Stability for Nonlinear Filtering

Continuous Time Noncompact Case

Alexander Veretennikov[1] Marina Kleptsyna[2]

[1] University of Leeds, UK

[2] University of Le Mans, France

October, 20th, 2008 / Brest

Outline

Outline

Statement of the problem

The model

Nonobservable ergodic diffusion process (X_t) with

- values in $\mathbb{R}^d$;
- observations (Y_t) from $\mathbb{R}^\ell$;
- initial distribution μ_0 (of X_0) known with some error.

The question

Is this error forgotten by the optimal filtering algorithm in the long run?

A question for discussion

What does it mean "the optimal filtering algorithm" ?

Statement of the problem

The model

Nonobservable ergodic diffusion process (X_t) with

- values in $\mathbb{R}^d$;
- observations (Y_t) from $\mathbb{R}^\ell$;
- initial distribution μ_0 (of X_0) known with some error.

The question

Is this error forgotten by the optimal filtering algorithm in the long run?

A question for discussion

What does it mean "the optimal filtering algorithm" ?

Statement of the problem

The model

Nonobservable ergodic diffusion process (X_t) with

- values in $\mathbb{R}^d$;
- observations (Y_t) from $\mathbb{R}^\ell$;
- initial distribution μ_0 (of X_0) known with some error.

The question

Is this error forgotten by the optimal filtering algorithm in the long run?

A question for discussion

What does it mean "the optimal filtering algorithm" ?

The observation model
the precise definition

- Markov diffusion process:

$$dX_t = b(X_t)dt + dW_t, \quad (t \geq 0),$$

- observation:

$$dY_t = h(X_t)dt + dV_t \quad (t \geq 0),$$

- where
  - (W_t, V_t) is $\mathbb{R}^{d+\ell}$ valued Wiener process;
  - $b : \mathbb{R}^d \to \mathbb{R}^d$;
  - $h : \mathbb{R}^d \to \mathbb{R}^\ell$;

The observation model
the precise definition

- Markov diffusion process:

$$dX_t = b(X_t)dt + dW_t, \quad (t \geq 0),$$

- observation:

$$dY_t = h(X_t)dt + dV_t \quad (t \geq 0),$$

- where
  - (W_t, V_t) is $\mathbb{R}^{d+\ell}$ valued Wiener process;
  - $b : \mathbb{R}^d \to \mathbb{R}^d$;
  - $h : \mathbb{R}^d \to \mathbb{R}^\ell$;

Stating the main question

- The true conditional probability:

$$\mathbf{P}_t^{\mu_0,Y}(\cdot) = P_{\mu_0}(X_t \in \cdot \mid \mathcal{F}_t^Y),$$

  - with $\mathcal{F}_t^Y = \sigma(Y_s : 0 \le s \le t)$,
  - with the initial measure μ_0.

- The strange conditional probability:

$$\mathbf{P}_t^{\nu_0,Y}(\cdot) = \mathbf{P}_t^{\mu_0,Y}(\cdot) \mid \mu_0 = \nu_0.$$

  - with μ_0 replaced by ν_0.

The question for discussion:

Why $\mathbf{P}_t^{\nu_0,Y}(\cdot)$ is well defined?

Stating the main question

- The true conditional probability:

$$\mathbf{P}_t^{\mu_0, Y}(\cdot) = P_{\mu_0}(X_t \in \cdot \mid \mathcal{F}_t^Y),$$

  - with $\mathcal{F}_t^Y = \sigma(Y_s : 0 \le s \le t)$,
  - with the initial measure μ_0.

- The strange conditional probability:

$$\mathbf{P}_t^{\nu_0, Y}(\cdot) = \mathbf{P}_t^{\mu_0, Y}(\cdot) \mid \mu_0 = \nu_0.$$

  - with μ_0 replaced by ν_0.

The question for discussion:

Why $\mathbf{P}_t^{\nu_0, Y}(\cdot)$ is well defined?

Stating the main question

- The true conditional probability:

$$\mathbf{P}_t^{\mu_0,Y}(\cdot) = P_{\mu_0}(X_t \in \cdot \mid \mathcal{F}_t^Y),$$

  - with $\mathcal{F}_t^Y = \sigma(Y_s : 0 \le s \le t)$,
  - with the initial measure μ_0.
- The strange conditional probability:

$$\mathbf{P}_t^{\nu_0,Y}(\cdot) = \mathbf{P}_t^{\mu_0,Y}(\cdot) \mid \mu_0 = \nu_0.$$

  - with μ_0 replaced by ν_0.

The question for discussion:

Why $\mathbf{P}_t^{\nu_0,Y}(\cdot)$ is well defined?

The main question, formulation

The main question:

True or false:

$$\lim_{t\to\infty} E_{\mu_0} \|\mathbf{P}_t^{\mu_0,Y}(\cdot) - \mathbf{P}_t^{\nu_0,Y}(\cdot)\|_{TV} = 0?$$

Stability of filters

Stability of filters

True or false:

$$\lim_{t\to\infty} E_{\mu_0}(\pi_t^{\mu_0,Y}(f) - \pi_t^{\nu_0,Y}(f))^2 = 0? \quad \forall f \in C_b$$

where

- the true conditional expectation:

$$\pi_t^{\mu_0,Y}(f) = E_{\mu_0}(f(X_t) \mid \mathcal{F}_t^Y)$$

- the strange conditional expectation:

$$\pi_t^{\nu_0,Y}(f) = E_{\mu_0}(f(X_t) \mid \mathcal{F}_t^Y) \mid \mu_0 = \nu_0.$$

Stability of filters

Stability of filters

True or false:

$$\lim_{t\to\infty} E_{\mu_0}(\pi_t^{\mu_0,Y}(f) - \pi_t^{\nu_0,Y}(f))^2 = 0? \quad \forall f \in C_b$$

where

- the true conditional expectation:

$$\pi_t^{\mu_0,Y}(f) = E_{\mu_0}(f(X_t) \mid \mathcal{F}_t^Y)$$

- the strange conditional expectation:

$$\pi_t^{\nu_0,Y}(f) = E_{\mu_0}(f(X_t) \mid \mathcal{F}_t^Y) \mid \mu_0 = \nu_0.$$

Outline

The first time

D.Blackwell, 1957

The model

Nonobservable **stationary** ergodic **finite** state Markov chain (X_n)

- observations $Y_n = \Phi(X_n)$
- Φ is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

A question for discussion

What is the connection with the subject of the talk?

The first time

D.Blackwell, 1957

The model

Nonobservable **stationary** ergodic **finite** state Markov chain (X_n)

- observations $Y_n = \Phi(X_n)$
- Φ is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

A question for discussion

What is the connection with the subject of the talk?

The first time

D.Blackwell, 1957

The model

Nonobservable **stationary** ergodic **finite** state Markov chain (X_n)

- observations $Y_n = \Phi(X_n)$
- Φ is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

A question for discussion

What is the connection with the subject of the talk?

The first time

D.Blackwell, 1957

The model

Nonobservable **stationary** ergodic **finite** state Markov chain (X_n)

- observations $Y_n = \Phi(X_n)$
- Φ is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

A question for discussion

What is the connection with the subject of the talk?

Stability and uniqueness

Two related questions

- Blackwell: Is a stationary measure unique (only Q)?
- We: Is the filter stable?

Fact

Stability *of filter* $\Rightarrow$ ***uniqueness*** *of stationary measure. (A. Budhiraja, H.J.Kushner).*

A.Budhiraja (2008) - link between different properties of the nonlinear filter process:

- Stability of the filter with respect to initial conditions
- Uniqueness of the invariant measure of the filter
- "Finite memory" property of the filter

Stability and uniqueness

Two related questions

- Blackwell: Is a stationary measure unique (only Q)?
- We: Is the filter stable?

Fact

Stability *of filter* $\Rightarrow$ ***uniqueness*** *of stationary measure.* *(A. Budhiraja, H.J.Kushner).*

A.Budhiraja (2008) - link between different properties of the nonlinear filter process:

- Stability of the filter with respect to initial conditions
- Uniqueness of the invariant measure of the filter
- "Finite memory" property of the filter

The first time: **often** the answer is **"yes"**

- **1971, 1991, H. Kunita: "yes" in diffusion model.**
  - **Model:**
    Signal X_t — ergodic Markov process valued in a locally compact space.
  - **Observations:**

$$dY_t = h(X_t)dt + dW_t$$

  - **Claim:**

$$\lim_{t\to\infty} E_{\mu_0}(f(X_t) - \pi_t^{\mu_0,Y}(f))^2$$

    does non depend on μ_0,
    the invariant measure of the filtering process is unique.
- **L.Stettner, 1989, 1991:** generalization; discrete time included.

The first time: **often** the answer is **"yes"**

- **1971, 1991, H. Kunita: "yes" in diffusion model.**
  - **Model:**
    Signal X_t — ergodic Markov process valued in a locally compact space.
  - **Observations:**

$$dY_t = h(X_t)dt + dW_t$$

  - **Claim:**

$$\lim_{t\to\infty} E_{\mu_0}(f(X_t) - \pi_t^{\mu_0,Y}(f))^2$$

    does non depend on μ_0,
    the invariant measure of the filtering process is unique.
- **L.Stettner, 1989, 1991:** generalization; discrete time included.

The first time: **often** the answer is **"yes"**

- **1971, 1991, H. Kunita: "yes" in diffusion model.**
  - **Model:**
    Signal X_t — ergodic Markov process valued in a locally compact space.
  - **Observations:**

$$dY_t = h(X_t)dt + dW_t$$

  - **Claim:**

$$\lim_{t\to\infty} E_{\mu_0}(f(X_t) - \pi_t^{\mu_0,Y}(f))^2$$

    does non depend on μ_0,
    the invariant measure of the filtering process is unique.
- **L.Stettner, 1989, 1991:** generalization; discrete time included.

The first time: **often** the answer is **"yes"**

- **1971, 1991, H. Kunita: "yes" in diffusion model.**
  - **Model:**
    Signal X_t — ergodic Markov process valued in a locally compact space.
  - **Observations:**

    $$dY_t = h(X_t)dt + dW_t$$

  - **Claim:**

    $$\lim_{t\to\infty} E_{\mu_0}(f(X_t) - \pi_t^{\mu_0,Y}(f))^2$$

    does non depend on μ_0,
    the invariant measure of the filtering process is unique.
- **L.Stettner, 1989, 1991:** generalization; discrete time included.

The first time: **often** the answer is **"yes"**

- **1971, 1991, H. Kunita: "yes" in diffusion model.**
  - **Model:**
    Signal X_t — ergodic Markov process valued in a locally compact space.
  - **Observations:**

    $$dY_t = h(X_t)dt + dW_t$$

  - **Claim:**

    $$\lim_{t\to\infty} E_{\mu_0}(f(X_t) - \pi_t^{\mu_0,Y}(f))^2$$

    does non depend on μ_0,
    the invariant measure of the filtering process is unique.
- **L.Stettner, 1989, 1991:** generalization; discrete time included.

The first time, sometimes the answer is **"no"**

1974, Kaijser : a counter-example

- X_n - an **ergodic** Markov chain with $\mathbb{S} = \{1, 2, 3, 4\}$
- transition matrix

$$\Lambda = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

- observation (**noiseless**): $\mathbf{Y}_n = \mathbf{1}_{X_n=1} + \mathbf{1}_{X_n=3}$

Result: there is no uniqueness, no stability

$$\lim_{n\to\infty} E_{\mu_0}(\pi_n^{\mu_0,Y}(x) - \pi_n^{\nu_0,Y}(x))^2 \geq C(\mu_0,\nu_0) > 0.$$

The first time, sometimes the answer is **"no"**

1974, Kaijser : a counter-example

- X_n - an **ergodic** Markov chain with $\mathbb{S} = \{1,2,3,4\}$
- transition matrix

$$\Lambda = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

- observation (**noiseless**): $\mathbf{Y}_n = \mathbf{1}_{X_n=1} + \mathbf{1}_{X_n=3}$

Result: there is no uniqueness, no stability

$$\lim_{n\to\infty} E_{\mu_0}(\pi_n^{\mu_0,Y}(x) - \pi_n^{\nu_0,Y}(x))^2 \geq C(\mu_0,\nu_0) > 0.$$

At the same time, independently, I

- **1991, Delyon** & **Zeitouni** :
  - consider finite state space ergodic signal or linear case;
  - introduce the term "memory length" of filters;
  - propose a programme of analysis of exponential stability of filters **using Lyapounov exponents**.
- **1996, D.Ocone, E.Pardoux** :
  - consider Kunita's model.
  - **Claim**: The optimal filter is stable:

$$\lim_{n\to\infty} E_{\mu_0}(\pi_n^{\mu_0,Y}(f) - \pi_n^{\nu_0,Y}(f))^2 = 0 \quad \forall f \in C_b,\ \nu_0 \sim \mu_0.$$

(the proof is **crucially** based on the H. Kunita result)

At the same time, independently, I

- **1991, Delyon** & **Zeitouni** :
  - consider finite state space ergodic signal or linear case;
  - introduce the term "memory length" of filters;
  - propose a programme of analysis of exponential stability of filters **using Lyapounov exponents**.
- **1996, D.Ocone, E.Pardoux** :
  - consider Kunita's model.
  - **Claim**: The optimal filter is stable:

$$\lim_{n\to\infty} E_{\mu_0}(\pi_n^{\mu_0,Y}(f) - \pi_n^{\nu_0,Y}(f))^2 = 0 \quad \forall f \in C_b,\ \nu_0 \sim \mu_0.$$

(the proof is **crucially** based on the H. Kunita result)

At the same time, independently, I

- **1991, Delyon** & **Zeitouni** :
  - consider finite state space ergodic signal or linear case;
  - introduce the term "memory length" of filters;
  - propose a programme of analysis of exponential stability of filters **using Lyapounov exponents**.
- **1996, D.Ocone, E.Pardoux** :
  - consider Kunita's model.
  - **Claim**: The optimal filter is stable:

$$\lim_{n\to\infty} E_{\mu_0}(\pi_n^{\mu_0,Y}(f) - \pi_n^{\nu_0,Y}(f))^2 = 0 \quad \forall f \in C_b,\ \nu_0 \sim \mu_0.$$

(the proof is **crucially** based on the H. Kunita result)

At the same time, independently, II

- **1997, Atar & Zeitouni**
  - consider discrete and **continuous** time, **compact** valued Markov signal;
  - use **Birkhoff contraction** principle.
- **1998, Atar**
  - considers **continuous time, one dimensional non-compact** case, with **linear** observations and sufficiently small noise in observations.

At the same time, independently, II

- **1997, Atar & Zeitouni**
  - consider discrete and **continuous** time, **compact** valued Markov signal;
  - use **Birkhoff contraction** principle.
- **1998, Atar**
  - considers **continuous time, one dimensional non-compact** case, with **linear** observations and sufficiently small noise in observations.

A bit later

2004 P.Baxendale, P.Chiganskii, R.Liptser
Serious gap in Kunita's proof.

The Kunita's proof was based on the following:

True or false

$$\bigcap_{n\geq 1} \mathcal{F}^{Y}_{[0,\infty)} \bigvee \mathcal{F}^{X}_{[n,\infty)} = \mathcal{F}^{Y}_{[0,\infty)}$$

for an **ergodic** Markov process X_t?

A bit later

2004 P.Baxendale, P.Chiganskii, R.Liptser
Serious gap in Kunita's proof.
The Kunita's proof was based on the following:

True or false

$$\bigcap_{n\geq 1} \mathcal{F}^{Y}_{[0,\infty)} \bigvee \mathcal{F}^{X}_{[n,\infty)} = \mathcal{F}^{Y}_{[0,\infty)}$$

for an **ergodic** Markov process X_t?

Nothing is clear

Counterexample for the proof

an **ergodic** Markov process X_t with

- state space $\mathbb{S} = \{1, 2, 3, 4\}$;
- transition intensity matrix:

$$\Lambda = \frac{1}{2}\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix};$$

- **noiseless** observation: $\mathbf{Y}_n = \mathbf{1}_{X_n=1} + \mathbf{1}_{X_n=3}$.

Result: the answer is "False". Also, filter is unstable, the invariant measure of the filtering process is not unique.

Nothing is clear

Counterexample for the proof

an **ergodic** Markov process X_t with

- state space $\mathbb{S} = \{1, 2, 3, 4\}$;
- transition intensity matrix:

$$\Lambda = \frac{1}{2}\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix};$$

- **noiseless** observation: $\mathbf{Y}_n = \mathbf{1}_{X_n=1} + \mathbf{1}_{X_n=3}$.

Result: the answer is "False". Also, filter is unstable, the invariant measure of the filtering process is not unique.

Today

- Le Gland & Oudjane (2003) — discrete time, stability properties under uniform ergodicity conditions.
- Budhiraja & Ocone (1999), Oudjane & Rubenthaler (2005) — discrete time, small observation noise.
- Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
- Liptser, Chigansky (2005,2006,2007) – **continuous time compact case**, exponential stability via Lyapounov exponents.
- van Handel (2008)- Kunita's proof is revised

Today

- Le Gland & Oudjane (2003) — discrete time, stability properties under uniform ergodicity conditions.
- Budhiraja & Ocone (1999), Oudjane & Rubenthaler (2005) — discrete time, small observation noise.
- Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
- Liptser, Chigansky (2005,2006,2007) – **continuous time compact case**, exponential stability via Lyapounov exponents.
- van Handel (2008)- Kunita's proof is revised

Today

- Le Gland & Oudjane (2003) — discrete time, stability properties under uniform ergodicity conditions.
- Budhiraja & Ocone (1999), Oudjane & Rubenthaler (2005) — discrete time, small observation noise.
- Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
- Liptser, Chigansky (2005,2006,2007) – **continuous time compact case**, exponential stability via Lyapounov exponents.
- van Handel (2008)- Kunita's proof is revised

Today

- Le Gland & Oudjane (2003) — discrete time, stability properties under uniform ergodicity conditions.
- Budhiraja & Ocone (1999), Oudjane & Rubenthaler (2005) — discrete time, small observation noise.
- Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
- Liptser, Chigansky (2005,2006,2007) – **continuous time compact case**, exponential stability via Lyapounov exponents.
- van Handel (2008)- Kunita's proof is revised

Reminder

The model

- Markov diffusion process: $dX_t = b(X_t)dt + dW_t$,
- observation: $dY_t = h(X_t)dt + dV_t$.

The question

True or false:

$$\lim_{t\to\infty} E_{\mu_0}\|\mathbf{P}_t^{\mu_0,Y}(\cdot) - \mathbf{P}_t^{\nu_0,Y}(\cdot)\|_{TV} = 0?$$

Assumptions, I

(A0) b is locally bounded;

($A1_p$) : the signal is **recurrent** (Khasminskii-Veretennikov conditions):

$$p = 0 : \qquad \limsup_{|x|\to\infty} \left\langle b(x), \frac{x}{|x|} \right\rangle \le -r, \ r > 0$$

or

$$p = 1 : \qquad \lim_{|x|\to\infty} \langle b(x), x \rangle = -\infty.$$

Examples

$$(p = 0) : b(x) = -sign(x), \ b(x) = -x; \ \ldots$$

$$(p = 1) : \ b(x) = -\frac{\arctan(x)}{\sqrt{1 + |x|}}; \ \ldots$$

Assumptions, I

(A0) b is locally bounded;

$(A1_p)$: the signal is **recurrent** (Khasminskii-Veretennikov conditions):

$$p = 0: \qquad \limsup_{|x|\to\infty} \left\langle b(x), \frac{x}{|x|} \right\rangle \le -r, \; r > 0$$

or

$$p = 1: \qquad \lim_{|x|\to\infty} \langle b(x), x \rangle = -\infty.$$

Examples

$$(p = 0): b(x) = -sign(x),\; b(x) = -x; \; \dots$$

$$(p = 1): \; b(x) = -\frac{\arctan(x)}{\sqrt{1+|x|}}; \; \dots$$

Assumptions, I

(A0) b is locally bounded;

($A1_p$) : the signal is **recurrent** (Khasminskii-Veretennikov conditions):

$$p = 0 : \qquad \limsup_{|x|\to\infty} \left\langle b(x), \frac{x}{|x|} \right\rangle \le -r,\ r > 0$$

or

$$p = 1 : \qquad \lim_{|x|\to\infty} \langle b(x), x \rangle = -\infty.$$

Examples

$$(p = 0) : b(x) = -sign(x),\ b(x) = -x;\ \ldots$$

$$(p = 1) :\ b(x) = -\frac{\arctan(x)}{\sqrt{1 + |x|}};\ \ldots$$

Assumptions, II

(A2) The function h is smooth enough:

$$h \in C^2, \quad \& \quad \|\nabla h\|_{C^1} < \infty.$$

(A3) : Initial data is **absolutely continuous.**

$$\left\| \frac{d\mu_0}{d\nu_0} \right\|_{L_\infty(\nu_0)} < \infty.$$

(A4) Initial moments are finite:

$$\int e^{c|x|} \mu_0(dx) < \infty.$$

Assumptions, II

(A2) The function h is smooth enough:

$$h \in C^2, \quad \& \quad \|\nabla h\|_{C^1} < \infty.$$

(A3) : Initial data is **absolutely continuous.**

$$\left\| \frac{d\mu_0}{d\nu_0} \right\|_{L_\infty(\nu_0)} < \infty.$$

(A4) Initial moments are finite:

$$\int e^{c|x|} \mu_0(dx) < \infty.$$

Assumptions, II

(A2) The function h is smooth enough:

$$h \in C^2, \quad \& \quad \|\nabla h\|_{C^1} < \infty.$$

(A3) : Initial data is **absolutely continuous.**

$$\left\| \frac{d\mu_0}{d\nu_0} \right\|_{L_\infty(\nu_0)} < \infty.$$

(A4) Initial moments are finite:

$$\int e^{c|x|} \mu_0(dx) < \infty.$$

Stability with bounds – main result

Theorem

Under Assumptions (A0) – (A4) the following bounds hold:

$$E_{\mu_0}\|\boldsymbol{P}_t^{\mu_0,Y}(\cdot)-\boldsymbol{P}_t^{\nu_0,Y}(\cdot)\|_{TV} \le \begin{cases} C_m t^{-m}, & p=1, \quad \forall m>0, \\ C\exp(-ct), & p=0. \end{cases}$$

Outline

Conditional distribution, a particular case

A pair

Let the pair (X, Y) be the solution of:

$$\begin{cases} dX_s = b(Y, X_s)\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

with independent (W, B)

Its first component

and let X_s^ψ be s.t. (with **deterministic** ψ) :

$$dX_s^\psi = b(\psi, X_s^\psi)\, ds + dW_s$$

Then the **conditional** law $\mathcal{L}(X \mid Y)$ is just the **law** of X^ψ with a **substitution** $\psi = Y$.

Conditional distribution, a particular case

A pair

Let the pair (X, Y) be the solution of:

$$\begin{cases} dX_s = b(Y, X_s)\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

with independent (W, B)

Its first component

and let X_s^ψ be s.t. (with **deterministic** ψ) :

$$dX_s^\psi = b(\psi, X_s^\psi)\, ds + dW_s$$

Then the **conditional** law $\mathcal{L}(X \mid Y)$ is just the **law** of X^ψ with a **substitution** $\psi = Y$.

Conditional distribution, a particular case

A pair

Let the pair (X, Y) be the solution of:

$$\begin{cases} dX_s = b(Y, X_s)\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

with independent (W, B)

Its first component

and let X_s^ψ be s.t. (with **deterministic** ψ) :

$$dX_s^\psi = b(\psi, X_s^\psi)\, ds + dW_s$$

Then the **conditional** law $\mathcal{L}(X \mid Y)$ is just the **law** of X^ψ with a **substitution** $\psi = Y$.

Cauchy problem

Diffusion process

$$dX_t^{\psi} = b(\psi(t), X_t^{\psi})dt + dW_t, \quad (t \geq 0),$$

Then $E_x f(X_t^{\psi}) \exp[\int_0^t c(s, X_s^{\psi})ds] = u^{\psi}(0, x)$ is the solution of:

Cauchy problem

$$u_s + \Delta u/2 + b(\psi(s), x)\nabla u + c(s, x)u = 0, \quad u(t, x) = f(x)$$

Continuity properties of the solution w.r.t ψ are known.

Cauchy problem

Diffusion process

$$dX_t^\psi = b(\psi(t), X_t^\psi)dt + dW_t, \quad (t \geq 0),$$

Then $E_x f(X_t^\psi) \exp[\int_0^t c(s, X_s^\psi)ds] = u^\psi(0, x)$ is the solution of:

Cauchy problem

$$u_s + \Delta u/2 + b(\psi(s), x)\nabla u + c(s, x)u = 0, \quad u(t, x) = f(x)$$

Continuity properties of the solution w.r.t ψ are known.

Cauchy problem

Diffusion process

$$dX_t^{\psi} = b(\psi(t), X_t^{\psi})dt + dW_t, \quad (t \geq 0),$$

Then $E_x f(X_t^{\psi}) \exp[\int_0^t c(s, X_s^{\psi})ds] = u^{\psi}(0, x)$ is the solution of:

Cauchy problem

$$u_s + \Delta u/2 + b(\psi(s), x)\nabla u + c(s, x)u = 0, \quad u(t, x) = f(x)$$

Continuity properties of the solution w.r.t ψ are known.

The first boundary problem

A bounded domain

Let

$$D_0 := \{ \sup_{0\le s\le 1} |X_s^\psi| < R+1 \}.$$

Then $E_x\left(1(D_0)f(X_1^\psi)\right)\exp[\int_0^1 c(s,X_s^\psi)ds] = u^\psi(0,x)$ – solution of

The first boundary problem

$$u_s + \frac{1}{2}\Delta u + b(\psi,x)\nabla u + c(s,x)u = 0,$$

$$u^\psi(1,x) = f(x); \quad u^\psi(s,x) = 0,\ 0 < s < 1,\ |x| = R+1.$$

The first boundary problem

A bounded domain

Let

$$D_0 := \{ \sup_{0\le s\le 1} |X^\psi_s| < R+1 \}.$$

Then $E_x \left(1(D_0) f(X^\psi_1)\right) \exp[\int_0^1 c(s, X^\psi_s)ds] = u^\psi(0,x)$ – solution of

The first boundary problem

$$u_s + \frac{1}{2}\Delta u + b(\psi, x)\nabla u + c(s,x)u = 0,$$

$$u^\psi(1,x) = f(x); \quad u^\psi(s,x) = 0,\ 0 < s < 1,\ |x| = R+1.$$

Outline

Harnack's inequality

Let $u \geq 0$ be a solution of the first boundary problem with

- **uniformly bounded** coefficients,
- final condition in a cylinder $\{(t, x) : 0 < t < 1; |x| < \mathbf{R+1}, \}$.

Variant of Harnack's inequality: Krylov, Safonov, 1980

$$\sup_{|x|,|z| \leq \mathbf{R}} \frac{u(0, x)}{u(0, z)} \leq C_R$$

where C_R depends only on R and on the upper bounds of the coefficients.

Harnack's inequality

Let $u \geq 0$ be a solution of the first boundary problem with

- **uniformly bounded** coefficients,
- final condition in a cylinder $\{(t, x) : 0 < t < 1; |x| < \mathbf{R{+}1}, \}$.

Variant of Harnack's inequality: Krylov, Safonov, 1980

$$\sup_{|x|,|z| \leq \mathbf{R}} \frac{u(0, x)}{u(0, z)} \leq C_R$$

where C_R depends only on R and on the upper bounds of the coefficients.

Outline

Birkhoff metric

Definition

The Birkhoff distance between positive measures:

$$\rho(\mu,\nu) = \begin{cases} \ln\sup(d\mu/d\nu) + \ln\sup(d\nu/d\mu), & \text{if finite,} \\ +\infty, & \text{otherwise.} \end{cases}$$

Remark. It is a pseudo-distance, measuring the difference between directions.

Birkhoff metric

Definition

The Birkhoff distance between positive measures:

$$\rho(\mu, \nu) = \begin{cases} \ln \sup(d\mu/d\nu) + \ln \sup(d\nu/d\mu), & \text{if finite,} \\ +\infty, & \text{otherwise.} \end{cases}$$

Remark. It is a pseudo-distance, measuring the difference between directions.

Two important properties, I

Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)

- For normalized measures μ and ν:

$$\|\mu - \nu\|_{TV} \leq \rho(\mu, \nu)$$

- The converse statement does not hold.

Example

$$\begin{aligned} q_\mu(x) &= \mathbf{1}\left(x \in [-1/2, 1/2]\right) \\ q_\nu(x) &= \frac{1}{2} \cdot \mathbf{1}\left(|x| \in [\varepsilon, 1/2]\right) + C \cdot \mathbf{1}\left(x \in [-\varepsilon, \varepsilon]\right) \end{aligned}$$

Then $\|\mu - \nu\|_{TV} = 1 - 2\varepsilon$, $\rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$

Two important properties, I

Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)

- For normalized measures μ and ν:

$$\|\mu - \nu\|_{TV} \leq \rho(\mu, \nu)$$

- The converse statement does not hold.

Example

$$q_\mu(x) = \mathbf{1}(x \in [-1/2, 1/2])$$

$$q_\nu(x) = \frac{1}{2} \cdot \mathbf{1}(|x| \in [\varepsilon, 1/2]) + C \cdot \mathbf{1}(x \in [-\varepsilon, \varepsilon])$$

Then $\|\mu - \nu\|_{TV} = 1 - 2\varepsilon$, $\rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$

Two important properties, I

Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)

- For normalized measures μ and ν:

$$\|\mu - \nu\|_{TV} \le \rho(\mu, \nu)$$

- The converse statement does not hold.

Example

$$q_\mu(x) = \mathbf{1}(x \in [-1/2, 1/2])$$
$$q_\nu(x) = \frac{1}{2} \cdot \mathbf{1}(|x| \in [\varepsilon, 1/2]) + C \cdot \mathbf{1}(x \in [-\varepsilon, \varepsilon])$$

Then $\|\mu - \nu\|_{TV} = 1 - 2\varepsilon$, $\rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$

Two important properties, I

Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)

- For normalized measures μ and ν:

$$\|\mu - \nu\|_{TV} \le \rho(\mu, \nu)$$

- The converse statement does not hold.

Example

$$\begin{aligned} q_\mu(x) &= \mathbf{1}(x \in [-1/2, 1/2]) \\ q_\nu(x) &= \frac{1}{2} \cdot \mathbf{1}(|x| \in [\varepsilon, 1/2]) + C \cdot \mathbf{1}(x \in [-\varepsilon, \varepsilon]) \end{aligned}$$

Then $\|\mu - \nu\|_{TV} = 1 - 2\varepsilon,\ \rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$

Two important properties, II

Birkhoff contraction for nonnegative kernels:

Let $\mathbf{Q} : \mathcal{M}(R^d) \to \mathcal{M}(R^d)$ s.t.: $\mu\mathbf{Q}(dy) = \int\limits_{R^d} Q(x,dy)\mu(dx)$.

Contraction

$$\rho(\mu\mathbf{Q}, \nu\mathbf{Q}) \le \frac{C^2-1}{C^2+1}\rho(\mu,\nu), \text{ with}$$

- (Krasnosel'skii, Lifshits, Sobolev)

$$C = \sup_{x,z,y} \frac{q(x,y)}{q(z,y)}, \; Q(x,dy) = q(x,y)dy.$$

- (Le Gland, Oudjane)

$$C = \sup_{x,z,A} \frac{Q(x,A)}{Q(z,A)}.$$

Two important properties, II

Birkhoff contraction for nonnegative kernels:
Let $\mathbf{Q} : \mathcal{M}(R^d) \to \mathcal{M}(R^d)$ s.t.: $\mu\mathbf{Q}(dy) = \int\limits_{R^d} Q(x, dy)\mu(dx)$.

Contraction

$$\rho(\mu\mathbf{Q}, \nu\mathbf{Q}) \le \frac{C^2 - 1}{C^2 + 1}\rho(\mu, \nu), \text{ with}$$

- (Krasnosel'skii, Lifshits, Sobolev)

$$C = \sup_{x,z,y} \frac{q(x,y)}{q(z,y)}, \; Q(x, dy) = q(x,y)dy.$$

- (Le Gland, Oudjane)

$$C = \sup_{x,z,A} \frac{Q(x,A)}{Q(z,A)}.$$

Two important properties, II

Birkhoff contraction for nonnegative kernels:

Let $\mathbf{Q} : \mathcal{M}(R^d) \to \mathcal{M}(R^d)$ s.t.: $\mu\mathbf{Q}(dy) = \int\limits_{R^d} Q(x, dy)\mu(dx)$.

Contraction

$$\rho(\mu\mathbf{Q}, \nu\mathbf{Q}) \le \frac{C^2 - 1}{C^2 + 1}\rho(\mu, \nu), \text{ with}$$

- (Krasnosel'skii, Lifshits, Sobolev)

$$C = \sup_{x,z,y} \frac{q(x,y)}{q(z,y)}, \; Q(x, dy) = q(x,y)dy.$$

- (Le Gland, Oudjane)

$$C = \sup_{x,z,A} \frac{Q(x,A)}{Q(z,A)}.$$

Outline

Hitting time estimates (A. Veretennikov, 1987):

For $\hat{\tau} = \inf(t \geq 0 : |X_t| \leq R)$

$$\begin{cases} E_x \hat{\tau}^k \leq C_m(1 + |x|^m) & (\forall\ m > 2k; p = 1), \\ E_x \exp(\alpha\hat{\tau}) \leq C \exp(c|x|) & (p = 0). \end{cases}$$

Corollary

Let $\Lambda(X)_R := \sum_{k=0}^{n-1} \mathbf{1}(|X_k| \leq R)$.

Then ($\forall\ 0 < \varepsilon < 1$ *and for* R *large enough*)

$$E_{\mu_0} 1(\Lambda(X)_R < \varepsilon n) \leq \begin{cases} C_m n^{-m}, & (p = 1), \\ C \exp(-cn), & (p = 0) \end{cases}$$

Ergodic processes in $\mathbb{R}^d$, properties

Hitting time estimates (A. Veretennikov, 1987):

For $\hat{\tau} = \inf(t \geq 0 : \ |X_t| \leq R)$

$$\begin{cases} E_x \hat{\tau}^k \leq C_m(1 + |x|^m) & (\forall\ m > 2k; p = 1), \\ E_x \exp(\alpha\hat{\tau}) \leq C \exp(c|x|) & (p = 0). \end{cases}$$

Corollary

Let $\Lambda(X)_R := \sum_{k=0}^{n-1} \mathbf{1}(|X_k| \leq R)$.

Then ($\forall\, 0 < \varepsilon < 1$ *and for R large enough)*

$$E_{\mu_0} 1(\Lambda(X)_R \quad < \quad \varepsilon n) \quad \leq \quad \begin{cases} C_m n^{-m}, & (p = 1), \\ C \exp(-cn), & (p = 0) \end{cases}$$

Outline

The Bayes formula, part 1

The classical Bayes formula, I

$$P(X_t \in \cdot \mid \mathcal{F}_t^Y) = \frac{\widehat{E}(\mathbf{1}(X_t \in \cdot) L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y)}{\widehat{E} L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y)},$$

with

$$L_t(\overline{X}, \overline{Y}) = \frac{dP}{d\widehat{P}}(\overline{X}, \overline{Y})$$

The Bayes formula, II

Changing of measure

Using **Girsanov's transformations and integration by parts** we change the measure:

$$\frac{dP}{d\widehat{P}} = \exp[\sum_{k=1}^{[t]} h^*(X_k)(Y_k - Y_{k-1}) + h^*(X_t)(Y_t - Y_{[t]})$$

$$+\frac{1}{2}\int_0^t c(X_s, Y)\, ds],$$

with

$$c(s, x, Y) = \|(Y_s - Y_{[s]})^* \nabla h(x)\|^2 - 2(Y_s - Y_{[s]})^* \Delta h(x) - \|h\|^2(x).$$

Exact filtering algorithm – explanation

- The density has a special form:

$$L_n = \prod_{k=1}^{n} \exp[h^*(X_k)(Y_k - Y_{k-1}) + \frac{1}{2}\int_{k-1}^{k} c(X_s, Y)\, ds],$$

- The transformed process (X, Y) (w.r.t $\hat{P}$) is nice:

$$\begin{cases} dX_s = (b(X_s) - (Y_s - Y_{[s]})^* \nabla h(X_s))\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

with independent W and B.

- We are in the situation "Conditional distribution, particular case"
- Now we can choose the continuous (w.r.t Y) version of the conditional measure.
- Hence, we can use the first boundary problem.

Exact filtering algorithm – explanation

- The density has a special form:

$$L_n = \prod_{k=1}^{n} \exp[h^*(X_k)(Y_k - Y_{k-1}) + \frac{1}{2}\int_{k-1}^{k} c(X_s, Y)\, ds],$$

- The transformed process (X, Y) (w.r.t $\hat{P}$) is nice:

$$\begin{cases} dX_s = (b(X_s) - (Y_s - Y_{[s]})^* \nabla h(X_s))\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

  with independent W and B.

- We are in the situation "Conditional distribution, particular case"
- Now we can choose the continuous (w.r.t Y) version of the conditional measure.
- Hence, we can use the first boundary problem.

Exact filtering algorithm – explanation

- The density has a special form:

$$L_n = \prod_{k=1}^{n} \exp[h^*(X_k)(Y_k - Y_{k-1}) + \frac{1}{2}\int_{k-1}^{k} c(X_s, Y)\, ds],$$

- The transformed process (X, Y) (w.r.t $\hat{P}$) is nice:

$$\begin{cases} dX_s = (b(X_s) - (Y_s - Y_{[s]})^* \nabla h(X_s))\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

  with independent W and B.
- We are in the situation "Conditional distribution, particular case"
- Now we can choose the continuous (w.r.t Y) version of the conditional measure.
- Hence, we can use the first boundary problem.

Exact filtering algorithm – explanation

- The density has a special form:

$$L_n = \prod_{k=1}^{n} \exp[h^*(X_k)(Y_k - Y_{k-1}) + \frac{1}{2}\int_{k-1}^{k} c(X_s, Y)\, ds],$$

- The transformed process (X, Y) (w.r.t $\hat{P}$) is nice:

$$\begin{cases} dX_s = (b(X_s) - (Y_s - Y_{[s]})^* \nabla h(X_s))\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

with independent W and B.

- We are in the situation "Conditional distribution, particular case"
- Now we can choose the continuous (w.r.t Y) version of the conditional measure.
- Hence, we can use the first boundary problem.

Exact filtering algorithm – explanation

- The density has a special form:

$$L_n = \prod_{k=1}^{n} \exp[h^*(X_k)(Y_k - Y_{k-1}) + \frac{1}{2}\int_{k-1}^{k} c(X_s, Y)\, ds],$$

- The transformed process (X, Y) (w.r.t $\hat{P}$) is nice:

$$\begin{cases} dX_s = (b(X_s) - (Y_s - Y_{[s]})^* \nabla h(X_s))\, ds + dW_s, \\ dY_s = dB_s, \end{cases}$$

  with independent W and B.
- We are in the situation "Conditional distribution, particular case"
- Now we can choose the continuous (w.r.t Y) version of the conditional measure.
- Hence, we can use the first boundary problem.

Exact filtering algorithm

- Exact filtering algorithm via a nonlinear integral operator

$$\bar{\mu}_t(\cdot;\mu_0) = \mathbf{P}_t^{\mu_0,Y}(\cdot) =: \mu_0\mathbf{Q}_t^Y(\cdot)$$

- Its explicit form $\bar{\mu}(A;\mu_0) = c_t^{\mu_0}\int_{R^d} Q_t(x_0,A)\,d\mu_0(x_0)$, with
  - $$Q_t(x_0,A) = \widehat{E}_{x_0}(\mathbf{1}(X_t\in A)L_t(\overline{X},\overline{Y}) \mid \mathcal{F}_t^Y)$$

    $Q_t(x_0,A)$ can be found from the Cauchy problem.
  - $c_t^{\mu_0}$ - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).

Exact filtering algorithm

- Exact filtering algorithm via a nonlinear integral operator

$$\bar{\mu}_t(\cdot;\mu_0) = \mathbf{P}_t^{\mu_0,Y}(\cdot) =: \mu_0 \mathbf{Q}_t^Y(\cdot)$$

- Its explicit form $\bar{\mu}(A;\mu_0) = c_t^{\mu_0} \int\limits_{R^d} Q_t(x_0, A)\, d\mu_0(x_0)$, with
  - $$Q_t(x_0, A) = \widehat{E}_{x_0}(\mathbf{1}(X_t \in A) L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y)$$

    $Q_t(x_0, A)$ can be found from the Cauchy problem.
  - $c_t^{\mu_0}$ - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).

Exact filtering algorithm

- Exact filtering algorithm via a nonlinear integral operator

$$\bar{\mu}_t(\cdot;\mu_0) = \mathbf{P}_t^{\mu_0,Y}(\cdot) =: \mu_0 \mathbf{Q}_t^Y(\cdot)$$

- Its explicit form $\bar{\mu}(A;\mu_0) = c_t^{\mu_0} \int_{R^d} Q_t(x_0, A)\, d\mu_0(x_0)$, with
  - $$Q_t(x_0, A) = \widehat{E}_{x_0}(\mathbf{1}(X_t \in A) L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y)$$

    $Q_t(x_0, A)$ can be found from the Cauchy problem.
  - $c_t^{\mu_0}$ - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).

Exact filtering algorithm

- Exact filtering algorithm via a nonlinear integral operator

$$\bar{\mu}_t(\cdot;\mu_0) = \mathbf{P}_t^{\mu_0,Y}(\cdot) =: \mu_0 \mathbf{Q}_t^Y(\cdot)$$

- Its explicit form $\bar{\mu}(A;\mu_0) = c_t^{\mu_0} \int\limits_{R^d} Q_t(x_0, A)\, d\mu_0(x_0)$, with
  - $$Q_t(x_0, A) = \widehat{E}_{x_0}(\mathbf{1}(X_t \in A) L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y)$$

    $Q_t(x_0, A)$ can be found from the Cauchy problem.
  - $c_t^{\mu_0}$ - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).

Exact filtering algorithm

- Exact filtering algorithm via a nonlinear integral operator

$$\bar{\mu}_t(\cdot;\mu_0) = \mathbf{P}_t^{\mu_0,Y}(\cdot) =: \mu_0 \mathbf{Q}_t^Y(\cdot)$$

- Its explicit form $\bar{\mu}(A;\mu_0) = c_t^{\mu_0} \int\limits_{R^d} Q_t(x_0, A)\, d\mu_0(x_0)$, with
  -
  $$Q_t(x_0, A) = \widehat{E}_{x_0}(\mathbf{1}(X_t \in A) L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y)$$
  $Q_t(x_0, A)$ can be found from the Cauchy problem.
  - $c_t^{\mu_0}$ - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).

Main question - reformulation

Strange conditional probability via the same operator:

$$\mathbf{P}_t^{\nu_0,Y}(\cdot) =: \nu_0 \mathbf{Q}_t^Y(\cdot) = c_t^{\nu_0} \int_{R^d} Q_t(x_0, A)\, d\nu_0(x_0).$$

Main question - reformulation

True or false:

$$\lim_{t\to\infty} E_{\mu_0} \| \mu_0 \mathbf{Q}_t^Y(\cdot) - \nu_0 \mathbf{Q}_t^Y(\cdot) \|_{TV} = 0?$$

Main question - reformulation

Strange conditional probability via the same operator:

$$\mathbf{P}_t^{\nu_0,Y}(\cdot) =: \nu_0\mathbf{Q}_t^Y(\cdot) = c_t^{\nu_0}\int_{R^d} Q_t(x_0, A)\, d\nu_0(x_0).$$

Main question - reformulation

True or false:

$$\lim_{t\to\infty} E_{\mu_0}\|\mu_0\mathbf{Q}_t^Y(\cdot) - \nu_0\mathbf{Q}_t^Y(\cdot)\|_{TV} = 0?$$

Outline

Coupling, doubling the space

Consider **independent** couples (X, Y) and $(\tilde X, \tilde Y)$ with initial laws $\mathcal{L}(X_0) = \mu_0$, $\mathcal{L}(\tilde X_0) = \nu_0$.

Doubling the operators, I

- New operators on the space of measures on R^{2d}
  $\bar\mu_t(A \times B; (\mu_0, \nu_0)) =$
  $C_t^{\mu_0} C_t^{\nu_0} \int_{R^{2d}} Q_t(x_0, \tilde x_0; A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde x_0).$
- with

$$Q_t(x_0, \tilde x_0; A \times B) = \widehat{E}_{x_0, \tilde x_0}(\mathbf{1}(X_t \in A, \tilde X_t \in B) \times L_t(X, Y) L_t(\tilde X, \tilde Y) \mid \mathcal{F}_t^{Y, \tilde Y})\Big|_{\tilde Y = Y}$$

Coupling, doubling the space

Consider **independent** couples (X, Y) and $(\tilde X, \tilde Y)$ with initial laws $\mathcal{L}(X_0) = \mu_0$, $\mathcal{L}(\tilde X_0) = \nu_0$.

Doubling the operators, I

- New operators on the space of measures on R^{2d}
$\bar\mu_t(A \times B; (\mu_0, \nu_0)) =$
$c_t^{\mu_0} c_t^{\nu_0} \int\limits_{R^{2d}} Q_t(x_0, \tilde x_0;\ A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde x_0).$
- with

$$Q_t(x_0, \tilde x_0;\ A \times B) = \widehat{E}_{x_0, \tilde x_0}\big(\mathbf{1}(X_t \in A, \tilde X_t \in B) \times L_t(X, Y) L_t(\tilde X, \tilde Y) \mid \mathcal{F}_t^{Y, \tilde Y}\big)\Big|_{\tilde Y = Y}$$

Coupling, doubling the space

Consider **independent** couples (X, Y) and $(\tilde X, \tilde Y)$ with initial laws $\mathcal{L}(X_0) = \mu_0$, $\mathcal{L}(\tilde X_0) = \nu_0$.

Doubling the operators, I

- New operators on the space of measures on R^{2d}
  $\bar\mu_t(A \times B; (\mu_0, \nu_0)) =$
  $c_t^{\mu_0} c_t^{\nu_0} \int\limits_{R^{2d}} Q_t(x_0, \tilde x_0;\ A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde x_0).$
- with

$$Q_t(x_0, \tilde x_0;\ A \times B) = \widehat{E}_{x_0, \tilde x_0}\Big(\mathbf{1}(X_t \in A, \tilde X_t \in B) \times L_t(X, Y) L_t(\tilde X, \tilde Y) \mid \mathcal{F}_t^{Y, \tilde Y}\Big)\Big|_{\tilde Y = Y}$$

Doubling the operators, II

Remark. The substitutions are well defined.

Comparison of measures

The following properties hold:

- $\bar{\mu}_t(A;\mu_0) = \bar{\mu}_t(A\times R^d;(\mu_0,\nu_0))$
- $\bar{\mu}_t(A;\nu_0) = \bar{\mu}_t(A\times R^d;(\nu_0,\mu_0))$

Comparison of distances

$$\|\bar{\mu}_t(\cdot;\mu_0) - \bar{\mu}_t(\cdot;\nu_0)\|_{TV} \le \|\bar{\mu}_t(\cdot;(\mu_0,\nu_0)) - \bar{\mu}_t(\cdot;(\nu_0,\mu_0))\|_{TV}$$

Doubling the operators, II

Remark. The substitutions are well defined.

Comparison of measures

The following properties hold:

- $\bar{\mu}_t(A; \mu_0) = \bar{\mu}_t(A \times R^d; (\mu_0, \nu_0))$
- $\bar{\mu}_t(A; \nu_0) = \bar{\mu}_t(A \times R^d; (\nu_0, \mu_0))$

Comparison of distances

$$\|\bar{\mu}_t(\cdot; \mu_0) - \bar{\mu}_t(\cdot; \nu_0)\|_{TV} \le \|\bar{\mu}_t(\cdot; (\mu_0, \nu_0)) - \bar{\mu}_t(\cdot; (\nu_0, \mu_0))\|_{TV}$$

Doubling the operators, II

Remark. The substitutions are well defined.

Comparison of measures

The following properties hold:

- $\bar{\mu}_t(A; \mu_0) = \bar{\mu}_t(A \times R^d; (\mu_0, \nu_0))$
- $\bar{\mu}_t(A; \nu_0) = \bar{\mu}_t(A \times R^d; (\nu_0, \mu_0))$

Comparison of distances

$$\|\bar{\mu}_t(\cdot; \mu_0) - \bar{\mu}_t(\cdot; \nu_0)\|_{TV} \le \|\bar{\mu}_t(\cdot; (\mu_0, \nu_0)) - \bar{\mu}_t(\cdot; (\nu_0, \mu_0))|_{TV}$$

Separation

Partition of unity

For fixed R, n, and any non-random vector $\delta \in \Delta = \{0;1\}^{n+1}$ define

$$1_\delta(X,\tilde X) := \prod_{i=0}^{n-1} (\mathbf{1}(D_i))^{\delta_i} \times (1 - \mathbf{1}(D_i))^{1-\delta_i},$$

where

$$D_i := \Big\{ \max\left(|X_i|, |\tilde X_i|\right) \le R; \\ \max\left(\sup_{i\le s\le i+1}|X_s|, \sup_{i\le s\le i+1}|\tilde X_s|\right) < R+1 \Big\}$$

Separation

Partition of unity

For fixed R, n, and any non-random vector $\delta \in \Delta = \{0; 1\}^{n+1}$ define

$$1_\delta(X, \tilde X) := \prod_{i=0}^{n-1} (\mathbf{1}(D_i))^{\delta_i} \times (1 - \mathbf{1}(D_i))^{1-\delta_i},$$

where

$$D_i := \Big\{\max\left(|X_i|, |\tilde X_i|\right) \le R;$$

$$\max\left(\sup_{i\le s\le i+1} |X_s|, \sup_{i\le s\le i+1} |\tilde X_s|\right) < R+1\Big\}$$

Partition of unity, II

Multiplicative decomposition

$$1_\delta(X,\tilde X) := \prod_{i=0}^{n-1} \mathbf{1}_{\delta_i}(D_i)$$

with

$$\mathbf{1}_{\delta_i}(D_i) = 1(\delta_i = 1)\mathbf{1}(D_i) + 1(\delta_i = 0)(1 - \mathbf{1}(D_i)).$$

Partition of unity

$$\mathbf{1} = \sum_{\delta\in\Delta} 1_\delta(X,\tilde X)$$

Partition of unity, II

Multiplicative decomposition

$$1_\delta(X,\tilde X) := \prod_{i=0}^{n-1} \mathbf{1}_{\delta_i}(D_i)$$

with

$$\mathbf{1}_{\delta_i}(D_i) = 1(\delta_i = 1)\mathbf{1}(D_i) + 1(\delta_i = 0)(1 - \mathbf{1}(D_i)).$$

Partition of unity

$$\mathbf{1} = \sum_{\delta\in\Delta} 1_\delta(X,\tilde X)$$

Separation of pairs

Denote by $\#1(\delta)$ the total number of **ones** in δ and by

$$\#1(X)_R := \sum_{k=0}^{n-1} \mathbf{1}(|X_k| \le R, \sup_{k\le s\le k+1} |X_s| < R+1,)$$

The following inequalities hold:

Separation of pairs, I

$$\sum_{\delta:\,\#1(\delta)<\varepsilon n} \mathbf{1}_\delta(X,\tilde X) \le \mathbf{1}(\#1(X)_R < \frac{1+\varepsilon}{2} n) + \mathbf{1}(\#1(\tilde X)_R < \frac{1+\varepsilon}{2} n)$$

Separation of pairs

Denote by $\#1(\delta)$ the total number of **ones** in δ and by

$$\#1(X)_R := \sum_{k=0}^{n-1} \mathbf{1}(|X_k| \le R, \sup_{k\le s\le k+1} |X_s| < R+1,)$$

The following inequalities hold:

Separation of pairs, I

$$\sum_{\delta:\, \#1(\delta)<\varepsilon n} \mathbf{1}_\delta(X,\tilde{X}) \le \mathbf{1}(\#1(X)_R < \frac{1+\varepsilon}{2} n) + \mathbf{1}(\#1(\widetilde{X})_R < \frac{1+\varepsilon}{2} n)$$

Separation of pairs, II

Then ($\forall\, \varepsilon > \frac{1}{2}$ and for R large enough)

Separation of pairs,II

$$E_{\mu_0} 1((\#1(X)_R \;<\; \varepsilon n) \;\le\; \begin{cases} C_m n^{-m}, & (p=1), \\ C\exp(-cn), & (p=0) \end{cases}$$

The proof is based on the hitting time estimates, exponential Chebyshev's inequality and the fact that

$$q = \sup_{x:|x|\le R} P_x(\sup_{0\le s\le +1} |X_s| \ge R+1) < 1/2.$$

Separation of pairs, II

Then ($\forall\, \varepsilon > \frac{1}{2}$ and for R large enough)

Separation of pairs,II

$$E_{\mu_0} 1((\#1(X)_R \;<\; \varepsilon n) \;\le\; \begin{cases} C_m n^{-m}, & (p=1), \\ C\exp(-cn), & (p=0) \end{cases}$$

The proof is based on the hitting time estimates, exponential Chebyshev's inequality and the fact that

$$q = \sup_{x:|x|\le R} P_x(\sup_{0\le s\le +1} |X_s| \ge R+1) < 1/2.$$

Outline

An estimate to prove

Our goal is to prove the following inequality:

The main inequality

$$E_{\mu_0}\|\bar{\mu}_t(\cdot;\mu_0) - \bar{\mu}_t(\cdot;\nu_0)\|_{TV} \le C \sum_{\delta\in\Delta} \kappa_R^{\#1(\delta)} E_{\mu_0,\nu_0} e_{[t]}^{Y;\delta;\mu_0,\nu_0},$$

$$\kappa_R := \frac{C_R^2 - 1}{C_R^2 + 1} < 1,$$

$$C_R = \sup_{|x|,|\tilde{x}|,|z|,|\tilde{z}|\le R} \frac{u(0,x,\tilde{x})}{u(0,z,\tilde{z})}$$

with $u(s,x,\tilde{x})$ - the solution of the first boundary problem.

An estimate to prove

Our goal is to prove the following inequality:

The main inequality

$$E_{\mu_0}\|\bar\mu_t(\cdot;\mu_0)-\bar\mu_t(\cdot;\nu_0)\|_{TV}\le C\sum_{\delta\in\Delta}\kappa_R^{\#1(\delta)}E_{\mu_0,\nu_0}e_{[t]}^{Y;\delta;\mu_0,\nu_0},$$

$$\kappa_R := \frac{C_R^2-1}{C_R^2+1}<1,$$

$$C_R = \sup_{|x|,|\tilde x|,|z|,|\tilde z|\le R}\frac{u(0,x,\tilde x)}{u(0,z,\tilde z)}$$

with $u(s,x,\tilde x)$ - the solution of the first boundary problem.

First boundary problem, II

First boundary problem

$$u_s + \frac{1}{2} u_{xx} + \frac{1}{2} u_{\tilde{x}\tilde{x}}$$

$$+(b(x) - (\psi_s - \psi_0)^* \nabla h(x))u_x + (b(\tilde{x}) - (\psi_s - \psi_0)^* \nabla h(\tilde{x}))u_{\tilde{x}}$$

$$+\frac{1}{2} c(x, \tilde{x}, \psi) u = 0,$$

$$u(1, x, \tilde{x}) = (\mathbf{1}(x \in A, \tilde{x} \in B)$$

$$\times \exp[h^*(x)(\psi_1 - \psi_0) + h^*(\tilde{x})(\psi_1 - \psi_0)]$$

$$u(s, x, \tilde{x}) = 0, \ \forall\, 0 < s < 1, \ \max(|x|, |\tilde{x}| = R + 1),$$

with a replacement $\psi = Y$.

Separating events, 2

The term $e_t^{Y;\delta;\mu_0,\nu_0}$ in the main inequality is defined by:

Probability separator, definition

$$e_t^{Y;\delta;\mu_0,\nu_0} := E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y, \tilde{Y})\Big|_{\tilde{Y}=Y}.$$

Remark. This term will be the normalizing coefficient.

Separating events, 2

The term $e_t^{Y;\delta;\mu_0,\nu_0}$ in the main inequality is defined by:

Probability separator, definition

$$e_t^{Y;\delta;\mu_0,\nu_0} := E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y, \tilde{Y})\Big|_{\tilde{Y}=Y}.$$

Remark. This term will be the normalizing coefficient.

Theorem 1, sketch of the proof, 1

We split the sum in the main inequality ($\forall \varepsilon > 0$):

$$\sum_{\delta:\,\#1(\delta)\geq\varepsilon n} + \sum_{\delta:\,\#1(\delta)<\varepsilon n}$$

and we estimate both terms:

- $$\sum_{\delta:\,\#1(\delta)\geq\varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0} e_n^{Y;\delta;\mu_0,\nu_0} \leq \kappa_R^{\varepsilon n}$$

- $$\sum_{\delta:\,\#1(\delta)<\varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right)$$
$$\leq \sum_{\delta:\,\#1(\delta)<\varepsilon n} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right).$$

Theorem 1, sketch of the proof, 1

We split the sum in the main inequality ($\forall \varepsilon > 0$):

$$\sum_{\delta:\,\#1(\delta)\geq\varepsilon n} + \sum_{\delta:\,\#1(\delta)<\varepsilon n}$$

and we estimate both terms:

- $$\sum_{\delta:\,\#1(\delta)\geq\varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0} e_n^{Y;\delta;\mu_0,\nu_0} \leq \kappa_R^{\varepsilon n}$$

- $$\sum_{\delta:\,\#1(\delta)<\varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y,\tilde{Y})\Big|_{\tilde{Y}=Y}\right)$$

$$\leq \sum_{\delta:\,\#1(\delta)<\varepsilon n} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y,\tilde{Y})\Big|_{\tilde{Y}=Y}\right).$$

Theorem 1, sketch of the proof, 1

We split the sum in the main inequality ($\forall\varepsilon > 0$):

$$\sum_{\delta:\,\#1(\delta)\ge\varepsilon n} + \sum_{\delta:\,\#1(\delta)<\varepsilon n}$$

and we estimate both terms:

- $$\sum_{\delta:\,\#1(\delta)\ge\varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0} e_n^{Y;\delta;\mu_0,\nu_0} \le \kappa_R^{\varepsilon n}$$

- $$\sum_{\delta:\,\#1(\delta)<\varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right)$$
$$\le \sum_{\delta:\,\#1(\delta)<\varepsilon n} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right).$$

Theorem 1, sketch of the proof, 1

We split the sum in the main inequality ($\forall \varepsilon > 0$):

$$\sum_{\delta\colon \#1(\delta)\geq \varepsilon n} + \sum_{\delta\colon \#1(\delta)< \varepsilon n}$$

and we estimate both terms:

- $$\sum_{\delta\colon \#1(\delta)\geq \varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0} e_n^{Y;\delta;\mu_0,\nu_0} \leq \kappa_R^{\varepsilon n}$$

- $$\sum_{\delta\colon \#1(\delta)< \varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0} \left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y, \tilde Y)\Big|_{\tilde Y = Y}\right)$$

$$\leq \sum_{\delta\colon \#1(\delta)< \varepsilon n} E_{\mu_0} \left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y, \tilde Y)\Big|_{\tilde Y = Y}\right).$$

Theorem 1, sketch of the proof, 2

- We can finish the proof:

$$E_{\mu_0}\left(E_{\mu_0,\nu_0}\left(\sum_{\delta:\,\#1(\delta)<\varepsilon n} 1_\delta(X,\tilde X)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right)$$

$$\le E_{\mu_0}\left(E_{\mu_0}\left(1(\#1(X)_R<\frac{1+\varepsilon}{2}n)\mid Y\right)\right)$$

$$+E_{\mu_0}\left(E_{\nu_0}\left(1(\#1(\tilde X)_R<\frac{1+\varepsilon}{2}n)\mid \tilde Y\right)\Big|_{\tilde Y=Y}\right)$$

(because X does not depend on $\tilde Y$, nor $\tilde X$ depends on Y).

- the inequality "separation of pairs" has been used.

Theorem 1, sketch of the proof, 3

We estimate the first term

$$E_{\mu_0}\left(E_{\mu_0}\left(1(\#1(X)_R < \frac{1+\varepsilon}{2}n) \mid Y\right)\right)$$

$$= E_{\mu_0}\left(1(\#1(X)_R < \frac{1+\varepsilon}{2}n)\right).$$

we can use the "Separation of pairs, II".

Theorem 1, sketch of the proof, 3

We estimate the first term

$$E_{\mu_0}\left(E_{\mu_0}\left(1(\#1(X)_R < \frac{1+\varepsilon}{2} n) \mid Y\right)\right)$$

$$= E_{\mu_0}\left(1(\#1(X)_R < \frac{1+\varepsilon}{2} n)\right).$$

we can use the "Separation of pairs, II”.

Theorem 1, sketch of the proof, 4

Next, we estimate the other term, using the **absolute continuity of the initial measures**:

$$
\begin{aligned}
& E_{\mu_0}\left(E_{\nu_0}\left(1(\#1(\tilde X)_R < \frac{1+\varepsilon}{2}n) \mid \tilde Y\right)\Big|_{\tilde Y = Y}\right) \\
\le\ & C_2\, E_{\nu_0}\left(E_{\nu_0}\left(1(\#1(\tilde X)_R < \frac{1+\varepsilon}{2}n) \mid \tilde Y\right)\Big|_{\tilde Y = Y}\right) \\
=\ & C_2\, E_{\nu_0}\left(1(\#1(\tilde X)_R < \frac{1+\varepsilon}{2}n)\right),
\end{aligned}
$$

Again, the **Separation of pairs, II.**

Theorem 1, sketch of the proof, 4

Next, we estimate the other term, using the **absolute continuity of the initial measures**:

$$E_{\mu_0}\left(E_{\nu_0}\left(1(\#1(\tilde X)_R < \frac{1+\varepsilon}{2}n) \mid \tilde Y\right)\Big|_{\tilde Y = Y}\right)$$

$$\le C_2\, E_{\nu_0}\left(E_{\nu_0}\left(1(\#1(\tilde X)_R < \frac{1+\varepsilon}{2}n) \mid \tilde Y\right)\Big|_{\tilde Y = Y}\right)$$

$$= C_2\, E_{\nu_0}\left(1(\#1(\tilde X)_R < \frac{1+\varepsilon}{2}n)\right),$$

Again, the **Separation of pairs, II.**

Theorem 1, sketch of the proof, 4

Next, we estimate the other term, using the **absolute continuity of the initial measures**:

$$
\begin{aligned}
& E_{\mu_0}\left(E_{\nu_0}\left(1(\#1(\tilde{X})_R < \frac{1+\varepsilon}{2} n) \mid \tilde{Y}\right)\Big|_{\tilde{Y}=Y}\right) \\
& \leq C_2\, E_{\nu_0}\left(E_{\nu_0}\left(1(\#1(\tilde{X})_R < \frac{1+\varepsilon}{2} n) \mid \tilde{Y}\right)\Big|_{\tilde{Y}=Y}\right) \\
& = C_2\, E_{\nu_0}\left(1(\#1(\tilde{X})_R < \frac{1+\varepsilon}{2} n)\right),
\end{aligned}
$$

Again, the **Separation of pairs, II.**

Outline

How can we prove the main inequality?

Define new **linear** operators on the space of non-normalized measures on $\mathbb{R}^{2d}$

$$\mu_t(A \times B; (\mu_0, \nu_0)) = \int_{R^{2d}} Q_t(x_0, \tilde{x}_0;\, A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde{x}_0).$$

with the same kernel Q_t:

$$Q_t(x_0, \tilde{x}_0;\, A \times B) = \widehat{E}_{x_0, \tilde{x}_0}(\mathbf{1}(X_t \in A, \tilde{X}_t \in B) \times L_t(X, Y) L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}})\Big|_{\tilde{Y}=Y}$$

We have

$$\bar{\mu}_t(A \times B; (\mu_0, \nu_0)) = c_t^{\mu_0} c_t^{\nu_0} \mu_t(A \times B; (\mu_0, \nu_0))$$

Coupling method, part 1

New operators, 2

How can we prove the main inequality?
Define new **linear** operators on the space of non-normalized measures on $\mathbb{R}^{2d}$

$$\mu_t(A \times B; (\mu_0, \nu_0)) = \int_{R^{2d}} Q_t(x_0, \tilde{x}_0;\ A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde{x}_0).$$

with the same kernel Q_t:

$$Q_t(x_0, \tilde{x}_0;\ A \times B) = \widehat{E}_{x_0, \tilde{x}_0}(\mathbf{1}(X_t \in A, \tilde{X}_t \in B) \times L_t(X, Y) L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}})\Big|_{\tilde{Y}=Y}$$

We have

$$\bar{\mu}_t(A \times B; (\mu_0, \nu_0)) = c_t^{\mu_0} c_t^{\nu_0} \mu_t(A \times B; (\mu_0, \nu_0))$$

Coupling method, part 1

New operators, 2

How can we prove the main inequality?
Define new **linear** operators on the space of non-normalized measures on $\mathbb{R}^{2d}$

$$\mu_t(A \times B; (\mu_0, \nu_0)) = \int_{R^{2d}} Q_t(x_0, \tilde{x}_0;\ A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde{x}_0).$$

with the same kernel Q_t:

$$Q_t(x_0, \tilde{x}_0;\ A \times B) = \widehat{E}_{x_0, \tilde{x}_0}(\mathbf{1}(X_t \in A, \tilde{X}_t \in B) \times L_t(X, Y) L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}})\Big|_{\tilde{Y}=Y}$$

We have

$$\bar{\mu}_t(A \times B; (\mu_0, \nu_0)) = c_t^{\mu_0} c_t^{\nu_0} \mu_t(A \times B; (\mu_0, \nu_0))$$

Coupling method, part 1

New operators, 2

How can we prove the main inequality?
Define new **linear** operators on the space of non-normalized measures on $\mathbb{R}^{2d}$

$$\mu_t(A \times B; (\mu_0, \nu_0)) = \int_{R^{2d}} Q_t(x_0, \tilde{x}_0;\ A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde{x}_0).$$

with the same kernel Q_t:

$$Q_t(x_0, \tilde{x}_0;\ A \times B) = \widehat{E}_{x_0, \tilde{x}_0}(\mathbf{1}(X_t \in A, \tilde{X}_t \in B) \times L_t(X, Y) L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}})\Big|_{\tilde{Y}=Y}$$

We have

$$\bar{\mu}_t(A \times B; (\mu_0, \nu_0)) = c_t^{\mu_0} c_t^{\nu_0} \mu_t(A \times B; (\mu_0, \nu_0))$$

Separating events, 2

New operators, 3

Using the partition of unit we obtain the following decomposition:

$$\mu_t(A \times B; (\mu_0, \nu_0)) = \sum_{\delta \in \Delta} \mu_t^\delta(A \times B; (\mu_0, \nu_0))$$

with

$$\mu_t^\delta(A \times B; (\mu_0, \nu_0)) = \int_{R^{2d}} Q_t^\delta(x_0, \tilde{x}_0;\ A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde{x}_0).$$

and with the kernel Q_t^δ:

$$Q_t^\delta(x_0, \tilde{x}_0;\ A \times B) = \widehat{E}_{x_0, \tilde{x}_0}(1(X_t \in A, \tilde{X}_t \in B) 1_\delta(X, \tilde{X}) \times L_t(X, Y) L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}}) \Big|_{\tilde{Y} = Y}$$

Separating events, 2

New operators, 3

Using the partition of unit we obtain the following decomposition:

$$\mu_t(A \times B; (\mu_0, \nu_0)) = \sum_{\delta \in \Delta} \mu_t^{\delta}(A \times B; (\mu_0, \nu_0))$$

with

$$\mu_t^{\delta}(A \times B; (\mu_0, \nu_0)) = \int_{R^{2d}} Q_t^{\delta}(x_0, \tilde{x}_0;\, A \times B)\, d\mu_0(x_0)\, d\nu_0(\tilde{x}_0).$$

and with the kernel Q_t^{δ}:

$$Q_t^{\delta}(x_0, \tilde{x}_0;\, A \times B) = \widehat{E}_{x_0, \tilde{x}_0}(\mathbf{1}(X_t \in A, \tilde{X}_t \in B) 1_{\delta}(X, \tilde{X}) \times L_t(X, Y) L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}}) \Big|_{\tilde{Y} = Y}$$

Probability separator, again

We see that the normalizing coefficient is exactly the $e_t^{Y;\delta;\mu_0,\nu_0}$:

Probability separator, II

$$e_t^{Y;\delta;\mu_0,\nu_0} := E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y, \tilde{Y})\Big|_{\tilde{Y}=Y}$$

$$= c_t^{\mu_0} c_t^{\nu_0} \mu_t^{\delta}(R^{2d}; (\mu_0, \nu_0))$$

Recursion, I

Using the Markov property of X_t we find the recursion (with $Z_t = (X_t, \widetilde{X}_t)$):

$$\mu_n^{\delta_n}(dz_n) = \int_{R^{2d}} Q^{\delta_n}(z_{n-1}, dz_n) d\mu_{n-1}^{\delta_{n-1}}(z_{n-1}),$$

with

$$Q^{\delta_n}(z_{n-1}, D) = E_{z_{n-1}} \mathbf{1}(Z_n \in D) \mathbf{1}_{\delta_n}(D_n) \exp[\int_{n-1}^{n} c(s, Z_s, Y)\, ds]$$

with

$$D_n := \left(|Z_{n-1}| \le R, \sup_{n-1 \le s \le n} |Z_s| < R+1 \right),$$

Comparison of distance,2

We can estimate the total variation norm:

$$
\begin{aligned}
&\|\bar{\mu}_t(\cdot;(\mu_0,\nu_0)) - \bar{\mu}_t(\cdot;(\nu_0,\mu_0))\|_{TV} \\
&\quad \le c_t^{\mu_0} c_t^{\nu_0} \sum_{\delta\in\Delta} \|\mu_t^{\delta}(\mu_0,\nu_0) - \mu_t^{\delta}(\nu_0,\mu_0)\|_{TV} \\
&\quad = \sum_{\delta\in\Delta} e_n^{Y;\delta;\mu_0,\nu_0} \|\hat{\mu}_t^{\delta}(\mu_0,\nu_0) - \hat{\mu}_t^{\delta}(\nu_0,\mu_0)\|_{TV}
\end{aligned}
$$

with normalization

$$
\hat{\mu}_t^{\delta}(\nu_0,\mu_0) = \frac{\mu_t^{\delta}(\mu_0,\nu_0)}{\mu_t^{\delta}(R^{2d};\mu_0,\nu_0)}
$$

Comparison of distance,2

We can estimate the total variation norm:

$$\|\bar\mu_t(\cdot;(\mu_0,\nu_0)) - \bar\mu_t(\cdot;(\nu_0,\mu_0))\|_{TV}$$
$$\le c_t^{\mu_0} c_t^{\nu_0} \sum_{\delta\in\Delta} \|\mu_t^{\delta}(\mu_0,\nu_0) - \mu_t^{\delta}(\nu_0,\mu_0)\|_{TV}$$
$$= \sum_{\delta\in\Delta} e_n^{Y;\delta;\mu_0,\nu_0} \|\hat\mu_t^{\delta}(\mu_0,\nu_0) - \hat\mu_t^{\delta}(\nu_0,\mu_0)\|_{TV}$$

with normalization

$$\hat\mu_t^{\delta}(\nu_0,\mu_0) = \frac{\mu_t^{\delta}(\mu_0,\nu_0)}{\mu_t^{\delta}(R^{2d};\mu_0,\nu_0)}$$

Using the Birkhoff metric, 1

Using the properties of the Birkhoff metric we see that

Birkhoff metric, 1st property.

$$\|\hat\mu_t^{\delta}(\mu_0,\nu_0) - \hat\mu_t^{\delta}(\nu_0,\mu_0)\|_{TV} \le \rho(\hat\mu_t^{\delta}(\mu_0,\nu_0); \hat\mu_t^{\delta}(\nu_0,\mu_0)).$$

Using the Birkhoff metric, 2

and that

Birkhoff metric, 2nd property.

$$\rho(\hat{\mu}_n^{\delta}(\mu_0,\nu_0); \hat{\mu}_n^{\delta}(\nu_0,\mu_0))$$

$$\equiv \rho\left(\mu_n^{\delta}(\mu_0,\nu_0); \mu_n^{\delta}(\nu_0,\mu_0)\right)$$

$$\le \kappa_R^{\delta_n}\rho\left(\mu_{n-1}^{\delta}(\mu_0,\nu_0); \mu_{n-1}^{\delta}(\nu_0,\mu_0)\right)$$

$$\le C\kappa_R^{k},$$

with

$$k = \#1(\delta)$$

which gives the desired inequality. QED