UNIVERSITÀ DEGLI STUDI DI PADOVA

$H(P\,|\,R) < +\infty$

Giovanni Conforti

02/07/2025

Based on works by C.Léonard

Setting the scene

The canonical space and process

- $\Omega = C([0,T];\mathbb{R}^d)$
- $X_t(\omega) = \omega_t, \quad \omega \in \Omega, t \le T$

The entropy

$$H(P\,|\,R) = \begin{cases} \int \log\left(\frac{dP}{dR}\right) dP & P \ll R \\ +\infty & \text{otherwise} \end{cases},$$

The (only) hypothesis (H)

Let

- R be a reference measure (maybe a nice diffusion)
- P a path measure be such that

$$H(P\,|\,R) < +\infty$$

Setting the scene

Martingale problem

Let $\mathscr{C}$ be a class of functions with values in $[0,T] \times \mathbb{R}^d$ and let

$$\mathscr{L} : \mathscr{C} \longrightarrow \text{Adapted stochastic processes}$$

We say that P is a solution of the martingale problem $MP(\mathscr{C}, \mathscr{L}; \mu_0)$ if

- $P_0 = \mu_0$
- For all $f \in \mathscr{C}$ the process

$$f(t, X_t) - f(0,X_0) - \int_0^t (\mathscr{L}f)(s, X_s)ds$$

is a local P-martingale.

Setting the scene

Q-local extended generator of P

As before but…

- v is defined Q-a.s.
- $$u(t, X_t) - u(0, X_0) - \int_0^t v(s, X_s)ds$$

is a Q-local P-martingale. That is, the localizing sequence of stopping times $(\tau_k)_{k\geq 1}$ satisfies

$$\lim_k \tau_k = +\infty \quad Q\text{-a.s.}$$

Equivalence with the notion of stochastic derivative (Nelson'67)

Girsanov's theory

$$\boxed{(\mathscr{L}^R f)_t = \partial_t f_t(X_t) + \frac{1}{2}\mathrm{Tr}(\nabla_x^2 f(X_t)A(X_t)) + b(X_t)\cdot\nabla_x f(X_t)}$$

The form of the density and expression of relative entropy

Let (H) hold and R be the unique solution of $MP(C_c^\infty, \mathscr{L}^R; R_0)$. Then

$$\frac{dP}{dR} = \mathbf{1}_{\frac{dP}{dR}>0}\frac{dP_0}{dR_0}(X_0)\ \exp\Big(\int_0^T u_t\cdot(dX_t - b(X_t)dt) - \frac{1}{2}\int_0^T u_t A(X_t)u_t\,dt\Big)$$

Moreover we have

$$H(P\,|\,R) = H(P_0\,|\,R_0) + \mathbb{E}\Big[\int_0^T u_t A(dt)u_t\Big]$$

is a $P-$Brownian motion.

Feynman-Kac formula

$$\boxed{(\mathscr{L}^R f)_t = \partial_t f_t(X_t) + \frac{1}{2}\mathrm{Tr}(\nabla^2 f(X_t)A(X_t)) + b(X_t)\cdot\nabla f(X_t)}$$

Theorem

Let P be defined by

$$\frac{dP}{dR} = f(X_0)\,\exp\Big(-\int_0^T V(X_s)ds\Big)\,g(X_T)$$

and assume $H(P\,|\,R) < +\infty$. Then

$$g_t(x) = \mathbb{E}_R\Big[\exp\Big(\int_t^T V(s,X_s)ds\Big)g_T(X_T)\Big|X_t = x\Big]$$

is in the domain of the P-local extension of R and satisfies

$$[\mathscr{L}^{R,P} + V]g(t,X_t) = 0 \quad dt\otimes P\text{-a.s.}$$

Feynman-Kac formula

Extended gradient

For any $u \in \mathrm{Dom}\mathscr{L}^{R,P}$ there exists a $dtP_t(dx)$-a.e. defined vector field $\tilde{\nabla}^{R,P} f$ such that Itô, formula holds, i.e.

$$df(t, X_t) = \mathscr{L}^{R,P} f(t, X_t)dt + \tilde{\nabla}^{R,P} f(t, X_t)dM_t^P$$

Theorem (Hamilton-Jacobi-Bellman equation)

The function $\psi_t = \log g_t$ is in the domain of the P-local extended generator of R $\mathscr{L}^{R,P}$ and it satisfies the extended pathwise HJB equation

$$\left[\mathscr{L}^{R,P}\psi_t + \frac{1}{2} |\tilde{\nabla}^{R,P}\psi_t|_A^2 + V\right](t, X_t) = 0, \quad dt \otimes dP - \text{a.e.}$$

Università degli Studi di Padova

A short story about time-reversal

"Je veux serrer les boulons dans un papier de Föllmer…"
CL, 05/2014

- H. Föllmer. Time reversal on Wiener space. In Stochastic Processes - Mathematics and Physics, volume 1158 of Lecture Notes in Math., pages 119–129. Springer, Berlin, 1986.

While "les boulons" were being tightened, time-reversal for diffusions revolutionized generative AI …

Published as a conference paper at ICLR 2021

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Yang Song*
Stanford University
yangsong@cs.stanford.edu

Jascha Sohl-Dickstein
Google Brain
jaschasd@google.com

Diederik P. Kingma
Google Brain
durk@google.com

Abhishek Kumar
Google Brain
abhishk@google.com

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

Ben Poole
Google Brain
pooleb@google.com

But in the end, 8 years later…

UNIVERSITÀ DEGLI STUDI DI PADOVA

Time-reversal

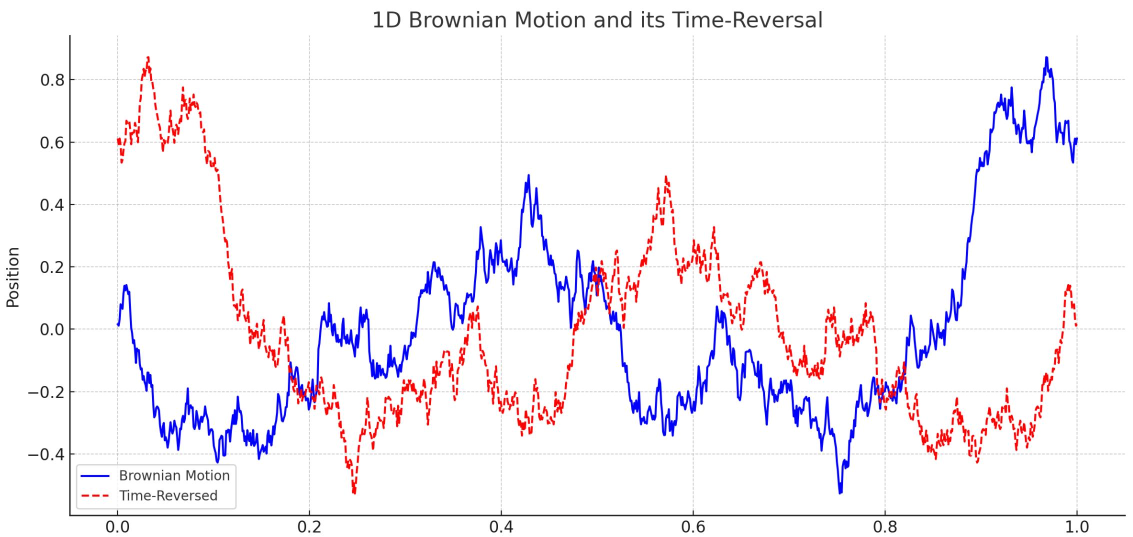
$$\boxed{(\mathscr{L}^{\overrightarrow{P}} f)_t = \frac{1}{2}\mathrm{Tr}(\nabla^2 f(X_t)A(X_t)) + \overrightarrow{b}_t(X_t) \cdot \nabla f(X_t)}$$

Assumptions

- R reversible w.r.t. m
- A is invertible
- Lyapunov condition
- And of course…. $H(\overrightarrow{P} \,|\, R) < +\infty$

Time-reversal of $\overrightarrow{P}$

$$\overleftarrow{P}[(X_t)_{t\in[0,T]} \in A] = \overrightarrow{P}[(X_{T-t})_{t\in[0,T]} \in A]$$



Time-reversal

Theorem

The time-reversal $\overleftarrow{P}$ is the unique solution of the martingale problem $\mathrm{MP}(\mathscr{C}_c^\infty, \mathscr{L}^{\overleftarrow{P}}; P_T)$ where

$$(\mathscr{L}^{\overleftarrow{P}} f)_t = \frac{1}{2}\mathrm{Tr}(\nabla^2 f(X_t)\, A(X_t)) + \overleftarrow{b}(X_t) \cdot \nabla f(X_t)$$

$$\overleftarrow{b}_t(x) = -\overrightarrow{b}_{T-t}(x) + \frac{\nabla \cdot (\mu_{T-t} A(x))}{\mu_{T-t}(x)}, \quad dtP_{T-t}(dx)\text{-a.e.}$$

where the divergence is in the sense of distributions.

Connection with Γ calculus and integration by parts

“Serrer les boulons” strikes again

Università degli Studi di Padova

10/2012: *“J’ai voulu serrer les boulons sur un papier de Csziszar”*

The Annals of Probability
1975, Vol. 3, No. 1, 146–158

I-DIVERGENCE GEOMETRY OF PROBABILITY DISTRIBUTIONS AND MINIMIZATION PROBLEMS

By I. Csiszár

Mathematical Institute of the Hungarian Academy of Sciences

Some geometric properties of PD’s are established, Kullback’s I-divergence playing the role of squared Euclidean distance. The minimum discrimination information problem is viewed as that of projecting a PD onto a convex set of PD’s and useful existence theorems for and characterizations of the minimizing PD are arrived at. A natural generalization of known iterative algorithms converging to the minimizing PD in special situations is given; even for those special cases, our convergence proof is more generally valid than those previously published. As corollaries of independent interest, generalizations of known results on the existence of PD’s or nonnegative matrices of a certain form are obtained. The Lagrange multiplier technique is not used.

10/2013: *Schrödinger problem and Sinkhorn algorithm arrive in ML!*

SINKHORN DISTANCES: LIGHTSPEED COMPUTATION OF OPTIMAL TRANSPORTATION DISTANCES

MARCO CUTURI

Abstract. Optimal transportation distances are a fundamental family of parameterized distances for histograms. Despite their appealing theoretical properties, excellent performance in retrieval tasks and intuitive formulation, their computation involves the resolution of a linear program whose cost is prohibitive whenever the histograms’ dimension exceeds a few hundreds. We propose in this work a new family of optimal transportation distances that look at transportation problems from a maximum-entropy perspective. We smooth the classical optimal transportation problem with an entropic regularization term, and show that the resulting optimum is also a distance which can be computed through Sinkhorn-Knopp’s matrix scaling algorithm at a speed that is several orders of magnitude faster than that of transportation solvers. We also report improved performance over classical optimal transportation distances on the MNIST benchmark problem.

“Serrer les boulons” never dies

UNIVERSITÀ DEGLI STUDI DI PADOVA

10/2012: *I started by “serrer les boulons” on Rüdiger’s work on reciprocal characteristics*

- STREATER, Raymond Frederick. *Lost Causes in and beyond Physics*. Springer Science & Business Media, 2007.

10/2022: *“Flow matching and stochastic interpolants algorithms”*

- Diffusion flow matching algorithms are **generative models** that compute the *Markovian projection* of a *reciprocal process*

Bonne retraite Christian!!