Probability and Mathematical Statistics

PhD days in Mathematics

Dario Trevisan

http://people.dm.unipi.it/trevisan/

January 22, 2025

General description

Randomness is essential for *modeling* real-world challenges
and a powerful ally in big data *analysis*.

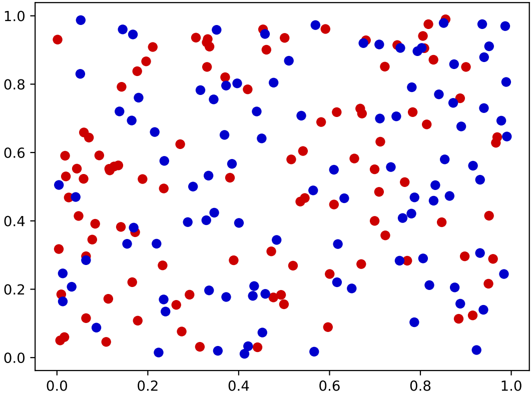
Two general research directions:

- **Probability** $\Rightarrow$ stochastic evolution or variational problems
- **Statistics** $\Rightarrow$ mathematics for statistical learning

Our group

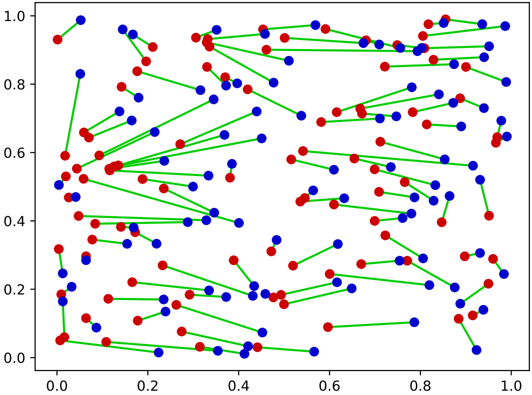
- Francesco Grotto
- Mario Maurelli
- Katerina Papagiannuoli
- Marco Romito
- Dario Trevisan

Random matching problem



Matching Problem: $$\min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|$$

Random matching problem



$$\text{Matching Problem:} \quad \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} \left| X_i - Y_{\sigma(i)} \right|$$

Problem (Talagrand)

Show that for i.i.d. points on $[0,1]^2$ uniformly distributed the limit *exists*:

$$\lim_{n\to\infty} \mathbb{E}\left[\min_{\sigma\in\mathcal{S}_n} \frac{1}{n}\sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|\right] / \sqrt{\log n/n}$$

- Predictions by statistical physics of disordered systems (Caracciolo-Lucibello-Sicuro-Parisi, 2014).

From matching to p-Laplacian (F. Vitillaro)

Theorem (≈ Ambrosio-*Vitillaro*-T., 2024)

For any $p > 1$ for i.i.d. points on $[0,1]^2$ uniformly distributed the limit

$$\lim_{n\to\infty} \mathbb{E}\left[\min_{\sigma\in\mathcal{S}_n} \frac{1}{n}\sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p\right] / (\log n/n)^{p/2}$$

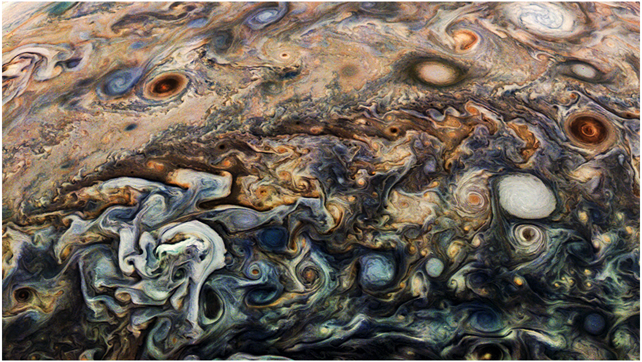
exists if and only if there exists

$$\lim_{n\to\infty} \mathbb{E}\left[\int_{[0,1]^2} |\nabla u|^q\right] / (\log n/n)^{p/2}$$

where $q = p/(p-1)$ and u solves the random q-Poisson equation

$$\mathrm{div}(|\nabla u|^{q-2}\nabla u) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} - \delta_{Y_i}$$

Stochastic Fluid Dynamics



Statistical theory of turbulence (K41): *random* forces or initial conditions yield predictions in turbulent fluids.

Lo scienziato Martin Hairer: “Ora vi spiego la matematica del caffellatte”

Piotr Cieslinski

Professore all’Imperial College di Londra, già medaglia Fields, lo studioso austriaco applica la teoria ai problemi pratici. Qui ci illustra come

18 GENNAIO 2025 ALLE 18:52

Problem (Clay Millennium Prize)

Prove or disprove *well-posedness* for 3D Navier-Stokes equation (incompressible non-viscous fluid):

$$\partial_t v + (v \cdot \nabla) v + \nabla p = \Delta v$$
$$\operatorname{div} v = 0$$

What about *probability*?

- stochastic terms improve numerical simulations
- ill-posed PDE's may become well-posed under addition of noise.

Noise prevents blow-up (M. Bagnara)

Theorem (*Bagnara*-Maurelli-Xu 2023)

Superlinear noise can suppress blow-up for 3D Euler equation:

$$\partial_t u + (u \cdot \nabla) u + \nabla p = \|u\|_{H^6}^{3/2+\epsilon} \dot{W}$$

is globally well-posed.

- By analogy: ODE with superlinear drift can blow up

$$\dot{x} = x^2, \quad \Rightarrow x(t) = \frac{1}{1-t}$$

- noise can suppress blow-up:

$$\dot{x} = x^2 + |x|^{3/2+\epsilon} \dot{W} \quad \Rightarrow \text{no blow-up!}$$

Dissipation for rough shear flows (L. Roveri)

Advection of a passive scalar f by a velocity field u (with $\operatorname{div} u = 0$)

$$\begin{cases} \partial_t f + u\partial_x f = \nu \partial_y^2 f, \\ f|_{t=0} = f_0, \end{cases}$$

Standard dissipation with rate ν due to the effect of viscosity:

$$\|f(t)\|_{L^2}^2 \leq \mathrm{e}^{-t\nu} \|f_0\|_{L^2}^2$$

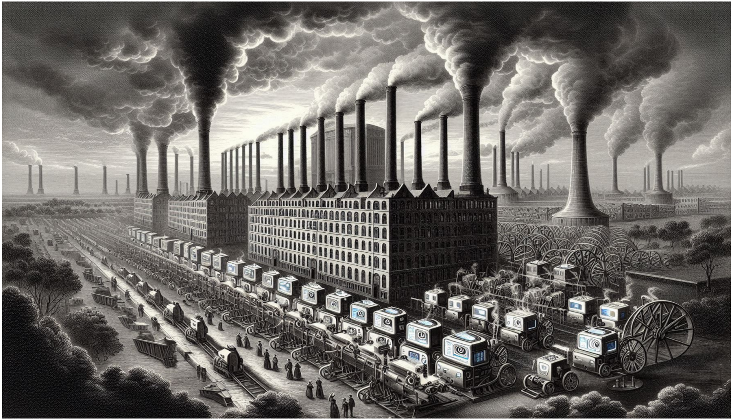
Theorem (Romito-*Roveri* 2024)

Generic Besov fields $u \in B^a_{1,\infty}$ $(-0.5 < a < 0)$ enhance mixing:

$$\|f(t)\|_{L^2}^2 \leq \mathrm{e}^{-t\nu^{a/(a+2)}} \|f_0\|_{L^2}^2.$$

- Improves upon previous results by Colombo, Galeati et al.

Statistical Learning



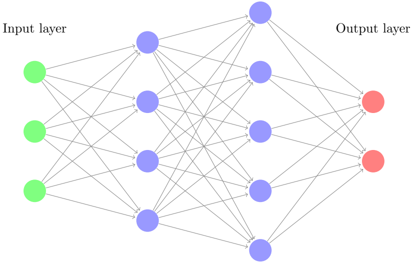
Bing Image Creator (DALL-E 3): *A XIX style engraving of a factory where instead of textile there are supercomputers running AI. Background shows polluting chimneys.*

Deep learning often compared to the industrial revolution:
a triumph of *engineering* $\Rightarrow$ new *science* (thermodynamics).

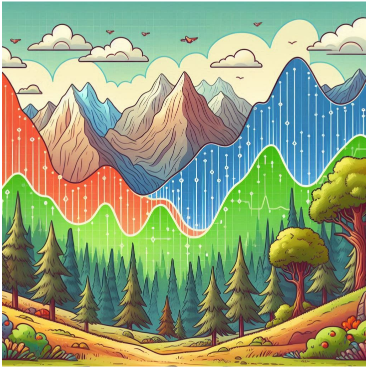
- Math research across fields: probability and statistics, statistical physics, numerical analysis, optimization...

Problem

Build "the" mathematics of AI.



Above: DNN (MLP). Right: Bing Image Creator (DALL-E 3): *A cartoonist representation of gradient descent algorithm in a landscape with montains and valleys.*



Neural Networks as Gaussian Processes (E. Mosig)

- Analysis of *wide limit* of DNNs randomly initialized and trained.
- At first order, DNNs are equivalent to a *Kernel* reflecting the architecture (NNGP) and training (NTK).
- The probabilistic analogue of kernels are *Gaussian Processes*.

Theorem (Agazzi-*Mosig*-T. 2024+)

Deep Neural Networks (MLP) with random initialized parameters and trained via *gradient descent* are *quantitatively close* in the wide limit to Gaussian processes:

$$d(\mathcal{N}\mathcal{N}(t), \mathcal{G}\mathcal{P}(t)) \lesssim \frac{1}{\sqrt{\text{width}}}.$$

Effective equations for ADAM (F. Triggiano)

Stochastic GDs in actual training of DNNs: adaptive learning rate

$$\begin{cases} \theta_{k+1} = \theta_k - \eta \frac{\nabla_\gamma R(\theta_k)}{(\sqrt{v_k}+\epsilon)}, \\ v_{k+1} = \beta v_k + (1-\beta)\left(\nabla_\gamma R(\theta_k)\right)^2 . \end{cases}$$

Theorem (Romito-*Triggiano*, 2024+)

Continuous stochastic equation in the small learning rate limit:

$$\begin{cases} \dot\theta = -\frac{\nabla R(\theta)}{\sqrt{v}+\epsilon} + \sqrt{\eta}\frac{\Sigma(\theta)^{1/2}}{\sqrt{v}+\epsilon}\dot W \\ \dot v = (-v + \nabla R(\theta)^2 + \Sigma(\theta)) - 2\sqrt{\eta}(\mathbf{M}^{\frac{1}{2}}(\theta) W + \dots)\dot W \end{cases}$$

- additional stochastic terms w.r.t. vanilla and momentum SGD (works by Li et al., E et al, Gess et al.)
- S. Saviozzi (ML for finance)