



# On a Wasserstein-type distance between solutions to stochastic differential equations

Jocelyne Bion-Nadal, Denis Talay

## ► To cite this version:

Jocelyne Bion-Nadal, Denis Talay. On a Wasserstein-type distance between solutions to stochastic differential equations. The Annals of Applied Probability, 2019, 29 (3), pp.1609-1639. 10.1214/18-AAP1423 . hal-01943863

**HAL Id: hal-01943863**

**<https://inria.hal.science/hal-01943863>**

Submitted on 4 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On a Wasserstein-type distance between solutions to stochastic differential equations

Jocelyne Bion–Nadal, Denis Talay

*CMAP, Ecole Polytechnique  
91128 Palaiseau, France*

*e-mail: [jocelyne.bion-nadal@polytechnique.edu](mailto:jocelyne.bion-nadal@polytechnique.edu)*

*INRIA*

*2004 route des Lucioles  
BP 93*

*06902 Sophia Antipolis, France  
e-mail: [denis.talay@inria.fr](mailto:denis.talay@inria.fr)*

**Abstract:** In this paper we introduce a Wasserstein-type distance on the set of the probability distributions of strong solutions to stochastic differential equations. This new distance is defined by restricting the set of possible coupling measures. We prove that it may also be defined by means of the value function of a stochastic control problem whose Hamilton–Jacobi–Bellman equation has a smooth solution, which allows one to deduce a priori estimates or to obtain numerical evaluations. We exhibit an optimal coupling measure and characterize it as a weak solution to an explicit stochastic differential equation, and we finally describe procedures to approximate this optimal coupling measure.

A notable application concerns the following modeling issue: given an exact diffusion model, how to select a simplified diffusion model within a class of admissible models under the constraint that the probability distribution of the exact model is preserved as much as possible?

**MSC 2010 subject classifications:** Primary 60J60, 93E20.

**Keywords and phrases:** Stochastic differential equations, Wasserstein distance, Stochastic control.

In this paper we propose a Wasserstein-type distance on the set of probability distributions which are solutions to martingale problems. This distance is defined as the solution to a stochastic control problem which allows a priori estimates and numerical approximations.

Our motivation comes from modeling issues. In many situations where stochastic modeling is used, one desires to choose the coefficients of a stochastic differential equation which represents the reality as simply as possible. This typically is the case in Physics when the analytical structure of the model is imposed by physical laws and therefore may present singularities or have a large algebraic complexity which makes the analysis or the simulation extremely difficult.

---

\*The first author’s research is partially supported by the ANR project CAESARS (ANR-15-CE05-0024)

Different methods may be used to this end such as smoothing techniques, homogenization, projection of the coefficients on a space of suitable functions, etc.

The choice of a ‘good’ simplified model should be related to the objective of the modeling process. In this paper, we consider the case where a diffusion model with high complexity coefficients is relevant and one desires to simplify it in order to be able to get theoretical estimates on its probability distribution or to develop accurate numerical approximation methods with low computational cost.

A natural way to set the present problem up consists in minimizing a distance between the probability distributions of a given class of simplified models and the probability distribution of the ‘perfect’ model. This distance should be small when the simplified model captures the essential of the target probability distribution. One therefore needs to consider distances which metrize the weak convergence topology, e.g. the Wasserstein  $\mathcal{W}^2$  distance defined as follows. Consider the map  $d_2(\omega, \bar{\omega}) := (\int_0^T |\omega_u - \bar{\omega}_u|^2 du)^{\frac{1}{2}}$  and the set of probability measures on the Polish space  $(L_2([0, T], \mathbb{R}^d), d_2)$ . The Wasserstein distance

$$\mathcal{W}^2(\mathbb{P}; \bar{\mathbb{P}}) := \left\{ \inf_{\pi \in \Pi(\mathbb{P}; \bar{\mathbb{P}})} \int_{\Omega} \int_0^T |\omega_s - \bar{\omega}_s|^2 ds \pi(d\omega, d\bar{\omega}) \right\}^{\frac{1}{2}}, \quad (0.1)$$

where  $\Pi(\mathbb{P}; \bar{\mathbb{P}})$  is the set of all the probability distributions  $\pi$  on  $L_2([0, T], \mathbb{R}^{2d})$  with marginal distributions  $\mathbb{P}$  and  $\bar{\mathbb{P}}$ . This distance metrizes the weak topology on the set of probability measures  $\pi$  on  $(L_2([0, T], \mathbb{R}^d), d_2)$  such that  $\mathbb{E}^\pi \int_0^T |\omega_s|^2 ds < \infty$ . Unfortunately, the numerical computation of  $\mathcal{W}^2$  or any other Wasserstein distance on an infinite dimensional space is impossible. To overcome this difficulty we here introduce a Wasserstein-type distance on the set of the probability distributions of strong solutions to stochastic differential equations. We show that this new distance has the advantage to be characterized as the value function of a stochastic control problem and thus can be computed by discretizing the related Hamilton–Jacobi–Bellman equation. We emphasize that, for the sake of simplicity, we limit ourselves to handle with a distance which is a variant of  $\mathcal{W}^2$ . One can easily extend our result to many variants of other Wasserstein distances provided that the Hamilton–Jacobi–Bellman equation (2.3) below admits a smooth enough solution.

The organization of the paper is as follows. In Section 1 we define the metric  $\widetilde{\mathcal{W}}^2$  and prove first properties of it. In Section 2 we prove that it may also be defined by means of the value function of a stochastic control problem whose Hamilton–Jacobi–Bellman equation has a smooth solution. We also exhibit an optimal coupling measure solution to a martingale problem. In Section 3 we prove that the optimal coupling measure can be approximated by means of continuous or smooth feedback controls. In Section 4 we prove a few elementary convex analysis results which are used in the preceding sections.

**Notation** In all the sequel  $\mathbf{M}_d$  denotes the space of  $d \times d$  matrices and  $\text{Id}_d$  denotes the  $d \times d$  Identity matrix. We denote by  $\mathbf{C}_d$  the set of  $d$ -dimensional

correlation matrices:

$$\mathbf{C}_d := \{C \in \mathbf{M}_d; \text{ there exist } \mathbb{R}^d \text{ valued centered random variables } X \text{ and } Y \text{ s.t.} \\ \mathbb{E}(XX^\top) = \mathbb{E}(YY^\top) = \text{Id}_d, \quad C = \mathbb{E}(XY^\top)\}. \quad (0.2)$$

We denote by  $\mathbf{O}_d$  be the set of orthogonal matrices in  $\mathbf{M}_d$ , and we set

$$\mathbf{SO}_d := \{O \in \mathbf{O}_d, \quad \det(O) = 1\}.$$

For any  $T > 0$  and  $0 < \alpha < 1$  the Hölder space  $\mathcal{C}^{0,\alpha}([0, T] \times \mathbb{R}^d)$  is the set of bounded continuous functions such that  $\phi(t, \cdot)$  is Hölder continuous of order  $\alpha$  for every  $t$  in  $[0, T]$  equipped with the norm

$$\|\phi\|_{0,\alpha} := \sup_{t \in [0, T]} \|\phi(t, \cdot)\|_\alpha := \sup_{t \in [0, T]} \left( \|\phi(t, \cdot)\|_\infty + \sup_{x, y \in \mathbb{R}^d, x \neq y} \frac{|\phi(t, x) - \phi(t, y)|}{|x - y|^\alpha} \right).$$

Similarly, the Hölder space  $\mathcal{C}^{\alpha/2,0}([0, T] \times \mathbb{R}^d)$  is the set of bounded continuous functions such that  $\phi(\cdot, x)$  is Hölder continuous of order  $\frac{\alpha}{2}$  for every  $x$  in  $\mathbb{R}^d$ . We equip the space

$$\mathcal{C}^{\frac{\alpha}{2},\alpha}([0, T] \times \mathbb{R}^d) := \mathcal{C}^{\alpha/2,0}([0, T] \times \mathbb{R}^d) \cap \mathcal{C}^{0,\alpha}([0, T] \times \mathbb{R}^d)$$

with the norm

$$\|\phi\|_{\frac{\alpha}{2},\alpha} := \|\phi\|_{\alpha/2,0} + \|\phi\|_{0,\alpha}.$$

Finally,  $\mathcal{C}^{1+\frac{\alpha}{2},2+\alpha}([0, T] \times \mathbb{R}^d)$  is the set of continuous bounded functions  $\phi$  of class  $\mathcal{C}^{1,2}([0, T] \times \mathbb{R}^d)$  with bounded derivatives such that  $\partial_t \phi$  and  $\partial_{x_i, x_j}^2 \phi$  are in  $\mathcal{C}^{\frac{\alpha}{2},\alpha}([0, T] \times \mathbb{R}^d)$  for all  $1 \leq i, j, \leq d$ . We equip this space with the norm

$$\|\phi\|_{1+\frac{\alpha}{2},2+\alpha} := \|\phi\|_\infty + \sum_{i=1}^d \|\partial_{x_i} \phi\|_\infty + \|\partial_t \phi\|_{\frac{\alpha}{2},\alpha} + \sum_{i,j=1}^d \|\partial_{x_i, x_j}^2 \phi\|_{\frac{\alpha}{2},\alpha}.$$

## 1. The $\widetilde{\mathcal{W}}^2$ distance

### 1.1. Definition of $\widetilde{\mathcal{W}}^2$

Denote by  $\Omega := \mathcal{C}(0, T; \mathbb{R}^d)$  the  $d$ -dimensional canonical space of continuous functions from  $[0, T]$  to  $\mathbb{R}^d$ . Equipped with the locally uniform convergence metric this space is Polish. The canonical filtration is denoted by  $(\mathcal{F}_s, 0 \leq s \leq T)$  and its Borel  $\sigma$ -field is denoted by  $\mathcal{F} := \bigvee_{0 \leq s \leq T} \mathcal{F}_s$ .

**Definition 1.1.** Let  $\mathbf{P}$  be the set of probability measures  $\mathbb{P}$  on the  $d$ -dimensional canonical space which satisfy: there exist  $x_0$  in  $\mathbb{R}^d$  and bounded Lipschitz functions  $\mu$  and  $\sigma$  with  $\sigma$  being uniformly strongly elliptic in the following sense:

$$\exists \lambda > 0, \quad \forall 0 \leq s \leq T, \quad \forall \xi \in \mathbb{R}^d, \quad \forall x \in \mathbb{R}^d, \quad \sum_{i,j=1}^d (\sigma(x)\sigma(x)^\top)^{ij} \xi^i \xi^j \geq \lambda |\xi|^2,$$

such that  $\mathbb{P} \equiv \mathbb{P}_{x_0}^{\mu, \sigma}$  is the probability distribution of the unique strong solution to the stochastic differential with coefficients  $\mu$  and  $\sigma$  and initial condition  $x_0$ .

The definition of the following Wasserstein-type distance on the set  $\mathbf{P}$  results from the obvious but important observation that not any coupling measure of two probability distributions in  $\mathbf{P}$  can be represented as the solution of a 2d-dimensional martingale problem. We thus modify the definition of the standard  $\mathcal{W}^2$  distance by restricting the set of possible coupling measures.

**Definition 1.2.** Given two probability measures  $\mathbb{P}_{x_0}^{\mu, \sigma}$  and  $\mathbb{P}_{\bar{x}_0}^{\bar{\mu}, \bar{\sigma}}$  belonging to  $\mathbf{P}$ , let  $\tilde{\Pi}(\mathbb{P}^{\mu, \sigma}; \mathbb{P}^{\bar{\mu}, \bar{\sigma}})$  be the set of the probability laws  $\tilde{\mathbb{P}}$  on  $(\Omega^{\otimes 2}, \mathcal{F}^{\otimes 2})$  which enjoy the following properties:

- (i) On some probability space equipped with a pair  $(W, \bar{W})$  of independent Brownian motions and some filtration  $\mathcal{G}$ , there exist  $\mathcal{G}$ -predictable processes  $(C_t)$  and  $(D_t)$  taking values in the space of correlation matrices  $\mathbf{C}_d$  and an  $\mathcal{G}$ -adapted solution  $(X^C, \bar{X})$  to the following system of SDEs:

$$\begin{cases} X_t^C = x_0 + \int_0^t \mu(X_s^C) ds + \int_0^t \sigma(X_s^C) (C_s d\bar{W}_s + D_s dW_s), \\ \bar{X}_t = \bar{x}_0 + \int_0^t \bar{\mu}(\bar{X}_s) ds + \int_0^t \bar{\sigma}(\bar{X}_s) d\bar{W}_s, \end{cases} \quad (1.1)$$

where  $D_s D_s^\top + C_s C_s^\top = \text{Id}_d$  for any  $0 \leq s \leq T$ .

- (ii)  $\tilde{\mathbb{P}}$  is the joint probability law of  $(X^C, \bar{X})$  and therefore the first and second marginal distributions of  $\tilde{\mathbb{P}}$  respectively are  $\mathbb{P}_{x_0}^{\mu, \sigma}$  and  $\mathbb{P}_{\bar{x}_0}^{\bar{\mu}, \bar{\sigma}}$ .

We define the map  $\tilde{\mathcal{W}}^2$  on  $\mathbf{P} \times \mathbf{P}$  as follows: Given  $\mathbb{P}_{x_0}^{\mu, \sigma}$  and  $\mathbb{P}_{\bar{x}_0}^{\bar{\mu}, \bar{\sigma}}$ ,

$$\tilde{\mathcal{W}}^2(\mathbb{P}_{x_0}^{\mu, \sigma}; \mathbb{P}_{\bar{x}_0}^{\bar{\mu}, \bar{\sigma}}) := \left\{ \inf_{\tilde{\mathbb{P}} \in \tilde{\Pi}(\mathbb{P}_{x_0}^{\mu, \sigma}; \mathbb{P}_{\bar{x}_0}^{\bar{\mu}, \bar{\sigma}})} \int_{\Omega^{\otimes 2}} \int_0^T |\omega_s - \bar{\omega}_s|^2 ds \tilde{\mathbb{P}}(d\omega, d\bar{\omega}) \right\}^{\frac{1}{2}}. \quad (1.2)$$

**Remark 1.3.** If the process  $(C_t)$  in (1.1) is predictable w.r.t the (non completed) filtration of the Brownian motion  $(W, \bar{W})$ , there exists a  $\mathcal{F}$ -predictable map  $\Psi$  such that

$$\forall 0 \leq s \leq T, \quad C_s = \Psi(s, \{(W_\theta, \bar{W}_\theta), 0 \leq \theta \leq s\}) \quad (1.3)$$

(see e.g. Claisse et al. [2, Prop.10]).

**Remark 1.4.** When the coefficients  $\mu, \sigma, \bar{\mu}$  and  $\bar{\sigma}$  are Lipschitz, it is easy to prove that for every  $(C_t)$  the system (1.1) has a pathwise unique square integrable strong solution (see, e.g., Fleming and Soner [4, Appendix D]).

We below prove that  $\tilde{\mathcal{W}}^2$  is a distance on  $\mathbf{P}$  which has the following desirable property for modeling and calibration purposes: it metrizes the weak convergence on the space of probability distributions of solutions to SDEs with coefficients sharing the same  $L^\infty$ -norm and Lipschitz constant. Theorem 2.2 shows that  $\tilde{\mathcal{W}}^2$  can be interpreted in terms of a stochastic control problem whose corresponding Hamilton-Jacobi-Bellman equation admits a unique smooth solution.

We will need the following elementary property of  $\tilde{\mathcal{W}}^2$ . Recall that  $\mathcal{W}^2$  was defined in (0.1).

**Proposition 1.5.** *One has*

$$\mathcal{W}^2(\mathbb{P}_{x_0}^{\mu,\sigma}, \mathbb{P}_{\bar{x}_0}^{\bar{\mu},\bar{\sigma}}) \leq \widetilde{\mathcal{W}}^2(\mathbb{P}_{x_0}^{\mu,\sigma}, \mathbb{P}_{\bar{x}_0}^{\bar{\mu},\bar{\sigma}}). \quad (1.4)$$

*Proof.* The process

$$\int_0^t \mathcal{C}_s d\bar{W}_s + \int_0^t \mathcal{D}_s dW_s$$

is a Brownian motion. In view of (1.1), for every  $\tilde{\mathbb{P}}$  in  $\tilde{\Pi}(\mathbb{P}_{x_0}^{\mu,\sigma}, \mathbb{P}_{\bar{x}_0}^{\bar{\mu},\bar{\sigma}})$  the first (resp. second) marginal distribution of  $\tilde{\mathbb{P}}$  is  $\mathbb{P}_{x_0}^{\mu,\sigma}$  (resp.  $\mathbb{P}_{\bar{x}_0}^{\bar{\mu},\bar{\sigma}}$ ). Thus  $\tilde{\Pi}(\mathbb{P}_{x_0}^{\mu,\sigma}, \mathbb{P}_{\bar{x}_0}^{\bar{\mu},\bar{\sigma}}) \subset \Pi(\mathbb{P}_{x_0}^{\mu,\sigma}, \mathbb{P}_{\bar{x}_0}^{\bar{\mu},\bar{\sigma}})$ .  $\square$

### 1.2. $\widetilde{\mathcal{W}}^2$ defines a metric on $\mathbf{P}$

**Proposition 1.6.** *The map  $\widetilde{\mathcal{W}}^2$  defines a distance on  $\mathbf{P}$ .*

*Proof.* We have to prove the triangle inequality

$$\forall \mathbb{P}, \bar{\mathbb{P}}, \underline{\mathbb{P}} \in \mathbf{P}, \quad \widetilde{\mathcal{W}}^2(\mathbb{P}; \underline{\mathbb{P}}) \leq \widetilde{\mathcal{W}}^2(\mathbb{P}; \bar{\mathbb{P}}) + \widetilde{\mathcal{W}}^2(\bar{\mathbb{P}}; \underline{\mathbb{P}}). \quad (1.5)$$

Consider a probability space which supports three independent Brownian motions  $(W_t, \bar{W}_t, \underline{W}_t)$ . Let  $\bar{\mathbb{P}} \in \tilde{\Pi}(\mathbb{P}; \bar{\mathbb{P}})$ .

Let  $\mu$  and  $\sigma$  be the coefficients of the SDE satisfied by the process  $(\underline{X}_t)$  with probability distribution  $\underline{\mathbb{P}}$ . Consider the new system of SDEs

$$\begin{cases} dX_s = \mu(X_s) ds + \sigma(X_s) (\mathcal{C}_s d\bar{W}_s + \mathcal{D}_s dW_s), \\ d\bar{X}_s = \bar{\mu}(\bar{X}_s) ds + \bar{\sigma}(\bar{X}_s) d\bar{W}_s, \\ d\underline{X}_s = \underline{\mu}(\underline{X}_s) ds + \underline{\sigma}(\underline{X}_s) (\underline{\mathcal{C}}_s d\bar{W}_s + \underline{\mathcal{D}}_s d\underline{W}_s), \end{cases} \quad (1.6)$$

where  $\underline{\mathcal{D}}_s \underline{\mathcal{D}}_s^\top + \underline{\mathcal{C}}_s \underline{\mathcal{C}}_s^\top = \text{Id}_d$ . Denote by  $\mathbb{E}$  the expectation under the underlying probability. We obviously have

$$\left\{ \mathbb{E} \int_0^T |X_s - \underline{X}_s|^2 ds \right\}^{1/2} \leq \left\{ \mathbb{E} \int_0^T |X_s - \bar{X}_s|^2 ds \right\}^{1/2} + \left\{ \mathbb{E} \int_0^T |\bar{X}_s - \underline{X}_s|^2 ds \right\}^{1/2}. \quad (1.7)$$

Notice that the process  $W^\sharp$

$$W^\sharp := \left( \int_0^t \underline{\mathcal{C}}_s d\bar{W}_s + \int_0^t \underline{\mathcal{D}}_s d\underline{W}_s, 0 \leq t \leq T \right)$$

is a Brownian motion independent of  $W$ . The process

$$W^b := \left( \int_0^t \mathcal{C}_s d\bar{W}_s + \int_0^t \mathcal{D}_s dW_s, 0 \leq t \leq T \right)$$

is also a Brownian motion. Thus on a possibly extended probability space there exist a Brownian motion  $B$  independent of  $W^\sharp$  and a predictable process  $\mathcal{C}^\sharp$  such that

$$W^b := \left( \int_0^t \mathcal{C}_s^\sharp dW_s^\sharp + \int_0^t \mathcal{D}_s^\sharp dB_s, 0 \leq t \leq T \right)$$

with  $\mathcal{D}_s^\sharp (\mathcal{D}_s^\sharp)^\top + \mathcal{C}_s^\sharp (\mathcal{C}_s^\sharp)^\top = \text{Id}_d$ . We thus have

$$\begin{cases} dX_s = \mu(X_s) ds + \sigma(X_s) (\mathcal{C}_s^\sharp dW_s^\sharp + \mathcal{D}_s^\sharp dB_s), \\ d\underline{X}_s = \underline{\mu}(\underline{X}_s) ds + \underline{\sigma}(\underline{X}_s) dW_s^\sharp. \end{cases}$$

Successively minimizing the right-hand side of (1.7) w.r.t.  $(C_s)$  and  $(\underline{C}_s)$ , and its left-hand side w.r.t.  $(C_s^\sharp)$ , allows us to deduce (1.5).  $\square$

### 1.3. Weak topology metrization with $\widetilde{\mathcal{W}}^2$

In this subsection we need the notion of martingale problem. Given a probability  $\mathbb{P}$  on  $\Omega$ , let  $\mathbb{F}$  denote the augmented and  $\mathbb{P}$ -completed Borel filtration of  $\Omega$ . As Karatzas and Shreve [7, Chap.5, Def.4.5] we here consider that the definition of a solution  $\mathbb{P}$  to a martingale problem involves the  $\mathbb{P}$ -completed canonical filtration denoted by  $(\mathcal{F}_s^\mathbb{P}, 0 \leq s \leq T)$ .

**Definition 1.7.** *A probability measure  $\mathbb{P}$  on  $(\Omega, \mathcal{F})$  is the unique solution to the martingale problem with Borel measurable drift coefficient  $\mu$ , Borel measurable diffusion coefficient  $\sigma$  and initial condition  $x_0$  if for every real valued function  $f$  of class  $C^\infty([0, T] \times \mathbb{R}^d)$  with compact support,*

$$f(t, x_t) - f(0, x_0) - \int_0^t \left( \sum_{i=1}^n \mu^i(x_s) \frac{\partial f}{\partial x_i}(s, x_s) \right) ds - \frac{1}{2} \sum_{i,j} (\sigma(x_s) \sigma(x_s)^\top)^{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(s, x_s) ds = 0 \quad (1.8)$$

*is a  $(\mathbb{P}, (\mathcal{F}_t^\mathbb{P}))$  martingale null at time 0, and if there is no other probability measure on  $(\Omega, \mathcal{F})$  satisfying that property.*

Given positive numbers  $\Lambda$  and  $A$ , let  $\mathcal{A}_{A,\Lambda}$  be the set of functions  $\mu$  and  $\sigma$ , respectively from  $\mathbb{R}^d$  to  $\mathbb{R}^d$  and from  $\mathbb{R}^d$  to  $\mathbf{M}_d$ , which satisfy

$$\begin{cases} |\mu(x)| \leq A, \quad |\sigma(x)| \leq A, \quad \forall x \in \mathbb{R}^d, \\ |\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq \Lambda |x - y|, \quad \forall x, y \in \mathbb{R}^d. \end{cases} \quad (1.9)$$

Denote by  $\mathbf{P}_{A,\Lambda}$  the set of the probability distributions  $\mathbb{P}_{x_0}^{\mu,\sigma}$  of pathwise unique solutions to SDEs with coefficients in  $\mathcal{A}_{A,\Lambda}$  and initial condition  $x_0$  in a given compact subset  $\mathcal{K}$  of  $\mathbb{R}^d$ .

Equipped with the uniform norm topology  $\|\omega\|_\infty := \sup_{0 \leq t \leq T} |\omega(t)|$  the  $d$ -dimensional canonical space is a Polish space. In all the following the weak topology on the set of probability measures on this space is called the weak topology for the  $L_\infty$ -norm.

We will also consider the weak topology for the  $L_2$ -norm  $\|\omega\|_2 := \left\{ \int_0^T |\omega(s)|^2 ds \right\}^{1/2}$ .

**Proposition 1.8.** *Given  $A$  and  $\Lambda$  the set  $\mathbf{P}_{A,\Lambda}$  is a compact subset of the set of probability measures on the  $d$ -dimensional canonical space for the weak topology for the  $L_\infty$ -norm and also for the weak topology for the  $L_2$  norm.*

*Proof.* To simplify the notation we here limit ourselves to the case where the compact set  $\mathcal{K}$  in the definition of  $\mathbf{P}_{A,\Lambda}$  is reduced to the single point  $\{x_0\}$ .

- The first step of the proof consists in proving that the set  $\mathcal{A}_{A,\Lambda}$  is compact in the set of continuous functions for the topology of uniform convergence on compact subsets. The topology of uniform convergence on compact subsets of  $\mathbb{R}^d$  is metrizable. From Bolzano-Weierstrass' theorem it is thus enough to prove that every sequence in  $\mathcal{A}_{A,\Lambda}$  admits a convergent subsequence. Let  $(\mu_k, \sigma_k)$  be a sequence in  $\mathcal{A}_{A,\Lambda}$ . The maps in  $\mathcal{A}_{A,\Lambda}$  are uniformly bounded and uniformly equicontinuous. It follows from Ascoli's theorem that the restriction of  $\mathcal{A}_{A,\Lambda}$  to every compact subset of  $\mathbb{R}^d$  is compact for the uniform convergence. Proceeding then by induction, for every  $n$ , one can construct a subsequence  $(\mu_k^n, \sigma_k^n)$  of  $(\mu_k, \sigma_k)$  such that for all  $n$ ,  $(\mu_k^n, \sigma_k^n)$  is a subsequence of  $(\mu_k^{n-1}, \sigma_k^{n-1})$  and  $(\mu_k^n, \sigma_k^n)$  is uniformly convergent on  $\{(t, x) \mid t \leq T, |x| \leq n\}$ . It follows that the diagonal sequence  $(\mu_n^n, \sigma_n^n)$  admits a limit for the topology of uniform convergence on compact sets of  $\mathbb{R}^d$ . Denote  $(\mu, \sigma)$  its limit. It is then easy to see that  $(\mu, \sigma) \in \mathcal{A}_{A,\Lambda}$ .
- The set of probability measures on the  $d$ -dimensional canonical space endowed with the weak topology for the  $L_\infty$ -norm is metrizable and once again we can apply Bolzano-Weierstrass' theorem. Let  $(\mathbb{P}_k)$  be a sequence in  $\mathbf{P}_{A,\Lambda}$ . Such a family of probability measures  $(\mathbb{P}_k)$  is obviously tight for the weak topology for the  $L_\infty$ -norm (see e.g. [7, Chap.2, Pb.4.11]). Therefore there exists a subsequence  $(\mathbb{P}_{\varphi(k)})$  of  $(\mathbb{P}_k)$  converging to some probability measure  $\mathbb{P}$  for the weak topology for the  $L_\infty$ -norm. For all  $k$ ,  $\mathbb{P}_{\varphi(k)} = \mathbb{P}^{\mu_k, \sigma_k}$  with  $(\mu_k, \sigma_k) \in \mathcal{A}_{A,\Lambda}$ . It follows from the first step of the proof that there exists a subsequence  $(\tilde{\mu}_k, \tilde{\sigma}_k)$  of  $(\mu_k, \sigma_k)$  converging to  $(\mu, \sigma) \in \mathcal{A}_{A,\Lambda}$  uniformly on compact sets. For every  $k$  one has that  $\mathbb{P}_{\varphi(k)}$  is the probability distribution of the SDE with coefficients  $\mu_k$  and  $\sigma_k$  and initial condition  $x_0$ . Choose a subsequence of  $(\tilde{\mu}_k, \tilde{\sigma}_k)$  with limit  $(\mu, \sigma)$  and pass to the limit in (1.8) for every function  $f$  in  $\mathcal{C}_\infty(\mathbb{R}^d)$  with compact support. It results that  $\mathbb{P}$  is a solution to the martingale problem with coefficients  $\mu$  and  $\sigma$  and initial condition  $x_0$ . From the uniqueness of the solution we get  $\mathbb{P} = \mathbb{P}^{\mu, \sigma}$ . This proves the compactness of  $\mathbf{P}_{A,\Lambda}$  for the weak topology for the  $L_\infty$ -norm.
- By definition of the weak topology, for every bounded continuous function  $\psi$  on  $(\mathcal{C}([0, T], \mathbb{R}^d), |\cdot|_\infty)$ ,  $\mathbb{E}^\mathbb{P}(\psi) = \lim \mathbb{E}^{\mathbb{P}_{\varphi(k)}}(\psi)$ . For all paths  $\omega, \omega'$ ,  $|\omega - \omega'|_{L_2} \leq T|\omega - \omega'|_{L_\infty}$ . It follows that the restriction to  $\mathcal{C}([0, T], \mathbb{R}^d)$  of every continuous function  $\psi$  on  $L_2([0, T], \mathbb{R}^d)$  is also continuous on  $\mathcal{C}([0, T], \mathbb{R}^d)$  for the uniform norm topology. This proves that  $\mathbb{P}_{\varphi(k)}$  converges to  $\mathbb{P}$  for the weak topology for the  $L_2$  norm.

□

**Proposition 1.9.**  $\widetilde{\mathcal{W}}^2$  metrizes the weak topology on  $\mathbf{P}_{A,\Lambda}$  for the  $L^2$ -norm.

*Proof.* In all this proof, we will shortly write ‘weak topology’ instead of ‘weak



topology for the  $L^2$ -norm'. To simplify the notation we again limit ourselves to the case where the compact set  $\mathcal{K}$  in the definition of  $\mathbf{P}_{A,\Lambda}$  is reduced to the single point  $\{x_0\}$ .

The set of probability measures on  $L_2([0, T], \mathbb{R}^d)$  endowed with the weak topology is a Polish space, thus the subset  $\mathbf{P}_{A,\Lambda}$  is metrizable separable for the weak topology. Therefore it is enough to prove that for all  $(\mathbb{P}_k)_{k \in \mathbb{N}}, \mathbb{P}$  in  $\mathbf{P}_{A,\Lambda}$ ,  $\mathbb{P}$  is the limit of  $\mathbb{P}_k$  for the weak topology if and only if  $\widetilde{\mathcal{W}}^2(\mathbb{P}_k, \mathbb{P}) \rightarrow 0$ .

• Step 1.

Let  $\mathbb{P}_k \in \mathbf{P}_{A,\Lambda}$  converging to  $\mathbb{P}$  for the weak topology. It follows from Proposition 1.8 that the probability measure  $\mathbb{P}$  belongs to  $\mathbf{P}_{A,\Lambda}$ . We will prove by contradiction that  $\widetilde{\mathcal{W}}_2(\mathbb{P}_k, \mathbb{P})$  tends to 0.

Assume that this is false. It follows from the proof of Proposition 1.8 that there is  $\eta > 0$  and a subsequence  $\widetilde{\mathbb{P}}_k = \mathbb{P}^{\mu_k, \sigma_k}$  of  $\mathbb{P}_k$  such that  $(\mu_k, \sigma_k)$  is converging to  $(\mu, \sigma) \in \mathcal{A}_{A,\Lambda}$  uniformly on compact sets,  $\mathbb{P} = \mathbb{P}^{\mu, \sigma}$  and such that  $\inf_k \widetilde{\mathcal{W}}_2(\widetilde{\mathbb{P}}_k, \mathbb{P}) = \eta > 0$ . For every  $k$  let  $Y^k$  and  $Y$  be the stochastic processes taking values in  $\mathbb{R}^d$  respectively solutions to

$$\begin{cases} Y_t^k = x_0 + \int_r^t \mu_k(Y_s^k) ds + \int_r^t \sigma_k(Y_s^k) d\overline{W}_s, \\ Y_t = x_0 + \int_r^t \mu(Y_s) ds + \int_r^t \sigma(Y_s) d\overline{W}_s. \end{cases} \quad (1.10)$$

Using Cauchy-Schwartz and Burkholder-Davis-Gundy's inequalities we get

$$\begin{aligned} \mathbb{E} \sup_{r \leq s \leq t} |Y_s^k - Y_s|^2 &\leq 2(t-r) \mathbb{E} \int_r^t |\mu_k(Y_s^k) - \mu(Y_s)|^2 ds \\ &\quad + 8 \mathbb{E} \int_r^t |\sigma_k(Y_s^k) - \sigma(Y_s)|^2 ds. \end{aligned} \quad (1.11)$$

As  $\mu_k$  and  $\sigma_k$  belong to  $\mathcal{A}_{A,\Lambda}$  it follows from (1.11) that

$$\begin{aligned} \mathbb{E}(\sup_{r \leq s \leq t} |Y_s^k - Y_s|^2) &\leq (4(t-r) + 16) \Lambda^2 \mathbb{E} \int_r^t |Y_s^k - Y_s|^2 ds \\ &\quad + 4(t-r) \mathbb{E} \int_r^t |\mu_k(Y_s) - \mu(Y_s)|^2 ds \\ &\quad + 16 \mathbb{E} \int_r^t |\sigma_k(Y_s) - \sigma(Y_s)|^2 ds. \end{aligned} \quad (1.12)$$

We have  $\lim_{K \rightarrow \infty} \mathbb{E}\{\sup_{s \leq T} |Y_s| \geq K\} \rightarrow 0$ . On the other hand, the functions  $\mu_k$ , respectively  $\sigma_k$ , converge uniformly to  $\mu$ , respectively  $\sigma$ , on every compact set of  $\mathbb{R}^d$  and are uniformly bounded on  $\mathbb{R}^d$ . It follows that for all  $\epsilon > 0$  there is  $k_0$  such that for all  $k \geq k_0$ ,  $F_k(t) \leq \epsilon + M \int_r^t F_k(s) ds$  with  $F_k(s) = \mathbb{E} \sup_{u \leq s} |Y_u^k - Y_u|^2$ . By Gronwall's Lemma it comes that  $F_k(t) \leq \epsilon \exp(M(t-r))$ . From the definition (1.2) of  $\widetilde{\mathcal{W}}^2$  it follows that  $\widetilde{\mathcal{W}}^2(\mathbb{P}_k, \mathbb{P}) \rightarrow 0$ . This proves the desired result by contradiction.

- Step 2.

Let  $\mathbb{P}_k, \mathbb{P}$  in  $\mathbf{P}_{A,\Lambda}$  such that  $\widetilde{\mathcal{W}}^2(\mathbb{P}_k, \mathbb{P}) \rightarrow 0$ . It follows from (1.4) that  $\mathcal{W}^2(\mathbb{P}_k, \mathbb{P}) \rightarrow 0$ . The distance  $\mathcal{W}^2$  metrizes the weak topology, therefore  $\mathbb{P}$  is the limit of  $\mathbb{P}_k$  for the weak topology.

□

## 2. Interpretation of $\widetilde{\mathcal{W}}^2(\mathbb{P}_x^{\mu,\sigma}; \mathbb{P}_{\bar{x}}^{\bar{\mu},\bar{\sigma}})$ in terms of stochastic control

This section is aimed to prove that the value of  $\widetilde{\mathcal{W}}^2(\mathbb{P}_x^{\mu,\sigma}; \mathbb{P}_{\bar{x}}^{\bar{\mu},\bar{\sigma}})$  can be obtained by solving the following stochastic control problem.

Recall that  $\mathbf{C}_d$  (resp.  $\mathbf{O}_d$ ) denotes the space of  $d \times d$  correlation (resp. orthogonal) matrices. For every  $0 \leq t \leq T$  let  $\mathbf{Ad}(t, T)$  denote the set of admissible controls between  $t$  and  $T$ , that is, the set of  $\mathcal{G}$ -predictable processes on  $[t, T]$  which take values in  $\mathbf{C}_d$  and are independent of  $\mathcal{G}_t$ .

Given a time origin  $0 \leq t \leq T$  and a control process  $(C_\theta)$  in  $\mathbf{Ad}(t, T)$  depending in a Borel measurable way on the initial conditions  $x$  and  $\bar{x}$ , in view of Remark 1.4 the hypotheses made below (see Theorem 2.1) ensure that there exists a pathwise unique solution  $(X_\theta^C(t, x, \bar{x}), \bar{X}_\theta(t, x))$  to

$$\forall t \leq \theta \leq T, \quad \begin{cases} X_\theta^C = x + \int_t^\theta \mu(X_s^C) ds + \int_t^\theta \sigma(X_s^C) (C_s d\bar{W}_s + \mathcal{D}_s dW_s), \\ \bar{X}_\theta = \bar{x} + \int_t^\theta \bar{\mu}(\bar{X}_s) ds + \int_t^\theta \bar{\sigma}(\bar{X}_s) d\bar{W}_s, \end{cases} \quad (2.1)$$

where  $\mathcal{D}_s \mathcal{D}_s^\top + C_s C_s^\top = \text{Id}_d$  for any  $t \leq s \leq T$ . Choosing the time origin  $t = 0$  we consider the objective function

$$\min_{(C_\theta) \in \mathbf{Ad}(0, T)} \mathbb{E} \int_0^T |X_\theta^C(0, x, \bar{x}) - \bar{X}_\theta(0, \bar{x})|^2 d\theta. \quad (2.2)$$

The corresponding Hamilton–Jacobi–Bellman equation is the following:

$$\begin{cases} \partial_t V(t, x, \bar{x}) + \mathcal{L}V(t, x, \bar{x}) + H(t, x, \bar{x}, V) = -|x - \bar{x}|^2, & 0 \leq t < T, \\ V(T, x, \bar{x}) = 0, \end{cases} \quad (2.3)$$

where

$$\begin{aligned} \mathcal{L}V(t, x, \bar{x}) &:= \sum_{i=1}^d \mu^i(x) \partial_{x_i} V(t, x, \bar{x}) + \sum_{i=1}^d \bar{\mu}^i(\bar{x}) \partial_{\bar{x}_i} V(t, x, \bar{x}) \\ &+ \frac{1}{2} \sum_{i,j=1}^d (\sigma \sigma^\top(x))^{ij} \partial_{x_i x_j}^2 V(t, x, \bar{x}) + \frac{1}{2} \sum_{i,j=1}^d (\bar{\sigma}(\bar{x}) \bar{\sigma}(\bar{x})^\top)^{ij} \partial_{\bar{x}_i \bar{x}_j}^2 V(t, x, \bar{x}) \end{aligned} \quad (2.4)$$

and

$$H(t, x, \bar{x}, V) := \min_{C \in \mathbf{C}_d} \sum_{i,j=1}^d (\sigma(x) C \bar{\sigma}(\bar{x})^\top)^{ij} \partial_{x_i \bar{x}_j}^2 V(t, x, \bar{x}).$$

In view of (4.4) below we have

$$H(t, x, \bar{x}, V) = \min_{C \in \mathbf{O}_d} \sum_{i,j=1}^d (\sigma(x) C \bar{\sigma}(\bar{x})^\top)^{ij} \partial_{x_i, \bar{x}_j}^2 V(t, x, \bar{x}). \quad (2.5)$$

We aim to prove:  $V(0, x, \bar{x})$  coincides with the objective function (2.2) and with  $(\widehat{\mathcal{W}}^2(\mathbb{P}^{\mu, \sigma}; \mathbb{P}^{\bar{\mu}, \bar{\sigma}}))^2$  and there exists an optimal admissible control. As explained in the introduction, this result can be used to get the numerical value of  $\widehat{\mathcal{W}}^2(\mathbb{P}^{\mu, \sigma}; \mathbb{P}^{\bar{\mu}, \bar{\sigma}})$  by numerically solving the PDE (2.3) and thus is relevant for the selection of a simplified model which approximates a given exact model  $\mathbb{P}^{\bar{\mu}, \bar{\sigma}}$  in  $\mathbf{P}$ .

The two main results are the following.

**Theorem 2.1.** *Suppose:*

- (i) *The functions  $\mu, \bar{\mu}, \sigma$  and  $\bar{\sigma}$  are in the Hölder space  $\mathcal{C}^\alpha(\mathbb{R}^d)$  with  $0 < \alpha < 1$ .*
- (ii) *The matrix-valued functions  $a(x) := \sigma(x)\sigma(x)^\top$  and  $\bar{a}(x) := \bar{\sigma}(x)\bar{\sigma}(x)^\top$  satisfy the strong ellipticity condition*

$$\exists \lambda > 0, \forall (\xi, \bar{\xi}, x) \in \mathbb{R}^{3d}, \quad \sum_{i,j=1}^d a^{ij}(x) \xi^i \xi^j + \sum_{i,j=1}^d \bar{a}^{ij}(x) \bar{\xi}^i \bar{\xi}^j \geq \lambda(|\xi|^2 + |\bar{\xi}|^2).$$

*Then there exists a solution  $V(t, x, \bar{x})$  to (2.3) in  $\mathcal{C}^{1,2}([0, T] \times \mathbb{R}^d)$  such that  $\frac{V(t, x, \bar{x})}{1+|x|^2+|\bar{x}|^2}$  is in  $\mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}([0, T] \times \mathbb{R}^{2d})$ .*

To prove the preceding theorem, we have to face several technical difficulties. We cannot apply classical results such as, e.g., Fleming and Soner [4, Thm.IV-4-2], because the control is involved in the diffusion coefficient and, which is a much more critical issue, this coefficient is possibly degenerate (e.g., in the 1D-case the optimal control is  $\mathcal{C}_t^* \equiv 1$ : See Subsection 2.1 below). We will circumvent the degeneracy by considering that the  $\partial_{x_i, \bar{x}_j}^2$  part is a perturbation of a strongly elliptic operator.

The next theorem shows that  $V(t, x, \bar{x})$  is the value function of a stochastic control problem for which an optimal control exists.

**Theorem 2.2.** *Under the hypotheses of Theorem 2.1 there exist a filtered probability space equipped with two independent standard Brownian motions  $W$  and  $\bar{W}$  and a predictable process  $(\mathcal{C}^*, \mathcal{D}^*)$  such that there exists an adapted and continuous solution  $(X^*(t, x, \bar{x}), \bar{X}(t, \bar{x}))$  on  $[t, T]$  to the system*

$$\forall t \leq \theta \leq T, \quad \begin{cases} X_\theta^* = x + \int_t^\theta \mu(X_s^*) ds + \int_t^\theta \sigma(X_s^*) C_s^* d\bar{W}_s + \int_t^\theta \sigma(X_s^*) \mathcal{D}_s^* dW_s, \\ \bar{X}_\theta = \bar{x} + \int_t^\theta \bar{\mu}(\bar{X}_s) ds + \int_t^\theta \bar{\sigma}(\bar{X}_s) d\bar{W}_s, \\ C_s^* \in \arg \min_{C \in \mathbf{C}_d} \sum_{i,j=1}^d \left( \sigma(X_s^*) C \bar{\sigma}(\bar{X}_s)^\top \right)^{ij} \partial_{x_i, \bar{x}_j}^2 V(s, X_s^*, \bar{X}_s), \\ \mathcal{C}_s^* (\mathcal{C}_s^*)^\top + \mathcal{D}_s^* (\mathcal{D}_s^*)^\top = \text{Id}_d, \end{cases} \quad (2.6)$$

which satisfies

$$V(0, x, \bar{x}) = \mathbb{E} \int_0^T |X_\theta^*(0, x, \bar{x}) - \bar{X}_\theta(0, \bar{x})|^2 d\theta = (\widetilde{\mathcal{W}}^2(\mathbb{P}_x^{\mu, \sigma}; \mathbb{P}_{\bar{x}}^{\bar{\mu}, \bar{\sigma}}))^2.$$

In addition, there exists a sequence  $(\mathbb{P}^{\mu^m, \sigma^m})$  of solutions to martingale problems with bounded continuous coefficients  $\mu^m$  and  $\sigma^m$  such that

$$\int_\Omega \int_0^T |\omega_s - \bar{\omega}_s|^2 ds \mathbb{P}^{\mu^m, \sigma^m}(d\omega, d\bar{\omega})$$

converges to  $(\widetilde{\mathcal{W}}^2(\mathbb{P}^{\mu, \sigma}; \mathbb{P}^{\bar{\mu}, \bar{\sigma}}))^2$ .

We start with examining the one-dimensional case  $d = 1$  which allows simple arguments and permits to see the difficulties to overcome to obtain the general statement of Theorem 2.2.

### 2.1. The one-dimensional case

Suppose that the functions  $\mu, \sigma, a, \bar{\mu}, \bar{\sigma}$  and  $\bar{a}$  satisfy the hypotheses of Theorem 2.1. Consider the family of stochastic differential equations

$$\forall t \leq \theta \leq T, \quad X_\theta^* = x + \int_t^\theta \mu(X_s^*) ds + \int_t^\theta \sigma(X_s^*) d\bar{W}_s.$$

Denote by  $(X_\theta^*(t, x))$  the corresponding pathwise unique stochastic flows. Under our assumptions it is well-known (see Kunita [9]) that a.s. the map  $x \mapsto X_\theta^*(t, x)$  is differentiable for any  $\theta$  and for any  $t \leq \theta \leq T$  one has

$$\frac{d}{dx} X_\theta^*(t, x) = \exp \left( \int_t^\theta (\mu' + \frac{1}{2}(\sigma')^2)(X_s^*(t, x)) ds + \int_t^\theta \sigma'(X_s^*(t, x)) d\bar{W}_s \right). \quad (2.7)$$

Of course, a similar result holds true for the flows  $(\bar{X}_\theta(t, \bar{x}), t \leq \theta \leq T)$  solutions to

$$\forall t \leq \theta \leq T, \quad \bar{X}_\theta = \bar{x} + \int_t^\theta \bar{\mu}(\bar{X}_s) ds + \int_t^\theta \bar{\sigma}(\bar{X}_s) d\bar{W}_s.$$

Now consider the function

$$V^*(t, x, \bar{x}) := \mathbb{E} \int_t^T (X_\theta^*(t, x) - \bar{X}_\theta(t, \bar{x}))^2 d\theta.$$

Under our hypotheses this function is the unique classical solution to the following parabolic PDE on  $[0, T] \times \mathbb{R}^2$ :

$$\begin{cases} \frac{\partial V^*}{\partial t}(t, x, \bar{x}) + \mu(x) \frac{\partial V^*}{\partial x}(t, x, \bar{x}) + \bar{\mu}(\bar{x}) \frac{\partial V^*}{\partial \bar{x}}(t, x, \bar{x}) + \frac{1}{2} \sigma^2(x) \frac{\partial^2 V^*}{\partial x^2}(t, x, \bar{x}) \\ \quad + \frac{1}{2} \bar{\sigma}^2(\bar{x}) \frac{\partial^2 V^*}{\partial \bar{x}^2}(t, x, \bar{x}) + \sigma(x) \bar{\sigma}(\bar{x}) \frac{\partial^2 V^*}{\partial x \partial \bar{x}}(t, x, \bar{x}) = -(x - \bar{x})^2, \quad t < T, \\ V^*(T, x, \bar{x}) = 0. \end{cases}$$

Notice that

$$\frac{\partial^2 V^*}{\partial x \partial \bar{x}}(t, x, \bar{x}) = -2 \int_t^T \mathbb{E} \left[ \frac{d}{dx} X_s^*(t, x) \frac{d}{d\bar{x}} \bar{X}_s(\bar{x}) \right] ds.$$

In view of (2.7) we thus have

$$\forall t, x, \bar{x}, \quad \frac{\partial^2 V^*}{\partial x \partial \bar{x}}(t, x, \bar{x}) < 0. \quad (2.8)$$

Therefore we have exhibited a classical solution  $V^*(t, x, \bar{x})$  to the following Hamilton-Jacobi-Bellman equation on  $[0, T] \times \mathbb{R}^2$ :

$$\begin{cases} \frac{\partial V}{\partial t}(t, x, \bar{x}) + \mu(x) \frac{\partial V}{\partial x}(t, x, \bar{x}) + \bar{\mu}(\bar{x}) \frac{\partial V}{\partial \bar{x}}(t, x, \bar{x}) + \frac{1}{2} \sigma^2(x) \frac{\partial^2 V}{\partial x^2}(t, x, \bar{x}) \\ + \frac{1}{2} \bar{\sigma}^2(\bar{x}) \frac{\partial^2 V}{\partial \bar{x}^2}(t, x, \bar{x}) + \min_{C \in [-1, 1]} \left( C \sigma(x) \bar{\sigma}(\bar{x}) \frac{\partial^2 V}{\partial x \partial \bar{x}}(t, x, \bar{x}) \right) = -(x - \bar{x})^2, \quad t < T, \\ V(T, x, \bar{x}) = 0. \end{cases}$$

Observe that here the admissible controls in  $\mathbf{Ad}(\mathbf{t}, \mathbf{T})$  are valued in  $[-1, 1]$ . Thus the constant correlation process  $C_t^* \equiv 1$  ( $0 \leq t \leq T$ ) is admissible. Therefore standard verification theorems in stochastic control theory (see e.g. Fleming and Soner [4, Thm.IV.3.1]) lead to

$$V^*(t, x, \bar{x}) = \min_{(C_\theta) \in \mathbf{Ad}(\mathbf{t}, \mathbf{T})} \mathbb{E} \int_t^T (X_\theta^C(t, x) - \bar{X}_\theta(t, x))^2 d\theta.$$

Since  $X_\theta^{C^*}(t, x) = X_\theta^*(t, x)$  for all  $t, \theta, x$ , the function  $V^*$  also satisfies

$$V^*(t, x, \bar{x}) = \mathbb{E} \int_t^T (X_\theta^{C^*}(t, x) - \bar{X}_\theta(t, x))^2 d\theta.$$

Notice that the probability distribution of  $(X_\theta^{C^*}(t, x))$  solves a martingale problem with coefficients  $\mu$  and  $\sigma$  and therefore belongs to  $\mathbf{P}$ . In view of the definition (1.2) we deduce that

$$V^*(0, x_0, x_0) = (\widehat{\mathcal{W}}^2(\mathbb{P}^{\mu, \sigma}; \mathbb{P}^{\bar{\mu}, \bar{\sigma}}))^2.$$

In the preceding, we have exhibited an optimal control which does not depend on the solution to the Hamilton-Jacobi-Bellman equation, and for which the cost function  $V^*$  is smooth. In the multi-dimensional case the situation is not so simple as we now see.

## 2.2. The multi-dimensional case: Preliminaries

For the reader's convenience we recall a classical result in PDE analysis (see e.g. Ladyzenskaja et al. [10, Chap.IV, Thm.5.1] or Lunardi [11, Thm.5.1.8]).

**Theorem 2.3.** *If the elliptic operator*

$$\mathcal{A} := \sum_{i=1}^d b^i(x) \partial_{x_i} + \frac{1}{2} \sum_{i=1}^d \gamma^{ij}(x) \partial_{x_i x_j}^2 + c(x)$$

*with coefficients in  $\mathcal{C}^\alpha(\mathbb{R}^d)$  ( $0 < \alpha < 2$ ,  $\alpha \neq 1$ ) is uniformly strongly elliptic, the source term  $f$  is in  $\mathcal{C}^{\frac{\alpha}{2}, \alpha}(\mathbb{R}^d)$  and the initial condition  $\psi$  is in  $\mathcal{C}^{2+\alpha}(\mathbb{R}^d)$ , then the parabolic problem on  $[0, T] \times \mathbb{R}^d$*

$$\begin{cases} \partial_t u = \mathcal{A}u + f & \text{in } [0, T] \times \mathbb{R}^d, \\ u(0, \cdot) = \psi(\cdot) & \text{in } \mathbb{R}^d, \end{cases} \quad (2.9)$$

*has a unique solution in  $\mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}([0, T] \times \mathbb{R}^d)$ . Furthermore, denoting by  $\|\cdot\|_{\frac{\alpha}{2}, \alpha}(s)$  the Hölder norm of  $\mathcal{C}^{\frac{\alpha}{2}, \alpha}([0, s] \times \mathbb{R}^d)$ , we have*

$$\forall 0 \leq t < T, \quad \|u\|_{1+\frac{\alpha}{2}, 2+\alpha}(t) \leq K(T) \|\psi\|_{2+\alpha} + K(T) \int_0^t \|f\|_{\frac{\alpha}{2}, \alpha}(s) ds, \quad (2.10)$$

*where the positive increasing function  $K$  depends on the coefficients of  $\mathcal{A}$  only.*

We easily deduce the following corollary.

**Corollary 2.4.** *Suppose that the coefficients of  $\mathcal{A}$  satisfy the hypotheses of Theorem 2.3. Suppose that the function  $f$  satisfies: there exist  $C > 0$ ,  $r \geq 1$ ,  $0 < \alpha < 1$  such that*

$$\begin{cases} \forall x, x' \in \mathbb{R}^d, \forall 0 \leq t, t' \leq T, \\ |f(t, x)| \leq C (1 + |x|^2)^r, \\ |f(t, x) - f(t', x')| \leq C (|t - t'|^\frac{\alpha}{2} + |x - x'|^\alpha) (1 + |x|^2 + |x'|^2)^r. \end{cases} \quad (2.11)$$

*Then there exists a unique function  $u(t, x)$  which satisfies (2.9) and is such that  $\tilde{u}(t, x) := \frac{u(t, x)}{(1 + |x|^2)^r}$  is in  $\mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}([0, T] \times \mathbb{R}^d)$ . In addition,*

$$\forall 0 \leq t < T, \quad \|\tilde{u}\|_{1+\frac{\alpha}{2}, 2+\alpha}(t) \leq \tilde{K}(T) \|\tilde{\psi}\|_{2+\alpha} + \tilde{K}(T) \int_0^t \|\tilde{f}\|_{\frac{\alpha}{2}, \alpha}(s) ds, \quad (2.12)$$

*where the positive increasing function  $\tilde{K}$  depends on the coefficients of  $\mathcal{A}$  only and the functions  $\tilde{\psi}$  and  $\tilde{f}$  are respectively defined as  $\tilde{\psi}(x) := \frac{\psi(x)}{(1 + |x|^2)^r}$  and  $\tilde{f}(t, x) := \frac{f(t, x)}{(1 + |x|^2)^r}$ .*

*Proof.* Define the differential operator  $\tilde{\mathcal{A}}$  by

$$\tilde{\mathcal{A}}\phi(x) := \frac{1}{(1 + |x|^2)^r} \mathcal{A}(\phi(x) (1 + |x|^2)^r).$$

For some functions  $\tilde{b}^i$  and  $\tilde{c}$  which can easily be explicited in terms of the coefficients of  $\mathcal{A}$  one has

$$\tilde{\mathcal{A}} := \sum_{i=1}^d \tilde{b}^i(x) \partial_{x_i} + \frac{1}{2} \sum_{i=1}^d \gamma^{ij}(x) \partial_{x_i x_j}^2 + \tilde{c}(x).$$

Notice that the leading coefficients of  $\tilde{\mathcal{A}}$  are identical to those of  $\mathcal{A}$  and thus  $\tilde{\mathcal{A}}$  is strongly uniformly elliptic. Now consider the parabolic equation

$$\begin{cases} \partial_t \tilde{u}(t, x) = \tilde{\mathcal{A}}\tilde{u}(t, x) + \frac{f(t, x)}{(1+|x|^2)^r} & \text{in } [0, T] \times \mathbb{R}^d, \\ \tilde{u}(0, x) = \frac{\psi(x)}{(1+|x|^2)^r} & \text{in } \mathbb{R}^d. \end{cases} \quad (2.13)$$

In order to be in a position to apply Theorem 2.3 it clearly suffices to prove that the coefficients  $\tilde{b}^i$  and  $\tilde{c}$  are in  $\mathcal{C}^\alpha(\mathbb{R}^d)$  ( $0 < \alpha < 2$ ,  $\alpha \neq 1$ ) and the function  $\frac{f(t, x)}{(1+|x|^2)^r}$  is in  $\mathcal{C}^{\frac{\alpha}{2}, \alpha}(\mathbb{R}^d)$ . We only consider the last function since similar arguments apply to the other ones.

W.l.o.g. we may and do suppose that  $|x| \leq |x'|$ . Denoting by  $C$  any constant independent of  $x, x', t, t'$ , one has

$$\begin{aligned} \left| \tilde{f}(t, x) - \tilde{f}(t, x') \right| &\leq \frac{|f(t, x) - f(t, x')|}{(1+|x'|^2)^r} + \frac{(1+|x'|^2)^r - (1+|x|^2)^r}{(1+|x|^2)^r (1+|x'|^2)^r} |f(t, x)| \\ &\leq C|x' - x|^\alpha + C|x' - x|^\alpha \frac{|x' - x|^{1-\alpha}(|x'| + |x|)(1+|x'|^2)^{r-1}}{(1+|x'|^2)^r} \\ &\leq C|x' - x|^\alpha. \end{aligned}$$

We also obviously have

$$\left| \tilde{f}(t', x) - \tilde{f}(t, x) \right| \leq C|t' - t|^{\frac{\alpha}{2}}.$$

Setting  $\tilde{u}(t, x) := \frac{u(t, x)}{(1+|x|^2)^r}$  we have obtained the desired existence and uniqueness result.  $\square$

**Remark 2.5.** It is well known (see e.g. Friedman [5, Chap.1, Sec.7, Eq.(7.6)]) that, under the hypotheses of Theorem 2.3, one has

$$u(t, x) = \int_{\mathbb{R}^d} \Gamma(t, x; 0, \xi) \psi(\xi) d\xi - \int_0^t \int_{\mathbb{R}^d} \Gamma(t, x; s, \xi) f(s, \xi) d\xi ds, \quad (2.14)$$

where  $\Gamma$  is the fundamental solution associated to the operator  $\mathcal{A}$ . By using classical properties of  $\Gamma$  (see Friedman [5, Chap.9, Sec.4, Thm.2 and Ineq.(4.19)] and Ladyzenskaja et al. [10, Chap.IV, Sect.13]) one can check that this representation also holds true under the hypotheses of Corollary 2.4. We will not need this property.

### 2.3. Proof of Theorem 2.1

Let  $\tilde{v}$  be a function in  $\mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}([0, T] \times \mathbb{R}^{2d})$ . Define the function  $v$  as  $v(t, x, \bar{x}) := \tilde{v}(T - t, x, \bar{x}) (1 + |x|^2 + |\bar{x}|^2)$ . Notice that then  $v$  necessarily satisfies (2.11). The functions  $\partial_{x_i, \bar{x}_j}^2 v(t, x, \bar{x})$  satisfy the inequalities (2.11). Apply Proposition 4.8 in our Appendix. It follows that  $H(t, x, \bar{x}, v)$  satisfies the inequalities (2.11). In

addition, there exists  $\tilde{C}_1 > 0$  depending on Hölder norms of  $\sigma$  and  $\bar{\sigma}$  only such that

$$\forall 0 \leq t \leq T, \quad \left\| \frac{H(t, x, \bar{x}, (1 + |x|^2 + |\bar{x}|^2) \tilde{v})}{1 + |x|^2 + |\bar{x}|^2} \right\|_{\frac{\alpha}{2}, \alpha} (t) \leq \tilde{C}_1 \|\tilde{v}\|_{1+\frac{\alpha}{2}, 2+\alpha}(t).$$

To transform the equation (2.3) into a time forward equation similar to (2.9) one changes  $t$  into  $T - t$ . Observe that

$$H(t, x, \bar{x}, (1 + |x|^2 + |\bar{x}|^2) \tilde{v}) = H(T - t, x, \bar{x}, v).$$

This leads one to consider the following PDE:

$$\begin{cases} \partial_t \tilde{V}(t, x, \bar{x}) = \tilde{\mathcal{L}} \tilde{V}(t, x, \bar{x}) + \frac{H(t, x, \bar{x}, (1 + |x|^2 + |\bar{x}|^2) \tilde{v})}{1 + |x|^2 + |\bar{x}|^2} + \frac{|x - \bar{x}|^2}{1 + |x|^2 + |\bar{x}|^2}, & 0 < t \leq T, \\ \tilde{V}(0, x, \bar{x}) = 0, \end{cases} \quad (2.15)$$

where the coefficients of the differential operator  $\tilde{\mathcal{L}}$  can be explicited in terms of those of  $\mathcal{L}$  by proceeding as in the proof of Corollary 2.4. Set

$$\tilde{C}_0 := \left\| \frac{|x - \bar{x}|^2}{1 + |x|^2 + |\bar{x}|^2} \right\|_{\alpha}.$$

Corollary 2.4 shows that (2.15) has a unique solution denoted by  $\tilde{\psi}(\tilde{v})$  such that

$$\begin{aligned} \|\tilde{\psi}(\tilde{v})\|_{1+\frac{\alpha}{2}, 2+\alpha}(t) &\leq \tilde{K}(T) \int_0^t \left\| \frac{H(s, x, \bar{x}, (1 + |x|^2 + |\bar{x}|^2) \tilde{v})}{1 + |x|^2 + |\bar{x}|^2} \right\|_{\frac{\alpha}{2}, \alpha} (s) ds \\ &\quad + \tilde{K}(T) \tilde{C}_0 t, \\ &\leq \tilde{K}(T) \tilde{C}_1 \int_0^t \|\tilde{v}\|_{1+\frac{\alpha}{2}, 2+\alpha}(s) ds + \tilde{K}(T) \tilde{C}_0 t, \end{aligned} \quad (2.16)$$

where  $\tilde{K}(T)$  depends on  $T$  and the coefficients of  $\mathcal{L}$  only.

Let

$$\mathbb{K} := \{\tilde{g} \in \mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}([0, T] \times \mathbb{R}^d);$$

$$\forall 0 \leq t \leq T, \quad \|\tilde{g}\|_{1+\frac{\alpha}{2}, 2+\alpha}(t) \leq \frac{\tilde{C}_0}{\tilde{C}_1} (\exp(\tilde{K}(T) \tilde{C}_1 t) - 1)\}.$$

Let  $\tilde{v}$  belong to  $\mathbb{K}$ . In view of (2.16) we get that  $\tilde{\psi}(\tilde{v})$  belongs to  $\mathbb{K}$ .

Let us now prove that the map  $\tilde{\psi}$  is continuous on  $\mathbb{K}$ . Let  $\tilde{v}_n$  and  $\tilde{v}$  in  $\mathbb{K}$  such that  $\|\tilde{v}_n - \tilde{v}\|_{1+\frac{\alpha}{2}, 2+\alpha}$  tends to 0 when  $n$  goes to infinity. Let us show that  $\tilde{\psi}(\tilde{v})$  is the unique accumulation point of  $(\tilde{\psi}(\tilde{v}_n))$ . Consider a converging subsequence  $(\tilde{w}_n)$  of  $(\tilde{\psi}(\tilde{v}_n))$  (such a subsequence exists since  $\mathbb{K}$  is a compact subset of  $\mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}([0, T] \times \mathbb{R}^{2d})$ ) and let  $\tilde{w}$  be such that  $\|\tilde{w} - \tilde{w}_n\|_{1+\frac{\alpha}{2}, 2+\alpha}$  tends to 0 when  $n$  goes to infinity. Observe that  $\partial_t \tilde{w}_n$ , respectively  $\tilde{\mathcal{L}} \tilde{w}_n$ , converges to



$\partial_t \tilde{w}$ , respectively  $\mathcal{L}\tilde{w}$ , uniformly on compact sets. In addition, Proposition 4.8 implies that  $H(t, x, \bar{x}, \tilde{v}_n(1+|x|^2+|\bar{x}|^2))$  converges to  $H(t, x, \bar{x}, \tilde{v}(1+|x|^2+|\bar{x}|^2))$  uniformly on compact sets. It therefore holds that

$$\begin{cases} \partial_t \tilde{w}(t, x, \bar{x}) = \mathcal{L}\tilde{w}(t, x, \bar{x}) + \frac{H(t, x, \bar{x}, (1+|x|^2+|\bar{x}|^2) \tilde{v})}{1+|x|^2+|\bar{x}|^2} + \frac{|x-\bar{x}|^2}{1+|x|^2+|\bar{x}|^2}, & 0 < t \leq T, \\ \tilde{w}(0, x, \bar{x}) = 0. \end{cases}$$

By uniqueness of the solution to the preceding PDE we deduce that  $\tilde{w} = \psi(\tilde{v})$ , which provides the desired continuity of the mapping  $\psi$ .

As the set  $\mathbb{K}$  is convex and compact we are in a position to apply the Kakutani-Fan-Glicksberg theorem (see e.g. Aliprantis and Border [1, Thm.16.51]) and thus get that the map  $\tilde{\psi}$  admits a fixed point  $\tilde{V}$  in  $\mathbb{K}$ .

We conclude by observing that the function  $V(t, x, \bar{x}) := \tilde{V}(T-t, x, \bar{x}) (1+|x|^2+|\bar{x}|^2)$  is a smooth solution to the PDE (2.3).

#### 2.4. Proof of Theorem 2.2

In view of (2.5) one can choose a Borel measurable selection  $C^*(s, x, \bar{x})$  such that

$$C^*(s, x, \bar{x}) \in \arg \min_{C \in \mathbf{O}_d} \sum_{i,j=1}^d \left( \sigma(x) C \bar{\sigma}(\bar{x})^{\top} \right)^{ij} \partial_{x_i \bar{x}_j}^2 V(s, x, \bar{x}) \quad (2.17)$$

(see e.g. Fleming and Rishel [3, Appendix B]). The map  $C^*(t, x, \bar{x})$  is uniquely defined and continuous if and only if for every  $t, x$  and  $\bar{x}$ , the matrix  $(\partial_{x_i \bar{x}_j}^2 V(t, x, \bar{x}))^{i,j}$  is invertible: See Proposition 4.7 below. Proposition 4.7 also shows that then one can explicitly construct the optimal feedback control  $(C_s^*)$  taking values in the set of orthogonal matrices. In addition,  $\mathcal{D}_s^* = 0$  for every  $s$ .

Unfortunately it seems unlikely that this property holds true except in the one-dimensional case (see (2.8)) and possibly in very particular multi-dimensional situations (e.g. when  $\sigma(x)\bar{\sigma}(\bar{x}) = \bar{\sigma}(\bar{x})\sigma(x)$  for all  $x, \bar{x}$ ).

In our next section (see Corollary 3.2) we will show that for every  $m > 0$  there exists a continuous map  $C^m$  with values in the set of correlation matrices such that

$$H(s, x, \bar{x}, V) \leq \sum_{i,j=1}^d (\sigma(x) C^m(s, x, \bar{x}) \bar{\sigma}(\bar{x})^{\top})^{ij} \partial_{x_i \bar{x}_j}^2 V(s, x, \bar{x}) \leq H(s, x, \bar{x}, V) + \frac{1}{m} \quad (2.18)$$

for all  $s, x, \bar{x}$ . We then consider a continuous function  $D^m(s, x, \bar{x})$  such that

$$C^m(s, x, \bar{x}) C^m(s, x, \bar{x})^{\top} + D^m(s, x, \bar{x}) D^m(s, x, \bar{x})^{\top} = \text{Id}_d.$$

From now on, to simplify the notation we fix the time origin at  $t = 0$  and the initial conditions at  $x$  and  $\bar{x}$ . It follows from the continuity of  $C^m$  and Stroock

and Varadhan's results [13] that there is a weak solution  $(Y_\theta^m, \bar{Y}_\theta, \tilde{B}_\theta^m, B_\theta, \bar{B}_\theta, 0 \leq \theta \leq T)$  to

$$\begin{cases} Y_\theta^m = x + \int_0^\theta \mu(Y_s^m) ds + \int_0^\theta \sigma(Y_s^m) [C^m(s, Y_s^m, \bar{Y}_s) d\bar{B}_s + D^m(s, Y_s^m, \bar{Y}_s) dB_s], \\ \bar{Y}_\theta = \bar{x} + \int_0^\theta \bar{\mu}(\bar{Y}_s) ds + \int_0^\theta \bar{\sigma}(\bar{Y}_s) d\bar{B}_s, \\ \tilde{B}_\theta^m = \int_0^\theta [C^m(s, Y_s^m, \bar{Y}_s) d\bar{B}_s + D^m(s, Y_s^m, \bar{Y}_s) dB_s], \end{cases} \quad (2.19)$$

where  $(B, \bar{B})$  are independent standard Brownian motions. (The process  $\tilde{B}^m$  will be needed in a convergence argument below).

As the coefficients are bounded one easily gets that the family indexed by  $m$  of the probability distribution of the vector  $(Y^m, \bar{Y}, B, \bar{B}, \tilde{B}^m)$  is tight, and any component  $Z^m$  of this vector satisfies

$$\exists C, \forall m \geq 1, \forall 0 \leq t \leq \theta \leq r \leq T, \mathbb{E}|Z_r^m - Z_\theta^m|^2 \leq C(r - \theta).$$

We apply two celebrated results due to Skorokhod (see Krylov [8, Chap.2, Sec.6, Lem.2 and Lem.3]) which assert that there exists a filtered probability space, a sequence of numbers  $n$ , a process

$$\xi := (X^*, \bar{X}, \bar{W}, W, \tilde{W}),$$

and a sequence  $\xi^n := (X^n, \bar{X}^n, \bar{W}^n, W^n, \tilde{W}^n)$  defined on this probability space, such that all the finite-dimensional distributions of  $\xi^n$  coincide with the corresponding finite-dimensional distributions of  $(Y^n, \bar{Y}, \bar{B}, B, \tilde{B}^n)$  and  $\xi_\theta^n$  converges in probability to  $\xi_\theta$  for every  $0 \leq \theta \leq T$ . In addition,

$$\int_0^\theta \sigma(X_s^n) [C^n(s, X_s^n, \bar{X}_s^n) d\bar{W}_s^n + D^n(s, X_s^n, \bar{X}_s^n) dW_s^n]$$

converges in probability to  $\int_0^\theta \sigma(X_s^*) d\tilde{W}_s$  as  $n$  tends to infinity for every  $0 \leq \theta \leq T$ . We emphasize that the process  $(\tilde{W}, \bar{W})$  is a possibly correlated Brownian motion.

Reasoning as in Krylov [8, Chap.2, Sec.6] one easily gets that a.s. for any  $0 \leq \theta \leq T$ ,

$$\begin{cases} X_\theta^n = x + \int_0^\theta \mu(X_s^n) ds + \int_0^\theta \sigma(X_s^n) [C^n(s, X_s^n, \bar{X}_s^n) d\bar{W}_s^n + D^n(s, X_s^n, \bar{X}_s^n) dW_s^n], \\ \bar{X}_\theta^n = \bar{x} + \int_0^\theta \bar{\mu}(\bar{X}_s^n) ds + \int_0^\theta \bar{\sigma}(\bar{X}_s^n) d\bar{W}_s^n, \\ \tilde{W}_\theta^n = \int_0^\theta [C^n(s, X_s^n, \bar{X}_s^n) d\bar{W}_s^n + D^n(s, X_s^n, \bar{X}_s^n) dW_s^n], \end{cases} \quad (2.20)$$

from which

$$\forall 0 \leq \theta \leq T, \quad \begin{cases} X_\theta^* = x + \int_0^\theta \mu(X_s^*) ds + \int_0^\theta \sigma(X_s^*) d\tilde{W}_s, \\ \bar{X}_\theta = \bar{x} + \int_0^\theta \bar{\mu}(\bar{X}_s) ds + \int_0^\theta \bar{\sigma}(\bar{X}_s) d\bar{W}_s. \end{cases} \quad (2.21)$$

Now, choose a subsequence of  $(\xi^n)$  (still denoted by  $(\xi^n)$ ) which converges

a.s. and apply Itô's formula to  $V(\theta, X_\theta^n, \bar{X}_\theta^n)$ :

$$\begin{aligned}
& V(\theta, X_\theta^n, \bar{X}_\theta^n) - V(0, x, \bar{x}) \\
&= \int_0^\theta \left( \frac{\partial V}{\partial s} + \mathcal{L}V \right) (s, X_s^n, \bar{X}_s^n) ds \\
&+ \sum_{i,j=1}^d \int_0^\theta (\sigma(X_s^n) C^n(s, X_s^n, \bar{X}_s^n) \bar{\sigma}(\bar{X}_s^n))^{ij} \partial_{x_i \bar{x}_j}^2 V(s, X_s^n, \bar{X}_s^n) ds \\
&+ \int_0^\theta \nabla_x V(s, X_s^n, \bar{X}_s^n) \cdot \sigma(X_s^n) [C^n(s, X_s^n, \bar{X}_s^n) d\bar{W}_s^n + D^n(s, X_s^n, \bar{X}_s^n) dW_s^n] \\
&+ \int_0^\theta \nabla_{\bar{x}} V(s, X_s^n, \bar{X}_s^n) \cdot \bar{\sigma}(X_s^n) d\bar{W}_s^n \\
&=: a^n + b^n + c^n + d^n.
\end{aligned}$$

Every function  $C^n$  satisfies Inequality (2.18). Thus we have

$$\left| b_n - \int_0^\theta H(s, X_s^n, \bar{X}_s^n, V) ds \right| \leq \frac{\theta}{n},$$

from which

$$a_n + b_n = - \int_0^\theta |X_s^n - \bar{X}_s|^2 ds + o\left(\frac{1}{n}\right).$$

Let  $n$  tend to infinity. In view of the Skorokhod lemma in [8, Chap.2, Sec.6, Lem.3] the sequence of stochastic integrals

$$\int_0^\theta \nabla_x V(s, X_s^n, \bar{X}_s^n) \cdot \sigma(X_s^n) [C^n(s, X_s^n, \bar{X}_s^n) d\bar{W}_s^n + D^n(s, X_s^n, \bar{X}_s^n) dW_s^n]$$

converges to

$$\int_0^\theta \nabla_x V(s, X_s^*, \bar{X}_s) \cdot \sigma(X_s^*) d\widetilde{W}_s.$$

We deduce that

$$\begin{aligned}
& V(\theta, X_\theta^*, \bar{X}_\theta) - V(0, x, \bar{x}) + \int_0^\theta |X_s^* - \bar{X}_s|^2 ds \\
&= \int_0^\theta \nabla_x V(s, X_s^*, \bar{X}_s) \cdot \sigma(X_s^*) d\widetilde{W}_s + \int_0^\theta \nabla_{\bar{x}} V(s, X_s^*, \bar{X}_s) \cdot \bar{\sigma}(X_s^*) d\bar{W}_s,
\end{aligned}$$

so that the left-hand side is a martingale.

In view of the proof of the predictable representation theorem (see e.g. Karatzas and Shreve [7, Chap.3, Thm.4.2] or Revuz and Yor [12, Chap.V, Thm.3.9]) there exists a correlation matrix-valued predictable process  $(C_t^*)$  such that

$$\frac{d}{dt} \langle X^*, \bar{X} \rangle_t = \sigma(X_t^*) C_t^* \bar{\sigma}(\bar{X}_t^*).$$

Now apply Itô's formula to the martingale

$$V(\theta, X_\theta^*, \bar{X}_\theta^*) + \int_0^\theta |X_s^* - \bar{X}_s^*|^2 ds.$$

The finite variation part is equal to

$$- \int_0^\theta H(s, X_s^*, \bar{X}_s^*, V) ds + \sum_{i,j=1}^d \int_0^\theta (\sigma(X_s^*) C_s^* \bar{\sigma}(\bar{X}_s^*))^{ij} \partial_{x_i \bar{x}_j}^2 V(s, X_s^*, \bar{X}_s^*) ds,$$

which, in view of (2.5), shows that

$$C_s^* \in \arg \min_{C \in \mathbf{C}_d} \sum_{i,j=1}^d \left( \sigma(X_s^*) C \bar{\sigma}(\bar{X}_s^*) \right)^{ij} \partial_{x_i \bar{x}_j}^2 V(s, X_s^*, \bar{X}_s^*).$$

Finally, similarly as at the end of the proof of Proposition 1.6 we extend our probability space to construct a Brownian motion  $W$  independent of  $\bar{W}$  such that

$$\widetilde{W} = \left( \int_0^t C_s^* d\bar{W}_s + \int_0^t \mathcal{D}_s^* dW_s, 0 \leq t \leq T \right)$$

with  $\mathcal{D}_s^* (\mathcal{D}_s^*)^\top + C_s^* (C_s^*)^\top = \text{Id}_d$  and  $(C_s^*, \mathcal{D}_s^*)$  is predictable. We thus have shown the existence of a solution to (2.6) and are in a position to apply the verification theorem IV.3.1 in Fleming and Soner [4] to get

$$V(0, x, \bar{x}) = (\widetilde{W}^2(\mathbb{P}_x^{\mu, \sigma}; \mathbb{P}_{\bar{x}}^{\bar{\mu}, \bar{\sigma}}))^2.$$

### 2.5. When is a weak solution to (2.6) unique?

When the function  $C^*(t, x, \bar{x})$  is uniquely determined (see the discussion at the beginning of Section 2.4) and, in addition, is Hölder continuous, then  $C_s^* = C^*(s, X_s, \bar{X}_s)$  and one gets the uniqueness of the weak solution to (2.6) with the following standard argument. Let  $\psi$  be an arbitrary  $\mathcal{C}^\infty([0, T] \times \mathbb{R}^{2d})$  function with compact support. Apply Theorem 2.3 with  $\mathcal{A}$  equal to the infinitesimal generator of  $(X^*, \bar{X})$  and  $f \equiv 0$  and denote by  $u$  the unique smooth solution to the corresponding equation (2.9). Itô's formula shows that

$$u(t, x, \bar{x}) = \mathbb{E} \psi(X_T^*(t, x, \bar{x}), \bar{X}_T(t, \bar{x})),$$

which implies the desired weak uniqueness.

### 3. On $\epsilon$ -optimal feedback controls

The objective of this section is two-fold. First, we show that for every  $\epsilon > 0$  there exists a continuous control  $C_\epsilon(s, x, \bar{x})$  taking values in the set of correlation matrices which is pointwise  $\epsilon$ -optimal:

$$\forall s, x, \bar{x}, \quad \sum_{i,j=1}^d (\sigma(x) C_\epsilon(s, x, \bar{x}) \bar{\sigma}(\bar{x}))^{ij} \partial_{x_i \bar{x}_j}^2 V(t, x, \bar{x}) \leq H(t, x, \bar{x}, V) + \epsilon. \quad (3.1)$$

Second, under an additional hypothesis on  $\sigma$  and  $\bar{\sigma}$ , we exhibit a family  $U_\epsilon(s, x, \bar{x})$  of infinitely differentiable maps taking values in the set of orthogonal matrices and converging in  $L^p$ -norm to  $C^*(s, x, \bar{x})$  defined as in (2.17).

**Proposition 3.1.** *Let  $A$  be a continuous map defined on a paracompact space  $E$  with values in  $\mathbf{M}_d$ . For all  $\epsilon > 0$  and  $x \in E$  let*

$$\mathcal{C}_\epsilon(x) := \{O \in \mathbf{C}_d, \text{Tr}(OA(x)) \leq \inf_{C \in \mathbf{C}_d} \text{Tr}(CA(x)) + \epsilon\}. \quad (3.2)$$

*The correspondence  $\mathcal{C}_\epsilon$  is lower hemicontinuous and admits a continuous selector.*

*Proof.* Let  $x_0 \in E$ . To prove that the correspondence  $\mathcal{C}_\epsilon$  is lower hemicontinuous at  $x_0$  we have to show that the lower inverse image  $\mathcal{C}_\epsilon^\ell(\mathcal{U})$  is a neighborhood of  $x_0$  (cf. [1, Sec.16.2]), that is: for every open set  $\mathcal{U}$  in  $\mathbf{C}_d$  with  $\mathcal{U} \cap \mathcal{C}_\epsilon(x_0) \neq \emptyset$ , there exists a neighborhood  $U(x_0)$  of  $x_0$  such that for any  $x$  in  $U(x_0)$  one has  $\mathcal{U} \cap \mathcal{C}_\epsilon(x) \neq \emptyset$ , that is, there exists a correlation matrix  $C(x)$  which belongs to  $\mathcal{U}$  and satisfies

$$\text{Tr}(C(x) A(x)) \leq \inf_{C \in \mathbf{C}_d} \text{Tr}(CA(x)) + \epsilon.$$

To this end, set  $S(x) := \inf_{C \in \mathbf{C}_d} \text{Tr}(CA(x))$  and choose  $C(\frac{\epsilon}{2}, x_0)$  in  $\mathbf{C}_d$  such that

$$\text{Tr}(C(\frac{\epsilon}{2}, x_0)A(x_0)) < S(x_0) + \frac{\epsilon}{2}.$$

We also choose an arbitrary matrix  $C(\mathcal{U}, x_0)$  in the set  $\mathcal{U} \cap \mathcal{C}_\epsilon(x_0)$ .

The set of correlation matrices is convex and  $\mathcal{U}$  is open, thus there exists  $\eta > 0$  such that for every  $0 < \alpha < \eta$  one has that  $(1 - \alpha) C(\mathcal{U}, x_0) + \alpha C(\frac{\epsilon}{2}, x_0)$  belongs to  $\mathcal{U} \cap \mathbf{C}_d$ . Choose any  $0 < \alpha < \eta$  and then consider the neighborhood  $U(x_0)$  of  $x_0$  defined by

$$U(x_0) := \left\{ x \in E; \begin{aligned} &|\text{Tr}(C(\frac{\epsilon}{2}, x_0)(A(x) - A(x_0)))| < \frac{\epsilon \alpha}{4}, \\ &|\text{Tr}(C(\mathcal{U}, x_0)(A(x) - A(x_0)))| < \frac{\epsilon \alpha}{4}, \quad |S(x) - S(x_0)| < \frac{\epsilon \alpha}{4} \end{aligned} \right\}.$$

Finally set  $C(x) := C^\sharp := (1 - \alpha) C(\mathcal{U}, x_0) + \alpha C(\frac{\epsilon}{2}, x_0)$ . Notice that  $C^\sharp$  is in  $\mathcal{U}$ . It is easy to obtain that

$$\text{Tr}(C^\sharp A(x)) \leq S(x_0) + \epsilon - \frac{\epsilon \alpha}{4} \leq S(x) + \epsilon,$$

which is the desired property.

Furthermore, for every  $x \in E$  the set  $\mathcal{C}_\epsilon(x)$  is non empty, convex and closed. By using the Michael Selection Theorem (see e.g. [1, Thm.16.61]) we conclude that the correspondence  $\mathcal{C}_\epsilon$  admits a continuous selector.  $\square$

**Corollary 3.2.** *For every  $\epsilon > 0$  there is a continuous map  $C_\epsilon(s, x, \bar{x})$  such that (3.1) is satisfied.*

We now show that, under an additional hypothesis on  $\sigma$  and  $\bar{\sigma}$  (see Corollary 3.10 below), for any Borel measurable selection  $C^*(s, x, \bar{x})$  satisfying (2.17) there exists a family  $U_\epsilon(s, x, \bar{x})$  of infinitely differentiable controls which take values in the set of orthogonal matrices and converges to  $C^*(s, x, \bar{x})$  in  $L^p$ -norm. The results below do not bring further information, either to the distance  $\widetilde{W}^2$ , or to its stochastic control representation. However, it may be interesting for numerical purposes, notably if one is interested in simulating a simplified model by using a standard discretization method such as the Euler scheme whose convergence requires the coefficients are at least continuous.

**Lemma 3.3.** *Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}^d$ . Let  $p \in [1, \infty)$ . Let  $O : [0, T] \times \mathbb{R}^d \rightarrow \mathbf{O}_d$  be a measurable map. For every  $\epsilon > 0$ , there is a simple map  $S : [0, T] \times \mathbb{R}^d \rightarrow \mathbf{O}_d$ ,  $S = \sum_{j=1}^N O_j \mathbb{I}_{A_j}$  where  $(A_j)_{1 \leq j \leq N}$  is a partition of  $[0, T] \times \mathbb{R}^d$ , such that  $|S - O|_{L^p([0, T] \times \mathcal{K})} < \epsilon$ .*

*Proof.* Let  $\epsilon' = \frac{\epsilon}{m([0, T] \times \mathcal{K})}$ , where  $m$  is the Lebesgue measure on  $[0, T] \times \mathbb{R}^d$ .  $\mathbf{O}_d$  is a compact set, thus there is a finite number of  $O_j \in \mathbf{O}_d$  such that  $\mathbf{O}_d \subset \bigcup_{j=1}^N \mathcal{B}(O_j, \epsilon')$ . The map

$$S := \sum_{j=1}^{N-1} O_j \mathbb{I}_{A_j} + \text{Id}_d \mathbb{I}_{[0, T] \times \mathcal{K}^c}$$

with

$$A_1 := \{(s, x) \in [0, T] \times \mathcal{K} \mid O(s, x) \in \mathcal{B}(O_1, \epsilon')\}$$

and

$$A_j := \{(s, x) \in [0, T] \times \mathcal{K} - \bigcup_{1 \leq k \leq j-1} A_k \mid O(s, x) \in \mathcal{B}(O_j, \epsilon')\}, \quad 1 < j < N,$$

satisfies the required condition.  $\square$

Recall the following result.

**Lemma 3.4.** *Every  $O \in \mathbf{O}_d$  admits the following block diagonalization: There is  $U \in \mathbf{O}_d$  such that  $O = U \cdot D(r, l, \theta_1, \dots, \theta_j) \cdot U^\top$ , where for all  $r, l \in \mathbb{N}$  and  $\theta_i \in \mathbb{R}$  we have set*

$$D(r, l, \theta_1, \dots, \theta_j) = \begin{pmatrix} \text{Id}_r & 0 & 0 & \dots & 0 \\ 0 & -\text{Id}_l & 0 & \dots & 0 \\ 0 & 0 & R_{\theta_1} & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & R_{\theta_j} \end{pmatrix}, \quad R_\theta := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

**Lemma 3.5.** *Assume that  $d$  is even. For any  $O \in \mathbf{SO}_d$  there are  $V \in \mathbf{O}_d$  and  $(\theta_i)_{1 \leq i \leq \frac{d}{2}}$  such that  $0 = VD(0, 0, \theta_1, \dots, \theta_{\frac{d}{2}})V^\top$ .*

*Proof.* It follows from Lemma 3.4 that  $O = UD(r, l, \theta_1, \dots, \theta_j)U^\top$ . The matrix  $O$  is in  $\mathbf{SO}_d$ , thus  $l$  is even, from which  $r$  is also even. Notice that  $-\text{Id}_2 = R_\pi$  and  $\text{Id}_2 = R_0$ . The result follows.  $\square$

Similarly we get the following result:

**Lemma 3.6.** *Assume that  $d$  is odd. For every  $O \in \mathbf{SO}_d$  there is  $V \in \mathbf{O}_d$  and  $(\theta_i)_{1 \leq i \leq \frac{d}{2}}$  such that  $0 = VD(1, 0, \theta_1, \dots, \theta_{\frac{d-1}{2}})V^\top$ .*

A simple computation leads to the next lemma.

**Lemma 3.7.** *Every rotation matrix  $R_\theta$  can be written*

$$R_\theta = J \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} J^\top$$

where  $J$  is the following unitary matrix:

$$J := \frac{1}{\sqrt{2}} \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix}.$$

**Proposition 3.8.** *Let  $S : [0, T] \times \mathbb{R}^d \rightarrow \mathbf{SO}_d$  be a simple map. For every  $\epsilon > 0$ , there is a continuous map  $T : [0, T] \times \mathbb{R}^d \rightarrow \mathbf{SO}_d$  such that  $m(\{(s, x), S(s, x) \neq T(s, x)\}) < \epsilon$ .*

*Proof. The case where  $d$  is even.*

Denote by  $J_d \in \mathbf{SO}_d$  the block diagonal matrix

$$J_d := \begin{pmatrix} J & 0 & \dots & 0 \\ \dots & & & \\ 0 & \dots & 0 & J \end{pmatrix}$$

Set

$$\mathcal{D}_0(\theta_1, \dots, \theta_{\frac{d}{2}}) := \begin{pmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & -\theta_1 & \dots & 0 \\ \dots & & & \\ 0 & \dots & \theta_{\frac{d}{2}} & 0 \\ 0 & \dots & 0 & -\theta_{\frac{d}{2}} \end{pmatrix}.$$

It follows from Lemmas 3.4 and 3.5 that

$$O_j = V_j J_d \exp\{i\mathcal{D}_0(\theta_{j,1}, \dots, \theta_{j,\frac{d}{2}})\} J_d^\top V_j^\top = \exp\{V_j J_d \{i\mathcal{D}_0(\theta_{j,1}, \dots, \theta_{j,\frac{d}{2}})\} J_d^\top V_j^\top\},$$

where  $\exp(M) := \sum_{0 \leq n} \frac{M^n}{n!}$ .

By hypothesis the map  $S$  can be written as  $S = \sum_{j=1}^N 0_j \mathbb{I}_{A_j}$ , where  $(A_j)_{1 \leq j \leq N}$  is a partition of  $[0, T] \times \mathbb{R}^d$ . It follows that

$$\begin{aligned} S &= \sum_{j=1}^N 0_j \mathbb{I}_{A_j} = \sum_{j=1}^N \exp\{V_j J_d \{i\mathcal{D}_0(\theta_{j,1}, \dots, \theta_{j,\frac{d}{2}})\} J_d^\top V_j^\top\} \mathbb{I}_{A_j} \\ &= S_1 S_2 \dots S_N, \end{aligned} \tag{3.3}$$

where  $S_j = \exp\{V_j J_d \{i\mathcal{D}_0(\theta_{j,1} \mathbb{I}_{A_j}, \dots, \theta_{j,\frac{d}{2}} \mathbb{I}_{A_j})\} J_d^\top V_j^\top\}$ . For every measurable set  $A_j$  there exist a compact set  $K_j$  and an open set  $W_j$  such that  $K_j \subset A_j \subset W_j$

and  $m(W_j - K_j) < \frac{\epsilon}{N}$ . It therefore exists a continuous function  $f_j$  with compact support,  $0 \leq f_j \leq 1$ , such that  $f_j = 1$  on  $K_j$  and  $f_j = 0$  on  $W_j^c$ .

Let  $T := T_1 T_2 \dots T_N$ , with

$$T_j := \exp\{V_j J_d \{i\mathcal{D}_0(\theta_{j,1}f_j, \dots, \theta_{j,\frac{d}{2}}f_j)\} J_d^\top V_j^\top\}.$$

For  $(s, x) \notin \bigcup_{1 \leq j \leq N} (W_j - K_j)$  one has  $T(s, x) = S(s, x)$ . Furthermore, for any  $(s, x) \in [0, T] \times \mathbb{R}^d$ ,  $H_j(s, x) = V_j J_d \{i\mathcal{D}_0(\theta_{j,1}f_j(s, x), \dots, \theta_{j,\frac{d}{2}}f_j(s, x))\} J_d^\top V_j^\top$  is an antihermitian matrix ( $\overline{H_j}^\top(s, x) = -H_j(s, x)$ ). It follows that  $T_j(s, x) = \exp(H_j(s, x))$  is a unitary matrix and  $T(s, x)$  is also a unitary matrix for all  $(s, x)$ . Note that the matrices  $(S_j)_{1 \leq j \leq N}$  commute but the matrices  $T_j$  do not commute.

Moreover, for any  $(s, x)$  one has

$$J_d \exp\{i\mathcal{D}_0(\theta_{j,1}f_j(s, x), \dots, \theta_{j,\frac{d}{2}}f_j(s, x))\} J_d^\top = D(0, 0, \theta_{j,1}f_j(s, x), \dots, \theta_{j,\frac{d}{2}}f_j(s, x)),$$

where we have used the same notation as in Lemma 3.4. Thus this matrix belongs to  $M_d(\mathbb{R})$ . It follows that the continuous maps  $T_j$  and  $T$  take values in  $\mathbf{O}_d$ . Furthermore,  $\det(T_j(s, x)) = 1$  if  $(s, x) \notin W_j$ . This proves that  $T_j$  and then  $T$  take values in  $\mathbf{SO}_d$ .

**The case where case  $d$  is odd.** The proof proceeds similarly as above by using Lemmas 3.4 and 3.6.  $\square$

We now are in a position to show that any measurable selection  $C^*(s, x, \bar{x})$  in  $\mathbf{SO}_d$  can be approximated in  $L^p$ -norm by smooth functions taking values in  $\mathbf{O}_d$ .

**Proposition 3.9.** *Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}^d$ . Let  $p \in [1, \infty)$ . Let  $O : [0, T] \times \mathbb{R}^n \rightarrow \mathbf{SO}_d$  be a measurable map. For every  $\epsilon > 0$ , there is a  $\mathcal{C}_\infty$  map  $U : [0, T] \times \mathbb{R}^d \rightarrow \mathbf{O}_d$  such that  $|U - O|_{L^p([0, T] \times \mathcal{K})} < \epsilon$ .*

*Proof.* It follows from Lemma 3.3 and Proposition 3.8 that there is a continuous map  $T = T_1 T_2 \dots T_N$  with

$$T_j(s, x) := \exp\{V_j J_d \{i\mathcal{D}_0(\theta_{j,1}f_j(s, x), \dots, \theta_{j,\frac{d}{2}}f_j(s, x))\} J_d^\top V_j^\top\}$$

such that  $|T - O|_{L^p([0, T] \times \mathcal{K})} < \frac{\epsilon}{2}$ . Every  $f_j$  is continuous with compact support and therefore is the uniform limit of a sequence  $f_{j,n} = f_j * \phi_{\epsilon_n}$  of  $\mathcal{C}^\infty(\mathbb{R}^d)$  functions, where  $\phi_{\epsilon_n}$  is a sequence of standard smooth mollifiers. Therefore, for any large enough  $n_j$  the map  $U = U_1 U_2 \dots U_N$  with

$$U_j := \exp\{V_j J_d \{i\mathcal{D}_0(\theta_{j,1}f_{j,n_j}, \dots, \theta_{j,\frac{d}{2}}f_{j,n_j})\} J_d^\top V_j^\top\}$$

satisfies all the required conditions.  $\square$

We now are in a position to get the desired approximation by smooth controls result.

**Corollary 3.10.** *Suppose that  $\sigma$  and  $\bar{\sigma}$  are such that one can choose  $C^*(s, x, \bar{x})$  in  $\mathbf{SO}_d$  satisfying (2.17) for all  $s, x, \bar{x}$ . Then there exists a family  $U_\epsilon(s, x, \bar{x})$  of infinitely differentiable maps which take values in the set of orthogonal matrices and converge to  $C^*(s, x, \bar{x})$  in  $L^p$ -norm.*



#### 4. Appendix: Extreme points in the set of correlation matrices and an optimization problem over orthogonal matrices

The goal of this section is to prove the linear algebra results we needed in the preceding sections, notably the convexity of the set  $\mathbf{C}_d$  and the characterization of its extreme points, and regularity properties of our optimization problem over orthogonal matrices. The first proposition is classical.

**Proposition 4.1.** *The set  $\mathbf{C}_d$  is convex.*

*Proof.* Let  $C_1$  and  $C_2$  be in  $\mathbf{C}_d$ , and let  $C = \alpha C_1 + (1 - \alpha)C_2$  with  $0 \leq \alpha \leq 1$ . Let  $X_1, Y_1$  be  $\mathbb{R}^d$  valued random variables as in (0.2) such that  $C_1 = \mathbb{E}(X_1 Y_1^\top)$ . Let  $(X_2, Y_2)$  independent of  $(X_1, Y_1)$  as in (0.2) such that  $C_2 = \mathbb{E}(X_2 Y_2^\top)$ . Finally, let  $X_\alpha := \sqrt{\alpha}X_1 + \sqrt{1-\alpha}X_2$  and  $Y_\alpha := \sqrt{\alpha}Y_1 + \sqrt{1-\alpha}Y_2$ . We have  $\mathbb{E}(X_\alpha Y_\alpha^\top) = \alpha C_1 + (1 - \alpha)C_2 = C$  and  $\mathbb{E}(X_\alpha X_\alpha^\top) = \mathbb{E}(Y_\alpha Y_\alpha^\top) = I$ .  $\square$

Now let us recall the definition of extreme points.

**Definition 4.2.** *A point  $C$  in a convex set  $\mathbf{C}$  is extreme if*

$$\exists \alpha \in [0, 1], \exists C_1, C_2 \in \mathbf{C}, C = \alpha C_1 + (1 - \alpha)C_2 \Rightarrow C = C_1 \text{ or } C = C_2.$$

Before characterizing the extreme points in  $\mathbf{C}_d$  we need to prove several technical lemmas.

**Lemma 4.3.** *Let  $U$  be an orthogonal matrix. The matrix  $C$  belongs to  $\mathbf{C}_d$  if and only if  $CU$  (respectively,  $UC$ ) belongs to  $\mathbf{C}_d$ .*

*Proof.* Let  $C \in \mathbf{C}_d$ . Let  $X, Y$  be  $\mathbb{R}^d$  valued random variables as in (0.2). Let  $\tilde{Y} = U^\top Y$ ,  $CU = \mathbb{E}(X \tilde{Y}^\top)$ . From (0.2) it follows that  $\mathbb{E}(\tilde{Y} \tilde{Y}^\top) = U^\top U = \text{Id}_d$ . Thus  $CU$  and  $UC$  belong to  $\mathbf{C}_d$ . The converse implications follow from the equality  $UU^\top = \text{Id}_d$ .  $\square$

**Lemma 4.4.** *For every matrix  $C$  in  $\mathbf{C}_d$  there is a matrix  $R$  in  $\mathbf{C}_d$  such that  $\text{Id}_d = C^\top C + R^\top R$ .*

*Proof.* Let  $C$  be in  $\mathbf{C}_d$  and  $X = (X_i)_{1 \leq i \leq d}$ ,  $Y = (Y_j)_{1 \leq j \leq d}$  be as in (0.2). Let  $\mathcal{V}$  be a  $2d$  dimensional vector subspace of the set of real valued random variables containing the random variables  $X_i$  and  $Y_j$ ,  $1 \leq i, j \leq d$ . The orthonormal set  $(X_i)_{1 \leq i \leq d}$  (for the scalar product  $\langle x, y \rangle := \mathbb{E}(x \cdot y)$ ) can be extended into an orthonormal basis  $(e_i)_{1 \leq i \leq 2d}$  of  $\mathcal{V}$  with  $e_i = X_i$  for all  $1 \leq i \leq d$ . Thus  $C_{kj} = \langle e_k, Y_j \rangle$  for all  $1 \leq k, j \leq d$ . Let  $R := (R_{k,j})$  with  $R_{k,j} := \langle e_{d+k}, Y_j \rangle$  for all  $1 \leq k, j \leq d$ . Thus  $Y_j = \sum_{1 \leq k \leq d} C_{k,j} e_k + R_{k,j} e_{d+k}$ . From the equality  $\mathbb{E}(Y Y^\top) = \text{Id}_d$  it results that  $\text{Id}_d = C^\top C + R^\top R$ .  $\square$

**Lemma 4.5.** *The unit matrix  $\text{Id}_d$  is an extreme point in  $\mathbf{C}_d$ .*

*Proof.* Let  $C_1$  and  $C_2$  be in  $\mathbf{C}_d$ . Let  $0 < \alpha < 1$  be such that

$$\text{Id}_d = \alpha C_1 + (1 - \alpha)C_2, \tag{4.1}$$

from which

$$d = \alpha^2 \operatorname{Tr}(C_1^\top C_1) + (1 - \alpha)^2 \operatorname{Tr}(C_2^\top C_2) + 2\alpha(1 - \alpha) \operatorname{Tr}(C_1^\top C_2). \quad (4.2)$$

For any  $\beta \in \mathbb{R}$  the second order polynomial function

$$\beta \mapsto \operatorname{Tr}((\beta C_1 + C_2)^\top (\beta C_1 + C_2))$$

is positive. Therefore

$$(\operatorname{Tr}(C_1^\top C_2 + C_2^\top C_1))^2 \leq 4 \operatorname{Tr}(C_1^\top C_1) \operatorname{Tr}(C_2^\top C_2).$$

Notice that the left-hand side is equal to  $4(\operatorname{Tr}(C_1^\top C_2))^2$ . We thus have shown that

$$(\operatorname{Tr}(C_1^\top C_2))^2 \leq \operatorname{Tr}(C_1^\top C_1) \operatorname{Tr}(C_2^\top C_2). \quad (4.3)$$

Now observe that Lemma 4.4 implies that  $\operatorname{Tr}(C^\top C) \leq d$  for all  $C$  in  $\mathbf{C}_d$ . It thus follows from (4.3) that the above equality (4.2) holds true only if  $\operatorname{Tr}(C_1^\top C_2) = \operatorname{Tr}(C_1^\top C_1) = \operatorname{Tr}(C_2^\top C_2) = d$ . Again applying Lemma 4.4 we get  $C_1^\top C_1 = C_2^\top C_2 = \operatorname{Id}_d$ . In view of (4.1) we also have

$$\operatorname{Id}_d = (\alpha C_1 + (1 - \alpha)C_2) (\alpha C_1 + (1 - \alpha)C_2)^\top,$$

from which  $C_1^\top C_2 + C_2^\top C_1 = 2\operatorname{Id}_d$ . Thus  $(C_1^\top - C_2^\top)(C_1 - C_2) = 0$ , which implies that  $C_1 = C_2$ , from which we get  $C_1 = C_2 = \operatorname{Id}_d$  by again using (4.1).  $\square$

**Proposition 4.6.** *The extreme points in  $\mathbf{C}_d$  are in  $\mathbf{O}_d$ .*

*Proof.* Let  $C$  be an extreme point in  $\mathbf{C}_d$ .

*Step 1: Reduction to the case of diagonal matrices  $C^\top C$ .* The matrix  $C^\top C$  is symmetric and thus there exists an orthogonal matrix  $U$  such that  $U^\top C^\top C U$  is diagonal. It follows from Lemma 4.3 that  $C$  is an extreme point in  $\mathbf{C}_d$  if and only if  $CU$  is an extreme point in  $\mathbf{C}_d$ . Thus without loss of generality, one can assume that  $C^\top C$  is a diagonal matrix.

*Step 2. The eigenvalues of  $C^\top C$  are all between 0 and 1.* From Lemma 4.4, there is a matrix  $R$  such that  $\operatorname{Id}_d = C^\top C + R^\top R$ . As we now suppose that  $C^\top C$  is diagonal the matrix  $R^\top R$  is also diagonal. Furthermore both matrices are non negative. Therefore the eigenvalues of  $C^\top C$  are all between 0 and 1.

*Step 3. The largest eigenvalue of  $D := C^\top C$  is equal to 1.* Permutation matrices are orthogonal matrices. Possibly replacing  $C$  by  $CP$  where  $P$  is a permutation matrix (which is allowed by Lemma 4.3), one can thus assume  $\mathcal{D}^{1,1} \geq \mathcal{D}^{i,i}$  for all  $i$ . As it is easily seen that the null matrix is not an extreme point in  $\mathbf{C}_d$ , it suffices to assume that  $0 < \mathcal{D}^{1,1} < 1$  and exhibit a contradiction.

Let  $\alpha := \sqrt{\mathcal{D}^{1,1}}$ . Let  $X, Y$  be as in (0.2). Let  $(e_i)$  and  $R$  be as in the proof of Lemma 4.4. Set

$$\tilde{Y}_j := \sum_{1 \leq i \leq d} \left[ \frac{\langle e_i, Y_j \rangle}{\alpha} e_i + \gamma_j < e_{d+i}, Y_j > e_{d+i} \right],$$

where  $\gamma_j \geq 0$  is choosen such that

$$\frac{(C^\top C)^{jj}}{\alpha^2} + \gamma_j^2 (R^\top R)^{jj} = 1.$$

As  $(\mathbb{E}(\tilde{Y}\tilde{Y}^\top))_{j,k} = \frac{\sum_i C^{ij}C^{ik}}{\alpha^2} + \gamma_j\gamma_k \sum_i R^{ij}R^{ik}$  and the matrices  $C^\top C$ ,  $R^\top R$  are diagonal, we get  $(\mathbb{E}(\tilde{Y}\tilde{Y}^\top)) = \text{Id}_d$ . It follows that the matrix  $\tilde{C} = \mathbb{E}(X\tilde{Y}) = \frac{C}{\alpha}$  belongs to  $\mathbf{C}_d$ . Thus  $C$  is a non trivial convex combination of  $\tilde{C}$  and the null matrix.

*Step 4: Induction.* The matrix  $C$  being an extreme point in  $\mathbf{C}_d$ , all the eigenvalues of the matrix  $C^\top C$  are equal to 1. We proceed inductively w.r.t. the dimension of  $C$ . Notice that it follows from Step 3 that  $C^\top C = I_1$  for  $d = 1$ .

Assume that  $C$  is an extreme point in  $\mathbf{C}_d$  for  $d \geq 2$ . In view of Step 3 and  $1 = \mathcal{D}_{1,1} = (C^\top C)_{1,1}$ , one has that  $1 = \sum_{1 \leq i \leq d} \langle e_i, Y_1 \rangle^2$ . From  $\mathbb{E}(Y_1 Y_1^\top) = 1$  it follows that  $Y_1$  belongs to the  $d$ -dimensional vector space  $E_d$  spanned by  $(e_i)_{1 \leq i \leq d}$ . Let  $(\hat{e}_i)$  be an orthonormal basis of  $E_d$  such that  $\hat{e}_1 := Y_1$ . Let  $\hat{X} := (\hat{e}_i)$  and  $\hat{C} := \mathbb{E}(\hat{X}Y^\top)$ . The matrix  $\hat{C}$  is equal to  $UC$  where  $U$  is the orthogonal matrix associated to the change of basis from  $(e_i)_{1 \leq i \leq n}$  to  $(\hat{e}_i)_{1 \leq i \leq n}$ , thus  $\hat{C}$  is an extreme point in  $\mathbf{C}_d$ . As  $\hat{e}_1 = Y_1$  we have

$$\hat{C} = \begin{pmatrix} 1 & 0 \\ 0 & M \end{pmatrix}$$

where the matrix  $M$  necessarily is an extreme point in  $\mathbf{C}_{d-1}$ . It then remains to apply the inductive hypothesis.  $\square$

We now characterize the solutions to the minimization problem we are interested in.

**Proposition 4.7.** *For any  $A \in \mathbf{M}_d$  one has*

$$\inf_{O \in \mathbf{C}_d} \text{Tr}(O A) = \inf_{O \in \mathbf{O}_d} \text{Tr}(O A). \quad (4.4)$$

*In addition, the minimization problem*

$$\inf_{O \in \mathbf{O}_d} \text{Tr}(O A) \quad (4.5)$$

*admits a solution  $O^* = -Q^{-1}$  where  $Q$  is an orthogonal matrix such that  $A = (AA^\top)^{\frac{1}{2}}Q$ . Furthermore the solution is unique and equal to  $-A^{-1}(AA^\top)^{\frac{1}{2}}$  if and only if  $A$  is invertible.*

*Proof.* The preceding proposition combined with Proposition 4.1 and a classical result in Convex Optimization (see e.g. Hiriart-Urruty and Lemaréchal [6, Chap.III, Prop.2.4.6]) implies (4.4).

We now turn to the second part of the statement and start with examining the case where  $A$  is invertible. It then exists a unique representation  $A = SQ$

with  $S$  symmetric definite positive and  $Q$  orthogonal. This representation is given by  $S = (AA^\top)^{\frac{1}{2}}$  and  $Q := S^{-1}A = (AA^\top)^{-\frac{1}{2}}A$ . Thus  $A = SQ$ . The matrix  $S$  is symmetric definite positive and thus there is a diagonal matrix  $D$  with strictly positive diagonal terms  $(\lambda_i)_{1 \leq i \leq d}$  and an orthogonal matrix  $U$  such that  $S = UDU^{-1}$ .

Thus  $\text{Tr}(OA) = \text{Tr}(OUUDU^{-1}Q) = \text{Tr}(DU^{-1}QOU)$ . The map  $O \rightarrow U^{-1}QOU$  is a one to one map from  $\mathbf{O}_d$  to  $\mathbf{O}_d$ . Thus the problem is reduced to compute  $\min_{O \in \mathbf{O}_d} \text{Tr}(OD)$ , where  $D$  is the diagonal matrix with strictly positive diagonal terms  $(\lambda_i)_{1 \leq i \leq d}$ .

Notice that  $\text{Tr}(OD) = \sum_{1 \leq i \leq n} \lambda_i O^{ii}$ . As the matrix  $O$  is orthogonal we have

$$|O^{ii}| \leq \left( \sum_j |O^{ij}|^2 \right)^{\frac{1}{2}} \leq 1.$$

It follows that  $\text{Tr}(OD) \geq -\sum_i \lambda_i$ . Furthermore the equality is satisfied if and only if  $O = -\text{Id}_d$ . Thus  $O^*$  is a solution to the minimization problem (4.5) if and only if  $U^{-1}QO^*U = -\text{Id}_d$ , that is, if  $O^* = -Q^{-1} = -A^{-1}(AA^\top)^{\frac{1}{2}}$ .

We now treat the case of non invertible matrices  $A$ . Then the decomposition  $A = SQ$  with  $S$  symmetric semi-definite positive and  $Q$  orthogonal exists but is not unique. The matrix  $S$  is non negative but non definite and we have  $\lambda_i \geq 0$ . Therefore  $O^* = -Q^{-1}$  is a solution but is not the only one. Indeed, observe that  $S$  can be written as

$$S = U \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} U^{-1}$$

for some  $d - k$  diagonal matrix  $D$  and orthogonal matrix  $U$ . Let  $V$  be an  $k \times k$  orthogonal matrix different from  $\text{Id}_k$ . Set

$$\bar{Q} := U \begin{pmatrix} \text{Id}_{d-k} & 0 \\ 0 & V \end{pmatrix} U^{-1} Q.$$

We both have  $A = S\bar{Q}$  and  $Q \neq \bar{Q}$ . □

**Proposition 4.8.** *Assume that the  $d \times d$  matrix-valued function  $A$  on  $\mathbb{R}^d$  satisfies*

$$\exists 0 < \gamma \leq 1, \forall x, y \in \mathbb{R}^d, \|A(x) - A(y)\| \leq K(x, y) |x - y|^\gamma,$$

where  $\|M\| := (\text{Tr}(M) \text{Tr}(M^\top))^{1/2}$  and  $K(x, y)$  is a given symmetric function. Given any non-empty subset  $\mathcal{M}$  of the unit ball of  $\mathbf{M}_d$ , the map  $T_A(x) := \inf_{O \in \mathcal{M}} \text{Tr}(OA(x))$  satisfies

$$|T_A(x) - T_A(y)| \leq K(x, y) |x - y|^\gamma.$$

*Proof.* For all  $O$  in  $\mathcal{M}$  and  $x, y$  in  $\mathbb{R}^d$  it obviously holds

$$|\text{Tr}(OA(x)) - \text{Tr}(OA(y))| \leq K(x, y) |x - y|^\gamma,$$

from which

$$\inf_{U \in \mathcal{M}} \text{Tr}(UA(x)) \leq \text{Tr}(OA(y)) + K(x, y) |x - y|^\gamma.$$

We deduce

$$\inf_{U \in \mathcal{M}} \text{Tr}(UA(x)) \leq \inf_{O \in \mathcal{M}} \text{Tr}(OA(y)) + K(x, y) |x - y|^\gamma.$$

We then get the desired result by exchanging  $x$  and  $y$ .  $\square$

## 5. Conclusion and perspectives

We here have proposed a Wasserstein-type distance on the set of the probability distributions of pathwise unique solutions to stochastic differential equations. We have proven that this distance is characterized as the value function of a stochastic control problem and thus can be computed by discretizing the related Hamilton–Jacobi–Bellman equation or by approximating optimal feedback controls.

For the sake of simplicity, we have limited ourselves to handle with a distance which is a variant of  $\mathcal{W}^2$ . One can easily extend our approach to many variants of other Wasserstein distances provided that the Hamilton–Jacobi–Bellman equation (2.3) admits a smooth enough solution. For example, the function  $|x - \bar{x}|^2$  can be replaced with a function  $\rho$  such that for some integer  $p$  the map  $\frac{\rho(x, \bar{x})}{1 + |x|^p + |\bar{x}|^p}$  is of class  $\mathcal{C}^{2+\alpha}(\mathbb{R}^d)$ .

Our result opens the way to design practical methods to optimally choose a simplified diffusion model within a wide class of diffusion models close to a perfect but complex one, that is, to choose a model which minimizes the  $\widetilde{W}^2$  distance to the probability distribution of the complex one. We plan to address this issue in a future work.

## Acknowledgement

We thank an anonymous referee for having pointed out that an inequality in the original proof of Theorem 2.1 was questionable, which led us to modify an argument in this proof.

## References

- [1] ALIPRANTIS, C. and BORDER, K. (1999). *Infinite Dimensional Analysis*. Springer–Verlag.
- [2] CLAISSE, J., TALAY, D. and TAN, X. (2016). A Pseudo-Markov property for controlled diffusion processes. *SIAM J. Control Optim.* 54(2), 1017–1029.
- [3] FLEMING, W.H. and RISHEL, R.W. (1975). *Deterministic and Stochastic Optimal Control*. Springer–Verlag.
- [4] FLEMING, W.H. and SONER, M. (2006). *Controlled Markov Processes and Viscosity Solutions*. Volume 25 of *Stochastic Modelling and Applied Probability*. Springer.
- [5] FRIEDMAN, A. (1964). *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Inc., Englewood Cliffs, N.J.

- [6] HIRIART-URRUTY, J-B. and LEMARÉCHAL, C. (1993). *Convex Analysis and Minimization Algorithms I*. Grundlehren der mathematische Wissenschaften. Springer-Verlag.
- [7] KARATZAS, I. and SHREVE, S.E. (1991). *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.
- [8] KRYLOV, N.V. (1987). *Nonlinear Elliptic and Parabolic Equations of the Second Order*. Reidel, Dordrecht.
- [9] KUNITA, H. (1984). Stochastic differential equations and stochastic flows of diffeomorphisms. *École d'été de probabilités de Saint-Flour, XII-1982*. Lecture Notes in Math. 1097, 143–303, Springer, Berlin.
- [10] LADYZENSKAJA, O.A., SOLONNIKOV, V.A. and URAL'CEVA, N.N. (1967). *Linear and Quasi-linear Equations of Parabolic Type*. Translations of Mathematical Monographs 23, American Mathematical Society, Providence.
- [11] LUNARDI, A. (1995). *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Volume 16 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Verlag.
- [12] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*. Volume 293 in *Fundamental Principles of Mathematical Sciences Series*. Springer-Verlag, Berlin.
- [13] STROOCK, D.W. and VARADHAN, S.R.S. (1979). *Multidimensional Diffusion Processes*. Volume 233 in *Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin.