PARTICLE SYSTEMS AND KINETIC EQUATIONS MODELING INTERACTING AGENTS IN HIGH DIMENSION

M. FORNASIER[†‡], J. HAŠKOVEC[‡], AND J. VYBÍRAL[‡]

Abstract. In this paper we explore how concepts of high-dimensional data compression via random projections onto lower-dimensional spaces can be applied for tractable simulation of certain dynamical systems modeling complex interactions. In such systems, one has to deal with a large number of agents (typically millions) in spaces of parameters describing each agent of high dimension (thousands or more). Even with todays powerful computers, numerical simulations of such systems are prohibitively expensive. We propose an approach for the simulation of dynamical systems governed by functions of *adjacency matrices* in high dimension, by random projections via Johnson-Lindenstrauss embeddings, and recovery by compressed sensing techniques. We show how these concepts can be generalized to work for associated kinetic equations, by addressing the phenomenon of the delayed curse of dimension, known in information-based complexity for optimal numerical integration problems in high dimensions.

Key words. Dimensionality reduction, dynamical systems, flocking and swarming, Johnson-Lindenstrauss embedding, compressed sensing, high-dimensional kinetic equations, delayed curse of dimension, optimal integration of measures in high dimension.

AMS subject classifications.

1. Introduction. The dimensionality scale of problems arising in our modern information society became very large and finding appropriate methods for dealing with them is one of the great challenges of today's numerical simulation. The most notable recent advances in data analysis are based on the observation that in many situations, even for very complex phenomena, the intrinsic dimensionality of the data is significantly lower than the ambient dimension. Remarkable progresses have been made in data compression, processing, and acquisition. We mention, for instance, the use of *diffusion maps* for data clouds and graphs in high dimension [5, 6, 17, 18, 19] in order to define low-dimensional local representations of data with small distance distortion, and meaningful automatic clustering properties. In this setting the embedding of data is performed by a *highly nonlinear* procedure, obtained by computing the eigenfunctions of suitable normalized diffusion kernels, measuring the probability of transition from one data point to another over the graph.

Quasi-isometrical *linear* embeddings of high-dimensional point clouds into low-dimensional spaces of parameters are provided by the well-known Johnson-Lindenstrauss Lemma [1, 22, 34]: any cloud of $\mathcal{N}$ points in $\mathbb{R}^d$ can be embedded by a random linear projection M nearly isometrically into $\mathbb{R}^k$ with $k = \mathcal{O}(\varepsilon^{-2}\log(\mathcal{N}))$ (a precise statement will be given below). This embedding strategy is simpler than the use of diffusion maps, as it is linear, however it is "blind" to the specific geometry and local dimensionality of the data, as the embedding dimension k depends exclusively on the number of points in the cloud. In many applications, this is sufficient, as the number of points $\mathcal{N}$ is supposed to be a power of the dimension d, and the embedding produces an effective reduction to $k = \mathcal{O}(\varepsilon^{-2}\log(\mathcal{N})) = \mathcal{O}(\varepsilon^{-2}\log(d))$ dimensions. As clarified in [3, 36], the Johnson-Lindenstrauss Lemma is also at the basis of the possibility

[†]Faculty of Mathematics, Technical University Munich, Boltzmannstr. 3, D-85748 Garching, Germany

[‡]Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstrasse 69, A-4040 Linz, Austria

of performing optimal compressed and nonadaptive acquisition of high-dimensional data. In *compressed sensing* [12, 24, 27] a vector $x \in \mathbb{R}^d$ is encoded in a vector $y \in \mathbb{R}^k$ by applying a random projection M, which is modeling a linear acquisition device with random sensors, i.e., $y = Mx$. From y it is possible to decode x approximatively (see Theorem 3.7 below) by solving the convex optimization problem

$$x^\# = \arg \min_{Mz=y} \left(\|z\|_{\ell_1^d} := \sum_{i=1}^d |z_i| \right),$$

with the error distortion

$$\|x^\# - x\|_{\ell_1^d} \leq C\sigma_K(x)_{\ell_1^d},$$

where $\sigma_K(x)_{\ell_1^d} = \inf_{z:\#\text{supp }(z)\leq K} \|z - x\|_{\ell_1^d}$ and $K = \mathcal{O}(k/(\log(d/k) + 1))$. We denote $\Sigma_K = \{z \in \mathbb{R}^d : \#\text{supp }(z) \leq K\}$ the set of K-sparse vectors, i.e., the union of K-dimensional coordinate subspaces in $\mathbb{R}^d$. In particular, if $x \in \Sigma_K$, then $x^\# = x$. Hence, not only M is a Johnson-Lindenstrauss embedding, quasi-isometrical on point clouds and K-dimensional coordinate subspaces, but also allows for the recovery of the most relevant components of high-dimensional vectors, from low-dimensional encoded information. A recent work [4, 47] extends the quasi-isometrical properties of the Johnson-Lindenstrauss embedding from point clouds and K-dimensional coordinate subspaces to smooth compact Riemannian manifolds with bounded curvature.

Besides these relevant results in compressing and coding-deconding high-dimensional "stationary" data, dimensionality reduction of complex dynamical systems and high-dimensional partial differential equations is a subject of recent intensive research. Several tools have been employed, for instance, the use of diffusion maps for dynamical systems [38], tensor product bases and sparse grids for the numerical solution of linear high-dimensional PDEs [23, 10, 29, 30], the reduced basis method for solving high-dimensional parametric PDEs [7, 9, 37, 42, 43, 45].

In this paper we shall further explore the connection between data compression and *tractable* numerical simulation of dynamical systems, and solutions of associated high-dimensional kinetic equations. We are specially interested in dynamical systems of the type

$$\dot{x}_i(t) = f_i(\mathcal{D}x(t)) + \sum_{j=1}^N f_{ij}(\mathcal{D}x(t))x_j(t), \tag{1.1}$$

where we use the following notation:

- $N \in \mathbb{N}$ - number of agents,
- $x(t) = (x_1(t), \ldots, x_N(t)) \in \mathbb{R}^{d\times N}$, where $x_i : [0,T] \to \mathbb{R}^d$, $i = 1, \ldots, N$,
- $f_i : \mathbb{R}^{N\times N} \to \mathbb{R}^d, \quad i = 1, \ldots, N$,
- $f_{ij} : \mathbb{R}^{N\times N} \to \mathbb{R}, \quad i, j = 1, \ldots, N$,
- $\mathcal{D} : \mathbb{R}^{d\times N} \to \mathbb{R}^{N\times N}$, $\mathcal{D}x := (\|x_i - x_j\|_{\ell_2^d})_{i,j=1}^N$ is the *adjacency matrix* of the point cloud x.

We shall assume that the governing functions f_i and f_{ij} are Lipschitz, but we shall specify the details later on. The system (1.1) describes the dynamics of multiple complex agents $x(t) = (x_1(t), \ldots, x_N(t)) \in \mathbb{R}^{d\times N}$, interacting on the basis of their mutual "social" distance $\mathcal{D}x(t)$, and its general form includes several models for swarming and collective motion of animals and micro-organisms, aggregation of cells, etc. Several relevant effects can be included in the model by means of the functions f_i and f_{ij}, in particular, fundamental binary mechanisms of *attraction, repulsion, aggregation* and *alignment* [13, 14, 20, 21, 40, 35]. Moreover, possibly adding stochastic terms of random noise may also allow to consider *diffusion* effects [8, 14]. However, these models and motion mechanisms are mostly derived borrowing a leaf from physics, by assuming the agents (animals, micro-organisms, cells etc.) as pointlike and exclusively determined by their spatial position and velocity in $\mathbb{R}^{3+3}$. In case we wished to extend such models of social interaction to more "sophisticated" agents, described by many parameters ($d \gg 3+3$), the simulation may become computationally prohibitive. Our motivation for considering high-dimensional situations stems from the modern development of communication technology and Internet, for which we witness the development of larger and larger communities accessing information (interactive databases), services (financial market), social interactions (social networks) etc. For instance, we might be interested to simulate the behavior of certain subsets of the financial market where the agents are many investors, who are characterized by their portfolios of several hundreds of investments. The behavior of each individual investor depends on the dynamics of others according to a suitable social distance determined by similar investments. Being able of producing meaningful simulations and learning processes of such complex dynamics is an issue, which might be challenged by using suitable compression/dimensionality reduction techniques.

The idea we develop in this paper is to project randomly the system and its initial condition by Johnson-Lindenstrauss embeddings to a lower-dimensional space where an independent simulation can be performed with significantly reduced complexity. We shall show that the use of multiple projections and parallel computations allows for an approximative reconstruction of the high-dimensional dynamics, by means of compressed sensing techniques. After we explored the tractable simulation of the dynamical systems (1.1) when the dimension d of the parameter space is large, we also address the issue of whether we can perform tractable simulations when also the number N of agents is getting very large. Unlike the control of a finite number of agents, the numerical simulation of a rather large population of interacting agents ($N \gg 0$) can constitute a serious difficulty which stems from the accurate solution of a possibly very large system of ODEs. Borrowing the strategy from the kinetic theory of gases, we want instead to consider a density distribution of agents, depending on their d-parameters, which interact with stochastic influence (corresponding to classical collisional rules in kinetic theory of gases) – in this case the influence is "smeared" since two individuals may interact also when they are far apart in terms of their "social distance" $\mathcal{D}x$. Hence, instead of simulating the behavior of each individual agent, we shall describe the collective behavior encoded by a density distribution μ, whose evolution is governed by one sole mesoscopic partial differential equation. We shall show that, under realistic assumptions on the concentration of the measure μ on sets of lower dimension, we can also acquire information on the properties of the high-dimensional measure solution μ of the corresponding kinetic equation, by considering random projections to lower dimension, which is determined according to a combination of optimal integration principles for the high-dimensional measure μ and the results previously achieved for particle dynamical systems.

1.1. Fundamental assumptions. We introduce the following notation for ℓ_p-norms of vectors $v \in \mathbb{R}^d$,

$$\|v\|_{\ell_p^d} := \left(\sum_{i=1}^{d} |v_i|^p\right)^{1/p} \qquad \text{for } 1 \le p < \infty,$$

and

$$\|v\|_{\ell_\infty^d} := \max_{i=1,\dots,d} |v_i|.$$

For matrices $x \in \mathbb{R}^{n\times m}$ we consider the mixed norm

$$\|x\|_{\ell_p^m(\ell_q^n)} := \|(\|x_i\|_{\ell_p^n})_{i=1}^m\|_{\ell_q^m},$$

where $x_i \in \mathbb{R}^n$ is the i^{th}-column of the matrix x.

For the rest of the paper we impose three fundamental assumptions about Lipschitz and boundedness properties of f_i and f_{ij},

$$|f_i(a) - f_i(b)| \le L\|a-b\|_{\ell_\infty^N(\ell_\infty^N)}, \quad i = 1,\dots,N \tag{1.2}$$

$$\max_{i=1,\dots,N} \sum_{j=1}^{N} |f_{ij}(a)| \le L', \tag{1.3}$$

$$\max_{i=1,\dots,N} \sum_{j=1}^{N} |f_{ij}(a) - f_{ij}(b)| \le L''\|a-b\|_{\ell_\infty^N(\ell_\infty^N)}, \tag{1.4}$$

for every $a, b \in \mathbb{R}^{N\times N}$. Unfortunately, models of real-life phenomena would not always satisfy these conditions, for instance models of financial markets or socio-economic interactions can be expected to exhibit severely discontinuous behavior. However, these assumptions are reasonable in certain regimes and allow us to prove the concept we are going to convey in this paper, i.e., the possibility of simulating high-dimensional dynamics by multiple independent simulations in low dimension.

1.2. Euler scheme, a classical result of stability and convergence, and its complexity. We shall consider the system of ordinary differential equations of the form (1.1) with the initial condition

$$x_i(0) = x_i^0, \qquad i = 1,\dots,N. \tag{1.5}$$

The Euler method for this system is given by (1.5) and

$$x_i^{n+1} := x_i^n + h\left[f_i(\mathcal{D}x^n) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}x^n)x_j^n\right], \quad n = 0,\dots,n_0-1. \tag{1.6}$$

where $h > 0$ is the time step and $n_0 := T/h$ is the number of iterations. We consider here the *explicit* Euler scheme exclusively for the sake of simplicity, for more sophisticated integration methods might be used. We start with a classical result, which we report in detail for the sake of the reader, and for simplicity we assume $f_{ij} = 0$ for all $i, j = 1, \dots N$.

THEOREM 1.1 (Stability and convergence of the Euler scheme). *Fix $x^0 \in \mathbb{R}^{d\times N}$ and let $x(t)$ be the unique solution of the ODE*

$$\dot{x}(t) = f(\mathcal{D}x(t))\,, \qquad x(0) = x^0\,, \tag{1.7}$$

on the interval $[0,T]$, $T > 0$, for $f = (f_i)_{i=1}^N$ satisfying (1.2). Moreover, fix $h > 0$ and let $t^n := nh$ and $\tilde{x}^n$ be the approximate solution obtained by the explicit Euler method, i.e.,

$$\tilde{x}^{n+1} = \tilde{x}^n + hf(\mathcal{D}\tilde{x}^n)\,, \qquad \tilde{x}_0 = \tilde{x}^0\,,$$

for $n = 0,\dots,n_0-1$. Note that we allow different initial conditions x^0 and $\tilde{x}^0$ for the continuous and, resp., discrete solutions. Then, we have the error estimate

$$\mathcal{E}^n \le \exp(2Lt^n)\left(\mathcal{E}^0 + ht^n\frac{\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}}{2}\right)\,,$$

where $\mathcal{E}^n = \|x(t^n) - \tilde{x}^n\|_{\ell_\infty^N(\ell_2^d)}$.

Proof. For the sake of the proof, we extend $\tilde{x}$ to the full interval $[0,T]$ by linear interpolation between the grid points t^n, i.e.,

$$\tilde{x}(t^n + s) = \tilde{x}(t^n) + sf(\mathcal{D}\tilde{x}(t^n)) \qquad \text{for } s \in [0,h]\,,$$

such that $\tilde{x}$ is a continuous, piecewise linear function on $[0,T]$.

For a fixed n and $t := t^n$, let us consider the exact and approximate solutions in the interval $[t, t+\tau]$ with $\tau \in [0,h]$:

$$x(t+\tau) = x(t) + \int_0^\tau f(\mathcal{D}x(t+s))\,\mathrm{d}s\,, \tag{1.8}$$

$$\tilde{x}(t+\tau) = \tilde{x}(t) + \int_0^\tau f(\mathcal{D}\tilde{x}(t))\,\mathrm{d}s\,. \tag{1.9}$$

Subtracting (1.9) from (1.8) and using (1.2), we obtain

$$\begin{aligned}
\|x(t+\tau) - \tilde{x}(t+\tau)\|_{\ell_\infty^N(\ell_2^d)} &\le \|x(t) - \tilde{x}(t)\|_{\ell_\infty^N(\ell_2^d)} + \int_0^\tau \|f(\mathcal{D}x(t+s)) - f(\mathcal{D}\tilde{x}(t))\|_{\ell_\infty^N(\ell_2^d)}\,\mathrm{d}s \\
&\le \|x(t) - \tilde{x}(t)\|_{\ell_\infty^N(\ell_2^d)} + L\int_0^\tau \|\mathcal{D}x(t+s) - \mathcal{D}\tilde{x}(t)\|_{\ell_\infty^N(\ell_\infty^N)}\,\mathrm{d}s \\
&\le \|x(t) - \tilde{x}(t)\|_{\ell_\infty^N(\ell_2^d)} + 2L\int_0^\tau \|x(t+s) - \tilde{x}(t)\|_{\ell_\infty^N(\ell_2^d)}\,\mathrm{d}s\,.
\end{aligned}$$

Moreover, for $s \in [0,h]$,

$$\begin{aligned}
\|x(t+s) - \tilde{x}(t)\|_{\ell_\infty^N(\ell_2^d)} &\le \|x(t+s) - \tilde{x}(t+s)\|_{\ell_\infty^N(\ell_2^d)} + \|\tilde{x}(t+s) - \tilde{x}(t)\|_{\ell_\infty^N(\ell_2^d)} \\
&= \|x(t+s) - \tilde{x}(t+s)\|_{\ell_\infty^N(\ell_2^d)} + s\|f(\mathcal{D}\tilde{x}(t))\|_{\ell_\infty^N(\ell_2^d)}\,.
\end{aligned}$$

The term $\|f(\mathcal{D}\tilde{x}(t))\|_{\ell_\infty^N(\ell_2^d)} = \|f(\mathcal{D}\tilde{x}^n)\|_{\ell_\infty^N(\ell_2^d)}$ is bounded by $(1+2Lh)^n\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}$, which can be seen from the simple induction

$$\begin{aligned}
\|f(\mathcal{D}\tilde{x}^n)\|_{\ell_\infty^N(\ell_2^d)} &\le \|f(\mathcal{D}\tilde{x}^n) - f(\mathcal{D}\tilde{x}_{n-1})\|_{\ell_\infty^N(\ell_2^d)} + \|f(\mathcal{D}\tilde{x}_{n-1})\|_{\ell_\infty^N(\ell_2^d)} \\
&\le L\|\mathcal{D}\tilde{x}_n - \mathcal{D}\tilde{x}_{n-1}\|_{\ell_\infty^N(\ell_\infty^N)} + \|f(\mathcal{D}\tilde{x}_{n-1})\|_{\ell_\infty^N(\ell_2^d)} \\
&\le 2L\|\tilde{x}_n - \tilde{x}_{n-1}\|_{\ell_\infty^N(\ell_2^d)} + \|f(\mathcal{D}\tilde{x}_{n-1})\|_{\ell_\infty^N(\ell_2^d)} \\
&= (1+2Lh)\|f(\mathcal{D}\tilde{x}_{n-1})\|_{\ell_\infty^N(\ell_2^d)} \le (1+2Lh)^n\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}\,.
\end{aligned}$$

Consequently, defining $\mathcal{E}(t+\tau) := \|x(t+\tau) - \tilde{x}(t+\tau)\|_{\ell_\infty^N(\ell_2^d)}$, we obtain

$$\begin{aligned}\mathcal{E}(t+\tau) &\leq \mathcal{E}(t) + 2L\int_0^\tau \left(\mathcal{E}(t+s) + s(1+2Lh)^n\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}\right)\,\mathrm{d}s\\ &\leq \mathcal{E}(t) + 2L\int_0^\tau \mathcal{E}(t+s)\,\mathrm{d}s + \frac{h^2}{2}(1+2Lh)^n\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}\,.\end{aligned}$$

An application of the Gronwall lemma yields

$$\mathcal{E}(t+h) \leq \left(\mathcal{E}(t) + \frac{h^2}{2}(1+2Lh)^n\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}\right)\exp(2Lh)\,.$$

By another simple induction we obtain

$$\mathcal{E}^n \leq \exp(2Lnh)\mathcal{E}^0 + \left(\sum_{k=1}^n \exp(2Lkh)(1+2Lh)^{n-k}\right)\frac{h^2}{2}\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}\,,$$

where we turned back to the notation $\mathcal{E}^n = \mathcal{E}(t^n)$. Using $(1+2Lh)^{n-k} \leq \exp(2Lh(n-k))$, we have

$$\mathcal{E}^n \leq \exp(2Lnh)\mathcal{E}^0 + \exp(2Lnh)n\frac{h^2}{2}\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}\,,$$

and, finally, writing t^n for nh, we conclude

$$\mathcal{E}^n \leq \exp(2Lt^n)\left(\mathcal{E}^0 + ht^n\frac{\|f(\mathcal{D}\tilde{x}^0)\|_{\ell_\infty^N(\ell_2^d)}}{2}\right)\,.$$

□

The simulation of the dynamical system (1.7) has a complexity which is at least the one of computing the *adjacency matrix* $\mathcal{D}\tilde{x}^n$ at each discrete time t^n, i.e., $\mathcal{O}(d\times N^2)$. The scope of the next sections is to show that, up to an ε-distortion, we can approximate the dynamics of (1.1) by projecting the system into lower dimension and by executing in parallel computations with reduced complexity. Computation of the adjacency matrix in the new dimension requires only $\mathcal{O}(\varepsilon^{-2}\log(N)\times N^2)$ operations. Especially if the distortion parameter $\varepsilon > 0$ is not too small and the number of agents is of a polynomial order in d, we reduce the complexity of computing the adjacency matrix to $\mathcal{O}(\log(d)\times N^2)$.

2. Projecting the Euler method: dimensionality reduction of discrete dynamical systems.

2.1. Johnson-Lindenstrauss embedding. We wish to project the dynamics of (1.1) into a lower-dimensional space by employing a well-known result of Johnson and Lindenstrauss [34], which we informally rephrase for our purposes as follows.

LEMMA 2.1 (Johnson and Lindenstrauss). *Let $\mathcal{P}$ be an arbitrary set of $\mathcal{N}$ points in $\mathbb{R}^d$. Given a distortion parameter $\varepsilon > 0$, there exists a constant*

$$k_0 = \mathcal{O}(\varepsilon^{-2}\log(\mathcal{N})),$$

such that for all integers $k \geq k_0$, there exists a $k\times d$ matrix M for which

$$(1-\varepsilon)\|x-\tilde{x}\|_{\ell_2^d}^2 \leq \|Mx - M\tilde{x}\|_{\ell_2^k}^2 \leq (1+\varepsilon)\|x-\tilde{x}\|_{\ell_2^d}^2, \tag{2.1}$$

for all $x, \tilde{x} \in \mathcal{P}$. It is easy to see that the condition

$$(1-\varepsilon)\|p\|_{\ell_2^d}^2 \leq \|Mp\|_{\ell_2^k}^2 \leq (1+\varepsilon)\|p\|_{\ell_2^d}^2, \tag{2.2}$$

implies

$$(1-\varepsilon)\|p\|_{\ell_2^d} \leq \|Mp\|_{\ell_2^k} \leq (1+\varepsilon)\|p\|_{\ell_2^d} \tag{2.3}$$

for $0 < \varepsilon < 1$, which will be used in the following sections. On the other hand, (2.3) implies (2.2) with 3ε instead of ε.

Our aim is to apply this lemma to dynamical systems. As the mapping M from Lemma 2.1 is linear and almost preserves distances between the points (up to the $\varepsilon > 0$ distortion as described above), we restrict ourselves to dynamical systems which are linear or whose non-linearity depends only on the mutual distances of the points involved, as in (1.1).

Let us define the additional notation, which is going to be fixed throughout the paper:

- $d \in \mathbb{N}$ - dimension (large),
- $\varepsilon > 0$ - the distortion parameter from Lemma 2.1,
- $k \in \mathbb{N}$ - new dimension (small),
- $M \in \mathbb{R}^{k\times d}$ - randomly generated matrix as described below.

The only constructions of a matrix M as in Lemma 2.1 known up to now are stochastic, i.e., the matrix is randomly generated and has the quasi-isometry property (2.1) with high probability. We refer the reader to [22] and [1, Theorem 1.1] for two typical versions of the Johnson-Lindenstrauss Lemma.

We briefly collect below some well-known instances of random matrices, which satisfy the statement of Lemma 2.1 with high probability:

- $k \times d$ matrices M whose entries $m_{i,j}$ are independent realizations of Gaussian random variables

$$m_{i,j} \sim \mathcal{N}\left(0, \frac{1}{k}\right);$$

- $k \times d$ matrices M whose entries are independent realizations of $\pm$ Bernoulli random variables

$$m_{i,j} := \begin{cases} +\frac{1}{\sqrt{k}}, & \text{with probability } \frac{1}{2} \\ -\frac{1}{\sqrt{k}}, & \text{with probability } \frac{1}{2} \end{cases}$$

Several other random projections suitable for Johnson-Lindenstrauss embeddings can be constructed following Theorem 3.6 recalled below, and we refer the reader to [36] for more details.

2.2. Uniform estimate for a general model. If $M \in \mathbb{R}^{k\times d}$ is a matrix, we consider the projected Euler method in $\mathbb{R}^k$ associated to the high-dimensional system (1.5)-(1.6), namely

$$y_i^0 := Mx_i^0, \tag{2.4}$$

$$y_i^{n+1} := y_i^n + h\left[Mf_i(\mathcal{D}'y^n) + \sum_{j=1}^N f_{ij}(\mathcal{D}'y^n)y_j^n\right], \quad n = 0, \ldots, n_0 - 1. \tag{2.5}$$

We denote here $\mathcal{D}' : \mathbb{R}^{k\times N} \to \mathbb{R}^{N\times N}$, $\mathcal{D}'y := (\|y_i - y_j\|_{\ell_2^k})_{i,j=1}^N$, the *adjacency matrix* of the agents $y = (y_1, \dots, y_N)$ in $\mathbb{R}^{k\times N}$. The first result of this paper reads as follows.

THEOREM 2.2. *Let the sequences*

$$\{x_i^n, i = 1, \dots, N \text{ and } n = 0, \dots, n_0\} \quad \text{and} \quad \{y_i^n, i = 1, \dots, N \text{ and } n = 0, \dots, n_0\}$$

be defined by (1.5)-(1.6) *and* (2.4)-(2.5) *with f_i and f_{ij} satisfying* (1.2)–(1.4) *and a matrix $M \in \mathbb{R}^{k\times d}$ with*

$$\|Mf_i(\mathcal{D}'y^n) - Mf_i(\mathcal{D}x^n)\|_{\ell_2^k} \le (1+\varepsilon)\,\|f_i(\mathcal{D}'y^n) - f_i(\mathcal{D}x^n)\|_{\ell_2^d}\,, \tag{2.6}$$

$$\|Mx_j^n\|_{\ell_2^k} \le (1+\varepsilon)\|x_j^n\|_{\ell_2^d}, \tag{2.7}$$

$$(1-\varepsilon)\|x_i^n - x_j^n\|_{\ell_2^d} \le \|Mx_i^n - Mx_j^n\|_{\ell_2^k} \le (1+\varepsilon)\|x_i^n - x_j^n\|_{\ell_2^d} \tag{2.8}$$

for all $i, j = 1, \dots, N$ and all $n = 0, \dots, n_0$. Moreover, let us assume that

$$\alpha \ge \max_j \|x_j^n\|_{\ell_2^d} \quad \text{for all} \quad n = 0, \dots, n_0, \quad j = 1, \dots, N.$$

Let

$$e_i^n := \|y_i^n - Mx_i^n\|_{\ell_2^k},\ i = 1, \dots, N \text{ and } n = 0, \dots, n_0 \tag{2.9}$$

and set $\mathcal{E}^n := \max_i e_i^n$. Then

$$\mathcal{E}^n \le \varepsilon h n B \exp(hnA), \tag{2.10}$$

where $A := L' + 2(1+\varepsilon)(L + \alpha L'')$ and $B := 2\alpha(1+\varepsilon)(L+\alpha L'')$. We remark that conditions (2.6)-(2.8) are in fact satisfied as soon as M is a suitable Johnson-Lindenstrauss embedding as in Lemma 2.1.

Proof. Using (2.9) and (1.5)-(1.6) and (2.4)-(2.5) combined with (2.6) and (2.7), we obtain

$$\begin{aligned}
e_i^{n+1} &\le e_i^n + h\,\|Mf_i(\mathcal{D}'y^n) - Mf_i(\mathcal{D}x^n)\|_{\ell_2^k} + h\left\|\sum_{j=1}^N f_{ij}(\mathcal{D}'y^n)y_j^n - f_{ij}(\mathcal{D}x^n)Mx_j^n\right\|_{\ell_2^k}\\
&\le e_i^n + h(1+\varepsilon)\,\|f_i(\mathcal{D}'y^n) - f_i(\mathcal{D}x^n)\|_{\ell_2^d}\\
&\quad + h\sum_{j=1}^N\Big(\|f_{ij}(\mathcal{D}'y^n)y_j^n - f_{ij}(\mathcal{D}'y^n)Mx_j^n\|_{\ell_2^k} + \|f_{ij}(\mathcal{D}'y^n)Mx_j^n - f_{ij}(\mathcal{D}x^n)Mx_j^n\|_{\ell_2^k}\Big)\\
&\le e_i^n + h(1+\varepsilon)\,\|f_i(\mathcal{D}'y^n) - f_i(\mathcal{D}x^n)\|_{\ell_2^d}\\
&\quad + h\sum_{j=1}^N\Big(|f_{ij}(\mathcal{D}'y^n)|e_j^n + (1+\varepsilon)\|x_j^n\|_{\ell_2^d}\cdot|f_{ij}(\mathcal{D}'y^n) - f_{ij}(\mathcal{D}x^n)|\Big).
\end{aligned}$$

Taking the maximum on both sides, this becomes

$$\begin{aligned}
\mathcal{E}^{n+1} &\le \mathcal{E}^n + h(1+\varepsilon)\max_i \|f_i(\mathcal{D}'y^n) - f_i(\mathcal{D}x^n)\|_{\ell_2^d}\\
&\quad + h\mathcal{E}^n \max_i \sum_{j=1}^N |f_{ij}(\mathcal{D}'y^n)| + h(1+\varepsilon)\alpha\cdot\max_i\sum_{j=1}^N |f_{ij}(\mathcal{D}'y^n) - f_{ij}(\mathcal{D}x^n)|.
\end{aligned}$$

We use (1.2)–(1.4) for $a = \mathcal{D}'y^n$ and $b = \mathcal{D}x^n$ to estimate all the terms on the right-hand side. This gives

$$\begin{aligned}\mathcal{E}^{n+1} &\le \mathcal{E}^n + h(1+\varepsilon)L\|\mathcal{D}'y^n - \mathcal{D}x^n\|_{\ell_\infty^N(\ell_\infty^N)} + h\mathcal{E}^n L' + h(1+\varepsilon)\alpha L''\|\mathcal{D}'y^n - \mathcal{D}x^n\|_{\ell_\infty^N(\ell_\infty^N)}\\ &\le \mathcal{E}^n(1+hL') + h(1+\varepsilon)(L+\alpha L'')\left[\|\mathcal{D}'y^n - \mathcal{D}'Mx^n\|_{\ell_\infty^N(\ell_\infty^N)} + \|\mathcal{D}'Mx^n - \mathcal{D}x^n\|_{\ell_\infty^N(\ell_\infty^N)}\right]\\ &\le \mathcal{E}^n(1+hL') + 2h(1+\varepsilon)(L+\alpha L'')(\mathcal{E}^n + \alpha\varepsilon),\end{aligned}$$

where we used (2.8) in the last line. This, together with $\mathcal{E}^0 = 0$, leads to

$$\mathcal{E}^n \le \varepsilon hnB\exp(hnA),$$

where $A := L' + 2(1+\varepsilon)(L+\alpha L'')$ and $B := 2\alpha(1+\varepsilon)(L+\alpha L'')$. ◻

2.3. Uniform estimate for the Cucker-Smale model. As a relevant example, let us now show that Theorem 2.2 can be applied to the well-known Cucker-Smale model, introduced and analyzed in [20, 21], which is described by

$$\dot{x}_i = v_i \in \mathbb{R}^d, \tag{2.11}$$

$$\dot{v}_i = \frac{1}{N}\sum_{j=1}^N g(\|x_i - x_j\|_{\ell_2^d})(v_j - v_i), \quad i = 1,\dots,N. \tag{2.12}$$

The function $g : [0,\infty) \to \mathbb{R}$ is given by $g(s) = \frac{G}{(1+s^2)^\beta}$, for $\beta > 0$, and bounded by $g(0) = G > 0$. This model describes the *emerging of consensus* in a group of interacting agents, trying to *align* (also in terms of abstract consensus) with their neighbors. One of the motivations of the model from Cucker and Smale was to describe the formation and evolution of languages [21, Section 6], although, due to its simplicity, it has been eventually related mainly to the description of the *emergence of flocking* in groups of birds [20]. In the latter case, in fact, spatial and velocity coordinates are sufficient to describe a pointlike agent ($d = 3+3$), while for the evolution of languages, one would have to take into account a much broader dictionary of parameters, hence a higher dimension $d \gg 3+3$ of parameters, which is in fact the case of our interest in the present paper.

Let us show that the model is indeed of the type (1.1). We interprete the system as a group of $2N$ agents in $\mathbb{R}^d$, whose dynamics is given by the following equations

$$\begin{aligned}\dot{x}_i &= \sum_{j=1}^N f_{ij}^x v_j \in \mathbb{R}^d,\\ \dot{v}_i &= \sum_{j=1}^N f_{ij}^v(\mathcal{D}x)v_j, \quad i = 1,\dots,N\end{aligned}$$

with $f_{ij}^x := \delta_{ij}$, $f_{ii}^v(\mathcal{D}x) := -\dfrac{1}{N}\displaystyle\sum_{k=1}^N g(\|x_i - x_k\|_{\ell_2^d})$, and $f_{ij}^v(\mathcal{D}x) := \dfrac{1}{N}g(\|x_i - x_j\|_{\ell_2^d})$, for $i \ne j$. The condition (1.2) is empty, (1.3) reads

$$L' \ge \max(1, 2G) \ge \max_i\left\{1, \frac{2}{N}\sum_{k=1}^N g(\|x_i^n - x_k^n\|_{\ell_2^d})\right\}.$$

Finally,

$$\begin{aligned}
\max_i \frac{2}{N}\sum_{j=1}^N \Big| g(\|x_i^n - x_j^n\|_{\ell_2^d}) - g(\|y_i^n - y_j^n\|_{\ell_2^k})\Big| \\
&\le \max_i \frac{2\|g\|_{\mathrm{Lip}}}{N}\cdot \sum_{j=1}^N \Big| \|x_i^n - x_j^n\|_{\ell_2^d} - \|y_i^n - y_j^n\|_{\ell_2^k}\Big| \\
&\le 2\|g\|_{\mathrm{Lip}} \cdot \|\mathcal{D}' y^n - \mathcal{D} x^n\|_{\ell_\infty^N(\ell_\infty^N)}
\end{aligned}$$

shows that $L'' \le 2\|g\|_{\mathrm{Lip}}$.

2.4. Least-squares estimate of the error for the Cucker-Smale model. The formula (2.10) provides the estimate of the maximum of the individual errors, i.e., $\mathcal{E}^n := \|(y_i^n - Mx_i^n)_{i=1}^N\|_{\ell_\infty^N(\ell_2^k)}$. In this section we address the stronger $\ell_2^N(\ell_2^k)$-estimate for the error. For generic dynamical systems (1.1) such estimate is not available in general, and one has to perform a case-by-case analysis. As a typical example of how to proceed, we restrict ourselves to the Cucker-Smale model, just recalled in the previous section. The foward Euler discretization of (2.11)–(2.12) is given by

$$\begin{aligned}
x_i^{n+1} &= x_i^n + hv_i^n, \\
v_i^{n+1} &= v_i^n + \frac{h}{N}\sum_{j=1}^N g(\|x_i^n - x_j^n\|_{\ell_2^d})(v_j^n - v_i^n)
\end{aligned} \tag{2.13}$$

with initial data x_i^0 and v_i^0 given. Let M be again a suitable random matrix in the sense of Lemma 2.1. The Euler method of the projected system is given by the initial conditions $y_i^0 = Mx_i^0$ and $w_i^0 = Mv_i^0$ and the formulas

$$\begin{aligned}
y_i^{n+1} &= y_i^n + hw_i, \\
w_i^{n+1} &= w_i^n + \frac{h}{N}\sum_{j=1}^N g(\|y_i^n - y_j^n\|_{\ell_2^k})(w_j^n - w_i^n).
\end{aligned} \tag{2.14}$$

We are interested in the estimates of the following quantities

$$\begin{aligned}
e_{x,i}^n &:= \|y_i^n - Mx_i^n\|_{\ell_2^k}, \quad \mathcal{E}_x^n := \sqrt{\frac{1}{N}\sum_{i=1}^N (e_{x,i}^n)^2} = \frac{\|(y_i^n - Mx_i^n)_{i=1}^N\|_{\ell_2^N(\ell_2^k)}}{\sqrt{N}}, \\
e_{v,i}^n &:= \|w_i^n - Mv_i^n\|_{\ell_2^k}, \quad \mathcal{E}_v^n := \sqrt{\frac{1}{N}\sum_{i=1}^N (e_{v,i}^n)^2} = \frac{\|(w_i^n - Mv_i^n)_{i=1}^N\|_{\ell_2^N(\ell_2^k)}}{\sqrt{N}}.
\end{aligned}$$

Using (2.13) and (2.14), we obtain

$$e_{x,i}^{n+1} \le e_{x,i}^n + he_{v,i}^n \quad \text{and} \quad \mathcal{E}_x^{n+1} \le \mathcal{E}_x^n + h\mathcal{E}_v^n.$$

To bound the quantity $\mathcal{E}_v^n$ we have to work more. Another application of (2.13) and

(2.14) leads to

$$
\begin{aligned}
e_{v,i}^{n+1} &\le e_{v,i}^n + \frac{h}{N}\sum_{j=1}^N \Big(\|g(\|y_i^n - y_j^n\|_{\ell_2^k})(w_j^n - w_i^n) \pm g(\|y_i^n - y_j^n\|_{\ell_2^k})(Mv_j^n - Mv_i^n) \\
&\qquad - g(\|x_i^n - x_j^n\|_{\ell_2^d})(Mv_j^n - Mv_i^n)\|_{\ell_2^k}\Big) \\
&\le e_{v,i}^n + \frac{h}{N}\sum_{j=1}^N g(\|y_i^n - y_j^n\|_{\ell_2^k})(e_{v,j}^n + e_{v,i}^n) \qquad (2.15)\\
&\qquad + \frac{(1+\varepsilon)h\|g\|_{\mathrm{Lip}}}{N}\cdot \sum_{j=1}^N \|v_j^n - v_i^n\|_{\ell_2^d}\cdot \big|\|x_i^n - x_j^n\|_{\ell_2^d} - \|y_i^n - y_j^n\|_{\ell_2^k}\big|.
\end{aligned}
$$

We estimate the first summand in (2.15)

$$
\frac{h}{N}\sum_{j=1}^N g(\|y_i^n - y_j^n\|_{\ell_2^k})(e_{v,j}^n + e_{v,i}^n) \le \frac{hG}{N}\big[Ne_{v,i}^n + \sum_{j=1}^N e_{v,j}^n\big] = hGe_{v,i}^n + \frac{hG}{N}\sum_{j=1}^N e_{v,j}^n
$$

and its ℓ_2-norm with respect to i with the use of Hölder inequality

$$
h\sqrt{N}G\mathcal{E}_v^n + \frac{hG}{N}\left(\sum_{i=1}^N\left(\sum_{j=1}^N e_{v,j}^n\right)^2\right)^{1/2} \le 2h\sqrt{N}G\mathcal{E}_v^n. \qquad (2.16)
$$

To estimate the second summand in (2.15), let us set $V := \max_{i,j,n}\|v_i^n - v_j^n\|_{\ell_2^d}$ and make use of

$$
\begin{aligned}
&\big|\|x_i^n - x_j^n\|_{\ell_2^d} - \|y_i^n - y_j^n\|_{\ell_2^k}\big| \\
&\quad \le \big|\|x_i^n - x_j^n\|_{\ell_2^d} - \|Mx_i^n - Mx_j^n\|_{\ell_2^k}\big| + \big|\|Mx_i^n - Mx_j^n\|_{\ell_2^k} - \|y_i^n - y_j^n\|_{\ell_2^k}\big| \\
&\quad \le \varepsilon\|x_i^n - x_j^n\|_{\ell_2^d} + e_{x,i}^n + e_{x,j}^n.
\end{aligned}
$$

We arrive at

$$
\begin{aligned}
&\frac{(1+\varepsilon)h\|g\|_{\mathrm{Lip}}}{N}\sum_{j=1}^N \|v_j^n - v_i^n\|_{\ell_2^d}(\varepsilon\|x_i^n - x_j^n\|_{\ell_2^d} + e_{x,i}^n + e_{x,j}^n) \\
&\qquad \le \frac{(1+\varepsilon)h\|g\|_{\mathrm{Lip}}V}{N}\left\{\varepsilon\sum_{j=1}^N \|x_i^n - x_j^n\|_{\ell_2^d} + Ne_{x,i}^n + \sum_{j=1}^N e_{x,j}^n\right\}.
\end{aligned}
$$

The ℓ_2-norm of this expression with respect to i is bounded by

$$
\begin{aligned}
&\frac{(1+\varepsilon)h\|g\|_{\mathrm{Lip}}V}{N}\left\{\varepsilon\Big(\sum_{i=1}^N\Big(\sum_{j=1}^N\|x_i^n - x_j^n\|_{\ell_2^d}\Big)^2\Big)^{1/2} + N\Big(\sum_{i=1}^N (e_{x,i}^n)^2\Big)^{1/2} + \sqrt{N}\sum_{j=1}^N e_{x,j}^n\right\} \\
&\le (1+\varepsilon)h\|g\|_{\mathrm{Lip}}V\sqrt{N}(\varepsilon X + 2\mathcal{E}_x^n), \qquad (2.17)
\end{aligned}
$$

where $X := \max_{i,j,n}\|x_i^n - x_j^n\|_{\ell_2^d}$. Combining (2.15) with (2.16) and (2.17) leads to the recursive estimate

$$
\begin{aligned}
\mathcal{E}_x^{n+1} &\le \mathcal{E}_x^n + h\mathcal{E}_v^n, \qquad (2.18)\\
\mathcal{E}_v^{n+1} &\le \mathcal{E}_v^n + 2hG\mathcal{E}_v^n + h(1+\varepsilon)\|g\|_{\mathrm{Lip}}V\left\{\varepsilon X + 2\mathcal{E}_x^n\right\},
\end{aligned}
$$

which we put into the matrix form

$$\begin{pmatrix} \mathcal{E}_x^{n+1} \\ \mathcal{E}_v^{n+1} \end{pmatrix} = \mathcal{A} \begin{pmatrix} \mathcal{E}_x^{n} \\ \mathcal{E}_v^{n} \end{pmatrix} + \begin{pmatrix} 0 \\ (1+\varepsilon)\varepsilon h \|g\|_{\mathrm{Lip}} VX \end{pmatrix}, \tag{2.19}$$

where $\mathcal{A}$ is a 2×2 matrix given by

$$\mathcal{A} = \mathcal{A}_1 + h\mathcal{A}_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + h \begin{pmatrix} 0 & 1 \\ 2(1+\varepsilon)\|g\|_{\mathrm{Lip}} V & 2G \end{pmatrix}.$$

Taking the norms on both sides of (2.19) leads to

$$\sqrt{(\mathcal{E}_x^{n+1})^2 + (\mathcal{E}_v^{n+1})^2} \le (1 + h\|\mathcal{A}_2\|)\sqrt{(\mathcal{E}_x^{n})^2 + (\mathcal{E}_v^{n})^2} + \varepsilon(1+\varepsilon)h\|g\|_{\mathrm{Lip}} VX$$

and the least-squares error estimate finally reads as follows.

$$\sqrt{(\mathcal{E}_x^{n})^2 + (\mathcal{E}_v^{n})^2} \le \varepsilon(1+\varepsilon)hn\|g\|_{\mathrm{Lip}} VX \exp(hn\|\mathcal{A}_2\|).$$

3. Dimensionality reduction for continuous dynamical systems.

3.1. Uniform estimates for continuous dynamical systems. In this section we shall establish the analogue of the above results for the continuous time setting of dynamical systems of the type (1.1),

$$\dot{x}_i = f_i(\mathcal{D}x) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}x)x_j\,, \qquad i = 1, \ldots, N\,, \tag{3.1}$$

$$x_i(0) = x_i^0\,, \qquad i = 1, \ldots, N\,. \tag{3.2}$$

We adopt again the assumptions about Lipschitz continuity and boundedness of the right-hand side made in Section 2, namely (1.2), (1.3) and (1.4).

THEOREM 3.1. *Let $x(t) \in \mathbb{R}^{d\times N}$, $t \in [0,T]$, be the solution of the system* (3.1)–(3.2) *with f_i's and f_{ij}'s satisfying* (1.2)–(1.4)*, such that*

$$\max_{t\in[0,T]} \max_{i,j} \|x_i(t) - x_j(t)\|_{\ell_2^d} \le \alpha\,. \tag{3.3}$$

Let us fix $k \in \mathbb{N}$, $k \le d$, and a matrix $M \in \mathbb{R}^{k\times d}$ such that

$$(1-\varepsilon)\|x_i(t) - x_j(t)\|_{\ell_2^d} \le \|Mx_i(t) - Mx_j(t)\|_{\ell_2^k} \le (1+\varepsilon)\|x_i(t) - x_j(t)\|_{\ell_2^d}\,, \tag{3.4}$$

for all $t \in [0,T]$ and i, $j = 1, \ldots, N$. Let $y(t) \in \mathbb{R}^{k\times N}$, $t \in [0,T]$ be the solution of the projected system

$$\dot{y}_i = Mf_i(\mathcal{D}'y) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}'y)y_j\,, \qquad i = 1, \ldots, N\,,$$

$$y_i(0) = Mx_i^0\,, \qquad i = 1, \ldots, N\,, \tag{3.5}$$

such that for a suitable $\beta > 0$,

$$\max_{t\in[0,T]} \|y(t)\|_{\ell_\infty^N(\ell_2^d)} \le \beta\,. \tag{3.6}$$

Let us define the column-wise ℓ_2-error $e_i(t) := \|y_i - Mx_i\|_{\ell_2^k}$ for $i = 1, \dots, N$ and

$$\mathcal{E}(t) := \max_{i=1,\dots,N} e_i(t) = \|y - Mx\|_{\ell_\infty^N(\ell_2^k)} .$$

Then we have the estimate

$$\mathcal{E}(t) \leq \varepsilon\alpha t(L\,\|M\| + L''\beta) \exp\left[(2L\,\|M\| + 2\beta L'' + L')t\right] . \tag{3.7}$$

Proof. Due to (1.2)–(1.4), we have for every $i = 1, \dots, N$ the estimate

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} e_i = \frac{\langle y_i - Mx_i, \frac{\mathrm{d}}{\mathrm{d}t}(y_i - Mx_i)\rangle}{\|y_i - Mx_i\|_{\ell_2^k}} &\leq \left\| \frac{\mathrm{d}}{\mathrm{d}t}(y_i - Mx_i) \right\|_{\ell_2^k} \\
&\leq \|Mf_i(\mathcal{D}'y) - Mf_i(\mathcal{D}x)\|_{\ell_2^k} + \sum_{j=1}^N \|f_{ij}(\mathcal{D}'y)y_j - f_{ij}(\mathcal{D}x)Mx_j\|_{\ell_2^k} \\
&\leq L\,\|M\|\,\|\mathcal{D}'y - \mathcal{D}x\|_{\ell_\infty^N(\ell_\infty^N)} + \sum_{j=1}^N \Big(\|f_{ij}(\mathcal{D}x)(Mx_j - y_j)\|_{\ell_2^k} + \|(f_{ij}(\mathcal{D}x) - f_{ij}(\mathcal{D}'y))y_j\|_{\ell_2^k} \Big) \\
&\leq L\,\|M\|\,\|\mathcal{D}'y - \mathcal{D}x\|_{\ell_\infty^N(\ell_\infty^N)} + L'\,\|Mx - y\|_{\ell_\infty^N(\ell_2^k)} + L''\,\|\mathcal{D}x - \mathcal{D}'y\|_{\ell_\infty^N(\ell_\infty^N)}\,\|y\|_{\ell_\infty^N(\ell_2^k)} .
\end{aligned}$$

The term $\|\mathcal{D}'y - \mathcal{D}x\|_{\ell_\infty^N(\ell_\infty^N)} \leq \|\mathcal{D}'y - \mathcal{D}'Mx\|_{\ell_\infty^N(\ell_\infty^N)} + \|\mathcal{D}'Mx - \mathcal{D}x\|_{\ell_\infty^N(\ell_\infty^N)}$ is estimated by

$$\begin{aligned}
\|\mathcal{D}'y - \mathcal{D}Mx\|_{\ell_\infty^N(\ell_\infty^N)} &= \max_{i,j} \Big| \|y_i - y_j\|_{\ell_2^k} - \|Mx_i - Mx_j\|_{\ell_2^k} \Big| \\
&\leq \max_{i,j} \|y_i - Mx_i\|_{\ell_2^k} + \|y_j - Mx_j\|_{\ell_2^k} \leq 2\mathcal{E}(t)\,,
\end{aligned}$$

and, using the assumption (3.4),

$$\|\mathcal{D}'Mx - \mathcal{D}x\|_{\ell_\infty^N(\ell_\infty^N)} = \max_{i,j} \Big| \|Mx_i - Mx_j\|_{\ell_2^k} - \|x_i - x_j\|_{\ell_2^d} \Big| \leq \varepsilon \max_{i,j} \|x_i - x_j\|_{\ell_2^k} = \varepsilon\,\|\mathcal{D}x\|_{\ell_\infty^N(\ell_\infty^N)} .$$

Finally, with the a priori estimate (3.3) for $\|\mathcal{D}x\|_{\ell_\infty^N(\ell_\infty^N)}$ and (3.6) for $\|y\|_{\ell_\infty^N(\ell_2^d)}$, we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} e_i &\leq L\,\|M\|\,(2\mathcal{E}(t) + \varepsilon\alpha) + L'\mathcal{E}(t) + L''\beta(2\mathcal{E}(t) + \varepsilon\alpha) \\
&= (2L\,\|M\| + 2\beta L'' + L')\mathcal{E}(t) + \varepsilon\alpha(L\,\|M\| + L''\beta) .
\end{aligned}$$

Now, let us split the interval $[0, T)$ into a union of finite disjoint intervals $I_j = [t_{j-1}, t_j)$, $j = 1, \dots, K$ for a suitable $K \in \mathbb{N}$, such that $\mathcal{E}(t) = e_{i(j)}(t)$ for $t \in I_j$. Consequently, on every I_j we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(t) = \frac{\mathrm{d}}{\mathrm{d}t} e_{i(j)}(t) \leq (2L\,\|M\| + 2\beta L'' + L')\mathcal{E}(t) + \varepsilon\alpha(L\,\|M\| + L''\beta)\,,$$

and the Gronwall lemma yields

$$\mathcal{E}(t) \leq \left[\varepsilon\alpha(L\,\|M\| + L''\beta)(t - t_{j-1}) + \mathcal{E}(t_{j-1})\right] \exp\left((2L\,\|M\| + 2\beta L'' + L')(t - t_{j-1})\right)$$

for $t \in [t_{j-1}, t_j)$. A concatenation of these estimates over the intervals I_j leads finally to the expected error estimate

$$\mathcal{E}(t) \leq \varepsilon\alpha t(L\,\|M\| + L''\beta) \exp\left[(2L\,\|M\| + 2\beta L'' + L')t\right] .$$

□

3.2. A continuous Johnson-Lindenstrauss Lemma. Let us now go through the assumptions we made in the formulation of Theorem 3.1 and discuss how they restrict the validity and applicability of the result. First of all, let us mention that (3.3) and (3.6) can be easily proven to hold for locally Lipschitz right-hand sides f_i and f_{ij} on finite time intervals. Obviously, the critical point for the applicability of Theorem 3.1 is the question how to find a matrix M satisfying the condition (3.4), i.e., being a quasi-isometry along the trajectory solution $x(t)$ for *every* $t \in [0,T]$. The answer is provided by the following generalization of the Johnson-Lindenstrauss Lemma (Lemma 2.1) for rectifiable $\mathcal{C}^1$-curves, by a suitable continuity argument. Let us stress that our approach resembles the "sampling and ϵ-net" argument in [3, 4, 47] for the extension of the quasi-isometry property of Johnson-Lindenstrauss embeddings to smooth Riemmanian manifolds. From this point of view the following result can be viewed as a specification of the work [4, 47].

We first prove an auxiliary technical result:

LEMMA 3.2. *Let $0 < \varepsilon < \varepsilon' < 1$, $a \in \mathbb{R}^d$ and let $M : \mathbb{R}^d \to \mathbb{R}^k$ be a linear mapping such that*

$$(1-\varepsilon)\|a\|_{\ell_2^d} \le \|Ma\|_{\ell_2^k} \le (1+\varepsilon)\|a\|_{\ell_2^d}.$$

Let $x \in \mathbb{R}^d$ satisfy

$$\|a - x\| \le \frac{(\varepsilon' - \varepsilon)\|a\|_{\ell_2^d}}{\|M\| + 1 + \varepsilon'}. \tag{3.8}$$

Then

$$(1-\varepsilon')\|x\|_{\ell_2^d} \le \|Mx\|_{\ell_2^k} \le (1+\varepsilon')\|x\|_{\ell_2^d}. \tag{3.9}$$

Proof. If $a = 0$, the statement is trivial. If $a \neq 0$, we denote the right-hand side of (3.8) by $\tau > 0$ and estimate with the triangle inequality

$$\begin{aligned}\frac{\|Mx\|_{\ell_2^k}}{\|x\|_{\ell_2^d}} &= \frac{\|M(x-a)+Ma\|_{\ell_2^k}}{\|x-a+a\|_{\ell_2^d}} \le \frac{\|M\|\cdot\|x-a\|_{\ell_2^d} + (1+\varepsilon)\|a\|_{\ell_2^d}}{\|a\|_{\ell_2^d} - \|x-a\|_{\ell_2^d}}\\ &\le \frac{\|M\|\cdot\tau + (1+\varepsilon)\|a\|_{\ell_2^d}}{\|a\|_{\ell_2^d} - \tau} \le 1+\varepsilon'.\end{aligned}$$

A similar chain of inequalities holds for the estimate from below. □

Now we are ready to establish a continuous version of Lemma 2.1.

THEOREM 3.3. *Let $\varphi : [0,1] \to \mathbb{R}^d$ be a $\mathcal{C}^1$ curve. Let $0 < \varepsilon < \varepsilon' < 1$,*

$$\gamma := \max_{\xi\in[0,1]} \frac{\|\varphi'(\xi)\|_{\ell_2^d}}{\|\varphi(\xi)\|_{\ell_2^d}} < \infty \quad \textit{and} \quad \mathcal{N} \ge (\sqrt{d}+2)\cdot\frac{\gamma}{\varepsilon'-\varepsilon}.$$

Let k be such that a randomly chosen (and properly normalized) projector M satisfies the statement of the Johnson-Lindenstrauss Lemma 2.1 with ε, d, k and $\mathcal{N}$ arbitrary points with high probability. Without loss of generality we assume that $\|M\| \le \sqrt{d/k}$ within the same probability (this is in fact the case, e.g., for the examples of Gaussian and Bernoulli random matrices reported in Section 2).

Then

$$(1-\varepsilon')\|\varphi(t)\|_{\ell_2^d} \le \|M\varphi(t)\|_{\ell_2^k} \le (1+\varepsilon')\|\varphi(t)\|_{\ell_2^d}, \ \textit{for all } t \in [0,1] \tag{3.10}$$

holds with the same probability.

Proof. Let $t_i = i/\mathcal{N}$, $i = 0,\dots,\mathcal{N}$ and put

$$T_i := \arg\max_{\xi\in[t_i,t_{i+1}]} \|\varphi'(\xi)\|_{\ell_2^d}, \quad i = 0,\dots,\mathcal{N}-1.$$

Let $M : \mathbb{R}^d \to \mathbb{R}^k$ be the randomly chosen and normalized projector (see Lemma 2.1). Hence $\|M\| \le \sqrt{d/k}$ and

$$(1-\varepsilon')\|\varphi(T_i)\|_{\ell_2^d} \le \|M(\varphi(T_i))\|_{\ell_2^k} \le (1+\varepsilon')\|\varphi(T_i)\|_{\ell_2^d}, \qquad i = 1,\dots,\mathcal{N} \tag{3.11}$$

with high probability. We show that (3.10) holds with (at least) the same probability.

This follows easily from (3.11) and the following estimate, which holds for every $t \in [t_i, t_{i+1}]$,

$$\begin{aligned}\|\varphi(t)-\varphi(T_i)\|_{\ell_2^d} &\le \int_t^{T_i} \|\varphi'(s)\|_{\ell_2^d} ds \le \frac{\|\varphi'(T_i)\|_{\ell_2^d}}{\mathcal{N}} \le \frac{\|\varphi'(T_i)\|_{\ell_2^d}(\varepsilon'-\varepsilon)}{\gamma(\sqrt{d}+2)} \\ &\le \frac{\|\varphi(T_i)\|_{\ell_2^d}(\varepsilon'-\varepsilon)}{\sqrt{d}+2} \le \frac{\|\varphi(T_i)\|_{\ell_2^d}(\varepsilon'-\varepsilon)}{\|M\|+1+\varepsilon'}.\end{aligned}$$

The proof is then finished by a straightforward application of Lemma 3.2. □

REMARK 1. *We show now that the condition*

$$\gamma := \max_{\xi\in[0,1]} \frac{\|\varphi'(\xi)\|_{\ell_2^d}}{\|\varphi(\xi)\|_{\ell_2^d}} < \infty$$

is necessary, hence it is a restriction to the type of curves one can quasi-isometrically project. Let d ≥ 3. *It is known that there is a continuous curve* $\varphi : [0,1] \to [0,1]^{d-1}$, *such that* $\varphi([0,1]) = [0,1]^{d-1}$, *i.e.* φ *goes* onto $[0,1]^{d-1}$. *The construction of such a* space-filling *curve goes back to Peano and Hilbert. After a composition with suitable dilations and d-dimensional spherical coordinates we observe that there is also a* surjective *continuous curve* $\varphi : [0,1] \to \mathbb{S}^{d-1}$, *where* $\mathbb{S}^{d-1}$ *denotes the* ℓ_2^d *unit sphere in* $\mathbb{R}^d$.

As M *was supposed to be a projection,* (3.10) *cannot hold for all* t*'s with* $\varphi(t) \in \ker M \neq \emptyset$.

Obviously, the key condition for applicability of Theorem 3.3 for finding a projection matrix M satisfying (3.4) is that

$$\sup_{t\in[0,T]} \max_{i,j} \frac{\|\dot{x}_i-\dot{x}_j\|_{\ell_2^d}}{\|x_i-x_j\|_{\ell_2^d}} \le \gamma < \infty\,. \tag{3.12}$$

This condition is, for instance, trivially satisfied when the right-hand sides f_i's and f_{ij}'s have the following Lipschitz continuity:

$$\begin{aligned}\|f_i(\mathcal{D}x)-f_j(\mathcal{D}x)\|_{\ell_2^d} &\le L'''\|x_i-x_j\|_{\ell_2^d} \qquad \text{for all } i,j = 1,\dots,N\,,\\ |f_{i,k}(\mathcal{D}x)-f_{j,k}(\mathcal{D}x)| &\le L''''\|x_i-x_j\|_{\ell_2^d} \qquad \text{for all } i,j,k = 1,\dots,N.\end{aligned}$$

We will show in the examples below how condition (3.12) is verified in cases of dynamical systems modeling standard social mechanisms of *attraction, repulsion, aggregation* and *alignment*.

3.3. Applicability to fundamental examples of dynamical systems describing social dynamics. In this section we show the applicability of our dimensionality reduction theory to well-known dynamical systems driven by "social forces" of *alignment, attraction, repulsion* and *aggregation*. Although these models were proposed as descriptions of *group motion in physical space*, the fundamental social effects can be considered as building blocks in the more abstract context of many parameter social dynamics. It has been shown [14, 40] that these models are able to produce meaningful *patterns*, for instance *mills* in two spatial dimensions (see Figure 3.1), reproducing the behavior of certain biological species. However, we should expect that

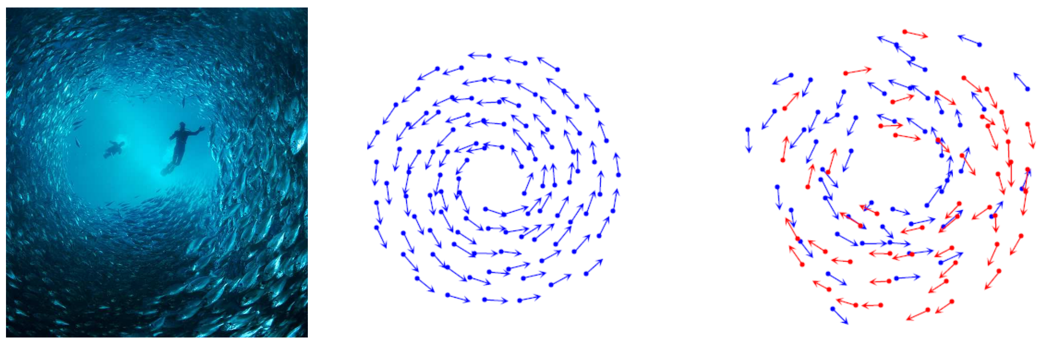

FIG. 3.1. *Mills in nature and in models*

in higher dimension the possible patterns produced by the combination of fundamental effects can be much more complex.

3.3.1. The Cucker-Smale system (alignment effect). As shown in Section 2, the Cucker and Smale flocking model (2.11)–(2.12) is of the type (1.1) and satisfies the Lipschitz continuity assumptions (1.2)–(1.4). Therefore, to meet all the assumptions of Theorem 3.1, we only need to check that it also satisfies the condition (3.12). However, for this we need to consider a slightly different framework than in Section 2.3; instead of considering the $2N$ d-dimensional variables (N position variables and N velocity variables), we need to arrange the model as N variables in $\mathbb{R}^{2d}$, each variable consisting of the position part (first d entries) and of the velocity part (the other d entries). We have then

$$\begin{aligned}
\|\dot{x}_i - \dot{x}_j\|_{\ell_2^d} + \|\dot{v}_i - \dot{v}_j\|_{\ell_2^d} &\leq \|v_i - v_j\|_{\ell_2^d} + \frac{1}{N}\sum_{k=1}^{N}\big|g(\|x_i - x_k\|_{\ell_2^d}) - g(\|x_j - x_k\|_{\ell_2^d})\big|\|v_k\|_{\ell_2^d}\\
&\leq \|v_i - v_j\|_{\ell_2^d} + \frac{\|g\|_{Lip}}{N}\sum_{k=1}^{N}\big|\|x_i - x_k\|_{\ell_2^d} - \|x_j - x_k\|_{\ell_2^d}\big|\|v_k\|_{\ell_2^d}\\
&\leq \|v_i - v_j\|_{\ell_2^d} + \frac{\|g\|_{Lip}}{N}\left(\sum_{k=1}^{N}\|v_k\|_{\ell_2^d}\right)\|x_i - x_j\|_{\ell_2^d}\\
&\leq \|v_i - v_j\|_{\ell_2^d} + c\|x_i - x_j\|_{\ell_2^d}\,,
\end{aligned}$$

for a suitable constant c depending on the initial data. We used here the a-priori boundedness of the term $\frac{1}{N}\left(\sum_{k=1}^{N}\|v_k\|_{\ell_2^d}\right)$, see [21] or [32] for details. Consequently, we can satisfy (3.12) with $\gamma = \max(1, c)$.

3.3.2. D'Orsogna model, gravitational and electrostatic interaction (attraction and repulsion effects). Another practically relevant model which fits into the class given by (1.1) is the so-called D'Orsogna model of flocking, [40]:

$$\dot{x}_i = v_i \,, \tag{3.13}$$

$$\dot{v}_i = (a - b\|v_i\|^2_{\ell_2^d})v_i - \frac{1}{N}\sum_{j\neq i} \nabla_{x_i} U(\|x_i - x_j\|_{\ell_2^d}) \,, \qquad i = 1, \dots, N, \tag{3.14}$$

where a and b are positive constants and $U : [0, \infty) \to \mathbb{R}$ is a smooth potential. We denote $u(s) = U'(s)/s$ and assume that u is a bounded, Lipschitz continuous function. We again arrange the model as a system of N variables in $\mathbb{R}^{2d}$, each variable consisting of the position part (first d entries) and of the velocity part (the other d entries). Consequently, the model can be put into a form compliant with (1.1) as follows:

$$\begin{aligned}
\dot{x}_i &= \sum_{j=1}^{N} f_{ij}^{xv} v_j \,, \\
\dot{v}_i &= \sum_{j=1}^{N} f_{ij}^{vv}(\mathcal{D}v) v_j + \sum_{j=1}^{N} f_{ij}^{vx}(\mathcal{D}x) x_j \,,
\end{aligned}$$

with $f_{ij}^{xv} = \delta_{ij}$, $f_{ii}^{vx}(\mathcal{D}x) = -\frac{1}{N}\sum_{j\neq i} u(\|x_i - x_j\|_{\ell_2^d})$ and $f_{ij}^{vx}(\mathcal{D}x) = \frac{1}{N}u(\|x_i - x_j\|_{\ell_2^d})$ for $i \neq j$. Moreover, we may set $f_{ij}^{vv}(\mathcal{D}v) = \delta_{ij}(a - b\|v_i\|^2_{\ell_2^d})$ by introducing an auxiliary, noninfluential constant zero particle $(x_0, v_0) = (0, 0)$ with null dynamics, i.e., $f_0^{*\star} = 0$ and $f_{0j}^{*\star} = 0$, $*, \star \in \{x, v\}$. Then, (1.2) is void, while (1.3) is satisfied by

$$\begin{aligned}
\max_i \sum_j (|f_{ij}^{xv}(\mathcal{D}x, \mathcal{D}v)| &+ |f_{ij}^{vx}(\mathcal{D}x, \mathcal{D}v)| + |f_{ij}^{vv}(\mathcal{D}x, \mathcal{D}v)|) \\
&\leq 1 + a + b \max_i \|v_i\|^2_{\ell_2^d} + 2\,\|u\|_{L_\infty} \leq L' \,,
\end{aligned}$$

since the theory provides an apriori bound on $\beta_v := \sup_{t\in[0,T]} \max_i \|v_i\|_{\ell_2^d}$, see [40]. Condition (1.4) for f_{ij}^{xv} is void, while for f_{ij}^{vv} it is satisfied by

$$\begin{aligned}
\max_i \sum_j \left|f_{ij}^{vv}(\mathcal{D}v) - f_{ij}^{vv}(\mathcal{D}w)\right| &\leq b \max_i \left| \|v_i\|^2_{\ell_2^d} - \|w_i\|^2_{\ell_2^d} \right| \\
&\leq b \max_i \left(\|v_i\|_{\ell_2^d} + \|w_i\|_{\ell_2^d} \right) \|v_i - w_i\|_{\ell_2^d} \\
&\leq L'' \, \|\mathcal{D}v - \mathcal{D}w\|_{\ell_\infty^N(\ell_\infty^N)} \,,
\end{aligned}$$

where we again use the apriori boundedness of β_v. For f_{ij}^{vx} is (1.4) satisfied by

$$\begin{aligned}
\max_i \sum_j \left|f_{ij}^{vx}(\mathcal{D}x) - f_{ij}^{vx}(\mathcal{D}y)\right| &\leq \max_i \frac{2}{N} \sum_{j\neq i} \left| u(\|x_i - x_j\|_{\ell_2^d}) - u(\|y_i - y_j\|_{\ell_2^d}) \right| \\
&\leq \max_i \frac{2}{N} \|u\|_{\mathrm{Lip}} \sum_{j\neq i} \left| \|x_i - x_j\|_{\ell_2^d} - \|y_i - y_j\|_{\ell_2^d} \right| \\
&\leq 2\,\|u\|_{\mathrm{Lip}} \, \|\mathcal{D}x - \mathcal{D}y\|_{\ell_\infty^N(\ell_\infty^N)} \,.
\end{aligned}$$

Finally, it can be easily checked that condition (3.12) is satisfied by

$$\|\dot{x}_i - \dot{x}_j\|_{\ell_2^d} + \|\dot{v}_i - \dot{v}_j\|_{\ell_2^d} \le (1 + a + 3b\beta_v^2)\|v_i - v_j\|_{\ell_2^d} + \Big(\|u\|_{L_\infty} + 2\beta_x \|u\|_{\mathrm{Lip}}\Big) \|x_i - x_j\|_{\ell_2^d} \,,$$

where $\beta_x := \sup_{t\in[0,T]} \max_i \|x_i\|_{\ell_2^d}$.

In fact, the D'Orsogna model is a generalization of the classical model of interacting particles through a potential U,

$$\begin{aligned} \dot{x}_i &= v_i \,, \qquad i = 1, \dots, N \,, \\ \dot{v}_i &= -\frac{1}{N}\sum_{j\neq i} \nabla U(\|x_i - x_j\|_{\ell_2^d}) \,, \qquad i = 1, \dots, N \,, \end{aligned}$$

for instance, gravitational or electrostatic interaction. However, in these cases the function $u(s) = U'(s)/s$ does not meet the assumptions of boundedness and Lipschitz continuity that are needed for the applicability of our method. Consequently, we only can consider models with regular enough potentials.

3.4. Recovery of the dynamics in high dimension from multiple simulations in low dimension. The main message of Theorem 3.1 is that, under suitable assumptions on the governing functions f_i, f_{ij}, the trajectory of the solution $y(t)$ of the *projected* dynamical system (3.5) is at an ε error from the trajectory of the *projection* of the solution $x(t)$ of the dynamical system (3.1)-(3.2), i.e.,

$$y_i(t) \approx Mx_i(t) \text{ or, more precisely, } \|Mx_i(t) - y_i(t)\|_{\ell_2^k} \le C(t)\varepsilon, \quad t \in [0,T]. \tag{3.15}$$

We wonder whether this approximation property can allow us to "learn" something about the original trajectory $x(t)$ in high dimension.

3.4.1. Sparse recovery. To address this issue we recall first some relevant and useful concepts from the field of *compressed sensing* [25, 27]. Again a central role here is played by (random) matrices with the so-called *Restricted Isometry Property* RIP, cf. [11].

Definition 3.4 (Restricted Isometry Property). *A $k \times d$ matrix M is said to have the Restricted Isometry Property of order $K \le d$ and level $\delta \in (0,1)$ if*

$$(1-\delta)\|x\|_{\ell_2^d}^2 \le \|Mx\|_{\ell_2^k}^2 \le (1+\delta)\|x\|_{\ell_2^d}^2$$

for all K-sparse $x \in \Sigma_K = \{z \in \mathbb{R}^d : \#supp\ (z) \le K\}$.

Both the typical matrices used in Johnson-Lindenstrauss embeddings (cf. Lemma 2.1) and matrices with RIP used in compressed sensing are usually generated at random. It was observed by [3] and [36], that there is an intimate connection between these two notions. A simple reformulation of the arguments of [3] yields the following.

Theorem 3.5 (Baraniuk, Davenport, DeVore, and Wakin). *Let M be a $k \times d$ matrix drawn at random which satisfies*

$$(1-\delta/2)\|x\|_{\ell_2^d}^2 \le \|Mx\|_{\ell_2^k}^2 \le (1+\delta/2)\|x\|_{\ell_2^d}^2, \quad x \in \mathcal{P}$$

for every set $\mathcal{P} \subset \mathbb{R}^d$ with $\#\mathcal{P} \le \left(\frac{12ed}{\delta K}\right)^K$ with probability $0 < \nu < 1$. Then M satisfies the Restricted Isometry Property of order K and level $\delta/3$ with probability at least equal to ν.

Combined with several rather elementary constructions of Johnson-Lindenstrauss embedding matrices available in literature, cf. [1] and [22], this result provides a simple construction of RIP matrices. The converse direction, namely the way from RIP matrices to matrices suitable for Johnson-Lindenstrauss embedding was discovered only recently in [36].

THEOREM 3.6 (Krahmer and Ward). *Fix $\eta > 0$ and $\varepsilon > 0$, and consider a finite set $\mathcal{P} \subset \mathbb{R}^d$ of cardinality $|\mathcal{P}| = \mathcal{N}$. Set $K \geq 40 \log \frac{4\mathcal{N}}{\eta}$, and suppose that the $k \times d$ matrix $\tilde{M}$ satisfies the Restricted Isometry Property of order K and level $\delta \leq \varepsilon/4$. Let $\xi \in \mathbb{R}^d$ be a Rademacher sequence, i.e., uniformly distributed on $\{-1,1\}^d$. Then with probability exceeding $1-\eta$,*

$$(1-\varepsilon)\|x\|_{\ell_2^d}^2 \leq \|Mx\|_{\ell_2^k}^2 \leq (1+\varepsilon)\|x\|_{\ell_2^d}^2.$$

uniformly for all $x \in \mathcal{P}$, where $M := \tilde{M}\operatorname{diag}(\xi)$, where $\operatorname{diag}(\xi)$ is a $d \times d$ diagonal matrix with ξ on the diagonal.

We refer to [41] for additional details.

REMARK 2.

Notice that M as constructed in Theorem 3.6 is both a Johnson-Lindenstrauss embedding and a matrix with RIP, because

$$\begin{aligned}(1-\delta)\|x\|_{\ell_2^d}^2 = (1-\delta)\|\operatorname{diag}(\xi)x\|_{\ell_2^d}^2 &\leq \|\underbrace{\tilde{M}\operatorname{diag}(\xi)}_{:=M}x\|_{\ell_2^k}^2 \\ &\leq (1+\delta)\|\operatorname{diag}(\xi)x\|_{\ell_2^d}^2 = (1+\delta)\|x\|_{\ell_2^d}^2.\end{aligned}$$

The matrices considered in Section 2 satisfy with high probability the RIP with

$$K = \mathcal{O}\left(\frac{k}{1+\log(d/k)}\right).$$

Equipped with the notion of RIP matrices we may state the main result of the theory of compressed sensing, which we shall use for the recovery of the dynamical system in $\mathbb{R}^d$.

THEOREM 3.7. *Assume that the matrix $M \in \mathbb{R}^{k\times d}$ has the RIP of order $2K$ and level*

$$\delta_{2K} < \frac{2}{3+\sqrt{7/4}} \approx 0.4627.$$

Then the following holds for all $x \in \mathbb{R}^d$. Let the low-dimensional approximation $y = Mx + \eta$ be given with $\|\eta\|_{\ell_2^k} \leq C\varepsilon$. Let $x^{\#}$ be the solution of

$$\min_{z\in\mathbb{R}^d} \|z\|_{\ell_1^d} \quad \text{subject to } \|Mz - y\|_{\ell_2^k} \leq \|\eta\|_{\ell_2^k}. \tag{3.16}$$

Then

$$\|x - x^{\#}\|_{\ell_2^d} \leq C_1\varepsilon + C_2\frac{\sigma_K(x)_{\ell_1^d}}{\sqrt{K}}$$

for some constants $C_1, C_2 > 0$ that depend only on δ_{2K}, and $\sigma_K(x)_{\ell_1^d} = \inf_{z:\#supp\ (z)\leq K}\|z-x\|_{\ell_1^d}$ is the best-K-term approximation in ℓ_1^d. This result says that provided the stability relationship (3.15), we can approximate the individual trajectories $x_i(t)$, for each $t \in [0,T]$ fixed, by a vector $x_i^{\#}(t)$ solution of an optimization problem of the type (3.16), and the accuracy of the approximation depends on the best-K-term approximation error $\sigma_K(x_i(t))_{\ell_1^d}$. Actually, when $x_i(t)$ is a vector in $\mathbb{R}^d$ with few large entries in absolute value, then $x_i^\#(t) \approx x_i(t)$ is a very good approximation, up to the ineliminable ε-distortion. However, if the vector $x_i(t)$ has many relevant entries, then this approximation will be rather poor. One possibility to improve the recovery error is to increase the dimension k (leading to a smaller distortion parameter $\varepsilon > 0$ in the Johnson-Lindenstrauss embedding). But we would like to explore another possibility, namely of projecting and simulating *in parallel and independently* the dynamical system L-times in the lower dimension k

$$\dot{y}_i^\ell = M^\ell f_i(\mathcal{D}' y^\ell) + \sum_{j=1}^N f_{ij}(\mathcal{D}' y^\ell) y_j^\ell\,, \qquad y_i^\ell(0) = M^\ell x_i^0\,, \quad \ell = 1, \ldots, L. \tag{3.17}$$

Let us give a brief overview of the corresponding error estimates. The number of points needed in every of the cases is $\mathcal{N} \approx N \times n_0$, where N is the number of agents and $n_0 = T/h$ is the number of iterations.

- We perform 1 projection and simulation in $\mathbb{R}^k$: Then $\varepsilon = \mathcal{O}\Big(\sqrt{\frac{\log \mathcal{N}}{k}}\Big)$, $K = \mathcal{O}\left(\frac{k}{1+\log(d/k)}\right)$ and an application of Theorem 3.7 leads to

$$\|x_i(t) - x_i^\#(t)\|_{\ell_2^d} \le C'(t) \left(\sqrt{\frac{\log \mathcal{N}}{k}} + \frac{\sigma_K(x_i(t))_{\ell_1^d}}{\sqrt{K}} \right). \tag{3.18}$$

  Here, $C'(t)$ combines both the constants from Theorem 3.7 and the time-dependent $C(t)$ from (3.15). So, to reach the precision of order $C'(t)\epsilon > 0$, we have to chose $k \in \mathbb{N}$ large enough, such that $\sqrt{\frac{\log \mathcal{N}}{k}} \le \epsilon$ and $\frac{\sigma_K(x_i(t))_{\ell_1^d}}{\sqrt{K}} \le \epsilon$. We then need $k \times N^2$ operations to evaluate the adjacency matrix.
- We perform 1 projection and simulation in $\mathbb{R}^{L\times k}$: Then $\varepsilon' = \mathcal{O}\Big(\sqrt{\frac{\log \mathcal{N}}{Lk}}\Big)$ and $K' = \mathcal{O}\left(\frac{Lk}{1+\log(d/Lk)}\right)$ and an application of Theorem 3.7 leads to

$$\|x_i(t) - x_i^\#(t)\|_{\ell_2^d} \le C'(t) \left(\sqrt{\frac{\log \mathcal{N}}{Lk}} + \frac{\sigma_{K'}(x_i(t))_{\ell_1^d}}{\sqrt{K'}} \right). \tag{3.19}$$

  The given precision of order $C'(t)\epsilon > 0$, may be then reached by to chosing $k, L \in \mathbb{N}$ large enough, such that $\sqrt{\frac{\log \mathcal{N}}{Lk}} \le \epsilon$ and $\frac{\sigma_{K'}(x_i(t))_{\ell_1^d}}{\sqrt{K'}} \le \epsilon$. We then need $Lk \times N^2$ operations to evaluate the adjacency matrix.
- We perform L independent and paralel projections and simulations in $\mathbb{R}^k$: Then we assemble the following system corresponding to (3.17)

$$\mathcal{M}x = \begin{pmatrix} M^1 \\ M^2 \\ \cdots \\ \cdots \\ M^L \end{pmatrix} x_i = \begin{pmatrix} y_i^1 \\ y_i^2 \\ \cdots \\ \cdots \\ y_i^L \end{pmatrix} - \begin{pmatrix} \eta_i^1 \\ \eta_i^2 \\ \cdots \\ \cdots \\ \eta_i^L \end{pmatrix},$$

  where for all $\ell = 1, \ldots, L$ the matrices $M^\ell \in \mathbb{R}^{k\times d}$ are (let us say) random matrices with each entry generated independently with respect to the properly normalised Gaussian distribution as described in Section 2. Then $\mathcal{M}/\sqrt{L}$ is a $Lk \times d$ matrix with Restricted Isometry Property of order $K' = \mathcal{O}\left(\frac{Lk}{1+\log(d/Lk)}\right)$ and level $\delta < 0.4627$. The initial disstortion of each of the projections is still $\varepsilon = \mathcal{O}\Big(\sqrt{\frac{\log \mathcal{N}}{k}}\Big)$. Therefore, by applying Theorem 3.7, we can compute $x_i^{\#}(t)$ such that

$$\|x_i(t) - x_i^{\#}(t)\|_{\ell_2^k} \le C'(t)\left(\sqrt{\frac{\log \mathcal{N}}{k}} + \frac{\sigma_{K'}(x_i(t))_{\ell_1^d}}{\sqrt{K'}}\right). \tag{3.20}$$

The larger is the number L of projections we perform, the bigger is K' and the smaller is the second summand in (3.20); actually $\sigma_{K'}(x_i(t))_{\ell_1^d}$ vanishes for $K' \ge d$. Unfortunately, the paralelisation can not help to reduce the initial distortion $\varepsilon > 0$. To reach again the precision of order $C'(t)\epsilon > 0$, we have to chose $k \in \mathbb{N}$ large enough, such that $\sqrt{\frac{\log \mathcal{N}}{k}} \le \epsilon$. Then we chose $L \ge 1$ large enough such that $\frac{\sigma_{K'}(x_i(t))_{\ell_1^d}}{\sqrt{K'}} \le \epsilon$. We again need $k \times N^2$ operations to evaluate the adjacency matrix.

In all three cases, we obtain the estimate

$$\|x_i(t) - x_i^{\#}(t)\|_{\ell_2^d} \le C'(t)\left(\varepsilon + \frac{\sigma_K(x_i(t))_{\ell_1^d}}{\sqrt{K}}\right), \tag{3.21}$$

where the corresponding values of $\varepsilon > 0$ and K together with the number of operations needed to evaluate the adjacency matrix may be found in the following table.

	ε	K	number of operations
1 projection into $\mathbb{R}^k$	$\mathcal{O}\Big(\sqrt{\frac{\log \mathcal{N}}{k}}\Big)$	$\mathcal{O}\left(\frac{k}{1+\log(d/k)}\right)$	$k \times N^2$
1 projection into $\mathbb{R}^{L\times k}$	$\mathcal{O}\Big(\sqrt{\frac{\log \mathcal{N}}{Lk}}\Big)$	$\mathcal{O}\left(\frac{Lk}{1+\log(d/Lk)}\right)$	$Lk \times N^2$
L projections into $\mathbb{R}^k$	$\mathcal{O}\Big(\sqrt{\frac{\log \mathcal{N}}{k}}\Big)$	$\mathcal{O}\left(\frac{Lk}{1+\log(d/Lk)}\right)$	$k \times N^2$

3.4.2. Manifold recovery. In recent papers [4, 47, 33], the concepts of compressed sensing and sparse recovery were extended to vectors on smooth manifolds. These methods could become very useful in our context if (for any reason) we would have an apriori knowledge that the trajectories $x_i(t)$ keep staying on or near such a smooth manifold. We leave this direction open for future research.

3.5. Numerical experiments. In this section we illustrate the practical use and performaces of our projection method for the Cucker-Smale system (2.11)–(2.12). As already mentioned, this system models the emergence of consensus in a group of interacting agents, trying to align with their neighbors. The qualitative behavior of its solutions is formulated by this well known result [20, 21, 32]:

THEOREM 3.8. *Let $(x_i(t), v_i(t))$ be the solutions of (2.11)–(2.12). Let us define the fluctuation of positions around the center of mass $x_c(t) = \frac{1}{N}\sum_{i=1}^N x_i(t)$, and, resp., the fluctuation of the rate of change around its average $v_c(t) = \frac{1}{N}\sum_{i=1}^N v_i(t)$ as*

$$\Lambda(t) = \frac{1}{N}\sum_{i=1}^N \|x_i(t) - x_c(t)\|_{\ell_2^d}^2, \qquad \Gamma(t) = \frac{1}{N}\sum_{i=1}^N \|v_i(t) - v_c(t)\|_{\ell_2^d}^2.$$

Then if either $\beta \leq 1/2$ or the initial fluctuations $\Lambda(0)$ and $\Gamma(0)$ are small enough (see [20] for details), then $\Gamma(t) \to 0$ as $t \to \infty$.

The phenomenon of $\Gamma(t)$ tending to zero as $t \to \infty$ is called *flocking* or *emergence of consensus*. If $\beta > 1/2$ and the initial fluctuations are not small, it is not known whether a given initial configuration will actually lead to flocking or not, and the only way to find out the possible formation of *consensus patterns* is to perform numerical simulations. However, these can be especially costly if the number of agents N and the dimension d are large; the algorithmic complexity of the calculation is $\mathcal{O}(d \times N^2)$. Therefore, a significant reduction of the dimension d, which can be achieved by our projection method, would lead to a corresponding reduction of the computational cost.

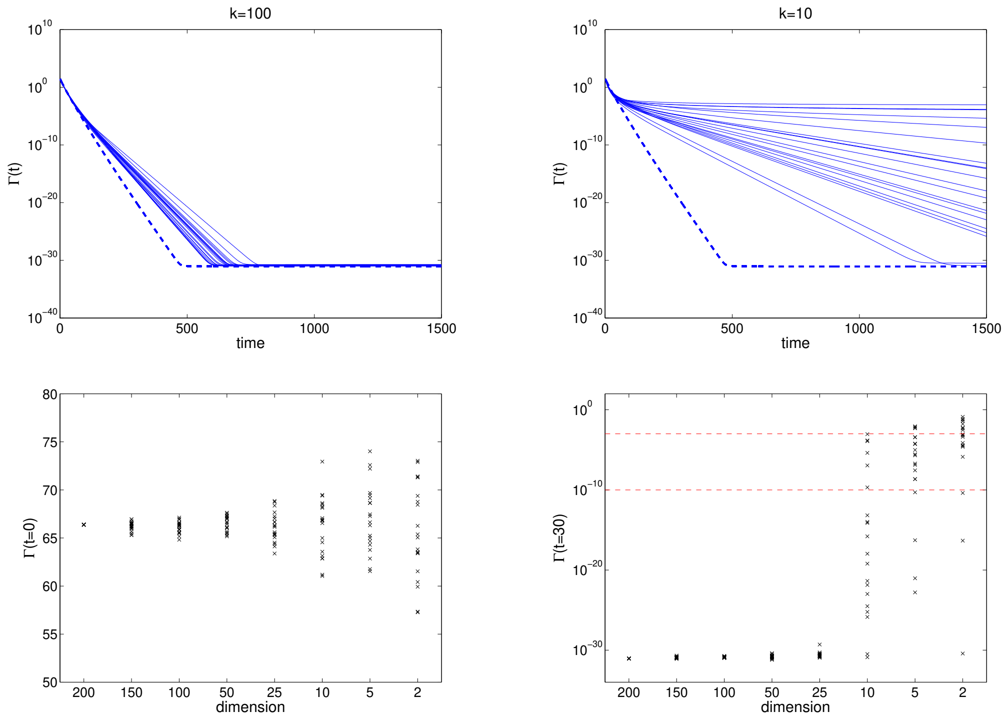

Fig. 3.2. *Numerical results for $\beta = 1.5$: First row shows the evolution of $\Gamma(t)$ of the system projected to dimension $k = 100$ (left) and $k = 10$ (right) in the twenty realizations, compared to the original system (bold dashed line). Second row shows the initial values $\Gamma(t = 0)$ and final values $\Gamma(t = 30)$ in all the performed simulations.*

We illustrate this fact by a numerical experiment, where we choose $N = 1000$ and $d = 200$, i.e., every agent i is determined by a 200-dimensional vector x_i of its state and a 200-dimensional vector v_i giving the rate of change of its state. The initial datum (x^0, v^0) is generated randomly, every component of x^0 being drawn independently from the uniform distribution on $[0, 1]$ and every component of v^0 being drawn independently from the uniform distribution on $[-1, 1]$. We choose $\beta = 1.5$, 1.62 and 1.7, and with every of these values we perform the following set of simulations:

1. Simulation of the original system in 200 dimensions.

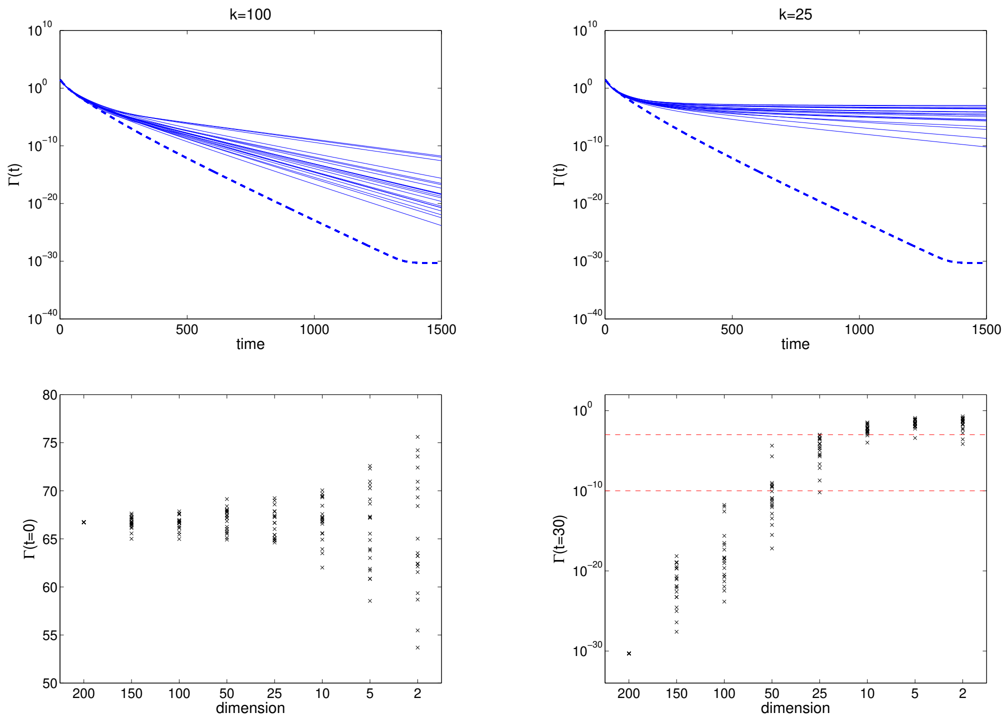

FIG. 3.3. *Numerical results for $\beta = 1.62$: First row shows the evolution of $\Gamma(t)$ of the system projected to dimension $k = 100$ (left) and $k = 25$ (right) in the twenty realizations, compared to the original system (bold dashed line). Second row shows the initial values $\Gamma(t = 0)$ and final values $\Gamma(t = 30)$ in all the performed simulations.*

2. Simulations in lower dimensions k: the initial condition (x^0, v^0) is projected into the k-dimensional space with a random Johnson-Lindenstrauss projection matrix M with Gaussian entries. The dimension k takes the values 150, 100, 50, 25, 10, 5, and 2. For every k, we perform the simulation twenty times, each time with a new random projection matrix M.

All the simulations were implemented in MATLAB, using 1500 steps of the forward Euler method with time step size 0.02. The paths of $\Gamma(t)$ from the twenty experiments with $k = 100$ and $k = 25$ or $k = 10$ are shown in the first rows of Figs. 3.2, 3.3 and, resp., 3.4 for $\beta = 1.5$, 1.62 and, resp., 1.7.

The information we are actually interested in is whether flocking takes place, in other words, whether the fluctuations of velocities $\Gamma(t)$ tend to zero. Typically, after an initial phase, the graph of $\Gamma(t)$ gives a clear indication either about exponentially fast convergence to zero (due to rounding errors, "zero" actually means values of the order 10^{-30} in the simulations) or about convergence to a positive value. However, in certain cases the decay may be very slow and a very long simulation of the system would be needed to see if the limiting value is actually zero or not. Therefore, we propose the following heuristic rules to decide about flocking from numerical simulations:

- If the value of Γ at the final time $t = 30$ is smaller than 10^{-10}, we conclude that flocking took place.

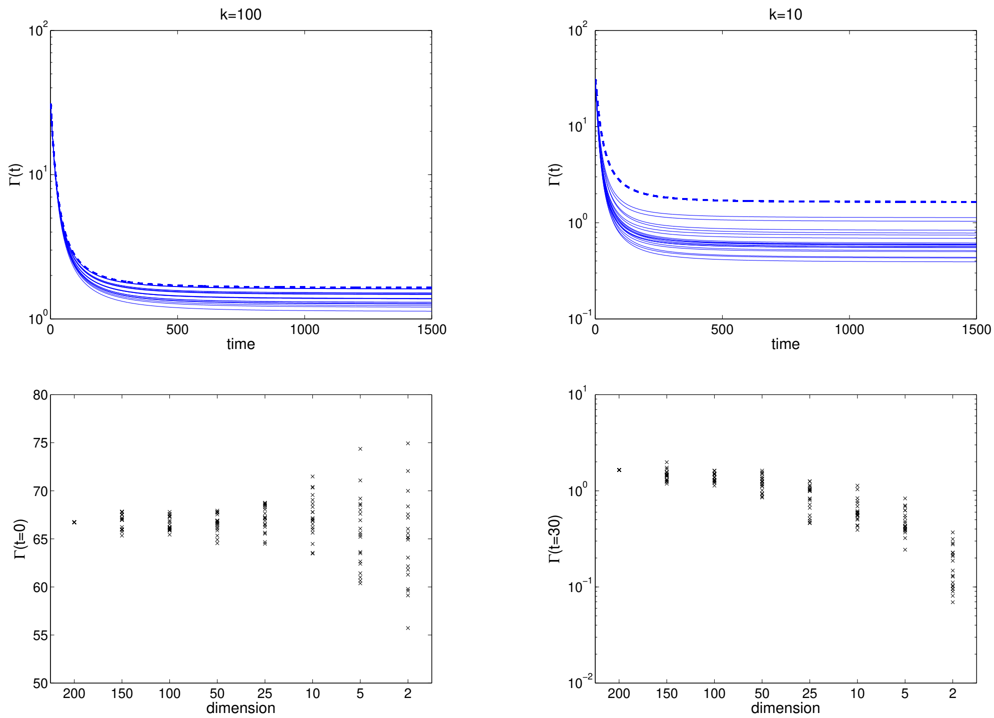

FIG. 3.4. *Numerical results for $\beta = 1.7$: First row shows the evolution of $\Gamma(t)$ of the system projected to dimension $k = 100$ (left) and $k = 10$ (right) in the twenty realizations, compared to the original system (bold dashed line). Second row shows the initial values $\Gamma(t = 0)$ and final values $\Gamma(t = 30)$ in all the performed simulations.*

- If the value of $\Gamma(30)$ is larger than 10^{-3}, we conclude that flocking did not take place.
- Otherwise, we do not make any conclusion.

In the second rows of Figs. 3.2, 3.3 and 3.4 we present the initial and final values of Γ of the twenty simulations for all the dimensions k, together with the original dimension $d = 200$. In accordance with the above rules, flocking takes place if the final value of Γ lies below the lower dashed line, does not take place if it lies above the upper dashed line, otherwise the situation is not conclusive. The results are summarized in Table 3.1.

Experience gained with a large amount of numerical experiments shows the following interesting fact: The flocking behavior of the Cucker-Smale system is very stable with respect to the Johnson-Lindenstrauss projections. Usually, the projected systems show the same flocking behavior as the original one, even if the dimension is reduced dramatically, for instance from $d = 200$ to $k = 10$ (see Figs 3.2 and 3.4). This stability can be roughly explained as follows: Since the flocking behavior depends mainly on the initial values of Γ and Λ, which are statistical properties of the random distributions used for the generation of initial data, and since N is sufficiently large, the concentration of measure phenomenon takes place. Its effect is that the initial values of the fluctuations of the projected data are very close to the original ones, and

$\beta = 1.5$			
dim	pos	neg	??
200	1	0	0
150	20	0	0
100	20	0	0
50	20	0	0
25	20	0	0
10	14	0	6
5	4	4	12
2	3	8	9

$\beta = 1.62$			
dim	pos	neg	??
200	1	0	0
150	20	0	0
100	20	0	0
50	13	0	7
25	1	1	18
10	0	18	2
5	0	19	1
2	0	18	2

$\beta = 1.7$			
dim	pos	neg	??
200	0	1	0
150	0	20	0
100	0	20	0
50	0	20	0
25	0	20	0
10	0	20	0
5	0	20	0
2	0	20	0

TABLE 3.1

Statistics of the flocking behaviors of the systems in the original dimension $d = 200$ *and in the projected dimensions. With* $\beta = 1.5$ *and* $\beta = 1.62$*, the original system (*$d = 200$*) exhibited flocking behavior. With* $\beta = 1.5$*, even after random projections into* 25 *dimensions, the system exhibited flocking in all* 20 *repetitions of the experiment, and still in* 14 *cases in dimension* 10*. With* $\beta = 1.62$*, the deterioration of the flocking behavior with decreasing dimension was much faster, and already in dimension* 25 *the situation was not conclusive. This is related to the fact that the value* $\beta = 1.62$ *was chosen to intentionally bring the system close to the borderline between flocking and non-flocking. Finally, with* $\beta = 1.7$*, the original system did not flock, and, remarkably, all the projected systems (even to two dimensions) exhibit the same behavior.*

thus the flocking behavior is (typically) the same. There is only a narrow interval of values of β (in our case this interval is located around the value $\beta = 1.62$), which is a borderline region between flocking and non-flocking, and the projections to lower dimensions spoil the flocking behavior, see Fig 3.3. Let us note that in our simulations we were only able to detect cases when flocking took place in the original system, but did not take place in some of the projected ones. Interestingly, we never observed the inverse situation, a fact which we are not able to explain satisfactorily. In fact, one can make other interesting observations, deserving further investigation. For instance, Figs. 3.2 and 3.3 show that if the original system exhibits flocking, then the curves of $\Gamma(t)$ of the projected systems tend to lie above the curve of $\Gamma(t)$ of the original one. The situation is reversed if the original system does not flock, see Fig. 3.4.

From a practical point of view, we can make the following conclusion: To obtain an indication about the flocking behavior of a highly dimensional Cucker-Smale system, it is typically satisfactory to perform a limited number of simulations of the system projected into a much lower dimension, and evaluate the statistics of their flocking behavior. If the result is the same for the majority of simulations, one can conclude that the original system very likely has the same flocking behavior as well.

4. Mean-field limit and kinetic equations in high dimension. In the previous sections we were concerned with tractable simulation of the dynamical systems of the type (1.1) when the dimension d of the parameter space is large. Another source of possible intractability in numerical simulations appears in the situation where the number of agents N is very large. Therefore, in the next sections we consider the so-called *mean-field limit* of (1.1) as $N \to \infty$, where the evolution of the system is described by time-dependent probability measures $\mu(t)$ on $\mathbb{R}^d$, representing the density distribution of agents, and satisfying mesoscopic partial differential equations of the type (4.1). This strategy originated from the kinetic theory of gases, see [16] for classical references. We show how our projection method can be applied for dimensionality reduction of the corresponding kinetic equations and explain how the probability measures can be approximated by atomic measures. Using the concept of *delayed curse of dimension* known from optimal integration problems in high dimension, we show that under the assumption that the measure concentrates along low-dimensional hyperplanes (and more in general alone low-dimensional sets or manifolds), it can be approximated by atomic measures with sub-exponential (with respect to d) number of atoms. Therefore, we can possibly approximate the (intractable) solution of the original system by a tractable simulation of the corresponding system of atoms as well as lower-dimensional kinetic equations.

4.1. Formal derivation of mean-field equations. In this section we briefly explain how the mean-field limit description corresponding to (1.1) can be derived. This is given, under suitable assumptions on the family of the governing functions $\mathcal{F}_N = \{f_i, f_{ij} : i, j = 1, \dots N\}$, by the general formula

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mathcal{H}_{\mathcal{F}}[\mu]\mu) = 0, \tag{4.1}$$

where $\mathcal{H}_{\mathcal{F}}[\mu]$ is a field in $\mathbb{R}^d$, determined by the sequence $\mathcal{F} = (\mathcal{F}_N)_{N\in\mathbb{N}}$.

In order to provide an explicit example, we show how to formally derive the mean field limit of systems of the type

$$\dot{x}_i = v_i\,, \tag{4.2}$$

$$\dot{v}_i = \sum_{j=1}^N f_{ij}^{vv}(\mathcal{D}x, \mathcal{D}v) v_j + \sum_{j=1}^N f_{ij}^{vx}(\mathcal{D}x) x_j\,, \tag{4.3}$$

with

$$f_{ij}^{vx}(\mathcal{D}x) = -\frac{\delta_{ij}}{N} \sum_{k\neq i} u(\|x_i - x_k\|_{\ell_2^d}) + \frac{1-\delta_{ij}}{N} u(\|x_i - x_j\|_{\ell_2^d})\,,$$

$$f_{ij}^{vv}(\mathcal{D}x, \mathcal{D}v) = \delta_{ij}\left(h(\|v_i\|^2_{\ell_2^d}) - \frac{1}{N}\sum_{k=1}^N g(\|x_i - x_k\|_{\ell_2^d}) \right) + \frac{1-\delta_{ij}}{N} g(\|x_i - x_j\|_{\ell_2^d})\,.$$

Note that for suitable choices of the functions h, g, u this formalism includes both the Cucker-Smale model (2.11)–(2.12) and D'Orsogna model (3.13)–(3.14). We define the empirical measure associated to the solutions $x_i(t)$, $v_i(t)$ of (4.2)–(4.3) as

$$\mu^N(t) := \mu^N(t, x, v) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i(t)}(x)\delta_{v_i(t)}(v)\,.$$

Taking a smooth, compactly supported test function $\xi \in C_0^\infty(\mathbb{R}^{2d})$ and using (4.2)–(4.3), one easily obtains by a standard formal calculation (see [14])

$$\begin{aligned} \frac{d}{dt}\langle \mu^N(t), \xi\rangle &= \frac{d}{dt}\left(\frac{1}{N}\sum_{i=1}^N \xi(x_i(t), v_i(t)) \right) \\ &= \int_{\mathbb{R}^{2d}} \nabla_x \xi(x,v)\cdot v \,\mathrm{d}\mu^N(t,x,v) + \int_{\mathbb{R}^{2d}} \nabla_v \xi(x,v)\cdot \mathcal{H}[\mu^N(t)](x,v)\,\mathrm{d}\mu^N(t,x,v)\,, \end{aligned} \tag{4.4}$$

with

$$\mathcal{H}[\mu](x,v) = h(\|v\|_{\ell_2^d})v + \int_{\mathbb{R}^{2d}} g(\|x-y\|_{\ell_2^d})(w-v)\,\mathrm{d}\mu(y,w) + \int_{\mathbb{R}^{2d}} u(\|x-y\|_{\ell_2^d})(y-x)\,\mathrm{d}\mu(y,w)\,.$$

We now assume weak convergence of a subsequence of $(\mu^N(t))_{N\in\mathbb{N}}$ to a time-dependent measure $\mu(t) = \mu(t,x,v)$ and boundedness of its first order moment, which indeed can be established rigorously for the Cucker-Smale and D'Orsogna systems (see [32], [40]). Then, passing to the limit $N \to \infty$ in (4.4), one obtains in the strong formulation that μ is governed by

$$\frac{\partial \mu}{\partial t}(t,x,v) + v \cdot \nabla_x \mu(t,x,v) + \nabla_v \cdot (\mathcal{H}[\mu(t)](x,v)\mu(t,x,v)) = 0\,,$$

which is an instance of the general prototype (4.1).

Using the same formal arguments as described above, one can easily derive mean field limit equations corresponding to (1.1) with different choices of the family $\mathcal{F}$.

4.2. Monge-Kantorovich-Rubinstein distance and stability. In several relevant cases, including the Cucker-Smale and D'Orsogna systems [13], solutions of equations of the type (4.1) are stable with respect to suitable distances. We consider the space $\mathcal{P}_1(\mathbb{R}^d)$, consisting of all probability measures on $\mathbb{R}^d$ with finite first moment. In $\mathcal{P}_1(\mathbb{R}^d)$ and for solutions of (4.1), a natural metric to work with is the so-called *Monge-Kantorovich-Rubinstein distance* [46],

$$W_1(\mu,\nu) := \sup\{|\langle \mu - \nu, \xi\rangle| = \left|\int_{\mathbb{R}^d} \xi(x) d(\mu-\nu)(x)\right|, \xi \in \mathrm{Lip}(\mathbb{R}^d), \mathrm{Lip}(\xi) \le 1\}. \tag{4.5}$$

We further denote $\mathcal{P}_c(\mathbb{R}^d)$ the space of compactly supported probability measures on $\mathbb{R}^d$. In particular, throughout the rest of this paper, we will assume that for two any compactly supported measure valued weak solutions $\mu(t), \nu(t) \in C([0,T], \mathcal{P}_c(\mathbb{R}^d))$ of (4.1) we have the following stability inequality

$$W_1(\mu(t),\nu(t)) \le C(t) W_1(\mu(0),\nu(0)), \quad t \in [0,T], \tag{4.6}$$

where $C(t)$ is a positive increasing function of t with $C(0) > 0$, independent of the dimension d. We address the interested reader to [13, Section 4] for a sample of general conditions on the vector field $\mathcal{H}[\mathcal{F}](\mu)$ which guarantee stability (4.6) for solutions of equations (4.1).

4.3. Dimensionality reduction of kinetic equations. Provided a high dimensional measure valued solution to the equation

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mathcal{H}_{\mathcal{F}}[\mu]\mu) = 0, \quad \mu(0) = \mu_0 \in \mathcal{P}_c(\mathbb{R}^d)\,, \tag{4.7}$$

we will study the question whether its solution can be approximated by suitable projections in lower dimension.

Given a probability measure $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, its projection into $\mathbb{R}^k$ by means of a matrix $M : \mathbb{R}^d \to \mathbb{R}^k$ is given by the *push-forward* measure $\mu_M := M\#\mu$,

$$\langle \mu_M, \varphi\rangle := \langle \mu, \varphi(M\cdot)\rangle \quad \text{for all } \varphi \in \mathrm{Lip}(\mathbb{R}^k). \tag{4.8}$$

Let us mention two explicit examples:

- If $\mu^N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ is an atomic measure, we have $\langle \mu^N_M, \varphi\rangle = \langle \mu^N, \varphi(M\cdot)\rangle = \frac{1}{N}\sum_{i=1}^N \varphi(Mx_i)$. Therefore,

$$\mu^N_M = \frac{1}{N}\sum_{i=1}^N \delta_{Mx_i}\,.$$

- If μ is absolutely continuous with respect to the Lebesgue measure, i.e., it is a function in $L^1(\mathbb{R}^d)$, the calculation requires a bit more effort: Let us consider $M^\dagger$ the pseudo-inverse matrix of M. Recall that $M^\dagger = M^*(MM^*)^{-1}$ is a right inverse of M, and $M^\dagger M$ is the orthogonal projection onto the range of M^*. Moreover, $x = M^\dagger Mx + \xi_x$, where $\xi_x \in \ker M$ for all $x \in \mathbb{R}^d$. According to these observations, we write

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(Mx)\mu(x)dx &= \int_{\mathbb{R}^d} \varphi(Mx)\mu(M^\dagger Mx + \xi_x)dx \\ &= \int_{\mathrm{ran} M^* \oplus \ker M} \varphi(Mx)\mu(M^\dagger Mx + \xi_x)dx \\ &= \int_{\mathrm{ran} M^*} \int_{\ker M} \varphi(Mv)\mu(M^\dagger Mv + v^\perp)dv^\perp dv\end{aligned}$$

Note now that $M_{|\mathrm{ran}M^*} : \mathrm{ran}M^* \to \mathrm{ran}M \approx \mathbb{R}^k$ is an isomorphism, hence $y = Mv$ implies the change of variables $dv = \det(M_{|\mathrm{ran}M^*})^{-1}dy = \det(MM^*)^{-1/2}dy$. Consequently, we have

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(Mx)\mu(x)dx &= \int_{\mathbb{R}^d} \varphi(Mx)\mu(M^\dagger Mx + \xi_x)dx \\ &= \int_{\mathrm{ran} M^*} \int_{\ker M} \varphi(Mv)\mu(M^\dagger Mv + v^\perp)dv^\perp dv \\ &= \int_{\mathbb{R}^k} \left(\frac{1}{\det(MM^*)^{1/2}} \int_{\ker M} \mu(M^\dagger y + v^\perp)dv^\perp \right) \varphi(y)dy\,,\end{aligned}$$

and

$$\mu_M(y) = \frac{1}{\det(MM^*)^{1/2}} \int_{\ker M} \mu(M^\dagger y + v^\perp)dv^\perp.$$

According to the notion of push-forward, we can consider the measure valued function $\nu \in C([0,T], \mathcal{P}_c(\mathbb{R}^k))$, solution of the equation

$$\frac{\partial \nu}{\partial t} + \nabla \cdot (\mathcal{H}_{\mathcal{F}_M}[\nu]\nu) = 0, \quad \nu(0) = (\mu_0)_M \in \mathcal{P}_c(\mathbb{R}^k), \tag{4.9}$$

where $(\mu_0)_M = M\#\mu_0$ and $\mathcal{F}_M = (\{Mf_i, f_{ij}, i,j = 1,\dots,N\})_{N\in\mathbb{N}}$. As for the dynamical system (3.5), also equation (4.9) is fully defined on the lower-dimensional space $\mathbb{R}^k$ and depends on the original high-dimensional problem exclusively by means of the initial condition.

The natural question at this point is whether the solution ν of (4.9) provides information about the solution μ of (4.7). In particular, similarly to the result of Theorem 3.1, we will examine whether the approximation

$$\nu(t) \approx \mu_M(t), \quad t \in [0,T],$$

in Monge-Kantorovich-Rubinstein distance is preserved in finite time. We depict the expected result by the following diagram:

$$\begin{array}{ccc} \mu(0) & \xrightarrow{t} & \mu(t) \\ \downarrow M & & \downarrow M \\ \nu(0) = (\mu_0)_M & \xrightarrow{t} & \nu(t) \approx \mu_M(t) \end{array} \quad .$$

This question will be addressed by approximation of the problem with atomic measures and by an application of Theorem 3.1 for the corresponding dynamical system, as concisely described by

$$\begin{array}{ccc} \mu & \overset{W_1(\mu,\mu^N)\lesssim\varepsilon}{\longrightarrow} & \mu^N \\ \downarrow M & & \downarrow M \\ \nu \approx \mu_M & \overset{W_1(\nu,\nu^N)\lesssim\varepsilon}{\longrightarrow} & \nu^N \approx \mu_M^N \end{array}$$

Let us now recall the framework and general assumptions for this analysis to be performed. We assume again that for all $N \in \mathbb{N}$ the family $\mathcal{F}_N = \{f_i, f_{ij} : i,j = 1,\dots N\}$ is composed of functions satisfying (1.2)-(1.4). Moreover, we assume that associated to $\mathcal{F} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ and to

$$\dot{x}_i(t) = f_i(\mathcal{D}x(t)) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}x(t))x_j(t), \tag{4.10}$$

we can define a mean-field equation

$$\frac{\partial\mu}{\partial t} + \nabla\cdot(\mathcal{H}[\mathcal{F}](\mu)\mu) = 0, \quad \mu(0) = \mu_0 \in \mathcal{P}_c(\mathbb{R}^d), \tag{4.11}$$

such that for any two compactly supported measure valued weak solutions $\mu(t), \nu(t) \in C([0,T], \mathcal{P}_c(\mathbb{R}^d))$ of (4.1) we have the following stability

$$W_1(\mu(t),\nu(t)) \le C(t)W_1(\mu(0),\nu(0)), \quad t \in [0,T], \tag{4.12}$$

where $C(t)$ is a positive increasing function of t, independent of the dimension d. We further require that corresponding assumptions, including stability, hold for the projected system (2.5) and kinetic equation (4.9). Then we have the following approximation result:

Theorem 4.1. *Let us assume that $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and there exist points $\{x_1^0,\dots,x_N^0\} \subset \mathbb{R}^d$, for which the atomic measure $\mu_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i^0}$ approximates μ_0 up to $\varepsilon > 0$ in Monge-Kantorovich-Rubinstein distance, in the following sense*

$$W_1(\mu_0,\mu_0^N) \le \varepsilon, \quad N = \mathcal{N}^{\overline{k}(\varepsilon)} \text{ for } \overline{k}(\varepsilon) \le d \text{ and } \overline{k}(\varepsilon) \to d \text{ for } \varepsilon \to 0. \tag{4.13}$$

Requirement (4.13) *is in fact called the* delayed curse of dimension *as explained below in detail in Section 4.5. Depending on $\varepsilon > 0$ we fix also*

$$k = k(\varepsilon) = \mathcal{O}(\varepsilon^{-2}\log(N)) = \mathcal{O}(\varepsilon^{-2}\log(\mathcal{N})\overline{k}(\varepsilon)).$$

Moreover, let $M : \mathbb{R}^d \to \mathbb{R}^k$ be a linear mapping which is a continuous Johnson-Lindenstrauss embedding *as in* (3.4) *for continuous in time trajectories $x_i(t)$ of* (4.10) *with initial datum $x_i(0) = x_i^0$. Let $\nu \in C([0,T], \mathcal{P}_c(\mathbb{R}^k))$ be the weak solution of*

$$\frac{\partial\nu}{\partial t} + \nabla\cdot(\mathcal{H}[\mathcal{F}_M](\nu)\nu) = 0, \tag{4.14}$$

$$\nu(0) = (\mu_0)_M \in \mathcal{P}_c(\mathbb{R}^k), \tag{4.15}$$

where $(\mu_0)_M = M\#\mu_0$. Then

$$W_1(\mu_M(t),\nu(t)) \le \mathcal{C}(t)\|M\|\varepsilon, \quad t \in [0,T], \tag{4.16}$$

where $\mathcal{C}(t)$ is an increasing function of t, with $\mathcal{C}(0) > 0$, which is at most polynomially growing with the dimension d.

Proof. Let us define $\nu^N(t)$ the solution to equation (4.14) with initial datum $\nu^N(0) = (\mu_0^N)_M$, or, equivalently,

$$\nu^N(t) = \frac{1}{N} \sum_{i=1}^{n} \delta_{y_i(t)},$$

where $y_i(t)$ is the solution of

$$\begin{aligned} \dot{y}_i &= f_i(\mathcal{D}'y) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}'y) y_j\,, \qquad i = 1, \ldots, N\,, \\ y_i(0) &= M x_i^0\,, \qquad i = 1, \ldots, N\,. \end{aligned}$$

We estimate

$$W_1(\mu_M(t), \nu(t)) \leq W_1(\mu_M(t), (\mu^N(t))_M) + W_1((\mu^N(t))_M, \nu^N(t)) + W_1(\nu^N(t), \nu(t)).$$

By using the definition of push-forward (4.8) and (4.13), the first term can be estimated by

$$\begin{aligned} W_1(\mu_M(t), (\mu^N(t))_M) &= \sup\{\langle \mu_M(t) - (\mu^N(t))_M, \varphi\rangle : \mathrm{Lip}(\varphi) \leq 1\} \\ &= \sup\{\langle \mu(t) - \mu^N(t), \varphi(M\cdot)\rangle : \mathrm{Lip}(\varphi) \leq 1\} \\ &\leq \|M\| W_1(\mu(t), \mu^N(t)) \leq \|M\| C(t)\varepsilon. \end{aligned}$$

We estimate now the second term

$$\begin{aligned} W_1((\mu^N(t))_M, \nu^N(t)) &= \sup\{\langle (\mu^N(t))_M - \nu^N(t), \varphi\rangle : \mathrm{Lip}(\varphi) \leq 1\} \\ &= \sup\{\frac{1}{N} \sum_{i=1}^{N} (\varphi(Mx_i(t)) - \varphi(y_i(t))) : \mathrm{Lip}(\varphi) \leq 1\} \\ &\leq \frac{1}{N} \sum_{i=1}^{N} \|Mx_i(t) - y_i(t)\|_{\ell_2^k}. \end{aligned}$$

We recall the uniform approximation of Theorem 3.1,

$$\|Mx_i(t) - y_i(t)\|_{\ell_2^k} \leq D(t)\varepsilon\,, \qquad i = 1, \ldots, N,$$

where $D(t)$ is the time-dependent function on the right-hand-side of (3.7). Hence

$$W_1(\mu_M(t), (\mu^N(t))_M) \leq D(t)\varepsilon.$$

We address now the upper estimate of the third term, by the assumed stability of the lower dimensional equation (4.9)

$$\begin{aligned} W_1(\nu^N(t), \nu(t)) &\leq C(t) W_1(\nu^N(0), \nu(0)) \\ &= C(t) W_1((\mu_0^N)_M, (\mu_0)_M) \\ &\leq C(t)\|M\| W(\mu_0^N, \mu_0) \leq C(t)\|M\|\varepsilon. \end{aligned}$$

We can fix $\mathcal{C}(t) = 2C(t)\|M\| + D(t)$, and, as observed in Theorem 3.3, we can assume without loss of generality that $\|M\| \leq \sqrt{\frac{d}{k}}$. Hence, $\mathcal{C}(t)$ depends at most polynomially with respect to the dimension d. $\square$

4.4. Approximation of probability measures by atomic measures and optimal integration. In view of the fundamental requirement (4.13) in Theorem 4.1, given $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$, we are interested to establish an upper bound to the best possible approximation in Monge-Kantorovich-Rubinstein distance by means of atomic measures $\mu_0^N = \frac{1}{N}\sum_{i=0}^{N-1}\delta_{x_i^0}$ with N atoms, i.e.,

$$\begin{aligned} \mathcal{E}_N(\mu_0) &:= \inf_{\mu_0^N = \frac{1}{N}\sum_{i=0}^{N-1}\delta_{x_i^0}} W_1(\mu_0, \mu_0^N) \\ &= \inf_{\{x_0^0,\dots,x_{N-1}^0\}\subset\mathbb{R}^d} \sup\Big\{\Big|\int_{\mathbb{R}^d}\xi(x)d\mu_0(x) - \frac{1}{N}\sum_{i=0}^{N-1}\xi(x_i^0)\Big| : \xi \in \mathrm{Lip}(\mathbb{R}^d), \mathrm{Lip}(\xi)\le 1\Big\}. \end{aligned} \tag{4.17}$$

In fact, once we identify the optimal points $\{x_0^0, \dots, x_{N-1}^0\}$, we can use them as initial conditions $x_i(0) = x_i^0$ for the dynamical system (4.10), and by using the stability relationship (4.6), we obtain

$$W_1(\mu(t), \mu^N(t)) \le C(T) W_1(\mu_0, \mu_0^N), \quad t \in [0, T], \tag{4.18}$$

where $\mu^N(t) = \frac{1}{N}\sum_{i=0}^{N-1}\delta_{x_i(t)}$, meaning that the solution of the partial differential equation (4.1) keeps optimally close to the particle solution of (4.10) also for successive time $t > 0$. Note that estimating (4.17) as a function of N is in fact a very classical problem in numerical analysis well-known as *optimal integration* with its high-dimensional behaviour being a relevant subject of the field of *Information Based Complexity* [39, 44].

The numerical integration of Lipschitz functions with respect to the Lebesgue measure and the study of its high-dimensional behaviour goes to Bakhvalov [2], but much more is known nowadays. We refer to [28] and [31] for the state of the art of quantization of probability distributions.

The scope of this section is to present some facets of these estimates and to reformulate them in terms of W_1 and $\mathcal{E}_N$. We emphasize that here and in what follows, we consider generic compactly supported probability measures μ, not necessarily absolutely continuous with respect to the Lebesgue measure. We start first by assuming $d = 1$, i.e., we work with a univariate measure $\mu \in \mathcal{P}_c(\mathbb{R})$ with support $\operatorname{supp}\mu \subset [a, b]$ and $\sigma := b - a > 0$. We define the points $x_0, \dots, x_{N-1}$ as the *quantiles* of the probability measure μ, i.e., $x_0 := a$ and

$$\frac{i}{N} = \int_{-\infty}^{x_i} d\mu(x), \quad i = 1, \dots, N-1. \tag{4.19}$$

This is notationally complemented by putting $x_N := b$. Note that by definition $\int_{x_i}^{x_{i+1}} d\mu(x) = \frac{1}{N}, i = 0, \dots, N-1$, and we have

$$\begin{aligned} \left|\int_{\mathbb{R}}\xi(x)d\mu(x) - \frac{1}{N}\sum_{i=0}^{N-1}\xi(x_i)\right| &= \left|\sum_{i=0}^{N-1}\int_{x_i}^{x_{i+1}}(\xi(x) - \xi(x_i))d\mu(x)\right| \\ &\le \sum_{i=0}^{N-1}\int_{x_i}^{x_{i+1}}|\xi(x) - \xi(x_i)|\, d\mu(x) \\ &\le \frac{\mathrm{Lip}(\xi)}{N}\sum_{i=0}^{N-1}(x_{i+1} - x_i) = \frac{\sigma\mathrm{Lip}(\xi)}{N}. \end{aligned} \tag{4.20}$$

Hence it is immediate to see that

$$\mathcal{E}_N(\mu) = \inf_{\mu^N = \frac{1}{N}\sum_{i=0}^{N-1}\delta_{x_i^0}} W_1(\mu, \mu^N) \le \frac{\sigma}{N}.$$

We would like to extend this estimate to higher dimension $d > 1$. However, for multivariate measures μ there is no such an easy upper bound and we necessarily need to restrict our class of measures μ to certain special cases. For simplicity and as a typical situation, here we address tensor product measures and sums of tensor products.

LEMMA 4.2. *Let $\mu^1, \dots, \mu^d \in \mathcal{P}_1(\mathbb{R})$ with $W_1(\mu^j, \mu^{j,N^j}) \le \varepsilon_j$, $j = 1, \dots, d$ for some $N^1, \dots, N^d \in \mathbb{N}$, $\varepsilon_1, \dots, \varepsilon_d > 0$ and $\mu^{j,N^j} := \frac{1}{N^j}\sum_{i=0}^{N^j-1}\delta_{x_i^j}$. Let $N = \prod_{i=1}^d N^i$. Then*

$$W_1(\mu^1 \otimes \dots \otimes \mu^d, \mu^N) \le \sum_{j=1}^d \varepsilon_j,$$

where

$$\mu^N := \frac{1}{N}\sum_{x\in X}\delta_x \quad \text{and} \quad X := \prod_{j=1}^d \{x_0^j, \dots, x_{N^j-1}^j\}.$$

Proof. The proof is based on a simple argument using a telescopic sum. For $j = 1, \dots, d+1$ we put

$$V_j := \frac{1}{\prod_{i=j}^d N^i}\sum_{i_j=0}^{N^j-1}\cdots\sum_{i_d=0}^{N^d-1}\int_{\mathbb{R}^{j-1}}\xi(x_1, \dots, x_{j-1}, x_{i_j}^j, \dots, x_{j_d}^d)d\mu^1(x_1)\dots d\mu^{j-1}(x_{j-1}).$$

Of course, if $j = 1$, then the integration over $\mathbb{R}^{j-1}$ is missing and if $j = d+1$ then the summation becomes empty. Now

$$\int_{\mathbb{R}^d}\xi(x)d\mu(x) - \frac{1}{\prod_{i=1}^d N^i}\sum_{i_1=0}^{N^1-1}\cdots\sum_{i_d=0}^{N^d-1}\xi(x_{i_1}^1, \dots, x_{i_d}^d) = \sum_{j=1}^d (V_{j+1} - V_j)$$

together with the estimate $|V_{j+1} - V_j| \le \varepsilon_j$ finishes the proof. □

Lemma 4.2 says, roughly speaking, that the tensor products of sampling points of univariate measures are good sampling points for the tensor product of the univariate measures. Next lemma deals with sums of measures.

LEMMA 4.3. *Let $\mu_1, \dots, \mu_L \in \mathcal{P}_1(\mathbb{R}^d)$ with $W_1(\mu_l, \mu_l^N) \le \varepsilon_l$, $l = 1, \dots, L$ for some $N \in \mathbb{N}$, $\varepsilon_1, \dots, \varepsilon_L > 0$ and $\mu_l^N := \frac{1}{N}\sum_{i=0}^{N-1}\delta_{x_{l,i}}$. Then*

$$W_1\Big(\frac{\mu_1 + \dots + \mu_L}{L}, \mu^{LN}\Big) \le \frac{1}{L}\sum_{l=1}^L \varepsilon_l,$$

where

$$\mu^{LN} := \frac{1}{LN}\sum_{x\in X}\delta_x = \frac{1}{L}\sum_{l=1}^L \mu_l^{N_l} \quad \text{and} \quad X := \bigcup_{l=1}^L \{x_{l,0}, \dots, x_{l,N-1}\}.$$

Proof. We use the homogeneity of the Monge-Kantorovich-Rubinstein distance $W_1(a\mu, a\nu) = aW_1(\mu,\nu)$ for $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $a \geq 0$ combined with its subadditivity $W_1(\mu_1+\mu_2,\nu_1+\nu_2) \leq W_1(\mu_1,\nu_1)+W_1(\mu_2,\nu_2)$ for $\mu_1,\mu_2,\nu_1,\nu_2 \in \mathcal{P}_1(\mathbb{R}^d)$. We obtain

$$W_1\Big(\frac{\mu_1+\cdots+\mu_L}{L}, \frac{\mu_1^N+\cdots+\mu_L^N}{L}\Big) \leq \frac{1}{L}\sum_{l=1}^L W_1(\mu_l,\mu_l^N) \leq \frac{1}{L}\sum_{l=1}^L \varepsilon_l.$$

□

Next corollary follows directly from Lemma 4.2 and Lemma 4.3.

COROLLARY 4.4. *(i) Let $\mu^1,\dots,\mu^d \in \mathcal{P}_1(\mathbb{R})$ and $N^1,\dots,N^d \in \mathbb{N}$. Then*

$$\mathcal{E}_N(\mu^1\otimes\cdots\otimes\mu^d) \leq \sum_{j=1}^d \mathcal{E}_{N^j}(\mu^j), \quad \textit{where} \quad N := N^1\cdots N^d.$$

(ii) Let $\mu_1,\dots,\mu_L \in \mathcal{P}_1(\mathbb{R}^d)$ and $N \in \mathbb{N}$. Then

$$\mathcal{E}_{LN}\Big(\frac{\mu_1+\cdots+\mu_L}{L}\Big) \leq \frac{1}{L}\sum_{l=1}^L \mathcal{E}_N(\mu_l).$$

4.5. Delayed curse of dimension. Although Lemma 4.2, Lemma 4.3 and Corollary 4.4 give some estimates of the Monge-Kantorovich-Rubinstein distance between general and atomic measures, the number of atoms needed may still be too large to allow the assumption (4.13) in Theorem 4.1 to be fulfilled. Let us for example consider the case, when $\mu^1 = \cdots = \mu^d$ in Lemma 4.2 and $\varepsilon_1 = \cdots = \varepsilon_d =: \varepsilon$. Then, of course, $N^1 = \cdots = N^d =: \mathcal{N}$ and we observe, that the construction given in Lemma 4.2 gives an atomic measure, which approximates μ up to the error $d\varepsilon$ using $\mathcal{N}^d$ atoms, hence with an exponential dependence on the dimension d. This effect is another instance of the well-known phenomenon of the *curse of dimension*.

However, in many real-life high-dimensional applications the objects of study (in our case the measure $\mu \in \mathcal{P}_c(\mathbb{R}^d)$) concentrate along low-dimensional hyperplanes (or, more general, along low-dimensional manifolds) [5, 6, 17, 18, 19]. The number of atoms necessary to approximate these measures behaves in a much better way, allowing the application of (4.13) and Theorem 4.1. To clarify this effect, let us consider $\mu = \mu^1\otimes\cdots\otimes\mu^d$ with supp $\mu^j \subset [a^j, b^j]$ and define $\sigma^j = b^j - a^j$. Let us assume, that $\sigma^1 \geq \sigma^2 \geq \cdots \geq \sigma^d > 0$ is a rapidly decreasing sequence. Furthermore, let $\varepsilon > 0$. Then we define $\overline{k} := \overline{k}(\varepsilon)$ to be the smallest natural number, such that

$$\sum_{k=\overline{k}(\varepsilon)+1}^d \sigma^k \leq \varepsilon/2$$

and put $N^k = 1$ for $k \in \{\overline{k}(\varepsilon)+1,\dots,d\}$. The numbers $N^1 = \cdots = N^{\overline{k}(\varepsilon)} = \mathcal{N}$ are chosen large enough so that

$$\frac{1}{\mathcal{N}}\sum_{k=1}^{\overline{k}(\varepsilon)} \sigma^k \leq \varepsilon/2.$$

Then Lemma 4.2 together with (4.19) state that there is an atomic measure μ^N with $N = \mathcal{N}^{\overline{k}(\varepsilon)}$ atoms, such that

$$W_1(\mu, \mu^N) \leq \sum_{k=1}^{d} \frac{\sigma^k}{N^k} \leq \varepsilon/2 + \varepsilon/2. \tag{4.21}$$

Hence, at the cost of assuming that the tensor product measure μ is concentrated along a k-dimensional coordinate subspace, we can always approximate the measure μ with accuracy ε by using an atomic measure supported on points whose number depends exponentially on $\overline{k} = \overline{k}(\varepsilon) \ll d$. However, if we liked to have $\varepsilon \to 0$, then $\overline{k}(\varepsilon) \to d$ and again we are falling under the curse of dimension. This delayed kicking in of the need of a large number of points for obtaining high accuracy in the approximation (4.21) is in fact the so-called *delayed curse of dimension*, expressed by assumption (4.13), a concept introduced first by Curbera in [15], in the context of optimal integration with respect to Gaussian measures in high dimension.

Let us only remark, that the discussion above may be easily extended (with help of Lemma 4.3) to sums of tensor product measures. In that case we obtain as atoms the so-called *sparse grids*, cf. [10]. Using suitable change of variables, one could also consider measures concentrated around (smooth) low-dimensional manifolds, but this goes beyond the scope of this work and we leave this direction open for future research.

Acknowledgments. We acknowledge the financial support provided by the START award "Sparse Approximation and Optimization in High Dimensions" no. FWF Y 432-N15 of the Fonds zur Förderung der wissenschaftlichen Forschung (Austrian Science Foundation). We would also like to thank Erich Novak for a discussion about optimal integration and for pointing us to some of the references given in Section 4.4.

REFERENCES

[1] D. Achlioptas, *Database-friendly random projections: Johnson-Lindenstrauss with binary coins*, J. Comput. Syst. Sci., 66 (2003), pp. 671–687.
[2] N. S. Bakhvalov, *On approximate computation of integrals*, Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem., 4 (1959), pp. 3–18.
[3] R. G. Baraniuk, M. Davenport, R. A. DeVore and M. Wakin, *A simple proof of the Restricted Isometry Property for random matrices*, Constr. Approx., 28 (2008), pp. 253–263.
[4] R. G. Baraniuk and M. B. Wakin, *Random projections of smooth manifolds*, Found. Comput. Math., 9 (2009), pp. 51–77.
[5] M. Belkin and P. Niyogi, *Laplacian eigenmaps and spectral techniques for embedding and clustering*, in: Advances in Neural Information Processing Systems 14 (NIPS 2001), MIT Press, Cambridge, 2001.
[6] M. Belkin and P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation*, *Neural Computation*, 6 (2003), pp. 1373–1396.
[7] P. Binev, A. Cohen, W. Dahmen, G. Petrova and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis methods*, preprint, 2010.
[8] F. Bolley, J. A. Canizo and J. A. Carrillo, *Stochastic mean-field limit: non-Lipschitz forces and swarming*, Math. Models Methods Appl. Sci., to appear.
[9] A. Buffa, Y. Maday, A. T. Patera, C. Prudhomme and G. Turinici, *A priori convergence of the greedy algorithm for the parameterized reduced basis*, preprint.
[10] H. Bungartz and M. Griebel, *Sparse grids*, Acta Numer., 13 (2004), pp. 147–269.
[11] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, Compte Rendus de l'Academie des Sciences, Paris, Serie I, 346 (2008), pp. 589–592.
[12] E. J. Candès, T. Tao and J. Romberg, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[13] J. A. Canizo, J. A. Carrillo and J. Rosado, *A well-posedness theory in measures for some kinetic models of collective motion*, Math. Models Methods Appl. Sci., to appear.

[14] J. A. Carrillo, M. Fornasier, G. Toscani and F. Vecil, *Particle, kinetic, hydrodynamic models of swarming*, in: Mathematical modeling of collective behavior in socio-economic and life-sciences, Birkhäuser, 2010.

[15] F. Curbera, *Delayed curse of dimension for Gaussian integration*, J. Complexity, 16 (2000), pp. 474–506.

[16] C. Cercignani, R. Illner and M. Pulvirenti, *The mathematical theory of dilute gases*, Springer series in Applied Mathematical Sciences 106, Springer, 1994.

[17] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure denition of data: Diffusion maps, part I.*, Proc. of Nat. Acad. Sci., 102 (2005), pp. 7426–7431.

[18] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure denition of data: Diffusion maps, part II.*, Proc. of Nat. Acad. Sci., 102 (2005), pp. 7432–7438.

[19] R. R. Coifman and S. Lafon, *Diffusion maps*, Appl. Comp. Harm. Anal., 21 (2006), pp. 5–30.

[20] F. Cucker and S. Smale, *Emergent behavior in flocks*, IEEE Trans. Automat. Control, 52 (2007), pp 852–862.

[21] F. Cucker and S. Smale, *On the mathematics of emergence*, Japan J. Math., 2 (2007), pp. 197–227.

[22] S. Dasgupta and A. Gupta, *An elementary proof of a theorem of Johnson and Lindenstrauss*, Random. Struct. Algorithms, 22 (2003), pp. 60–65.

[23] T. Dijkema, C. Schwab and R. Stevenson, *An adaptive wavelet method for solving high-dimensional elliptic PDEs*, Constr. Approx., 30 (2009), pp. 423–455.

[24] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[25] S. Foucart, *A note on ensuring sparse recovery via ℓ_1-minimization*, Appl. Comput. Harmon. Anal., 29 (2010), pp. 97–103.

[26] M. Fornasier, *Numerical methods for sparse recovery*, in: Theoretical Foundations and Numerical Methods for Sparse Recovery (ed. M. Fornasier), Volume 9 of Radon Series Comp. Appl. Math., deGruyter, pp. 93–200, 2010.

[27] M. Fornasier and H. Rauhut, *Compressive Sensing*, in: Handbook of Mathematical Methods in Imaging (ed. O. Scherzer), Springer, 2010.

[28] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics, 1730, Springer-Verlag, Berlin, 2000.

[29] M. Griebel and S. Knapek, *Optimized tensor-product approximation spaces*, Constr. Approx., 16 (2000), pp. 525–540.

[30] M. Griebel and P. Oswald, *Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems*, Adv. Comput. Math., 4 (1995), pp. 171–206.

[31] P. M. Gruber, *Optimum quantization and its applications*, Adv. Math. 186 (2004), pp. 456–497.

[32] S.-Y. Ha and E. Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinetic and Related models, 1 (2008), pp. 315–335.

[33] M. Iwen and M. Maggioni, *Approximation of points on low-dimensional manifolds via compressive measurements*, in preparation.

[34] W. B. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Contem. Math., 26 (1984), pp. 189–206.

[35] E. F. Keller and L. A. Segel, *Initiation of slime mold aggregation viewed as an instability*, J. Theoret. Biol. 26 (1970), pp. 399–415.

[36] F. Krahmer and R. Ward, *New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property*, preprint, 2010.

[37] Y. Maday, A. T. Patera and G. Turinici, *Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations*, C. R. Acad. Sci., Paris, Ser. I, Math., 335 (2002), pp. 289–294.

[38] R. C. B. Nadler, S. Lafon and I. Kevrekidis, *Diffusion maps, spectral clustering and the reaction coordinates of dynamical systems*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 113–127.

[39] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems Volume II: Standard Information for Functionals*, Eur. Math. Society, EMS Tracts in Mathematics, Vol 12, 2010.

[40] M. R. D'Orsogna, Y. L. Chuang, A. L. Bertozzi and L. Chayes, *Self-propelled particles with soft-core interactions: patterns, stability, and collapse*, Phys. Rev. Lett. 96 (2006).

[41] H. Rauhut, *Compressive sensing and structured random matrices*, in: Theoretical Foundations

and Numerical Methods for Sparse Recovery (ed. M. Fornasier), Volume 9 of Radon Series Comp. Appl. Math., deGruyter, pp. 1–92, 2010.

[42] G. Rozza, D. B. P. Huynh and A. T. Patera, *Reduced basis approximation and a posteriori error estimation for afinely parametrized elliptic coercive partial diferential equations, application to transport and continuum mechanics*, Arch. Comput Method E, 15 (2008), pp. 229–275.

[43] S. Sen, *Reduced-basis approximation and a posteriori error estimation for many-parameter heat conduction problems*, Numer. Heat Tr. B-Fund, 54 (2008), pp. 369–389.

[44] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, *Information-based Complexity, Computer Science and Scientific Computing*, Academic Press, Inc., Boston, MA, 1988.

[45] K. Veroy, C. Prudhomme, D. V. Rovas and A. T. Patera, *A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations*, in: Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, 2003.

[46] C. Villani, *Topics in Optimal transportation*, Graduate Studies in Mathematics, 58, American Mathematical Society, Providence, RI, 2003.

[47] M. B. Wakin, *Manifold-based signal recovery and parameter estimation from compressive measurements*, preprint, 2008.