Noidy Conmunixatipn: On the Convergence of the Averaging Population Protocol

Frederik Mallmann-Trenn
MIT, CSAIL, US
mallmann@mit.edu

Yannic Maus
Department of Computer Science, Technion, Haifa, Israel,
yannic.maus@cs.technion.ac.il

Dominik Pajak
Faculty of Fundamental Problems of Technology, Wroclaw University of Science and Technology, Tooploox, Wroclaw, Poland
dominik.pajak@pwr.edu.pl

Abstract

We study a process of *averaging* in a distributed system with *noisy communication*. Each of the agents in the system starts with some value and the goal of each agent is to compute the average of all the initial values. In each round, one pair of agents is drawn uniformly at random from the whole population, communicates with each other and each of these two agents updates their local value based on their own value and the received message. The communication is noisy and whenever an agent sends any value v, the receiving agent receives $v+N$, where N is a zero-mean Gaussian random variable. The two quality measures of interest are (i) the total sum of squares $TSS(t)$, which measures the sum of square distances from the average load to the *initial average* and (ii) $\bar{\phi}(t)$, which measures the sum of square distances from the average load to the *running average* (average at time t).

It is known that the simple averaging protocol—in which an agent sends its current value and sets its new value to the average of the received value and its current value—converges eventually to a state where $\bar{\phi}(t)$ is small. It has been observed that $TSS(t)$, due to the noise, eventually diverges and previous research—mostly in control theory—has focused on showing eventual convergence w.r.t. the running average. We obtain the first probabilistic bounds on the convergence time of $\bar{\phi}(t)$ and precise bounds on the drift of $TSS(t)$ that show that although $TSS(t)$ eventually diverges, for a wide and interesting range of parameters, $TSS(t)$ stays small for a number of rounds that is polynomial in the number of agents. Our results extend to the synchronous setting and settings where the agents are restricted to discrete values and perform rounding.

2012 ACM Subject Classification Theory of computation → Distributed computing models

Keywords and phrases population protocols, noisy communication, distributed averaging

Digital Object Identifier 10.4230/LIPIcs.ICALP.2019.143

Category Regular Paper Track C

Related Version See [?] (`http://arxiv.org/abs/1904.10984`) for the full version with all proofs.

Acknowledgements Frederik Mallmann-Trenn was partly supported by NSF Award Numbers CCF-1461559, CCF-0939370, and CCF-18107. Yannic Maus was partly supported by ERC Grant No. 336495 (ACDC). Dominik Pajak was partly supported by Polish National Science Center grant UMO-2018/29/B/ST6/02969 and NSF Award Numbers CCF-1461559, CCF-0939370, CCF-18107.

1 Introduction

We consider the problem of distributed averaging by a group of agents (*e.g.,* sensors), initialized with values that represent, for example, different temperature measurements. The agents' goal is to compute the average of all the initial values using the following simple

46th International Colloquium on Automata, Languages, and Programming (ICALP 2019).
Editors: Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi;
Article No. 143; pp. 143:1–143:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

dynamic: In each discrete round, two agents are drawn uniformly at random from the whole population, communicate their values to each other and set their new values to the average of their old value and the received value. Converging to the average plays a key role in many applications, e.g., for sensor networks [58, 52], social insects [10], and robotics [21, 31]. In all of these applications, the agents (sensors, ants, and robots) are very simple and are therefore limited in both memory and communication. Moreover, communication is often erroneouees.[1] This motivates the study of the aforementioned simple averaging dynamic in a setting where the agents only remember one value, do not use any additional memory, and the communication is subject to noise. We model the noise in the communication as follows: Whenever an agent sends any value v, the receiving agent receives $v + N$, where random variable N is distributed according to some zero-mean probability distribution $\aleph$, *e.g.,* a normal distribution. The agents update their values as follows: whenever two agents communicate, each agent sets its new value to the average of their old value and the received value; note that—due to the noise—the two agents might have distinct new values.

The values of the n nodes in step t of the process are denoted by $X_1^{(t)}, X_2^{(t)}, \ldots, X_n^{(t)}$. We consider the following models: (i) the *sequential setting* where one pair of agents is chosen uniformly at random and (ii) the *synchronous setting* where each agent is matched to exactly one other agent chosen uniformly at random. The two quality measures of the convergence used in this work are (i) the total sum of squares $TSS(t) = \sum_i (X_i^{(t)} - \varnothing^{(0)})^2$, where $\varnothing^{(0)} = \sum_i X_i^{(0)}/n$ is the initial average and (ii) the sum of squared distances to the running average $\bar{\phi}(t) = \sum_i (X_i^{(t)} - \varnothing^{(t)})^2$, where $\varnothing^{(t)} = \sum_i X_i^{(t)}/n$ is the *running average.* Our contributions can be informally summarized as follows:

(i) We give, under mild assumptions on the noise, the first bounds on the convergence time of the running average $\bar{\phi}(t)$ in the noisy gossip-based communication setting. The bounds we obtain are—up to a constant factor—tight. In particular, the potential converges to a value that is linear in n and the second moment of the noise $\mathbb{E}\left[N^2\right]$; which is tight. So far it was only known that the process eventually converges to a state where $\bar{\phi}(t)$ is small (*e.g.,* [56]), but precise bounds were not known. (Thm. 1)

(ii) We show that, in contrast to the current belief, one can hope to converge to the *initial* average in addition to convergence to the *running* average as long as the number of rounds are bounded: It was known that $TSS(t)$, due to the noise, eventually diverges (the running average diverges from the initial average) and for this reason related research—mostly in control theory—has focused on showing eventual convergence w.r.t. $\bar{\phi}(t)$; leaving $TSS(t)$ aside. Since we give precise bounds on the convergence time of the running average, we can show the following. Under mild assumptions on the noise, $TSS(t)$ converges to almost the same value as $\bar{\phi}(t)$ as long as the number of time steps t is bounded by $O(n^2)$, where n is the number of nodes. (Cor. 2)

(iii) We pioneer in the discrete setting in which the agents can store only integer values and the noise is also an integer. In this setting the agents in our algorithm perform randomized rounding. We show that this only causes a negligible difference from the continuous case. (Cor. 3)

(iv) We study both the sequential and the synchronous setting and show that there is no significant difference (up to a scaling of time) between the models. (Cor. 4)

(v) We perform simulations in the setting where nodes are limited in storage, *i.e.,* they can only store values from a bounded range. This leads to a much faster (by order

[1] Consult Sec. 1.1 for a more detailed review of these applications including the limitation of agents and further motivation. Sec. 1.1 also contains related work on the averaging protocol.

of magnitude) divergence between the running average and the initial average. Our simulations also seem to indicate strong bounds on the distribution of distances to the running average in our main model (unbounded values). (Sec. 5)

The convergence time of the averaging processes in the gossip-based communication setting *without* noise has been studied before (e.g., [39]). However, to the best of our knowledge, no bounds on the convergence time are known in the gossip-based communication setting with noise. We continue with a detailed motivation for studying noise in the simple averaging dynamic and related work.

1.1 Motivation and Related Work

Converging to the average plays a key role in many applications in which agents have limited computational and communication power, e.g.,

(i) sensor networks [58, 52]: here there is a wide range of application including terrain monitor applications [53], computing an average temperature, PIR sensors measuring the infrared light radiation emitted from objects, and many more applications. In such scenarios links are often faded [48, 14],

(ii) social insects: for ants, values could represent the individuals' different assessments of nest qualities when house hunting [10] or the deficit of workers at a given task [43], and

(iii) robotics [21, 31] and in particular memory-limited robots, *e.g.,* Kilobots exploring the percentage of white tiles in an area [22], or microbots measuring the concentration of chemicals.

In all of these applications the agents (representing sensors, ants or robots) are very simple and severely limited in both memory and communication. Moreover, the communication is often not only limited but also erroneous (e.g., consider wireless communication with obstacles between robots), or received messages are subject to interpretation (e.g., when insects communicate through gestures [41]). Motivated by this unreliable communication in applications we study the simple averaging dynamic where the communication is subject to noise.

We continue with related work. The problem of distributed values converging to the average (often without noise) has been studied in various areas reaching back to early versions studied in statistics [19, 27, 32]. However, to the best of our knowledge, none of the studied models match our model. We review the related work by areas: (i) average consensus and its applications, (ii) gossip-based communication models, (iii) consensus protocols in population protocols, (iv) biological distributed algorithms, (v) noise and failures in sensor networks.

Average consensus and its applications. Consensus has been studied intensively in various settings in general network topologies, much of it under the name of *average consensus* [57, 55]. Most of this work is orthogonal to our work: First, due to the general network topology and the fact that, in each step of the studied algorithms, the agents update their values with a weighted average of <u>all</u> of their neighbors' values whereas in our averaging dynamic, an agent can only access a single other value per interaction. Second, while the potential functions in these works and the noise, if any, are usually identically or similarly defined as in our work the main goal of these papers is—just as in the classic works—to study under which circumstances the processes eventually converge to a state with a small potential function [57], whereas we are interested in the number of interactions until our process obtains a small potential. Recent papers [47, 11, 42, 15] consider the convergence rate of the weighted averaging process, but only in the noiseless setting. Average consensus has also been studied in networks with time-varying topologies [46, 51]. Variants with noisy communication were studied [57, 38], but they only consider additive noise and assume it to be zero-mean with unit variance (as mentioned before, only convergence in the limit is shown).

The noisy version of the problem also received ample attention in control theory [54, 50, 49]. Already in the early works on average consensus immediate applications of converging to the average were discovered and intensively studied, e.g., applications to load balancing between parallel machines [9, 18] or to coordinate distributed mobile agents [9, 36, 24]. For a more detailed overview on average linear consensus consult the survey [28].

Gossip-based communication models. Much closer to our work is the study of aggregating information in gossip-based model. In this model, each node can contact one of its neighbors in the network in each round and exchange information with it. Even though a node can be contacted by many neighbors in a single round, this model, if applied to the complete graph, is very similar to our synchronous model. On the complete graph [39] shows that $O(n \cdot \ln n)$ interactions are enough to approximate the average well with high probability. On the one hand they consider more general graphs (in some sense we consider the complete graph); on the other hand they do not consider noise, which simplifies their analysis of the convergence time significantly.

Consensus protocols in population protocols, biological distributed algorithms. Motivated by biological applications, population protocols have also been studied in the noisy setting in the context of biological distributed algorithms. The authors of [25] study rumor spreading and consensus in extremely faulty networks where a bit in a message can be flipped with probability $1/2 - \varepsilon$. This was later generalized in [26] to plurality consensus. The authors of [8] study the differences between pull and push rumor spreading in the noisy setting. Reaching consensus to an opinion in population protocols in the noiseless setting has received much attention (see *e.g.,* [4, 23, 1, 2, 5, 6, 20, 7, 40, 30, 29, 37]).

Noise and failures in sensor networks. The problem of converging to the average (and similar problems) have also been studied in (noisy) sensor networks [58, 52] where nodes again can interact with all their neighbors. In these networks another type of unreliable communication, i.e., packages might be dropped, has received ample attention, e.g., [12] studies the broadcast problem and [13] develops a framework to transform certain algorithms for failure free networks to also work in faulty sensor networks.

An interesting type of failure has been studied in [33]. There failures do not happen during the communication but the algorithm itself might be faulty, i.e., a state machine run at an agent might switch to a wrong state.

1.2 Formal Results

We now formally state our main theorems. For the ease of presentation, in the discussion we assume that noise is normally distributed with unit variance, $N \sim \mathcal{N}(0,1)$, but our results hold for general variance σ^2. Let $\phi_0 = \bar{\phi}(\mathbf{X}^{(0)})$ be the initial potential. Our first theorem shows that the agents converge to a small value of $\bar{\phi}(t) = O(n)$ after parallel time[2] that is logarithmic in ϕ_0/n. In particular, if we use b to denote the initial imbalance ($b = \max_{i,j}\{x_i^{(0)} - x_j^{(0)}\}$), then it takes $O(\ln b)$ parallel steps for the potential to become $\bar{\phi}(t) = O(n)$. Note that $\bar{\phi}(t) = O(n)$ means that the 'average' difference between the values of any two agents is constant and we show that the constant hidden in the O-notation is actually very small. It is worth mentioning that this is tight in two senses: (i) In expectancy we have $\bar{\phi}(t) = \Omega(n)$ for any fixed time step $t \geq n$, (i.e., after one parallel time step). Even in the case where all nodes initially have the same value, our results show that the potential increases

[2] Recall that in parallel time we scale time by a factor of n for a fair comparison with the synchronous time model.

after n interactions in expectation by $\Omega(n\mathbb{E}\left[N^2\right]) = \Omega(n)$. (ii) At least $\Omega(\ln b)$ parallel time steps are required[3] to decrease the potential to $O(n)$, since the potential only drops in expectation by a constant factor in each parallel step. The formal statement is as follows.

▷ Theorem 1 (Convergence to Running Avg.). Consider any noise-distribution $\aleph$ with (at least) exponential-decay[4]. Fix any $\delta \in \mathbb{R}$. Let $n = n(\delta)$ be large enough. The following hold:

(i) for any $t = \Omega\left(n \ln\left(\frac{\phi_0}{\delta\sigma^2 n}\right)\right)$ with probability at least $1-\delta$ we have $\bar{\phi}(\mathbf{X}^{(t)}) = O(\sigma^2 n \ln(1/\delta))$,

(ii) for any $t \geq n$ (parallel time) with constant probability we have $\bar{\phi}(\mathbf{X}^{(t)}) = \Omega(\sigma^2 n)$ and

(iii) even without noise, for any $t = o\left(n \ln\left(\frac{\phi_0}{\sigma^2 n}\right)\right)$ we have $\mathbb{E}\left[\bar{\phi}(\mathbf{X}^{(t)})\right] = \omega(\sigma^2 n)$.

While the above theorem shows a quick convergence to the running average, this does not imply convergence to the initial average. In fact, as time progresses the distance to the initial average ($TSS(\mathbf{X}^{(t)})$) is likely to increase. Nonetheless, in the case of the Gaussian white noise model we can bound the drift of the running average from the initial average in a time window of $O(n^2)$ steps (see Lem. 17). Thm. 1 roughly says that after at least $t = \Omega(n \log n)$ steps the distance to the running average is small if we start with a potential that is polynomial in n. Thus, as long as $t = \Omega(n \log n)$ and $t = O(n^2)$ we obtain $TSS(\mathbf{X}^{(t)}) = O(n)$. After the $O(n^2)$ step time window the potential starts to increase again, which, is unavoidable, due to the noise causing drift of the running average; in Gaussian white noise model, the running average after t steps diverges with constant probability from the initial average by $\frac{\sqrt{t}}{n}$ (Lem. 17). This in turn implies that $TSS(\mathbf{X}^{(t)}) \geq t/n$.

▷ Corollary 2 ((Bounded) Divergence from Initial Avg.). In the case of Gaussian white noise model, for any $\delta \in \mathbb{R}$ and large enough $n = n(\delta)$ and all $t = \Omega\left(n \ln\left(\frac{\bar{\phi}(\mathbf{X}^{(0)})}{\delta\sigma^2 n}\right)\right)$ we have

(i) '*non-divergence* for $O(n^2)$ steps', i.e., $TSS(\mathbf{X}^{(t)}) = O\left(\left(\frac{t}{n} + n\right)\sigma^2 \ln(1/\delta)\right)$ with probability at least $1 - \delta$ and

(ii) '*divergence* for $\omega(n^2)$ steps', i.e., $TSS(\mathbf{X}^{(t)}) = \Omega\left(\left(\frac{t}{n} + n\right)\sigma^2\right)$ with constant probability.

If one can bound the divergence between the running average and the initial average for a general noise-distribution $\aleph$ with (at least) exponential-decay[5] the following remark is useful to obtain a similar bound for the $TSS(\mathbf{X}^{(t)})$ as in Cor. 2. Recall that $\varnothing^{(t)} = \sum_i X_i^{(t)}/n$ and in particular, $\varnothing^{(0)}$ denotes the initial average.

▶ Remark 2. Fix any $\delta \in \mathbb{R}$. Let $n = n(\delta)$ be large enough. For any fixed $t = \Omega\left(n \ln\left(\frac{\phi_0}{\delta\sigma^2 n}\right)\right)$ with probability at least $1 - \delta$ we have $TSS(\mathbf{X}^{(t)}) = \Theta\left(n\left(\varnothing^{(t)} - \varnothing^{(0)}\right)^2 + \sigma^2 n \ln(1/\delta)\right)$.

Remark 2 follows by rewriting $TSS(t) = \bar{\phi}(\mathbf{X}^{(t)}) + n \cdot \left(\varnothing^{(0)} - \varnothing^{(t)}\right)^2$ (cf. Fact 9) and plugging in the first part of Thm. 1. Cor. 2 then follows by plugging in the bounded deviation of the running average from the initial average for the Gaussian white noise model (cf. Lem. 17).

The Influence of Rounding. Agents with limited computational power might not be able to store real values. Motivated by this we also consider the setting where agents can only store integers. In particular, we consider the case that the averaging protocol is augmented with the following rounding procedure: Assume that the noise $N \sim \aleph$ takes only integer

[3] For the case where constant fraction of the values are at distance b.
[4] In fact we only require the function to be smooth, which we define later. This class is much broader and contains most of the famous distributions including the normal distribution, geometric distribution and the Poisson distribution.
[5] Again, we only require the function to be smooth, which we define in Sec. 3.

variables. After a node i receives the value from node j, the node averages it as before and then rounds up or down with equal probability. In the full version we show how to relate the setting of rounding to the original setting allowing us to derive the following corollary.

▷ Corollary 3. The bounds of Thm. 1 and Cor. 2 hold even if rounding is used.

The Synchronous Model. In the full version, we show how our results extend to the synchronous setting. It turns out that the results are the same up to a rescaling of time.

▷ Corollary 4 (Synchronous Setting). The bounds of Thm. 1 and Cor. 2 hold even in the synchronous setting, where time is rescaled by a factor of $2/n$.

Experimental Results. In Sec. 5, we simulate the averaging dynamic in various settings. In the first setting, we consider the distribution of the distances between agents' values and the running average. Our simulations show that these distances seem to follow an exponential law, i.e., the concentration is even stronger than what Thm. 1 implies.

Due to the limited memory of agents it would be desirable to obtain similar results as in Thm. 1 for the averaging dynamic in the setting where agents can only store values from a bounded range. However, our simulations in Sec. 5 show that this setting leads to a much faster (by order of magnitude) divergence between the running average and the initial average.

1.3 Technical Contributions

While it is not hard to show that in expectation the potentials $TSS(t)$ and $\bar{\phi}(t)$ decrease in one step as long as their value is large, it is surprisingly challenging to derive probabilistic bounds on either potential at an arbitrary point in time, *i.e.,* bounds of the type $\mathbb{P}\left[\bar{\phi}(t) \geq b\right] \leq p(b)$. Two of the reasons are as follows. (i) The potential decreases (expectedly) only conditioned on the fact that it is large enough. In fact, when the potential is small, then due to the noise it will increase in expectation. (ii) Since we study general distributions and in particular the normal distribution, the noise in a given round can be arbitrarily large leading to an arbitrarily large increase in $\bar{\phi}(t)$; if the protocol runs long enough (possibly exponentially long in n) we, indeed, will have encountered some time steps with a very large potential increase. There are surprisingly few analytical tools for using potentials as $\bar{\phi}(t)$ with challenges (i) and (ii). One notable exception is Hajek's theorem [34], which can be used to bound the value of such a potential at a given time t. However, in our setting—with our potential function—the results obtained are very weak.[6]

Instead, we use a more sophisticated approach that at its core has a decomposition of the potential change in a single time step into three additive (but dependent) random variables. We iterate this decomposition over time throughout some interval $\mathcal{I} = (t_0, t_1]$ and sum the respective variables which we will denote as $S^-(\mathcal{I})$, $S'(\mathcal{I})$, and $S^*(\mathcal{I})$. Then (cf. Pro. 12) we are able to bound the potential change at the end of the interval as

$$\bar{\phi}(\mathbf{X}^{(t_1)}) \leq \left(1 - \frac{S^-(\mathcal{I})}{t_1 - t_0}\right)^{t_1 - t_0} \cdot \bar{\phi}(\mathbf{X}^{(t_0)}) + S'(\mathcal{I}) + S^*(\mathcal{I}). \tag{1}$$

Due to the dependencies between the three variables we use strong Martingale concentration bounds to separately upper bound $S'(\mathcal{I}) + S^*(\mathcal{I})$ and lower bound $S^-(\mathcal{I})$ (cf. Lem. 13). We

[6] Hajek's theorem considers the moment generating function of the potential. In order to apply the theorem to our potential, it seems that one would need to consider a logarithmic version of the potential, which together with the moment generating function results in bound that is weaker than a simple union bound.

then use a union bound—to circumvent the dependencies—to bound each of these variables allowing us to get a bound on Eq. 1. It is critical that we define the random variable S^- in such a way that it always has an expected decrease. This is in stark contrast to the entire potential, which, as we mentioned before in (i), only decreases in expectation when it is large. Having an unconditional decrease of S^- allows us to consider arbitrarily large intervals. With these bounds at hand one can use Eq. 1 to obtain probabilistic bounds on the potential at any given point time t_1. However, due to the bound on $S'(\mathcal{I})+S^*(\mathcal{I})$ the total bound becomes very weak for large intervals. As a remedy, we carefully trace the change in the potential in different regimes (with several phases in each regime) and we separately apply the aforementioned analysis with a fresh (small) interval in each phase. The intervals (and thus also the phases) have variable length—decreasing geometrically or even exponentially, depending on the regime.

All missing proofs can be found in the full version [?].

2 Model

In this section we present the model including all assumptions. We have a collection of n agents that have initial values $X_1^{(0)}, X_2^{(0)}, \ldots, X_n^{(0)}$. Time is discrete and $X_i^{(t)}$ denotes the value of agent $i \in [n]$ at time t. Recall that $\varnothing^{(t)} = \sum_i X_i^{(t)}/n$ denotes the average value at time t; in particular, $\varnothing^{(0)}$ denotes the initial average. For two random variables X and Y we write $X \stackrel{d}{=} Y$ if they have the same (probability) distribution. Next, we define the communication models.

▶ **Definition 5** (Communication Models). We consider two communication models.

(i) *Sequential model*: At every discrete time step two of the agents i, j are chosen uniformly at random (with replacement[7]) and send their current values x_i and x_j to each other, where the values received are $x_i + N_i$ and $x_j + N_j$, where $N_i, N_j \stackrel{d}{=} N$.

(ii) *Synchronous model*: At every discrete time step a perfect matching is chosen u.a.r. among all perfect matchings on the n agents[8]. All matched agents interchange their values as in the sequential model.

We use the *parallel time*, which was first defined in [3], to denote the time step t/n in the sequential model. This notion eases the comparison of results in both models, as the total number of interactions is up to a factor of 2 equal.

▶ **Definition 6** (Noise Models). Let v be the value sent by an agent. The value received is $v+N$, where N is distributed according to some zero-mean noise distribution $\aleph$ with $\sigma^2 = \mathrm{Var}[N]$.

We consider general noise distributions and our results depend on the moments of N. The following two models are of special interest in this paper.

(i) *Gaussian white noise model* where $\aleph = \mathcal{N}(0, \sigma^2)$ for an arbitrary σ.

(ii) *Discrete white noise model* where $\aleph = \mathcal{D}(p)$, with $\mathbb{P}[N = i] = \frac{1}{2}p(1-p)^{|i|}$, for $i \in \mathbb{Z} \setminus \{0\}$ and $\mathbb{P}[N = 0] = p$, where $p \in (0, 1]$. Note that $\mathrm{Var}[N] = \frac{1-p}{p^2}$.

From now on we assume that the noise N is distributed according to a fixed noise distribution $\aleph$ that is independent of n.

▶ **Definition 7** (Averaging Dynamic). We consider the real valued and the discrete valued algorithm. A node with value v at time receiving the input w sets its new value to

[7] This is not crucial to our results, but simplifies the calculations slightly.

[8] Again, we allow matchings of the kind (i, i) for simplicity. It is easy but slightly less aesthetic to modify our results to exclude matchings (i, i).

(i) $v' = (v+w)/2$ in the *real valued model.*

(ii) $v' = \begin{cases} \lceil (v+w)/2 \rceil & \text{w.p. } \frac{1}{2} \\ \lfloor (v+w)/2 \rfloor & \text{otherwise} \end{cases}$ in the *discrete valued model.*

A probability distribution $\mathcal{D}$ is called *sub-Gaussian* if for $X \sim \mathcal{D}$ we have that there exists positive constants c_1, c_2 such that for every x we have $\mathbb{P}[\,|X| \geq x\,] \leq c_1 \exp(-c_2 x^2)$.

Whenever we calculate the new values $\mathbf{X}^{(t+1)}$ by conditioning on the current state, $\mathbf{X}^{(t)} = \mathbf{x}^{(t)}$ we use small letters $x_i^{(t)}$ to denote fixed values and capitalized letters $X_i^{(t+1)}$ to denote random variables. Furthermore, we use bold-face to denote vectors. Throughout the paper we will assume that the number of agents n is large enough and in particular $n\mathbb{E}\left[N^2\right] \geq 1$.

We define the following potentials which are essential in all our proofs and formal results.

▶ **Definition 8** (Potentials).

$$TSS(\mathbf{x}^{(t)}) = \sum_i \left(x_i^{(t)} - \varnothing^{(0)}\right)^2, \quad \bar{\phi}(\mathbf{x}^{(t)}) = \sum_i \left(x_i^{(t)} - \varnothing^{(t)}\right)^2, \quad \phi(\mathbf{x}^{(t)}) = \sum_{i,j} \left(x_i^{(t)} - x_j^{(t)}\right)^2.$$

When clear from the context we drop the time index t and we write $\mathbf{x}$ instead of $\mathbf{x}^{(t)}$, x_i instead of $x_i^{(t)}$, etc. Similarly we will use the following short forms $TSS(t) = TSS(\mathbf{x}^{(t)})$ and $\bar{\phi}(t) = \bar{\phi}(\mathbf{x}^{(t)})$. We emphasize that the difference between $\bar{\phi}(\mathbf{x})$ and $TSS(t)$ is that the former measures the squared distance w.r.t. the *running average* and the latter w.r.t. *initial average*. Initially, we have $\bar{\phi}(\mathbf{x}^{(0)}) = TSS(0)$. In Appendix B we prove the following fact that shows how $\bar{\phi}(\mathbf{X}^{(t)})$ relates to $TSS(t)$ and how $\bar{\phi}$ relates to ϕ.

▷ Fact 9. We have that (i) $TSS(t) = \bar{\phi}(\mathbf{X}^{(t)}) + n \cdot \left(\varnothing^{(0)} - \varnothing^{(t)}\right)^2$ and (ii) $\phi(\mathbf{x}) = 2n \cdot \bar{\phi}(\mathbf{x})$.

Note that many alternative ways to define the potential at a time t such as the max distance and $\ell 1$ norm give only a very partial picture: The max distance to the mean for example does not distinguish between just one node being far and all nodes being far. On the other hand, the $\ell 1$ norm does not 'punish' outliers enough: there is no difference between n nodes being off by 1 from the average and one node being off by n.

Notation

We use $X \sim \mathcal{D}$ to denote that X is distributed according to probability distribution $\mathcal{D}$. For two random variables X and Y we write $X \leq^{\text{st}} Y$ if X is *stochastically dominated* by Y, i.e., $\mathbb{P}[\,X \geq x\,] \leq \mathbb{P}[\,Y \geq x\,]$ for all $x \in \mathbb{R}$. We use $\|\mathbf{x}\|_2$ to denote the $L2$-norm. In the sequential model we have two random variables $N_1^{(t)}$ and $N_2^{(t)}$ for the noise of the channel at time step t (recall that $N_1^{(t)}$ and $N_2^{(t)}$ are distributed according to $\aleph$). We define the following two random variables $N'^{(t)}$ and $N^{*(t)}$ that will play a key role in our analysis:

$$N'^{(t)} = \left(N_1^{(t)}\right)^2 + \left(N_2^{(t)}\right)^2, \qquad\qquad N^{*(t)} = N_1^{(t)} + N_2^{(t)}\ .$$

▷ Fact 10. In the Gaussian noise model, we have $N^{*(t)} \sim \mathcal{N}(0, 2\sigma^2)$ and $N'^{(t)} \sim \Gamma(1, 2\sigma^2)$, where $\Gamma(\cdot,\cdot)$ denotes the gamma distribution.

When clear from the context we simply write N' and N^* instead of $N'^{(t)}$ and $N^{*(t)}$, respectively. We use $\mathcal{F}_t$ to denote the filtration at time t, which encapsulates all randomness up to time t as well as the initial values of the nodes; hence it defines the state at time t completely.

3 The Sequential Setting: Convergence towards the Running Average

Conditioning on all the randomness until time t, *i.e.,* conditioning on $\mathcal{F}_t$, we define $\Delta^{(t+1)} = \begin{cases} \frac{\left(x_i^{(t)}-x_j^{(t)}\right)^2}{2\bar{\phi}(\mathbf{x}^{(t)})} & \text{for } \bar{\phi}(\mathbf{x}^{(t)}) > 0 \\ 1/n & \text{otherwise} \end{cases}$, where i and j are the chosen in round t.

▷ Lemma 11 (One Step Bound). Fix an arbitrary potential at time t. Suppose the pair i, j was chosen to communicate and condition on the filtration $\mathcal{F}_t$ (all events that happened up to round t). Then, the following holds

$$\bar{\phi}(\mathbf{X}^{(t+1)}) - \bar{\phi}(\mathbf{x}^{(t)}) \leq -\Delta^{(t+1)}\bar{\phi}(\mathbf{x}^{(t)}) + \frac{N'^{(t+1)}}{4} + N^{*(t+1)}\left(\frac{x_i^{(t)} + x_j^{(t)}}{2} - \varnothing^{(t)}\right).$$

Further we have $\mathbb{E}\left[\, \Delta^{(t+1)} \mid \mathcal{F}_t \,\right] = \frac{1}{n}$.

In order to prove the statement, we first calculate the exact expected change in one step (which we do in the full version). We then majorize (stochastic dominance) with the slightly more convenient statement above.

For an arbitrary time interval $\mathcal{I}$ define

$$S'(\mathcal{I}) = \sum_{\tau\in\mathcal{I}} N'^{(\tau)}/4, \quad S^*(\mathcal{I}) = \sum_{\tau\in\mathcal{I}} N^{*(\tau)}\left(\frac{x_i^{(\tau-1)} + x_j^{(\tau-1)}}{2} - \varnothing^{(\tau)}\right), \quad S^-(\mathcal{I}) = \sum_{\tau\in\mathcal{I}} \Delta^{(\tau)}.$$

Note that, in the definition of S^*, we sum up over all time steps τ in the interval $\mathcal{I}$ and we consider the pair i and j that is chosen in round τ (in each round a different pair i and j can be chosen). With Lem. 11 and the definitions of S', S^* and S^- we can deduce the following decomposed bound on the potential for an arbitrary interval.

▷ Proposition 12 (Decomposition of Potential). Fix arbitrary t_0, t_1 and consider the interval $\mathcal{I} = (t_0, t_1]$. For $t = t_1 - t_0$ we have that

$$\bar{\phi}(\mathbf{X}^{(t_1)}) \leq \left(1 - \frac{S^-(\mathcal{I})}{t}\right)^t \bar{\phi}(\mathbf{X}^{(t_0)}) + S'(\mathcal{I}) + S^*(\mathcal{I}). \tag{1}$$

Our results only hold for smooth noise distributions, which we define in the following. Let $m_{t,\delta} = \arg\max_{\ell} \left\{\mathbb{P}\left[\max\left(\left\{N'^{(t_0)}, \ldots, N'^{(t_0+t)}\right\} \cup \left\{N^{*(t_0)}, \ldots, N^{*(t_0+t)}\right\}\right) \leq \ell\right] \geq 1-\delta\right\}$.

▶ **Definition 13.** A noise distribution $\aleph$ is *smooth* if for all $\delta > 0$ and all $t > 0$ we have $m_{t,\delta} \leq \left(\frac{t}{\delta}\right)^{1/20}$.

Any (sub-)linear probability distribution and even some inverse polynomial distributions are smooth. Thus many practically relevant distributions such as Gaussian, binomial and Poisson distributions are smooth. For example, for the standard normal distribution ($N \sim \mathcal{N}(0,1)$) we have $m_{t,\delta} = \log(t/\delta)$, since in each time step the probability that the N^2 exceeds $\log(t/\delta)$ is equal to the probability that N exceeds $\sqrt{\log(t/\delta)}$ which happens w.p. at most δ/t. Taking union bound over all t steps shows that it is smooth.

Using strong martingale concentration bounds (Thm. 22 and Thm. 23) and bounding the variance, we deduce the following upper bound on $S^* + S'$ and lower bound on S^-.

▷ Lemma 14. Let t_0, t_1 be such that $t_1 > t_0$ and consider the interval $\mathcal{I} = (t_0, t_1]$.
(i) With probability $1-\delta$ we have

$$S^*(\mathcal{I}) + S'(\mathcal{I}) \leq$$

$$\frac{t}{4}\mathbb{E}\left[\, N' \,\right] + 5\sqrt{\frac{t}{n}}\left(\ln(4t/\delta)m^*_{t,\delta/4}\right)^2 \left(2 + \mathbb{E}\left[\, N' \,\right]\right)\sqrt{\bar{\phi}(\mathbf{x}^{(t_0)}) + 9t\mathbb{E}\left[\, N' \,\right] + 2}\,.$$

(ii) For any $\gamma < 1$, w.p. at least $1 - \exp\left(-\frac{3\gamma^2 t}{8n}\right)$ we have $S^-(\mathcal{I}) \geq (1-\gamma)\frac{t}{n}$.

The following proposition almost directly implies Thm. 1.

▷ Proposition 15. Fix any $\delta \in (0, 1]$ and assume that the noise distribution is smooth. There exists a constant c such that for a time step t_0 with potential $\bar{\phi}(\mathbf{x}^{(t_0)})$ we have

$$\mathbb{P}\left[\bar{\phi}(\mathbf{X}^{(t^*)}) \geq \ln(1/\delta)n\mathbb{E}\left[N'\right] + b \mid \mathcal{F}_{t_0}\right] \leq \delta,$$

where $t^* = t_0 + cn\ln\left(\frac{\bar{\phi}(\mathbf{x}^{(t_0)})}{\mathbb{E}[N']n\delta}\right)$ and $b = 2\left(1 + \mathbb{E}\left[N'\right]\right)\left(\ln(1/\delta)\right)^9 n^{9/10}$.

Proof Sketch. We only sketch the proof idea for a simplified setting; during the sketch we assume that $N \sim \mathcal{N}(0, 1)$ (with $\mathbb{E}\left[N'\right] = O(1)$) and also that δ is at least $1/n^3$. The main ingredients for the proof are Pro. 12 and Lem. 13. For an interval $\mathcal{I} = (t_0, t_1]$ Pro. 12 upper bounds the potential at time t_1 by

$$\bar{\phi}(\mathbf{X}^{(t_1)}) \leq \left(1 - \frac{S^-(\mathcal{I})}{t}\right)^t \bar{\phi}(\mathbf{X}^{(t_0)}) + S'(\mathcal{I}) + S^*(\mathcal{I}), \tag{2}$$

where t is the length of the interval. Lem. 13 lower bounds $S^-(\mathcal{I})$ and upper bounds the sum $S'(\mathcal{I}) + S^*(\mathcal{I})$. To prove Pro. 15 we have to show that the initial potential $\bar{\phi}(\mathbf{x}^{(t_0)})$ decreases to $O(n)$ after $O(n \cdot \log\bar{\phi}(\mathbf{x}^{(t_0)}))$ time steps with probability $1 - \delta$. Optimally, we would use a single application of Pro. 12 to upper bound the potential as in Eq. 2 and then bound the terms $S^-(\mathcal{I})$ and $S'(\mathcal{I}) + S^*(\mathcal{I})$ via Lem. 13. However, the bounds on S^- and $S' + S^*$ given by Lem. 13 are too loose to yield the desired result via a single application of Pro. 12 and Lem. 13 with the whole time interval $\mathcal{I} = [t_0, t_0 + O(n\log\bar{\phi}(\mathbf{x}^{(t_0)}))]$. For example, the bound on $S' + S^*$ inherently has a term of order $\sqrt{\bar{\phi}}$, where $\bar{\phi}$ is the potential at the start of the interval for which Lem. 13, (i) is applied. Thus a one shot proof as described above can never reach a potential below $\sqrt{\bar{\phi}}$. This is not sufficient if the initial potential is large, e.g., say for $\bar{\phi} \gg n^{8/3}$.

To circumvent this problem we apply Pro. 12 and Lem. 13 several times for smaller time intervals: More detailed, we split the proof of Pro. 15 into two regimes. In regime 2 we use several phases to decrease the potential to $\Theta(n^{4/3})$. If the potential is $\bar{\phi}$ at the beginning of a phase a single application of Pro. 12 and Lem. 13 reduces the potential to $\bar{\phi}^{3/4}$. The length of each such phase is geometrically decreasing by a factor $3/4$ where the first phase is of length $O\left(n\ln\left(\frac{\bar{\phi}(\mathbf{x}^{(t_0)})}{n\delta}\right)\right)$. After the last phase of regime 2 the potential is of order $n^{4/3}$.

Then, in regime 1 the potential reduces from $\Theta(n^{4/3})$ to $O(n)$, again through several phases. If the first phase of regime 1 starts with a potential of size B, the phase has length $t = O(n\ln(B))$. If there was no additive increase due to the noise, then this would reduce the potential to 0. However, there is an additive increase of $\Theta(t) = \Theta(n\ln(B))$ which leaves us with a potential of size $O(n\ln(B))$. The next phase will therefore be of length $n\ln\ln(B)$ etc. This is repeated for $\ln^*(B)$ phases until the potential reduces to $O(n)$, which, as we explained in Sec. 1.2, is the furthest the potential can be decreased .

Putting everything together, we get that after $O\left(n\ln\left(\frac{\bar{\phi}(\mathbf{x}^{(t_0)})}{n\delta}\right)\right)$ rounds the potential reduces to $O(n)$. ◀

The full proof of Pro. 15 handles general $\mathbb{E}\left[N'\right]$ and general δ and thus it is significantly more technical. It can be found in the full version. From Pro. 15 we are able to derive Thm. 1.

4 Deviation from the Initial Average

An informal argument for the statements in this section in the special case of $\sigma = 1$ can be found in [56]. Before we state our results we need the following result on the standard normal distribution.

▷ Theorem 16 ([17]). Let $\Phi(x)$ denote the cumulative distribution function of the standard normal distribution. We have for $x \geq 0$:

$$\frac{1}{\sqrt{2\pi}} \frac{x}{x^2+1} \exp\left(-x^2/2\right) \leq \Phi(x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left(-x^2/2\right).$$

We can now state and prove the main results of this section.

▷ Lemma 17. For any t and any $\delta < 1$, we have $\varnothing^{(t)} - \varnothing^{(0)} \sim \frac{\sum_{\tau=1}^{2t} N^{(\tau)}}{2n}$ with probability at least $1-\delta$, where $N^{(\tau)}$ is the noise of the channel. In particular, for the Gaussian white noise model setting where $N \sim \mathcal{N}(0,\sigma^2)$ we have $\sum_{\tau=1}^{2t} N^{(\tau)} \sim \mathcal{N}(0, 2t\sigma^2)$. Thus

(i) $|\varnothing^{(t)} - \varnothing^{(0)}| \leq \frac{\sigma\sqrt{t\ln(1/\delta)}}{n}$ w.p. at least $1-\delta$

(ii) $|\varnothing^{(t)} - \varnothing^{(0)}| \geq \frac{\sigma\sqrt{t\ln(1/\delta)}}{n}$ w.p. at least $\frac{\delta}{2\sqrt{2\ln(1/\delta)}}$.

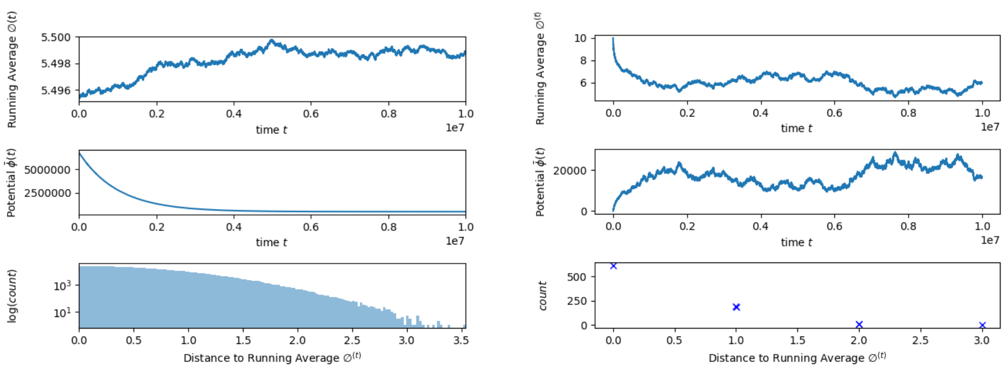
Using the Berry-Esseen theorem, one can easily prove similar bounds for any distribution with bounded third moment including *discrete white noise*.

In the following we consider the potential $(\varnothing_t)_{t\geq 0}$ as a Martingale allowing us to use Thm. 22 to derive the desired concentration bounds. The following bound is weaker than the aforementioned bounds, however, it is useful whenever the noise is such that $m_{t,\delta/(2t)}$ is small.

▷ Proposition 18. For any $t \geq 2$ and any $\delta < 1$, we have $-m_{t,\delta/(2t)}\sigma\sqrt{2t} \leq \varnothing^{(t)} - \varnothing^{(0)} \leq m_{t,\delta/(2t)}\sigma\sqrt{2t}$ with probability at least $1-\delta$.

5 Experimental Results

The goal of this section is twofold. First, we seek to better understand the distribution $\mathcal{D}$ of the distances $x_i^{(t)} - \varnothing^{(t)}$. Second, we simulate a setting in which the range of values is bounded, motivated by computational and storage limited agents. All results in this section are based on an implementation of the simple averaging dynamic. The code (python3) for the experiments can be found here [44].



(a) The setting of this example is: $n = 10^6$, initial distribution of values is uniformly at random in the range $[1, n^2]$, $10n$ iterations, Gaussian white noise with variance 1, unbounded range.

(b) The setting of this example is: $n = 1000$, all values equal to 10, using discrete white noise model $\mathcal{D}(0.8)$ (see Definition 6), bounded range in the interval $[1, 10]$, $10^4 n$ iterations. The avg. of the values drifts from 10 to 6.

Figure 1 The figure depicts the distribution of distances as well as the bounded value setting.

5.1 The Distribution of the Distances

The experiments suggest that the distance decays at least exponentially. Note that the experiments only show a single iteration, however, this phenomena was observable in every single run. The bound on $\mathbb{E}\left[\bar{\phi}\left(\mathbf{X}^{(t)}\right)\right]$ we obtained in Thm. 1 only implies that $\mathcal{D}$ is at most $O(1/d^3)$. However, we conjecture, for sub-Gaussian noise that $\mathbb{P}\left[|X_i^{(t)} - \varnothing^{(t)}| \geq x\right] = O(\exp^{-x})$ (cf. Fig. 1a). Showing this rigorously is challenging due to the dependencies among the values. Nonetheless, such bounds are very important since they immediately bound the maximum difference and we consider this the most important open question.

5.2 The Bounded Values Setting

One of the motivations for the very simple averaging dynamic arises in the setting of limited computational power of the interacting agents. So far we assumed that agents can store and transmit (intermediate) values from an unbounded range. For many applications and in particular motivated by agents with bounded memory one would hope for similar results if there is a maximum and a minimum value that can be stored or transmitted. The formal definition is as follows: values can only be from the range $[v_{min}, v_{max}]$ ($= [1, 10]$ in our experiments). We assume noise of the channel cannot produce values larger than v_{max} or smaller than v_{min}, which can be motivated as follows in the setting where the values correspond to amplitudes: here v_{max} and v_{min} are simply the amplitudes (high amplitude and no amplitude) where the signal-to-noise ratio is very large, and noise becomes negligible. An equivalent model is that the agents know the range of possible communication values, and hence, they can simply correct every value larger than v_{max} to v_{max}. In particular when agents only have limited storage, the communication range will often be bounded, and even rounding might become necessary (see the full version).

We refer to these equivalent models as the model with *cutoffs*. While the experiments indicate that values still converge towards the running average, there is a clear drift of the running average from the initial average if the input values are chosen unsuitably. In our experiments, we set the range of values to $[1, 10]$, use the noise described in the discrete noise model together with rounding. Initially, all agents have value 10. We see a drastic drift of the running average (see Fig. 1b). Even though the initial average is 10, the running average appears to approach the midpoint of the range, i.e., 5. The histogram of distances to the initial average shows even more clearly that the values are not concentrated around the initial average. Although the experiments only show a single iteration, this phenomena was observable in every single run. We believe that the reason for this is simply that the noise is no longer symmetric and no longer zero-mean due to the cutoffs $[1, 10]$. Proving convergence to the running-average in this model seems challenging and interesting.

We believe that the insights in bounding this potential might be useful in similar problems.

6 Conclusion and Open Problems

In this paper we showed bounds on the convergence time for the unbounded setting. Our simulations in Sec. 5 yield two interesting open problems: (i) study the setting where the values are restricted to some interval (in this case the noise is no longer symmetrical) and (ii) prove tail bounds on the distance distribution w.r.t. to the running or initial average. Another interesting research direction is to move away from zero-mean noise and consider biased noise models: how quickly can the bias(es) be estimated and is convergence still feasible by compensating for the (learned) bias?

References

1. Dan Alistarh, James Aspnes, David Eisenstat, Rati Gelashvili, and Ronald L. Rivest. Time-space trade-offs in population protocols. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 2560–2579, 2017. URL: https://doi.org/10.1137/1.9781611974782.169, doi:10.1137/1.9781611974782.169.
2. Dan Alistarh, James Aspnes, and Rati Gelashvili. Space-optimal majority in population protocols. *CoRR*, abs/1704.04947, 2017. URL: http://arxiv.org/abs/1704.04947, arXiv:1704.04947.
3. Dan Alistarh, Rati Gelashvili, and Milan Vojnović. Fast and exact majority in population protocols. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing*, PODC '15, pages 47–56, New York, NY, USA, 2015. ACM. URL: http://doi.acm.org/10.1145/2767386.2767429, doi:10.1145/2767386.2767429.
4. Dana Angluin, James Aspnes, Zoë Diamadi, Michael J. Fischer, and René Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing*, 18(4):235–253, 2006. URL: https://doi.org/10.1007/s00446-005-0138-3, doi:10.1007/s00446-005-0138-3.
5. Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, Riccardo Silvestri, and Luca Trevisan. Simple dynamics for plurality consensus. *Distributed Computing*, 30(4):293–306, 2017. URL: https://doi.org/10.1007/s00446-016-0289-4, doi:10.1007/s00446-016-0289-4.
6. Petra Berenbrink, Andrea E. F. Clementi, Robert Elsässer, Peter Kling, Frederik Mallmann-Trenn, and Emanuele Natale. Ignore or comply?: On breaking symmetry in consensus. In *Proceedings of the ACM Symposium on Principles of Distributed Computing, PODC 2017, Washington, DC, USA, July 25-27, 2017*, pages 335–344, 2017. URL: https://doi.org/10.1145/3087801.3087817, doi:10.1145/3087801.3087817.
7. Petra Berenbrink, Robert Elsässer, Tom Friedetzky, Dominik Kaaser, Peter Kling, and Tomasz Radzik. A population protocol for exact majority with o(log5/3 n) stabilization time and theta(log n) states. In *32nd International Symposium on Distributed Computing, DISC 2018, New Orleans, LA, USA, October 15-19, 2018*, pages 10:1–10:18, 2018. URL: https://doi.org/10.4230/LIPIcs.DISC.2018.10, doi:10.4230/LIPIcs.DISC.2018.10.
8. Lucas Boczkowski, Ofer Feinerman, Amos Korman, and Emanuele Natale. Limits for rumor spreading in stochastic populations. In *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, pages 49:1–49:21, 2018. URL: https://doi.org/10.4230/LIPIcs.ITCS.2018.49, doi:10.4230/LIPIcs.ITCS.2018.49.
9. J. E. Boillat. Load balancing and poisson equation in a graph. *Concurrency: Pract. Exper.*, 2(4):289–313, November 1990. URL: http://dx.doi.org/10.1002/cpe.4330020403, doi:10.1002/cpe.4330020403.
10. Henrik Brumm. *Animal communication and noise*, volume 2. Springer, 2013.
11. Jingjing Bu, Maryam Fazel, and Mehran Mesbahi. Accelerated consensus with linear rate of convergence. In *2018 Annual American Control Conference (ACC)*, pages 4931–4936. IEEE, 2018.
12. Keren Censor-Hillel, Bernhard Haeupler, D. Ellis Hershkowitz, and Goran Zuzic. Broadcasting in noisy radio networks. In *Proceedings of the ACM Symposium on Principles of Distributed Computing, PODC 2017, Washington, DC, USA, July 25-27, 2017*, pages 33–42, 2017. URL: http://doi.acm.org/10.1145/3087801.3087808, doi:10.1145/3087801.3087808.
13. Keren Censor-Hillel, Bernhard Haeupler, D. Ellis Hershkowitz, and Goran Zuzic. Erasure correction for noisy radio networks. *CoRR*, abs/1805.04165, 2018. URL: http://arxiv.org/abs/1805.04165, arXiv:1805.04165.
14. Biao Chen, Ruixiang Jiang, Teerasit Kasetkasem, and Pramod K Varshney. Channel aware decision fusion in wireless sensor networks. *IEEE Transactions on Signal Processing*, 52(12):3454–3458, 2004.

15 Ge Chen, Le Yi Wang, Chen Chen, and George Yin. Critical connectivity and fastest convergence rates of distributed consensus with switching topologies and additive noises. *IEEE Transactions on Automatic Control*, 62(12):6152–6167, 2017.

16 Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Math.*, 3(1):79–127, 2006. URL: https://projecteuclid.org:443/euclid.im/1175266369.

17 John Cook. Upper and lower bounds for the normal distribution function. https://www.johndcook.com/blog/norm-dist-bounds/, 2018. [Online; accessed 6-September-2018].

18 G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. *J. Parallel Distrib. Comput.*, 7(2):279–301, October 1989. URL: http://dx.doi.org/10.1016/0743-7315(89)90021-X, doi:10.1016/0743-7315(89)90021-X.

19 Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

20 David Doty and David Soloveichik. Stable leader election in population protocols requires linear time. *Distributed Computing*, 31(4):257–271, 8 2018. URL: https://doi.org/10.1007/s00446-016-0281-z, doi:10.1007/s00446-016-0281-z.

21 Julia T. Ebert, Melvin Gauci, and Radhika Nagpal. Multi-feature collective decision making in robot swarms. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, Stockholm, Sweden, July 10-15, 2018*, pages 1711–1719, 2018. URL: http://dl.acm.org/citation.cfm?id=3237953.

22 Julia T. Ebert, Melvin Gauci, and Radhika Nagpal. Multi-feature collective decision making in robot swarms. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, Stockholm, Sweden, July 10-15, 2018*, pages 1711–1719, 2018. URL: http://dl.acm.org/citation.cfm?id=3237953.

23 Robert Elsässer, Tom Friedetzky, Dominik Kaaser, Frederik Mallmann-Trenn, and Horst Trinker. Efficient k-party voting with two choices. *CoRR*, abs/1602.04667, 2016. URL: http://arxiv.org/abs/1602.04667, arXiv:1602.04667.

24 J Alexander Fax and Richard M Murray. Information flow and cooperative control of vehicle formations. *IEEE transactions on automatic control*, 49(9):1465–1476, 2004.

25 Ofer Feinerman, Bernhard Haeupler, and Amos Korman. Breathe before speaking: Efficient information dissemination despite noisy, limited and anonymous communication. In *Proceedings of the 2014 ACM Symposium on Principles of Distributed Computing*, PODC '14, pages 114–123, New York, NY, USA, 2014. ACM. URL: http://doi.acm.org/10.1145/2611462.2611469, doi:10.1145/2611462.2611469.

26 Pierre Fraigniaud and Emanuele Natale. Noisy rumor spreading and plurality consensus. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing, PODC 2016, Chicago, IL, USA, July 25-28, 2016*, pages 127–136, 2016. URL: http://doi.acm.org/10.1145/2933057.2933089, doi:10.1145/2933057.2933089.

27 Simon French. *Group consensus probability distributions: A critical survey.* University of Manchester. Department of Decision Theory, 1983.

28 Federica Garin and Luca Schenato. *A Survey on Distributed Estimation and Control Applications Using Linear Consensus Algorithms*, pages 75–107. Springer London, London, 2010. URL: https://doi.org/10.1007/978-0-85729-033-5_3, doi:10.1007/978-0-85729-033-5_3.

29 Leszek Gasieniec and Grzegorz Stachowiak. Fast space optimal leader election in population protocols. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 2653–2667, 2018. URL: https://doi.org/10.1137/1.9781611975031.169, doi:10.1137/1.9781611975031.169.

30 Leszek Gasieniec, Grzegorz Stachowiak, and Przemyslaw Uznanski. Almost logarithmic-time space optimal leader election in population protocols. *CoRR*, abs/1802.06867, 2018. URL: http://arxiv.org/abs/1802.06867, arXiv:1802.06867.

31 Melvin Gauci, Monica E. Ortiz, Michael Rubenstein, and Radhika Nagpal. Error cascades in collective behavior: A case study of the gradient algorithm on 1000 physical agents. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017*, pages 1404–1412, 2017. URL: http://dl.acm.org/citation.cfm?id=3091319.

32 Gustavo L Gilardoni and Murray K Clayton. On reaching a consensus using degroot's iterative pooling. *The Annals of Statistics*, pages 391–401, 1993.

33 Seth Gilbert and Calvin Newport. Symmetry breaking with noisy processes. In *Proceedings of the ACM Symposium on Principles of Distributed Computing, PODC 2017, Washington, DC, USA, July 25-27, 2017*, pages 273–282, 2017. URL: http://doi.acm.org/10.1145/3087801.3087814, doi:10.1145/3087801.3087814.

34 B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(3):502–525, 1982.

35 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. URL: http://www.jstor.org/stable/2282952.

36 Ali Jadbabaie, Jie Lin, and A Stephen Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on automatic control*, 48(6):988–1001, 2003.

37 Varun Kanade, Frederik Mallmann-Trenn, and Thomas Sauerwald. On coalescence time in graphs: When is coalescing as fast as meeting?: Extended abstract. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 956–965, 2019. URL: https://doi.org/10.1137/1.9781611975482.59, doi:10.1137/1.9781611975482.59.

38 Soummya Kar and José MF Moura. *IEEE Transactions on Signal Processing*, 57(1):355–369, 2009.

39 D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 482–491, 10 2003. doi:10.1109/SFCS.2003.1238221.

40 Adrian Kosowski and Przemyslaw Uznanski. Brief announcement: Population protocols are fast. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing, PODC 2018, Egham, United Kingdom, July 23-27, 2018*, pages 475–477, 2018. URL: https://dl.acm.org/citation.cfm?id=3212788.

41 Sara Diana Leonhardt, Florian Menzel, Volker Nehring, and Thomas Schmitt. Ecology and evolution of communication in social insects. *Cell*, 164(6):1277–1287, 2016.

42 Zhongkui Li and Jie Chen. Robust consensus of linear feedback protocols over uncertain network graphs. *IEEE Transactions on Automatic Control*, 62(8):4251–4258, 2017.

43 Romain Libbrecht, Miguel Corona, Franziska Wende, Dihego O Azevedo, Jose E Serrão, and Laurent Keller. Interplay between insulin signaling, juvenile hormone, and vitellogenin regulates maternal effects on polyphenism in ants. *Proceedings of the National Academy of Sciences*, 110(27):11050–11055, 2013.

44 Frederik Mallmann-Trenn, Yannic Maus, and Dominik Pajak. Code of the experiments. URL: https://bitbucket.org/frederikmallmann/noisy-communication-code/.

45 Colin McDiarmid. *Concentration*, pages 195–248. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. URL: https://doi.org/10.1007/978-3-662-12788-9_6, doi:10.1007/978-3-662-12788-9_6.

46 Luc Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on automatic control*, 50(2):169–182, 2005.

47 Angelia Nedić and Ji Liu. On convergence rate of weighted-averaging dynamics for consensus problems. *IEEE Transactions on Automatic Control*, 62(2):766–781, 2017.

48 Theodore S Rappaport et al. *Wireless communications: principles and practice*, volume 2. prentice hall PTR New Jersey, 1996.

49 Wei Ren and Randal W Beard. *Distributed consensus in multi-vehicle cooperative control.* Springer, 2008.

50 Wei Ren, Randal W Beard, and Ella M Atkins. A survey of consensus problems in multi-agent coordination. In *American Control Conference, 2005. Proceedings of the 2005*, pages 1859–1864. IEEE, 2005.

51 David Saldana, Amanda Prorok, Shreyas Sundaram, Mario FM Campos, and Vijay Kumar. Resilient consensus for time-varying networks of dynamic agents. In *American Control Conference (ACC), 2017*, pages 252–258. IEEE, 2017.

52 Ioannis D Schizas, Alejandro Ribeiro, and Georgios B Giannakis. Consensus-based distributed parameter estimation in ad hoc wireless sensor networks with noisy links. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 2, pages II–849. IEEE, 2007.

53 Slobodan N Simić and Shankar Sastry. Distributed environmental monitoring using random sensor networks. In *Information Processing in Sensor Networks*, pages 582–592. Springer, 2003.

54 Behrouz Touri and Angelia Nedic. Distributed consensus over network with noisy links. In *Information Fusion, 2009. FUSION'09. 12th International Conference on*, pages 146–154. IEEE, 2009.

55 Lin Xiao and S. Boyd. Fast linear iterations for distributed averaging. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475)*, volume 5, pages 4997–5002 Vol.5, 12 2003. doi:10.1109/CDC.2003.1272421.

56 Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.

57 Lin Xiao, Stephen Boyd, and Seung-Jean Kim. Distributed average consensus with least-mean-square deviation. *J. Parallel Distrib. Comput.*, 67(1):33–46, January 2007. URL: http://dx.doi.org/10.1016/j.jpdc.2006.08.010, doi:10.1016/j.jpdc.2006.08.010.

58 Lin Xiao, Stephen Boyd, and Sanjay Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pages 63–70. IEEE, 2005.

A Appendix

All missing proofs can be found in the full version [?]. Here, we only present the type of generalized Hoeffding bounds that we use and the proof of Fact 9. We use the following slightly generalized versions of the Hoeffding bound (see [35]).

▷ Theorem 19 ([35]). Let $X = \sum_{i=1}^{m} X_i$ be a sum of m independent random variables with $a_i \le X_i \le b_i$ for all i. Then

$$\mathbb{P}\left[\,|X - \mathbb{E}\left[\,X\,\right]| \ge b\,\right] \le \exp\left(-\frac{2b^2}{\sum_{i=1}^{m}(b_i - a_i)^2}\right). \tag{3}$$

The following Theorem finds its origins in the work of [45].

▷ Theorem 20 ([16, Theorem 6.1]). Let X be the martingale associated with a filter $\mathcal{F}$ satisfying

(i) $\mathrm{Var}\left[\,X_i \mid \mathcal{F}_{i-1}\,\right] \le \sigma_i^2$, for $1 \le i \le m$;
(ii) $|X_i - X_{i-1}| \le M$, for $1 \le i \le m$.

Then we have

$$\mathbb{P}\left[\,X - \mathbb{E}\left[\,X\,\right] \ge b\,\right] \le \exp\left(-\frac{b^2}{2\left(\sum_{i=1}^{m}\sigma_i^2 + Mb/3\right)}\right).$$

▷ Theorem 21 ([16, Theorem 6.5]). Let X be the martingale associated with a filter $\mathcal{F}$ satisfying

(i) $\mathrm{Var}\left[\,X_i \mid \mathcal{F}_{i-1}\,\right] \le \sigma_i^2$, for $1 \le i \le m$;
(ii) $X_{i-1} - M - a_i \le X_i$, for $1 \le i \le m$.

Then we have

$$\mathbb{P}\left[\, X \le \mathbb{E}\left[\, X \,\right] - b \,\right] \le \exp\left(-\frac{b^2}{2\left(\sum_{i=1}^{m}(\sigma_i^2 + a_i^2) + Mb/3\right)}\right).$$

Throughout this paper we will frequently make use of the fact that the sum of independent variables is a martingale.

Proof of Fact 9. Consider part (i).

$$\begin{aligned} TSS(t) &= \sum_i \left(x_t^{(i)} - \varnothing^{(0)}\right)^2 = \sum_i \left(x_i - \varnothing^{(t)} + \varnothing^{(t)} - \varnothing^{(0)}\right)^2 \\ &= \sum_i \left(\left(x_t^{(i)} - \varnothing^{(t)}\right)^2 + 2(x_t^{(i)} - \varnothing^{(t)})(\varnothing^{(0)} - \varnothing^{(t)}) + \left(\varnothing^{(0)} - \varnothing^{(t)}\right)^2\right) \\ &= \bar{\phi}(\mathbf{X}^{(t)}) + 2\left(\sum_i x_t^{(i)} - n\varnothing^{(t)}\right)(\varnothing^{(0)} - \varnothing^{(t)}) + n\left(\varnothing^{(0)} - \varnothing^{(t)}\right)^2 \\ &= \bar{\phi}(\mathbf{X}^{(t)}) + n\left(\varnothing^{(0)} - \varnothing^{(t)}\right)^2. \end{aligned}$$

Consider part (ii).

$$\begin{aligned} \phi(\mathbf{x}) &= \sum_{i,j}(x_i - x_j)^2 = 2n\sum_i x_i^2 - 2\sum_{i,j} x_i x_j = 2n\sum_i x_i^2 - 2n\varnothing\sum_i x_i \\ &= 2n\left(\sum_i x_i^2 - \sum_i x_i\varnothing\right) = 2n\left(\sum_i x_i^2 - 2\sum_i x_i\varnothing + n\varnothing^2\right) = 2n\sum_i (x_i - \varnothing)^2 \\ &= 2n \cdot \bar{\phi}(\mathbf{x}). \end{aligned}$$

◀