Markov processes conditioned on large deviations

Hugo Touchette

National Institute for Theoretical Physics (NITheP)
Stellenbosch, South Africa

44th Dutch Stochastics Meeting
Lunteren, The Netherlands
9 November 2015

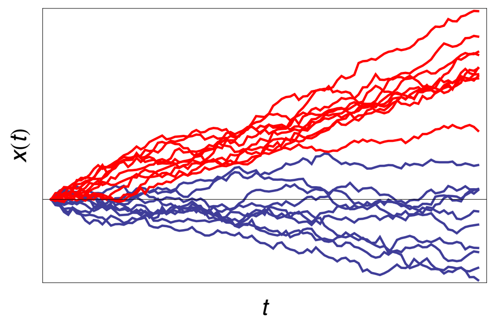
Work with

- Raphaël Chetrite (Nice, France)
- Florian Angeletti (post-doc, NITheP)
- Pelerine Tsobgni (PhD student, NITheP)

Problem

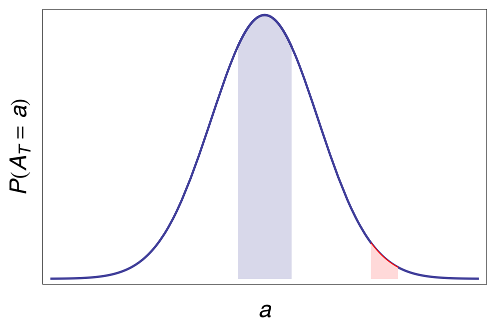
Physical

- Stochastic process: X_t
- Observable: $A_T[x]$
- Look at trajectories leading to $A_T = a$
- Find an effective process that describes these trajectories



Mathematical

- Markov process: $\{X_t\}_{t=0}^T$
- Conditioned process: $X_t | A_T = a$
- Is it a Markov process?
- Construct its generator



Markov conditioning

- State conditioning [Doob 1957]

$$X_t \mid X_T \in \mathcal{A} \qquad \text{target point or set}$$

- Schrödinger bridge [Schrödinger 1931]

$$X_t \mid p(x, T) = q(x) \qquad \text{target distribution}$$

- Quasi-stationary distributions

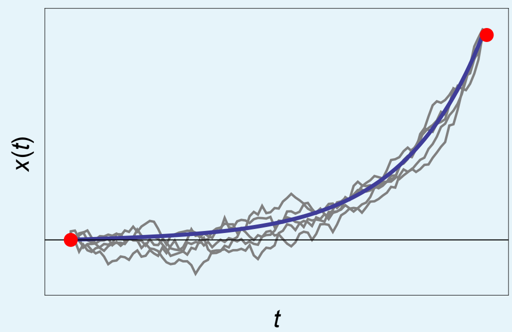
$$\underbrace{X_t}_{\text{absorbing}} \mid \text{not reaching absorbing state} \equiv \underbrace{\hat{X}_t}_{\text{non-absorbing}}$$

Here

- $X_t | A_T$ with A_T defined on $[0, T]$
- Requires generalization of Doob's transform
- Asymptotic equivalence

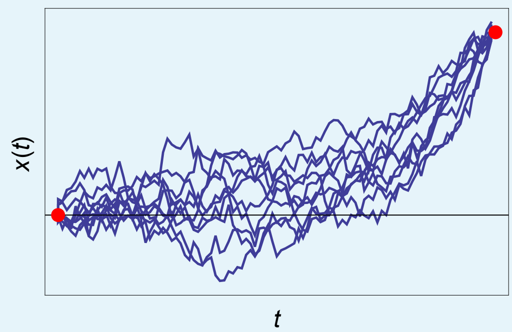
Comparaison with optimal paths

Low noise



- Fluctuation created by single path
- Reactive path or instanton
- Freidlin-Wentzell 70s
- Onsager-Machlup 50s

Not low noise



- Fluctuation created by many paths
- No reactive path
- Whole fluctuation process
- Effective process

Process

- Markov process: X_t
  - One or many particles
  - Equilibrium or nonequilibrium
  - Includes external forces, reservoirs
- Master (Fokker-Planck) equation:

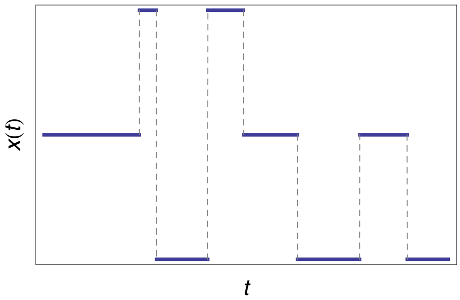
$$\partial_t p(x,t) = L^\dagger p(x,t)$$

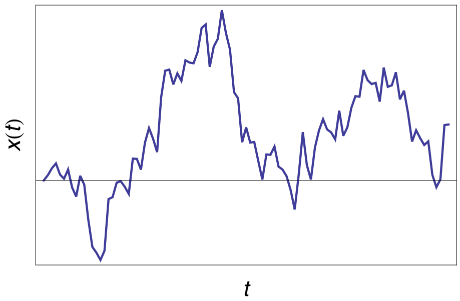
- Generator:

$$\partial_t E_x[f(X_t)] = E_x[Lf(X_t)]$$

- Path measure:

$$P[x] = P(\{x_t\}_{t=0}^T)$$





Examples of Markov processes

Pure jump process

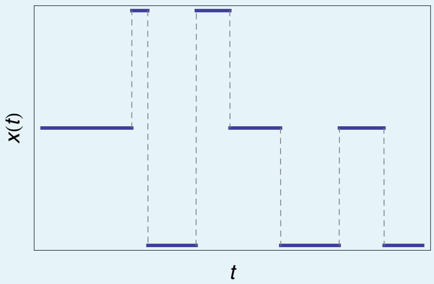
- Transition rates:

$$W(x,y) = P(x \to y \text{ in } dt)/dt$$

- Escape rates:

$$\lambda(x) = \sum_y W(x,y) = (W1)(x)$$

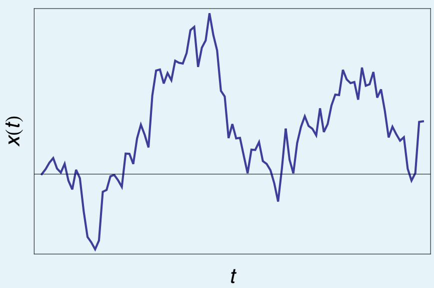
- Generator: $L = \underbrace{W}_{\text{off-diag}} - \underbrace{\lambda}_{\text{diag}}$



Pure diffusion

- SDE: $dX_t = F(X_t)dt + \sigma dW_t$
- Generator:

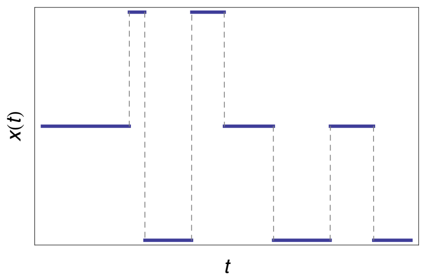
$$L = F \cdot \nabla + \frac{D}{2}\nabla^2, \qquad D = \sigma\sigma^T$$



Conditioning observable

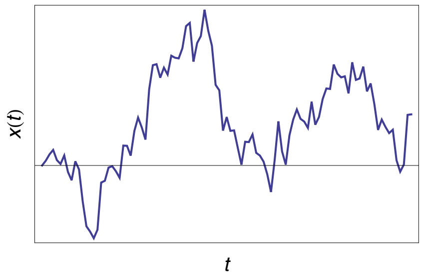
- Random variable: $A_T[x]$
- Jump processes:

$$A_T = \frac{1}{T}\int_0^T f(X_t)\,dt + \frac{1}{T}\sum_{\Delta X_t \neq 0} g(X_{t^-}, X_{t^+})$$



- Diffusions:

$$A_T = \frac{1}{T}\int_0^T f(X_t)\,dt + \frac{1}{T}\int_0^T g(X_t)\circ dX_t$$



Examples

- Occupation time $X_t \in \Delta$
- Mean number jumps (activity), current
- Work, heat, entropy production,...

Rare event conditioning

Large deviation principle

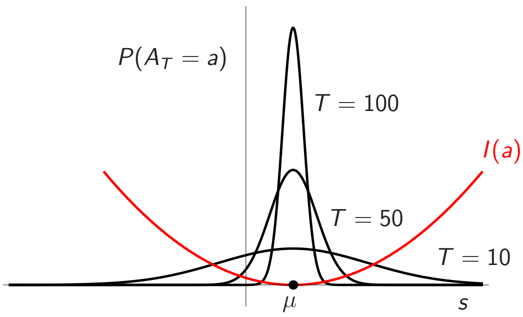
$$P(A_T = a) \approx e^{-TI(a)}$$

- Meaning of $\approx$:

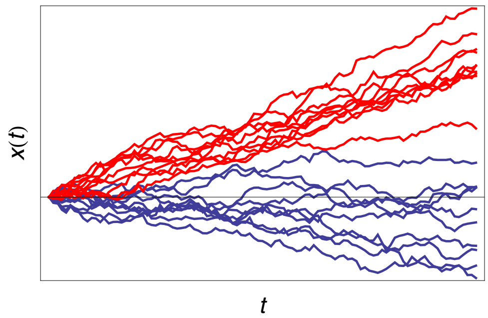
$$\lim_{T\to\infty} -\frac{1}{T}\ln P(A_T = a) = I(a), \qquad P(A_T = a) = e^{-TI(a)+o(T)}$$

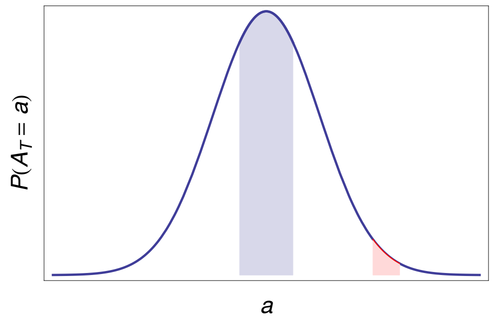
- Rate function: $I(a)$

- Exponentially rare fluctuations
- Applies to many process / observables
- Zero of I = Law of Large Numbers
- Small fluctuations = CLT



Conditioned process





- Conditioned process: $X_t | A_T = a$
- Path distribution:

$$P^a[x] = P[x|A_T = a] = \frac{P[x, A_T = a]}{P(A_T = a)} = P[x]\,\frac{\delta(A_T[x] - a)}{P(A_T = a)}$$

- Path microcanonical ensemble

- Not necessarily Markov for $T < \infty$
- Becomes equivalent to Markov process as $T \to \infty$
- Driven process $\hat{X}_t$

Spectral elements

Scaled cumulant function

$$\Lambda_k = \lim_{T\to\infty} \frac{1}{T} \ln E[e^{TkA_T}]$$

- $k \in \mathbb{R}$

Gärtner-Ellis Theorem

Λ_k differentiable, then

1. LDP for A_T
2. $I(a) = \sup_k \{ka - \Lambda_k\}$

Perron-Frobenius

$$\mathcal{L}_k r_k = \Lambda_k r_k$$

- Tilted (twisted) operator: $\mathcal{L}_k$
- Dominant eigenvalue: Λ_k
- Dominant eigenfunction: r_k

Jump processes

$$\mathcal{L}_k = We^{kg} - \lambda + kf$$

Diffusions

$$\mathcal{L}_k = F\cdot(\nabla + kg) + \frac{D}{2}(\nabla + kg)^2 + kf$$

Driven process $\hat{X}_t$

Generator

$$L_k = r_k^{-1}\mathcal{L}_k r_k - r_k^{-1}(\mathcal{L}_k r_k)$$

- Generalization of Doob's transform (1957)
- Action:

$$(L_k h)(x) = \frac{1}{r_k(x)}(\mathcal{L}_k r_k h)(x) - \Lambda_k h(x)$$

- Markov operator: $(L_k 1) = 0$
- Path distribution:

$$\underbrace{P_k^{\text{driven}}[x]}_{\text{new}} = r_k^{-1}(X_0)\, e^{T(kA_T - \Lambda_k)}\, r_k(X_T)\ \underbrace{P[x]}_{\text{original}}$$

Main result

Hypotheses

- A_T satisfies LDP
- Rate function $I(a)$ convex
- Other properties of spectral elements (gap, regular r_k)

Result

Conditioned		Driven	
$X_t \vert A_T = a$	$\overset{T\to\infty}{\cong}$	$\hat{X}_t$	$k(a) = I'(a)$
$P^a[x]$	$\asymp$	$P_{k(a)}^{\text{driven}}[x]$	almost everywhere
$A_T = a$		$A_T \to a$	in probability

- Driven process realizes conditioning $A_T = a$
- Same typical states for other observables

Interpretation of equivalence

Measure equivalence

$$\lim_{T\to\infty} \frac{1}{T} \ln \frac{dP^a}{dP_k^{\text{driven}}}(\omega) = 0, \qquad P^a \text{ or } P_k^{\text{driven}}\text{-a.e.}$$

- $P^a[x] = P_k^{\text{driven}}[x]\, e^{o(T)}$
- Same path measures on log scale

Observable equivalence

Conditioned		Driven
$X_t \| A_T = a$		$\hat{X}_t$
$B_T \to b^*$	$\Leftrightarrow$	$B_T \to b^*$

- Same typical states
- Different fluctuations (LDPs) in general

Idea of the proof

Microcanonical	Canonical	Driven
$X_t \| A_T = a$		$\hat{X}_t$
	$P_k^{\text{cano}}[x] = \dfrac{e^{kTA_T[x]}}{E[e^{kTA_T}]} P[x]$	
$P^a[x] = P[x \| A_T = a]$		$P_k^{\text{driven}}[x]$

Driven $\to$ canonical

- $P_k^{\text{driven}}[x] \asymp P_k^{\text{cano}}[x]$
- Same large deviations

Microcanonical $\to$ canonical

- $P^a[x] \asymp P_k^{\text{cano}}[x]$ if $I(a)$ convex
- Same typical states
- General result about conditioning vs tilting

Driven process: Explicit form

Jump process

- Original process: $W(x, y)$
- Driven process:

$$W_k(x, y) = r_k^{-1}(x)\, W(x, y)\, e^{kg(x,y)}\, r_k(y), \qquad k = I'(a)$$

- [Evans PRL 2004, Jack and Sollich PTPS 2010]

Diffusion

- Reference SDE:

$$dX_t = F(X_t)dt + \sigma dW_t$$

- Driven SDE:

$$dY_t = F_k(Y_t)dt + \sigma dW_t$$

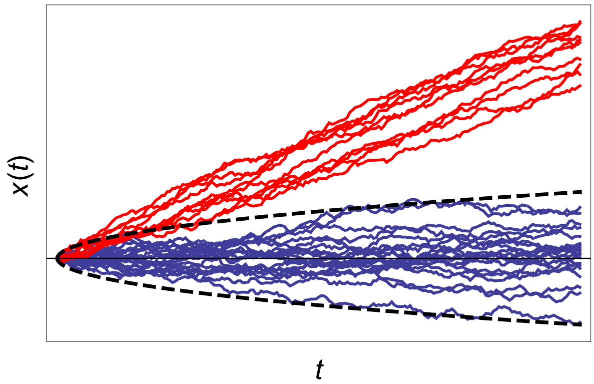
- Modified drift:

$$F_k(y) = F(y) + D(kg + \nabla \ln r_k), \qquad k = I'(a)$$

Application: Brownian motion

- Process: W_t
- Typical trajectories: $w_t \sim \sqrt{t}$
- Atypical trajectories:

$$A_T = \frac{W_T}{T} = \frac{1}{T}\int_0^T dW_t$$



Effective process

$$W_t | A_T = a \quad \overset{T\to\infty}{\cong} \quad \hat{X}_t = W_t + \underbrace{at}_{\text{added drift}}$$

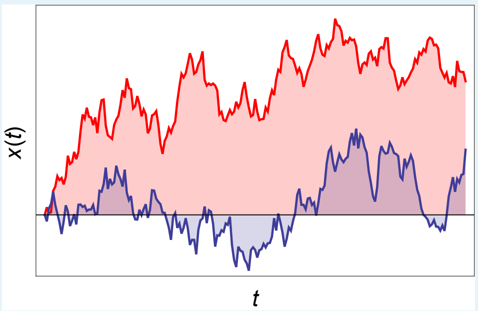
Langevin equation

$$dX_t = -\gamma X_t dt + \sigma dW_t$$

Area

$$A_T = \frac{1}{T}\int_0^T X_t dt$$

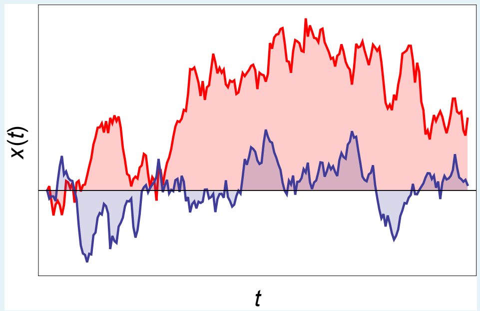
- $F_a(x) = -\gamma x + \gamma a$
- Modified drift



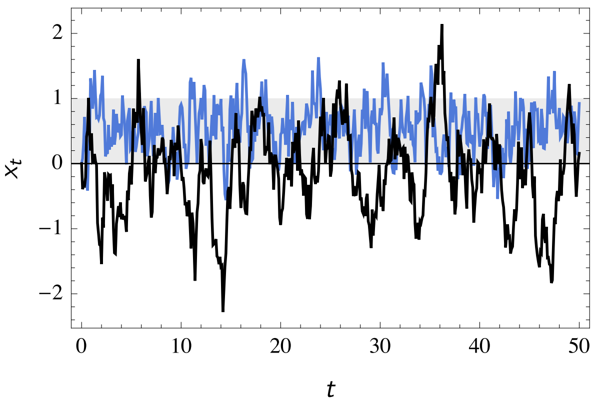
Empirical variance

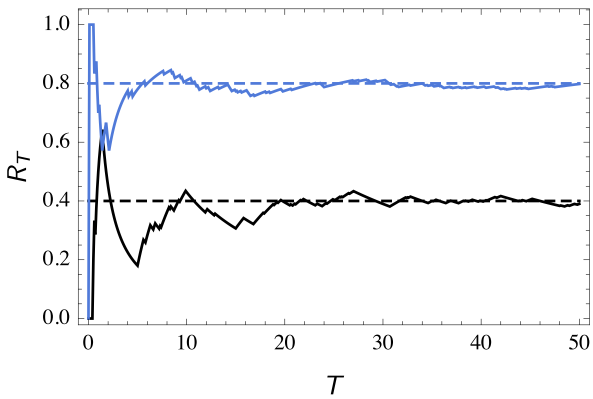
$$A_T = \frac{1}{T}\int_0^T X_t^2 dt$$

- $F_a = -\sigma^2 x/(2a)$
- Modified friction



Occupation conditioning





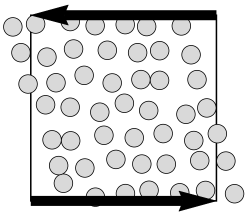
- Occupation set: S
- Occupation measure:

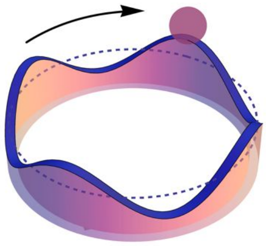
$$R_T = \frac{1}{T}\int_0^T \mathbb{1}_S(X_t)dt$$

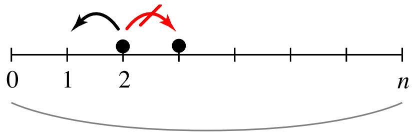
- Conditioning: $X_t | R_T = r$
- Related to: Brownian meander, quasi-stationary distributions
  [Angeletti & HT arxiv:1510.04893]

Applications

- Sheared fluids
- Interacting particle systems
  - TASEP, ZRP, Glauber-Ising, rotators
  - Conditioning: Occupation, current
- Diffusions
  - Occupation, current
- Chemical reactions
  - Occupation, concentration
- Random walks on graphs
  - Occupation, degree, etc.





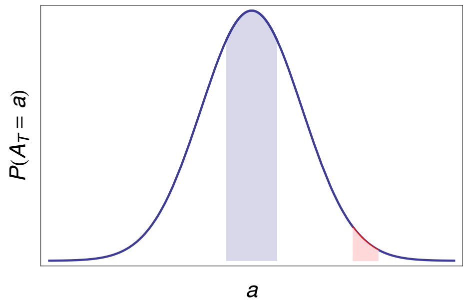


Conditioning induces non-local forces/long-range interactions

Large deviation simulations

- $P(A_T = a) \sim e^{-TI(a)}$
- Direct sampling:

$$\text{sample size } = L \sim e^{T}$$



Importance sampling - reweighting

- Change process: $X_t \to Y_t$
- Make $A_T = a$ typical
- Change of measure:

$$P(A_T = a) = E_X[\delta(A_T - a)] = E_Y\left[\frac{dP_X}{dP_Y}\delta(A_T - a)\right]$$

- Estimator:

$$\hat{P}_L(a) = \frac{1}{L}\sum_{j=1}^{L}\delta(A_T^{(j)} - a)\frac{dP_X}{dP_Y}$$

Large deviation simulations (cont'd)

- Driven process $\hat{X}_t$ makes $A_T = a$ typical
- Good change of process
- Optimal: estimator has asymptotic zero variance

Problem: $\hat{X}_t$ requires r_k, Λ_k and $I(a)$

Learning algorithm

- Direct sampling + feedback $\rightarrow$ iterative estimation of r_k
- Adaptive importance sampling
- Controlled sampling leading to driven process
- Markov chains: [Borkar 2008]
- Diffusions: [work with Florian Angeletti]
- Similar to [Nemoto-Sasa 2014]
- Different from cloning/splitting

Conclusion

$$\underbrace{X_t \,|\, A_T = a}_{\text{conditioned microcanonical}} \quad \stackrel{T\to\infty}{\cong} \quad \underbrace{\hat{X}_t}_{\text{driven canonical}}$$

- Effective Markov dynamics for fluctuations
- Process (ensemble) equivalence

Other links

- Variational principles (maxent, caliber)
- Stochastic control
- Quasi-stationary distributions
- Conditional limit theorems (Gibbs conditioning)

Ongoing work

- Nonequilibrium systems
- Simulation of rare events

References

R. Chetrite, H. Touchette
Nonequilibrium microcanonical and canonical ensembles...
Phys. Rev. Lett. **111**, 120601, 2013

R. Chetrite, H. Touchette
Nonequilibrium Markov processes conditioned on large deviations
Ann. Henri Poincaré **16**, 2005, 2015

R. Chetrite, H. Touchette
Variational and optimal control representations...
arxiv:1506.05291

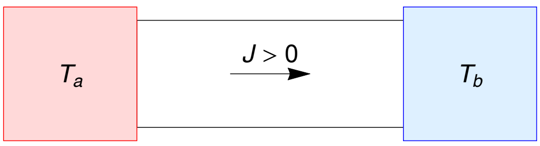
F. Angeletti, H. Touchette
Diffusions conditioned on occupation measures
arxiv:1510.04893

`www.physics.sun.ac.za/~htouchette`

Funded by Stellenbosch University and NRF

Nonequilibrium systems

Nonequilibrium

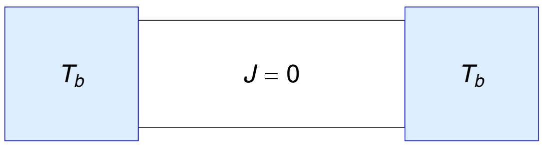


- Microscopic dynamics:

$$W^{\text{noneq}}(x \to y)?$$

- Many models possible

Equilibrium



- Microscopic dynamics known
- Detailed balance:

$$\frac{W^{\text{eq}}(x \to y)}{W^{\text{eq}}(y \to x)} = e^{\beta \Delta E}$$

Evans's hypothesis [PRL 2004; JPA 2005]

$$W^{\text{noneq}}(x \to y) = W^{\text{eq}}(x \to y | J)$$

- Nonequilibrium = conditioning of equilibrium

Conditional limit theorems

- Sequence of RVs:

$$X_1, X_2, \ldots, X_n, \qquad X_i \sim P(x)$$

- Sample mean:

$$S_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$$

Conditional marginal

$$\lim_{n\to\infty} P(X_i = x | S_n = s) = \frac{e^{kf(x)}}{E[e^{kf(X)}]} P(x)$$

- Modified marginal under conditioning
- Similar results for Markov chains [Csiszar, Cover and Choi]
- No results for general Markov processes

Control representations of PDEs

[Fleming, Sheu, Whittle, Dupuis-Ellis: 80s and 90s]

PDE $\phi(x,t)$ $\partial_t \phi = L\phi$ $\overset{I=-\ln\phi}{\longrightarrow}$ Hamilton-Jacobi equation (Hopf-Cole)

$\downarrow$

Dynamic programming

$\downarrow$

Optimal stochastic control = Doob transform

- Optimal control for specific PDEs = driven process
- Provides control characterization of conditioning
- No conditioning in Fleming
- No asymptotic equivalence (conditioning = driven)