The Annals of Probability
1991, Vol. 19, No. 2, 538–561

SOME ESTIMATES OF THE TRANSITION DENSITY OF A NONDEGENERATE DIFFUSION MARKOV PROCESS[1]

BY SHUENN-JYI SHEU

Academia Sinica

Dedicated to Professor W. H. Fleming

In this paper we study the transition density $P_t(x, y)$ of a nondegenerate diffusion process by using the stochastic control method invented by Fleming and the idea of stochastic parallel translation. We obtain a two-sided estimate for $P_t(x, y)$ as well as some bounds for the derivatives of $\log P_t(x, y)$.

1. Introduction. Let $X(\cdot)$ be the diffusion Markov process on $\mathbb{R}^d$ with the generator L:

$$Lf(x) = \frac{1}{2}\sum a_{ij}(x)\frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum b_j(x)\frac{\partial f(x)}{\partial x_j},$$

for $f \in C_0^2(R^d)$, the space of functions on $\mathbb{R}^d$ with compact support which is continuous up to second-order derivatives. Here the summation is taken over $i, j = 1, \dots, d$. This convention will be used throughout the paper. In many places we also omit Σ if no confusion arises. Here we assume (A):

(A1) $a_{ij}(\cdot), b_j(\cdot)$ are bounded.

(A2) $(a_{ij}(x))$ is symmetric and $(a_{ij}(x)) \geq c_0 I_{d\times d}$ for all $x \in R^d$. c_0 is some positive constant and $I_{d\times d}$ is the $d \times d$ identity matrix.

(A3) $a_{ij}(\cdot), b_j(\cdot)$ are of Hölder β for some β, $0 < \beta \leq 1$.

The existence of a diffusion Markov process $x(\cdot)$ was proved in [16]. Moreover, $x(\cdot)$ is strongly Markovian and has a transition density $P_t(x, y)$ (see [8]). In this paper, we give some estimates for $P_t(x, y)$ which include a lower and upper bound for $P_t(x, y)$ and bounds for the derivatives of $\log P_t(x, y)$ if $a_{ij}(\cdot), b_j(\cdot)$ are assumed to be smooth.

As we know, $P_t(x, y)$ is the fundamental solution of the parabolic equation

$$\frac{\partial}{\partial t} f(t, x) = Lf(t, x), \qquad t > 0, x \in R^d,$$
$$f(0, x) = f_0(x). \tag{1.1}$$

Received August 1989; revised February 1990.

[1]Research partially supported by NSC 78-0208-M001-08.

AMS 1980 *subject classifications*. Primary 60H10, 60J35; secondary 60J60.

Key words and phrases. Diffusion process, transition density, control method, stochastic parallel translation.

In the theory of partial differential equations, the following properties of $P_t(x, y)$ were derived by using the method of parametrix under our assumption (A):

$$P_t(x,y) \le K_1 t^{-d/2} \exp\left(-c_1 \frac{|y-x|^2}{t}\right),$$

$$(1.2) \qquad P_t(x,y) \ge K_2 t^{-d/2} \exp\left(-c_2 \frac{|y-x|^2}{t}\right) - K_3 t^{-(d-\lambda)/2} \exp\left(-c_3 \frac{|y-x|^2}{t}\right),$$

where K_1, K_2, K_3, c_1, c_2, c_3 and λ are positive constants.

$$(1.3) \qquad |D_x^m P_t(x,y)| \le K_m t^{-(d+|m|)/2} \exp\left(-c_1 \frac{|y-x|^2}{t}\right), \qquad |m| \le 2,$$

where $m = (m_1, \dots, m_d)$, $|m| = m_1 + \cdots + m_d$, $t > 0$, $x, y \in R^d$,

$$D_x^m f(x) = \frac{\partial^{|m|} f(x)}{\partial x_1^{m_1} \cdots \partial x_d^{m_d}}.$$

See [8], page 229. If $a_{ij}(\cdot)$, $b_j(\cdot)$ are smooth, then (1.3) still holds for large $|m|$. On the other hand, the following result was obtained in [1] by using a quite different argument if L has divergence form

$$L = \frac{1}{2} \sum \frac{\partial}{\partial x_i} a_{ij}(x) \frac{\partial}{\partial x_j} + \sum b_j(x) \frac{\partial}{\partial x_j}$$

and satisfies the conditions (A1) and (A2):

$$(1.4) \qquad K_2 t^{-d/2} \exp\left(-c_1 \frac{|y-x|^2}{t}\right) \le P_t(x,y) \le K_1 t^{-d/2} \exp\left(-c_2 \frac{|y-x|^2}{t}\right).$$

See also [12], [13] and [15].

One can also use probabilistic arguments to study $P_t(x, y)$. In [11], Malliavin calculus was applied quite successfully to this problem, where an estimate of type (1.4) was discussed even for diffusions which allow degeneracy of $(a_{ij}(x))$. The basic idea in this approach is to consider $x(t)$ as a "smooth" functional of Brownian motion (Wiener process) $B(\cdot)$. Since we need the smoothness of $a_{ij}(\cdot)$ and $b_j(\cdot)$ to prove that $x(\cdot)$ is a smooth functional, this approach is not applicable if we do not assume $a_{ij}(\cdot)$, $b_j(\cdot)$ to be smooth.

Here we propose another approach. The basic ingredient is the idea of Fleming's logarithmic transformation. That is, we consider $J_t(x, y) = -\log P_t(x, y)$. If we fix y and consider this as a function of (t, x), we find that J satisfies a nonlinear equation which turns out to be the dynamic programming equation of a stochastic control problem, that is,

$$(1.5)\qquad \begin{aligned} \frac{\partial J_t(x,y)}{\partial t} &= \frac{1}{2}\sum a_{ij}(x)\frac{\partial^2 J_t(x,y)}{\partial x_i\,\partial x_j} + \sum b_j(x)\frac{\partial J_t(x,y)}{\partial x_j} \\ &\quad - \frac{1}{2}\sum a_{ij}(x)\frac{\partial J_t(x,y)}{\partial x_i}\frac{\partial J_t(x,y)}{\partial x_j} \\ &= \inf_{u\in R^d}\left[L_x^u J_t(x,y) + k(x,u)\right]. \end{aligned}$$

Here

$$(1.6)\qquad \begin{aligned} L^u f(x) &= \frac{1}{2}\sum a_{ij}(x)\frac{\partial^2 f(x)}{\partial x_i\,\partial x_j} + \sum u_j\frac{\partial f}{\partial x_j}, \\ k(x,u) &= \frac{1}{2}\sum g_{ij}(x)\big(b_i(x) - u_i\big)\big(b_j(x) - u_j\big), \\ \big(g_{ij}(x)\big) &= \big(a_{ij}(x)\big)^{-1}. \end{aligned}$$

We expect that $J_T(x,y)$ is equal to

$$(1.7)\qquad \inf_{u\in\mathscr{F}} E\left[\int_0^T k(\eta(t),u(t))\,dt + J_0(\eta(T),y)\right].$$

Here $u\in\mathscr{F}$ is a measurable function: $[0,T]\times R^d\to R^d$ such that

$$d\eta(t) = u(t,\eta(t))\,dt + \sigma(\eta(t))\,dB(t), \qquad \eta(0) = x$$

has a weak solution and

$$E\left[\int_0^T |u(t)|^2\,dt\right] < \infty, \qquad u(t) \equiv u(t,\eta(t)),$$

$$\sigma(\cdot) = \text{square root of } \big(a_{ij}(\cdot)\big) = a(\cdot).$$

See [6]. Then a lower and upper estimate of $J_T(x,y)$ [hence of $P_T(x,y)$] may be derived by choosing a suitable $u\in\mathscr{F}$. We observe that there is difficulty using this approach due to the fact that $J_0(\eta,y)$ is not well behaved. See [7]. In order to justify this approach we need to use the penalty argument. See [4]. In [14], this idea was used to get the following result.

THEOREM A. *If we assume condition* (A), *then there are* $c_1(\cdot), c_2(\cdot), k_1(\cdot), k_2(\cdot) > 0$ *such that*

$$(1.8)\qquad \begin{aligned} &\left(\frac{1}{\sqrt{2\pi T}}\right)^d \frac{1}{\sqrt{\det a(y)}} k_2(T)\exp(-c_2(T)I(T,x,y)) \\ &\le P_T(x,y) \\ &\le \left(\frac{1}{\sqrt{2\pi T}}\right)^d \frac{1}{\sqrt{\det a(y)}} - k_1(T)\exp(-c_1(T)I(T,x,y)), \end{aligned}$$

where $c_i(\cdot)$, $k_i(\cdot)$, $i = 1, 2$, are bounded above and bounded below away from 0 on bounded intervals. They depend only on the bounds and Hölder constants of $a_{ij}(\cdot)$, $b_j(\cdot)$ as well as c_0 in (A2). *$I(T, x, y)$ is given by*

$$I(T,x,y) = \inf\left\{ \tfrac{1}{2}\int_0^T g_{ij}(\phi(t))\big(\dot\phi(t) - b(\phi(t))\big)_i \times\big(\dot\phi(t) - b(\phi(t))\big)_j \, dt; \phi(0) = x, \phi(T) = y \right\}. \tag{1.9}$$

COROLLARY A. *Let $x^\varepsilon(\cdot)$ be the diffusion process generated by L^ε:*

$$L^\varepsilon f(x) = \frac{\varepsilon}{2}\sum a_{ij}(x)\frac{\partial^2 f(x)}{\partial x_i\,\partial x_j} + \sum b_j(x)\frac{\partial f(x)}{\partial x_j}, \qquad f \in C_0^2(R^d).$$

We assume that condition (A) *holds and $(\partial/\partial x_i)b_j(\cdot)$, $i, j = 1, \dots, d$, are of Hölder β. Denote $P_t^\varepsilon(x, y)$ to be the transition density of $x^\varepsilon(\cdot)$. Then there are $c_i(\cdot)$, $k_i(\cdot)$ as in Theorem* A, *such that*

$$\left(\frac{1}{\sqrt{2\pi T\varepsilon}}\right)^d k_2(T)\exp\left(-c_2(T)\frac{I(T,x,y)}{\varepsilon}\right) \le P_T^\varepsilon(x,y) \le \left(\frac{1}{\sqrt{2\pi T\varepsilon}}\right)^d k_1(T)\exp\left(-c_1(T)\frac{I(T,x,y)}{\varepsilon}\right). \tag{1.10}$$

The proof of these results was carried out in [14]. For the convenience of the reader, we will sketch the main idea in Section 3.

At first glance, it seems difficult to estimate $D_x^m \log P_t(x, y)$ since we are asked to estimate $D_x^m P_t(x, y)/P_t(x, y)$ instead of $D_x^m P_t(x, y)$ as was done in the literature. However, our approach to attack the problem is quite straightforward. For example, in order to estimate $(\partial/\partial x_i)\log P_t(x, y)$, we differentiate (1.5), with respect to x_α to get

$$\begin{aligned}\frac{\partial}{\partial t}D_\alpha J_t(x,y) &= \frac{1}{2}\sum a_{ij}(x)D_iD_jD_\alpha J_t(x,y) \\ &\quad + \sum \big(b(x) - a(x)\nabla J_t(x,y)\big)_j D_j D_\alpha J_t(x,y) \\ &\quad + H_\alpha(t,x).\end{aligned} \tag{1.11}$$

Here $\nabla J_t(x, y)$ is the gradient of $J_t(x, y)$ as a function of x and, for fixed y,

$$\begin{aligned}H_\alpha(t,x) &= \frac{1}{2}\sum D_\alpha a_{ij}(x)D_iD_jJ_t(x,y) \\ &\quad + \sum D_\alpha b_j(x)D_jJ_t(x,y) \\ &\quad - \frac{1}{2}\sum D_\alpha a_{ij}(x)D_iJ_t(x,y)D_jJ_t(x,y).\end{aligned} \tag{1.12}$$

That is, we see $D_\alpha J_t(x, y)$, as a function of (t, x), is a solution of the inhomogeneous parabolic equation (1.11). A difficulty arises since H_α depends on the first and the second derivatives of $J_t(x,y)$. However, quite remarkably, this difficulty will be removed when we perform covariant differentiation along the stochastic curve $\xi(\cdot)$ solving the SDE

$$d\xi(t) = u^*(t,\xi(t))\,dt + \sigma(\xi(t))\,dB(t), \quad \xi(0) = x, \tag{1.13}$$

if we consider R^d as a curved manifold with Riemannian metric $g_{ij}\,dx_i, dx_j$. Here

$$u^*(t,x) = b(x) - a(x)\,\nabla J_{T-t}(x,y) \tag{1.14}$$

is the optimal control for the stochastic control problem (1.7). The process $\xi(\cdot)$ is in fact the process $x(\cdot)$ conditioned by $x(0) = x$, $x(T) = y$. This argument leads to an expression for $D_i J_T(x,y)$ which is similar to the one obtained by Bismut (see [2], Theorem 2.14). Now we state our main result whose proof is given in Section 5.

THEOREM B. *Assume* (A) *and* $a_{ij}(\cdot), b_j(\cdot) \in C_b^\infty(\mathbb{R}^d)$. *Then we have*

$$|D_i \log P_T(x,y)| \le c\frac{1}{T^{1/2}}(I(T,x,y)+1)^{1/2}, \tag{1.15}$$

$$|D_x^m \log P_T(x,y)| \le c\left(\frac{1}{T}\right)^{|m|/2}(I(T,x,y)+1)^{1+(3/2)(|m|-2)}, \qquad |m| \ge 2. \tag{1.16}$$

This implies

$$|D_x^m P_T(x,y)| \le c\left(\frac{1}{T}\right)^{|m|/2}(I(T,x,y)+1)^{1+(3/2)(|m|-2)}P_T(x,y)$$

for $|m| \ge 2$ *and in turn implies the classical estimation* (1.3) *for the derivatives of* $P_T(x,y)$ *from Theorem* A.

We now give the notation which will be used in the rest of this paper.

Throughout the rest of this paper Σ will be omitted unless there arises confusion, for example, $a_{ij}u_iu_j$ means $\sum a_{ij}u_iu_j$.

$C_b(R^d)$ is the space of bounded continuous functions from $\mathbb{R}^d$ to R.

$C_b^\infty(\mathbb{R}^d)$ is the space of bounded smooth functions from $\mathbb{R}^d$ to R whose derivatives are in $C_b(\mathbb{R}^d)$.

For $f \in C_b(\mathbb{R}^d)$, $\|f\| = \sup_x|f(x)|$. If f is of Hölder β, we denote $\|f\|_\beta = \|f\| + \sup_{x\neq y}|f(x)-f(y)|/|x-y|^\beta$. $D^m f(x) = D_1^{m_1}\cdots D_d^{m_d}f(x)$ if $m = (m_1,\dots,m_d)$. Here $D_i f(x) = (\partial/\partial x_i)f(x)$. $\nabla f(x) = (D_1 f(x),\dots,D_d f(x))$. $|m| = m_1 + \cdots + m_d$ if $m = (m_1,\dots,m_d)$, m_i, $i = 1,\dots,d$, are nonnegative integers.

For $x \in \mathbb{R}^d$, $\sigma(x) = (\sigma_{ij}(x))$ is a positive definite matrix which is the square root of $a(x) = (a_{ij}(x))$. $\sigma(\cdot)$ is of Hölder β if we assume (A) (see [9], volume 1, page 128).

$$(g_{ij}(x)) = a(x)^{-1},$$

$$\Gamma^k_{ij}(x) = \tfrac{1}{2} a_{ks}(x)\big(D_j g_{si} + D_i g_{sj} - D_s g_{ij}\big)(x).$$

For ϕ: $[0, T] \to \mathbb{R}^d$ with each component being absolutely continuous, $I_T(\phi) = \int_0^T k(\phi(t), \dot{\phi}(t))\, dt$. $I(T, x, y) \equiv \inf\{I_T(\phi);\ \phi(0) = x,\ \phi(T) = y\}$. Here $k(\cdot, \cdot)$ is given in (1.6).

$\mathscr{F}_{T,x}$ = family of u: $[0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ being measurable such that

$$d\eta(t) = u(t, \eta(t))\, dt + \sigma(\eta(t))\, dB(t),$$
$$\eta(0) = x,$$

has a weak solution $\eta(\cdot)$ which satisfies

$$E\left[\int_0^T |u(t, \eta(t))|^2\, dt\right] < \infty,$$

with $B(t) = (B_1(t), \dots, B_d(t))$ d-dim Brownian motion.

$\int_0^T u(s)\, dB(s)$ is the Itô integral and $\int_0^T u(s) \circ dB(s)$ is the Stratonovich integral (see [10]).

2. Logarithmic transformation. Assume (A). Then for a function $f \in C_b(\mathbb{R}^d)$ of Hölder β, the function

$$f(t, x) = \int P_t(x, y) f(y)\, dy = E_x[f(x(t))]$$

has the properties that $D_i D_j f(t, x)$, $D_j f(t, x)$, $(\partial/\partial t) f(t, x)$ are continuous and satisfy

$$\frac{\partial}{\partial t} f(t, x) = Lf(t, x), \qquad t > 0,\ x \in R^d,$$
$$f(0, x) = f(x);$$

see [8].

Especially for a fixed $y_0 \in \mathbb{R}^d$, take

$$f_\alpha(y) = \left(\frac{1}{\sqrt{2\pi\alpha}}\right)^d \frac{1}{\sqrt{\det a(y_0)}} \exp\left(-\frac{1}{2\alpha} g_{ij}(y_0)(y - y_0)_i (y - y_0) j\right). \tag{2.1}$$

Let

$$P^\alpha(t, x) = E_x[f_\alpha(x(t))]. \tag{2.2}$$

Remark that

$$\lim_{\alpha\to 0} P^\alpha(t,x) = P_t(x,y_0), \tag{2.3}$$

$J^\alpha(t,x) = -\log P^\alpha(t,x)$ satisfies the nonlinear PDE

$$\begin{aligned} \frac{\partial}{\partial t}J^\alpha(t,x) &= \frac{1}{2}a_{ij}(x)D_iD_jJ^\alpha(t,x) + b_j(x)D_jJ^\alpha(t,x) \\ &\quad - \frac{1}{2}a_{ij}(x)D_iJ^\alpha(t,x)D_jJ^\alpha(t,x), \qquad t>0,\ x\in R^d, \\ J^\alpha(0,x) &= \frac{d}{2}\log 2\pi\alpha + \frac{1}{2}\log\det a(y_0) \\ &\quad + \frac{1}{2\alpha}g_{ij}(y_0)(x-y_0)_i(x-y_0)_j. \end{aligned} \tag{2.4}$$

By a standard argument using the Itô formula, we have the following. (See [6].)

LEMMA 2.1.

$$J^\alpha(T,x) = \inf_{u\in\mathscr{F}_{T,x}} E\left[\int_0^T k(\eta(t),u(t))\,dt + J^\alpha(0,\eta(T))\right].$$

For the notation see (1.6) *and* (1.7). *Moreover the* inf *can be attained at* u^*_α, *where*

$$u^*_\alpha(t,x) = b(x) - a(x)\nabla J^\alpha(T-t,x). \tag{2.5}$$

COROLLARY 2.2. *Let*

$$J_T(x,y) = -\log P_T(x,y).$$

Then

$$J_T(x,y_0) = \lim_{\alpha\to 0}\inf_{u\in\mathscr{F}_{T-\alpha,x}} E\left[\int_0^{T-\alpha} k(\eta(t),u(t))\,dt + J^\alpha(0,\eta(T-\alpha))\right].$$

3. Two-sided estimates. In order to obtain a lower and upper bound for $P_T(x_0,y_0)$ stated in Theorem A, we only need to obtain a lower and upper bound for $J^\alpha(T-\alpha,x_0)$ which is independent of α. We will sketch the proof assuming $a_{ij}(\cdot), b_j(\cdot)\in C_b^\infty(R^d)$. One may find the details of the proof for the general case in [14].

First, we treat the upper bound. We pick ϕ: $[0,T]\to\mathbb{R}^d$ to be absolutely continuous, $\phi(0)=x_0$, $\phi(T)=y_0$ such that $I_T(\phi)=I(T,x,y)$. Define

$$u(t,x) = \dot\phi(t) - \frac{x-\phi(t)}{T-t}, \qquad 0\le t\le T-\alpha. \tag{3.1}$$

It is easy to see $u\in\mathscr{F}_{T-\alpha,x_0}$, since $u(t,x)$ is Lipschitz continuous in x.

Therefore

$$(3.2)\quad J^{\alpha}(T-\alpha, x_0) \le E\left[\int_0^{T-\alpha} k(\eta(t), u(t))\, dt + J^{\alpha}(0, \eta(T-\alpha))\right]$$

by Lemma 2.1. Here

$$(3.3)\quad \begin{aligned} d\eta(t) &= u(t, \eta(t))\, dt + \sigma(\eta(t))\, dB(t),\\ \eta(0) &= x_0. \end{aligned}$$

From (3.3),

$$d(\eta(t) - \phi(t)) = -\frac{\eta(t) - \phi(t)}{T-t} dt + \sigma(\eta(t))\, dB(t).$$

Therefore

$$(3.4)\quad \frac{\eta(t) - \phi(t)}{T-t} = \int_0^t \frac{1}{T-s} \sigma(\eta(s))\, dB(s).$$

This implies

$$(3.5)\quad \begin{aligned} &E\left[|\eta(t) - \phi(t)|^2\right] \le c(T-t), \qquad 0 \le t \le T-\alpha,\\ &E\left[(\eta(t) - \phi(t))_i (\eta(t) - \phi(t))_j\right]\\ &\quad = (T-t)^2 E\left\{\int_0^t \left(\frac{1}{T-s}\right)^2 a_{ij}(\eta(s))\, ds\right\}. \end{aligned}$$

Here $c > 0$ is some constant.

We now use the Taylor expansion of $k(\eta(t), u(t))$ around $(\phi(t), \dot{\phi}(t))$,

$$(3.6)\quad \begin{aligned} k(\eta(t), u(t)) &\doteq k(\phi(t), \dot{\phi}(t)) + D_{x_i} k\, \Delta_i^1(t) + D_{u_i} k\, \Delta_i^2(t)\\ &\quad + \tfrac{1}{2}\Big(D_{x_i x_j} k\, \Delta_i^1(t)\, \Delta_j^1(t) + 2 D_{x_i u_j} k\, \Delta_i^1(t)\, \Delta_j^2(t)\\ &\qquad + D_{u_i u_j} k\, \Delta_i^2(t)\, \Delta_j^2(t)\Big). \end{aligned}$$

Here

$$\Delta^1(t) = \eta(t) - \phi(t),$$

$$\Delta^2(t) = u(t) - \dot{\phi}(t) = \frac{\eta(t) - \phi(t)}{T-t} = -\frac{1}{T-t}\Delta^1(t),$$

$$D_{x_i} k = \frac{\partial}{\partial x_i} k(\phi(t), \dot{\phi}(t))$$

and similarly for $D_{u_i} k$, $D_{x_i x_j} k$, $D_{x_i u_j} k$ and $D_{u_i u_j} k$.

It is easy to see

$$(3.7)\quad E\left[\int_0^{T-\alpha} \left(D_{x_i} k\, \Delta_i^1(t) + D_{u_i} k\, \Delta_i^2(t)\right) dt\right] = 0$$

by the martingale property of $[\eta(t) - \phi(t)]/(T-t)$ [see (3.4)]. Now consider the term

$$\tfrac{1}{2}E\left[\int_0^{T-\alpha} D_{u_i u_j} k\, \Delta_j^2(t)\Delta_j^2(t)\, dt\right].$$

Since $D_{u_i u_j} k = g_{ij}(\phi(t))$, this term is equal to

$$\begin{aligned}(3.8)\qquad &\frac{1}{2}\int_0^{T-\alpha} g_{ij}(\phi(t))\,E\left[\left(\frac{\eta(t)-\phi(t)}{T-t}\right)_i\left(\frac{\eta(t)-\phi(t)}{T-t}\right)_j\right] dt\\ &= \frac{1}{2}\int_0^{T-\alpha} E\left[\int_0^t \left(\frac{1}{T-s}\right)^2 g_{ij}(\phi(t))a_{ij}(\eta(s))\, ds\right] dt \quad [\text{from } (3.4)].\end{aligned}$$

Writing

$$\begin{aligned}a_{ij}(\eta(s)) &= \big(a_{ij}(\eta(s)) - a_{ij}(\phi(s))\big)\\ &\quad + \big(a_{ij}(\phi(s)) - a_{ij}(\phi(t))\big) + a_{ij}(\phi(t)),\end{aligned}$$

and noticing $(g_{ij}(\cdot)) = (a_{ij}(\cdot))^{-1}$ and $a_{ij}(\cdot) = a_{ji}(\cdot)$, we see by some calculation using (3.5),

$$\begin{aligned}(3.9)\qquad (3.8) &\le \frac{d}{2}\int_0^{T-\alpha}\int_0^t \left(\frac{1}{T-s}\right)^2 ds\, dt + c_1 I_T(\phi)^{1/2} + c_2\\ &= \frac{d}{2}(\log T - \log\alpha) + c_1 I_T(\phi)^{1/2} + c_2.\end{aligned}$$

It is not difficult to estimate the other terms in (3.6) as follows:

$$\begin{aligned}(3.10)\qquad &E\left[\int_0^{T-\alpha} D_{x_i x_j} k\, \Delta_i^1(t)\, \Delta_j^1(t)\, dt\right] \le c_1 I_T(\phi) + c_2,\\ &E\left[\int_0^{T-\alpha} D_{x_i u_j} k\, \Delta_i^1(t)\, \Delta_j^2(t)\, dt\right] \le c_1 I_T(\phi)^{1/2} + c_2.\end{aligned}$$

Finally,

$$\begin{aligned}(3.11)\qquad &E[J^\alpha(0, \eta(T-\alpha))]\\ &\quad = \frac{d}{2}\log 2\pi\alpha + \frac{1}{2}\log\det a(y_0)\\ &\qquad + \frac{1}{2\alpha} g_{ij}(y_0) E\big[(\eta(T-\alpha) - y_0)_i(\eta(T-\alpha) - y_0)_j\big].\end{aligned}$$

Writing

$$\eta(T-\alpha) - y_0 = \eta(T-\alpha) - \phi(T-\alpha) + \phi(T-\alpha) - \phi(T),$$

we have from (3.4), (3.5), and the Schwarz inequality,

$$\begin{aligned}(3.12)\qquad &E\big[(\eta(T-\alpha) - y_0)_i(\eta(T-\alpha) - y_0)_j\big]\\ &\quad = E\big[(\eta(T-\alpha) - \phi(T-\alpha))_i(\eta(T-\alpha) - \phi(T-\alpha)_j)\big]\\ &\qquad + (\phi(T-\alpha) - \phi(T))_i(\phi(T-\alpha) - \phi(T)_j)\big]\\ &\quad \le c\alpha + c\alpha I_T(\phi).\end{aligned}$$

Putting (3.6)–(3.12) into (3.2), we get

$$J^\alpha(T-\alpha, x_0) \le \frac{d}{2}\log 2\pi T + \frac{1}{2}\log\det a(y_0) + c_2(T) I(T, x_0, y_0) + k_2(T),$$

with $c_2(T), k_2(T)$ being independent of α. This also gives a lower bound for $P_T(x_0, y_0)$ in Theorem A, except we have to explain how one can handle the remaining part due to the approximation of $k(\eta(t), u(t))$ in (3.6). The difference between the two sides of (3.6) is given by

$$\int_0^1\int_0^1 \big(D_{x_ix_j}k(\lambda\mu) - D_{x_ix_j}k(0)\big)\,\Delta_i^1(t)\,\Delta_j^1(t)\lambda\,d\mu\,d\lambda$$
$$+ 2\int_0^1\int_0^1 \big(D_{x_iu_j}k(\lambda\mu) - D_{x_iu_j}k(0)\big)\,\Delta_i^1(t)\,\Delta_j^2(t)\lambda\,d\lambda\,d\mu$$
$$+ \int_0^1\int_0^1 \big(D_{u_iu_j}k(\lambda\mu) - D_{u_iu_j}k(0)\big)\,\Delta_i^2(t)\,\Delta_j^2(t)\lambda\,d\lambda\,d\mu.$$

Here, for example,

$$D_{u_iu_j}k(\lambda) = D_{u_iu_j}k\big(\phi(t) + \lambda\,\Delta^1(t), \dot\phi(t) + \lambda\,\Delta^2(t)\big).$$

Then

$$\left|E\left[\int_0^1\int_0^1 \big(D_{u_iu_j}k(\lambda\mu) - D_{u_iu_j}k(0)\big)\,\Delta_i^2(t)\,\Delta_j^2(t)\lambda\,d\lambda\,d\mu\right]\right|$$
$$\le cE\left[|\Delta^1(t)|\,|\Delta^2(t)|^2\right]$$
$$\le c(T-t)^{-1/2}.$$

Then integrate this with respect to t, from 0 to $T-\alpha$ to show that the error term due to this is bounded by $cT^{1/2}$. Similarly, we can estimate the error due to the remaining terms.

Now we treat the lower bound for $J^\alpha(T-\alpha, x_0)$. Let u_α^* be as in Lemma 2.1,

$$d\xi(t) = u_\alpha^*(t, \xi(t))\,dt + \sigma(\xi(t))\,dB(t),$$
$$\xi(0) = x_0.$$

Then

$$J^\alpha(T-\alpha, x_0) = E\left[\int_0^{T-\alpha} k(\xi(t), u^*(t))\,dt + J^\alpha(0, \xi(T-\alpha))\right]. \tag{3.13}$$

Here $u^*(t) \equiv u_\alpha^*(t, \xi(t))$.

We define the random path $\psi(\cdot)$ which is the solution of

$$\dot\psi(t) = \frac{\xi(t) - \psi(t)}{T-t} + u^*(t), \qquad 0 \le t \le T-\alpha,$$
$$\psi(0) = x_0. \tag{3.14}$$

Then

$$\frac{\xi(t)-\psi(t)}{T-t} = \int_0^t \frac{1}{T-s}\sigma(\xi(s))\,dB(s). \tag{3.15}$$

Similar to (3.5), this implies

$$\begin{aligned} &E\left[|\xi(t)-\psi(t)|^2\right] \le c(T-t),\\ &E\left[(\xi(t)-\psi(t))_i(\xi(t)-\psi(t))_j\right]\\ &\quad = (T-t)^2 E\left[\int_0^t \left(\frac{1}{T-s}\right)^2 a_{ij}(\xi(s))\,ds\right]. \end{aligned} \tag{3.16}$$

Moreover, (3.13) implies

$$E\left[J^\alpha(0,\xi(T-\alpha))\right] \le J^\alpha(T-\alpha,x_0).$$

Therefore

$$E\left[|\xi(T-\alpha)-y_0|^2\right] \le c_1\alpha J^\alpha(T-\alpha,x_0) - c_2\alpha\log\alpha. \tag{3.17}$$

Together with (3.16), we see that $\psi(T-\alpha)$ is close to y_0 as $\alpha\to 0$, at least in probability. Therefore, in probability, $I(T,x_0,y_0)$ will be a lower bound for $\int_0^{T-\alpha}k(\psi(t),\dot\psi(t))\,dt$ as $\alpha\to 0$. Taking this into account, using the Taylor expansion of $k(\xi(t),u^*(t))$ around $(\psi(t),\dot\psi(t))$ as in (3.6) and estimating each term from below in terms of Z_α as before,

$$Z_\alpha = E\left[\int_0^{T-\alpha}k(\psi(t),\dot\psi(t))\,dt\right],$$

letting $\alpha\to 0$, we may obtain a lower bound for $J_T(x_0,y_0)$. In fact, we may rewrite the right side of (3.13) as $I_1+I_2+I_3+I_4+I_5+I_6$, where

$$I_1 = \frac{1}{2}E\left[\int_0^{T-\alpha} g_{ij}(\xi(t))(b(\psi(t))-\dot\psi(t))_i(b(\psi(t))-\dot\psi(t))_j\,dt\right],$$

$$I_2 = E\left[\int_0^{T-\alpha} g_{ij}(\xi(t))(b(\psi(t))-\dot\psi(t))_i\left(b(\xi(t))-b(\psi(t)) + \frac{\xi(t)-\psi(t)}{T-t}\right)_j dt\right],$$

$$I_3 = \frac{1}{2}E\left[\int_0^{T-\alpha} g_{ij}(\xi(t))(b(\xi(t))-b(\psi(t)))_i(b(\xi(t))-b(\psi(t)))_j\,dt\right],$$

$$I_4 = E\left[\int_0^{T-\alpha} g_{ij}(\xi(t))(b(\xi(t))-b(\psi(t)))_i(\xi(t)-\psi(t))_j\frac{1}{T-t}dt\right],$$

$$I_5 = \frac{1}{2}E\left[\int_0^{T-\alpha} g_{ij}(\xi(t))(\xi(t)-\psi(t))_i(\xi(t)-\psi(t))_j\left(\frac{1}{T-t}\right)^2 dt\right],$$

$$I_6 = E\left[J^\alpha(0,\xi(T-\alpha))\right].$$

It is easy to get

$$\begin{aligned}
&I_1 \geq cZ_\alpha,\\
&I_3 \geq 0,\\
&I_4 \geq -cT^{1/2},\\
&I_5 \geq -\frac{d}{2}\ln\frac{\alpha}{T} - \frac{d}{2}\frac{T-\alpha}{T} - cT^{1/2}\left(1 + Z_\alpha^{1/2}\right)
\end{aligned}$$

by using an idea similar to that of proving (3.9).

The term I_2 causes the main difficulty. This is due to the fact that (3.8) is no longer true. This term can be rewritten as $L_1 + L_2 + L_3 + L_4$. Here

$$\begin{aligned}
L_1 &= E\left[\int_0^{T-\alpha} g_{ij}(\xi(t))\big(b(\psi(t)) - \dot{\psi}(t)\big)_i\big(b(\xi(t)) - b(\psi(t))\big)_j\,dt\right]\\
&\geq -cT^{1/2}Z_\alpha^{1/2},\\
L_2 &= E\left[\int_0^{T-\alpha} g_{ij}(y_0)\big(b(\psi(t)) - \dot{\psi}(t)\big)_i\big(\xi(t) - \psi(t)\big)_j\frac{1}{T-t}\,dt\right]\\
&= E\left[\int_0^{T-\alpha} g_{ij}(y_0)b_i(\psi(t))\big(\xi(t) - \psi(t)\big)_j\frac{1}{T-t}\,dt\right]\\
&\qquad - g_{ij}(y_0)E\left[\big(\psi(T-\alpha) - y_0\big)_i\big(\xi(T-\alpha) - \psi(T-\alpha)\big)_j\right]\frac{1}{\alpha}.
\end{aligned}$$

Here we use the martingale property of $(\xi(t) - \psi(t))_i 1/(T-t)$. From this and (3.16)–(3.17) we see L_2 is bounded from below by

$$\begin{aligned}
&-cT^{1/2} - c - c\left(\frac{1}{\alpha}E\left[g_{ij}(y_0)\big(\xi(T-\alpha) - y_0\big)_i\big(\xi(T-\alpha) - y_0\big)_j\right]\right)^{1/2}\\
&\qquad \geq -cT^{1/2} - c - \frac{1}{2\alpha}E\left[g_{ij}(y_0)\big(\xi(T-\alpha) - y_0\big)_i\big(\xi(T-\alpha) - y_0\big)_j\right].
\end{aligned}$$

$$\begin{aligned}
L_3 &= E\Bigg[\int_0^{T-\alpha}\big(g_{ij}(\xi(t)) - g_{ij}(\psi(t))\big)\big(b(\psi(t)) - \dot{\psi}(t)\big)_i\\
&\qquad\qquad \times\big(\xi(t) - \psi(t)\big)_j\frac{1}{T-t}\,dt\Bigg]\\
&\geq -cT^{1/2}Z_\alpha^{1/2},\\
L_4 &= E\Bigg[\int_0^{T-\alpha}\big(g_{ij}(\psi(t)) - g_{ij}(y_0)\big)\big(b(\psi(t)) - \dot{\psi}(t)\big)_i\\
&\qquad\qquad \times\big(\xi(t) - \psi(t)\big)_j\frac{1}{T-t}\,dt\Bigg].
\end{aligned}$$

We now use

$$
\begin{aligned}
&|g_{ij}(\psi(t)) - g_{ij}(y_0)| \\
&\le \left| g_{ij}(\psi(t)) - g_{ij}\left(\psi(T-\alpha) - \int_t^{T-\alpha} b(\psi(s))\,ds\right)\right| \\
&\quad + \left| g_{ij}\left(\psi(T-\alpha) - \int_t^{T-\alpha} b(\psi(s))\,ds\right) - g_{ij}(y_0)\right| \\
&\le c\left|\psi(t) - \psi(T-\alpha) + \int_t^{T-\alpha} b(\psi(s))\,ds\right|^\beta \\
&\quad + c\left|\int_t^{T-\alpha} b(\psi(s))\,ds\right|^\beta + c|\psi(T-\alpha) - y_0|^\beta \\
&\le c\left(\int_t^{T-\alpha} |\dot\psi - b(\psi(s))|\,ds\right)^\beta + c(T-t)^\beta + c|\psi(T-\alpha) - y_0|^\beta,
\end{aligned}
$$

for a β with $0 < 2\beta < 1$, and (3.16) and (3.17) as well as some simple calculation to get that

$$L_4 \ge -cT^\beta Z_\alpha^{1/2} - cZ_\alpha^{1/2+(\beta/2)}T^{\beta/2}.$$

Finally, putting them together, using $Z_\alpha^\gamma \le \delta Z_\alpha + c(\delta,\gamma)$, for $0 \le \gamma < 1$ and $\delta > 0$, noticing that $\liminf_{\alpha\to 0} Z_\alpha \ge I(T, x_0, y_0)$, we may conclude that

$$J_T(x_0, y_0) \ge \tfrac{1}{2}\log 2\pi T + \tfrac{1}{2}\det a(y_0) + c_1(T)I(T, x_0, y_0) - k_1(T).$$

This also gives an upper bound for $P_T(x_0, y_0)$ as in Theorem A.

4. Stochastic parallel translation. In this section we assume (A) and $a_{ij}(\cdot), b_j(\cdot) \in C_b^\infty(\mathbb{R}^d)$. Let $c(\cdot)$: $[0,T] \to \mathbb{R}^d$ be a smooth curve in $\mathbb{R}^d$. The solution of

$$\frac{d}{dt}u^i(t) + \Gamma_{kl}^i(c(t))u^k(t)\dot c^l(t) = 0 \tag{4.1}$$

gives the matrix $(E_j^i(t))$ [depending on $c(\cdot)$],

$$u^i(t) = E_j^i(t)u^j(0).$$

$u(t) = (u^1(t), \dots, u^d(t))$ [viewed as a vector at $c(t)$] is the parallel translation of $u(0)$ along $c(\cdot)$.

We consider also the equation

$$\frac{d}{dt}v_i(t) - \Gamma_{i,l}^k(c(t))v_k(t)\dot c^l(t) = 0, \tag{4.2}$$

which gives the matrix $(F_j^i(t))$, $0 \le t \le T$,

$$v_i(t) = F_i^j(t)v_j(0).$$

$v(t) = (v_1(t), \dots, v_d(t))$ [viewed as a covector at $c(t)$] is also called the parallel translation of $v(0)$ along $c(\cdot)$. The following are several basic properties of $(E_j^i(t)), (F_j^i(t))$. (See [3], Chapter 7.)

LEMMA 4.1.

(i) $(E^i_j(t))$ *is the inverse of* $(F^i_j(t))$.

(ii) $g_{ij}(c(t))u^i(t)u^i(t)$ $[a_{ij}(c(t))v_i(t)v_j(t)]$ *is independent of* t *if* $u(\cdot)$ $[v(\cdot)]$ *is a solution of* (4.1) [(4.2), *resp*.].

This implies, in particular, that

(iii) *there is a constant* $c > 0$ *such that for all* t, $|E^i_j(t)| \le c$, $|F^i_j(t)| \le c$.

We may consider an equation which is analogous to (4.1) and (4.2) with $c(\cdot)$ being replaced by a random curve.

LEMMA 4.2. *Assume* $\xi(t) = (\xi^1(t), \dots, \xi^d(t))$, *a stochastic process on* $(\Omega, \mathscr{F}, P)$, *is such that* $\xi^i(t)$, $i = 1, \dots, d$, *are semimartingales w.r.t. a filtration* $(\mathscr{F}_t)_{t \ge 0}$. *Then for each* $u, v \in \mathbb{R}^d$, *there exist unique semimartingales* $u^i(t)$ *and* $v_i(t)$, $1 \le i \le d$, *such that*

$$du^i + \Gamma^i_{kl}(\xi(t))u^k(t) \circ d\xi^l(t) = 0, \qquad u^i(0) = u^i, \tag{4.3}$$

$$dv_i + \Gamma^k_{i,l}(\xi(t))v_k(t) \circ d\xi^l(t) = 0, \qquad v_i(0) = v_i. \tag{4.4}$$

We may write

$$\begin{aligned} u^i(t) &= E^i_j(t)u^j(0), \\ v_i(t) &= F^j_i(t)v_j(0). \end{aligned} \tag{4.5}$$

For each j, $E^i_j(\cdot)[F^j_i(\cdot)]$ *satisfies* (4.3) [(4.4), *resp*.] *and*

$$E^i_j(0) = \delta_{ij} = F^i_j(0). \tag{4.6}$$

Properties (*i*), (*ii*) *and* (*iii*) *in Lemma* 4.1 *hold in this case.*

PROOF. Existence and uniqueness of the solution for equations (4.3) and (4.4) follow from [10], Theorem 2.1, page 103 (see also the proof of the corollary on page 106 there). The rest follows as that of Lemma 4.1 if we note that the Stratonovich integral obeys the ordinary differential rule. □

$u(t)$ $[v(t)]$ is called a stochastic parallel translation (or displacement) of $u(0)$ $[v(0)]$ along a stochastic process $\xi(\cdot)$. One can find more on the subject in [10].

5. Estimation of the derivatives. In this section, we assume condition (A) and $a_{ij}(\cdot), b_j(\cdot) \in C^\infty_b(\mathbb{R}^d)$. Our main object is to prove Theorem B.

In the rest, $T > 0$ and $x_0, y_0 \in \mathbb{R}^d$ are fixed.

$$J(t,x) = -\log P_t(x, y_0),$$

$$u^*(t,x) = b(x) - a(x)\nabla J(T-t,x),$$

$J(t,x)$ is the same as $J_t(x, y_0)$ in Section 1. The stochastic process $\xi(t)$ solves

SDE (1.13), that is,

$$d\xi(t) = u^*(t, \xi(t))\,dt + \sigma(\xi(t))\,dB(t),$$
$$\xi(0) = x.$$

$E_x[\cdots]$ is the expectation w.r.t. the distribution of $\xi(\cdot)$ with $\xi(0) = x$.

We first consider $D_i J(T, x)$. We start with a lemma.

LEMMA 5.1. *$\xi(\cdot)$ exists in a weak sense in $[0, T)$. The following holds*:

$$\begin{aligned} J(T, x_0) = {} & E_{x_0}\left[\frac{1}{2}\int_0^t a_{ij}(\xi(s)) D_i J(T-s, \xi(s)) D_j J(T-s, \xi(s))\,ds\right] \\ & + E_{x_0}[J(T-t, \xi(t))], \qquad t<T. \end{aligned} \tag{5.1}$$

COROLLARY 5.2 (Prior estimate). *There is a constant $c > 0$ depending on T such that*

$$E_{x_0}\left[\int_0^{T/2} |\nabla J(T-s, \xi(s))|^2\,ds\right] \le c(I(T, x_0, y_0) + 1). \tag{5.2}$$

PROOFS. (5.2) follows easily from (5.1), condition (A) and Theorem A. For showing (5.1), we have, by the Itô rule,

$$\begin{aligned} dJ(T-t, \xi(t)) = {} & -\tfrac{1}{2}a_{ij}(\xi(t)) D_i J(T-t, \xi(t)) D_j J(T-t, \xi(t))\,dt \\ & + D_i J(T-t, \xi(t))\sigma_{ij}(\xi(t))\,dB_j(t), \end{aligned}$$

for $t < T$ up to the explosion time τ of $\xi(\cdot)$. In particular,

$$\begin{aligned} & E_{x_0}\left[\frac{1}{2}\int_0^{t\wedge\tau_N} a_{ij}(\xi(s)) D_i J(T-s, \xi(s)) D_j J(T-s, \xi(s))\,ds\right] \\ & \quad = J(T, x_0) - E_{x_0}[J(T - t\wedge\tau_N, \xi(t\wedge\tau_N))], \end{aligned} \tag{5.3}$$

where $\tau_N = \inf\{t < T;\ |\xi(t)| \ge N\}$, $t < T$. By Theorem A, the right-hand side of (5.3) has an upper bound which is independent of N. This implies, for each $t < T$,

$$E_{x_0}\left[\int_0^{t\wedge\tau} |\nabla J(T-s, \xi(s))|^2\,ds\right] \le c \tag{5.4}$$

for some c depending on t. Taking into account (1.13), we conclude that $\tau > t$ a.e. Therefore $\xi(\cdot)$ exists in $[0, T)$, and (5.4) implies

$$E_{x_0}\left[\int_0^t |\nabla J(T-s, \xi(s))|^2\,ds\right] \le c,$$

which further implies $|\xi(t)|$ is square integrable. From this, Theorem A and an inequality $I(t, x, y_0) \le c((1/t)|x - y_0|^2 + 1)$, we may apply the dominated convergence theorem when $N \to \infty$ to obtain (5.1). This ends the proof of Lemma 5.1. □

LEMMA 5.3. *For every integer $m > 0$, there is $c > 0$ such that*

$$E_{x_0}\left[I(T-t,\xi(t),y_0)^m\right] \le c\left(I(T,x_0,y_0)+1\right)^m, \qquad t \le \frac{T}{2}.$$

PROOF. By Theorem A, there are $c_1 > 0$, $\alpha > 0$ such that

$$\hat{J}(t,x) \equiv J(T,x) - \frac{d}{2}\log 2\pi T + c_1$$

has the property

$$\hat{J}(t,x) \le k_1 I(t,x,y_0) + c_2,$$
$$\hat{J}(t,x) \ge k_1 I(t,x,y_0), \qquad \frac{T}{2} \le t \le T,$$

for some k_1, k_2 and $c_2 > 0$. Since

$$d\hat{J}(T-t,\xi(t)) = -\tfrac{1}{2}a_{ij}(\xi(t))D_iJ(T-t,\xi(t))D_jJ(T-t,\xi(t))\,dt + D_\alpha J(T-t,\xi(t))\sigma_{\alpha\beta}(\xi(t))\,dB_\beta(t) \quad \text{[from (1.5)]},$$

we have

$$\begin{aligned} d\hat{J}^m(T-t,\xi(t)) = \Big(&\frac{m(m-1)}{2}a_{ij}(\xi(t))D_iJD_jJ\hat{J}^{m-2}(T-t,\xi(t)) \\ &-\frac{m}{2}a_{ij}(\xi(t))D_iJD_jJ\hat{J}^{m-1}(T-t,\xi(t))\Big)\,dt \\ &+ dM(t), \qquad t \le \frac{T}{2}, \end{aligned} \tag{5.5}$$

$M(\cdot)$ a martingale. This implies

$$\begin{aligned} &\frac{m}{2}E_{x_0}\left[\int_0^{T/2}\hat{J}^{m-1}a_{ij}D_iJD_jJ(T-t,\xi(t))\,dt\right] \\ &= \frac{m(m-1)}{2}E_{x_0}\left[\int_0^{T/2}\hat{J}^{m-2}a_{ij}D_iJD_jJ(T-t,\xi(t))\,dt\right] \\ &\quad + \hat{J}^m(T,x_0) - E_{x_0}\left[\hat{J}^m\left(\frac{T}{2},\xi\left(\frac{T}{2}\right)\right)\right] \\ &= \cdots \\ &= \sum_{k=1}^m \frac{m!}{k!}\left(\hat{J}^k(T,x_0) - E_{x_0}\left[\hat{J}^k\left(\frac{T}{2},\xi\left(\frac{T}{2}\right)\right)\right]\right) \quad \text{(from Lemma 5.1)} \\ &\le c\left(I(T,x_0,y_0)+1\right)^m \quad \text{(by the positivity of } \hat{J}). \end{aligned}$$

By (5.5) again,

$$\begin{aligned} &E_{x_0}\left[\hat{J}^m(T-t,\xi(t))\right] \\ &\le \hat{J}^m(T,x_0) + \frac{m(m-1)}{2}E_{x_0}\left[\int_0^{T/2}\hat{J}^{m-2}a_{ij}D_iJD_jJ(T-t,\xi(t))\,dt\right] \\ &\le c\left(I(T,x_0,y_0)+1\right)^m. \end{aligned}$$

This implies the inequality we want. □

The following gives the required estimate for $D_iJ(T,x)$ and a prior estimate for $D_{ij}J(t,x)$ in an average form which plays a similar role as that of Corollary 5.2 to $D_iJ(t,x)$.

LEMMA 5.4.

$$|D_iJ(T,x_0)| \le c\frac{1}{T^{1/2}}\big(I(T,x_0,y_0)+1\big)^{1/2}, \tag{5.6}$$

$$E_{x_0}\left[\int_0^{T/2}|D_{i_1i_2}J(T-t,\xi(t))|^2\,dt\right] \le c\frac{1}{T}\big(I(T,x_0,y_0)+1\big). \tag{5.7}$$

PROOF. By the Itô rule and (1.11),

$$dD_\alpha J(T-t,\xi(t)) = -H_\alpha(T-t,\xi(t))\,dt + dM_\alpha(t),$$

$$M_\alpha(t) = \int_0^t D_iD_\alpha J(T-s,\xi(s))\sigma_{ij}(\xi(s))\,dB_j(s).$$

Now we perform covariant differentiation of $D_\alpha J(T-t,\xi(t))$ along the stochastic curve $\xi(\cdot)$. Here we consider $\mathbb{R}^d$ as a curved manifold with Riemannian metric $g_{ij}\,dx_i\,dx_j$. That is, we consider $I_\beta(t) = E_{\beta\alpha}(t)D_\alpha(J(T-t,\xi(t)))$, where

$$dE_{\beta\alpha}(t) = -E_{\beta k}(t)\Gamma^\alpha_{kl}(\xi(t))\circ d\xi_l(t),$$

$$E_{\beta\alpha}(0) = \delta_{\beta\alpha},$$

and in Itô type,

$$\begin{aligned} dE_{\beta\alpha}(t) = &-\tfrac{1}{2}\big(E_{\beta k}(t)D_p\Gamma^\alpha_{kl}(\xi(t)) - E_{\beta m}(t)\Gamma^k_{mp}(\xi(t))\Gamma^\alpha_{kl}(\xi(t))\big)a_{pl}(\xi(t))\,dt \\ &- E_{\beta k}(t)\Gamma^\alpha_{kl}(\xi(t))\,d\xi_l(t). \end{aligned}$$

See Lemma 4.2. We have

$$\begin{aligned} dI_\beta(t) &= E_{\beta\alpha}(t)\circ dD_\alpha J(T-t,\xi(t)) \\ &\quad - E_{\beta k}(t)\Gamma^\alpha_{kl}(\xi(t))D_\alpha J(T-t,\xi(t))\circ d\xi_l(t) \\ &= \big\{-E_{\beta\alpha}(t)H_\alpha(T-t,\xi(t)) \\ &\quad -E_{\beta k}(t)\Gamma^\alpha_{kl}(\xi(t))D_\alpha J(T-t,\xi(t))\big(b_l(\xi(t)) \\ &\quad\quad -a_{lp}(\xi(t))D_pJ(T-t,\xi(t))\big)\big\}\,dt + \tfrac{1}{2}d\langle E_{\beta\alpha}, D_\alpha J\rangle_t \\ &\quad -\tfrac{1}{2}d\langle E_{\beta k}\Gamma^\alpha_{kl}D_\alpha J, \xi_l\rangle_t + dN_\beta(t), \end{aligned} \tag{5.8}$$

$$\begin{aligned} N_\beta(t) = \int_0^t &\big(E_{\beta\alpha}(s)D_\alpha D_lJ(T-s,\xi(s)) \\ &-E_{\beta k}(s)\Gamma^\alpha_{kl}(\xi(s))D_\alpha J(T-s,\xi(s))\big)\sigma_{lj}(\xi(s))\,dB_j(s), \end{aligned}$$

$$\begin{aligned} d\langle E_{\beta\alpha}, D_\alpha J\rangle_t &= -E_{\beta k}(t)\Gamma^\alpha_{kl}(\xi(t))a_{lp}(\xi(t))D_pD_\alpha J(T-t,\xi(t))\,dt \\ &= \tfrac{1}{2}E_{\beta k}(t)D_ka_{lp}(\xi(t))D_pD_lJ(T-t,\xi(t))\,dt \end{aligned} \tag{5.9}$$

and

$$
\begin{aligned}
& d\langle E_{\beta k}\Gamma^{\alpha}_{kl}D_{\alpha}J, \xi_l\rangle \\
&\quad = \big\{-E_{\beta m}(t)\Gamma^{k}_{mp}(\xi(t))\Gamma^{\alpha}_{kl}(\xi(t))D_{\alpha}J(T-t,\xi(t))a_{pl}(\xi(t)) \\
&\qquad + E_{\beta k}(t)D_p\Gamma^{\alpha}_{kl}(\xi(t))D_{\alpha}J(T-t,\xi(t))a_{pl}(\xi(t)) \\
&\qquad + E_{\beta k}(t)\Gamma^{\alpha}_{kp}(\xi(t))D_pD_{\alpha}J(T-t,\xi(t))a_{pl}(\xi(t))\big\}\,dt \\
&\quad = -E_{\beta m}(t)\big(\Gamma^{k}_{mp}(\xi(t))\Gamma^{\alpha}_{kl}(\xi(t)) - D_p\Gamma^{\alpha}_{ml}(\xi(t))\big) \\
&\qquad \times D_{\alpha}J(T-t,\xi(t))a_{pl}(\xi(t))\,dt \\
&\qquad - \tfrac{1}{2}E_{\beta k}(t)D_k a_{lp}(\xi(t))D_pD_lJ(T-t,\xi(t))\,dt.
\end{aligned} \tag{5.10}
$$

Here we use

$$\Gamma^{\alpha}_{kl}a_{lp}D_pD_{\alpha}f = -\tfrac{1}{2}D_k a_{lp}D_pD_l f, \tag{5.11}$$

which follows by using the definition of Γ^{k}_{ij} in Section 1. In fact,

$$\Gamma^{\alpha}_{kl}a_{lp}D_pD_{\alpha}f = \tfrac{1}{2}a_{\alpha m}(D_k g_{lm} + D_l g_{km} - D_m g_{kl})a_{lp}D_pD_{\alpha}f.$$

Using the fact that (g_{ij}) is the inverse of (a_{ij}), we have

$$
\begin{aligned}
D_k g_{lm} a_{lp} &= -g_{lm}D_k a_{lp}, \\
\tfrac{1}{2}a_{\alpha m}D_k g_{lm}a_{lp}D_pD_{\alpha}f &= -\tfrac{1}{2}a_{\alpha m}g_{lm}D_k a_{lp}D_pD_{\alpha}f \\
&= \text{right-hand side of } (5.11).
\end{aligned}
$$

And notice that the terms $\frac{1}{2}a_{\alpha m}D_l g_{km}a_{lp}D_pD_{\alpha}f$, $\frac{1}{2}a_{\alpha m}D_m g_{kl}a_{lp}D_pD_{\alpha}f$ are equal to each other if we replace the indexes l by m, m by l, p by α and α by p in the latter term. Then we get (5.11).

Put these into (5.8) and simplify to get

$$
\begin{aligned}
dI_{\beta}(t) &= A_{\beta}(t)\,dt + dN_{\beta}(t), \\
A_{\beta}(t) &= -E_{\beta m}(t)\big\{D_m b_{\alpha}(\xi(t)) + \Gamma^{\alpha}_{ml}(\xi(t))b_l(\xi(t)) \\
&\qquad - \big(\tfrac{1}{2}\Gamma^{k}_{mp}(\xi(t))\Gamma^{\alpha}_{kl}(\xi(t)) - D_p\Gamma^{\alpha}_{ml}(\xi(t))\big)a_{pl}(\xi(t))\big\} \\
&\qquad \times D_{\alpha}J(T-t,\xi(t)) \\
&= \text{a linear combination of } D_{\alpha}J(T-t,\xi(t)) \\
&\qquad \text{with bounded coefficients.}
\end{aligned} \tag{5.12}
$$

Therefore

$$
\begin{aligned}
D_{\beta}J(T,x_0) &= E_{x_0}[I_{\beta}(t)] - E_{x_0}\left[\int_0^t A_{\beta}(s)\,ds\right] \\
&= \frac{2}{T}E_{x_0}\left[\int_0^{T/2} I_{\beta}(t)\,dt\right] - \frac{2}{T}E_{x_0}\left[\int_0^{T/2} A_{\beta}(s)\left(\frac{T}{2} - s\right)ds\right].
\end{aligned}
$$

(5.6) follows from this and Corollary 5.2 and Lemma 4.2(iii).

To prove (5.7), we have from relation (5.12),

$$E_{x_0}\left[N_\beta^2\left(\frac{T}{2}\right)\right] = E_{x_0}\left[\left(I_\beta\left(\frac{T}{2}\right) - I_\beta(0) - \int_0^{T/2} A_\beta(t)\,dt\right)^2\right]$$
$$\le cE_{x_0}\left[\left|I_\beta\left(\frac{T}{2}\right)\right|^2 + T\int_0^{T/2} |\nabla J(T-t,\xi(t))|^2\,dt\right]$$
$$\le c\frac{1}{T}(I(T,x_0,y_0)+1).$$

Here we use (5.6) and Lemma 5.3. This gives

$$E_{x_0}\left[\int_0^{T/2} |E_{\beta k}(t) D_k D_l J(T-t,\xi(t)) - E_{\beta k}(t)\Gamma_{kl}^\alpha(\xi(t)) D_\alpha J(T-t,\xi(t))|^2\,dt\right]$$
$$\le c\frac{1}{T}(I(T,x_0,y_0)+1),$$

which in turn implies (5.7) if we use Corollary 5.2 and Lemma 4.2. This ends the proof. □

In the rest, we consider $n \ge 2$. We will prove the following by induction on n:

$$(5.13)_n \quad |D_{i_1,\dots,i_n} J(T,x_0)| \le c\left(\frac{1}{T}\right)^{n/2} (I(T,x_0,y_0)+1)^{1+(3/2)(n-2)},$$

$$(5.14)_n \quad E_{x_0}\left[\int_0^{T/2} |D_{i_1,\dots,i_{n+1}} J(T-t,\xi(t))|^2\,dt\right] \le c\left(\frac{1}{T}\right)^n (I(T,x_0,y_0)+1)^{1+3(n-1)}$$

holds for all $1 \le i_k \le d$, $k = 1,\dots,n+1$. Theorem B follows from these.

Before we start, we remark about the basic idea used in proving these assertions. We will see that $(5.14)_n$ follows from $(5.13)_n$ together with an expression for the semimartingale $D_{i_1,\dots,i_n}J(T-t,\xi(t))$. Then $(5.14)_n$ serves as a prior estimate when we want to prove $(5.13)_{n+1}$. Lemma 5.1, which follows from optimality of $\xi(\cdot)$, is the starting point of this iteration procedure.

We differentiate (1.5) with respect to $D_{i_1},\dots,D_{i_n}$,

$$(5.15)\quad \begin{aligned} \frac{\partial}{\partial t} D_{i_1,\dots,i_n} J &= \tfrac{1}{2} a_{ij} D_{ij} D_{i_1,\dots,i_n} J + (b_j - a_{ij} D_i J) D_j D_{i_1,\dots,i_n} J \\ &\quad + \sum_{u=1}^n \tfrac{1}{2} D_{i_u} a_{ij} D_{iji_1,\dots,\hat{i}_u,\dots,i_n} J + f_{i_1,\dots,i_n} \\ &\quad + g_{i_1,\dots,i_n} + h_{i_1,\dots,i_n}. \end{aligned}$$

Here $D_{iji_1,\dots,\hat{i}_u,\dots,i_n}J$ means $D_{iji_1,\dots,i_{u-1},i_{u+1},\dots,i_n}J$.

$$f_{i_1,\dots,i_n}(t,x) = \tfrac{1}{2}D_{k_1,\dots,k_m}a_{ij}(x)D_{ijl_1,\dots,l_{n-m}}J(t,x)$$

with summation being taken over $2 \le m \le n$, $\{k_1,\dots,k_m,\ l_1,\dots,l_{n-m}\} = \{i_1,\dots,i_n\}$.

$$g_{i_1,\dots,i_n}(t,x) = D_{k_1,\dots,k_m}b_j(x)D_{jl_1,\dots,l_{n-m}}J(t,x)$$

with summation being taken over $1 \le m \le n$, $\{k_1,\dots,k_m,\ l_1,\dots,l_{n-m}\} = \{i_1,\dots,i_n\}$.

$$h_{i_1,\dots,i_n}(t,x) = -\tfrac{1}{2}D_{k_1,\dots,k_m}a_{ij}(x)D_{ip_1,\dots,p_u}J(t,x)D_{jq_1,\dots,q_v}J(t,x)$$

with summation being taken over $1 \le m \le n$, or $m = 0$ and $1 \le u \le n-1$, $\{k_1,\dots,k_m,p_1,\dots,p_u,q_1,\dots,q_v\} = \{i_1,\dots,i_n\}$. We note that $f_{i_1,\dots,i_n}, g_{i_1,\dots,i_n}$ are linear combinations of the derivatives of J of order up to n.

As before, we consider $E_{\alpha_1 i_1(t)},\dots,E_{\alpha_n i_n}(t)D_{i_1,\dots,i_n}J(T-t,\xi(t))$:

$$\begin{aligned}
dE_{\alpha_1 i_1}(t)&\cdots E_{\alpha_n i_n}(t)D_{i_1,\dots,i_n}J(T-t,\xi(t))\\
&= E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots E_{\alpha_n i_n}(t)D_{i_1,\dots,i_n}J(T-t,\xi(t))\,dE_{\alpha_u i_u}(t)\\
&\quad + E_{\alpha_1 i_1}(t)\cdots E_{\alpha_n i_n}(t)\,dD_{i_1,\dots,i_n}J(T-t,\xi(t))\\
&\quad + E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots \hat{E}_{\alpha_v i_v}(t)\cdots E_{\alpha_n, i_n}(t)D_{i_1,\dots,i_n}J(T-t,\xi(t))\\
&\qquad \times d\langle E_{\alpha_u i_u}, E_{\alpha_v i_v}\rangle_t\\
&\quad + E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots E_{\alpha_n, i_n}(t)\,d\langle E_{\alpha_u i_u}, D_{i_1,\dots,i_n}J\rangle_t.
\end{aligned}$$

Here $(\cdots \hat{E}_{\alpha_u i_u}\cdots)$ means that $E_{\alpha_u i_u}$ is deleted in the expression. We write

$$\begin{aligned}
dE_{\alpha_1 i_1}(t)&\cdots E_{\alpha_n i_n}(t)D_{i_1,\dots,i_n}J(T-t,\xi(t))\\
&= \big(F_{\alpha_1,\dots,\alpha_n}(t) + G_{\alpha_1,\dots,\alpha_n}(t) + H_{\alpha_1,\dots,\alpha_n}(t)\big)\,dt\\
&\quad + \sigma_{\alpha_1,\dots,\alpha_n;k}(t)\,dB_k(t),
\end{aligned} \tag{5.16}$$

where

$$\begin{aligned}
F_{\alpha_1,\dots,\alpha_n}(t) &= -E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots\\
&\qquad \times E_{\alpha_n i_n}(t)\bigg(E_{\alpha_u p}(t)\Gamma^{i_u}_{pq}(\xi(t))b_q(\xi(t))\\
&\qquad\qquad + \frac{1}{2}\frac{d}{dt}\langle \xi_q, E_{\alpha_u p}(t)\Gamma^{i_u}_{pq}\rangle_t\bigg)D_{i_1,\dots,i_n}J(T-t,\xi(t))\\
&\qquad - E_{\alpha_1 i_1}(t)\cdots E_{\alpha_n i_n}(t)\big(f_{i_1,\dots,i_n} + g_{i_1,\dots,i_n}\big)(T-t,\xi(t))\\
&\qquad + E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots \hat{E}_{\alpha_v i_v}(t)\cdots E_{\alpha_n i_n}(t)E_{\alpha_u p_1}(t)\\
&\qquad \times \Gamma^{i_u}_{p_1 q_1}(\xi(t))E_{\alpha_v p_2}(t)\Gamma^{i_v}_{p_2 q_2}(\xi(t))\\
&\qquad \times a_{q_1 q_2}(\xi(t))D_{i_1,\dots,i_n}J(T-t,\xi(t))\\
&= \text{a linear combination of the derivatives of } J \text{ of order up to } n,
\end{aligned}$$

$$G_{\alpha_1,\dots,\alpha_n}(t) = E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots E_{\alpha_n i_n}(t)E_{\alpha_u p}(t)\Gamma^{i_u}_{pq}(\xi(t))$$
$$\times a_{iq}(\xi(t))D_iJD_{i_1,\dots,i_n}J(T-t,\xi(t))$$
$$- E_{\alpha_1 i_1}(t)\cdots E_{\alpha_n i_n}(t)h_{i_1,\dots,i_n}(T-t,\xi(t)),$$
$$H_{\alpha_1,\dots,\alpha_n}(t) = -\tfrac{1}{2}E_{\alpha_1 i_1}(t)\cdots E_{\alpha_n i_n}(t)D_{i_u}a_{ij}(\xi(t))D_{iji_1,\dots,\hat{i}_u,\dots,i_n}J(T-t,\xi(t))$$
$$- E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots E_{\alpha_n i_n}(t)E_{\alpha_u p}(t)$$
$$\times\Gamma^{i_u}_{pq}(\xi(t))a_{q\beta}(\xi(t))D_{\beta i_1,\dots,i_n}J(T-t,\xi(t)),$$
$$\sigma_{\alpha_1,\dots,\alpha_n;k}(t) = -E_{\alpha_1 i_1}(t)\cdots \hat{E}_{\alpha_u i_u}(t)\cdots E_{\alpha_n i_n}(t)E_{\alpha_u p}(t)\Gamma^{i_u}_{pq}(\xi(t))$$
$$\times\sigma_{qk}(\xi(t))D_{i_1,\dots,i_n}J(T-t,\xi(t))$$
$$+ E_{\alpha_1 i_1}(t)\cdots E_{\alpha_n i_n}(t)D_{\beta i_1,\dots,i_n}J(T-t,\xi(t))\sigma_{\beta k}(\xi(t)).$$

Using (5.11), it is not difficult to show

$$H_{\alpha_1,\dots,\alpha_n}(t) \equiv 0$$

and

$$G_{\alpha_1,\dots,\alpha_n}(t) = \tfrac{1}{2}E_{\alpha_1 i_1}(t)\cdots E_{\alpha_n i_n}(t)\big\{D_{i_u}a_{ij}(\xi(t))D_iJD_{ji_1,\dots,\hat{i}_u,\dots,i_n}J(T-t,\xi(t))$$
$$+D_{k_1,\dots,k_m}a_{ij}(\xi(t))D_{ip_1,\dots,p_u}JD_{jq_1,\dots,q_v}J(T-t,\xi(t))\big\}$$

with summation over $1 \le m \le n$ or $m = 0$ and $1 \le u \le n-1$, $1 \le v \le n-1$ and $\{i_1,\dots,i_n\} = \{k_1,\dots,k_m,p_1,\dots,p_u,q_1,\dots,q_v\}$.

Let us first consider the case $n = 2$.

$$\tag{5.17} \begin{aligned} &dE_{\alpha_1 i_1}(t)E_{\alpha_2 i_2}(t)D_{i_1 i_2}J(T-t,\xi(t)) \\ &\quad = \big(F_{\alpha_1\alpha_2}(t) + G_{\alpha_1\alpha_2}(t)\big)\,dt + \sigma_{\alpha_1\alpha_2;k}(t)\,dB_k(t), \end{aligned}$$

$$\tag{5.18} \begin{aligned} &|F_{\alpha_1\alpha_2}(t)| \le c\sum|D_{i_1 i_2}J(T-t,\xi(t))| + \sum|D_iJ(T-t,\xi(t))|, \\ &|G_{\alpha_1\alpha_2}(t)| \le c\Big(\sum|D_{i_1 i_2}J(T-t,\xi(t))| + \sum|D_iJ(T-t,\xi(t))|\Big)^2. \end{aligned}$$

Then

$$D_{\alpha_1\alpha_2}J(T,x_0) = E_{x_0}\big[E_{\alpha_1 i_1}(t)E_{\alpha_2 i_2}(t)D_{i_1 i_2}J(T-t,\xi(t))\big]$$
$$- E_{x_0}\left[\int_0^t\big(F_{\alpha_1\alpha_2}(s) + G_{\alpha_1\alpha_2}(s)\big)\,ds\right].$$

Integrate both sides with respect to t from 0 to $T/2$, divide by $T/2$, then apply Lemma 5.4 and Corollary 5.2 to get

$$|D_{\alpha_1\alpha_2}J(T,x_0)| \le c\left(\frac{1}{T}\right)^{1/2}\left(\frac{1}{T}(I(T,x_0,y_0)+1)\right)^{1/2} + cT^{1/2}\left(\frac{1}{T}(I(T,x_0,y_0)+1)\right)^{1/2} + c\frac{1}{T}(I(T,x_0,y_0)+1)$$
$$\le c\frac{1}{T}(I(T,x_0,y_0)+1).$$

This proves (5.13) for $n=2$.

Using (5.17) again, we have

$$E_{x_0}\left[\int_0^{T/2}\sigma^2_{\alpha_1\alpha_2;k}(t)\,dt\right] \le cE_{x_0}\left[\left|D_{i_1i_2}J\left(\frac{T}{2},\xi\left(\frac{T}{2}\right)\right)\right|^2\right] + cTE_{x_0}\left[\int_0^{T/2}\left(|F_{\alpha_1\alpha_2}(t)|^2 + G_{\alpha_1\alpha_2}(t)|^2\right)dt\right].$$

By (5.13) and Lemma 5.3,

$$E_{x_0}\left[\left|D_{i_1i_2}J\left(\frac{T}{2},\xi\left(\frac{T}{2}\right)\right)\right|^2\right] \le c\frac{1}{T^2}(I(T,x_0,y_0)+1)^2.$$

By Lemma 5.4,

$$E_{x_0}\left[\int_0^{T/2}|F_{\alpha_1\alpha_2}(t)|^2\,dt\right] \le c\frac{1}{T}(I(T,x_0,y_0)+1).$$

By (5.18) and (5.13) and Lemmas 5.3 and 5.4, we have

$$E_{x_0}\left[|G_{\alpha_1\alpha_2}(t)|^2\right] \le c\frac{1}{T^4}(I(T,x_0,y_0)+1)^4.$$

Combine these to get

$$E_{x_0}\left[\int_0^{T/2}\sigma^2_{\alpha_1\alpha_2;k}(t)\,dt\right] \le c\frac{1}{T^2}(I(T,x_0,y_0)+1)^4.$$

By the form of $\sigma_{\alpha_1\alpha_2;k}(\cdot)$, we may deduce from the above relation that

$$E_{x_0}\left[\int_0^{T/2}|D_{i_1i_2i_3}J(T-t,\xi(t))|^2\,dt\right] \le c\frac{1}{T^2}(I(T,x_0,y_0)+1)^4.$$

We have thus proved (5.13) and (5.14) for $n=2$.

Assume that (5.13) and (5.14) hold for $2 \le n \le N-1$. We need to prove that they are true for $n=N$. Since

$$|G_{\alpha_1,\dots,\alpha_N}(t)| \le c|D_{p_1,\dots,p_u}J(T-t,\xi(t))|\,|D_{q_1,\dots,q_v}J(T-t,\xi(t))|$$

with summation being taken over $1 \le u \le N$, $1 \le v \le N$, $u+v \le N+1$ or

$2 \le u, v$, but $u + v = N + 2$. Using a similar argument as above, we show

$$
\begin{aligned}
&|D_{\alpha_1,\dots,\alpha_N} J(T, x_0)| \\
&\quad \le c \frac{1}{T^{1/2}} \left(\frac{1}{T}\right)^{(N-1)/2} (I(T, x_0, y_0) + 1)^{[1+3(N-2)]/2} \\
&\qquad + c T^{1/2} \left(\frac{1}{T}\right)^{(N-1)/2} (I(T, x_0, y_0) + 1)^{[1+3(N-2)]/2} \\
&\qquad + c \left(\frac{1}{T}\right)^{(u-1)/2} (I(T, x_0, y_0) + 1)^{[1+3(u-2)]/2} \left(\frac{1}{T}\right)^{(v-1)/2} \\
&\qquad\quad \times (I(T, x_0, y_0) + 1)^{[1+3(v-2)]/2} \qquad (u \ge 2, v \ge 2, u + v = N + 2) \\
&\qquad + c (I(T, x_0, y_0) + 1)^{1/2} \left(\frac{1}{T}\right)^{(v-1)/2} \\
&\qquad\quad \times (I(T, x_0, y_0) + 1)^{[1+3(v-2)]/2} \qquad (v \le N) \\
&\quad \le c \left(\frac{1}{T}\right)^{N/2} (I(T, x_0, y_0) + 1)^{[1+3(N-2)]/2}.
\end{aligned}
$$

This proves (5.13) for $n = N$.

From (5.16),

$$
\begin{aligned}
E_{x_0}\left[\int_0^{T/2} \sigma^2_{\alpha_1,\dots,\alpha_N;k}(t)\,dt\right] &\le c E_{x_0}\left[\left|D_{i_1,\dots,i_N} J\left(\frac{T}{2}, \xi\left(\frac{T}{2}\right)\right)\right|^2\right] \\
&\quad + c T E_{x_0}\left[\int_0^{T/2} |F_{\alpha_1,\dots,\alpha_N}(t)|^2 + |G_{\alpha_1,\dots,\alpha_N}(t)|^2\,dt\right],
\end{aligned}
$$

$$
E_{x_0}\left[\left|D_{i_1,\dots,i_N} J\left(\frac{T}{2}, \xi\left(\frac{T}{2}\right)\right)\right|^2\right] \le c \left(\frac{1}{T}\right)^N (I(T, x_0, y_0) + 1)^{2+3(N-2)},
$$

$$
E_{x_0}\left[|F_{\alpha_1,\dots,\alpha_N}(t)|^2\right] \le c \left(\frac{1}{T}\right)^N (I(T, x_0, y_0) + 1)^{2+3(N-2)}, \qquad 0 \le t \le \frac{T}{2},
$$

$$
\begin{aligned}
E_{x_0}\left[|G_{\alpha_1,\dots,\alpha_N}(t)|^2\right] &\le c \left(\frac{1}{T}\right)^{u+v} (I(T, x_0, y_0) + 1)^{4+3(u+v-4)} \\
&\qquad\qquad (u \ge 2, v \ge 2, u + v = N + 2) \\
&\quad + c \left(\frac{1}{T}\right)^{1+v} (I(T, x_0, y_0) + 1)^{3(v-1)} \qquad (v \le N) \\
&\le c \left(\frac{1}{T}\right)^{N+2} (I(T, x_0, y_0) + 1)^{1+3(N-1)}.
\end{aligned}
$$

Therefore

$$
E_{x_0}\left[\int_0^{T/2} \alpha_{i_1,\dots,i_N;k}(t)\,dt\right] \le c \left(\frac{1}{T}\right)^N (I(T, x_0, y_0) + 1)^{1+3(N-1)}
$$

and (5.14) follows from this easily. This ends the proof of Theorem B.

Acknowledgments. The idea of using the so-called "control method" (or logarithmic transformation) was initiated by Professor W. H. Fleming and has been applied successfully to various problems, for example, the problems related to small random perturbed diffusion processes, the large deviation theory of Markov processes, the theory of reciprocal diffusion processes. The problems discussed in this paper serve as another example of where this idea could be useful. I would like to dedicate this paper to Professor Fleming on the occasion of his sixtieth birthday for teaching me this fundamental idea. I would also like to thank the referee for his effective help in revising this article.

REFERENCES

[1] ARONSON, D. G. (1967). Bounds on the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* **73** 890–896.

[2] BISMUT, J. M. (1984). *Large Deviations and the Malliavin Calculus*. Birkhäuser, Boston.

[3] BOOTHBY, W. M. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic, New York.

[4] FLEMING, W. H. (1978). Exit probabilities and optimal stochastic control. *Appl. Math. Optim.* **4** 329–346.

[5] FLEMING, W. H. (1982). Logarithmic transformations and stochastic control. *Lecture Notes in Control and Inform. Sci.* **42** 131–141. Springer, New York.

[6] FLEMING, W. H. and RISHEL, R. W. (1975). *Deterministic and Stochastic Optimal Control*. Springer, New York.

[7] FLEMING, W. H. and SHEU, S. J. (1985). Stochastic variational formula for fundamental solutions of parabolic P.D.E. *Appl. Math. Optim.* **13** 193–204.

[8] FRIEDMAN, A. (1964). *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, N.J.

[9] FRIEDMAN, A. (1975). *Stochastic Differential Equations and Applications* **1**. Academic, New York.

[10] IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam.

[11] KUSUOKA, S. and STROOCK, D. (1987). Application of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sec. IA Math.* **34** 391–442.

[12] KUSUOKA, S. and STROOCK, D. (1988). Long time estimates for the heat kernel associated with a uniformly subelliptic symmetric second order operator. *Ann. of Math. (2)* **127** 165–189.

[13] PORPER, F. O. and EIDEL'MAN, S. D. (1984). Two-sided estimates of fundamental solutions of second-order parabolic equations, and some applications. *Russian Math. Surveys* **39** 119–178.

[14] SHEU, S.-J. (1990). Two-sided estimates of fundamental solutions of second-order parabolic equations via logarithmic transformation. Technical Report, Institute of Mathematics, Academia Sinica.

[15] STROOCK, D. (1988). Diffusion semigroups corresponding to uniformly elliptic divergence form operators. *Lecture Notes in Math.* **1321** 316–348. Springer, New York.

[16] STROOCK, D. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Process*. Springer, New York.

INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
NANKANG, TAIPEI
TAIWAN
REPUBLIC OF CHINA