Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
J. Math. Pures Appl. 89 (2008) 278–296
JOURNAL DE MATHÉMATIQUES PURES ET APPLIQUÉES
www.elsevier.com/locate/matpur

Flocking in noisy environments

Felipe Cucker [a,*], Ernesto Mordecki [b]

[a] *Department of Mathematics, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong*
[b] *Centro de Matemática, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay*

Received 5 July 2007

Available online 5 December 2007

Abstract

In recent years, a number of articles proposed mathematical models for emergent phenomena. This is the case, for instance for the flocking of birds or the schooling of fish. In particular, in [F. Cucker, S. Smale, Emergent behavior in flocks, IEEE Trans. on Autom. Control 52 (2007) 852–862], a model was proposed for flocking and it was proved that under certain conditions on the initial positions and velocities of the birds, flocking occurs. In this paper we modify this model by adding random noise to it. We prove that, under conditions similar to those just mentioned, (nearly) flocking occurs in finite time with a certain confidence.

Résumé

Au cours de ces dernières années un certain nombre d'articles ont été consacrés à la modélisation mathématique des phénomènes émergents, par exemple dans le cas des rassemblements d'oiseaux ou de poissons. En particulier dans [F. Cucker, S. Smale, Emergent behavior in flocks, IEEE Trans. on Autom. Control 52 (2007) 852–862] un modèle a été proposé pour ces phénomènes, dans lequel il a été prouvé que le rassemblement des oiseaux se produit sous certaines conditions de positions et de vitesses initiales. Dans cet article nous modifions ce modèle par ajout d'un bruit aléatoire et nous montrons que, sous des conditions similaires, un quasi-rassemblement s'opère en temps fini avec un niveau élevé de probabilité.

MSC: 70K75; 70Q05

Keywords: Flocking; Emergent behavior; Consensus

1. Introduction

The problem of reaching a consensus in a group of autonomous agents has been the object of study in a number of situations ranging from linguistics [11,18,20] to distributed computing [25,26] and from physics [27] to animal behavior [6,13,23,24].

Research on the latter attempts to explain, by appropriately modeling it, the observed behavior of a group of animals, say a flock of birds, whose velocities converge to a common one. An influential model for this behavior has been postulated in [27] by Vicsek and collaborators and studied in [17] where convergence is shown under some

* Corresponding author.
E-mail addresses: macucker@math.cityu.edu.hk (F. Cucker), mordecki@cmat.edu.uy (E. Mordecki).

0021-7824/$ – see front matter
doi:10.1016/j.matpur.2007.12.002

conditions on the sequence of states of the flock. Also, the work of Buhl et al. [5] showed that Vicsek's model accurately describes the convergence to order in swarms of locusts. A different model for the same phenomenon was proposed in [9] and extended in [10] (to situations other than flocking) and in [22] (to flocking situations where a hierarchical leadership structure is present). A main feature of these papers is that, in contrast with the results in [17,27], convergence results are established conditioned to the *initial* state of the flock only.

The model in [9] postulates the following behavior: every bird adjusts its velocity by adding to it a weighted average of the differences of its velocity with those of the other birds. That is, at time t, and for bird i,

$$\boldsymbol{v}_i(t+h)=\boldsymbol{v}_i(t)+h\sum_{j=1}^{k}a_{ij}\big(\boldsymbol{v}_j(t)-\boldsymbol{v}_i(t)\big), \tag{1}$$

where the weights $\{a_{ij}\}$ quantify the way the birds influence each other, and $h>0$ is the time step. Vicsek's model is more general in the sense that allows the obtained $\boldsymbol{v}_i(t+h)$ to be perturbed by some centered noise. More precisely, it replaces (1) by:

$$\boldsymbol{v}_i(t+h)=\boldsymbol{v}_i(t)+h\sum_{j=1}^{k}a_{ij}\big(\boldsymbol{v}_j(t)-\boldsymbol{v}_i(t)\big)+h\boldsymbol{H}_i(t), \tag{2}$$

where $\boldsymbol{H}_i(t)\in\mathbb{E}$ is some centered random variable modeling the noise. Here $\mathbb{E}$ denotes 3-dimensional Euclidean space.

It is reasonable to assume that the influence weights a_{ij} are a function of the distance between birds. This is the case in [9,27] and the major difference between these models is in the choice of the a_{ij}. In this paper we will follow [9] but slightly depart from it and take the *adjacency matrix* A_x to have entries,

$$a_{ij}=\frac{K}{(1+\|\boldsymbol{x}_i-\boldsymbol{x}_j\|)^{\alpha}}, \tag{3}$$

for some fixed $K>0$ and $\alpha\geqslant 0$.

We can write the set of equalities (1) in a more concise form. Let D_x be the $k\times k$ diagonal matrix whose ith diagonal entry is $d_i=\sum_{j\leqslant k}a_{ij}$ and $L_x=D_x-A_x$. Then (cf. [9]),

$$\boldsymbol{v}(t+h)-\boldsymbol{v}(t)=-hL_x\boldsymbol{v}(t)+h\boldsymbol{H}(t),$$

where $\boldsymbol{H}=(\boldsymbol{H}_1,\dots,\boldsymbol{H}_k)$.

Note that the matrix notation $L_x\boldsymbol{v}(t)$ does not have the usual meaning of a $k\times k$ matrix acting on $\mathbb{R}^k$. Instead, the matrix L_x is acting on $\mathbb{E}^k$ by mapping $(\boldsymbol{v}_1,\dots,\boldsymbol{v}_k)$ to $((L_x)_{i1}\boldsymbol{v}_1+\dots+(L_x)_{ik}\boldsymbol{v}_k)_{i\leqslant k}$.

Adding a natural equation for the change of positions we obtain the discrete dynamical system:

$$\begin{aligned}\boldsymbol{x}(t+h)&=\boldsymbol{x}(t)+h\boldsymbol{v}(t),\\ \boldsymbol{v}(t+h)&=(\mathrm{Id}-hL_x)\boldsymbol{v}(t)+h\boldsymbol{H}(t).\end{aligned} \tag{D}$$

We also consider evolution for continuous time. The corresponding model is obtained by taking limits for $h\to 0$ and can be given by the system of differential equations:

$$\begin{aligned}\boldsymbol{x}'&=\boldsymbol{v},\\ \boldsymbol{v}'&=-L_x\boldsymbol{v}+\boldsymbol{H}.\end{aligned} \tag{C}$$

The main result in [9] shows, in the unperturbed case, for both discrete and continuous time, convergence to the alignment of the velocities. More precisely, convergence to a common velocity when the initial positions and velocities of the flock are not too dissimilar (for $\alpha\geqslant 1$, otherwise, convergence holds unconditionally). For systems (D) and (C), due to the presence of noise, we can not expect convergence to a common velocity. Once the velocities $\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}$ are similar enough compared with the noise the latter will, with positive probability, outdo the contractive character of the system. Perfect alignment as in [9] should therefore be replaced by "nearly-alignment". A formal measure of similarity (and with it a definition of nearly-alignment) will be given soon in Section 2. A description of the forms of noise we consider in this paper will be given in Sections 3.1 and 4.2. We nevertheless state now an informal version of our main results (see Theorems 1 and 3 for precise statements).

Main result. *Assume that at time* 0 *the positions and velocities of the flock are not both too dissimilar* (*for* $\alpha \geqslant 1$, *otherwise, no assumption is needed*) *and that the time step h is small enough* (*in case of discrete time*). *Then nearly-alignment is* (*quickly*) *reached with a certain probability and we exhibit a lower bound for this probability in terms of the initial similarity of positions and velocities, the variance of* $\boldsymbol{H}$, *and the parameters* k, K *and* α.

The critical value $\alpha = 1$ is sharp as shown in [9] (cf. Remark 6 below).

The proof of our main result closely follows the proofs in [9]. Some changes had to be made to make room for the noise and in doing so we did a few simplifications as well.

2. Some preliminaries

2.1. Laplacians

Given a nonnegative, symmetric, $k \times k$ matrix A the *Laplacian* L of A is defined to be

$$L = D - A,$$

where $D = \mathsf{diag}(d_1, \dots, d_k)$ and $d_\ell = \sum_{j=1}^k a_{\ell j}$. Some features of L are immediate. It is symmetric and it does not depend on the diagonal entries of A.

The matrix L_x in (D) and (C) is thus the Laplacian of A_x. It satisfies that for all $u \in \mathbb{E}$, $L_x(u, \dots, u) = 0$. In addition, it is positive semidefinite.

The smallest eigenvalue of L_x is zero. Its second eigenvalue is called the *Fiedler number* of A_x. We will denote it by ϕ_x.

Proposition 1. *(See [10, Proposition 1].) Let A be a $k \times k$ nonnegative, symmetric matrix, $L = D - A$ its Laplacian, ϕ its Fiedler number, and $\mu = \min_{i \neq j} a_{ij}$. Then $\phi \geqslant k\mu$.*

2.2. Similarity and nearly-alignment

The inner product on $\mathbb{E}$ naturally induces an inner product on $\mathbb{E}^k$. Let Δ be the diagonal of $\mathbb{E}^k$, i.e.,

$$\Delta = \big\{(u, u, \dots, u) \;\big|\; u \in \mathbb{E}\big\},$$

and $\Delta^\perp$ be the orthogonal complement of Δ in $\mathbb{E}^k$. Then, every point $\boldsymbol{v} \in \mathbb{E}^k$ decomposes in a unique way as $\boldsymbol{v} = \boldsymbol{v}_\Delta + \boldsymbol{v}_\perp$ with $\boldsymbol{v}_\Delta \in \Delta$ and $\boldsymbol{v}_\perp \in \Delta^\perp$. This decomposition has a simple explicit form. Denote by,

$$\boldsymbol{m} = \frac{1}{k}\sum_{i=1}^k \boldsymbol{v}_i,$$

the mean of the $\boldsymbol{v}_i$. Then $\boldsymbol{v}_\Delta = (\boldsymbol{m}, \dots, \boldsymbol{m})$ and $\boldsymbol{v}_\perp = (\boldsymbol{v}_1 - \boldsymbol{m}, \dots, \boldsymbol{v}_k - \boldsymbol{m})$. This follows immediately from the equality,

$$\begin{aligned}\langle \boldsymbol{v}_\Delta, \boldsymbol{v}_\perp\rangle &= \sum_{i=1}^k \big\langle \boldsymbol{m}, (\boldsymbol{v}_i - \boldsymbol{m})\big\rangle = \left\langle \boldsymbol{m}, \sum_{i=1}^k \boldsymbol{v}_i - \boldsymbol{m}\right\rangle \\ &= \left\langle \boldsymbol{m}, \left(\sum_{i=1}^k \boldsymbol{v}_i\right) - k\boldsymbol{m}\right\rangle = \langle \boldsymbol{m}, 0\rangle = 0.\end{aligned}$$

We can look at the evolution of the velocities $\boldsymbol{v}_i(t)$ decomposing into the evolution of their mean $\boldsymbol{m}(t)$ and that of the distances to that mean $\boldsymbol{v}_\perp = (\boldsymbol{v}_1 - \boldsymbol{m}, \dots, \boldsymbol{v}_k - \boldsymbol{m})$ and a key observation at this stage is the fact that convergence to a common velocity (or nearly-alignment) is a feature of the second evolution only. More precisely, the condition "the velocities $\boldsymbol{v}_i(t)$ tend to alignment" is equivalent to the condition "$\boldsymbol{v}_\perp(t) \to 0$". We are thus interested on the projection $(\boldsymbol{x}_\perp(t), \boldsymbol{v}_\perp(t))$ over $\Delta^\perp \times \Delta^\perp$ of the solutions $(\boldsymbol{x}(t), \boldsymbol{v}(t))$ of the system (D) (or (C)). It is easy to show (see [9]) that these projections are the solutions of the restriction of (D) (respectively, (C)) to $\Delta^\perp \times \Delta^\perp$.

More precisely, they are the solutions of:

$$\begin{aligned}\boldsymbol{x}(t+h)_\perp &= \boldsymbol{x}(t)_\perp + h\boldsymbol{v}(t)_\perp,\\ \boldsymbol{v}(t+h)_\perp &= (\mathrm{Id} - hL_{x_\perp})\boldsymbol{v}(t)_\perp + h\boldsymbol{H}_\perp.\end{aligned}$$

Hence, in what follows, we will consider positions in,

$$X := \mathbb{E}^k/\Delta \simeq \Delta^\perp,$$

and velocities in

$$V := \mathbb{E}^k/\Delta \simeq \Delta^\perp.$$

For $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{E}^k$ we will denote $x = \boldsymbol{x}_\perp$ and $v = \boldsymbol{v}_\perp$. Finally, we will denote $H = \boldsymbol{H}_\perp$. With these notations, our focus is now on the system:

$$\begin{aligned}x(t+h) &= x(t) + hv(t),\\ v(t+h) &= (\mathrm{Id} - hL_x)v(t) + hH(t),\end{aligned} \tag{D}$$

and with continuous time,

$$\begin{aligned}x' &= v,\\ v' &= -L_x v + H.\end{aligned} \tag{C}$$

Note that we still refer to these systems as (D) and (C).

It is natural now to take the norm $\|\boldsymbol{x}_\perp\|$ of the projection $\boldsymbol{x}_\perp$ as the *dissimilarity* of $\boldsymbol{x}$ and similarly for $\|\boldsymbol{v}_\perp\|$. In the case of $\boldsymbol{x}$ we may call this measure the *dispersion* of the flock. It relates with its "diameter".

Lemma 1. *For all* $\boldsymbol{x} \in \mathbb{E}^k$, $\max_{i\neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leqslant \sqrt{2}\|\boldsymbol{x}_\perp\|$.

Proof. Write $\boldsymbol{x} = \boldsymbol{x}_\Delta + \boldsymbol{x}_\perp = (\tilde{u}, \ldots, \tilde{u}) + ((\boldsymbol{x}_\perp)_1, \ldots, (\boldsymbol{x}_\perp)_k)$. Then, for all $i \neq j$, $\boldsymbol{x}_i - \boldsymbol{x}_j = (\boldsymbol{x}_\perp)_i - (\boldsymbol{x}_\perp)_j$ and

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_{\mathbb{E}} = \big\|(\boldsymbol{x}_\perp)_i - (\boldsymbol{x}_\perp)_j\big\|_{\mathbb{E}} \leqslant \big\|(\boldsymbol{x}_\perp)_i\big\|_{\mathbb{E}} + \big\|(\boldsymbol{x}_\perp)_j\big\|_{\mathbb{E}} \leqslant \sqrt{2}\|\boldsymbol{x}_\perp\|_{\mathbb{E}^k}. \qquad \square$$

The notion of similarity leads to the following definition:

Definition 1. Let $\nu > 0$. We say that the flock $\{1, \ldots, k\}$ is ν-*nearly-aligned* (or simply *nearly-aligned*) when $\|\boldsymbol{v}_\perp\| \leqslant \nu$.

2.3. A few functions of the initial state

The initial state of the flock is characterized by the pair $(\boldsymbol{x}(0), \boldsymbol{v}(0))$. For convergence to alignment (or to nearly-alignment) to hold one needs to require that the dissimilarities of these two vectors are not both large.

We close this section with a few quantities related to these initial dissimilarities which will occur when describing the conditions ensuring convergence. These are:

$$\boldsymbol{a} = \frac{2\sqrt{2}}{kK}\big\|v(0)\big\|, \qquad \boldsymbol{b} = 1 + \sqrt{2}\big\|x(0)\big\|;$$

$$U_0 = \begin{cases} \max\{(2\boldsymbol{a})^{1/(1-\alpha)}, 2\boldsymbol{b}\} & \text{if } \alpha < 1,\\ \frac{\boldsymbol{b}}{1-\boldsymbol{a}} & \text{if } \alpha = 1,\\ \frac{\alpha}{\alpha-1}\boldsymbol{b} & \text{if } \alpha > 1;\end{cases}$$

$$B_0 = \frac{U_0 - 1}{\sqrt{2}}, \quad \text{and} \quad \mathcal{H}_0 = \frac{2^{-\alpha-1}kK}{U_0^\alpha}.$$

3. Discrete time

Assume the initial state for (D) is at time 0. Then the sequence of states is $\{x(th), v(th)\}_{t\in\mathbb{N}}$. To simplify notation we will denote $x(th)$ simply by $x[t]$ and similarly for v.

3.1. Statement of the result

Recall, the random noise in (D) has the form $\boldsymbol{H} = (\boldsymbol{H}_1, \ldots, \boldsymbol{H}_k)$ and we have $\boldsymbol{H} = \boldsymbol{H}_\Delta + H$. Note that the component $\boldsymbol{H}_\Delta$ of $\boldsymbol{H}$ corresponds to the perturbation of the common velocity $\boldsymbol{v}_\Delta$ within Δ and is therefore of no consequence regarding convergence to alignment or nearly-alignment.

In what follows we assume that, for all $i \in \{1, \ldots, k\}$, and for all $t \in \mathbb{N}$,

$$\boldsymbol{H}_i(t) = \big(e_i^{(1)}(t), e_i^{(2)}(t), e_i^{(3)}(t)\big),$$

where the $e_i^{(\ell)}(t)$ are one-dimensional random variables, the coordinates of the perturbation.

We consider two possible laws for the distribution of $\boldsymbol{H}$:

$$\textbf{Uniform:}\quad \boldsymbol{H} \simeq U_{3k}(0, r),$$

where, for some $r > 0$, $U_{3k}(0, r)$ is the uniform distribution in $B(0, r) \subset \mathbb{R}^{3k}$, and

$$\textbf{Gaussian:}\quad \boldsymbol{H} \simeq N\big(0, \sigma^2 \operatorname{Id}_{3k}\big),$$

a $3k$-dimensional centered Gaussian distribution with covariance matrix $\sigma^2 \operatorname{Id}_{3k}$. As a consequence, in the Gaussian case, the random variables $e_i^{(\ell)}$ are independent.

Our main result for discrete time is the following.

Theorem 1. *Consider the system* (D) *with adjacency matrix given by* (3)*. Assume that h satisfies,*

$$h < \min\left\{\frac{1}{2(k-1)\sqrt{kK}}, \frac{1}{2\sqrt{2}\|v(0)\|}\left(\frac{kK}{2\mathcal{H}_0}\right)^{1/\alpha}\right\}.$$

Assume also that one of the three following hypothesis holds:

(i) $\alpha < 1$,
(ii) $\alpha = 1$, *and* $\|v(0)\| < \frac{kK}{2\sqrt{2}}$,
(iii) $\alpha > 1$, *and*

$$\left(\frac{1}{\alpha \boldsymbol{a}}\right)^{1/(\alpha-1)} \frac{\alpha - 1}{\alpha} > \boldsymbol{b} + 2kKh\boldsymbol{a}.$$

Then ν-nearly-alignment for some $\nu < \|v(0)\|$ occurs in a number of iterations bounded by:

$$T_0 := \frac{2U_0^\alpha}{hkK} \ln\left(\frac{\|v(0)\|}{\nu}\right),$$

with probability at least

$$\left(\frac{\mathcal{H}_0 \nu}{r}\right)^{3kT_0},$$

in the uniform case (1 *if* $r \leqslant \mathcal{H}_0\nu$)*, and with probability at least*

$$\left(\int_0^{\sqrt{\mathcal{H}_0\nu/(2\sigma)}} \frac{t^{\frac{3k-5}{2}}}{\Gamma(\frac{3k-3}{2})} \mathrm{e}^{-t}\, \mathrm{d}t\right)^{T_0},$$

in the Gaussian case.

Remark 1. For each of the cases (i), (ii), and (iii) we can replace U_0 and $\mathcal{H}_0$ by their respective values. In case (iii) and with uniform noise, for instance, this yields:

$$T_0 = \frac{2}{hkK}\ln\left(\frac{\|v(0)\|}{\nu}\right)\left(\left(1+\sqrt{2}\|x(0)\|\right)\frac{\alpha}{\alpha-1}\right)^{\alpha}$$

and

$$\frac{\mathcal{H}_0\nu}{r} = \frac{2^{-\alpha-1}kK}{r}\left(\frac{\alpha-1}{\alpha(1+\sqrt{2}\|x(0)\|)}\right)^{\alpha}.$$

Note that this means that for

$$T_0 = \frac{1}{hkK}\ln\left(\frac{\|v(0)\|}{\nu}\right)\mathcal{O}\left(\|x(0)\|^{\alpha}\right),$$

we have:

$$\mathsf{Prob}\{\text{nearly-align in at most } T_0 \text{ iterations}\} \geqslant \left(\frac{kK}{r\mathcal{O}(\|x(0)\|^{\alpha})}\right)^{3kT_0}.$$

From these expressions it is easy to read the role of the deterministic setting parameters h, k and K, the probabilistic r, the radius ν, and the initial dissimilarities $\|x(0)\|$, $\|v(0)\|$ both in the time required to reach nearly-alignment and in the confidence with which this occurs.

Remark 2. The integral in the bound for the probability in the Gaussian case satisfies, when k is odd and writing $n = \frac{3k-3}{2}$, the equality:

$$\int_0^x \frac{t^{n-1}}{\Gamma(n)}\mathrm{e}^{-t}\,\mathrm{d}t = 1 - \mathrm{e}^{-x}\left(\frac{x^{n-1}}{(n-1)!} + \cdots + \frac{x}{1!} + 1\right).$$

For $x = \sqrt{\mathcal{H}_0\nu/(2\sigma)}$ and for small σ this probability bound is equivalent to,

$$1 - \frac{T_0}{\Gamma(n)}\mathrm{e}^{-x}x^{n-1}.$$

By L'Hôpital's rule, this equivalence holds as well when k is even.

3.2. Bounded noise

Fix a solution (x, v) of (D). At a time $t \in \mathbb{N}$, $x[t]$ and $v[t]$ are elements in X and V, respectively. In particular, $x[t]$ determines an adjacency matrix $A_{x[t]}$. For notational simplicity we will denote its Laplacian and Fiedler number by L_t and ϕ_t, respectively.

Lemma 2. *For all $x \in X$,*

$$\|L_x\| \leqslant 2(k-1)\sqrt{k}K.$$

In particular, if $h < \frac{1}{2(k-1)\sqrt{k}K}$ then $h\|L_x\| \in (0, 1]$.

Proof. For all $i, j \leqslant k$, $a_{ij} \leqslant K$. Therefore,

$$\|L_x\|_{\max} = \max_{i\leqslant k}\sum_{j=1}^{k}\left|(L_x)_{ij}\right| \leqslant 2(k-1)K.$$

Now use that $\|L_x\| \leqslant \sqrt{k}\|L_x\|_{\max}$ [14, Table 6.2] to deduce the result. □

Proposition 2. *Assume that* $h < \frac{1}{2(k-1)\sqrt{kK}}$. *Assume also that, for all* $0 \leqslant t < T$, $\|H\| \leqslant \mathcal{H}_0\|v[t]\|$. *Then, for all* $t < T$,

$$\|v[t+1]\| \leqslant (1 - h\phi_t + h\mathcal{H}_0)\|v[t]\|.$$

In particular, $\|v\|$ *is decreasing as a function of* t *for* $t < T$.

Proof. The linear map $\mathrm{Id} - hL_t$ is self-adjoint and its eigenvalues are in the interval $(0, 1)$. Its largest eigenvalue is $1 - h\phi_t$. Therefore

$$\begin{aligned}\|v[t+1]\| &= \|(\mathrm{Id} - hL_t)v[t] + hH\| \leqslant \|\mathrm{Id} - hL_t\|\|v[t]\| + h\|H\| \\ &\leqslant (1 - h\phi_t)\|v[t]\| + h\mathcal{H}_0\|v[t]\| = (1 - h\phi_t + h\mathcal{H}_0)\|v[t]\|. \qquad \square\end{aligned}$$

Corollary 1. *In the hypothesis of Proposition* 2, *for all* $t \in \{0, \ldots, T-1\}$ *we have*:

$$\|v[t]\| \leqslant \|v[0]\| \prod_{i=0}^{t-1}(1 - h\phi_i + h\mathcal{H}_0).$$

A proof of the following lemma is in [8, Lemma 7].

Lemma 3. *Let* $c_1, c_2 > 0$ *and* $s > q > 0$. *Then the equation,*

$$F(z) = z^s - c_1 z^q - c_2 = 0$$

has a unique positive zero z_*. *In addition,*

$$z_* \leqslant \max\{(2c_1)^{1/(s-q)}, (2c_2)^{1/s}\}$$

and $F(z) \leqslant 0$ *for* $0 \leqslant z \leqslant z^*$.

Theorem 2. *Let* $T \in \mathbb{N} \cup \{+\infty\}$. *Assume that, for all* $0 \leqslant t < T$, $\|H\| \leqslant \mathcal{H}_0\|v[t]\|$, *and that* h *satisfies,*

$$h < \min\left\{\frac{1}{2(k-1)\sqrt{kK}}, \frac{1}{2\sqrt{2}\|v[0]\|}\left(\frac{kK}{2\mathcal{H}_0}\right)^{1/\alpha}\right\},$$

where $\mathcal{H}_0$ *is as in Theorem* 1. *Assume also that one of the three following hypothesis holds*:

(i) $\alpha < 1$,
(ii) $\alpha = 1$, *and* $\|v[0]\| < \frac{kK}{2\sqrt{2}}$,
(iii) $\alpha > 1$, *and*

$$\left(\frac{1}{\alpha \boldsymbol{a}}\right)^{1/(\alpha-1)} \frac{\alpha - 1}{\alpha} > \boldsymbol{b} + 2kKh\boldsymbol{a}.$$

Then $1 - h\frac{kK}{2U_0^\alpha} \in (0, 1)$, *for all* $0 \leqslant t < T$, $\|x[t]\| \leqslant B_0$ *and*

$$\|v[t]\| \leqslant \|v[0]\|\left(1 - h\frac{kK}{2U_0^\alpha}\right)^t.$$

In particular, when $T = \infty$, $\|v[t]\| \to 0$ *for* $t \to \infty$.

Proof. Let

$$\Upsilon = \left\{t \in \{0, \ldots, T-1\} \;\middle|\; (1 + \sqrt{2}\|x(t)\|)^\alpha \leqslant \frac{kK}{2\mathcal{H}_0}\right\}.$$

Note that in all three cases ((i), (ii), and (iii)) the definition of $\mathcal{H}_0$ implies that $0 \in \Upsilon$ and hence, that $\Upsilon \neq \emptyset$. Assume that $\Upsilon \neq \{0, \ldots, T-1\}$ and let $\hat{t} = \min\{\{0, \ldots, T-1\} \setminus \Upsilon\}$.

For $t < T$, let t^* be the point maximizing $\|x\|$ in $\{0, 1, \ldots, t\}$. Then, by Proposition 1 and Lemma 1, for $i \in \{0, 1, \ldots, t\}$,

$$\phi_i \geqslant \frac{kK}{(1+\sqrt{2}\|x[i]\|)^\alpha} \geqslant \frac{kK}{(1+\sqrt{2}\|x[t^*]\|)^\alpha}.$$

Moreover, since $t^* \leqslant t < \hat{t}$ we have:

$$\phi_i - \mathcal{H}_0 \geqslant \frac{kK}{(1+\sqrt{2}\|x[t^*]\|)^\alpha} - \mathcal{H}_0 \geqslant \frac{kK}{2(1+\sqrt{2}\|x[t^*]\|)^\alpha} =: R(t^*).$$

Using Corollary 1 we obtain, for all $\tau \leqslant t$,

$$\begin{aligned}
\|x[\tau]\| &\leqslant \|x[0]\| + \sum_{j=0}^{\tau-1}\|x[j+1]-x[j]\| \leqslant \|x[0]\| + h\sum_{j=0}^{\tau-1}\|v[j]\| \\
&\leqslant \|x[0]\| + h\left(\|v[0]\| + \sum_{j=1}^{\tau-1}\|v[j]\|\right) \\
&\leqslant \|x[0]\| + h\left(\|v[0]\| + \sum_{j=1}^{\tau-1}\|v[0]\|\prod_{i=1}^{j}(1-h\phi_i+h\mathcal{H}_0)\right) \\
&\leqslant \|x[0]\| + h\|v[0]\|\sum_{j=0}^{\tau-1}\big(1-hR(t^*)\big)^j \\
&\leqslant \|x[0]\| + h\frac{1}{hR(t^*)}\|v[0]\| \\
&= \|x[0]\| + \frac{2(1+\sqrt{2}\|x[t^*]\|)^\alpha}{kK}\|v[0]\|.
\end{aligned}$$

Multiplying by $\sqrt{2}$ and taking $\tau = t^*$, the inequality above takes the following equivalent form:

$$\sqrt{2}\|x[t^*]\| \leqslant \sqrt{2}\|x[0]\| + \frac{2\sqrt{2}(1+\sqrt{2}\|x[t^*]\|)^\alpha}{kK}\|v[0]\|,$$

or yet

$$\big(1+\sqrt{2}\|x[t^*]\|\big) \leqslant \big(1+\sqrt{2}\|x[0]\|\big) + \frac{2\sqrt{2}(1+\sqrt{2}\|x[t^*]\|)^\alpha}{kK}\|v[0]\|. \tag{4}$$

Let $z = 1+\sqrt{2}\|x[t^*]\|$. Then (4) can be rewritten as $F(z) \leqslant 0$, with

$$F(z) = z - \boldsymbol{a}z^\alpha - \boldsymbol{b}.$$

(i) Assume $\alpha < 1$. By Lemma 3, $F(z) \leqslant 0$ implies that $(1+\sqrt{2}\|x[t^*]\|) \leqslant U_0$. Since U_0 is independent of t we deduce that, for all $t < \hat{t}$,

$$\|x[t]\| \leqslant \frac{U_0 - 1}{\sqrt{2}} = B_0.$$

Therefore, for all $t < \hat{t}$,

$$\big(1+\sqrt{2}\|x[t]\|\big)^\alpha \leqslant \big(1+\sqrt{2}\|x[t^*]\|\big)^\alpha \leqslant U_0^\alpha \leqslant 2^{-\alpha}\frac{kK}{2\mathcal{H}_0},$$

the last by the definition of $\mathcal{H}_0$. It follows that

$$\|x[\hat{t}]\| = \|x[\hat{t}-1]\| + h\|v[\hat{t}-1]\| \leqslant \|x[\hat{t}-1]\| + h\|v[0]\|,$$

and therefore

$$\begin{aligned}\left(1+\sqrt{2}\|x[\hat{t}\,]\|\right) &\leqslant 1+\sqrt{2}\|x[\hat{t}-1]\|+\sqrt{2}h\|v[0]\| \\ &\leqslant \left(2^{-\alpha}\frac{kK}{2\mathcal{H}_0}\right)^{1/\alpha}+\sqrt{2}h\|v[0]\| \leqslant \left(\frac{kK}{2\mathcal{H}_0}\right)^{1/\alpha},\end{aligned}$$

the last by our hypothesis on h. This is in contradiction with the definition of $\hat{t}$ and shows that no such $\hat{t}$ exists. That is, for all $t < T$, $\|x[t]\| \leqslant B_0$, and

$$\phi_t - \mathcal{H}_0 \geqslant F_0 := \frac{kK}{2(1+\sqrt{2}B_0)^\alpha} = \frac{kK}{2U_0^\alpha}.$$

By Corollary 1, for $t < T$,

$$\|v[t]\| \leqslant \|v[0]\| \prod_{i=0}^{t-1}(1-h\phi_i+h\mathcal{H}_0) \leqslant (1-hF_0)^t\|v[0]\|.$$

The convergence results for the case $T=\infty$ now readily follow (cf. [9, Theorem 3]).

(ii) Assume now $\alpha = 1$. Then (4) takes the form

$$\left(1+\sqrt{2}\|x[t^*]\|\right)\left(1-\frac{2\sqrt{2}}{kK}\|v[0]\|\right)-\left(1+\sqrt{2}\|x[0]\|\right) \leqslant 0.$$

By hypothesis, $\|v[0]\| < \frac{kK}{2\sqrt{2}}$. This implies that

$$\|x[t^*]\| \leqslant kK\frac{1+\sqrt{2}\|x[0]\|}{\sqrt{2}kK-4\|v[0]\|}-1 = B_0.$$

We conclude that, for all $t < \hat{t}$,

$$\left(1+\sqrt{2}\|x[t]\|\right)^\alpha \leqslant 1+\sqrt{2}\|x[t^*]\| \leqslant kK\frac{1+\sqrt{2}\|x[0]\|}{kK-2\sqrt{2}\|v[0]\|} \leqslant \frac{kK}{4\mathcal{H}_0},$$

by the definition of $\mathcal{H}_0$. We now proceed as in case (i).

(iii) Assume finally $\alpha > 1$. The derivative $F'(z) = 1-\alpha\boldsymbol{a}z^{\alpha-1}$ has a unique zero at $z_* = (\frac{1}{\alpha\boldsymbol{a}})^{1/(\alpha-1)}$, and

$$\begin{aligned}F(z_*) &= \left(\frac{1}{\alpha\boldsymbol{a}}\right)^{1/(\alpha-1)} - \boldsymbol{a}\left(\frac{1}{\alpha\boldsymbol{a}}\right)^{\alpha/(\alpha-1)} - \boldsymbol{b} \\ &= \left(\frac{1}{\alpha}\right)^{1/(\alpha-1)}\left(\frac{1}{\boldsymbol{a}}\right)^{1/(\alpha-1)} - \left(\frac{1}{\alpha}\right)^{\alpha/(\alpha-1)}\left(\frac{1}{\boldsymbol{a}}\right)^{1/(\alpha-1)} - \boldsymbol{b} \\ &= \left(\frac{1}{\boldsymbol{a}}\right)^{1/(\alpha-1)}\left(\frac{1}{\alpha}\right)^{1/(\alpha-1)}\frac{\alpha-1}{\alpha} - \boldsymbol{b} \\ &> 0\end{aligned}$$

the last by our hypothesis. Since $F(0) = -\boldsymbol{b} < 0$, $F''(z) = \alpha(\alpha-1)\boldsymbol{a}z^{\alpha-2} > 0$ for all $z > 0$, and $F(z) \to -\infty$ when $z \to \infty$, we deduce that the shape of F is as follows (see Fig. 1).

For $t \in \mathbb{N}$ let $z(t) = 1+\sqrt{2}\|x[t^*]\|$. When $t=0$ we have $t^*=0$ as well and

$$z(0) \leqslant 1+\sqrt{2}\|x[0]\| = \boldsymbol{b} < \left(\frac{1}{\boldsymbol{a}}\right)^{1/(\alpha-1)}\left(\frac{1}{\alpha}\right)^{1/(\alpha-1)} = z_*.$$

This implies that $z(0) < z_\ell$. Assume that there exists $t < T$ such that $z(t) \geqslant z_u$ and let r be the first such t. Then $r = r^* \geqslant 1$ and, for all $t < r$,

$$1+\sqrt{2}\|x[t]\| \leqslant z(r-1) \leqslant z_\ell.$$

Let z_0 be the intersection of the z axis with the line segment joining $(0,-\boldsymbol{b})$ and $(z_*, F(z_*))$ (see Fig. 1). The line where this segment lies has equation,

$$y+\boldsymbol{b} = z\frac{z_*-\boldsymbol{a}z_*^\alpha}{z_*},$$

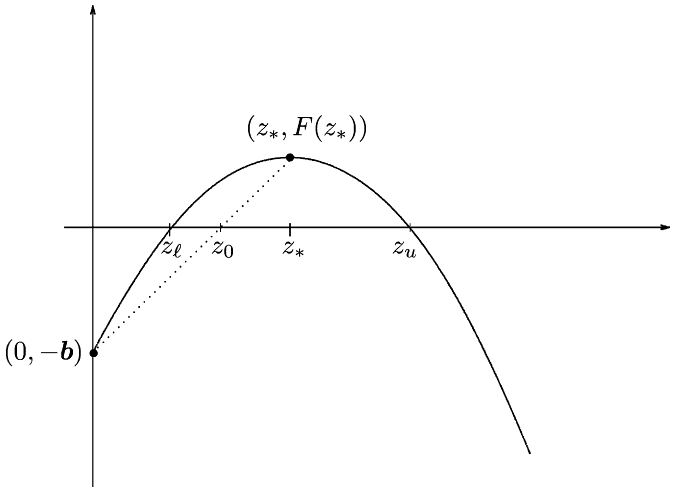

Fig. 1.

from which it follows that

$$z_0 = \frac{\boldsymbol{b}}{1 - \boldsymbol{a} z_*^{\alpha-1}} = \left(1 + \sqrt{2}\|x(0)\|\right)\frac{\alpha}{\alpha - 1}.$$

It follows that, for all $t < r$,

$$\|x[t]\| \leqslant \frac{1}{\sqrt{2}}(z_\ell - 1) \leqslant \frac{1}{\sqrt{2}}(z_0 - 1) = B_0.$$

In particular,

$$\|x[r-1]\| \leqslant \frac{1}{\sqrt{2}}(z_\ell - 1).$$

For r instead, we have

$$\|x[r]\| \geqslant \frac{1}{\sqrt{2}}(z_u - 1).$$

This implies

$$\|x[r] - x[r-1]\| \geqslant \|x[r]\| - \|x[r-1]\| \geqslant \frac{1}{\sqrt{2}}(z_u - z_\ell) \geqslant \frac{1}{\sqrt{2}}(z_* - z_\ell). \tag{5}$$

From the intermediate value theorem, there is $\xi \in [z_\ell, z_*]$ such that $F(z_*) = F'(\xi)(z_* - z_\ell)$. But $F'(\xi) \geqslant 0$ and $F'(\xi) = 1 - a\alpha\xi^{\alpha-1} \leqslant 1$. Therefore,

$$z_* - z_\ell \geqslant F(z_*),$$

and it follows from (5) that

$$\|x[r] - x[r-1]\| \geqslant \frac{1}{\sqrt{2}}F(z_*). \tag{6}$$

But,

$$\|x[r] - x[r-1]\| = h\|v[r-1]\| \leqslant h\|v[0]\|,$$

the last since $\|v\|$ is decreasing for $t < T$. Putting this inequality together with (6) shows that

$$F(z_*) \leqslant \sqrt{2}h\|v[0]\|$$

or equivalently,

$$\left(\frac{1}{\boldsymbol{a}}\right)^{1/(\alpha-1)}\left(\frac{1}{\alpha}\right)^{1/(\alpha-1)}\frac{\alpha-1}{\alpha}-\boldsymbol{b}\leqslant\sqrt{2}h\big\|v[0]\big\|,$$

which contradicts our hypothesis. This shows that, for all $t<T$, $z(t)<z_\ell$ and hence, for all $t<\hat{t}$,

$$\big(1+\sqrt{2}\big\|x[t]\big\|\big)^\alpha\leqslant z_0^\alpha=\left(\big(1+\sqrt{2}\big\|x[0]\big\|\big)\frac{\alpha}{\alpha-1}\right)^\alpha\leqslant 2^{-\alpha}\frac{kK}{2\mathcal{H}_0},$$

the last by the definition of $\mathcal{H}_0$. We now proceed as in case (i). □

3.3. Proof of Theorem 1

Proposition 3. *For $\varepsilon>0$, let $p(\varepsilon)=\mathsf{Prob}\{\|H\|\leqslant\varepsilon\}$. Then, in the uniform case, we have the bound,*

$$p(\varepsilon)\geqslant\left(\frac{\varepsilon}{r}\right)^{3k}$$

($p(\varepsilon)=1$ if $r\leqslant\varepsilon$) while in the Gaussian case $\|H/\sigma\|^2$ has a Chi-square distribution with $3k-3$ degrees of freedom, and in consequence

$$p(\varepsilon)=\int_0^{\sqrt{\varepsilon/(2\sigma)}}\frac{t^{(3k-5)/2}}{\Gamma(\frac{3k-3}{2})}\mathrm{e}^{-t}\,\mathrm{d}t.$$

Proof. In the uniform case, since $\|H\|\leqslant\|\boldsymbol{H}\|$, we have:

$$\mathsf{Prob}\big(\|H\|\leqslant\varepsilon\big)\geqslant\mathsf{Prob}\big(\|\boldsymbol{H}\|\leqslant\varepsilon\big)=\left(\frac{\varepsilon}{r}\right)^{3k}.$$

In the Gaussian case, the decomposition $\boldsymbol{H}=\boldsymbol{H}_\Delta+\boldsymbol{H}_\perp$ takes the form $\boldsymbol{H}_\Delta=(\boldsymbol{m},\dots,\boldsymbol{m})$, with $\boldsymbol{m}=\frac{1}{k}\sum_{j=1}^k\boldsymbol{H}_j$, and $\boldsymbol{H}_\perp=\boldsymbol{H}-(\boldsymbol{m},\dots,\boldsymbol{m})$. Consequently

$$\|H\|^2=\|\boldsymbol{H}_\perp\|^2=\sum_{j=1}^k(\boldsymbol{H}_j-\boldsymbol{m})^2=\sum_{\ell=1}^3\sum_{j=1}^k\big(e_j^{(\ell)}-\boldsymbol{m}^{(\ell)}\big)^2,$$

where $\boldsymbol{m}=(\boldsymbol{m}^{(1)},\boldsymbol{m}^{(2)},\boldsymbol{m}^{(3)})$. A standard result in statistics (see [2, p. 219]) states that $\frac{1}{\sigma^2}\sum_{j=1}^k(e_j^{(\ell)}-\boldsymbol{m}^{(\ell)})^2$ has a Chi-square distribution with $k-1$ degrees of freedom. Therefore, by independence, $\|H/\sigma\|^2$ has a Chi-square distribution with $3k-3$ degrees of freedom. The expression for $p(\varepsilon)$ follows from the form of the density of the Chi-square random variable. □

We can now give the proof of Theorem 1.

Let $T>0$ and assume that the hypothesis of Theorem 2 holds for T. Then,

$$\big\|v[t]\big\|\leqslant\big\|v[0]\big\|\left(1-h\frac{kK}{2(1+\sqrt{2}B_0)^\alpha}\right)^t,$$

where B_0 is as in Theorem 2. Therefore

$$\begin{aligned}\big\|v[T]\big\|\leqslant\nu\quad&\Longleftarrow\quad\big\|v[0]\big\|\left(1-h\frac{kK}{2(1+\sqrt{2}B_0)^\alpha}\right)^T\leqslant\nu\\&\Longleftrightarrow\quad T\geqslant\left(\ln\left(1-h\frac{kK}{2(1+\sqrt{2}B_0)^\alpha}\right)\right)^{-1}\ln\left(\frac{\nu}{\|v[0]\|}\right)\\&\Longleftarrow\quad T\geqslant\frac{2(1+\sqrt{2}B_0)^\alpha}{hkK}\ln\left(\frac{\|v[0]\|}{\nu}\right)=T_0.\end{aligned}$$

At each iteration t, Theorem 2 requires that $\|H(t)\| \leqslant \mathcal{H}_0\|v[t]\|$ for a quantity $\mathcal{H}_0$ which depends on the initial conditions and on the case ((i), (ii), or (iii)) at hand. If nearly-alignment has not occurred, then $\|v[t]\| \geqslant \nu$ and therefore,

$$\mathsf{Prob}\{\|H(t)\| \leqslant \mathcal{H}_0\|v[t]\|\} \geqslant \mathsf{Prob}\{\|H(t)\| \leqslant \mathcal{H}_0\nu\} \geqslant p(\mathcal{H}_0\nu).$$

It follows that

$$\mathsf{Prob}\{\|H(t)\| \leqslant \mathcal{H}_0\|v[t]\| \text{ for } t = 0, \ldots, T_0 - 1\} \geqslant p(\mathcal{H}_0\nu)^{T_0}$$

and therefore the claimed bounds in the uniform and Gaussian cases.

4. Continuous time

The goal of this section is to show a continuous time version of Theorem 1. In contrast with the discrete time setting, though, the description of the noise $\boldsymbol{H}$ is not straightforward. There is no obvious continuous time version of a sequence of independent and identically distributed random variables. We thus begin by discussing our model of noise.

4.1. Continuous time stochastic processes

A *continuous time stochastic process* $\{X(t) \mid t \geqslant 0\}$, or for short a *stochastic process*, is a family of random variables $X(t)$ defined in a common probability space $(\Omega, \mathcal{F}, \mathsf{Prob})$. More precisely, the stochastic process depends on two arguments, t and ω, and when necessary we denote it by $X(t, \omega)$. In order to describe our assumptions, we make a brief comparison with the discrete time situation.

In the discrete time situation it is natural to assume that each perturbation $X(t)$ is a centered random variable (i.e., $\mathbf{E}(X(t)) = 0$), that for all values of t the random variables $X(t)$ are mutually independent, and that they have a common distribution. Making some additional assumption on the distribution of each random variable, typically assuming normality with a given standard deviation σ, completes the specification of the noise probabilistic structure. Let us call this structure a *Gaussian white noise sequence*.

Unfortunately, in continuous time we do not have a reasonable tractable model that shares the properties of the Gaussian white noise sequence. To understand why this is so we need to consider the properties of the *trajectories* of the processes, i.e., the curves obtained when $\omega \in \Omega$ is fixed and t ranges in the interval $[0, \infty)$. Assuming independence of the random variables of the stochastic process for close values of t makes the trajectories of the process to have an extremely irregular behaviour. Consequently, if we want the trajectories of the process to be continuous, or differentiable, we cannot assume the independence of the random variables for pairs of close values of t.

To analyze this difficulty consider first, in the discrete time case, the accumulated perturbation produced by a Gaussian white noise sequence $\{X(t) \mid t \in \mathbb{N}\}$. That is, consider the sequence of sums:

$$S(0) = 0, \qquad S(t) = X(1) + \cdots + X(t), \quad t = 1, 2, \ldots.$$

This random sequence associated to the Gaussian white noise sequence, called *Gaussian random walk*, does have a natural counterpart in the continuous time case, known as *Wiener process* or *Brownian motion*. A Wiener process $W = \{W(t) \mid t \geqslant 0\}$ is a continuous time stochastic process satisfying the following properties, that are natural extensions to continuous time of the properties of the sequence $\{S(t) \mid t \in \mathbb{N}\}$:

(a) $W(0) = 0$.
(b) The increments of the process are independent random variables. That is, for all $0 \leqslant t_0 < t_1 < \cdots < t_n$ the random variables,

$$W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1}),$$

are independent.
(c) The increments are homogeneous and Gaussian. That is, for all $t, h \geqslant 0$, the random variable,

$$W(t + h) - W(t),$$

has a centered Gaussian distribution with variance h.

(d) Finally, the trajectories of W are continuous. That is, the curves $\{W(\cdot,\omega)\}$ obtained when $\omega \in \Omega$ is fixed and $t \geqslant 0$ are continuous functions of t for almost all $\omega \in \Omega$.

In the discrete time case the Gaussian white noise sequence can be recovered from the sums $S(t)$ by taking differences $X(t) = S(t) - S(t-1)$. In the continuous time case we would like to have a formula like $X(t) = \dot{W}(t)$, but it is not immediate to give a sense to this last time derivative, as it is known that the trajectories of the Wiener process are nowhere differentiable [4]. A possible way out is to first take a differentiable approximation of the Wiener process and then take the time derivative of this approximation as a model of noise in the continuous time case. Another, alternative, way out is described in Remark 5 below.

The approximation is obtained by convolution with a smooth kernel. Let $\psi : \mathbb{R} \to \mathbb{R}_+$ be a $\mathscr{C}^1$ function, with compact support, say $\mathsf{supp}(\psi) \subset [-1/2, 1/2]$ and such that $\int_{\mathbb{R}} \psi(x)\,\mathrm{d}x = 1$. For $\delta > 0$ consider,

$$\psi_\delta(x) = \frac{1}{\delta}\psi\left(\frac{x}{\delta}\right),$$

that has $\mathsf{supp}(\psi_\delta) \subset [-\delta/2, \delta/2]$.

The approximation $W^\delta = \{W^\delta(t) \mid t \geqslant 0\}$ of the Wiener process is obtained by convolution with ψ_δ in the following way:

$$W^\delta(t) = (\psi_\delta * W)(t) = \int_{\mathbb{R}} \psi_\delta(t-s)W(s)\,\mathrm{d}s = \int_{\mathbb{R}} \psi_\delta(-w)W(t+w)\,\mathrm{d}w, \tag{7}$$

where, if $s < 0$, we replace $W(s)$ in the integrand by $\widehat{W}(-s)$, where $\widehat{W}$ is another Wiener process independent of W.

Observe that the process W^δ inherits the regularity properties of ψ. In particular, it has $\mathscr{C}^1$ trajectories. We now define a *noise process* X^δ as the time derivative of W^δ,

$$X^\delta(t) = \frac{\mathrm{d}}{\mathrm{d}t}W^\delta(t) = \int_{\mathbb{R}} \frac{\partial}{\partial t}\big(\psi_\delta(t-s)\big)W(s)\,\mathrm{d}s \tag{8}$$

$$= \frac{1}{\delta}\int_{\mathbb{R}} \dot{\psi}(-w)\big(W(t+\delta w) - W(t-\delta/2)\big)\,\mathrm{d}w \tag{9}$$

$$= \int_{\mathbb{R}} \psi_\delta(t-s)\,\mathrm{d}W(s). \tag{10}$$

Here the second equality in (8) is obtained by differentiation under the integral sign, (9) after a change of variables $s - t = \delta w$ and using that $\int_{\mathbb{R}} \dot{\psi}(-w)\,\mathrm{d}w = 0$, and (10) by integration by parts departing from (8) (note, this involves a stochastic integral [21]).

Using (10) and Itô's isometry for the stochastic integral [21, Theorem 4.2] we obtain (we write X instead of X^δ), for $h \geqslant 0$,

$$\mathbf{E}\big(X(t+h)X(t)\big) = \mathbf{E}\left(\int_{\mathbb{R}} \psi_\delta(t+h-s)\,\mathrm{d}W(s)\int_{\mathbb{R}} \psi_\delta(t-s)\,\mathrm{d}W(s)\right)$$

$$= \int_{\mathbb{R}} \psi_\delta(t+h-s)\psi_\delta(t-s)\,\mathrm{d}s. \tag{11}$$

Taking $h = 0$, we obtain:

$$\mathbf{E}\big(X(t)^2\big) = \int_{\mathbb{R}} \psi_\delta(t-s)\psi_\delta(t-s)\,\mathrm{d}s = \|\psi_\delta\|^2 = \frac{1}{\delta}\|\psi\|^2. \tag{12}$$

Taking $h \geqslant \delta$ in (11) and using that $\mathsf{supp}(\psi_\delta) \subset [-\delta/2, \delta/2]$ we obtain:

$$\mathbf{E}\big(X(t+h)X(t)\big) = 0. \tag{13}$$

Furthermore, from the different expressions in (8)–(10) we obtain the following properties of the process $\{X(t) \mid t \geqslant 0\}$:

(a) It is a *centered Gaussian* process with variance $\|\psi_\delta\|^2 = (1/\delta)\|\psi\|^2$. That is, $X(t) \sim N(0, (1/\delta)\|\psi\|^2)$ for all $t \in [0, T]$.

(b) It is a *stationary* process. That is, for all times $t_1, \dots, t_n$, intervals $I_1, \dots, I_n$, and time increment h, we have:

$$\mathsf{Prob}\big(X(t_1) \in I_1, \dots, X(t_n) \in I_n\big) = \mathsf{Prob}\big(X(t_1 + h) \in I_1, \dots, X(t_n + h) \in I_n\big).$$

(c) It is δ-*dependent*. That is, the two sets of random variables,

$$\big\{X(s) \bigm| s \leqslant t\big\} \quad \text{and} \quad \big\{X(s) \bigm| s \geqslant t + \delta\big\},$$

are independent for each t.

The first property can be obtained from (8) since the integral there is the limit of a linear combination of Gaussian random variables. Indeed, such a linear combination remains Gaussian and the limit of the resulting random variables preserves Gaussianity as well. This variable is centered since all the involved variables are centered, and the limit defining the integral preserves the expectation. The value for the variance follows from (12).

The stationarity property of $X(t)$ is inherited from the stationarity of the increments of the Wiener process. In order to see it we use the representation (9). The probability distribution of the stochastic process $\{W(t + w) - W(t - \delta/2) \mid -\delta/2 \leqslant w \leqslant \delta/2\}$ does not depend on the value of t, or more precisely, the probability distribution of $\{W(t_i + w) - W(t_i - \delta/2) \mid -\delta/2 \leqslant w \leqslant \delta/2, i = 1, \dots, n\}$ coincides with the probability distribution of $\{W(t_i + h + w) - W(t_i + h - \delta/2) \mid -\delta/2 \leqslant w \leqslant \delta/2, i = 1, \dots, n\}$. This makes the process probabilities invariant under a shift of h, i.e. the process satisfy the definition of stationarity in (b) above.

Finally, the δ-dependency is a consequence of (13). This equality give us non-correlation, when the lag $h \geqslant \delta$. The independence follows since a Gaussian vector without correlation has independent components. It should be noticed that this property is not essential to our development below; it simply mimics the discrete time independence. Furthermore, it is possible to derive similar results in this discrete time case for a weakly dependent noise.

Remark 3. The contents of this section is not new. It is exposed in certain detail for ease of the reader. Regarding the equalities in (8)–(10), it can be seen that any centered Gaussian process admits such a representation with an adequate kernel. General results on Gaussian processes and their diverse applications can be found for instance in [1,3,7].

4.2. Statement of the main result

We can now describe the noise $\boldsymbol{H}$ in (C) and state a continuous version of Theorem 1.

We assume that $\boldsymbol{H}_i(t)$ is a three-dimensional Gaussian centered, stationary stochastic process, that satisfies a δ-dependence condition for some $\delta > 0$, has $\mathscr{C}^1$ trajectories, and independent coordinates. More precisely, we assume that $\boldsymbol{H}_i(t) = (e_i^{(1)}(t), e_i^{(2)}(t), e_i^{(3)}(t))$, where each coordinate is given by:

$$e_i^{(\ell)}(t) = \sigma\sqrt{\delta} \int_{\mathbb{R}} \psi_\delta(t - s)\,\mathrm{d}W_i^{(\ell)}(s), \tag{14}$$

where ψ_δ is a kernel as in Section 4.1, $\sigma > 0$, and $\{W_i^{(\ell)}(t) \mid t \geqslant 0\}$ is a set of $3k$ independent Wiener processes. That is, each coordinate of $\boldsymbol{H}_i$ is of the form $\sigma\sqrt{\delta}X^\delta$ with X^δ as in Section 4.1. Note that the variance $\mathsf{Var}(e_i^{(\ell)}(t)) = \sigma^2$ for all $t \geqslant 0$.

Theorem 3. *Consider the system* (C) *with adjacency matrix given by* (3) *and noise given by* (14)*. Let $x_0, v_0 \in \mathbb{E}$. Then, there exists a unique solution $(x(t), v(t))$ of* (C)*, defined for all $t \in \mathbb{R}$, with initial conditions $x(0) = x_0$ and $v(0) = v_0$. Assume that one of the three following hypothesis holds*:

(i) $\alpha < 1$,
(ii) $\alpha = 1$, *and* $\|v(0)\| < \frac{kK}{2\sqrt{2}}$,
(iii) $\alpha > 1$, *and*

$$\left(\frac{1}{\alpha\boldsymbol{a}}\right)^{1/(\alpha-1)} \frac{\alpha - 1}{\alpha} > \boldsymbol{b}.$$

Then ν-nearly-alignment for some $\nu < \|v(0)\|$ occurs before time,

$$T_0 := \frac{U_0^\alpha}{kK} \ln\left(\frac{\|v(0)\|}{\nu}\right),$$

with probability at least,

$$\left\{2\boldsymbol{\Phi}\left(\frac{\nu\mathcal{H}_0}{\sigma\sqrt{3k}}\right) - \frac{2T_0\sigma\|\dot\psi\|}{\delta\sqrt{2\pi}}\varphi\left(\frac{\nu\mathcal{H}_0}{\sigma\sqrt{3k}}\right) - 1\right\}^{3k}, \tag{15}$$

where $\boldsymbol{\Phi}(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \mathrm{e}^{-u^2/2}\,\mathrm{d}u$ *is the standard normal distribution and* $\varphi = \boldsymbol{\Phi}'$ *its density.*

Remark 4. Using the identity,

$$1 - \boldsymbol{\Phi}(x) = \frac{\varphi(x)}{x}\left(1 + \mathcal{O}\left(\frac{1}{x^2}\right)\right),$$

and performing some elementary computations we obtain that the bound in (15) is equivalent, for small σ, to

$$1 - 6k\sigma\left(\frac{\sqrt{3k}}{\nu\mathcal{H}_0} + \frac{T_0\|\dot\psi\|}{\delta\sqrt{2\pi}}\right)\varphi\left(\frac{\nu\mathcal{H}_0}{\sigma\sqrt{3k}}\right).$$

Remark 5. An alternative way to model the noise in our system relies on the similarity of the Gaussian random walk and the Wiener process. Integrating the second equation in (C) we obtain:

$$\boldsymbol{v}(t) = \boldsymbol{v}(0) - \int_0^t L_x\boldsymbol{v}(s)\,\mathrm{d}s + \int_0^t \boldsymbol{H}(s)\,\mathrm{d}s.$$

The last term in the right-hand side is the accumulated noise for which we noted in Section 4.1 that the natural continuous time version is the Wiener process. Multiplying the latter by $\sigma > 0$ (as we did at the beginning of Section 4.2 to obtain $\mathsf{Var}(\boldsymbol{H}) = \sigma^2\,\mathrm{Id}_{3k}$) we obtain:

$$\boldsymbol{v}(t) = \boldsymbol{v}(0) - \int_0^t L_x\boldsymbol{v}(s)\,\mathrm{d}s + \sigma W(t),$$

and integral equation often written in its "differential form"

$$\mathrm{d}\boldsymbol{v}(t) = -L_x\boldsymbol{v}(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t).$$

Hence, an alternative to (C) is the system of stochastic differential equations:

$$\begin{aligned}\mathrm{d}\boldsymbol{x} &= \boldsymbol{v}\,\mathrm{d}t,\\ \mathrm{d}\boldsymbol{v} &= -L_x\boldsymbol{v}\,\mathrm{d}t + \sigma\,\mathrm{d}W.\end{aligned} \tag{SDE}$$

The construction of a solution for this system relies on the *stochastic calculus* developed by Itô [16].

We note that, while it is possible to prove that $W^\delta \to W$ when $\delta \to 0$, it is not generally true (cf. [19]) that the solution of a system of stochastic differential equations driven by a smoothed noise W^δ converges towards the solution of the corresponding system driven by the original noise W. Investigating whether this is the case for (C) and (SDE) would take us out of the scope of the present work.

To prove Theorem 3 we follow the steps in the proof of Theorem 1.

4.3. Bounded noise

For $x \in X$ we denote $\Gamma(x) = \|x\|^2$ and for $v \in V$ we denote $\Lambda(v) = \|v\|^2$.

In this section we fix $T \in (0,\infty) \cup \{+\infty\}$ and a solution (x, v) of (C) (which we assume exists and is, almost surely, differentiable in $[0, T)$). The meaning of expressions like ϕ_t, L_t, $\Lambda(t)$, or $\Gamma(t)$ is as described in Section 3.2.

Denote $\Phi_t = \min_{\tau\in[0,t]} \phi_\tau$.

Proposition 4. *Assume that, for all* $0 \leqslant t < T$, $\|H_t\| \leqslant \|v(t)\|\mathcal{H}_0$. *Then, for all* $0 \leqslant t < T$,

$$\|v(t)\| \leqslant \|v(0)\| \mathrm{e}^{-t(\Phi_t - \mathcal{H}_0)}.$$

Proof. Let $\tau \in [0, t]$. Then

$$\begin{aligned}
\Lambda'(\tau) &= \frac{\mathrm{d}}{\mathrm{d}\tau}\langle v(\tau), v(\tau)\rangle \\
&= 2\langle v'(\tau), v(\tau)\rangle \\
&= -2\langle L_\tau v(\tau) + H_\tau, v(\tau)\rangle \\
&= -2\langle L_\tau v(\tau), v(\tau)\rangle - 2\langle H_\tau, v(\tau)\rangle \\
&\leqslant -2\phi_{x(\tau)}\Lambda(\tau) + 2\|H_\tau\|\|v(\tau)\| \\
&\leqslant -2\Lambda(\tau)\big(\phi_{x(\tau)} - \mathcal{H}_0\big).
\end{aligned}$$

Here we used that L_τ is symmetric positive semidefinite on V. Using this inequality,

$$\ln\big(\Lambda(\tau)\big)\Big|_0^t = \int_0^t \frac{\Lambda'(\tau)}{\Lambda(\tau)}\,\mathrm{d}\tau \leqslant \int_0^t -2(\phi_\tau - \mathcal{H}_0)\,\mathrm{d}\tau \leqslant -2t(\Phi_t - \mathcal{H}_0),$$

i.e.,

$$\ln\big(\Lambda(t)\big) - \ln(\Lambda_0) \leqslant -2t(\Phi_t - \mathcal{H}_0),$$

from which the statement follows. □

Proposition 5. *Assume that* $\Phi_t > \mathcal{H}_0$ *for all* $0 \leqslant t < T$. *Then, for all* $0 \leqslant t < T$,

$$\|x(t)\| \leqslant \|x(0)\| + \frac{\|v(0)\|}{\Phi_t - \mathcal{H}_0}.$$

Proof. For $\tau \leqslant t$ we have $|\Gamma'(\tau)| = |2\langle v(\tau), x(\tau)\rangle| \leqslant 2\|v(\tau)\|\|x(\tau)\|$. But $\|x(\tau)\| = \Gamma(\tau)^{1/2}$ and $\|v(\tau)\|^2 = \Lambda(\tau) \leqslant \Lambda_0 \mathrm{e}^{-2\tau(\Phi_\tau - \mathcal{H}_0)}$, by Proposition 4. Therefore,

$$\Gamma'(\tau) \leqslant \big|\Gamma'(\tau)\big| \leqslant 2\big(\Lambda_0 \mathrm{e}^{-2\tau(\Phi_\tau - \mathcal{H}_0)}\big)^{1/2}\Gamma(\tau)^{1/2}, \tag{16}$$

and using that $\tau \mapsto \Phi_\tau$ is non-increasing and that $\Phi_\tau - \mathcal{H}_0 > 0$ for all $\tau \leqslant t$,

$$\begin{aligned}
\int_0^t \frac{\Gamma'(\tau)}{\Gamma(\tau)^{1/2}}\,\mathrm{d}\tau &\leqslant 2\int_0^t \big(\Lambda_0 \mathrm{e}^{-2\tau(\Phi_\tau - \mathcal{H}_0)}\big)^{1/2}\,\mathrm{d}\tau \\
&\leqslant 2\int_0^t \Lambda_0^{1/2}\mathrm{e}^{-\tau(\Phi_t - \mathcal{H}_0)}\,\mathrm{d}\tau \\
&= 2\Lambda_0^{1/2}\left(-\frac{1}{\Phi_t - \mathcal{H}_0}\right)\mathrm{e}^{-\tau(\Phi_t - \mathcal{H}_0)}\Big|_0^t \leqslant \frac{2\Lambda_0^{1/2}}{\Phi_t - \mathcal{H}_0};
\end{aligned}$$

the last inequality because $\Phi_t > \mathcal{H}_0$. This implies:

$$\Gamma(\tau)^{1/2}\big|_0^t = \frac{1}{2}\int_0^t \frac{\Gamma'(\tau)}{\Gamma(\tau)^{1/2}}\,\mathrm{d}\tau \leqslant \frac{\Lambda_0^{1/2}}{\Phi_t - \mathcal{H}_0},$$

from which it follows that

$$\Gamma(t)^{1/2} \leqslant \Gamma_0^{1/2} + \frac{\Lambda_0^{1/2}}{\Phi_t - \mathcal{H}_0}. \qquad \square$$

The main result in this section is the following:

Theorem 4. *Assume that, for all* $0 \leqslant t < T$, $\|H(t)\| \leqslant \|v(t)\|\mathcal{H}_0$. *Assume also that one of the three following hypothesis hold*:

(i) $\alpha < 1$,
(ii) $\alpha = 1$, *and* $\|v(0)\| < \frac{kK}{2\sqrt{2}}$,
(iii) $\alpha > 1$, *and*

$$\left(\alpha \frac{1}{\boldsymbol{a}}\right)^{1/(\alpha-1)} \frac{\alpha - 1}{\alpha} > \boldsymbol{b}.$$

Then, for all $0 \leqslant t < T$, $\|x(t)\| \leqslant B_0$, *and*

$$\|v(t)\| \leqslant \|v(0)\| \mathrm{e}^{-kK/U_0^\alpha t}.$$

In particular, when $T = \infty$, $\|v(t)\| \to 0$ *for* $t \to \infty$ *and there exists* $\hat{x} \in X$ *such that* $x(t) \to \hat{x}$ *when* $t \to \infty$.

Proof. Let

$$\Upsilon = \left\{ t \in [0, T) \;\middle|\; \left(1 + \sqrt{2}\|x(t)\|\right)^\alpha \leqslant \frac{kK}{2\mathcal{H}_0} \right\}.$$

Note that in all three cases ((i), (ii), and (iii)) the definition of $\mathcal{H}_0$ implies that $0 \in \Upsilon$ and hence, that $\Upsilon \neq \emptyset$. Assume that $\Upsilon \neq [0, T)$ and let $\hat{t} = \inf\{[0, T) \setminus \Upsilon\}$. Clearly, $1 + \sqrt{2}\|x(\hat{t})\| = \frac{kK}{2\mathcal{H}_0}$.

By Proposition 1 and Lemma 1, for all $x \in X$,

$$\phi_x \geqslant \frac{kK}{(1 + \max_{i \neq j} \|x_i - x_j\|)^\alpha} \geqslant \frac{kK}{(1 + \sqrt{2}\|x\|)^\alpha}.$$

Let $t < \hat{t}$ and $t^* \in [0, t]$ be the point maximizing $\|x\|$ in $[0, t]$. Then

$$\Phi_t = \min_{\tau \in [0,t]} \phi_\tau \geqslant \min_{\tau \in [0,t]} \frac{kK}{(1 + \sqrt{2}\|x(\tau)\|)^\alpha} \geqslant \frac{kK}{(1 + \sqrt{2}\|x(t^*)\|)^\alpha}.$$

Moreover, since $t^* \leqslant t < \hat{t}$, $t^* \in \Upsilon$ and we have:

$$\Phi_t - \mathcal{H}_0 \geqslant \frac{kK}{(1 + \sqrt{2}\|x(t^*)\|)^\alpha} - \mathcal{H}_0 \geqslant \frac{kK}{2(1 + \sqrt{2}\|x(t^*)\|)^\alpha} > 0. \tag{17}$$

Hence, we may apply Proposition 5 to obtain:

$$\begin{aligned} \|x(t)\| &\leqslant \|x(0)\| + \|v(0)\| \frac{1}{\Phi_t - \mathcal{H}_0} \\ &\leqslant \|x(0)\| + \frac{2\|v(0)\|(1 + \sqrt{2}\|x(t^*)\|)^\alpha}{kK}. \end{aligned} \tag{18}$$

Since t^* maximizes Γ in $[0, t]$ it also does so in $[0, t^*]$. Thus, for $t = t^*$, (18) takes the form,

$$\left(1 + \sqrt{2}\|x(t^*)\|\right) - 2\sqrt{2}\|v(0)\| \frac{(1 + \sqrt{2}\|x(t^*)\|)^\alpha}{kK} - \left(1 + \sqrt{2}\|x(0)\|\right) \leqslant 0. \tag{19}$$

Let $z = 1 + \sqrt{2}\|x(t^*)\|$. Then (19) can be rewritten as $F(z) \leqslant 0$ with $F(z) = z - \boldsymbol{a}z^\alpha - \boldsymbol{b}$. One can now finish the proof by dividing in cases as in Theorem 4 and following the steps in its proof. □

4.4. Proof of Theorem 3

We begin with a result on the behaviour of the maximum of the processes $e_i^{(\ell)}$ described in Section 4.2.

Proposition 6. *Fix $T > 0$. Denote $e(t) = e_i^{(\ell)}(t)$ and*

$$p(x) = \mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T}\big|e(t)\big| < x\Big).$$

Then

$$1 - p(x) = \mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T}\big|e(t)\big| \geqslant x\Big) \leqslant 2\bigg[\frac{T\sigma\|\dot{\psi}\|}{\delta\sqrt{2\pi}}\varphi\Big(\frac{x}{\sigma}\Big) + 1 - \boldsymbol{\Phi}\Big(\frac{x}{\sigma}\Big)\bigg].$$

Proof. The proof is an application of Davies's inequality [12] (see also Chapter 4 in [3]):

$$\mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T} e(t) \geqslant x\Big) \leqslant \frac{1}{\sqrt{2\pi}}\varphi\Big(\frac{x}{\sigma}\Big)\int_0^T \sqrt{r_{11}(t,t)}\,\mathrm{d}t + 1 - \boldsymbol{\Phi}\Big(\frac{x}{\sigma}\Big),$$

where $r_{11}(t,t) = \frac{\partial^2}{\partial t^2}\mathbf{E}(e(t)^2)$ and therefore, using (11),

$$r_{11}(t,t) = \frac{\sigma^2}{\delta^2}\|\dot{\psi}\|^2.$$

The conclusion now follows from the trivial bound:

$$\mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T}\big|e(t)\big| \geqslant x\Big) \leqslant 2\,\mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T} e(t) \geqslant x\Big). \qquad \square$$

We can now give the proof of Theorem 3. The existence of a unique solution follows, for each $\omega \in \Omega$, from [15, Chapter 8].

Using that $\|\boldsymbol{H}(t)\| \leqslant \sqrt{3k}\|\boldsymbol{H}(t)\|_\infty$, and Proposition 6 for one coordinate we obtain:

$$\begin{aligned}
\mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T}\big\|H(t)\big\| < \varepsilon\Big) &\geqslant \mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T}\big\|\boldsymbol{H}(t)\big\| < \varepsilon\Big)\\
&\geqslant \mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T}\max_{1\leqslant j\leqslant k}\max_{1\leqslant \ell\leqslant 3}\big|e_i^{(\ell)}(t)\big| < \frac{\varepsilon}{\sqrt{3k}}\Big)\\
&= \mathsf{Prob}\Big(\max_{0\leqslant t\leqslant T}\big|e(t)\big| < \frac{\varepsilon}{\sqrt{3k}}\Big)^{3k}\\
&\geqslant \bigg\{1 - 2\bigg[\frac{T\sigma\|\dot{\psi}\|}{\delta\sqrt{2\pi}}\varphi\Big(\frac{\varepsilon}{\sigma\sqrt{3k}}\Big) + 1 - \boldsymbol{\Phi}\Big(\frac{\varepsilon}{\sigma\sqrt{3k}}\Big)\bigg]\bigg\}^{3k}\\
&= \bigg\{2\boldsymbol{\Phi}\Big(\frac{\varepsilon}{\sigma\sqrt{3k}}\Big) - \frac{2T\sigma\|\dot{\psi}\|}{\delta\sqrt{2\pi}}\varphi\Big(\frac{\varepsilon}{\sigma\sqrt{3k}}\Big) - 1\bigg\}^{3k}.
\end{aligned}$$

Similarly as in the proof of Theorem 1, but taking into account that now $v(t)$ is a continuous function we define:

$$T(\omega) = \inf\big\{t \geqslant 0 \;\big|\; \big\|v(t)\big\| \leqslant \nu\big\}, \qquad T_0 = \frac{U_0^\alpha}{kK}\ln\Big(\frac{\|v(0)\|}{\nu}\Big).$$

Taking $\varepsilon = \nu\mathcal{H}_0$ we now see that $T(\omega) \leqslant T_0$ with probability at least,

$$\bigg\{2\boldsymbol{\Phi}\Big(\frac{\nu\mathcal{H}_0}{\sigma\sqrt{3k}}\Big) - \frac{2T_0\sigma\|\dot{\psi}\|}{\delta\sqrt{2\pi}}\varphi\Big(\frac{\nu\mathcal{H}_0}{\sigma\sqrt{3k}}\Big) - 1\bigg\}^{3k}. \tag{20}$$

Let us then take ω in the set $\max_{0\leqslant t\leqslant T_0}\|H(t)\| \leqslant \nu\mathcal{H}_0$. By our previous computation, this set has a probability not smaller than the bound (20). If $T(\omega) > T_0$, then $\|v(t)\| > \nu$ on the interval $[0, T_0]$, Theorem 4 holds for $T = T_0$, and we obtain $\|v(t)\| \leqslant \nu$ for some $t \leqslant T_0$, obtaining a contradiction. This concludes the proof.

Remark 6. The value $\alpha = 1$ at which the behavior of the system changes is sharp. Indeed, in [9, Section IV] the situation with two birds flying on a line and with no noise, i.e., taking $\mathbb{E} = \mathbb{R}$ (instead of $\mathbb{R}^3$), $k = 2$, and $\boldsymbol{H} = 0$ is considered. In this case, the corresponding system (C) can be explicitly solved and it is shown that, for all $\alpha > 1$ there exist initial states leading to dispersion. That is, initial states for which there is no convergence to flocking.

References

[1] R.J. Adler, An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes, IMS, Hayward, CA, 1990.
[2] S.F. Arnold, Mathematical Statistics, Prentice Hall, Englewood Cliffs, N.J., 1990.
[3] J.M. Azaïs, M. Wschebor, Level Sets and Extrema of Random Processes and Fields, book in preparation.
[4] P. Billingsley, Probability and Measure, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995. A Wiley-Interscience Publication.
[5] J. Buhl, D.J.T. Sumpter, I.D. Couzin, J.J. Hale, E. Despland, E.R. Miller, S.J. Simpson, From disorder to order in marching locusts, Science 312 (2006) 1402–1406.
[6] Y.L. Chuang, Y.R. Huang, M.R. D'Orsogna, A.L. Bertozzi, Multi-vehicle flocking: Scalability of cooperative control algorithms using pairwise potentials, preprint, 2006.
[7] H. Cramér, M.R. Leadbetter, Stationary and Related Stochastic Processes, Dover, 2004, reprint.
[8] F. Cucker, S. Smale, Best choices for regularization parameters in learning theory, Found. Comput. Math. 2 (2002) 413–428.
[9] F. Cucker, S. Smale, Emergent behavior in flocks, IEEE Trans. on Autom. Control 52 (2007) 852–862.
[10] F. Cucker, S. Smale, On the mathematics of emergence, Japan J. Math. 2 (2007) 197–227.
[11] F. Cucker, S. Smale, D.X. Zhou, Modeling language evolution, Found. Comput. Math. 4 (2004) 315–343.
[12] R.B. Davies, Hypothesis testing when a nuisance parameter is present only under the alternative, Biometrika 64 (1977) 247–254.
[13] G. Flierl, D. Grünbaum, S. Levin, D. Olson, From individuals to aggregations: the interplay between behavior and physics, J. Theor. Biol. 196 (1999) 397–454.
[14] N. Higham, Accuracy and Stability of Numerical Algorithms, SIAM, 1996.
[15] M. Hirsch, S. Smale, Differential Equations, Dynamical Systems, and Linear Algebra, Pure and Applied Mathematics, vol. 60, Academic Press, 1974.
[16] K. Itô, Essentials of Stochastic Processes, Translations of Mathematical Monographs, vol. 231, AMS, Providence, RI, 2006. Translated from the 1957 Japanese original by Yuji Ito.
[17] A. Jadbabaie, J. Lin, A.S. Morse, Coordination of groups of mobile autonomous agents using nearest neighbor rules, IEEE Trans. on Autom. Control 48 (2003) 988–1001.
[18] J. Ke, J. Minett, C.-P. Au, W.S.-Y. Wang, Self-organization and selection in the emergence of vocabulary, Complexity 7 (2002) 41–54.
[19] T.G. Kurtz, P. Protter, Weak limit theorems for stochastic integrals and stochastic differential equations, Ann. Probab. 19 (1991) 1035–1070.
[20] P. Niyogi, The Computational Nature of Language Learning and Evolution, The MIT Press, 2006.
[21] B. Oksendal, Stochastic Differential Equations, Universitext, sixth ed., Springer-Verlag, Berlin, 2003.
[22] J. Shen, Cucker–Smale flocking under hierarchical leadership, preprint, available at http://www.math.umn.edu/~jhshen, 2007.
[23] H.G. Tanner, A. Jadbabaie, G.J. Pappas, Stable flocking of mobile agents, Part I: Fixed topology, and Part II: Dynamic topology, in: Proceedings of the 42nd IEEE Conference on Decision and Control, 2003, pp. 2010–2021.
[24] C.M. Topaz, A.L. Bertozzi, M.A. Lewis, A nonlocal continuum model for biological aggregation, Bulletin of Mathematical Biology 68 (2006) 1601–1623.
[25] J.N. Tsitsiklis, Problems in decentralized decision making and computation, PhD thesis, Department of EECS, MIT, 1984.
[26] J.N. Tsitsiklis, D.P. Bertsekas, M. Athans, Distributed asynchronous deterministic and stochastic gradient optimization algorithms, IEEE Trans. on Autom. Control 31 (1986) 803–812.
[27] T. Vicsek, A. Czirók, E. Ben-Jacob, O. Shochet, Novel type of phase transition in a system of self-driven particles, Phys. Rev. Lett. 75 (1995) 1226–1229.