HAL
open science

Distributed Alignment Processes with Samples of Group Average

Amos Korman, Robin Vacus

To cite this version:

Amos Korman, Robin Vacus. Distributed Alignment Processes with Samples of Group Average. IEEE Transactions on Control of Network Systems, 2022, 10.1109/TCNS.2022.3212640 . hal-03124213v4

HAL Id: hal-03124213

https://hal.science/hal-03124213v4

Submitted on 20 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed Alignment Processes with Samples of Group Average *†

Amos Korman, Robin Vacus

Abstract

This paper studies a stochastic alignment problem assuming that agents can sense the general tendency of the system. More specifically, we consider n agents, each being associated with a real number. In each round, each agent receives a noisy measurement of the system's average value and then updates its value. This value is then perturbed by random drift. We assume that both noise and drift are Gaussian. We prove that a distributed weighted-average algorithm optimally minimizes the deviation of each agent from the average value, and for every round. Interestingly, this optimality holds even in the centralized setting, where a master agent can gather all the agents' measurements and instruct a move to each one. We find this result surprising since it can be shown that the set of measurements obtained by all agents contains strictly more information about the deviation of Agent i from the average value, than the information contained in the measurements obtained by Agent i alone. Although this information is relevant for Agent i, it is not processed by it when running a weighted-average algorithm. Finally, we also analyze the drift of the center of mass and show that no distributed algorithm can achieve drift that is as small as the one that can be achieved by the best centralized algorithm.

1 Introduction

1.1 Background and Motivation

Reaching agreement, or approximate agreement, is fundamental to many distributed systems, including: computer networks, mobile sensor systems, animal groups, and neural networks [24, 26, 31]. In the natural world, one fascinating example occurs during *cooperative transport* in ants, in which a group of ants join forces to carry a large or heavy food load [10, 11]. While carrying the load, improved consensus on the direction of forces applied by the ants amounts to faster movement of the carried load, since less forces are canceling each other. Another beautiful example concerns flocking birds (or, e.g., schooling fish) that manage to maintain a cohesive direction of movement by constantly aligning their directions of flights [31]. In these contexts, as well as in multiple other contexts, such as during clock synchronization [27, 28], the space in which the process occurs is continuous, measurements are noisy and the output needs to be maintained over time, while being subject to drift.

In several contexts, distributed consensus algorithms can be designed by letting each agent repeatedly execute the following two steps.

1. obtain an estimate regarding the average "opinion" of other agents in the system.
2. shift the "opinion" by a certain extent towards the estimated average.

The first step assumes that agents are somehow able to sense the general behavior of the system. Such an ability may be quite limited in many contexts, but not in all. We shall soon provide several examples for scenarios in which agents are able to measure global parameters of the system, either directly, or indirectly.

*©2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published article can be found at `https://ieeexplore.ieee.org/document/9913713`.

†This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 648032).

However, before we list these examples, we would like to stress that the current paper is not concerned with the mechanism behind implementing such samples, and instead focuses on optimizing the second step, corresponding to the shifts that should be made after obtaining such samples.

Examples of global parameter samplings

We next provide several examples for settings in which agents can sample global parameters. The emphasis is on examples from the biological world. We classify these examples into three categories.

- *Using an external entity as a relay of information.* In several contexts, agents may use an external entity (or a "leader" agent) to gather information from agents, compute a certain function of it (e.g., the average value), and then communicate this information to all others. Indirectly, this happens, for example, during the process of *cooperative transport*, in which agents physically carry a large or heavy object, as mentioned above. In this process the carried object itself serves as a medium entity through which the global tendency of the group could be sensed by each carrying agent. In the context of ants, for example, the authors of [10, 11, 17, 20] assumed that each ant uses the carried object to sense the sum of the forces exerted by all carrying ants. After measuring this sum (which represents the velocity vector of the carried object), each ant decides whether, and by how much, it should align its force to this measurement. Cooperative transport following a similar approach was also studied in the context of robotics, see, e.g., [4, 32]. Other natural examples include the behavior of a swarm of flying midges [13], and during firefly synchronization [14]. In such swarms, each midge (respectively, firefly) is affected by the acoustic (respectively, visual) field produced by all others. In these examples, the environment plays the role of relaying global information to agent.

- *Broadcast in clique networks.* Obtaining measurements of the whole system is also relevant for relatively packed systems, so that the communication is essentially being broadcasted over a clique. In the natural world, the clique abstraction can find relevance in small flocks of birds, or in other small animal groups [7, 29].

- *Well-mixed populations.* The assumption that agents sense global parameters of the system also finds relevance in highly mobile systems. In particular, when the population is well-mixed, the pattern of communication may be captured by a random process, in which the set of neighbors is constantly being replaced by agents chosen uniformly at random. In such cases, sensing the current set of neighbors can serve as a fresh random sample of the global population. This perspective has recently been adopted by multiple works in theoretical computer science (although in contrast to this paper, the typical assumption in these works is that the set of opinions is discrete), see, e.g., [2, 3, 5, 8, 9, 12]. For example, in *Majority Dynamics*, each agent pulls the opinions of a few random agents, and changes its opinion to the majority opinion among these samples [2, 3, 8, 12]. As such, these protocols can be viewed as being composed of two parts: (1) obtain a noisy sample of the majority opinion in the population based on the opinions of the pulled agents, and (2) align to this majority.

Informally, the *alignment* problem we study, considers a group of n agents positioned on the real line, aiming to be located as close as possible to one another. We stress that the setting on the line does not necessarily aim to capture the physical locations of the agents, which may actually be scattered in arbitrary domains (e.g., two or three dimensional Euclidean spaces). Instead, the positions on the line aim to capture the "opinions" of the agents, which, in this preliminary work, are restricted to a one dimensional space. For example, in view of general alignment problems, such as flocking, the position of an agent aims to model its direction. When flocks of birds fly, the altitude axis often remains stable, and hence the flight could be viewed as two-dimensional, implying that the direction could be represented as one-dimensional. A similar observation holds in many contexts of two-dimensional navigation, for example, during *cooperative transport* by ants. In view of clock-synchronization [27, 28] or firefly synchronization [14], the position of an agent models its clock (i.e., time), which again can be viewed as a one-dimensional space [1].

[1] Depending on the application, the actual domain may be bounded, or periodic. For example, when modeling directions, the domain is $[-\pi, \pi]$ and when modeling clock synchronization, the domain may be $[0, T]$ for some phase duration T. Since we are interested in the cases where agents are more or less aligned, approximating an interval domain with the real line is not expected to reduce the generality of our results.

Initially, agents' positions are sampled from a Gaussian distribution around 0. Execution proceeds in discrete rounds, where in each round, each agent receives a noisy measurement of its current deviation from the average position of others. Then, governed by the rules of its algorithm, each agent performs a *move* to re-adjust its position. Subsequently, before the next round begins, the position of each agent is perturbed following random drift. The drift component may model unreliability due to external conditions (e.g., perturbation of direction by wind), or the inability of an agent to precisely adjust itself as it wishes. Both noises in measurements and random drifts are governed by Gaussian distributions. We are mostly interested in the following questions:

- Which re-adjustment rule should agents adopt if their goal is to minimize expected distance of each agent from the average position?
- Could further communication between agents (e.g., by sharing measurements) help?
- Which protocols minimize the drift of the center of mass?

Importantly, we assume that agents are unaware of the actual value of their current positions, and of the realizations of the random drifts, and instead, must base their movement decisions only on noisy measurements of relative positions. This lack of global "sense of orientation" prevents the implementation of the trivial distributed protocol in which all agents simply move to a predetermined point, say 0.

One trivial algorithm is the "fully responsive" protocol, where in each round, each agent moves all the way to its current measurement of the average position of others. When drift is large compared to measurement noise, and the number of agents is large, this protocol can become highly efficient. However, when measurement noise is non-negligible compared to the drift, it is expected that incorporating past measurements could enhance the cohesion, even though drift may have changed the configuration considerably.

Perhaps the simplest class of algorithms that take such past information into account are *weighted-average* algorithms. By weighing the current position against the measured position in a non-trivial way, such algorithms can potentially exploit the fact that the current position implicitly encodes information from past measurements. Indeed, in a centralized setting, when a single agent aims to estimate a fixed target relying on noisy Gaussian measurements, a weighted-average algorithm is known to be optimal, in the sense that it minimizes the expected distance between the agent and the target [1]. However, here the setting is more complex since it is distributed, and the objective goal is to estimate (and get closer to) the average position, which is a function of the agents decisions.

1.2 The alignment problem

We consider n agents located on the real line $\mathbb{R}$. Let $I = \{1, \ldots, n\}$ be the set of agents. We denote by $\theta_i^{(t)} \in \mathbb{R}$ the position of Agent i at round t, where initially, the position of each agent is chosen independently according to a normal distribution around 0, with variance σ_0^2, that is, for each Agent i, $\theta_i^{(0)} \sim \mathcal{N}\left(0, \sigma_0^2\right)$. As mentioned, we assume that an agent is unaware of the actual value of its current position.

Execution proceeds in discrete rounds. At round t, each agent i receives a noisy measurement of the deviation from the current average position of all other agents. Specifically, denote the average of the positions of all agents except i by:

$$\langle \theta_{-i}^{(t)} \rangle = \tfrac{1}{n-1} \sum_{j \in I \setminus i} \theta_i^{(t)}.$$

Let $\overline{\theta}_i^{(t)} = \langle \theta_{-i}^{(t)} \rangle - \theta_i^{(t)}$ denote the *stretch* of Agent i. At any round t, for every $i \in I$, a noisy measurement of the stretch of Agent i is sampled:

$$\mathrm{Y}_i^{(t)} = \overline{\theta}_i^{(t)} + N_{m,i}^{(t)}, \tag{1}$$

where $N_{m,i}^{(t)} \sim \mathcal{N}(0, \sigma_m^2)$. In response, Agent i makes a move $d\theta_i^{(t)}$ and may update its memory state (if it has any). Finally, the position of Agent i at the next round is obtained by adding a drift:

$$\theta_i^{(t+1)} = \theta_i^{(t)} + d\theta_i^{(t)} + N_{d,i}^{(t)}, \tag{2}$$

where $N_{d,i}^{(t)} \sim \mathcal{N}(0, \sigma_d^2)$. All random perturbations $(N_{m,i}^{(t)})_{i \in I}$ and $(N_{d,i}^{(t)})_{i \in I}$ are mutually independent, and $\sigma_m, \sigma_d > 0$.

Definition 1 (*Cost*)**.** *The* cost *of Agent i at a given time t is the absolute value of its expected stretch at that time* [2]*, i.e.,*

$$\mathcal{C}_i^{(t)} := \mathbb{E}\left(\left|\overline{\theta}_i^{(t)}\right|\right).$$

Note that the cost depends on the algorithm used by i and on the algorithms used by others. As these algorithms will be clear from the context, we omit mentioning them in notations.

Definition 2 (*Optimality*)**.** *We say that an algorithm is* optimal *if, for every $i \in I$ and every round t, no algorithm can achieve a strictly smaller cost $\mathcal{C}_i^{(t)}$. Note that the definition refers to a very strong notion of optimality — it implies that if Algorithm A is optimal, then for any given round t and for any given Agent i, no algorithm B can achieve better performances for Agent i at round t than the performances guaranteed by A, regardless of what B did in previous rounds.*

Definition 3 (*Weighted-average algorithms*)**.** *A* weighted-average *algorithm is a distributed algorithm that is characterized by a responsiveness parameter $\rho^{(t)}$ for each round t, indicating the weight given to the measurement at that round. Formally, an agent i following the* weighted-average *algorithm $\mathcal{W}(\rho^{(t)})$ at round t, sets*

$$d\theta_i^{(t)} = \rho^{(t)} Y_i^{(t)}. \tag{3}$$

Full communication (or centralized) model. When executing a weighted-average algorithm, an agent bases its decisions solely on its own measurements. A main question we ask is whether, and if so to what extent, can performances be improved if agents could communicate with each other to share their measurements. In order to study the impact of communication, we compare the performances of the best weighted-average algorithm to the performances of the best algorithm in the *full-communication model*, where agents are free to share their measurements with all other agents at no cost. In the case that agents have identities, this setting is essentially equivalent to the following centralized setting: Consider a *master* agent that is external to the system. The master agent receives, at any round t, the stretch measurements of all agents, i.e., the collection $\{Y_j^{(t)}\}_{j\in I}$, where these measurements are noisy in the same manner as described in Eq. (1). Analyzing these measurements at round t, the master agent then instructs each agent i to move by a quantity $d\theta_i^{(t)}$. After moving, the agents are subject to drift, as described in Eq. (2). Note that the master agent is unable to "see" the positions of the agents, and its information regarding their locations is based only on the measurements it gathers from the agents, and on the movements it instructs. The goal of the master agent is to minimize the maximal cost of an agent, (where the maximum is over the agents), per round. In particular, an algorithm is said to be optimal in the centralized setting if it satisfies Definition 2, (where optimality is with respect to all algorithms in the centralized setting).

1.3 Our results

We prove that a distributed *weighted-average* algorithm, termed $\mathcal{W}^\star$, optimally minimizes the expected stretch of all agents at all rounds, in the strong sense of Definition 2. The optimality of this weighted-average algorithm holds even in the centralized (full-communication) model. We find this result surprising since it can be shown that the measurements made by Agent i contain strictly less information about the agent's relative position to the average position than the information contained in the collection of all measurements $\{Y_j^{(t)}\}_{j\in I}$. Indeed, although the $\{Y_j^{(t)}\}_{j\in I}$ measurements are not centered around the stretch $\overline{\theta}_i^{(t)}$ of Agent i, they still contain useful information about the stretch of Agent i. For example, it can be shown that

$$-\sum_{j\in I\setminus i} Y_j^{(t)} = \overline{\theta}_i^{(t)} - \sum_{j\in I\setminus i}^{n} N_{m,j}^{(t)}, \tag{4}$$

thus representing an additional "fresh" estimation of $\overline{\theta}_i^{(t)}$. Hence, the measurements obtained by Agent i are not sufficient statistics with respect to all measurements, when aiming to estimate the stretch of Agent i.

[2] Another natural cost measure is the deviation from the average position of all agents (including the agent), i.e., $\mathcal{C}_i^{(t)'} := \mathbb{E}(|\frac{1}{n}\sum_{j\in I}\theta_j^{(t)} - \theta_i^{(t)}|)$. These two measures are effectively equivalent. Indeed, $\mathcal{C}_i^{(t)'} = \frac{n-1}{n}\mathcal{C}_i^{(t)}$, thus an algorithm minimizing one measure will also minimize the other.

Therefore, at a first glance, it may appear that in the centralized setting the stretch of agents could potentially be reduced by letting the master agent process all measurements. Nevertheless, when operating under $\mathcal{W}^\star$, Agent i processes merely its own measurements, and still its stretch is minimized, as it could best be minimized by utilizing all measurements. The reason why this happens, is that other agents manage to fully process the information not processed by Agent i in a way that also benefits Agent i, by shifting the center of mass towards it.

Finally, we also analyze the drift of the center of mass and show that no distributed algorithm can achieve as small of a drift as can be achieved by the best centralized algorithm. In light of this, we also show that the drift associated with $\mathcal{W}^\star$ incurs a relatively small overhead over the best possible drift in the centralized setting.

Intuition: The case of Two Agents in One round. In order to obtain preliminary understanding, let us first examine the case of two agents operating in a single round, and compare the situation between the centralized and the distributed setting. This simple example is informative since it already demonstrates the aforementioned surprising phenomenon.

To really limit this example to "one round", we assume that the initial distribution of the stretches has a very large variance (formally, $\sigma_0^2 \gg \sigma_m^2$). Therefore, we can put aside the Bayesian conflict between the a priory knowledge about the stretches, and the information contained in the new measurement – and only work with the latter.

When considering two agents, the *stretch* $\overline{\theta}_i$ of Agent $i \in \{1, 2\}$ is simply its relative distance to the other agent. Assume that their positions are initially set to $\theta_1 = 0$ and $\theta_2 = 1$, so that $\overline{\theta}_1 = +1$ and $\overline{\theta}_2 = -1$. Assume further that the respective measurement noises are $N_{m,1} = 0.1$ and $N_{m,2} = 0.2$. In this case, the measurements of the stretches are:

$$\begin{cases} Y_1 = \overline{\theta}_1 + N_{m,1} = +1.1 \\ Y_2 = \overline{\theta}_2 + N_{m,2} = -0.8. \end{cases} \tag{5}$$

<u>*Centralized setting.*</u> Here, a central entity gathers both measurements Y_1 and Y_2, and, based on these measurements, instructs each agent of its move. Since we assumed that the initial distribution of the stretches has a very large variance compared to σ_m^2, the most accurate estimates of the relative stretches between the agents depends only on the measurements

$$\begin{cases} \hat{\theta}_1^{\text{cent}} = \frac{1}{2}(Y_1 - Y_2) = \frac{+1.9}{2} = +0.95 \\ \hat{\theta}_2^{\text{cent}} = \frac{1}{2}(Y_2 - Y_1) = \frac{-1.9}{2} = -0.95. \end{cases} \tag{6}$$

Recall that the goal of the central entity is to minimize the absolute stretch between the agents. Note that for this purpose, it has a degree of freedom, since, once it moves each of them, shifting both by the same amount would not change their stretch. However, if the central entity also wishes to minimize the drift of the center of mass, then it should instruct the agents to move towards the estimated center, shifting them by $\hat{\theta}_i/2$ towards each-other. (This type of algorithm will later be called "meet at the center".) Specifically, the instructed moves would be $d\theta_1^{\text{cent}} = \frac{1}{2}\,\hat{\theta}_1^{\text{cent}} = +0.475$ and $d\theta_2^{\text{cent}} = \frac{1}{2}\,\hat{\theta}_2^{\text{cent}} = -0.475$. By definition, the new positions of the agents (before drift applies) would be:

$$\begin{cases} \theta_1'^{\text{cent}} = \theta_1 + d\theta_1^{\text{cent}} = 0.475 \\ \theta_2'^{\text{cent}} = \theta_2 + d\theta_2^{\text{cent}} = 0.525 \end{cases} \quad \text{so} \begin{cases} \overline{\theta}_1'^{\text{cent}} = +0.05, \\ \overline{\theta}_2'^{\text{cent}} = -0.05. \end{cases} \tag{7}$$

As can be seen, the center of mass is unchanged. In fact, this is no coincidence and would be the case for any measurement noises (before drift takes place).

<u>*Distributed setting.*</u> In the *distributed* setting, an agent must base its estimate of the stretch on its measurement alone. In this case, we have:

$$\begin{cases} \hat{\theta}_1^{\text{dist}} = Y_1 = +1.1, \\ \hat{\theta}_2^{\text{dist}} = Y_2 = -0.8. \end{cases}$$

Of course, the quality of these estimates is less than in the centralized setting (compare with Eq. (6)). Once again, in attempting to "meet at the center", the best that the agents can do is to go half way to

their respective estimate of the location of the other agent. That is, $d\theta_1{}^{\text{dist}} = \frac{1}{2}\,\hat{\theta}_1{}^{\text{dist}} = \frac{1}{2}\cdot Y_1 = +0.55$ and $d\theta_2{}^{\text{dist}} = \frac{1}{2}\,\hat{\theta}_2{}^{\text{dist}} = \frac{1}{2}\cdot Y_2 = -0.4$. By definition, their new locations (before drift applies) would be:

$$\begin{cases}\theta_1'{}^{\text{dist}} = \theta_1 + d\theta_1{}^{\text{dist}} = 0.55\\ \theta_2'{}^{\text{dist}} = \theta_2 + d\theta_2{}^{\text{dist}} = 0.6\end{cases}, \text{ yielding } \begin{cases}\overline{\theta}_1'{}^{\text{dist}} = +0.05\\ \overline{\theta}_2'{}^{\text{dist}} = -0.05.\end{cases}$$

Evidently, in contrast to the centralized case (Eq. (7)), here the center of mass moved as a result of the agents movements. Importantly, however, the relative positions of the agents (and therefore, their costs) are identical to the ones obtained in the centralized protocol. Again, this is no coincidence, and holds for any measurement noises. The reason is, that the two locations of the agents in the distributed setting differ by the same shift, compared to their location in the centralized setting. Indeed, we have:

$$\begin{cases}\theta_1'{}^{\text{dist}} - \theta_1'{}^{\text{cent}} = d\theta_1{}^{\text{dist}} - d\theta_1{}^{\text{cent}} = \frac{1}{4}(Y_1+Y_2)\\ \theta_2'{}^{\text{dist}} - \theta_2'{}^{\text{cent}} = d\theta_2{}^{\text{dist}} - d\theta_2{}^{\text{cent}} = \frac{1}{4}(Y_1+Y_2).\end{cases}$$

This shift, i.e., $\frac{1}{4}(Y_1+Y_2)$, represents the "mistake" made by distributed agents in attempting to meet at the center, compared with the centralized algorithm. However, since both agents in the distributed setting are shifted by the same quantity compared to their respective positions in the centralized setting, their relative stretch remains the same in both settings. A main challenge of this paper is to show that this phenomenon holds also in the multi-round scenario, and with arbitrary number of agents.

Weighted-average algorithms with fixed responsiveness. As a warm-up towards understanding multi-round scenarios with multiple agents, we first investigate weighted-average algorithms in which all agents have the same responsiveness ρ, that furthermore remains fixed throughout the execution (see Eq. (3)). The proofs of both Theorems 1 and 2 below are given in Section 2.

Theorem 1. *Assume that all agents execute $\mathcal{W}(\rho)$, for a fixed $0 \le \rho \le 1$. Then for every $i \in I$ and every $t \in \mathbb{N}$, the stretch $\overline{\theta}_i^{(t)}$ is normally distributed, and*

$$\lim_{t\to+\infty} \operatorname{Var}\left(\overline{\theta}_i^{(t)}\right) = \frac{\frac{n}{n-1}(\rho^2\sigma_m^2+\sigma_d^2)}{1-(1-\frac{n}{n-1}\rho)^2},$$

with the convention that $\lim_{t\to+\infty}\operatorname{Var}(\overline{\theta}_i^{(t)}) = +\infty$ if the denominator $1-(1-\frac{n}{n-1}\rho)^2 = 0$.

If all agents run $\mathcal{W}(\rho)$, then Agent i's cost at steady state is captured by $\operatorname{Var}(\rho) := \lim_{t\to+\infty}\operatorname{Var}\left(\overline{\theta}_i^{(t)}\right)$. Indeed, for every i, since $\overline{\theta}_i^{(t)}$ is normally distributed, $\lim_{t\to+\infty}\mathbb{E}\left(\left|\overline{\theta}_i^{(t)}\right|\right) = \sqrt{\frac{2}{\pi}\operatorname{Var}(\rho)}$. The minimal value of this is achieved when taking $\operatorname{argmin}_\rho \operatorname{Var}(\rho)$ as the responsiveness parameter.

Theorem 2. *The weighted-average algorithm that optimizes group variance among all weighted-average algorithms $\mathcal{W}(\rho)$ (that use the same responsiveness parameter ρ at all rounds) is $\mathcal{W}(\rho^\star)$, where*

$$\rho^\star = \frac{\sigma_d\sqrt{4\sigma_m^2+\left(\frac{n}{n-1}\sigma_d\right)^2} - \frac{n}{n-1}\sigma_d^2}{2\sigma_m^2}. \tag{8}$$

When n is large, Eq. (8) becomes $\rho^\star \approx \frac{\sigma_d\sqrt{4\sigma_m^2+\sigma_d^2}-\sigma_d^2}{2\sigma_m^2}$. Note that the role played by the measurement noise is very different than the role played by the drift. For example, if $\sigma_m \gg \sigma_d$, then $\rho^\star \approx 0$. However, if $\sigma_m \ll \sigma_d$ then $\rho^\star \approx 1$. Interestingly, if $\sigma_m = \sigma_d$ then $\rho^\star \approx \frac{\sqrt{5}-1}{2}$, which is highly related to the golden ratio. Moreover, for large n, the minimal $\operatorname{Var}(\rho)$ is

$$\operatorname{Var}(\rho^\star) = \tfrac{1}{2}\sigma_d\left(\sqrt{4\sigma_m^2+\sigma_d^2}+\sigma_d\right). \tag{9}$$

Note that when the measurements are perfect, i.e., $\sigma_m = 0$, we have $\mathrm{Var}(\rho^\star) = \sigma_d^2$, which is the best achievable value that an agent can hope for, since no algorithm can overcome the drift-noise.

The impact of communication. Our next goal is to understand whether, and if so, to what extent, can performances be improved if further communication between agents is allowed. For this purpose, we compare the performances of $\mathcal{W}(\rho^\star)$ to the performances of the best algorithm in the centralized (full-communication) setting.

A natural candidate for an optimal algorithm in the centralized setting is the "meet at the center" algorithm. This algorithm first obtains, for each agent, the best possible estimate of the distance from the agent's position to the average position $\langle\theta\rangle$, based on all measurements, and then instructs the agent to move by this quantity (towards the estimated average position). However, it is not immediate to figure out the distances to the average position, and furthermore, quantify the performances of this algorithm. To this end, we make use of the Kalman filter tool [1], by adapting it to our setting. Solving the Kalman filter system associated with the centralized version of our alignment problem, we obtain an estimate of the relative distance of each agent i from the average position (based on all measurements). To describe these estimates we first define the following.

Definition 4 (Uncertainty)**.** *We inductively define the sequence $(\alpha_t)_{t=0}^{\infty}$. Let $\alpha_0 = n\sigma_0^2/(n-1)$, and for every t, let*

$$\alpha_{t+1} = \frac{\sigma_m^2 \alpha_t}{\frac{n}{n-1}\alpha_t + \sigma_m^2} + \frac{n}{n-1}\sigma_d^2. \tag{10}$$

The value of α_t aims to represent the variance of the (random) error made when estimating the stretch of the agents at round t. When thinking of n as very large, the formula for α_{t+1} is reminiscent of the resistance formula in electrical circuits, where we add in parallel to the current resistor with resistance α_t a resistor with resistance σ_m^2 (standing for the measurement noise), and another resistor in a series, with resistance σ_d^2 (standing for the drift).

Definition 5 (Optimal weight)**.** *For every integer t, let*

$$\rho_\star^{(t)} = \frac{\alpha_t}{\frac{n}{n-1}\alpha_t + \sigma_m^2}.$$

The value of $\rho_\star^{(t)}$ aims to indicate by how much to weigh the measurements of round t in order to move optimally: if $\sigma_m^2 \ll \alpha_t$, then $\rho_\star^{(t)} \approx 1$, meaning that the agent should entirely rely on the measurement. Conversely, if $\sigma_m^2 \gg \alpha_t$, then $\rho_\star^{(t)} \approx 0$, and the agent should not take the measurement into account.

At each round t, the Kalman filter returns an estimate of the relative distance of each agent i from the average position, which turns out to be

$$\rho_\star^{(t)}\Big(\tfrac{n-1}{n}Y_i^{(t)} - \tfrac{1}{n}\sum_{j\in I\setminus i} Y_j^{(t)}\Big).$$

As guaranteed by the properties of the Kalman filter, these estimates minimize the expected sum of square-errors, which can be translated to our desired measure of minimizing the agents' costs. The "Meet-at-the-center" algorithm is given by Algorithm 1 below, and the following theorem stating its optimality is proved in Section 3.5.

Theorem 3. *Meet-at-the-center is optimal in the centralized setting.*

Note that in the centralized setting, once we have an optimal algorithm A, we can derive another optimal algorithm B by simply shifting all agents, at each round t, by a fixed quantity λ_t. Indeed, such shifts do not influence the relative positions between the agents. Conversely, we show in Appendix D.1 that all (optimal) deterministic algorithms in the centralized setting are shifts of one another (note that the shifts λ_t are not necessarily the same for all rounds t).

As it happens, another solution that follows from the Kalman filter estimations can be described as a distributed weighted-average algorithm (see Algorithm 2). As such, this algorithm, henceforth called $\mathcal{W}^\star$, also yields optimal stretches. The proof of the following theorem is given in Section 3.6.

Algorithm 1: Meet-at-the-center

foreach *round* t **do**
  Consider all measurements at round t, $\{Y_j^{(t)} \mid 1 \le j \le n\}$;
  foreach *agent* i **do**
    Set $d\theta_i^{(t)} = \rho_\star^{(t)} \left(\frac{n-1}{n} Y_i^{(t)} - \frac{1}{n} \sum_{j \neq i} Y_j^{(t)} \right)$; `/* Output an estimate of` $\langle \theta \rangle - \theta_i$ `*/`
  end
end

Algorithm 2: Algorithm $\mathcal{W}^\star$

foreach *round* t **do**
  Consider all measurements at round t, $\{Y_j^{(t)} \mid 1 \le j \le n\}$;
  foreach *agent* i **do**
    Set $d\theta_i^{(t)} = \rho_\star^{(t)} Y_i^{(t)}$;
  end
end

Theorem 4. *Algorithm* $\mathcal{W}^\star = \mathcal{W}(\rho_\star^{(t)})$ *is optimal in the centralized setting.*

Since both Meet-at-the-center and $\mathcal{W}^\star$ are optimal deterministic algorithms they must be shifts of one another. Indeed, as validated in [19] (Appendix D.2), by adding the shift $\lambda_t = \frac{1}{n} \rho_\star^{(t)} \sum_{i \in I} Y_i^{(t)}$ to the agents in Meet-at-the-center we obtain Algorithm $\mathcal{W}^\star$.

Note that in contrast to $\mathcal{W}(\rho^\star)$, Algorithm $\mathcal{W}^\star$ uses a different responsiveness $\rho_\star^{(t)}$ at each round t. We next argue that the sequences (α_t) and $(\rho_\star^{(t)})$ converge. Not surprisingly, at the limit, we recover the optimal responsiveness as stated in Theorem 2. Due to lack of space the proof of the following claim is deferred to Appendix B.2 of [19].

Claim 1. *The sequence* α_t *converges to*

$$\alpha_\infty := \frac{1}{2} \left(\sigma_d \sqrt{4\sigma_m^2 + \left(\frac{n}{n-1} \sigma_d \right)^2} + \frac{n}{n-1} \sigma_d^2 \right).$$

Moreover, $\lim_{t \to +\infty} \rho_\star^{(t)} = \rho_\star$.

Drift of the center of mass. In this section we analyze the drift of the center of mass. Essentially, we show that with respect to this drift (and in contrast to the measure of stretch as used in Definition 1), there is a gap between the centralized and the distributed settings. Specifically, we first show that Meet-at-the-center is "drift-optimal", in sense that it achieves the best possible drift among all algorithms in the centralized setting. In contrast, we prove that there is no distributed algorithm that can achieve as small of a drift as the one obtained by the "meet at the center" algorithm. Consistent with this result, Algorithm $\mathcal{W}^\star$ is not "drift-optimal", yet we show that the overhead it incurs in terms of this drift is relatively small, especially when n is large compared to t or when σ_m is small compared to σ_d.

Let us now describe these results in more detail. Denote $\Delta_t = \langle \theta^{(t)} \rangle - \langle \theta^{(0)} \rangle$. By definition (Eq. (2)), for a given algorithm, the center of mass changes in round t as follows:

$$\langle \theta^{(t+1)} \rangle - \langle \theta^{(t)} \rangle = \frac{1}{n} \sum_{i \in I} \left(d\theta_i^{(t)} + N_{d,i}^{(t)} \right). \tag{11}$$

The drift of the center of mass after t rounds is therefore

$$\Delta_t = \frac{1}{n} \sum_{s=0}^{t} \sum_{i \in I} \left(d\theta_i^{(s)} + N_{d,i}^{(s)} \right).$$

Let us define *inherent-drift* as $\Delta_t^{\text{inh}} := \frac{1}{n}\sum_{s=0}^{t}\sum_{i\in I} N_{d,i}^{(s)}$ and *relative-drift* as $\Delta_t^{\text{rel}} := \frac{1}{n}\sum_{s=0}^{t}\sum_{i\in I} d\theta_i^{(s)}$. With these notations, we have:

$$\Delta_t = \Delta_t^{\text{rel}} + \Delta_t^{\text{inh}}. \tag{12}$$

Due to lack of space, the proof of the following lemma appears in Appendix E of [19].

Lemma 1. *For every protocol and for every round t, Δ_t^{inh} and Δ_t^{rel} are independent.*

A direct consequence of Lemma 1 is that

$$\mathbb{E}\left(\Delta_t^2\right) = \mathbb{E}\left({\Delta_t^{\text{rel}}}^2\right) + \text{Var}\left(\Delta_t^{\text{inh}}\right) \geq \text{Var}(\Delta_t^{\text{inh}}), \tag{13}$$

justifying the choice of the term "inherent". Since $\text{Var}(\Delta_t^{\text{inh}}) = \text{Var}\left(\frac{1}{n}\sum_{s=0}^{t}\sum_{i\in I} N_{d,i}^{(s)}\right) = \frac{t+1}{n}\sigma_d^2$, we obtain the following.

Theorem 5. *For every protocol in the centralized setting and for every round t, the drift of the center of mass satisfies $\mathbb{E}\left(\Delta_t^2\right) \geq \frac{t+1}{n}\sigma_d^2$.*

Definition 6. *An algorithm is called* drift-optimal*, if for every round t, the drift of the center of mass satisfies $\mathbb{E}\left(\Delta_t^2\right) = \text{Var}(\Delta_t^{\text{inh}}) = \frac{t+1}{n}\sigma_d^2$.*

Theorem 6. *Meet-at-the-center is drift-optimal.*

Proof. When the agents follow Algorithm Meet-at-the-center, for every $t \in \mathbb{N}$

$$\frac{1}{n}\sum_{i\in I} d\theta_i^{(t)} = \frac{1}{n}\sum_{i\in I}\rho_\star^{(t)}\left(\frac{n-1}{n}Y_i^{(t)} - \frac{1}{n}\sum_{j\in I\setminus i} Y_j^{(t)}\right) = \frac{\rho_\star^{(t)}}{n}\left(\frac{n-1}{n}\sum_{i\in I} Y_i^{(t)} - \frac{1}{n}\sum_{i\in I}(n-1)\cdot Y_i^{(t)}\right) = 0.$$

Hence, by definition, $\Delta_t^{\text{rel}} = 0$ and by Eq. (13), the drift of the center of mass in round t is limited to Δ_t^{inh}, which concludes the proof of Theorem 6. □

Next, we analyze the drift of $\mathcal{W}^\star$.

Theorem 7. *When Algorithm $\mathcal{W}^\star$ is used, the drift of the center of mass satisfies $\mathbb{E}\left(\Delta_t^2\right) = \frac{t+1}{n}\sigma_d^2 + \frac{1}{n}\sum_{s=0}^{t}{\rho_\star^{(s)}}^2\sigma_m^2$.*

Proof. Under Algorithm $\mathcal{W}^\star$, we have for every $t \in \mathbb{N}$

$$\frac{1}{n}\sum_{i\in I} d\theta_i^{(t)} = \frac{1}{n}\sum_{i\in I}\rho_\star^{(t)}\left(\overline{\theta}_i^{(t)} + N_{m,i}^{(t)}\right) = \frac{1}{n}\sum_{i\in I}\rho_\star^{(t)} N_{m,i}^{(t)},$$

where the last equality is by Lemma 2. Using Eq. (13), and summing over the rounds $s = 0, \ldots, t$, we conclude the proof of Theorem 7. □

Note that since $\rho_\star^{(s)} \leq 1$, the overhead incurred by Algorithm $\mathcal{W}^\star$ is at most $\frac{t+1}{n}\sigma_m^2$.

Finally, we argue that the sub-optimality of Algorithm $\mathcal{W}^\star$ (in terms of the drift of the center of mass) is not coincidental. The proof of the following claim is deferred to Appendix D.2.

Claim 2. *No algorithm in the distributed setting can be simultaneously optimal (Definition 2) and drift-optimal (Definition 6).*

1.4 Related works

The Kalman filter algorithm is a prolific tool commonly used in control theory and statistics, with numerous technological applications [1]. In this paper, we used the Kalman filter to investigate the centralized setting, where all relative measurements are gathered at one master agent that processes them and instructs agents how to move. Fusing relative measurements of multiple agents is often referred to in the literature as *distributed Kalman filter*, see surveys in [6, 22, 23]. However, there, the typical setting is that each agent invokes a separate Kalman filter to process its measurements, and then the resulted outputs are integrated using communication. Moreover, works in that domain often assume that observations are attained by static observers (i.e., the agents are residing on a fixed underlying graph) and that the measured target is external to the system. In contrast, here we consider a self-organizing system, with mobile agents that measure a target (average position) that is a function of their positions. Closer works to our setting where a group of mobile agents aim to gather together are [16, 30]. However, among the several differences with our paper, these papers assume that there is no drift and that agents communicate their absolute position, rather than their relative position.

Similarly to our findings, the authors of [18] also emphasize the effectiveness of performing a weighted-average between the opinion of the individual and the sample it receives. However, we find this result less surprising under their model, which features one-to-one communications and no drift.

Another related self-organization problem is *clock synchronization*, where the goal is to maintain a common notion of time in the absence of a global time source. The main difficulty is handling the fact that clocks count time at slightly different rates and that messages arrive with some delays. Variants of this problem were studied in wireless network contexts, mostly by control theoreticians [25, 27, 28]. A common technique uses oscillating models [15, 21].

2 Warm-up proofs

The goal of this section is to prove Theorems 1 and 2. These theorems assume all agents execute a weighted-average algorithm with the same responsiveness parameter ρ, that furthermore remains fixed throughout the execution (see Eq. (3)). The proofs of all intermediate results can be found in Appendix B. We start with the following observation.

Lemma 2. *For any integer t, and whatever the positions of the agents, we have $\sum_{i\in I} \overline{\theta}_i^{(t)} = 0$.*

Next, we compute how the stretch $\overline{\theta}_i^{(t)}$ of Agent i changes when all agents perform weighted-average moves with constant responsiveness parameter ρ.

Lemma 3. *Assume that all agents execute $\mathcal{W}(\rho)$, for some $0 \le \rho \le 1$. Let $\mathcal{E}_j^{(t)} = \rho N_{m,j}^{(t)} + N_{d,j}^{(t)}$. Then for every $i \in I$ and every $t \in \mathbb{N}$,*

$$\overline{\theta}_i^{(t+1)} = \left(1 - \tfrac{n}{n-1}\rho\right)\overline{\theta}_i^{(t)} + \tfrac{1}{n-1}\sum_{j\in I\setminus i} \mathcal{E}_j^{(t)} - \mathcal{E}_i^{(t)}\,.$$

Now we can prove that the stretch of each agent is normally distributed at every round, and compute its variance.

Lemma 4. *Assume that all agents execute $\mathcal{W}(\rho)$, for some $0 \le \rho \le 1$. Then, for every $i \in I$ and every $t \in \mathbb{N}$, the stretch $\overline{\theta}_i^{(t)}$ is normally distributed, $\mathbb{E}\left(\overline{\theta}_i^{(t)}\right) = 0$, and*

$$\mathrm{Var}\left(\overline{\theta}_i^{(t+1)}\right) = (1 - \tfrac{n\rho}{n-1})^2 \cdot \mathrm{Var}\left(\overline{\theta}_i^{(t)}\right) + \frac{n(\rho^2\sigma_m^2 + \sigma_d^2)}{n-1}.$$

Before proving Theorem 1, we need a small technical result.

Claim 3. *Let $a, b \ge 0$. Consider the sequence $\{u_n\}_{n=0}^{\infty}$ defined by letting $u_0 \in \mathbb{R}$ and for every integer n, $u_{n+1} = au_n + b$. If $a < 1$, then $\{u_n\}_{n=0}^{\infty}$ converges and $\lim_{n\to+\infty} u_n = b/(1-a)$. If $a = 1$, and $b > 0$, then $\lim_{n\to+\infty} u_n = +\infty$.*

Proof. (of Theorem 1.) First, apply Lemma 4. Hence, by Claim 3, $\mathrm{Var}\left(\overline{\theta}_i^{(t)}\right)$ converges to the limit as stated. Note that the variance is infinite if either (1) the responsiveness is equal to 0, in which case the drift adds up endlessly, or (2) the responsiveness is equal to 1 and $n = 2$, in which case the agents "swap" at each round, producing the same result. This completes the proof of Theorem 1. □

Proof. (of Theorem 2.) Consider the function

$$\mathrm{Var}(\rho) = \frac{\frac{n}{n-1}(\rho^2\sigma_m^2 + \sigma_d^2)}{1 - (1 - \frac{n}{n-1}\rho)^2} = \frac{\rho^2\sigma_m^2 + \sigma_d^2}{2\rho - \frac{n}{n-1}\rho^2}.$$

Note that this function evaluates to $+\infty$ when $\rho = 0$, or when $\rho = 1$ and $n = 2$. Computing its derivative yields:

$$\mathrm{Var}'(\rho) = \frac{2\rho\sigma_m^2\left(2\rho - \frac{n}{n-1}\rho^2\right) - (\rho^2\sigma_m^2 + \sigma_d^2)\left(2 - 2\frac{n}{n-1}\rho\right)}{\left(2\rho - \frac{n}{n-1}\rho^2\right)^2}.$$

Hence, $\mathrm{Var}'(\rho) = 0$ if and only if:

$$\rho^2\sigma_m^2\left(2 - \tfrac{n}{n-1}\rho\right) - (\rho^2\sigma_m^2 + \sigma_d^2)\left(1 - \tfrac{n}{n-1}\rho\right) = 0 \iff \rho^2\sigma_m^2 - \sigma_d^2\left(1 - \tfrac{n}{n-1}\rho\right) = 0.$$

This equation has a unique solution in the interval $[0, 1]$, which is $\rho = \frac{\sigma_d\sqrt{4\sigma_m^2 + \left(\frac{n}{n-1}\sigma_d\right)^2} - \frac{n}{n-1}\sigma_d^2}{2\sigma_m^2}$. Checking that this is a minimum concludes the proof. □

3 Solving the alignment problem

We denote $\mathcal{N}(\mu, \Sigma)$ the multivariate normal distribution with mean vector $\mu \in \mathbb{R}^n$ and co-variance matrix $\Sigma \in \mathbb{R}^{n\times n}$, and by I the identity matrix. Letting $\mathbb{1}$ be the matrix whose all coefficients are equal to 1, we denote

$$\mathcal{M}(a, b) = b\mathbb{1} + (a - b)I,$$

the matrix having diagonal coefficients equal to a, and other coefficients equal to b.

3.1 Rephrasing the alignment problem as a linear filtering problem

The goal of this section is to write Eqs. (1) and (2) in a matrical form. Let $\overline{\theta}^{(t)}$, $d\theta^{(t)}$, $Y^{(t)}$, $N_m^{(t)}$, and $N_d^{(t)}$ denote the vectors (indexed by the set of agents) corresponding to each variable. (E.g., $\overline{\theta}^{(t)} = (\overline{\theta}_1^{(t)}, \ldots, \overline{\theta}_n^{(t)})$, and similarly for the other vectors.)

Measurement rule Recall that, as stated in Eq. (1), the measurement of Agent i at time t is given by $Y_i^{(t)} = \overline{\theta}_i^{(t)} + N_{m,i}^{(t)}$. We simply rewrite this equation as:

$$Y^{(t)} = \overline{\theta}^{(t)} + N_m^{(t)}, \tag{14}$$

where, by definition, $N_m^{(t)} \sim \mathcal{N}\left(0, \sigma_m^2 I\right)$.

Update rule We recall the update equation of the stretch, which follows from Eq. (2) (see also Eq. (29)):

$$\begin{aligned}\overline{\theta}_i^{(t+1)} &= \overline{\theta}_i^{(t)} - d\theta_i^{(t)} - N_{d,i}^{(t)} + \left(\langle\theta_{-i}^{(t+1)}\rangle - \langle\theta_{-i}^{(t)}\rangle\right)\\ &= \overline{\theta}_i^{(t)} - d\theta_i^{(t)} - N_{d,i}^{(t)} + \tfrac{1}{n-1}\sum_{j\in I\setminus i}\left(d\theta_j^{(t)} + N_{d,j}^{(t)}\right).\end{aligned} \tag{15}$$

We define the matrix

$$M_n = \mathcal{M}\left(-1, \tfrac{1}{n-1}\right).$$

Let $\tilde{N}_d^{(t)} = M_n N_d^{(t)}$. With these definitions and Eq. (15), we write Eq. (2) in the following vector notation:

$$\overline{\theta}^{(t+1)} = \overline{\theta}^{(t)} + M_n d\theta^{(t)} + \tilde{N}_d^{(t)}. \tag{16}$$

By definition, $N_d^{(t)} \sim \mathcal{N}\left(0, \sigma_d^2 I\right)$, so by Claim 4, $\tilde{N}_d^{(t)} \sim \mathcal{N}\left(0, Q\right)$ where $Q = \sigma_d^2 \cdot M_n I M_n^\top = \sigma_d^2 M_n^2$.

3.2 Applying the Kalman filter

We use the Kalman filter as a tool for analysis purposes of the process. We stress, however, that the Kalman filter is not actually applied by any agent.

We denote by $\hat{\theta}_t$ the estimate of $\theta^{(t)}$ produced by the Kalman filter *after* the measurements at round t were obtained (Eq. (14)) and before the update of round t occurs (Eq. (16)). Let P_t denote the error co-variance matrix associated with this estimate. Specifically, $P_t = \mathbb{E}\left((\theta^{(t)} - \hat{\theta}_t)(\theta^{(t)} - \hat{\theta}_t)^\top\right)$. We add the superscript "-" to these notations (for example P_t^-), to denote the same quantities *before* the measurement obtained at round t.

So, in a sense, round t can be divided into the following four consecutive time steps: (a) the filter produces an estimation $\hat{\theta}_t^-$ of $\theta^{(t)}$, (b) a new measurement vector $Y^{(t)}$ of $\theta^{(t)}$ is obtained, (c) the filter produces an estimation $\hat{\theta}_t$ of $\theta^{(t)}$ given the new measurement, and (d) $\theta^{(t)}$ is updated to $\theta^{(t+1)}$. Recall that at round 0, i.e., at the initialization stage, the agents are normally distributed around 0. For technical reasons, we define the Kalman filter estimate at round 0 to be zero, that is, $\hat{\theta}_0^- = 0$.

Measurement update In order to produce (c), the filter incorporates the measurement in (b) to the estimation in (a). Specifically, the filter computes the "Kalman gain":

$$K_t = P_t^- \left(P_t^- + \sigma_m^2 I\right)^{-1}. \tag{17}$$

Then, it produces the following estimate, as required in (c):

$$\hat{\theta}_t = \hat{\theta}_t^- + K_t \left(Y^{(t)} - \hat{\theta}_t^-\right). \tag{18}$$

The new error co-variance matrix can then be written as:

$$P_t = (I - K_t)\, P_t^-. \tag{19}$$

Time update The estimation for (a) for the following round is then given by:

$$\hat{\theta}_{t+1}^- = \hat{\theta}_t + M_n d\theta^{(t)}. \tag{20}$$

Finally, the new error co-variance matrix that is used in the Kalman gain corresponding to round $t+1$ (Eq. (17)) is:

$$P_{t+1}^- = P_t + \sigma_d^2 \cdot M_n^2. \tag{21}$$

3.3 Optimality

Definition 7. *A* history $\mathcal{H}_t$ *is a realization of all measurements and all moves up to time* $(t-1)$.

The following is a well-known result, see e.g., [1].

Theorem 8. *For every round t, every history $\mathcal{H}_t$ and every vector $x \in \mathbb{R}^n$,*

$$\sum_{i \in I} \mathbb{E}\left((\theta_i^{(t)} - \hat{\theta}_{t,i}^-)^2 \mid \mathcal{H}_t\right) \leq \sum_{i \in I} \mathbb{E}\left((\theta_i^{(t)} - x_i)^2 \mid \mathcal{H}_t\right).$$

In particular, for every round t, every history $\mathcal{H}_t$, every $i \leq n$ and every scalar $x \in \mathbb{R}$,

$$\mathbb{E}\left((\theta_i^{(t)} - \hat{\theta}_{t,i}^-)^2 \mid \mathcal{H}_t\right) \leq \mathbb{E}\left((\theta_i^{(t)} - x)^2 \mid \mathcal{H}_t\right).$$

Definition 8. *We say that an algorithm for the alignment problem is* Kalman-perfect *if it always produces a sequence of moves* $(d\theta^{(t)})_{t\geq 0}$*, such that for every integer* $t \geq 1$*,* $\hat{\theta}_t^- = 0$*.*

The proof of the following result is provided in Appendix A.

Proposition 1. *If there exists a Kalman-perfect algorithm for the alignment problem, then this algorithm is optimal in the centralized setting (in the sense of Definition 2). Moreover, any other optimal (deterministic) algorithm is Kalman-perfect.*

3.4 Kalman gain and the error co-variance matrix

Recall the sequences $(\alpha_t)_{t\geq 0}$ and $(\rho_\star^{(t)})_{t\geq 0}$ introduced in Definitions 4 and 5. This section is dedicated to proving the following lemma, which provides a formula for the error co-variance matrix associated with the estimate $\hat{\theta}_t^-$, as well as for the Kalman gain.

Lemma 5. *For every integer* t*,* $P_t^- = \mathcal{M}\left(\alpha_t, \frac{-\alpha_t}{n-1}\right) = -\alpha_t M_n$*, and* $K_t = \frac{-\frac{n-1}{n}\alpha_t M_n}{\frac{n}{n-1}\sigma_m^2 + \alpha_t} = -\rho_\star^{(t)} M_n$*.*

Proof. We prove the first part of the claim by induction, and prove that for every round t, the second part of the claim (regarding K_t) follows from the first part (regarding P_t^-).

By Claim 4 (see Appendix C), $P_0^- = \sigma_0^2 M_n^2$. By Claim 5 (see Appendix C),

$$M_n^2 = \mathcal{M}\left(\tfrac{n}{n-1}, -\tfrac{n}{(n-1)^2}\right) = \tfrac{-n}{n-1} M_n.$$

Therefore, $P_0^- = -\frac{n}{n-1}\sigma_0^2 M_n$, and so the first part of the claim holds at round 0 since $\alpha_0 = \frac{n}{n-1}\sigma_0^2$.

Now, let us assume that the first part of the claim holds for some $t \in \mathbb{N}$. It follows that $P_t^- + \sigma_m^2 I = \mathcal{M}\left(\alpha_t + \sigma_m^2, \frac{-\alpha_t}{n-1}\right)$. By Claim 6 (see Appendix C), since $\sigma_m > 0$, we have

$$\left(P_t^- + \sigma_m^2 I\right)^{-1} = \frac{\mathcal{M}\left(\alpha_t + \sigma_m^2 - (n-2)\cdot\frac{\alpha_t}{n-1}, \frac{\alpha_t}{n-1}\right)}{\left(\alpha_t + \sigma_m^2 + \frac{\alpha_t}{n-1}\right)\left(\alpha_t + \sigma_m^2 - (n-1)\cdot\frac{\alpha_t}{n-1}\right)} = \frac{\mathcal{M}\left(\frac{\alpha_t}{n-1} + \sigma_m^2, \frac{\alpha_t}{n-1}\right)}{\sigma_m^2\left(\frac{n}{n-1}\alpha_t + \sigma_m^2\right)}.$$

By Claim 5 again, we can compute the "Kalman gain":

$$\begin{aligned}
K_t &= P_t^-\left(P_t^- + \sigma_m^2 I\right)^{-1} \\
&= \frac{1}{\sigma_m^2\left(\frac{n}{n-1}\alpha_t + \sigma_m^2\right)} \cdot \mathcal{M}\left(\frac{\alpha_t^2}{n-1} + \sigma_m^2\alpha_t - \frac{(n-1)\cdot\alpha_t^2}{(n-1)^2}, \frac{-\alpha_t^2}{(n-1)^2} - \frac{\sigma_m^2\alpha_t}{n-1} + \frac{\alpha_t^2}{n-1} - (n-2)\cdot\frac{\alpha_t^2}{(n-1)^2}\right) \\
&= \frac{\mathcal{M}\left(\sigma_m^2\alpha_t, \frac{-\sigma_m^2\alpha_t}{n-1}\right)}{\sigma_m^2\left(\frac{n}{n-1}\alpha_t + \sigma_m^2\right)} = \frac{-\alpha_t M_n}{\frac{n}{n-1}\alpha_t + \sigma_m^2},
\end{aligned}$$

This proves that the second part of the claim holds at round t. Next, we compute the error co-variance matrix after the measurement:

$$\begin{aligned}
P_t &= (I - K_t)\cdot P_t^- = \frac{\mathcal{M}\left(\frac{\alpha_t}{n-1} + \sigma_m^2, \frac{\alpha_t}{n-1}\right)}{\frac{n}{n-1}\alpha_t + \sigma_m^2} \cdot P_t^- \\
&= \frac{1}{\frac{n}{n-1}\alpha_t + \sigma_m^2} \cdot \mathcal{M}\left(\frac{\alpha_t^2}{n-1} + \sigma_m^2\alpha_t - (n-1)\cdot\frac{\alpha_t^2}{(n-1)^2}, -\frac{\alpha_t^2}{(n-1)^2} - \frac{\sigma_m^2\alpha_t}{n-1} + \frac{\alpha_t^2}{n-1} - (n-2)\cdot\frac{\alpha_t^2}{(n-1)^2}\right) \\
&= \frac{\mathcal{M}\left(\sigma_m^2\alpha_t, \frac{-\sigma_m^2\alpha_t}{n-1}\right)}{\frac{n}{n-1}\alpha_t + \sigma_m^2} = -\frac{\frac{n-1}{n}\sigma_m^2\alpha_t}{\alpha_t + \frac{n-1}{n}\sigma_m^2} M_n.
\end{aligned}$$

Eventually, we compute the error co-variance matrix at round $t+1$, before the measurement:

$$P_{t+1}^- = P_t + \sigma_d^2 M_n^2 = P_t - \tfrac{n}{n-1}\sigma_d^2 M_n.$$

Plugging in the expression of P_t, we get

$$P^-_{t+1} = -\frac{\frac{n-1}{n}\sigma_m^2\alpha_t}{\alpha_t + \frac{n-1}{n}\sigma_m^2} M_n - \frac{n}{n-1}\sigma_d^2 M_n = -\left(\frac{\frac{n-1}{n}\sigma_m^2\alpha_t}{\frac{n-1}{n}\sigma_m^2 + \alpha_t} + \frac{n}{n-1}\sigma_d^2\right) M_n = -\alpha_{t+1} M_n,$$

which concludes the induction proof. □

3.5 Meet-at-the-center

The goal of this section is to prove Theorem 3 stating that Algorithm Meet-at-the-center is optimal in the centralized setting (in the sense of Definition 2). The choice of the name comes from the fact that this algorithm minimizes the drift of the center of mass (see Theorem 6).

Proof. (of Theorem 3.) By Proposition 1, it is sufficient to prove that Algorithm Meet-at-the-center is Kalman-perfect, that is, that for every integer $t \geq 1$, the Kalman filter associated with the moves it produces gives the estimate $\hat{\theta}^-_t = 0$.

For this purpose, assume that all agents run Algorithm Meet-at-the-center. We prove by induction that for every integer $t \geq 0$, the Kalman filter produces the estimate $\hat{\theta}^-_t = 0$. This holds at $t = 0$ since we assumed that the Kalman filter estimates 0 at round 0, i.e., $\hat{\theta}^-_0 = 0$. Next, let us assume that $\hat{\theta}^-_t = 0$ for some integer $t \geq 0$. We have by definition,

$$\hat{\theta}^-_{t+1} = \hat{\theta}_t + M_n d\theta^{(t)}, \tag{22}$$

and

$$\hat{\theta}_t = \hat{\theta}^-_t + K_t\left(Y^{(t)} - \hat{\theta}^-_t\right) = K_t Y^{(t)}, \tag{23}$$

where the second equality in Eq. (23) is by the induction hypothesis. By Lemma 5, $K_t = -\rho_\star^{(t)} M_n$. Note that, by definition of Algorithm Meet-at-the-center, $d\theta^{(t)} = -\frac{n-1}{n}\rho_\star^{(t)} M_n Y^{(t)} = \frac{n-1}{n}\hat{\theta}_t$. Finally, by the equations above, we rewrite Eq. (22) as:

$$\hat{\theta}^-_{t+1} = \left(I + \tfrac{n-1}{n}M_n\right)\hat{\theta}_t = \left(I + \tfrac{n-1}{n}M_n\right) K_t Y^{(t)} = -\left(M_n + \tfrac{n-1}{n}M_n^2\right)\rho_\star^{(t)} Y^{(t)}.$$

Since $M_n^2 = \frac{-n}{n-1}M_n$, we have $\hat{\theta}^-_{t+1} = 0$. This concludes the induction, and completes the proof of the theorem. □

3.6 Algorithm $\mathcal{W}^\star$ is optimal in the centralized setting

The goal of this section is to prove Theorem 4. The proof follows the same line of arguments as the proof of Theorem 3.

Proof. (of Theorem 4) Our goal is to prove that Algorithm $\mathcal{W}^\star$ is Kalman-perfect, that is, for every integer $t \geq 1$, the Kalman filter associated with the moves produced by Algorithm $\mathcal{W}^\star$ gives the estimate $\hat{\theta}^-_t = 0$. This would conclude the proof of the theorem, by Proposition 1.

For this purpose, assume that all agents run Algorithm $\mathcal{W}^\star$. We prove by induction that for every integer $t \geq 0$, the Kalman filter produces the estimate $\hat{\theta}^-_t = 0$.

The base case, where $t = 0$, holds since we assumed that the Kalman filter estimates zero at round zero, i.e., $\hat{\theta}^-_0 = 0$. Next, let us assume that $\hat{\theta}^-_t = 0$ for some integer $t \geq 0$, and consider $t+1$. We have by definition,

$$\hat{\theta}^-_{t+1} = \hat{\theta}_t + M_n d\theta^{(t)}, \tag{24}$$

and

$$\hat{\theta}_t = \hat{\theta}^-_t + K_t\left(Y^{(t)} - \hat{\theta}^-_t\right) = K_t Y^{(t)}, \tag{25}$$

where the second equality in Eq. (25) is by induction hypothesis. By Lemma 5, $K_t = -\rho_\star^{(t)} M_n$. Hence, Eq. (24) rewrites

$$\hat{\theta}^-_{t+1} = M_n\left(-\rho_\star^{(t)} Y^{(t)} + d\theta^{(t)}\right).$$

Finally, by the definition of $\mathcal{W}^\star$, we have $d\theta^{(t)} = \rho_\star^{(t)} Y^{(t)}$, so $\hat{\theta}^-_{t+1} = 0$, concluding the induction step. □

4 Discussion and Future Work

We introduced the distributed *alignment* problem, aiming to study basic algorithmic aspects that govern approximate agreement processes in unreliable stochastic environments, while relying on measurements of the overall group tendency. We proposed a weighted-average algorithm, termed $\mathcal{W}^\star$, and proved that it achieves optimal stretches, that is, it minimizes the expected distance of each agent from the center of mass of others, at each round. In fact, the optimality of $\mathcal{W}^\star$ holds also in the full-communication (or centralized) model, where communication of internal states is unrestricted. Algorithm $\mathcal{W}^\star$ is based on having a different responsiveness parameter that is applied by each agent at each round. We also find the optimal weighted-average algorithm among the ones that use the same responsiveness at all times. In contrast to $\mathcal{W}^\star$ that minimizes the expected stretch at every round separately, this algorithm becomes optimal only at steady-state.

In our setting, the cost of an individual is defined as its expected distance from the average position of other agents. Minimizing this quantity is equivalent to minimizing the expected distance from the average position of all agents (see Footnote 2). Another interesting measure is the expected *diameter* of the group, defined as the maximal distance between any two agents, at steady state. It would not come as a surprise if Algorithm $\mathcal{W}^\star$ would turn out to be optimal also with respect to this measure, however, analyzing its expected diameter would require handling further dependencies between agents, and therefore remains for future work.

It would also be interesting to see how our work could be extended to scenarios in which each agent can observe a subset of randomly chosen agents at each round. As mentioned in the introduction, averaging the values of the agents in such a subset can serve as a noisy sample of average value in the population. However, in contrast to our model, the noise in this sample is not fixed, and depends on the distribution of values in the population and on the size of the sampled subset. In many cases, one can expect that sampling a subset containing a logarithmic number of agents could suffice to obtain a relatively good sample of the population. We find this encouraging towards generalizing our work to such scenarios.

Acknowledgment The authors thank Yongcan Cao for a helpful discussion regarding related works on distributed Kalman filter, as well as anonymous reviewers whose critical comments helped improve the presentation of the paper significantly.

References

[1] Brian DO Anderson and John B Moore. *Optimal filtering*. Courier Corporation, 2012.

[2] Luca Becchetti, Andrea Clementi, and Emanuele Natale. Consensus dynamics: An overview. *ACM SIGACT News*, 51(1):58–104, 2020.

[3] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, Riccardo Silvestri, and Luca Trevisan. Simple dynamics for plurality consensus. *Distributed Computing*, 30(4):293–306, 2017.

[4] Spring Berman, Quentin Lindsey, Mahmut Selman Sakar, Vijay Kumar, and Stephen Pratt. Study of group food retrieval by ants as a model for multi-robot collective transport strategies. *Robotics: Sci. Syst. VI*, page 259, 2011.

[5] Lucas Boczkowski, Amos Korman, and Emanuele Natale. Minimizing message size in stochastic communication patterns: fast self-stabilizing protocols with 3 bits. *Distributed Comput.*, 32(3):173–191, 2019.

[6] Yongcan Cao, Wenwu Yu, Wei Ren, and Guanrong Chen. An overview of recent progress in the study of distributed multi-agent coordination. *IEEE Transactions on Industrial informatics*, 9(1):427–438, 2012.

[7] Larissa Conradt and Timothy J Roper. Consensus decision making in animals. *Trends in ecology & evolution*, 20(8):449–456, 2005.

[8] Benjamin Doerr, Leslie Ann Goldberg, Lorenz Minder, Thomas Sauerwald, and Christian Scheideler. Stabilizing consensus with the power of two choices. In *Proceedings of the twenty-third annual ACM symposium on Parallelism in algorithms and architectures*, pages 149–158, 2011.

[9] Ofer Feinerman, Bernhard Haeupler, and Amos Korman. Breathe before speaking: efficient information dissemination despite noisy, limited and anonymous communication. *Distributed Computing*, 30(5):339–355, 2017.

[10] Aviram Gelblum, Itai Pinkoviezky, Ehud Fonio, Abhijit Ghosh, Nir Gov, and Ofer Feinerman. Ant groups optimally amplify the effect of transiently informed individuals. *Nature communications*, 6:7729, 2015.

[11] Aviram Gelblum, Itai Pinkoviezky, Ehud Fonio, Nir S Gov, and Ofer Feinerman. Emergent oscillations assist obstacle negotiation during ant cooperative transport. *Proceedings of the National Academy of Sciences*, 113(51):14615–14620, 2016.

[12] Mohsen Ghaffari and Johannes Lengler. Nearly-tight analysis for 2-choice and 3-majority consensus dynamics. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing*, pages 305–313, 2018.

[13] Dan Gorbonos, Reuven Ianconescu, James G Puckett, Rui Ni, Nicholas T Ouellette, and Nir S Gov. Long-range acoustic interactions in insect swarms: an adaptive gravity model. *New Journal of Physics*, 18(7):073042, 2016.

[14] Rachid Guerraoui and Alexandre Maurer. Byzantine fireflies. In *International Symposium on Distributed Computing*, pages 47–59. Springer, 2015.

[15] Yao-Win Hong and Anna Scaglione. A scalable synchronization protocol for large scale sensor networks and its applications. *IEEE Journal on Selected Areas in Communications*, 23(5):1085–1099, 2005.

[16] Minyi Huang and Jonathan H Manton. Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. *SIAM Journal on Control and Optimization*, 48(1):134–161, 2009.

[17] Amos Korman, Oran Ayalon, Yigal Sternklar, Ehud Fonio, Nir Gov, and Ofer Feinerman. Sequential decision-making in ants and implications to the evidence accumulation decision model. *Frontiers in Applied Mathematics and Statistics*, 7:37, 2021.

[18] Amos Korman, Efrat Greenwald, and Ofer Feinerman. Confidence sharing: An economic strategy for efficient information flows in animal groups. *PLoS Computational Biology*, 10(10):e1003862, 2014.

[19] Amos Korman and Robin Vacus. Distributed Alignment Processes with Samples of Group Average. https://hal.archives-ouvertes.fr/hal-03124213, February 2022.

[20] Helen F McCreery and MD Breed. Cooperative transport in ants: a review of proximate mechanisms. *Insectes sociaux*, 61(2):99–110, 2014.

[21] Renato E Mirollo and Steven H Strogatz. Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics*, 50(6):1645–1662, 1990.

[22] Reza Olfati-Saber. Distributed kalman filtering for sensor networks. In *2007 46th IEEE Conference on Decision and Control*, pages 5492–5498. IEEE, 2007.

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

[24] Craig W Reynolds. *Flocks, herds and schools: A distributed behavioral model*, volume 21. ACM, 1987.

[25] Osvaldo Simeone, Umberto Spagnolini, Yeheskel Bar-Ness, and Steven H Strogatz. Distributed synchronization in wireless networks. *IEEE Signal Processing Magazine*, 25(5):81–97, 2008.

[26] Andrew M Simons. Many wrongs: the advantage of group navigation. *Trends in ecology & evolution*, 19(9):453–455, 2004.

[27] Fikret Sivrikaya and Bülent Yener. Time synchronization in sensor networks: a survey. *IEEE network*, 18(4):45–50, 2004.

[28] Bharath Sundararaman, Ugo Buy, and Ajay D Kshemkalyani. Clock synchronization for wireless sensor networks: a survey. *Ad hoc networks*, 3(3):281–323, 2005.

[29] Staffan Tamm. Bird orientation: single homing pigeons compared with small flocks. *Behavioral Ecology and Sociobiology*, 7(4):319–322, 1980.

[30] Huaibin Tang and Tao Li. Continuous-time stochastic consensus: Stochastic approximation and kalman–bucy filtering based protocols. *Automatica*, 61:146–155, 2015.

[31] Tamás Vicsek and Anna Zafeiris. Collective motion. *Physics Reports*, 517(3):71 – 140, 2012. Collective motion.

[32] Zijian Wang and Mac Schwager. Multi-robot manipulation without communication. In *Distributed autonomous robotic systems*, pages 135–149. Springer, 2016.

A Proof of Proposition 1

Proposition 1 is a direct consequence of the following result.

Lemma 6. *Fix $i \in I$ and an integer $t \geq 1$, and consider a history $\mathcal{H}_t$ such that the Kalman filter produces $\hat{\theta}^{-}_{t,i} = 0$ on $\mathcal{H}_t$. Let $d\vartheta^{(t-1)}$ be an alternative move for the agents at round $(t-1)$, and consider history $\mathcal{H}'_t$ obtained from $\mathcal{H}_t$ by replacing $d\theta^{(t-1)}$ by $d\vartheta^{(t-1)}$. Then,*

$$\mathbb{E}\left({\theta_i^{(t)}}^2 \mid \mathcal{H}'_t\right) \geq \mathbb{E}\left({\theta_i^{(t)}}^2 \mid \mathcal{H}_t\right).$$

Proof. Let c be the ith coordinate of the vector $M_n(d\theta^{(t)} - d\vartheta^{(t)})$. By Eq. (2), c is the difference between the position of Agent i in the beginning of round t if it moves by $d\theta^{(t)}$ instead of moving by $d\vartheta^{(t)}$ (conditioning on having the same drift at the end of round $t-1$). In other words,

$$\mathbb{E}\left({\theta_i^{(t)}}^2 \mid \mathcal{H}'_t\right) = \mathbb{E}\left((\theta_i^{(t)} - c)^2 \mid \mathcal{H}_t\right). \tag{26}$$

By Theorem 8,

$$\mathbb{E}\left((\theta_i^{(t)} - c)^2 \mid \mathcal{H}_t\right) \geq \mathbb{E}\left((\theta_i^{(t)} - \hat{\theta}_{t,i})^2 \mid \mathcal{H}_t\right). \tag{27}$$

By assumption, $\hat{\theta}_{t,i} = 0$, so

$$\mathbb{E}\left((\theta_i^{(t)} - \hat{\theta}_{t,i})^2 \mid \mathcal{H}_t\right) = \mathbb{E}\left({\theta_i^{(t)}}^2 \mid \mathcal{H}_t\right). \tag{28}$$

Combining Eqs. (26), (27) and (28) establishes the result. □

B More Results on Weighted-Average Algorithms

Proof. (of Lemma 2.)

$$\begin{aligned}\sum_{i\in I} \overline{\theta}_i^{(t)} &= \sum_{i\in I}\left(\langle\theta_{-i}^{(t)}\rangle - \theta_i^{(t)}\right) = \tfrac{1}{n-1}\sum_{i\in I}\sum_{j\in I\setminus i}\theta_i^{(t)} - \sum_{i\in I}\theta_i^{(t)} \\ &= \tfrac{1}{n-1}\cdot(n-1)\sum_{i\in I}\theta_i^{(t)} - \sum_{i\in I}\theta_i^{(t)} = 0.\end{aligned}$$

□

Proof. (of Lemma 3.) Agent i's stretch at round $t+1$ is:

$$\begin{aligned}\overline{\theta}_i^{(t+1)} &= \langle\theta_{-i}^{(t+1)}\rangle - \theta_i^{(t+1)} = \langle\theta_{-i}^{(t+1)}\rangle - (\theta_i^{(t)} + d\theta_i^{(t)} + N_{d,i}^{(t)}) \\ &= \left(\langle\theta_{-i}^{(t+1)}\rangle - \langle\theta_{-i}^{(t)}\rangle\right) + \left(\langle\theta_{-i}^{(t)}\rangle - \theta_i^{(t)}\right) - d\theta_i^{(t)} - N_{d,i}^{(t)} \\ &= \left(\langle\theta_{-i}^{(t+1)}\rangle - \langle\theta_{-i}^{(t)}\rangle\right) + \overline{\theta}_i^{(t)} - d\theta_i^{(t)} - N_{d,i}^{(t)},\end{aligned} \tag{29}$$

where the first equality is by definition, and the second equality is a consequence of (2). Let us break down the first term:

$$\begin{aligned}\langle\theta_{-i}^{(t+1)}\rangle - \langle\theta_{-i}^{(t)}\rangle &= \tfrac{1}{n-1}\sum_{j\in I\setminus i}\left(d\theta_j^{(t)} + N_{d,j}^{(t)}\right) \\ &= \tfrac{1}{n-1}\sum_{j\in I\setminus i}\left(\rho\left(\overline{\theta}_j^{(t)} + N_{m,j}^{(t)}\right) + N_{d,j}^{(t)}\right),\end{aligned}$$

where the second equality is because $d\theta_j^{(t)} = \rho Y_j^{(t)} = \rho\left(\overline{\theta}_j^{(t)} + N_{m,j}^{(t)}\right)$. By Lemma 2, $\sum_{j\in I\setminus i}\overline{\theta}_j^{(t)} = -\overline{\theta}_i^{(t)}$, so we can rewrite the last equation as

$$\begin{aligned}\langle\theta_{-i}^{(t+1)}\rangle - \langle\theta_{-i}^{(t)}\rangle &= \tfrac{-\rho}{n-1}\overline{\theta}_i^{(t)} + \tfrac{1}{n-1}\sum_{j\in I\setminus i}\rho N_{m,j}^{(t)} + N_{d,j}^{(t)} \\ &= \tfrac{-\rho}{n-1}\overline{\theta}_i^{(t)} + \tfrac{1}{n-1}\sum_{j\in I\setminus i}\mathcal{E}_j^{(t)}\,.\end{aligned} \tag{30}$$

Plugging Eq. (30) into Eq. (29) gives

$$\overline{\theta}_i^{(t+1)} = \left(1 - \tfrac{\rho}{n-1}\right)\overline{\theta}_i^{(t)} + \tfrac{1}{n-1}\sum_{j\in I\setminus i}\mathcal{E}_j^{(t)} - d\theta_i^{(t)} - N_{d,i}^{(t)}. \tag{31}$$

By assumption, $d\theta_i^{(t)} = \rho Y_i^{(t)} = \rho\left(\overline{\theta}_i^{(t)} + N_{m,i}^{(t)}\right)$. Plugging this expression into Eq. (31) gives the result. □

Proof. (of Lemma 4.) We prove that the stretch is normally distributed with mean 0 by induction on t. By construction, for every i, $\overline{\theta}_i^{(0)}$ is normally distributed with mean 0. Let us assume that $\overline{\theta}_i^{(t)}$ is normally distributed with mean 0 for some round t, and consider round $t+1$. Recall that Lemma 3 gives

$$\overline{\theta}_i^{(t+1)} = \left(1 - \tfrac{n}{n-1}\rho\right)\overline{\theta}_i^{(t)} + \tfrac{1}{n-1}\sum_{j\in I\setminus i}\mathcal{E}_j^{(t)} - \mathcal{E}_i^{(t)}\,. \tag{32}$$

Since by definition, for every j, $\mathcal{E}_j^{(t)}$ is normally distributed around 0, and by induction $\overline{\theta}_i^{(t)}$ is normally distributed around 0, then $\overline{\theta}_i^{(t+1)}$ is also normally distributed around 0. This concludes the induction. Noting that $\mathrm{Var}(\mathcal{E}_j^{(t)}) = \rho^2\sigma_m^2 + \sigma_d^2$ and taking the variance in Eq. (32) concludes the proof. □

Proof. (of Claim 3.) First, consider the case that $a < 1$. Let $\lambda = b/(a-1)$. Consider the sequence defined by $v_n = u_n + \lambda$. We have $v_{n+1} = u_{n+1} + \lambda = au_n + b + \lambda = au_n + (a-1)\lambda + \lambda = au_n + a\lambda = a(u_n + \lambda) = av_n$. Since $|a| < 1$, $\lim v_n = 0$, and so $\lim u_n = -\lambda = b/(1-a)$.

Now, if $a = 1$, then we have for every $n \in \mathbb{N}$, $u_n = u_0 + nb$. If $b > 0$ then $\lim u_n = +\infty$. □

C Useful linear algebra claims

Claim 4. *If $X \sim \mathcal{N}(\mu, \Sigma)$, then for every $c \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n\times n}$, $c + BX \sim \mathcal{N}\left(c + B\mu, B\Sigma B^\top\right)$.*

Claim 4 above is well-known, and Claim 5 below follows from a straightforward matrix multiplication.

Claim 5. *For every $a, b, a', b' \in \mathbb{R}$, $\mathcal{M}(a,b)\mathcal{M}(a',b') = \mathcal{M}(aa' + (n-1)bb', ab' + a'b + (n-2)bb')$.*

Claim 6. *For every $a, b \in \mathbb{R}$ such that $a \neq b$ and $a \neq -(n-1)b$, the matrix $\mathcal{M}(a,b)$ is invertible, and $\mathcal{M}(a,b)^{-1} = \frac{\mathcal{M}(a+(n-2)b,-b)}{(a-b)(a+(n-1)b)}$.*

Proof. Note that $\mathbb{1}^2 = n\mathbb{1}$. Let $A = \mathcal{M}(a,b)$. We have

$$\begin{aligned}A^2 &= (b\mathbb{1} + (a-b)I)^2 = b^2\mathbb{1}^2 + 2b(a-b)\mathbb{1} + (a-b)^2 I \\ &= (nb + 2(a-b))(b\mathbb{1}) + (a-b)^2 I \\ &= (2a + (n-2)b)(b\mathbb{1} + (a-b)I) \\ &\quad - (nb + 2(a-b))(a-b)I + (a-b)^2 I \\ &= (2a + (n-2)b)A + (a-b)((a-b) - nb - 2(a-b))I \\ &= (2a + (n-2)b)A - (a-b)(a + (n-1)b)I.\end{aligned}$$

Hence $A(A - (2a + (n-2)b)I) = -(a-b)(a + (n-1)b)I$, from which we conclude the proof of Claim 6. □

D Additional claims about optimal algorithms

D.1 All optimal algorithms are shifts of one another

In this section, we characterize all optimal (deterministic) algorithms for the alignment problem in the centralized setting. We show that each of these algorithms can be obtained from $\mathcal{W}^\star$, by shifting all the agents by the same quantity λ_t.

Theorem 9. *A deterministic algorithm is optimal in the centralized setting if and only if for every round t, there exists λ_t such that for every $i \in I$, $d\theta_i^{(t)} = \rho_\star^{(t)} Y_i^{(t)} + \lambda_t$.*

Proof. We have already established that Algorithm $\mathcal{W}^\star$ is Kalman-perfect. Therefore, by Proposition 1, any other deterministic algorithm is optimal in the centralized setting if and only if it is Kalman-perfect. In other words, it is optimal if and only if it produces a sequence of moves such that for every round t, the Kalman-filter estimator operating on the corresponding process yields

$$\begin{aligned}
\hat{\theta}_{t+1}^{-} = 0 &\iff \hat{\theta}_t + M_n d\theta^{(t)} = 0 \\
&\iff K_t Y^{(t)} + M_n d\theta^{(t)} = 0 \\
&\iff -\rho_\star^{(t)} M_n Y^{(t)} + M_n d\theta^{(t)} = 0 \\
&\iff M_n \left(-\rho_\star^{(t)} Y^{(t)} + d\theta^{(t)} \right) = 0 \\
&\iff -\rho_\star^{(t)} Y^{(t)} + d\theta^{(t)} \in \ker(M_n).
\end{aligned}$$

Writing $\mathbf{1}$ to denote the vector whose coefficients are all equal to 1, we observe that $\mathbf{1} \in \ker(M_n)$. Since $\operatorname{rank}(M_n) = n-1$, $\dim(\ker(M_n)) = 1$, so for every round t,

$$\begin{aligned}
&-\rho_\star^{(t)} Y^{(t)} + d\theta^{(t)} \in \ker(M_n) \\
&\qquad \iff \exists \lambda_t \in \mathbb{R},\ -\rho_\star^{(t)} Y^{(t)} + d\theta^{(t)} = \lambda_t \cdot \mathbf{1} \\
&\qquad\qquad \iff \exists \lambda_t \in \mathbb{R},\ d\theta^{(t)} = \lambda_t \cdot \mathbf{1} + \rho_\star^{(t)} Y^{(t)}.
\end{aligned}$$

□

D.2 Proof of Claim 2

Let A be an optimal algorithm. Again, we denote the variables involved in the execution of A by $[\cdot]^A$. We have shown in Appendix D.1 that all optimal algorithms are shifts of one another, i.e., for all t we can find a random variable λ_t s.t. for all i, $\left[d\theta_i^{(t)}\right]^A = \left[d\theta_i^{(t)}\right]^{\mathrm{MatC}} + \lambda_t$. It implies that $\frac{1}{n}\sum_{i\in I}\left[d\theta_i^{(t)}\right]^A = \lambda_t + \frac{1}{n}\sum_{i\in I}\left[d\theta_i^{(t)}\right]^{\mathrm{MatC}}$, which is equal to λ_t (the second term is 0 following the definition of Algorithm 1). Therefore, $\left[\Delta_t^{\mathrm{rel}}\right]^A = \frac{1}{n}\sum_{s=0}^{t}\sum_{i\in I}\left[d\theta_i^{(s)}\right]^A = \sum_{s=0}^{t}\lambda_s$. If A is different than Algorithm 1, we can define $t_0 := \min\{t \in \mathbb{N}, \mathbb{E}\left(\lambda_t^2\right) > 0\} < +\infty$, and have

$$\mathbb{E}\left(\left[\Delta_{t_0}^{\mathrm{rel}}\right]^{A^2}\right) = \mathbb{E}\left(\left(\sum_{s=0}^{t_0}\lambda_s\right)^2\right) \geq \sum_{s=0}^{t_0}\mathbb{E}\left(\lambda_s^2\right) > 0.$$

By Eq. (13), we conclude that, among optimal algorithm, "meet at the center" is the only one minimizing Δ_t for every t. Note that it is not implementable in the distributed setting, because it requires for each agent to know all measurements. This concludes the proof of Claim 2.