The Projection Method for Dynamical Systems of Interacting aAgents in High-Dimension

Massimo Fornasier, Jan Haškovec, and Jan Vybíral

Johann Radon Institute for Computational and Applied Mathematics (RICAM)
Österreichische Akademie der Wissenschaften
massimo.fornasier@oeaw.ac.at
http://www.ricam.oeaw.ac.at/people/page/fornasier/

Sparse and Low Rank Approximation
Banff 2011

Outline

The START-Project “Sparse Approximation and Optimization in High Dimensions”

Outline

The START-Project "Sparse Approximation and Optimization in High Dimensions"

Numerical simulation of dynamical systems in high-dimension

- Dynamical systems in high-dimension
- Dimensionality reduction via Johnson-Lindenstrauss embeddings
- Restricted Isometry Property and Johnson-Lindenstrauss embeddings
- Projection of the dynamics in lower dimension

Outline

Bridging compression and simulation, beyond signal coding-decoding.

Outline

The application framework

First, some notation:

- $d \in \mathbb{N}$ - dimension (very large!!),
- $N \in \mathbb{N}$ - number of agents, typically $N = d^\alpha$, $\alpha > 0$;
- $x = \{x_1, \ldots, x_N\} \in \mathbb{R}^{N\times d}$, where $x_i \in \mathbb{R}^d$, $i = 1, \ldots, N$,
- $\mathcal{D} : \mathbb{R}^{N\times d} \to \mathbb{R}^{N\times N}$, $\mathcal{D}x = (|x_i - x_j|)_{i,j=1}^N$ is the adjacency matrix of x;
- $f_i : \mathbb{R}^{N\times N} \to \mathbb{R}^d, \quad i = 1, \ldots, N,$;
- $f_{ij} : \mathbb{R}^{N\times N} \to \mathbb{R}, \quad i, j = 1, \ldots, N.$

The application framework

First, some notation:

- $d \in \mathbb{N}$ - dimension (very large!!),
- $N \in \mathbb{N}$ - number of agents, typically $N = d^\alpha$, $\alpha > 0$;
- $x = \{x_1, \dots, x_N\} \in \mathbb{R}^{N \times d}$, where $x_i \in \mathbb{R}^d$, $i = 1, \dots, N$,
- $\mathcal{D} : \mathbb{R}^{N \times d} \to \mathbb{R}^{N \times N}$, $\mathcal{D}x = (|x_i - x_j|)_{i,j=1}^N$ is the adjacency matrix of x;
- $f_i : \mathbb{R}^{N \times N} \to \mathbb{R}^d, \quad i = 1, \dots, N,$;
- $f_{ij} : \mathbb{R}^{N \times N} \to \mathbb{R}, \quad i,j = 1, \dots, N.$

We are interested in the

- numerical simulation

The application framework

First, some notation:

- $d \in \mathbb{N}$ - dimension (very large!!),
- $N \in \mathbb{N}$ - number of agents, typically $N = d^\alpha$, $\alpha > 0$;
- $x = \{x_1, \ldots, x_N\} \in \mathbb{R}^{N \times d}$, where $x_i \in \mathbb{R}^d$, $i = 1, \ldots, N$,
- $\mathcal{D} : \mathbb{R}^{N \times d} \to \mathbb{R}^{N \times N}$, $\mathcal{D}x = (|x_i - x_j|)_{i,j=1}^N$ is the adjacency matrix of x;
- $f_i : \mathbb{R}^{N \times N} \to \mathbb{R}^d, \quad i = 1, \ldots, N,$;
- $f_{ij} : \mathbb{R}^{N \times N} \to \mathbb{R}, \quad i, j = 1, \ldots, N.$

We are interested in the

- numerical simulation
- automatic learning/training

of dynamical systems of the type

$$\dot{x}_i(t) = f_i(\mathcal{D}x(t)) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}x(t)) x_j(t), \quad x(0) = x^0 \in \mathbb{R}^{N \times d},$$

describing the dynamics of multiple complex agents, interacting on the basis of their mutual “social” distance.

The application framework

First, some notation:

- $d \in \mathbb{N}$ - dimension (very large!!),
- $N \in \mathbb{N}$ - number of agents, typically $N = d^\alpha$, $\alpha > 0$;
- $x = \{x_1, \dots, x_N\} \in \mathbb{R}^{N \times d}$, where $x_i \in \mathbb{R}^d$, $i = 1, \dots, N$,
- $\mathcal{D} : \mathbb{R}^{N \times d} \to \mathbb{R}^{N \times N}$, $\mathcal{D}x = (|x_i - x_j|)_{i,j=1}^N$ is the adjacency matrix of x;
- $f_i : \mathbb{R}^{N \times N} \to \mathbb{R}^d, \quad i = 1, \dots, N,$;
- $f_{ij} : \mathbb{R}^{N \times N} \to \mathbb{R}, \quad i, j = 1, \dots, N.$

We are interested in the

- numerical simulation
- automatic learning/training

of dynamical systems of the type

$$\dot{x}_i(t) = f_i(\mathcal{D}x(t)) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}x(t)) x_j(t), \quad x(0) = x^0 \in \mathbb{R}^{N \times d},$$

describing the dynamics of multiple complex agents, interacting on the basis of their mutual “social” distance.

The application framework

First, some notation:

- $d \in \mathbb{N}$ - dimension (very large!!),
- $N \in \mathbb{N}$ - number of agents, typically $N = d^\alpha$, $\alpha > 0$;
- $x = \{x_1, \ldots, x_N\} \in \mathbb{R}^{N \times d}$, where $x_i \in \mathbb{R}^d$, $i = 1, \ldots, N$,
- $\mathcal{D} : \mathbb{R}^{N \times d} \to \mathbb{R}^{N \times N}$, $\mathcal{D}x = (|x_i - x_j|)_{i,j=1}^N$ is the adjacency matrix of x;
- $f_i : \mathbb{R}^{N \times N} \to \mathbb{R}^d, \quad i = 1, \ldots, N,$;
- $f_{ij} : \mathbb{R}^{N \times N} \to \mathbb{R}, \quad i, j = 1, \ldots, N.$

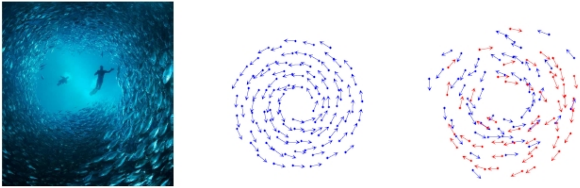
We are interested in the

- numerical simulation
- automatic learning/training

M. Fornasier, K. Schnass, and J. Vybíral, *Learning functions of few arbitrary linear parameters in high dimensions*, preprint, 2010.

K. Schnass, and J. Vybíral, *Compressed learning of high-dimensional sparse functions*, ICASSP, 2011.

An example inspired by nature



Mills in nature and in our simulations.

J. A. Carrillo, M. Fornasier, G. Toscani, and F. Vecil, *Particle, kinetic, hydrodynamic models of swarming*, within the book "Mathematical modeling of collective behavior in socio-economic and life-sciences", Birkhäuser (Eds. Lorenzo Pareschi, Giovanni Naldi, and Giuseppe Toscani), 2010.

The application framework

With the development of communication technology and internet, larger and larger groups of people will access

- information (interactive database access, trends in scientific literature and in newspapers ...)
- services (Google, the financial market ...)
- social interactions (social networks ...)

The application framework

With the development of communication technology and internet, larger and larger groups of people will access

- information (interactive database access, trends in scientific literature and in newspapers ...)
- services (Google, the financial market ...)
- social interactions (social networks ...)

Our aim is to provide innovative tools for analyzing, simulating, even predicting and controlling the behavior of such large crowds, as one today can already do with weather forecast.

The application framework

With the development of communication technology and internet, larger and larger groups of people will access

- information (interactive database access, trends in scientific literature and in newspapers ...)
- services (Google, the financial market ...)
- social interactions (social networks ...)

Our aim is to provide innovative tools for analyzing, simulating, even predicting and controlling the behavior of such large crowds, as one today can already do with weather forecast.

We are facing very difficult challenges due to the "curse of dimensionality", as our individuals are not physical particles and needs a large number d of degrees of freedom to be described.

Some assumptions

We assume the following Lipschitz properties of f_i and f_{ij}, namely

$$
\begin{aligned}
|f_i(a) - f_i(b)| &\le L\|a - b\|_\infty, \\
\max_i \sum_j |f_{ij}(a)| &\le L', \\
\max_i \sum_j |f_{ij}(a) - f_{ij}(b)| &\le L''\|a - b\|_\infty,
\end{aligned}
$$

for every $a, b \in \mathbb{R}^{N\times N}$. Here, $\|a - b\|_\infty := \max_{i,j} |a_{ij} - b_{ij}|$.

A classical result

Theorem (Convergence of the Euler scheme)

Assume $f_{ij} = 0$. *Fix* $x^0 \in \mathbb{R}^{N \times d}$ *and let* $x(t)$ *be the unique solution of the ODE*

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

on the interval $[0, T]$, $T > 0$.

A classical result

Theorem (Convergence of the Euler scheme)

Assume $f_{ij} = 0$. *Fix* $x^0 \in \mathbb{R}^{N \times d}$ *and let* $x(t)$ *be the unique solution of the ODE*

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

on the interval $[0, T]$, $T > 0$.

Fix $h > 0$ *and* $t_n := nh$:

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

for $n = 1, 2, \ldots$.

A classical result

Theorem (Convergence of the Euler scheme)

Assume $f_{ij} = 0$*. Fix* $x^0 \in \mathbb{R}^{N\times d}$ *and let* $x(t)$ *be the unique solution of the ODE*

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

on the interval $[0, T]$*,* $T > 0$*.*

Fix $h > 0$ *and* $t_n := nh$*:*

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

for $n = 1, 2, \ldots$ *.*

Then, we have the estimate for $e_n = |x(t_n) - \tilde{x}_n|$*,*

$$e_n \leq \exp(Lt_n)\left(e_0 + ht_n\frac{|f(\mathcal{D}\tilde{x}^0)|}{2}\right).$$

Exponential complexity reduction in d

The complexity of this algorithm stems from the evaluation of $f(\mathcal{D}x)$ which can be (generically) estimated by

$$\mathcal{O}(d \times N^2).$$

Our first aim is to reduce the dimensionality of the problem to a log-factor in d, and consequently the complexity to

$$\mathcal{O}(\log(d) \times N^2).$$

Outline

Dimensionality reduction via Johnson-Lindenstrauss embeddings

Again some notation

- $\varepsilon > 0$ - a distortion parameter from J-L Lemma, see below,
- $n_0 \in \mathbb{N}$ - number of iterations,
- $\mathcal{N} = n_0 N$ - number of iterations times number of agents
- $k = \mathcal{O}(\varepsilon^{-2} \log(\mathcal{N}))$, new lower dimension - see below,
- $M \in \mathbb{R}^{k \times d}$ - randomly generated matrix, see below,
- $\mathcal{D} : \mathbb{R}^{N \times d} \to \mathbb{R}^{N \times N}$, $\mathcal{D}x = (|x_i - x_j|)_{i,j=1}^N$ is the adjacency matrix in high-dimension and similarly defined $\mathcal{D}' : \mathbb{R}^{N \times k} \to \mathbb{R}^{N \times N}$, $\mathcal{D}' y = (|y_i - y_j|)_{i,j=1}^N$, the one in low-dimension.

Dimensionality reduction via Johnson-Lindenstrauss embeddings

Lemma (Johnson and Lindenstrauss)

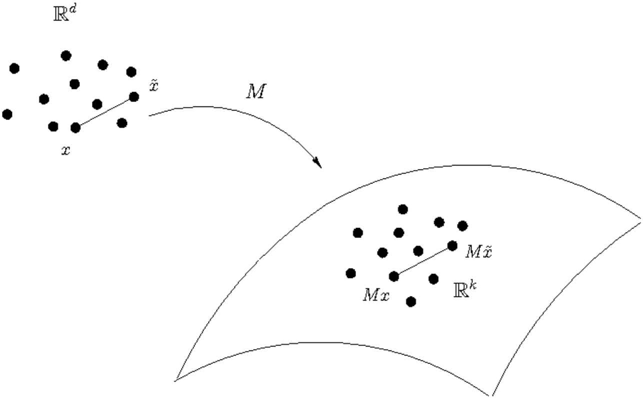
Let $\mathcal{P}$ be an arbitrary set of $\mathcal{N}$ points in $\mathbb{R}^d$. Given $\varepsilon > 0$, there exists

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

such that for all integers $k \geq k_0$, there exists a $k \times d$ random matrix M for which with high probability, for all $x, \tilde{x} \in \mathcal{P}$

$$(1-\varepsilon)|x-\tilde{x}|^2 \leq |Mx - M\tilde{x}|^2 \leq (1+\varepsilon)|x-\tilde{x}|^2.$$

Dimensionality reduction via Johnson-Lindenstrauss embeddings



Outline

Restricted Isometry Property

Definition

A $k \times d$ matrix $\tilde{M}$ is said to have the Restricted Isometry Property of order $K \leq d$ and level $\delta \in (0, 1)$ if

$$(1 - \delta)|x|^2 \leq |\tilde{M}x|^2 \leq (1 + \delta)|x|^2$$

for all K-sparse $x \in \mathbb{R}^d$.

Restricted Isometry Property

Definition

A $k \times d$ matrix $\tilde{M}$ is said to have the Restricted Isometry Property of order $K \leq d$ and level $\delta \in (0, 1)$ if

$$(1 - \delta)|x|^2 \leq |\tilde{M}x|^2 \leq (1 + \delta)|x|^2$$

for all K-sparse $x \in \mathbb{R}^d$.

Theorem (Krahmer, 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 $1 - \eta$,

$$(1 - \varepsilon)|x|^2 \leq |Mx|^2 \leq (1 + \varepsilon)|x|^2.$$

uniformly for all $x \in \mathcal{P}$, where $M := \tilde{M}\,\mathrm{diag}(\xi)$.

Outline

Projection of the dynamical system

We consider the system of ordinary differential equations in the fixed form with the initial condition

$$x_i(0) = x_i^0\,, \qquad i = 1, \ldots, N\,.$$

The Euler method for this system is given by this initial condition and

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

where $h > 0$ is the time step and $n_0 := T/h$ is the number of iterations.

Projection of the dynamical system

We consider the system of ordinary differential equations in the fixed form with the initial condition

$$x_i(0) = x_i^0 \,, \qquad i = 1, \ldots, N \,.$$

The Euler method for this system is given by this initial condition and

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

where $h > 0$ is the time step and $n_0 := T/h$ is the number of iterations.
If $M \in \mathbb{R}^{k \times d}$ is a matrix, we may consider the associated Euler method in $\mathbb{R}^k$, namely

$$y_i^0 := M x_i^0,$$

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

A first surprising result

Theorem (Fornasier, Haskovec, Vybiral)

Given a matrix $M \in \mathbb{R}^{k\times d}$ such that

$$\left|Mf_i(\mathcal{D}'y^n) - Mf_i(\mathcal{D}x^n)\right| \leq (1+\varepsilon)\left|f_i(\mathcal{D}'y^n) - f_i(\mathcal{D}x^n)\right|,$$
$$|Mx_j^n| \leq (1+\varepsilon)|x_j^n|,$$
$$(1-\varepsilon)|x_i^n - x_j^n| \leq |Mx_i^n - Mx_j^n| \leq (1+\varepsilon)|x_i^n - x_j^n|$$

for all $i, j = 1, \dots, N$ and all $n = 0, \dots, n_0$. Let us also assume, that $\alpha \geq \max_j |x_j^n|$ for all $n = 0, \dots, n_0$, $j = 1, \dots, N$. Let

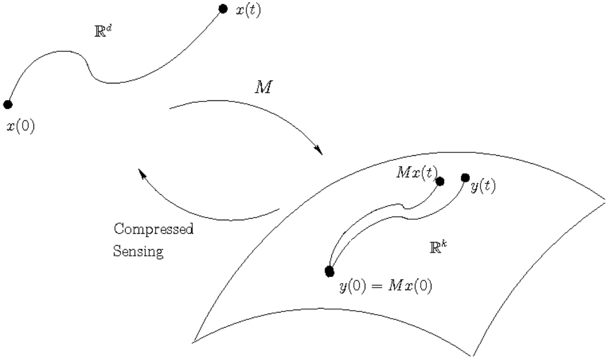
$$e_i^n := |y_i^n - Mx_i^n|, \; i = 1, \dots, N \text{ and } n = 0, \dots, n_0$$

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

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

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

Visual explanation



A continuous Johnson-Lindenstrauss Lemma

Theorem (Fornasier, Haskovec, Vybiral)

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

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

Let k *be such a dimension, that a randomly chosen (and properly normalized) projector* M *satisfies the statement of the Johnson-Lindenstrauss Lemma with* ε, d, k *and* $\mathcal{N}$ *arbitrary points. Then*

$$(1 - \varepsilon')|\varphi(t)| \leq |M\varphi(t)| \leq (1 + \varepsilon')|\varphi(t)|, \qquad t \in [0,1]$$

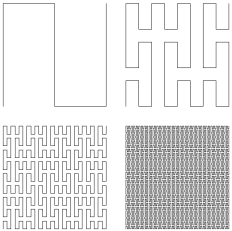
holds with the same probability.

A continuous Johnson-Lindenstrauss Lemma

The condition

$$\gamma := \max_{\xi \in [0,1]} \frac{|\dot{\varphi}(\xi)|}{|\varphi(\xi)|} < \infty \quad \text{and} \quad \mathcal{N} \geq (\sqrt{d} + 1) \cdot \frac{\gamma}{\varepsilon' - \varepsilon}$$

is necessary.



Peano's space-filling curve

By lifting a suitable parametrization a Peano's space-filling curve on the unit sphere S^{d-1}, one generates a curve with infinite speed (i.e., the condition does not hold), and at the same time it generates any possible vector including those in the kernel of M, hence

$$(1 - \varepsilon')|\varphi(t)| \leq |M\varphi(t)|$$

cannot hold!

Projecting the continuous system

Theorem (Fornasier, Haskovec, Vybiral)

Let $x(t) \in \mathbb{R}^{d\times N}$, $t \in [0,T]$, *be the solution of the given ODE system, such that* $\max_{t\in[0,T]} \max_{i,j} |x_i(t) - x_j(t)| \le \alpha$. *Let us fix* $k \in \mathbb{N}$, $k \le d$, *and a matrix* $M \in \mathbb{R}^{k\times d}$ *such that*

$$(1-\varepsilon)|x_i(t) - x_j(t)| \le |Mx_i(t) - Mx_j(t)| \le (1+\varepsilon)|x_i(t) - x_j(t)|\,,$$

for all $t \in [0,T]$ *and* $i,j = 1,\dots,N$. *Let* $y(t) \in \mathbb{R}^{k\times N}$, $t \in [0,T]$ *be the solution of the projected (continuous) system such that for a suitable* $\beta > 0$, $\max_{t\in[0,T]} \max_i |y_i(t)| \le \beta$. *Let us define the columnwise* ℓ^2*-error* $e_i(t) := |y_i(t) - Mx_i(t)|$ *for* $i = 1,\dots,N$ *and*

$$\mathcal{E}(t) := \max_{i=1,\dots,N} e_i(t)\,.$$

Then we have the estimate

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

Verifying the crucial condition

According to our continuous Johnson-Lindenstrauss Lemma

$$(1-\varepsilon)|x_i(t)-x_j(t)| \leq |Mx_i(t)-Mx_j(t)| \leq (1+\varepsilon)|x_i(t)-x_j(t)|\,,$$

for all $t \in [0, T]$ and i, $j = 1, \ldots, N$, is verified if the necessary condition

$$\sup_{t\in[0,T]} \max_{i,j} \frac{|\dot{x}_i(t)-\dot{x}_j(t)|}{|x_i(t)-x_j(t)|} \leq \gamma < \infty\,,$$

holds.

Verifying the crucial condition

According to our continuous Johnson-Lindenstrauss Lemma

$$(1-\varepsilon)|x_i(t)-x_j(t)| \leq |Mx_i(t)-Mx_j(t)| \leq (1+\varepsilon)|x_i(t)-x_j(t)|\,,$$

for all $t \in [0, T]$ and $i, j = 1, \ldots, N$, is verified if the necessary condition

$$\sup_{t\in[0,T]} \max_{i,j} \frac{|\dot{x}_i(t)-\dot{x}_j(t)|}{|x_i(t)-x_j(t)|} \leq \gamma < \infty\,,$$

holds. It is, for instance, trivially satisfied when the right hand sides f_i, f_{ij} have the following Lipschitz continuity:

$$\begin{aligned} |f_i(\mathcal{D}x)-f_j(\mathcal{D}x)| &\leq L'''|x_i-x_j| && \text{for all } i,j=1,\ldots,N\,, \\ |f_{i\ell}(\mathcal{D}x)-f_{j\ell}(\mathcal{D}x)| &\leq L''''|x_i-x_j| && \text{for all } i,j,\ell=1,\ldots,N\,. \end{aligned}$$

Verifying the crucial condition

According to our continuous Johnson-Lindenstrauss Lemma

$$(1-\varepsilon)|x_i(t)-x_j(t)| \le |Mx_i(t)-Mx_j(t)| \le (1+\varepsilon)|x_i(t)-x_j(t)|\,,$$

for all $t \in [0, T]$ and $i, j = 1, \ldots, N$, is verified if the necessary condition

$$\sup_{t\in[0,T]} \max_{i,j} \frac{|\dot{x}_i(t)-\dot{x}_j(t)|}{|x_i(t)-x_j(t)|} \le \gamma < \infty\,,$$

holds. It is, for instance, trivially satisfied when the right hand sides f_i, f_{ij} have the following Lipschitz continuity:

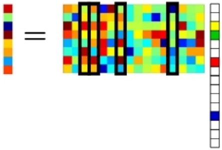
$$\begin{aligned} |f_i(\mathcal{D}x) - f_j(\mathcal{D}x)| &\le L'''|x_i - x_j| \qquad \text{for all } i,j = 1,\ldots,N\,, \\ |f_{i\ell}(\mathcal{D}x) - f_{j\ell}(\mathcal{D}x)| &\le L''''|x_i - x_j| \qquad \text{for all } i,j,\ell = 1,\ldots,N\,. \end{aligned}$$

We will show examples below for which the condition is verified.

Compressed sensing enters the picture

Theorem

Given a matrix $M \in \mathbb{R}^{k \times d}$ *with the RIP of order* $2K$ *and level* $\delta < 0.4$, *and*

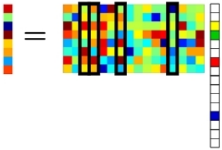


$$y = Mx + \eta \in \mathbb{R}^k, \quad |\eta| \leq \varepsilon$$

Compressed sensing enters the picture

Theorem

Given a matrix $M \in \mathbb{R}^{k\times d}$ *with the RIP of order* $2K$ *and level* $\delta < 0.4$, *and*



$$y = Mx + \eta \in \mathbb{R}^k, \quad |\eta| \le \varepsilon$$

The vector $\hat{x}$ *computed by* $\hat{x} = \arg\min_{|Mz-y|\le\varepsilon} |z|_1 := \sum_{i=1}^d |z_i|$, *has the approximation property*

$$|\hat{x} - x| \le C_1 \frac{\sigma_K(x)_1}{\sqrt{K}} + C_2 \varepsilon,$$

where $\sigma_K(z)_1 = |z - z_{[K]}|_1$, *best-*$K$*-term approx. error.*

A second surprising algorithmic result

As a consequence of this theorem, by projecting and simulating in parallel the dynamical system d_k-times, $d_k \leq \frac{d}{k}$ in lower dimension

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

A second surprising algorithmic result

As a consequence of this theorem, by projecting and simulating in parallel the dynamical system d_k-times, $d_k \leq \frac{d}{k}$ in lower dimension

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

we can assemble the following system

$$\begin{pmatrix} M^1 \\ M^2 \\ \ldots \\ \ldots \\ M^{d_k} \end{pmatrix} x_i = \begin{pmatrix} y_i^1 \\ y_i^2 \\ \ldots \\ \ldots \\ y_i^{d_k} \end{pmatrix} - \begin{pmatrix} \eta_i^1 \\ \eta_i^2 \\ \ldots \\ \ldots \\ \eta_i^{d_k} \end{pmatrix}$$

A second surprising algorithmic result

As a consequence of this theorem, by projecting and simulating in parallel the dynamical system d_k-times, $d_k \leq \frac{d}{k}$ in lower dimension

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

we can assemble the following system

$$\begin{pmatrix} M^1 \\ M^2 \\ \dots \\ \dots \\ M^{d_k} \end{pmatrix} x_i = \begin{pmatrix} y_i^1 \\ y_i^2 \\ \dots \\ \dots \\ y_i^{d_k} \end{pmatrix} - \begin{pmatrix} \eta_i^1 \\ \eta_i^2 \\ \dots \\ \dots \\ \eta_i^{d_k} \end{pmatrix}$$

Therefore we can compute $\hat{x}_i$ such that

$$|\hat{x}_i - x_i| \leq C_1 \frac{\sigma_{d_k K}(x_i)_1}{\sqrt{d_k K}} + C_2 \varepsilon.$$

A second surprising algorithmic result

As a consequence of this theorem, by projecting and simulating in parallel the dynamical system d_k-times, $d_k \leq \frac{d}{k}$ in lower dimension

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

we can assemble the following system

$$\begin{pmatrix} M^1 \\ M^2 \\ \dots \\ \dots \\ M^{d_k} \end{pmatrix} x_i = \begin{pmatrix} y_i^1 \\ y_i^2 \\ \dots \\ \dots \\ y_i^{d_k} \end{pmatrix} - \begin{pmatrix} \eta_i^1 \\ \eta_i^2 \\ \dots \\ \dots \\ \eta_i^{d_k} \end{pmatrix}$$

Therefore we can compute $\hat{x}_i$ such that

$$|\hat{x}_i - x_i| \leq C_1 \frac{\sigma_{d_k K}(x_i)_1}{\sqrt{d_k K}} + C_2 \varepsilon.$$

The computation of $\hat{x}_i$ can be parallelized!

M. Fornasier, *Domain decomposition methods for linear inverse problems with sparsity constraints*, Inverse Problems, Vol. 23, 2007, pp. 2505-2526.

Interesting examples

Our theory apply, for instance, on

Interesting examples

Our theory apply, for instance, on

- the Cucker-Smale model, which is given by

$$\dot{x}_i = v_i \in \mathbb{R}^d,$$
$$\dot{v}_i = \frac{1}{N}\sum_{j=1}^{N} g(|x_i - x_j|)(v_j - v_i).$$

The function $g : \mathbb{R} \to \mathbb{R}$ is given by $g(t) = \frac{G}{(1+t^2)^\beta}$, $t > 0$ and bounded by $g(0) = G > 0$.

Interesting examples

Our theory apply, for instance, on

- the Cucker-Smale model, which is given by

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

$$\dot{v}_i = \frac{1}{N}\sum_{j=1}^{N} g(|x_i - x_j|)(v_j - v_i).$$

  The function $g : \mathbb{R} \to \mathbb{R}$ is given by $g(t) = \frac{G}{(1+t^2)^\beta}$, $t > 0$ and bounded by $g(0) = G > 0$.

- the D'Orsogna-Chuang-Bertozzi-Chayes model, which is given by

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

$$\dot{v}_i = (a - b|v_i|^2)v_i - \frac{1}{N}\sum_{j \neq i} \nabla U(|x_i - x_j|),$$

  where a and b are positive constants and $U : \mathbb{R} \to \mathbb{R}$ is a smooth potential.

Interesting examples

In principle, we can also consider

Interesting examples

In principle, we can also consider

- the Keller-Segel model, given by

$$dx_i(t) = -c \sum_{j \neq i} \frac{x_i - x_j}{|x_i - x_j|^d} dt + \sqrt{2} dB_i ,$$

  where $B_i(t)$, $i = 1, \ldots, N$ are mutually independent d-dimensional Brownian motions and c is a positive constant.

Interesting examples

In principle, we can also consider

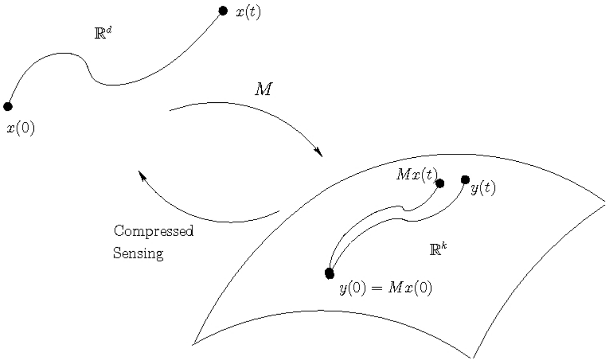
- the Keller-Segel model, given by

$$dx_i(t) = -c \sum_{j \neq i} \frac{x_i - x_j}{|x_i - x_j|^d} dt + \sqrt{2} dB_i \,,$$

where $B_i(t)$, $i = 1, \ldots, N$ are mutually independent d-dimensional Brownian motions and c is a positive constant.

In this case, though, the matrix M should be better a partial orthogonal random matrix (for instance a random partial Fourier matrix), as $MB_i(t)$, $i = 1, \ldots, N$ are mutually independent k-dimensional Brownian motions!

Summarizing



Outline

Passage to kinetic equations $N \to \infty$

We specify

$$\dot{x}_i(t) = f_i(\mathcal{D}x(t)) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}x(t))x_j(t), \quad x(0) = x^0 \in \mathbb{R}^{N \times d},$$

for the Cucker-Smale model.

Passage to kinetic equations $N \to \infty$

We specify

$$\dot{x}_i(t) = f_i(\mathcal{D}x(t)) + \sum_{j=1}^{N} f_{ij}(\mathcal{D}x(t))x_j(t), \quad x(0) = x^0 \in \mathbb{R}^{N \times d},$$

for the Cucker-Smale model.

The Cucker–Smale flocking model:

$$\begin{cases} \dot{x}_i = v_i \in \mathbb{R}^d, \\ \dot{v}_i = \dfrac{1}{N} \displaystyle\sum_{j=1}^{N} g(|x_i - x_j|)(v_j - v_i), \end{cases}$$

for $i = 1, \ldots, N$, where f is the *communication rate*.

S.-Y. Ha and E. Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinetic and Related Models 1(3) (2008) 415-435.

J. A. Carrillo, M. Fornasier, J. Rosado, and G. Toscani, *Asymptotic flocking dynamics for the kinetic Cucker-Smale model*, SIAM. J. Math. Anal., 42(1) (2010) 218-236.

Passage to kinetic equations

The Cucker–Smale model projected in low-dimension:

$$\begin{cases} M\dot{x}_i = Mv_i \in \mathbb{R}^k, \\ \\ M\dot{v}_i \approx \frac{1}{N}\sum_{j=1}^{N} g(|Mx_i - Mx_j|)(Mv_j - Mv_i) \ , \end{cases}$$

for $i = 1, \dots, N$.
Substituting $y_i = Mx_i \in \mathbb{R}^k$ and $w_i = Mv_i \in \mathbb{R}^k$, we define

$$\begin{cases} \dot{y}_i = w_i, \\ \\ \dot{w}_i = \frac{1}{N}\sum_{j=1}^{N} g(|y_i - y_j|)(w_j - w_i) \ , \end{cases}$$

Mean-field limit

Define the empirical distribution density associated to a solution $(x(t), v(t))$ of the Cucker-Smale model

$$\mu^n(x, v, t) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i(t))\delta(v - v_i(t)) \ ,$$

Mean-field limit

Define the empirical distribution density associated to a solution $(x(t), v(t))$ of the Cucker-Smale model

$$\mu^n(x, v, t) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i(t))\delta(v - v_i(t)) \ ,$$

This probability measure formally satisfies the following equation in weak sense

$$\frac{\partial \mu^n}{\partial t} + v \cdot \nabla_x \mu^n = \nabla_v \cdot [\xi(\mu^n)\mu^n]$$

$$\downarrow n \to \infty$$

$$\frac{\partial \mu}{\partial t} + v \cdot \nabla_x \mu = \nabla_v \cdot [\xi(\mu)\mu] \ ,$$

where $\xi(\mu)(x, v, t) = [(g(x)\nabla_v U(v)) * \mu]$ and $U(v) = \frac{1}{2}|v|^2$ and $*$ is the (x, v)-convolution.

Outline

Stability, approximation properties, and optimal integration

Such approximation is evaluated in terms of dual Wasserstein-type distances:

$$W(\mu,\nu) = \sup\left\{ \int \varphi d(\mu-\nu) : \varphi \in \mathrm{Lip}, \|\varphi\|_{\mathrm{Lip}} \leq 1 \right\}.$$

In particular

$$W(\mu,\mu^n) = \sup_{\varphi\in\mathrm{Lip}(\mathbb{R}^{d\times d}),\|\varphi\|_{\mathrm{Lip}}\leq 1} \left\{ \int \varphi d\mu - \frac{1}{n}\sum_{i=1}^{n}\varphi(x_i,v_i) \right\}.$$

Stability, approximation properties, and optimal integration

Such approximation is evaluated in terms of dual Wasserstein-type distances:

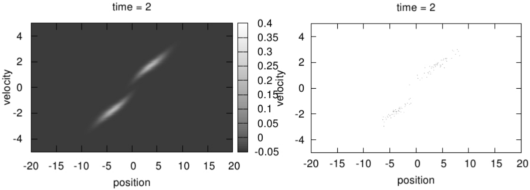
$$W(\mu,\nu) = \sup\left\{ \int \varphi d(\mu-\nu) : \varphi \in \mathsf{Lip}, \|\varphi\|_{\mathsf{Lip}} \leq 1 \right\}.$$

In particular

$$W(\mu,\mu^n) = \sup_{\varphi\in \mathsf{Lip}(\mathbb{R}^{d\times d}), \|\varphi\|_{\mathsf{Lip}}\leq 1} \left\{ \int \varphi d\mu - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i,v_i) \right\}.$$

Stability (e.g, Ha-Liu|Carrillo-Canizo-Rosado):

$$W(\mu(t),\mu^n(t)) \leq C(T) W(\mu(0),\mu^n(0)),$$



Stability, approximation properties, and optimal integration

Let us consider:

$$W(\mu, \mu^n) = \sup_{\varphi \in \mathrm{Lip}(\mathbb{R}^{d \times d}), \|\varphi\|_{\mathrm{Lip}} \leq 1} \left\{ \int \varphi d\mu - \frac{1}{n} \sum_i \varphi(x_i, v_i) \right\}.$$

Which are the optimal and universal integration points (x_i, v_i), $i = 1, \ldots, n$ such that

$$W(\mu, \mu^n) = \mathcal{O}(n^{-\gamma}),$$

for the largest possible $\gamma > 0$?

Stability, approximation properties, and optimal integration

Let us consider:

$$W(\mu, \mu^n) = \sup_{\varphi \in \mathrm{Lip}(\mathbb{R}^{d\times d}), \|\varphi\|_{\mathrm{Lip}} \leq 1} \left\{ \int \varphi d\mu - \frac{1}{n} \sum_i \varphi(x_i, v_i) \right\}.$$

Which are the optimal and universal integration points (x_i, v_i), $i = 1, \ldots, n$ such that

$$W(\mu, \mu^n) = \mathcal{O}(n^{-\gamma}),$$

for the largest possible $\gamma > 0$? This question belongs to the realm of Information Based Complexity.

Stability, approximation properties, and optimal integration

Let us assume $d = 1$ and that μ is a nice function supported on $[x_0, x_n]$ and $\sqrt{\sigma} := |x_n - x_0|$. We define x_i the quatiles such that

$$\int_{-\infty}^{x_i} \mu(x) dx = \frac{i}{n}.$$

Then, for $\varphi \in \mathrm{Lip}(\mathbb{R})$,

$$\left| \int_{\mathbb{R}} \varphi(x)\mu(x) dx - \frac{1}{n} \sum_{i=1}^{n} \varphi(x_i) \right| \leq \frac{\sqrt{\sigma}}{n} = \mathcal{O}(n^{-1})$$

Stability, approximation properties, and optimal integration

Let us assume $d = 1$ and that μ is a nice function supported on $[x_0, x_n]$ and $\sqrt{\sigma} := |x_n - x_0|$. We define x_i the quatiles such that

$$\int_{-\infty}^{x_i} \mu(x)dx = \frac{i}{n}.$$

Then, for $\varphi \in \mathrm{Lip}(\mathbb{R})$,

$$\left| \int_{\mathbb{R}} \varphi(x)\mu(x)dx - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i) \right| \leq \frac{\sqrt{\sigma}}{n} = \mathcal{O}(n^{-1}) \Rightarrow \gamma = 1.$$

Stability, approximation properties, and optimal integration

Let us assume $d = 1$ and that μ is a nice function supported on $[x_0, x_n]$ and $\sqrt{\sigma} := |x_n - x_0|$. We define x_i the quatiles such that

$$\int_{-\infty}^{x_i} \mu(x)dx = \frac{i}{n}.$$

Then, for $\varphi \in \text{Lip}(\mathbb{R})$,

$$\left| \int_{\mathbb{R}} \varphi(x)\mu(x)dx - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i) \right| \leq \frac{\sqrt{\sigma}}{n} = \mathcal{O}(n^{-1}) \Rightarrow \gamma = 1.$$

Assume now $d \gg 1$ and $\mu = \mu_1 \otimes \cdots \otimes \mu_d$, μ_i univariate compactly supported with corresponding support size $\sqrt{\sigma}_i$, then

$$\left| \int_{\mathbb{R}^d} \varphi(x)\mu(x)dx - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i) \right| \leq C_d \sum_{i=1}^{d} \frac{\sqrt{\sigma}_i}{n_i},$$

and $n := \prod_{i=1}^{d} n_i$ is the number of optimal sampling points.

Stability, approximation properties, and optimal integration

Hence, if $n_i = N$, for all $i = 1, \ldots, d$, then

$$n = N^d,$$

but

$$\left| \int_{\mathbb{R}^d} \varphi(x)\mu(x)dx - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i) \right| \leq C_d \frac{\sum_{i=1}^{d} \sqrt{\sigma}_i}{N} = \mathcal{O}(n^{-1/d}).$$

Stability, approximation properties, and optimal integration

Hence, if $n_i = N$, for all $i = 1, \ldots, d$, then

$$n = N^d,$$

but

$$\left| \int_{\mathbb{R}^d} \varphi(x)\mu(x)dx - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i) \right| \leq C_d \frac{\sum_{i=1}^{d} \sqrt{\sigma}_i}{N} = \mathcal{O}(n^{-1/d}).$$

The number of points depends EXPONENTIALLY, $\gamma = 1/d$, on the dimension d!! $\Rightarrow$ curse of dimensionality.

Stability, approximation properties, and optimal integration

Hence, if $n_i = N$, for all $i = 1, \dots, d$, then

$$n = N^d,$$

but

$$\left| \int_{\mathbb{R}^d} \varphi(x)\mu(x)dx - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i) \right| \leq C_d \frac{\sum_{i=1}^{d} \sqrt{\sigma}_i}{N} = \mathcal{O}(n^{-1/d}).$$

The number of points depends EXPONENTIALLY, $\gamma = 1/d$, on the dimension d!! $\Rightarrow$ curse of dimensionality.

One way to improve this approximation is by assuming $\sum_{i\in\mathcal{X}} \sqrt{\sigma}_i \leq \varepsilon$, for a suitable set $\mathcal{X} \subset \{1, \dots, d\}$, such that $\#\mathcal{X} \leq d - k, \quad k = k(\varepsilon)$:

$$n_i = N, i \in \mathcal{X}^c \text{ and } n_i = 1, \quad i \in \mathcal{X} \Rightarrow n = N^k,$$

$$k = k(\varepsilon) \to d, \quad \varepsilon \to 0.$$

Outline

Projecting the PDE in low-dimension

The natural projection in low-dimension of a measure is given again by duality: $\nu_M := M\#\mu \Rightarrow$ (push-forward):

$$\langle \nu_M, \varphi \rangle := \langle \mu, \varphi(M\cdot) \rangle, \quad \forall \varphi \in \mathrm{Lip}(\mathbb{R}^{k\times k}).$$

If $\mu \in L_1$ then, we have a *generalized Radon transform*:

$$\nu_M(y, w, t) = C_M \int_{\ker M} \mu(M^\dagger(y, w) + \xi, t) d\xi.$$

Projecting the PDE in low-dimension

The natural projection in low-dimension of a measure is given again by duality: $\nu_M := M\#\mu \Rightarrow$ (push-forward):

$$\langle \nu_M, \varphi \rangle := \langle \mu, \varphi(M\cdot) \rangle, \quad \forall \varphi \in \mathrm{Lip}(\mathbb{R}^{k\times k}).$$

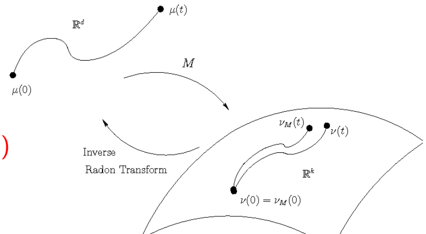
If $\mu \in L_1$ then, we have a *generalized Radon transform*:

$$\nu_M(y,w,t) = C_M \int_{\ker M} \mu(M^\dagger(y,w) + \xi, t) d\xi.$$

The equation in low-dimension with projected datum:

$$\frac{\partial \nu}{\partial t} + w \cdot \nabla_y \nu = \nabla_w \cdot [\xi(\nu)\nu] \ , \quad \nu(y,w,0) = \nu_M(y,w,0)$$

$$\begin{array}{ccc} \mu(0) & \xrightarrow{t} & \mu(t) \\ \downarrow M & & \downarrow M \\ \nu(0) = \nu_M(0) & \xrightarrow{t} & \nu(t) \approx \nu_M(t) \end{array}$$



Outline

Delayed curse of dimensionality

The way to prove it:

$$\begin{array}{ccc} \mu & \xrightarrow{W(\mu,\mu^n)\lesssim n^{-\gamma}=\varepsilon} & \mu^n \\ \downarrow M & & \downarrow M \\ \nu & \xrightarrow{W(\nu,\nu^n)\lesssim n^{-\gamma}=\varepsilon} & \nu^n \end{array}$$

This good approximation property can be ensured only if

$$W(\mu,\mu^n) \lesssim n^{-\gamma} = \varepsilon, \text{ for } \gamma = 1/k \gg 1/d,$$

and this is possible if μ "concentrates" on manifolds of dimension k!.

Delayed curse of dimensionality

The way to prove it:

$$\begin{array}{ccc} \mu & \overset{W(\mu,\mu^n)\lesssim n^{-\gamma}=\varepsilon}{\longrightarrow} & \mu^n \\ \downarrow M & & \downarrow M \\ \nu & \overset{W(\nu,\nu^n)\lesssim n^{-\gamma}=\varepsilon}{\longrightarrow} & \nu^n \end{array}$$

This good approximation property can be ensured only if

$$W(\mu,\mu^n) \lesssim n^{-\gamma} = \varepsilon, \text{ for } \gamma = 1/k \gg 1/d,$$

and this is possible if μ "concentrates" on manifolds of dimension k!.

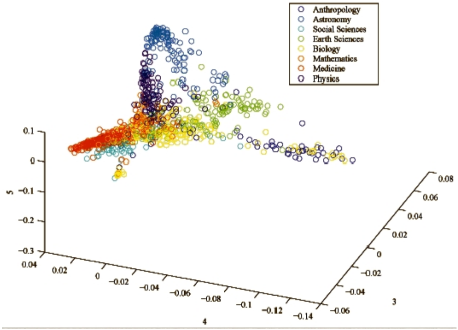
Typical non compactly supported example:

$$\mu(x,v) = \prod_{i=1}^{d} (2\pi\sigma_i)^{-1/2} \exp\left(x_i^2/(2\sigma_i)\right) \times \prod_{j=1}^{d} (2\pi\tilde{\sigma}_j)^{-1/2} \exp\left(v_i^2/(2\tilde{\sigma}_j)\right),$$

for $\sum_{i\in\mathcal{X}} \sqrt{\sigma}_i + \sum_{i\in\mathcal{V}} \sqrt{\tilde{\sigma}_j} \le \varepsilon$, for suitable sets $\mathcal{X},\mathcal{V} \subset \{1,\ldots,d\}$,
$\#\mathcal{X} + \#\mathcal{V} \le d-k, \quad k = k(\varepsilon)$.

Is this concentration property a realistic assumption?

About 1100 Science News articles, from 8 different categories. We compute about 1000 coordinates, i-th coordinate of document d represents frequency in document d of the i-th word in a dictionary.



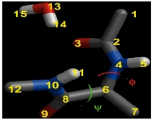
Clustering in high-dimension.

Is this concentration property a realistic assumption?

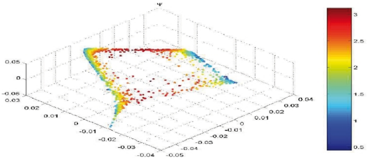
The dynamics of a small protein in a bath of water molecules is approximated by a Langevin system of stochastic equations

$$\dot{x} = \nabla U(x) + \dot{w}.$$

The set of states of the protein is a noisy set of points in $\mathbb{R}^{36}$.



Protein in a water bath.



. Projection in $\mathbb{R}^3$ of the states in high-dimension.

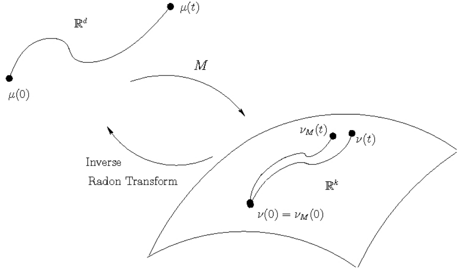
Conclusion

- The projection of a nonlinear PDE associated to a dynamical system governed by the adjacency matrix can furnish good approximations in lower dimension only if the initial value measure is concentrated;

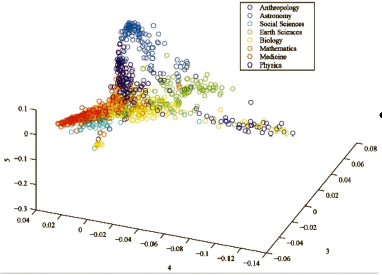
Conclusion

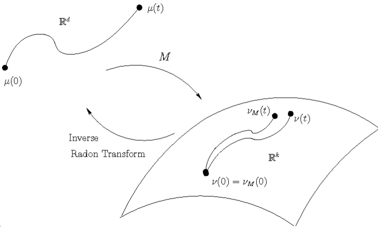
- The projection of a nonlinear PDE associated to a dynamical system governed by the adjacency matrix can furnish good approximations in lower dimension only if the initial value measure is concentrated;
- The recovery of the higher dimensional measure from lower dimensional simulations is obtained by the inversion of the following generalized Radon-type transform:

$$\nu(y, w, t) \approx C_M \int_{\ker M} \mu(M^\dagger(y, w, t) + \xi) d\xi.$$



Visual summary





Bridging compression and simulation, beyond signal coding-decoding.