Reweighting samples under covariate shift using a Wasserstein distance criterion

Julien Reygner and Adrien Touboul

ABSTRACT. Considering two random variables with different laws to which we only have access through finite size i.i.d samples, we address how to reweight the first sample so that its empirical distribution converges towards the true law of the second sample as the size of both samples goes to infinity. We study an optimal reweighting that minimizes the Wasserstein distance between the empirical measures of the two samples, and leads to an expression of the weights in terms of Nearest Neighbors. The consistency and some asymptotic convergence rates in terms of expected Wasserstein distance are derived, and do not need the assumption of absolute continuity of one random variable with respect to the other. These results have some application in Uncertainty Quantification for decoupled estimation and in the bound of the generalization error for the Nearest Neighbor regression under covariate shift.

1. Introduction

1.1. **Regression under covariate shift.** This article is dedicated to the study of a method aimed at approximating the law of a random variable

$$Y = f(X, \Theta), \tag{1}$$

where $X \in \mathbb{R}^d$, $\Theta \in \mathbf{\Theta}$ are independent random variables, with respective laws denoted by μ_X and μ_Θ, and $f : \mathbb{R}^d \times \mathbf{\Theta} \to \mathbb{R}^e$ is a measurable function. The space $\mathbf{\Theta}$ is only assumed to be measurable. The specificity of the problem at stake is that we assume to be provided with:

- a *training* sample $(X'_j, Y'_j)_{j\in[\![1,m]\!]}$ of i.i.d observations $Y'_j = f(X'_j, \Theta_j)$ where Θ_j has law μ_Θ and is independent from X'_j, but the law $\mu_{X'}$ of X'_j may differ from μ_X;
- an *evaluation* sample $(X_i)_{i\in[\![1,n]\!]}$ with i.i.d observations distributed according to μ_X.

This situation is known as *covariate shift* in the statistical learning literature [8, 33].

This problem is motivated by the study of decomposition-based uncertainty quantification (UQ) methods in complex industrial systems, as is detailed in Subsection 5.2 below. In this context, the overall objective is to approximate a *quantity of interest* of the form

$$\mathrm{QI} = \mathbb{E}[\phi(Y)] \tag{2}$$

for some function $\phi : \mathbb{R}^e \to \mathbb{R}$. Following previous works in this direction [2, 3, 4], our estimator of QI assumes the form

$$\widehat{\mathrm{QI}}_{m,n} = \frac{1}{m}\sum_{j=1}^{m} w_j \phi(Y'_j), \tag{3}$$

Key words and phrases. Reweighting; Covariate shift; Wasserstein distance; Uncertainty Quantification; Nearest neighbor regression; Nearest neighbor distance.

This research work has been carried out under the leadership of the Technological Research Institute SystemX, and therefore granted with public funds within the scope of the French Program "Investissements d'Avenir".

where the vector of *weights* $\mathbf{w}_m = (w_1, \dots, w_m)$ is chosen so that the *weighted empirical measure*

$$\widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m} := \frac{1}{m} \sum_{j=1}^{m} w_j \delta_{X'_j}$$

of the training sample $\mathbf{X}'_m = (X'_1, \dots, X'_m)$ be close, in a sense which will be made precise below, to the empirical measure

$$\widehat{\mu}_{\mathbf{X}_n} := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

of the evaluation sample $\mathbf{X}_n = (X_1, \dots, X_n)$. Such a reweighting procedure is a standard approach to the problem of *density ratio estimation* in the statistical learning literature [35], the purpose of which is to estimate the density $\mathrm{d}\mu_X / \mathrm{d}\mu_{X'}$ from the samples $\mathbf{X}_n$ and $\mathbf{X'}_m$, without estimating separately the measures μ_X and $\mu_{X'}$. While the theoretical analysis of such methods almost always requires this density to exist, in the UQ context which motivates the present study it is desirable not to assume that any of the measures μ_X and $\mu_{X'}$ be absolutely continuous with respect to the other, see in particular Remark 5.3.

The first step of our work is thus the computation of *optimal* weights $\mathbf{w}_m$ for the problem

$$\min_{\mathbf{w}_m} W_q \left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m} \right), \tag{4}$$

$$\forall j \in [\![1, m]\!], \quad w_j \geq 0, \qquad \text{and} \quad \sum_{j=1}^{m} w_j = m, \tag{5}$$

where W_q denotes the *Wasserstein distance* of order $q \geq 1$ on $\mathbb{R}^d$. The reason for the choice of this distance is that unlike criteria already studied in the density ratio estimation literature, such as moment/kernel matching, L^2 distance, Kullback–Leibler divergence (see [35] and the references therein), it is not sensitive to absolute continuity conditions and therefore it is well suited to our UQ motivation. On the other hand, this choice makes the problem closely related to the fields of *optimal quantization* [23] and *Nearest Neighbor* (NN) estimation [5].

More precisely, for any $k \in [\![1, m]\!]$, denote by $\widehat{\psi}_m^{(k)}(x)$ the k-NN estimator of the *regression function*

$$\psi(x) := \mathbb{E}\left[\phi(Y) | X = x\right] = \mathbb{E}\left[\phi(f(x, \Theta))\right], \tag{6}$$

defined from the observation of the training sample, and then consider the Monte Carlo estimator

$$\widehat{\mathrm{QI}}_{m,n}^{(k)} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\psi}_m^{(k)}(X_i)$$

of QI. Then the vector of weights $\mathbf{w}_m$ which are optimal for (4)–(5) turns out not to depend on the value of q, and the associated estimator $\widehat{\mathrm{QI}}_{m,n}$ defined by (3) coincides with the 1-NN estimator $\widehat{\mathrm{QI}}_{m,n}^{(1)}$. For this reason, we shall denote by $\mathbf{w}_m^{(1)}$ the vector of optimal weights for (4)–(5), and more generally by $\mathbf{w}_m^{(k)}$ the vector of weights induced by the k-NN estimator of ψ.

The main results of this paper describe the asymptotic behavior, as the respective sizes m and n of the training and evaluation samples grow to infinity, of both the Wasserstein distance $W_q(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m^{(k)}})$ and the estimator $\widehat{\mathrm{QI}}_{m,n}^{(k)}$ of QI. While taking $k = 1$ is optimal for the convergence of $\widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m^{(k)}}$ to $\widehat{\mu}_{\mathbf{X}_n}$, one may expect from the theory of NN regression that the estimator $\widehat{\mathrm{QI}}_{m,n}^{(k)}$ display better convergence properties if k is chosen to grow to infinity with m. Therefore we shall study both regimes $k = 1$ and $k = k_m \to +\infty$.

1.2. **Outline of the article.** The derivation of the Wasserstein optimal weights $\mathbf{w}_m$ is detailed in Section 2, where we also highlight connections between our results and various topics in numerical probability and statistical learning. The asymptotic behavior of $W_q(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m^{(k)}})$ and $\widehat{\mathrm{QI}}_{m,n}^{(k)}$ are respectively studied in Sections 3 and 4. Applications to decomposition-based UQ and the generalization error for NN regression under covariate shift, as well as numerical illustrations, are presented in Section 5.

1.3. **Notation.** We denote by $\mathbb{N}$ the set of the natural integers including zero and by $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ the set of positive integers. Given two integers $n_1 \leq n_2$, the set of the integers between n_1 and n_2 is written $[\![n_1, n_2]\!] = \{n_1, \dots, n_2\}$. For $x \in \mathbb{R}$, $\lceil x \rceil$ (resp. $\lfloor x \rfloor$) is the unique integer verifying $x \leq \lceil x \rceil < x + 1$ (resp. $x - 1 < \lfloor x \rfloor \leq x$). For $(x, y) \in \mathbb{R}^2$, we use the join and meet notation $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$. Last, we denote by $(x)_+ := 0 \vee x$ and $(x)_- := 0 \vee (-x)$ the nonnegative and nonpositive parts of $x \in \mathbb{R}$.

We fix a norm $|\cdot|$ on $\mathbb{R}^d$, which need not be the Euclidean norm. The supremum norm of $\phi : \mathbb{R}^d \to \mathbb{R}$ is denoted by $\|\phi\|_\infty = \sup_{x \in \mathbb{R}^d} |\phi(x)|$. The distance between a point $x \in \mathbb{R}^d$ and a subset $A \subset \mathbb{R}^d$ is denoted by $\mathrm{dist}(x, A)$. Last, for all $x \in \mathbb{R}^d$ and $r \geq 0$, we denote $B(x, r) := \{x' \in \mathbb{R}^d : |x - x'| \leq r\}$, and recall that the *support* of a probability measure $\nu \in \mathcal{P}(\mathbb{R}^d)$ is defined by

$$\mathrm{supp}(\nu) := \left\{ x \in \mathbb{R}^d : \forall r > 0, \nu(B(x, r)) > 0 \right\}.$$

2. Wasserstein distance minimization and NN regression

2.1. **Optimal weights for Wasserstein distances.** We begin by recalling the definition of the Wasserstein distance.

Definition 2.1 (Wasserstein distance)**.** *Let $\mathcal{P}(\mathbb{R}^d)$ be the set of probability measures on $\mathbb{R}^d$ and, for any $q \in [1, +\infty)$, let*

$$\mathcal{P}_q(\mathbb{R}^d) = \left\{ \nu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^q \mathrm{d}\nu(x) < +\infty \right\}.$$

The Wasserstein distance of order q between μ and $\nu \in \mathcal{P}_q(\mathbb{R}^d)$ is defined as

$$W_q(\mu, \nu) = \inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - x'|^q \mathrm{d}\gamma(x, x') : \gamma \in \Pi(\mu, \nu) \right\}^{1/q},$$

where $\Pi(\mu, \nu)$ is the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν.

We refer to [38, Section 6] for a general introduction to Wasserstein distances.

This definition allows for an explicit resolution of the minimization problem (4)–(5), which relies on the notion of *Nearest Neighbor* (NN). For $x \in \mathbb{R}^d$ and $k \in [\![1, m]\!]$, we denote by $\mathrm{NN}_{\mathbf{X}'_m}^{(k)}(x)$ the *k-th Nearest Neighbor* (k-NN) of x among the sample $\mathbf{X}'_m$, that is to say the k-th closest point to x among $X'_1, \dots, X'_m$ for the norm $|\cdot|$. If there are several such points, we define $\mathrm{NN}_{\mathbf{X}'_m}^{(k)}(x)$ to be the point X'_j with lowest index j. We omit the superscript notation (k) when referring to the 1-NN, *i.e.*

$$\mathrm{NN}_{\mathbf{X}'_m}(x) = \mathrm{NN}_{\mathbf{X}'_m}^{(1)}(x).$$

In the next statement, for any $i \in [\![1, n]\!]$ and $l \in [\![1, m]\!]$, we denote by $j_i^{(l)}$ the (lowest) index j such that $X'_j = \mathrm{NN}_{\mathbf{X}'_m}^{(l)}(X_i)$.

Proposition 2.2 (Optimal vector of weights)**.** *Let the k-NN vector of weights $\mathbf{w}_m^{(k)} = (w_1^{(k)}, \dots, w_m^{(k)})$ be defined by, for all $j, k \in [\![1, m]\!]$,*

$$w_j^{(k)} := \frac{m}{kn} \sum_{i=1}^{n} \sum_{l=1}^{k} \mathbb{1}_{\{j = j_i^{(l)}\}}. \tag{7}$$

The vector $\mathbf{w}_m^{(k)}$ satisfies (5) and verifies, for all $q \in [1, +\infty)$,

$$W_q^q \left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m^{(k)}} \right) \leq \frac{1}{kn} \sum_{i=1}^{n} \sum_{l=1}^{k} \left| X_i - \mathrm{NN}_{\mathbf{X}'_m}^{(l)}(X_i) \right|^q. \tag{8}$$

For $k = 1$, the equality is reached

$$W_q^q \left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m^{(1)}} \right) = \frac{1}{n} \sum_{i=1}^{n} \left| X_i - \mathrm{NN}_{\mathbf{X}'_m}(X_i) \right|^q, \tag{9}$$

and the vector is optimal for (4) in the sense that for any $\mathbf{w}_m = (w_1, \dots, w_m)$ which also satisfies (5), we have

$$W_q \left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m^{(1)}} \right) \leq W_q \left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m} \right). \tag{10}$$

In other words, for a given $j \in [\![1, m]\!]$, $w_j^{(k)}$ is proportional to the number of points X_i of which X'_j is one of the first k NN. We refer to [27] for a numerical illustration of the use of the vector of weights $\mathbf{w}_m^{(1)}$ in the context of classification under covariate shift.

Proof. For a general vector of weights $\mathbf{w}_m = (w_1, \dots, w_m)$ which satisfies (5), the Wasserstein distance $W_q^q(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m})$ is the solution of the following optimal transport problem

$$\begin{aligned}
&\inf_{(\gamma_{i,j})_{(i,j) \in [\![1,n]\!] \times [\![1,m]\!]}} \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{i,j} |X_i - X'_j|^q, \\
&\forall i \in [\![1, n]\!], \quad \sum_{j=1}^{m} \gamma_{i,j} = \frac{1}{n} \quad \text{(marginal condition on } \widehat{\mu}_{\mathbf{X}_n}), \\
&\forall j \in [\![1, m]\!], \quad \sum_{i=1}^{n} \gamma_{i,j} = \frac{w_j}{m} \quad \text{(marginal condition on } \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m}), \\
&\forall (i, j) \in [\![1, n]\!] \times [\![1, m]\!], \quad \gamma_{i,j} \geq 0,
\end{aligned} \tag{11}$$

where $\gamma_{i,j}$ is the coefficient of the discrete transport plan between δ_{X_i} and $\delta_{X'_j}$. For the k-NN vector of weights $\mathbf{w}_m^{(k)}$ defined by (7), the transport plan

$$\gamma_{i,j}^{(k)} = \frac{1}{kn} \sum_{l=1}^{k} \mathbb{1}_{\{j = j_i^{(l)}\}}$$

satisfies the two marginal conditions. Reordering the terms in the associated cost gives the upper bound of Equation (8).

We now prove the equality (9) and optimality (10) of $\mathbf{w}_m^{(1)}$ at the same time. On the one hand, it is clear that for any vector of weights $\mathbf{w}_m = (w_1, \dots, w_m)$ and any transport plan $(\gamma_{i,j})_{(i,j) \in [\![1,n]\!] \times [\![1,m]\!]}$

between $\widehat{\mu}_{\mathbf{X}_n}$ and $\widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_n}$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{i,j}|X_i - X'_j|^q \geq \sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{i,j}|X_i - \mathrm{NN}_{\mathbf{X}'_m}(X_i)|^q$$
$$= \frac{1}{n}\sum_{i=1}^{n}|X_i - \mathrm{NN}_{\mathbf{X}'_m}(X_i)|^q,$$

therefore taking the infimum over all transport plans yields

$$W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m}\right) \geq \frac{1}{n}\sum_{i=1}^{n}|X_i - \mathrm{NN}_{\mathbf{X}'_m}(X_i)|^q.$$

On the other hand, taking $\mathbf{w}_m = \mathbf{w}_m^{(1)}$ in the left-hand side and combining this inequality with (8) for $k = 1$, we obtain both the equality (9) and optimality (10). □

Remark 2.3. *In order to alleviate notation, we now write* $\widehat{\mu}_{\mathbf{X}'_m}^{(k)} = \widehat{\mu}_{\mathbf{X}'_m}^{\mathbf{w}_m^{(k)}}$.

2.2. **Comments on Proposition 2.2.** In this subsection, we discuss the relation between the result of Proposition 2.2 and other fields in numerical probability and statistical learning, as well a generalization of this result to a more general framework.

2.2.1. *Link with optimal quantization.* It is clear from Proposition 2.2 that $\widehat{\mu}_{\mathbf{X}'_m}^{(1)}$ is the pushforward of $\widehat{\mu}_{\mathbf{X}_n}$ by $\mathrm{NN}_{\mathbf{X}'_m}$, and that this transport map yields an optimal coupling between $\widehat{\mu}_{\mathbf{X}_n}$ and $\widehat{\mu}_{\mathbf{X}'_m}^{(1)}$ in Definition 2.1, for any $q \geq 1$. The idea to associate each X_i with $\mathrm{NN}_{\mathbf{X}'_m}(X_i) = X'_{j_i^{(1)}}$ is the basis of the theory of *optimal quantization* [23, 24, 30]. In this context, the sample $\mathbf{X}'_m$ plays the role of the *quantization grid*, and $\mathrm{NN}_{\mathbf{X}'_m}$ is known to be the *optimal quantization function*. The right-hand side of (9) then corresponds to the L^q *mean quantization error* induced by the grid $\mathbf{X}'_m$ for the measure $\widehat{\mu}_{\mathbf{X}_n}$.

2.2.2. *Link with geometric inference.* When $q = 2$, the right-hand side of (8) rewrites

$$\int_{x\in\mathbb{R}^d} \mathrm{d}^2_{\widehat{\mu}_{\mathbf{X}'_m},k/m}(x)\widehat{\mu}_{\mathbf{X}_n}(\mathrm{d}x),$$

where $\mathrm{d}_{\mu,\alpha}(\cdot)$ is the *distance function to* μ *with parameter* α introduced by Chazal, Cohen-Steiner and Mérigot in [11, Definition 3.2] in order to perform geometric and topological inference for set estimation, see also [12, 9] for robust inference. In particular, [11, Proposition 3.3] shows that for any $x \in \mathbb{R}^d$,

$$\mathrm{d}^2_{\widehat{\mu}_{\mathbf{X}'_m},k/m}(x) = \inf_{\mathbf{w}_m} W_2^2\left(\frac{1}{m}\sum_{j=1}^{m} w_j\delta_{X'_j}, \delta_x\right), \tag{12}$$
$$\mathbf{w}_m = (w_1,\ldots,w_m) \text{ satisfies (5) and } w_j \leq m/k \text{ for all } j.$$

This result may be directly compared with the estimates (8) and (9), at least in the case where $\mu_X = \delta_x$. In this case, if $k = 1$, then the supplementary constraint $w_j \leq m/k$ is necessarily implied by (5) and therefore, combining (12) with (9), we recover the optimality result (10). For arbitrary values of k, the combination of (12) with (8) shows that the vector $\mathbf{w}_m^{(k)}$ need not be optimal for (10), but yields a solution which is lower than any solution with the supplementary constraint that $w_j \leq m/k$.

2.2.3. *A more general problem.* Proposition 2.2 may appear as a specific instance, restricted to empirical measures, of the following problem: given two probability measures μ and ν on $\mathbb{R}^d$, and assuming that $\mu \in \mathcal{P}_q(\mathbb{R}^d)$, compute the infimum of $W_q(\mu, \rho \mathrm{d}\nu)$ over all probability densities ρ with respect to ν. Similar questions were recently addressed in [10]. First, it is clear that if μ is absolutely continuous with respect to ν, then taking $\rho = \mathrm{d}\mu/\mathrm{d}\nu$ shows that this minimum is 0. Next, following the proof of Proposition 2.2, it is easily seen that for any ρ,

$$W_q^q(\mu, \rho \mathrm{d}\nu) \geq \mathbb{E}\left[\mathrm{dist}(X, \mathrm{supp}(\nu))^q\right], \qquad X \sim \mu.$$

To show that the right-hand side actually matches with the infimum of the left-hand side when ρ varies, we keep following the proof of Proposition 2.2. Since $\mathrm{supp}(\nu)$ is closed, for any $x \in \mathbb{R}^d$ the set $\Psi(x) := \{x' \in \mathrm{supp}(\nu) : |x - x'| = \mathrm{dist}(x, \mathrm{supp}(\nu))\}$ is nonempty and closed. Besides, the multifunction Ψ is weakly measurable[1], therefore by the Kuratowski–Ryll-Nardzewski theorem it admits a measurable selection which we denote by nn_ν. We denote by μ^* the associated pushforward of μ by nn_ν. Then it is clear that $\mathrm{supp}(\mu^*) \subset \mathrm{supp}(\nu)$ on the one hand, and that

$$W_q^q(\mu, \mu^*) \leq \mathbb{E}\left[\mathrm{dist}(X, \mathrm{supp}(\nu))^q\right], \qquad X \sim \mu,$$

on the other hand. We finally deduce from the approximation result stated in Lemma 2.4 below that

$$\inf_\rho W_q^q(\mu, \rho \mathrm{d}\nu) = W_q^q(\mu, \mu^*) = \mathbb{E}\left[\mathrm{dist}(X, \mathrm{supp}(\nu))^q\right], \qquad X \sim \mu,$$

which thereby generalizes the results of Proposition 2.2. Notice that there may not exist a minimizer ρ for this problem as the measure μ^* need not be absolutely continuous with respect to ν.

Lemma 2.4 (W_q approximation by absolutely continuous measures)**.** *Let μ^* and ν be two probability measures on $\mathbb{R}^q$ such that $\mathrm{supp}(\mu^*) \subset \mathrm{supp}(\nu)$. For any $\epsilon > 0$, there exists a probability density ρ with respect to ν such that, for all $q \geq 1$, $W_q(\mu^*, \rho \mathrm{d}\nu) < \epsilon$.*

Proof. Let $\epsilon > 0$ and $X \in \mathbb{R}^d$ be a random variable with law μ^*. Almost surely, $X \in \mathrm{supp}(\mu^*) \subset \mathrm{supp}(\nu)$ and therefore $\nu(B(X, \epsilon)) > 0$, which allows to draw Y with conditional distribution

$$\mathrm{d}\nu(y | B(X, \epsilon)) := \mathbb{1}_{\{y \in B(X,\epsilon)\}} \frac{\mathrm{d}\nu(y)}{\nu(B(X, \epsilon))}.$$

On the one hand, the random variable Y has density

$$\rho(y) := \int_{x \in \mathbb{R}^d} \mathbb{1}_{\{y \in B(x,\epsilon)\}} \frac{\mathrm{d}\mu^*(x)}{\nu(B(x, \epsilon))}$$

with respect to ν, and on the other hand we have $|X - Y| < \epsilon$, almost surely, which ensures that $W_q(\mu^*, \rho \mathrm{d}\nu) < \epsilon$ for any $q \geq 1$. □

[1] Let us denote $S = \mathrm{supp}(\nu)$ and fix $U \subset \mathbb{R}^d$ an open set, which we write as the countable union of closed sets $(F_n)_{n \geq 1}$. Then $\{x \in \mathbb{R}^d : \Psi(x) \cap U \neq \varnothing\} = \cup_{n \geq 1} \{x \in \mathbb{R}^d : \mathrm{dist}(x, (S \cap F_n)) = \mathrm{dist}(x, S)\}$ and it is easily seen that each set in the right-hand side is measurable.

3. Analysis of the Wasserstein distance $W_q(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k)}_{\mathbf{X}'_m})$

In this section, we study the asymptotic behavior of $\mathbb{E}[W_q^q(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k)}_{\mathbf{X}'_m})]$ when $n, m \to +\infty$. To this aim, we first notice that by Proposition 2.2, we have

$$\begin{aligned} \mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right)\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \mathrm{NN}_{\mathbf{X}'_m}(X_i)\right|^q\right] \\ &= \mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_m}(X)\right|^q\right], \end{aligned} \tag{13}$$

for $k = 1$, and

$$\begin{aligned} \mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k)}_{\mathbf{X}'_m}\right)\right] &\leq \mathbb{E}\left[\frac{1}{kn}\sum_{i=1}^{n}\sum_{l=1}^{k}\left|X_i - \mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X_i)\right|^q\right] \\ &= \frac{1}{k}\sum_{l=1}^{k}\mathbb{E}\left[\left|X - \mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X)\right|^q\right], \end{aligned} \tag{14}$$

for $k \geq 2$. Observe that the right-hand sides of both (13) and (14) no longer depend on n.

3.1. Consistency.

Our first main result is a consistency result. Before stating it in Theorem 3.3, we formulate two crucial assumptions.

Assumption 3.1 (Support condition)**.** *We have* $\mathrm{supp}(\mu_X) \subset \mathrm{supp}(\mu_{X'})$.

Assumption 3.2 (Min-integrability)**.** *There exists an integer* $m_0 \geq 1$ *such that*

$$\mathbb{E}\left[\min_{j\in[\![1,m_0]\!]}|X'_j|\right] < +\infty.$$

Theorem 3.3 (Consistency)**.** *Let Assumptions 3.1 and 3.2 hold. For all* $q \in [1, +\infty)$ *such that* $\mathbb{E}[|X|^q] < +\infty$*, and any sequence of positive integers* $(k_m)_{m\geq 1}$ *such that* $k_m/m \to 0$ *when* $m \to +\infty$*, we have*

$$\lim_{m\to+\infty}\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k_m)}_{\mathbf{X}'_m}\right)\right] = 0,$$

uniformly in n.

Theorem 3.3 is proved in Subsection 3.3.

Remark 3.4 (On Assumption 3.2)**.** *Assumption 3.2 is obviously satisfied if* X' *has a finite first order moment, but also for some heavy-tailed distributions. It writes under the equivalent form*

$$\int_0^{+\infty}\mathbb{P}(|X'| > r)^{m_0}\mathrm{d}r < +\infty,$$

which may be easier to check. An example of a random variable which does not satisfy this assumption, in dimension $d = 1$*, is* $X' = \exp(1/U)$ *where* U *is a uniform random variable on* $[0, 1]$.

Remark 3.5 (Limit without support condition)**.** *If the support condition of Assumption 3.1 does not hold, then the proof of Theorem 3.3 may be adapted to show that*

$$\lim_{m,n\to+\infty}\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right)\right] = \mathbb{E}\left[\mathrm{dist}(X, \mathrm{supp}(\mu_{X'}))^q\right].$$

3.2. **Rates of convergence.** The next step of our study consists in complementing Theorem 3.3 with a rate of convergence. We first discuss the case $k = 1$. Following (13), we start by writing

$$\begin{aligned}\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right)\right] &= \mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_m}(X)\right|^q\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_m}(X)\right|^q \middle| X\right]\right],\end{aligned} \tag{15}$$

and observe that for any $x \in \mathrm{supp}(\mu_X)$, $|x - \mathrm{NN}_{\mathbf{X}'_m}(x)| = \min_{j\in[\![1,m]\!]} |x - X'_j|$. If there is an open set U of $\mathbb{R}^d$ containing x and such that $\mu_{X'}(\cdot \cap U)$ has a density $p_{X'}$ with respect to the Lebesgue measure which is continuous at x, then an elementary computation shows that, for all $r \geq 0$,

$$\lim_{m\to+\infty} \mathbb{P}\left(m^{1/d} \min_{j\in[\![1,m]\!]} |x - X'_j| > r\right) = \exp\left(-r^d v_d p_{X'}(x)\right),$$

where v_d denotes the volume of the unit sphere of $\mathbb{R}^d$ for the norm $|\cdot|$. If $p_{X'}(x) > 0$ then this indicates that the correct order of convergence in Theorem 3.3 should be $m^{-q/d}$. If $p_{X'}(x) = 0$, or if the measure $\mu_{X'}(\cdot \cap U)$ is not absolutely continuous with respect to the Lebesgue measure, it is easy to construct elementary examples yielding different rates of convergence; see also [5, Chapter 2] for the singular case. We leave these peculiarities apart and work under the following strengthening of the support condition of Assumption 3.1.

Assumption 3.6 (Strong support condition)**.** *There exists an open set $U \subset \mathbb{R}^d$ which contains $\mathrm{supp}(\mu_X)$ and such that:*

(i) the measure $\mu_{X'}(\cdot \cap U)$ has a density $p_{X'}$ with respect to the Lebesgue measure;

(ii) the density $p_{X'}$ is continuous and positive on U;

(iii) there exist $\kappa \in (0, 1]$ and $r_\kappa > 0$ such that, for any $x \in U$, for any $r \in [0, r_\kappa]$,

$$\mathbb{P}\left(X' \in B(x, r)\right) \geq \kappa p_{X'}(x) v_d r^d.$$

Obviously, Assumption 3.6 implies Assumption 3.1 because then $\mathrm{supp}(\mu_X) \subset U \subset \mathrm{supp}(\mu_{X'})$. Part (iii) of Assumption 3.6 was introduced in [20] in the context of Nearest Neighbor classification, and called *Strong minimal mass assumption* there. Similar assumptions are commonly used in set estimation, geometric inference and quantization, such as *standardness* [14] or *Ahlfors regularity* [24].

Under Assumption 3.6, for all $x \in \mathrm{supp}(\mu_X)$, a positive random variable Z such that $\mathbb{P}(Z > r) = \exp(-r^d v_d p_{X'}(x))$ has moments

$$\mathbb{E}\left[Z^q\right] = \frac{\Gamma(1 + q/d)}{(v_d p_{X'}(x))^{q/d}},$$

where Γ denotes Euler's Gamma function. Therefore, as soon as the sequence

$$m^{q/d} \min_{j\in[\![1,m]\!]} \left|X - X'_j\right|^q$$

is uniformly integrable, the normalized quantity

$$m^{q/d}\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right)\right]$$

converges to

$$\frac{\Gamma(1 + q/d)}{v_d^{q/d}} \mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right],$$

when m goes to infinity. This statement appears for example in the literature of stochastic optimal quantization [23, Theorem 9.1]. Here, we provide an explicit moment condition ensuring uniform integrability.

Assumption 3.7 (Moments)**.** *In addition to Assumption 3.6, the condition*

$$\mathbb{E}\left[\frac{1+|X|^q}{p_{X'}(X)^{q/d}}\right] < +\infty$$

holds.

Assumptions 3.6 and 3.7 are discussed in more detail below. We now state our second main result.

Theorem 3.8 (Convergence rates for $k = 1$)**.** *Let Assumptions 3.2 and 3.6 hold, and let $q \in [1, +\infty)$ be such that Assumption 3.7 holds. Then we have*

$$\lim_{m\to+\infty} m^{q/d}\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right)\right] = \frac{\Gamma(1+q/d)}{v_d^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right].$$

Theorem 3.8 is proved in Subsection 3.3.

We now discuss the estimation of $\widehat{\mu}_{\mathbf{X}_n}$ by the weighted empirical measure $\widehat{\mu}^{(k)}_{\mathbf{X}'_m}$ for an arbitrary $k \in [\![1, m]\!]$. By (10), we first observe that we always have

$$W_q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k)}_{\mathbf{X}'_m}\right) \geq W_q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right),$$

so that the estimation of $\widehat{\mu}_{\mathbf{X}_n}$ is deteriorated by increasing the number of neighbors. Still, in the asymptotic regime of Theorem 3.3, a bound of the same order of magnitude as Theorem 3.8 may be obtained.

Corollary 3.9 (Convergence rates for k-NN)**.** *Under the assumptions of Theorem 3.8, for any nondecreasing sequence of positive integers $(k_m)_{m\geq 1}$ such that $k_m/m \to 0$ when $m \to +\infty$, we have*

$$\limsup_{m\to+\infty}\left(\frac{m}{k_m}\right)^{q/d}\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k_m)}_{\mathbf{X}'_m}\right)\right] \leq c_{d,q}\frac{\Gamma(1+q/d)}{v_d^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right],$$

with some constant $c_{d,q} > 1$.

Corollary 3.9 is proved in Subsection 3.3, where the expression of the constant $c_{d,q}$ is also given.

Remark 3.10 (Optimal choice of $\mu_{X'}$)**.** *When X has a density p_X with respect to the Lebesgue measure, an interesting fact is that the minimum of the quantity $\mathbb{E}\left[1/p_{X'}(X)^{q/d}\right]$ over the probability measure $p_{X'}$ is not reached when $p_{X'} = p_X$. Instead, according to* [41]*, the minimum is attained when $p_{X'}(x) \propto p_X(x)^{d/(q+d)}$.*

Remark 3.11 (NN distance without covariate shift)**.** *In the case where $\mu_X = \mu_{X'}$, the quantity*

$$\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right)\right] = \mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_m}(X)\right|^q\right]$$

is called Nearest Neighbor distance*. It naturally arises in the theoretical study of NN regression and classification* [5, Chapter 2]*. Previous works on the topic focus mainly on the convergence when $q = 2$ and assume that X has a bounded support* [5, 17, 26, 32]*. Some works* [13, 25] *consider some random variables X with unbounded support in the context of k-NN regression, but make the assumption of a bounded regression function ψ.*

In this perspective, a direct corollary from Theorem 3.8 is the following statement: let X have a density p_X for which the strong minimal mass assumption 3.6 (iii) holds with $U = \mathbb{R}^d$ and

$$\int_{\mathbb{R}^d}(1+|x|^q)p_X(x)^{1-q/d}\mathrm{d}x < +\infty.$$

Let $\mathbf{X}_n = (X_1, \dots, X_n)$ be an i.i.d sample from p_X, independent from X, and let $\mathrm{NN}_{\mathbf{X}_n}$ denote the NN among $\mathbf{X}_n$. We have

$$\mathbb{E}\left[|X - \mathrm{NN}_{\mathbf{X}_n}(X)|^q\right] \underset{n\to+\infty}{\sim} \frac{\Gamma(1+q/d)}{v_d^{q/d} n^{q/d}} \int_{\mathbb{R}^d} p_X(x)^{1-q/d} \mathrm{d}x.$$

This extends the results of the literature by ensuring the asymptotic equivalence for random variables with unbounded support.

Let us conclude this subsection with some comments on Assumptions 3.6 and 3.7. When X has a compact support, Assumptions 3.6 and 3.7 are verified as soon as $\mu_{X'}$ has a continuous density $p_{X'}$ which is bounded from below and above on an open set U' containing the support of μ_X. Indeed, in that case there exist $\epsilon > 0$ and an open subset U of U' such that U contains $\mathrm{supp}(\mu_X)$ and U' contains the ϵ-neighborhood of U. Then, Assumption 3.6 (iii) is verified with $r_\kappa < \epsilon$ and $\kappa = \inf_{x\in U'} p_{X'}(x) / \sup_{x\in U'} p_{X'}(x)$.

Assumptions 3.6 and 3.7 also hold in some nontrivial noncompact cases. An example of a sufficient condition for Assumption 3.6, which does not depend on μ_X, is given in the next statement and is proved in Subsection 3.3.

Lemma 3.12 (Radial density - Sufficient condition for Assumption 3.6)**.** *Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$, induced by an inner product and not necessarily identical to $|\cdot|$. If $\mu_{X'}$ has a density $p_{X'}$ with respect to the Lebesgue measure on $\mathbb{R}^d$, which writes $p_{X'}(x) = h(\|x - x_0\|)$ for some $x_0 \in \mathbb{R}^d$ and $h : [0, +\infty) \to \mathbb{R}$ continuous, positive and nonincreasing, then Assumption 3.6 holds with $U = \mathbb{R}^d$.*

We also refer to [20, Section 2.4] for a discussion of this assumption.

Assumption 3.7 gives a relationship between μ_X and $p_{X'}$ to ensure the convergence. In essence, it asserts that the tail of μ_X must be quite lightweight compared to the tail of $p_{X'}$. For instance, if X and X' are centered Gaussian vectors with respective covariance $\sigma^2 I_d$ and $\sigma'^2 I_d$, then by Lemma 3.12, Assumption 3.6 is satisfied with $U = \mathbb{R}^d$, and it is easy to check that for $q \in [1, +\infty)$, Assumption 3.7 holds if and only if $\sigma'^2 > \sigma^2 q/d$.

3.3. **Proofs.** In this subsection, we present the proofs of Theorems 3.3 and 3.8, Corollary 3.9 and Lemma 3.12.

Proof of Theorem 3.3. We begin our proof with the constant case $k_m = 1$ for all m and then extend it to the general case. We recall that by (13),

$$\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m}\right)\right] = \mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_m}(X)\right|^q\right] = \mathbb{E}\left[\min_{j\in[\![1,m]\!]} |X - X'_j|^q\right].$$

By Assumption 3.1, $X \in \mathrm{supp}(\mu_{X'})$ almost surely, so that we deduce from Lemma 2.2 in [5, Chapter 2] that

$$\min_{j\in[\![1,m]\!]} |X - X'_j|^q \xrightarrow[m\to+\infty]{a.s.} 0.$$

Let m_0 be the integer given by Assumption 3.2, we have

$$\min_{j\in[\![1,m]\!]} |X - X'_j|^q \le 2^{q-1}\left(|X|^q + \min_{j\in[\![1,m]\!]} |X'_j|^q\right).$$

The random variable $|X|^q$ is integrable by assumption and for $m \geq \lceil q \rceil m_0$, the inequality

$$\begin{aligned}\mathbb{E}\left[\min_{j\in[\![1,m]\!]} \left|X'_j\right|^q\right] &\leq \mathbb{E}\left[\min_{j\in[\![1,m]\!]} \left|X'_j\right|^{\lceil q\rceil}\right]^{q/\lceil q\rceil} \\ &\leq \mathbb{E}\left[\min_{j\in[\![1,m_0]\!]} \left|X'_j\right| \min_{j\in[\![m_0+1,2m_0]\!]} \left|X'_j\right| \cdots \min_{j\in[\![(\lceil q\rceil-1)m_0+1,\lceil q\rceil m_0]\!]} \left|X'_j\right|\right]^{q/\lceil q\rceil} \\ &\leq \mathbb{E}\left[\min_{j\in[\![1,m_0]\!]} \left|X'_j\right|\right]^q < +\infty\end{aligned}$$

holds. Then by the dominated convergence theorem,

$$\mathbb{E}\left[\min_{j\in[\![1,m]\!]} |X - X'_j|^q\right] \underset{m\to+\infty}{\longrightarrow} 0.$$

For the general case $k_m/m \to 0$, we adapt directly the proof of [5, Theorem 2.4] to the context $\mu_X \neq \mu_{X'}$. Let us fix $l \in [\![1, m/2]\!]$ and partition the set $\{X'_1, \ldots, X'_m\}$ into $2l$ sets of size $m_1, \ldots, m_{2l}$ with, for all $j \in [\![1, 2l]\!]$,

$$\lfloor m/2l \rfloor \leq m_j \leq \lfloor m/2l \rfloor + 1.$$

We denote by $\mathrm{NN}^{(1,j)}_{\mathbf{X}'_m}$ the 1-NN among the subset j. By the definition of $\mathrm{NN}^{(l)}_{\mathbf{X}'_m}$, there are at least l subsets j for which

$$|X - \mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X)| \leq |X - \mathrm{NN}^{(1,j)}_{\mathbf{X}'_m}(X)|,$$

therefore

$$|X - \mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X)|^q \leq \frac{1}{l}\sum_{j=1}^{2l} |X - \mathrm{NN}^{(1,j)}_{\mathbf{X}'_m}(X)|^q,$$

and consequently

$$\mathbb{E}\left[\left|X - \mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X)\right|^q\right] \leq 2\mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_{\lfloor m/2l\rfloor}}(X)\right|^q\right].$$

Finally, we deduce from (14) that, as soon as $k_m \leq m/2$,

$$\begin{aligned}\mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k_m)}_{\mathbf{X}'_m}\right)\right] &\leq \frac{1}{k_m}\sum_{l=1}^{k_m} \mathbb{E}\left[\left|X - \mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X)\right|^q\right] \\ &\leq \frac{2}{k_m}\sum_{l=1}^{k_m} \mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_{\lfloor m/2l\rfloor}}(X)\right|^q\right] \\ &\leq 2\mathbb{E}\left[\left|X - \mathrm{NN}_{\mathbf{X}'_{\lfloor m/2k_m\rfloor}}(X)\right|^q\right],\end{aligned} \tag{16}$$

which goes to 0 as a consequence of the first part of the proof when $m/2k_m$ goes to infinity. □

Proof of Theorem 3.8. By (13), we have

$$\begin{aligned}\mathbb{E}\left[m^{q/d} W_q^q(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m})\right] &= \mathbb{E}\left[\mathbb{E}\left[m^{q/d} \min_{j\in[\![1,m]\!]} \left|X - X'_j\right|^q \middle| X\right]\right] \\ &= \int_{\mathbb{R}^d}\int_0^{+\infty} \mathbb{P}\left(m^{q/d} \min_{j\in[\![1,m]\!]} |x - X'_j|^q > t\right) \mathrm{d}t\mathrm{d}\mu_X(x) \\ &= \int_{\mathbb{R}^d}\int_0^{+\infty} \mathbb{P}(m^{q/d}|x - X'|^q > t)^m \mathrm{d}t\mathrm{d}\mu_X(x),\end{aligned} \tag{17}$$

by independence of the X'_j. The proof consists in computing the pointwise limit of $\mathbb{P}(m^{p/d}|x - Y|^p > t)^m$ for $(x, t) \in \mathrm{supp}(\mu_X) \times \mathbb{R}^+$ and then establishing the convergence of the integral via the dominated convergence theorem.

Pointwise convergence. We have

$$\begin{aligned}\mathbb{P}(m^{q/d}|x-X'|^q > t)^m &= \left(1-\mathbb{P}(|x-X'| \leq t^{1/q}/m^{1/d})\right)^m \\ &= \exp\left(m\log\left(1-\mathbb{P}(|x-X'| \leq t^{1/q}/m^{1/d})\right)\right).\end{aligned}$$

By Assumption 3.6, we have

$$\mathbb{P}(|x-X'| \leq t^{1/q}/m^{1/d}) = p_{X'}(x)v_d t^{d/q}/m + \mathrm{o}(1/m),$$

with v_d the volume of the unit sphere. Thus

$$m\log\left(1-\mathbb{P}(|x-X'| \leq t^{1/q}/m^{1/d})\right) = -p_{X'}(x)v_d t^{d/q} + \mathrm{o}(1),$$

and we conclude that

$$\mathbb{P}(m^{q/d}|x-X'|^q > t)^m \underset{m\to+\infty}{\longrightarrow} \exp\left(-p_{X'}(x)v_d t^{d/q}\right).$$

Dominated convergence. Let $r_\kappa > 0$ be given by Assumption 3.6. We split the integral in the right-hand side of (17) and study each term separately

$$\int_{\mathbb{R}^d}\int_0^{+\infty} \mathbb{P}(m^{q/d}|x-X'|^q > t)^m \mathrm{d}t\mathrm{d}\mu_X(x) = \mathrm{I} + \mathrm{II}$$

with

$$\begin{aligned}\mathrm{I} &:= \int_{\mathbb{R}^d}\int_0^{r_\kappa^q m^{q/d}} \mathbb{P}(|x-X'| > t^{1/q}/m^{1/d})^m \mathrm{d}t\mathrm{d}\mu_X(x), \\ \mathrm{II} &:= \int_{\mathbb{R}^d}\int_{r_\kappa^q m^{q/d}}^{+\infty} \mathbb{P}(|x-X'| > t^{1/q}/m^{1/d})^m \mathrm{d}t\mathrm{d}\mu_X(x).\end{aligned}$$

Convergence of I. For $t \in [0, r_\kappa^q m^{q/d}]$, we have $t^{1/q}/m^{1/d} \leq r_\kappa$ and thus

$$\begin{aligned}\mathbb{P}(|x-X'| > t^{1/q}/m^{1/d})^m &= \left(1-\mathbb{P}(|x-X'| \leq t^{1/q}/m^{1/d})\right)^m \\ &\leq \left(1-\frac{p_{X'}(x)v_d\kappa t^{d/q}}{m}\right)^m\end{aligned}$$

by Assumption 3.6.

Using the elementary inequality $(1-a/n)^n \leq \exp(-a)$ for $a \leq n$, we can write

$$\mathbb{P}(|x-X'| > t^{1/q}/m^{1/d})^m \leq \exp(-\kappa v_d p_{X'}(x)t^{d/q}).$$

This bound does not depend on m and the integral

$$\begin{aligned}\int_{\mathbb{R}^d}\int_0^{+\infty} \exp\left(-\kappa v_d p_{X'}(x)t^{d/q}\right)\mathrm{d}t\mathrm{d}\mu_X(x) &= \int_{\mathbb{R}^d} \frac{\Gamma(1+q/d)}{(\kappa v_d p_{X'}(x))^{q/d}}\mathrm{d}\mu_X(x) \\ &= \frac{\Gamma(1+q/d)}{(\kappa v_d)^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right]\end{aligned}$$

is finite by Assumption 3.7. We therefore deduce from the dominated convergence theorem that

$$\begin{aligned}\lim_{m\to+\infty}\mathrm{I} &= \int_{\mathbb{R}^d}\int_0^{+\infty} \exp\left(-p_{X'}(x)v_d t^{d/q}\right)\mathrm{d}t\mathrm{d}\mu_X(x) \\ &= \frac{\Gamma(1+q/d)}{v_d^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right].\end{aligned}$$

Convergence of II. Let $m \geq 2(q+1)m_0$. Using the change of variable $r^q = t/m^{q/d}$, we have

$$\begin{aligned}\mathrm{II} &= q \int_{\mathbb{R}^d} \int_{r_\kappa}^{+\infty} m^{q/d} r^{q-1} \mathbb{P}(|x - X'| > r)^m \mathrm{d}r \mathrm{d}\mu_X(x) \\ &\leq q \int_{\mathbb{R}^d} \int_{r_\kappa}^{+\infty} V_m(x,r) \mathrm{d}r \mathrm{d}\mu_X(x),\end{aligned}$$

with

$$V_m(x,r) := m^{q/d} \mathbb{P}(|x - X'| > r_\kappa)^{m-(q+1)m_0} r^{q-1} \mathbb{P}(|x - X'| > r)^{(q+1)m_0}.$$

As $\mathbb{P}(|x - X'| > r_\kappa) < 1$ for all x in U, by Assumption 3.6, $V_m(x,r)$ is pointwise convergent to 0 on the support of μ_X. We check that $V_m(x,r)$ is bounded from above by an integrable function which does not depend on m. Let us denote $m' = m - (q+1)m_0 \geq m/2$ and rewrite

$$\begin{aligned}m^{q/d} \mathbb{P}(|x - X'| > r_\kappa)^{m-(q+1)m_0} &= \left(\frac{m}{m'}\right)^{q/d} m'^{q/d} \mathbb{P}(|x - X'| > r_\kappa)^{m'} \\ &\leq 2^{q/d} m'^{q/d} \left(1 - \mathbb{P}(|x - X'| \leq r_\kappa)\right)^{m'} \\ &\leq 2^{q/d} m'^{q/d} \exp\left(-m' \kappa p_{X'}(x) v_d r_\kappa^d\right),\end{aligned}$$

where we have used Assumption 3.6 and the elementary above inequality at the third line. We deduce that

$$m^{q/d} \mathbb{P}(|x - X'| > r_\kappa)^{m-(q+1)m_0} \leq \frac{C_1}{p_{X'}(x)^{q/d}}, \quad C_1 := \frac{2^{q/d}}{(\kappa v_d r_\kappa^d)^{q/d}} \sup_{u \geq 0}(u^{q/d} e^{-u}),$$

so that

$$V_m(x,r) \leq \widetilde{V}(x,r) := \frac{C_1}{p_{X'}(x)^{q/d}} r^{q-1} \mathbb{P}(|x - X'| > r)^{(q+1)m_0}. \tag{18}$$

To complete the proof, we verify that $\widetilde{V}(x,r)$ is integrable on $U \times [r_\kappa, +\infty)$. We first fix $x \in \mathbb{R}^d$ and estimate the integral of $\widetilde{V}(x,r)$ in r. Using the fact that if $|x - X'| > r$ then $|X'| > r - |x|$, we first write

$$\begin{aligned}\int_{r_\kappa}^{+\infty} r^{q-1} \mathbb{P}(|x - X'| > r)^{(q+1)m_0} \mathrm{d}r &\leq \int_0^{+\infty} r^{q-1} \mathbb{P}(|X'| > r - |x|)^{(q+1)m_0} \mathrm{d}r \\ &= \int_{-|x|}^{+\infty} (r + |x|)^{q-1} \mathbb{P}(|X'| > r)^{(q+1)m_0} \mathrm{d}r.\end{aligned}$$

On the interval $[-|x|, 0]$, we have

$$\int_{-|x|}^0 (r + |x|)^{q-1} \mathbb{P}(|X'| > r)^{(q+1)m_0} \mathrm{d}r = \int_{-|x|}^0 (r + |x|)^{q-1} \mathrm{d}r = \frac{|x|^q}{q}.$$

On the interval $[0, +\infty)$, we first rewrite

$$\int_0^{+\infty} (r + |x|)^{q-1} \mathbb{P}(|X'| > r)^{(q+1)m_0} \mathrm{d}r = \int_0^{+\infty} (r + |x|)^{q-1} \mathbb{P}\left(\min_{j \in [\![1, m_0]\!]} |X'_j| > r\right)^{q+1} \mathrm{d}r,$$

and recall from Assumption 3.2 that $C_2 := \mathbb{E}[\min_{j \in [\![1,m_0]\!]} |X'_j|] < +\infty$. As a consequence, we deduce from Markov's inequality that the right-hand side in the previous equality is bounded from above by

$$\int_0^{|x| \vee 1} (r + |x|)^{q-1} \mathrm{d}r + C_2^{q+1} \int_{|x| \vee 1}^{+\infty} \frac{(r + |x|)^{q-1}}{r^{q+1}} \mathrm{d}r.$$

If $|x| \leq 1$ then this expression is bounded from above. If $|x| > 1$, then we have

$$\int_0^{|x|} (r + |x|)^{q-1} \mathrm{d}r \leq 2^{q-1} |x|^q$$

on the one hand, and

$$\int_{|x|}^{+\infty} \frac{(r+|x|)^{q-1}}{r^{q+1}}\mathrm{d}r = \frac{1}{|x|}\int_{1}^{+\infty} \frac{(u+1)^{q-1}}{u^{q+1}}\mathrm{d}u,$$

which is bounded from above, on the other hand. Overall, we conclude that there exists a constant C_3 such that

$$\int_{r_\kappa}^{+\infty} r^{q-1}\mathbb{P}(|x-X'|>r)^{(q+1)m_0}\mathrm{d}r \leq C_3(1+|x|^q). \tag{19}$$

As a consequence, the combination of (18) and (19) yields

$$\int_{\mathbb{R}^d}\int_{r_\kappa}^{+\infty} \widetilde{V}(x,r)\mathrm{d}r\mathrm{d}\mu_X(x) \leq C_1C_3\mathbb{E}\left[\frac{1+|X|^q}{p_{X'}(X)^{q/d}}\right],$$

which by Assumption 3.7 allows to apply the dominated convergence theorem to show that II goes to 0, and thereby completes the proof. □

Proof of Corollary 3.9. We start from the second line of Equation (16) and estimate its right-hand side

$$\begin{aligned}
\mathbb{E}\left[W_q^q(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k_m)}_{\mathbf{X}'_m})\right] &\leq \frac{2}{k_m}\sum_{l=1}^{k_m}\mathbb{E}\left[\left|X-\mathrm{NN}_{\mathbf{X}'_{\lfloor m/2l\rfloor}}(X)\right|^q\right] \\
&= \frac{2}{k_m}\sum_{l=1}^{k_m}\left(\frac{2k_m}{m}\frac{l}{k_m}\frac{m}{2l}\right)^{q/d}\mathbb{E}\left[\left|X-\mathrm{NN}_{\mathbf{X}'_{\lfloor m/2l\rfloor}}(X)\right|^q\right] \\
&= \left(\frac{k_m}{m}\right)^{q/d}\frac{2^{q/d+1}}{k_m}\sum_{l=1}^{k_m}\left(\frac{l}{k_m}\right)^{q/d}F\left(\frac{m}{2l}\right)
\end{aligned}$$

with $F(u) = u^{q/d}\mathbb{E}[|X-\mathrm{NN}_{\mathbf{X}'_{\lfloor u\rfloor}}(X)|^q]$. Let $\epsilon>0$. By Theorem 3.8, there exists $u_\epsilon \geq 0$ such that, for all $u \geq u_\epsilon$,

$$\left|F(u) - \frac{\Gamma(1+q/d)}{v_d^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right]\right| \leq \epsilon.$$

We can remark that for $m\in\mathbb{N}^*$ and $l\in[\![1,k_m]\!]$,

$$\frac{m}{2l} \geq \frac{m}{2k_m} \xrightarrow[m\to+\infty]{} +\infty.$$

Thus, if we take m_ϵ such that for all $m\geq m_\epsilon$, $\left\lfloor \frac{m}{2k_m}\right\rfloor \geq u_\epsilon$, we have

$$\left|F\left(\frac{m}{2l}\right) - \frac{\Gamma(1+q/d)}{v_d^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right]\right| \leq \epsilon$$

for any $m\geq m_\epsilon$ and $l\leq k_m$. Consequently,

$$\left|\frac{1}{k_m}\sum_{l=1}^{k_m}\left(\frac{l}{k_m}\right)^{q/d}\left(F\left(\frac{m}{2l}\right) - \frac{\Gamma(1+q/d)}{v_d^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right]\right)\right| \leq \epsilon\left|\frac{1}{k_m}\sum_{l=1}^{k_m}\left(\frac{l}{k_m}\right)^{q/d}\right| \leq \epsilon,$$

so that

$$\lim_{m\to+\infty}\frac{2^{q/d+1}}{k_m}\sum_{l=1}^{k_m}\left(\frac{l}{k_m}\right)^{q/d}F\left(\frac{m}{2l}\right) = c_{d,q}\frac{\Gamma(1+q/d)}{v_d^{q/d}}\mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right],$$

where

$$c_{d,q} := \lim_{m\to+\infty} \frac{2^{q/d+1}}{k_m} \sum_{l=1}^{k_m} \left(\frac{l}{k_m}\right)^{q/d}$$
$$= \begin{cases} \dfrac{2^{q/d+1}}{k} \displaystyle\sum_{l=1}^{k} \left(\frac{l}{k}\right)^{q/d} & \text{if } \sup_{m\geq 1} k_m = k < +\infty, \\ 2^{q/d+1} \displaystyle\int_0^1 u^{q/d} \mathrm{d}u = \frac{2^{q/d+1}}{q/d+1} & \text{if } \sup_{m\geq 1} k_m = +\infty, \end{cases}$$

because k_m is nondecreasing. This concludes the proof. □

Proof of Lemma 3.12. Obviously, it suffices to check that $p_{X'}$ satisfies (iii) in Assumption 3.6. Let us denote by $\langle\cdot,\cdot\rangle$ and $\mathcal{B}(x,r)$ respectively the inner product and the ball of center x and radius r associated to $\|\cdot\|$. We set $x_0 = 0$ without loss of generality. As h is positive and nonincreasing, we may fix $r_0 > 0$ and define

$$\overline{\kappa} := \frac{h(r_0)}{h(0)} \in (0,1].$$

If $\|x\| \leq r_0/2$, then for all $y \in \mathcal{B}(0, r_0/2)$, the monotonicity of h ensures that $p_{X'}(x+y) \geq \overline{\kappa} p_{X'}(x)$. By the equivalence of the norms, there exist $C \geq c > 0$ such that for any $x \in \mathbb{R}^d$ and any $r \geq 0$, $\mathcal{B}(x, cr) \subset B(x,r) \subset \mathcal{B}(x, Cr)$. Thus

$$\forall r \leq r_0/2c, \qquad \mathbb{P}\left(X' \in B(x,r)\right) \geq \mathbb{P}\left(X' \in \mathcal{B}(x,cr)\right) \geq (c/C)^d v_d \overline{\kappa} p_{X'}(x) r^d.$$

If $\|x\| > r_0/2$, let us introduce the half-cone

$$\mathcal{C}_x = \left\{ x' \in \mathbb{R}^d : \langle x' - x, -x\rangle \geq \frac{\|x'-x\|\|x\|}{2} \right\},$$

and notice that for all $r \leq r_0/2$ and $x' \in \mathcal{C}_x \cap \mathcal{B}(x,r)$,

$$\begin{aligned} \|x'\|^2 &= \|x\|^2 + \|x'-x\|^2 + 2\langle x'-x, x\rangle \\ &\leq \|x\|^2 + \|x'-x\|^2 - \|x'-x\|\|x\| \\ &\leq \|x\|^2 + \|x'-x\|^2 - \|x'-x\|^2 = \|x\|^2. \end{aligned}$$

Thus, for all $x' \in \mathcal{C}_x \cap \mathcal{B}(x,r)$, $p_{X'}(x') \geq p_{X'}(x)$. For a given r, the sets $\mathcal{C}_x \cap \mathcal{B}(x,r)$ have the same volume for all x, which we denote by $\alpha v_d r^d$ for some $\alpha \in (0, 1/C^d)$. Finally, we have

$$\forall r \leq r_0/2c, \qquad \mathbb{P}(X' \in B(x,r)) \geq \mathbb{P}(X' \in \mathcal{B}(x,cr) \cap \mathcal{C}_x) \geq \alpha c^d v_d p_{X'}(x) r^d.$$

If we take $\kappa = (c/C)^d \min(\alpha C^d, \overline{\kappa})$ and $r_\kappa = r_0/2c$, we obtain the point (iii) of Assumption 3.6. □

4. Convergence of $\widehat{\mathrm{QI}}_{m,n}^{(k)}$ to QI

This section is dedicated to the study of the convergence of $\widehat{\mathrm{QI}}_{m,n}^{(k)}$ to QI. As a preliminary step, we complement the results from Section 3 by deriving rates of convergence for the Wasserstein distance between $\widehat{\mu}_{\mathbf{X}'_m}^{(k)}$ and μ_X in Subsection 4.1. We then distinguish between the *noiseless case* in which $Y = f(X)$, addressed in Subsection 4.2, and the *noisy case* $Y = f(X,\Theta)$, addressed in Subsection 4.3.

4.1. **Convergence of $\widehat{\mu}^{(k)}_{\mathbf{X}'_m}$ to μ_X.** Let us fix $q \in [1, +\infty)$ and use Jensen's inequality to write, for $k = k_m \in [\![1, m]\!]$,

$$\mathbb{E}\left[W_q^q\left(\mu_X, \widehat{\mu}^{(k_m)}_{\mathbf{X}'_m}\right)\right] \leq 2^{q-1}\left(\mathbb{E}\left[W_q^q\left(\mu_X, \widehat{\mu}_{\mathbf{X}_n}\right)\right] + \mathbb{E}\left[W_q^q\left(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(k_m)}_{\mathbf{X}'_m}\right)\right]\right). \tag{20}$$

Under the assumptions of Corollary 3.9, the second term has order of magnitude at most $(k_m/m)^{q/d}$. The study of the first term, namely the rate of convergence of the expected W_q distance (taken to the power q) between the empirical measure of iid realizations and their common distribution, has been the subject of several works. Under the condition that there exists $s > 2q$ such that $\mathbb{E}[|X|^s] < +\infty$, we have from [19, Theorem 1]

$$\mathbb{E}\left[W_q^q\left(\mu_X, \widehat{\mu}_{\mathbf{X}_n}\right)\right] = \begin{cases} \mathrm{O}\left(n^{-1/2}\right) & \text{if } q > d/2, \\ \mathrm{O}\left(n^{-1/2}\log(1+n)\right) & \text{if } q = d/2, \\ \mathrm{O}\left(n^{-q/d}\right) & \text{if } q < d/2. \end{cases} \tag{21}$$

These estimates may be improved if more assumptions are made on μ_X. For example, if this measure possesses a lower and upper bounded density on some bounded subset of $\mathbb{R}^d$, then the rate is known to be $n^{-q/d}$ even if $q > d/2$ [21]. This rate may even be improved if μ_X concentrates on a low-dimensional submanifold of $\mathbb{R}^d$ [39, 16], which is particularly relevant in the UQ context which motivates this study, see Remark 5.3. In order to make the use of our results as flexible as possible, from now on we shall denote by $(\tau_{q,d}(n))_{n\geq 1}$ a sequence such that

$$\mathbb{E}\left[W_q^q\left(\mu_X, \widehat{\mu}_{\mathbf{X}_n}\right)\right] = \mathrm{O}\left(\tau_{q,d}(n)\right),$$

and thus

$$\mathbb{E}\left[W_q^q\left(\mu_X, \widehat{\mu}^{(k_m)}_{\mathbf{X}'_m}\right)\right] = \mathrm{O}\left(\tau_{q,d}(n)\right) + \mathrm{O}\left(\left(\frac{k_m}{m}\right)^{q/d}\right).$$

As is sketched in the discussion above, the precise order of $\tau_{q,d}(n)$ depends on properties of the measure μ_X.

In the sequel, where we study the convergence of $\widehat{\mathrm{QI}}^{(k)}_{m,n}$ to QI, the W_1 distance plays a specific role, due to the Kantorovitch duality formula [38, Remark 6.5]

$$W_1(\mu, \nu) = \sup_{|\varphi|_{\mathrm{Lip}}\leq 1}\left\{\int_{\mathbb{R}^d}\varphi(x)\mathrm{d}\mu(x) - \int_{\mathbb{R}^d}\varphi(x)\mathrm{d}\nu(x)\right\}, \tag{22}$$

where $|\varphi|_{\mathrm{Lip}}$ denotes the Lipschitz constant of φ.

We shall need the following estimate.

Lemma 4.1 (W_1^q estimate)**.** *If $\mathbb{E}[|X|^q] < +\infty$ with $q \geq 2$, then*

$$\mathbb{E}\left[W_1^q\left(\mu_X, \widehat{\mu}_{\mathbf{X}_n}\right)\right] = \mathrm{O}\left(\frac{1}{n^{q/2}} + \tau_{1,d}(n)^q\right).$$

Proof. For any vector $(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$, let us define

$$\begin{aligned} W(x_1, \ldots, x_n) &:= W_1\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \mu_X\right) \\ &= \sup_{|\varphi|_{\mathrm{Lip}}\leq 1}\left\{\frac{1}{n}\sum_{i=1}^n \varphi(x_i) - \int_{\mathbb{R}^d}\varphi(x)\mu_X(\mathrm{d}x)\right\} \end{aligned}$$

thanks to (22). Then for any $i \in [\![1,n]\!]$ and $x_i^\flat \in \mathbb{R}^d$, using the identity above and the fact that $|\sup f - \sup g| \leq \sup |f-g|$, we get

$$\begin{aligned}
&\left|W(x_1,\dots,x_n) - W(x_1,\dots,x_{i-1},x_i^\flat,x_{i+1},\dots,x_n)\right| \\
&\leq \sup_{|\varphi|_{\mathrm{Lip}}\leq 1} \left| \frac{\varphi(x_1)+\cdots+\varphi(x_n)}{n} - \frac{\varphi(x_1)+\cdots+\varphi(x_i^\flat)+\cdots+\varphi(x_n)}{n}\right| \\
&= \frac{1}{n} \sup_{|\varphi|_{\mathrm{Lip}}\leq 1} \left|\varphi(x_i) - \varphi(x_i^\flat)\right| \\
&= \frac{1}{n}|x_i - x_i^\flat|.
\end{aligned}$$

As a consequence, letting $\mathbf{X}_n = (X_1,\dots,X_n)$ and $\mathbf{X}_n^\flat = (X_1^\flat,\dots,X_n^\flat)$ be two independent samples from μ_X, we deduce that the random variables V^+ and V^- defined by

$$V^\pm := \mathbb{E}\left[\sum_{i=1}^n \Big(W(X_1,\dots,X_n) - W(X_1,\dots,X_{i-1},X_i^\flat,X_{i+1},\dots,X_n)\Big)_\pm^2 |\mathbf{X}_n\right]$$

satisfy the bound

$$V^\pm \leq \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[|X_i - X_i^\flat|^2 |\mathbf{X}_n\right].$$

As a consequence, for any $q \geq 2$ we have by Jensen's inequality

$$\begin{aligned}
\mathbb{E}\left[(V^\pm)^{q/2}\right] &\leq \frac{1}{n^{q/2}}\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[|X_i - X_i^\flat|^2|X\right]\right)^{q/2}\right] \\
&\leq \frac{1}{n^{q/2}}\mathbb{E}\left[|X - X^\flat|^q\right] \leq \frac{2^{q-1}}{n^{q/2}}\mathbb{E}\left[|X|^q\right].
\end{aligned}$$

We therefore deduce from the higher-order Efron–Stein inequality [7, Theorem 2] that there exists a universal constant C_q such that

$$\mathbb{E}\left[|W(\mathbf{X}_n) - \mathbb{E}\left[W(\mathbf{X}_n)\right]|^q\right] \leq C_q \mathbb{E}\left[(V^-)^{q/2} + (V^+)^{q/2}\right] \leq \frac{2^q C_q}{n^{q/2}}\mathbb{E}\left[|X|^q\right].$$

We conclude the proof by writing, using Jensen's inequality again,

$$\begin{aligned}
\mathbb{E}\left[W(\mathbf{X}_n)^q\right] &= \mathbb{E}\left[|W(\mathbf{X}_n) - \mathbb{E}\left[W(\mathbf{X}_n)\right] + \mathbb{E}\left[W(\mathbf{X}_n)\right]|^q\right] \\
&\leq 2^{q-1}\left(\mathbb{E}\left[|W(\mathbf{X}_n) - \mathbb{E}\left[W(\mathbf{X}_n)\right]|^q\right] + \mathbb{E}\left[W(\mathbf{X}_n)\right]^q\right),
\end{aligned}$$

which yields the claimed estimate. □

Remark 4.2. *With the estimates given by (21), $\tau_{1,d}(n)$ is of order $n^{-1/d}$ as soon as $d \geq 3$. Thus, in this case, we get from Lemma 4.1 that $\mathbb{E}[W_1^q(\mu_X, \widehat{\mu}_{\mathbf{X}_n})]$ is of order $n^{-q/d}$. When $q > d/2$, this is a faster rate of decay to 0 than what one would have obtained bounding W_1 by W_q.*

4.2. **Rate of convergence of $\widehat{\mathrm{QI}}_{m,n}^{(k_m)}$ in the noiseless case.** We assume that $Y = f(X)$ and study the rate of convergence of $\widehat{\mathrm{QI}}_{m,n}^{(k_m)}$ to QI. When $\phi \circ f$ is L-Lipschitz continuous, we deduce from (22) that

$$\begin{aligned}
\left|\mathrm{QI} - \widehat{\mathrm{QI}}_{m,n}^{(k_m)}\right|^q &= \left|\int_{\mathbb{R}^d} \phi\circ f(x)\mathrm{d}\mu_X(x) - \int_{\mathbb{R}^d}\phi\circ f(x)\mathrm{d}\widehat{\mu}_{\mathbf{X}'_m}^{(k_m)}(x)\right|^q \\
&\leq L^q W_1^q\left(\mu_X, \widehat{\mu}_{\mathbf{X}'_m}^{(k_m)}\right).
\end{aligned}$$

We therefore obtain the following result.

Proposition 4.3 (Rates of convergence in the noiseless case)**.** *Assume that:*

(i) the function f does not depend on Θ,

(ii) the function $\phi \circ f$ is globally Lipschitz continuous,

and let the assumptions of Corollary 3.9 hold for some $q \geq 2$. Then

$$\mathbb{E}\left[\left|\mathrm{QI} - \widehat{\mathrm{QI}}_{m,n}^{(k_m)}\right|^q\right]^{1/q} = \mathrm{O}\left(\frac{1}{\sqrt{n}} + \tau_{1,d}(n)\right) + \mathrm{O}\left(\left(\frac{k_m}{m}\right)^{1/d}\right). \tag{23}$$

There is no need for k_m to go to infinity and thus $k_m = 1$ is optimal.

These computations can be adapted to cases other than $\phi \circ f$ Lipschitz continuous. For instance, if $A \subset \mathbb{R}^e$, $\phi(y) = \mathbb{1}_{\{y \in A\}}$ and f is globally Lipschitz continuous, it is possible to use the margin assumption of [37] to deduce theoretical rates of convergence in the estimation of $\mathrm{QI} = \mathbb{P}(Y \in A)$.

4.3. **Rate of convergence of $\widehat{\mathrm{QI}}_{m,n}^{(k_m)}$ in the noisy case.** We now study the convergence of $\widehat{\mathrm{QI}}_{m,n}^{(k_m)}$ to QI when $Y = f(X, \Theta)$. A first striking result is then that even under the assumptions of Theorem 3.3, the estimator $\widehat{\mathrm{QI}}_{m,n}^{(1)}$ need not be consistent. Indeed, consider the case where X is actually deterministic and always equal to some $x_0 \in \mathbb{R}^d$. Then we have

$$\widehat{\mathrm{QI}}_{m,n}^{(1)} = \frac{1}{n}\sum_{i=1}^{n} \phi(Y'_{j_i^{(1)}}),$$

where $j_i^{(1)}$ is the index of the closest X'_j to X_i. But since $X_i = x_0$ for all i, all indices $j_i^{(1)}$ are equal to some $j^{(1)}$ and the estimator rewrites

$$\widehat{\mathrm{QI}}_{m,n}^{(1)} = \phi(Y'_{j^{(1)}}) = \phi(f(X'_{j^{(1)}}, \Theta_{j^{(1)}})).$$

While Assumption 3.1 ensures that $X'_{j^{(1)}}$ converges to x_0 when $m \to +\infty$, in general the corresponding sequence of $\Theta_{j^{(1)}}$ does not converge.

As is evidenced on this example, the presence of an atom in the law of X makes the estimator $\widehat{\mathrm{QI}}_{m,n}^{(1)}$ depend on a single realization of Θ and therefore prevents this estimator from displaying an averaging behavior with respect to the law of Θ. In Proposition 4.4, we clarify this point by exhibiting a necessary and sufficient condition for the estimator $\widehat{\mathrm{QI}}_{m,n}^{(1)}$ to be consistent, while in Proposition 4.5, we show that replacing $\widehat{\mathrm{QI}}_{m,n}^{(1)}$ with $\widehat{\mathrm{QI}}_{m,n}^{(k_m)}$ with $k_m \to +\infty$ allows to recover such an averaging behavior and makes the estimator consistent, even when μ_X has atoms. In the latter case, we also provide rates of convergence in Proposition 4.6.

We recall that $\psi(x) = \mathbb{E}[\phi(f(x, \Theta))]$ is defined in Equation (6). In the next statement, we denote by $\mathcal{A}_X$ the set of atoms of μ_X, that is to say the set of $x \in \mathbb{R}^d$ such that $\mathbb{P}(X = x) > 0$, and introduce the notation

$$\vartheta(x) := \mathrm{Var}(\phi(f(x, \Theta))).$$

Proposition 4.4 (Consistency of the 1-NN in the noisy case)**.** *Assume that:*

(i) the function ϕ is bounded,

(ii) the function ψ is globally Lipschitz continuous,

(iii) the function ϑ is continuous,

and let the assumptions of Theorem 3.3 hold. We have

$$\mathbb{E}\left[\left|\widehat{\mathrm{QI}}^{(1)}_{m,n} - \mathrm{QI}\right|\right] \xrightarrow[m,n\to+\infty]{} 0$$

if and only if

$$\forall x \in \mathcal{A}_X, \qquad \vartheta(x) = 0.$$

In particular, under the above assumptions, if the law of X has no atom, *i.e.* $\mathcal{A}_X = \varnothing$, then $\widehat{\mathrm{QI}}^{(1)}_{m,n}$ converges to QI.

Proof. Let us write

$$\widehat{\mathrm{QI}}^{(1)}_{m,n} - \mathrm{QI} = \left(\widehat{\mathrm{QI}}^{(1)}_{m,n} - \widetilde{\mathrm{QI}}^{(1)}_{m,n}\right) + \left(\widetilde{\mathrm{QI}}^{(1)}_{m,n} - \mathrm{QI}\right),$$

with

$$\widetilde{\mathrm{QI}}^{(1)}_{m,n} = \frac{1}{m}\sum_{j=1}^{m} w_j^{(1)} \psi(X'_j).$$

Using the Lipschitz continuity of ψ, the duality formula (22) and Theorem 3.3, we get that $\widetilde{\mathrm{QI}}^{(1)}_{m,n} - \mathrm{QI}$ converges to 0 when $m, n \to +\infty$, in L^1. Therefore, $\widehat{\mathrm{QI}}^{(1)}_{m,n} - \mathrm{QI}$ converges to 0 if and only if $\widehat{\mathrm{QI}}^{(1)}_{m,n} - \widetilde{\mathrm{QI}}^{(1)}_{m,n}$ converges to 0.

Let us rewrite

$$\begin{aligned}\widehat{\mathrm{QI}}^{(1)}_{m,n} - \widetilde{\mathrm{QI}}^{(1)}_{m,n} &= \frac{1}{m}\sum_{j=1}^{m} w_j^{(1)} \left(\phi\left(f\left(X'_j, \Theta_j\right)\right) - \psi\left(X'_j\right)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \left(\phi\left(f\left(X'_{j_i^{(1)}}, \Theta_{j_i^{(1)}}\right)\right) - \psi\left(X'_{j_i^{(1)}}\right)\right),\end{aligned}$$

introduce the notation

$$\mathcal{A}_X^+ := \{x \in \mathcal{A}_X : \vartheta(x) > 0\},$$

and denote

$$\begin{aligned}e_1 &:= \frac{1}{n}\sum_{i=1}^{n} \left(\phi(f(X'_{j_i^{(1)}}, \Theta_{j_i^{(1)}})) - \psi(X'_{j_i^{(1)}})\right) \mathbb{1}_{\{X_i \notin \mathcal{A}_X^+\}}, \\ e_2 &:= \frac{1}{n}\sum_{i=1}^{n} \left(\phi(f(X'_{j_i^{(1)}}, \Theta_{j_i^{(1)}})) - \psi(X'_{j_i^{(1)}})\right) \mathbb{1}_{\{X_i \in \mathcal{A}_X^+\}}.\end{aligned}$$

In Step 1 below, we prove that

$$\mathbb{E}\left[|e_1|\right] \xrightarrow[n,m\to+\infty]{} 0,$$

demonstrating at the same time the direct implication of the convergence when $\mathcal{A}_X^+ = \varnothing$. In Step 2, we show that if $\mathcal{A}_X^+ \neq \varnothing$ then $\mathbb{E}[|e_2|]$ does not converge to 0, which implies that in this case, $\widehat{\mathrm{QI}}^{(1)}_{m,n} - \widetilde{\mathrm{QI}}^{(1)}_{m,n}$ does not converge to 0 in L^1.

In both steps, we shall use the following preliminary remark: given a measurable subset $\mathcal{A}$ of $\mathbb{R}^d$, taking the conditional expectation with respect to $(\mathbf{X}_n, \mathbf{X}'_m)$ it is easy to see that for $i \in [\![1, n]\!]$,

$$\mathbb{E}\left[(\phi(f(X'_{j_i^{(1)}}, \Theta_{j_i^{(1)}})) - \psi(X'_{j_i^{(1)}}))\mathbb{1}_{\{X_i \in \mathcal{A}\}}\right] = 0,$$

and for $(i_1, i_2) \in [\![1, n]\!]^2$,

$$\mathbb{E}\left[(\phi(f(X'_{j_{i_1}^{(1)}}, \Theta_{j_{i_1}^{(1)}})) - \psi(X'_{j_{i}^{(1)}{}_1}))\mathbb{1}_{\{X_{i_1}\in\mathcal{A}\}}(\phi(f(X'_{j_{i_2}^{(1)}}, \Theta_{j_{i_2}^{(1)}})) - \psi(X'_{j_{i_2}^{(1)}}))\mathbb{1}_{\{X_{i_2}\in\mathcal{A}\}}\right]$$
$$= \mathbb{E}\left[\mathbb{1}_{\{j_{i_1}^{(1)}=j_{i_2}^{(1)}\}}\vartheta(X'_{j_{i_1}^{(1)}})\mathbb{1}_{\{X_{i_1}\in\mathcal{A},X_{i_2}\in\mathcal{A}\}}\right].$$

Therefore,

$$\mathbb{E}[|e_1|^2] = \frac{1}{n^2}\left(\sum_{i=1}^n \mathbb{E}\left[\vartheta(X'_{j_i^{(1)}})\mathbb{1}_{\{X_i\notin\mathcal{A}_X^+\}}\right] + \sum_{i_1\neq i_2}\mathbb{E}\left[\mathbb{1}_{\{j_{i_1}^{(1)}=j_{i_2}^{(1)}\}}\vartheta(X'_{j_{i_1}^{(1)}})\mathbb{1}_{\{X_{i_1}\notin\mathcal{A}_X^+,X_{i_2}\notin\mathcal{A}_X^+\}}\right]\right)$$
$$= \frac{1}{n}\mathbb{E}\left[\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\notin\mathcal{A}_X^+\}}\right] + \frac{n-1}{n}\mathbb{E}\left[\mathbb{1}_{\{j_1^{(1)}=j_2^{(1)}\}}\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\notin\mathcal{A}_X^+,X_2\notin\mathcal{A}_X^+\}}\right],$$

and a similar expression holds for $\mathbb{E}[|e_2|^2]$.

Step 1. Thanks to the boundedness of ϕ, and thus of ϑ, it is immediate that

$$\frac{1}{n}\mathbb{E}[\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\notin\mathcal{A}_X^+\}}] \xrightarrow[n\to+\infty]{} 0,$$

uniformly in m. Therefore, to show that $\mathbb{E}[|e_1|^2]$ converges to 0, it suffices to prove that

$$\mathbb{E}\left[\mathbb{1}_{\{j_1^{(1)}=j_2^{(1)}\}}\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\notin\mathcal{A}_X^+,X_2\notin\mathcal{A}_X^+\}}\right] \xrightarrow[m\to+\infty]{} 0.$$

In this purpose, let us first write

$$\mathbb{E}\left[\mathbb{1}_{\{j_1^{(1)}=j_2^{(1)}\}}\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\notin\mathcal{A}_X^+,X_2\notin\mathcal{A}_X^+\}}\right] \leq \mathbb{E}\left[\mathbb{1}_{\{\mathrm{NN}_{\mathbf{X}'_m}(X_1)=\mathrm{NN}_{\mathbf{X}'_m}(X_2)\}}\vartheta(\mathrm{NN}_{\mathbf{X}'_m}(X_1))\mathbb{1}_{\{X_1\notin\mathcal{A}_X^+\}}\right],$$

and recall that, by Assumption 3.1 and Lemma 2.2 in [5, Chapter 2], $\mathrm{NN}_{\mathbf{X}'_m}(X_1)$ converges to X_1 and $\mathrm{NN}_{\mathbf{X}'_m}(X_2)$ converges to X_2, almost surely. As a consequence, if $X_1 \in \mathcal{A}_X \setminus \mathcal{A}_X^+$ then $\vartheta(X_1) = 0$ and by the continuity of ϑ and the boundedness of ϕ, the dominated convergence theorem shows that

$$\mathbb{E}\left[\mathbb{1}_{\{X_1\in\mathcal{A}_X\setminus\mathcal{A}_X^+\}}\mathbb{1}_{\{\mathrm{NN}_{\mathbf{X}'_m}(X_1)=\mathrm{NN}_{\mathbf{X}'_m}(X_2)\}}\vartheta(\mathrm{NN}_{\mathbf{X}'_m}(X_1))\right] \xrightarrow[m\to+\infty]{} 0.$$

On the other hand, if $X_1 \notin \mathcal{A}_X$, then almost surely $X_1 \neq X_2$, and therefore $\mathbb{1}_{\{\mathrm{NN}_{\mathbf{X}'_m}(X_1)=\mathrm{NN}_{\mathbf{X}'_m}(X_2)\}}$ converges to 0 almost surely. Using the boundedness of ϕ and the dominated convergence theorem again, we deduce that

$$\mathbb{E}\left[\mathbb{1}_{\{X_1\notin\mathcal{A}_X\}}\mathbb{1}_{\{\mathrm{NN}_{\mathbf{X}'_m}(X_1)=\mathrm{NN}_{\mathbf{X}'_m}(X_2)\}}\vartheta(\mathrm{NN}_{\mathbf{X}'_m}(X_1))\right] \xrightarrow[m\to+\infty]{} 0,$$

which shows that $\mathbb{E}[|e_1|^2]$, and thus $\mathbb{E}[|e_1|]$, converge to 0.

Step 2. Let us now assume that $\mathcal{A}_X^+$ is nonempty and show that e_2 does not converge to 0 in L^1. We shall actually prove that e_2 does not converge to 0 in L^2: since e_2 is bounded then this prevents the convergence from occuring in L^1. From the preliminary remark, we write

$$\mathbb{E}[|e_2|^2] = \frac{1}{n}\mathbb{E}\left[\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\in\mathcal{A}_X^+\}}\right] + \frac{n-1}{n}\mathbb{E}\left[\mathbb{1}_{\{j_1^{(1)}=j_2^{(1)}\}}\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\in\mathcal{A}_X^+,X_2\in\mathcal{A}_X^+\}}\right],$$

and we prove that

$$\liminf_{m\to+\infty}\mathbb{E}\left[\mathbb{1}_{\{j_1^{(1)}=j_2^{(1)}\}}\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\in\mathcal{A}_X^+,X_2\in\mathcal{A}_X^+\}}\right] > 0.$$

Let $x \in \mathcal{A}_X^+$. Obviously,

$$\mathbb{E}\left[\mathbb{1}_{\{j_1^{(1)}=j_2^{(1)}\}}\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1\in\mathcal{A}_X^+,X_2\in\mathcal{A}_X^+\}}\right] \geq \mathbb{E}\left[\mathbb{1}_{\{j_1^{(1)}=j_2^{(1)}\}}\vartheta(X'_{j_1^{(1)}})\mathbb{1}_{\{X_1=X_2=x\}}\right]$$
$$= \mathbb{E}\left[\vartheta(\mathrm{NN}_{\mathbf{X}'_m}(x))\mathbb{1}_{\{X_1=X_2=x\}}\right].$$

By Assumption 3.1 and Lemma 2.2 in [5, Chapter 2] again, $\mathrm{NN}_{\mathbf{X}'_m}(x)$ converges to x almost surely, therefore using the continuity and boundedness assumptions on ϑ, the dominated convergence theorem shows that

$$\mathbb{E}\left[\vartheta(\mathrm{NN}_{\mathbf{X}'_m}(x))\mathbb{1}_{\{X_1=X_2=x\}}\right] \xrightarrow[m\to+\infty]{} \vartheta(x)\mu_X(\{x\})^2 > 0,$$

which completes the proof. □

We now study the estimator $\widehat{\mathrm{QI}}_{m,n}^{(k_m)}$ and show that it is unconditionnally consistent as soon as $k_m \to +\infty$. We provide L^2 convergence rates in Proposition 4.6.

Proposition 4.5 (Consistency in the noisy case)**.** *Assume that*

(i) the function ϕ is bounded,

(ii) the function ψ is globally Lipschitz continuous,

and let the assumptions of Theorem 3.3 hold with $q = 2$. As soon as k_m goes to infinity with m and $k_m/m \to 0$, we have

$$\mathbb{E}\left[\left|\widehat{\mathrm{QI}}_{m,n}^{(k_m)} - \mathrm{QI}\right|^2\right]^{1/2} \xrightarrow[n,m\to+\infty]{} 0.$$

Proof. We decompose the error as

$$\widehat{\mathrm{QI}}_{m,n}^{(k_m)} - \mathrm{QI} = \left(\widehat{\mathrm{QI}}_{m,n}^{(k_m)} - \widetilde{\mathrm{QI}}_{m,n}^{(k_m)}\right) + \left(\widetilde{\mathrm{QI}}_{m,n}^{(k_m)} - \mathrm{QI}\right), \tag{24}$$

with

$$\widetilde{\mathrm{QI}}_{m,n}^{(k_m)} = \frac{1}{m}\sum_{j=1}^{m} w_j^{(k_m)}\psi(X'_j).$$

As ψ is globally Lipschitz continuous and does not depend on Θ, we have

$$\begin{aligned}\mathbb{E}\left[\left(\widetilde{\mathrm{QI}}_{m,n}^{(k_m)} - \mathrm{QI}\right)^2\right]^{1/2} &\leq L\mathbb{E}\left[W_1^2\left(\mu_X, \widehat{\mu}_{\mathbf{X}'_m}^{(k_m)}\right)\right]^{1/2} \\ &\leq 2L\left(\mathbb{E}\left[W_1^2\left(\mu_X, \widehat{\mu}_{\mathbf{X}_n}\right)\right] + \mathbb{E}\left[W_1^2(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{(k_m)})\right]\right)^{1/2}\end{aligned}$$

by Jensen's inequality, with L the Lipschitz constant of ψ. The second term is bounded from above by $\mathbb{E}[W_2^2(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}_{\mathbf{X}'_m}^{(k_m)})]$, which goes to 0 by Theorem 3.3. For the first term, the same arguments as in the proof of Lemma 4.1 show that

$$\mathbb{E}\left[W_1^2\left(\mu_X, \widehat{\mu}_{\mathbf{X}_n}\right)\right] = \mathrm{O}\left(\frac{1}{n} + \mathbb{E}\left[W_1\left(\mu_X, \widehat{\mu}_{\mathbf{X}_n}\right)\right]^2\right).$$

Since $W_1(\mu_X, \widehat{\mu}_{\mathbf{X}_n})$ converges to 0 in probability [31] and the assumption that $\mathbb{E}[|X|^2] < +\infty$ ensures that this sequence is uniformly integrable, we deduce that its expectation converges to 0 [6, Section 5]. Thus, the second part of the right-hand side of (24) converges to 0 in L^2.

Let us consider the first part in the right-hand side of (24). We write the quadratic error

$$\begin{aligned}
&\mathbb{E}\left[\left|\widehat{\mathrm{QI}}_{m,n}^{(k_m)} - \widetilde{\mathrm{QI}}_{m,n}^{(k_m)}\right|^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^m w_j^{(k_m)}(\psi(X_j') - \phi(f(X_j',\Theta_j)))\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{m^2}\sum_{j=1}^m w_j^{(k_m)2}\left(\psi(X_j') - \phi(f(X_j',\Theta_j))\right)^2\right] \\
&+ \mathbb{E}\left[\frac{m-1}{m^2}\sum_{j\neq l} w_j^{(k_m)} w_l^{(k_m)}\left(\psi(X_j') - \phi(f(X_j',\Theta_j))\right)\left(\psi(X_l') - \phi(f(X_l',\Theta_l))\right)\right].
\end{aligned}$$

Using the fact that $\mathbb{E}[w_j^{(k_m)} f(X_j',\Theta_j)|\mathbf{X}_n,\mathbf{X}_m'] = w_j^{(k_m)}\psi(X_j')$ by definition and the independence of the Θ_j, the cross terms vanish. The remaining quadratic term is

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{m^2}\sum_{j=1}^m w_j^{(k_m)2}\left(\psi(X_j') - \phi(f(X_j',\Theta_j))\right)^2\right] \\
&= \frac{1}{m^2}\sum_{j=1}^m \mathbb{E}\left[w_j^{(k_m)2}(\psi(X_j') - \phi(f(X_j',\Theta)))^2\right] \\
&\leq \frac{4}{m^2}\sum_{j=1}^m \mathbb{E}\left[\left(w_j^{(k_m)}\right)^2\right]\|\phi\|_\infty^2.
\end{aligned} \tag{25}$$

We remark that

$$\sum_{j=1}^m \left(w_j^{(k_m)}\right)^2 = \frac{m^2}{n^2 k_m^2}\sum_{i_1,i_2=1}^n \sum_{l_1,l_2=1}^{k_m} \mathbb{1}_{\{j_{i_1}^{(l_1)} = j_{i_2}^{(l_2)}\}}$$

and that for some fixed i_1,i_2 and l_1, there exists exactly one $l_2 \in [\![1,m]\!]$ such that $j_{i_1}^{(l_1)} = j_{i_2}^{(l_2)}$ as $(j_{i_2}^{(l)})_{1\leq l\leq m}$ is a permutation of $[\![1,m]\!]$. Therefore, there exists at most one $l_2 \in [\![1,k_m]\!]$ verifying this property and, consequently,

$$\sum_{j=1}^m \left(w_j^{(k_m)}\right)^2 \leq \frac{m^2}{n^2 k_m^2}\sum_{i_1,i_2=1}^n \sum_{l_1=1}^{k_m} 1 = \frac{m^2}{k_m}.$$

We can then bound the second term by

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^m w_j^{(k_m)}(\psi(X_j') - \phi(f(X_j',\Theta_j)))\right)^2\right]^{1/2} \leq \frac{2}{k_m^{1/2}}\|\phi\|_\infty,$$

which converges to 0 when k_m goes to infinity. □

In order to complement Proposition 4.5 with a rate of convergence, we restart from the decomposition (24). Under the additional assumptions of Corollary 3.9 with $q = 2$, the same arguments as in the proof of Proposition 4.3 yield

$$\mathbb{E}\left[\left(\widetilde{\mathrm{QI}}_{m,n}^{(k_m)} - \mathrm{QI}\right)^2\right]^{1/2} = \mathrm{O}\left(\frac{1}{\sqrt{n}} + \tau_{1,d}(n)\right) + \mathrm{O}\left(\left(\frac{k_m}{m}\right)^{1/d}\right),$$

while we still have

$$\mathbb{E}\left[\left|\widehat{\mathrm{QI}}_{m,n}^{(k_m)} - \widetilde{\mathrm{QI}}_{m,n}^{(k_m)}\right|^2\right]^{1/2} = \mathrm{O}\left(\frac{1}{\sqrt{k_m}}\right)$$

from the proof of Proposition 4.5. As a consequence,

$$\mathbb{E}\left[\left|\mathrm{QI} - \widehat{\mathrm{QI}}_{m,n}^{(k_m)}\right|^2\right]^{1/2} = \mathrm{O}\left(\frac{1}{\sqrt{n}} + \tau_{1,d}(n)\right) + \mathrm{O}\left(\left(\frac{k_m}{m}\right)^{1/d} + \frac{1}{\sqrt{k_m}}\right).$$

Optimizing in k_m, we get the following statement.

Proposition 4.6 (Rates of convergence in the noisy case)**.** *Under the assumptions of Proposition 4.5 and Corollary 3.9 with $q = 2$, we have*

$$\mathbb{E}\left[\left|\mathrm{QI} - \widehat{\mathrm{QI}}_{m,n}^{(k_m)}\right|^2\right]^{1/2} = \mathrm{O}\left(\frac{1}{\sqrt{n}} + \tau_{1,d}(n)\right) + \mathrm{O}\left(\frac{1}{m^{1/(d+2)}}\right)$$

for $k_m \sim m^{2/(d+2)}$.

The loss of convergence order with respect to Proposition 4.3 is similar to the NNR context, in which it deteriorates from the rate $1/d$ in the noiseless case to the rate of $1/(d+2)$ in the noisy case [5, Section 14.6 and Section 15.3].

5. Applications and numerical illustration

We present a reformulation of our results in a standard framework for k-NN regression in Subsection 5.1, and then provide a detailed account of the original motivation of this work by decomposition-based UQ in Subsection 5.2. Last, numerical illustrations of our main results in a simple setting are reported in Subsection 5.3; we refer to [36, Chapter 11] for an application in an industrial context.

5.1. Generalization error of k-NN regression under covariate shift.

In this subsection, we address the k-NN regression problem under covariate shift from the following more standard point of view: the quantity of interest is directly the regression function

$$r(x) := \mathbb{E}\left[Y|X = x\right] = \mathbb{E}\left[f(x, \Theta)\right],$$

and the k-NN estimator of $r(x)$ is defined from the training set by

$$\widehat{r}_m^{(k)}(x) = \frac{1}{k}\sum_{l=1}^{k} Y'_{j^{(l)}(x)},$$

where $j^{(l)}(x)$ denotes the (smallest) index j such that $X'_j = \mathrm{NN}_{\mathbf{X}'_m}^{(l)}(x)$. We are no longer interested in some quantity $\mathbb{E}[\phi(Y)]$ but rather in the *(L^2) generalization error* under covariate shift

$$\mathbb{E}\left[\left|r(X) - \widehat{r}_m^{(k)}(X)\right|^2\right], \qquad X \sim \mu_X.$$

For the sake of simplicity we assume that Y, $r(X)$, etc. take scalar values.

Theorem 5.1 (L^2 generalization error of the k-NN regression under covariate shift)**.** *Let μ_X and $\mu_{X'}$ verify the assumptions of Theorem 3.8 for $q = 2$. Assume in addition that f is Lipschitz continuous in x, uniformly in Θ, and that $\mathrm{Var}(f(x, \Theta)) \leq \sigma^2 < +\infty$ for all $x \in \mathrm{supp}(\mu_{X'})$.*

When the sequence $(k_m)_{m\geq 1}$ satisfies the assumptions of Corollary 3.9, with $k_m \sim m^{2/(d+2)}$, there exists $C \geq 0$ such that

$$\limsup_{m\to+\infty} m^{1/(2+d)} \mathbb{E}\left[\left(r(X) - \widehat{r}_m^{(k_m)}(X)\right)^2\right]^{1/2} \leq \sigma + C\mathbb{E}\left[\frac{1}{p_{X'}(X)^{2/d}}\right]^{1/2}.$$

We retrieve essentially the same orders of convergence as in the case without covariate shift. The quantity $\mathbb{E}[1/p_{X'}(X)^{2/d}]^{1/2}$ seems to be the relevant bound of the loss due to the use of $\mu_{X'}$ instead of μ_X and we expect that the greater this quantity is, the slower the convergence will be.

Proof. The proof is an adaptation of [5, Theorem 14.5], using elements of the proofs of Theorem 3.8 and Corollary 3.9. We can decompose the L^2 error

$$\mathbb{E}\left[\left(r(X) - \widehat{r}_m^{(k_m)}(X)\right)^2\right]^{1/2} \leq \mathbb{E}\left[\left(r(X) - \widetilde{r}_m^{(k_m)}(X)\right)^2\right]^{1/2} + \mathbb{E}\left[\left(\widetilde{r}_m^{(k_m)}(X) - \widehat{r}_m^{(k_m)}(X)\right)^2\right]^{1/2}$$

with

$$\widetilde{r}_m^{(k_m)}(x) = \frac{1}{k_m}\sum_{l=1}^{k_m} \mathbb{E}\left[f(\mathrm{NN}^{(l)}_{\mathbf{X}'_m}(x), \Theta)\right].$$

By Jensen's inequality, the first term can be bounded by

$$\mathbb{E}\left[\left(r(X) - \widetilde{r}_m^{(k_m)}(X)\right)^2\right] \leq L^2\left(\frac{1}{k_m}\sum_{l=1}^{k_m}\mathbb{E}\left[\left(X - \mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X)\right)^2\right]\right),$$

where L is the Lipschitz constant of f, and then following the proof of Corollary 3.9, we get

$$\limsup_{m\to+\infty}\left(\frac{m}{k_m}\right)^{1/d}\mathbb{E}\left[\left(r(X) - \widetilde{r}_m^{(k_m)}(X)\right)^2\right]^{1/2} \leq C\mathbb{E}\left[\frac{1}{p_{X'}(X)^{2/d}}\right]^{1/2},$$

with $C := L(c_{d,2}\Gamma(1+2/d)/v_d^{2/d})^{1/2}$. The second term is bounded by

$$\begin{aligned}
&\mathbb{E}\left[\left(\widetilde{r}_m^{(k_m)}(X) - \widehat{r}_m^{(k_m)}(X)\right)^2\right]^{1/2} \\
&= \frac{1}{k_m}\mathbb{E}\left[\sum_{l=1}^{k_m}\left(f(\mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X), \Theta_{j^{(l)}(X)}) - \mathbb{E}[f(\mathrm{NN}^{(l)}_{\mathbf{X}'_m}(X), \Theta)|X]\right)^2\right]^{1/2} \\
&\leq \frac{1}{k_m^{1/2}}\sigma.
\end{aligned}$$

The optimal rate is $k_m \sim m^{2/(2+d)}$, leading to

$$\limsup_{m\to+\infty} m^{1/(2+d)} \mathbb{E}\left[\left(r(X) - \widehat{r}_m^{(k_m)}(X)\right)^2\right]^{1/2} \leq \sigma + C\mathbb{E}\left[\frac{1}{p_{X'}(X)^{2/d}}\right]^{1/2},$$

which completes the proof. □

5.2. **Application to decomposition-based UQ.** In the UQ context, the relation (1) represents a computer simulation [18, 15]: the random variable X is the *input* of the simulation, the random variable Θ describes the set of its *parameters*, the function f is the *numerical model* and the random variable Y is the *output* of the simulation. The function ϕ involved in the definition (2) of the quantity of interest QI is the *observable*.

The fact that we assume that both X and Θ may be random, but with distinct sources of uncertainty (which is modeled by their statistical independence), comes from the study of uncertainty propagation in complex networks of numerical models [1, 3, 28, 34, 29]. In this context, several computer codes,

representing various disciplines, are connected with each other by the fact that the outputs of certain codes are taken as inputs of other codes. Then f represents one discipline, with 'internal' uncertain parameters Θ whose law is known by the agent in charge of the simulation, and 'external' uncertain parameters X which are the output of possibly several upstream numerical simulations. Independently from our complex system context, assuming that the internal parameter Θ may be random is a standard practice to take aleatoric or epistemic uncertainty into account [15, 22].

If X is deterministic then the computation of a global quantity of interest can be treated by the so-called Collaborative Optimization methods [8, 40] in Multidisciplinary Analysis and Optimization. However, if X is random, a direct Monte Carlo evaluation of QI is often impossible to implement in practice. Indeed, if the number of interacting disciplines is large and each code evaluation is costly, then one cannot wait for a sample $X_1, \dots, X_n$ to be generated by the upstream simulations before starting running one's own simulation. Therefore, decomposition-based UQ methods have been introduced in the literature in order to allow disciplines to run their numerical simulations independently. Basically, these methods work in two phases. In an *offline* phase, each discipline generates it own *synthetic* sample $X'_1, \dots, X'_m$ according to some user-chosen probability measure $\mu_{X'}$ on $\mathbb{R}^d$ (or possibly other designs of experiment). The numerical model f is then evaluated on the sample $(X'_1, \Theta_1), \dots, (X'_m, \Theta_m)$ to obtain a corresponding set of realizations $Y'_1, \dots, Y'_m$. Once actual realizations $X_1, \dots, X_n$ become available in a subsequent *online* phase, they have to be used in combination with the synthetic sample to construct an estimator of QI, but evaluations of the numerical model f are no longer allowed.

We refer to [2, 3, 4] for examples and background on these methods. The k-NN reweighting scheme introduced in the present article is yet another possible approach to this problem. More general non-parametric regression methods, such as Nadaraya–Watson estimators, may also be considered. A more systematic study of such approaches, based on linear reweighting, as well as their generalization to the estimation of quantities of interest defined on a *graph* of numerical models, may be found in [36, Chapter 11] and will be the object of a future publication.

Remark 5.2 (Stochastic simulators)**.** *Our framework is also suited to the situation where Θ does not represent well-identified* parameters*, but must rather be interpreted as the inherent randomness of the numerical model. In the UQ literature, such models are called* stochastic simulator *(see for instance* [42] *and the references therein) and their emulation is closely related with the regression problem addressed in this article, interpreting Θ as a* noise *term.*

Remark 5.3 (A simple example with low-dimensionally supported data)**.** *In the multidisciplinary context introduced above, consider the simple setting in which a random variable X^1 is taken as an input by two distinct disciplines, represented by two numerical models f^1 and f^2. Assume in addition that the output $Y^1 = f^1(X^1)$ of the former is taken as an input by the latter, so that $Y^2 = f^2(X^1, Y^1)$. In the decomposition-based approach, the second discipline has to design a synthetic sample $(X'^1_j, Y'^1_j)_{j \in [\![1,m]\!]}$ without the actual knowledge of f^1. Therefore it is unlikely that this sample be absolutely continuous with respect to the true law of (X^1, Y^1), the support of which lies in the manifold $\{(x, f^1(x))\}$.*

5.3. **Numerical illustrations.** This subsection investigates numerically the influence of the choice of the synthetic distribution $\mu_{X'}$ on the quality of the respective approximations of μ_X by $\widehat{\mu}^{(1)}_{\mathbf{X}'_m}$, and of QI by $\widehat{\mathrm{QI}}^{(k_m)}_{m,n}$.

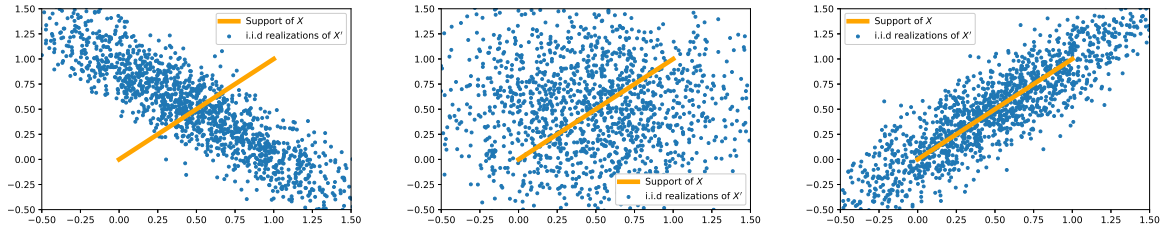
5.3.1. *Influence of $\mu_{X'}$ on the convergence of $\widehat{\mu}^{(1)}_{\mathbf{X}'_m}$.* We investigate how the relationship between μ_X and $\mu_{X'}$ impacts the convergence of $\widehat{\mu}^{(1)}_{\mathbf{X}'_m}$ presented in Subsection 4.1. In this numerical experiment, we set the dimension $d = 2$, choose

$$X = (U, U), \quad U \sim \mathcal{U}([0,1]),$$

and

$$X' \sim \mathcal{N}\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \sigma^2 \begin{pmatrix}1 & s_{\mathrm{corr}}\\ s_{\mathrm{corr}} & 1\end{pmatrix}\right),$$

with $\mu = 0.5$, $\sigma = 0.3$ and various s_{corr} in $(-1, 1)$. Intuitively, the closer s_{corr} is to 1, the closer $\mu_{X'}$ is to μ_X, as illustrated in Figure 1.



(a) $s_{\mathrm{corr}} = -0.9$ (b) $s_{\mathrm{corr}} = 0$ (c) $s_{\mathrm{corr}} = 0.9$

FIGURE 1. Plot of the support of X and 1500 i.i.d realizations of X' for different values of s_{corr}.

As a first 'purely visual' indication of the quality of the approximation of μ_X by $\widehat{\mu}^{(1)}_{\mathbf{X}'_m}$, we plot on Figure 2 the trace of a kernel smoothing of $\widehat{\mu}^{(1)}_{\mathbf{X}'_m}$ on the segment $\{(u,u), u \in [0,1]\}$. We can see that the greater s_{corr} is, the better the reconstruction looks like. From a more quantitative point of view, this observation is confirmed in Figure 3, where we plot the evolution of $\mathbb{E}[W_2^2(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m})]$ as a function of m. We can see that although this quantity converges at the theoretical rate m^{-1}, the multiplicative constant decreases with s_{corr}.

5.3.2. *Influence of $\mu_{X'}$ on the convergence of $\widehat{\mathrm{QI}}^{(k_m)}_{m,n}$.* We now concentrate on the impact on the efficiency of $\widehat{\mathrm{QI}}^{(k_m)}_{m,n}$. We keep the framework of the previous paragraph, and we try to estimate the quantity of interest

$$\mathrm{QI} = \mathbb{E}[\phi(f(X,\Theta))], \quad f((x_1,x_2),\theta) = \sin(2\pi x_1)\sin(2\pi x_2)(1+\theta)$$

with $\Theta \sim \mathcal{U}([-1,1])$ and $\phi(y) = y$. The L^2 error

$$\mathbb{E}\left[\left|\widehat{\mathrm{QI}}^{(k_m)}_{m,n} - \mathrm{QI}\right|^2\right]^{1/2} = \mathbb{E}\left[\left|\widehat{\mathrm{QI}}^{(k_m)}_{m,n} - 0.5\right|^2\right]^{1/2}$$

is computed by Monte Carlo estimation. As highlighted in Figure 4, the closeness of μ_X to $\mu_{X'}$ is an important factor for the efficiency of the estimator.

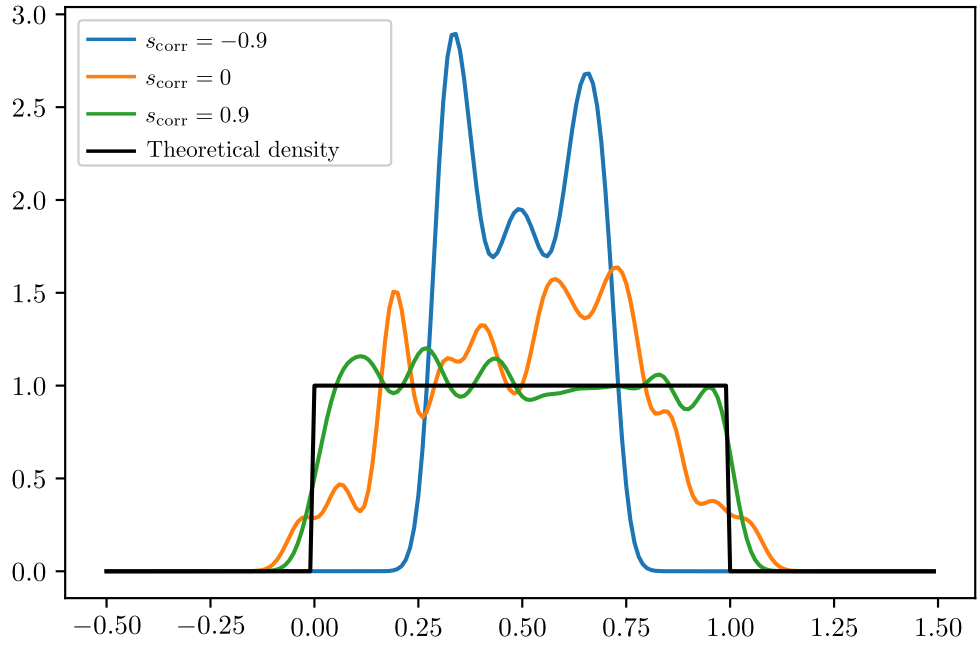


FIGURE 2. Trace of a kernel smoothing of $\widehat{\mu}^{(1)}_{\mathbf{X}'_m}$ on the segment $\{(u,u), u \in [0,1]\}$ for different values of s_{corr}.

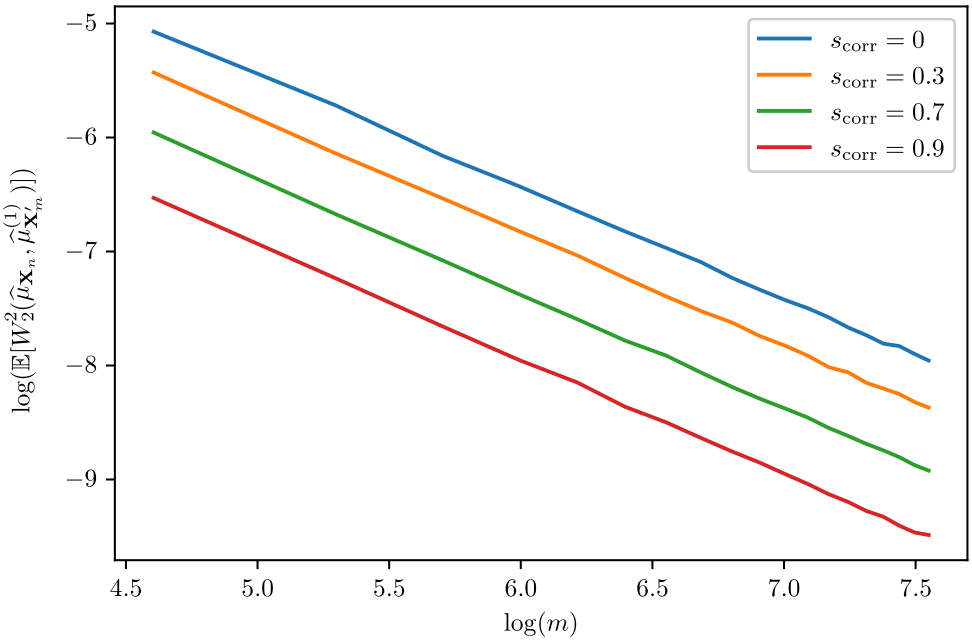


FIGURE 3. Evolution of $\mathbb{E}[W_2^2(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m})]$ with respect to m, for various values of s_{corr}. For each value of m and s_{corr}, the quantity $W_2^2(\widehat{\mu}_{\mathbf{X}_n}, \widehat{\mu}^{(1)}_{\mathbf{X}'_m})$ is computed (with $n = 100$) thanks to the formula (9), and its expectation is estimated through Monte Carlo average over 700 replications. The theoretical rate of $m^{-q/d} = m^{-1}$ is attained for each experiment.

Acknowledgements. This work was motivated by a collaboration with P. Benjamin, F. Mangeant and M. Yagoubi. We also benefited from fruitful discussions with G. Biau and A. Guyader. Last, we thank two anonymous referees for their careful reading of the article, and their numerous suggestions which allowed to greatly improve the presentation of this work.

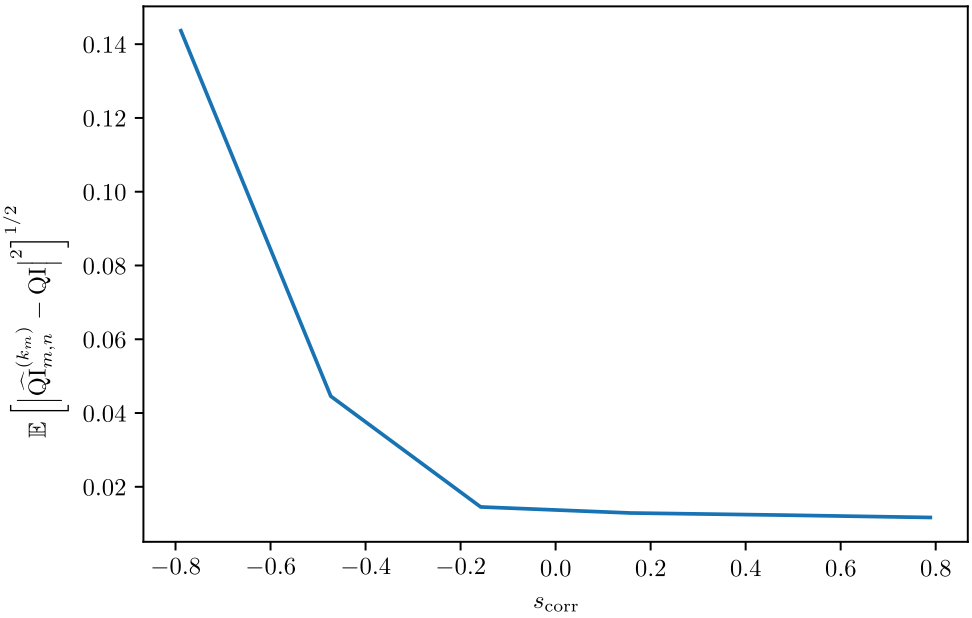


FIGURE 4. Estimation of the L^2 error with respect to s_{corr} for $n = m = 900$ and $k_m = 4$, averaged on 2000 replications.

References

[1] D. Allaire and K. Willcox. Surrogate modeling for uncertainty assessment with application to aviation environmental system models. *AIAA J.*, 48(8):1791–1803, 2010.

[2] S. Amaral, D. Allaire, and K. Willcox. A decomposition approach to uncertainty analysis of multidisciplinary systems. In *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2012.

[3] S. Amaral, D. Allaire, and K. Willcox. A decomposition-based approach to uncertainty analysis of feed-forward multicomponent systems. *Internat. J. Numer. Methods Engrg.*, 100(13):982–1005, 2014.

[4] S. Amaral, D. Allaire, and K. Willcox. Optimal L_2-norm empirical importance weights for the change of probability measure. *Stat. Comput.*, 27(3):625–643, 2017.

[5] G. Biau and L. Devroye. *Lectures on the nearest neighbor method*. Springer Series in the Data Sciences. Springer, Cham, 2015.

[6] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[7] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 2005.

[8] R. D. Braun and I. M. Kroo. Development and application of the collaborative optimization architecture in a multidisciplinary design environment. 1995.

[9] C. Brécheteau and C. Levrard. A k-points-based distance for robust geometric inference. *Bernoulli*, 26(4):3017–3050, 2020.

[10] G. Buttazzo, G. Carlier, and M. Laborde. On the Wasserstein distance between mutually singular measures. *Adv. Calc. Var.*, 13(2):141–154, 2020.

[11] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Found. Comput. Math.*, 11(6):733–751, 2011.

[12] F. Chazal, P. Massart, and B. Michel. Rates of convergence for robust geometric inference. *Electron. J. Stat.*, 10(2):2243–2286, 2016.

[13] G. H. Chen and D. Shah. *Explaining the success of nearest neighbor methods in prediction*. Now Publishers, Inc., 2018.

[14] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *Ann. Statist.*, 25(6):2300–2312, 1997.

[15] E. De Rocquigny, N. Devictor, and S. Tarantola. *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. John Wiley & Sons, 2008.

[16] V. Divol. Measure estimation on manifolds: an optimal transport approach. *Probab. Theory Related Fields*, 2022.

[17] D. Evans, A. J. Jones, and W. M. Schmidt. Asymptotic moments of near-neighbour distance distributions. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 458(2028):2839–2849, 2002.

[18] K.-T. Fang, R. Li, and A. Sudjianto. *Design and modeling for computer experiments*. CRC press, 2005.

[19] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.

[20] S. Gadat, T. Klein, and C. Marteau. Classification in general finite dimensional spaces with the k-nearest neighbor rule. *Ann. Statist.*, 44(3):982–1009, 2016.
[21] N. García Trillos and D. Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canad. J. Math.*, 67(6):1358–1383, 2015.
[22] R. Ghanem, D. Higdon, and H. Owhadi. *Handbook of uncertainty quantification*, volume 6. Springer, 2017.
[23] S. Graf and H. Luschgy. *Foundations of quantization for probability distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000.
[24] B. Kloeckner. Approximation by finitely supported measures. *ESAIM Control Optim. Calc. Var.*, 18(2):343–359, 2012.
[25] M. Kohler, A. Krzyzak, and H. Walk. Rates of convergence for partitioning and nearest neighbor regression estimates with unbounded data. *J. Multivariate Anal.*, 97(2):311–323, 2006.
[26] E. Liitiäinen, A. Lendasse, and F. Corona. Bounds on the mean power-weighted nearest neighbour distance. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 464(2097):2293–2301, 2008.
[27] M. Loog. Nearest neighbor-based importance weighting. In *2012 IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6. IEEE, 2012.
[28] S. Marque-Pucheu, G. Perrin, and J. Garnier. Efficient sequential experimental design for surrogate modeling of nested codes. *ESAIM Probab. Stat.*, 23:245–270, 2019.
[29] D. Ming and S. Guillas. Integrated emulators for systems of computer models. Preprint arXiv:1912.09468.
[30] G. Pagès. Introduction to vector quantization and its applications for numerics. In *CEMRACS 2013—modelling and simulation of complex systems: stochastic and deterministic approaches*, volume 48 of *ESAIM Proc. Surveys*, pages 29–79. EDP Sci., Les Ulis, 2015.
[31] V. M. Panaretos and Y. Zemel. Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.*, 6:405–431, 2019.
[32] M. D. Penrose and J. E. Yukich. Laws of large numbers and nearest neighbor distances. In *Advances in directional and linear statistics*, pages 189–199. Physica-Verlag/Springer, Heidelberg, 2011.
[33] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence. *Dataset shift in machine learning*. The MIT Press, 2009.
[34] F. Sanson, O. Le Maitre, and P. M. Congedo. Systems of Gaussian process models for directed chains of solvers. *Comput. Methods Appl. Mech. Engrg.*, 352:32–55, 2019.
[35] M. Sugiyama, T. Suzuki, and T. Kanamori. *Density ratio estimation in machine learning*. Cambridge University Press, Cambridge, 2012. With a foreword by Thomas G. Dietterich.
[36] A. Touboul. *Model of margin, margin sensitivity analysis and uncertainty quantification in graphs of functions in complex industrial systems.* PhD thesis, École des Ponts ParisTech, 2021.
[37] A. B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *Ann. Statist.*, 32(1):135–166, 2004.
[38] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
[39] J. Weed and F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.
[40] W. Yao, X. Chen, W. Luo, M. van Tooren, and J. Guo. Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Prog. Aerosp. Sci.*, 47(6):450–479, 2011.
[41] P. L. Zador. *Development and evaluation of procedures for quantizing multivariate distributions*. ProQuest LLC, Ann Arbor, MI, 1964. Thesis (Ph.D.)–Stanford University.
[42] X. Zhu and B. Sudret. Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. *Reliab. Eng. Syst. Safe.*, 214:107815, 2021.

Julien Reygner
CERMICS, Ecole des Ponts, Marne-la-Vallée, France
Email address: julien.reygner@enpc.fr

Adrien Touboul
CERMICS, Ecole des Ponts, Marne-la-Vallée, France
IRT SystemX, Paris-Saclay, France
Email address: adrien.touboul@enpc.fr