Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Regression on particle systems connected to mean-field SDEs with applications

Denis Belomestny[1], John G. M. Schoenmakers[2]

submitted: December 20, 2017

[1] Duisburg-Essen University
Thea-Leymann-Str. 9
45127 Essen
Germany
E-Mail: denis.belomestny@uni-due.de

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: john.schoenmakers@wias-berlin.de

No. 2464
Berlin 2017

2010 *Mathematics Subject Classification.* 60H10, 60K35, 62L15.

Key words and phrases. McKean-Vlasov equations, optimal stopping, regression, Bellman principle.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Regression on particle systems connected to mean-field SDEs with applications

Denis Belomestny, John G. M. Schoenmakers

Abstract

In this note we consider the problem of using regression on interacting particles to compute conditional expectations for McKean-Vlasov SDEs. We prove general result on convergence of linear regression algorithms and establish the corresponding rates of convergence. Application to optimal stopping and variance reduction are considered.

1 Introduction

McKean-Vlasov or mean-field SDEs are a class of stochastic differential equations where the drift and diffusion depend on the current position along the path and on the current distribution. They were derived to describe propagation of chaos in a system of particles that interact only by their empirical mean in the limit of large number of particles.

Let $[0,T]$ be a finite time interval and $(\Omega,\mathcal{F},\mathrm{P})$ be a complete probability space, where a standard m-dimensional Brownian motion W is defined. We consider a class of McKean-Vlasov SDEs, i.e., stochastic differential equations (SDE), whose drift and diffusion coefficients may depend on the current distribution of the process, of the form:

$$\left\{\begin{array}{ll} X_t & = \xi + \int_0^t \int_{\mathbb{R}^d} a(X_s,y)\mu_s(dy)ds + \int_0^t \int_{\mathbb{R}^d} b(X_s,y)\mu_s(dy)dW_s \\ \mu_t & = \mathrm{Law}(X_t), \quad t \geq 0, \quad X_0 \sim \mu_0 \end{array}\right. \tag{1}$$

where μ_0 is a distribution in $\mathbb{R}^d$, $a:\ \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ and $b:\ \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$.

A popular way of simulating the MVSDE (1) is to sample from the N-particle interacting diffusion model, or particle system for short,

$$X_t^{i,N} = \xi^i + \frac{1}{N}\sum_{j=1}^N \int_0^t a(X_s^{i,N}, X_s^{j,N})\, ds + \frac{1}{N}\sum_{j=1}^N \int_0^t b(X_s^{i,N}, X_s^{j,N})\, dW_s^i \tag{2}$$

for $i=1,\ldots,N$, where ξ^i, $i=1,\ldots,N$, are i.i.d copies of a r.v. ξ, distributed according the law μ_0, and W^i, $i=1,...,N$, are independent copies of W. Due to [1] one has that

$$\left\| \sup_{0\leq r\leq T} \left| X_r^{\cdot,N} - X_r^{\cdot} \right| \right\|_p \leq C_p N^{-1/2}. \tag{3}$$

In reality of course of course, N-dimensional SDE the system (2) cannot be exactly solved either and one has to approximate it by some suitable numerical integration scheme such as the Euler method,

DOI 10.20347/WIAS.PREPRINT.2464

leading to a next approximation $\mathbf{X}_t^{N,h} = \left(X_t^{1,N,h}, ..., X_t^{N,N,h}\right)$ if h is the size of each Euler time step. Following [1], one then has

$$\left\| \sup_{0\le r\le T} \left| X_r^{\cdot,N,h} - X_r^{\cdot,N} \right| \right\|_p \lesssim \sqrt{h}, \tag{4}$$

where $\lesssim$ involves a constant that does not depend on N and h.

Remark 1. *In order to focus on our main ideas and to avoid a notational blow up, we assume in this paper that the system $\mathbf{X}_t^N$ (cf. (2)) is constructed exactly, hence we neglect the numerical integration error (4) due to the Euler scheme for example. On the other hand, due to (4) it will be clear how several results in this paper have to be adapted in case (2) is approximated using the Euler scheme.*

The central problem in this paper is the computation of functionals of the form

$$w(x) = \mathsf{E}\left[f\left(X_T\right) \mid X_t = x\right], \quad x \in \mathbb{R}^d, \tag{5}$$

for fixed $t \ge 0$ and $T > t$, globally in space, where X is the solution to (1). In this respect we propose a regression approach based on the particle system (2) and analyze its convergence properties.

2 Regression for expected functionals on particle systems

Let $\mathbf{X}_t^N := (X_t^{1,N}, ..., X_t^{N,N})$, $t \ge 0$ be a particle system (2). Let for each $K \in \mathbb{N}$, $\mathcal{H}_K$ be a K-dimensional linear space of functions $h : \mathbb{R}^d \to \mathbb{R}$, where the dimension K may depend on N. Next consider

$$\widetilde{w}_N\left(\cdot\right) = \operatorname*{arg\,min}_{h\in\mathcal{H}_K} \left\{ \frac{1}{N}\sum_{i=1}^N \left(f(X_T^{i,N}) - h(X_t^{i,N}) \right)^2 \right\} \tag{6}$$

as a least-squares estimate of (5). In this section we are going to analyze the properties of the estimate $\widetilde{w}_N$. Note that the paths $X_\cdot^{1,N}, \ldots, X_\cdot^{N,N}$ are generally dependent, so that the known results from regression analysis (see, e.g. [3]) can not be applied directly. At the same time let $\mathbf{X}_t = \left(X_t^1, ..., X_t^N\right)$ be a vector of i.i.d. copies of the exact solution to (1), and define for a fixed $t > 0$,

$$w_N(\cdot) = \operatorname*{arg\,min}_{h\in\mathcal{H}_K} \left\{ \frac{1}{N}\sum_{i=1}^N \left(f(X_T^{i}) - h(X_t^{i}) \right)^2 \right\}.$$

Now let $(\psi_k)_{k=1,2,\ldots}$ be a sequence of linearly independent basis functions and let $\mathcal{H}_K :=$ span $\{\psi_1, ..., \psi_K\}$. Further let us denote by $V, \widetilde{V} \in \mathbb{R}^N$ the column vectors with coordinates

$$V_i = \frac{f(X_T^i)}{\sqrt{N}}, \quad \widetilde{V}_i = \frac{f(X_T^{i,N})}{\sqrt{N}}, \quad i = 1, \ldots, N,$$

respectively, and consider the $\mathbb{R}^{N\times K}$ matrices

$$\widetilde{Z} = \left(\psi_k(X_t^{i,N})/\sqrt{N},\, i = 1, \ldots, N,\, k = 1, \ldots, K\right),$$
$$Z = \left(\psi_k\left(X_t^{i}\right)/\sqrt{N},\, i = 1, \ldots, N,\, k = 1, \ldots, K\right).$$

Then we have

$$\widetilde{w}_N(\cdot) = \widetilde{\beta}_N^\top \boldsymbol{\psi}_K(\cdot), \quad \widetilde{\beta}_N = \left(\widetilde{Z}^\top \widetilde{Z}\right)^{-1} \widetilde{Z}^\top \widetilde{V} = \widetilde{Z}^\dagger \widetilde{V}$$

and

$$w_N(\cdot) = \beta_N^\top \boldsymbol{\psi}_K(\cdot), \quad \beta_N = \left(Z^\top Z\right)^{-1} Z^\top V = Z^\dagger V$$

with $\boldsymbol{\psi}_K = (\psi_1, \ldots, \psi_K)^\top$. Let us now consider the truncated versions of the estimates $\widetilde{w}^N$ and w^N defined as $T_M \widetilde{w}^N$ and $T_M w^N$, respectively, where T_M is a truncation operator of the form:

$$T_M f = \begin{cases} M, & f > M, \\ f, & -M \le f \le M, \\ -M, & f < -M. \end{cases} \tag{7}$$

The following theorem, proved in Section 2, relays on perturbation analysis in the context of linear regression carried out in Section 4.

Theorem 2. *Suppose that* $\sup_{x\in\mathbb{R}^d} |f(x)| \le C_f$, $\sup_{x\in\mathbb{R}^d} |w(x)| \le M$, *and that*

$$\sigma^2 := \sup_{x\in\mathbb{R}^d} \operatorname{Var}\left[f(X_T)\right| X_t = x] < \infty,$$

for some constants $C_f > 0$, $M > 0$, *and* $\sigma > 0$, *respectively. Further assume that all functions* ψ_k, $k = 1, 2, \ldots$ *and* f *are Lipschitz continuous, i.e.,*

$$|\psi_k(x) - \psi_k(y)| \le L_k |x - y|, \quad |f(x) - f(y)| \le L_f |x - y|$$

for all $x, y \in \mathbb{R}^d$ *and some constants* L_f, L_k, $k = 1, 2, \ldots$, *and that*

$$\frac{1}{K} \sum_{k=1}^{K} \int \psi_k^2(x) \mu_t(dx) \le D_\psi^2 \tag{8}$$

for some constant D_ψ *not depending on* K. *Finally, suppose that*

$$0 < \varkappa_\circ \le \lambda_{\min}(\Sigma_K) < \lambda_{\max}(\Sigma_K) \le \varkappa^\circ < \infty$$

for all $K \in \mathbb{N}$ *with*

$$\Sigma_K = \left(\int \psi_k(x)\, \psi_l(x)\, \mu_t(dx),\ k, l = 1, \ldots, K\right),$$

where $\varkappa_\circ, \varkappa^\circ$ *do not depend on* K. *If* $K/N \to 0$ *as* $N \to \infty$, *then it holds*

$$\mathsf{E}\left[\int (T_M \widetilde{w}_N(x) - w(x))^2 \mu_t(dx)\right] \lesssim \frac{K}{N}\left[\sum_{k=1}^{K} L_k^2 + \log(N)\right] + \exp\left(-\frac{N}{D}\right) + \inf_{h\in\mathcal{H}_K}\left(\int (h(x) - w(x))^2 \mu_t(dx)\right) \tag{9}$$

for constant D *depending on* $\varkappa_\circ$, $\varkappa^\circ$ *only, where* $\lesssim$ *stands for* $\le$ *up to a constant depending on* M, σ, L_f, C_f *and* D_ψ.

2.1 Error bounds for piecewise polynomial regression

There are different ways to choose the basis functions $\psi_1, \ldots, \psi_K$. In this section we describe piecewise polynomial partitioning estimates and present L^2-upper bounds for the estimation error. We fix some $p \in \mathbb{N}$, which will denote the maximal degree of polynomials involved in our basis functions. The piecewise polynomial partitioning estimate of w works as follows: consider some $R > 0$ and an equidistant partition of $[-R, R]^d$ in S^d cubes $Q_1, \ldots, Q_{S^d}$, where $S \in \mathbb{N}$ denotes the number of equidistant subintervals of $[-R, R]$. Further, consider the basis functions $\psi_{k,1}, \ldots, \psi_{k,c_{p,d}}$ with $k \in \{1, \ldots, S^d\}$ and $c_{p,d} := \binom{p+d}{d}$ such that $\psi_{k,1}(x), \ldots, \psi_{k,c_{p,d}}(x)$ are polynomials with degree less than or equal to p for $x \in Q_k$ and $\psi_{k,1}(x) = \ldots = \psi_{k,c_{p,d}}(x) = 0$ for $x \notin Q_k$. Then we obtain the least squares regression estimate $\widetilde{w}_N(x)$ for $x \in \mathbb{R}^d$ as described in the previous section, based on $K = S^d c_{p,d} = O(S^d p^d)$ basis functions. In particular, we have $\widetilde{w}_N(x) = 0$ for any $x \notin [-R, R]^d$. We note that the cost of computing $\widetilde{w}_N$ is $O(N_r S^d p^{2d})$ rather than $O(N_r S^{2d} p^{2d})$ due to a block diagonal matrix structure of $\left(\widetilde{Z}^\top \widetilde{Z}\right)^{-1} \widetilde{Z}^\top$. An equivalent approach, which leads to the same estimator $\widetilde{w}_N$, is to perform separate regressions for each cube $Q_1, \ldots, Q_{S^d}$. Here, the number of basis functions at each regression is of order p^d so that the overall cost is of order $N_r S^d p^{2d}$, too. For $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $h \in [1, \infty)$, we will use the notations

$$|x|_h \doteq \left(\sum_{i=1}^d |x_i|^h\right)^{1/h}, \quad |x|_\infty \doteq \max_{i=1,\ldots,d} |x_i|.$$

Let us define the operator D^α as follows

$$D^\alpha g(x) \doteq \frac{\partial^{|\alpha|} g(x)}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}, \tag{10}$$

where g is a real-valued function, $\alpha \in \mathbb{N}_0^d$ and $|\cdot|$ means the cardinality of a set. For $s \in \mathbb{N}_0$, $C > 0$ and $h \in [1, \infty]$, we say that a function $g\colon \mathbb{R}^d \to \mathbb{R}$ is $(s+1, C)$*-smooth w.r.t. the norm* $|\cdot|_h$ whenever, for all α with $|\alpha| = \sum_{i=1}^d \alpha_i = s$, we have

$$|D^\alpha g(x) - D^\alpha g(y)| \leq C|x - y|_h, \quad x, y \in \mathbb{R}^d,$$

i.e. the function $D^\alpha g$ is globally Lipschitz with the Lipschitz constant C with respect to the norm $|\cdot|_h$ on $\mathbb{R}^d$ (cf. Definition 3.3 in [3]). We assume that, for some constant $h \in [1, \infty]$ and some positive constants C_h, ν, B_ν, it holds:

(A1) w is $(p+1, C_h)$-smooth w.r.t. the norm $|\cdot|_h$,

(A2) $\sup_{t \in [0,T]} \mathsf{P}(|X_t| > R) \leq B_\nu R^{-\nu}$ for all $R > 0$.

Theorem 3. *Suppose that the conditions of Theorem 2 hold, then under (A1) and (A2) we have*

$$\begin{aligned} \mathsf{E}\left[\int \left(T_M \widetilde{w}_N(x) - w(x)\right)^2 \mu_t(dx)\right] &\lesssim \frac{S^d c_{p,d}}{N} \left[\sum_{k=1}^{S^d c_{p,d}} L_k^2 + \log(N)\right] \\ &\quad + \exp\left(-\frac{N}{Q}\right) + \frac{8\, C_h^2}{(p+1)!^2 d^{2-2/h}} \left(\frac{Rd}{S}\right)^{2(p+1)} + 8A^2 B_\nu R^{-\nu}. \end{aligned} \tag{11}$$

Moreover, if $L_k \lesssim k^\rho$, $k \to \infty$ for some $\rho \geq 0$, then under a proper choice of K and S depending on N we get

$$\mathsf{E}\left[\int \left(T_M \widetilde{w}_N(x) - w(x)\right)^2 \mu_t(dx)\right] \lesssim N^{-\frac{2\nu(p+1)}{((2\rho+1)d+1)(\nu+2(p+1))+2\nu(p+1)}}, \quad N \to \infty.$$

3 Applications

3.1 Optimal stopping

Let $Z : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be a measurable reward map and $0 < t_1 < \ldots < t_n := T$, $T > 0$, be a given sequence of n exercise dates. Let us take n fixed (for the time being) and consider the stopping problem

$$V_0 := \sup_{\mathbb{F}\text{-stopping times } \tau} \mathsf{E}\left[Z\left(\tau, X_\tau\right)\right]. \tag{12}$$

In (12) a generic $\mathbb{F}$-stopping time τ takes values in the set $\{t_1, \ldots, t_n\}$ and satisfies

$$\{\tau \leq t_k\} \in \mathcal{F}_{t_k}, \quad k = 1, \ldots, n,$$

where the filtration $\mathcal{F}_t$ is generated by W and augmented in the usual way. In fact, the MVSDE (1) may be considered as a usual non-autonomous, Markovian diffusion SDE, since $\{\mu_s : 0 \leq s \leq T\}$ is some deterministic flow of distributions, although not explicitly known beforehand. Therefore, connected with (12) the standard notions of Snell envelope and Bellman principle apply. That is, we may introduce the Snell envelope,

$$V_j := \sup_{\mathbb{F}\text{-stopping times } \tau,\ \tau \geq t_j} \mathsf{E}\left[\left. Z\left(\tau, X_\tau\right)\right| \mathcal{F}_{t_j}\right], \quad j = 0, ..., n \tag{13}$$

(with $\sup$ standing for the essential supremum), as being a discrete time Markovian process that satisfies the Bellman principle,

$$V_j = \max\left(Z_j, \mathsf{E}\left[\left. V_{j+1}\right| \mathcal{F}_{t_j}\right]\right) = \max\left(Z_j, \mathsf{E}\left[\left. V_{j+1}\right| X_{t_j}\right]\right), \quad j = 0, \ldots, n-1. \tag{14}$$

with $Z_j := Z\left(t_j, X_{t_j}\right)$.

Due to the Bellman principle (14) and the Markov property of (1), it is natural to treat the stopping problem (12) by regression based simulation methods in the spirit of [2], [5], and [6]. There exist functions $\mathcal{V}_j : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, $j = 0, \ldots, n$, that satisfies

$$V_j = \mathcal{V}_j\left(X_{t_j}\right).$$

Our goal is a backward construction of the functions $\mathcal{V}_j$, for $j = n, n-1, \ldots, 0$. Let us first suppose that we are given an independent (and identically distributed) set of trajectories $\left(X_t^i : 0 \leq t \leq T\right)$, that solve (1) for $i = 1, \ldots, N$. We then may consider the following backward pseudo-algorithm in the spirit of [2], [5], and [6]:

3.1.1 Backward pseudo-algorithm

- $j = n$: The function $\mathcal{V}_n$ is trivially known, i.e., $\mathcal{V}_n(\cdot) = Z\left(t_n, \cdot\right)$.
- Suppose for some j, $0 < j < n$, an approximation $\overline{\mathcal{V}}_{j+1}$ to $\mathcal{V}_{j+1}$ is established. We then, in principle, aim at the estimation of a function $\mathcal{C}_j$ representing the conditional expectation

$$\mathcal{C}_j(X_{t_j}) := \mathsf{E}\left[\left. \overline{\mathcal{V}}_{j+1}\left(X_{t_{j+1}}\right)\right| X_{t_j}\right] \tag{15}$$

via a regression algorithm based on the sample $(X^1_{t_j}, \overline{\mathcal{V}}_{j+1}(X^1_{t_{j+1}})), \ldots, (X^N_{t_j}, \overline{\mathcal{V}}_{j+1}(X^N_{t_{j+1}}))$ on a suitable system of basis functions $(\psi_1, \ldots, \psi_K)$ with $\psi_k : \mathbb{R}^d \to \mathbb{R}$. One thus obtains a function

$$\widehat{\mathcal{C}}_j = \sum_{k=1}^{K} \widehat{\beta}_k \psi_k, \quad \text{where} \tag{16}$$

$$\widehat{\beta} := \operatorname*{arg\,min}_{\beta \in \mathbb{R}^K} \sum_{i=1}^{N} \left(\overline{\mathcal{V}}_{j+1}(X^i_{t_{j+1}}) - \sum_{k=1}^{K} \beta_k \psi_k(X^i_{t_j}) \right)^2 ,$$

and then proceed with

$$\overline{\mathcal{V}}_j := \max \left(Z\left(t_j, \cdot\right), \widehat{\mathcal{C}}_j\left(\cdot\right) \right).$$

- The above step may be repeated backwardly until $j = 1$, and then an approximation for (12) is finally obtained via straight forward Monte Carlo,

$$\overline{V}_0 := \sum_{i=1}^{N} \overline{\mathcal{V}}_1(X^i_{t_1}).$$

In reality the implementation (3.1.1) in the context of the MVSDE (1) is not possible, since the independent solution trajectories $(X^i_t : 0 \leq t \leq T)$ are not available. Rather, we suppose that we are given an approximation to (1) by the particle system (2) and, in view of Theorem 2, consider to replace the independent solutions $(X^i_t : 0 \leq t \leq T)$ with the (generally dependent) approximations $\left(X^{i,N}_t : 0 \leq t \leq T \right), i = 1, ..., N$, in the regression procedure (16). That is, we propose the following algorithm.

3.1.2 Backward algorithm for (12) based on (2)

- $j = n : \mathcal{V}_n(\cdot) = Z\left(t_n, \cdot\right).$
- Suppose for some j, $0 < j < n$, an approximation $\overline{\mathcal{V}}_{j+1}$ to $\mathcal{V}_{j+1}$ is established. Then, estimate the function $\mathcal{C}_j$ representing the conditional expectation

$$\mathcal{C}_j(x) := \mathsf{E}\left[\overline{\mathcal{V}}_{j+1}\left(X_{t_{j+1}}\right) \middle| X_{t_j} = x \right]$$

via regression of the sample $\left(X^{1,N}_{t_j}, \overline{\mathcal{V}}_{j+1}\left(X^{1,N}_{t_{j+1}}\right)\right), ..., \left(X^{N,N}_{t_j}, \overline{\mathcal{V}}_{j+1}\left(X^{N,N}_{t_{j+1}}\right)\right)$ on the system of basis functions $(\psi_1, ..., \psi_K)$ with $\psi_k : \mathbb{R}^d \to \mathbb{R}$ as spelled out in Section 2 below. That is, set

$$\widehat{\mathcal{C}}_j = \sum_{k=1}^{K} \widehat{\beta}_k \psi_k, \quad \text{where}$$

$$\widehat{\beta} := \operatorname*{arg\,min}_{\beta \in \mathbb{R}^K} \sum_{i=1}^{N} \left(\overline{\mathcal{V}}_{j+1}\left(X^{i,N}_{t_{j+1}}\right) - \sum_{k=1}^{K} \beta_k \psi_k\left(X^{i,N}_{t_j}\right) \right)^2 ,$$

and then proceed with

$$\overline{\mathcal{V}}_j := \max \left(Z\left(t_j, \cdot\right), T_M \widehat{\mathcal{C}}_j\left(\cdot\right) \right),$$

where, T_M is a truncation operator (7) and M is some a-priori upper bound for the function $\mathcal{C}_j$.

- Repeat the above step backwardly until $j=1$, and then approximate (12) finally obtained via,

$$\overline{V}_0 := \sum_{i=1}^{N} \overline{\mathcal{V}}_1(X_{t_1}^{i,N}).$$

3.2 Variance reduction

Given the solution μ_t that satisfies the nonlinear Fokker-Planck equation

$$\partial_t \mu_t(y) = -\sum_{i=1}^{d} \partial_{y^i} \left(\mu_t(y) \int a^i(y,v)\mu_t(v)dv \right) \\ +\frac{1}{2}\sum_{i,j=1}^{d} \partial_{y^i y^j} \left(\mu_t(y) \sum_{l=1}^{m} \int\int b^{il}(y,u)b^{jl}(y,v)\mu_t(u)\mu_t(v)dudv \right) \tag{17}$$

with initial density $\mu_0(y)\,dy = \delta(y-x_0)dy$, with $\delta(\cdot)$ being the dirach delta function, (1) may be considered as a non-autonomous standard SDE in fact. We thus may consider the martingale

$$w(t, X_t^{0,x_0}) = \mathsf{E}_{\mathcal{F}_t}\left[\int f\left(X_T^{0,x_0}, u\right) \mu_T^{0,x_0}(du) \right]. \tag{18}$$

Applying Itô to (18) yields,

$$dw(t, X_t^{0,x_0}) = \sum_{i=1}^{d} \partial_{y^i} w(t, X_t^{0,x_0}) \left[\sum_{l=1}^{m} \int b^{il}(X_t^{0,x_0}, u)\mu_t^{0,x_0}(u)\,du\,dW_t^l \right]$$

since $w(t, X_t)$ is a martingale. Thus, if

$$Y_t = w(t, X_t^{0,x_0}) + \int_0^t F(s, X_s^{0,x_0}) \cdot dW_s$$

for some $F : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \to \mathbb{R}^m \times \mathbb{R}^d$, it holds

$$Y_t = \mathbb{E}_{\mathcal{F}_t}\left[\int f\left(X_T^{0,x_0}, u\right) \mu_T^{0,x_0}(du) + \int_0^T F(s, X_s^{0,x_0}) \cdot dW_s \right].$$

That is, if

$$F^l(t,y) = -\sum_{i=1}^{d} \partial_{y^i} w(t,y) \int b^{il}(y,u)\mu_t^{0,x_0}(u)du \tag{19}$$

we have

$$w(0,x_0) = w(T, X_T^{0,x_0}) + \int_0^T F(s, X_s^{0,x_0}) \cdot dW_s \\ = \int f\left(X_T^{0,x_0}, u\right) \mu_T^{0,x_0}(du) + \int_0^T F(s, X_s^{0,x_0}) \cdot dW_s \quad \text{almost surely.}$$

3.2.1 Practical variance reduction

By a pre-simulation of the particle system of size N_r, compute for (several) fixed t the regression estimate,

$$(\widehat{\alpha}_t^1,\ldots,\widehat{\alpha}_t^{K_w}) := \underset{\alpha\in\mathbb{R}^{K_w}}{\arg\min}\frac{1}{N_r}\sum_{i=1}^{N_r}\left(\frac{1}{N_r}\sum_{j=1}^{N_r} f(X_T^{i,N_r},X_T^{j,N_r}) - \sum_{l=0}^{K_w}\alpha^l\varphi_l(X_t^{i,N_r})\right)^2,$$

where $\varphi_1,\ldots,\varphi_{K_w}$ is the set of basis functions. Hence we can define

$$w(t,y)\;\approx\sum_{r=0}^{K_w}\widehat{\alpha}_t^r\varphi_r(y),\quad\text{and}\quad \partial_{y^i}w(t,y)\;\approx\sum_{r=0}^{K_w}\widehat{\alpha}_t^r\partial_{y^i}\varphi_r(y). \tag{20}$$

Next, based on a new simulation with N particles we consider in view of (19) and (20), the control functional

$$\widetilde{F}^l(t,y) = -\sum_{r=0}^{K_w}\widehat{\alpha}_t^r\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{d}\partial_{y^i}\varphi_r(y)b^{il}(y,X_t^j). \tag{21}$$

To analyze the variance reduction effect of this control functional, we can use Theorem 3.

4 Perturbation analysis for linear regression

Consider a least squares problem of the form

$$\beta^\circ = \underset{\beta\in\mathbb{R}^d}{\arg\min}\sum_{i=1}^{n}(Y_i-\beta^\top U_i)^2, \tag{22}$$

where for $i=1,...,n$, (Y_i,U_i) are i.i.d. pairs of a random variable Y_i and a random (column) vector $U_i\in\mathbb{R}^d$. With $U := (U_1,\ldots,U_n)\in\mathbb{R}^{d\times n}$, $Z = n^{-1/2}U^\top$, and $V = n^{-1/2}\left(Y_1,\ldots,Y_n\right)^\top$, the solution of the problem (22) can be written in terms of pseudo inverses (denoted with †),

$$\beta^\circ = \left(UU^\top\right)^{-1}UY = \left(Z^\top Z\right)^{-1}Z^\top V = Z^\dagger V. \tag{23}$$

Consider now the least squares problem (22) due to a perturbation $\left(\widetilde{Y}_i,\widetilde{U}_i\right)$ of the pairs (Y_i,U_i), and define $\widetilde{Z}$ and $\widetilde{V}$ accordingly. We so consider (cf. (23))

$$\widetilde{\beta}^\circ = \left(\widetilde{Z}^\top\widetilde{Z}\right)^{-1}\widetilde{Z}^\top\widetilde{V} = \widetilde{Z}^\dagger\widetilde{V} \tag{24}$$

and set

$$\widetilde{Z} = Z+E,\quad \widetilde{V} = V+F. \tag{25}$$

While the rows of Z and the components of V are independent, the rows of the perturbation matrix E and the components of the perturbation vector F are generally dependent. Also we note that we don't assume any kind of independence between the perturbations E and F and the matrix Z and vector V, respectively.

Theorem 4. *Consider the least squares problem (22) with solution (23), and its perturbation due to (25) with solution (24), respectively. Assume that* $U_1, \ldots, U_n$ *in (22) are i.i.d. random vectors in* $\mathbb{R}^d$ *such that for some* $\gamma > 0,$

$$\mathsf{E}\left[\exp\left(\alpha^\top U_1\right)\right] \leq \exp\left(|\alpha|^2 \gamma/2\right),$$

for all $\alpha \in \mathbb{R}^d$*, and set*

$$\mathsf{E}\left[U_1 U_1^\top\right] = \Sigma,$$

hence

$$Z^\top Z = \frac{1}{n} U U^\top = \frac{1}{n} \sum_{i=1}^n U_i U_i^\top.$$

Let $\lambda_{\min}(\Sigma)$ *be the smallest eigenvalue, and* $\lambda_{\max}(\Sigma)$ *be the largest eigenvalue of* $\Sigma,$ *respectively. Then for any* $\rho \in (0, \lambda_{\min}(\Sigma))$ *and* $\varepsilon \in (0, (\lambda_{\min}(\Sigma) - \rho) \wedge \gamma)$ *we have on the set* $\mathcal{C} := \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4$ *with*

$$\begin{aligned}
\mathcal{C}_1 :&\quad \lambda_{\max}\left(Z^\top Z\right) < \lambda_{\max}(\Sigma) + \varepsilon, \\
\mathcal{C}_2 :&\quad \lambda_{\min}\left(Z^\top Z\right) > \lambda_{\min}(\Sigma) - \varepsilon, \\
\mathcal{C}_3 :&\quad \lambda_{\min}(\Sigma) - \left(2\sqrt{\lambda_{\max}(\Sigma) + \varepsilon} + 1\right) |E| > \rho + \varepsilon, \\
\mathcal{C}_4 :&\quad |E| < 1.
\end{aligned}$$

that

$$\left|\widetilde{\beta}^\circ - \beta^\circ\right| \leq c_1(\Sigma, \varepsilon, \rho)\, |E|\, |V| + c_2(\Sigma, \varepsilon, \rho)\, |F|,$$

where

$$\begin{aligned}
c_1(\Sigma, \varepsilon, \rho) &:= \frac{1}{\rho} + \frac{2\left(\lambda_{\max}(\Sigma) + \varepsilon\right) + \sqrt{\lambda_{\max}(\Sigma) + \varepsilon}}{\rho^2} \quad \text{and} \\
c_2(\Sigma, \varepsilon, \rho) &:= c_1(\Sigma, \varepsilon, \rho) + \frac{\sqrt{\lambda_{\max}(\Sigma) + \varepsilon}}{\lambda_{\min}(\Sigma) - \varepsilon}.
\end{aligned}$$

For the probability of $\mathcal{C}$ *we have that*

$$\begin{aligned}
\mathsf{P}[\mathcal{C}] \geq\ & 1 - 2 \cdot 9^d \exp\left(-\frac{n}{144} \varepsilon^2/\gamma^2\right) \\
& - C_p \left(\left(\frac{2\sqrt{\lambda_{\max}(\Sigma) + \varepsilon} + 1}{\lambda_{\min}(\Sigma) - \varepsilon - \rho}\right)^p + 1\right),
\end{aligned}$$

where $C_p := \mathsf{E}\left[|E|^p\right]$ *is small enough (such that the above bound is positive).*

Proof. Note that $\mathcal{C} := \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4$ implies (28) in Lemma 6 and so by this Lemma,

$$\begin{aligned}
\left|(Z + E)^\dagger - Z^\dagger\right| &\leq \frac{|E|}{\rho}\left[1 + \frac{(2|Z| + 1)|Z|}{\rho}\right] \\
&\leq \frac{|E|}{\rho}\left[1 + \frac{2\left(\lambda_{\max}(\Sigma) + \varepsilon\right) + \sqrt{\lambda_{\max}(\Sigma) + \varepsilon}}{\rho}\right] \\
&= c_1(\Sigma, \varepsilon, \rho)\, |E|
\end{aligned}$$

Thus, on $\mathcal{C}$ one has also,

$$\begin{aligned}\left|(Z+E)^{\dagger}\right| &\leq \left|(Z+E)^{\dagger}-Z^{\dagger}\right|+\left|Z^{\dagger}\right| \\ &\leq c_1(\Sigma,\varepsilon,\rho)+\frac{\sqrt{\lambda_{\max}(\Sigma)+\varepsilon}}{\lambda_{\min}(\Sigma)-\varepsilon}=c_2(\Sigma,\varepsilon,\rho),\end{aligned}$$

using that $\left|Z^{\dagger}\right| \leq \left|\left(Z^{\top}Z\right)^{-1}\right| |Z|$. So on $\mathcal{C}$ we get,

$$\begin{aligned}\left|\widetilde{\beta}^{\circ}-\beta^{\circ}\right| &= \left|\widetilde{Z}^{\dagger}\widetilde{V}-Z^{\dagger}V\right| = \left|(Z+E)^{\dagger}(V+F)-Z^{\dagger}V\right| \\ &\leq c_1(\Sigma,\varepsilon,\rho)\,|E|\,|V|+c_2(\Sigma,\varepsilon,\rho)\,|F|\,.\end{aligned}$$

For the probability of $\mathcal{C}$ one has

$$\mathsf{P}\left[\mathcal{C}\right] \geq 1-\mathsf{P}\left[\Omega\backslash\mathcal{C}_1\cup\Omega\backslash\mathcal{C}_2\right]-\mathsf{P}\left[\Omega\backslash\mathcal{C}_3\right]-\mathsf{P}\left[\Omega\backslash\mathcal{C}_3\right]. \tag{26}$$

For the term $\mathsf{P}\left[\Omega\backslash\mathcal{C}_1\cup\Omega\backslash\mathcal{C}_2\right]$ we are going to apply Lemma 7. A straightforward calculation shows that (32) can be transformed into

$$\begin{aligned}\delta &= 2\cdot 9^d \exp\left(-\frac{n}{4}\frac{\varepsilon^2/\gamma^2}{\left(\sqrt{\varepsilon/\gamma+8}+2\sqrt{2}\right)^2}\right) \\ &\leq 2\cdot 9^d \exp\left(-\frac{n}{128}\frac{\varepsilon^2/\gamma^2}{1+\frac{1}{8}\varepsilon/\gamma}\right) \\ &\leq 2\cdot 9^d \exp\left(-\frac{n}{144}\varepsilon^2/\gamma^2\right)\end{aligned} \tag{27}$$

since by assumption $\varepsilon/\gamma<1$. By using (27), Lemma 7 now yields that

$$\mathsf{P}\left[\Omega\backslash\mathcal{C}_1\cup\Omega\backslash\mathcal{C}_2\right] \leq 2\cdot 9^d \exp\left(-\frac{n}{144}\varepsilon^2/\gamma^2\right).$$

Further,

$$\begin{aligned}\mathsf{P}\left[\Omega\backslash\mathcal{C}_3\right] &= \mathsf{P}\left[\lambda_{\min}(\Sigma)-\left(2\sqrt{\lambda_{\max}(\Sigma)+\varepsilon}+1\right)|E| \leq \rho+\varepsilon\right] \\ &= \mathsf{P}\left[\frac{\lambda_{\min}(\Sigma)-\varepsilon-\rho}{2\sqrt{\lambda_{\max}(\Sigma)+\varepsilon}+1} \leq |E|\right] \\ &\leq \left(\frac{2\sqrt{\lambda_{\max}(\Sigma)+\varepsilon}+1}{\lambda_{\min}(\Sigma)-\varepsilon-\rho}\right)^p \mathsf{E}\left[|E|^p\right],\end{aligned}$$

and

$$\mathsf{P}\left[\Omega\backslash\mathcal{C}_4\right] = \mathsf{P}\left[|E| \geq 1\right] \leq \mathsf{E}\left[|E|^p\right].$$

The statement now follows from (26). $\square$

Corollary 5. *Take* $\rho = \lambda_{\min}(\Sigma)/4, \varepsilon = \gamma \wedge (\lambda_{\min}(\Sigma)/4)$. *Then with*

$$c_1(\Sigma) := \frac{1}{\lambda_{\min}(\Sigma)/4} + \frac{2(\lambda_{\max}(\Sigma) + \gamma \wedge (\lambda_{\min}(\Sigma)/4)) + \sqrt{\lambda_{\max}(\Sigma) + \gamma \wedge (\lambda_{\min}(\Sigma)/4)}}{(\lambda_{\min}(\Sigma)/4)^2} \quad \text{and}$$

$$c_2(\Sigma) := c_1(\Sigma) + \frac{\sqrt{\lambda_{\max}(\Sigma) + \gamma \wedge (\lambda_{\min}(\Sigma)/4)}}{\lambda_{\min}(\Sigma) - \gamma \wedge (\lambda_{\min}(\Sigma)/4)},$$

we have on $\mathcal{C}$,

$$\left|\widetilde{\beta}^\circ - \beta^\circ\right| \le c_1(\Sigma)\,|E|\,|V| + c_2(\Sigma)\,|F|,$$

with probability

$$\mathsf{P}[\mathcal{C}] \ge 1 - 2 \cdot 9^d \exp\left(-\frac{n}{144}\left(1 \wedge \frac{\lambda_{\min}^2(\Sigma)}{16\gamma^2}\right)\right) - C_p\left(\left(\frac{2\sqrt{\lambda_{\max}(\Sigma) + \lambda_{\min}(\Sigma)/4} + 1}{\lambda_{\min}(\Sigma)/2}\right)^p + 1\right).$$

5 Proofs

5.1 Proof of Theorem 2

Proof. By using that

$$\left|T_M \widetilde{w}^N(x) - T_M w^N(x)\right| \le \left|\widetilde{w}^N(x) - w^N(x)\right|$$

almost surely, one has for any event $\mathcal{C} \in \mathcal{F}$

$$\begin{aligned}
\left(\mathsf{E}\left[\int \left(T_M \widetilde{w}^N(x) - w(x)\right)^2 \mu_t(dx)\right]\right)^{1/2} \le \\
\left(\mathsf{E}\left[\int 1_{\mathcal{C}} \left(\widetilde{w}^N(x) - w^N(x)\right)^2 \mu_t(dx)\right]\right)^{1/2} + 2M\left(\mathsf{P}[\Omega \backslash \mathcal{C}]\right)^{1/2} \\
+ \left(\mathsf{E}\left[\int \left(T_M w^N(x) - w(x)\right)^2 \mu_t(dx)\right]\right)^{1/2} \\
\le D_{\boldsymbol{K}}^{\psi} \left(\mathsf{E}\left[\left|\widetilde{\beta}_N - \beta_N\right|^2 1_{\mathcal{C}}\right]\right)^{1/2} \\
+2M\left(\mathsf{P}[\Omega \backslash \mathcal{C}]\right)^{1/2} \\
+ \left(\mathsf{E}\left[\int \left(T_M w^N(x) - w(x)\right)^2 \mu_t(dx)\right]\right)^{1/2}.
\end{aligned}$$

Now let

$$\begin{aligned}
V &= \left((f(X_t^{i,N})/\sqrt{N},\, i = 1, \ldots, N\right)^\top \in \mathbb{R}^N \\
E &= \left((\psi_k(X_t^{i,N}) - \psi_k(X_t^i))/\sqrt{N},\, i = 1, \ldots, N,\, k = 1, \ldots, K,\right) \in \mathbb{R}^{N \times K}, \\
F &= \left((f(X_t^{i,N}) - f(X_t^i))/\sqrt{N},\, i = 1, \ldots, N\right)^\top \in \mathbb{R}^N
\end{aligned}$$

and then Theorem 4 implies via Corollary 5 that

$$
\begin{aligned}
\left|\widetilde{\beta}_N - \beta_N\right|^2 &\leq 2c_1^2(\Sigma)\,|E|^2\,|V|^2 + 2c_2^2(\Sigma)\,|F| \\
&\leq 2d_1^2\,|E|^2\,C_f^2 + 2d_2^2\,|F|
\end{aligned}
$$

on a set $\mathcal{C}$ with probability

$$
\begin{aligned}
\mathsf{P}\left[\mathcal{C}\right] \geq 1 &- 2\cdot 9^K \exp\left(-\frac{N}{144}\left(1\wedge\frac{\varkappa_\circ^2}{16\gamma^2}\right)\right) \\
&- \mathsf{E}[|E|^p]\left(\left(\frac{\sqrt{5\varkappa^\circ}+1}{\varkappa_\circ/2}\right)^p + 1\right)
\end{aligned}
$$

where the constants $d_1,\, d_2$ depend only on $\varkappa_\circ,\, \varkappa^\circ$. In particular we may take

$$
\begin{aligned}
d_1 &:= \frac{4}{\varkappa_\circ} + 8\frac{5\varkappa^\circ + \sqrt{5\varkappa^\circ}}{\varkappa_\circ^2}, \quad \text{and} \\
d_2 &:= d_1 + \frac{\sqrt{5\varkappa^\circ}}{\varkappa_\circ} = \frac{4+\sqrt{5\varkappa^\circ}}{\varkappa_\circ} + 8\frac{5\varkappa^\circ + \sqrt{5\varkappa^\circ}}{\varkappa_\circ^2}.
\end{aligned}
$$

As a result

$$
\begin{aligned}
\mathsf{E}\left[\left|\widetilde{\beta}_N - \beta_N\right|^2 1_{\mathcal{C}}\right] &\leq 2d_1^2 C_f^2 \mathsf{E}\left[|E|^2\right] + 2d_2^2 \mathsf{E}\left[|F|^2\right] \\
&\leq 2d_1^2 C_f^2 \left(\mathsf{E}\left[\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K\left(\psi_k\left(X_t^{i,N}\right) - \psi_k\left(X_t^i\right)\right)^2\right]\right) \\
&+ 2d_2^2 \mathsf{E}\left[\frac{1}{N}\sum_{i=1}^N\left(f\left(X_t^{i,N}\right) - f\left(X_t^i\right)\right)^2\right] \\
&\leq \left(2d_1^2 C_f^2 \sum_{k=1}^K L_k^2 + 2d_2^2 L_f^2\right)\mathsf{E}\left[\left|X_t^{\cdot,N} - X_t^{\cdot}\right|^2\right]
\end{aligned}
$$

We further have for $p \geq 2,$

$$
\begin{aligned}
\left(\mathsf{E}\left[|E|^p\right]\right)^{1/p} &\leq \left(\mathsf{E}\left[\left(\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K\left(\psi_k\left(X_t^{i,N}\right) - \psi_k\left(X_t^i\right)\right)^2\right)^{p/2}\right]\right)^{1/p} \\
&= \left(\left(\mathsf{E}\left[\left(\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K\left(\psi_k\left(X_t^{i,N}\right) - \psi_k\left(X_t^i\right)\right)^2\right)^{p/2}\right]\right)^{2/p}\right)^{1/2} \\
&\leq \left(\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K\left(\mathsf{E}\left[\left(\psi_k\left(X_t^{i,N}\right) - \psi_k\left(X_t^i\right)\right)^p\right]\right)^{2/p}\right)^{1/2} \\
&\leq \left(\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K L_k^2\left(\mathsf{E}\left[\left|X_t^{i,N} - X_t^i\right|^p\right]\right)^{2/p}\right)^{1/2} \\
&= \sqrt{\sum_{k=1}^K L_k^2}\left(\mathsf{E}\left[\left|X_t^{\cdot,N} - X_t^{\cdot}\right|^p\right]\right)^{1/p}.
\end{aligned}
$$

Combining the latter bounds with (3) and Theorem 11.3 from [3], and taking $p = 2$ for simplicity, we get

$$\left(\mathsf{E}\left[\int \left(T_M \widetilde{w}^N(x) - w(x)\right)^2 \mu_t(dx)\right]\right)^{1/2} \leq$$
$$\leq D^{\psi}_{\boldsymbol{K}} \left(2d_1^2 C_f^2 \sum_{k=1}^{K} L_k^2 + 2d_2^2 L_f^2\right)^{1/2} \frac{C_2}{\sqrt{N}}$$
$$+4M \cdot 9^{K/2} \exp\left(-\frac{N}{288}\left(1 \wedge \frac{\varkappa_\circ^2}{16\gamma^2}\right)\right)$$
$$+2\sqrt{2}M\left(\sum_{k=1}^{K} L_k^2\right)^{1/2} \left(\left(\frac{\sqrt{5\varkappa^\circ}+1}{\varkappa_\circ/2}\right)^2 + 1\right)^{1/2} \frac{C_2}{\sqrt{N}}$$
$$+c_1 \max(\sigma, M) \frac{\sqrt{1+\log N}\sqrt{K}}{\sqrt{N}}$$
$$+c_2 \inf_{h \in \mathcal{H}_K} \left(\int (h(x) - w(t,x))^2 \mu_t(dx)\right)^{1/2}$$

for universal constants c_1, c_2. Summarizing we obtain (9) (using (8)). □

6 Appendix

Lemma 6. *Let $\rho > 0$ and the matrix $Z \in \mathbb{R}^{n\times d}$ be of full rank with $n > d$. Let Z and $E \in \mathbb{R}^{n\times d}$ be such that*

$$\lambda_{\min}\left(Z^\top Z\right) - (2|Z|+1)|E| > \rho, \quad |E| < 1. \tag{28}$$

Then we have

$$\left|(Z+E)^\dagger - Z^\dagger\right| \leq \frac{|E|}{\rho}\left[1 + \frac{(2|Z|+1)|Z|}{\rho}\right]. \tag{29}$$

Proof. Denote

$$\Delta = Z^\top E + E^\top Z + E^\top E,$$

then using the identity

$$\begin{aligned}\left((Z+E)^\top (Z+E)\right)^{-1} - \left(Z^\top Z\right)^{-1} &= -\left((Z+E)^\top (Z+E)\right)^{-1} \Delta \left(Z^\top Z\right)^{-1} \\ &= -\left(Z^\top Z + \Delta\right)^{-1} \Delta \left(Z^\top Z\right)^{-1},\end{aligned}$$

we derive

$$\begin{aligned}\left|\left(\left((Z+E)^\top (Z+E)\right)^{-1} - \left(Z^\top Z\right)^{-1}\right) Z^\top\right| &\leq \left|\left(Z^\top Z + \Delta\right)^{-1}\right| \left|\left(Z^\top Z\right)^{-1}\right| (2|Z|+1)|E||Z| \\ &\leq \frac{(2|Z|+1)|E||Z|}{\rho^2},\end{aligned} \tag{30}$$

since we have $\left|\left(Z^\top Z\right)^{-1}\right| = \lambda_{\min}^{-1}\left(Z^\top Z\right) < \rho^{-1}$ and

$$\begin{aligned}\lambda_{\min}\left(Z^\top Z+\Delta\right) &= \inf_{|x|=1} x^\top\left(Z^\top Z+\Delta\right)x \geq \inf_{|x|=1} x^\top Z^\top Zx + \inf_{|x|=1} x^\top \Delta x\\ &\geq \lambda_{\min}\left(Z^\top Z\right) - |\Delta| \geq \lambda_{\min}\left(Z^\top Z\right) - (2\,|Z|+1)\,|E|\\ &> \rho > 0.\end{aligned}$$

Analogously we have

$$\left|\left((Z+E)^\top(Z+E)\right)^{-1}E\right| = \left|\left(Z^\top Z+\Delta\right)^{-1}E\right| \leq \frac{|E|}{\rho}, \tag{31}$$

and then (29) follows by (30), (31), and the triangle inequality. □

Lemma 7. *Let $X_1,\ldots,X_n$ be independent random vectors in $\mathbb{R}^d$ such that*

$$\mathsf{E}\left[X_iX_i^\top\right] = \Sigma,$$

and for some $\gamma>0$,

$$\mathsf{E}\left[\exp\left(\alpha^\top X_i\right)\right] \leq \exp\left(|\alpha|^2\gamma/2\right)$$

for all $\alpha\in\mathbb{R}^d$ and all $i=1,\ldots,n$. Then for all $\delta\in(0,1)$,

$$\mathbb{P}\left(\left\{\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^\top\right) > \lambda_{\max}(\Sigma)+\varepsilon_{\delta,n}\right\}\cup \left\{\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^\top\right) < \lambda_{\min}(\Sigma)-\varepsilon_{\delta,n}\right\}\right) \leq \delta,$$

where

$$\varepsilon_{\delta,n}/\gamma = 8\sqrt{2}\sqrt{\frac{d\log 9+\log(2/\delta)}{n}} + 4\frac{d\log 9+\log(2/\delta)}{n}. \tag{32}$$

Proof. See [4]. □

References

[1] Fabio Antonelli, Arturo Kohatsu-Higa, et al. Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *The Annals of Applied Probability*, 12(2):423–476, 2002.

[2] J.F. Carriere. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insur. Math. Econ.*, 19(1):19–30, 1996.

[3] László Györfi, Michael Kohler, Adam Krzyżak, and Harro Walk. *A distribution-free theory of non-parametric regression*. Springer Series in Statistics. Springer-Verlag, New York, 2002.

[4] Daniel Hsu, Sham M Kakade, and Tong Zhang. Dimension-free tail inequalities for sums of random matrices. *arXiv preprint arXiv:1104.1672*, 2011.

[5] F.A. Longstaff and E.S. Schwartz. Valuing american options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001.

[6] J. Tsitsiklis and B. Van Roy. Regression methods for pricing complex american style options. *IEEE Trans. Neural. Net.*, 12(14):694–703, 2001.