Ergodic Filters

Veretennikov, Kleptsyna

Introduction
Problem statement
Historical survey

Assumptions and main result

Auxiliaries
Parabolic equations and diffusion processes
Harnack inequality
Birkhoff metric
Ergodic processes
The Bayes approach

Sketch of the proof
Coupling and separation
The main inequality
Sketch of the proof, part 2

Stability for Nonlinear Filtering
Continuous Time Noncompact Case

Alexander Veretennikov¹  Marina Kleptsyna²

¹University of Leeds, UK
²University of Le Mans, France

October, 20th, 2008 / Brest
Outline

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and main result

3. Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in $\mathbb{R}^d$
   - Reformulation of the problem, the Bayes approach

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
Outline

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and main result

3. Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in $\mathbb{R}^d$
   - Reformulation of the problem, the Bayes approach

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
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 \(\mu_0\) (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 \(\mu_0\) (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 \(\mu_0\) (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:
  \[ P_{t, Y}^{\mu_0} (\cdot) = P_{\mu_0}(X_t \in \cdot \mid \mathcal{F}_t^Y), \]
  - with \( \mathcal{F}_t^Y = \sigma(Y_s : 0 \leq s \leq t) \),
  - with the initial measure \( \mu_0 \).

- The **strange** conditional probability:
  \[ P_{t, Y}^{\nu_0} (\cdot) = P_{t}^{\mu_0, Y} (\cdot) \mid \mu_0 = \nu_0. \]
  - with \( \mu_0 \) replaced by \( \nu_0 \).

The question for discussion:

Why \( P_{t}^{\nu_0, Y} (\cdot) \) is well defined?
Stating the main question

- The true conditional probability:

\[ P_{t}^{\mu_0, Y}(\cdot) = P_{\mu_0}(X_t \in \cdot | \mathcal{F}_t^Y), \]

- with \( \mathcal{F}_t^Y = \sigma(Y_s : 0 \leq s \leq t) \),
- with the initial measure \( \mu_0 \).

- The strange conditional probability:

\[ P_{t}^{\nu_0, Y}(\cdot) = P_{t}^{\mu_0, Y}(\cdot) | \mu_0 = \nu_0. \]

- with \( \mu_0 \) replaced by \( \nu_0 \).

The question for discussion:

Why \( P_{t}^{\nu_0, Y}(\cdot) \) is well defined?
Stating the main question

- The true conditional probability:
  \[ P_{t}^{\mu_0, Y}(\cdot) = P_{\mu_0}(X_t \in \cdot \mid \mathcal{F}_t^Y), \]
  - with \( \mathcal{F}_t^Y = \sigma(Y_s : 0 \leq s \leq t) \),
  - with the initial measure \( \mu_0 \).

- The strange conditional probability:
  \[ P_{t}^{\nu_0, Y}(\cdot) = P_{t}^{\mu_0, Y}(\cdot) \mid \mu_0 = \nu_0. \]
  - with \( \mu_0 \) replaced by \( \nu_0 \).

The question for discussion:
Why \( P_{t}^{\nu_0, Y}(\cdot) \) is well defined?
The main question, formulation

The main question:

True or false:

$$\lim_{t \to \infty} E_{\mu_0} \| P_{t}^{\mu_0, Y}(\cdot) - P_{t}^{\nu_0, Y}(\cdot) \|_{TV} = 0?$$
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:
  \[
  \pi_t^{\mu_0,Y}(f) = E_{\mu_0}(f(X_t) | \mathcal{F}^Y_t)
  \]

- the strange conditional expectation:
  \[
  \pi_t^{\nu_0,Y}(f) = E_{\mu_0}(f(X_t) | \mathcal{F}^Y_t) | \mu_0 = \nu_0.
  \]
Stability of filters

True or false:

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

where

- the **true** conditional expectation:
  \[ \pi_{t, Y}^{\mu_0}(f) = E_{\mu_0}(f(X_t) \mid \mathcal{F}_t^Y) \]

- the **strange** conditional expectation:
  \[ \pi_{t, Y}^{\nu_0}(f) = E_{\mu_0}(f(X_t) \mid \mathcal{F}_t^Y) \mid \mu_0 = \nu_0. \]
Outline

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and main result

3. Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in $\mathbb{R}^d$
   - Reformulation of the problem, the Bayes approach

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
D. Blackwell, 1957

**The model**

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

- observations \(Y_n = \Phi(X_n)\)
- \(\Phi\) 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?
D. Blackwell, 1957

The model

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

- observations \(Y_n = \Phi(X_n)\)
- \(\Phi\) 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?
D. Blackwell, 1957

The model

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

- observations \(Y_n = \Phi(X_n)\)
- \(\Phi\) 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 model

Nonobservable stationary ergodic finite state Markov chain \((X_n)\)
- observations \(Y_n = \Phi(X_n)\)
- \(\Phi\) 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

\textbf{Stability of filter} $\Rightarrow$ \textbf{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

- **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 $\mu_0$,
  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 $\mu_0$,
    the invariant measure of the filtering process is unique.

- **L. Stettner, 1989, 1991:** generalization; discrete time included.

Model:
Signal $X_t$ — ergodic Markov process valued in a locally compact space.

Observations:

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

Claim:

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

does not depend on $\mu_0$, the invariant measure of the filtering process is unique.

The first time: often the answer is "yes"

  - Model:
    Signal $X_t$ — ergodic Markov process valued in a locally compact space.
  - Observations:
    $dY_t = h(X_t)dt + dW_t$
  - Claim:
    $\lim_{t \to \infty} E_{\mu_0}(f(X_t) - \pi_{t, Y}(f))^2$
    does non depend on $\mu_0$, the invariant measure of the filtering process is unique.

The first time: **often** the answer is "**yes**"
The first time, sometimes the answer is "no".

1974, Kaijser: a counter-example

- $X_n$ - an ergodic Markov chain with $\mathcal{S} = \{1, 2, 3, 4\}$
- 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): $Y_n = 1_{X_n=1} + 1_{X_n=3}$

Result: there is no uniqueness, no stability

$$\lim_{n \to \infty} E_{\mu_0}(\pi^{\mu_0}_n, Y(x) - \pi^{\nu_0}_n, Y(x))^2 \geq C(\mu_0, \nu_0) > 0.$$
1974, Kajisier: a counter-example

- $X_n$ - an **ergodic** Markov chain with $\mathcal{S} = \{1, 2, 3, 4\}$
- 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**): $Y_n = 1_{X_n=1} + 1_{X_n=3}$

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 Lyapunov exponents.

- **1996, D.Ocone, E.Pardoux**:
  - consider Kunita’s model.
  - **Claim**: The optimal filter is stable:

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

(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:
    \[
    \lim_{n \to \infty} E_{\mu_0} (\pi^{\mu_0}_n, Y(f) - \pi^{\nu_0}_n, Y(f))^2 = 0 \quad \forall f \in C_b, \nu_0 \sim \mu_0.
    \]

  (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.
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

\[ \bigcap_{n \geq 1} \mathcal{F}^Y_{[0,\infty)} \bigvee_{n=0} \mathcal{F}^X_{[n,\infty)} = \mathcal{F}^Y_{[0,\infty)} \]

for an ergodic Markov process \( X_t \)?
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 \textbf{ergodic} Markov process \( X_t \)?
Counterexample for the proof

an **ergodic** Markov process $X_t$ with

- state space $S = \{1, 2, 3, 4\}$;
- 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: $Y_n = 1_{X_n=1} + 1_{X_n=3}$.

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

an ergodic Markov process $X_t$ with
- state space $\mathcal{S} = \{1, 2, 3, 4\}$;
- 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: $Y_n = 1_{X_n=1} + 1_{X_n=3}$.

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


Stannat (2005) Continuous time case, a gradient type drift and linear observation part under additional assumptions.


van Handel (2008) - Kunita’s proof is revised
Today

- Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
- van Handel (2008)- Kunita’s proof is revised
Today

- Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
- van Handel (2008)- Kunita’s proof is revised.
Today

- Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
- 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:

\[
\lim_{t \to \infty} E_{\mu_0} \| P_{t}^{\mu_0, Y}(\cdot) - P_{t}^{\nu_0, Y}(\cdot) \|_{TV} = 0?
\]
Assumptions, I

(A0) \( b \) is locally bounded;

(A1\(p \)) : the signal is **recurrent**

(Khasminskii-Veretennikov conditions):

\[
p = 0 : \quad \limsup_{|x| \to \infty} \left\langle b(x), \frac{x}{|x|} \right\rangle \leq -r, \ r > 0
\]

or

\[
p = 1 : \quad \lim_{|x| \to \infty} \left\langle b(x), x \right\rangle = -\infty.
\]

Examples

\((p = 0) : b(x) = -\text{sign}(x), \ b(x) = -x; \ldots\)

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

(A0) \( b \) is locally bounded;

(A1_p) : the signal is \textbf{recurrent}

(Khasminskii-Veretennikov conditions):

\[
p = 0 : \quad \limsup_{|x| \to \infty} \left\langle b(x), \frac{x}{|x|} \right\rangle \leq -r, \quad r > 0
\]

or

\[
p = 1 : \quad \lim_{|x| \to \infty} \left\langle b(x), x \right\rangle = -\infty.
\]

\[\](p = 0) : b(x) = -\text{sign}(x), \quad b(x) = -x; \ldots\]

\[\](p = 1) : b(x) = -\frac{\arctan(x)}{\sqrt{1 + |x|^2}}; \ldots\]
Assumptions, I

\((A0)\) \(b\) is locally bounded;
\((A1_p)\) \(p\) : the signal is \textbf{recurrent}
(Khasminskii-Veretennikov conditions):

\[
p = 0 : \quad \limsup_{|x| \to \infty} \left\langle b(x), \frac{x}{|x|} \right\rangle \leq -r, \quad r > 0
\]

or

\[
p = 1 : \quad \lim_{|x| \to \infty} \left\langle b(x), x \right\rangle = -\infty.
\]

Examples

\((p = 0)\) : \(b(x) = -\text{sign}(x), \quad b(x) = -x; \ldots\)

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

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

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and main result

3. Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in $\mathbb{R}^d$
   - Reformulation of the problem, the Bayes approach

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
Conditional distribution, a particular case

A pair

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

\[
\begin{align*}
\frac{dX_s}{ds} &= b(Y, X_s) \, ds + dW_s, \\
\frac{dY_s}{ds} &= dB_s,
\end{align*}
\]

with independent \((W, B)\)

Its first component

and let \(X_s^\psi\) be s.t. (with deterministic \(\psi\)) :

\[
\frac{dX_s^\psi}{ds} = b(\psi, X_s^\psi) \, ds + dW_s
\]

Then the conditional law \(\mathcal{L}(X \mid Y)\) is just the law of \(X_s^\psi\) 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 \(X^\psi_s\) be s.t. (with deterministic \(\psi\)) :

\[
dX^\psi_s = b(\psi, X^\psi_s) \, ds + dW_s
\]

Then the conditional law \(\mathcal{L}(X \mid Y)\) is just the law of \(X^\psi\) with a substitution \(\psi = Y\).
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 \(X_s^\psi\) be s.t. (with \textit{deterministic} \(\psi\)):

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

Then the \textbf{conditional} law \(\mathcal{L}(X \mid Y)\) is just the \textbf{law} of \(X_s^\psi\) with a \textbf{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 \( \psi \) 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 \( \psi \) 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 \( \psi \) are known.
A bounded domain

Let

\[ D_0 := \{\sup_{0 \leq s \leq 1} |X_s^\psi| < R + 1\}. \]

Then

\[
E_x \left( 1(D_0) f(X_1^\psi) \right) \exp [\int_0^1 c(s, X_s^\psi) ds] = u(0, x) - \text{solution of}
\]

The first boundary problem

\[
\begin{align*}
& u_s + \frac{1}{2} \Delta u + b(\psi, x) \nabla u + c(s, x) u = 0, \\
& u^\psi(1, x) = f(x); \quad u^\psi(s, x) = 0, \quad 0 < s < 1, \ |x| = R + 1.
\end{align*}
\]
The first boundary problem

A bounded domain

Let

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

Then $E_x \left( 1(D_0)f(X_1^\psi) \right) \exp[\int_0^1 c(s, X_s^\psi)ds] = u^\psi(0, x) - \text{solution of the first boundary problem}$

The first boundary problem

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

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

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

- **uniformly bounded** coefficients,
- final condition in a cylinder

$\{(t, x) : 0 < t < 1; |x| < R+1, \}$.

**Variant of Harnack’s inequality: Krylov, Safonov, 1980**

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

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
  \[ \{(t, x) : 0 < t < 1; |x| < R+1, \} \].

**Variant of Harnack’s inequality: Krylov, Safonov, 1980**

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

where \( C_R \) depends only on \( R \) and on the upper bounds of the coefficients.
Outline

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and main result

3. Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in \( \mathbb{R}^d \)
   - Reformulation of the problem, the Bayes approach

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
Definition

The Birkhoff distance between positive measures:

\[ \rho(\mu, \nu) = \begin{cases} 
\ln \sup \left( \frac{d\mu}{d\nu} \right) + \ln \sup \left( \frac{d\nu}{d\mu} \right), & \text{if finite,} \\
+\infty, & \text{otherwise.} 
\end{cases} \]

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

The Birkhoff distance between positive measures:

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

\textbf{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 $\mu$ and $\nu$:
  \[
  \|\mu - \nu\|_{TV} \leq \rho(\mu, \nu)
  \]

- The converse statement does not hold.

Example

\[
q_\mu(x) = \begin{cases} 
1 & (x \in [-1/2, 1/2]) \\
0 & \text{otherwise}
\end{cases}
\]

\[
q_\nu(x) = \frac{1}{2} \cdot \begin{cases} 
1 & (|x| \in [\varepsilon, 1/2]) \\
0 & \text{otherwise}
\end{cases} + C \cdot \begin{cases} 
1 & (x \in [-\varepsilon, \varepsilon]) \\
0 & \text{otherwise}
\end{cases}
\]

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 $\mu$ and $\nu$:

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

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 $\mu$ and $\nu$:

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

- The converse statement does not hold.

Example

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

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

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 $\mu$ and $\nu$:

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

- The converse statement does not hold.

Example

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

Then $\|\mu - \nu\|_{\text{TV}} = 1 - 2\varepsilon$, $\rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$
Birkhoff contraction for nonnegative kernels: Let $Q : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ s.t.: $\mu Q(dy) = \int_{\mathbb{R}^d} Q(x, dy) \mu(dx)$.

Contraction

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

- (Krasnosel'skii, Lifshits, Sobolev)

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

- (Le Gland, Oudjane)

$$C = \sup_{x,z,A} \frac{Q(x, A)}{Q(z, A)}.$$
Birkhoff contraction for nonnegative kernels:
Let $Q : \mathcal{M}(R^d) \rightarrow \mathcal{M}(R^d)$ s.t.: $\mu Q(dy) = \int_{R^d} Q(x, dy) \mu(dx)$.

**Contraction**

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

- (Krasnosel’skii, Lifshits, Sobolev)

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

- (Le Gland, Oudjane)

$$C = \sup_{x,z,A} \frac{Q(x, A)}{Q(z, A)}.$$
Birkhoff contraction for nonnegative kernels:
Let $Q : \mathcal{M}(\mathbb{R}^d) \rightarrow \mathcal{M}(\mathbb{R}^d)$ s.t.: $\mu Q(dy) = \int_{\mathbb{R}^d} Q(x, dy) \mu(dx)$.

\[ \rho(\mu Q, \nu Q) \leq \frac{C^2 - 1}{C^2 + 1} \rho(\mu, \nu), \quad \text{with} \]

- (Krasnosel’skii, Lifshits, Sobolev)
  \[ C = \sup_{x, z, y} \frac{q(x, y)}{q(z, y)}, \quad Q(x, dy) = q(x, y)dy. \]

- (Le Gland, Oudjane)
  \[ C = \sup_{x, z, A} \frac{Q(x, A)}{Q(z, A)}. \]
Ergodic Filters
Veretennikov, Kleptsyna

Introduction
Problem statement
Historical survey

Assumptions and main result

Auxiliaries
Parabolic equations and diffusion processes
Harnack inequality
Birkhoff metric
Ergodic processes in $\mathbb{R}^d$
Reformulation of the problem, the Bayes approach

Sketch of the proof
Coupling and separation
The main inequality
Sketch of the proof, part 2
Hitting time estimates (A. Veretennikov, 1987):

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

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

Corollary

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

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)

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

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and main result

3. Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in $\mathbb{R}^d$
   - Reformulation of the problem, the Bayes approach

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
The Bayes formula, part 1

The classical Bayes formula, 1

\[ P(\mathcal{F}^Y_t \in \cdot) = \frac{\hat{E}(\mathcal{F}^Y_t \in \cdot) L_t(\mathcal{X}, \mathcal{Y})}{\hat{E}L_t(\mathcal{X}, \mathcal{Y}) \mid \mathcal{F}^Y_t} \]

with

\[ L_t(\mathcal{X}, \mathcal{Y}) = \frac{d\mu}{d\hat{\mu}}(\mathcal{X}, \mathcal{Y}) \]
The Bayes formula, II

Changing of measure

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

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

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

with

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

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

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:

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

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:

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

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.
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 \((X, Y) (\text{w.r.t } \hat{P})\) is nice:

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

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

\[
\bar{\mu}_t(\cdot; \mu_0) = P_{\mu_0}^{\mu_0, Y}(\cdot) =: \mu_0 Q_{\mu}^{Y}(\cdot)
\]

- Its explicit form \(\bar{\mu}(A; \mu_0) = c_{\mu_0}^\mu \int_{R^d} Q_t(x_0, A) \, d\mu_0(x_0),\) with

\[
Q_t(x_0, A) = \widehat{E}_{x_0}(1(X_t \in A)L_t(\bar{X}, \bar{Y}) \mid \mathcal{F}_t^Y)
\]

\(Q_t(x_0, A)\) can be found from the Cauchy problem.

- \(c_{\mu_0}^\mu\) - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).
Exact filtering algorithm via a nonlinear integral operator

\[ \overline{\mu}_t(\cdot; \mu_0) = P_{t}^{\mu_0, Y}(\cdot) =: \mu_0 Q_t^Y(\cdot) \]

Its explicit form \( \overline{\mu}(A; \mu_0) = c_t^{\mu_0} \int_{\mathbb{R}^d} Q_t(x_0, A) \, d\mu_0(x_0), \) with

\[ Q_t(x_0, A) = \widehat{E}_{x_0}(1(X_t \in A)L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y) \]

\( Q_t(x_0, A) \) 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

\[ \bar{\mu}_t(\cdot; \mu_0) = P_t^{\mu_0,Y}(\cdot) =: \mu_0 Q_t^Y(\cdot) \]

- 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

\[ Q_t(x_0, A) = \hat{E}_{x_0}(1(X_t \in A) L_t(\bar{X}, \bar{Y}) | \mathcal{F}_t^Y) \]

\( Q_t(x_0, A) \) 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

\[ \bar{\mu}_t(\cdot; \mu_0) = P_t^{\mu_0, Y}(\cdot) =: \mu_0 Q_t^{Y}(\cdot) \]

- 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

\[ Q_t(x_0, A) = \hat{E}_{x_0}(1(X_t \in A)L_t(\bar{X}, \bar{Y}) \mid \mathcal{F}_t^{Y}) \]

\( Q_t(x_0, A) \) 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

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

- Its explicit form $$\bar{\mu}(A; \mu_0) = c^{\mu_0}_t \int_{R^d} Q_t(x_0, A) \, d\mu_0(x_0),$$ with

  $$Q_t(x_0, A) = \hat{E}_{x_0}(\mathbf{1}(X_t \in A) L_t(\overline{X}, \overline{Y}) \mid \mathcal{F}_t^Y)$$

  $$Q_t(x_0, A)$$ can be found from the Cauchy problem.

- $$c^{\mu_0}_t$$ - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).
Strange conditional probability via the same operator:

\[
P_t^{\nu_0, Y}(\cdot) =: \nu_0 Q_t^Y(\cdot) = c_t^{\nu_0} \int_{\mathbb{R}^d} Q_t(x_0, A) \, d\nu_0(x_0).
\]

True or false:

\[
\lim_{t \to \infty} E_{\mu_0} \|\mu_0 Q_t^Y(\cdot) - \nu_0 Q_t^Y(\cdot)\|_{TV} = 0?
\]
Strange conditional probability via the same operator:

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

Main question - reformulation

True or false:

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

1 Introduction
   - Problem statement
   - Historical survey

2 Assumptions and main result

3 Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in $\mathbb{R}^d$
   - Reformulation of the problem, the Bayes approach

4 Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
Using coupling method, I
A. Veretennikov, Lecture Notes, 2004

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 \(\mathbb{R}^{2d}\)

\[
\bar{\mu}_t(A \times B; (\mu_0, \nu_0)) = c_t^{\mu_0} c_t^{\nu_0} \int_{\mathbb{R}^{2d}} Q_t(x_0, \tilde{x}_0; A \times B) \, d\mu_0(x_0) \, d\nu_0(\tilde{x}_0).
\]

- with

\[
Q_t(x_0, \tilde{x}_0; A \times B) = \tilde{E}_{x_0, \tilde{x}_0} (1(X_t \in A, \tilde{X}_t \in B) \times L_t(X, Y) L_t(\tilde{X}, \tilde{Y}) | \mathcal{F}_t^{Y, \tilde{Y}}) \bigg|_{\tilde{Y} = Y},
\]
Using coupling method, I
A. Veretennikov, Lecture Notes, 2004

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
  \[
  Q_t(x_0, \tilde{x}_0; A \times B) = \mathbb{E}_{x_0, \tilde{x}_0} (1(X_t \in A, \tilde{X}_t \in B) \\
  \times L_t(X, Y)L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}}) \bigg|_{\tilde{Y} = Y}
  \]
Using coupling method, I
A. Veretennikov, Lecture Notes, 2004

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 \(R^{2d}\)
  \[
  \bar{\mu}_t(A \times B; (\mu_0, \nu_0)) = \int_{R^{2d}} Q_t(x_0, \tilde{x}_0; A \times B) \, d\mu_0(x_0) \, d\nu_0(\tilde{x}_0).
  \]
  with
  \[
  Q_t(x_0, \tilde{x}_0; A \times B) = \mathbb{E}_{x_0, \tilde{x}_0}(1(X_t \in A, \tilde{X}_t \in B) \times L_t(X, Y)L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t)_{\tilde{Y} = Y}
  \]

- **Coupling, doubling the space**

- **Doubling the operators, I**
**Remark.** The substitutions are well defined.

**Comparison of measures**

The following properties hold:

\[ \bar{\mu}_t(A; \mu_0) = \bar{\mu}_t(A \times \mathbb{R}^d; (\mu_0, \nu_0)) \]

\[ \bar{\mu}_t(A; \nu_0) = \bar{\mu}_t(A \times \mathbb{R}^d; (\nu_0, \mu_0)) \]

**Comparison of distances**

\[ \| \bar{\mu}_t(\cdot; \mu_0) - \bar{\mu}_t(\cdot; \nu_0) \|_{TV} \leq \| \bar{\mu}_t(\cdot; (\mu_0, \nu_0)) - \bar{\mu}_t(\cdot; (\nu_0, \mu_0)) \|_{TV} \]
Remark. The substitutions are well defined.

Comparison of measures

The following properties hold:

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

Comparison of distances

\[
\| \bar{\mu}_t(\cdot; \mu_0) - \bar{\mu}_t(\cdot; \nu_0) \|_{TV} \leq \| \bar{\mu}_t(\cdot; (\mu_0, \nu_0)) - \bar{\mu}_t(\cdot; (\nu_0, \mu_0)) \|_{TV}
\]
Remark. The substitutions are well defined.

Comparison of measures

The following properties hold:

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

Comparison of distances

\[ \| \overline{\mu}_t(\cdot; \mu_0) - \overline{\mu}_t(\cdot; \nu_0) \|_{TV} \leq \| \overline{\mu}_t(\cdot; (\mu_0, \nu_0)) - \overline{\mu}_t(\cdot; (\nu_0, \mu_0)) \|_{TV} \]
Separation

Partition of unity

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

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

where

$$D_i := \left\{ \max \left( |X_i|, |\tilde{X}_i| \right) \leq R; \right\}$$

$$\max \left( \sup_{i \leq s \leq i+1} |X_s|, \sup_{i \leq s \leq i+1} |\tilde{X}_s| \right) < R + 1$$
Partition of unity

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

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

where

$$D_i := \left\{ \max \left( |X_i|, |\tilde{X}_i| \right) \leq R; \right. \right.$$ 

$$\left. \max \left( \sup_{i \leq s \leq i+1} |X_s|, \sup_{i \leq s \leq i+1} |\tilde{X}_s| \right) < R + 1 \right\}$$
Partition of unity, II

Multiplicative decomposition

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

with

\[ 1_{\delta_i}(D_i) = 1(\delta_i = 1)1(D_i) + 1(\delta_i = 0)(1 - 1(D_i)). \]

Partition of unity

\[ 1 = \sum_{\delta \in \Delta} 1_{\delta}(X, \tilde{X}) \]
Partition of unity, II

**Multiplicative decomposition**

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

with

\[ 1_{\delta_i}(D_i) = 1(\delta_i = 1)1(D_i) + 1(\delta_i = 0)(1 - 1(D_i)). \]

**Partition of unity**

\[ 1 = \sum_{\delta \in \Delta} 1_\delta(X, \tilde{X}) \]
Denote by \( \#1(\delta) \) the total number of ones in \( \delta \) and by

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

The following inequalities hold:

\[
\sum_{\delta: \#1(\delta) < \varepsilon n} 1_{\delta}(X, \tilde{X}) \leq 1(\#1(X)_R < \frac{1 + \varepsilon}{2} n) + 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n)
\]
Separation of pairs

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

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

The following inequalities hold:

Separation of pairs, I

$$\sum_{\delta: \#1(\delta) < \varepsilon n} 1_{\delta}(X, \tilde{X}) \leq 1(\#1(X)_R < \frac{1 + \varepsilon}{2} n) + 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n)$$
Separation of pairs, II

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

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

\[
q = \sup_{x:|x| \leq R} P_x(\sup_{0 \leq s \leq +1} |X_s| \geq R + 1) < 1/2.
\]
Then \((\forall \varepsilon > \frac{1}{2} \text{ and for } R \text{ large enough})\)

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

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

\[
q = \sup_{x:|x| \leq R} \sup_{0 \leq s \leq +1} P_x(\sup_{0 \leq s \leq +1} |X_s| \geq R + 1) < 1/2.
\]
Outline

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and main result

3. Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in \( \mathbb{R}^d \)
   - Reformulation of the problem, the Bayes approach

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
Our goal is to prove the following inequality:

**The main inequality**

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

where

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

\[ C_R = \sup_{|x|,|\tilde{x}|,|z|,|\tilde{z}| \leq R} \frac{u(0, x, \tilde{x})}{u(0, z, \tilde{z})} \]

with \( u(s, x, \tilde{x}) \) - the solution of the first boundary problem.
Our goal is to prove the following inequality:

**The main inequality**

\[
E_{\mu_0} \| \bar{\mu}_t (\cdot; \mu_0) - \bar{\mu}_t (\cdot; \nu_0) \|_{TV} \leq C \sum_{\delta \in \Delta} \kappa_R \#^1(\delta) E_{\mu_0, \nu_0} e^Y_{\delta; \mu_0, \nu_0},
\]

where

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

\[
C_R = \sup_{|x|, |\tilde{x}|, |z|, |\tilde{z}| \leq R} \frac{u(0, x, \tilde{x})}{u(0, z, \tilde{z})}
\]

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

\[
\begin{align*}
&u_s + \frac{1}{2}u_{xx} + \frac{1}{2}u_{\tilde{x}\tilde{x}} \\
&+ (b(x) - (\psi_s - \psi_0)^\ast \nabla h(x))u_x + (b(\tilde{x}) - (\psi_s - \psi_0)^\ast \nabla h(\tilde{x}))u_{\tilde{x}} \\
&+ \frac{1}{2}c(x, \tilde{x}, \psi)u = 0, \\
&u(1, x, \tilde{x}) = (1(x \in A, \tilde{x} \in B) \\
&\times \exp[h^\ast(x)(\psi_1 - \psi_0) + h^\ast(\tilde{x})(\psi_1 - \psi_0)] \\
u(s, x, \tilde{x}) = 0, \quad \forall \quad 0 < s < 1, \quad \max(|x|, |\tilde{x}| = R + 1),
\end{align*}
\]

with a replacement \(\psi = Y\).
The term $e_t^{Y;\delta;\mu_0,\nu_0}$ in the main inequality is defined by:

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

Remark. This term will be the normalizing coefficient.
The term $e_t^{Y;\delta;\mu_0,\nu_0}$ in the main inequality is defined by:

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

Remark. This term will be the normalizing coefficient.
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:

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

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

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

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

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

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

$$
\sum_{\delta: \#1(\delta) < \varepsilon n} E_{\mu_0} \left( E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg\vert_{\tilde{Y} = Y} \right)
\leq \sum_{\delta: \#1(\delta) < \varepsilon n} E_{\mu_0} \left( E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg\vert_{\tilde{Y} = Y} \right).
$$
Theorem 1, sketch of the proof, 1

We split the sum in the main inequality (∀ε > 0):

\[ \sum_{\delta: \#1(\delta) \geq \varepsilon n} + \sum_{\delta: \#1(\delta) < \varepsilon n} \]

and we estimate both terms:

\[ \sum_{\delta: \#1(\delta) \geq \varepsilon n} \kappa_R^{\#1(\delta)} E_{\mu_0} e_{\varepsilon n}^{Y; \delta; \mu_0, \nu_0} \leq \kappa_{\varepsilon n} \]

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

\[ \leq \sum_{\delta: \#1(\delta) < \varepsilon n} E_{\mu_0} \left( E_{\mu_0, \nu_0}(1_{\delta}(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg| \tilde{Y} = Y \right) . \]
Theorem 1, sketch of the proof, 2

- We can finish the proof:

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

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

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

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

- the inequality "separation of pairs" has been used.
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".
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".
Next, we estimate the other term, using the absolute continuity of the initial measures:

\[
E_{\mu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n) \mid \tilde{Y} \right) \bigg| \tilde{Y} = Y \right)
\leq C_2 E_{\nu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n) \mid \tilde{Y} \right) \bigg| \tilde{Y} = Y \right)
\]

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

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:

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

\[
\leq C_2 E_{\nu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n) \mid \tilde{Y} \right) \mid \tilde{Y} = Y \right)
\]

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

Again, the **Separation of pairs, II.**
Next, we estimate the other term, using the absolute continuity of the initial measures:

\[
E_{\mu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n) \mid \tilde{Y} \right) \mid \tilde{Y} = Y \right) \\
\leq C_2 E_{\nu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n) \mid \tilde{Y} \right) \mid \tilde{Y} = Y \right) \\
= C_2 E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{1 + \varepsilon}{2} n) \right),
\]

Again, the Separation of pairs, II.
Outline

1 Introduction
   - Problem statement
   - Historical survey

2 Assumptions and main result

3 Auxiliaries for the proof
   - Parabolic equations and diffusion processes
   - Harnack inequality
   - Birkhoff metric
   - Ergodic processes in $\mathbb{R}^d$
   - Reformulation of the problem, the Bayes approach

4 Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
How can we prove the main inequality?

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

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

with the same kernel $Q_t$:

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

We have

$$
\bar{\mu}_t(A \times B; (\mu_0, \nu_0)) = c_t^{\mu_0} c_t^{\nu_0} \mu_t(A \times B; (\mu_0, \nu_0))
$$
How can we prove the main inequality?
Define new **linear** operators on the space of non-normalized measures on $\mathbb{R}^{2d}$

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

with the same kernel $Q_t$:

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

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 $\mathbb{R}^{2d}$

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

with the same kernel $Q_t$:

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

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

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

with the same kernel $Q_t$:

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

We have

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

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

and with the kernel $Q_t^\delta$:

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

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

and with the kernel $Q_t^\delta$:

$$Q_t^\delta(x_0, \tilde{x}_0; A \times B) = \mathbb{E}_{X_0, \tilde{X}_0}(1(X_t \in A, \tilde{X}_t \in B)1_{\delta}(X, \tilde{X})$$

$$\times L_t(X, Y)L_t(\tilde{X}, \tilde{Y}) \mid \mathcal{F}_t^{Y, \tilde{Y}}) \bigg|_{\tilde{Y} = Y}$$
We see that the normalizing coefficient is exactly the $e_t^{Y;\delta;\mu_0,\nu_0}$.

Probability separator, II

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

$$= c_t^{\mu_0} c_t^{\nu_0} \mu_t^\delta(R^{2d}; (\mu_0, \nu_0))$$
Using the Markov property of $X_t$ we find the recursion (with $Z_t = (X_t, \tilde{X}_t)$):

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

with

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

with

$$
D_n := \left( |Z_{n-1}| \leq R, \sup_{n-1 \leq s \leq n} |Z_s| < R + 1 \right)
$$
We can estimate the total variation norm:

\[
\left\| \bar{\mu}_t(\cdot; (\mu_0, \nu_0)) - \bar{\mu}_t(\cdot; (\nu_0, \mu_0)) \right\|_{TV} 
\leq c_t^{\mu_0} c_t^{\nu_0} \sum_{\delta \in \Delta} \left\| \mu_t^\delta(\mu_0, \nu_0) - \mu_t^\delta(\nu_0, \mu_0) \right\|_{TV}
\]

\[
= \sum_{\delta \in \Delta} e^{Y;\delta;\mu_0,\nu_0} \left\| \hat{\mu}_t^\delta(\mu_0, \nu_0) - \hat{\mu}_t^\delta(\nu_0, \mu_0) \right\|_{TV}
\]

with normalization

\[
\hat{\mu}_t^\delta(\nu_0, \mu_0) = \frac{\mu_t^\delta(\mu_0, \nu_0)}{\mu_t^\delta(\mathbb{R}^{2d}; \mu_0, \nu_0)}
\]
We can estimate the total variation norm:

\[
\|\bar{\mu}_t(\cdot; (\mu_0, \nu_0)) - \bar{\mu}_t(\cdot; (\nu_0, \mu_0))\|_{TV} \\
\leq c_t^{\mu_0} c_t^{\nu_0} \sum_{\delta \in \Delta} \|\mu_t^\delta(\mu_0, \nu_0) - \mu_t^\delta(\nu_0, \mu_0)\|_{TV} \\
= \sum_{\delta \in \Delta} e_{n,Y;\delta;\mu_0,\nu_0} \|\hat{\mu}_t^\delta(\mu_0, \nu_0) - \hat{\mu}_t^\delta(\nu_0, \mu_0)\|_{TV}
\]

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 properties of the Birkhoff metric we see that

**Birkhoff metric, 1st property.**

\[
\| \hat{\mu}_t^\delta (\mu_0, \nu_0) - \hat{\mu}_t^\delta (\nu_0, \mu_0) \|_{TV} \leq \rho(\hat{\mu}_t^\delta (\mu_0, \nu_0); \hat{\mu}_t^\delta (\nu_0, \mu_0)).
\]
Using the Birkhoff metric, 2

and that

Birkhoff metric, 2nd property.

\[ \rho(\hat{\mu}_n(\mu_0, \nu_0); \hat{\mu}_n(\nu_0, \mu_0)) \]
\[ \equiv \rho \left( \mu_n^\delta(\mu_0, \nu_0); \mu_n^\delta(\nu_0, \mu_0) \right) \]
\[ \leq \kappa_n^{\delta} \rho \left( \mu_{n-1}^\delta(\mu_0, \nu_0); \mu_{n-1}^\delta(\nu_0, \mu_0) \right) \]
\[ \leq C \kappa_R^k, \]

with

\[ k = \#1(\delta) \]

which gives the desired inequality.

QED