Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2023, Vol. 59, No. 4, 1778–1812
https://doi.org/10.1214/23-AIHP1389
This research was funded, in whole or in part, by [European Research Council, 817257]. A CC BY 4.0 license is applied to this article arising from this submission, in accordance with the grant's open access conditions

ANNALES
DE L'INSTITUT
HENRI
POINCARÉ
PROBABILITÉS
ET STATISTIQUES

Lipschitz continuity of probability kernels in the optimal transport framework

E. Dolera[1,a] and E. Mainini[2,b]

[1] *Department of Mathematics, University of Pavia, Pavia, 27100, Italy,* [a]*emanuele.dolera@unipv.it*
[2] *Department of Machanical Engineering, University of Genova, Genova, 16145, Italy,* [b]*edoardo.mainini@unige.it*

Received 14 July 2022; accepted 28 March 2023

Abstract. In Bayesian statistics, a continuity property of the posterior distribution with respect to the observable variable is crucial as it expresses well-posedness, i.e., stability with respect to errors in the measurement of data. Essentially, this requires analyzing the continuity of a probability kernel or, equivalently, of a conditional probability distribution with respect to the conditioning variable.

Here, we tackle this problem from a theoretical point of view. Let $(\mathbb{X}, \mathrm{d}_{\mathbb{X}})$ be a metric space, and let $\mathscr{B}(\mathbb{R}^d)$ denote the Borel σ-algebra on $\mathbb{R}^d$. Let $\pi(\cdot|\cdot) : \mathscr{B}(\mathbb{R}^d) \times \mathbb{X} \to [0, 1]$ be a dominated probability kernel, i.e. of the form $\pi(\mathrm{d}\theta|x) = g(x, \theta)\pi(\mathrm{d}\theta)$ for some suitable function $g : \mathbb{X} \times \mathbb{R}^d \to [0, +\infty)$. We provide general conditions ensuring the Lipschitz continuity of the mapping $\mathbb{X} \ni x \mapsto \pi(\cdot|x) \in \mathcal{P}(\mathbb{R}^d)$ when the space of probability measures $\mathcal{P}(\mathbb{R}^d)$ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ is endowed with a metric arising within the optimal transport framework, such as a Wasserstein metric. In particular, we prove explicit upper bounds for the Lipschitz constant in terms of Fisher-information functionals and weighted Poincaré constants, obtained by exploiting the dynamic formulation of the optimal transport.

Finally, we give some illustrations on noteworthy classes of probability kernels, and we apply the main results to improve on some open questions in Bayesian statistics, dealing with the approximation of posterior distributions by mixtures and posterior consistency.

Résumé. En statistique bayésienne, une propriété de continuité de la distribution a posteriori par rapport à la variable observée est cruciale puisque'elle exprime le caractère bien posé du problème, c'est-à-dire la stabilité par rapport aux erreurs de mesure dans les données. Cela nécessite essentiellement d'analyser la continuité d'un noyau de probabilité ou, de manière équivalente, d'une distribution de probabilité conditionnelle par rapport à la variable de conditionnement.

Ici, nous abordons ce problème d'un point de vue théorique. Soit $(\mathbb{X}, \mathrm{d}_{\mathbb{X}})$ un espace métrique, et soit $\mathscr{B}(\mathbb{R}^d)$ la tribu borélienne sur $\mathbb{R}^d$. Soit $\pi(\cdot|\cdot) : \mathscr{B}(\mathbb{R}^d) \times \mathbb{X} \to [0, 1]$ un noyau de probabilité dominé, c'est-à-dire de la forme $\pi(\mathrm{d}\theta|x) = g(x, \theta)\pi(\mathrm{d}\theta)$ pour une fonction appropriée $g : \mathbb{X} \to [0, +\infty)$. Nous fournissons des conditions générales assurant la continuité lipschitzienne de l'application $x \in \mathbb{X} \mapsto \mathcal{P}(\mathbb{R}^d)$ lorsque que l'espace des mesures de probabilités $\mathcal{P}(\mathbb{R}^d)$ sur $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ est muni d'une métrique issue d'un cadre de transport optimal, telle qu'une métrique de Wasserstein. En particulier, nous prouvons des bornes supérieures explicites pour la constante de Lipschitz en termes de fonctionnelles d'information de Fisher et de constantes de Poincaré pondérées, obtenues en exploitant la formulation dynamique du transport optimal.

Enfin, nous donnons quelques illustrations sur des classes remarquables de noyaux de probabilité, et nous appliquons nos résultats principaux pour améliorer certaines questions ouvertes en statistique bayésienne, traitant de l'approximation de distributions a posteriori par des mélanges et la consistance a posteriori.

MSC2020 subject classifications: Primary 60F10; secondary 62F15

Keywords: Bayes formula; Bayesian consistency; Benamou–Brenier formula; Probability Kernel; Optimal transport; Wasserstein distance

1. Introduction

1.1. *Formulation of the problem and main contributions*

Several problems in probability and statistics involve mappings of the form $x \mapsto \pi(\cdot|x)$, where $\pi(\cdot|\cdot) : \mathscr{B}(\mathbb{R}^d) \times \mathbb{X} \to [0, 1]$ is a *probability kernel*. In general, $\mathbb{X}$ is a metric space endowed with distance $\mathrm{d}_{\mathbb{X}}$ and Borel σ-algebra $\mathscr{X}$, while $\mathscr{B}(\mathbb{R}^d)$ stands for the usual Borel σ-algebra on $\mathbb{R}^d$. Being any probability kernel $\pi(\cdot|\cdot)$ conceivable as a mapping from $\mathbb{X}$

into the space $\mathcal{P}(\mathbb{R}^d)$ of all probability measures (p.m.'s) on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, our main goal is to provide general conditions for getting a global form of Lipschitz continuity, namely

$$\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}\big(\pi(\cdot|x_1), \pi(\cdot|x_2)\big) \le L\, \mathrm{d}_{\mathbb{X}}(x_1, x_2) \quad \forall x_1, x_2 \in \mathbb{X}, \tag{1.1}$$

where $\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}$ is a suitable distance on $\mathcal{P}(\mathbb{R}^d)$ and $L \ge 0$. Of course, the problem is strongly influenced by the choice of the distance $\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}$ which, at least at the level of abstract theory, can be done in several ways. See, e.g., the review [50]. Here, we will focus only on the *p-Wasserstein distance* $\mathcal{W}_p$, $p \ge 1$, and the *total variational distance* d_{TV} because of their mathematical tractability, their clever conception as minimal transport distances, and their relationships with other probability metrics. For the sake of clarity, we recall that

$$\mathrm{d}_{\mathrm{TV}}(\mu, \nu) := \sup_{A \in \mathscr{B}(\mathbb{R}^d)} \big|\mu(A) - \nu(A)\big| \quad \forall \mu, \nu \in \mathcal{P}\big(\mathbb{R}^d\big),$$

$$\mathcal{W}_p(\mu, \nu) := \inf_{\eta \in \mathcal{F}(\mu,\nu)} \left(\int_{\mathbb{R}^{2d}} |\theta_1 - \theta_2|^p \eta(\mathrm{d}\theta_1 \mathrm{d}\theta_2) \right)^{1/p} \quad \forall \mu, \nu \in \mathcal{P}_p\big(\mathbb{R}^d\big),$$

where $\mathcal{P}_p(\mathbb{R}^d) := \{\zeta \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |\theta|^p \zeta(\mathrm{d}\theta) < +\infty\}$ and $\mathcal{F}(\mu, \nu)$ denotes the class of all p.m.'s on $(\mathbb{R}^{2d}, \mathscr{B}(\mathbb{R}^{2d}))$ with first marginal μ and second marginal ν. See, e.g., [5,82,83] for further information about the p-Wasserstein distance.

Our main results are concerned with dominated kernels of the form

$$\pi(B|x) := \int_B g(x, \theta)\pi(\mathrm{d}\theta) \quad \forall B \in \mathscr{B}\big(\mathbb{R}^d\big), \forall x \in \mathbb{X}, \tag{1.2}$$

with some measurable, non-negative function g and some measure π on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. In this setting, we provide novel contributions in different directions. First, we formulate a general theory aimed at solving (1.1), with emphasis on estimates for the Lipschitz constant L. See Theorem 2.2 and its extensions in Section 4. Second, we illustrate the new methods on some well-known classes of probability kernels, such as exponential families (see Section 2.2) and certain truncation families (see Section 4.3). Third, we show the usefulness of estimate (1.1) for the solution of other allied questions, mainly of statistical nature (see Section 3). We emphasize the following strength points of our theory: the generality of the kernels under consideration, which are not constrained to belong to specific classes; the estimates for the constant L given in terms of some well-known functionals involving g and π; the focus on the 2-Wasserstein distance, for which we will take advantage of the dynamic formulation, recalled in Section 4.1; the inclusion of the non-standard case of kernels with a support that varies with x (see Section 4.2).

1.2. *Basic motivations from probability and Bayesian statistics*

A basic motivation for the analysis of a property like (1.1) comes from the theory of (regular) conditional distributions and its applications. In fact, probability kernels arise naturally in connection with the *disintegration problem*, within the abstract measure-theoretic formulation due to Kolmogorov. See, e.g., Theorems 6.3 and 6.4 in [57], and Chapters 1-5 of [73] for an overview. For clarity, we recall the notion of disintegration, with the same notation of Section 1.1: given a random vector (X, Z) on a probability space $(\Omega, \mathscr{A}, \mathsf{P})$ with values in $\mathbb{X} \times \mathbb{R}^d$, we say that a probability kernel $\pi(\cdot|\cdot) : \mathscr{B}(\mathbb{R}^d) \times \mathbb{X} \to [0, 1]$ solves the disintegration problem if $\mathsf{E}[\pi(B|X)\mathbb{1}_A(X)] = \mathsf{P}[X \in A, Z \in B]$ holds for any $B \in \mathscr{B}(\mathbb{R}^d)$ and any $A \in \mathscr{X}$, where $\mathbb{1}_A$ denotes the indicator function. The well-known issue of non-uniqueness of solutions to the disintegration problem (in the sense that if $\pi_1(\cdot|\cdot)$ is a solution, then $\pi_2(\cdot|\cdot)$ is also a solution as soon as $\mathsf{P}[\pi_1(\cdot|X) \ne \pi_2(\cdot|X)] = 0$) introduces a remarkable gap between theory and practice, since it entails that conditional probabilities of the form $\mathsf{P}[\cdot|X = x]$ are in general meaningless for a single $x \in \mathbb{X}$ such that $\mathsf{P}[X = x] = 0$. See the discussion about the so-called *Borel paradox* in [73]. However, the necessity of pointwise evaluations usually emerges in Bayesian inference (see [25] and the reference therein), statistical mechanics (see e.g. [60]) and theory of stochastic processes (see e.g. [59]), where x stands for some really observed datum and the observer would like to evaluate a conditional probability exactly at x. This foundational mismatch could be overcome by introducing suitable additional conditions that grant uniqueness in the disintegration problem: in fact, we recall that, if (the distribution of) X has full support in $\mathbb{X}$, then there exists at most one probability kernel $\pi(\cdot|\cdot)$ satisfying both the disintegration and the property that $\mathbb{X} \ni x \mapsto \pi(\cdot|x) \in \mathcal{P}(\mathbb{R}^d)$ is continuous with respect to the topology of weak convergence on $\mathcal{P}(\mathbb{R}^d)$. The existence of such a continuous representative, under additional conditions on the joint distribution of (X, Z), was first analyzed in [85]. See also Chapter 9 of [80]. In this respect, a stronger form of continuity like (1.1) expresses a quantitative stability of conditional distributions with respect to small deviations of the observed point, in analogy with the classical notion of well-posedness introduced by Hadamard. However, a general formalization seems still lacking, and deserves deeper investigations.

A specific situation of interest arises in Bayesian statistics in the case that the joint distribution of (X, Z) turns out to be absolutely continuous with respect to a product measure, say $\lambda \otimes \pi$, on $(\mathbb{X} \times \mathbb{R}^d, \mathscr{X} \otimes \mathscr{B}(\mathbb{R}^d))$, with density $f : \mathbb{X} \times \mathbb{R}^d \to [0, +\infty)$. When a (jointly) continuous density f is assigned as starting point of the analysis, the conditional distribution of Z given $X = x$ emerges more naturally from the well-known *Bayes formula*, rather than a disintegration. Precisely, $\mathsf{P}[Z \in \cdot | X = x]$ is given by a kernel of the form (1.2) with

$$g(x, \theta) = \frac{f(x, \theta)}{\int_{\mathbb{R}^d} f(x, \tau)\pi(\mathrm{d}\tau)} \tag{1.3}$$

for any $x \in \mathbb{X}$ such that $\int_{\mathbb{R}^d} f(x, \tau)\pi(\mathrm{d}\tau) > 0$. In this framework, very basic results aimed at proving a local form of (1.1) are contained in our recent paper [35], which is confined to the choice of the total variation distance. In the present paper, we will improve on the results of [35] by relaxing the regularity assumptions, by providing global Lipschitz continuity, and most importantly by considering the Wasserstein distance. Concerning other quantitative estimates like (1.1), the literature is relatively scant. A fairly general approach can be found in the work [77] by A.M. Stuart, who minted the expression *Bayesian well-posedness* for a local version of (1.1). See Section 4.2 of [77]. See also the discussion about well-posedness in [61].

Another strong motivation from Bayesian inference is the following. Let us consider again the evaluation of the conditional probability $\mathsf{P}[Z \in \cdot | X = x]$. Besides disposing of a specific datum $x \in \mathbb{X}$, we assume the presence of some noise in the process of observation. This leads us to interpret x as a realization of $\varphi_\varepsilon(X)$ rather than of X itself, where $\varphi_\varepsilon : \mathbb{X} \to \mathbb{X}$ is some random perturbation of the identity map, stochastically independent of (X, Z). If we dispose of some *apriori* bound (pointwise or in the mean) on the discrepancy between φ_ε and the identity map, we could exploit a property like (1.1) to get a bound on the discrepancy between the conditional distributions $\mathsf{P}[Z \in \cdot | X = x]$ and $\mathsf{P}[Z \in \cdot | \varphi_\varepsilon(X) = x]$. That is, (1.1) highlights the impact of the perturbation of the data in inference. This remark is of some relevance in the recent studies on *differential privacy*. See, e.g., [14,58].

1.3. *Further motivations and applications*

We present a short list of problems that further motivate our analysis and represent the main applications of our theory. We shall provide new explicit solutions to such problems in Section 3, by stressing the key role of (1.1). We also mention some related works in the literature, that often make use a property like of (1.1) only as a technical tool.

(a) *Bayesian well-posedness*. In the same spirit of [77], by Bayesian well-posedness we mean the validity of a local version of (1.1) along with (1.2)-(1.3). This notion have been investigated in the context of Bayesian inverse problems in [29,30,56,61,77,78,81]. Due to their specific focus, these papers only deal with kernels arising from linear regression problems, which are in exponential form. In Sections 2.2 and 3.1 we also analyze Bayesian well-posedness with exponential kernels and, by applying our main results from Section 2, we provide new estimates for the Lipschitz constant. We also consider the customary situation of an inference process with multiple exchangeable observations. Other new results on Bayesian well-posedness will be given in Section 4.3 where we analyze Pareto-like statistical models.

(b) *Approximation of posterior distributions by mixtures*. This problem arises in Bayesian inference when the posterior is not expressible in closed form. To carry out the inferential procedures, a possible strategy is to approximate the prior by a mixture of conjugate prior (conjugation being referred to the statistical model), leading to an approximated posterior which is again in the form of a mixture. Here, property (1.1) yields a bound for the error in approximating the posterior, besides a more precise characterization of the posterior weights. See Section 3.2. See also [84] for developments in parametric settings, [74] for the nonparametric approach, and [76] for density estimation. In particular, Proposition 2 of [74] is an evident application of (1.1).

(c) *Bayesian consistency*. The foundational topic of frequency validation of Bayesian procedures (see [31] and [47, Chapter 6]) can be rewritten as an approximation problem between posterior distributions. See Section 3.3 along with our recent contributions [22,33,34], where (1.1) is at the core of the main argument.

Finally, we foresee a number of other interesting applications that, for reason of space, are not developed in this paper. Thus, we just mention the field of: *Bayesian robustness* (see [67]); *Bayesian deconvolution and empirical Bayes methods* (see [37]); *theory of computability* (see [1]). Hopefully, a general theory of Bayesian well-posedness could bring novel contributions also to these fields.

2. Main results

2.1. *Lipschitz estimates in terms of* $\mathrm{d_{TV}}$, $\mathcal{W}_1$ *and* $\mathcal{W}_2$

The results of this subsection are concerned with kernels of the form (1.2) which fulfill the following

Assumptions 2.1. Let π be a p.m. on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that $\operatorname{supp}(\pi) = \overline{\Theta}$ for some (nonempty) connected open set $\Theta \subseteq \mathbb{R}^d$, and $\pi(\partial\Theta) = 0$. Let $\mathbb{X}$ be a convex open subset of $\mathbb{R}^m$, endowed with the reference σ-algebra $\mathscr{X}$ of all Lebesgue-measurable subsets of $\mathbb{X}$. Finally, let the function g be an element of $L^1_{\mathcal{L}^m\otimes\pi}(\mathbb{X}\times\Theta)$ with $\int_\Theta g(x,\theta)\pi(\mathrm{d}\theta) = 1$ for all $x \in \mathbb{X}$, where $\mathcal{L}^m$ denotes the m-dimensional Lebesgue measure.

In the main theorem, we will also assume that $g \in L^1_\pi(\Theta; W^{1,1}_{\mathrm{loc}}(\mathbb{X}))$, meaning that the distributional gradient of the mapping $x \mapsto g(x,\theta)$ (denoted by ∇_x) belongs to $L^1_{\mathrm{loc}}(\mathbb{X})$ for π-a.e. $\theta \in \Theta$ and that $\nabla_x g \in L^1_{\mathcal{L}^m\otimes\pi}(\tilde{\mathbb{X}}\times\Theta)$ for any open set $\tilde{\mathbb{X}}$ compactly contained in $\mathbb{X}$. In such a case, if $g(x,\cdot) > 0$ for π-a.e. $\theta \in \Theta$, we define the *Fisher functional* of g relative to π as

$$\mathcal{J}_\pi\big[g(x,\cdot)\big] := \left(\int_\Theta \frac{|\nabla_x g(x,\theta)|^2}{g(x,\theta)}\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}. \tag{2.1}$$

Another key assumption of the theory will be the validity of the so-called *weighted Poincaré-Wirtinger inequalities*. We say that a Radon measure μ on $(\Theta, \mathscr{T})$ satisfies a weighted Poincaré-Wirtinger inequality of order $q \in [1,+\infty)$ if a constant $\mathcal{C}_q$ exists such that

$$\inf_{a\in\mathbb{R}}\left(\int_\Theta \big|\psi(\theta) - a\big|^q \mu(\mathrm{d}\theta)\right)^{\frac{1}{q}} \le \mathcal{C}_q\left(\int_\Theta \big|\nabla\psi(\theta)\big|^q \mu(\mathrm{d}\theta)\right)^{\frac{1}{q}} \tag{2.2}$$

holds for every $\psi \in C^1_c(\overline{\Theta})$. Here, $\psi \in C^1_c(\overline{\Theta})$ means that ψ is the restriction to $\overline{\Theta}$ of a C^1 compactly supported function on $\mathbb{R}^d$. We denote by $\mathcal{C}_q[\mu]$ the best constant in such inequality and we put $\mathcal{C}[\mu] := \mathcal{C}_2[\mu]$. Further details are contained in Section A.3.

We also consider (unweighted) Sobolev–Poincaré inequalities. Let either $1 \le p < d$ or $1 = p = d$. We say that Θ satisfies a Sobolev–Poincaré inequality of order p if a constant $\mathcal{S}_p$ exists such that

$$\inf_{a\in\mathbb{R}} \|\zeta - a\|_{L^{p^*}(\Theta)} \le \mathcal{S}_p\|\nabla\zeta\|_{L^p(\Theta)} \quad \text{for any } \zeta \in C^1_c(\overline{\Theta}), \tag{2.3}$$

where $p^* = \frac{dp}{d-p}$ if $1 \le p < d$ and $p^* = +\infty$ if $p = d = 1$. We denote by $\mathcal{S}_p(\Theta)$ the corresponding best constant.

Theorem 2.2. *For a given kernel $\pi(\cdot|\cdot)$ in the form (1.2), let Assumptions 2.1 be in force. Let also $g \in L^1_\pi(\Theta; W^{1,1}_{\mathrm{loc}}(\mathbb{X}))$. When $\mathcal{W}_p$ is involved, assume further that $\int_\Theta |\theta|^p\pi(\mathrm{d}\theta|x) < +\infty$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. The following statements hold.*

(i) *Suppose that*

$$K := \operatorname*{ess\,sup}_{x\in\mathbb{X}}\big\|\nabla_x g(x,\cdot)\big\|_{L^1_\pi(\Theta)} < +\infty.$$

Then, there exists a $\mathrm{d_{TV}}$-Lipschitz version of $\pi(\cdot|\cdot)$, satisfying (1.1) with $L = K/2$.

(ii) *For $1 < p \le +\infty$ and $q = \frac{p}{p-1}$, suppose that*

$$K := \pi(\Theta)^{1/q}\mathcal{C}_q[\pi]\operatorname*{ess\,sup}_{x\in\mathbb{X}}\big\|\nabla_x g(x,\cdot)\big\|_{L^p_\pi(\Theta)} < +\infty.$$

Then, there exists a $\mathcal{W}_1$-Lipschitz version of $\pi(\cdot|\cdot)$, satisfying (1.1) with $L = K$.

(iii) *If $g > 0$ $(\mathcal{L}^m \otimes \pi)$-a.e. and*

$$K := \operatorname*{ess\,sup}_{x\in\mathbb{X}}\mathcal{C}\big[g(x,\cdot)\pi\big]\mathcal{J}_\pi\big[g(x,\cdot)\big] < +\infty,$$

then there exists a $\mathcal{W}_2$-Lipschitz version of $\pi(\cdot|\cdot)$, satisfying (1.1) with $L = K$.

(iv) *If* $g > 0$ $(\mathcal{L}^m \otimes \pi)$*-a.e.*, $\pi = \mathcal{L}^d \llcorner \Theta$ *and*

$$K := \mathcal{S}_p(\Theta) \operatorname*{ess\,sup}_{x\in\mathbb{X}} \left\| \frac{1}{g(x,\cdot)} \right\|^{1/2}_{L^{\frac{p}{2-p}}(\Theta)} \left\| \nabla_x g(x,\cdot) \right\|_{L^{\frac{r}{r-1}}(\Theta)} < +\infty,$$

then there exists a $\mathcal{W}_2$*-Lipschitz version of* $\pi(\cdot|\cdot)$*, satisfying* (1.1) *with* $L = K$.

We notice that the assumption $g > 0$ $(\mathcal{L}^m \otimes \pi)$-a.e., made in points (iii)–(iv), entails that support of the p.m. $\pi(\cdot|x)$ coincides with $\overline{\Theta}$ for $\mathcal{L}^m$-a.e. x in $\mathbb{X}$. We will refer to this fact in the sequel by saying we are in the case of *fixed domains*. In Section 4.2 we will provide other new results that generalize points (iii)–(iv) to the situation of *moving domains*, meaning that the support of $\pi(\cdot|x)$ is allowed to vary smoothly with x. Concerning the first two points of Theorem 2.2, we remark that the assumptions of point (ii) imply the ones of point (i), which is formally the limit case $p = 1$ of point (ii). On the other hand, if Θ is bounded, the elementary inequality $\mathcal{W}_1 \le 2\mathrm{diam}(\Theta)\mathrm{d}_{\mathrm{TV}}$ allows to deduce an estimate for the $\mathcal{W}_1$ distance under the assumptions of point (i). Moreover, we notice that point (iv) holds whenever Θ is a domain for which the Sobolev–Poincaré inequality (2.3) is satisfied. Therefore, point (iv) applies for instance if Θ is the whole of $\mathbb{R}^d$ (and $\mathcal{S}_p(\mathbb{R}^d)$ is explicit, see [7,79]) or if Θ is a $W^{1,p}$ extension domain with $\mathcal{L}^d(\Theta) < +\infty$ (see, e.g., [62, Chapter 12]). More generally, it applies if Θ is a John domain (see [21,26,55]), including the half space and domains with compact Lipschitz boundary. If $d = 1$, (2.3) holds on any interval $\Theta \subseteq \mathbb{R}$ with $\mathcal{S}_1(\Theta) = 1$.

We conclude with a brief discussion about the best constant in the weighted Poincaré-Wirtinger inequality (2.2). The most classical Poincaré inequalities hold by taking μ to be the d-dimensional Lebesgue measure on a bounded domain Θ with Lipschitz boundary and $q = 2$: the reciprocal square of $\mathcal{C}[\mathcal{L}^d \llcorner \Theta]$ is the first nontrivial eigenvalue of the Neumann Laplacian on Θ. If μ is the d-dimensional Lebesgue measure on a bounded convex set $\Theta \subset \mathbb{R}^d$, the classical result by Payne and Weinberger [70] shows that $\mathcal{C}[\mathcal{L}^d \llcorner \Theta]$ is proportional to the diameter of Θ, see also [2,12,38] for $q \neq 2$. Explicit estimates for star-shaped domains are found in [42]. According to the Bakry–Emery condition in the Euclidean setting (see [10]), if μ is a p.m. on a convex set $\Theta \subseteq \mathbb{R}^d$ and $V \in C^2(\Theta)$ exists such that

$$\mu(\mathrm{d}\theta) = e^{-V(\theta)}\,\mathrm{d}\theta, \qquad \langle \mathrm{Hess}[V]\xi, \xi \rangle \ge \alpha > 0 \quad \text{in } \Theta \text{ for any } \xi \in \mathbb{R}^d, \tag{2.4}$$

then (2.2) holds with $\mathcal{C}[\mu] \le 1/\sqrt{\alpha}$. See for instance [66] or [6, Chapitre 5], see also [24,65]. On the other hand, it is shown in [44] that for any log-concave measure μ on a bounded convex domain Θ of $\mathbb{R}^d$ (i.e., for any convex V), the constant $\mathcal{C}[\mu]$ can be bounded explicitly by $\mathrm{diam}(\Theta)/\pi$. Therefore, the Poincaré best constant can be improved by the presence of a log-concave weight with $\alpha > 0$ (the unweighted case corresponding here to $V = 0$). Let us also mention the result by Bobkov [16] which allows to estimate the Poincaré constant of a log-concave measure μ on $\mathbb{R}^d$ in terms of the variance, i.e., $\mathcal{C}[\mu] \le 12\sqrt{3}(\int_{\mathbb{R}^d} |\theta - \overline{\mu}|^2 \mu(\mathrm{d}\theta))^{1/2}$ with $\overline{\mu} := \int_{\mathbb{R}^d} \theta \mu(\mathrm{d}\theta)$. The fundamental Bakry–Emery citerion admits other generalizations. For instance, (2.2) holds on $\mathbb{R}^d$ if the condition $\frac{1}{2}|\nabla V(\theta)|^2 - \Delta V(\theta) \ge c > 0$ is satisfied for any large enough $|\theta|$, see for instance [8,9]. Different results are also available for measures of the form $\mu = e^{-V}\nu$, where ν itself satisfies (2.2), the most simple instance being the Holley–Stroock [54] perturbation principle $\mathcal{C}^2[\mu] \le \exp\{\sup V - \inf V\}\mathcal{C}^2[\nu]$. See e.g. [6, Théorème 3.4.1] or [66]. Further statements in this direction are contained in [23, Proposition 4.1], where it is assumed that μ satisfies the stronger log-Sobolev inequality.

2.2. *A remarkable example*: *Exponential models*

A useful rephrasing of the main results from Theorem 2.2 can be obtained, in Bayesian statistical inference, by considering a statistical model in the following form:

$$f(x|\theta) = e^{\Phi(x,\theta)}h(x), \qquad g(x,\theta) = \frac{f(x|\theta)}{\rho(x)} = \frac{e^{\Phi(x,\theta)}}{\int_\Theta e^{\Phi(x,\tau)}\pi(\mathrm{d}\tau)} \tag{2.5}$$

for some measurable functions $h : \mathbb{X} \to (0, +\infty)$ and $\Phi : \mathbb{X} \times \Theta \to \mathbb{R}$. Here, π denotes the prior probability measure and $\rho(x) = h(x)\int_\Theta e^{\Phi(x,\theta)}\pi(\mathrm{d}\theta) > 0$ for any $x \in \mathbb{X}$. The function g in (2.5) is therefore obtained by applying the Bayes formula. Under the formulation (2.5), the Fisher functional $\mathcal{J}_\pi[g]$ defined in (2.1) can be formally rewritten as

$$\mathcal{J}_\pi\big[g(z,\cdot)\big] = \left(\int_\Theta \big|\Psi(z,\theta)\big|^2 g(z,\theta)\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}, \quad \text{where } \Psi(x,\theta) := \nabla_x\Phi(x,\theta) - \int_\Theta \nabla_x\Phi(x,\tau)g(x,\tau)\pi(\mathrm{d}\tau). \tag{2.6}$$

It is worth noticing that the mapping $\theta \mapsto \Psi(x,\theta)$ satisfies the null-mean property, i.e. $\int_\Theta \Psi(x,\theta)g(x,\theta)\pi(\mathrm{d}\theta) = 0$ for any $x \in \mathbb{X}$, which allows a further application of the Poincaré inequality (2.2). Therefore, Theorem 2.2–(iii) can be revisited as follows.

Corollary 2.3. *Let $\Phi \in C^1(\mathbb{X} \times \Theta)$ be such that $\theta \mapsto \nabla_x \Phi(x,\theta)$ is Lipschitz, for any $x \in \mathbb{X}$. Given a positive and measurable function $h : \mathbb{X} \to \mathbb{R}$, let f, g, Ψ be defined by (2.5) and (2.6). If $\int_\Theta |\theta|^2 g(x,\theta)\pi(\mathrm{d}\theta) < +\infty$ holds for every $x \in \mathbb{X}$ and*

$$K := \operatorname*{ess\,sup}_{x\in\mathbb{X}} \big(\mathcal{C}\big[g(x,\cdot)\pi\big]\big)^2 \bigg(\int_\Theta \big|\nabla_\theta \Psi(x,\theta)\big|^2 g(x,\theta)\pi(\mathrm{d}\theta)\bigg)^{\frac{1}{2}} < +\infty, \tag{2.7}$$

then the probability kernel $\pi(\cdot|\cdot)$, defined by (1.2) and (2.5), satisfies (1.1) with $\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)} = \mathcal{W}_2$ and $L = K$.

Proof. The regularity of Φ ensures that $\nabla_x \int_\Theta e^{\Phi(x,\theta)}\pi(\mathrm{d}\theta) = \int_\Theta e^{\Phi(x,\theta)} \nabla_x \Phi(x,\theta)\pi(\mathrm{d}\theta)$, and thanks to this property a computation immediately shows that (2.6) holds for every $x \in \mathbb{X}$. As already mentioned, $\int_\Theta \Psi(x,\theta)g(x,\theta)\pi(\mathrm{d}\theta) = 0$ for any $x \in \mathbb{X}$, thus an application of the Poincaré inequality (2.2) yields

$$\mathcal{C}\big[g(x,\cdot)\pi\big]\mathcal{J}_\pi\big[g(x,\cdot)\big] \le \big(\mathcal{C}\big[g(x,\cdot)\pi\big]\big)^2 \bigg(\int_\Theta \big|\nabla_\theta \Psi(x,\theta)\big|^2 g(x,\theta)\pi(\mathrm{d}\theta)\bigg)^{\frac{1}{2}}$$

for every $x \in \mathbb{X}$. The conclusion follows from Theorem 2.2–(iii). □

This corollary allows us to easily deal with statistical models $f(\cdot|\cdot)$ belonging to the well-known *exponential family*. See, e.g., [11,20] for a comprehensive treatment of the exponential family from the point of view of classical statistics, and [32] for a Bayesian approach. For the canonical exponential family, we consider two measurable functions $T : \mathbb{X} \to \mathbb{R}^d$ and $h : \mathbb{X} \to (0,+\infty)$. Upon putting $\Theta = \{\theta \in \mathbb{R}^d : \int_\mathbb{X} e^{T(x)\cdot\theta} h(x)\,\mathrm{d}x < +\infty\}$, the function Φ assumes the form

$$\Phi(x,\theta) = T(x)\cdot\theta - M(\theta), \quad \text{with } M(\theta) := \log \int_\mathbb{X} e^{T(x)\cdot\theta} h(x)\,\mathrm{d}x. \tag{2.8}$$

In addition, we recall the standard *regularity conditions* for the canonical exponential family: Θ is a nonempty open subset of $\mathbb{R}^d$ and the interior of the convex hull of the support of $h \circ T^{-1}$ is assumed to be nonempty. Under such conditions Θ proves to be convex, while $M : \Theta \to \mathbb{R}$ turns out to be strictly convex, analytic and steep (cf. Definition 3.2 of [20]). These considerations allows further estimates on the Poincaré constant in (2.7), according to the discussion of Section 2.1. Finally, if the function T belongs to $C_b^1(\mathbb{X};\mathbb{R}^d)$, the integral term in (2.7) is formally re-written according to

$$\bigg(\int_\Theta \big|\nabla_\theta \Psi(x,\theta)\big|^2 g(x,\theta)\pi(\mathrm{d}\theta)\bigg)^{\frac{1}{2}} = \big|\nabla T(x)\big|.$$

In this setting, we can further refine Corollary 2.3, thanks to the Bakry–Emery criterion (2.4), by stating the following

Proposition 2.4. *Consider a statistical model from the exponential family with a Lipschitz-continuous T, $h : \mathbb{X} \to (0,+\infty)$, Θ and M as above. Let $\pi(\mathrm{d}\theta) = e^{-W(\theta)}\,\mathrm{d}\theta$ with $W \in C^2(\Theta)$. If $\mathrm{Hess}[M+W] \ge \alpha I$ on Θ in the sense of quadratic forms for some $\alpha > 0$, and $\int_\Theta |\theta|^2 \exp\{T(x)\cdot\theta - M(\theta) - W(\theta)\}\,\mathrm{d}\theta < +\infty$ for every $x \in \mathbb{X}$, then the posterior distribution $\pi(\cdot|\cdot)$, defined by (1.2) and (2.5), satisfies (1.1) with distance $\mathcal{W}_2$ and $K = \mathrm{Lip}(T)/\alpha$.*

Remark 2.5. Because of their frequent use in practical statistical context, the exponential family is often rewritten under different re-parametrizations, both of the parameter and the data. Of course, property (1.1) depends crucially on the specific parametrization, and can fail after a re-parametrization. For example, the re-parametrization of the parameter in terms of the mean (see, for example, Chapter 3 of [20]) preserves the Lipschitz continuity if $\nabla M : \Theta \to \mathbb{R}^d$ is itself Lipschitz. Apropos of the re-parametrization of the data, very often the sufficient statistics T is itself viewed as the datum, which leads to a simpler problem. See Section 3.1 below.

Remark 2.6. If $\Theta = \mathbb{R}^d$ and $\alpha = 0$ in Proposition 2.4, an alternative estimate of the $\mathcal{W}_2$-Lipschitz constant in (1.1) is $L \le 12\sqrt{3}\mathrm{Lip}(T)\mathrm{Var}(e^{M+V})$, in view of an already recalled result by Bobkov [16]. Further variants can be obtained by applying Proposition A.5 in the Appendix.

3. Applications

3.1. *Statistical inference with n exchangeable observations*

In concrete statistical applications, it is customary to consider the observed datum x as a vector $(x_1,\dots,x_n)$ containing the outcomes of n experiments. Accordingly, the space $\mathbb{X}$ mentioned in Section 1.1 becomes a product space, say $\mathbb{X}_1^n$. In the Bayesian approach, the vector $(x_1,\dots,x_n)$ is viewed as the realization of some random vector, say $(X_1,\dots,X_n)$, and the core of the analysis hinges on the stochastic dependence between the components of this random vector. In particular, when the experiments are performed under "ideally similar physical conditions" the order in which the outcomes are collected becomes irrelevant. This intuitive, practical observation is captured by the notion of *exchangeability*, introduced by B. de Finetti. See [3] for a comprehensive reference on exchangeability, and Section 2.12 of [48] for a statistical perspective.

Here, we illustrate how to apply our theory of Lipschitz-continuous kernels within the field of statistical inference with n exchangeable observations, lending our results a more statistical flavour and giving a deeper insight into the concept of "Bayesian well-posedness". First, we recall that a sequence $\{X_i\}_{i\ge 1}$ of $\mathbb{X}_1$-valued random variables, defined on $(\Omega,\mathscr{A},\mathsf{P})$, is exchangeable if the identity $\mathsf{P}[X_1\in A_1,\dots,X_n\in A_n]=\mathsf{P}[X_1\in A_{\sigma_n(1)},\dots,X_n\in A_{\sigma_n(n)}]$ is fulfilled for any $n\in\mathbb{N}$, permutation $\sigma_n:\{1,\dots,n\}\to\{1,\dots,n\}$ and $A_1,\dots,A_n\in\mathscr{X}_1$, where $\mathscr{X}_1$ is a σ-algebra on $\mathbb{X}_1$. Under fairly general assumptions (e.g., when $\mathbb{X}_1$ is a Polish metric space and $\mathscr{X}_1$ coincides with its Borel σ-algebra), de Finetti's representation theorem states that the law of the observations can be written as $\mathsf{P}[X_1\in A_1,\dots,X_n\in A_n]=\int_{\mathbb{T}}[\prod_{i=1}^n \nu(A_i|\theta)]\pi(\mathrm{d}\theta)$, where $(\mathbb{T},\mathscr{T})$ is a suitable measurable space (the *parameter space*), π is a prior p.m. on $(\mathbb{T},\mathscr{T})$, and $\nu:\mathscr{X}_1\times\mathbb{T}\to[0,1]$ is a kernel representing the statistical model for any single observation. If we suppose that the family $\{\nu(\cdot|\theta)\}_{\theta\in\mathbb{T}}$ of p.m.'s is dominated by some σ-finite measure λ_1 on $(\mathbb{X}_1,\mathscr{X}_1)$, with relative density $f(\cdot|\theta)$, then, by resorting to the Bayes formula (1.3), the posterior distribution of the random parameter given the observations can be written as

$$\pi_n(\mathrm{d}\theta|x_1,\dots,x_n):=\frac{[\prod_{i=1}^n f(x_i|\theta)]\pi(\mathrm{d}\theta)}{\int_{\mathbb{T}}[\prod_{i=1}^n f(x_i|\tau)]\pi(\mathrm{d}\tau)} \tag{3.1}$$

for any $(x_1,\dots,x_n)\in\mathbb{X}_1^n$ such that $\int_{\mathbb{T}}[\prod_{i=1}^n f(x_i|\tau)]\pi(\mathrm{d}\tau)>0$. Moreover, from a classical perspective, the product $\prod_{i=1}^n f(x_i|\theta)$, when viewed as a function of θ, represents the *likelihood function* $L_n(\theta;x_1,\dots,x_n)$. Hence, with a view to highlighting the role of our theory, we focus on the appealing situation in which there exists a *classical sufficient statistics*. By the well-known Fisher–Neyman factorization criterion, we recall that a measurable mapping $\mathsf{t}_n:(\mathbb{X}_1^n,\mathscr{X}_1^n)\to(\mathbb{S},\mathscr{S})$ is named a classical sufficient statistics whenever there exist a measurable space $(\mathbb{S},\mathscr{S})$ and two measurable functions $\overline{g}:\mathbb{S}\times\mathbb{T}\to[0,+\infty)$ and $\overline{h}:\mathbb{X}_1^n\to[0,+\infty)$ such that $L_n(\theta;x_1,\dots,x_n)=\overline{g}(\mathsf{t}_n(x_1,\dots,x_n);\theta)\overline{h}(x_1,\dots,x_n)$ holds for every $(x_1,\dots,x_n)\in\mathbb{X}_1^n$. We also notice that, in the exchangeable case, any classical sufficient statistics t_n turns out to be a symmetric function of $x_1,\dots,x_n$. A remarkable example is obtain when $\mathbb{X}_1$ is endowed with some metric structure and the mapping $x\mapsto f(x|\theta)$ is continuous and positive for every $\theta\in\mathbb{T}$. In fact, $(\mathbb{S},\mathscr{S})$ can be chosen as the space of all probability densities on $(\mathbb{X}_1,\mathscr{X}_1,\lambda_1)$, endowed with the topology of weak (narrow) convergence and ensuing Borel σ-algebra $\mathscr{S}$, and $\mathsf{t}_n(x_1,\dots,x_n)$ as the *empirical measure* $\frac{1}{n}\sum_{i=1}^n\delta_{x_i}$. In this case, (3.1) can be rewritten by replacing the product $\prod_{i=1}^n f(x_i|\theta)$ with $\overline{g}(\mathsf{t}_n(x_1,\dots,x_n);\theta)$ for any $(x_1,\dots,x_n)\in\mathbb{X}_1^n$ such that $\int_{\mathbb{T}}\overline{g}(\mathsf{t}_n(x_1,\dots,x_n);\tau)\pi(\mathrm{d}\tau)>0$. This identity is crucial to notice that, in the case of n exchangeable observations, it seems more natural to investigate the Lipschitz-continuity of the posterior distribution with respect to the variable t_n, rather than the original vector $(x_1,\dots,x_n)$. Thus, a natural reformulation of (1.1) becomes

$$\mathrm{d}_{\mathcal{P}(\mathbb{T})}\big(\pi_n(\mathrm{d}\theta|x_1,\dots,x_n),\pi_n(\mathrm{d}\theta|y_1,\dots,y_n)\big)\le K\,\mathrm{d}_{\mathbb{S}}\big(\mathsf{t}_n(x_1,\dots,x_n),\mathsf{t}_n(y_1,\dots,y_n)\big) \tag{3.2}$$

with some suitable distance $\mathrm{d}_{\mathbb{S}}$ on $\mathbb{S}$. This reformulation is in harmony with the original assumption of exchangeability, since the RHS of (3.2) is invariant after a permutation of the data $(x_1,\dots,x_n)$ or $(y_1,\dots,y_n)$, unlike the (product) distance between $(x_1,\dots,x_n)$ and $(y_1,\dots,y_n)$, which is not preserved by permutation. As already noted in [27, Section 2.3] and [28], these considerations provide a new geometrical perspective on the basic formulation of Bayesian inference.

To illustrate the last consideration, we restrict to the case in which the above density $f(\cdot|\theta)$ has the exponential form as in (2.5) and (2.8). Thus, under the same standard regularity conditions for Θ of Section 2.2, we can take $\mathbb{T}$ equal to $\overline{\Theta}$. In this framework, we have at our disposal the classical sufficient statistics $\mathsf{t}_n(x_1,\dots,x_n)=\frac{1}{n}\sum_{i=1}^n T(x_i)$ which is an element of the interior Λ of the convex hull of the support of $h\circ T^{-1}$. Indeed, we recall that $\nabla M:\Theta\to\Lambda$ is a smooth diffeomorphism and $\hat{\theta}_n:=(\nabla M)^{-1}(\mathsf{t}_n(x_1,\dots,x_n))$ coincides with the maximum likelihood estimator (MLE). Thus, we will study the Lipschitz-continuity of the posterior distribution of the random parameter with respect to t_n which, due to the recalled relation with the MLE, establishes an interesting link between Bayesian and classical statistics.

Proposition 3.1. *Consider a statistical model from the exponential family* (2.5), *with* $T:\mathbb{X}_1\to\mathbb{R}^d$, $h:\mathbb{X}_1\to(0,+\infty)$, Θ *and* M *as in* Section 2.2. *Let* $\pi(\mathrm{d}\theta)=e^{-W(\theta)}\,\mathrm{d}\theta$ *with* $W\in C^2(\Theta)$. *If* $\mathrm{Hess}[M]\ge\alpha I$ *and* $\mathrm{Hess}[W]\ge\lambda_* I$ *on* Θ *in the sense of quadratic forms*, *for some* $\alpha>0$ *and* $\lambda_*\in\mathbb{R}$, *and*

$$\int_\Theta|\theta|^2\exp\big\{n\big[\mathsf{t}_n(x_1,\dots,x_n)\cdot\theta-M(\theta)\big]-W(\theta)\big\}\,\mathrm{d}\theta<+\infty$$

for every $n \in \mathbb{N}$ *and* $(x_1, \ldots, x_n) \in \mathbb{X}_1^n$, *then the posterior* $\pi_n(\cdot|\cdot)$ *satisfies* (3.2) *for every* $n \geq \max\{1, -\lambda_*/\alpha\}$, *with* $\mathbb{T} = \overline{\Theta}$, $\mathrm{d}_{\mathcal{P}(\mathbb{T})} = \mathcal{W}_2$, $\mathrm{d}_{\mathbb{S}}$ *equal to the Euclidean distance on* $\mathbb{R}^d$, *and* $K = \frac{n}{n\alpha+\lambda_*}$. *In addition, if* ∇M *is Lipschitz-continuous with constant* ℓ, *then*

$$\mathcal{W}_2\big(\pi_n(\mathrm{d}\theta|x_1, \ldots, x_n), \pi_n(\mathrm{d}\theta|y_1, \ldots, y_n)\big) \leq \frac{n\ell}{n\alpha + \lambda_*}\big|\hat{\theta}_n(x_1, \ldots, x_n) - \hat{\theta}_n(y_1, \ldots, y_n)\big|.$$

holds for every $n \geq \max\{1, -\lambda_*/\alpha\}$ *and* $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{X}_1^n$.

Proof. We just notice that the function Φ of Section 2.2 becomes $\Phi(\mathfrak{t}_n, \theta) = n[\mathfrak{t}_n - M(\theta)]$, and apply Proposition 2.4. □

3.2. *Approximation of posterior distributions by mixtures*

This subsection is referred to the setting of Section 1.1-1.2, with the further assumption that $(\mathbb{X}, \mathrm{d}_{\mathbb{X}})$ is totally bounded. A joint p.m. γ is given on $(\mathbb{X} \times \mathbb{R}^d, \mathscr{X} \otimes \mathscr{B}(\mathbb{R}^d))$, with first marginal χ. The probability kernel $\pi(\cdot|\cdot) : \mathscr{B}(\mathbb{R}^d) \times \mathbb{X} \to [0, 1]$ is thought of as a distinguished solution of the disintegration problem, that is $\int_A \pi(B|x)\chi(\mathrm{d}x) = \gamma(A \times B)$ for any $A \in \mathscr{X}$ and $B \in \mathscr{B}(\mathbb{R}^d)$. For simplicity, we assume that the support of χ coincides with the whole of $\mathbb{X}$. Now, we briefly describe an approximation procedure due to Renyi [75]. See also [71,74] and references therein. Fix $\epsilon > 0$ arbitrarily. By total boundedness, there is a finite partition of $\mathbb{X}$, denoted by $\{A_1, \ldots, A_{k(\epsilon)}\}$, satisfying

(i) $A_i \cap A_j = \varnothing$, for every $i, j \in \{1, \ldots, k(\epsilon)\}$ with $i \neq j$
(ii) $\bigcup_{j=1}^{k(\epsilon)} A_j = \mathbb{X}$
(iii) $\chi(A_j) > 0$ for every $j \in \{1, \ldots, k(\epsilon)\}$
(iv) $\chi(\partial A_j) = 0$ for every $j \in \{1, \ldots, k(\epsilon)\}$
(v) $\mathrm{diam}(A_j) \leq \epsilon$.

The number $k(\epsilon)$ is usually referred to as the ϵ-covering number of $(\mathbb{X}, \mathrm{d}_{\mathbb{X}})$, and it is related to the dimension of $\mathbb{X}$. We consider the following approximation of $\pi(\cdot|\cdot)$, given by

$$\pi_\epsilon(B|x) := \sum_{j=1}^{k(\epsilon)} \frac{\gamma(A_j \times B)}{\chi(A_j)} \mathbb{1}_{A_j}(x)$$

for any $B \in \mathscr{B}(\mathbb{R}^d)$ and $x \in \mathbb{X}$. Finally, we endow the space $\mathcal{P}(\mathbb{R}^d)$ of all p.m.'s on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ with the Borel σ-algebra $\mathscr{P}(\mathbb{R}^d)$ originated by the weak convergence of p.m.'s. We have the following

Proposition 3.2. *Let* $\epsilon > 0$ *and* $\{A_1, \ldots, A_{k(\epsilon)}\}$ *be given as above. Let* $\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}$ *be any distance which is convex and* $\mathscr{P}(\mathbb{R}^d) \otimes \mathscr{P}(\mathbb{R}^d) \setminus \mathscr{B}([0, +\infty))$*-measurable. Let the kernel* $\pi(\cdot|\cdot)$ *satisfy* (1.1) *with such distance* $\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}$. *Then,*

$$\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}\big(\pi(\cdot|x), \pi_\epsilon(\cdot|x)\big) \leq L\epsilon, \quad \forall x \in \mathbb{X}. \tag{3.3}$$

Proof. Fix $x \in \mathbb{X}$. Then, $x \in A_{j(x)}$ for some $j(x) \in \{1, \ldots, k(\epsilon)\}$ and $\pi_\epsilon(\cdot|x) = \frac{1}{\chi(A_{j(x)})}\int_{A_{j(x)}} \pi(\cdot|y)\chi(\mathrm{d}y)$. Since $\pi(\cdot|x) = \frac{1}{\chi(A_{j(x)})}\int_{A_{j(x)}} \pi(\cdot|x)\chi(\mathrm{d}y)$, exploit the convexity of $\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}$ to obtain

$$\mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}\big(\pi(\cdot|x), \pi_\epsilon(\cdot|x)\big) \leq \frac{1}{\chi(A_{j(x)})}\int_{A_{j(x)}} \mathrm{d}_{\mathcal{P}(\mathbb{R}^d)}\big(\pi(\cdot|x), \pi(\cdot|y)\big)\chi(\mathrm{d}y).$$

Combination of this last inequality with (1.1) leads immediately to (3.3). □

The above proposition can be used to tackle the following question, which occurs very frequently in Bayesian inference. See [74] and [63] for formalizations within the Bayesian nonparametric setting and the parametric setting obtained by the classical exponential family, respectively. Let $\nu(\cdot|\cdot) : \mathscr{X} \times \mathbb{R}^d \to [0, 1]$ be a probability kernel representing the statistical model, not necessarily dominated. Given some prior π on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, suppose that the posterior is not computable in a closed form, so that very little can be said beyond its existence. This phenomenon usually happens in a semi-parametric or nonparametric setting. In any case, π can be well approximated by mixtures of the form $\sum_{j=1}^N \lambda_j\pi_j$, where $\pi_1, \ldots, \pi_N$ are prior measures on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, usually belonging to some distinguished class, and $\lambda_1, \ldots, \lambda_N \in [0, 1]$ with $\sum_{j=1}^N \lambda_j = 1$. Now, assume that the posterior $\pi_j(\cdot|\cdot) : \mathscr{B}(\mathbb{R}^d) \times \mathbb{X} \to [0, 1]$, relative to the prior π_j, is actually computable in a closed form. Thus, it can be shown that the posterior $\pi_*(\cdot|\cdot) : \mathscr{B}(\mathbb{R}^d) \times \mathbb{X} \to [0, 1]$, relative to the prior $\sum_{j=1}^N \lambda_j \pi_j$, is equal to

$$\pi_*(\cdot|x) = \lambda_j(x)\pi_j(\cdot|x) \quad \text{with } \lambda_j(x) := \frac{\lambda_j \int_{\mathbb{R}^d} f(x|\tau)\pi_j(\mathrm{d}\tau)}{\sum_{i=1}^N \lambda_i \int_{\mathbb{R}^d} f(x|\tau)\pi_i(\mathrm{d}\tau)}.$$

Following [63,74], we observe that the above Proposition 3.2 can be used to compute the degree of approximation of the true posterior $\pi(\cdot|\cdot)$ by $\pi_*(\cdot|\cdot)$, uniformly with respect to the observed value x. For instance, our Proposition 3.2 improves on Proposition 2 of [74] by providing an explicit rate of convergence.

3.3. *Bayesian consistency*

In the problem of consistency, we start by considering a sequence of exchangeable observations, say $\{X_i\}_{i\geq 1}$, whose probability distribution is given by the identity $\mathsf{P}[X_1 \in A_1, \dots, X_n \in A_n] = \int_{\mathbb{T}}[\prod_{i=1}^n \nu(A_i|\theta)]\pi(\mathrm{d}\theta)$, as explained in Section 3.1. In this subsection, we confine ourselves to case of real-valued X_i's, so that $A_1, \dots, A_n \in \mathscr{B}(\mathbb{R})$, with reference measure $\lambda_1 = \mathcal{L}^1$. Moreover, we let Θ be an open subset of $\mathbb{R}^d$, and π a p.m. with support equal to $\overline{\Theta}$ with $\pi(\partial\Theta) = 0$. Hence, the above space $\mathbb{T}$ coincides with $\overline{\Theta}$. Lastly, we suppose that, for all $\theta \in \Theta$, $\nu(\cdot|\theta)$ is absolutely continuous with respect to λ_1 with density $f(\cdot|\cdot) > 0$, and that the mapping $x \mapsto f(x|\theta)$ is continuous. In this framework, the posterior distribution is given by the Bayes formula (3.1), while the likelihood can be written as $\exp\{n \int_{\mathbb{R}} \log f(y|\theta)\mathfrak{e}_n^x(\mathrm{d}y)\}$ where $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\mathfrak{e}_n^x(\cdot) := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}(\cdot)$ denotes the *empirical measure*. In the theory of Bayesian consistency, one fixes $\theta_0 \in \Theta$ and generates a sequence $\{\xi_i\}_{i\geq 1}$ of i.i.d. random variables from the p.m. $\nu(\cdot|\theta_0)$ given by the density $f(\cdot|\theta_0)$. The objective is to prove that the posterior piles up near the true value θ_0, i.e. that $\pi_n(U_0^c|\xi_1, \dots, \xi_n) \to 0$ as $n \to \infty$ for every neighborhood $U_0 \in \mathscr{B}(\Theta)$ of θ_0, where convergence is intended in probability. See [31] and [48, Chapter 4] for foundational motivations. Now, with the help of the theory developed in this paper, we are able to provide a *posterior contraction rate* at θ_0, i.e. a sequence $\{\epsilon_n\}_{n\in\mathbb{N}}$ of positive numbers for which

$$\pi_n\big(\big\{\theta \in \Theta : |\theta - \theta_0| \geq M_n\epsilon_n\big\}|\xi_1, \dots, \xi_n\big) \xrightarrow{\mathsf{P}} 0, \quad \text{as } n \to \infty, \tag{3.4}$$

holds for every diverging sequence $\{M_n\}_{n\geq 1}$ of positive numbers, where $\xrightarrow{\mathsf{P}}$ denotes convergence in probability. Cfr. Definition 8.1 in [47]. Now, we further assume that both $\nu(\cdot|\theta_0)$ and $\nu_1(\cdot)$ belong to $\mathcal{P}_1(\mathbb{R})$, where $\nu_1(A) := \int_A \int_\Theta f(x|\theta)\,\mathrm{d}x\pi(\mathrm{d}\theta)$. Thus, we can put

$$\epsilon_n = \mathsf{E}\big[\mathcal{W}_1\big(\pi_n(\mathrm{d}\theta|\xi_1, \dots, \xi_n); \delta_{\theta_0}\big)\big] \tag{3.5}$$

and notice that this choice actually provides a posterior contraction rate at θ_0, highlighting the relevant role played by the Wasserstein distance in this theory. In fact, an application of the Markov inequality yields

$$\pi_n\big(\big\{\theta \in \Theta : |\theta - \theta_0| \geq M_n\epsilon_n\big\}|\xi_1, \dots, \xi_n\big) \leq \frac{1}{M_n\epsilon_n}\mathcal{W}_1\big(\pi_n(\mathrm{d}\theta|\xi_1, \dots, \xi_n); \delta_{\theta_0}\big)$$

and the conclusion displayed in (3.4) follows by taking expectation of both sides of the above inequality, after recalling the suitable choice of ϵ_n made in (3.5), Now, for any distribution function F on $\mathbb{R}$, we introduce the probability kernel

$$\pi_n^*(\mathrm{d}\theta|F) := \frac{\exp\{n\int_{\mathbb{R}} \log f(y|\theta)\,\mathrm{d}F(y)\}}{\int_\Theta \exp\{n\int_{\mathbb{R}} \log f(y|t)\,\mathrm{d}F(y)\}\pi(\mathrm{d}t)}\pi(\mathrm{d}\theta) = \frac{\exp\{n\int_0^1 \log f(F^{-1}(u)|\theta)\,\mathrm{d}u\}}{\int_\Theta \exp\{n\int_0^1 \log f(F^{-1}(u)|t)\,\mathrm{d}u\}\pi(\mathrm{d}t)}\pi(\mathrm{d}\theta),$$

where, in the first line, integrals on $\mathbb{R}$ are intended in Riemann–Stieltjes sense, while, in the second line, $F^{-1}(u) := \inf\{y \in \mathbb{R}|F(y) \geq u\}$. In this notation, we have $\pi_n(\mathrm{d}\theta|x) = \pi_n^*(\mathrm{d}\theta|\hat{F}_n^x)$, where $\hat{F}_n^x(y) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[x_i,+\infty)}(y)$ denotes the empirical distribution function. Thanks to the triangle inequality for the Wasserstein distance, we can provide the following useful bound for the expression of ϵ_n given in (3.5), namely

$$\epsilon_n \leq \mathcal{W}_1\big(\pi_n^*(\mathrm{d}\theta|F_0); \delta_{\theta_0}\big) + \mathsf{E}\big[\mathcal{W}_1\big(\pi_n^*(\mathrm{d}\theta|F_0); \pi_n^*\big(\mathrm{d}\theta|\hat{F}_n^\xi\big)\big)\big],$$

where $F_0(y) := \int_{-\infty}^y f(x|\theta_0)\,\mathrm{d}x$ and $\hat{F}_n^\xi(y) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[\xi_i,+\infty)}(y)$. Apropos of the former term on the above RHS, we notice that $\mathcal{W}_1(\pi_n^*(\mathrm{d}\theta|F_0); \delta_{\theta_0}) = \int_\Theta |\theta - \theta_0|\pi_n^*(\mathrm{d}\theta|F_0)$. Then, combining the definitions of *Kullback–Leibler divergence*

$K(\theta|\theta_0) := \int_{\mathbb{R}} \log(\frac{f(y|\theta_0)}{f(y|\theta)}) f(y|\theta_0)\,\mathrm{d}y$ with that of $\pi_n^*(\mathrm{d}\theta|F_0)$, we can write

$$\mathcal{W}_1\big(\pi_n^*(\mathrm{d}\theta|F_0);\delta_{\theta_0}\big) = \frac{\int_\Theta |\theta-\theta_0| e^{-nK(\theta|\theta_0)}\pi(\mathrm{d}\theta)}{\int_\Theta e^{-nK(\theta|\theta_0)}\pi(\mathrm{d}\theta)}.$$

Here, we confine ourselves to dealing with *regular models* (Cfr. [43, Chapter 18]), meaning that the Fisher information matrix $\mathrm{I}[\theta_0]$ at θ_0, given by

$$\mathrm{I}[\theta_0] := \left(-\int_{\mathbb{R}}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_i} f(x|\theta)\right]_{\theta=\theta_0} f(x|\theta_0)\,\mathrm{d}x\right)_{ij}$$

is strictly positive definite. Thus, with the quadratic form notation as in (2.4), we have that $K(\theta|\theta_0) = \frac{1}{2}\langle \mathrm{I}[\theta_0](\theta-\theta_0), (\theta-\theta_0)\rangle + o(|\theta-\theta_0|^2)$ as $\theta\to\theta_0$, and that $\inf\{K(\theta|\theta_0)|\theta\in\Theta, |\theta-\theta_0|\geq\epsilon\} > 0$ for all sufficiently small $\epsilon>0$. Now, an application of Theorem 41 in [19] shows that

$$\int_\Theta e^{-nK(\theta|\theta_0)}\pi(\mathrm{d}\theta) \sim \left(\frac{2\pi}{n}\right)^{d/2}\frac{1}{\sqrt{\mathrm{I}[\theta_0]}},$$

while Theorem 43 of the same reference gives

$$\int_\Theta |\theta-\theta_0| e^{-nK(\theta|\theta_0)}\pi(\mathrm{d}\theta) \sim \left(\frac{2}{n}\right)^{(d+1)/2}\frac{1}{2}\Gamma\left(\frac{d+1}{2}\right)\frac{\int_{\mathbb{S}^{d-1}}(\langle \mathrm{I}[\theta_0]^{-1}z,z\rangle)^{1/2}\,\mathrm{d}\sigma(z)}{\sqrt{\mathrm{I}[\theta_0]}},$$

where $\mathbb{S}^{d-1}$ stands for the surface of the ball of radius equal to 1 and centered at the origin of $\mathbb{R}^d$. In conclusion, for regular models, we get $\mathcal{W}_1(\pi_n^*(\mathrm{d}\theta|F_0);\delta_{\theta_0}) \sim \frac{1}{\sqrt{n}}$ as $n\to+\infty$. At this stage, if we were able to show that the mapping $F\mapsto\pi_n^*(\mathrm{d}\theta|F)$ is Lipschitz-continuous, in the sense that

$$\mathcal{W}_1\big(\pi_n^*(\mathrm{d}\theta|F_1);\pi_n^*(\mathrm{d}\theta|F_2)\big) \leq L(f,\pi)\mathcal{W}_2(\mu_1;\mu_2) \tag{3.6}$$

with $\mu_i((-\infty,y]) = F_i(y)$ for $i=1,2$, for some constant $L(f,\pi)\geq 0$ independent of n, then we would conclude that

$$\mathsf{E}\big[\mathcal{W}_1\big(\pi_n^*(\mathrm{d}\theta|F_0);\pi_n^*\big(\mathrm{d}\theta|\hat{F}_n^\xi\big)\big)\big] \leq L(f,\pi)\mathsf{E}\big[\mathcal{W}_2\big(\mathfrak{e}_n^\xi;\nu(\cdot|\theta_0)\big)\big],$$

establishing in this way a very interesting connection. In fact, the term $\mathsf{E}[\mathcal{W}_2(\mathfrak{e}_n^\xi;\nu(\cdot|\theta_0))]$ is well-known in the probabilistic literature as *speed of mean Glivenko–Cantelli convergence*, or *monopartite matching problem*. See, for example, [17,36,45]. In particular, for one-dimensional distributions, if $\nu(\cdot|\theta_0)\in\mathcal{P}_2(\mathbb{R})$ satisfies also

$$\int_{\mathbb{R}} \frac{\nu((-\infty,x]|\theta_0)\nu((x,+\infty)|\theta_0)}{f(x|\theta_0)}\,\mathrm{d}x < +\infty \tag{3.7}$$

we have $\mathsf{E}[\mathcal{W}_2(\mathfrak{e}_n^\xi;\nu(\cdot|\theta_0))] \sim \frac{1}{\sqrt{n}}$ as $n\to+\infty$, which again represent the optimal rate. Cfr. [17, Theorem 5.1].

To prove (3.6), we bring the theory developed in Section 2 into the game. We start from a well-known identity by Dall'Aglio, according to which

$$\mathcal{W}_2(\mu_1,\mu_2) = \|F_1^{-1}-F_2^{-1}\|_{\mathrm{L}^2(0,1)} := \left(\int_0^1 \big|F_1^{-1}(u)-F_2^{-1}(u)\big|^2\,\mathrm{d}u\right)^{1/2}.$$

Thanks to this fact, we can apply point (ii) or (iii) of Theorem 2.2—or, more precisely, their infinite-dimensional reformulations, stated as point (ii) or (iii) of Theorem 4.5 below, with $\mathbb{V} = \mathrm{L}^2(0,1)$,

$$\mathbb{X} = \big\{H:(0,1)\to\mathbb{R}|H(u) = \inf\big\{y\in\mathbb{R}|\mu((-\infty,y])\geq u\big\} \text{ for some } \mu\in\mathcal{P}_2(\mathbb{R})\big\} \tag{3.8}$$

and

$$g_n(H,\theta) := \frac{\exp\{n\int_0^1 \log f(H(u)|\theta)\,\mathrm{d}u\}}{\int_\Theta \exp\{n\int_0^1 \log f(H(u)|t)\,\mathrm{d}u\}\pi(\mathrm{d}t)} = \frac{e^{n\Phi(H,\theta)}}{\int_\Theta e^{n\Phi(H,t)}\pi(\mathrm{d}t)},\qquad H\in\mathbb{X},$$

where $\Phi(H,\theta) := \int_0^1 \log f(H(u)|\theta)\,\mathrm{d}u$. Indeed, we notice that $\pi_n^*(\mathrm{d}\theta|F) = g_n(F^{-1},\theta)\pi(\mathrm{d}\theta)$ for any distribution function F with $F^{-1} \in L^2(0,1)$. We show an explicit solution based on Theorem 4.5–(iii). The evaluation of the Fisher functional starts from the evaluation of the Gateaux derivative of the mapping $H \mapsto g_n(H,\theta)$, namely

$$\begin{aligned}\nabla_H g_n(H,\theta) &= n\frac{\nabla_H \Phi(H,\theta)e^{n\Phi(H,\theta)}(\int_\Theta e^{n\Phi(H,t)}\pi(\mathrm{d}t)) - e^{n\Phi(H,\theta)}(\int_\Theta \nabla_H\Phi(H,t)e^{n\Phi(H,t)}\pi(\mathrm{d}t))}{(\int_\Theta e^{n\Phi(H,t)}\pi(\mathrm{d}t))^2}\\ &= n g_n(H,\theta)\bigg[\nabla_H\Phi(H,\theta) - \int_\Theta \nabla_H \Phi(H,t) g_n(H,t)\pi(\mathrm{d}t)\bigg].\end{aligned}$$

This computation yields

$$\mathcal{J}_\pi\big[g(H,\cdot)\big] = n\bigg(\int_\Theta \bigg\|\nabla_H\Phi(H,\theta) - \int_\Theta \nabla_H\Phi(H,t)g_n(\cdot,t)\pi(\mathrm{d}t)\bigg\|^2_{\mathrm{L}^2(0,1)} g_n(H,\theta)\pi(\mathrm{d}\theta)\bigg)^{1/2}.$$

Moreover, we notice that $\langle \nabla_H\Phi(H,\theta),\Psi\rangle_{\mathrm{L}^2(0,1)} = \int_0^1 \frac{\partial_x f(H(u)|\theta)}{f(H(u)|\theta)}\Psi(u)\,\mathrm{d}u$ which, by resorting once again to the Poincaré inequality, entails

$$\begin{aligned}&\bigg(\int_\Theta \bigg\|\nabla_H\Phi(H,\theta) - \int_\Theta \nabla_H\Phi(H,t)g_n(\cdot,t)\pi(\mathrm{d}t)\bigg\|^2_{\mathrm{L}^2(0,1)} g_n(H,\theta)\pi(\mathrm{d}\theta)\bigg)^{1/2}\\ &\quad\le \mathcal{C}\big[g_n(H,\theta)\pi(\mathrm{d}\theta)\big]\bigg(\int_\Theta\bigg\|\nabla_H\frac{\partial_x f(H(\cdot)|\theta)}{f(H(\cdot)|\theta)}\bigg\|^2_{\mathrm{L}^2(0,1)} g_n(H,\theta)\pi(\mathrm{d}\theta)\bigg)^{1/2}.\end{aligned}$$

We assume that the following scaling estimate holds

$$\mathcal{C}^2\big[g_n(H,\cdot)\pi(\cdot)\big] \le \frac{\tilde{C}(H;f,\pi)}{n}, \tag{3.9}$$

where $\tilde{C}(H;f,\pi)$ is a constant independent of n. Finally, we define

$$\mathcal{E}(H;f,\pi) := \bigg(\sup_{n\in\mathbb{N}}\int_\Theta\bigg\|\nabla_H\frac{\partial_x f(H(\cdot)|\theta)}{f(H(\cdot)|\theta)}\bigg\|^2_{\mathrm{L}^2(0,1)} g_n(H,\theta)\pi(\mathrm{d}\theta)\bigg)^{1/2}$$

and $L(f;\pi) := \sup_{H\in L^2(0,1)} \tilde{C}(H;f,\pi)\mathcal{E}(H;f,\pi)$. We can now condense this line of reasoning in the following

Theorem 3.3. *Suppose that*:

(i) $f(x|\theta) > 0$ *for all* $(x,\theta) \in \mathbb{R}\times\Theta$ *and* $x \mapsto f(x|\theta) \in C^2(\mathbb{R})$ *for all* $\theta \in \Theta$;
(ii) $\nu_1(\cdot) \in \mathcal{P}_2(\mathbb{R})$, *where* $\nu_1(A) := \int_A\int_\Theta f(x|\theta)\,\mathrm{d}x\pi(\mathrm{d}\theta)$;
(iii) *for fixed* $\theta_0 \in \Theta$, $\{f(\cdot|\theta)\}_{\theta\in\Theta}$ *defines a* C^2*-regular model at* θ_0, *as stated, e.g., in* [43, *Chapter* 18];
(iv) $\nu(\cdot|\theta_0) \in \mathcal{P}_2(\mathbb{R})$ *satisfies* (3.7);
(v) $\nabla_H \frac{\partial_x f(H(\cdot)|\theta_0)}{f(H(\cdot)|\theta_0)} \in L^2(0,1)$ *for any* $H \in \mathbb{X}$, *where* $\mathbb{X}$ *is defined by* (3.8);
(vi) *there exists* $\tilde{C}(H;f,\pi)$ *such that Poincaré constant of the posterior satisfies the bound* (3.9) *for any* $n\in\mathbb{N}$;
(vii) $L(f;\pi) < +\infty$.

Then the posterior is consistent at θ_0, *with the optimal posterior rate* $1/\sqrt{n}$.

The validity of the estimate (3.9), which is here an assumption, is natural under suitable conditions like the ones in Proposition A.5 in the Appendix. The above assumptions (vi)–(vii) are therefore a rephrasing of the assumption in Theorem 4.5–(iii), which is stated later in Section 4 (as a generalization of Theorem 2.2–(iii)) and can be invoked for proving Theorem 3.3. For extensions and sharpening of this approach to Bayesian consistency and of Theorem 3.3, including rigorous proofs, we refer to the recent contribution [34], where we also show novel applications.

4. Some extensions and other new results

4.1. *Wasserstein distance*: *The PDE approach*

Here, we briefly describe the techniques we shall exploit when considering the 2-Wasserstein distance, in order to establish (1.1) under Assumptions 2.1 for a probability kernel of the form (1.2) and such that $\pi(\cdot|x)$ has finite second moment for any $x \in \mathbb{X}$. Indeed, it will be convenient to take advantage of the following dynamical formulation and to resort to the ensuing PDE approach.

Letting $C_c^\infty(\overline{\Theta})$ denote the space of restrictions to $\overline{\Theta}$ of $C_c^\infty(\mathbb{R}^d)$ functions, the dynamic formulation of the 2-Wasserstein distance is based on the *continuity equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\overline{\Theta}}\psi(\theta)\mu_t(\mathrm{d}\theta) = \int_{\overline{\Theta}}\langle\nabla\psi(\theta), \mathbf{w}_t(\theta)\rangle\mu_t(\mathrm{d}\theta) \quad \forall\psi \in C_c^\infty(\overline{\Theta}), \tag{4.1}$$

where $[0,1] \ni t \mapsto \mu_t \in \mathcal{P}_2(\overline{\Theta})$ is a narrowly continuous curve and $\overline{\Theta} \ni \theta \mapsto \mathbf{w}_t(\theta) \in$ is a time-dependent velocity vector field. The *Benamou–Brenier formula* [13] asserts that the Wasserstein distance between μ_0 and μ_1 can be computed as

$$\mathcal{W}_2(\mu_0,\mu_1) = \inf\int_0^1\left(\int_{\overline{\Theta}}\left|\mathbf{w}_t(\theta)\right|^2\mu_t(\mathrm{d}\theta)\right)^{\frac{1}{2}}\mathrm{d}t,$$

where the infimum is taken among all narrowly continuous curves from μ_0 to μ_1 in $\mathcal{P}_2(\overline{\Theta})$ and all Borel functions $[0,1]\times\overline{\Theta} \ni (t,\theta) \mapsto \mathbf{w}_t(\theta) \in \mathbb{R}^d$ such that $\mathbf{w}_t \in L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)$ for a.e. $t \in (0,1)$ and such that (4.1) holds.

By looking at the map $x \mapsto \pi(\cdot|x) \in \mathcal{P}_2(\overline{\Theta})$ associated to a probability kernel in the form (1.2), let us fix two points $x_1, x_2 \in \mathbb{X}$. We notice that a continuous curve $[0,1] \ni t \mapsto \alpha_{x_1,x_2}(t) \in \mathbb{X}$ such that $\alpha_{x_1,x_2}(0) = x_1$, $\alpha_{x_1,x_2}(1) = x_2$, naturally induces a curve on $\mathcal{P}_2(\overline{\Theta})$ defined by

$$[0,1] \ni t \mapsto \pi\left(\cdot|\alpha_{x_1,x_2}(t)\right) \in \mathcal{P}_2(\overline{\Theta}).$$

We use this curve for bounding the Wasserstein distance, as the computation of associated velocity vector fields $\mathbf{w}_t^{x_1,x_2}$ yields a direct estimate by means of the Benamou–Brenier formula. Indeed, if the vector field $\mathbf{w}_t^{x_1,x_2} \in L^2_{\pi(\cdot|\alpha_{x_1,x_2}(t))}(\overline{\Theta})$ satisfies the continuity equation in coupling with the curve $\pi(\cdot|\alpha_{x_1,x_2}(t))$, for every fixed $x_1, x_2 \in \mathbb{X}$, then the Benamou–Brenier formula entails

$$\mathcal{W}_2\big(\pi(\cdot|x_1),\pi(\cdot|x_2)\big) \le \int_0^1\left(\int_{\overline{\Theta}}\left|\mathbf{w}_t^{x_1,x_2}(t,\theta)\right|^2\pi\left(\mathrm{d}\theta|\alpha_{x_1,x_2}(t)\right)\right)^{\frac{1}{2}}\mathrm{d}t.$$

Therefore, if we can further prove that $K \ge 0$ exist such that

$$\int_0^1\left(\int_{\overline{\Theta}}\left|\mathbf{w}_t^{x_1,x_2}(t,\theta)\right|^2\pi\left(\mathrm{d}\theta|\alpha_{x_1,x_2}(t)\right)\right)^{\frac{1}{2}}\mathrm{d}t \le K|x_1-x_2|, \tag{4.2}$$

then we obtain (1.1) with the $\mathcal{W}_2$ distance and $L = K$. In this regard, if $\alpha_{x_1,x_2}(t)$ is chosen to be a line segment, the velocity vector field scales as $|\alpha'_{x_1,x_2}(t)| = |x_1 - x_2|$. Henceforth, we restrict indeed to the case of the line segment (which is related to the choice of $\mathbb{X}$ as a convex set), that is, we let

$$\alpha_{x_1,x_2}(t) = \mathbf{s}_{x_1,x_2}(t) := (1-t)x_1 + tx_2, \quad x_1, x_2 \in \mathbb{X}, t \in [0,1]. \tag{4.3}$$

In order to obtain an estimate like (4.2), taking account of the time-scaling induced by the choice (4.3), we analyze the dual norm

$$\sup\left\{\partial_\nu\int_{\overline{\Theta}}\psi(\theta)\pi(\mathrm{d}\theta|x) : \psi \in C_c^\infty(\overline{\Theta}), \int_{\overline{\Theta}}\left|\nabla\psi(\theta)\right|^2\pi(\mathrm{d}\theta|x) \le 1\right\}, \tag{4.4}$$

where $x \in \mathbb{X}$, ν is a unit vector in $\mathbb{R}^m$ and ∂_ν denotes the associated directional derivative. We note that (4.4) is the dual expression of the $L^2_{\pi(\cdot|x)}(\overline{\Theta})$ norm of the solution $\mathbf{w}_x^\nu$ to

$$\partial_\nu\int_{\overline{\Theta}}\psi(\theta)\pi(\mathrm{d}\theta|x) = \int_{\overline{\Theta}}\langle\mathbf{w}_x^\nu(\theta), \nabla\psi(\theta)\rangle\pi(\mathrm{d}\theta|x) \quad \forall\psi \in C_c^\infty(\overline{\Theta}). \tag{4.5}$$

For $x = \mathbf{s}_{x_1,x_2}(t)$ and $\nu = \frac{x_2-x_1}{|x_2-x_1|}$, we get indeed $\mathbf{w}_t^{x_1,x_2} = |x_2 - x_1|\mathbf{w}_x^\nu$. Therefore, a crucial step towards the desired estimate (4.2) will be an estimate for the norm (4.4). Indeed, if $\|\mathbf{w}_x^\nu\|_{L^2_{\pi(\cdot|x)}(\Theta)} \le K$ for some constant K that is independent of x and ν, then (4.2) holds.

It is natural to look for a solution to (4.5) in the form of a gradient vector field $\mathbf{w}_x^\nu = \nabla u_x^\nu$, thus providing optimality of the $L^2_{\pi(\cdot|x)}$ norm (as we detail in Section 5.1). Therefore, by recalling the general form (1.2) of the probability kernel, we formally interpret equation (4.5) as a family of degenerate elliptic problems (where we write $g_x(\cdot) = g(x,\cdot)$, hinting at the fact that here $x \in \mathbb{X}$ plays the role of parameter)

$$\begin{cases} -\mathrm{div}\big(g_x\pi\nabla u_x^\nu\big) = \partial_\nu g_x\pi & \text{in } \Theta, \\ g_x\nabla u_x \cdot \mathbf{n} = 0 & \text{on } \partial\Theta, \end{cases} \tag{4.6}$$

where $\mathbf{n}$ denotes the normal to the boundary.

Existence, regularity and estimation of weak solutions to degenerate elliptic equations (see [40,41]) are related to the validity of a weighted Poincaré inequality such as (2.2), the weights being given here by the p.m.'s $\pi(\cdot|x)$ as x varies in $\mathbb{X}$. In view of the above discussion, the result in Theorem 2.2–(iii) has a clear PDE interpretation: $g \in L^1_\pi(\Theta; W^{1,1}_{\mathrm{loc}}(\mathbb{X}))$ is a regularity assumption that allows to take the ∂_ν-derivative under the integral sign in (4.5), while the condition involving both the Poincaré constant and the Fisher functional appears as an estimate of the solution to (4.6). A similar interpretation holds for Theorem 2.2–(iv).

4.2. *Estimates of $\mathcal{W}_2$ on moving domains*

Also in this subsection, we keep the mathematical setting of Assumptions 2.1 and we confine ourselves to treating kernels of the form (1.2). We provide two other results, in which we get rid of the positivity restriction on g appearing in Theorem 2.2–(iii) and of the Sobolev assumption on g in the x variable. This task requires the introduction of some new notation, along with the assumption that π admits a density q with respect to the Lebesgue measure $\mathcal{L}^d$. Thus, without loss of generality, we fix $\pi = \mathcal{L}^d \llcorner \Theta$ in (1.2), throughout this subsection. For $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$, we assume that $\Theta_x := \{g(x,\cdot) > 0\}$ is, up to a $\mathcal{L}^d$-null set, an open connected subset of Θ with locally Lipschitz boundary. Moreover, for any direction $\nu \in \mathbb{S}^{m-1}$, we consider the following Neumann boundary value problem

$$\begin{cases} -\mathrm{div}\big(g_x\nabla u_x^\nu\big) = \partial_\nu \tilde{g}_x & \text{in } \Theta_x, \\ g_x\nabla u_x^\nu \cdot \mathbf{n}_x = g_x\mathbf{V}_x^\nu \cdot \mathbf{n}_x & \text{on } \partial\Theta_x, \end{cases} \tag{4.7}$$

where g_x is a shorthand for $g(x,\cdot)$ and $\mathbf{n}_x$ denotes the exterior unit normal to $\partial\Theta_x$. This problem represents of course a generalization of (4.6). The map $(\mathbb{X},\Theta) \ni (x,\theta) \mapsto \tilde{g}(x,\theta) = \tilde{g}_x(\theta)$ is a Sobolev map extending $g(x,\cdot)$ to the whole of Θ, while ∂_ν denotes the derivative in the direction $\nu \in \mathbb{S}^{m-1}$. More precisely, we assume that there exists $\tilde{g} \in L^1_{\mathrm{loc}}(\mathbb{X}\times\Theta)$ such that $\tilde{g} \in W^{1,1}(\tilde{\mathbb{X}}\times\Theta)$ for any open set $\tilde{\mathbb{X}}$ compactly contained in $\mathbb{X}$ and such that $g(x,\theta) = \tilde{g}(x,\theta)\mathbb{1}_{\Theta_x}(\theta)$ for $(\mathcal{L}^m \otimes \mathcal{L}^d)$-a.e. $(x,\theta) \in \mathbb{X}\times\Theta$. We note this extension guarantees that the right hand side in the first equation of (4.7) belongs to $L^1(\Theta_x)$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. Moreover, $\mathbf{V}_x^\nu : \Theta_x \to \mathbb{R}^d$ is the vector field representing the velocity of Θ_x in Θ, when the parameter x varies along the ν direction. Here, Θ_x is assumed to be the image of a reference connected open set with locally Lipschitz boundary, say $\Theta_* \subset \mathbb{R}^d$, through Φ_x, where $\{\Phi_x\}_{x\in\mathbb{X}}$ is a smooth family of diffeomorphisms. In such a case, we say that the positivity set of $g(x,\cdot)$ varies according to a $\mathbb{X}$-*regular motion*. The detailed notion of $\mathbb{X}$-regular motion will be given in Definition 5.8, in Section 5.2. Then, we put $\mathbf{V}_x^\nu = \partial_\nu\Phi_x \circ \Phi_x^{-1}$ and we introduce the notation $\mathbf{V}_x$ for the matrix $\nabla_x\Phi_x \circ \Phi_x^{-1}$. If $g \in L^1_{\mathrm{loc}}(\mathbb{X}\times\Theta)$ satisfies all the above conditions, we say that g admits a *regular extension*. Again, a detailed notion of regular extension will be given in Definition 5.9, in Section 5.2.

The way is now paved for the formulation of a first abstract result, where we refer to weak solutions to problem (4.7). For clarity, a weak solution u_x^ν is defined in the usual way, through integration by parts, as an element of the weighted Sobolev space $H^1(\Theta_x, g_x)$. See Definition 5.5 below.

Theorem 4.1. *Let $\pi(\cdot|\cdot)$ be a kernel in the form (1.2), with $\pi = \mathcal{L}^d \llcorner \Theta$ and with $g \in L^1_{\mathrm{loc}}(\mathbb{X}\times\Theta)$ admitting a regular extension and satisfying $\int_\Theta |\theta|^2 g(x,\theta)\,\mathrm{d}\theta < +\infty$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. For any $\nu \in \mathbb{S}^{m-1}$, suppose there exists a weak solution $u_x^\nu \in H^1(\Theta_x, g_x)$ to the problem (4.7), for $\mathcal{L}^{m-1}$-a.e. $x \in \mathbb{X}$, and that*

$$K := \sup_{\nu\in\mathbb{S}^{m-1}} \operatorname*{ess\,sup}_{x\in\mathbb{X}} \left(\int_{\Theta_x} \big|\nabla u_x^\nu(\theta)\big|^2 g(x,\theta)\,\mathrm{d}\theta\right)^{1/2} < +\infty. \tag{4.8}$$

Then, there exists a $\mathcal{W}_2$-Lipschitz version of $\pi(\cdot|\cdot)$ satisfying (1.1) with $L = K$.

The next theorem will provide an estimate of the solution to problem (4.7). For a kernel $\pi(\cdot|\cdot)$ in the form (1.2) with $\pi = \mathcal{L}^d \llcorner \Theta$, we will use the shorthand $\mathcal{C}[g(x,\cdot)]$ to denote the Poincaré constant of the p.m. $g(x,\cdot)\mathcal{L}^d \llcorner \Theta$ on $(\Theta, \mathscr{T})$. Denoting as usual by ∇_x the gradient in the x-variable and by ∇ the gradient in the θ-variable, we introduce the *Fisher functionals* associated to (the regular extension of) g as

$$\mathcal{J}_1\big[\tilde{g}(x,\cdot)\big] := \left(\int_{\Theta_x} \frac{|\nabla_x \tilde{g}(x,\theta)|^2}{\tilde{g}(x,\theta)}\,\mathrm{d}\theta\right)^{\frac{1}{2}}, \qquad \mathcal{J}_2\big[\tilde{g}(x,\cdot)\big] := \left(\int_{\Theta_x} \frac{|\nabla \tilde{g}(x,\theta)|^2}{\tilde{g}(x,\theta)}\,\mathrm{d}\theta\right)^{\frac{1}{2}}. \tag{4.9}$$

Theorem 4.2. *Let $\pi(\cdot|\cdot)$ be a kernel in the form (1.2), with $\pi = \mathcal{L}^d \llcorner \Theta$ and with $g \in L^1_{\mathrm{loc}}(\mathbb{X}\times\Theta)$ admitting a regular extension $\tilde{g}$ and satisfying $\int_\Theta |\theta|^2 g(x,\theta)\,\mathrm{d}\theta < +\infty$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. If*

$$K := \operatorname*{ess\,sup}_{x\in\mathbb{X}}\big\{\|\mathbf{V}_x(\cdot)\|_{W^{1,\infty}(\Theta_x)}\big(1 + \mathcal{C}\big[g(x,\cdot)\big]\big(1 + \mathcal{J}_2\big[\tilde{g}(x,\cdot)\big]\big)\big) + \mathcal{C}\big[g(x,\cdot)\big]\mathcal{J}_1\big[\tilde{g}(x,\cdot)\big]\big\} < +\infty$$

is valid, then there exists a $\mathcal{W}_2$-Lipschitz version of $\pi(\cdot|\cdot)$ satisfying (1.1) with $L = K$.

In the derivation of (4.7) from the continuity equation (see Section 5), we handle the derivative of the integral on the left-hand side of (4.5) by making use of the *Reynolds transport formula* from continuum mechanics (see Lemma A.1 in the Appendix). This explains the role of the vector $\mathbf{V}_x^\nu = \partial_\nu \Phi_x \circ \Phi_x^{-1}$ that represents the spatial velocity, defined on the deformed configuration Θ_x, whereas Θ_* is the reference configuration. This approach is, in a sense, alternative (although less general) to the optimal transport formalism and the Monge–Ampere equation, but suitable to the statistical framework, where probability densities are often defined by truncation. In this context, the extension map $\tilde{g}_x(\cdot)$ is actually given a-priori, as in the examples dealing with Pareto-type statistical models, in Section 4.3. Of course, if $\Theta_x = \Theta$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$, Theorem 4.2 is reduced to a particular instance of Theorem 2.2–(iii), as problem (4.7) is reduced to (4.6).

4.3. *A remarkable example*: *Pareto statistical models*

Pareto statistical models, which are considered in the next two propositions, provide a paradigmatic application of Theorem 4.1, as they gives rise to a moving support of probability densities that are defined by truncation.

Proposition 4.3. *Consider the one-dimensional Pareto statistical model*

$$\mathbb{X} = (1,+\infty), \qquad \Theta = (1,\theta_0), \qquad \theta_0 \in (1,+\infty], \qquad f(x|\theta) = \frac{\theta}{x^2}\mathbb{1}\{\theta < x\}.$$

Suppose we are given a prior distribution π, whose support is $\overline{\Theta}$, admitting a density $q \in L^1(\Theta)$, and let $Q(\theta) := \theta q(\theta)$. Assume further that $Q \in W^{1,1}(\Theta)$ and $1/Q \in L^1(\Theta)$, and let $C_Q(\theta) := Q(\theta)(\int_1^\theta Q(\tau)\,\mathrm{d}\tau)^{-1/2}(\int_1^\theta \frac{\mathrm{d}\tau}{Q(\tau)})^{1/2}$. Then the posterior distribution $\pi(\cdot|\cdot)$, defined by means of (1.2) and (1.3), satisfies (1.1) with distance $\mathcal{W}_2$ and $L = K$, where $K := \sup_{x\in\mathbb{X}} C_Q(x \wedge \theta_0)$.

Proof. We give the proof in case $\theta_0 = +\infty$ (minor variants are required if $\theta_0 < +\infty$). The proof is a direct application of Theorem 4.1. With respect to the notation therein, we drop the apex ν as the directional derivative is reduced to the derivative in the x-variable. We have by Bayes formula (1.3)

$$\rho(x) = \frac{1}{x^2}\int_1^x Q(\tau)\,\mathrm{d}\tau, \qquad g_x(\theta) = \frac{f(x|\theta)q(\theta)}{\rho(x)} = \frac{Q(\theta)}{\int_1^x Q(\tau)\,\mathrm{d}\tau}\mathbb{1}\{\theta < z\},$$

so that $\Theta_x = (1,x)$ is the positivity set of g_x. It is clear that the function $\mathbb{X}\times\Theta \ni (x,\theta) \mapsto g_x(\theta)$ satisfies $\int_\Theta g_x(\theta)\,\mathrm{d}\theta = 1$ and $\int_\Theta \theta^2 g(x,\theta)\,\mathrm{d}\theta < +\infty$ for every $x \in \mathbb{X}$. Moreover, as required by Theorem 4.1, it admits a regular extension according to Definition 5.9 which is found later in Section 5.3. Indeed, we may define $\Phi_x(\cdot) : \Theta_* \to \Theta_x$ by $\Theta_* = (1,2)$ and $\Phi_x(\theta) = (\theta-1)(x-1)+1$. As a consequence, we have $\partial_x\Phi_x \circ (\Phi_x)^{-1}(\theta) = \frac{\theta-1}{x-1}$ and $\Phi_x(\theta)$ satisfies the conditions of Definition 5.8. Moreover, we may consider the natural extension

$$\tilde{g}_x(\theta) := \frac{Q(\theta)}{\int_1^x Q(\tau)\,\mathrm{d}\tau}, \qquad x \in (1,+\infty), \theta \in (1,+\infty).$$

By the assumptions on q, the map $(t,\theta)\mapsto \tilde g_t(\theta)$ belongs to $W^{1,1}((\alpha,\beta)\times\Theta)$ for any $1<\alpha<\beta<+\infty$. It is indeed a regular extension in the sense of Definition 5.9. Therefore, problem (4.7) is reduced to

$$\begin{cases} -\big(g_x u_x'\big)' = \partial_x \tilde g_x & \text{in } (1,x),\\ g_x(x)u_x'(x) = g_x(x), & \\ g_x(1)u_x(1) = 0, & \end{cases} \tag{4.10}$$

where the $'$ stands for the derivative in the θ variable. By taking into account that $\partial_x \tilde g_x(\theta) = -Q(\theta)Q(x)(\int_1^x Q(\tau)\,\mathrm{d}\tau)^{-2}$, the solution u_x to problem (4.10) satisfies for any $x\in(1,+\infty)$

$$u_x'(\theta) = \frac{Q(x)}{Q(\theta)}\bigg(\int_1^x Q(\tau)\,\mathrm{d}\tau\bigg)^{-1}\int_1^\theta Q(\tau)\,\mathrm{d}\tau.$$

We obtain

$$\bigg(\int_1^x \big|u_x'(\theta)\big|^2 g_x(\theta)\,\mathrm{d}\theta\bigg)^{1/2} \le Q(x)\bigg(\int_1^x Q(\tau)\,\mathrm{d}\tau\bigg)^{-1/2}\bigg(\int_1^x \frac{\mathrm{d}\tau}{Q(\tau)}\bigg)^{1/2} = C_Q(x).$$

By the assumptions on q it easily follows that $C_Q(x)$ is bounded on $\mathbb{X}$, so that the assumptions of Theorem 4.1 are satisfied, thus we conclude by taking $\sup_{x\in\mathbb{X}} C_Q(x)$ as bound for the Lipschitz constant. □

Proposition 4.4. *Let us consider the statistical model*

$$\mathbb{X} = (1,+\infty), \qquad (\theta,\varepsilon)\in\Theta = (1,2)^2, \qquad f(x|\theta,\varepsilon) = \frac{\varepsilon\theta^\varepsilon}{x^{1+\varepsilon}}\mathbb{1}\{\theta<x\}$$

along with a prior probability density $q\in C^2(\Theta)$ such that $0<c_q\le q(\theta,\varepsilon)$ for any $(\theta,\varepsilon)\in\Theta$. Then there exists an explicit positive constant Z_q, only depending on c_q and $\|q\|_2 := \sup_\Theta q + \sup_\Theta|\nabla q| + \sup_\Theta|\nabla^2 q|$, such that the posterior distribution satisfies (1.1) *with distance $\mathcal{W}_2$ and $L = Z_q$.*

Proof. We first look at the values of $x\in(1,2]$. Let $C_q := \sup_\Theta q$. A comutation shows that for $\theta\in(1,2)$, $\varepsilon\in(1,2)$,

$$g_x(\theta,\varepsilon) = \frac{f(\theta,\varepsilon|x)q(\theta,\varepsilon)}{\rho(x)} = \frac{\varepsilon\theta^\varepsilon q(\theta,\varepsilon)\mathbb{1}\{\theta<x\}(\theta,\varepsilon)}{x^{\varepsilon+1}\rho(x)}, \quad \text{where } \rho(x) = \int_1^x\int_1^2 \frac{\sigma\tau^{\sigma+1}}{x^{\sigma+1}} q(\tau,\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\tau.$$

An easy estimate shows that for any $x\in[1,2]$

$$8C_q(x-1)\ge\rho(x)\ge c_q(x-1), \qquad \big|\rho'(x)\big|\le 26C_q, \quad \text{and} \quad \big|\rho'(x)\big| + \big|\rho''(x)\big| + \big|\rho'''(x)\big| \le M\|q\|_2, \tag{4.11}$$

where M is a suitable numerical constant. Moreover

$$\tilde g_x(\theta,\varepsilon) = \frac{\varepsilon\theta^\varepsilon q(\theta,\varepsilon)}{x^{\varepsilon+1}\rho(x)} \quad \text{and} \quad \partial_x\tilde g_x(\theta,\varepsilon) = -A_x(\varepsilon)\theta^\varepsilon q(\theta,\varepsilon), \quad \text{where } A_x(\varepsilon) := \frac{\varepsilon(\varepsilon+1)}{x^{\varepsilon+2}\rho(x)} + \frac{\varepsilon\rho'(x)}{x^{\varepsilon+1}\rho^2(x)},$$

so that from (4.11) we deduce that there exists a positive constant K_q, depending only on c_q and C_q, such that

$$\big|A_x(\varepsilon)\big| \le \frac{K_q}{(x-1)^2} \quad \text{for any } \varepsilon\in(1,2) \text{ and any } x\in(1,2]. \tag{4.12}$$

We see that the map $(x,\theta,\varepsilon)\mapsto\tilde g_x(\theta,\varepsilon)$ belongs to $W^{1,1}((\alpha,\beta)\times(1,2)\times(1,2))$ for any $1<\alpha<\beta<2$. Moreover, for any $x\in(1,2)$ the map $(\theta,\varepsilon)\mapsto\partial_x\tilde g_x(\theta,\varepsilon)$ belongs to $L^1(\Theta_x)$ where $\Theta_x := (1,x)\times(1,2) = \Phi_x(\Theta_2)$ and $\Phi_x(\theta,\varepsilon) = ((\theta-1)(x-1)+1,\varepsilon)$ has first component as in the proof of Proposition 4.3. In this way, we see that indeed $(x,\theta,\varepsilon)\mapsto \tilde g_x(\theta,\varepsilon)$ is a regular extension of the function $(x,\theta,\varepsilon)\mapsto g_x(\theta,\varepsilon)$ on $(1,2)\times\Theta$, according to Definition 5.9. In order to conclude, we apply Theorem 4.1. The corresponding Neumann boundary value problem is posed on a rectangle, and precisely it is (again we drop the apex ν as $\mathbb{X}$ is one-dimensional)

$$\begin{cases} -\mathrm{div}(g_x\nabla u_x) = \partial_x\tilde g_x & \text{in } \Theta_x,\\ \partial_\theta u_x = 1 & \text{on } \Gamma_x := \{x\}\times(1,2),\\ \nabla u_x\cdot\mathbf{n}_x = 0 & \text{on } \partial\Theta_x\setminus\Gamma_x. \end{cases}$$

We define $G_x(\theta,\varepsilon) := \int_1^\theta \partial_x \tilde{g}_x(\tau,\varepsilon)\,\mathrm{d}\tau = -A_x(\varepsilon)\int_1^\theta \tau^\varepsilon q(\tau,\varepsilon)\,\mathrm{d}\tau$, and we proceed with the estimate of the $H^1(\Theta_x, g_x)$ norm of u_x by duality and using the notion of weak solution to the above problem (see Definition 5.5). We obtain after an integration by parts in the θ-variable

$$\begin{aligned}\int_{\Theta_x} |\nabla u_x|^2 g_x &= \sup_{\|\psi\|_x=1}\int_{\Theta_x}\nabla u_x\cdot\nabla\psi g_x = \sup_{\|\psi\|_x=1}\left(\int_{\Theta_x}\psi(\theta,\varepsilon)\partial_x\tilde{g}_x(\theta,\varepsilon)\,\mathrm{d}\theta\,\mathrm{d}\varepsilon + \int_1^2 \psi(x,\varepsilon)g_x(x,\varepsilon)\,\mathrm{d}\varepsilon\right)\\ &= \sup_{\|\psi\|_x=1}\left(\int_{\Theta_x} -\partial_\theta\psi(\theta,\varepsilon)G_x(\theta,\varepsilon)\,\mathrm{d}\theta\,\mathrm{d}\varepsilon + \int_1^2\psi(x,\varepsilon)\big(G_x(x,\varepsilon)+g_x(x,\varepsilon)\big)\,\mathrm{d}\varepsilon\right),\end{aligned}$$

where the supremum is taken among test functions $\psi \in C^1_{g_x}(\overline{\Theta}_x)$ and $\|\psi\|_x$ is a shorthand for the norm $(\int_{\Theta_x}|\nabla\psi|^2 g_x)^{1/2}$ on $C^1_{g_x}(\overline{\Theta}_x)$. Therefore, with the notation $H_x(\varepsilon) := G_x(x,\varepsilon)+g_x(x,\varepsilon)$ and with the divergence theorem we obtain

$$\begin{aligned}\int_{\Theta_x}|\nabla u_x|^2 g_x &= \sup_{\|\psi\|_x=1}\left(\int_{\Theta_x} -G_x(\theta,\varepsilon)\partial_\theta\psi(\theta,\varepsilon)\,\mathrm{d}\theta\,\mathrm{d}\varepsilon + \int_{\Theta_x}\operatorname{div}\left(\psi(\theta,\varepsilon)\frac{\theta-1}{x-1}\big(H_x(\varepsilon),0\big)\right)\mathrm{d}\theta\,\mathrm{d}\varepsilon\right)\\ &= \sup_{\|\psi\|_x=1}\left(\int_{\Theta_x}\partial_\theta\psi\left(\frac{\theta-1}{x-1}H_x(\varepsilon)-G_x(\theta,\varepsilon)\right)\mathrm{d}\theta\,\mathrm{d}\varepsilon + \int_{\Theta_x}\frac{\psi(\theta,\varepsilon)H_x(\varepsilon)}{x-1}\,\mathrm{d}\theta\,\mathrm{d}\varepsilon\right)\\ &\le \sup_{\|\psi\|_x=1}\left(\int_{\Theta_x}\big|\nabla\psi(\theta,\varepsilon)\big|\big|G_x(\theta,\varepsilon)+H_x(\varepsilon)\big|\,\mathrm{d}\theta\,\mathrm{d}\varepsilon + \int_{\Theta_x}\frac{|\psi(\theta,\varepsilon)H_x(\varepsilon)|}{x-1}\,\mathrm{d}\theta\,\mathrm{d}\varepsilon\right).\end{aligned}$$

By Cauchy–Schwarz inequality and (2.2), if $\mathcal{C}[g_x]$ is the Poincaré constant of the probability measure $g_x\mathcal{L}^2 \llcorner \Theta_x$, we get

$$\int_{\Theta_x}|\nabla u_x|^2 g_x \le \left(\left(\int_{\Theta_x}\frac{(G_x+H_x)^2}{g_x}\right)^{\frac12} + \frac{\mathcal{C}[g_x]}{x-1}\left(\int_{\Theta_x}\frac{H_x^2}{g_x}\right)^{\frac12}\right). \tag{4.13}$$

Let us now compute suitable bounds for the terms in the above right hand side. From (4.11) and (4.12) we get for any $\theta\in(1,2)$, any $\varepsilon\in(1,2)$ and any $x\in(1,2]$ that $|G_x(\theta,\varepsilon)| \le 4C_q|A_x(\varepsilon)|(x-1) \le \frac{4C_qK_q}{x-1}$ and thus

$$\int_{\Theta_x}\frac{G_x(\theta,\varepsilon)^2}{g_x(\theta,\varepsilon)}\,\mathrm{d}\theta\,\mathrm{d}\varepsilon \le 2^{10}c_q^{-1}C_q^3K_q^2, \quad \text{for every } x\in(1,2]. \tag{4.14}$$

Since $H_x(\varepsilon) = \frac{N_\varepsilon(x)}{x^{\varepsilon+2}\rho^2(x)}$, where $N_\varepsilon(x) := (-\varepsilon(\varepsilon-1)\rho(x) - \varepsilon x\rho'(x))\int_1^x \tau^\varepsilon q(\tau,\varepsilon)\,\mathrm{d}\tau + \varepsilon x^{1+\varepsilon}\rho(x)q(x,\varepsilon)$, we have $N_\varepsilon(1)=0$, and since $N_\varepsilon'(x)$ also vanishes at $x=1$, by a Taylor expansion in x we have $N_\varepsilon(x) = \frac12(x-1)^2N_\varepsilon''(\xi_\varepsilon)$ for some $\xi_\varepsilon\in[1,x]$ and a computation exploiting (4.11) shows that for any $x\in(1,2]$ there holds $|N_\varepsilon''(x)|\le 2U\|q\|_2$ for a universal constant U, so that we deduce $H_x(\varepsilon)\le c_q^{-2}U\|q\|_2$ for any $\varepsilon\in(1,2)$ and any $x\in(1,2]$. As a consequence, (4.11) implies that for every $x\in(1,2]$

$$\frac{1}{x-1}\left(\int_{\Theta_x}\frac{H_x^2(\varepsilon)}{g_x(\theta,\varepsilon)}\,\mathrm{d}\theta\,\mathrm{d}\varepsilon\right)^{\frac12} \le 8C_q^{-3/2}c_q^{-1/2}U\|q\|_2. \tag{4.15}$$

Let us moreover treat $\mathcal{C}[g_x]$ by invoking the Holley–Stroock estimate (see Section 2): letting $V_x(\theta,\varepsilon) := -\log g_x(\theta,\varepsilon)$, we have $\mathcal{C}^2[g_x]\le\exp\{\sup_{\Theta_x}V_x - \inf_{\Theta_x}V_x\}\mathcal{C}^2[\mathcal{L}^2\llcorner\Theta_x]$, where $\mathcal{C}[\mathcal{L}^2\llcorner\Theta_x]$ enjoys the standard estimate [70] in terms of $\operatorname{diam}(\Theta_x)/\pi\le\sqrt{2}/\pi$, for any $1<x\le 2$. Since a direct estimate shows that $\sup_{\Theta_x}V_x - \inf_{\Theta_x}V_x \le \log(C_q/c_q)+6\log 2$ for every $x\in(1,2]$, we eventually get $\mathcal{C}^2[g_x]\le 2^7\pi^{-2}C_q/c_q$. This estimate can be inserted in (4.13), together with (4.15) and (4.14), and we deduce that for a suitable explicit constant Z_q, only depending on c_q and $\|q\|_2$, there holds $\int_{\Theta_x}|\nabla u_x|^2g_x \le Z_q^2$ for any $z\in(1,2]$. In conclusion, Theorem 4.1 yields the validity of (1.1) with $\mathcal{W}_2$ distance and $L = Z_q$ for any $x_1,x_2\in(1,2]$. Since the density of $\pi(\cdot,\cdot|x)$ is $g_x(\cdot,\cdot)$, and since the latter is given by $g_x(\theta,\varepsilon) = \varepsilon\theta^\varepsilon q(\theta,\varepsilon)/\mathcal{Q}$ for any $x>2$, where $\mathcal{Q} := \int_1^2\int_1^2 \sigma\tau^\sigma q(\tau,\sigma)\,\mathrm{d}\tau\,\mathrm{d}\sigma$, the estimate (1.1) trivially extends to all $x_1,x_2\in\mathbb{X}$. □

4.4. *Infinite-dimensional sample space*

Here, generalizing the setting displayed in Assumptions 2.1, we extend Theorem 2.2 to the case in which $\mathbb{X}$ is a convex set of a real separable Banach space $(\mathbb{V},\|\cdot\|_{\mathbb{V}})$, endowed with a σ-finite reference measure λ. Upon denoting by $\mathcal{H}^1$ the 1-dimensional *Hausdorff measure* on $(\mathbb{V},\mathscr{B}(\mathbb{V}))$, we prescribe the following properties for λ:

(λ1) $\lambda(A) > 0$ for every nonempty open set $A \in \mathscr{B}(\mathbb{V})$;
(λ2) if $\lambda(A) = 0$ for some $A \in \mathscr{B}(\mathbb{V})$, then $\lambda(\alpha A + v) = 0$ for all $\alpha > 0$ and $v \in \mathbb{V}$;
(λ3) for a set $A \in \mathscr{B}(\mathbb{V})$, the condition $\lambda(\{v \in \mathbb{V} | \mathcal{H}^1([0, v] \cap A) > 0\}) = 0$ entails $\lambda(A) = 0$, where $[0, v] := \{\alpha v | \alpha \in [0, 1]\}$.

We plainly observe that $\mathbb{V}$ can be taken equal to any Euclidean space $\mathbb{R}^d$ for any $d \in \mathbb{N}$, and that the standard d-dimensional Lebesgue measure fulfills the conditions (λ1)–(λ2)–(λ3). We notice that (λ1) entails that the complement of any λ-null set is dense in $\mathbb{V}$. Non-degenerate Gaussian measures on $\mathbb{V}$ also satisfy (λ1)–(λ2)–(λ3). See [72].

In the next theorem, we again confine ourselves to treating kernels of the form (1.2), but now we stress that $g : \mathbb{X} \times \Theta \to \mathbb{R}$ is defined pointwise. Therefore, g and π determine pointwise the probability kernel $\pi(\cdot|\cdot)$. Since we let $\mathbb{X}$ be a convex subset of $\mathbb{V}$, we have $\mathrm{d}_{\mathbb{X}}(x_1, x_2) := \|x_1 - x_2\|_{\mathbb{V}}$. For a λ-null subset $\mathbb{Z}$ of $\mathbb{X}$, we define

$$\mathbb{B}(\mathbb{Z}) := \{(x_1, x_2) \in \mathbb{X}^2 : x_1, x_2 \in \mathbb{X} \setminus \mathbb{Z}, \mathcal{H}^1([x_1, x_2] \cap \mathbb{Z}) = 0\}. \tag{4.16}$$

We let D_x denote the Gateaux differential operator with respect to the x-variable and $\mathbb{V}'$ be the dual space of $\mathbb{V}'$ with operator norm $\|\cdot\|_{\mathbb{V}'}$. Finally, the Fisher functional relative to π is now defined as

$$\mathcal{J}_\pi[g(x, \cdot)] := \left(\int_\Theta \frac{\|D_x g(x, \theta)\|^2_{\mathbb{V}'}}{g(x, \theta)} \pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}.$$

The way is now paved for the formulation of the following

Theorem 4.5. *Suppose there exists a λ-null set $\mathbb{Z} \subset \mathbb{X}$ such that, for π-a.e. $\theta \in \Theta$, the mapping $x \mapsto g(x, \theta)$ is Gateaux-differentiable at any $x \in \mathbb{X} \setminus \mathbb{Z}$ and absolutely continuous on any segment $[x_1, x_2]$ with $(x_1, x_2) \in \mathbb{B}(\mathbb{Z})$. When $\mathcal{W}_p$ is involved, it is further assumed that $\int_\Theta |\theta|^p \pi(\mathrm{d}\theta|x) < +\infty$ for any $x \in \mathbb{X}$. Then, the following statements hold.*

(i) *Suppose that*

$$K := \lambda\text{-}\operatorname*{ess\,sup}_{x \in \mathbb{X}} \int_\Theta \|D_x g(x, \theta)\|_{\mathbb{V}'} \pi(\mathrm{d}\theta) < +\infty.$$

Then, there exists a d_{TV}-Lipschitz version of $\pi(\cdot|\cdot)$, satisfying (1.1) with $L = K/2$.

(ii) *Let $1 \le q < +\infty$ and let p be the Hölder conjugate exponent of q. If π admits a Poincaré constant $\mathcal{C}_q[\pi]$ and if*

$$K := \pi(\Theta)^{\frac{1}{q}} \mathcal{C}_q[\pi] \, \lambda\text{-}\operatorname*{ess\,sup}_{x \in \mathbb{X}} \big\| \|D_x g(x, \cdot)\|_{\mathbb{V}'} \big\|_{L^p_\pi(\Theta)} < +\infty,$$

then there exists a $\mathcal{W}_1$-Lipschitz version of $\pi(\cdot|\cdot)$, satisfying (1.1) with $L = K$.

(iii) *Let $g > 0$ $(\lambda \otimes \pi)$-a.e. in $\mathbb{X} \times \Theta$ and let $\int_0^1 \int_\Theta \|D_x g(\mathbf{s}_{x_1,x_2}(t), \theta)\|_{\mathbb{V}'} \pi(\mathrm{d}\theta) < +\infty$ for any $(x_1, x_2) \in \mathbb{B}(\mathbb{Z})$. If*

$$K := \lambda\text{-}\operatorname*{ess\,sup}_{x \in \mathbb{X}} \mathcal{C}[g(x, \cdot)\pi] \mathcal{J}_\pi[g(x, \cdot)] < +\infty,$$

then there exists a $\mathcal{W}_2$-Lipschitz version of $\pi(\cdot|\cdot)$, satisfying (1.1) with $L = K$.

If $\mathbb{V}$ is infinite-dimensional, a total variation distance estimate like the one of Theorem 4.5–(i) can be found under stronger assumptions like the following: the map $x \mapsto g(x, \theta)$ is Lipschitz with a constant L_θ satisfying $\int_\Theta L_\theta \pi(\mathrm{d}\theta) < +\infty$ and λ is a Gaussian measure. In particular, these last assumptions imply Gateaux differentiability, according to [64, Theorem 1.1] and references therein.

5. Proofs

5.1. *The dynamic formulation of the Wasserstein distance*

Here, we provide the theoretical framework for the estimate of the 2-Wasserstein distance. We start by introducing some facts about the dynamic formulation by means of the continuity equation and the Benamou–Brenier [13] formula, which are related to the geometry of the space of probability measures. This theory is established in the seminal paper by Otto [68], see also [69], in the books of Villani [82,83], as well as the book by Ambrosio, Gigli and Savaré [5], see also [4]. Then, we give most emphasis to the continuity equation as a family of degenerate elliptic boundary value problems, parametrized by time.

We let Θ be an open connected subset of $\mathbb{R}^d$. We recall that by $C^1_c(\overline{\Theta})$ we denote the space of functions in $C^1(\overline{\Theta})$ whose support is a compact set contained in $\overline{\Theta}$. Of course, $C^1_c(\overline{\Theta}) \equiv C^1(\overline{\Theta})$ if Θ is bounded. The space $C^1_c(\overline{\Theta})$ is separable with respect to the $C^1(\overline{\Theta})$ norm $\|\psi\|_{C^1(\overline{\Theta})} := \sup_{\overline{\Theta}} |\psi| + \sup_{\overline{\Theta}} |\nabla\psi|$ and contains the space of C^1 functions with compact support in Θ.

We shall consider Borel families of measures $\{\mu_t\}_{t\in[0,1]} \subset \mathcal{P}(\overline{\Theta})$, i.e., $[0,1] \ni t \mapsto \mu_t(A)$ is Borel measurable for any Borel set $A \subseteq \overline{\Theta}$. Moreover, $[0,1] \ni t \mapsto \mu_t \in \mathcal{P}(\overline{\Theta})$ is said to be a narrowly continuous curve if $t, t_0 \in [0,1]$ and $t \to t_0$ imply the narrow convergence of μ_t to μ_{t_0}. In the following, we say that a narrowly continuous curve $[0,1] \ni t \mapsto \mu_t \in \mathcal{P}(\overline{\Theta})$ satisfies the continuity equation on $\overline{\Theta}$, in coupling with a family of vector fields $\{\mathbf{v}_t\}_{t\in[0,1]}$ such that $[0,1] \times \overline{\Theta} \ni (t,\theta) \mapsto \mathbf{v}_t(\theta) \in \mathbb{R}^d$ is Borel measurable, if

$$\int_0^1 \int_{\overline{\Theta}} \big(\partial_t\varphi(t,\theta) + \nabla\varphi(t,\theta)\cdot\mathbf{v}_t(\theta)\big)\mu_t(\mathrm{d}\theta)\,\mathrm{d}t = 0 \quad \text{for any } \varphi \in C^1_c\big((0,1)\times\overline{\Theta}\big). \tag{5.1}$$

Here and in the following, ∇ denotes the gradient in the θ variable.

We start by giving sufficient conditions on the curve $[0,1] \ni t \mapsto \mu_t$ in order to apply the Benamou–Brenier formula and estimate the 2-Wasserstein distance between μ_0 and μ_1.

Theorem 5.1. *Let $\mu_0 \in \mathcal{P}_2(\overline{\Theta})$, $\mu_1 \in \mathcal{P}_2(\overline{\Theta})$, and let $\{\mu_t\}_{t\in[0,1]} \subset \mathcal{P}(\overline{\Theta})$ be a Borel family of probability measures. Let $\mathcal{D}$ be a countable dense subset of $C^1_c(\overline{\Theta})$ (with respect to the $C^1(\overline{\Theta})$ norm). Suppose that*

(i) *the map $[0,1] \ni t \mapsto \int_{\overline{\Theta}} \psi(\theta)\mu_t(\mathrm{d}\theta)$ is absolutely continuous for any $\psi \in \mathcal{D}$,*
(ii) *$\Psi \in L^1(0,1)$, where $\Psi(t) := \sup\{\frac{\mathrm{d}}{\mathrm{d}s}\int_{\overline{\Theta}} \psi(\theta)\mu_s(\mathrm{d}\theta)|_{s=t} : \psi \in \mathrm{span}\mathcal{D}, \int_{\overline{\Theta}} |\nabla\psi(\theta)|^2\mu_t(\mathrm{d}\theta) \le 1\}$.*

Then, for a.e. $t \in (0,1)$, there exists a unique vector field $\mathbf{w}_t \in \overline{\{\nabla\psi : \psi \in C^1_c(\overline{\Theta})\}}^{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)}$ which is solution to

$$\langle \mathbf{w}_t, \nabla\psi\rangle_{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)} = \frac{\mathrm{d}}{\mathrm{d}s}\int_{\overline{\Theta}} \psi(\theta)\mu_s(\mathrm{d}\theta)\bigg|_{s=t} \quad \forall\psi \in \mathrm{span}\mathcal{D}. \tag{5.2}$$

Moreover, $\|\mathbf{w}_t\|_{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)} = \Psi(t)$ holds for a.e. $t \in (0,1)$, $\mu_t \in \mathcal{P}_2(\overline{\Theta})$ for any $t \in (0,1)$ and

$$\mathcal{W}_2(\mu_{t_1}, \mu_{t_2}) \le \int_{t_1}^{t_2} \|\mathbf{w}_t\|_{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)}\,\mathrm{d}t \quad \textit{for any } 0 \le t_1 < t_2 \le 1. \tag{5.3}$$

Proof. We preliminarily notice that assumption *i)* implies that the curve $[0,1] \ni t \mapsto \mu_t$ is narrowly continuous, in view of the Portmanteau theorem (see, e.g. [15, Section 2]). Moreover, since $\mathcal{D}$ is countable, there exists a $\mathcal{L}^1$-null set $N \subset (0,1)$ such that the mapping $t \mapsto \int_\Theta \psi(\theta)\mu_t(\mathrm{d}\theta)$ is differentiable at $t \in (0,1) \setminus N$ for any $\psi \in \mathrm{span}\mathcal{D}$. Then the supremum in assumption *ii)* is well defined (and nonnegative) for every t in $(0,1) \setminus N$, hence for a.e. $t \in (0,1)$.

A gradient vector field in $\{\nabla\psi : \psi \in \mathrm{span}\mathcal{D}\}$ admits a potential on $\overline{\Theta}$ which is unique up to a constant, therefore by mass conservation we see that for a.e. $t \in (0,1)$

$$\mathcal{T}_t[\nabla\psi] := \frac{\mathrm{d}}{\mathrm{d}s}\int_{\overline{\Theta}} \psi(\theta)\mu_s(\mathrm{d}\theta)\bigg|_{s=t}$$

defines indeed a linear functional on $\{\nabla\psi : \psi \in \mathrm{span}\mathcal{D}\}$. Moreover, since for any $\psi \in C^1_c(\overline{\Theta})$ and any $t \in (0,1)$ there holds $\|\nabla\psi\|_{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)} \le \sup_{\overline{\Theta}} |\nabla\psi|$, we see that $\{\nabla\psi : \psi \in \mathrm{span}\mathcal{D}\}$ is dense in the $L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)$ closure of the linear space $\{\nabla\psi : \psi \in C^1_c(\overline{\Theta})\}$ for any $t \in (0,1)$. Therefore, assumption *ii)* shows that, for a.e. $t \in (0,1)$, the operator $\mathcal{T}_t$ uniquely extends to a bounded linear functional on the space $\overline{\{\nabla\psi : \psi \in C^1_c(\overline{\Theta})\}}^{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)}$, where we find by Riesz representation theorem a unique vector field $\mathbf{w}_t(\cdot)$ such that

$$\mathcal{T}_t[\nabla\psi] = \int_{\overline{\Theta}} \big\langle \mathbf{w}_t(\theta), \nabla\psi(\theta)\big\rangle \mu_t(\mathrm{d}\theta) \quad \forall\psi \in \mathrm{span}\mathcal{D}$$

and such that $\|\mathbf{w}_t\|_{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)} = \Psi(t)$, thus $\mathbf{w}_t$ is the desired solution to (5.2).

At this stage, it is possible to prove (we refer to [5, Theorem 8.3.1]) that there exists of a Borel map $[0,1] \times \overline{\Theta} \ni (t,\theta) \mapsto \mathbf{v}(t,\theta) \in \mathbb{R}^d$ such that $\|\mathbf{v}(t,\cdot)\|_{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)} = \|\mathbf{w}_t\|_{L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)}$ for a.e. $t \in (0,1)$ and such that the couple $(\mathbf{v}(t,\cdot), \mu_t)$ satisfies the continuity equation (5.1). As a consequence, we may invoke the Benamou–Brenier formula (see [5, Theorem 8.3.1] and [4, Proposition 3.30]) to get $\mu_t \in \mathcal{P}_2(\overline{\Theta})$ for every $t \in (0,1)$ and the validity of the estimate (5.3). □

Remark 5.2 (Tangency condition). Following [5, Section 8.4], we define the tangent space to a measure μ in $\mathcal{P}_2(\overline{\Theta})$ by

$$\mathcal{TAN}_\mu\big(\mathcal{P}_2(\overline{\Theta})\big) := \overline{\big\{\nabla\psi : \psi \in C_c^1(\overline{\Theta})\big\}}^{L^2_\mu(\overline{\Theta};\mathbb{R}^d)}.$$

Therefore, in Theorem 5.1, we conclude that $\mathbf{w}_t \in \mathcal{TAN}_{\mu_t}(\mathcal{P}_2(\overline{\Theta}))$ for a.e. $t \in (0,1)$. This is equivalent to saying that $\mathbf{w}_t$ has minimal $L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)$ norm among $L^2_{\mu_t}(\overline{\Theta};\mathbb{R}^d)$ solutions to (5.2), see also [4, Section 3.3.2].

We notice that μ_t can be either supported on the whole of $\overline{\Theta}$ or on a subset which possibly depends on t. The rest of this section is devoted to a further analysis of the case of mobile support, starting with some more definitions and notation.

Definition 5.3 (Regular motion). Let $\Theta_* \subseteq \mathbb{R}^d$ be a nonempty open connected set with locally Lipschitz boundary. We say that a smooth mapping $[a,b] \times \Theta_* \ni (t,\theta) \mapsto \Phi_t(\theta) \in \mathbb{R}^d$ is a *regular motion* in Θ if the following conditions hold. For any $t \in [a,b]$, Φ_t is a diffeomorphism between Θ_* and a nonempty open connected set with locally Lipschitz boundary $\Theta_t := \Phi_t(\Theta_*) \subseteq \Theta$. Further, there exist positive constants k_1, k_2 such that for any $t \in [a,b]$ and any $\theta \in \Theta_*$

$$\big|\partial_t\Phi_t(\theta)\big| + \big|\nabla\Phi_t(\theta)\big| + \big|\nabla\partial_t\Phi_t(\theta)\big| \le k_2 \quad \text{and} \quad k_1 \le \det\nabla\Phi_t(\theta).$$

We notice that under the assumptions of Definition 5.3, Θ_* is bounded if and only if Θ_t is bounded for every $t \in [a,b]$. A typical example of a family of diffeomorphisms that yields a regular motion is $\Phi_t(\theta) = \theta + t\mathbf{v}(\theta)$, where $\mathbf{v} \in W^{1,\infty}(\Theta_*) \cap C^1(\Theta_*)$, $t \in [0,1]$ and $\sup_{\theta\in\Theta_*}|\nabla\mathbf{v}(\theta)| < 1$.

We next apply the above definition to positivity sets of probability densities. In view of the next definition, we say that $f \in \mathrm{ACL}([0,1]\times\Theta)$ (in short, that f has the ACL property) if for every coordinate direction ν of $\mathbb{R}\times\mathbb{R}^d$ and for $\mathcal{L}^d$-almost any line ℓ_ν in the direction of ν, f is absolutely continuous on any closed segment contained in $\ell_\nu \cap ([0,1]\times\Theta)$. More details about the ACL property will be given in the next subsection. Furthermore, we will denote by $\mathbb{1}_A$ the indicator function of a set $A \subseteq \Theta$ (i.e., $\mathbb{1}_A(\theta)$ is equal to 1 if $\theta \in A$ and it is equal to 0 otherwise).

Definition 5.4. Let $(t,\theta) \mapsto g_t(\theta) \in \mathbb{R}$ be a nonnegative $L^1_{\mathrm{loc}}((0,1)\times\Theta)$ function such that $\int_\Theta g_t(\theta)\,\mathrm{d}\theta = 1$ for a.e. $t \in (0,1)$. We say that it admits a regular extension if the following conditions are satisfied:

(i) for a.e. t in $(0,1)$, the positivity set $\{\theta \in \Theta : g_t(\theta) > 0\}$ coincides (up to a $\mathcal{L}^d$-null set) with a nonempty open connected set $\Theta_t = \Phi_t(\Theta_*) \subseteq \Theta$ with locally Lipschitz boundary, where $[0,1]\times\Theta_* \ni (t,\theta) \mapsto \Phi_t(\theta)$ is a regular motion according to Definition 5.3;

(ii) there exists a $W^{1,1}((0,1)\times\Theta) \cap \mathrm{ACL}([0,1]\times\Theta)$ function $[0,1]\times\Theta \ni (t,\theta) \mapsto \tilde{g}_t(\theta) \in \mathbb{R}$ such that

$$g_t(\theta) = \tilde{g}(t,\theta)\mathbb{1}_{\Theta_t}(\theta) \quad \text{for } \big(\mathcal{L}^1 \otimes \mathcal{L}^d\big)\text{-a.e. } (t,\theta) \in (0,1)\times\Theta.$$

As a consequence of the latter definition, we notice that $\partial_t\tilde{g}_t \in L^1(\Theta)$ for a.e. $t \in (0,1)$ and $g_t \in W^{1,1}(\Theta_t)$ for a.e. $t \in (0,1)$, with a L^1_{loc} trace on $\partial\Theta_t$ thanks to the standard characterization [46] of traces of $W^{1,1}$ functions, see for instance [62, Chapter 15].

The next result provides an estimate of the Wasserstein distance in terms of weak solutions to Neumann boundary value problems (on time-dependent domain). This is an important step towards the proof of Theorem 4.1 which will be provided in Section 5.2. Indeed, to a regular extension of a function $g \in L^1_{\mathrm{loc}}((0,1)\times\Theta)$ according to Definition 5.4 we associate the family (parametrized by t) of Neumann boundary value problems

$$\begin{cases} -\mathrm{div}(g_t\nabla u_t) = \partial_t\tilde{g}_t & \text{in } \Theta_t, \\ g_t\nabla u_t \cdot \mathbf{n}_t = g_t\partial_t\Phi_t \circ \Phi_t^{-1} \cdot \mathbf{n}_t & \text{on } \partial\Theta_t, \end{cases} \tag{5.4}$$

Here, for given t, $\mathbf{n}_t$ denotes the ($\mathcal{H}^{d-1}$- a.e. existing on $\partial\Theta_t$) outer unit normal to $\partial\Theta_t$. For such moving domains, a natural calculus tool is the Reynolds transport formula (see Lemma A.1 in the Appendix): notice that $\partial_t\Phi_t \circ \Phi_t^{-1}$ represents the velocity of the boundary. It is possible that such velocity vanishes on some part of the boundary or on the whole boundary (in particular, if Φ_t does not depend on t, then the domain is fixed, i.e., $\Theta_t \equiv \Theta$ for any t). In fact, by means of Definition 5.4 we require two properties: a regularly moving domain and the existence of a global Sobolev extension. Such properties will ensure the applicability of the Reynolds transport formula from Lemma A.1 (which also implies the standard compatibility condition for the Neumann problem (5.4), thanks to the mass conservation property $\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Theta_t} g_t(\theta)\,\mathrm{d}\theta = 0$). For given t, a weighted Sobolev space on Θ_t (with weight g_t that is positive $\mathcal{L}^d$-a.e. on Θ_t) is the natural framework for a notion of weak solution to problem (5.4). Moreover, if Θ_t is bounded we complement (5.4) with the null mean condition $\int_{\Theta_t} u_t(\theta) g_t(\theta)\,\mathrm{d}\theta = 0$ (instead if Θ_t is unbounded we complement (5.4) with a vanishing condition at infinity). Therefore, if Θ_t is bounded we let $C^1_{g_t}(\overline{\Theta}_t)$ be the space of functions $\psi \in C^1(\overline{\Theta}_t)$ such that $\int_{\Theta_t} \psi(\theta) g_t(\theta)\,\mathrm{d}\theta = 0$, while if Θ_t is unbounded we just let $C^1_{g_t}(\overline{\Theta}_t) := C^1_c(\overline{\Theta}_t)$. We give the following

Definition 5.5 (Weak solution). Let g satisfy all the conditions in Definition 5.4. Fix $t \in (0,1)$ such that $g_t \in W^{1,1}(\Theta_t)$ and $\partial_t \tilde{g}_t \in L^1(\Theta)$. The weighted Sobolev space $H^1(\Theta_t, g_t)$ is then defined as the completion of $C^1_{g_t}(\overline{\Theta}_t)$ w.r.t. the norm $\|\psi\|_{H^1(\Theta_t, g_t)} := (\int_{\Theta_t} |\nabla\psi(\theta)|^2 g_t(\theta)\,\mathrm{d}\theta)^{1/2}$. We say that $u_t \in H^1(\Theta_t, g_t)$ is a weak solution to problem (5.4) if for every $\psi \in C^1_{g_t}(\overline{\Theta}_t)$ there holds

$$\int_{\Theta_t} \nabla\psi(\theta)\cdot\nabla u_t(\theta) g_t(\theta)\,\mathrm{d}\theta = \int_{\Theta_t} \psi(\theta)\partial_t\tilde{g}_t(\theta)\,\mathrm{d}\theta + \int_{\partial\Theta_t} \psi(\sigma) g_t(\sigma)\partial_t\Phi_t\big(\Phi_t^{-1}(\sigma)\big)\cdot \mathbf{n}_t(\sigma)\mathcal{H}^{d-1}(\mathrm{d}\sigma).$$

Theorem 5.6. *Let g satisfy all the conditions in* Definition 5.4. *For any $t \in [0,1]$, let $\mu_t := \tilde{g}_t\mathcal{L}^d \llcorner \Theta_t$ and suppose that $\mu_0 \in \mathcal{P}_2(\overline{\Theta})$ and $\mu_1 \in \mathcal{P}_2(\overline{\Theta})$. Suppose that, for a.e. $t \in (0,1)$, $u_t \in H^1(\Theta_t, g_t)$ is a weak solution to problem* (5.4), *and that the map $(0,1) \ni t \mapsto (\int_\Theta |\nabla u_t(\theta)|^2 g_t(\theta)\,\mathrm{d}\theta)^{1/2}$ belongs to $L^1(0,1)$. Then, there hold $\mu_t \in \mathcal{P}_2(\overline{\Theta})$ for any $t \in (0,1)$, $\nabla u_t \in \mathcal{TAN}_{\mu_t}(\mathcal{P}_2(\overline{\Theta}))$ for a.e. $t \in (0,1)$ and*

$$\mathcal{W}_2(\mu_{t_1}, \mu_{t_2}) \le \int_{t_1}^{t_2} \left(\int_\Theta \left|\nabla u_t(\theta)\right|^2 g_t(\theta)\,\mathrm{d}\theta\right)^{\frac{1}{2}} \mathrm{d}t \quad \textit{for any } 0 \le t_1 < t_2 \le 1. \tag{5.5}$$

Proof. Let $\mathcal{D}$ be a countable dense subset of $C^1_c(\overline{\Theta})$ (in the $C^1(\overline{\Theta})$ norm). By Lemma A.1 in the Appendix, $\tilde{g}$ is such that $[0,1] \ni t \mapsto \int_{\Theta_t} \psi(\theta)\tilde{g}_t(\theta)\,\mathrm{d}\theta = \int_\Theta \psi(\theta)\mu_t(\mathrm{d}\theta)$ is absolutely continuous, and since $\mathcal{D}$ is countable, the null set $N \in (0,1)$ of its nondifferentiability points can be assumed to be independent on $\psi \in \mathrm{span}\mathcal{D}$ (in particular, assumption *i)* of Theorem 5.1 is satisfied by the Borel family $\{\mu_t\}_{t\in[0,1]} \subset \mathcal{P}(\overline{\Theta})$). Moreover, with the notation

$$\overline{\psi}_t(\cdot) := \begin{cases} \psi(\cdot) - \int_\Theta \psi(\theta) g_t(\theta)\,\mathrm{d}\theta & \text{if } \Theta_t \text{ is bounded,} \\ \psi(\cdot) & \text{otherwise,} \end{cases}$$

we have $\overline{\psi}_t \in C^1_{g_t}(\overline{\Theta}_t)$ and, for any $t \in (0,1) \setminus N$ and any $\psi \in \mathrm{span}\mathcal{D}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}r}\int_\Theta \psi(\theta)\mu_t(\mathrm{d}\theta)\bigg|_{r=t} = \frac{\mathrm{d}}{\mathrm{d}r}\int_{\Theta_r} \psi(\theta)\tilde{g}(r,\theta)\,\mathrm{d}\theta\bigg|_{r=t} = \frac{\mathrm{d}}{\mathrm{d}r}\int_{\Theta_r} \overline{\psi}_t(\theta)\tilde{g}(r,\theta)\,\mathrm{d}\theta\bigg|_{r=t}.$$

Therefore, still by making use of Lemma A.1, we apply Reynolds transport formula we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}r}\int_{\Theta_r} \psi(\theta)\tilde{g}_r(\theta)\,\mathrm{d}\theta\bigg|_{r=t} &= \int_{\Theta_t} \overline{\psi}_t(\theta)\partial_t\tilde{g}(t,\theta)\,\mathrm{d}\theta + \int_{\partial\Theta_t} \overline{\psi}_t(\sigma) g_t(\sigma)\partial_t\Phi_t\big(\Phi_t^{-1}(\sigma)\big)\cdot\mathbf{n}_t(\sigma)\mathcal{H}^{d-1}(\mathrm{d}\sigma) \\ &= \int_{\Theta_t} \nabla\overline{\psi}_t(\theta)\cdot\mathbf{w}_t(\theta) g_t(\theta)\,\mathrm{d}\theta = \int_{\Theta_t} \nabla\psi(\theta)\cdot\mathbf{w}_t(\theta) g_t(\theta)\,\mathrm{d}\theta\end{aligned}$$

for any $t \in (0,1) \setminus N$ and any $\psi \in \mathrm{span}\mathcal{D}$, where we used the fact that u_t is a weak solution to (5.4) and the notation $\mathbf{w}_t := \nabla u_t$. By assumption we have $u_t \in H^1(\Theta_t, g_t)$ for a.e. $t \in (0,1)$, thus $\mathbf{w}_t \in \overline{\{\nabla\psi : \psi \in C^1_c(\overline{\Theta}_t)\}}^{L^2_{g_t}(\Theta_t;\mathbb{R}^d)}$. Since any $C^1_c(\overline{\Theta}_t)$ function can be extended to a function in $C^1_c(\overline{\Theta})$, by truncation and extension the spaces $\overline{\{\nabla\psi : \psi \in C^1_c(\overline{\Theta}_t)\}}^{L^2_{g_t}(\Theta_t;\mathbb{R}^d)}$ and $\overline{\{\nabla\psi : \psi \in C^1_c(\overline{\Theta})\}}^{L^2_{g_t}(\Theta;\mathbb{R}^d)}$ are isometric. Thus, for a.e. $t \in (0,1)$, $\mathbf{w}_t$ is the unique solution in $\overline{\{\nabla\psi : \psi \in C^1_c(\overline{\Theta})\}}^{L^2_{g_t}(\Theta;\mathbb{R}^d)}$ to problem (5.2) and by Riesz isomorphism it satisfies $\|\mathbf{w}_t\|_{L^2_{g_t}(\Theta;\mathbb{R}^d)} = \Psi(t)$ for a.e. $t \in (0,1)$, where Ψ is defined in Theorem 5.1. Hence, condition *ii)* of Theorem 5.1 is also satisfied. Therefore, (5.5) follows along with $\mathbf{w}_t \in \mathcal{TAN}_{\mu_t}(\mathcal{P}_2(\overline{\Theta}))$ for a.e. $t \in (0,1)$. □

5.2. *Basic estimates of* d_{TV}, $\mathcal{W}_1$ *and* $\mathcal{W}_2$: *Proof of Theorem* 2.2

In this subsection, we provide the proofs of the basic results on a finite dimensional sample space under the validity of Assumptions 2.1. Let us introduce some further notation. For $\nu \in \mathbb{S}^{m-1}$, let P_ν be the projection operator onto $\{\nu\}^\perp :=$

$\{z \in \mathbb{R}^m : z \cdot \nu = 0\}$ and let $\mathbb{X}_\nu := P_\nu(\mathbb{X})$. We notice that $\mathbb{X}_\nu$ inherits the convexity of $\mathbb{X}$. We introduce the line segment $I_{\xi,\nu} := \{\xi + t\nu : t \in \mathbb{R}\} \cap \mathbb{X}$, and cleary $I_{\xi,\nu} \neq \varnothing$ if $\xi \in \mathbb{X}_\nu$. The following is the standard ACL characterization of Sobolev functions (see for instance [18,39,52,62]): $G \in W^{1,1}_{\mathrm{loc}}(\mathbb{X})$ if and only if it admits a representative such that, for any coordinate direction ν in $\mathbb{R}^m$, the restriction to $I_{\xi,\nu}$ is locally absolutely continuous for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$. In such case, the ACL property holds with respect to any direction, and it can be rephrased as follows. Given $\nu \in \mathbb{S}^{m-1}$, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$, the $L^1(0,1)$ map $t \mapsto G(\mathbf{s}_{x_1,x_2}(t))$ is (up to having modified G on a $\mathcal{L}^m$-null set) absolutely continuous on $(0,1)$, where x, y are any two distinct points of $I_{\xi,\nu}$, and

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(G\big(\mathbf{s}_{x_1,x_2}(t)\big)\big) = |x_1 - x_2|\partial_\nu G\big(\mathbf{s}_{x_1,x_2}(t)\big) = \nabla G\big(\mathbf{s}_{x_1,x_2}(t)\big) \cdot (x_2 - x_1) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0,1). \tag{5.6}$$

The weak ν-directional derivative of G coincides with the pointwise $\mathcal{L}^m$-a.e. classical ν-directional derivative. Before the proof of the main theorems, we state the following simple lemma.

Lemma 5.7. *Let $\mathbb{Y} \subset \mathbb{R}^m$ be open. Let $\psi \in L^\infty_\pi(\Theta)$. Let $g \in L^1_{\mathcal{L}^m \otimes \pi}(\mathbb{Y} \times \Theta)$. If $\int_\Theta \|g(\cdot,\theta)\|_{W^{1,1}(\mathbb{Y})}\pi(\mathrm{d}\theta) < +\infty$, then*

$$G_\psi(\cdot) := \int_\Theta \psi(\theta) g(\cdot,\theta)\pi(\mathrm{d}\theta) \tag{5.7}$$

belongs to $W^{1,1}(\mathbb{Y})$ and

$$\nabla_x G_\psi(x) = \int_\Theta \psi(\theta)\nabla_x g(x,\theta)\pi(\mathrm{d}\theta) \quad \textit{for } \mathcal{L}^m\textit{-a.e. } x \in \mathbb{Y}. \tag{5.8}$$

Proof. By the assumptions, $g \in L^1_{\mathcal{L}^m \otimes \pi}(\mathbb{Y} \times \Theta)$, $\nabla_x g \in L^1_{\mathcal{L}^m \otimes \pi}(\mathbb{Y} \times \Theta)$ and for π-a.e. $\theta \in \Theta$ the mapping $x \mapsto g(x,\theta)$ belongs to $W^{1,1}(\mathbb{Y})$. We apply Fubini's theorem to get

$$\begin{aligned}\int_{\mathbb{Y}} G_\psi(x)\nabla_x \zeta(x)\,\mathrm{d}x &= \int_\Theta \psi(\theta)\left(\int_{\mathbb{Y}} g(x,\theta)\nabla_x\zeta(x)\,\mathrm{d}x\right)\pi(\mathrm{d}\theta) = -\int_\Theta \psi(\theta)\left(\int_{\mathbb{Y}} \nabla_x g(x,\theta)\zeta(x)\,\mathrm{d}x\right)\pi(\mathrm{d}\theta)\\ &= -\int_{\mathbb{Y}} \zeta(x)\left(\int_\Theta \psi(\theta)\nabla_x g(x,\theta)\pi(\mathrm{d}\theta)\right)\mathrm{d}x\end{aligned}$$

for any $\zeta \in C^\infty_c(\mathbb{Y})$ and

$$\left|\int_{\mathbb{Y}}\left(\int_\Theta \psi(\theta)\nabla_x g(x,\theta)\pi(\mathrm{d}\theta)\right)\mathrm{d}x\right| \le \|\psi\|_{L^\infty_\pi(\Theta)}\int_\Theta \big\|g(\cdot,\theta)\big\|_{W^{1,1}(\mathbb{Y})}\pi(\mathrm{d}\theta) < +\infty.$$

Therefore, the right-hand-side in (5.8) belongs to $L^1(\mathbb{Y})$ and it is the weak gradient of G_ψ. □

We proceed to the proof of the main results. We start with the most direct proof concerning the estimate in total variation distance. We also refer to [35] for further results involving the total variation distance. We recall the dual formulation of the total variation distance. For $\mu, \nu \in \mathcal{P}(\overline{\Theta})$ there holds

$$\mathrm{d}_{\mathrm{TV}}(\mu,\nu) = \frac{1}{2}\sup_{\substack{\psi \in C_c(\overline{\Theta})\\ |\psi| \le 1}}\left(\int_{\overline{\Theta}} \psi(\theta)\mu(\mathrm{d}\theta) - \int_{\overline{\Theta}} \psi(\theta)\nu(\mathrm{d}\theta)\right),$$

where $C_c(\overline{\Theta})$ is the set of continuous functions on $\overline{\Theta}$ having compact support contained in $\overline{\Theta}$. We notice that due to the separability of $C_c(\overline{\Theta})$ it is possible to compute the above supremum on a countable dense subset (w.r.t. the sup norm).

Proof of Theorem 2.2-(i). We first claim that for any bounded continuous function ψ on $\overline{\Theta}$, the function G_ψ from (5.7), which is in $W^{1,1}_{\mathrm{loc}}(\mathbb{X})$ by Lemma 5.7, belongs to $W^{1,\infty}(\mathbb{X})$. Indeed, since $\pi(\Theta|x) = \int_\Theta g(x,\theta)\pi(\mathrm{d}\theta) = 1$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$, it is clear that $|G_\psi(x)| \le \sup_{\overline{\Theta}}|\psi|$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. By Lemma 5.7 and by assumption, we get

$$\big|\nabla G_\psi(x)\big| \le \sup_{\overline{\Theta}}|\psi|\int_\Theta \big|\nabla_x g(x,\theta)\big|\pi(\mathrm{d}\theta) = K\sup_{\overline{\Theta}}|\psi|$$

for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. The claim is proved. In particular, for any $\psi \in C_c(\overline{\Theta})$, G_ψ has a Lipschitz representative on $\mathbb{X}$.

Let $\mathcal{D}$ denote a countable dense subset (in the sup norm) of $\{\psi \in C_c(\overline{\Theta}) : |\psi| \le 1 \text{ on } \overline{\Theta}\}$. Let $\hat{g}$ be a representative (according to the $(\mathcal{L}^m \otimes \pi)$-a.e. identification) of g such that $\int_\Theta \hat{g}(x,\theta)\pi(\mathrm{d}\theta) = 1$ for every $x \in \mathbb{X}$. Therefore, $\hat{\pi}(\mathrm{d}\theta|x) := \hat{g}(x,\theta)\pi(\mathrm{d}\theta)$ is a representative of the kernel defined by (1.2), and $\hat{G}_\psi(x) := \int_\Theta \psi(\theta)\hat{g}(x,\theta)\pi(\mathrm{d}\theta)$ is a representative of G_ψ for any $\psi \in \mathcal{D}$. Moreover, for $\psi \in \mathcal{D}$, $\hat{G}_\psi$ agrees $\mathcal{L}^m$-a.e. with a Lipschitz function on $\mathbb{X}$, i.e., there exists a $\mathcal{L}^m$-null set $\mathbb{Z}_\psi \subset \mathbb{X}$ such that $|\hat{G}_\psi(x_2) - \hat{G}_\psi(x_1)| \le K|x_2 - x_1|$ for any $x_1, x_2 \in \mathbb{X} \setminus \mathbb{Z}_\psi$. Since $\mathcal{D}$ is countable, there exists a $\mathcal{L}^m$-null set $\mathbb{Z} \subset \mathbb{X}$ such that for every $\psi \in \mathcal{D}$ the restriction of $\hat{G}_\psi$ to $\mathbb{X} \setminus \mathbb{Z}$ is Lipschitz (with Lipschitz constant bounded by K). Therefore

$$\mathrm{d_{TV}}\big(\hat{\pi}(\cdot|x_2), \hat{\pi}(\cdot|x_1)\big) = \frac{1}{2}\sup_{\psi\in\mathcal{D}}\big(\hat{G}_\psi(x_2) - \hat{G}_\psi(x_1)\big) \le \frac{1}{2}\sup_{\psi\in\mathcal{D}}\|\nabla G_\psi\|_{L^\infty(\mathbb{X})}|x_2 - x_1| \le \frac{K}{2}|x_2 - x_1|$$

for any $x_1, x_2 \in \mathbb{X} \setminus \mathbb{Z}$. Note that $\hat{\pi}(\cdot|x) \in \mathcal{P}(\Theta)$ for any $x \in \mathbb{X} \setminus \mathbb{Z}$. Since $\mathbb{X} \setminus \mathbb{Z}$ is dense in $\mathbb{X}$ and since $(\mathcal{P}(\Theta), d_{\mathrm{TV}})$ is a complete metric space, the mapping $\mathbb{X} \setminus \mathbb{Z} \ni x \mapsto \hat{\pi}(\cdot|x) \in \mathcal{P}(\Theta)$ admits a unique Lipschitz continuous extension (with respect to the total variation distance) to the whole of $\mathbb{X}$ with the same Lipschitz constant $K/2$. □

For the proof of Theorem 2.2–(ii), we take advantage of the Kantorovich–Rubinstein dual fomulation of the 1-Wasserstein distance. See [82, Section 1.2]. For any $\mu, \nu \in \mathcal{P}_1(\overline{\Theta})$, there holds

$$\mathcal{W}_1(\mu,\nu) = \sup_{\substack{\psi\in C^1_c(\overline{\Theta})\\ \mathrm{Lip}(\psi)\le 1}}\left(\int_{\overline{\Theta}}\psi(\theta)\mu(\mathrm{d}\theta) - \int_{\overline{\Theta}}\psi(\theta)\nu(\mathrm{d}\theta)\right).$$

Again, the separability of $C^1_c(\overline{\Theta})$ allows to take the above supremum on a countable dense set (in the sup norm).

Proof of Theorem 2.2-(ii). We claim that the function G_ψ from (5.7) belongs to $W^{1,\infty}(\mathbb{X})$ for any $\psi \in C^1_c(\overline{\Theta})$. Indeed, as seen in the proof of Theorem 2.2–(i), we have $|G_\psi(x)| \le \sup_{\overline{\Theta}}|\psi|$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. Moreover, by Lemma 5.7, by Hölder inequality and by the Poincaré inequality (2.2), since $\nabla G_\psi = \nabla G_{\psi - a}$ for any $a \in \mathbb{R}$, we get

$$\begin{aligned}\big|\nabla G_\psi(x)\big| &= \inf_{a\in\mathbb{R}}\left|\int_\Theta\big(\psi(\theta) - a\big)\nabla_x g(x,\theta)\pi(\mathrm{d}\theta)\right| \le \inf_{a\in\mathbb{R}}\left(\int_\Theta\big|\psi(\theta) - a\big|^q\pi(\mathrm{d}\theta)\right)^{\frac{1}{q}}\left(\int_\Theta\big|\nabla_x g(x,\theta)\big|^p\pi(\mathrm{d}\theta)\right)^{\frac{1}{p}}\\ &\le \mathcal{C}_q[\pi]\left(\int_\Theta\big|\nabla\psi(\theta)\big|^q\pi(\mathrm{d}\theta)\right)^{\frac{1}{q}}\left(\int_\Theta\big|\nabla_x g(x,\theta)\big|^p\pi(\mathrm{d}\theta)\right)^{\frac{1}{p}} \le K\,\mathrm{Lip}(\psi)\end{aligned}$$

for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$, thus proving the claim.

Let $\mathcal{D}$ be a countable dense subset (in the sup norm) of $\{\psi \in C^1_c(\overline{\Theta}) : \mathrm{Lip}(\psi) \le 1\}$. By the same argument as in the proof of Theorem 2.2–(i), we obtain a $\mathcal{L}^m$-null set $\mathbb{Z}$ in $\mathbb{X}$ and a representative (still denoted by $\pi(\cdot|\cdot)$) of the kernel defined by (1.2) such that, for any $x_1, x_2 \in \mathbb{X} \setminus \mathbb{Z}$

$$\mathcal{W}_1\big(\pi(\cdot|x_1), \pi(\cdot|x_2)\big) \le \sup_{\psi\in\mathcal{D}}\|\nabla G_\psi\|_{L^\infty(\mathbb{X})}|x_2 - x_1| \le K|x_2 - x_1|.$$

Since $\mathbb{X} \setminus \mathbb{Z}$ is dense in $\mathbb{X}$ and since $(\mathcal{P}_1(\overline{\Theta}), \mathcal{W}_1)$ is complete, there exists a unique map $\mathbb{X} \ni x \mapsto \pi^*(\cdot|x) \in \mathcal{P}_1(\overline{\Theta})$ that satisfies the above Lipschitz estimate on the whole of $\mathbb{X}$, with the same Lipscthitz constant K, and such that $\pi^*(\cdot|x) \equiv \pi(\cdot|x)$ for any $x \in \mathbb{X} \setminus \mathbb{Z}$. Since the assumptions of Theorem 2.2 are also satisfied, $x \mapsto \pi^*(\cdot|x)$ is also continuous with respect to the total variation distance, therefore $\pi^*(\overline{\Theta}|x) = \pi^*(\Theta|x) = 1$ for any $x \in \mathbb{X}$. □

Proof of Theorem 2.2-(iii). Once more, we start by claiming that, for any $\psi \in C^1_c(\overline{\Theta})$, the function G_ψ from (5.7) belongs to $W^{1,\infty}(\mathbb{X})$. Indeed, we have as usual $\|G_\psi\|_{L^\infty(\mathbb{X})} \le \sup_{\overline{\Theta}}|\psi|$ and again by Cauchy–Schwarz inequality and by (2.2), by the positivity of g and by assumption, we get

$$\begin{aligned}\big|\nabla_x G_\psi(x)\big| &= \inf_{a\in\mathbb{R}}\left|\int_\Theta\big(\psi(\theta) - a\big)\nabla_x g(x,\theta)\pi(\mathrm{d}\theta)\right| \le \inf_{a\in\mathbb{R}}\left(\int_\Theta\big|\psi(\theta) - a\big|^2 g(x,\theta)\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}\left(\int_\Theta\frac{|\nabla_x g(x,\theta)|^2}{g(x,\theta)}\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}\\ &\le \mathcal{C}\big[g(x,\cdot)\pi\big]\left(\int_\Theta\big|\nabla\psi(\theta)\big|^2 g(x,\theta)\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}\mathcal{J}_\pi\big[g(x,\cdot)\big] \le K\left(\int_\Theta\big|\nabla\psi(\theta)\big|^2 g(x,\theta)\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}} \le K\,\mathrm{Lip}(\psi) \qquad (5.9)\end{aligned}$$

for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$, thus proving the claim. As seen in the proof of Theorem 2.2–(ii), this shows that there exists a $\mathcal{W}_1$-Lipschitz map $\mathbb{X} \ni x \mapsto \pi^*(\cdot|x) \in \mathcal{P}_1(\overline{\Theta})$, with Lipschitz constant K, which is a version of the kernel defined by (1.2). We are left to check that $\pi^*(\cdot|x)$ is Lipschitz with respect to $\mathcal{W}_2$ as well. Note that by assumption the second moment of $\pi^*(\cdot|x)$ is finite for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$.

We first notice that $G^*_\psi(x) := \int_{\overline{\Theta}} \psi(\theta)\pi^*(\mathrm{d}\theta|x)$ is the Lipschitz-continuous representative of G_ψ, for any $\psi \in C^1_c(\overline{\Theta})$. Indeed, the $\mathcal{W}_1$-Lipschitz estimate entails

$$\begin{aligned}\left|G^*_\psi(x_1) - G^*_\psi(x_2)\right| &= \left|\int_{\overline{\Theta}} \psi(\theta)\pi^*(\mathrm{d}\theta|x_1) - \int_{\overline{\Theta}} \psi(\theta)\pi^*(\mathrm{d}\theta|x_2)\right| \le \int_{\overline{\Theta}\times\overline{\Theta}} \left|\psi(\theta_1) - \psi(\theta_2)\right| \eta_{x_1,x_2}(\mathrm{d}\theta_1\,\mathrm{d}\theta_2) \\ &\le \mathrm{Lip}(\psi)\mathcal{W}_1\big(\pi^*(\cdot|x_1), \pi^*(\cdot|x_2)\big) \le K\,\mathrm{Lip}(\psi)|x_1 - x_2|\end{aligned}$$

for any $x_1, x_2 \in \mathbb{X}$, where $\eta_{x_1,x_2} \in \mathcal{P}(\overline{\Theta} \times \overline{\Theta})$ is an optimal coupling between $\pi^*(\cdot|x_1)$ and $\pi^*(\cdot|x_2)$ for the 1-Wasserstein distance. In particular for any $x_1, x_2 \in \mathbb{X}$, the map $[0, 1] \ni t \mapsto G^*_\psi(\mathbf{s}_{x_1,x_2}(t))$ is absolutely continuous for any $\psi \in C^1_c(\overline{\Theta})$, so that assumption *i)* of Theorem 5.1 is satisfied by the narrowly continuous curve $[0, 1] \ni t \mapsto \pi^*(\cdot|\mathbf{s}_{x_1,x_2}(t)) \in \mathcal{P}(\overline{\Theta})$.

Let $\mathcal{D}$ be a countable dense subset of $C^1_c(\overline{\Theta})$ (in the $C^1(\overline{\Theta})$ norm). Let $\nu \in \mathbb{S}^{m-1}$. We take advantage of the fact that any $\mathcal{L}^m$-null subset $\mathbb{Z}$ of $\mathbb{X}$ has the following property (by Fubini theorem): for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ there holds $\mathcal{L}^1(\mathbb{Z} \cap I_{\xi,\nu}) = 0$, and the $\mathcal{L}^{m-1}$-null set of ξ's where this property fails can be taken to be independent of $\psi \in \mathcal{D}$, since $\mathcal{D}$ is countable. Therefore, thanks to (5.9), for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$, we have the following:

$$\left|\nabla_x G^*_\psi\big(\mathbf{s}_{x_1,x_2}(t)\big)\right| = \inf_{a\in\mathbb{R}}\left|\int_\Theta \big(\psi(\theta) - a\big)\nabla_x g\big(\mathbf{s}_{x_1,x_2}(t), \theta\big)\pi(\mathrm{d}\theta)\right| \le K\left(\int_\Theta \left|\nabla\psi(\theta)\right|^2 g\big(\mathbf{s}_{x_1,x_2}(t), \theta\big)\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}$$

for a.e. $t \in (0, 1) \setminus N$ and any $\psi \in \mathcal{D}$, where N is a null set which is again independent of $\psi \in \mathcal{D}$. Moreover, for any $t \in (0, 1) \setminus N$ the latter inequality also holds for any $\psi \in \mathrm{span}\mathcal{D}$, due to the linearity of $\psi \mapsto G^*_\psi(x)$.

As a consequence, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}r}\int_{\overline{\Theta}} \psi(\theta)\pi^*\big(\mathrm{d}\theta|\mathbf{s}_{x_1,x_2}(r)\big)\bigg|_{r=t} &= \frac{\mathrm{d}}{\mathrm{d}r}\big(G^*_\psi\big(\mathbf{s}_{x_1,x_2}(r)\big)\big)\bigg|_{r=t} \le |x_1 - x_2| \cdot \left|\nabla_x G^*_\psi\big(\mathbf{s}_{x_1,x_2}(t)\big)\right| \\ &\le K|x_1 - x_2|\left(\int_\Theta \left|\nabla\psi(\theta)\right|^2 g(x, \theta)\pi(\mathrm{d}\theta)\right)^{\frac{1}{2}}\end{aligned}$$

for any $t \in (0, 1) \setminus N$ and any $\psi \in \mathrm{span}\mathcal{D}$. Whence,

$$\Psi_{x_1,x_2}(t) := \sup_{\psi\in\mathrm{span}\mathcal{D}}\left\{\frac{\mathrm{d}}{\mathrm{d}r}\int_{\overline{\Theta}} \psi(\theta)\pi^*\big(\mathrm{d}\theta|\mathbf{s}_{x_1,x_2}(r)\big)\bigg|_{r=t} : \int_{\overline{\Theta}} \left|\nabla\psi(\theta)\right|^2 \pi^*(\mathrm{d}\theta|g_{\mathbf{s}_{x_1,x_2}(t)}) \le 1\right\} \le K|x_1 - x_2|$$

for a.e. $t \in (0, 1)$. Here, $(0, 1) \ni t \mapsto \Psi_{x_1,x_2}(t)$ is measurable, being the supremum of the linear span of countably many measurable functions. Moreover, we deduce from the latter estimate that $\int_0^1 \Psi_{x,y}(t)\,\mathrm{d}t \le K|x - y|$ for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$. In particular, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$, assumption *ii)* of Theorem 5.1 is satisfied by the curve $[0, 1] \ni t \mapsto \pi^*(\cdot|\mathbf{s}_{x_1,x_2}(t)) \in \mathcal{P}(\overline{\Theta})$, and we also notice that (since $\pi^*(\cdot|x) \in \mathcal{P}_2(\overline{\Theta})$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$) we have $\pi^*(\cdot|\mathbf{s}_{x_1,x_2}(t)) \in \mathcal{P}_2(\overline{\Theta})$ for a.e. $t \in (0, 1)$ up to another $\mathcal{L}^{m-1}$-null set of ξ's in $\mathbb{X}_\nu$. Therefore, by applying Theorem 5.1, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$ we get that both $\pi^*(\cdot|x_1)$ and $\pi^*(\cdot|x_2)$ belong to $\mathcal{P}_2(\overline{\Theta})$ and that the bound in (1.1) is fulfilled for such x_1, x_2, with the $\mathcal{W}_2$ distance and with $L = K$. By the arbitrariness of ν and by the narrow continuity of $\mathbb{X} \ni x \mapsto \pi^*(\cdot|x) \in \mathcal{P}(\overline{\Theta})$, the $\mathcal{W}_2$-Lipschitz estimate extends to any $x_1, x_2 \in \mathbb{X}$. Indeed, given generic $x_1, x_2 \in \mathbb{X}$ with $x_1 \ne x_2$, letting $\nu := \frac{x_2 - x_1}{|x_2 - x_1|}$, it is enough to take sequences $x_{1,n} \to x_1$ and $x_{2,n} \to x_2$ such that, for every $n \in \mathbb{N}$, $\frac{x_{2,n} - x_{1,n}}{|x_{2,n} - x_{1,n}|} = \nu$ and such that (1.1) applies for any couple of points on the line $I_{P_\nu(x_{2,n}),\nu}$. Then, (1.1) applies for the couple $x_{1,n}, x_{2,n}$, for any n, and it passes to the limit by the narrow lower semicontinuity of $\mathcal{W}_2$, according to [5, Proposition 7.13]. □

Proof of Theorem 2.2-(iv). The proof is very similar to the previous ones. We first show that G_ψ from (5.7) belongs to $W^{1,\infty}(\mathbb{X})$ for any $\psi \in C^1_c(\overline{\Theta})$. It belongs indeed to $W^{1,1}_{\mathrm{loc}}(\mathbb{X})$ by Lemma 5.7, and to $L^\infty(\mathbb{X})$ with the same argument of the proof of Theorem 2.2–(i). Moreover, combining Lemma 5.7, the Sobolev inequality (2.3) with critical exponent

$p^* = \frac{dp}{d-p}$ ($p^* = +\infty$ if $p = d = 1$), and Hölder's inequality, we get

$$\big|\nabla_x G_\psi(x)\big| \le \inf_{a\in\mathbb{R}} \int_\Theta \big|\psi(\theta) - a\big|\big|\nabla_x g(x,\theta)\big|\,\mathrm{d}\theta \le \inf_{a\in\mathbb{R}} \|\psi - a\|_{L^{p^*}(\Theta)} \big\|\nabla_x g(x,\cdot)\big\|_{L^{\frac{p^*}{p^*-1}}(\Theta)}$$

$$\le \mathcal{S}_p(\Theta)\bigg(\int_\Theta \big|\nabla\psi(\theta)\big|^2 g(x,\theta)\,\mathrm{d}\theta\bigg)^{\frac{1}{2}} \bigg\|\frac{1}{g(x,\cdot)}\bigg\|^{\frac{1}{2}}_{L^{\frac{p}{2-p}}(\Theta)} \big\|\nabla_x g(x,\cdot)\big\|_{L^{\frac{p^*}{p^*-1}}(\Theta)} \tag{5.10}$$

for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. By assumption and by (5.10) we conclude that $\|\nabla_x G_\psi\|_{L^\infty(\mathbb{X})} \le K\mathrm{Lip}(\psi)$. As seen in the proof of Theorem 2.2–(ii), it follows that the probability kernel $\pi(\cdot|\cdot)$ defined by (1.2) admits a $\mathcal{W}_1$-Lipschitz continuous version $\mathbb{X} \ni x \mapsto \pi^*(\cdot|x) \in \mathcal{P}(\overline{\Theta})$. With the same argument of the proof of Theorem 2.2–(iii), the proof concludes by showing that $x \mapsto \pi^*(\cdot|x) \in \mathcal{P}_2(\overline{\Theta})$ is also $\mathcal{W}_2$-Lipschitz-continuous, with Lipschitz constant not exceeding K. ☐

5.3. *Moving domains*: *Proof of* Theorem 4.1 *and of* Theorem 4.2

We deal with solutions to nonhomogeneous Neumann boundary value problems, following the line of Theorem 5.6. Given a propability kernel $\pi(\cdot|\cdot)$, the curve $[0,1] \ni t \mapsto \pi(\cdot|\mathbf{s}_{x_1,x_2}(t)) \in \mathcal{P}(\overline{\Theta})$ depends on the two parameters x_1, x_2. Accordingly, we specify the notion of regular motion, which is essentially the same as Definition 5.3.

Definition 5.8 ($\mathbb{X}$-regular motion). Let $\Theta_* \subseteq \mathbb{R}^d$ and $\Theta_x \subseteq \Theta$ be nonempty open connected sets with locally Lipschitz boundary, for any $x \in \mathbb{X}$. We say that a smooth mapping $\mathbb{X} \times \Theta_* \ni (x,\theta) \mapsto \Phi_x(\theta)$ is a $\mathbb{X}$-regular motion if $[0,1] \times \Theta_* \ni (t,\theta) \mapsto \Phi_{\mathbf{s}_{x_1,x_2}(t)}(\theta)$ is regular motion according to Definition 5.3 for any $x_1, x_2 \in \mathbb{X}$ and $\Theta_x = \Phi_x(\Theta_*)$ for any $x \in \mathbb{X}$. In such assumptions, we further define for any $x \in \mathbb{X}$ and any $\nu \in \mathbb{S}^{m-1}$ the function $\mathbf{V}_x^\nu : \Theta_x \to \mathbb{R}^d$ (resp. $\mathbf{V}_x : \Theta_x \to \mathbb{R}^{d\times d}$) by $\mathbf{V}_x^\nu := \partial_\nu \Phi_x \circ \Phi_x^{-1}$ (resp. $\mathbf{V}_x := \nabla_x \Phi_x \circ \Phi_x^{-1}$).

Definition 5.9 (Regular extension). Let $g \in L^1_{\mathrm{loc}}(\mathbb{X} \times \Theta)$ satisfy $\int_\Theta g(x,\theta)\,\mathrm{d}\theta = 1$ for a.e. $x \in \mathbb{X}$. We say that g admits a regular extension if the following conditions are satisfied:

(i) there is a $\mathbb{X}$-regular motion $\Phi_x : \Theta_* \to \Theta_x$ according to Definition 5.8 such that $\Theta_x \equiv \{g(x,\cdot) > 0\}$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$;

(ii) there exists $\tilde{g} \in L^1_{\mathrm{loc}}(\mathbb{X} \times \Theta)$ such that $\tilde{g} \in W^{1,1}(\tilde{\mathbb{X}} \times \Theta)$ for any open set $\tilde{\mathbb{X}}$ compactly contained in $\mathbb{X}$ and such that

$$\tilde{g}(x,\theta)\mathbb{1}_{\Theta_x}(\theta) = g(x,\theta) \quad \text{for } \big(\mathcal{L}^m \otimes \mathcal{L}^d\big)\text{-a.e. } (x,\theta) \text{ in } \mathbb{X} \times \Theta. \tag{5.11}$$

Of course, for fixed x, the above identification $\equiv$ is understood up to $\mathcal{L}^d$-null sets of Θ. As $\tilde{g}$ from Definition 5.9 is in $W^{1,1}(\tilde{\mathbb{X}} \times \Theta)$, we shall use Sobolev regularity on linear submanifolds (see also [18, Theorem 2.5.3]). We summarize some basic facts in the following proposition.

Proposition 5.10. *Let $\tilde{g} \in W^{1,1}(\tilde{\mathbb{X}} \times \Theta)$ for any open set $\tilde{\mathbb{X}}$ compactly contained in $\mathbb{X}$. Let $\nu \in \mathbb{S}^{m-1}$. For $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any two distinct points $x_1, x_2 \in I_{\xi,\nu}$, the map $(t,\theta) \mapsto \tilde{g}(\mathbf{s}_{x_1,x_2}(t),\theta)$ belongs to $W^{1,1}((0,1) \times \Theta)$ and for $(\mathcal{L}^1 \otimes \mathcal{L}^d)$-a.e. $(t,\theta) \in (0,1) \times \Theta$ there hold*

$$\frac{\mathrm{d}}{\mathrm{d}t}\big\{\tilde{g}\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\big\} = |x_2 - x_1|\partial_\nu \tilde{g}\big(\mathbf{s}_{x_1,x_2}(t),\theta\big), \qquad \bigg|\frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\bigg| \le |x_2 - x_1|\big|\nabla_x \tilde{g}\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\big|, \tag{5.12}$$

and, in particular, $\partial_\nu \tilde{g}(x,\cdot) \in L^1(\Theta)$ for $\mathcal{L}^1$-a.e. $x \in I_{\xi,\nu}$.

Proof. The ACL representative of $\tilde{g}$ (here not relabeled) has the ACL property on almost any lower dimensional hyperplane intersecting $\tilde{\mathbb{X}} \times \Theta$. Let $\nu \in \mathbb{S}^{m-1}$; for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$, the map $(t,\theta) \mapsto g(\xi + t\nu, \theta)$ belongs therefore to $W^{1,1}((t_1,t_2) \times \Theta)$ for any $t_1 < t_2$ such that $\xi + t_i\nu \in \mathbb{X}$, $i = 1,2$. The map $(t,\theta) \mapsto \tilde{g}(\mathbf{s}_{x_1,x_2}(t),\theta)$ belongs to $W^{1,1}((0,1) \times \Theta)$ as the composition of the latter with the segment parametrization $[0,1] \ni t \mapsto \mathbf{s}_{x_1,x_2}(t)$, where $x = \xi + t_1\nu$ and $y = \xi + t_2\nu$. Then, (5.12) follows from the fact that $\tilde{g}$ has classical ν-directional derivative almost everywhere, coinciding with the scalar product of ν with the gradient. See for instance [49, Theorem 4, pp. 200]. ☐

Associated to a function g as in Definition 5.9, we consider the boundary value problem (4.7), where $\mathbf{V}_x^\nu := \partial_\nu \Phi_x \circ (\Phi_x)^{-1}$. The precise notation for (4.7) is the same of problem (5.4), apart from the $\mathbb{X}$-valued index x instead of t.

Therefore, for those couples x, ν such that $g(x,\cdot) \in W^{1,1}(\Theta_x)$ and $\partial_\nu \tilde{g}(x,\cdot) \in L^1(\Theta_x)$, we may define a weak solution $u_x^\nu \in H^1(\Theta_x, g(x,\cdot))$ to (4.7) by means of Definition 5.5.

Proof of Theorem 4.1. Throughout this proof, for notational ease, we shorten the expression $\mathbf{s}_{x_1,x_2}(t)$ to $\mathbf{s}(t)$, whenever it is clear which couple (x_1, x_2) we are referring to. We start by preliminarily observing that, given $\nu \in \mathbb{S}^{m-1}$, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any couple of distinct points $x_1, x_2 \in I_{\xi,\nu}$, the map $(t,\theta) \mapsto \tilde{g}(\mathbf{s}(t),\theta)$ belongs to $W^{1,1}((0,1)\times\Theta)$, by Proposition 5.10. This fact entails that $g(\mathbf{s}(t),\cdot) \in W^{1,1}(\Theta_{\mathbf{s}(t)})$ and $\frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}(\mathbf{s}(t),\cdot) \in L^1(\Theta_{\mathbf{s}(t)})$ for a.e. $t \in (0,1)$. Therefore, we may take advantage of the notion of weak solution to problem (5.4) as given in Definition 5.5, with $g_t(\cdot)$ therein replaced by $g(\mathbf{s}(t),\cdot)$, and Φ_t therein replaced by $\Phi_{\mathbf{s}(t)}$.

The proof is an application of Theorem 5.6, for almost every line in $\mathbb{X}$ in any given direction. Indeed, let us consider an ACL representative of the regular extension $\tilde{g}$, that we still denote by $\tilde{g}$. Of course, combining the assumptions on g with (5.11), we have $\int_{\Theta_x} \tilde{g}(x,\theta)\,\mathrm{d}\theta = 1$ and $\int_{\Theta_x} |\theta|^2 \tilde{g}(x,\theta)\,\mathrm{d}\theta < +\infty$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. Let $\nu \in \mathbb{S}^{m-1}$. For $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$, we apply Theorem 5.6 to obtain $\tilde{\pi}(\cdot|x_1) \in \mathcal{P}_2(\overline{\Theta})$, $\tilde{\pi}(\cdot|x_2) \in \mathcal{P}_2(\overline{\Theta})$ and

$$\mathcal{W}_2\big(\tilde{\pi}(\cdot|x_1), \tilde{\pi}(\cdot|x_2)\big) \le |x_1 - x_2| \int_0^1 \left(\int_\Theta \big|\nabla u^\nu_{\mathbf{s}(t)}(\theta)\big|^2 g\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right)^{\frac{1}{2}} \mathrm{d}t, \tag{5.13}$$

where $\tilde{\pi}(\cdot|x) := \tilde{g}(x,\cdot)\mathcal{L}^d \llcorner \Theta_x$ is a representative of the kernel $\pi(\cdot|\cdot)$, in view of (5.11). We notice that the appearance of the factor $|x_1 - x_2|$ is due to (5.12) and to the identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\Phi_{\mathbf{s}(t)}(\theta)\big) = |x_1 - x_2| \partial_\nu \Phi_{\mathbf{s}(t)}(\theta). \tag{5.14}$$

As a consequence, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any couple of distinct points $x_1, x_2 \in I_{\xi,\nu}$, we get $\mathcal{W}_2(\tilde{\pi}(\cdot|x_1), \tilde{\pi}(\cdot|y_2)) \le K|x_1 - x_2|$. The last inequality follows from (5.13) and (4.8): we bound once more the $\mathcal{L}^1$-essential supremum on $(0,1)$ with the $\mathcal{L}^m$-essential supremum on $\mathbb{X}$, for all but a $\mathcal{L}^{m-1}$-null set of lines in a given direction.

Now, let $\psi \in C_c^1(\overline{\Theta})$ and $\tilde{G}_\psi(x) := \int_{\Theta_x} \psi(\theta)\tilde{g}(x,\theta)\,\mathrm{d}\theta$, so that, by Definition 5.9, we get $|\tilde{G}_\psi(x)| \le \sup_\Theta |\psi|$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. By performing the same estimate of the proof of Theorem 5.6, also taking (5.12) and (5.14) into account, we have the following: given any $\nu \in \mathbb{S}^{m-1}$, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any couple of distinct points $x_1, x_2 \in I_{\xi,\nu}$ there holds

$$\frac{\mathrm{d}}{\mathrm{d}r}\tilde{G}_\psi\big(\mathbf{s}(r)\big)\bigg|_{r=t} = |x_1 - x_2| \int_{\Theta_{\mathbf{s}(t)}} \nabla\psi(\theta) \cdot \nabla u^\nu_{\mathbf{s}(t)}(\theta) g\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta \tag{5.15}$$

for a.e. $t \in (0,1)$, where we have used the definition of u_z^ν as solution to the boundary value problem (4.7). Taking (5.6) into account, (5.15) can be rephrased as follows: $\nu \cdot \nabla_x \tilde{G}_\psi(x) = \int_{\Theta_x} \nabla\psi(\theta) \cdot \nabla u_x^\nu(\theta) g(x,\theta)\,\mathrm{d}\theta$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. Since $\int_{\Theta_x} \tilde{g}(x,\theta)\,\mathrm{d}\theta = 1$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$, we further estimate by the Cauchy–Schwarz inequality and (4.8), to get

$$\big|\nu \cdot \nabla_x \tilde{G}_\psi(x)\big| \le \left(\int_{\Theta_x} \big|\nabla\psi(\theta)\big|^2 g(x,\theta)\,\mathrm{d}\theta\right)^{\frac{1}{2}} \left(\int_{\Theta_x} \big|\nabla u_x^\nu(\theta)\big|^2 g(x,\theta)\,\mathrm{d}\theta\right)^{\frac{1}{2}} \le K \mathrm{Lip}(\psi)$$

for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. Here, K is independent of ν by assumption, hence $|\nabla_x \tilde{G}_\psi(x)| \le K\mathrm{Lip}(\psi)$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. Note that $\tilde{G}_\psi$ is a representative of the $L^\infty(\mathbb{X})$ function $G_\psi(\cdot) := \int_\Theta \psi(\theta) g(\cdot,\theta)\,\mathrm{d}\theta$. Having shown that $|G_\psi(x)| \le \sup_\Theta |\psi|$ and $|\nabla_x G_\psi(x)| \le K\mathrm{Lip}(\psi)$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$, by the same argument as in the proof of Theorem 2.2–(ii), we obtain the existence of a $\mathcal{W}_1$-Lipschitz representative $\pi^*(\cdot|\cdot)$ for the probability kernel $\pi(\cdot|\cdot)$.

Since $G_\psi^*(x) := \int_\Theta \psi(\theta)\pi^*(\mathrm{d}\theta|x)$ is Lipschitz on $\mathbb{X}$ and $\tilde{G}_\psi$ is ACL, the two functions coincide pointwise everywhere on almost every segment in a given direction $\nu \in \mathbb{S}^{m-1}$. Taking a countable dense subset $\mathcal{D}$ of $\psi \in C_c^1(\overline{\Theta})$ (in the $C^1(\overline{\Theta})$ norm) shows that $\pi^*(\cdot|\cdot)$ coincides with $\tilde{\pi}(\cdot|\cdot)$ on almost every line segment in the same direction ν. Therefore, given $\nu \in \mathbb{S}^{m-1}$, $\pi^*(\cdot|\cdot)$ itself satisfies $\mathcal{W}_2(\pi^*(\cdot|x_1), \pi^*(\cdot|x_2)) \le K|x_1 - x_2|$ for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any couple of distinct points $x_1, x_2 \in I_{\xi,\nu}$. The result follows by the same argument at the end of the proof of Theorem 2.2–(iii). $\square$

In most situations a solution to (4.7) is not at disposal. Therefore, with some stronger assumptions we try to give an estimate of the norm of the solution in its dual formulation as seen in Theorem 5.1. This is done in Theorem 4.2. We recall that the definition of the Fisher functionals $\mathcal{J}_1$ and $\mathcal{J}_2$ is given in (4.9).

Proof of Theorem 4.2. Throughout this proof, as in the previous one, we simplify the notation by writing $\mathbf{s}(t)$ in place of $\mathbf{s}_{x_1,x_2}(t)$, since no ambiguity arises. We apply Reynolds transport formula. Let us consider an ACL representative of

$\tilde{g}$, still denoted by $\tilde{g}$. Let $\psi \in C_c^1(\overline{\Theta})$. Given any $\nu \in \mathbb{S}^{m-1}$, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ we take any $x_1, x_2 \in I_{\xi,\nu}$ and we obtain the absolute continuity of the map $[0,1] \ni t \mapsto \int_{\Theta_{\mathbf{s}(t)}} \psi(\theta)\tilde{g}(\mathbf{s}(t),\theta)\,\mathrm{d}\theta$, along with

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}r}\int_{\Theta_{\mathbf{s}(t)}} \psi(\theta)\tilde{g}\big(\mathbf{s}(r),\theta\big)\,\mathrm{d}\theta\bigg|_{r=t} \\
&\quad= \frac{\mathrm{d}}{\mathrm{d}r}\int_{\Theta} \overline{\psi}_{\mathbf{s}(t)}(\theta)\tilde{g}\big(\mathbf{s}(r),\theta\big)\,\mathrm{d}\theta\bigg|_{r=t} \\
&\quad= |x_1-x_2|\int_{\Theta_{\mathbf{s}(t)}} \overline{\psi}_{\mathbf{s}(t)}(\theta)\partial_\nu\tilde{g}\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta + |x_1-x_2|\int_{\Theta_{\mathbf{s}(t)}} \mathrm{div}\big(\overline{\psi}_{\mathbf{s}(t)}(\theta)\tilde{g}\big(\mathbf{s}(t),\theta\big)\mathbf{V}_{\mathbf{s}(t)}(\theta)\big)\,\mathrm{d}\theta
\end{aligned} \tag{5.16}$$

for a.e. $t \in (0,1)$. Here, we have used Lemma A.1 and (5.12), and we have introduced the function

$$\overline{\psi}_{\mathbf{s}(t)}(\cdot) := \psi(\cdot) - \int_{\Theta_{\mathbf{s}(t)}} \psi(\theta)\tilde{g}\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta, \quad t \in [0,1].$$

Let us proceed by estimating the two terms in the right hand side of (5.16). The first term in the right hand side of (5.16) can be treated as in the proof of Theorem 2.2–(iii), so that by Cauchy–Schwarz inequality

$$\int_{\Theta_{\mathbf{s}(t)}} \overline{\psi}_{\mathbf{s}(t)}(\theta)\partial_\nu\tilde{g}\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta \le \left(\int_\Theta \big|\overline{\psi}_{\mathbf{s}(t)}(\theta)\big|^2 g\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right)^{\frac{1}{2}} \mathcal{J}_1\big(\tilde{g}\big(\mathbf{s}(t),\cdot\big)\big)$$

and then the Poincaré inequality (2.2) implies

$$\int_{\Theta_{\mathbf{s}(t)}} \overline{\psi}_{\mathbf{s}(t)}(\theta)\partial_\nu\tilde{g}\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta \le \mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\left(\int_\Theta \big|\nabla\psi(\theta)\big|^2 g\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right)^{\frac{1}{2}} \mathcal{J}_1\big[\tilde{g}\big(\mathbf{s}(t),\cdot\big)\big]. \tag{5.17}$$

Note that $\tilde{g}(\mathbf{s}(t),\cdot)\mathbb{1}_{\Theta_{\mathbf{s}(t)}}(\cdot)$ and $g(\mathbf{s}(t),\cdot)$ coincide for a.e. $t \in (0,1)$ as $L^1(\Theta)$ functions. The divergence term in (5.16) can be estimated by Cauchy–Schwarz and Poincaré's inequalities: indeed, since $\int_\Theta g(x,\theta)\,\mathrm{d}\theta = 1$ and $|\mathbf{V}_z^\nu| \le |\mathbf{V}_z|$, making use of the shorthand $\mathcal{A}_t := \|\mathbf{V}_{\mathbf{s}(t)}\|_{W^{1,\infty}(\Theta_{\mathbf{s}(t)})}$, there holds

$$\begin{aligned}
&\int_{\Theta_{\mathbf{s}(t)}} \mathrm{div}\big(\overline{\psi}_{\mathbf{s}(t)}(\theta)\tilde{g}\big(\mathbf{s}(t),\theta\big)\mathbf{V}_{\mathbf{s}(t)}(\theta)\big)\,\mathrm{d}\theta \\
&\quad\le \mathcal{A}_t\left(\int_{\Theta_{\mathbf{s}(t)}} \big|\nabla\big(\overline{\psi}_{\mathbf{s}(t)}(\theta)\tilde{g}\big(\mathbf{s}(t),\theta\big)\big)\big|\,\mathrm{d}\theta + \int_{\Theta_{\mathbf{s}(t)}} \big|\overline{\psi}_{\mathbf{s}(t)}(\theta)\big|\tilde{g}\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right) \\
&\quad\le \mathcal{A}_t\left[\left(\int_{\Theta_{\mathbf{s}(t)}} |\nabla\overline{\psi}_{\mathbf{s}(t)}|^2\tilde{g}\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right)^{\frac{1}{2}} + \int_{\Theta_{\mathbf{s}(t)}} |\overline{\psi}_{\mathbf{s}(t)}|\big(\tilde{g}\big(\mathbf{s}(t),\theta\big) + \big|\nabla\tilde{g}\big(\mathbf{s}(t),\theta\big)\big|\big)\,\mathrm{d}\theta\right] \\
&\quad\le \mathcal{A}_t\left(\int_\Theta \big|\nabla\psi(\theta)\big|^2 g\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right)^{\frac{1}{2}}\big(1 + \mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big] + \mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\mathcal{J}_2\big(\tilde{g}\big(\mathbf{s}(t),\cdot\big)\big)\big).
\end{aligned}$$

By plugging (5.17) and the latter estimate into (5.16), we get the following: given any $\nu \in \mathbb{S}^{m-1}$, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any couple $x_1, x_2 \in I_{\xi,\nu}$, there holds

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}r}\int_{\Theta_{\mathbf{s}(t)}} \psi(\theta)\tilde{g}\big(\mathbf{s}(r),\theta\big)\,\mathrm{d}\theta\bigg|_{r=t} \\
&\quad\le \mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\left(\int_\Theta \big|\nabla\psi(\theta)\big|^2 g\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right)^{\frac{1}{2}} \mathcal{J}_1\big[\tilde{g}\big(\mathbf{s}(t),\cdot\big)\big]|x_1-x_2| \\
&\qquad + \mathcal{A}_t\left(\int_{\Theta_{\mathbf{s}(t)}} \big|\nabla\psi(\theta)\big|^2 g\big(\mathbf{s}(t),\theta\big)\,\mathrm{d}\theta\right)^{\frac{1}{2}}\big(1 + \mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\big(1 + \mathcal{J}_2\big(\tilde{g}\big(\mathbf{s}(t),\cdot\big)\big)\big)\big)|x_1-x_2|
\end{aligned} \tag{5.18}$$

for a.e. $t \in (0,1)$. Now, let $G_\psi(\cdot) := \int_\Theta g(\cdot,\theta)\,\mathrm{d}\theta$, so that $|G_\psi(x)| \le \sup_{\overline{\Theta}}|\psi|$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. However, by (5.18), by assumption and by the same argument as in the proof of Theorem 4.1, we get $|\nabla G_\psi(x)| \le K\,\mathrm{Lip}(\psi)$ for $\mathcal{L}^m$-a.e. $x \in \mathbb{X}$. Again this shows that there exists a $\mathcal{W}_1$-Lipscthiz representative $\pi^*(\cdot|\cdot)$ of the kernel $\pi(\cdot|\cdot)$.

We now let $\tilde{\pi}(\cdot|x) := \tilde{g}(x,\cdot)\mathcal{L}^d \llcorner \Theta_x$, which gives a representative of the kernel $\pi(\cdot|\cdot)$. Let $\mathcal{D}$ be a countable dense subset of $C^1_c(\overline{\Theta})$ (in the $C^1(\overline{\Theta})$ norm) and let $\nu \in \mathbb{S}^{m-1}$. For $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$, from (5.18) we get

$$\begin{aligned}\Psi_{x_1,x_2}(t) &:= \sup\left\{\frac{\mathrm{d}}{\mathrm{d}r}\int_\Theta \psi(\theta)\tilde{\pi}\big(\mathrm{d}\theta|\mathbf{s}(t)\big)\Big|_{r=t} : \psi \in \operatorname{span}\mathcal{D}, \int_\Theta \big|\nabla\psi(\theta)\big|^2\tilde{\pi}\big(\mathrm{d}\theta|\mathbf{s}(t)\big) \le 1\right\} \\ &\le |x_1-x_2|\|\mathbf{V}_{\mathbf{s}(t)}\|_{W^{1,\infty}(\Theta_{\mathbf{s}(t)})}\big(1+\mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\big(1+\mathcal{J}_2\big[g\big(\mathbf{s}(t),\cdot\big)\big]\big)\big) + |x_1-x_2|\mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\mathcal{J}_1\big[g\big(\mathbf{s}(t),\cdot\big)\big]\end{aligned}$$

for a.e. $t \in (0,1)$. Therefore, the estimate

$$\begin{aligned}\int_0^1 \Psi_{x_1,x_2}(t)\,\mathrm{d}t &\le |x_1-x_2|\int_0^1 \|\mathbf{V}_{\mathbf{s}(t)}\|_{W^{1,\infty}(\Theta_{\mathbf{s}(t)})}\big(1+\mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\big(1+\mathcal{J}_2\big[g\big(\mathbf{s}(t),\cdot\big)\big]\big)\big)\,\mathrm{d}t \\ &\quad + |x_1-x_2|\int_0^1 \mathcal{C}\big[g\big(\mathbf{s}(t),\cdot\big)\big]\mathcal{J}_1\big[g\big(\mathbf{s}(t),\cdot\big)\big]\,\mathrm{d}t \le K|x_1-x_2|\end{aligned}$$

holds for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$. By invoking Theorem 5.1, we deduce that $\tilde{\pi}(\cdot|x_1)$ and $\tilde{\pi}(\cdot|x_2)$ are in $\mathcal{P}_2(\overline{\Theta})$ and that $\mathcal{W}_2(\tilde{\pi}(\cdot|x_1),\tilde{\pi}(\cdot|x_2)) \le K|x_1-x_2|$, for $\mathcal{L}^{m-1}$-a.e. $\xi \in \mathbb{X}_\nu$ and any $x_1, x_2 \in I_{\xi,\nu}$. By the same argument as in the proof of Theorem 4.1, the same conclusion holds for $\pi^*(\cdot|\cdot)$, which identifies with $\tilde{\pi}(\cdot|\cdot)$ on almost every line in any given direction. But $x \mapsto \pi^*(\cdot|x)$ is $\mathcal{W}_1$-Lipschitz on the whole of $\mathbb{X}$, so that we conclude by the argument already explained at the end of the proof of Theorem 2.2–(iii). □

5.4. *Infinite-dimensional sample space*: *Proof of Theorem* 4.5

We next provide the proof of the results that deal with infinite-dimensional sample space from Section 4.4. In this case we shall prove that a Lipschitz estimate holds for 'good couples' $(x_1,x_2) \in \mathbb{B}(\mathbb{Z})$ and then we invoke the Lipschitz extension result from Lemma A.4 in the Appendix.

Proof of Theorem 4.5. We start by proving point (i). Let $\psi \in C^1_c(\overline{\Theta})$ and $G_\psi(x) := \int_\Theta g(x,\theta)\pi(\mathrm{d}\theta)$. Given a couple $(x_1,x_2) \in \mathbb{B}(\mathbb{Z})$, for π-a.e. $\theta \in \Theta$, the map $[0,1] \ni t \mapsto g(\mathbf{s}_{x_1,x_2}(t),\theta)$ has, by assumption, the following properties: it is absolutely continuous, and for a.e. $t \in (0,1)$ the point $(\mathbf{s}_{x_1,x_2}(t),\theta)$ is a Gateaux-differentiability point of g with respect to the x-variable. Therefore, by an application of Fubini's theorem, we get

$$\begin{aligned}G_\psi(x_2)-G_\psi(x_1) &= \int_\Theta \psi(\theta)\big(g\big(\mathbf{s}_{x_1,x_2}(1),\theta\big)-g\big(\mathbf{s}_{x_1,x_2}(0),\theta\big)\big)\pi(\mathrm{d}\theta) = \int_\Theta \psi(\theta)\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\big(g\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\big)\,\mathrm{d}t\,\pi(\mathrm{d}\theta) \\ &= \int_0^1\int_\Theta \psi(\theta)\big\langle D_x g\big(\mathbf{s}_{x_1,x_2}(t),\theta\big), x_2-x_1\big\rangle\pi(\mathrm{d}\theta)\,\mathrm{d}t \\ &\le \|x_2-x_1\|_{\mathbb{V}}\sup_\Theta|\psi|\operatorname*{ess\,sup}_{t\in(0,1)}\int_\Theta \big\|D_x g\big(\mathbf{s}_{x_2,x_1}(t),\theta\big)\big\|_{\mathbb{V}'}\pi(\mathrm{d}\theta) \\ &\le \|x_2-x_1\|_{\mathbb{V}}\sup_\Theta|\psi|\,\lambda\text{-}\operatorname*{ess\,sup}_{x\in\mathbb{X}}\int_\Theta \big\|D_x g(x,\theta)\big\|_{\mathbb{V}'}\pi(\mathrm{d}\theta).\end{aligned}$$

In the last inequality we have used the fact that any function $F : \mathbb{X} \to \mathbb{R}$ satisfies

$$\operatorname*{ess\,sup}_{t\in(0,1)} F\big(\mathbf{s}_{x_1,x_2}(t)\big) \le \lambda\text{-}\operatorname*{ess\,sup}_{x\in\mathbb{X}} F(x), \tag{5.19}$$

since $\mathcal{H}^1([x_1,x_2]\cap\mathbb{Z}) = 0$ as $(x_1,x_2) \in \mathbb{B}(\mathbb{Z})$. Therefore, we get

$$2\,\mathrm{d}_{\mathrm{TV}}\big(\pi(\cdot|x_2),\pi(\cdot|x_1)\big) = \sup_{\substack{\psi\in C_c(\Theta)\\|\psi|\le 1}}\big(G_\psi(x_2)-G_\psi(x_1)\big) \le \sup_{\substack{\psi\in C_c(\Theta)\\|\psi|\le 1}}\|\nabla G_\psi\|_{L^\infty(\mathbb{X})}\|x_1-x_2\|_{\mathbb{V}} \le K\|x_1-x_2\|_{\mathbb{V}}$$

for any $(x_1,x_2) \in \mathbb{B}(\mathbb{Z})$. By invoking the extension result from Lemma A.4, the mapping $x \mapsto \pi(\cdot|x) \in \mathcal{P}(\overline{\Theta})$ admits a Lipschitz continuous extension $\mathbb{X} \ni x \mapsto \pi^*(\cdot|x) \in \mathcal{P}(\overline{\Theta})$ with respect to the total variation distance. Since $\pi(\Theta) = 1$, the continuity in total variation also shows that $\pi^*(\Theta|x) = 1$ for any $x \in \mathbb{X}$. This ends the proof of (i).

The result in point (ii) is obtained by introducing the Poincaré inequality (2.2) in the computations of the proof of point (i), as done in the proof of Theorem 2.2–(ii).

Let us conclude by proving (iii). The proof is similar to the one of Theorem 2.2–(iii). Let $\mathcal{D}$ denote a countable dense subset of $C_c^1(\overline{\Theta})$ (in the $C^1(\overline{\Theta})$ norm), let $G_\psi(x) := \int_\Theta \psi(\theta)g(x,\theta)\pi(\mathrm{d}\theta)$. Let us consider a couple $(x_1,x_2) \in \mathbb{B}(\mathbb{Z})$. Thanks to the assumptions, we have the absolute continuity of the map $t \mapsto g(\mathbf{s}_{x_1,x_2}(t),\theta)$ for π-a.e. $\theta \in \Theta$ and we may apply Lemma A.2 from the Appendix, so that the map $t \mapsto \int_\Theta \psi(\theta)g(\mathbf{s}_{x_1,x_2}(t),\theta)\pi(\mathrm{d}\theta)$ is absolutely continuous for any $\psi \in \mathcal{D}$ and we may differentiate under integral sign to get for a.e. $t \in (0,1)$

$$\frac{\mathrm{d}}{\mathrm{d}r}\int_\Theta \psi(\theta)g\big(\mathbf{s}_{x_1,x_2}(r),\theta\big)\pi(\mathrm{d}\theta)\bigg|_{r=t} = \int_\Theta \psi(\theta)\bigg(\frac{\mathrm{d}}{\mathrm{d}t}g\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\bigg)\pi(\mathrm{d}\theta).$$

As usual, the $\mathcal{L}^1$-null set of non-differentiability points of the map $t \mapsto \int_\Theta \psi(\theta)g(\mathbf{s}_{x_1,x_2}(t),\theta)\,\mathrm{d}\theta$ is independent of $\psi \in \mathcal{D}$, since $\mathcal{D}$ is countable. By considering the Gateaux-differentiability property of g, we get, for a.e. $t \in (0,1)$,

$$\frac{\mathrm{d}}{\mathrm{d}r}\int_\Theta \psi(\theta)g\big(\mathbf{s}_{x_1,x_2}(r),\theta\big)\pi(\mathrm{d}\theta)\bigg|_{r=t} \le \|x_1-x_2\|_{\mathbb{V}}\int_\Theta \big|\psi(\theta)\big|\big\|D_x g\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\big\|_{\mathbb{V}'}\pi(\mathrm{d}\theta).$$

Hence, combining the Cauchy–Schwarz and the Poincaré inequality (2.2), we obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}r}\int_\Theta \psi(\theta)g\big(\mathbf{s}_{x_1,x_2}(r),\theta\big)\pi(\mathrm{d}\theta)\bigg|_{r=t}\\
&\quad= \inf_{a\in\mathbb{R}}\frac{\mathrm{d}}{\mathrm{d}r}\int_\Theta \big(\psi(\theta)-a\big)g\big(\mathbf{s}_{x_1,x_2}(r),\theta\big)\pi(\mathrm{d}\theta)\bigg|_{r=t}\\
&\quad\le \|x_1-x_2\|_{\mathbb{V}}\bigg(\int_\Theta \frac{\|D_x g(\mathbf{s}_{x_1,x_2}(t),\theta)\|^2_{\mathbb{V}'}}{g(\mathbf{s}_{x_1,x_2}(t),\theta)}\pi(\mathrm{d}\theta)\bigg)^{\frac12}\inf_{a\in\mathbb{R}}\bigg(\int_\Theta \big|\psi(\theta)-a\big|^2 g\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\pi(\mathrm{d}\theta)\bigg)^{\frac12}\\
&\quad\le \|x_1-x_2\|_{\mathbb{V}}\mathcal{C}\big[g\big(\mathbf{s}_{x_1,x_2}(t),\cdot\big)\pi\big]\bigg(\int_\Theta \big|\nabla\psi(\theta)\big|^2 g\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\pi(\mathrm{d}\theta)\bigg)^{\frac12}\mathcal{J}_\pi\big[g\big(\mathbf{s}_{x_1,x_2}(t),\cdot\big)\big].
\end{aligned}$$

Whence,

$$\begin{aligned}
\Psi_{x_1,x_2}(t) &:= \sup_{\psi\in\operatorname{span}\mathcal{D}}\bigg\{\frac{\mathrm{d}}{\mathrm{d}r}\int_\Theta \psi(\theta)g\big(\mathbf{s}_{x_1,x_2}(r),\theta\big)\pi(\mathrm{d}\theta)\bigg|_{r=t} : \int_\Theta \big|\psi(\theta)\big|^2 g\big(\mathbf{s}_{x_1,x_2}(t),\theta\big)\,\mathrm{d}\theta \le 1\bigg\}\\
&\le \mathcal{C}\big[g\big(\mathbf{s}_{x_1,x_2}(t),\cdot\big)\pi\big]\mathcal{J}_\pi\big[g\big(\mathbf{s}_{x_1,x_2}(t),\cdot\big)\big]\|x_1-x_2\|_{\mathbb{V}}.
\end{aligned}$$

Combining the latter estimate with (5.19), we conclude that for any $(x_1,x_2) \in \mathbb{B}(\mathbb{Z})$ there holds

$$\int_0^1 \Psi_{x_1,x_2}(t)\,\mathrm{d}t \le \|x_1-x_2\|_{\mathbb{V}}\int_0^1 \mathcal{C}\big[g\big(\mathbf{s}_{x_1,x_2}(t),\cdot\big)\pi\big]\mathcal{J}_\pi\big[g\big(\mathbf{s}_{x_1,x_2}(t),\cdot\big)\pi\big]\,\mathrm{d}t \le K\|x_1-x_2\|_{\mathbb{V}}.$$

Hence, an application of Theorem 5.1 shows that $x \mapsto \pi(\cdot|x)$ satisfies the desired estimate for any $(x_1,x_2) \in \mathbb{B}(\mathbb{Z})$. The Lipschitz extension property from Lemma A.4, applied to the complete metric space $(\mathcal{P}_2(\overline{\Theta}),\mathcal{W}_2)$, yields the result. □

Appendix

A.1. *A proof of Reynolds transport formula*

We give here a proof of some useful calculus formulae that are often needed through the paper. The following is a proof of Reynolds transport theorem, see also, for instance, [53, Théorèm 5.2.2] or [51, Section 10]. The proof is given for domains that vary according to a regular motion as defined in Section 5.1. In the following lemma, we make use of the notation $C_a^1(\overline{\Theta}) := \{\psi + a : a \in \mathbb{R}, \psi \in C_c^1(\overline{\Theta})\}$ where, as usual, Θ is an open connected subset of $\mathbb{R}^d$.

Lemma A.1. *Let $\tilde g \in W^{1,1}((a,b)\times\Theta)$. Let $[a,b]\times\Theta_* \ni (t,\theta) \mapsto \Phi_t(\theta) \in \mathbb{R}^d$ be a regular motion in Θ according to* Definition 5.3, *with $\Theta_t := \Phi_t(\Theta_*)$. Then (any ACL representative of) $\tilde g$ is such that $[a,b] \ni t \mapsto \int_{\Theta_t}\psi(\theta)\tilde g(t,\theta)\,\mathrm{d}\theta$ is absolutely continuous for any $\psi \in C_a^1(\overline{\Theta})$. Moreover, given $\psi \in C_a^1(\overline{\Theta})$ there holds for a.e. $t \in (a,b)$*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Theta_t}\psi(\theta)\tilde g(t,\theta)\,\mathrm{d}\theta = \int_{\Theta_t}\psi(\theta)\partial_t\tilde g(t,\theta)\,\mathrm{d}\theta + \int_{\Theta_t}\nabla\cdot\big(\psi(\theta)\tilde g(t,\theta)(\partial_t\Phi_t\circ\Psi_t)(\theta)\big)\,\mathrm{d}\theta,$$

where $\Psi_t : \Theta_t \to \Theta_*$ is the inverse of Φ_t. If $\psi \in C^1_c(\overline{\Theta})$ we also have for a.e. $t \in (a,b)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Theta_t}\psi(\theta)\tilde g(t,\theta)\,\mathrm{d}\theta = \int_{\Theta_t}\psi(\theta)\partial_t\tilde g(t,\theta)\,\mathrm{d}\theta + \int_{\partial\Theta_t}\psi(\sigma)\tilde g(t,\sigma)\mathbf{n}_t(\sigma)\cdot(\partial_t\Phi_t\circ\Psi_t)(\sigma)\mathcal{H}^{d-1}(\mathrm{d}\sigma),$$

where $\mathbf{n}_t$ denotes the exterior normal to Θ_t, and the $L^1_{\mathrm{loc}}(\partial\Theta_t)$ boundary trace of $\tilde g(t,\cdot)$ on $\partial\Theta_t$ appears in the last term.

Proof. As Φ_t is a global diffeomorphism of Θ_* onto Θ_t for any $t \in [a,b]$, then $(\mathbf{i},\Phi_t)$ is a global diffeomorphism of $(a,b)\times\Theta_*$ onto $\{(t,\theta)\in(a,b)\times\mathring{\Theta} : \theta\in\Theta_t\}$, whose Jacobian determinant is bounded away from 0 and $+\infty$ (tanks to the assumptions in Definition 5.3). Thus, $\tilde g\circ(\mathbf{i},\Phi_t)\in W^{1,1}((a,b)\times\Theta_*)$.

Let $\psi \in C^1_a(\overline{\Theta})$. By change of variables we have for a.e.$t \in (a,b)$

$$\int_{\Theta_t}\psi(\theta)\tilde g(t,\theta)\,\mathrm{d}\theta = \int_{\Theta_*}\tilde g\big(t,\Phi_t(\theta)\big)\psi\big(\Phi_t(\theta)\big)\det\nabla\Phi_t(\theta)\,\mathrm{d}\theta. \tag{A.1}$$

By distributional chain rule we have $\partial_t\tilde g(t,\Phi_t(\theta)) + \nabla\tilde g(t,\Phi_t(\theta))\cdot\partial_t\Phi_t(\theta) = \frac{\mathrm{d}}{\mathrm{d}t}\tilde g(t,\Phi_t(\theta))$ and similarly

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\big[\tilde g\big(t,\Phi_t(\theta)\big)\psi\big(\Phi_t(\theta)\big)\det\nabla\Phi_t(\theta)\big]\\
&\quad= \psi\big(\Phi_t(\theta)\big)\det\nabla\Phi_t(\theta)\frac{\mathrm{d}}{\mathrm{d}t}\tilde g\big(t,\Phi_t(\theta)\big)\\
&\qquad+ \tilde g\big(t,\Phi_t(\theta)\big)\det\nabla\Phi_t(\theta)\nabla\psi\big(\Phi_t(\theta)\big)\cdot\partial_t\Phi_t(\theta) + \tilde g\big(t,\Phi_t(\theta)\big)\psi\big(\Phi_t(\theta)\big)\mathrm{Tr}\big((\partial_t\nabla\Phi_t)(\nabla\Phi_t)^{-1}\big)\det\nabla\Phi_t,
\end{aligned}$$

where we used the identity $\frac{\mathrm{d}}{\mathrm{d}t}\det\nabla\Phi_t = \mathrm{Tr}((\partial_t\nabla\Phi_t)(\nabla\Phi_t)^{-1})\det\nabla\Phi_t$ (with Tr denoting matrix trace). Therefore, the assumptions in Definition 5.3 show that the map $(t,\theta)\mapsto h_\psi(t,\theta) := \tilde g(t,\Phi_t(\theta))\psi(\Phi_t(\theta))\det\nabla\Phi_t(\theta)$ belongs to $L^1((a,b)\times\Theta_*)$ together with $\partial_t h_\psi$. We conclude that, for a.e. $\theta\in\Theta_*$, the map $t\mapsto h_\psi(t,\theta)$ is in $W^{1,1}(a,b)$, hence absolutely continuous up to defining it through a representative of $\tilde g\circ(\mathbf{i},\Phi_t)\in W^{1,1}((a,b)\times\Theta_*)$ with the same property: then, by the fundamental theorem of calculus and by Fubini theorem we have for $a \le s < t \le b$

$$\int_{\Theta_*}\big(h_\psi(t,\theta)-h_\psi(s,\theta)\big)\,\mathrm{d}\theta = \int_{\Theta_*}\int_s^t\partial_t h_\psi(r,\theta)\,\mathrm{d}r\,\mathrm{d}\theta = \int_s^t\int_{\Theta_*}\partial_t h_\psi(r,\theta)\,\mathrm{d}\theta\,\mathrm{d}r \tag{A.2}$$

and this shows that $h_\psi(t,\cdot)\in L^1(\Theta_*)$ for any $t\in[a,b]$ and that $t\mapsto\int_{\Theta_*}h_\psi(t,\theta)\,\mathrm{d}\theta$ is absolutely continuous on $[a,b]$.

We claim that if $\tilde g$ is a (not relabeled) ACL representative of $\tilde g$, then $t\mapsto\int_{\Theta_t}\psi(\theta)\tilde g(t,\theta)\,\mathrm{d}\theta$ is indeed absolutely continuous on $[a,b]$. It is enough to check that it is continuous, since we have just shown that the right hand side of (A.1) has an absolutely continuous representative on $[a,b]$. Assuming wlog that $a\le s<t\le b$, we have $\tilde g(t,\cdot)\in L^1(\Theta)$ for any $t\in[a,b]$ as well as the absolute continuity of $t\mapsto\int_\Theta\tilde g(t,\theta)\,\mathrm{d}\theta$, since and $\partial_t\tilde g\in L^1((a,b)\times\Theta)$ and Fubini theorem implies as above

$$\int_\Theta\tilde g(t,\theta)\,\mathrm{d}\theta - \int_\Theta\tilde g(s,\theta)\,\mathrm{d}\theta = \int_s^t\int_\Theta\partial_t\tilde g(t,\theta)\,\mathrm{d}\theta\,\mathrm{d}t.$$

Then, again by Fubini theorem we get

$$\int_{\Theta_t}\psi(\theta)\tilde g(t,\theta)\,\mathrm{d}\theta - \int_{\Theta_s}\psi(\theta)\tilde g(s,\theta)\,\mathrm{d}\theta = \int_\Theta\psi(\theta)(\mathbb{1}_{\Theta_t}-\mathbb{1}_{\Theta_s})\tilde g(t,\theta)\,\mathrm{d}\theta + \int_s^t\int_{\Theta_s}\psi(\theta)\partial\tilde g(t,\theta)\,\mathrm{d}t,$$

where the last term vanishes as $s\to t$ since ψ is bounded and $\partial_t\tilde g\in L^1((a,b)\times\Theta)$. The first term in the right hand side vanishes as well as $s\to t$ thanks to dominated convergence, since $\tilde g(t,\cdot)\in L^1(\Theta)$ and since the pointwise converges of $\mathbb{1}_{\Theta_s}$ to $\mathbb{1}_{\Theta_t}$ easily follows from the assumptions in Definition 5.3. The claim is proved.

Dividing (A.2) by s and using the Lebesgue points theorem, we see that for $a.e.$ $t\in(a,b)$ there holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Theta_*}h_\psi(t,\theta)\,\mathrm{d}\theta = \int_{\Theta_*}\partial_t h_\psi(t,\theta)\,\mathrm{d}\theta.$$

Since we can take the time derivative inside the integral sign, we have by change of variables and by the identity $(\nabla\Phi_t)^{-1}\circ\Psi_t = \nabla\Psi_t$, and with the notation $J_t = \det\nabla\Phi_t$ (so that $\frac{\mathrm{d}}{\mathrm{d}t}J_t = \mathrm{Tr}((\partial_t\nabla\Phi_t)(\nabla\Phi_t)^{-1})\det\nabla\Phi_t$),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Theta_t}\psi(\theta)\tilde{g}(t,\theta)\,\mathrm{d}\theta &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Theta_t}h_\psi(t,\theta)\,\mathrm{d}\theta = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Theta_*}\tilde{g}\big(t,\Phi_t(\theta)\big)\psi\big(\Phi_t(\theta)\big)J_t(\theta)\,\mathrm{d}\theta\\ &= \int_{\Theta_*}\bigg(J_t(\theta)\frac{\mathrm{d}}{\mathrm{d}t}\big(\tilde{g}\big(t,\Phi_t(\theta)\big)\psi\big(\Phi_t(\theta)\big)\big) + \tilde{g}\big(t,\Phi_t(\theta)\big)\psi\big(\Phi_t(\theta)\big)\frac{\mathrm{d}}{\mathrm{d}t}J_t(\theta)\bigg)\,\mathrm{d}\theta\\ &= \int_{\Theta_t}\partial_t\tilde{g}(t,\theta)\psi(\theta)\,\mathrm{d}\theta + \int_{\Theta_t}\nabla\cdot\big(\tilde{g}(t,\theta)\psi(\theta)\partial_t\Phi_t\big(\Psi_t(\theta)\big)\big)\,\mathrm{d}\theta,\end{aligned}$$

for a.e. $t\in(a,b)$. By the divergence theorem, the proof is concluded. □

Of course, if $\Theta_t\equiv\Theta$ for all t, we have that Φ_t is the identity map for any t. Lemma A.1 holds and Reynolds transport formula reduces to differentiation under integral sign. However, in such case we may extend the result to general probability measures on Θ, without requiring a density. We have the following standard result.

Lemma A.2. *Let $(\Theta,\mathscr{T},\pi)$ be a measure space, with π a σ-finite measure. Let $g:[a,b]\times\Theta\to\mathbb{R}$. Suppose that*

(i) *$g(\cdot,\theta)\in AC([a,b])$ for π-a.e. $\theta\in\Theta$ and $g(t,\cdot)\in L^1_\pi(\Theta)$ for all $t\in[a,b]$;*
(ii) *$\int_a^b\int_\Theta|\partial_t g(t,\theta)|\pi(\mathrm{d}\theta)\,\mathrm{d}t < +\infty$.*

Then the map $t\mapsto\int_\Theta\psi(\theta)g(t,\theta)\pi(\mathrm{d}\theta)$ is absolutely continuous on $[a,b]$ for any bounded continuous function ψ on Θ. In particular, if ψ is a bounded continuous function on Θ there holds a.e. on (a,b)

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Theta\psi(\theta)g(t,\theta)\pi(\mathrm{d}\theta) = \int_\Theta\psi(\theta)\partial_t g(t,\theta)\pi(\mathrm{d}\theta).$$

Proof. By assumption (i), g is a Carathéodory function, hence $(\mathrm{d}t\otimes\pi)$-measurable, moreover g is (classically) partially differentiable with respect to t at $(\mathrm{d}t\otimes\pi)$-a.e. $(t,\theta)\in(a,b)\times\Theta$, and then, by assumption (ii), we have $\partial_t g\in L^1_{\mathrm{d}t\otimes\pi}((a,b)\times\Theta)$. For any $a\le t_1<t_2\le b$, assumption (i) entails that $g(t_2,\theta)-g(t_1,\theta)=\int_{t_1}^{t_2}\partial_t g(t,\theta)\,\mathrm{d}t$ holds for π-a.e. $\theta\in\Theta$, the mapping $t\mapsto\partial_t g(t,\theta)$ being in $L^1(a,b)$ for π-a.e $\theta\in\Theta$. Thanks to (ii) and the boundedness of ψ, the map $t\mapsto\int_\Theta\psi(\theta)\partial_t g(t,\theta)\pi(\mathrm{d}\theta)$ belongs to $L^1(a,b)$, and we may apply Fubini theorem to obtain for $a\le t_1<t_2\le b$

$$\int_\Theta\psi(\theta)\big(g(t_2,\theta)-g(t_1,\theta)\big)\pi(\theta) = \int_\Theta\psi(\theta)\bigg(\int_{t_1}^{t_2}\partial_t g(t,\theta)\bigg)\pi(\mathrm{d}\theta) = \int_{t_1}^{t_2}\int_\Theta\psi(\theta)\partial_t g(t,\theta)\pi(\mathrm{d}\theta)\,\mathrm{d}t.$$

This shows that the map $t\mapsto\int_\Theta\psi(\theta)g(t,\theta)\pi(\mathrm{d}\theta)$ is an $AC([a,b])$ map. Finally, since the map $t\mapsto\int_\Theta\psi(\theta)\partial_t g(t,\theta)\pi(\mathrm{d}\theta)$ is in $L^1(a,b)$ and $g(\cdot,\theta)\in AC([a,b])$ for π-a.e. $\theta\in\Theta$, we apply Fubini once more to get

$$\int_\Theta\psi(\theta)\partial_t g(t,\theta)\pi(\mathrm{d}\theta) = \frac{\mathrm{d}}{\mathrm{d}t}\int_\Theta\psi(\theta)\big(g(t,\theta)-g(t_0,\theta)\big)\pi(\mathrm{d}\theta) = \frac{\mathrm{d}}{\mathrm{d}t}\int_\Theta\psi(\theta)g(t,\theta)\pi(\mathrm{d}\theta),$$

for $t_0\in(a,b)$ and for $a.e.$ t in (a,b), which completes the proof. □

A.2. *Lipschitz continuous extensions*

We provide the proof of a Lipschitz extension result, namely Lemma A.4. It is stated in a general framework, where $\mathbb{V}$, $\mathbb{X}$ and λ are as in Section 4.4.

Lemma A.3. *Let $\mathbb{Z}$ be a subset of $\mathbb{X}$ with $\lambda(\mathbb{Z})=0$. Then, for any $x\in\mathbb{X}\setminus\mathbb{Z}$, one has $\lambda(\mathbb{K}_x)=0$, where $\mathbb{K}_x:=\{y\in\mathbb{X}\setminus\mathbb{Z}:\mathcal{H}^1([x,y]\cap\mathbb{Z})>0\}$.*

Proof. Suppose by contradiction that there exists $x\in\mathbb{X}\setminus\mathbb{Z}$ such that $\lambda(\mathbb{K}_x)>0$. Then there exists a bounded set $\mathbb{U}$ in $\mathbb{X}$ such that $\lambda(\mathbb{K}_x\cap\mathbb{U})>0$. Moreover, by definition of $\mathbb{K}_x$, for every $y\in\mathbb{K}_x$ there holds $\mathcal{H}^1([y,x]\cap\mathbb{Z})>0$, which together with $\lambda(\mathbb{K}_x \cap \mathbb{U}) > 0$ implies

$$0 < \int_{\mathbb{K}_x \cap \mathbb{U}} \mathcal{H}^1\big([y,x] \cap \mathbb{Z}\big)\lambda(\mathrm{d}y) = \int_{\mathbb{K}_x \cap \mathbb{U}} \|x-y\| \int_0^1 \mathbb{1}_{\mathbb{Z}}\big((1-t)y+tx\big)\,\mathrm{d}t\lambda(\mathrm{d}y)$$

$$\leq \sup_{y \in \mathbb{U}} \|x-y\| \int_0^1 \int_{\mathbb{K}_x \cap \mathbb{U}} \mathbb{1}_{\frac{1}{1-t}(\mathbb{Z}-tx)}(y)\lambda(\mathrm{d}y)\,\mathrm{d}t,$$

where the last inequality is due to Fubini's theorem. But this is a contradiction, since the right-hand side is equal to zero, being the set $\frac{1}{1-t}(\mathbb{Z}-tx)$ of zero λ-measure for all $t \in (0,1)$. □

Lemma A.4 (Lipschitz Extension). *Let $\mathbb{Z}$ be a λ-null subset of $\mathbb{X}$ and let $\mathbb{B}(\mathbb{Z})$ be defined by (4.16). Let (S, d_S) be a complete metric space. Let $f: \mathbb{X} \setminus \mathbb{Z} \to S$. If*

$$d_S\big(f(x), f(y)\big) \leq L\|x-y\| \quad \forall (x,y) \in \mathbb{B}(\mathbb{Z}) \tag{A.3}$$

holds for some $L \geq 0$, then f admits a Lipschitz extension to the whole of $\mathbb{X}$, with the same Lipschitz constant L.

Proof. For every $x \in \mathbb{X} \setminus \mathbb{Z}$, let $\mathbb{K}_x$ be defined as in Lemma A.3. First, we prove that f is Lipschitz-continuous on $\mathbb{X} \setminus \mathbb{Z}$. In fact, fix two points x, y in $\mathbb{X} \setminus \mathbb{Z}$, and choose a sequence $\{\xi_n\}_{n \geq 1} \subset \mathbb{K}_x^c \cap \mathbb{K}_y^c \cap \mathbb{Z}^c$ converging to x. This choice is possible since Lemma A.3 shows that $\lambda(\mathbb{K}_x \cup \mathbb{K}_y \cup \mathbb{Z}) = 0$, implying that $\mathbb{K}_x^c \cap \mathbb{K}_y^c \cap \mathbb{Z}^c$ is dense in $\mathbb{X}$. Moreover, notice that $(\xi_n, x) \in \mathbb{B}(\mathbb{Z})$ and $(\xi_n, y) \in \mathbb{B}(\mathbb{Z})$ for every n, since ξ_n belongs to both $\mathbb{K}_x^c \setminus \mathbb{Z}$ and $\mathbb{K}_y^c \setminus \mathbb{Z}$. Then, invoke (A.3) to obtain, for every $n \in \mathbb{N}$, that $\mathrm{d}_S(f(x), f(y)) \leq \mathrm{d}_S(f(\xi_n), f(x)) + \mathrm{d}_S(f(\xi_n), f(y)) \leq L(\|x-\xi_n\| + \|\xi_n - y\|)$. By taking the limit as $n \to +\infty$, we get the desired Lipschitz property on $\mathbb{X} \setminus \mathbb{Z}$. In conclusion, the existence of a Lipschitz extension with same constant L follows from the standard extension result with a dense domain, being (S, d_S) complete. □

A.3. *Scaling estimates of the Poincaré constant*

In view of the applications of our theorems, the log-concavity condition and its variants are the more natural tools, mostly when considering exponential statistical models (Section 2.2), exchangeability (Section 3.1) and Bayesian consistency (Section 3.3). Accordingly, we summarize some estimates of the Poincaré constant in the following statement, providing some extension of the results in [8]. In particular, in view of our results about Bayesian consistency in Section 3.3, in this statement we highlight some scaling properties of the Poincaré constant that arise by multiplying V by some large $n \in \mathbb{N}$.

Proposition A.5. *Let $V, U \in C^2(\mathbb{R}^d)$ be bounded from below and such that $\int_{\mathbb{R}^d} e^{-V(\theta)-U(\theta)}\,\mathrm{d}\theta < +\infty$. Let $\mu_n(\mathrm{d}\theta) := e^{-nV-U}\,\mathrm{d}\theta$, for any $n \in \mathbb{N}$. The following statements about the squared Poincaré constant of μ_n hold.*

(1) *Suppose that $\alpha > 0$ and $h \in \mathbb{R}$ exist such that* $\mathrm{Hess}(V) \geq \alpha I$ *on $\mathbb{R}^d$ and* $\mathrm{Hess}(U) \geq hI$ *on $\mathbb{R}^d$ in the sense of quadratic forms. Then $\mathcal{C}^2[\mu_n] \leq (n\alpha + h)^{-1}$ for every $n > -h/\alpha$.*

(2) *Suppose that there exist $\alpha > 0$, $c > 0$, $R > 0$, $h \in \mathbb{R}$, $\ell \in \mathbb{R}$ such that the following conditions hold*: $\mathrm{Hess}(V(\theta)) \geq \alpha I$ *and* $\mathrm{Hess}(U(\theta)) \geq hI$ *in the sense of quadratic forms whenever $|\theta| \leq R$, moreover $\theta \cdot \nabla V(\theta) \geq c|\theta|$ and $\theta \cdot \nabla U(\theta) \geq \ell|\theta|$ whenever $|\theta| \geq R$. Then, for every $n > (-h/\alpha) \vee ((d_R + 1 - \ell)/c)$,*

$$\mathcal{C}^2[\mu_n] \leq \frac{\alpha n + h + (cn + \ell - d_R + nV_R + U_R)C_R}{(\alpha n + h)(cn + \ell - 1 - d_R)},$$

where $d_R := (d-1)/R$, $V_R := \sup_{B_R} |\nabla V|$, $U_R := \sup_{B_R} |\nabla U|$ and C_R is an explicit universal constant only depending on R.

(3) *Suppose that there exist $\alpha > 0$, $c_1 > 0$, $c_2 > 0$, $R > 0$, $h \in \mathbb{R}$ such that the following conditions hold*: $\mathrm{Hess}(V(\theta)) \geq \alpha I$ *and* $\mathrm{Hess}(U(\theta)) \geq hI$ *in the sense of quadratic forms whenever $|\theta| \leq R$, and*

$$\big|\nabla V(\theta)\big|^2 \geq 2c_1 + c_2\big[\Delta V(\theta) + \nabla V(\theta) \cdot \nabla U(\theta)\big]_+ \tag{A.4}$$

whenever $|\theta| \geq R$. Then, for every $n > (1 + 1/c_2) \vee (-h/\alpha)$,

$$\mathcal{C}^2[\mu_n] \leq \frac{\alpha n + h + e^{\omega_R}(c_1 n + V_R^* + W_R)}{(\alpha n + h)c_1 n},$$

where $V_R^* := \sup_{B_R} |\Delta V|$, $W_R := \sup_{B_R} |\nabla U||\nabla V|$ and $\omega_R := \sup_{B_R} V - \inf_{\mathbb{R}^d} V$.

By choosing $n = 1$ and $U \equiv 0$ in Proposition A.5, we obtain a direct estimate for the Poincaré constant of a given finite measure of the form $\mu(\mathrm{d}\theta) = e^{-V}\,\mathrm{d}\theta$, where V is bounded from below (see also [8]). Besides the Bakry–Emery criterion, which requires the Hessian of V to be bounded away from zero, perturbations of convex functions are included. For instance $V(\theta) = \frac{1}{2}|\theta|^2 - 2\cos|\theta|$ satisfies the assumptions of points (2) and (3) with $R = 1$.

Proof of Proposition A.5. Point (1) is the Bakry–Emery criterion, see for instance [66, Theorem 3.1].

Let us consider point (2). We shall apply the arguments from [8]. Let $V_n := nV + U$. First of all, if $|\theta| \ge R$ we get from the assumptions

$$\theta \cdot V_n(\theta) \ge (cn + \ell)|\theta|. \tag{A.5}$$

Let $W(\theta)$ be a $C^2(\mathbb{R}^d)$ function such that $W \ge 1$ on $\mathbb{R}^d$ and such that $W(\theta) = e^{|\theta|}$ if $|\theta| \ge R$. Let $C_R := \sup_{B_r}|W| + \sup_{B_R}|\nabla W| + \sup_{B_R}|\Delta W|$. Let us introduce the diffusion operator $\mathcal{L}_{V_n}[\phi] := \Delta\phi - \langle\nabla\phi, \nabla V_n\rangle$. A computation shows that if $|\theta| \ge R$ there holds

$$\mathcal{L}_{V_n} W = \left(\frac{d-1}{|\theta|} + 1 - \frac{\theta}{|\theta|}\cdot\nabla V_n(\theta)\right) W(\theta).$$

Let $\tau_n := cn + \ell - 1 - d_R$, so that $\tau_n > 0$ as soon as $n > (d_R + 1 - \ell)/c$. If $|\theta| \ge R$, from (A.5) we deduce $\mathcal{L}_{V_n} W(\theta) \le -\tau_n W(\theta)$. If $|\theta| \le R$, we estimate as

$$\begin{aligned}\mathcal{L}_{V_n} W(\theta) &= -\tau_n W(\theta) + \tau_n W(\theta) + \mathcal{L}_{V_n} W(\theta) \le -\tau_n W(\theta) + \tau_n W(\theta) + \big|\Delta W(\theta)\big| + \big|\nabla W(\theta)\big|\big|\nabla V_n(\theta)\big| \\ &\le -\tau_n W(\theta) + C_R(1 + \tau_n + nV_R + U_R).\end{aligned}$$

All in all we have $W(\theta) \ge 1$ and $\mathcal{L}_{V_n} W(\theta) \le -\tau_n W(\theta) + b_n \chi_{B_R}(\theta)$ for every $\theta \in \mathbb{R}^d$, where $b_n := C_R(1 + \tau_n + nV_R + U_R)$. By [8, Theorem 1.4] we conclude that $\mathcal{C}^2(\mu_n) \le \frac{1}{\tau_n}(1 + b_n k_R)$, where k_R is the squared Poincaré constant of the measure $e^{-V_n}\mathcal{L}^d \llcorner B_R$. Since we have by assumption $\mathrm{Hess} V_n \ge (n\alpha + h)I$ in the sense of quadratic forms on B_R, the Bakry–Emery criterion yields $k_R \le (\alpha n + h)^{-1}$ as soon as $n > -h/\alpha$. The conclusion follows.

Let us prove point (3). The argument is similar. Let $W(\theta) = \exp\{V(\theta) - \inf_{\mathbb{R}^d} V\}$, $\theta \in \mathbb{R}^d$. Let once more $V_n := nV + U$ and $\mathcal{L}_{V_n}[\phi] := \Delta\phi - \langle\nabla\phi, \nabla V_n\rangle$. We have after a direct computation $\Delta W = W|\nabla V|^2 + W\Delta V$ and then

$$\mathcal{L}_{V_n} W = \big((1-n)\big|\nabla V(\theta)\big|^2 + \Delta V(\theta) - \nabla U(\theta)\cdot\nabla V(\theta)\big) W(\theta), \quad \theta \in \mathbb{R}^d. \tag{A.6}$$

Thanks to assumption (A.4) we have $(n-1)|\nabla V(\theta)|^2 \ge \Delta V(\theta) - \nabla V(\theta)\cdot\nabla U(\theta) + nc_1$ whenever $|\theta| \ge R$ and $n > 1 + 1/c_2$. Therefore, if $n > 1 + 1/c_2$, from (A.6) we obtain $\mathcal{L}_{V_n} W(\theta) \le -c_1 n W(\theta)$ whenever $|\theta| \ge R$. On the other hand, if $|\theta| \le R$ we easily estimate from (A.6) as

$$\mathcal{L}_{V_n} W(\theta) = -c_1 n W(\theta) + c_1 n W(\theta) + \mathcal{L}_{V_n} W(\theta) \le -c_1 n W(\theta) + e^{\omega_R}\big(c_1 n + V_R^* + W_R\big).$$

Then, we have $W(\theta) \ge 1$ and $\mathcal{L}_{V_n} W(\theta) \le -c_1 n W(\theta) + \tilde{b}_n \chi_{B_R}(\theta)$ for every $\theta \in \mathbb{R}^d$, where $\tilde{b}_n := +e^{\omega_R}(c_1 n + V_R^* + W_R)$. By invoking [8, Theorem 1.4] and the Bakry–Emery criterion, the conclusion follows by repeating the same argument in the end of the proof of point (2). □

Funding

The authors acknowledge support from the MIUR-PRIN project No 2017TEXA3H. E.M. acknowledges support from the INdAM-GNAMPA 2019 project *"Trasporto ottimo per dinamiche con interazione"*. E. D. received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 817257. E. D. gratefully acknowledges the financial support from the Italian Ministry of Education, University and Research (MIUR), "Dipartimenti di Eccellenza" grant 2018-2022.

References

[1] N. L. Ackerman, C. E. Freer and D. M. Roy. On computability and disintegration. *Math. Structures Comput. Sci.* **27** (2017) 1287–1314. MR3724454 https://doi.org/10.1017/S0960129516000098

[2] G. Acosta and R. G. Durán. An optimal Poincaré inequality in L^1 for convex domains. *Proc. Amer. Math. Soc.* **132** (2004) 195–202. MR2021262 https://doi.org/10.1090/S0002-9939-03-07004-7

[3] D. Aldous. *Exchangeability and Related Topics. Ecole d'Eté de Probabilités de Saint-Flour XIII. Lecture Notes in Math.* **1117**. Springer, Berlin, 1985. MR0883646 https://doi.org/10.1007/BFb0099421
[4] L. Ambrosio and N. Gigli. A user's guide to optimal transport. Modelling and optimisation of flows on networks. In *Lecture Notes in Math.* 1–155, **2062**. Fond. CIME/CIME Found. Subser., Springer, Heidelberg, 2013. MR3050280 https://doi.org/10.1007/978-3-642-32160-3_1
[5] L. Ambrosio, N. Gigli and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd edition. Birkhäuser, Basel, 2008. MR2401600
[6] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer. *Sur les Inégalités de Sobolev Logarithmiques*. *Panor. Synthèses* **10**. Soc. Math, France, 2000. MR1845806
[7] T. Aubin. Problèmes isopérimétriques et espaces de Sobolev. *J. Differential Geom.* **11** (1976) 573–598. MR0448404
[8] D. Bakry, F. Barthe, P. Cattiaux and A. Guillin. A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.* **13** (2008) 60–66. MR2386063 https://doi.org/10.1214/ECP.v13-1352
[9] D. Bakry, P. Cattiaux and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** (2008) 727–759. MR2381160 https://doi.org/10.1016/j.jfa.2007.11.002
[10] D. Bakry and M. Émery Diffusions hypercontractives. In *Séminaire de Probabilités, XIX, 1983/84* 177–206. *Lecture Notes in Math.* **1123**. Springer, Berlin, 1985. MR0889476 https://doi.org/10.1007/BFb0075847
[11] O. E. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. Wiley & Sons, Chichester, 1978. MR0489333
[12] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwend.* **22** (2003) 751–756. MR2036927 https://doi.org/10.4171/ZAA/1170
[13] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** (2000) 375–393. MR1738163 https://doi.org/10.1007/s002110050002
[14] G. Bernstein and D. Sheldon. Differentially private Bayesian inference for exponential families. *NeurIPS* **2018** (2018) 2924–2934.
[15] P. Billingsley. *Convergence of Probability Measures*, 2nd edition. Wiley & Sons, New York, 1999. MR1700749 https://doi.org/10.1002/9780470316962
[16] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27** (1999) 1903–1921. MR1742893 https://doi.org/10.1214/aop/1022874820
[17] S. G. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distance. *Mem. Amer. Math. Soc.* **261** (2019) 1259. MR4028181 https://doi.org/10.1090/memo/1259
[18] V. I. Bogachev. *Differentiable Measures and the Malliavin Calculus*. *Mathematical Surveys and Monographs* **164**. Amer. Math. Soc., Providence, 2010. MR2663405 https://doi.org/10.1090/surv/164
[19] K. W. Breitung. *Asymptotic Approximations for Probability Integrals*. Springer-Verlag, Berlin, 1994. MR1329727 https://doi.org/10.1007/BFb0073538
[20] L. D. Brown *Fundamentals of Statistical Exponential Families with Application in Statistical Decision Theory*. *Institute of Mathematical Statistics, Lecture Notes-Monograph Series* **9**. Hayward. MR0882001
[21] S. Buckley and P. Koskela. Sobolev–Poincaré implies John. *Math. Res. Lett.* **2** (1995) 577–593. MR1359964 https://doi.org/10.4310/MRL.1995.v2.n5.a5
[22] F. Camerlenghi, E. Dolera, S. Favaro and E. Mainini. Wasserstein posterior contraction rates in non-dominated Bayesian nonparametric models, 2022. Available at arXiv:2201.12225.
[23] P. Cattiaux. Hypercontractivity for perturbed diffusion semigroups. *Ann. Fac. Sci. Univ. Toulouse Sci. Math.* **XIV** (2005) 609–628. MR2188585
[24] D. Chafaï. Entropies, convexity, and functional inequalities: On Φ-entropies and Φ-Sobolev inequalities. *J. Math. Kyoto Univ.* **44** (2004) 325–363. MR2081075 https://doi.org/10.1215/kjm/1250283556
[25] J. T. Chang and D. Pollard. Conditioning as disintegration. *Stat. Neerl.* **51** (1997) 287–317. MR1484954 https://doi.org/10.1111/1467-9574.00056
[26] Z. Q. Chen, R. J. Williams and Z. Zhao. A Sobolev inequality and Neumann heat kernel estimate for unbounded domains. *Math. Res. Lett.* **1** (1994) 177–184. MR1266756 https://doi.org/10.4310/MRL.1994.v1.n2.a5
[27] D. M. Cifarelli, E. Dolera and E. Regazzini. Frequentistic approximations to Bayesian prevision of exchangeable random elements. *Internat. J. Approx. Reason.* **78** (2016) 138–152. MR3543878 https://doi.org/10.1016/j.ijar.2016.06.007
[28] D. M. Cifarelli, E. Dolera and E. Regazzini. Note on "Frequentistic approximations to Bayesian prevision of exchangeable random elements". *Internat. J. Approx. Reason.* **86** (2017) 26–27. MR3653971 https://doi.org/10.1016/j.ijar.2017.03.013
[29] S. L. Cotter, M. Dashti, J. C. Robinson and A. M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Probl.* **25** (2009) 115008. MR2558668 https://doi.org/10.1088/0266-5611/25/11/115008
[30] M. Dashti and A. M. Stuart The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification* 311–428. Springer, Cham, 2017. MR3839555
[31] P. Diaconis and D. Freedman. On consistency of Bayes estimates. *Ann. Statist.* **14** (1986) 1–26. MR0829555 https://doi.org/10.1214/aos/1176349830
[32] P. Diaconis and D. Yilvisaker. Conjugate priors exponential. *Ann. Statist.* **23** (1987) 397–423.
[33] E. Dolera. On an asymptotic property of posterior distributions. *Boll. Unione Mat. Ital. (9)* **6** (2013) 741–748. MR3202852
[34] E. Dolera, S. Favaro and E. Mainini. Strong posterior contraction rates via Wasserstein dynamics, 2022. Available at arXiv:2203.10754.
[35] E. Dolera and E. Mainini. On uniform continuity of posterior distributions. *Statist. Probab. Lett.* **157** (2020) 108627. MR4014341 https://doi.org/10.1016/j.spl.2019.108627
[36] E. Dolera and E. Regazzini. Uniform rates of the Glivenko–Cantelli convergence and their use in approximating Bayesian inferences. *Bernoulli* **25** (2019) 2982–3015. MR4003571 https://doi.org/10.3150/18-BEJ1077
[37] B. Efron. Empirical Bayes deconvolution estimates. *Biometrika* **103** (2016) 1–20. MR3465818 https://doi.org/10.1093/biomet/asv068
[38] L. Esposito, C. Nitsch and C. Trombetti. Best constants in Poincaré inequalities for convex domains. *J. Convex Anal.* **20** (2011) 253–264. MR3086452
[39] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*, Revised edition. CRC Press, Boca Raton, 2015. MR3409135
[40] E. Fabes, D. Jerison and C. Kenig. The Wiener test for degenerate elliptic equations. *Ann. Inst. Fourier (Grenoble)* **32** (1982) 151–182. MR0688024
[41] E. Fabes, C. Kenig and R. Serapioni. The local regularity of solutions of degenerate elliptic equations. *Comm. Partial Differential Equations* **7** (1982) 77–116. MR0643158 https://doi.org/10.1080/03605308208820218

[42] R. Farwig and V. Rosteck. Note on Friedrichs' inequality in N-star-shaped domains. *J. Math. Anal. Appl.* **435** (2016) 1514–1524. MR3429656 https://doi.org/10.1016/j.jmaa.2015.10.046
[43] T. S. Ferguson. *A Course in Large Sample Theory*. *Texts in Statistical Science Series*. Chapman & Hall, London, 1996. MR1699953 https://doi.org/10.1007/978-1-4899-4549-5
[44] V. Ferone, C. Nitsch and C. Trombetti. A remark on optimal weighted Poincaré inequalities for convex domains. *Atti Accad. Naz. Lincei, Rend. Lincei, Mat. Appl.* **23** (2012) 467–475. MR2999557 https://doi.org/10.4171/RLM/640
[45] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** (2015) 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7
[46] E. Gagliardo. Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in n variabili. *Rend. Semin. Mat. Univ. Padova* **27** (1957) 284–305. MR0102739
[47] S. Ghosal and A. W. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge, 2017. MR3587782 https://doi.org/10.1017/9781139029834
[48] J. K. Ghosh, M. Delampady and T. Samanta. *An Introduction to Bayesian Analysis. Theory and Methods*. Springer, New York, 2006. MR2247439
[49] M. Giaquinta, G. Modica and J. Souček. *Cartesian Currents in the Calculus of Variations*, **1**. Springer-Verlag, Berlin, 1998. MR1482275
[50] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *Int. Stat. Rev.* **70** (2002) 419–435.
[51] M. E. Gurtin. *An Introduction to Continuum Mechanics*. Academic Press, New York, 1981. MR0636255
[52] J. Heinonen. Nonsmooth calculus. *Bull. Amer. Math. Soc.* **44** (2007) 163–232. MR2291675 https://doi.org/10.1090/S0273-0979-07-01140-8
[53] A. Henrot and M. Pierre. *Variation et Optimisation de Formes*. *Mathématiques & Applications* **48**. Springer, Berlin, 2005. MR2512810 https://doi.org/10.1007/3-540-37689-5
[54] R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.* **46** (1987) 1159–1194. MR0893137 https://doi.org/10.1007/BF01011161
[55] R. Hurri-Syrjänen. Unbounded Poincaré domains. *Ann. Acad. Sci. Fenn., Ser. A 1 Math.* **17** (1992) 409–423. MR1190332 https://doi.org/10.5186/aasfm.1992.1725
[56] M. A. Iglesias, K. Lin and A. M. Stuart. Well-posed Bayesian geometric inverse problems arising in subsurface flow. *Inverse Probl.* **30** (2014) 114001. MR3274585 https://doi.org/10.1088/0266-5611/30/11/114001
[57] O. Kallenberg. *Foundations of Modern Probability*, 2nd edition. Springer, New York, 2002. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8
[58] T. Kulkarni, J. Jälkö, S. Kaski and A. Honkela. Locally Differentially Private Bayesian Inference, 2021. Available at arXiv:2110.14426.
[59] T. Lagatta. Continuous disintegrations of Gaussian processes. *Theory Probab. Appl.* **57** (2013) 151–162. MR3201645 https://doi.org/10.1137/S0040585X9798587X
[60] O. E. Lanford. Entropy and equilibrium states in classical statistical mechanics. In *Statistical Mechanics and Mathematical Problems*, A. Lenard (Ed.) *Lecture Notes in Physics* **20**. Springer, Berlin, Heidelberg, 1973. MR0395622
[61] J. Latz. On the well-posedness of Bayesian inverse problems. *SIAM/ASA J. Uncertain. Quantificat.* **8** (2020) 451–482. MR4075344 https://doi.org/10.1137/19M1247176
[62] G. Leoni. *A First Course in Sobolev Spaces*. American Mathematical Society, Providence, 2009. MR2527916 https://doi.org/10.1090/gsm/105
[63] A. Lijoi. Approximating priors by finite mixtures of conjugate distributions for an exponential family. *J. Statist. Plann. Inference* **113** (2003) 419–435. MR1965119 https://doi.org/10.1016/S0167-7152(01)00205-X
[64] J. Lindenstrauss and D. Preiss. On Fréchet differentiability of Lipschitz maps between Banach spaces. *Ann. Math.* **157** (2003) 257–288. MR1954267 https://doi.org/10.4007/annals.2003.157.257
[65] D. Matthes, A. Jüngel and G. Toscani. Convex Sobolev inequalities derived from entropy dissipation. *Arch. Ration. Mech. Anal.* **199** (2011) 563–596. MR2763034 https://doi.org/10.1007/s00205-010-0331-9
[66] G. Menz and A. Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.* **42** (2014) 1809–1884. MR3262493 https://doi.org/10.1214/14-AOP908
[67] J. V. Miller and D. B. Dunson. Robust Bayesian inference via coarsening. *J. Amer. Statist. Assoc.* **114** (2019) 1113–1125. MR4011766 https://doi.org/10.1080/01621459.2018.1469995
[68] F. Otto. The geometry of dissipative evolution equations: The porous-medium equation. *Comm. Partial Differential Equations* **26** (2001) 101–174. MR1842429 https://doi.org/10.1081/PDE-100002243
[69] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** (2000) 361–400. MR1760620 https://doi.org/10.1006/jfan.1999.3557
[70] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Ration. Mech. Anal.* **5** (1960) 286–292. MR0117419 https://doi.org/10.1007/BF00252910
[71] J. Pfanzagl. Conditional distributions as derivatives. *Ann. Probab.* **7** (1979) 1046–1050. MR0548898
[72] R. R. Phelps. Gaussian null sets and differentiability of Lipschitz map on Banach spaces. *Pacific J. Math.* **77** (1978) 523–531. MR0510938
[73] M. M. Rao *Conditional Measures and Applications*, 2nd edition. Chapman & Hall/CRC, Boca Raton, 2005. MR2149673 https://doi.org/10.1201/9781420027433
[74] E. Regazzini and V. V. Sazonov. Approximation of laws of random probabilities by mixtures of Dirichlet distributions with applications to nonparametric Bayesian inference. *Theory Probab. Appl.* **45** (2001) 93–110. MR1810976 https://doi.org/10.1137/S0040585X97978063
[75] A. Renyi. On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hung.* **6** (1955) 285–335. MR0081008 https://doi.org/10.1007/BF02024393
[76] C. Scricciolo. On rates of convergence for Bayesian density estimation. *Scand. J. Stat.* **34** (2007) 626–642. MR2368802 https://doi.org/10.1111/j.1467-9469.2006.00540.x
[77] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.* **19** (2010) 451–559. MR2652785 https://doi.org/10.1017/S0962492910000061
[78] A. M. Sullivan. Well-posedness of Bayesian inverse problems in quasi-Banach spaces with stable priors, 2017. Available at arXiv:1710.05610. MR3681971 https://doi.org/10.3934/ipi.2017040
[79] G. Talenti. Best constant in Sobolev inequality. *Ann. Mat. Pura Appl.* **110** (1976) 353–372. MR0463908 https://doi.org/10.1007/BF02418013
[80] T. Tjur. *Probability Based on Radon Measure*. Wiley, New York, 1980. MR0595868

[81] N. G. Trillos and D. Sanz-Alonso. The Bayesian formulation and well-posedness of fractional elliptic inverse problems. *Inverse Probl.* **33** (2017) 065006. MR3656667 https://doi.org/10.1088/1361-6420/aa711e
[82] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, Providence, 2003. MR1964483 https://doi.org/10.1090/gsm/058
[83] C. Villani. *Optimal Transport, Old and New*. Springer, Berlin, 2009. MR2914136 https://doi.org/10.1016/j.joems.2011.09.007
[84] M. West. Approximating posterior distributions by mixtures. *J. Roy. Statist. Soc. Ser. B* **55** (1979) 409–422. MR1224405
[85] S. L. Zabell. Continuous versions of regular conditional distributions. *Ann. Probab.* **7** (1979) 159–165. MR0515823