On the construction of set-valued dual processes

Laurent Miclo

Institut de Mathématiques de Toulouse
Toulouse School of Economics

Plan of the talk

Plan

Convergence to equilibrium

Consider $X := (X_t)_{t\geqslant 0}$ an ergodic Markov process on some state space V and let π be its invariant probability measure. One would like to estimate for large $t \geqslant 0$,

$$\|\mathcal{L}(X_t) - \pi\|_{\mathrm{tv}}$$

or any other quantity measuring the distance to equilibrium.
In practice, π is given and X is typically constructed via a Metropolis-type procedure. The deterministic **mixing time**

$$T_\epsilon \quad := \quad \inf\{t \,:\, \|\mathcal{L}(X_t) - \pi\|_{\mathrm{tv}} \leqslant \epsilon\}$$

enables to stop the algorithm X to get a r.v. X_{T_ϵ} sampled according to π, up to the precision $\epsilon \in (0, 1)$.

Strong stationary times

Strong stationary times provides a probabilistic approach to convergence to equilibrium: by looking at a given trajectory, one decides to stop it at a random time to get an *exact* sample of π.
A **strong stationary time** τ associated to X is a finite stopping time such that

$$\tau \perp\!\!\!\perp X_\tau \quad \text{and} \quad X_\tau \sim \pi$$

It can be used to deduce estimates on the speed of convergence:

$$\forall\ t \geqslant 0, \qquad \|\mathcal{L}(X_t) - \pi\|_{\mathrm{tv}} \leqslant \mathfrak{s}(\mathcal{L}(X_t), \mu) \leqslant \mathbb{P}[\tau > t]$$

in **total variation** and in **separation discrepancy**: for any probability measures μ and π on the same state space:

$$\mathfrak{s}(\mu, \pi) := \mathrm{esssup}_\pi 1 - \frac{d\mu}{d\pi} \geqslant \frac{1}{2} \left\| \frac{d\mu}{d\pi} - 1 \right\|_{\mathbb{L}^1(\pi)} =: \|\mu - \pi\|_{\mathrm{tv}}$$

Strong stationary times were introduced by Aldous and Diaconis [1986] to investigate the quantitative convergence to equilibrium of the top-to-random card shuffle.

How to obtain a strong stationary time?
Assume we can find an absorbed **dual** Markov process $\mathfrak{X} := (\mathfrak{X}_t)_{t \geqslant 0}$ on a state space $\mathfrak{V}$ such that there exist Λ a Markov kernel from $\mathfrak{V}$ to V satisfying the **intertwining relations**

$$\begin{aligned} \mathcal{L}(X_0) &= \mathcal{L}(\mathfrak{X}_0)\Lambda \\ \mathfrak{L}\Lambda &= \Lambda L \end{aligned}$$

where L and $\mathfrak{L}$ are the generators of X and $\mathfrak{X}$. Then there is a coupling of X and $\mathfrak{X}$ such that the absorption time for $\mathfrak{X}$ is a strong stationary time for X.
This method was developed by Diaconis and Fill [1990], at least for discrete time and finite state spaces V and $\mathfrak{V}$. The coupling was such that, for all $n \in \mathbb{Z}_+$,

$$\begin{aligned} \mathcal{L}(\mathfrak{X}_{[\![0,n]\!]}|X) &= \mathcal{L}(\mathfrak{X}_{[\![0,n]\!]}|X_{[\![0,n]\!]}) \\ \mathcal{L}(X_n|\mathfrak{X}_{[\![0,n]\!]}) &= \Lambda(\mathfrak{X}_n, \cdot) \end{aligned}$$

Set-valued dual processes

An interesting class of absorbed dual processes are those which are **set-valued**: $\mathfrak{V}$ is a nice subset of the set of the measurable subsets A of $\mathfrak{V}$ such that $\pi(A) > 0$ or A is a singleton. The kernel Λ corresponds to the **conditional expectation** under π: for any $A \in \mathfrak{V}$,

$$\Lambda(A, \cdot) \quad = \quad \begin{cases} \frac{\pi(A \cap \cdot)}{\pi(A)} & \text{, if } \pi(A) > 0 \\ \delta_x & \text{, if } A = \{x\} \end{cases}$$

Furthermore, the process $\mathfrak{X}$ is assumed to be absorbed at $V \in \mathfrak{V}$.

Pitman's theorem

A famous example is **Pitman's intertwining relation** between the Brownian motion and the Bessel process [Pitman 1975, Pitman and Rogers 1981]. Here $V = \mathbb{R}$, X is the Brownian motion starting from 0 and

$$\mathfrak{V} \quad := \quad \{[-r, r] \,:\, r \geqslant 0\}$$

The dual process $\mathfrak{X} := ([-R_t, R_t])_{t \geqslant 0}$ is given by

$$\forall\ t \geqslant 0, \qquad R_t \quad := \quad 2M_t^X - X_t$$

where $M^X := (M_t^X)_{t \geqslant 0}$ is the **maximum process**:

$$\forall\ t \geqslant 0, \qquad M_t^X \quad := \quad \max\{X_s \,:\, s \in [0, t]\}$$

The process $(R_t)_{t \geqslant 0}$ is known to be a **Bessel-3 process**, namely has the same law as the norm of a Brownian motion in dimension 3.

Pitman's theorem in picture

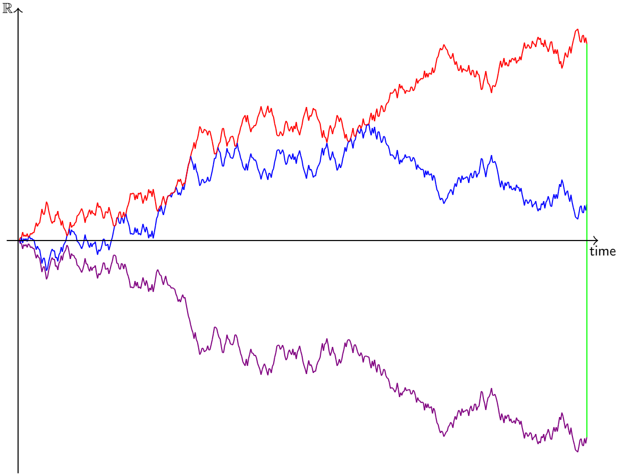

Figure: Trajectories: Brownian motion $B_{[0,t]}$, $R_{[0,t]}$, $-R_{[0,t]}$, and the segment-valued dual: $[-R_t, R_t]$

Plan

Random mappings (1)

Consider here the discrete time and finite state space situation. Let $P := (P(x,y))_{x,y\in V}$ be the transition matrix of $X := (X_n)_{n\in\mathbb{Z}_+}$. Assume P irreducible and let π be the invariant probability measure. The **adjoint transition matrix** $P^* := (P^*(x,y))$ is given by

$$\forall\, x,y \in V, \qquad P^*(x,y) \;\; := \;\; \frac{\pi(y)}{\pi(x)} P(y,x)$$

A random mapping $\psi \,:\, V \to V$ is said to be **associated to** P^* when

$$\forall\, x,x' \in V, \qquad \mathbb{P}[\psi(x) = x'] \;\; = \;\; P^*(x,x')$$

It is convenient to have at our disposal a family of random mappings $(\psi_S)_{S\in\bar{\mathfrak{V}}}$ associated to P^*, where $\bar{\mathfrak{V}} := \{A \,:\, A \subset V\}$.

Random mappings (2)

Such a family enables to define a random mapping Ψ from $\bar{\mathfrak{V}}$ to $\bar{\mathfrak{V}}$ via

$$\forall\ S \in \bar{\mathfrak{V}}, \qquad \Psi(S) \ := \ \{y \in V\ :\ \psi_S(y) \in S\}$$

Consider the transition matrix K from $\bar{\mathfrak{V}}$ to $\bar{\mathfrak{V}}$ given by

$$\forall\ S, S' \in \bar{\mathfrak{V}}, \qquad K(S, S') \ := \ \mathbb{P}[\Psi(S) = S']$$

as well as its **Doob transform**:

$$\forall\ S, S' \in \mathfrak{V}, \qquad \mathfrak{P}(S, S') \ := \ \frac{\pi(S')}{\pi(S)} K(S, S')$$

where $\mathfrak{V} = \bar{\mathfrak{V}} \backslash \{\emptyset\}$.

Conditioned random mappings

For $x, x' \in V$ with $P(x, x') > 0$ and $S \in \mathfrak{V}$ containing x, denote

$$\forall\ S' \in \mathfrak{V}, \qquad K_{x,x'}(S, S') \ := \ \mathbb{P}[\Psi(S) = S' | \psi_S(x') = x]$$

Note that the conditioning is non-degenerate, since $\mathbb{P}[\psi_S(x') = x] = P^*(x', x) > 0$. Consider

$$W \ := \ \{(x, S) \in V \times \mathfrak{V} \,:\, x \in S\}$$

and let $\mathcal{A}$ be the set of probability measures m on W of the form

$$\forall\ (x, S) \in W, \qquad m(x, S) \ = \ \mu(S)\Lambda(S, x)$$

where μ is the marginal of m on $\mathfrak{V}$. Define a Markov kernel Q on W via

$$\forall\ (x, S), (x', S') \in W, \qquad Q((x, S), (x', S')) \ := \ P(x, x')K_{x,x'}(S, S')$$

Principle of the construction via random mappings

Theorem 1

Let $(X_n, \mathfrak{X}_n)_{n\in\mathbb{Z}_+}$ be a Markov chain on W whose initial distribution $\mathcal{L}(X_0, \mathfrak{X}_0)$ belongs to $\mathcal{A}$ and whose transitions are given by Q. Then $\mathfrak{X} := (\mathfrak{X}_n)_{n\in\mathbb{N}}$ is a set-valued absorbed dual for $X := (X_n)_{n\in\mathbb{N}}$ whose transitions are given by $\mathfrak{P}$.

This result is related to the coupling-from-the-past algorithm of Propp and Wilson [1996] and to the evolving set process of Morris and Peres [2005]. There is an improvement based on random mappings weakly associated to P^*.

A corresponding algorithm

Let be given a trajectory $(x_n)_{n\in\mathbb{Z}_+}$ of X. We start with $\mathfrak{X}_0 := \{x_0\}$ (this can sometimes be improved, for instance one can take $\mathfrak{X}_0 = S$ if $\mathcal{L}(X_0) = \Lambda(S,\cdot)$). Assume next that $\mathfrak{X}_n$ has been constructed for some for $n \in \mathbb{Z}_+$. We consider a random mapping $\psi_{\mathfrak{X}_n}$ weakly associated to P^*, whose law may depend on $\mathfrak{X}_n$ (but not directly on $(x_m)_{m\in[\![0,n]\!]}$) and whose underlying randomness is independent from all that has been done before. We condition by the fact that $\psi_{\mathfrak{X}_n}(x_{n+1}) = x_n$ and we sample a corresponding mapping φ, to construct

$$\mathfrak{X}_{n+1} \quad := \quad \{y \in V \ : \ \varphi(y) \in \mathfrak{X}_n\}$$

Due to the conditioning, we are sure that $x_{n+1} \in \mathfrak{X}_{n+1}$.
This works for any random mapping associated to P^*, all the difficulty stays on relevant choices leading to (close to optimal) strong stationary times.

Plan

A denumerable state space example

Consider the transition "matrix" of the simple random walk on $\mathbb{Z}$:

$$\forall\ x, y \in \mathbb{Z}, \qquad P(x,y) \quad := \quad \begin{cases} 1/2 & , \text{if } |y-x| = 1 \\ 0 & , \text{otherwise} \end{cases}$$

An invariant measure π is the counting measure, it is even reversible for P, in the sense that $P^* = P$. The kernel Λ is still well-defined, if we take for $\mathfrak{V}$ the set of finite non-empty subsets of $\mathbb{Z}$.
Let $X := (X_n)_{n\in\mathbb{Z}_+}$ be a random walk with transition kernel P and starting from 0. Introduce the process $R^\vee := (R_n^\vee)_{n\in\mathbb{Z}_+}$ defined by

$$\forall\ n \in \mathbb{Z}_+, \qquad R_n^\vee \quad := \quad 2\max\{X_m \ :\ m \in [\![0,n]\!]\} - X_n$$

Finally consider $\mathfrak{X} := (\mathfrak{X}_n)_{n\in\mathbb{Z}_+}$ given by

$$\forall\ n \in \mathbb{Z}_+, \qquad \mathfrak{X}_n \quad := \quad \{X_n^\vee - 2m \ :\ m \in [\![0, X_n^\vee]\!]\}$$

Pitman [1975] has shown that $\mathfrak{X}$ is a set-valued dual for X.

A corresponding random mapping

Consider the function ψ given by

$$\forall\ S \in \mathfrak{V},\ \forall\ x \in \mathbb{Z},\ \forall\ b \in \{-1, 1\},$$
$$\psi(S, x, b) \quad := \quad \begin{cases} x + b & \text{, if } x > \max(S) \\ x - b & \text{, if } x \leqslant \max(S) \end{cases}$$

Consider a Rademacher variable B, i.e. such that $\mathbb{P}[B = -1] = \mathbb{P}[B = 1] = 1/2$ and for fixed $S \in \mathfrak{V}$, let ψ_S be the random mapping given by

$$\forall\ x \in \mathbb{Z}, \qquad \psi_S(x) \quad := \quad \psi(S, x, B)$$

It is clear that ψ_S is a random mapping associated to $P^* = P$. The discrete Pitman's theorem can be easily deduced from Theorem 1 with the family $(\psi_S)_{s \in \mathfrak{V}}$.

Schematic proof

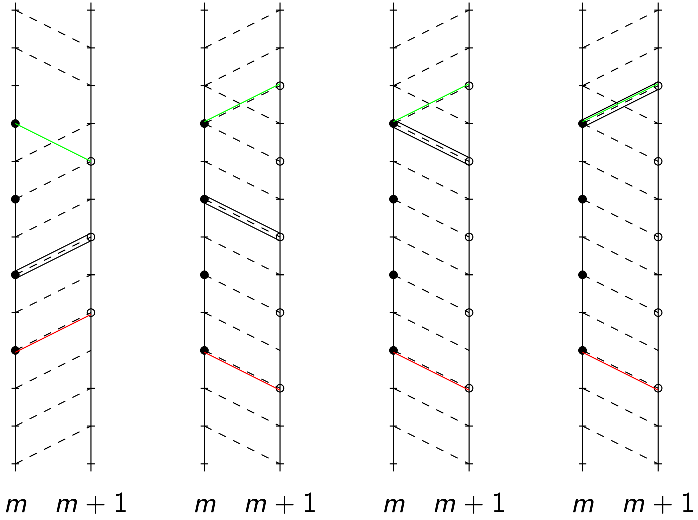

Figure: Schematic proof of the discrete Pitman theorem via random mappings

Plan

A card shuffle example

The **top-to-random shuffle** takes the top card of a deck and put it at a uniform random location inside. With N the number of cards, it leads to a Markov chain $X := (X_n)_{n\in\mathbb{Z}_+}$ on the symmetric group $V := \mathcal{S}_N$, starting from id and whose transition matrix $P := (P(\sigma,\sigma'))_{\sigma,\sigma'\in\mathcal{S}_N}$ is given by

$$P(\sigma,\sigma') := \begin{cases} 1/N & \text{, if there exists } l \in [\![1,N]\!] \text{ with} \\ & \sigma' = (1 \to l \to l-1 \to \cdots \to 2)\circ\sigma \\ 0 & \text{, otherwise} \end{cases}$$

P is irreducible and its invariant probability π is the uniform probability distribution on $\mathcal{S}_N$. The Markov chain X admits a famous set-valued dual process $\widetilde{\mathfrak{X}} := (\widetilde{\mathfrak{X}}_n)_{n\in\mathbb{Z}_+}$ defined by Aldous and Diaconis [1986]:

$$\forall\ n \in \mathbb{Z}_+, \qquad \widetilde{\mathfrak{X}}_n := A_{X_n,Y_n}$$

where $Y_n \in [\![1,N]\!]$ is the position of the initial last card and with for any $\sigma \in \mathcal{S}_N$ and $y \in [\![0,N]\!]$,

$$A_{\sigma,y} := \{\sigma' \in \mathcal{S}_N : \sigma'(1) = \sigma(1), \ldots, \sigma'(y) = \sigma(y)\}$$

A corresponding random mapping (1)

The associated strong stationary time $\widetilde{\tau}$ is first time the initial last card arrives at the top of the deck and is inserted. It is easy to check that $\mathbb{E}[\widetilde{\tau}] \sim N\ln(N)$.
Let us check that first time τ the initial last-but-one card arrives at the top of the deck and is inserted is also a strong stationary time, strictly better than $\widetilde{\tau}$, but since $\mathbb{E}[\widetilde{\tau}] - \mathbb{E}[\tau] = N$, the improvement is not very significant. The proof uses random mappings.
Here P^* is the transition matrix of the **random-to-top shuffle** and corresponds to taking a card of the deck at a uniform random location and putting it at the top. Consider for any $x \in [\![1, N]\!]$, the mapping $\psi^{(x)} : \mathcal{S}_N \to \mathcal{S}_N$ which acts on any permutation σ by removing the card x from the deck and putting it at the top.
Formally, we have

$$\forall\ \sigma \in \mathcal{S}_N, \qquad \psi^{(x)}(\sigma) \;=\; (1 \to 2 \to \cdots \to \sigma^{-1}(x)) \circ \sigma$$

($\sigma^{-1}(x)$ is the position of the card x).

A corresponding random mapping (2)

Let $(U_n)_{n\in\mathbb{N}}$ be a family of independent random variables uniformly distributed on $[\![1,N]\!]$. At any time $n\in\mathbb{N}$, consider the random mapping $\psi^{(U_n)}$, it is associated to P^*. Here there is no dependence on a subset $S\in\mathfrak{V}$.
Let be given a trajectory $x_{[\![0,n]\!]}$, for some fixed $n\in\mathbb{Z}_+$, starting from the identity, $x_0=\mathrm{id}$. For any $m\in[\![1,n]\!]$, let φ_m be the conditioning of $\psi^{(U_m)}$ by $\psi^{(U_m)}(x_m)=x_{m-1}$. As in the previous example, φ_m is deterministic, as we have $\varphi_m=\psi^{(x_{m-1}(1))}$.
It can be checked that

$$\forall\ n\in\mathbb{Z}_+,\qquad \mathfrak{X}_n \;=\; \{\sigma\in\mathcal{S}_N\ :\ \varphi_1\circ\varphi_2\circ\cdots\circ\varphi_n(\sigma)=\mathrm{id}\}$$

and that the corresponding absorption time at $\mathcal{S}_N$ is τ.

Plan

Extensions

Theorem 1 admits several extensions:

• the state space V can be a Polish space,

• the time can be continuous (for the moment, only in a regular diffusion framework).

The underlying theory is quite involved and misses the existence of convenient stochastic coalescing flows depending on subset evolutions, despite the works of Le Jan and Raymond [2004, 2006]. Let us just illustrate the use of stochastic coalescing flows on the classical Brownian motion:

Theorem 2

The process $\mathfrak{X} = ([-(L_0^X(t) + |X_t|), L_0^X(t) + |X_t|])_{t \geqslant 0}$ *is a set-valued dual for the Brownian motion* X *starting from 0, where* $L_0^X := (L_0^X(t))_{t \geqslant 0}$ *is the local time of* X *at 0.*

One recover Pitman's theorem, by taking into account that the two processes $(L_0^X + |X|, X)$ and $(2M^X - X, X)$ have the same law.

Stochastic coalescing flows

Let $B := (B_s)_{s \geqslant 0}$ be another Brownian motion. Consider the following system of equations, for all $t \geqslant 0$ and $y \in \mathbb{R}$,

$$\begin{cases} dY_s^{(t)}(y) = -\mathrm{sgn}(Y_s^{(t)}(y)) dB_s^{(t)}, & \forall\ s \in [0,t] \\ Y_0^{(t)}(y) = y \end{cases} \tag{1}$$

where sgn equals -1 on $(-\infty, 0]$ and 1 on $(0, +\infty)$ and where $B^{(t)} := (B_{t-s})_{s \in [0,t]}$ is the time-reversed process associated to B at time $t \geqslant 0$.

Le Jan and Raimond [2006] provide a (non-Wiener) coalescing stochastic flow solution to this system.

Define $\psi := (\psi_{s,t}(y))_{(s,t,y) \in \triangle \times \mathbb{R}}$ via

$$\forall\ x \in \mathbb{R},\ \forall\ 0 \leqslant s \leqslant t, \qquad \psi_{s,t}(y) := Y_{t-s}^{(t)}(y)$$

with $\triangle := \{(s,t) : 0 \leqslant s \leqslant t\}$. By monotonicity in y, there is a version of ψ which is càdlàg in y.

Conditioned stochastic coalescing flows (1)

Fix $t \geqslant 0$ and a Brownian trajectory $X_{[0,t]}$. Conditioning ψ by the event

$$\forall\ s \in [0,t], \qquad \psi_{s,t}(X_t) \ = \ X_s$$

implies in particular that $X^{(t)}$ is a solution to Tanaka's stochastic differential equation:

$$\forall\ s \in [0,t], \qquad dX_s^{(t)} \ = \ -\mathrm{sgn}(X_s^{(t)})dB_s^{(t)}$$

We deduce that the conditioned flow φ is given by

$$\forall\ 0 \leqslant s \leqslant t, \forall\ z \in \mathbb{R}, \qquad \varphi_{s,t}(z) \ := \ Z_s^{(t)}(z)$$

where

$$\begin{cases} dZ_s^{(t)}(z) & = & \mathrm{sgn}(Z_s^{(t)}(z))\mathrm{sgn}(X_s^{(t)})dX_s^{(t)} \\ Z_0^{(t)}(z) & = & z \end{cases}$$

This system is the same as (1), once $B^{(t)}$ is replaced by the standard Brownian motion $(-\int_0^s \mathrm{sgn}(X_v^{(t)})\, dX_v^{(t)})_{s\in[0,t]}$.

Conditioned stochastic coalescing flows (2)

Theorem 2 is deduced from the fact that $\mathfrak{X} := (\mathfrak{X}_t)_{t\geqslant 0}$, defined by

$$\forall\ t \geqslant 0, \qquad \mathfrak{X}_t \ := \ \varphi_{0,t}^{-1}(\{0\})$$

is a set-valued dual for X and that we can compute

$$\forall\ t > 0, \qquad \mathfrak{X}_t \ = \ [-(L_0^X(t) + |X_t|), L_0^X(t) + |X_t|)$$

In the underlying theory, one has to consider

$$\mathfrak{V} \ := \ \{[a, b)\ :\ a < b \in \mathbb{R}\}$$

to which is added the initial subset $\mathfrak{X}(0) = \{0\}$.

What we really want

We would like to have at our disposal a solution to the following system of equations, for all $0 \leqslant s \leqslant t$ and $y \in \mathbb{R}$,

$$\begin{cases} dY_s^{(t)}(y) &=& -\mathrm{sgn}(Y_s^{(t)}(y) - R_{t-s}^{\vee})dB_s^{(t)} + b(Y_s^{(t)}(y))ds \\ Y_0^{(t)}(y) &=& y \\ R_{t-s}^{\vee} &:=& \max\{z \in \mathbb{R} \,:\, Y_{t-s}^{(t-s)}(z) \in \mathfrak{X}_0\} \end{cases}$$

where $b : \mathbb{R} \to \mathbb{R}$ is a nice drift.
For $b = 0$, it would lead to a proof of Pitman's theorem similar to that given in the discrete situation. Above all, for $b \neq 0$, it would provide a direct coupling of the diffusion X solution of the s.d.e.

$$\forall\ t \geqslant 0, \qquad dX_t \quad = \quad dW_t + b(X_t)dt$$

(where $W := (W_t)_{t \geqslant 0}$ is a Brownian motion) with a non-trivial segment-valued dual process. It would open the way for multidimensional and hypo-elliptic extensions, which are the remote motivation for this work.

Plan

Primary references on intertwining relations

- James W. Pitman.
  One-dimensional Brownian motion and the three-dimensional Bessel process.
  Advances in Appl. Probability, 7(3):511–526, 1975.

- L. Chris G. Rogers and James W. Pitman.
  Markov functions.
  Ann. Probab., 9(4):573–582, 1981.

- David Aldous and Persi Diaconis.
  Shuffling cards and stopping times.
  Amer. Math. Monthly, 93(5):333–348, 1986.

- Persi Diaconis and James Allen Fill.
  Strong stationary times via a new form of duality.
  Ann. Probab., 18(4):1483–1522, 1990.

- David A. Levin, Yuval Peres, and Elizabeth L. Wilmer.
  Markov chains and mixing times.
  American Mathematical Society, Providence, RI, 2009.

References on some *geometric* processes

James Gary Propp and David Bruce Wilson.
Exact sampling with coupled Markov chains and applications to statistical mechanics.
In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 223–252, 1996.

Ben Morris and Yuval Peres.
Evolving sets, mixing and heat kernel bounds.
Probab. Theory Related Fields, 133(2):245–266, 2005.

Yves Le Jan and Olivier Raimond.
Flows, coalescence and noise.
Ann. Probab., 32(2):1247–1315, 2004.

Yves Le Jan and Olivier Raimond.
Flows associated to Tanaka's SDE.
ALEA Lat. Am. J. Probab. Math. Stat., 1:21–34, 2006.

Further references on Markov intertwinings

Philippe Carmona, Frédérique Petit and Marc Yor.
Beta-gamma random variables and intertwining relations between certain Markov processes.
Rev. Mat. Iberoamericana, 14(2):311–367, 1998.

Alexei Borodin and Grigori Olshanski.
Markov processes on the path space of the Gelfand-Tsetlin graph and on its boundary.
J. Funct. Anal., 263(1):248–303, 2012.

Soumik Pal and Mykhaylo Shkolnikov.
Intertwining diffusions and wave equations.
ArXiv e-prints, June 2013.

Laurent Miclo.
Strong stationary times for one-dimensional diffusions.
Ann. Inst. Henri Poincaré Probab. Stat., 53(2):957–996, 2017.