A Sharp Bound on the s-Energy and Its Applications to Averaging Systems

Bernard Chazelle

***Abstract*—The *s-energy* is a generating function of wide applicability in network-based dynamics. We derive an (essentially) optimal bound of $(3/\rho s)^{n-1}$ on the s-energy of an n-agent symmetric averaging system, for any positive real $s \leq 1$, where ρ is a lower bound on the nonzero weights. This is done by introducing the new dynamics of *twist systems*. We show how to use the new bound on the s-energy to tighten the convergence rates of systems in opinion dynamics, flocking, and synchronization.**

***Index Terms*—*s-energy*.**

I. Introduction

Averaging dynamics over time-varying networks is a process commonly observed in many well-studied multiagent systems. It has been used to model swarming, polarization, synchronization, gossip processes, and consensus formation in distributed systems [1], [8], [9]. Because of a dearth of general convergence techniques, results in the area often rely on network connectivity assumptions. The *s-energy* is a powerful analytical tool that allows us to overcome these restrictions [3]. It provides a global parametrized measure of the "footprint" of the system over an infinite horizon. This stands in sharp contrast with the local arguments (spectral or Lyapunov-based) typically used to prove fixed-point attraction.

The main result of this paper is an optimal bound on the s-energy of symmetric averaging systems. The new bound is used to tighten the convergence rates of various multiagent systems in opinion dynamics, flocking, and self-synchronization of coupled oscillators [1], [3], [7], [10], [11], [13]–[15], [17], [19]–[21].

Before moving to the technical discussion, we illustrate the role of the s-energy with a toy system. Fix $\rho \in (0, 1/2]$ and place n agents at $x_1, \ldots, x_n$ in [0, 1]. Given any $\varepsilon > 0$, for any integer $t > 0$, pick two agents i, j such that $x_j - x_i \geq \varepsilon$ (if any) and move them anywhere in the interval $[x_i + \delta, x_j - \delta]$, where $\delta = \rho(x_j - x_i)$. Repeat this process as long as possible. Note the high nondeterminism of the dynamics: not only can we choose the pair of agents at each step, but we can move them anywhere we please within the specified interval. Despite this freedom, the process always terminates in $O(\frac{1}{\rho n} \log \frac{1}{\varepsilon})^{n-1}$ steps, for any small enough $\varepsilon > 0$, and the bound is tight.[1] This result is a direct consequence of our new bound on the s-energy. The proof relies on a reduction to *twist systems*, a new type of multiagent dynamics that we define in the next section.

A. S-Energy

Let $(g_t)_{t=1}^{\infty}$ be an infinite sequence of graphs over a fixed vertex set $\{1, \ldots, n\}$. Each g_t is embedded in [0, 1], meaning that its vertices (the "agents") are represented by n real numbers between 0 and 1. Let $\mu_1, \ldots, \mu_k$ denote the lengths of the intervals formed by the union of the embedded edges of g_t, and put $\ell_t = \mu_1^s + \cdots + \mu_k^s$, for real or complex s.[2] The *s-energy* $\mathcal{E}(s)$ of the system is defined as the infinite sum $\sum_{t>0} \ell_t$. Because the s-energy follows an obvious scaling law, we note that embedding the graphs in the unit interval is not restrictive.

B. Averaging Systems

In a (symmetric) *averaging system*, g_t is undirected and supplied with self-loops at the vertices. To simplify the notation, we fix $t \in \mathbb{Z}^+$ and denote by x_i and y_i the positions of vertex i at times t and $t + 1$, respectively. Vertices are labeled so that $x_1 \leq \cdots \leq x_n$. For each $i \in \{1, \ldots, n\}$, write $r(i) = \max\{j \mid (i, j) \in g_t\}$ and $l(i) = \min\{j \mid (i, j) \in g_t\}$.[3] Fix $\rho \in (0, 1/2]$. The move of vertex i from x_i to y_i is subject to

$$x_{l(i)} + \delta_i \leq y_i \leq x_{r(i)} - \delta_i \tag{1}$$

where $\delta_i = \rho(x_{r(i)} - x_{l(i)})$. In other words, vertex i can move anywhere within the interval covered by its incident edges, but not too close to the endpoints. If $\rho = 0$, convergence is clearly impossible to ensure since i can easily oscillate periodically between two fixed vertices. We emphasize the high nondeterminism of the process: g_t is arbitrary and so is the motion of i within its allotted interval.

C. Results

Although the 0-energy is typically unbounded, it may come as a surprise that $\mathcal{E}(s)$ is always finite for any $s > 0$ [3]. In particular, the case $s = 1$ shows that it takes only a finite amount of ink to draw the infinite sequence of graphs g_t. We state the main result of this paper,[4] and prove it in Section III.

Theorem 1.1: The s-energy satisfies $\mathcal{E}(s) \leq (3/\rho s)^{n-1}$, for any $0 < \rho \leq 1/2$ and $0 < s \leq 1$.

We prove in item 1 in Section V that the bound of $O(1/\rho s)^{n-1}$ is optimal for $s = O(1/\log \frac{1}{\rho})$ and $\rho \leq 1/3$. These are the conditions we

Manuscript received July 21, 2018; revised January 2, 2019; accepted February 3, 2019. Date of publication February 14, 2019; date of current version September 25, 2019. Recommended by Associate Editor B. De Schutter. This work was supported by the Army Research Office and the Defense Advanced Research Projects Agency and was accomplished under Grant W911NF-17-1-0078. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, the Defense Advanced Research Projects Agency, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

The author is with the Department of Computer Science, Princeton University, Princeton, NJ 08540 USA (e-mail: chazelle@cs.princeton.edu).

Digital Object Identifier 10.1109/TAC.2019.2899509

[1] All logarithms are to the base 2.

[2] For example, if g_t consists of three edges embedded as [0, 0.2], [0.1, 0.3], [0.7, 0.9], and one self-loop at 0.5, then the union of the edges forms the three intervals [0, 0.3], [0.5, 0.5], [0.7, 0.9] and $\ell_t = (0.3)^s + (0.2)^s$.

[3] Because g_t is undirected and has self-loops, $l(i) \leq i \leq r(i)$, $(l \circ r)(i) \leq i \leq (r \circ l)(i)$. The notation l, r should not obscure the fact that both functions can be chosen differently for each graph g_t and its embedding $(x_i)_{i=1}^n$.

[4] We actually prove the slightly stronger bound of $2(2/\rho s)^{n-1}$ for $n > 2$.

encounter in practice, which is why we are able to provide tight bounds for all the applications discussed in this work. For $s = 1$, a quasi-optimal lower bound of $\Omega(1/\rho)^{\lfloor n/2 \rfloor}$ is already known [3]. Theorem 1.1 lowers the previous upper bound of $(1/s)^{n-1}(1/\rho)^{n^2+O(1)}$ [3].

The s-energy helps us bound the convergence rates of averaging network systems in full generality. To our knowledge, no other current technique can prove these results. The power of the s-energy is that it makes no connectivity requirements about the underlying dynamic networks. We use it typically to bound the *communication count* $\mathcal{C}_\varepsilon$, which is defined as the maximum number of steps t such that g_t has at least one edge of length $\varepsilon > 0$ or higher. From the inequality $\mathcal{C}_\varepsilon \leq \varepsilon^{-s}\mathcal{E}(s)$, setting $s = 1/\log\frac{1}{\varepsilon}$ and $s = n/\log\frac{1}{\varepsilon}$ in Theorem 1.1 yields:

Theorem 1.2: The communication count satisfies $\mathcal{C}_\varepsilon = O(\frac{1}{\rho}\log\frac{1}{\varepsilon})^{n-1}$ *for any* $2^{-n} \leq \varepsilon \leq 1/2$, *and* $\mathcal{C}_\varepsilon = O(\frac{1}{\rho n}\log\frac{1}{\varepsilon})^{n-1}$ *for* $0 < \varepsilon < 2^{-n}$.

This lowers the previous upper bound of $(1/\rho)^{n^2+O(1)}(\log 1/\varepsilon)^{n-1}$ [3]. We prove in item 2 in Section V that Theorem 1.2 is optimal for any positive $\varepsilon \leq \rho^{2n}$ and $\rho \leq 1/3$. We close this introduction with a few remarks about the results and their context.

1) The results extend to a large family of asymmetric averaging systems. Indeed, Theorems 1.1 and 1.2 hold for any infinite sequence of cut-balanced digraphs g_t: recall that a directed graph is said to be *cut-balanced* if its weakly connected components are also strongly connected.
2) The polylogarithmic factor $(\log 1/\varepsilon)^{n-1}$ in the convergence rate of Theorem 1.2 is a distinctive feature of time-varying network-based dynamics. Markov chains, for example, have convergence rates proportional to $\log 1/\varepsilon$.
3) Our definition of the s-energy differs slightly from the original formulation [3], which introduced the *total s-energy* as $\sum_{t>0}\sum_{(i,j)\in g_t} d_{ij}(t)^s$, where $d_{ij}(t)$ is the distance between the vertices i, j in the embedding of g_t. Up to a correction factor of at most $\binom{n}{2}$, our bounds apply to the total s-energy as well.
4) As noted in [6], the s-energy can be interpreted as a generalized Dirichlet series or, alternatively, as a partition function with s as the inverse temperature. Both interpretations have their own benefits, such as highlighting the lossless encoding properties of the s-energy or the usefulness of Legendre-transform arguments with the relevant thermodynamical quantities.

II. Twist Systems

We reduce averaging systems to a simpler kind of dynamics where agents keep the same ordering at all time. In a *twist system*, n points move within [0, 1] at discrete time steps. As before, we fix $t \in \mathbb{Z}^+$ and describe the motion of each point x_i at time t to its next position y_i at time $t + 1$. Unlike the averaging kind, twist systems preserve order; that is, assuming that $x_1 \leq \cdots \leq x_n$, then $y_1 \leq \cdots \leq y_n$. To describe the motion from t to $t + 1$, we choose two integers $1 \leq u < v \leq n$ and, for any i $(u \leq i \leq v)$, we define the *twist* of x_i as the interval within $[x_u, x_v]$ defined by

$$\tau_i = \left[x_u + \rho(x_{\min\{i+1,v\}} - x_u),\, x_v - \rho(x_v - x_{\max\{i-1,u\}})\right]. \quad (2)$$

Fixing $\rho \in (0, 1/2]$ ensures that all the twists are well-defined.[5] The only constraints on the dynamics are: 1) $y_1 \leq \cdots \leq y_n$; and 2) $y_i \in \tau_i$ for any $u \leq i \leq v$, and $y_i = x_i$ otherwise (see Fig. 1).

Observe that conditions (1, 2) are always feasible: for example, we can choose y_i to be the leftmost point in τ_i; of course, there is no need to do so and the expressive power of twist systems comes from the freedom they offer. Like their averaging counterparts, such systems are highly nondeterministic: at each step t, both the choice of u, v and the motion of the points are entirely arbitrary within the constraints (1, 2). Writing $\ell_t = (x_v - x_u)^s$, we define the s-energy of the twist system as $\mathcal{E}(s) = \sum_{t>0}\ell_t$. The next result justifies the introduction of twist systems.

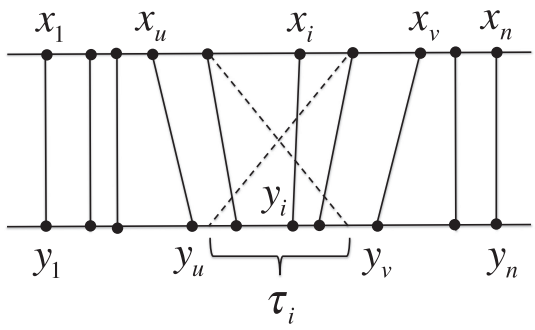

Fig. 1. Interval τ_i extends from a distance $\rho(x_{i+1} - x_u)$ to the right of x_u to a distance $\rho(x_v - x_{i-1})$ to the left of x_v: it thus *twists* $[x_{i-1}, x_{i+1}]$ into the allowed interval for y_i.

Theorem 2.1: Any averaging system can be viewed as a twist system with the same parameter ρ and the same s-energy.

Proof: Referring to our previous notation, recall that $\mu_1, \ldots, \mu_k$ denote the lengths of the intervals I_j formed by the union of the edge embeddings of g_t. We subdivide the time interval from t to $t + 1$ into k time windows and, for $j = 1, \ldots, k$, we process the motion within I_j during the jth window while keeping the other vertices fixed. All windows are treated similarly, so it suffices to explain the case $k = 1$. Let x_i (resp. x'_i) be the position of vertex i at time t (resp. $t + 1$) and let $y_1 \leq \cdots \leq y_n$ be the sequence of x'_i sorted in nondecreasing order.[6] Let $x_u, \ldots, x_v$ denote the positions within I_1; we may assume that $u < v$. The other vertices are kept fixed, so we have $y_i = x_i$ for $i < u$ or $i > v$. To show that the transition from x_i to y_i meets the conditions of a twist system, we need to prove that $y_i \in \tau_i$ for any i between u and v. By the symmetry of (2), it suffices to show that, for $u \leq i \leq v$

$$y_i \leq x_v - \rho(x_v - x_{\max\{i-1,u\}}). \quad (3)$$

Assume that $u < i \leq v$ and let $\bar{x}_j$ be shorthand for $\rho x_j + (1 - \rho)x_v$. The entire interval I_1 is covered by edges of g_t, so there must be at least one edge (a, b) that covers $[x_{i-1}, x_i]$, i.e., $b < i \leq a$. By (1), $x'_a \leq \rho x_{l(a)} + (1 - \rho)x_{r(a)}$, with $l(a) \leq b < i$ and $r(a) \leq v$; hence $x'_a \leq \bar{x}_{i-1}$. It also follows from (1) and the presence of self-loops that $x'_j \leq \rho x_{l(j)} + (1 - \rho)x_{r(j)} \leq \bar{x}_{i-1}$ for any j $(u \leq j < i)$; also $x'_j = x_j \leq \bar{x}_{i-1}$ for $j < u$. Putting it all together, this proves the existence of at least i indices $l \leq v$ such that $x'_l \leq \bar{x}_{i-1}$. It follows that $y_i \leq \bar{x}_{i-1}$; hence, (3) for $u < i \leq v$. To complete the proof of (3), we note that the case $i = u$ follows from $y_u \leq y_{u+1}$. The case $k > 1$ is handled by repeating the previous analysis for each interval I_j. The s-energy contributed by one step of the averaging system matches the energetic contribution of the k substeps of the twist system. ■

III. Bounding the s-Energy

The proof of Theorem 1.1 is unusual in the context of dynamics because it is *algorithmic*: it consists of a set of trading rules that allows money to be injected into the system and exchanged among the vertices to meet their needs. As the transactions take place, money is spent to pay for the s-energy expended along the way. If all of the energy can

[5] Indeed, we can check that $\tau_i = [a, b]$, where $a \leq b$. The terminology refers to the "twisting" of the interval $[x_{i-1}, x_{i+1}]$ around x_i into the interval τ_i around y_i.

[6] We break ties by using the index i. Note that the y_i's are sorted, so they are not the same as those used in the definition of averaging systems given above.

be accounted for in this manner, then the amount of money injected in the system is an upper bound on $\mathcal{E}(s)$. In our earlier work [3], we were able to pursue this approach only for the case $s = 1$. We show here how to extend it to all $s \in (0, 1]$. The idea was to supply each vertex with its own credit account and then let them trade credits to pay for the s-energy incrementally. This strategy does not work here because of its inability to cope with all the scales present in the system.[7] The remedy is to supply each *pair* of vertices with their own account. Only then are we able to accommodate all scales at once. By appealing to Theorem 2.1, we may substitute twist systems for averaging systems. We focus the analysis on the transition at time t from $x_1 \le \cdots \le x_n$ to $y_1 \le \cdots \le y_n$. Our only assumption is that, for some u, v ($1 \le u < v \le n$), we have $y_i \in \tau_i$ for any $u \le i \le v$, and $y_i = x_i$ otherwise.

For each pair (i, j) such that $1 \le i < j \le n$, we maintain an account $B_{i,j}$ consisting of $(x_j - x_i)^s A^{j-i}$ credits, where $A := 2/\rho s$ and one credit is used to pay for a single unit of s-energy. (Amounts paid need not be integers.) We show that updating each $B_{i,j}$ at time t to $B'_{i,j}$ at time $t + 1$ leaves us with enough unused money to pay for the s-energy $(x_v - x_u)^s$ released at that step.[8] No new money is needed past the initial injection at time 1, so the s-energy is at most the sum of all the $B_{i,j}$'s at the beginning: $\mathcal{E}(s) \le \sum_{i<j} A^{j-i} < (\frac{A}{A-1})^2 A^{n-1} < 2(2/\rho s)^{n-1}$, for $n > 2$. For $n = 2$, $\mathcal{E}(s) \le A$, hence Theorem 1.1. We begin with a few words of intuition.

1) We update $B_{i,j}$ to $B'_{i,j}$ by considering the pairs (i, j) in descending order of j–i, starting with $(1, n)$. In general, the update for (i, j) will rely on money released by the pairs $(i - 1, j)$ and $(i, j + 1)$, whose accounts will have already been updated. In turn, the pair (i, j) will then be expected to provide money to both $(i, j - 1)$ and $(i + 1, j)$: the donation will be made in two equal amounts.
2) How much money should (i, j) receive from its donors. For the sake of this informal discussion, let us focus on the case $u \le i < j \le v$. The account $B_{i,j}$ should receive enough to grow to $(x_v - x_u)^s A^{j-i}$. This typically exceeds its balance of $(x_j - x_i)^s A^{j-i}$ at time t, so an infusion of money is required. Of course, the amount actually needed for $B'_{i,j}$ is only $(y_j - y_i)^s A^{j-i}$, so this in turn frees $((x_v - x_u)^s - (y_j - y_i)^s)A^{j-i} \ge 0$, which can be then passed on to $(i, j - 1)$ and $(i + 1, j)$.
3) We pay for the energetic contribution at time t by spending the leftover money from the update for $(u, u + 1)$, which we show to be at least $(x_v - x_u)^s$, as required.

Proof of Theorem 1.1: We update $B_{i,j}$ by using $C_{i,j}$ credits supplied by the accounts $B_{i-1,j}$ and $B_{i,j+1}$. We show how this produces a leftover $D_{i,j}$, which can then be donated to $(i + 1, j)$ and $(i, j - 1)$ in equal amounts. Here are the details: for all $1 \le i < j \le n$ in descending order of $j - i = n - 1, \ldots, 1$, apply the following assignments (see Fig. 2):

$$\begin{cases} C_{i,j} \leftarrow \frac{1}{2}(D_{i-1,j} + D_{i,j+1}) \\ D_{i,j} \leftarrow B_{i,j} + C_{i,j} - B'_{i,j} \end{cases} \tag{4}$$

where $B_{i,j} = (x_j - x_i)^s A^{j-i}$, $B'_{i,j} = (y_j - y_i)^s A^{j-i}$, and $D_{i,j} = 0$ if $i < 1$ or $j > n$. The assignments denote transfers of money. This explains the factor of $1/2$, which keeps the money pool conserved: for example, one half of $D_{i,j}$ goes to $(i, j - 1)$ and the other half to $(i + 1, j)$. The soundness of the trading scheme rests entirely on the claimed nonnegativity of all the donations $D_{i,j}$. For any $i \in \{1, \ldots, n\}$, define $u(i) = u$ and $v(i) = v$ if $u \le i \le v$; and set $u(i) = v(i) = i$ otherwise. We prove by induction on $j - i > 0$ that, for $1 \le i < j \le n$

$$\begin{cases} B_{i,j} + C_{i,j} \ge (x_{v(j)} - x_{u(i)})^s A^{j-i} & (5) \\ D_{i,j} \ge 0. & (6) \end{cases}$$

The next inequality, which follows from $dz^s/dz \ge s$ for $s, z \in (0, 1]$, will prove useful in establishing (5, 6)

$$1 - (1 - x)^s \ge sx \quad \text{for any } s, x \in [0, 1]. \tag{7}$$

1) *Case $u \le i < j \le v$:* By affine invariance, we can always assume that $x_u = x_v - 1 = 0$. We begin with the case $u < i < j \le v$ and observe that $v(j) = v$ and $u(i - 1) = u(i) = u$. Because $y_{i-1} \in \tau_{i-1}$, we have $y_{i-1} \ge \rho x_i$. Using (7), we find that

$$\begin{aligned} D_{i-1,j} &= B_{i-1,j} + C_{i-1,j} - B'_{i-1,j} \\ &\ge ((x_{v(j)} - x_{u(i-1)})^s - (y_j - y_{i-1})^s)A^{j+1-i} \\ &\ge (1 - (1 - \rho x_i)^s)A^{j+1-i} \ge \rho s x_i A^{j+1-i}. \end{aligned} \tag{8}$$

If $i = u$, we have $x_i = 0$, hence (8) merely expresses nonnegativity, which holds inductively. We conclude that (8) obtains for any $u \le i < j \le v$. Likewise, by symmetry, $D_{i,j+1} \ge \rho s(1 - x_j)A^{j+1-i}$. It follows from (4) that

$$C_{i,j} \ge \frac{1}{2}\rho s(1 - (x_j - x_i))A^{j+1-i} \ge (1 - (x_j - x_i)^s)A^{j-i}$$

therefore

$$\begin{aligned} B_{i,j} + C_{i,j} &\ge (x_j - x_i)^s A^{j-i} + (1 - (x_j - x_i)^s)A^{j-i} \\ &= A^{j-i} = (x_{v(j)} - x_{u(i)})^s A^{j-i} \end{aligned}$$

which establishes (5). Since $x_{u(i)} = x_u \le y_i \le y_j \le x_v = x_{v(j)}$, this also proves that

$$D_{i,j} = B_{i,j} + C_{i,j} - B'_{i,j} \ge A^{j-i} - (y_j - y_i)^s A^{j-i} \ge 0$$

hence (6).

2) *Case $i < u \le j \le v$:* This time, we set $x_i = 0$ and $x_v = 1$ and note that $u(i) = i$ and $v(j) = v$. We begin with the case $j <$

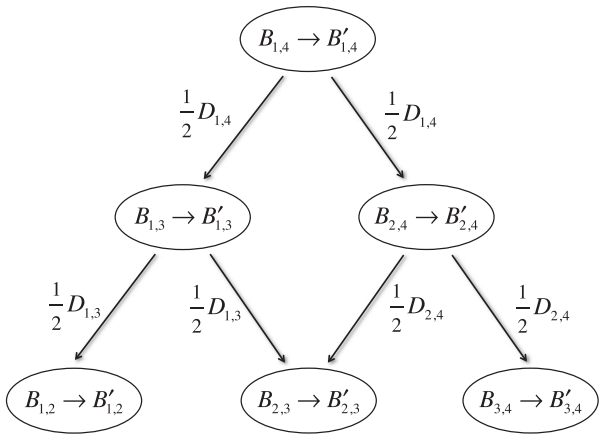

Fig. 2. Updating $B_{1,4}$ to its new value of $B'_{1,4}$ releases $D_{1,4}$ credits, which are passed on evenly to the pairs (1, 3) and (2, 4). With this scheme in place, updating $B_{2,3}$ to $B'_{2,3}$ can make use of $C_{2,3} = \frac{1}{2}(D_{1,3} + D_{2,4})$ credits.

[7] We illustrate the difficulty with a simple example. Set $n = 3$ and assign $x_i^s A^i$ credits to the account for vertex $i = 1, 2, 3$. Initialize the system with $x_1 = 0$, $x_2 = 1 - \varepsilon$, and $x_3 = 1$; set $\rho = 1/2$, with g_1 consisting of the single edge $(2, 3)$. Assume now that $y_1 = 0$ and $y_2 = y_3 = 1 - \varepsilon/2$. The account for vertex 3, the only one to release money, gives out only $(1 - (1 - \varepsilon/2)^s)A^3 \approx \frac{1}{2}s\varepsilon A^3$ credits. If $s < 1$ and $\varepsilon > 0$ is very small, this is not enough to cover the s-energy of ε^s needed for the first step. The problem is that the credit accounts do not operate at all scales.

[8] We refer to $B_{i,j}$ as both the account for (i, j) and its value.

v, which implies that $v(j+1)=v$. Using (4) and (5), $y_i = x_i$, $y_{j+1} \in \tau_{j+1}$, and (7) in this order, we find that

$$\begin{aligned} D_{i,j+1} &= B_{i,j+1} + C_{i,j+1} - B'_{i,j+1} \\ &\geq ((x_{v(j+1)} - x_{u(i)})^s - (y_{j+1} - x_i)^s)A^{j+1-i} \\ &\geq (1-(1-\rho(1-x_j))^s)A^{j+1-i} \\ &\geq \rho s(1-x_j)A^{j+1-i} = 2(1-x_j)A^{j-i}. \end{aligned} \tag{9}$$

Again, by induction, $D_{i-1,j} \geq 0$; therefore, by (4)

$$\begin{aligned} B_{i,j} + C_{i,j} &\geq B_{i,j} + \frac{1}{2}D_{i,j+1} \geq (x_j^s + (1-x_j))A^{j-i} \\ &\geq A^{j-i} = (x_{v(j)} - x_{u(i)})^s A^{j-i} \end{aligned}$$

hence (5). For the case $j = v$, again note that the lower bounds on $D_{i-1,j}$ and $D_{i,j+1}$ we just used still hold, and thus so does (5) for all $i < u \leq j \leq v$. Finally, $y_j \leq x_v$; hence, $x_{u(i)} = x_i = y_i \leq y_j \leq x_v = x_{v(j)}$, and (6) follows from (4 and (5).

The case $u \leq i \leq v < j$ is the mirror image of the last one, whereas the remaining three cases are trivial and require no account updates. We pay for the s-energy contribution at time t by tapping into $D_{u,u+1}$, which is unused. For this to work, it suffices to show that $D_{u,u+1} \geq (x_v - x_u)^s$. We have $y_i \in \tau_i$ $(i = u, u+1)$; hence

$$\begin{aligned} y_{u+1} - y_u &\leq x_v - \rho(x_v - x_u) - (x_u + \rho(x_{u+1} - x_u)) \\ &\leq \rho(x_u - x_{u+1}) + (1-\rho)(x_v - x_u) \\ &\leq (1-\rho)(x_v - x_u). \end{aligned}$$

Thus, it follows from (4), (5), and (7), together with $u(u) = u$ and $v(u+1) = v$, that

$$\begin{aligned} D_{u,u+1} &= B_{u,u+1} + C_{u,u+1} - B'_{u,u+1} \\ &\geq A(x_{v(u+1)} - x_{u(u)})^s - A(y_{u+1} - y_u)^s \\ &\geq (1-(1-\rho)^s)A(x_v - x_u)^s \\ &\geq \rho s A(x_v - x_u)^s \geq (x_v - x_u)^s. \end{aligned}$$

This completes the proof of Theorem 1.1 for twist systems. By Theorem 2.1, this also implies the same upper bound for averaging systems. ■

IV. Applications

A number of known convergence rates for various averaging systems can be sharpened by appealing to Theorems 1.1 and 1.2. We give a few examples below.

A. Asymmetric Averaging Systems

Symmetric averaging systems have been widely used to model backward products of the form $(A_t \cdots A_1 x)_{t>0}$, where each A_k is a type-symmetric stochastic matrix with positive diagonal and nonzero entries at least $\rho > 0$ [2], [8], [11], [13], [15], [17].[9] In other words, A_k is the matrix of a lazy random walk in an undirected graph g_k with a lower bound of ρ on the nonzero probabilities. A close examination of the proof of Theorem 2.1 shows that the graphs g_t may be directed as long as the vertices still have self-loops, and, for each $i = u+1, \ldots, v$, there exist edges "hovering" over i from both sides, i.e., (a, b) and (b', a'), with $a, a' < i \leq b, b'$. We note that this property holds if each directed graph g_t is cut-balanced.[10] This gives us a strict generalization of Theorems 1.1 and 1.2 to asymmetric averaging systems whose sequences of digraphs are cut-balanced. This result goes beyond the mere convergence of these systems, which was established in [12].

B. Opinion Dynamics

There has been considerable attention given to consensus formation in social dynamics [8]–[10]. Given a set of agents in high-dimensional space, where coordinates model opinions, one imagines that at each step a subset of them come into contact and, through a process of deliberation, adjust their opinions toward agreement. Will such a process converge to consensus, polarization, a mixture of both, or not at all? Mathematically, the agents are represented by their position in d-dimensional space: $x_1, \ldots, x_n$ in $[0,1]^d$. We fix $0 < \alpha \leq 1$ and iterate on the following process forever: (1) choose an arbitrary nonempty subset of the agents and move them anywhere inside the box $(1-\alpha)B + \alpha c$, where B is the smallest (axis-parallel) box enclosing the chosen agents and c is the center of B; (2) repeat. Intuitively, one "squeezes" the subset of agents together a little.

Theorem 4.1: For any positive $\varepsilon \leq 2^{-dn}$, at all but $O(\frac{1}{d\alpha n}\log\frac{1}{\varepsilon})^{n-1}$ time steps, the smallest box enclosing the chosen agents has volume less than ε.

Proof: We set up a symmetric averaging system as follows: g_t consists of n self-loops, together with the complete graph joining the agents of the chosen subset; along each axis, the dynamics obeys (1) with parameter $\rho = \alpha/2$. Let $\ell_t(j)$ be the length of the graph's projection onto the jth axis. By Theorem 1.1, we know that, for any $0 < r \leq 1$, $\sum_{t>0} \ell_t(j)^r \leq (6/\alpha r)^{n-1}$. Let V_t be the volume of the smallest box enclosing the agents picked at time t. By the generalized Hölder's inequality, for $0 < s \leq 1/d$

$$\sum_{t>0} V_t^s = \sum_{t>0}\prod_{j=1}^{d} \ell_t(j)^s \leq \prod_{j=1}^{d}\left(\sum_{t>0}\ell_t(j)^{ds}\right)^{1/d} \leq (6/d\alpha s)^{n-1}.$$

Set $s = n/\log\frac{1}{\varepsilon}$ and use Markov's inequality to complete the proof. ■

C. Flocking

Many models of bird flocking have been developed over the years and used to great effect in CGI for film and animation. Their mathematical analysis has lagged behind, however. In a simple, popular model tracing its roots back to Cucker and Smale, Vicsek, and ultimately Reynolds, a group of n birds is represented by two $n \times 3$ matrices $\mathbf{x}(t)$ and $\mathbf{v}(t)$, where the ith rows encode the location and velocity in $\mathbb{R}^3$ of the ith bird, respectively [7], [13], [21]. The dynamics obeys the relations

$$\begin{cases} \mathbf{x}(t) = \mathbf{x}(t-1) + v(t) \\ \mathbf{v}(t+1) = P(t)\,\mathbf{x}(t) \end{cases}$$

where $P(t)$ is an $n \times n$ stochastic matrix whose entry (i,j) is positive if and only if birds i and j are within a fixed distance R of each other. All entries are rationals over $O(\log n)$ bits. A tight bound on the convergence of the dynamical system was established in [4] and [5]: it was shown that steady state is always reached within a number of steps equal to a tower-of-twos of height proportional to $\log n$; even more amazing, this bound is optimal. The lead-up to steady-state consists of two phases: fragmentation and aggregation. The latter can feature only the merging of flocks while the (much shorter) fragmentation phase

[9] A matrix A is type-symmetric if A_{ij} and A_{ji} are both positive or both 0 for all i, j.

[10] A directed graph is cut-balanced if its weakly connected components are strongly connected.

can witness the repeated formation and breakup of flocks. Technically, a flock is defined as the birds in a given connected component of the network joining any two birds within distance R. It has been shown that the total number of network switches (i.e., the number of steps where the communication network changes) is $n^{O(n^2)}$. We improve this bound to $n^{O(n)}$ by using the s-energy. It was demonstrated in [4, p. 21:7] that the number of network switches is bounded by the communication count C_ε, for $\varepsilon \geq n^{-bn^2}$, $\rho \geq n^{-c}$ and constant $b, c > 0$. Our claim follows from Theorem 1.2. □

D. Self-Synchronizing Oscillators

The self-organized synchronization of coupled oscillators is a well-known phenomenon in physics and biology: it is observed in circadian neurons, firing fireflies, yeast cell suspensions, cardiac pacemaker cells, power plant grids, and even musical composition (e.g., Ligeti's *poème symphonique*). In the discrete Kuramoto model studied in [16]–[18], all oscillators share the same natural frequency and the phase of the ith one obeys the recurrence

$$\theta_i(t+1) = \theta_i(t) + \frac{K\Delta T}{|n_i(t)|} \sum_{j \in n_i(t)} \sin(\theta_j(t) - \theta_i(t))$$

where $n_i(t)$ is the set of vertices adjacent to i in g_t (which includes i). Following [17], we assume that all n phases start in the same open half-circle, which we can express as $\alpha - \pi/2 \leq \theta_i(0) \leq \pi/2$, for some arbitrarily small positive constant α. We find that $\sin(\theta_j(0) - \theta_i(0)) = a_{ij}(\theta_j(0) - \theta_i(0))$, where $c\alpha \leq a_{ij} \leq 1$, for constant $c > 0$. This condition holds for all t since averaging keeps the phases in the same open half-circle. The dynamics is that of a symmetric averaging system provided that we pick ρ small enough so that $b\rho n/\alpha \leq K\Delta T \leq 1$, for a suitable constant $b > 0$. By Theorem 1.2, for any $\varepsilon \leq 2^{-n}$, the number of steps where two oscillators are joined by an edge while their phases are off by ε or more is $O(\frac{1}{\alpha K \Delta T} \log \frac{1}{\varepsilon})^{n-1}$.

V. Lower Bound Proofs

1) We prove that the bound $O(1/\rho s)^{n-1}$ from Theorem 1.1 is optimal for $s = O(1/\log \frac{1}{\rho})$ and $\rho \leq 1/3$. A lower bound construction from [3, p. 1703] describes a system whose n-agent s-energy satisfies the recurrence $\mathcal{E}_n \geq \rho^s \mathcal{E}_{n-1} + (1-2\rho)^s \mathcal{E}_n + 1$ for $n > 1$; hence, for positive constant b, $\mathcal{E}_2 \geq b/\rho s$ and $\mathcal{E}_n \geq (b\rho^{s-1}/s)\mathcal{E}_{n-1}$ for $n > 2$. This shows that $\mathcal{E}_n \geq (b/\rho s)^{n-1} \rho^{s(n-2)} = \Omega(1/\rho s)^{n-1}$, for $s = O(1/\log \frac{1}{\rho})$, as claimed.
2) We prove that $\mathcal{C}_\varepsilon = \Omega(\frac{1}{\rho n} \log \frac{1}{\varepsilon})^{n-1}$, for any positive $\varepsilon \leq \rho^{2n}$ and $\rho \leq 1/3$. Note that ρ must be bounded away from $1/2$ (we choose $1/3$ for convenience): indeed, in the case of two vertices at distance 1 joined by an edge, we have the trivial bound $C_\varepsilon = 1$ for $\rho = 1/2$. The proof revisits an earlier construction [3] and modify its analysis to fit our purposes. If $n > 1$, the n vertices of g_1 are positioned at 0, except for $x_n = 1$. Besides the self-loops, the graph g_1 has the single edge $(n-1, n)$. At time 2, the vertices are all at 0 except for $x_{n-1} = \rho$ and $x_n = 1 - \rho$. The first $n-1$ vertices form a system that stays in place if $n = 2$ and, otherwise, proceeds recursively within $[0, \rho]$ until it converges to the fixed point $\rho/(n-1)$: this value is derived from the fact that each step keeps the mass center invariant. After convergence[11] of the vertices labeled 1 through $n-1$, the n-vertex system repeats the previous construction within $[\rho/(n-1), 1-\rho]$. Let $C(n, \varepsilon)$ denote the communication count for n agents: we have $C(n, \varepsilon) = 0$ if $n = 1$ or $\varepsilon > 1$; else

$$C(n,\varepsilon) \geq 1 + C\left(n-1, \frac{\varepsilon}{\rho}\right) + C\left(n, \frac{\varepsilon}{1 - \rho n/(n-1)}\right). \tag{10}$$

By expanding the recurrence and using monotonicity

$$C(n,\varepsilon) \geq k + k\, C\left(n-1, \frac{\varepsilon}{\rho(1-2\rho)^{k-1}}\right) \tag{11}$$

where we set

$$k = \left\lceil \frac{(\log \varepsilon)/n - \log \rho}{2 \log(1-2\rho)} \right\rceil.$$

Assume now that $\varepsilon \leq \rho^{2n}$. From our choice of k, we easily verify that

$$\rho(1-2\rho)^{k-1} \geq \varepsilon^{1/n}. \tag{12}$$

The recurrence (11) requires that $\varepsilon/(\rho(1-2\rho)^{k-1}) < 1$, which follows from (12). Since $\varepsilon^{1/2n} \leq \rho \leq 1/3$, we have $k \geq \frac{b}{\rho n} \log \frac{1}{\varepsilon}$, for constant $b > 0$. It follows that $C(2, \varepsilon) = \Omega(\frac{1}{\rho} \log \frac{1}{\varepsilon})$ and, for $n > 2$, by (12)

$$C(n,\varepsilon) \geq \left(\frac{b}{\rho n} \log \frac{1}{\varepsilon}\right)\, C(n-1, \varepsilon^{1-1/n}).$$

We verify that the condition $\varepsilon \leq \rho^{2n}$ holds recursively: $\varepsilon^{1-1/n} \leq \rho^{2(n-1)}$. By induction, it follows that $C(n, \varepsilon) \geq \Omega(\frac{1}{\rho n} \log \frac{1}{\varepsilon})^{n-1}$, as desired.

[11] We can use a limiting argument to break out of the infinite loop.

Acknowledgment

The author would like to thank the referees for their helpful comments and suggestions.

References

[1] F. Bullo, J. Cortés, and S. Martinez, *Distributed Control of Robotic Networks* Applied Mathematics Series. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[2] M. Cao, D. A.Spielman, and A. S. Morse, "A lower bound on convergence of a distributed network consensus algorithm," in *Proc. 44th IEEE Conf. Decision Control, Eur. Control Conf.*, Seville, Spain, 2005, pp. 2356–2361.
[3] B. Chazelle, "The total s-energy of a multiagent system," *SIAM J. Control Optim.*, vol. 49, pp. 1680–1706, 2011.
[4] B. Chazelle, "The convergence of bird flocking," *J. ACM*, vol. 61, vol. 21, pp. 1–35, 2014.
[5] B. Chazelle, "How many bits can a flock of birds compute?" *Theory Comput.*, vol. 10, pp. 421–451, 2014.
[6] B. Chazelle, "An algorithmic approach to collective behavior," *J. Statist. Phys.*, vol. 158, pp. 514–548, 2015.
[7] F. Cucker and S. Smale, "Emergent behavior in flocks," *IEEE Trans. Autom. Control*, vol. 52, no. 5, pp. 852–862, May 2007.
[8] F. Fagnani and P. Frasca, *Introduction to Averaging Dynamics over Networks*, Lecture Notes in Control and Information Sciences, vol. 472. New York, NY, USA: Springer, 2018.
[9] S. Fortunato, "On the consensus threshold for the opinion dynamics of Krause-Hegselmann," *Int. J. Modern Phys. C*, vol. 16, no. 2, pp. 259–270, 2005.
[10] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *J. Artif. Soc. Soc. Simul.*, vol. 5, no. 3, pp. 1–33, 2002.
[11] J. M. Hendrickx and V. D. Blondel, "Convergence of different linear and non-linear Vicsek models," in *Proc. 17th Int. Symp. Math. Theory Netw. Syst.*, Kyoto, Japan, Jul. 2006, pp. 1229–1240.
[12] J. M. Hendrickx and J. Tsitsiklis, "Convergence of type-symmetric and cut-balanced consensus seeking systems," *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 214–218, Jan. 2013.
[13] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, Jun. 2003.

[14] A. Jadbabaie, N. Motee, and M. Barahona, "On the stability of the Kuramoto model of coupled nonlinear oscillators," in *Proc. Am. Control Conf.*, 2004, vol. 5, pp. 4296–4301.

[15] J. Lorenz, "A stabilization theorem for dynamics of continuous opinions," *Physica A: Statistical Mech. Appl.*, vol. 355, pp. 217–223, 2005.

[16] S. Martinez, "A convergence result for multiagent systems subject to noise," in *Proc. Am. Control Conf.*, St. Louis, MO, USA, 2009, pp. 4537–4542.

[17] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Trans. Autom. Control*, vol. 50, no. 2, pp. 169–182, Feb. 2005.

[18] L. Scardovi, A. Sarlette, and R. Sepulchre, "Synchronization and balancing on the N-torus," *Syst. Control Lett.*, vol. 56, pp. 335–341, 2007.

[19] S. H. Strogatz, "From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators," *Physica D*, vol. 143, pp. 1–20, 2000.

[20] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 1984.

[21] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles," *Phys. Rev. Lett.*, vol. 75, pp. 1226–1229, 1995.