Reinforcement Learning via Nonparametric Smoothing in a Continuous-Time Stochastic Setting with Noisy Data

Chenyang Jiang[1], Bowen Hu[1], Yazhen Wang[1], Shang Wu[2]

University of Wisconsin-Madison[1] *and Fudan University*[2]

Supplementary Material

S1 Supplementary Appendix: Proofs

We fix notations to facilitate long proof arguments. Denote by C a generic constant whose value may change from appearance to appearance. We specify value functions under the deterministic and stochastic cases as follows.

1. $V_0^{\pi}(x)$ denotes the value function for the deterministic $X_t(0)$, namely, $V^{\pi}(x) = \int_0^{\infty} e^{-\beta t} r(X_t(0), \pi(X_t(0)))dt$ given $X_0(0) = x$. Let $V_0^{*,M}(x)$ be the corresponding optimal value function that is the maximum of $V_0^{\pi}(x)$ over $\pi \in \mathrm{Lip}_M$.

2. $V_\sigma^{\pi}(x)$ denotes the value function for the stochastic $X_t(\sigma)$, namely, $V_\sigma^{\pi}(x) = E\left[\int_0^{\infty} e^{-\beta t} r(X_t(\sigma), \pi(X_t(\sigma))dt \middle| X_0(\sigma) = x\right]$. Let $V_\sigma^{*,M}(x)$ be the corresponding optimal value function that is the maximum of $V_\sigma^{\pi}(x)$ over $\pi \in \mathrm{Lip}_M$.

We collect processes for discrete samples $\check{X}_{t_\ell}(0)$ and $\tilde{X}_{t_k}^h(\sigma)$ as follows.

1. $\check{X}_{t_\ell}(0)$, $\ell = 1, \cdots, N$, denotes the discrete trajectory sampled from $X_t(0)$ by using

algorithm (4.4) and (4.5). Let $\check{X}_t(0)$ be the continuous-time step process corresponding to $\check{X}_{t_\ell}(0)$ where $\check{X}_t(0)$ is equal to $\check{X}_{t_\ell}(0)$ for $t \in (t_{\ell-1}, t_\ell]$.

2. $\tilde{X}^h_{t_k}(\sigma)$, $k = 1, \cdots, n = N/m$, denotes the discrete smoothed samples obtained by the implementation procedure (4.6-(4.8), and $\tilde{X}^h_t(\sigma)$ denotes the continuous-time smoothed process corresponding to $\tilde{X}^h_{t_k}(\sigma)$.

In addition to the value functions $V^\pi_0(x)$ and $V^\pi_\sigma(x)$ we define the following value functions:

1. Let $\check{V}^{0,\pi}_0 = \int_0^\infty e^{-\beta t} r(\check{X}_t(0), \pi(\check{X}_t(0)))dt$ given $\check{X}_0(0) = x$, which is the value function for the deterministic $\check{X}_t(0)$ (or $\check{X}_{t_\ell}(0)$).
2. Let $\tilde{V}^{h,\pi}_\sigma(x) = E\left[\int_0^\infty e^{-\beta t} r(\tilde{X}^h_t(\sigma), \pi(\tilde{X}^h_t(\sigma)))dt \middle| \tilde{X}^h_0(\sigma) = x\right]$ be the value function for the stochastic $\tilde{X}^h_t(\sigma)$ (or $\tilde{X}^h_{t_k}(\sigma)$).
3. Define the optimal value functions with policies restricted in Lip_M in the deterministic and stochastic cases as follows. Let $V^{*,M}_0(x) = \sup_{\pi\in\mathrm{Lip}_M} V^\pi_0(x)$, $\check{V}^{0,*,M}_0(x) = \sup_{\pi\in\mathrm{Lip}_M} \check{V}^{0,\pi}_0(x)$, $V^{*,M}_\sigma(x) = \sup_{\pi\in\mathrm{Lip}_M} V^\pi_\sigma(x)$, and $\tilde{V}^{h,*,M}_\sigma(x) = \sup_{\pi\in\mathrm{Lip}_M} \tilde{V}^{h,\pi}_\sigma(x)$.

We define the optimal policies in Lip_M as follows.

1. $\pi^{*,M}_0(x)$ denotes an optimal policy for $X_t(0)$, namely, $\pi^{*,M}_0(x) = \underset{\pi\in\mathrm{Lip}_M}{\arg\max} V^\pi_0(x)$.
2. $\check{\pi}^{0,*,M}_0(x)$ denotes an optimal policy for $\check{X}_t(0)$, namely, $\check{\pi}^{0,*,M}_0(x) = \underset{\pi\in\mathrm{Lip}_M}{\arg\max} \check{V}^{0,\pi}_0(x)$.
3. $\pi^{*,M}_\sigma(x)$ denotes an optimal policy for $X_t(\sigma)$, namely, $\pi^{*,M}_\sigma(x) = \underset{\pi\in\mathrm{Lip}_M}{\arg\max} V^\pi_\sigma(x)$.
4. $\tilde{\pi}^{h,*,M}_\sigma(x)$ denotes an optimal policy for $\tilde{X}^h_t(\sigma)$, namely, $\tilde{\pi}^{h,*,M}_\sigma(x) = \underset{\pi\in\mathrm{Lip}_M}{\arg\max} \tilde{V}^{h,\pi}_\sigma(x)$.

Table 1: Processes notations

Process	Deterministic	Stochastic
Continuous-time	$X_t(0)$	$X_t(\sigma)$
	$\check{X}_t(0)$	
		$\tilde{X}_t^h(\sigma)$
Discrete-time	$X_{t_\ell}(0)$	$X_{t_\ell}(\sigma)$
	$\check{X}_{t_\ell}(0)$	
		$\tilde{X}_{t_k}^h(\sigma)$

Table 2: Value function and policy notations

	Deterministic		Stochastic	
	policy π	optimal in Lip_M	policy π	optimal in Lip_M
Continuous	$V_0^\pi(x)$	$V_0^{*,M}(x), \pi_0^{*,M}(x)$	$V_\sigma^\pi(x)$	$V_\sigma^{*,M}(x), \pi_\sigma^{*,M}(x)$
Discrete	$\check{V}_0^{0,\pi}(x)$	$\check{V}_0^{0,*,M}(x), \check{\pi}_0^{0,*,M}(x)$		
Smoothing			$\tilde{V}_\sigma^{h,\pi}(x)$	$\tilde{V}_\sigma^{h,*,M}(x), \tilde{\pi}_\sigma^{h,*,M}(x)$

To keep track all the notations for better illustration and clear comparison, we list these notations in Tables 1 and 2. For simplicity, let $\tilde{r}(x) = r(x, \pi(x))$.

S1.1 Proof of Theorem 1

We first state the well-known Grönwall inequality as follows.

Lemma 1 (Grönwall's inequality (Kloeden and Platen, 1995))**.** *Let α, β and u be real-valued functions defined on an interval I. Assume that β and u are continuous and that the negative part of α is integrable on every closed and bounded subinterval* *of I. If β is non-negative and if u satisfies the integral inequality, $u(t) \leq \alpha(t) + \int_{t_0}^t \beta(s)u(s)\mathrm{d}s,\ \forall t, t_0 \in I$, then*

$$u(t) \leq \alpha(t) + \int_{t_0}^t \alpha(s)\beta(s)\exp\left(\int_s^t \beta(r)\mathrm{d}r\right)\mathrm{d}s, \quad t, t_0 \in I.$$

Lemma 2. *$b(x, \pi(x))$ and $\sigma(x, \pi(x))$ are Lipschitz continuous, that is,*

$$|b(x, \pi(x)) - b(y, \pi(y))| \leq L|x - y|, \qquad |\sigma(x, \pi(x)) - \sigma(y, \pi(y))| \leq L|x - y|,$$

where $L = (M + 1)\max(L_1, L_2)$.

Proof. Because of similarity, we provide arguments only for $b(x, \pi(x))$ as follows. Assumption (A2) implies

$$|b(x, \pi(x)) - b(y, \pi(y))| \leq L_1[|x - y| + |\pi(x) - \pi(y)|] \leq L_1|x - y| + L_1 M|x - y| \leq L|x - y|.$$

Proof of Theorem 1

Proof. By Lemma 2, we have that $\forall \pi \in \mathrm{Lip}_M$,

$$\left[\int_0^t b(X_s(\varepsilon\varsigma), \pi(X_s(\varepsilon\varsigma))) - b(X_s(0), \pi(X_s(0)))ds\right]^2 \leq \left[\int_0^t L\big|X_s(\varepsilon\varsigma) - X_s(0)\big|ds\right]^2$$
$$\leq L^2 t \int_0^t \big|X_s(\varepsilon\varsigma) - X_s(0)\big|^2 ds,$$

where the second inequality comes from Cauchy-Schwarz inequality. Note that

$$X_t(\varepsilon\varsigma) - X_t(0) = \int_0^t b(X_s(\varepsilon\varsigma), \pi(X_s(\varepsilon\varsigma))) - b(X_s(0), \pi(X_s(0)))ds + \varepsilon\int_0^t \sigma(X_s(\varepsilon\varsigma), \pi(X_s(\varepsilon\varsigma)))dW_s.$$

Denote by $\mathbb{E}^x$ the conditional expectation given initial value x. Then, we obtain

$$\begin{aligned}
\mathbb{E}^x|X_t(\varepsilon\varsigma)-X_t(0)|^2 \le\ & 2\mathbb{E}^x\left[\Big|\int_0^t b(X_s(\varepsilon\varsigma),\pi(X_s(\varepsilon\varsigma)))-b(X_s(0),\pi(X_s(0)))ds\Big|^2\right.\\
&\left.+\varepsilon^2\left(\int_0^t \sigma(X_s(\varepsilon\varsigma),\pi(X_s(\varepsilon\varsigma)))dW_s\right)^2\right]\\
=\ & 2\mathbb{E}^x\left[\Big|\int_0^t b(X_s(\varepsilon\varsigma),\pi(X_s(\varepsilon\varsigma)))-b(X_s(0),\pi(X_s(0)))ds\Big|^2\right]\\
&+2\varepsilon^2\mathbb{E}^x\left[\int_0^t \sigma^2(X_s(\varepsilon\varsigma),\pi(X_s(\varepsilon\varsigma)))ds\right]\\
\le\ & 2\mathbb{E}^x\left[L^2t\int_0^t\big|X_s(\varepsilon\varsigma)-X_s(0)\big|^2ds\right]+2\varepsilon^2\mathbb{E}^x\left[\int_0^t \sigma^2(X_s(\varepsilon\varsigma),\pi(X_s(\varepsilon\varsigma)))ds\right]\\
=\ & 2L^2t\int_0^t\mathbb{E}^x\big|X_s(\varepsilon\varsigma)-X_s(0)\big|^2ds+2\varepsilon^2\mathbb{E}^x\left[\int_0^t \sigma^2(X_s(\varepsilon\varsigma),\pi(X_s(\varepsilon\varsigma)))ds\right].
\end{aligned}$$

Recall that $g_\varepsilon(t,x)=\mathbb{E}^x\left[\int_0^t \varsigma^2(X_s(\varepsilon\varsigma),\pi(X_s(\varepsilon\varsigma)))ds\right]$. Apply Grönwall's lemma, we have

$$\begin{aligned}
\mathbb{E}^x|X_t(\varepsilon\varsigma)-X_t(0)|^2 &\le 2\varepsilon^2 g_\varepsilon(t,x)+\int_0^t 2\varepsilon^2 g_\varepsilon(s,x)2L^2t\exp\left(\int_s^t L^2t\,dr\right)ds\\
&=2\varepsilon^2 g_\varepsilon(t,x)+2\varepsilon^2 2L^2t\int_0^t g_\varepsilon(s,x)\exp\left\{2L^2t(t-s)\right\}ds.
\end{aligned}$$

Since $g_\varepsilon(t,x)$ is non-negative and non-decreasing in t, we conclude

$$\begin{aligned}
&\sup_{0\le t\le T,\pi\in \mathrm{Lip}_M}\mathbb{E}^x|X_t(\varepsilon\varsigma)-X_t(0)|^2\\
&\le 2\varepsilon^2 g_\varepsilon(T,x)+2\varepsilon^2 2L^2T\int_0^T g_\varepsilon(s,x)\exp\left\{2L^2T(T-s)\right\}ds=\varepsilon^2M_T^x.
\end{aligned}$$

Finally, the boundedness of $g_\varepsilon(t,x)$ and M_T^x follows from their expressions and the boundedness of ζ.

□

S1.2 Proof of Theorem 2

Lemma 3. *As $\varepsilon\to 0$, we have $V_{\varepsilon\varsigma}^{\pi}(x)\to V_0^{\pi}(x)$ uniformly over $\pi\in Lip_M$.*

Proof. First we show that $\tilde{r}$ is Lipschitz continuous. By Assumption (A4) we have for $\pi \in \mathrm{Lip}_M$,

$$
\begin{aligned}
|\tilde{r}(x) - \tilde{r}(y)| =& |r(x, \pi(x)) - r(y, \pi(y))| \\
\leq& L_r \left[|x - y| + |\pi(x) - \pi(y)| \right] \\
\leq& L_r (M+1) |x - y|.
\end{aligned}
$$

Note that $\forall \eta > 0$, $\exists T$, s.t. $e^{-\beta T} \leq \frac{\eta}{4M}$. Next we will prove that for any $\pi \in \mathrm{Lip}_M$, there exists $\tau > 0$ such that $\forall \varepsilon \in (0, \tau)$, $|V^{\pi}_{\varepsilon\varsigma}(x) - V^{\pi}_0(x)| \leq \eta$. Indeed, with $\sigma = \varepsilon\zeta$ we have

$$
\begin{aligned}
&|V^{\pi}_{\sigma}(x) - V^{\pi}_0(x)| = \left| \mathbb{E}\left[\int_0^{\infty} e^{-\beta t} [\tilde{r}(X_t(\sigma)) - \tilde{r}(X_t(0))] dt \right| X_0(\sigma) = x, X_0(0) = x \right] \Bigg| \\
\leq& \left| \int_0^T e^{-\beta t} \mathbb{E}\left[\tilde{r}(X_t(\sigma)) - \tilde{r}(X_t(0)) \right| X_0(\sigma) = x, X_0(0) = x \right] dt \Bigg| \\
&+ \left| \int_T^{\infty} e^{-\beta t} \mathbb{E}\left[\tilde{r}(X_t(\sigma)) - \tilde{r}(X_t(0)) \right| X_0(\sigma) = x, X_0(0) = x \right] dt \Bigg| \\
\leq& \int_0^T e^{-\beta t} \mathbb{E}\left[|\tilde{r}(X_t(\sigma)) - \tilde{r}(X_t(0))| \,\middle|\, X_0(\sigma) = x, X_0(0) = x \right] dt + 2M e^{-\beta T} \\
\leq& L_r (M+1) \int_0^T e^{-\beta t} \mathbb{E}\left[|X_t(\sigma) - X_t(0)| \,\middle|\, X_0(\sigma) = x, X_0(0) = x \right] dt + \frac{\eta}{2} \\
\leq& L_r (M+1) \varepsilon \sqrt{M^x_T} + \frac{\eta}{2},
\end{aligned}
$$

where the equalities in the fifth and sixth lines of the array above are due to, respectively, the Lipschitz continuity of $\tilde{r}$ and the Cauchy-Schwarz inequality together with that Theorem 1 indicates that for $\pi \in \mathrm{Lip}_M$,

$$
\sup_{0 \leq t \leq T} \mathbb{E}\left[|X_t(\varepsilon\zeta) - X_t(0)|^2 \,\middle|\, X_0(\varepsilon\zeta) = x, X_0(0) = x \right] \leq \varepsilon^2 M^x_T.
$$

Note that for any given T, M^x_T is bounded over all x and ε. Choose τ so that with the selected T and for all x and ϵ, we have

$$
L_r (M+1) \tau \sqrt{M^x_T} \leq \frac{\eta}{2}.
$$

Hence, we conclude that for $\varepsilon \in (0, \tau)$,

$$|V^{\pi}_{\varepsilon\varsigma}(x) - V^{\pi}_0(x)| \leq L_r(M+1)\varepsilon\sqrt{M^x_T} + \frac{\eta}{2} \leq \frac{\eta}{2} + \frac{\eta}{2} = \quad \eta.$$

That is, we show that as $\varepsilon \to 0$, $V^{\pi}_{\varepsilon\varsigma}(x)$ converges to $V^{\pi}_0(x)$ uniformly over $\pi \in \mathrm{Lip}_M$. □

Lemma 4. *We have $V^{*,M}_{\varepsilon\varsigma}(x) \to V^{*,M}_0(x)$ as $\varepsilon \to 0$.*

Proof. If $V^{*,M}_{\varepsilon\varsigma}(x) \geq V^{*,M}_0(x)$,

$$\begin{aligned}\left|V^{*,M}_{\varepsilon\varsigma}(x) - V^{*,M}_0(x)\right| &= \sup_{\pi\in\mathrm{Lip}_M} V^{\pi}_{\varepsilon\varsigma}(x) - \sup_{\pi\in\mathrm{Lip}_M} V^{\pi}_0(x) \\ &\leq \sup_{\pi\in\mathrm{Lip}_M} \left(V^{\pi}_{\varepsilon\varsigma}(x) - V^{\pi}_0(x)\right) \\ &\leq \sup_{\pi\in\mathrm{Lip}_M} \left|\left(V^{\pi}_{\varepsilon\varsigma}(x) - V^{\pi}_0(x)\right)\right|.\end{aligned}$$

Similarly, if $V^{*,M}_{\varepsilon\varsigma}(x) \leq V^{*,M}(x)$, we can show

$$\left|V^{*,M}_{\varepsilon\varsigma}(x) - V^{*,M}_0(x)\right| \leq \sup_{\pi\in\mathrm{Lip}_M} \left|\left(V^{\pi}_{\varepsilon\varsigma}(x) - V^{\pi}_0(x)\right)\right|.$$

Lemma 3 shows that

$$\lim_{\varepsilon\to 0} \sup_{\pi\in\mathrm{Lip}_M} |V^{\pi}_{\varepsilon\varsigma}(x) - V^{\pi}_0(x)| = 0,$$

and therefore, we have $\left|V^{*,M}_{\varepsilon\varsigma}(x) - V^{*,M}_0(x)\right| \to 0$ as $\varepsilon \to 0$. □

Lemma 5. *$\pi^{*,M}_{\varepsilon\varsigma}(x) \to \pi^{*,M}_0(x)$ as $\varepsilon \to 0$.*

Proof. First we show that $V^{\pi}_{\sigma}(x)$ and $V^{\pi}_0(x)$ are Lipschitz continuous in π as follows. For any π_1 and π_2, define $\|\pi_1 - \pi_2\| = \max_x |\pi_1(x) - \pi_2(x)|$. Then we can show

$$|V^{\pi_1}_{\sigma}(x) - V^{\pi_2}_{\sigma}(x)| \leq L_v\|\pi_1 - \pi_2\|, \;\; |V^{\pi_1}_0(x) - V^{\pi_2}_0(x)| \leq L_v\|\pi_1 - \pi_2\|,$$

where L_v is a constant. Indeed, by Assumption (A4) we have that $r(x, a)$ is Lipschitz continuous; thus, we obtain

$$|r(X_t(\sigma), \pi_1(X_t(\sigma)) - r(X_t(\sigma), \pi_2(X_t(\sigma))| \leq L_r|\pi_1(X_t(\sigma)) - \pi_2(X_t(\sigma))| \leq L_r\|\pi_1 - \pi_2\|.$$

and

$$|V_\sigma^{\pi_1}(x) - V_\sigma^{\pi_2}(x)| \le \int_0^\infty e^{-\beta t} L_r \|\pi_1 - \pi_2\| dt = L_v \|\pi_1 - \pi_2\|,$$

where $L_v = L_r/\beta$. The same argument can show the Lipschitz continuity of V_0^π.

Assume that as $\varepsilon \to 0$, $\pi_{\varepsilon\varsigma}^{*,M}(x)$ does not convergence to any optimal policy $\pi_0^{*,M}(x)$. Since these policies belong to $\pi \in \mathrm{Lip}_M$, which is compact, there must exist a suboptimal policy $\pi_*(x) \neq \pi_0^{*,M}(x)$ and an optimal policy sequence $\pi_{\varepsilon_\ell \varsigma}^{*,M}(x)$, $\ell \ge 1$, such that as $\ell \to \infty$, $\varepsilon_\ell \to 0$, and $\pi_{\varepsilon_\ell\varsigma}^{*,M}(x) \to \pi_*(x)$. This convergence result together with the Lipschitz continuity of V_0^π in π lead us to that as $\ell \to \infty$,

$$V_0^{\pi_*}(x) = \lim_{\ell\to\infty} V_0^{\pi_{\varepsilon_\ell\varsigma}^{*,M}}(x) = \lim_{\ell\to\infty} V_{\varepsilon_\ell\zeta}^{\pi_{\varepsilon_\ell\varsigma}^{*,M}}(x) = \lim_{\ell\to\infty} V_{\varepsilon_\ell\zeta}^{*,M}(x) = V_0^{*,M}(x),$$

where the second, third, and fourth equalities are due to Lemma 3, the definition of $V_\sigma^{*,M}(x)$, and Lemma 4, respectively. The above equality shows that $\pi_*(x)$ is an optimal policy, which leads to a contraction that $\pi_*(x)$ is not optimal. □

Remark 1. Equations (3.11) and (3.4) indicate that the optimal policies for deterministic and stochastic cases obey the following relationships,

$$\pi_0^{*,M}(x) = \underset{\pi\in\mathrm{Lip}_M}{\arg\max} \left\{ r(x,\pi(x)) + b(x,\pi(x))^\dagger \frac{\partial V_0^{*,M}(x)}{\partial x} \right\},$$

$$\pi_\sigma^{*,M}(x) = \underset{\pi\in\mathrm{Lip}_M}{\arg\max} \left\{ r(x,\pi(x)) + b(x,\pi(x))^\dagger \frac{\partial V_\sigma^{*,M}(x)}{\partial x} + \frac{1}{2} tr\left(\sigma(x)\sigma^\dagger(x) \frac{\partial^2 V_\sigma^{*,M}(x)}{\partial x^2} \right) \right\}.$$

As $\sigma = \varepsilon\varsigma$ and $\varepsilon \to 0$, intuitively it is easy to see that $\pi_{\varepsilon\varsigma}^{*,M}(x)$ is close to $\pi_0^{*,M}(x)$.

S1.3 Proof of Theorem 3

For simplicity we use the expectation, variance, and covariance notations in the proof arguments below to stand for their conditional counterparts given $X_0(\sigma)$ and $X_0(0)$.

From the definition of Υ_1^{k-1} in equations (4.3)-(4.5) and the definition of $\Upsilon_{1,\sigma}^{k-1,i}$ in equations (4.6)-(4.9), we have

$$|\Upsilon_{1,\sigma}^{k-1,i}-\Upsilon_1^{k-1}|=\frac{T}{N}\left|b\left(\check{X}_{t_{k-1}+i}(\sigma),\pi(\tilde{X}^h_{t_{k-1}}(\sigma))\right)+\sigma\check{W}_i(t_{k-1})-b\left(X_{t_{k-1}}(0),\pi(X_{t_{k-1}}(0))\right)\right|$$
$$\leq\frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\sigma|\check{W}_i(t_{k-1})|\right).$$

Similarly, we can show

$$|\Upsilon_{2,\sigma}^{k-1,i}-\Upsilon_2^{k-1}|=\frac{T}{N}\left|b\left(\check{X}_{t_{k-1}+i}(\sigma)+\Upsilon_{1,\sigma}^{k-1,i}/2,\pi(\tilde{X}^h_{t_{k-1}}(\sigma))\right)+\sigma\check{W}_i(t_{k-1}+\frac{T}{2N})\right.$$
$$\left.-b\left(X_{t_{k-1}}(0)+\Upsilon_1^{k-1}/2,\pi(X_{t_{k-1}}(0))\right)\right|$$
$$\leq\frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)+\Upsilon_{1,\sigma}^{k-1,i}/2-X_{t_{k-1}}(0)-\Upsilon_1^{k-1}/2|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right)$$
$$\leq\frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{L}{2}|\Upsilon_{1,\sigma}^{k-1,i}-\Upsilon_1^{k-1}|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right)$$
$$\leq\frac{T}{N}\left(\left(L+\frac{TL^2}{2N}\right)|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{TL}{2N}\sigma|\check{W}_i(t_{k-1})|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right),$$

$$|\Upsilon_{3,\sigma}^{k-1,i}-\Upsilon_3^{k-1}|=\frac{T}{N}\left|b\left(\check{X}_{t_{k-1}+i}(\sigma)+\Upsilon_{2,\sigma}^{k-1,i}/2,\pi(\tilde{X}^h_{t_{k-1}}(\sigma))\right)+\sigma\check{W}_i(t_{k-1}+\frac{T}{2N})\right.$$
$$\left.-b\left(X_{t_{k-1}}(0)+\Upsilon_2^{k-1}/2,\pi(X_{t_{k-1}}(0))\right)\right|$$
$$\leq\frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)+\Upsilon_{2,\sigma}^{k-1,i}/2-X_{t_{k-1}}(0)-\Upsilon_2^{k-1}/2|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right)$$
$$\leq\frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{L}{2}|\Upsilon_{2,\sigma}^{k-1,i}-\Upsilon_2^{k-1}|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right)$$
$$\leq\frac{T}{N}\left(\left(L+\frac{TL^2}{2N}+\frac{T^2L^3}{4N^2}\right)|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{T^2L^2}{4N^2}\sigma|\check{W}_i(t_{k-1})|\right.$$
$$\left.+\frac{TL}{2N}\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right),$$

and

$$
\begin{aligned}
&|\Upsilon_{4,\sigma}^{k-1,i}-\Upsilon_4^{k-1}| = \frac{T}{N}\left|b\left(\check{X}_{t_{k-1}+i}(\sigma)+\Upsilon_{3,\sigma}^{k-1,i},\pi(\tilde{X}_{t_{k-1}}^h(\sigma))\right)+\sigma\check{W}_i(t_{k-1}+\frac{T}{N})\right.\\
&\left.-b\left(X_{t_{k-1}}(0)+\Upsilon_3^{k-1},\pi(X_{t_{k-1}}(0))\right)\right|\\
&\leq \frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)+\Upsilon_{3,\sigma}^{k-1,i}-X_{t_{k-1}}(0)-\Upsilon_3^{k-1}|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{N})|\right)\\
&\leq \frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+L|\Upsilon_{3,\sigma}^{k-1,i}-\Upsilon_3^{k-1}|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{N})|\right)\\
&\leq \frac{T}{N}\left(\left(L+\frac{TL^2}{N}+\frac{T^2L^3}{2N^2}+\frac{T^3L^4}{4N^3}\right)|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{T^3L^3}{4N^3}\sigma|\check{W}_i(t_{k-1})|\right.\\
&\left.+\frac{T^2L^2}{2N^2}\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|+\frac{TL}{N}\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{N})|\right).
\end{aligned}
$$

Therefore, we conclude

$$
\begin{aligned}
\operatorname{Cov}\left(\tilde{X}_{t_{k-1}}^h(\sigma),\Upsilon_{1,\sigma}^{k-1,i}\right) &= \operatorname{Cov}\left(\tilde{X}_{t_{k-1}}^h(\sigma),\frac{T}{N}(b(\check{X}_{t_{k-1}+i}(\sigma),\pi(\tilde{X}_{t_{k-1}}^h(\sigma)))+\sigma\check{W}_i(t_{k-1}))\right)\\
&= \frac{T}{N}\operatorname{Cov}\left(\tilde{X}_{t_{k-1}}^h(\sigma),b(\check{X}_{t_{k-1}+i}(\sigma),\pi(\tilde{X}_{t_{k-1}}^h(\sigma)))\right)\\
&\leq \frac{T}{N}\left\{\operatorname{Var}\left(\tilde{X}_{t_{k-1}}^h(\sigma)\right)\operatorname{Var}\left(b(\check{X}_{t_{k-1}+i}(\sigma),\pi(\tilde{X}_{t_{k-1}}^h(\sigma)))\right)\right\}^{1/2}\\
&= \mathcal{O}\left(\frac{1}{N}\right).
\end{aligned}
$$

Similar arguments lead us to conclude that $\operatorname{Cov}(\tilde{X}_{t_{k-1}}^h(\sigma),\Upsilon_{j,\sigma}^{k-1,i})=\mathcal{O}(\frac{1}{N})$ for $j=2,3,4$.

Hence, we have

$$
\begin{aligned}
&\operatorname{Cov}\left(\tilde{X}_{t_{k-1}}^h(\sigma)-X_{t_{k-1}}(0),\right.\\
&\quad\left.\frac{1}{6m}\sum_{i=0}^{m-1}\left((\Upsilon_{1,\sigma}^{k-1,i}-\Upsilon_1^{k-1})+2(\Upsilon_{2,\sigma}^{k-1,i}-\Upsilon_2^{k-1})+2(\Upsilon_{3,\sigma}^{k-1,i}-\Upsilon_3^{k-1})+(\Upsilon_{4,\sigma}^{k-1,i}-\Upsilon_4^{k-1})\right)\right)\\
&=\frac{1}{6m}\sum_{i=0}^{m-1}\operatorname{Cov}\left(\tilde{X}_{t_{k-1}}^h(\sigma),\Upsilon_{1,\sigma}^{k-1,i}+2\Upsilon_{2,\sigma}^{k-1,i}+2\Upsilon_{3,\sigma}^{k-1,i}+\Upsilon_{4,\sigma}^{k-1,i}\right)\\
&=\mathcal{O}\left(\frac{1}{N}\right),
\end{aligned}
$$

and

$$\mathbb{E}[\tilde{X}^h_{t_k}(\sigma) - X_{t_k}(0)]^2 = \mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0) + \right.$$
$$\left.\frac{1}{6m}\sum_{i=0}^{m-1}\Big((\Upsilon^{k-1,i}_{1,\sigma} - \Upsilon^{k-1}_1) + 2(\Upsilon^{k-1,i}_{2,\sigma} - \Upsilon^{k-1}_2) + 2(\Upsilon^{k-1,i}_{3,\sigma} - \Upsilon^{k-1}_3) + (\Upsilon^{k-1,i}_{4,\sigma} - \Upsilon^{k-1}_4)\Big)\right]^2$$
$$= \mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + 2\mathrm{Cov}\Big(\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0),$$
$$\frac{1}{6m}\sum_{i=0}^{m-1}\Big((\Upsilon^{k-1,i}_{1,\sigma} - \Upsilon^{k-1}_1) + 2(\Upsilon^{k-1,i}_{2,\sigma} - \Upsilon^{k-1}_2) + 2(\Upsilon^{k-1,i}_{3,\sigma} - \Upsilon^{k-1}_3) + (\Upsilon^{k-1,i}_{4,\sigma} - \Upsilon^{k-1}_4)\Big)\Big)$$
$$+ \mathbb{E}\left[\frac{1}{6m}\sum_{i=0}^{m-1}\Big((\Upsilon^{k-1,i}_{1,\sigma} - \Upsilon^{k-1}_1) + 2(\Upsilon^{k-1,i}_{2,\sigma} - \Upsilon^{k-1}_2) + 2(\Upsilon^{k-1,i}_{3,\sigma} - \Upsilon^{k-1}_3) + (\Upsilon^{k-1,i}_{4,\sigma} - \Upsilon^{k-1}_4)\Big)\right]^2$$
$$\leq \mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N}) +$$
$$\frac{1}{36m^2}\mathbb{E}\left[\sum_{i=0}^{m-1}\Big((\Upsilon^{k-1,i}_{1,\sigma} - \Upsilon^{k-1}_1) + 2(\Upsilon^{k-1,i}_{2,\sigma} - \Upsilon^{k-1}_2) + 2(\Upsilon^{k-1,i}_{3,\sigma} - \Upsilon^{k-1}_3) + (\Upsilon^{k-1,i}_{4,\sigma} - \Upsilon^{k-1}_4)\Big)\right]^2$$
$$\leq \mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N}) +$$
$$\frac{1}{36m^2}\mathbb{E}\left[\sum_{i=0}^{m-1}\Big(|\Upsilon^{k-1,i}_{1,\sigma} - \Upsilon^{k-1}_1| + 2|\Upsilon^{k-1,i}_{2,\sigma} - \Upsilon^{k-1}_2| + 2|\Upsilon^{k-1,i}_{3,\sigma} - \Upsilon^{k-1}_3| + |\Upsilon^{k-1,i}_{4,\sigma} - \Upsilon^{k-1}_4|\Big)\right]^2$$
$$\equiv \mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N}) + \aleph,$$

where

$$
\begin{aligned}
\aleph \leq & \frac{1}{36m^2}\mathbb{E}\left[\sum_{i=0}^{m-1}\left\{\frac{T}{N}\left(L|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\sigma|\check{W}_i(t_{k-1})|\right)\right.\right.\\
&+\frac{2T}{N}\left(\left(L+\frac{TL^2}{2N}\right)|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{TL}{2N}\sigma|\check{W}_i(t_{k-1})|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right)\\
&+\frac{2T}{N}\left(\left(L+\frac{TL^2}{2N}+\frac{T^2L^3}{4N^2}\right)|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{T^2L^2}{4N^2}\sigma|\check{W}_i(t_{k-1})|\right.\\
&\left.+\frac{TL}{2N}\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|\right)\\
&+\frac{T}{N}\left(\left(L+\frac{TL^2}{N}+\frac{T^2L^3}{2N^2}+\frac{T^3L^4}{4N^3}\right)|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|+\frac{T^3L^3}{4N^3}\sigma|\check{W}_i(t_{k-1})|\right.\\
&\left.\left.\left.+\frac{T^2L^2}{2N^2}\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|+\frac{TL}{N}\sigma|\check{W}_i(t_{k-1}+\frac{T}{2N})|+\sigma|\check{W}_i(t_{k-1}+\frac{T}{N})|\right)\right\}\right]^2\\
&=\frac{1}{36m^2}\mathbb{E}\left[\sum_{i=0}^{m-1}\frac{T}{N}\left(6L+\frac{3TL^2}{N}+\frac{T^2L^3}{N^2}+\frac{T^3L^4}{4N^3}\right)|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|\right]^2\\
&\qquad+\sum_{i=0}^{m-1}\mathcal{O}(\frac{1}{m^2N^2})\\
&=\mathcal{O}(\frac{1}{m^2N^2})\mathbb{E}\left[\sum_{i=0}^{m-1}|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|\right]^2+\mathcal{O}(\frac{1}{mN^2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=0}^{m-1}|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|\right]^2\leq\mathbb{E}\left(m\max_{0\leq i\leq m-1}|\check{X}_{t_{k-1}+i}(\sigma)-X_{t_{k-1}}(0)|\right)^2\\
&\leq 2m^2\left[\mathbb{E}\left(\max_{0\leq i\leq m-1}|\check{X}_{t_{k-1}+i}(\sigma)-\tilde{X}^h_{t_{k-1}}(\sigma)|\right)^2+\mathbb{E}\left(|\tilde{X}^h_{t_{k-1}}(\sigma)-X_{t_{k-1}}(0)|\right)^2\right]\\
&\leq 2m^2\mathbb{E}\left(\max_{0\leq i\leq m-1}|\sum_{j=0}^{i-1}\frac{1}{6}(\Gamma^{k-1,j}_{1,\sigma}+2\Gamma^{k-1,j}_{2,\sigma}+2\Gamma^{k-1,j}_{3,\sigma}+\Gamma^{k-1,j}_{4,\sigma})|\right)^2\\
&+2m^2\mathbb{E}\left(|\tilde{X}^h_{t_{k-1}}(\sigma)-X_{t_{k-1}}(0)|\right)^2\\
&\leq 2m^2\mathbb{E}\left(|\sum_{i=0}^{m-2}\frac{1}{6}(\Gamma^{k-1,i}_{1,\sigma}+2\Gamma^{k-1,i}_{2,\sigma}+2\Gamma^{k-1,i}_{3,\sigma}+\Gamma^{k-1,i}_{4,\sigma})|\right)^2+2m^2\mathbb{E}\left(|\tilde{X}^h_{t_{k-1}}(\sigma)-X_{t_{k-1}}(0)|\right)^2\\
&=\mathcal{O}(\frac{m^4}{N^2})+2m^2\mathbb{E}\left(|\tilde{X}^h_{t_{k-1}}(\sigma)-X_{t_{k-1}}(0)|\right)^2.
\end{aligned}
$$

Putting together above results we arrive at

$$\begin{aligned}
&\mathbb{E}\left[\tilde{X}^h_{t_k}(\sigma) - X_{t_k}(0)\right]^2 \leq \mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N}) + \\
&\mathcal{O}(\frac{1}{m^2N^2})\mathbb{E}\left[\sum_{i=0}^{m-1} |\check{X}_{t_{k-1}+i}(\sigma) - X_{t_{k-1}}(0)|\right]^2 + \mathcal{O}(\frac{1}{mN^2}) \\
&\leq \mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N}) + \mathcal{O}(\frac{1}{mN^2}) \\
&\mathcal{O}(\frac{1}{m^2N^2})\left\{\mathcal{O}(\frac{m^4}{N^2}) + 2m^2\mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2\right\} \\
&= \left[1 + \mathcal{O}(\frac{1}{N^2})\right]\mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N} + \frac{1}{mN^2} + \frac{m^2}{N^4}) \\
&= \left[1 + \mathcal{O}(\frac{1}{N^2})\right]\mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N}).
\end{aligned}$$

An application of the induction argument leads to

$$\begin{aligned}
&\mathbb{E}\left[\tilde{X}^h_{t_k}(\sigma) - X_{t_k}(0)\right]^2 \leq \left[1 + \mathcal{O}(\frac{1}{N^2})\right]\mathbb{E}\left[\tilde{X}^h_{t_{k-1}}(\sigma) - X_{t_{k-1}}(0)\right]^2 + \mathcal{O}(\frac{1}{N}) \\
&\leq \left[1 + \mathcal{O}(\frac{1}{N^2})\right]\left\{\left[1 + \mathcal{O}(\frac{1}{N^2})\right]\mathbb{E}\left[\tilde{X}^h_{t_{k-2}}(\sigma) - X_{t_{k-2}}(0)\right]^2 + \mathcal{O}(\frac{1}{N})\right\} + \mathcal{O}(\frac{1}{N}) \\
&= \left[1 + \mathcal{O}(\frac{1}{N^2})\right]^2\mathbb{E}\left[\tilde{X}^h_{t_{k-2}}(\sigma) - X_{t_{k-2}}(0)\right]^2 + \left[1 + \mathcal{O}(\frac{1}{N^2})\right]\mathcal{O}(\frac{1}{N}) + \mathcal{O}(\frac{1}{N}) \\
&\leq \left[1 + \mathcal{O}(\frac{1}{N^2})\right]^k\mathbb{E}\left[\tilde{X}^h_{t_0}(\sigma) - X_{t_0}(0)\right]^2 + \sum_{i=0}^{k-1}\left[1 + \mathcal{O}(\frac{1}{N^2})\right]^i\mathcal{O}(\frac{1}{N}) \\
&\leq C\left[1 + \mathcal{O}(\frac{1}{N^2})\right]^k + \sum_{i=0}^{k-1}(1 + \mathcal{O}(\frac{1}{N^2}))^i\mathcal{O}(\frac{1}{N}) \\
&= C\left[1 + \mathcal{O}(\frac{1}{N^2})\right]^k + \frac{1 - (1 + \mathcal{O}(\frac{1}{N^2}))^k}{1 - (1 + \mathcal{O}(\frac{1}{N^2}))}\mathcal{O}(\frac{1}{N}) \\
&= \mathcal{O}(N^2)\mathcal{O}(\frac{k}{N^2})\mathcal{O}(\frac{1}{N}) = \mathcal{O}(\frac{k}{N}) = \mathcal{O}(\frac{1}{m}),
\end{aligned}$$

where the last equality uses the facts that $1 \leq k \leq n = \frac{N}{m}$.

S1.4 Proof of Theorem 4

The proof idea is to combine Theorem 3 and the arguments for proving Theorem 2.

Lemma 6. *We have*

$$\sup_{\pi \in Lip_M} |\tilde{V}^{h,\pi}_{\sigma}(x) - \check{V}^{0,\pi}_0(x)| \to 0,$$

as $h \to 0$*,* $m \to \infty$ *and* $mh \to 0$*.*

Proof. Direct calculations show that for any $\eta > 0$,

$$\begin{aligned}
&|\tilde{V}^{h,\pi}_{\sigma}(x) - \check{V}^{0,\pi}_0(x)| = \left|\mathbb{E}\left[\int_0^\infty e^{-\beta t}\left[\tilde{r}(\tilde{X}^h_t(\sigma)) - \tilde{r}(\check{X}_t(0))\right] dt \middle| \tilde{X}^h_0(\sigma) = x, \check{X}_0(0) = x\right]\right| \\
&\leq \mathbb{E}\left[\int_0^T e^{-\beta t}\left|\tilde{r}(\tilde{X}^h_t(\sigma)) - \tilde{r}(\check{X}_t(0))\right| dt \middle| \tilde{X}^h_0(\sigma) = x, \check{X}_0(0) = x\right] + \frac{\eta}{2} \\
&\leq L_r(M+1)\mathbb{E}\left[\int_0^T e^{-\beta t}|\tilde{X}^h_t(\sigma) - \check{X}_t(0)| dt \middle| \tilde{X}^h_0(\sigma) = x, \check{X}_0(0) = x\right] + \frac{\eta}{2} \\
&\leq L_r(M+1)C \sup_{1 \leq n \leq N} \mathbb{E}\left[|\tilde{X}^h_{t_n}(\sigma) - X_{t_n}(0)| \middle| X_0(\sigma) = x, X_0(0) = x\right] \int_0^T e^{-\beta T} dt + \frac{\eta}{2} \\
&\leq L_r(M+1)Cm^{-1/2} + \frac{\eta}{2},
\end{aligned}$$

where the last inequality is from Theorem 3, and the inequalities in the third and second lines of the above array are due to, respectively, the Lipschitz continuity of $\tilde{r}(x)$ and the facts that $\tilde{r}(x)$ is bounded, and $\exists T$ s.t.

$$\mathbb{E}\left[\int_T^\infty e^{-\beta t}\left|\tilde{r}(\tilde{X}^h_t(\sigma)) - \tilde{r}(\check{X}_t(0))\right| dt \middle| \tilde{X}^h_0(\sigma) = x, \check{X}_0(0) = x\right] \leq C\int_T^\infty e^{-\beta t} dt \leq \frac{\eta}{2}.$$

Finally we complete the proof by letting $h \to 0$ and then $\eta \to 0$. □

Lemma 7. *We have*

$$\sup_{\pi \in Lip_M} |\check{V}^{0,\pi}_0(x) - V^{\pi}_0(x)| \to 0,$$

as $h \to 0$*,* $m \to \infty$ *and* $mh \to 0$*.*

Proof. Recall that

$$V^{\pi}_0(x) - \check{V}^{0,\pi}_0(x) = \int_0^\infty e^{-\beta t}\left[\tilde{r}(X_t(0)) - \tilde{r}(\check{X}_t(0))\right] dt \text{ given } \check{X}_0(0) = x \text{ and } X_0(0) = x.$$

Note that $\forall \eta > 0$, $\exists T$, s.t.

$$\int_T^\infty e^{-\beta t} \left|\tilde{r}(X_t(0)) - \tilde{r}(\check{X}_t(0))\right| dt \le C \int_T^\infty e^{-\beta t} dt \le \frac{\eta}{2}.$$

Hence, we conclude

$$\begin{aligned}
&|V_0^\pi(x) - \check{V}_0^{0,\pi}(x)| \le \int_0^T e^{-\beta t} \left|\tilde{r}(X_t(0)) - \tilde{r}(\check{X}_t(0))\right| dt + \frac{\eta}{2} \\
&\le L_r(M+1) \int_0^T e^{-\beta t} \left|X_t(0) - \check{X}_t(0)\right| dt + \frac{\eta}{2} \\
&\le L_r(M+1) \int_0^T e^{-\beta t} \left[|X_t(0) - X_{t^n}(0)| + |X_{t^n}(0) - \check{X}_t(0)|\right] dt + \frac{\eta}{2} \\
&\le L_r(M+1) \sup_{0\le t\le T} |X_t(0) - X_{t^n}(0)| \int_0^T e^{\beta t} dt + L_r(M+1) \int_0^T e^{-\beta t} |X_{t^n}(0) - \check{X}_t(0)| dt + \frac{\eta}{2} \\
&\le L_r(M+1) \sup_{0\le t\le T} |X_t(0) - X_{t^n}(0)| + L_r(M+1) \int_0^T e^{-\beta t} C h^4 dt + \frac{\eta}{2} \\
&\le L_r(M+1) \sup_{0\le t\le T} |b(X_t(0)), r(X_t(0)))| h + L_r(M+1) C h^4 + \frac{\eta}{2},
\end{aligned}$$

which can be arbitrarily small, where for each t, t^n is chosen to be one of the discrete points $t_k = kT/n$ such that $t^n \in [t, t+h)$, and the inequalities in the fifth and sixth lines of the above array are due to the fact that the RK scheme has approximation errors of the 4th order, and an application of the mean value theorem, respectively. Finally we complete the proof by letting $h \to 0$ and then $\eta \to 0$. □

The following proposition shows the first result (i.e. the value function convergence) in Theorem 4.

Proposition 1. *We have that as $h \to 0$, $k \to \infty$ and $mh \to 0$,*

$$\sup_{\pi \in Lip_M} |\tilde{V}_\sigma^{h,\pi}(x) - V_0^\pi(x)| \to 0.$$

Proof. The proposition immediately follows from Lemmas 6 and 7. □

The following proposition establishes the second result (i.e. the optimal value function convergence) in Theorem 4.

Proposition 2. *We have*

$$\tilde{V}_{\sigma}^{h,*,M}(x) \to V_0^{*,M}(x),$$

as $h \to 0$, $m \to \infty$ *and* $mh \to 0$.

Proof. We prove the proposition by using Proposition 1 and the same arguments for proving Lemma 4. □

The third result (i.e. the optimal policy convergence) is given by the following proposition.

Proposition 3. *We have* $\tilde{\pi}_{\sigma}^{h,*,M}(x) \to \pi_0^{*,M}(x)$ *as* $h \to 0$, $m \to \infty$ *and* $mh \to 0$.

Proof. We can prove the proposition by using Propositions 1 and 2 and the same arguments for proving Lemma 5. □