

## Regularization lemmas and convergence in total variation

Vlad Bally\*      Lucia Caramellino<sup>†</sup>      Guillaume Poly<sup>‡</sup>

### Abstract

We provide a simple abstract formalism of integration by parts under which we obtain some regularization lemmas. These lemmas apply to any sequence of random variables  $(F_n)$  which are smooth and non-degenerated in some sense and enable one to upgrade the distance of convergence from smooth Wasserstein distances to total variation in a quantitative way. This is a well studied topic and one can consult for instance [3, 11, 14, 20] and the references therein for an overview of this issue. Each of the aforementioned references share the fact that some non-degeneracy is required along the whole sequence. We provide here the first result removing this costly assumption as we require only non-degeneracy at the limit. The price to pay is to control the smooth Wasserstein distance between the Malliavin matrix of the sequence and its limit, which is particularly easy in the context of Gaussian limit as the Malliavin matrix is deterministic. We then recover, in a slightly weaker form, the main findings of [19]. Another application concerns the approximation of the semi-group of a diffusion process by the Euler scheme in a quantitative way and under the Hörmander condition.

**Keywords:** abstract Malliavin calculus; regularization techniques; total variation distance; Wasserstein distance.

**AMS MSC 2010:** 60B11; 60H07.

Submitted to EJP on October 7, 2019, final version accepted on June 8, 2020.

## 1 Introduction

The main historical application of Malliavin calculus, introduced in 1975 by Paul Malliavin, was a probabilistic proof of the Hörmander regularity criterion. But in the 40 last years it gave rise to a huge amount of various applications, and in particular it has been developed as a branch of stochastic analysis on the Wiener space, see the

---

\*Université Paris-Est, LAMA (UMR CNRS, UPEMLV, UPEC), MathRisk INRIA, F-77454 Marne-la-Vallée, France. E-mail: [bally@univ-mlv.fr](mailto:bally@univ-mlv.fr)

<sup>†</sup>Dipartimento di Matematica, Università di Roma “Tor Vergata”, and INdAM-GNAMPA, Via della Ricerca Scientifica 1, I-00133 Roma, Italy. E-mail: [caramell@mat.uniroma2.it](mailto:caramell@mat.uniroma2.it)

<sup>‡</sup>IRMAR, Université de Rennes 1, 263 avenue du Général Leclerc, CS 74205 35042 Rennes, France. E-mail: [guillaume.poly@univ-rennes1.fr](mailto:guillaume.poly@univ-rennes1.fr)

classical book of Nualart [22], as well as the more recent area of research pioneered by Nourdin and Peccati, see [17]. There is a major philosophical difference between the two aforementioned views of Malliavin calculus, as the so-called Malliavin-Stein's method (of Nourdin and Peccati), which has been intensively studied in a recent past, mixes the formalism of integration by parts provided by Malliavin calculus operators together with the Stein's method. Let us recall that the quintessence of Stein's method consists of identifying a suitable functional operator which characterizes a specific target and use it to prove convergence towards this target in a quantitative way. The most emblematic example is certainly the univariate standard Gaussian distribution  $\gamma$  which is characterized by the equation  $\langle f' - xf, \gamma \rangle = 0$  for every test function  $f$ . The link with Malliavin calculus appears in the identity:

$$\mathbb{E}(f'(X) - Xf(X)) = \mathbb{E}(f'(X)(1 - \Gamma[X, -\mathcal{L}^{-1}X]))$$

where  $\Gamma$  is the square field operator on the Wiener space and  $\mathcal{L}^{-1}$  is the pseudo-inverse of the Ornstein-Uhlenbeck operator. The quantity of interest is then  $\Gamma[X, -\mathcal{L}^{-1}X]$  which is different from the quantity  $\Gamma[X, X]$  which is standard in Malliavin calculus. Hence, although these two points of view are rather close as they both employ the Malliavin calculus to compute distances between distributions, they go towards different directions. Malliavin-Stein methods focus on specific targets with specific operators in order to provide rates of convergence whereas regularization lemmas focus on smoothness of distribution and upgrading distances of convergence. The present article explores this direction, namely we do not aim at proving limit theorems but instead of that, given a limit theorem, we explore the strongest probabilistic distances and the smoothness of the laws. In some sense, both approaches are complementary.

To do so, we introduce an abstract framework built on Dirichlet form theory in which such properties may be obtained by using some integration by parts techniques. Those techniques are very similar to the standard Malliavin calculus but are presented in a more general framework which goes far beyond the sole case of the Wiener space. In particular, we aim at providing a minimalist setting leading to our regularization lemma. Our unified framework includes the standard Malliavin calculus and different known versions: the calculus based on the splitting method developed and used in [3, 4, 6] as well as the  $\Gamma$  calculus in [2]. We also mention that our approach applies in the case of the Malliavin calculus for jump type processes as settled by Bichteler, Gravereaux and Jacod [10] and in the "lent particle" approach for Poisson point measures developed by Bouleau and Denis [12]. But we have to stress that in the framework of jump processes, another type of Malliavin calculus, based on chaotic decomposition has been developed (see [23, 24, 29, 26] and many others). In contrast with the framework presented in our paper, in this type of Malliavin calculus, the derivative operator does not verify the chain rule – so our approach does not apply.

The first aim of this paper is to present, in this unified framework, the following regularization lemma (see Theorem 3.1):

$$|\mathbb{E}(f(F)) - \mathbb{E}(f_\delta(F))| \leq C \|f\|_\infty \left( \mathbb{P}(\det \sigma_F \leq \eta) + \frac{\delta^q}{\eta^{2q}} \mathcal{C}_q(F) \right). \tag{1.1}$$

Here  $f_\delta = f * \phi_\delta$  is the regularization by convolution by means of a super kernel  $\phi_\delta$  (see (3.1)–(3.2) and (3.3)). We use Malliavin calculus (abstract version) for  $F$ : then  $\sigma_F$  is the Malliavin covariance matrix and  $\mathcal{C}_q(F)$  is a quantity which involves the Malliavin-Sobolev norms up to order  $q$  of  $F$ . This inequality holds for every  $\delta > 0, \eta > 0$  and every  $q \in \mathbb{N}$ . So one may play on these parameters according to the problem at hand. A more powerful variant of the above lemma involves derivatives of the test function  $f$ :

$$|\mathbb{E}((\partial_\gamma f)(F)) - \mathbb{E}((\partial_\gamma f_\delta)(F))| \leq C \left( \|f\|_{m,\infty} \mathbb{P}(\det \sigma_F \leq \eta) + \frac{\delta^q}{\eta^{2(q+m)}} \|f\|_\infty \mathcal{C}_{q+m}(F) \right).$$

Here  $\|f\|_{m,\infty} = \sum_{|\beta|=m} \|\partial_\beta f\|_\infty$  and  $m = |\gamma|$ . Such an inequality allows to handle convergence in distribution norms for the law of  $F_n$  to the law of  $F$ . Applications of such convergence results are given in [4, 6].

One important application of the regularization lemma consists in proving that, if a sequence  $F_n \rightarrow F$  in a distance involving smooth test functions (as for example for the Wasserstein distance) then it converges also in total variation distance. Of course, in order to get such a result, we need  $F_n$  to be smooth, in order to control  $\mathcal{C}_q(F_n)$ , and (more or less) non degenerated, in order to control  $\mathbb{P}(\det \sigma_{F_n} \leq \eta)$ . Actually, according to the non degeneracy properties, several variants of the convergence result are obtained.

Let us give an informal version of these results. Assume first that we have the uniform non degeneracy condition  $Q_p := \sup_n \mathbb{E}((\det \sigma_{F_n})^{-p}) < \infty$  for every  $p$ . Then we prove (see Lemma 3.9 for a precise statement) that, for every given  $\varepsilon > 0$

$$d_{TV}(F, F_n) \leq C d_W^{1-\varepsilon}(F, F_n) \tag{1.2}$$

where  $d_{TV}$  is the total variation distance and  $d_W$  is the Wasserstein distance. Here  $C$  is a constant which depends on the Sobolev norms and on the “non degeneracy” constant  $Q_p$  for some  $p$  large enough. Notice that we lose something, because we get the power  $1 - \varepsilon$  instead of 1 for  $d_W(F, F_n)$ . This is somehow a technical drawback of our method which is based on an optimization procedure. A more careful examination of this optimization procedure is likely to provide logarithmic losses but this would result in highly technical computations which fall beyond the scope of this paper. Let us emphasize that the previous estimate requires non-degeneracy assumptions along the whole sequence  $(F_n)$  which may be in general rather hard to check. Assumptions of non-degeneracy on the sequence  $(F_n)$  may sometimes be provided by classic anti-concentration estimates. For instance, when the underlying Gaussian functionals are polynomials, the Carbery-Wright estimate gives a kind of non-degeneracy but in a much weaker way. The reader can consult [11, 20] for results in this direction. Another reference of interest is [14] where convergence of densities is explored when the limit is Gaussian and under the same non-degeneracy assumption. Finally, let us mention the reference [25] which shows that in the particular setting of quadratic forms of Gaussian vectors, the central convergence automatically implies the required non-degeneracy assumptions and the previous results apply under the sole assumption of Gaussian convergence.

In order to bypass this major issue, we are able to obtain a variant of the above estimate without assuming anything on the non degeneracy of  $F_n$  (so  $Q_p$  may be infinite). In Proposition 3.11 we prove that, for every  $\varepsilon > 0$

$$d_{TV}(F, F_n) \leq C(d_W^{1-\varepsilon}(F, F_n) + d_W^{1-\varepsilon}(\det \sigma_F, \det \sigma_{F_n})) \tag{1.3}$$

where  $C$  depends on the Sobolev norms and  $\varepsilon$  only (and not on the non degeneracy constant  $Q_p$ ). And in Proposition 3.12 we prove the same result with  $\det \sigma_F$  replaced by a general non degenerate positive random variable, so one just needs that  $\sigma_{F_n}$  converges (non necessarily to  $\sigma_F$ ).

In concrete examples it may be difficult to precisely estimate  $d_W(\det \sigma_F, \det \sigma_{F_n})$ , but then one may use the standard upper bound  $d_W(\det \sigma_F, \det \sigma_{F_n}) \leq C \|DF - DF_n\|_1$ . Doing this is not innocent, because we replace “weak distances” with “strong” ones and this may induce a serious loss of accuracy: for example the weak distance is of order  $\frac{1}{n}$  while the strong distance is  $\frac{1}{\sqrt{n}}$ . However, the aforementioned result completely covers the case of central convergence as in this case  $\sigma_F$  is a deterministic matrix and the quantity  $d_W(F, F_n)$  is easy to estimate. Using this strategy we recover a central result of Nourdin-Peccati theory [19] establishing multivariate total variation estimates for suitable sequences converging to Gaussian. Proofs are completely different as the

proof of the aforementioned article employs tools of information theory and provides stronger results such as convergence in entropy. On the other hand, our result is more general and requires much less structural information on the sequence approximating the Gaussian law.

We also illustrate the above results in the framework of the approximation of the semi-group of a diffusion process by using the Euler scheme: if one assumes uniform ellipticity, then one has uniform non degeneracy for the Euler scheme and may use (1.2). But if one works under Hörmander condition, then the Euler scheme is degenerated so  $Q_p = \infty$ . In [9] this problem has been discussed and the authors have been obliged to work with a slightly regularized Euler scheme in order to bypass this difficulty. Now, one may use (1.3) and to get the convergence for the real Euler scheme (without regularization). But one loses accuracy: we pass from  $\frac{1}{n}$  to  $\frac{1}{\sqrt{n}}$ , so the result is not optimal.

A last type of results concerns the distance between density functions. This issue has already been discussed in [3]. Here, in Theorem 3.15 we prove the following: if  $F$  and  $G$  are smooth and non degenerated then the density functions  $p_F$  and  $p_G$  exists and are smooth. Moreover, for every multi index  $\alpha$  and for every  $\varepsilon > 0$

$$\|\partial^\alpha p_F - \partial^\alpha p_G\|_\infty \leq C d_W^{1-\varepsilon}(F, G). \tag{1.4}$$

This is a striking improvement with respect to the estimate obtained in [3], see (2.53) there.

## 2 Abstract framework

In this section we present an abstract framework which covers most of the known variants of Malliavin calculus and which allows to obtain the integration by parts formula that we need.

We consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and a subset  $\mathcal{E} \subset \cap_{p>1} L^p(\Omega; \mathbb{R})$ . The guiding example is  $\mathcal{E} = \mathcal{S}$  or also,  $\mathcal{E} = \mathbb{D}^\infty$  (the space of simple functionals respectively the space of smooth functionals in the classical Malliavin calculus). We assume that for every  $\phi \in C_p^\infty(\mathbb{R}^d)$  (smooth functions with polynomial growth) and every  $F \in \mathcal{E}^d$ , one has  $\phi(F) \in \mathcal{E}$ . In particular  $\mathcal{E}$  is an algebra. In the sequel we will also use the following consequence. For  $\eta > 0$  we denote by  $\Psi_\eta : (0, \infty) \rightarrow \mathbb{R}$  a smooth function which is equal to zero on  $(0, \eta/2)$  and to one on  $(\eta, \infty)$ . Then for every  $\eta > 0$ ,

$$F \in \mathcal{E} \implies \frac{1}{F} \Psi_\eta(F) \in \mathcal{E} \tag{2.1}$$

Moreover we consider

- ♣  $\Gamma : \mathcal{E} \times \mathcal{E} \rightarrow \mathcal{E}$  which is a symmetric bilinear form such that  $\Gamma(F, F) \geq 0$  and  $\Gamma(F, F) = 0$  iff  $F = 0$ .

In the language of Dirichlet forms  $\Gamma$  is the *carré du champ* operator. Notice that, since  $\Gamma(F, G) \in \mathcal{E}$  and  $\mathcal{E}$  is an algebra, if  $F, G, H \in \mathcal{E}$  then  $\Gamma(F, G)H \in \mathcal{E}$ . We also may define  $\Gamma(F, \Gamma(G, H))$ .

- ♣  $L : \mathcal{E} \rightarrow \mathcal{E}$  which is a linear operator.

We assume:

- ◆ **[Chain rule]** For every  $\phi \in C_p^\infty(\mathbb{R}^d)$  and  $F = (F_1, \dots, F_d) \in \mathcal{E}^d$

$$\Gamma(\phi(F), G) = \sum_{i=1}^d \partial_i \phi(F) \Gamma(F_i, G) \tag{2.2}$$

In particular, taking  $\phi(x, y) = xy$  we obtain

$$\Gamma(FH, G) = F\Gamma(H, G) + H\Gamma(F, G) \tag{2.3}$$

◆ **[Duality formula]** For every  $F, G \in \mathcal{E}$ ,

$$\mathbb{E}(\Gamma(F, G)) = \mathbb{E}(FLG) = \mathbb{E}(GLF). \tag{2.4}$$

Notice that we also have the following extension of the duality formula: using the duality first and the chain rule for the function  $\phi(x, y) = xy$  we get for every  $F, G, H \in \mathcal{E}$ :

$$\mathbb{E}(HFGL) = \mathbb{E}(\Gamma(HF, G)) = \mathbb{E}(H\Gamma(F, G) + F\Gamma(H, G))$$

so that

$$\mathbb{E}(\Gamma(F, G)H) = \mathbb{E}(F(HLG - \Gamma(H, G))) \tag{2.5}$$

This gives the standard integration by parts formula that we present now.

**Lemma 2.1.** Let  $F = (F_1, \dots, F_d) \in \mathcal{E}^d$  and let  $\sigma_F^{i,j} = \Gamma(F_i, F_j)$ ,  $i, j = 1, \dots, d$  be its Malliavin covariance matrix. We suppose that  $\sigma_F$  is invertible, we denote  $\gamma_F = \sigma_F^{-1}$ , and we assume that

$$\gamma_F^{k,i} \in \mathcal{E} \quad \forall i, k = 1, \dots, d. \tag{2.6}$$

Then for every  $\phi \in C_p^\infty(\mathbb{R}^d)$  and  $G \in \mathcal{E}$

$$\mathbb{E}(\partial_i \phi(F)G) = \mathbb{E}(\phi(F)H_i(F, G)) \tag{2.7}$$

with

$$H_i(F, G) = \sum_{k=1}^d G(\gamma_F^{k,i} LF_k - \Gamma(\gamma_F^{k,i}, F_k)) - \sum_{k=1}^d \gamma_F^{k,i} \Gamma(G, F_k). \tag{2.8}$$

Moreover, iterating this relation we get

$$\mathbb{E}(\partial_\alpha \phi(F)G) = \mathbb{E}(FH_\alpha(F, G)) \tag{2.9}$$

with  $H_\alpha(F, G)$  obtained by iterations: if  $\alpha = (\alpha_1, \dots, \alpha_m) \in \{1, \dots, d\}^m$  and  $\bar{\alpha} = (\alpha_1, \dots, \alpha_{m-1})$  then we define  $H_\alpha(F, G) = H_{\alpha_m}(F, H_{\bar{\alpha}}(F, G))$ .

**Remark 2.2.** If  $\det \sigma_F$  is almost surely invertible and  $(\det \sigma_F)^{-1} \in \mathcal{E}$  then (2.6) is verified

*Proof.* We use the chain rule and we get

$$\Gamma(\phi(F), F_k) = \sum_{i=1}^d \partial_i \phi(F) \Gamma(F_i, F_k) = \sum_{i=1}^d \partial_i \phi(F) \sigma_F^{i,k}.$$

This gives  $\nabla \phi(F) = \Gamma(\phi(F), F) \gamma_F$  which, on components reads

$$\partial_i \phi(F) = \sum_{k=1}^d \Gamma(\phi(F), F_k) \gamma_F^{k,i}.$$

Then, by (2.5)

$$\begin{aligned} \mathbb{E}(\partial_i \phi(F)G) &= \sum_{k=1}^d \mathbb{E}(\Gamma(\phi(F), F_k) \gamma_F^{k,i} G) \\ &= \sum_{k=1}^d \mathbb{E}(\phi(F) (\gamma_F^{k,i} GLF_k - \Gamma(\gamma_F^{k,i} G, F_k))). \quad \square \end{aligned}$$

The non degeneracy hypothesis (2.6) considered in Lemma 2.1 is sometimes too strong (this is the case in our framework). So we present now a localized version of the previous integration by parts formula. We recall that  $\Psi_\eta(x)$  is a smooth function which is null for  $x \leq \eta/2$  and equal to one for  $x > \eta$ . Notice that  $\sigma_F$  is invertible on the set  $\{\Psi_\eta(\det \sigma_F) > 0\}$  so we are able to define

$$\gamma_{F,\eta}^{i,j} := \gamma_F^{i,j} \Psi_\eta(\det \sigma_F).$$

And, by (2.1) we know that  $\gamma_{F,\eta}^{i,j} \in \mathcal{E}$ .

**Lemma 2.3.** *Let  $F = (F_1, \dots, F_d) \in \mathcal{E}^d$ . Then for every  $\phi \in C_p^\infty(\mathbb{R}^d)$  and  $G \in \mathcal{E}$*

$$\mathbb{E}(\partial_i \phi(F) G \Psi_\eta(\det \sigma_F)) = \mathbb{E}(F H_{\eta,i}(F, G)) \tag{2.10}$$

with

$$H_{\eta,i}(F, G) = \sum_{k=1}^d G(\gamma_{F,\eta}^{k,i} L F_k - \Gamma(\gamma_{F,\eta}^{k,i} G, F_k)). \tag{2.11}$$

Moreover, iterating this relation we get

$$\mathbb{E}(\partial_\alpha \phi(F) G \Psi_\eta(\det \sigma_F)) = \mathbb{E}(F H_{\eta,\alpha}(F, G)) \tag{2.12}$$

with  $H_{\eta,\alpha}(F, G)$  obtained by iterations: if  $\alpha = (\alpha_1, \dots, \alpha_m) \in \{1, \dots, d\}^m$  and  $\bar{\alpha} = (\alpha_1, \dots, \alpha_{m-1})$  then we define  $H_{\eta,\alpha}(F, G) = H_{\eta,\alpha_m}(F, H_{\eta,\bar{\alpha}}(F, G))$ .

*Proof.* The proof is almost the same as above. The only change is that in the first step we multiply with  $\Psi_\eta(\det \sigma_F)$  and we write

$$\Gamma(\phi(F), F_k) \Psi_\eta(\det \sigma_F) = \sum_{i=1}^d \partial_i \phi(F) \sigma_F^{i,k} \Psi_\eta(\det \sigma_F).$$

On the set  $\Psi_\eta(\det \sigma_F) > 0$  the matrix  $\sigma_F$  is invertible so we get

$$\partial_i \phi(F) \Psi_\eta(\det \sigma_F) = \sum_{k=1}^d \Gamma(\phi(F), F_k) \Psi_\eta(\det \sigma_F) \gamma_F^{k,i} = \sum_{k=1}^d \Gamma(\phi(F), F_k) \gamma_{F,\eta}^{k,i}$$

and then, by (2.5)

$$\begin{aligned} \mathbb{E}(\partial_i \phi(F) G \Psi_\eta(\det \sigma_F)) &= \sum_{k=1}^d \mathbb{E}(\Gamma(\phi(F), F_k) \gamma_{F,\eta}^{k,i} G) \\ &= \sum_{k=1}^d \mathbb{E}(\phi(F) (\gamma_{F,\eta}^{k,i} G L F_k - \Gamma(\gamma_{F,\eta}^{k,i} G, F_k))). \quad \square \end{aligned}$$

### 2.1 Norms

In order to be able to give estimates of  $H_\alpha(F, G)$  we need to assume that  $\Gamma$  is given by a derivative operator as follows (this is actually always true, see Mokobodzki [16]):

- ♣ We assume that there exists a separable Hilbert space  $\mathcal{H}$  and a linear application  $D : \mathcal{E} \rightarrow \cap_{p>1} L^p(\Omega; \mathcal{H})$  such that

- i)  $\Gamma(F, G) = \langle DF, DG \rangle_{\mathcal{H}}$
- ii)  $D_h F := \langle DF, h \rangle_{\mathcal{H}} \in \mathcal{E}$ .

We also assume that we have the chain rule: for  $\phi \in C^\infty(\mathbb{R}^d)$  and  $F \in \mathcal{E}^d$  we have

$$iii) \quad D\phi(F) = \sum_{i=1}^d \partial_i \phi(F) D F_i.$$

Then we may define higher order derivatives in the following way.  $D^2 : \mathcal{E} \rightarrow \cap_{p>1} L^p(\Omega; \mathcal{H}^{\otimes 2})$  is defined by  $\langle D^2 F, h_1 \otimes h_2 \rangle_{\mathcal{H}^{\otimes 2}} = D_{h_2} D_{h_1} F$ . So, if we denote  $D_{h_1, h_2}^2 F = \langle D^2 F, h_1 \otimes h_2 \rangle_{\mathcal{H}^{\otimes 2}}$  then  $D_{h_1, h_2}^2 F = D_{h_2} D_{h_1} F$ . In a similar way we define (by induction)

$$D_{h_1, \dots, h_k}^k F = D_{h_k} D_{h_1, \dots, h_{k-1}}^{k-1} F.$$

We introduce now the norms

$$|F|_{1,k} = \sum_{i=1}^k |D^i F|_{\mathcal{H}^{\otimes i}}, \quad |F|_k = |F| + |F|_{1,k} = |F| + \sum_{i=1}^k |D^i F|_{\mathcal{H}^{\otimes i}}$$

For  $F = (F_1, \dots, F_d) \in \mathcal{E}^d$  we define

$$|F|_{1,k} = \sum_{i=1}^d |F_i|_{1,k}, \quad |F|_k = \sum_{i=1}^d |F_i|_k.$$

Notice that since  $\mathcal{H}$  is separable we may take an orthonormal base  $(e_i)_{i \in \mathbb{N}}$  and denote  $D_i F = D_{e_i} F = \langle DF, e_i \rangle$ . Then  $DF = \sum_{i=1}^{\infty} D_i F \times e_i$  and more generally

$$D^k F = \sum_{i_1, \dots, i_k} D_{i_1, \dots, i_k} F \times \otimes_{j=1}^k e_{i_j}.$$

Moreover, on the set  $\{\det \sigma_F > 0\}$  we denote

$$\alpha_k = \frac{|F|_{1,k+1}^{2(d-1)} (|F|_{1,k+1} + |LF|_k)}{\det \sigma_F}, \quad \beta_k = \frac{|F|_{1,k+1}^{2d}}{\det \sigma_F} \tag{2.13}$$

and

$$\mathcal{K}_{n,k}(F) = (|F|_{1,k+n+1} + |LF|_{k+n})^n (1 + |F|_{1,k+n+1})^{2d(2n+k)}, \tag{2.14}$$

$$\mathcal{C}_n(F) = \mathcal{K}_{n,0}(F) = (|F|_{1,n+1} + |LF|_n)^n (1 + |F|_{1,n+1})^{4dn} \tag{2.15}$$

$$\mathcal{C}_{n,p}(F) = \|\mathcal{C}_n(F)\|_p. \tag{2.16}$$

Then, the following lemma is proved in the Appendix of [4]:

**Lemma 2.4. A.** *Let  $F \in \mathcal{E}^d$ . Suppose that  $\det \sigma_F(\omega) > 0$ . Then, for every  $k$  and  $n$  there exists a constant  $C$  such that for every multi-index  $\rho$  with  $|\rho| \leq n$  one has*

$$|H_\rho(F, G)|_k \leq C \alpha_{k+n}^n \sum_{p_1+p_2 \leq k+n} |G|_{p_2} (1 + \beta_{k+n})^{p_1}. \tag{2.17}$$

**B.** *For every  $\eta > 0$*

$$|H_\rho(F, \Psi_\eta(\det \sigma_F)G)|_k \leq \frac{C}{\eta^{2n+k}} \times \mathcal{K}_{n,k}(F) \times |G|_{k+n}. \tag{2.18}$$

As an immediate consequence of (2.18) and of (2.12) we get

**Corollary 2.5.** *Let  $F \in \mathcal{E}^d$  and  $\eta > 0$ . Then*

$$|\mathbb{E}(\partial_\alpha f(F) \Psi_\eta(\det \sigma_F))| \leq \frac{C}{\eta^{2|\alpha|}} \times \mathcal{C}_{|\alpha|,1}(F) \times \|f\|_\infty. \tag{2.19}$$

### 3 Regularization lemma

We go now on and we give the regularization lemma. We recall that a super kernel  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  is a function which belongs to the Schwartz space  $\mathbf{S}$  (infinitely differentiable functions with rapid decrease) and such that for every multi-indexes  $\alpha$  and  $\beta$ , one has

$$\int \phi(x)dx = 1 \text{ and } \int y^\alpha \phi(y)dy = 0 \text{ for } |\alpha| \geq 1, \tag{3.1}$$

$$\int |y|^m |\partial_\beta \phi(y)| dy < \infty. \tag{3.2}$$

As usual, for a multi-index  $\alpha = (\alpha_1, \dots, \alpha_m) \in \{1, \dots, d\}^m$  then  $|\alpha| = m$  and  $y^\alpha = \prod_{i=1}^m y_{\alpha_i}$ . Since super kernels play a crucial role in our approach we give here the construction of such an object. We do it in dimension  $d = 1$  and then we take tensor products. We take  $\psi \in \mathbf{S}$  which is symmetric and equal to one in a neighborhood of zero and we define  $\phi = \mathcal{F}^{-1}\psi$ , the inverse of the Fourier transform of  $\psi$ . Since  $\mathcal{F}^{-1}$  sends  $\mathbf{S}$  into  $\mathbf{S}$  the property (3.2) is verified. And we also have  $0 = \psi^{(m)}(0) = i^{-m} \int x^m \phi(x)dx$  so (3.1) holds as well. We finally normalize in order to obtain  $\int \phi = 1$ .

We fix a super kernel  $\phi$ . For  $\delta \in (0, 1)$  and for a function  $f$  we define

$$\phi_\delta(y) = \frac{1}{\delta^d} \phi\left(\frac{y}{\delta}\right) \text{ and } f_\delta = f * \phi_\delta, \tag{3.3}$$

the symbol  $*$  denoting convolution. Moreover, for  $f \in C_b^\infty(\mathbb{R}^d)$  we denote

$$\|f\|_{k,\infty} = \sum_{0 \leq |\alpha| \leq k} \|\partial_\alpha f\|_\infty. \tag{3.4}$$

**Lemma 3.1.** *Let  $f \in C_b^\infty(\mathbb{R}^d)$  and  $F \in \mathcal{E}^d$ . For every  $q, m \in \mathbb{N}$  there exists a universal constant  $C$  (depending on  $q$  and  $m$  only) such that for every multi-index  $\gamma$  with  $|\gamma| = m$ , every  $\delta > 0$  and every  $\eta > 0$*

$$|\mathbb{E}(\partial_\gamma f(F)) - \mathbb{E}(\partial_\gamma f_\delta(F))| \leq C \|\partial_\gamma f\|_\infty \mathbb{P}(\det \sigma_F \leq \eta) + \frac{\delta^q}{\eta^{2(q+m)}} \|f\|_\infty \mathcal{C}_{q+m,1}(F) \tag{3.5}$$

with  $\mathcal{C}_{q+m}(F)$  given in (2.15). In particular, taking  $m = 0$

$$|\mathbb{E}(f(F)) - \mathbb{E}(f_\delta(F))| \leq C \|f\|_\infty \left( \mathbb{P}(\det \sigma_F \leq \eta) + \frac{\delta^q}{\eta^{2q}} \mathcal{C}_{q,1}(F) \right) \tag{3.6}$$

**Remark 3.2.** A similar estimate holds for  $|\mathbb{E}(G\partial_\gamma f(F)) - \mathbb{E}(G\partial_\gamma f_\delta(F))|$  with  $G \in \mathcal{E}$ . But in this case one has to replace  $\mathbb{P}(\det \sigma_F \leq \eta)$  with  $\|G\|_2 \mathbb{P}^{1/2}(\det \sigma_F \leq \eta)$  and  $\mathcal{C}_{q+m,1}(F)$  by  $\| |G|_{q+m+1} \|_2 \mathcal{C}_{q+m,2}(F)$  in the right hand side of (3.5). The proof is the same.

*Proof.* The proof is given in [4] in a particular framework, but, for the convenience of the reader, we recall it here. Using Taylor expansion of order  $q$  (with integral remainder) and (3.1) we obtain

$$\partial_\gamma f(x) - \partial_\gamma f_\delta(x) = \int (\partial_\gamma f(x) - \partial_\gamma f(y)) \phi_\delta(x - y) dy = \int R_{\gamma,q}(x, y) \phi_\delta(x - y) dy$$

with

$$R_{\gamma,q}(x, y) = \frac{1}{q!} \sum_{|\alpha|=q} \int_0^1 \partial_\alpha \partial_\gamma f(x + \lambda(y - x)) (x - y)^\alpha (1 - \lambda)^q d\lambda.$$

By a change of variable we get

$$\int R_{\gamma,q}(x, y) \phi_\delta(x - y) dy = \frac{1}{q!} \sum_{|\alpha|=q} \int_0^1 \int dz \phi_\delta(z) \partial_\alpha \partial_\gamma f(x + \lambda z) z^\alpha (1 - \lambda)^q d\lambda.$$

So, we have

$$\begin{aligned} & \mathbb{E}(\Psi_\eta(\det \sigma_F)\partial_\gamma f(F)) - \mathbb{E}(\Psi_\eta(\det \sigma_F)\partial_\gamma f_\delta(F)) \\ &= \mathbb{E}\left(\int \Psi_\eta(\det \sigma_F)R_{\gamma,q}(F, y)\phi_\delta(F - y)dy\right) \\ &= \frac{1}{q!} \sum_{|\alpha|=q} \int_0^1 \int dz \phi_\delta(z)\mathbb{E}(\Psi_\eta(\det \sigma_F)\partial_\alpha \partial_\gamma f(F + \lambda z))z^\alpha(1 - \lambda)^q d\lambda. \end{aligned} \tag{3.7}$$

As a consequence of (2.19)

$$|\mathbb{E}(\Psi_\eta(\det \sigma_F)\partial_\alpha \partial_\gamma f(F + \lambda z))| \leq \frac{C}{\eta^{2(q+m)}}\mathcal{C}_{q+m,1}(F) \|f\|_\infty.$$

We also have, if  $|\alpha| = q$ ,

$$\int dz |\phi_\delta(z)z^\alpha| \leq \delta^q \int |\phi(z)||z|^{|\alpha|} dz \tag{3.8}$$

so we conclude that

$$I_1 = |\mathbb{E}(\Psi_\eta(\det \sigma_F)\partial_\gamma f(F)) - \mathbb{E}(\Psi_\eta(\det \sigma_F)\partial_\gamma f_\delta(F))| \leq \frac{C\delta^q}{\eta^{2(q+m)}}\mathcal{C}_{q+m,1}(F) \|f\|_\infty.$$

We write now

$$I_2 = |\mathbb{E}((1 - \Psi_\eta(\det \sigma_F))\partial_\gamma f(F))| \leq \|\partial_\gamma f\|_\infty \mathbb{P}(\det \sigma_F < \eta)$$

and similarly,

$$I_3 = |\mathbb{E}((1 - \Psi_\eta(\det \sigma_F))\partial_\gamma f_\delta(F))| \leq \|\partial_\gamma f_\delta\|_\infty \mathbb{P}(\det \sigma_F < \eta) \leq \|\partial_\gamma f\|_\infty \mathbb{P}(\det \sigma_F < \eta).$$

Since  $|\mathbb{E}(\partial^\gamma f(F)) - \mathbb{E}(\partial^\gamma f_\delta(F))| \leq I_1 + I_2 + I_3$ , (3.5) follows.  $\square$

**Remark 3.3.** The main contribution of Lemma 3.1 is to separate the quantities concerning the non degeneracy (that is,  $\mathbb{P}(\det \sigma_F \geq \eta)$ ) and the regularity (the Malliavin Sobolev norms contained in  $\mathcal{C}_{m,1}(F)$ ): on one hand we localize on the set  $\{\det \sigma_F \geq \eta\}$  where we have non degeneracy and there we can use the integration by parts formula which involves the Sobolev norms (multiplied by  $1/\eta$ ). This is  $I_1$ . On the other hand, on the set where  $\det \sigma_F \leq \eta$  we are not able to use integration by parts so we upper bound  $\partial_\gamma f$  by the infinite norm (put it otherwise, we do nothing). These are  $I_2$  and  $I_3$ .

Under a weak non degeneracy condition for  $F$ , we get the following immediate consequence.

**Corollary 3.4.** Let  $F \in \mathcal{E}^d$ . Suppose that for some  $\kappa > 0$  there exists a positive constant  $\theta_\kappa(F)$  (so, possibly depending on  $F$ ) such that

$$\mathbb{P}(\det \sigma_F \leq \eta) \leq \theta_\kappa(F)\eta^\kappa \quad \forall \eta > 0. \tag{3.9}$$

Then for every,  $q \in \mathbb{N}$  and  $\delta > 0$

$$|\mathbb{E}(f(F)) - \mathbb{E}(f_\delta(F))| \leq C \|f\|_\infty \mathcal{C}_{q,1}^{\frac{\kappa}{\kappa+2q}}(F)\theta_\kappa^{\frac{2q}{\kappa+2q}}(F) \times \delta^{\frac{\kappa q}{\kappa+2q}}. \tag{3.10}$$

*Proof.* One plugs (3.9) into (3.6). This gives

$$|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| \leq C \|f\|_\infty \left( \theta_\kappa(F)\eta^\kappa + \frac{\delta^q}{\eta^{2q}}\mathcal{C}_{q,1}(F) \right).$$

Then we optimize over  $\eta$  and  $\delta$ , that is, we choose  $\eta$  in order that the addenda in the above r.h.s. are equal. So,  $\eta = (\delta^q\mathcal{C}_{q,1}(F)/\theta_\kappa(F))^{\frac{1}{\kappa+2q}}$  and by inserting, (3.10) follows.  $\square$

**Example.** Let  $F = \Delta^2$  with  $\Delta$  a standard normal random variable. We use standard Malliavin calculus to see what (3.10) gives in this case. We have  $DF = 2\Delta$  so that  $\sigma_F = 4\Delta^2$  and then (3.9) reads

$$\mathbb{P}(4\Delta^2 \leq \eta) = \mathbb{P}\left(|\Delta| \leq \frac{1}{2}\sqrt{\eta}\right) \leq \eta^{1/2}.$$

So  $\kappa = \frac{1}{2}$  here and for every  $\varepsilon > 0$ , (3.10) gives (one takes  $q$  large enough)

$$|\mathbb{E}(f(F)) - \mathbb{E}(f_\delta(F))| \leq C_\varepsilon \|f\|_\infty \times \delta^{\frac{\kappa}{2} - \varepsilon} = C_\varepsilon \|f\|_\infty \times \delta^{\frac{1}{4} - \varepsilon}.$$

But some informal computations give

$$|\mathbb{E}(f(F)) - \mathbb{E}(f_\delta(F))| \sim C \|f\|_\infty \times \delta^{\frac{1}{2}}.$$

So our calculus is not sharp: we get  $\frac{1}{4}$  instead of  $\frac{1}{2}$ .

In the regularization Lemma 3.1 we have not assumed that  $\sigma_F$  is invertible but we preferred to keep  $\mathbb{P}(\det \sigma_F \leq \eta)$ . We give now a variant under a strong non degeneracy assumption for  $F$ : for every  $p \geq 1$

$$\mathbb{E}((\det \sigma_F)^{-p}) \leq C_p < \infty. \tag{3.11}$$

Denote

$$\mathcal{Q}_l(F) = \mathcal{C}_{l,2}(F) (\mathbb{E}(\det \sigma_F)^{-2l})^{1/2} \tag{3.12}$$

and  $\mathcal{C}_{l,2}(F)$  is given in (2.16).

**Lemma 3.5.** *Let  $f \in C_b^\infty(\mathbb{R}^d)$  and  $F \in \mathcal{E}^d$  such that (3.11) holds. For every  $q, m \in \mathbb{N}$  there exists a universal constant  $C$  (depending on  $q$  and  $m$  only) such that for every multi-index  $\gamma$  with  $|\gamma| \leq m$ , every  $\delta > 0$  and every  $\eta > 0$*

$$|\mathbb{E}(\partial_\gamma f(F)) - \mathbb{E}(\partial_\gamma f_\delta(F))| \leq C \delta^q \|f\|_\infty \mathcal{Q}_{q+m}(F). \tag{3.13}$$

*Proof.* We follow the same reasoning as in the previous proof and we come back to (3.7), but we do not multiply with  $\Psi_\eta(\det \sigma_F)$  anymore:

$$\begin{aligned} & \mathbb{E}(\partial_\gamma f(F)) - \mathbb{E}(\partial_\gamma f_\delta(F)) \\ &= \frac{1}{q!} \sum_{|\alpha|=q} \int_0^1 \int dz \phi_\delta(z) \mathbb{E}(\partial_\alpha \partial_\gamma f(F + \lambda z)) z^\alpha (1 - \lambda)^q d\lambda. \end{aligned}$$

Using the standard integration by parts formula (2.9) we obtain, for some  $p \geq 1$

$$|\mathbb{E}(\partial_\alpha \partial_\gamma f(F + \lambda z))| \leq (\mathbb{E}(\det \sigma_F)^{-2(q+m)})^{1/2} \mathcal{C}_{q+m,2}(F) \|f\|_\infty.$$

And by (3.8) we conclude that

$$|\mathbb{E}(\partial_\gamma f(F)) - \mathbb{E}(\partial_\gamma f_\delta(F))| \leq \mathcal{Q}_{q+m}(F) \|f\|_\infty \delta^q. \quad \square$$

In [4] one gives the following more sophisticated version of the regularization lemma for smooth functions with polynomial growth. More precisely we denote by  $C_p^\infty(\mathbb{R}^d)$  the space of smooth functions such that for every  $q \in \mathbb{N}$  there exists  $L_q(f)$  and  $l_q(f)$  such that, for every multi index with  $|\alpha| \leq q$  and every  $x \in \mathbb{R}^d$

$$|\partial_\alpha f(x)| \leq L_q(f)(1 + |x|)^{l_q(f)}.$$

Then we have the following result (see [4] Lemma 5.3)

**Lemma 3.6.** *Let  $F \in \mathcal{E}^d$  and  $q, m \in \mathbb{N}$ . There exists some constant  $C \geq 1$ , depending on  $d, m$  and  $q$  only, such that for every  $f \in C_{\text{pol}}^{q+m}(\mathbb{R}^d)$ , every multi-index  $\gamma$  with  $|\gamma| = m$  and every  $\eta, \delta > 0$  and  $p > 1$*

$$|\mathbb{E}(\partial_\gamma f(x + F)) - \mathbb{E}(\partial_\gamma f_\delta(x + F))| \leq C(1 + \|F\|_{p_{l_0}(f)})^{l_0(f)}(1 + |x|)^{l_m(f)} \times \left( L_m(f)c_{l_m(f)}2^{l_m(f)}\mathbb{P}^{(p-1)/p}(\det \sigma_F \leq \eta) + 2^{l_0(f)}c_{l_0(f)+q}L_0(f)\frac{\delta^q}{\eta^{2(q+m)}}C_{q+m,1}(F) \right). \tag{3.14}$$

**3.1 Convergence in total variation distance**

Let us introduce the following distances:

$$d_k(F, G) = \sup\{|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| : \sum_{|\alpha|=k} \|\partial^\alpha f\|_\infty \leq 1\} \tag{3.15}$$

In the case  $k = 0$  this means that the test functions  $f$  are just measurable and bounded and in this case  $d_0$  is the so called “total variation distance” that we will denote by  $d_{TV}$ . Another interesting distance is the “Wasserstein distance”

$$d_W(F, G) = d_1(F, G) = \sup\{|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| : \|\nabla f\|_\infty \leq 1\}.$$

In many problems the estimate of the error involves some Taylor type expansions and then the test functions have to be differentiable and the norms of the derivatives come on. So we are able to estimate  $d_k$  for some  $k$ . And then one asks about the possibility to obtain estimates for measurable test functions, as in total variation distance. And one may use the regularization lemma presented before in order to do it. We give several forms of such a result. But as we will show, the regularization lemma can be applied to other distances. As an example, in the sequel we consider also the following distance between random vectors in  $\mathbb{R}^d$ : we set

$$d_{CF}(F, G) = \sup\{|\mathbb{E}(e^{i\langle \vartheta, F \rangle}) - \mathbb{E}(e^{i\langle \vartheta, G \rangle})| : \vartheta \in \mathbb{R}^d\}. \tag{3.16}$$

So,  $d_{CF}$  is the maximum distance between the characteristic functions of  $F$  and  $G$ . There are many situations where it is easier to obtain bounds on the difference of characteristic functions, especially when the targets in consideration are infinitely divisible. One may for instance consult [1] for an introduction to Stein’s method theory for this kind of distribution. Again, the regularization lemma allows one to pass from such distance to the distance in total variation.

The key remark which allows to use the regularization lemma is the following.

**Lemma 3.7.** *Let  $\phi_\delta$  be the super-kernel introduced in (3.1)–(3.2) and let  $F, G$  denote random vectors in  $\mathbb{R}^d$ . We also fix a multi-index  $\beta$  with  $|\beta| = r$  (including the void multi-index, in which case  $r = 0$ ). Then for every  $k \in \mathbb{N}$  there exists a constant  $C$  depending on  $k$  and  $r$  only such that for every  $f \in C_b(\mathbb{R}^d)$*

$$|\mathbb{E}(\partial^\beta(f * \phi_\delta)(F)) - \mathbb{E}(\partial^\beta(f * \phi_\delta)(G))| \leq C\delta^{-(k+r)}d_k(F, G) \times \|f\|_\infty. \tag{3.17}$$

We also have, for a constant  $C > 0$  depending on  $r$  only,

$$|\mathbb{E}(\partial^\beta(f * \phi_\delta)(F)) - \mathbb{E}(\partial^\beta(f * \phi_\delta)(G))| \leq C\delta^{-(2d+r)}d_{CF}(F, G) \times \|f\|_1. \tag{3.18}$$

And moreover, for every  $\varepsilon > 0$  one may find  $C$  (depending on  $\varepsilon$  and  $r$ ) such that

$$|\mathbb{E}(\partial^\beta(f * \phi_\delta)(F)) - \mathbb{E}(\partial^\beta(f * \phi_\delta)(G))| \leq C\delta^{-(2d+r)}d_{CF}(F, G)^{1-\varepsilon} \times \|f\|_\infty. \tag{3.19}$$

*Proof.* (3.17) is an immediate consequence of the definition of  $d_k$  and of  $\|\partial^\alpha(f * \phi_\delta)\|_\infty \leq C\delta^{-|\alpha|} \times \|f\|_\infty$ .

Let us prove (3.18). Take first  $f \in \mathbf{S}$  (Schwartz space). Let  $\mathcal{F}f(\xi) = \int f(x)e^{-2\pi i(x,\xi)} dx$  denote the Fourier transform of  $f$  and  $\mathcal{F}_F(\xi) = \mathbb{E}(e^{i(F,\xi)})$  be the characteristic function of  $F$ . Then, using the inverse  $\mathcal{F}^{-1}$  of  $\mathcal{F}$

$$\mathbb{E}(f(F)) = \mathbb{E}(\mathcal{F}\mathcal{F}^{-1}f(F)) = \mathbb{E} \int \mathcal{F}^{-1}f(\xi)e^{-2\pi i(F,\xi)} d\xi = \int \mathcal{F}^{-1}f(\xi)\mathcal{F}_F(-2\pi\xi)d\xi.$$

Take  $\alpha$  such that  $\partial^\alpha = \partial_1^2 \dots \partial_d^2$  and notice that, by using integration by parts, one obtains  $\mathcal{F}^{-1}\partial^\alpha f(\xi) = (2\pi i)^{2d} \prod_{i=1}^d \xi_i^2 \mathcal{F}^{-1}f(\xi)$ . Setting  $B(0, 1)$  the ball of radius 1 centered at 0 in  $\mathbb{R}^d$ , we can write

$$\mathbb{E}(f(F)) = \int_{B(0,1)} \mathcal{F}^{-1}f(\xi)\mathcal{F}_F(-2\pi i\xi)d\xi + \frac{1}{(2\pi i)^{2d}} \int_{B(0,1)^c} \prod_{i=1}^d \xi_i^{-2} \mathcal{F}^{-1}\partial^\alpha f(\xi)\mathcal{F}_F(-2\pi i\xi)d\xi.$$

It follows that

$$\begin{aligned} |\mathbb{E}(f(F)) - \mathbb{E}(f(G))| &\leq C\|\mathcal{F}_F - \mathcal{F}_G\|_\infty (\|\mathcal{F}^{-1}f\|_\infty + \|\mathcal{F}^{-1}\partial^\alpha f\|_\infty) \\ &= C d_{CF}(F, G) (\|\mathcal{F}^{-1}f\|_\infty + \|\mathcal{F}^{-1}\partial^\alpha f\|_\infty). \end{aligned}$$

We will use this formula with  $f * \phi_\delta$  instead of  $f$ . One has

$$\begin{aligned} \|\mathcal{F}^{-1}\partial^\gamma(f * \phi_\delta)\|_\infty &= \delta^{-|\gamma|} \|\mathcal{F}^{-1}(f * \partial^\alpha \partial^\beta \phi)\|_\infty \leq \delta^{-|\gamma|} \|f * \partial^\gamma \phi\|_1 \\ &\leq \delta^{-|\gamma|} \|f\|_1 \|\partial^\gamma \phi\|_1 \leq C\delta^{-|\gamma|} \|f\|_1. \end{aligned}$$

So, by inserting above, (3.18) is proved. In order to prove (3.19) we take a truncation function  $\Psi_M \in C^\infty(\mathbb{R}^d)$  such that  $1_{B_M(0)} \leq \Psi_M \leq 1_{B_{M+1}(0)}$  and we use (3.18) for  $f\Psi_M$ . Since  $\|f\Psi_M\|_1 \leq CM^d \|f\|_\infty$  we obtain

$$|\mathbb{E}((f\Psi_M) * \partial^\beta \phi_\delta(F)) - \mathbb{E}((f\Psi_M) * \partial^\beta \phi_\delta(G))| \leq C\delta^{-(2d+r)} M^d \|f\|_\infty d_{CF}(F, G).$$

Notice that for  $q$  one has

$$\int |y|^q |\partial^\beta \phi_\delta(y)| dy = \delta^{q-r} \int |y|^q |\partial^\beta \phi(y)| dy.$$

Moreover, since  $F$  has finite moments of any order, one has,

$$\mathbb{P}(|F - y| \geq M + 1) \leq M^{-q} \mathbb{E}(|F - y|^q) \leq CM^{-q}(1 + |y|^q).$$

So, for every  $q$ ,

$$\begin{aligned} |\mathbb{E}((f(1 - \Psi_M)) * \partial^\beta \phi_\delta(F))| &\leq \|f\|_\infty \int \mathbb{P}(|F - y| \geq M + 1) |\partial^\beta \phi_\delta(y)| dy \\ &\leq C\delta^{-r} \|f\|_\infty M^{-q}. \end{aligned}$$

The same is true for  $G$  and then, combining the two previous estimates we obtain, for every  $q \in \mathbb{N}$

$$|\mathbb{E}(f * \partial^\beta \phi_\delta(F)) - \mathbb{E}(f * \partial^\beta \phi_\delta(G))| \leq C\delta^{-(2d+r)} \|f\|_\infty (d_{CF}(F, G)M^d + M^{-q}).$$

We optimize over  $M$ : we choose  $M$  such that  $d_{CF}(F, G)M^d = M^{-q}$ . We get  $M = d_{CF}^{-\frac{1}{q+d}}(F, G)$  and by inserting, we obtain (3.19).  $\square$

We first consider the case in which both  $F$  and  $G$  satisfy the weak non degeneracy condition in (3.9).

**Lemma 3.8.** *Let  $F, G \in \mathcal{E}^d$  be such that (3.9) holds true for some fixed  $\kappa > 0$ . Then for every  $k, q \in \mathbb{N}$*

$$d_{TV}(F, G) \leq C \times (C_{\kappa,q}(F) + C_{\kappa,q}(G))^{\frac{k}{\kappa+a}} d_k(F, G)^{\frac{a}{\kappa+a}} \tag{3.20}$$

with

$$C_{\kappa,q}(F) = \theta_{\kappa}(F)^{\frac{2q}{\kappa+2q}} \mathcal{C}_{q,2}(F)^{\frac{\kappa}{\kappa+2q}}, \quad a = \frac{\kappa q}{\kappa + 2q}. \tag{3.21}$$

Moreover, for every  $\varepsilon > 0$

$$d_{TV}(F, G) \leq C \times (C_{\kappa,q}(F) + C_{\kappa,q}(G))^{\frac{2d}{2d+a}} d_{CF}(F, G)^{\frac{a(1-\varepsilon)}{2d+a}} \tag{3.22}$$

*Proof.* By (3.17) (with  $r = 0$ )

$$|\mathbb{E}(f * \phi_{\delta}(F)) - \mathbb{E}(f * \phi_{\delta}(G))| \leq \|f\|_{\infty} d_k(F, G) \delta^{-k},$$

so, using (3.10) we get

$$|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| \leq C \|f\|_{\infty} ((C_{\kappa,q}(F) + C_{\kappa,q}(G)) \delta^{\frac{\kappa q}{\kappa+2q}} + d_k(F, G) \delta^{-k}).$$

We optimize over  $\delta$ : we choose  $\delta$  such that  $(C_{\kappa,q}(F) + C_{\kappa,q}(G)) \delta^{\frac{\kappa q}{\kappa+2q}} = d_k(F, G) \delta^{-k}$ . We get  $\delta = ((C_{\kappa,q}(F) + C_{\kappa,q}(G))/d_k(F, G))^{\frac{1}{a+k}}$ ,  $a$  being given in (3.21), and by inserting, (3.20) follows. The proof of (3.22) is the same, but one employs (3.19).  $\square$

We give now a result under the strong non degeneracy condition both for  $F$  and  $G$ :  $(\det \sigma_F)^{-1}, (\det \sigma_G)^{-1} \in \cap_{p \geq 1} L^p$ .

**Lemma 3.9.** *Let  $F, G \in \mathcal{E}$  be such that  $\mathcal{Q}_q(F) + \mathcal{Q}_q(G) < \infty$  for every  $q \in \mathbb{N}$  (see (3.12)). Then, for every  $q, k \in \mathbb{N}$  there exists a constant  $C$  (depending on  $q$  and  $k$  only) such that*

$$d_{TV}(F, G) \leq C (\mathcal{Q}_q(F) + \mathcal{Q}_q(G))^{\frac{k}{q+k}} \times d_k^{\frac{q}{q+k}}(F, G). \tag{3.23}$$

In particular, for every  $\varepsilon > 0$

$$d_{TV}(F, G) \leq C (\mathcal{Q}_{q(\varepsilon)}(F) + \mathcal{Q}_{q(\varepsilon)}(G))^{\varepsilon} \times d_k^{1-\varepsilon}(F, G) \tag{3.24}$$

with  $q(\varepsilon) = k(\frac{1}{\varepsilon} - 1)$ . Moreover, with  $q(\varepsilon) = 2d(\frac{1}{\varepsilon} - 1)$ ,

$$d_{TV}(F, G) \leq C (\mathcal{Q}_{q(\varepsilon)}(F) + \mathcal{Q}_{q(\varepsilon)}(G))^{\varepsilon} \times d_{CF}^{(1-\varepsilon)^2}(F, G) \tag{3.25}$$

*Proof.* Let  $f \in C_b^{\infty}(\mathbb{R})$  and  $\delta > 0$ . Using (3.13)

$$|\mathbb{E}(f(F)) - \mathbb{E}(f * \phi_{\delta}(F))| \leq C \delta^q \|f\|_{\infty} \mathcal{Q}_q(F)$$

and a similar estimate holds for  $G$ . Moreover by (3.17)

$$|\mathbb{E}(f_{\delta}(F)) - \mathbb{E}(f * \phi_{\delta}(G))| \leq \frac{C}{\delta^k} \|f\|_{\infty} d_k(F, G)$$

so that

$$|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| \leq C \|f\|_{\infty} \left( \delta^q (\mathcal{Q}_q(F) + \mathcal{Q}_q(G)) + \frac{1}{\delta^k} d_k(F, G) \right).$$

We choose  $\delta$  such that  $\delta^q (\mathcal{Q}_q(F) + \mathcal{Q}_q(G)) = \frac{1}{\delta^k} d_k(F, G)$ , that is,  $\delta = (d_k(F, G)/(\mathcal{Q}_q(F) + \mathcal{Q}_q(G)))^{\frac{1}{q+k}}$  and we get (3.23). Using (3.19), we obtain, in the same way, (3.25).  $\square$

**Remark 3.10.** Compare with Corollary 2.8 pg 11/33 in [3] – there we have  $d_k^{1/(k+1)}(F, F_n)$ . So it is much less good there. The reason is that we have now a much stronger regularization lemma. Compare the estimate (2.29) pg 8/33 in [3] with (3.6) here: here we have “for every  $q$ ” and this is what gives the much better result.

We finish this section with a variant of the previous Lemma: now we assume the strong non degeneracy condition  $(\det \sigma_F)^{-1} \in \cap_{p \geq 1} L^p$  for  $F$  but we assume no non degeneracy condition on  $G$ . Then we get the following:

**Proposition 3.11.** *Let  $F, G \in \mathcal{E}$  be such that  $\mathcal{Q}_q(F) + \mathcal{C}_{q,1}(G) < \infty$  for every  $q \in \mathbb{N}$  (see (3.12)). Then, for every  $p, p' \in \mathbb{N}$  and every  $\varepsilon > 0$  there exists a constant  $C$  (depending on  $\varepsilon, p, p'$ ) such that*

$$d_{TV}(F, G) \leq C \times C_\varepsilon(F, G) \times (d_p(F, G) + d_{p'}(\det \sigma_F, \det \sigma_G))^{1-\varepsilon} \tag{3.26}$$

with

$$C_\varepsilon(F, G) = 1 + \mathcal{Q}_{q(\varepsilon)}(F) + \mathcal{C}_{q(\varepsilon),1}(G), \quad q(\varepsilon) = \max\{[4p/\varepsilon] + 1, [p'/2\varepsilon] + 1\}. \tag{3.27}$$

Moreover, with  $q(\varepsilon) = \max\{[8d/\varepsilon] + 1, [p'/2\varepsilon] + 1\}$  one has

$$d_{TV}(F, G) \leq C \times C_\varepsilon(F, G) \times (d_{CF}(F, G) + d_{p'}(\det \sigma_F, \det \sigma_G))^{1-\varepsilon}. \tag{3.28}$$

*Proof.* We denote

$$d_{p,p'} = d_p(F, G) + d_{p'}(\det \sigma_F, \det \sigma_G).$$

Take  $\eta > 0$  and take  $\Phi_\eta \in C_b^\infty(\mathbb{R}_+)$  such that  $1_{(0,\eta)} \leq \Phi_\eta \leq 1_{(0,2\eta)}$  and  $\|\Phi_\eta^{(k)}\|_\infty \leq C\eta^{-k}$ . We recall that  $\sigma_G$  is the Malliavin covariance matrix of  $G$  and we write

$$\begin{aligned} & \mathbb{P}(\det \sigma_G \leq \eta) \\ & \leq \mathbb{E}(\Phi_\eta(\det \sigma_G)) \leq \mathbb{E}(\Phi_\eta(\det \sigma_F)) + |\mathbb{E}(\Phi_\eta(\det \sigma_G)) - \mathbb{E}(\Phi_\eta(\det \sigma_F))| \\ & \leq \mathbb{P}(\sigma_F \leq 2\eta) + C\eta^{-p'} d_{p'}(\det \sigma_F, \det \sigma_G) \\ & \leq C(\mathcal{Q}_\rho(F)\eta^\rho + \eta^{-p'} d_{p,p'}) \end{aligned} \tag{3.29}$$

the last inequality being true for every  $\rho$ . Then, using the regularization lemma, we get for every  $\delta > 0$  and every  $q \in \mathbb{N}$

$$\begin{aligned} |\mathbb{E}(f(G)) - \mathbb{E}(f_\delta(G))| & \leq C \|f\|_\infty \left( \mathbb{P}(\det \sigma_G \leq \eta) + \frac{\delta^q}{\eta^{2q}} \mathcal{C}_{q,1}(G) \right) \\ & \leq C \|f\|_\infty (\mathcal{Q}_\rho(F) + \mathcal{C}_q(G)) \left( \eta^\rho + \eta^{-p'} d_{p,p'} + \frac{\delta^q}{\eta^{2q}} \right). \end{aligned}$$

We also have  $\mathbb{P}(\sigma_F \leq \eta) \leq \mathcal{Q}_\rho(F)\eta^\rho$  for every  $\rho$  so that the regularization lemma for  $F$  gives

$$|\mathbb{E}(f(F)) - \mathbb{E}(f_\delta(F))| \leq C \|f\|_\infty \mathcal{Q}_\rho(F) \left( \eta^\rho + \frac{\delta^q}{\eta^{2q}} \right).$$

On the other hand

$$|\mathbb{E}(f_\delta(F)) - \mathbb{E}(f_\delta(G))| \leq C \|f\|_\infty \delta^{-p} d_p(F, G).$$

Putting these together

$$|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| \leq C \|f\|_\infty \left( 1 + \mathcal{Q}_\rho(F) + \mathcal{C}_q(G) \right) \left( \delta^{-p} d_{p,p'} + \eta^\rho + \eta^{-p'} d_{p,p'} + \frac{\delta^q}{\eta^{2q}} \right).$$

We fix now  $\varepsilon > 0$  and we choose  $\delta$  and  $\eta$ . First take  $\delta = d_{p,p'}^{2\varepsilon}$ , so that

$$\delta^{-p} d_{p,p'} = d_{p,p'}^{1-2p\varepsilon}.$$

Take  $\eta = d_{p,p'}^{\varepsilon/2}$ , so that  $d_{p,p'}^\varepsilon \times \eta^{-2} = 1$  and consequently

$$\frac{\delta^q}{\eta^{2q}} = \frac{d_{p,p'}^{2q\varepsilon}}{\eta^{2q}} = d_{p,p'}^{q\varepsilon} \times \frac{d_{p,p'}^{q\varepsilon}}{\eta^{2q}} = d_{p,p'}^{q\varepsilon}.$$

We also have

$$\eta^{-p'} d_{p,p'} = d_{p,p'}^{1-p'\varepsilon/2} \quad \text{and} \quad \eta^\rho = d_{p,p'}^{\rho\varepsilon/2}.$$

Choose  $q = 1/\varepsilon$  and  $\rho = 2/\varepsilon$  in order to obtain

$$\eta^\rho + \frac{\delta^q}{\eta^{2q}} \leq 2d_{p,p'}.$$

Then

$$|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| \leq C \|f\|_\infty (1 + \mathcal{Q}_{\lfloor 2/\varepsilon \rfloor + 1}(F) + \mathcal{C}_{\lfloor 1/\varepsilon \rfloor + 1, 1}(G)) (d_{p,p'}^{1-2p\varepsilon} + d_{p,p'}^{1-p'\varepsilon/2} + d_{p,p'}).$$

Take now  $\bar{\varepsilon} = \max\{2p\varepsilon, p'\varepsilon/2\}$  and  $q(\bar{\varepsilon}) = \max\{\lfloor 4p/\bar{\varepsilon} \rfloor + 1, \lfloor p'/2\bar{\varepsilon} \rfloor + 1\}$ . The above estimate reads

$$|\mathbb{E}(f(F)) - \mathbb{E}(f(G))| \leq C \|f\|_\infty (1 + \mathcal{Q}_{q(\bar{\varepsilon})}(F) + \mathcal{C}_{q(\bar{\varepsilon}), 1}(G)) \times d_{p,p'}^{1-\bar{\varepsilon}}$$

so (3.30) is proved. In order to prove (3.28), we proceed as before but we use (3.19) instead of (3.17).  $\square$

In inequality (3.29) one can replace  $\det \sigma_F$  with any other random variable  $H \geq 0$  such that  $\mathbb{P}(H \leq \eta) \leq c_\kappa \eta^\kappa$  for every  $\kappa$  and  $\eta > 0$ , that is,  $H^{-1} \in \cap_p L^p$ . So, Proposition 3.11 can be reformulated as follows:

**Proposition 3.12.** *Let  $F, G \in \mathcal{E}$  be such that  $\mathcal{Q}_q(F) + \mathcal{C}_{q,1}(G) < \infty$  for every  $q \in \mathbb{N}$  (see (3.12)). Let  $H \geq 0$  be a r.v. such that  $H^{-1} \in \cap_p L^p$ . Then, for every  $p, p' \in \mathbb{N}$  and every  $\varepsilon > 0$  there exists a constant  $C$  (depending on  $\varepsilon, p, p'$ ) such that*

$$d_{TV}(F, G) \leq C \times C_\varepsilon(F, G, H) \times (d_p(F, G) + d_{p'}(\det \sigma_F, H))^{1-\varepsilon}, \quad (3.30)$$

with  $C_\varepsilon(F, G, H) = C_\varepsilon(F, G) + \|H^{-1}\|_{2/\varepsilon}$  with  $C_\varepsilon(F, G)$  given in (3.27). Moreover,

$$d_{TV}(F, G) \leq C \times C_\varepsilon(F, G, H) \times (d_{CF}(F, G) + d_{p'}(\det \sigma_F, H))^{1-\varepsilon}. \quad (3.31)$$

**Remark 3.13.** Proposition 3.11 essentially says that if  $(F_n, \det \sigma_{F_n}) \rightarrow (F, \det \sigma_F)$  in some “smooth distance”  $d_p$  (for example in the Wasserstein distance  $d_W$ ) then  $F_n \rightarrow F$  in total variation distance. And Proposition 3.12 says that it is not necessary that  $\sigma_{F_n}$  converges to  $\sigma_F$ : one can also have  $(F_n, \det \sigma_{F_n}) \rightarrow (F, H)$ . In any case one obtains the estimate of the speed of convergence: one loses a little bit because we have the power  $1 - \varepsilon$ . The striking fact is the we do not need the non degeneracy condition for  $F_n$  but only for  $F$ .

**Remark 3.14.** For the use of Proposition 3.11, in concrete applications it may be difficult to compute  $d_{p'}(\det \sigma_F, \det \sigma_G)$ . Then we are obliged to come back to “strong distances”: we have  $d_1(\det \sigma_F, \det \sigma_G) \leq (\|DF\|^{d-1} + \|DG\|^{d-1}) \|DF - DG\|$  so we take  $p' = 1$  and we get

$$d_{TV}(F, G) \leq C \times C_\varepsilon(F, G) \times (d_p(F, G) + \|DF - DG\|)^{1-\varepsilon} \quad (3.32)$$

Notice however that here we lose the right order of convergence. For example, in the case of the convergence of the Euler scheme  $X_t^n$  to the diffusion process  $X_t$  we have  $d_1(X_t^n, X_t) \leq \frac{C}{n}$  but  $\|DX_t^n - DX_t\| \sim \frac{C}{n^{1/2}}$ . This is because we deal with the weak convergence in the first case and with the strong convergence in the second one. So we pass from  $\frac{1}{n}$  to  $\frac{1}{n^{1/2}}$ . If we have an ellipticity property, then we do not need to estimate  $\|DX_t^n - DX_t\|$  so we are at level  $\frac{1}{n}$ .

### 3.2 The distance between density functions

We give here an immediate application of the regularization lemma under the strong non degeneracy condition (Lemma 3.5): we estimate the distance between density functions.

**Proposition 3.15.** *Let  $F, G \in \mathcal{E}^d$  be such that  $\mathcal{Q}_q(F) + \mathcal{Q}_q(G) < \infty$  for every  $q \in \mathbb{N}$  (see (3.12)). Then for every  $k \in \mathbb{N}$ , every multi index  $\alpha = (\alpha_1, \dots, \alpha_m)$  and every  $\varepsilon > 0$  there exists some constants  $C$  and  $q$  (depending on  $\varepsilon, k$  and on  $m$ ) such that*

$$|\mathbb{E}(\partial^\alpha f(F)) - \mathbb{E}(\partial^\alpha f(G))| \leq C d_k^{1-\varepsilon}(F, G) \|f\|_\infty (\mathcal{Q}_q(F) + \mathcal{Q}_q(G))^\varepsilon \tag{3.33}$$

In particular if  $p_F$  and  $p_G$  are the density functions for  $F$  and  $G$  respectively, then for every  $x \in \mathbb{R}^d$

$$|\partial^\alpha p_F(x) - \partial^\alpha p_G(x)| \leq C d_k^{1-\varepsilon}(F, G) (\mathcal{Q}_q(F) + \mathcal{Q}_q(G))^\varepsilon \tag{3.34}$$

The same estimates hold with  $d_k^{1-\varepsilon}(F, G)$  replaced by  $d_{CF}^{1-\varepsilon}(F, G)$ .

**Remark 3.16.** Compare with the estimate (2.53) pg 14/33 in [3]: here the estimate is much better because, using  $d_1$  for example, we have just  $d_1^{1-\varepsilon}(F, G) \leq \mathbb{E}(|F - G|)^{1-\varepsilon}$  and the Sobolev norms of  $F - G$  are not involved (as it is the case in [3]).

*Proof.* We will prove just (3.33) because (3.34) follows by standard regularization methods. To begin we use (3.13) and we get

$$|\mathbb{E}(\partial^\alpha f(F)) - \mathbb{E}(\partial^\alpha (f * \phi_\delta)(F))| \leq C \delta^q \|f\|_\infty \mathcal{Q}_{q+m}(F)$$

and a similar estimate for  $G$ . Moreover, using (3.17)

$$|\mathbb{E}(\partial^\alpha (f * \phi_\delta)(F)) - \mathbb{E}(\partial^\alpha (f * \phi_\delta)(G))| \leq C \delta^{-(k+m)} d_k(F, G) \times \|f\|_\infty$$

so that

$$|\mathbb{E}(\partial^\alpha f(F)) - \mathbb{E}(\partial^\alpha f(G))| \leq C (\delta^{-(k+m)} d_k(F, G) + \delta^q (\mathcal{Q}_{q+m}(F) + \mathcal{Q}_{q+m}(G))) \|f\|_\infty.$$

We choose  $\delta$  such that  $\delta^{-(k+m)} d_k(F, G) = \delta^q (\mathcal{Q}_{q+m}(F) + \mathcal{Q}_{q+m}(G))$ , that is,  $\delta = (d_k(F, G) / (\mathcal{Q}_{q+m}(F) + \mathcal{Q}_{q+m}(G)))^{\frac{1}{k+m+q}}$ , so that

$$|\mathbb{E}(\partial^\alpha f(F)) - \mathbb{E}(\partial^\alpha f(G))| \leq C d_k^{\frac{q}{k+m+q}}(F, G) (\mathcal{Q}_{q+m}(F) + \mathcal{Q}_{q+m}(G))^{\frac{k+m}{k+m+q}} \|f\|_\infty.$$

Then we choose  $q$  large in order that  $\frac{k+m}{k+m+q} < \varepsilon$ , and we get (3.33). □

## 4 Examples

### 4.1 Euler scheme

In this section we discuss the convergence in total variation distance of the Euler scheme.

We consider the  $d$  dimensional diffusion process

$$X_t = x + \sum_{j=1}^m \int_0^t \sigma_j(X_s) dB_s^j + \int_0^t b(X_s) ds$$

where  $\sigma_j \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ ,  $j = 1, \dots, m$  and  $b \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ . We are concerned with the Euler scheme defined in the following way. We fix  $n \in \mathbb{N}$ , we define  $\tau_n(t) = \frac{k}{n}$  for  $\frac{k}{n} \leq t < \frac{k+1}{n}$  and then the Euler scheme is given by

$$X_t^n = x + \sum_{j=1}^m \int_0^t \sigma_j(X_{\tau_n(s)}^n) dB_s^j + \int_0^t b(X_{\tau_n(s)}^n) ds.$$

In this framework there are two types of errors which are of interest: first, the “strong error” is  $\|X_t - X_t^n\|_p \leq \frac{C}{\sqrt{n}}$ . And moreover, the weak error: for  $f \in C_b^\beta(\mathbb{R}^d)$ , Talay and Tubaro proved in [28] that

$$|\mathbb{E}(f(X_t)) - \mathbb{E}(f(X_t^n))| \leq C \|f\|_{6,\infty} \times \frac{1}{n}. \tag{4.1}$$

Then one is interested to obtain the above estimate for measurable and bounded functions, so to replace  $\|f\|_{6,\infty}$  by  $\|f\|_\infty$  in the above estimate (this means to obtain the estimate of the error in total variation distance). This has been done by Bally and Talay in [9] and by Guyon in [13] by using Malliavin calculus (another approach, based on the parametrix method has been given by Konackov and Memen [15], in the elliptic case). In order to do this one has to assume a non degeneracy hypothesis. We construct the Lie algebra associated to the coefficients of the above *SDE*:

$$\begin{aligned} \mathcal{A}_0 &= \{\sigma_1, \dots, \sigma_m\}, \\ \mathcal{A}_k &= \{[\sigma_1, \psi], \dots, [\sigma_m, \psi], [b, \psi] : \psi \in \mathcal{A}_{k-1}\} \end{aligned}$$

where  $[\phi, \psi] = \langle \phi, \nabla \psi \rangle - \langle \psi, \nabla \phi \rangle$  is the Lie bracket. We also denote  $\mathcal{A}_k(x) = \{\phi(x), \phi \in \mathcal{A}_k\}$ . Then we have two types of non degeneracy conditions: the ellipticity condition in  $x$  means that  $\sigma_1(x), \dots, \sigma_m(x)$  span  $\mathbb{R}^d$ . This is also equivalent with the fact that  $\sigma\sigma^*(x)$  is invertible. The second condition, much less strong, is that  $\cup_{k \in \mathbb{N}} \mathcal{A}_k(x)$  span  $\mathbb{R}^d$ . This is the so called Hörmander’s condition. And Hörmander’s theorem (proved by Malliavin by a probabilistic approach) say that under this condition the law of  $X_t(x)$  is absolutely continuous and has a smooth density.

Let us come back to the estimate of the weak error in total variation distance. J. Guyon proved in [13] that if the uniform ellipticity condition holds that is  $\sigma\sigma^*(x) \geq \lambda > 0$  for every  $x$ , then

$$|\mathbb{E}(f(X_t(x))) - \mathbb{E}(f(X_t^n(x)))| \leq C \|f\|_\infty \times \frac{1}{n}. \tag{4.2}$$

The estimate of the weak error in total variation distance, under the Hörmander condition, has been done in [9] under the “uniform Hörmander condition”: there is a  $k \in \mathbb{N}$  and some  $\lambda > 0$  such that

$$\Lambda(x) := \inf_{|\xi|=1} \sum_{\psi \in \mathcal{A}_k} \langle \psi(x), \xi \rangle^2 \geq \lambda.$$

But here a supplementary difficulty appears: the Hörmander assumption is not sufficient in order to guarantee that the Malliavin covariance matrix  $\sigma_{X_t^n(x)}$  of the Euler scheme  $X_t^n(x)$  is invertible (and this was a crucial ingredient in the proof). In [9] this difficulty has been bypassed by replacing  $X_t^n(x)$  by the “regularized version”  $\bar{X}_t^n(x) = X_t^n(x) + \varepsilon_n \Delta$  where  $\Delta$  is a standard normal random variable independent of the Brownian motion  $B$ . From a simulation point of view this is really not a problem because this just means to simulate one more random variable  $\Delta$ . And one proves that

$$\left| \mathbb{E}(f(X_t(x))) - \mathbb{E}(f(\bar{X}_t^n(x))) \right| \leq C \|f\|_\infty \times \frac{1}{n}. \tag{4.3}$$

Although from a practical point of view this has not big interest, the following theoretical question remained open: is it possible to prove (4.3) for the real Euler scheme  $X_t^n(x)$  (without the regularization factor  $\varepsilon_n \Delta$ ) under the Hörmander condition?

Let us see what we may obtain using the results from the previous sections. We use the standard Malliavin calculus so now  $\mathcal{E} = \mathbb{D}^\infty$  (see the notation in Nualart [22]). And under our assumptions on the coefficients ( $\sigma_j, b \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ ) standard estimates yield, for every  $q \in \mathbb{N}$

$$\mathcal{C}_q(X_t(x)) + \sup_n \mathcal{C}_q(X_t^n(x)) = \mathcal{C}_q < \infty.$$

Moreover, if the ellipticity condition holds, one proves that  $\det \sigma_{X_t(x)} \geq \lambda(x) > 0$  and  $\det \sigma_{X_t^n(x)} \geq \lambda(x) > 0$  with  $\lambda(x)$  independent of  $n$ . It follows that

$$\mathcal{Q}_q(X_t(x)) + \sup_n \mathcal{Q}_q(X_t^n(x)) = Q_q < \infty.$$

So, using (3.24) first and (4.1) then, we obtain for every  $\varepsilon > 0$

$$d_{TV}(X_t, X_t^n) \leq C \times d_6^{1-\varepsilon}(X_t, X_t^n) \leq \frac{C}{n^{1-\varepsilon}}. \tag{4.4}$$

Comparing with the result of Guyon (4.2) we see that we have lost a little bit because we have the power  $1 - \varepsilon$  instead of 1. This is a structural drawback of our method which is based on optimization. However there is a slight gain because we need just ellipticity in the starting point  $x$  and not uniform ellipticity.

Let us see now what we are able to say under Hörmander’s condition. We stress that we do not need the uniform Hörmander condition but only the condition in the starting point  $x$ : we just assume that  $Span\{\cup_{k \in \mathbb{N}} \mathcal{A}_k(x)\} = \mathbb{R}^d$ . This is sufficient in order to guarantee that  $\det \sigma_{X_t(x)} > 0$  and this is all we need. As we mentioned above, we are no more able to prove that  $\det \sigma_{X_t^n(x)} \geq \lambda(x) > 0$  so we have to use (3.26) (together with (4.1)):

$$\begin{aligned} d_{TV}(X_t, X_t^n) &\leq C \times (d_6(X_t, X_t^n) + d_p(\det \sigma_{X_t}, \det \sigma_{X_t^n}))^{1-\varepsilon} \\ &\leq C \times \left( \frac{1}{n} + d_p(\det \sigma_{X_t}, \det \sigma_{X_t^n}) \right)^{1-\varepsilon}. \end{aligned}$$

Now we have to estimate  $d_p(\det \sigma_{X_t}, \det \sigma_{X_t^n})$ . If we were able to prove that for some  $p \in \mathbb{N}$  one would have  $d_p(\det \sigma_{X_t}, \det \sigma_{X_t^n}) \leq \frac{C}{n}$ , then we come back to the same estimate as in the elliptic case. At a first glance this seems reasonable, but taking things seriously this is not so clear – we give up to answer this question here, and we just notice that easy standard arguments give  $\|\det \sigma_{X_t} - \det \sigma_{X_t^n}\|_1 \leq \frac{C}{\sqrt{n}}$  which yields  $d_1(\det \sigma_{X_t}, \det \sigma_{X_t^n}) \leq \frac{C}{n^{1/2}}$ . Finally we obtain:

$$d_{TV}(X_t, X_t^n) \leq \frac{C}{n^{\frac{1}{2}-\varepsilon}}. \tag{4.5}$$

We conclude that we are still able to prove that  $\lim_n d_{TV}(X_t(x), X_t^n(x)) = 0$  but we lose much on the speed of convergence.

**Remark 4.1.** We guess that one could obtain the rate  $(\ln n)^\alpha/n$  instead of  $1/n^{1-\varepsilon}$ . But this would require hard work and we are not able to do it at this time. However we believe that the method presented in our paper always loses a little bit and never gives  $1/n$ .

#### 4.2 Central Limit Theorem for Wiener chaoses

Let us fix  $1 \leq q_1 \leq q_2 \leq \dots \leq q_d$  a sequence of  $d$  positive integers. Let us consider here  $F_n = (F_{1,n}, \dots, F_{d,n})$  a sequence of random vectors such that for all  $i \in \{1, \dots, d\}$  and all  $n \geq 1$ , the random variable  $F_{i,n}$  belongs to the  $q_i$ -th Wiener chaos. We will further assume that the covariance matrix of  $F_n$  is the identity matrix for every  $n \geq 1$ . A central result of Nourdin-Peccati theory established in [19] provides an explicit bound in total variation between the distribution  $F_n$  and the distribution of a standard Gaussian vector, say  $N = (N_1, \dots, N_d)$ :

$$d_{TV}(F, N) \leq C \Phi \left( \mathbb{E}(|F_n|^4) - \mathbb{E}(|N|^4) \right), \quad \Phi(x) = |\log(x)|\sqrt{x}. \tag{4.6}$$

Let us mention that an entropic result is actually proved in [19] and the previous bound is the corresponding total variation estimate which is derived from Pinsker inequality.

The proof of (4.6) uses clever arguments from information theory which are nevertheless rather specific to Gaussian targets. Our goal here is to apply Proposition 3.12 to this situation and to compare the bounds. First, from [18] one has the following result regarding the Wasserstein distance:

$$d_W(F, N) \leq C \left[ \sum_{i,j=1}^d \mathbb{E} \left( \left( \delta_{i,j} - \frac{1}{q_i} \Gamma[F_{i,n}, F_{j,n}] \right)^2 \right) \right]^{\frac{1}{2}}, \quad (4.7)$$

while from [21] one gets the bound

$$\sum_{i,j=1}^d \mathbb{E} \left( \left( \delta_{i,j} - \frac{1}{q_i} \Gamma[F_{i,n}, F_{j,n}] \right)^2 \right) \leq \mathbb{E}(|F_n|^4) - \mathbb{E}(|N|^4). \quad (4.8)$$

Finally it is obvious that for some constant  $C$  only depending on  $d$  we get

$$d_W(\det \sigma_{F_n}, \det \mathbf{D}) \leq C \left[ \sum_{i,j=1}^d \mathbb{E} \left( \left( \delta_{i,j} - \frac{1}{q_i} \Gamma[F_{i,n}, F_{j,n}] \right)^2 \right) \right]^{\frac{1}{2}}, \quad (4.9)$$

with  $\mathbf{D} = \text{Diag}(q_1, q_2, \dots, q_d)$ . Since  $\mathbf{D}$  is a deterministic invertible matrix and  $F_n$  is uniformly bounded in  $\mathbb{D}^\infty$ , one can apply Proposition 3.12 with  $H = \det \mathbf{D}$ . For every  $\epsilon > 0$ , gathering the bounds (4.7), then inequalities (4.8) and (4.9) lead to

$$d_{TV}(F, N) \leq C_\epsilon \left( \mathbb{E}(|F_n|^4) - \mathbb{E}(|N|^4) \right)^{\frac{1}{2} - \epsilon} \quad (4.10)$$

which, up to an arbitrarily small loss, retrieves the correct order of magnitude in the total variation estimate of the so-called fourth moment Theorem.

## References

- [1] Arras, B. and Houdré, C.: *Applications to Sums of Independent Random Variables*. Springer, 2019.
- [2] Bakry, D., Gentil, I. and Ledoux, M.: *Analysis and geometry of Markov diffusion operators. Grundlehren der Mathematischen Wissenschaften* **348**, Springer, 2014. MR-3155209
- [3] Bally, V. and Caramellino, L.: On the distances between probability density functions. *Electron. J. Probab.* **19**, no. 110, (2014), 1–33. MR-3296526
- [4] Bally, V. and Caramellino, L.: Asymptotic development for the CLT in total variation distance. *Bernoulli* **22**, (2016), 2442–2485. MR-3498034
- [5] Bally, V., Caramellino, L. and Poly, G.: Convergence in distribution norms in the CLT for non identical distributed random variables. *Electron. J. Probab.* **23**, no. 45, (2018), 1–51. MR-3814239
- [6] Bally, V., Caramellino, L. and Poly, G.: Non universality for the variance of the number of real roots of random trigonometric polynomials. *Probab. Th. Rel. Fields* **174**, (2019), 887–927. MR-3980307
- [7] Bally, V. and Clément E.: Integration by parts formula and applications to equations with jumps. *Probab. Th. Rel. Fields* **151**, (2011), 613–657. MR-2851695
- [8] Bally, V. and Clément, E.: Integration by parts formula with respect to jump times for stochastic differential equations. *Stochastic Analysis 2010*, 7–29, Springer, Heidelberg, 2011. MR-2789077
- [9] Bally, V. and Talay, D.: The law of the Euler scheme for stochastic differential equations. *Probab. Th. Rel. Fields* **104**, (1996), 43–60. MR-1367666
- [10] Bichtler, K., Gravereaux, J.-B. and Jacod, J.: *Malliavin calculus for processes with jumps*. Stochastic Monographs **2**, Gordon and Breach Science Publishers, 1987. MR-1008471

- [11] Bogachev, V.I., Kosov, E.D. and Zelenov, G.I.: Fractional smoothness of distributions of polynomials and a fractional analog of the Hardy-Landau-Littlewood inequality. *Trans. Amer. Math. Soc.* **370**, (2018), 4401–4432. MR-3811533
- [12] Bouleau, N. and Denis, L.: *Dirichlet forms methods for Poisson point measures and Lévy processes. With emphasis on the creation-annihilation techniques*. Probability Theory and Stochastic Modelling **76**, Springer, 2015. MR-3444890
- [13] Guyon, J.: Euler scheme and tempered distributions. *Stochastic Process. Appl.* **116**, (2006), 877–904. MR-2254663
- [14] Hu, Y., Lu, F. and Nualart, D.: Convergence of densities of some functionals of Gaussian processes. *J. Funct. Anal.* **266**, (2014), 814–875. MR-3132731
- [15] Konakov, V. and Mammen, E.: Edgeworth type expansions for Euler schemes for stochastic differential equations. *Monte Carlo Methods Appl.* **8**, (2002), 271–285. MR-1931967
- [16] Mokobodzki, G.: *Sur l’algèbre contenue dans le domaine étendu d’un générateur infinitésimal*. Sémin. Théorie du Potentiel, Paris, No3, Lecture Notes in Math. **681**, Springer-Verlag, 1978. MR-0521785
- [17] Nourdin, I. and Peccati, G.: *Normal approximations with Malliavin calculus. From Stein’s method to universality*. Tracts in Mathematics **192**. Cambridge University Press, 2012. MR-2962301
- [18] Nourdin, I., Peccati, G. and Réveillac, A.: Multivariate normal approximation using Stein’s method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.* **46**, (2010), 45–58. MR-2641769
- [19] Nourdin, I., Peccati, G. and Swan, Y.: Entropy and the fourth moment phenomenon. *J. Funct. Anal.* **266**, (2014), 3170–3207. MR-3158721
- [20] Nourdin, I. and Poly, G.: Convergence in total variation on Wiener chaos. *Stochastic Process. Appl.* **123**, (2013), 651–674. MR-3003367
- [21] Nourdin, I. and Rosiński, J.: Asymptotic independence of multiple Wiener-Itô integrals and the resulting limit laws. *Ann. Probab.* **42**, (2014), 497–526. MR-3178465
- [22] Nualart, D.: *The Malliavin calculus and related topics. Second Edition*. Springer-Verlag, 2006. MR-2200233
- [23] Nualart, D. and Vives, J.: Anticipative calculus for the Poisson process based on the Fock space. Séminaire de Probabilités, XXIV, 1988/89, Lecture Notes in Math. **1426**, 154–165, Springer, 1990. MR-1071538
- [24] Nualart, D. and Vives, J.: A duality formula on the Poisson space and some applications. Seminar on Stochastic Analysis, Random Fields and Applications (Ascona 1993), Progr. Probab. **36**, 205–213, Birkhäuser, 1995. MR-1360277
- [25] Poly, G.: Regularization along central convergence on second and third Wiener chaoses. arXiv:1905.02784.
- [26] Privault, N.: Chaotic and variational calculus in discrete and continuous time for the Poisson process. *Stochastics Stochastics Rep.* **51**, (1994) 83–109. MR-1380764
- [27] Solé, J.L., Utzet, F. and Vives, J.: Canonical Lévy process and Malliavin calculus. *Stoch. Process. Appl.* **117**, (2007), 165–187. MR-2290191
- [28] Talay, D. and Tubaro, L.: Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8**, (1990), 483–509. MR-1091544
- [29] Vives, J.: Malliavin calculus for Lévy processes: a survey. *Rend. Sem. Mat. Univ. Politec. Torino* **71**, (2013), 261–272. MR-3345056

**Acknowledgments.** L.C. acknowledges the MIUR Excellence Department Project awarded to the Department of Mathematics, University of Rome Tor Vergata, CUP E83C18000100006, and the INdAM-GNAMPA Project “Stime asintotiche: principi di invarianza e grandi deviazioni”.

G.P. is supported by the ANR grant UNIRANDOM.