Collective dynamics in the social and data sciences

Marie-Therese Wolfram

SFB 1294 Colloquium

Table of contents

Collective dynamics in the social sciences

Collective dynamics

- Refers to the alignment of characteristic features in large interacting particle/agent systems.
- Leads to the formation of complex states, such as clusters, aggregates, ...
- Often observed in animal flocks as collective motion is more efficient and gives better protection against predators.
- Social sciences: observed in opinion formation or pedestrian dynamics.

Mathematical modelling

Microscopic interactions:

- *Individual interactions with others and the surrounding:* alignment of velocities, averaging of opinions,....
- *Common objective or goal:* reach a target - for example a food source, an exit, or the minimum of a function - as fast as possible.
- *Explore and wander around* - not necessarily at a constant rate in time.

Mathematical modelling

Microscopic interactions:

- *Individual interactions with others and the surrounding:* alignment of velocities, averaging of opinions,....

$\Rightarrow$ interaction potential U

- *Common objective or goal:* reach a target - for example a food source, an exit, or the minimum of a function - as fast as possible.

$\Rightarrow$ given potential V

- *Explore and wander around* - not necessarily at a constant rate in time.

$\Rightarrow$ noise

Coarse graining

microscopic	macroscopic/mean-field
particle position $\mathbf{X}_t^i$	particle density $\rho(x,t)$ wrt to the position x

Microscopic models

Consider N particles characterised by their position $\mathbf{X}_t^i(t) \in \mathbb{R}^d$, $\mathbf{X}_t = (\mathbf{X}_t^1, \dots, \mathbf{X}_t^N)$,

First order model: aka over-damped Langevin equations

$$\mathrm{d}\mathbf{X}_t^i = \sqrt{2}\sigma \mathrm{d}\mathbf{W}_t^i - \nabla_{\mathbf{X}_t^i} U(\mathbf{X}_t)\mathrm{d}t + \nabla V(\mathbf{X}_t^i)\mathrm{d}t,$$

where $\mathbf{W}_t^i$ is an independent Wiener process, V a given potential and particles interact through the pairwise interactions

$$U(\mathbf{X}_t) = \chi \sum_{i \le j < j \le N} u\left(\mathbf{X}_t^i, \mathbf{X}_t^j\right),$$

where χ and ℓ represent the strength and the range in space.

Examples:

- Bounded confidence models in opinion formation: Scaling $\chi = \frac{1}{N}$, interaction radius $R > 0$

$$u(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{1}_{|\mathbf{X}_t^i - \mathbf{X}_t^j| \le R}(\mathbf{X}_t^j - \mathbf{X}_t^i)$$

- Attraction-repulsion models

The Germknödel dispute

(a) The only possible way

(b) Just wrong

Let $\mathbf{X}_t^i \in [0,1]$ denote the opinion of the i-th agent at time t:

$$\mathrm{d}\mathbf{X}_t^i = \frac{1}{N}\sum_j \mathbf{1}_{|\mathbf{X}_t^i - \mathbf{X}_t^j| \leq R}(\mathbf{X}_t^j - \mathbf{X}_t^i)\mathrm{d}t + \sqrt{2}\sigma \mathrm{d}\mathbf{W}_t^i x$$

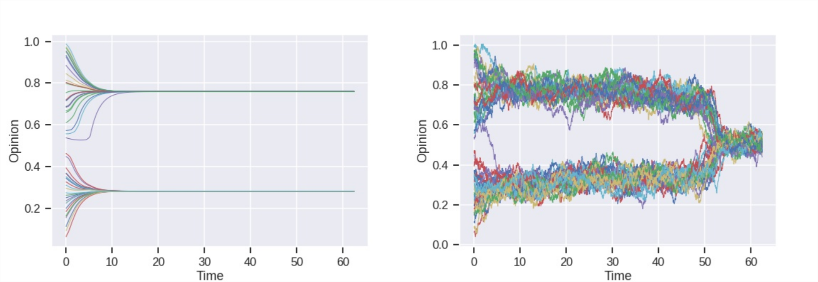


Figure 1: Evolution of 50 agents in time, with interaction radius $R = 0.25$ and $\sigma^2 = 0.04$

Opinion formation

Observations:

- Formation of clusters; their number depends on the interaction radius R.
- Averaging within each cluster.
- Noise changes the game completely.

Mean-field limit: the empirical distribution

$$\rho(x,t) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\mathbf{X}_t^j(t)}(x)$$

satisfies the nonlinear conservative equation

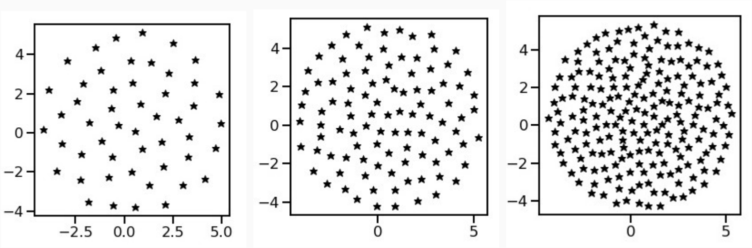
$$\partial_t \rho + \nabla \cdot (\rho(\nabla U * \rho)) = 0,$$

with $U(r) = \int_0^r s u(s)\,\mathrm{d}s$.

Short range repulsion and long range attraction

Consider N particles with position $\mathbf{X}_t^i$ in $\mathbb{R}^2$ and an interaction function corresponds to the Morse potential

$$U(r) = C_R e^{-\frac{r}{\ell_R}} - C_A e^{\frac{r}{\ell_A}}.$$



(a) 50 particles (b) 100 particles (c) 200 particles

Literature: Carrillo, Bertozzi, ...

Nonlinear conservation laws for the density ρ

The corresponding density $\rho = \rho(x, t)$ of many of these microscopic models satisfies a nonlinear conservation law:

$$\partial_t \rho = \nabla \cdot \left[M(\rho) \nabla \left(E' + V + U' * \rho \right) \right],$$

where E is an energy, V a potential, U an interaction energy and $M(\rho)$ a nonlinear mobility.

Examples:

- Heat equation: $M(\rho) = \rho$, $E = \int \rho \log \rho \, \mathrm{d}x$, $V \equiv 0$, $U \equiv 0$
- Porous medium equation $M(\rho) = \rho$, $E = \int \rho^m \mathrm{d}x$, $V \equiv 0$, $U \equiv 0$
- Fokker-Planck equation: $M(\rho) = \rho$, $E = \int \rho \log \rho \, \mathrm{d}x$, V given, $U \equiv 0$
  For example, if $V(\mathbf{x}) = x_1$, then $\nabla V = \mathbf{e}_1$; everybody is moving to the left.
- Aggregation equations: $M(\rho) = \rho$, $U(r) = \frac{r^a}{a} - \frac{r^b}{b}$ with $a < b$, $V \equiv 0$, $E \equiv 0$.

Nonlinear conservation laws for the density ρ

The corresponding density $\rho = \rho(x, t)$ of many of these microscopic models satisfies a nonlinear conservation law:

$$\partial_t \rho = \nabla \cdot \left[M(\rho) \nabla \left(E' + V + U' * \rho \right) \right],$$

where E is an energy, V a potential, U an interaction energy and $M(\rho)$ a nonlinear mobility.

Examples:

- Heat equation: $M(\rho) = \rho$, $E = \int \rho \log \rho \, \mathrm{d}x$, $V \equiv 0$, $U \equiv 0$
- Porous medium equation $M(\rho) = \rho$, $E = \int \rho^m \mathrm{d}x$, $V \equiv 0$, $U \equiv 0$
- Fokker-Planck equation: $M(\rho) = \rho$, $E = \int \rho \log \rho \, \mathrm{d}x$, V given, $U \equiv 0$
  For example, if $V(\mathbf{x}) = x_1$, then $\nabla V = \mathbf{e}_1$; everybody is moving to the left.
- Aggregation equations: $M(\rho) = \rho$, $U(r) = \frac{r^a}{a} - \frac{r^b}{b}$ with $a < b$, $V \equiv 0$, $E \equiv 0$.

What can we say something about the large-time behaviour of these equations?

Sliding down energy functionals

L^2-Wasserstein gradient flow: conservation law can be written as

$$\partial_t \rho(\mathbf{x}, t) = \nabla \cdot \left[M(\rho) \nabla \left(\frac{\delta \mathcal{E}(\rho)}{\delta \rho} \right) \right],$$

where $M(\rho)$ is a positive definite operator and the entropy is given by

$$\mathcal{E}(\rho) = \int \left(E(\rho) + \rho V(\rho) + \rho \left(U * \rho \right) \right) \mathrm{d}x.$$

L^2-Wasserstein space: Let (M, d) be a complete separable metric space, define

$$\mathcal{P}_2(M) = \{\mu \in \mathcal{P}(M) \colon \int d(x, x_0)^2 d\mu(x) < \infty\}$$

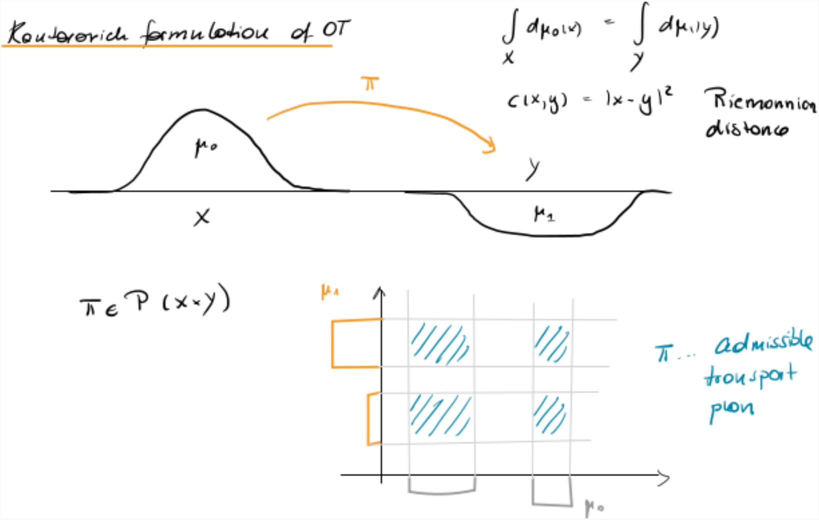
and define the so-called Wasserstein distance

$$d_W(\mu_0, \mu_1) = \inf_{\pi \in \Pi} \left[\int_{M \times M} d^2(x, y) d\pi(x, y) \right]^{\frac{1}{2}}$$

where Π denotes the set of admissible transportation plans satisfying

$$(P_X)_{\#}\pi = \mu_0 \text{ and } (P_Y)_{\#}\pi = \mu_1.$$

Connection to optimal transport



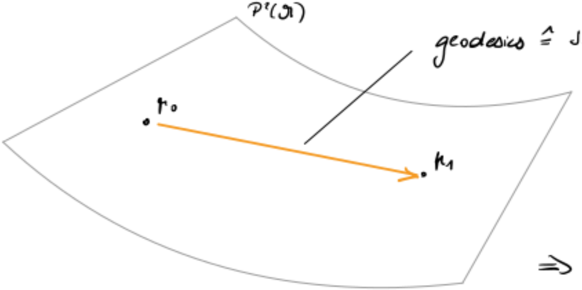
Dynamic formulation of OT

Benamou-Brenier:

action / kinetic energy

$$d_{W_2}(\mu_0, \mu_1) = \inf \int_0^1 \rho \|v\|_{L^2}^2 \, dt$$

such that $\partial_t \rho + \nabla \cdot (\rho v) = 0$ $\quad \rho(x,0) = \mu_0$, $\rho(x,1) = \mu_1$



geodesics $\hat{=}$ shortest path

$\Rightarrow$ notion of distance depends on chosen metric

$\Rightarrow$ generalised Wasserstein flows

Trend to equilibrium

The linear Fokker-Planck equation

$$\begin{aligned}\partial_t \rho &= \nabla \cdot (\nabla \rho + \rho \nabla V) \\ &= \nabla \cdot \left(\rho \nabla \left(\frac{\delta \mathcal{E}}{\delta \rho} \right) \right),\end{aligned}$$

for the entropy $\mathcal{E} = \int [\rho \log \rho + \rho V] \, \mathrm{d}x$.

Equilibrium solution ρ_∞ is a minimiser of the entropy $\mathcal{E}$

$$\log \rho_\infty + V = \mathrm{c} \Rightarrow \rho_\infty \propto e^{-V}.$$

Relative entropy or Kullback-Leibler divergence:

$$\mathcal{E}(\rho | \rho_\infty) = \mathcal{E}(\rho) - \mathcal{E}(\rho_\infty).$$

Evolution of the entropy

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{E} &= \int \partial_t \rho \, [\log \rho + V] + \rho \nabla \, [\log \rho + V] \, \mathrm{d}x \\ &= - \int \rho \nabla |\log \rho + V|^2 \mathrm{d}x \\ &:= -\mathcal{I}(\rho) \qquad \Leftarrow \text{Entropy production or Fisher-information}\end{aligned}$$

Trend to equilibrium

Entropy method or Bakry-Emery strategy:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{I} = -\lambda\mathcal{I} - \mathcal{R}(t),$$

with $\lambda > 0$ and $\mathcal{R}(t) \geq 0$ then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E} = \frac{1}{\lambda}\left(\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} + \mathcal{R}(t)\right)$$

for some $\lambda \geq 0$.

Since $\mathcal{R}$ is positive

$$\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} \leq \lambda\mathcal{I} \Rightarrow \mathcal{I}(\rho) \leq e^{-\lambda t}\mathcal{I}(\rho_0)$$

and therefore

$$\mathcal{E}(\rho) - \mathcal{E}(\rho_\infty) \leq \frac{1}{\lambda}\mathcal{I}(\rho_0).$$

Using that $\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t} = -\mathcal{I}$ we obtain

$$\frac{\mathrm{d}\left(\mathcal{E}(\rho(t)) - \mathcal{E}(\rho_\infty)\right)}{\mathrm{d}t} \leq \frac{1}{\lambda}\left(\mathcal{E}(\rho(t)) - \mathcal{E}(\rho_\infty)\right).$$

Trend to equilibrium

Entropy method or Bakry-Emery strategy:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{I} = -\lambda\mathcal{I} - \mathcal{R}(t),$$

with $\lambda > 0$ and $\mathcal{R}(t) \geq 0$ then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E} = \frac{1}{\lambda}\left(\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} + \mathcal{R}(t)\right)$$

for some $\lambda \geq 0$.

Since $\mathcal{R}$ is positive

$$\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} \leq \lambda\mathcal{I} \Rightarrow \mathcal{I}(\rho) \leq e^{-\lambda t}\mathcal{I}(\rho_0)$$

and therefore

$$\mathcal{E}(\rho) - \mathcal{E}(\rho_\infty) \leq \frac{1}{\lambda}\mathcal{I}(\rho_0).$$

Using that $\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t} = -\mathcal{I}$ we obtain

$$\frac{\mathrm{d}\left(\mathcal{E}(\rho(t)) - \mathcal{E}(\rho_\infty)\right)}{\mathrm{d}t} \leq \frac{1}{\lambda}\left(\mathcal{E}(\rho(t)) - \mathcal{E}(\rho_\infty)\right).$$

$\Rightarrow$ Exponential convergence in entropy towards equilibrium.

Geodesic convexity

Geodesic or displacement convexity: let (X, d) be a geodesic space and let $\lambda \in \mathbb{R}^+$. A functional $\mathcal{E} : X \to \mathbb{R} \cup \{\infty\}$ is called λ-geodesically convex, if for every two points $x_0, x_1 \in X$ there exists a constant speed geodesic γ connecting $\gamma(0) = x_0$ to $\gamma(1) = x_1$ and

$$\mathcal{E}(\gamma(t)) \leq (1-t)E(x_0) + tE(x_1) - \lambda \frac{t(1-t)}{2} d^2(x_0, x_1) \quad \forall t \in [0, 1].$$

- The λ-geodesic convexity of a functional guarantees the existence of the gradient flow in Wasserstein space, and determines the rate of convergence to equilibrium.
- Convergence in entropy also implies convergence in the Wasserstein distance (Talgrand inequality).
- Choosing different metrics (adapted to the problem considered) can improve convergence to equilibrium.

Collective dynamics in the data sciences

Collective dynamics: social vs. data sciences

Social sciences

- Complex individual interactions
- Very large number of individuals
- Finite size effects
- Second order microscopic models

Data sciences

- Mostly first order schemes
- Interactions less complex
- Goal is to sample from a distribution or find its minimum
- High dimensional setting
- What's the best metric?

Bayesian inverse problems

The inverse problem: Given observations $y \in \mathbb{R}^K$ infer $x \in \mathbb{R}^d$ based on noisy evaluations of $G(x)$:

$$y = G(x) + \xi.$$

- Assumption: noise $\xi \sim \mathcal{N}(0, \Gamma)$, with strictly positive-definite covariance $\Gamma \in \mathbb{R}^{K \times K}$.

Bayes rule: Imposing a Gaussian prior $x \sim \mathcal{N}(m, \Sigma)$, the posterior distribution is given by

$$\pi(x) \propto e^{-V(x)},$$
$$V(x) := \frac{1}{2}|y - G(x)|_\Gamma^2 + \frac{1}{2}|x - m|_\Sigma^2.$$

Goal: Sample or maximise negative log-likelihood.

Particle methods

Let's consider N particles with position $\mathbf{X}_t^i$.

Gradient based sampler: particles evolve according to over-damped Langevin equations

$$\mathrm{d}\mathbf{X}_t^i = -C(\mathbf{X}_t)\nabla V(\mathbf{X}_t^i)\mathrm{d}t + \sqrt{2C(\mathbf{X}_t)}\mathrm{d}\mathbf{W}_t^i,$$

where $C(\mathbf{X}_t)$ is a positive definite matrix.
Examples: Ensemble Langevin Sampler (ELS), Stochastic Stein Variational Gradient Descent (SVGD),....

Gradient free sampling: approximate high-dimensional and possibly very rough gradient in the Langevin dynamics

$$\mathrm{d}\mathbf{X}_t^i = -C\nabla \tilde{V}(\mathbf{X}_t)\mathrm{d}t + \sqrt{2C(\mathbf{X}_t)}\mathrm{d}\mathbf{W}_t^i$$

Examples: Ensemble Kalman Sampler (EKS), consensus based optimisation (CBO), ...

Ensemble Langevin Sampler (ELS)

ELS dynamics:

$$\mathrm{d}\mathbf{X}_t^i = -C(\mathbf{X}_t)\nabla V(\mathbf{X}_t^i)\,dt + \nabla_{\mathbf{x}^i}\cdot C(\mathbf{X}_t)\,dt + \sqrt{2C(\mathbf{X}_t)}\mathrm{d}\mathbf{W}_t^i.$$

Here $C : \mathbb{R}^{Nd} \to \mathbb{R}^{d\times d}$ denotes the empirical covariance function.

The mean field limit of the ELS is given by

$$dx_t = -\mathcal{C}(\rho)\nabla V(x_t) + \sqrt{2\mathcal{C}(\rho)}dW_t,$$

where function $\mathcal{C}(\cdot)$ is the covariance defined for densities.

Then the time-dependent density ρ of the process satisfies

$$\begin{aligned}\partial_t\rho &= \nabla\cdot\left(\mathcal{C}(\rho)\left(\nabla_x\rho + \nabla_x V\rho\right)\right).\\ &= \nabla\cdot\left(\rho\,\mathcal{C}(\rho)\nabla(\log\rho + V)\right)\end{aligned}$$

with $C(\rho) = \int(x - m(\rho))\otimes(x - m(\rho))\rho(x)\,\mathrm{d}x$ and $m(\rho) = \int x\rho(x)\,\mathrm{d}x$.

Ensemble Kalman sampler (EKS)

EKS comprises N coupled SDEs in $\mathbb{R}^d$, for X_t^i given by

$$\mathrm{d}\mathbf{X}_t^i = -\left(\frac{1}{N}\sum_{n=1}^{N}\langle G(\mathbf{X}_t^n) - \overline{G}_t, G(\mathbf{X}_t^i) - y\rangle_\Gamma \mathbf{X}_t^n\right)\mathrm{d}t - C_t\Sigma^{-1}(\mathbf{X}_t^i - m)\,\mathrm{d}t$$
$$+ \frac{d+1}{N}(\mathbf{X}_t^i - \overline{\mathbf{X}}_t)\,\mathrm{d}t + \sqrt{2C_t}\,\mathrm{d}\mathbf{W}_t^i;$$

here the W^i are standard independent Brownian motions in $\mathbb{R}^d$ and

$$\overline{\mathbf{X}}_t = \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}_t^n, \qquad \overline{G}_t = \frac{1}{N}\sum_{n=1}^{N}G(\mathbf{X}_t^n),$$
$$C_t = \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{X}_t^n - \overline{\mathbf{X}}_t\right)\otimes\left(X_t^n - \overline{\mathbf{X}}_t\right).$$

The mean field equation for the density of the process ρ is

$$\partial_t\rho = \nabla\cdot\left(\rho\,\mathcal{C}(\rho)\nabla\left(\log\rho + V\right)\right).$$

Covariance modulated OT

Consider covariance modulated FPE:

$$\partial_t \rho = \nabla \cdot \left(\rho\, \mathcal{C}(\rho) \nabla \left(\log \rho + V \right) \right).$$

- Carrillo and Vaes showed the following stability estimate for two solutions ρ^1 and ρ^2 in case of a linear mapping G:

  $$d_{W^2}(\rho^1, \rho^2) \leq C\gamma(t) d_{W^2}(\rho_0^1, \rho_0^2),$$

  where function $\gamma(t)$ converges to zero exponentially fast as $t \to \infty$.
- Ongoing work by M. Burger, F. Hoffmann, D. Matthes and A. Schlichting $\Rightarrow$ ask them for more information!

Consensus based optimisation (CBO)

Idea: Move towards the particle with the smallest function value.

Particle dynamics:

$$\mathrm{d}\mathbf{X}_t^i = -\lambda \left(\mathbf{X}_t^i - \bar{\mathbf{X}}_V\right) H^\epsilon(f(\mathbf{X}^i) - f(\bar{\mathbf{X}}_V))\mathrm{d}t + \sigma|\mathbf{X}^i - \bar{\mathbf{X}}_V|\mathrm{d}\mathbf{W}^i.$$

where $\bar{\mathbf{X}}_V$ is the weighted average (wrt to the function V) of the particles

$$\bar{\mathbf{X}}_V = \frac{\sum_j \mathbf{X}_t^j e^{-\beta V(\mathbf{X}_t^j)}}{\sum_j e^{-\beta V(\mathbf{X}_t^j)}}$$

with $\lambda > 0$, β sufficiently large and H^ϵ being a regularisation of the Heaviside function.

Mean field equation for the density of the process:

$$\partial_t \rho = \Delta\left(\kappa[\rho]\rho\right) + \nabla \cdot \left(\mu[\rho]\rho\right)$$

with $\kappa[\rho] = \sigma^2\left(x - \bar{x}_V[\rho]\right)^2$, $\mu[\rho] = -\lambda(x - \bar{x}_V[\rho])$ and the weighted average

$$\bar{x}_V[\rho] = \frac{1}{\int_{\mathbb{R}^d} e^{-\beta V}\rho\,\mathrm{d}y}\int_{\mathbb{R}^d} y e^{-\beta V}\rho\,\mathrm{d}y.$$

Ashley function

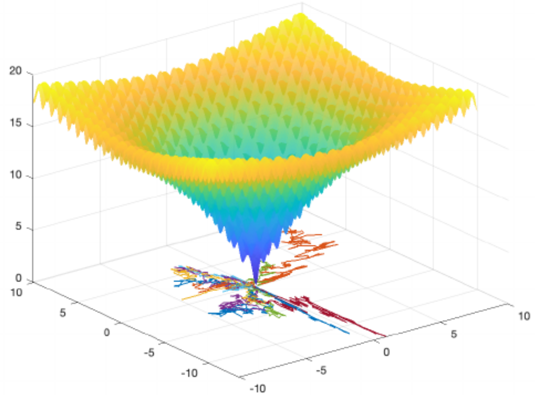


FIGURE 1. Plot of the Ackley [4] benchmark function for global optimization in two dimensions with trajectories of one realisation of (6) with 20 particles visualized in the xy-plane.

Image from C. Totzeck, *Trends in consensus based optimization*, Active Crowds 3, Springer, 2021.

CBO variants

Component-wise diffusion:

Consider component-wise geometric Brownian motion

$$\mathrm{d}\mathbf{X}_t^i = -\lambda \left(\mathbf{X}_t^i - \bar{\mathbf{X}}_V\right) H^\epsilon(f(\mathbf{X}^i) - f(\bar{\mathbf{X}}_V))\mathrm{d}t + \sigma \sum_{k=1}^{d} \left(\mathbf{X}^i - \bar{\mathbf{X}}_V\right)_k \mathrm{d}\mathbf{W}_k^i \mathbf{e}_k.$$

where $\mathbf{e}_k$ denotes the k-th unit vector in $\mathbb{R}^d$.
Advantage: If $2\lambda > \sigma^2$, then particles concentrate. In the original version convergence depended on d.

Random batches:

- Select q random subsets $J^\theta \subset \{1 \ldots N\}$ of size $|J^\theta| = M \ll N$ and compute the empirical expectation $V(\mathbf{X}^\theta)$ for each.
- Calculate the weighted mean using the empirical expectations and update the particle positions.

Stein Variational Gradient Descent

Idea: Evaluate the gradient of the function V at positions $\mathbf{X}_t^i$ for $i = 1, \dots N$ and calculate the respective smoothed gradient of the function V using Gaussian process regression.

SVGD dynamics:

$$\mathrm{d}\mathbf{X}_t^i = \frac{1}{N}\sum_{j=1}^{N}\left[-k(\mathbf{X}_t^i, \mathbf{X}_t^j)\nabla V(\mathbf{X}_t^j) + \nabla_{\mathbf{X}_t^j} k(\mathbf{X}_t^i, \mathbf{X}_t^j)\right]\mathrm{d}t + \sum_{j=1}^{N}\sqrt{2\mathbf{K}(\mathbf{X}^t)}\mathrm{d}\mathbf{W}_t^i.$$

where $\mathbb{K} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd\times Nd}$ is given by

$$\mathbf{K}(x) = \begin{pmatrix} K_{11}(x) & \dots & \dots K_{1N}(x) \\ \vdots & & \\ K_{N1}(x) & \dots & K_{NN}(x) \end{pmatrix}$$

with $K_{ij}(x) = \frac{1}{N}k(x_i, x_j)\mathbf{1}_{d\times d}$ is a sufficiently 'nice' kernel, such as the Gaussian kernel

$$k(x, y) = \lambda e^{-\frac{-|x-y|^2}{2\ell^2}}.$$

Stein-Wasserstein gradient flows

As $N \to \infty$ the empirical measure converges to the solution of the following Fokker-Planck type equation

$$\partial_t \rho = \nabla \cdot \left(\rho \int_{\mathbb{R}^d} k(x,y) \left[\nabla \rho(y,t) + \rho(y,t) \nabla V(y) \right] dy \right)$$

Stein-Wasserstein gradient flow:

$$\begin{aligned} \partial_t \rho(x,t) &= \nabla \cdot \left(\rho T_{k,\rho} \left(\nabla \log \rho + \nabla V() \right) \right) \\ &= \nabla \cdot \left(\rho T_{k,\rho} \nabla \frac{\delta \mathcal{E}}{\delta \rho} \right) \end{aligned}$$

where $T_{k,\rho}\phi = \int_{\mathbb{R}^d} k(\cdot,y)\phi(y)\mathrm{d}\rho(y)$ for all $\phi \in L^2(\rho)$.

For more information on Stein-Wassertstein gradient flows $\Rightarrow$ talk to A. Duncan, N. Nüsken and L. Szpruch.

Ensemble Gaussian process sampler

Idea: Calculate a smooth approximation of the potential V, using the particle positions and then calculate the gradient.

Approximation of the potential gradient

$$\nabla \tilde{V}_L(x;\sigma,\lambda,l) = \sum_{i,j=1}^{N} \nabla_x k(x,\mathbf{X}_t^i,\lambda,l) K(\mathbf{X};\sigma,\lambda,l)_{ij}^{-1} V_L(\mathbf{X}_t^j).$$

based on the assumption that the data misfit term/potential

$$V_L(x) = \frac{1}{2}\langle y - G(x), \Gamma^{-1}(y - G(x))\rangle \text{ is a Gaussian process.}$$

Ensemble particles evolve according to *over-damped Langevin dynamics*

$$d\mathbf{X}_t^i = -\nabla \tilde{V}_L(\mathbf{X}_t^i;\sigma,\lambda,l)\,dt - \Sigma^{-1}\mathbf{X}_t^i\,dt + \sqrt{2}\sigma\,\mathrm{d}\mathbf{W}_t.$$

- *Approximate gradient $\nabla \tilde{V}_L$ depends on the hyper-parameters* (σ,λ,l), which have to be updated/trained as the density evolves.
- Similar spirit as the work of Maoutsa et al (2020), who estimate the gradient of the logarithm of the particle density using a statistical estimator.

Ensemble Gaussian Process sampler

Mean-field equation for the density of the process:

$$\partial_t \rho = \sigma^2 \Delta \rho + \nabla \cdot \left(\rho \nabla \int k(x,y) \mu_V(y) \rho(y) \, \mathrm{d}y \right)$$

where μ_V solves

$$\int \left(k(x,y) + \sigma^2 \delta_{xy} \right) \mu_V(y) \, \mathrm{d}y = V(x).$$

Ensemble Gaussian Process sampler

Mean-field equation for the density of the process:

$$\partial_t \rho = \sigma^2 \Delta \rho + \nabla \cdot \left(\rho \nabla \int k(x,y) \mu_V(y) \rho(y) \, \mathrm{d}y \right)$$

where μ_V solves

$$\int \left(k(x,y) + \sigma^2 \delta_{xy} \right) \mu_V(y) \, \mathrm{d}y = V(x).$$

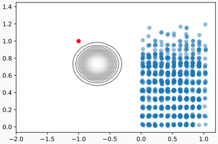
We have no idea what to do with this!

Linear forward model

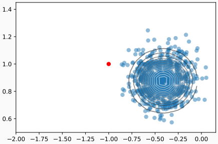
Forward map G_ϵ of the form, for $x = (x_1, x_2)$,

$$G_\epsilon(x) = G_0(x) + G_1(x/\epsilon),$$

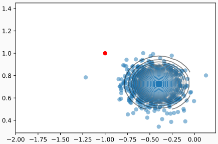
$$G_0(x) = Ax, \quad G_1(x) = [\sin(2\pi x_1), \sin(2\pi x_2)]^\top, \quad \text{with } A = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.$$



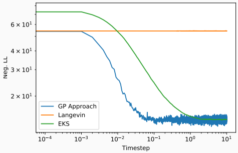
(d) Ensemble Langevin Samples



(e) Ensemble Kalman Sampler



(f) Ensemble GP sampler



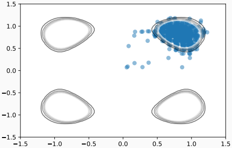
(g) Negative Log-Likelihood

Multi-modal posteriors

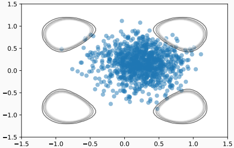
We consider a forward map for $x = (x_1, x_2)$ which is defined by

$$G_\epsilon(x) = G_0(x) + G_1(x/\epsilon),$$
$$G_0(x) = (x_1^2 - 1)^2 + (x_2^2 - 1)^2, \quad G_1(x) = \nu(\sin(2\pi x_1) + \sin(2\pi x_2)),$$

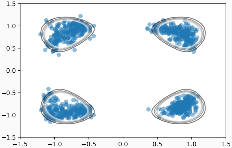
and where $\Gamma = \gamma^2 I$.



(h) ELS



(i) EKS



(j) EGPS

Take home message

- Generalised Wasserstein gradient flows are a powerful analytical tool to analyse and construct interacting particle methods.
- Consensus based optimisation is fun, but doesn't really fit into the picture yet.
- Lots of recent developments, ... and lots of open questions!

Take home message

- Generalised Wasserstein gradient flows are a powerful analytical tool to analyse and construct interacting particle methods.
- Consensus based optimisation is fun, but doesn't really fit into the picture yet.
- Lots of recent developments, ... and lots of open questions!
- Trust the Austrians and have home-made Germknödel with melted butter and Graumohn!

Thank you very much for your attention!

References

- S. Motsch and E. Tadmor, *Heterophilius dynamics enhances convergence*, SIAM Review, 2014
- J.A. Carrillo, S. Jin, L. Lei and Y. Zhu, *A consensus-based global optimisation method for high dimensional machine learning problems*, ESAIM Control, Optimisation and Calculus of Variations, 2021
- R. Pinnau, C. Totzeck, O. Tse and S. Martin, *A consensus-based model for global optimisation and its mean-field limit*, M3AS, 2017
- C. Totzeck, *Trends in consensus based optimization*, Active Crowds 3, Springer, 2021.
- J.A. Carrillo and U. Vaes, *Wasserstein stability estimates for covariance Fokker-Planck equations*, Nonlinearity, 2020
- A. Garbuno-Inigo, F. Hoffmann, W. Li and A.M. Stuart, *Interacting Langevin diffusions: Gradient Structure and ensemble Kalman sampler*, SIADS 19, 2020
- A. Duncan, N. Nüsken, L. Szpruch, *On the geometry of Stein variational gradient descent*, JLMR, 2021
- O.R.A. Dunbar, A.B. Duncan, A.M. Stuart and MTW, *Ensemble Inference Methods for Models with Noisy and Expensive Likelihoods*, to appear in SIADS, 2022
- D. Maoutsa, S. Reich and M. Opper, *Interacting Particle Solutions of Fokker-Planck Equations Through Gradient-Log-Density Estimation*, Entropy 2020