orecchia, L., Sachdeva, S., Vishnoi, N.K.: Approximating the exponential, the lanczos method and an õ(m)-time spectral algorithm for balanced separator. In: H.J. Karloff, T. Pitassi (eds.) Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012, pp. 1141–1160. ACM (2012). https://doi.org/10.1145/2213977.2214080
51. Osborne, M.R.: Finite Algorithms in Optimization and Data Analysis. John Wiley & Sons Inc, New York, NY, USA (1985)
52. Perko, L.: Differential equations and dynamical systems, 3rd edn. Springer Science & Business Media, Berlin (2000)
53. Rao, B.D., Kreutz-Delgado, K.: An affine scaling methodology for best basis selection. IEEE Trans. Signal Process. **47**(1), 187–200 (1999). https://doi.org/10.1109/78.738251
54. Sherman, J.: Nearly maximum flows in nearly linear time. In: 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA, pp. 263–269 (2013). https://doi.org/10.1109/FOCS.2013.36
55. Spielman, D.A.: Algorithms, graph theory, and the solution of laplacian linear equations. ICALP **2**, 24–26 (2012)
56. Spielman, D.A., Teng, S.: Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In: Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004, pp. 81–90 (2004). https://doi.org/10.1145/1007352.1007372
57. Stonick, V.L., Alexander, S.T.: A relationship between the recursive least squares update and homotopy continuation methods. IEEE Trans. Signal Process. **39**(2), 530–532 (1991). https://doi.org/10.1109/78.80849
58. Straszak, D., Vishnoi, N.K.: On a natural dynamics for linear programming. In: ACM Innovations in Theoretical Computer Science (2016)

59. Straszak, D., Vishnoi, N.K.: IRLS and slime mold: equivalence and convergence. CoRR. arXiv:1601.02712 (2016)
60. Straszak, D., Vishnoi, N.K.: Natural algorithms for flow problems. In: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10–12, 2016, pp. 1868–1883. https://doi.org/10.1137/1.9781611974331.ch131 (2016)
61. Teng, S.H.: The Laplacian paradigm: Emerging algorithms for massive graphs. In: TAMC, pp. 2–14 (2010)
62. Tero, A., Kobayashi, R., Nakagaki, T.: A mathematical model for adaptive transport network in path finding by true slime mold. J. Theor. Biol. **244**(4), 553 (2007)
63. Valiant, L.G.: Evolvability. J. ACM **56**(1), 3:1–3:21 (2009). https://doi.org/10.1145/1462153.1462156
64. Végh, L.A.: A strongly polynomial algorithm for a class of minimum-cost flow problems with separable convex objectives. SIAM J. Comput. **45**(5), 1729–1761 (2016). https://doi.org/10.1137/140978296
65. Vishnoi, N.K.: $Lx = b$. Foundat. Trends Theor. Comput. Sci. **8**(1–2), 1–141 (2012)
66. Vishnoi, N.K.: The speed of evolution. In: Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '15, pp. 1590–1601. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2015). URL http://dl.acm.org/citation.cfm?id=2722129.2722234
67. Wright, S.: Primal-Dual Interior-Point Methods. Society for Industrial and Applied Mathematics (1997)