Constructive Approximation (2022) 55:987–1036
https://doi.org/10.1007/s00365-022-09570-9

CONSTRUCTIVE APPROXIMATION

Sparse Approximation of Triangular Transports, Part II: The Infinite-Dimensional Case

Jakob Zech[1] · **Youssef Marzouk**[2]

Received: 12 June 2020 / Revised: 21 June 2021 / Accepted: 9 September 2021 /
Published online: 17 March 2022

Abstract

For two probability measures ρ and π on $[-1, 1]^{\mathbb{N}}$ we investigate the approximation of the triangular Knothe–Rosenblatt transport $T : [-1, 1]^{\mathbb{N}} \to [-1, 1]^{\mathbb{N}}$ that pushes forward ρ to π. Under suitable assumptions, we show that T can be approximated by rational functions without suffering from the curse of dimension. Our results are applicable to posterior measures arising in certain inference problems where the unknown belongs to an (infinite dimensional) Banach space. In particular, we show that it is possible to efficiently approximately sample from certain high-dimensional measures by transforming a lower-dimensional latent variable.

Keywords Transport maps · Sampling · Domains of holomorphy · Sparse approximation

Mathematics Subject Classification 62D05 · 32D05 · 41A10 · 41A25 · 41A46

Communicated by Albert Cohen.

This paper was written during the postdoctoral stay of JZ at MIT. JZ acknowledges support by the Swiss National Science Foundation under Early Postdoc Mobility Fellowship 184530. YM and JZ acknowledge support from the United States Department of Energy, Office of Advanced Scientific Computing Research, AEOLUS Mathematical Multifaceted Integrated Capability Center.

✉ Jakob Zech
jakob.zech@uni-heidelberg.de

Youssef Marzouk
ymarz@mit.edu

1 Heidelberg University, 69120 Heidelberg, Germany

2 Massachusetts Institute of Technology, Cambridge, MA 02139, USA

1 Introduction

In this paper we discuss the approximation of transport maps on infinite-dimensional domains. Our main motivation are inference problems, in which the unknown belongs to a Banach space Y. Two examples could be the following:

- **Groundwater flow** Consider a porous medium in a domain $\mathrm{D} \subseteq \mathbb{R}^3$. Given observations of the subsurface flow, we are interested in the permeability (hydraulic conductivity) of the medium in D. The physical system is described by an elliptic partial differential equation, and the unknown quantity describing the permeability can be modeled as a function $\psi \in L^\infty(\mathrm{D}) = Y$ [25].
- **Inverse scattering** Suppose that $\mathrm{D}_{\mathrm{scat}} \subseteq \mathbb{R}^3$ is filled by a perfect conductor and illuminated by an electromagnetic wave. Given measurements of the scattered wave, we are interested in the shape of the scatterer $\mathrm{D}_{\mathrm{scat}}$. Assume that this domain can be described as the image of some bounded reference domain $\mathrm{D} \subseteq \mathbb{R}^3$ under a bi-Lipschitz transformation $\psi : \mathrm{D} \to \mathbb{R}^3$, i.e., $\mathrm{D}_{\mathrm{scat}} = \psi(\mathrm{D})$. The unknown is then the function $\psi \in W^{1,\infty}(\mathrm{D}) = Y$. We describe the forward model in [17].

The Bayesian approach to these problems is to model ψ as a Y-valued random variable and to determine the distribution of ψ conditioned on a (typically noisy) observation of the system. Bayes' theorem can used to specify this "posterior" distribution via the prior and the likelihood. The prior is a measure on Y that represents our information on $\psi \in Y$ before making an observation. Mathematically speaking, assuming that the observation and the unknown follow some joint distribution, the prior is the marginal distribution of the unknown ψ. The goal is to explore the posterior and in this way to make inferences about ψ. We refer to [9] for more details on the general methodology of Bayesian inversion in Banach spaces.

For the analysis and implementation of such methods, instead of working with (prior and posterior) measures on the Banach space Y, it can be convenient to parameterize the problem and work with measures on $\mathbb{R}^{\mathbb{N}}$ instead. To demonstrate this, choose a sequence $(\psi_j)_{j\in\mathbb{N}}$ in Y and a measure μ on $\mathbb{R}^{\mathbb{N}}$. With $\boldsymbol{y}:=(y_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ and

$$\Phi(\boldsymbol{y}):=\sum_{j\in\mathbb{N}} y_j \psi_j \tag{1.1}$$

we can formally define a prior measure on Y as the pushforward $\Phi_\sharp\mu$. Instead of inferring $\psi \in Y$ directly, we may instead infer the coefficient sequence $\boldsymbol{y} = (y_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, in which case μ holds the prior information on the unknown coefficients. These viewpoints are equivalent in the sense that the conditional distribution of ψ given an observation is the pushforward, under Φ, of the conditional distribution of $\boldsymbol{y}$ given the observation. Under certain assumptions on the prior and the space Y, construction (1.1) arises naturally through the Karhunen–Loève expansion; see, e.g., [1, 22]. In this case the $y_j \in \mathbb{R}$ are uncorrelated random variables with unit variance, and the ψ_j are eigenvectors of the prior covariance operator, with their norms equal to the square root of the corresponding eigenvalues.

In this paper we concentrate on the special case where the coefficients y_j are known to belong to a bounded interval. Up to a shift and a scaling this is equivalent to $y_j \in [-1, 1]$, which will be assumed throughout. We refer to [9,Sect. 2] for the construction and further discussion of such (bounded) priors. The goal then becomes to determine and explore the posterior measure on $U:=[-1, 1]^{\mathbb{N}}$. Denote this measure by π and let μ be the prior measure on U such that $\pi \ll \mu$. Then the Radon-Nikodym derivative $f_\pi:=\frac{\mathrm{d}\pi}{\mathrm{d}\mu} : U \to [0, \infty)$ exists. Since the forward model (and thus the likelihood) only depends on $\Phi(\boldsymbol{y})$ in the Banach space Y, f_π must be of the specific type

$$f_\pi(\boldsymbol{y}) = \mathfrak{f}_\pi(\Phi(\boldsymbol{y})) = \mathfrak{f}_\pi\Big(\sum_{j\in\mathbb{N}} y_j\psi_j\Big) \tag{1.2}$$

for some $\mathfrak{f}_\pi : Y \to [0, \infty)$. We give a concrete example in Example 2.6 where this relation holds.

"Exploring" the posterior refers to computing expectations and variances w.r.t. π, or detecting areas of high probability w.r.t. π. A standard technique to do so in high dimensions is Monte Carlo—or in this context Markov chain Monte Carlo—sampling, e.g., [31]. Another approach is via transport maps [23]. Let ρ be another measure on U from which it is easy to sample. Then, a map $T : U \to U$ satisfying $T_\sharp\rho = \pi$ (i.e., $\pi(A) = \rho(\{\boldsymbol{y} : T(\boldsymbol{y}) \in A\})$ for all measurable A) is called a transport map that pushes forward ρ to π. Such a T has the property that if $\boldsymbol{y} \sim \rho$ then $T(\boldsymbol{y}) \sim \pi$, and thus samples from π can easily be generated once T has been computed. Observe that $\Phi \circ T : U \to Y$ will then transform a sample from ρ to a sample from $(\Phi \circ T)_\sharp\rho = \Phi_\sharp\pi$, which is the posterior in the Banach space Y. Thus, given T, we can perform inference on the quantity in the Banach space.

This motivates the setting we are investigating in this paper: for two measures ρ and π on U, such that their densities are of type (1.2) for a smooth (see Sect. 2) function $\mathfrak{f}_\pi$, and we are interested in the approximation of $T : U \to U$ such that $T_\sharp\rho = \pi$. More precisely, we will discuss the approximation of the so-called Knothe–Rosenblatt (KR) transport by rational functions. The reason for using rational functions (rather than polynomials) is to guarantee that the resulting approximate transport is a bijection from $U \to U$. The rate of convergence will in particular depend on the decay rate of the functions ψ_j. If (1.1) is a Karhunen–Loève expansion, this is the decay rate of the square root of the eigenvalues of the covariance operator of the prior. The faster this decay, the larger the convergence rate will be. The reason for analyzing the triangular KR transport is its wide use in practical algorithms [13, 16, 35, 38], and the fact that its concrete construction makes it amenable to a rigorous analysis.

Sampling from high-dimensional distributions by transforming a (usually lower-dimensional) "latent" variable into a sample from the desired distribution is a standard problem in machine learning. It is tackled by methods such as generative adversarial networks [15] and variational autoencoders [12]. In the setting above, the high-dimensional distribution is the posterior on Y. We will show that under the assumptions of this paper, it is possible to approximately sample from this distribution by transforming a low-dimensional latent variable and without suffering from the curse of dimensionality. While Bayesian inference is our motivation, for the rest of the manuscript the presentation remains in an abstract setting, and our results therefore have ramifications on the broader task of transforming high-dimensional distributions.

1.1 Contributions and Outline

In this manuscript we generalize the analysis of [41] to the infinite-dimensional case. Part of the proofs are based on the results in [41], which we recall in appendix where appropriate to improve readability.

In Sect. 2 we provide a short description of our main result. Section 3 discusses the KR map in infinite dimensions. Its well-definedness in infinite dimensions has been established in [4]. In Theorem 3.3 we additionally give a formula for the push-forward density assuming continuity of the densities w.r.t. the product topology. In Sect. 4 we analyze the regularity of the KR transport. The fact that a transport inherits the smoothness of the densities is known for certain function classes: for example, in the case of C^k densities, [11] shows that the optimal transport also belongs to C^k, and a similar statement holds for the KR transport; see for example [33,Remark 2.19]. In Proposition 4.2, assuming analytic densities we show analyticity of the KR transport. Furthermore, and more importantly, we carefully examine the domain of holomorphic extension to the complex numbers. These results are exploited in Sect. 5 to show convergence of rational function approximations to T in Theorem 5.2. This result proves a *dimension-independent* higher-order convergence rate for the transport of measures supported on infinite-dimensional spaces (which need not be supported on finite-dimensional subspaces). In this result, *all* occurring constants (not just the convergence rate) are controlled independently of the dimension. In Sect. 6 we show that this implies convergence of the pushforward measures (on U and on the Banach space Y) in the Hellinger distance, the total variation distance, the KL divergence, and the Wasserstein distance. These results are formulated in Theorems 6.1 and 6.4. To prove the latter, in Proposition 6.2 we slightly extend a statement from [32] to compact Polish spaces, to show that the Wasserstein distance between two pushforward measures can be bounded by the maximal distance of the two maps pushing forward the initial measure. Finally, we show that it is possible to compute approximate samples from the pushforward measure in the Banach space Y, by mapping a low-dimensional reference sample to the Banach space; see Corollary 6.5. All proofs can be found in the appendix.

2 Main Result

Let for $k \in \mathbb{N}$

$$U_k := [-1, 1]^k \qquad \text{and} \qquad U := [-1, 1]^{\mathbb{N}} \tag{2.1}$$

where these sets are equipped with the product topology and the Borel σ-algebra, which coincides with the product σ-algebra [3,Lemma 6.4.2 (ii)]. Additionally, let $U_0 := \emptyset$. Denote by λ the Lebesgue measure on $[-1, 1]$ and by

$$\mu = \bigotimes_{j \in \mathbb{N}} \frac{\lambda}{2} \tag{2.2}$$

the infinite product measure. Then μ is a (uniform) probability measure on U. By abuse of notation, for $k \in \mathbb{N}$ we additionally denote $\mu = \otimes_{j=1}^{k} \frac{\lambda}{2}$, where k will always be clear from context.

For a *reference* $\rho \ll \mu$ and a *target* measure $\pi \ll \mu$ on U, we investigate the smoothness and approximability of the KR transport $T : U \to U$ satisfying $T_\sharp \rho = \pi$; the notation $T_\sharp \rho$ refers to the pushforward measure defined by $T_\sharp \rho(A) := \rho(\{T(\boldsymbol{y}) \in A : \boldsymbol{y} \in U\})$ for all measurable $A \subseteq U$. While in general there exist multiple maps $T : U \to U$ pushing forward ρ to π, the KR transport is the unique such map satisfying *triangularity* and *monotonicity*. Triangularity refers to the kth component T_k of $T = (T_k)_{k\in\mathbb{N}}$ being a function of the variables $x_1, \ldots, x_k$ only, i.e., $T_k : U_k \to U_1$ for all $k \in \mathbb{N}$. Monotonicity means that $x_k \mapsto T_k(x_1, \ldots, x_{k-1}, x_k)$ is monotonically increasing on U_1 for every $k \in \mathbb{N}$ and every fixed $(x_1, \ldots, x_{k-1}) \in U_{k-1}$.

Absolute continuity of ρ and π w.r.t. μ implies existence of the Radon-Nikodym derivatives

$$f_\rho := \frac{\mathrm{d}\rho}{\mathrm{d}\mu} \qquad \text{and} \qquad f_\pi := \frac{\mathrm{d}\pi}{\mathrm{d}\mu}$$

which will also be referred to as the densities of these measures. Assuming for the moment existence of the KR transport T, approximating T requires approximating the *infinitely many* functions $T_k : U_k \to U_1$, $k \in \mathbb{N}$. This, and the fact that the domain U_k of T_k becomes increasingly high dimensional as $k \to \infty$, makes the problem quite challenging.

For these reasons, further assumptions on ρ and π are necessary. Typical requirements imposed on the measures guarantee some form of intrinsic low dimensionality. Examples include densities belonging to certain reproducing kernel Hilbert spaces, or to other function classes of sufficient regularity. In this paper we concentrate on the latter. As is well-known, if $T_k : U_k \to U_1$ belongs to C^k, then it can be uniformly approximated with the k-independent convergence rate of 1, for instance with multivariate polynomials. The convergence rate to approximate T_k then does not deteriorate with increasing k, but the constants in such error bounds usually still depend exponentially on k. Moreover, as $k \to \infty$, this line of argument requires the components of the map to become arbitrarily regular. For this reason, in the present work, where $T = (T_k)_{k\in\mathbb{N}}$, it is not unnatural to restrict ourselves to transports that are C^∞. More precisely, we in particular assume *analyticity* of the densities f_ρ and f_π, which in turn implies analyticity of T as we shall see. This will allow us to control all occurring constants *independent* of the dimension, and approximate the whole map $T : U \to U$ using only finitely many degrees of freedom in our approximation.

Assume in the following that Z is a Banach space with complexification $Z_\mathbb{C}$; see, e.g., [18, 27] for the complexification of Banach spaces. We may think of Z and $Z_\mathbb{C}$ as real- and complex-valued function spaces, e.g., $Z = L^2([0,1], \mathbb{R})$ and $Z_\mathbb{C} = L^2([0,1], \mathbb{C})$. To guarantee analyticity and the structure in (1.1) we consider densities f of the following type:

Assumption 2.1 For constants $p \in (0, 1)$, $0 < M \leq L < \infty$, a sequence $(\psi_j)_{j\in\mathbb{N}} \subseteq Z$, and a differentiable function $\mathfrak{f} : O_Z \to \mathbb{C}$ with $O_Z \subseteq Z_\mathbb{C}$ open, the following hold:

(a) $\sum_{j\in\mathbb{N}} \|\psi_j\|_Z^p < \infty$,

(b) $\sum_{j\in\mathbb{N}} y_j\psi_j \in O_Z$ for all $\boldsymbol{y} \in U$,
(c) $\mathfrak{f}(\sum_{j\in\mathbb{N}} y_j\psi_j) \in \mathbb{R}$ for all $\boldsymbol{y} \in U$,
(d) $M = \inf_{\psi\in O_Z} |\mathfrak{f}(\psi)| \le \sup_{\psi\in O_Z} |\mathfrak{f}(\psi)| = L$.

The function $f : U \to \mathbb{R}$ given by

$$f(\boldsymbol{y}) := \mathfrak{f}\Big(\sum_{j\in\mathbb{N}} \psi_j y_j\Big) \tag{2.3}$$

satisfies $\int_U f(\boldsymbol{y})\, \mathrm{d}\mu(\boldsymbol{y}) = 1$.

Assumption 2.2 For two sequences $(\psi_{*,j})_{j\in\mathbb{N}} \in Z$ with $(*, Z) \in \{(\rho, X), (\pi, Y)\}$, the functions

$$f_\rho(\boldsymbol{y}) = \mathfrak{f}_\rho\Big(\sum_{j\in\mathbb{N}} y_j\psi_{\rho,j}\Big), \qquad f_\pi(\boldsymbol{y}) = \mathfrak{f}_\pi\Big(\sum_{j\in\mathbb{N}} y_j\psi_{\pi,j}\Big)$$

both satisfy Assumption 2.1 for some fixed constants $p \in (0, 1)$ and $0 < M \le L < \infty$.

The summability parameter p determines the decay rate of the functions ψ_j—the smaller p the stronger the decay of the ψ_j. Because $p < 1$, the argument of $\mathfrak{f}$ in (2.3) is well-defined for $\boldsymbol{y} \in U$ since $\sum_{j\in\mathbb{N}} |y_j|\|\psi_j\|_Z < \infty$.

Our main result is about the existence and approximation of the KR-transport $T : U \to U$ satisfying $T_\sharp\rho = \pi$. We state the result here in a simplified form; more details will be given in Theorems 5.2, 6.1, and 6.4. We only mention that the trivial approximation $T_k(x_1, \ldots, x_k) \simeq x_k$ is interpreted as not requiring any degrees of freedom in the following theorem.

Theorem 2.3 *Let $f_\rho : U \to (0, \infty)$ and $f_\pi : U \to (0, \infty)$ be two probability densities as in Assumption 2.2 for some $p \in (0, 1)$. Then there exists a unique triangular, monotone, and bijective map $T : U \to U$ satisfying $T_\sharp\rho = \pi$.*

Moreover, for every $N \in \mathbb{N}$ there exists a space of rational functions employing N degrees of freedom, and a bijective, monotone, and triangular $\tilde{T} : U \to U$ in this space such that

$$\mathrm{dist}(\tilde{T}_\sharp\rho, \pi) \le CN^{-\frac{1}{p}+1}. \tag{2.4}$$

*Here C is a constant independent of N and "*dist*" may refer to the total variation distance, the Hellinger distance, the KL divergence, or the Wasserstein distance.*

Equation (2.4) shows a dimension-independent convergence rate (indeed our transport is defined on the infinite-dimensional domain $U = [-1, 1]^{\mathbb{N}}$), so that the curse of dimensionality is overcome. The rate of algebraic convergence becomes arbitrarily large as $p \in (0, 1)$ in Assumption 2.1 becomes small. The convergence rate $\frac{1}{p} - 1$ in Theorem 2.3 is well-known for the approximation of functions as in (2.3) by sparse polynomials, e.g., [6–8]; also see Remark 2.7. There is a key difference to earlier results dealing with the approximation of such functions: we do not approximate the function $f : U \to \mathbb{R}$ in (2.3), but instead we approximate the transport $T : U \to U$, i.e., an infinite number of functions. Our main observation in this paper is that the sparsity of the densities f_ρ and f_π carries over to the transport. Even though it has infinitely many components, T can still be approximated very efficiently if the ansatz space is carefully chosen and tailored to the specific densities. In addition to showing error convergence (2.4), in Theorem 5.2 we give concrete ansatz spaces achieving this convergence rate. These ansatz spaces can be computed in linear complexity and may be used in applications.

The main application for our result is to provide a method to sample from the target π or the pushforward $\Phi_\sharp\pi$ in the Banach space Y, where $\Phi(\boldsymbol{y}) = \sum_{j\in\mathbb{N}} y_j\psi_{\pi,j}$. Given an approximation $\tilde{T} = (\tilde{T}_j)_{j\in\mathbb{N}}$ to T, this is achieved via $\Phi(\tilde{T}(\boldsymbol{y}))$ for $\boldsymbol{y} \sim \rho$. It is natural to truncate this expansion, which yields

$$\sum_{j=1}^{s} \tilde{T}_j(y_1, \dots, y_j)\psi_{\pi,j}$$

for some truncation parameter $s \in \mathbb{N}$ and $(y_1, \dots, y_s) \in U_s$. This map transforms a sample from a distribution on the s-dimensional space U_s to a sample from an infinite-dimensional distribution on Y. In Corollary 6.5 we show that the error of this truncated representation in the Wasserstein distance converges with the same rate as given in Theorem 2.3.

Remark 2.4 The reference ρ is a "simple" measure whose main purpose is to allow for easy sampling. One possible choice for ρ (that we have in mind throughout this paper) is the uniform measure μ. It trivially satisfies Assumption 2.1 with $\mathfrak{f}_\rho : \mathbb{C} \to \mathbb{C}$ being the constant 1 function (and, e.g., $\psi_{\rho,j} = 0 \in \mathbb{C}$).

Remark 2.5 Even though we can think of ρ as being μ, we formulated Theorem 2.3 in more generality, mainly for the following reason: since the assumptions on ρ and π are the same, we may switch their roles. Thus Theorem 2.3 can be turned into a statement about the inverse transport $S{:=}T^{-1} : U \to U$, which can also be approximated at the rate $\frac{1}{p} - 1$.

Example 2.6 (Bayesian inference) For a Banach space Y ("parameter space") and a Banach space $\mathcal{X}$ ("solution space"), let $\mathfrak{u} : O_Y \to \mathcal{X}_\mathbb{C}$ be a complex differentiable *forward operator*. Here $O_Y \subseteq Y_\mathbb{C}$ is some nonempty open set. Let $G : \mathcal{X} \to \mathbb{R}^m$ be a bounded linear *observation operator*. For some unknown $\psi \in Y$ we are given a noisy observation of the system in the form

$$\varsigma = G(\mathfrak{u}(\psi)) + \eta \in \mathbb{R}^m,$$

where $\eta \sim \mathcal{N}(0, \Gamma)$ is a centered Gaussian random variable with symmetric positive definite covariance $\Gamma \in \mathbb{R}^{m\times m}$. The goal is to recover ψ given the measurement ς.

To formulate the Bayesian inverse problem, we first fix a prior: let $(\psi_j)_{j\in\mathbb{N}}$ be a summable sequence of linearly independent elements in Y. With

$$\Phi(\boldsymbol{y}):=\sum_{j\in\mathbb{N}} y_j\psi_j$$

and the uniform measure μ on U, we choose the prior $\Phi_\sharp\mu$ on Y. Determining ψ within the set $\{\Phi(\boldsymbol{y}) : \boldsymbol{y}\in U\}\subseteq Y$ is equivalent to determining the coefficient sequence $\boldsymbol{y}\in U$. Assuming independence of $\boldsymbol{y}\sim\mu$ and $\eta\sim\mathcal{N}(0,\Gamma)$, the distribution of $\boldsymbol{y}$ given ς (the posterior) can then be characterized by its density w.r.t. μ, which, up to a normalization constant, equals

$$\exp\left(\left(\varsigma - G\left(\mathfrak{u}\left(\sum_{j\in\mathbb{N}} y_j\psi_j\right)\right)\right)^\top \Gamma^{-1}\left(\varsigma - G\left(\mathfrak{u}\left(\sum_{j\in\mathbb{N}} y_j\psi_j\right)\right)\right)\right). \tag{2.5}$$

This posterior density is of form (2.3) and the corresponding measure π can be chosen as a target in Theorem 2.3. Given T satisfying $T_\sharp\rho=\pi$, we may then explore π to perform inference on the unknown $\boldsymbol{y}$; see for instance [41,Sect. 7.4]. For more details on the rigorous derivation of (2.5) we refer to [34] and in particular [9,Sect. 3].

Remark 2.7 Functions as in Assumption 2.1 belong to the set of so-called $(\boldsymbol{b},p,\varepsilon)$-holomorphic functions; see, e.g., [6]. This class contains infinite parametric functions that are holomorphic in each argument y_j, and exhibit some growth in the domain of holomorphic extension as $j\to\infty$. The results of the present paper and the key arguments remain valid if we replace Assumption 2.1 with the $(\boldsymbol{b},p,\varepsilon)$-holomorphy assumption. Since most relevant examples of such functions are of specific type (2.3), we restrict the discussion to this case in order to avoid technicalities.

3 The Knothe–Rosenblatt Transport in Infinite Dimensions

Recall that we consider the product topology on $U=[-1,1]^{\mathbb{N}}$. Assume that $f_\rho\in C^0(U,\mathbb{R}_+)$ and $f_\pi\in C^0(U,\mathbb{R}_+)$ are two positive probability densities. Here $\mathbb{R}_+:=(0,\infty)$, and $C^0(U,\mathbb{R}_+)$ denotes the continuous functions from $U\to\mathbb{R}_+$. We now recall the construction of the KR map.

For $\boldsymbol{y}=(y_j)_{j\in\mathbb{N}}\in\mathbb{C}^{\mathbb{N}}$ and $1\le k\le n<\infty$ let

$$\boldsymbol{y}_{[k]}:=(y_j)_{j=1}^k,\qquad \boldsymbol{y}_{[k:n]}:=(y_j)_{j=k}^n,\qquad \boldsymbol{y}_{[n:]}:=(y_j)_{j\ge n}. \tag{3.1}$$

For $*\in\{\rho,\pi\}$ and $\boldsymbol{y}\in U$ define

$$\hat{f}_{*,0}(\boldsymbol{y}):=1 \tag{3.2a}$$

and for $k \in \mathbb{N}$

$$\hat{f}_{*,k}(\boldsymbol{y}_{[k]}) := \int_U f_*(\boldsymbol{y}_{[k]}, \boldsymbol{t}) \, \mathrm{d}\mu(\boldsymbol{t}) > 0, \qquad f_{*,k}(\boldsymbol{y}_{[k]}) := \frac{\hat{f}_{*,k}(\boldsymbol{y}_{[k]})}{\hat{f}_{*,k-1}(\boldsymbol{y}_{[k-1]})} > 0. \tag{3.2b}$$

Then, $\boldsymbol{y}_{[k]} \mapsto \hat{f}_{\rho,k}(\boldsymbol{y}_{[k]})$ is the marginal density of ρ in the first k variables $\boldsymbol{y}_{[k]} \in U_k$, and we denote the corresponding measure on U_k by ρ_k. Similarly, $y_k \mapsto f_{\rho,k}(\boldsymbol{y}_{[k-1]}, y_k)$ is the conditional density of y_k given $\boldsymbol{y}_{[k-1]}$, and the corresponding measure on U_1 is denoted by $\rho_k^{\boldsymbol{y}_{[k-1]}}$. The same holds for the densities of π, and we use the analogous notation π_k and $\pi_k^{\boldsymbol{y}_{[k-1]}}$ for the marginal and conditional measures.

Recall that for two atomless measures η and ν on U_1 with distribution functions $F_\eta : U_1 \to [0, 1]$ and $F_\nu : U_1 \to [0, 1]$, $F_\eta^{-1} \circ F_\nu : U_1 \to U_1$ pushes forward ν to η, as is easily checked, e.g., [33,Theorem 2.5]. In case η and ν have positive densities on U_1, this map is the unique strictly monotonically increasing such function. With this in mind, the KR-transport can be constructed as follows: let $T_1 : U_1 \to U_1$ be the (unique) monotonically increasing transport satisfying

$$(T_1)_\sharp \rho_1 = \pi_1. \tag{3.3a}$$

Analogous to (3.1) denote $T_{[k]} := (T_j)_{j=1}^k : U_k \to U_k$. Let inductively for any $\boldsymbol{y} \in U$, $T_{k+1}(\boldsymbol{y}_{[k]}, \cdot) : U_1 \to U_1$ be the (unique) monotonically increasing transport such that

$$(T_{k+1}(\boldsymbol{y}_{[k]}, \cdot))_\sharp \rho_{k+1}^{\boldsymbol{y}_{[k]}} = \pi_{k+1}^{T_{[k]}(\boldsymbol{y}_{[k]})}. \tag{3.3b}$$

Note that $T_{k+1} : U_{k+1} \to U_1$ and thus $T_{[k+1]} = (T_j)_{j=1}^{k+1} : U_{k+1} \to U_{k+1}$. It can then be shown that for any $k \in \mathbb{N}$ [33,Proposition 2.18]

$$(T_{[k]})_\sharp \rho_k = \pi_k. \tag{3.4}$$

By induction this construction yields a map $T := (T_k)_{k \in \mathbb{N}}$ where each $T_k : U_k \to U_1$ satisfies that $T_k(\boldsymbol{y}_{[k-1]}, \cdot) : U_1 \to U_1$ is strictly monotonically increasing and bijective. This implies that $T : U \to U$ is bijective, as follows. First, to show *injectivity*: let $\boldsymbol{x} \neq \boldsymbol{y} \in U$ and $j = \operatorname{argmin}\{i : x_i \neq y_i\}$. Since $t \mapsto T_j(x_1, \ldots, x_{j-1}, t)$ is bijective, $T_j(x_1, \ldots, x_{j-1}, x_j) \neq T_j(x_1, \ldots, x_{j-1}, y_j)$ and thus $T(\boldsymbol{x}) \neq T(\boldsymbol{y})$. Next, to show *surjectivity:* fix $\boldsymbol{y} \in U$. Bijectivity of $T_1 : U_1 \to U_1$ implies existence of $x_1 \in U_1$ such that $T_1(x_1) = y_1$. Inductively choose x_j such that $T_j(x_1, \ldots, x_j) = y_j$. Then $T(\boldsymbol{x}) = \boldsymbol{y}$. Thus:

Lemma 3.1 *Let $T = (T_k)_{k \in \mathbb{N}} : U \to U$ be triangular. If $t \mapsto T_k(\boldsymbol{y}_{[k-1]}, t)$ is bijective from $U_1 \to U_1$ for every $\boldsymbol{y} \in U$ and $k \in \mathbb{N}$, then $T : U \to U$ is bijective.*

The continuity assumption on the densities guarantees that the marginal densities on U_k converge uniformly to the full density, as we show next. This indicates that in principle it is possible to approximate the infinite-dimensional transport map by restricting to finitely many dimensions.

Lemma 3.2 *Let $f \in C^0(U, \mathbb{R}_+)$, and let $\hat{f}_k$ and f_k be as in (3.2). Then*

(i) f is measurable and $f \in L^2(U, \mu)$,
(ii) $\hat{f}_k \in C^0(U_k, \mathbb{R}_+)$ and $f_k \in C^0(U_k, \mathbb{R}_+)$ for every $k \in \mathbb{N}$,
(iii) it holds

$$\lim_{k\to\infty} \sup_{y\in U} |\hat{f}_k(\boldsymbol{y}_{[k]}) - f(\boldsymbol{y})| = 0. \tag{3.5}$$

Throughout what follows T always stands for the KR transport defined in (3.3). Next we show that T indeed pushes forward ρ to π, and additionally we provide a formula for the transformation of densities. In the following $\partial_j g(\boldsymbol{x}):=\frac{\partial}{\partial x_j} g(\boldsymbol{x})$. Furthermore, we call $f : U \to \mathbb{R}$ a *positive probability density* if $f(\boldsymbol{y}) > 0$ for all $\boldsymbol{y} \in U$ and $\int_U f(\boldsymbol{y})\, \mathrm{d}\mu(\boldsymbol{y}) = 1$.

Theorem 3.3 *Let f_π, $f_\rho \in C^0(U, \mathbb{R}_+)$ be two positive probability densities. Then*

(i) $T = (T_k)_{k\in\mathbb{N}} : U \to U$ is measurable, bijective and satisfies $T_\sharp\rho = \pi$,
(ii) for each $k \in \mathbb{N}$ holds $\partial_k T_k(\boldsymbol{y}_{[k]}) \in C^0(U_k, \mathbb{R}_+)$ and

$$\det dT(\boldsymbol{y}):= \lim_{n\to\infty} \prod_{j=1}^{n} \partial_j T_j(\boldsymbol{y}_{[j]}) \in C^0(U, \mathbb{R}_+) \tag{3.6}$$

is well-defined (i.e., converges in $C^0(U, \mathbb{R}_+)$). Moreover

$$f_\pi(T(\boldsymbol{y})) \det dT(\boldsymbol{y}) = f_\rho(\boldsymbol{y}) \qquad \forall \boldsymbol{y} \in U.$$

Remark 3.4 Switching the roles of f_ρ and f_π, for $S = T^{-1}$ it holds $f_\rho(S(\boldsymbol{y})) \det dS(\boldsymbol{y}) = f_\pi(\boldsymbol{y})$ for all $\boldsymbol{y} \in U$, where $\det dS(\boldsymbol{y}):= \lim_{n\to\infty} \prod_{j=1}^{n} \partial_j S_j(\boldsymbol{y}_{[j]})$ is well-defined.

4 Analyticity of *T*

In this section we investigate the domain of analytic extension of T. To state our results, for $\delta > 0$ and $D \subseteq \mathbb{C}$ we introduce the complex sets

$$\mathcal{B}_\delta:=\{z \in \mathbb{C} : |z| < \delta\} \qquad \text{and} \qquad \mathcal{B}_\delta(D):=\{z + y : z \in \mathcal{B}_\delta,\ y \in D\},$$

and for $k \in \mathbb{N}$ and $\boldsymbol{\delta} \in (0, \infty)^k$

$$\mathcal{B}_{\boldsymbol{\delta}}:= \times_{j=1}^{k} \mathcal{B}_{\delta_j} \qquad \text{and} \qquad \mathcal{B}_{\boldsymbol{\delta}}(D):= \times_{j=1}^{k} \mathcal{B}_{\delta_j}(D),$$

which are subsets of $\mathbb{C}^k$. Their closures will be denoted by $\bar{\mathcal{B}}_\delta$, etc. If we write $\mathcal{B}_{\boldsymbol{\delta}}(U_1) \times U$ we mean elements $\boldsymbol{y} \in \mathbb{C}^{\mathbb{N}}$ with $y_j \in \mathcal{B}_{\delta_j}(U_1)$ for $j \le k$ and $y_j \in U_1$ otherwise. Subsets of $\mathbb{C}^{\mathbb{N}}$ are always equipped with the product topology.

In this section we analyze the domain of holomorphic extension of each component $T_k : U_k \to U_1$ of T to subsets of $\mathbb{C}^k$. The reason why we are interested in such statements, is that they allow to upper bound the expansion coefficients w.r.t. certain polynomial bases: for a multiindex $\boldsymbol{\nu} \in \mathbb{N}_0^k$ (where $\mathbb{N}_0 = \{0, 1, 2, \dots\}$) let $L_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{j=1}^k L_{\nu_j}(y_j)$ be the product of the one-dimensional Legendre polynomials normalized in $L^2(U_1, \mu)$. Then $(L_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathbb{N}_0^k}$ forms an orthonormal basis of $L^2(U_k, \mu)$. Hence we can expand $\partial_k T_k(\boldsymbol{y}_{[k]}) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} l_{k,\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y}_{[k]})$ for $\boldsymbol{y} \in U$ and with the Legendre coefficients

$$l_{k,\boldsymbol{\nu}} = \int_{U_k} \partial_k T_k(\boldsymbol{y}_{[k]}) L_{\boldsymbol{\nu}}(\boldsymbol{y}_{[k]}) \in \mathbb{R}. \tag{4.1}$$

Analyticity of T_k (and thus of $\partial_k T_k$) on the set $\mathcal{B}_{\boldsymbol{\delta}}(U_1)$ implies bounds of the type (see Lemma C.3)

$$|l_{k,\boldsymbol{\nu}}| \le C \|\partial_k T_k\|_{L^\infty(\mathcal{B}_{\boldsymbol{\delta}}(U_1))} \prod_{j=1}^k (1+\delta_j)^{-\nu_j}. \tag{4.2}$$

Here C in particular depends on $\min_j \delta_j > 0$. The exponential decay in each ν_j leads to exponential convergence of truncated sparse Legendre expansions. Once we have approximated $\partial_k T_k$, we integrate this term in x_k to obtain an approximation to T_k. The reason for not approximating T_k directly is explained after Proposition 4.2 below; see (4.5). The size of the holomorphy domain (the size of $\boldsymbol{\delta}$) determines the constants in these estimates—the larger the entries of $\boldsymbol{\delta}$, the smaller the upper bound (4.2) and the faster the convergence.

We are now in position to present our main technical tool to find suitable holomorphy domains of each T_k (or equivalently $\partial_k T_k$). We will work under the following assumption on the two densities $f_\rho : U \to (0, \infty)$ and $f_\pi : U \to (0, \infty)$. The assumption is a modification of [41,Assumption 3.5].

Assumption 4.1 For constants C_1, $M > 0$, $L < \infty$, $k \in \mathbb{N}$, and $\boldsymbol{\delta} \in (0, \infty)^k$, the following hold:

(a) $f \in C^0(\mathcal{B}_{\boldsymbol{\delta}}(U_1) \times U, \mathbb{C})$ and $f : U \to \mathbb{R}_+$ is a probability density,
(b) $\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y}) \in C^1(\mathcal{B}_{\boldsymbol{\delta}}(U_1), \mathbb{C})$ for all $\boldsymbol{y} \in U$,
(c) $M \le |f(\boldsymbol{y})| \le L$ for all $\boldsymbol{y} \in \mathcal{B}_{\boldsymbol{\delta}}(U_1) \times U$,
(d) $\sup_{\boldsymbol{y}\in\mathcal{B}_{\boldsymbol{\delta}}\times\{0\}^{\mathbb{N}}} |f(\boldsymbol{x}+\boldsymbol{y}) - f(\boldsymbol{x})| \le C_1$ for all $\boldsymbol{x} \in U$,
(e) $\sup_{\boldsymbol{y}\in\mathcal{B}_{\boldsymbol{\delta}_{[j]}}\times\{0\}^{\mathbb{N}}} |f(\boldsymbol{x}+\boldsymbol{y}) - f(\boldsymbol{x})| \le C_1\delta_{j+1}$ for all $\boldsymbol{x} \in U$ and $j \in \{1, \dots, k-1\}$.

Such densities yield certain holomorphy domains for T_k as we show in the next proposition, which is an infinite-dimensional version of [41,Theorem 3.6].

Proposition 4.2 *Let $k \in \mathbb{N}$, $\boldsymbol{\delta} \in (0, \infty)^k$ and $0 < M \le L < \infty$. There exist $C_1 > 0$, $C_2 \in (0, 1]$ and $C_3 > 0$ solely depending on M and L (but not on k or $\boldsymbol{\delta}$) such that if f_ρ and f_π satisfy Assumption 4.1 with C_1, M, L, and $\boldsymbol{\delta}$, then:*

With $\boldsymbol{\zeta} = (\zeta_j)_{j=1}^k$ *defined by*

$$\zeta_j := C_2\delta_j \qquad \forall j \in \{1, \dots, k\}, \tag{4.3}$$

it holds for all $j \in \{1, \dots, k\}$ *with* $R_j := \partial_j T_j$ *(with* T *as in (3.3)) that*

(i) $R_j \in C^1(\mathcal{B}_{\boldsymbol{\zeta}_{[j]}}(U_1), \mathcal{B}_{C_3}(1))$ *and* $\Re(R_j(\boldsymbol{x})) \geq \frac{1}{C_3}$ *for all* $\boldsymbol{x} \in \mathcal{B}_{\boldsymbol{\zeta}_{[j]}}(U_1)$,
(ii) if $j \geq 2$, $R_j : \mathcal{B}_{\boldsymbol{\zeta}_{[j-1]}}(U_1) \times U_1 \to \mathcal{B}_{\frac{C_3}{\delta_j}}(1)$.

Let us sketch how this result can be used to show that T_k can be approximated by polynomial expansions. In Appendix B.2 we will verify Assumption 4.1 for densities as in (2.3). Proposition 4.2 (i) then provides a holomorphy domain for $\partial_k T_k$, and together with (4.2) we can bound the expansion coefficients $l_{k,\boldsymbol{\nu}}$ of $\partial_k T_k = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} l_{k,\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y})$. However, there is a catch: in general one can find different $\boldsymbol{\delta}$ such that Assumption 4.1 holds. The difficulty is to choose $\boldsymbol{\delta}$ in a way that depends on $\boldsymbol{\nu}$ to obtain a possibly sharp bound in (4.2). To do so we will use ideas from, e.g., [6] where similar calculations were made.

The outlined argument based on Proposition 4.2 (i) suffices to prove convergence of sparse polynomial expansions in the finite-dimensional case; see [41,Theorem 4.6]. In the infinite-dimensional case where we want to approximate $T = (T_k)_{k\in\mathbb{N}}$ with only finitely many degrees of freedom we additionally need to employ Proposition 4.2 (ii): for $\boldsymbol{\nu} \in \mathbb{N}_0^k$ such that $\boldsymbol{\nu} \neq \boldsymbol{0} := (0)_{j=1}^k$ but $\nu_k = 0$, Proposition 4.2 (ii) together with (4.2) implies a bound of the type

$$|l_{k,\boldsymbol{\nu}}| = \left| \int_{U_k} (\partial_k T_k(\boldsymbol{y}_{[k]}) - 1) L_{\boldsymbol{\nu}}(\boldsymbol{y}_{[k]}) \,\mathrm{d}\mu(\boldsymbol{y}_{[k]}) \right| \leq C \frac{1}{\delta_k} \prod_{j=1}^k (1+\delta_j)^{-\nu_j}, \tag{4.4}$$

where the additional $\frac{1}{\delta_k}$ stems from $\|\partial_k T_k - 1\|_{L^\infty(\mathcal{B}_{\boldsymbol{\zeta}_{[j-1]}}(U_1)\times U_1)} \leq \frac{C_3}{\delta_k}$. Here we used the fact $\int_{U_k} L_{\boldsymbol{\nu}}(\boldsymbol{y}_{[k]}) \,\mathrm{d}\mu(\boldsymbol{y}_{[k]}) = 0$ for all $\boldsymbol{\nu} \neq \boldsymbol{0}$ by orthogonality of the $(L_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathbb{N}_0^k}$ and because $L_{\boldsymbol{0}} \equiv 1$. In case $\nu_k \neq 0$, then the factor $\frac{1}{1+\delta_k}$ occurs on the right-hand side of (4.2) . Hence, *all* coefficients $l_{k,\boldsymbol{\nu}}$ for which $\boldsymbol{\nu} \neq \boldsymbol{0}$ are of size $O(\frac{1}{\delta_k})$. In fact one can show that even $\sum_{\boldsymbol{\nu}\neq\boldsymbol{0}} |l_{k,\boldsymbol{\nu}}||L_{\boldsymbol{\nu}}(\boldsymbol{y}_{[k]})|$ is of size $O(\frac{1}{\delta_k})$. Thus

$$\partial_k T_k(\boldsymbol{y}_{[k]}) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} l_{k,\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y}_{[k]}) = l_{k,\boldsymbol{0}} L_{\boldsymbol{0}}(\boldsymbol{y}_{[k]}) + O\left(\frac{1}{\delta_k}\right).$$

Using $L_{\boldsymbol{0}} \equiv 1$

$$\begin{aligned} l_{k,\boldsymbol{0}} &= \int_{U_k} \partial_k T_k(\boldsymbol{y}_{[k]}) L_{\boldsymbol{0}}(\boldsymbol{y}_{[k]}) \,\mathrm{d}\mu(\boldsymbol{y}_{[k]}) \\ &= \int_{U_{k-1}} T_k(\boldsymbol{y}_{[k-1]}, 1) - T_k(\boldsymbol{y}_{[k-1]}, -1) \,\mathrm{d}\mu(\boldsymbol{y}_{[k-1]}) = 2, \end{aligned}$$

and therefore if δ_k is very large, since $L_{\mathbf{0}} \equiv 1$

$$T_k(\boldsymbol{y}_{[k]}) = -1 + \int_{-1}^{y_k} \partial_k T_k(\boldsymbol{y}_{[k-1]}, t)\, \mathrm{d}\mu(t)$$
$$\simeq -1 + \int_{-1}^{y_k} l_{k,\mathbf{0}} L_{\mathbf{0}}(\boldsymbol{y}_{[k-1]}, t)\, \mathrm{d}\mu(t) = y_k. \tag{4.5}$$

Hence, for large δ_k we can use the trivial approximation $T_k(\boldsymbol{y}_{[k]}) \simeq y_k$. To address this special role played by the kth variable for the kth component we introduce

$$\gamma(\boldsymbol{\varrho}, \boldsymbol{\nu}) := \varrho_k^{-\max\{1,\nu_k\}} \prod_{j=1}^{k-1} \varrho_j^{-\nu_j} \qquad \forall \boldsymbol{\varrho} \in (1,\infty)^{\mathbb{N}},\ \boldsymbol{\nu} \in \mathbb{N}_0^k, \tag{4.6}$$

which, up to constants, corresponds to the minimum of (4.2) and (4.4). This quantity can be interpreted as measuring the importance of the monomial $\boldsymbol{y}^{\boldsymbol{\nu}}$ in the ansatz space used for the approximation of T_k, and we will use it to construct such ansatz spaces.

Remark 4.3 To explain the key ideas, in this section we presented the approximation of T_k via a Legendre expansion of $\partial_k T_k$. For the proofs of our approximation results in Sect. 5 we instead approximate $\sqrt{\partial_k T_k} - 1$ with truncated Legendre expansions. This will guarantee the approximate transport to satisfy the monotonicity property as explained in Sect. 5.

5 Convergence of the Transport

We are now in position to state an algebraic convergence result for approximations of infinite-dimensional transport maps $T : U \to U$ associated to densities of type (2.3).

For a triangular approximation $\tilde{T} = (\tilde{T}_k)_{k\in\mathbb{N}}$ to T it is desirable that it retains the monotonicity and bijectivity properties, i.e., $\partial_k \tilde{T}_k > 0$ and $\tilde{T} : U \to U$ is bijective. The first guarantees that $\tilde{T}$ is injective and easy to invert (by subsequently solving the one-dimensional equations $x_k = \tilde{T}_k(y_1, \ldots, y_k)$ for y_k starting with $k = 1$), and for the purpose of generating samples, the second property ensures that for $\boldsymbol{y} \sim \rho$, the transformed sample $\tilde{T}(\boldsymbol{y}) \sim \tilde{T}_\sharp \rho$ also belongs to U. These constraints are hard to enforce for polynomial approximations. For this reason, we use the same rational parametrization we introduced in [41] for the finite-dimensional case: for a set of k-dimensional multiindices $\Lambda \subseteq \mathbb{N}_0^k$, define

$$\mathbb{P}_\Lambda := \mathrm{span}\{\boldsymbol{y}^{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \Lambda\}.$$

The dimension of this space is equal to the cardinality of Λ, which we denote by $|\Lambda|$. Let $p_k \in \mathbb{P}_\Lambda$ (where Λ remains to be chosen) be a polynomial approximation to $\sqrt{\partial_k T_k} - 1$. Set for $\boldsymbol{y} \in U_k$

$$\tilde{T}_k(\boldsymbol{y}) := -1 + 2\frac{\int_{-1}^{y_k} \int_{U_{k-1}} (p_k(\boldsymbol{y}_{[k-1]}, t) + 1)^2\, \mathrm{d}\mu(\boldsymbol{y}_{[k-1]})\, \mathrm{d}\mu(t)}{\int_{U_k} (p_k(\boldsymbol{y}) + 1)^2\, \mathrm{d}\mu(\boldsymbol{y})}. \tag{5.1}$$

It is easily checked that $\tilde{T}_k$ satisfies both monotonicity and bijectivity as long as $p_k \neq -1$. Thus we end up with a rational function $\tilde{T}_k$, but we emphasize that the use of rational functions instead of polynomials is not due to better approximation capabilities, but solely to guarantee bijectivity of $\tilde{T} : U \to U$.

Remark 5.1 Observe that $\Lambda = \emptyset$ gives the trivial approximation $p_k :=0 \in \mathbb{P}_\emptyset$ and $\tilde{T}_k(\boldsymbol{y}) = y_k$.

The following theorem yields an algebraic convergence rate *independent of the dimension* (since the dimension is infinity) in terms of the total number of degrees of freedom for the approximation of T. Therefore the curse of dimensionality is overcome for densities as in Assumption 2.1.

Theorem 5.2 *Let f_ρ, $f_\pi : U \to (0, \infty)$ be two probability densities satisfying Assumption 2.2 for some $p \in (0, 1)$. Set $b_j := \max\{\|\psi_{\rho,j}\|_Z, \|\psi_{\pi,j}\|_Z\}$, $j \in \mathbb{N}$.*

There exist $\alpha > 0$ and $C > 0$ such that the following holds: for $j \in \mathbb{N}$ set

$$\varrho_j := 1 + \frac{\alpha}{b_j}, \tag{5.2}$$

and with $\gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})$ as in (4.6) define

$$\Lambda_{\varepsilon,k} := \{\boldsymbol{\nu} \in \mathbb{N}_0^k : \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu}) \geq \varepsilon\} \qquad \forall k \in \mathbb{N}.$$

For each $k \in \mathbb{N}$ there exists a polynomial $p_k \in \mathbb{P}_{\Lambda_{\varepsilon,k}}$ such that with the components $\tilde{T}_{\varepsilon,k}$ as in (5.1), $\tilde{T}_\varepsilon = (\tilde{T}_{\varepsilon,k})_{k\in\mathbb{N}} : U \to U$ is a monotone triangular bijection. For all $\varepsilon > 0$, it holds that $N_\varepsilon := \sum_{k\in\mathbb{N}} |\Lambda_{\varepsilon,k}| < \infty$ and

$$\sum_{k\in\mathbb{N}} \|T_k - \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \leq C N_\varepsilon^{-\frac{1}{p}+1} \tag{5.3a}$$

and

$$\sum_{k\in\mathbb{N}} \|\partial_k T_k - \partial_k \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \leq C N_\varepsilon^{-\frac{1}{p}+1}. \tag{5.3b}$$

Remark 5.3 Fix $\varepsilon > 0$. Since $N_\varepsilon < \infty$, there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ holds $\Lambda_{\varepsilon,k} = \emptyset$ and thus $\tilde{T}_{\varepsilon,k}(\boldsymbol{y}_{[k]}) = y_k$, cp. Remark 5.1.

Switching the roles of ρ and π, Theorem 5.2 also yields an approximation result for the inverse transport $S = T^{-1}$ by some rational functions $\tilde{S}_k$ as in (5.1). Moreover, if $\tilde{T}$ is the rational approximation from Theorem 5.2, then its inverse $\tilde{T}^{-1} : U \to U$ (whose components are not necessarily rational functions) also satisfies an error bound of type (5.3) as we show next.

Corollary 5.4 *Consider the setting of Theorem 5.2. Denote* $S := T^{-1} : U \to U$ *and* $\tilde{S}_\varepsilon := \tilde{T}_\varepsilon^{-1} : U \to U$. *Then there exists a constant* C *such that for all* $\varepsilon > 0$

$$\sum_{k\in\mathbb{N}} \|S_k - \tilde{S}_{\varepsilon,k}\|_{L^\infty(U_k)} \le C N_\varepsilon^{-\frac{1}{p}+1} \tag{5.4a}$$

and

$$\sum_{k\in\mathbb{N}} \|\partial_k S_k - \partial_k \tilde{S}_{\varepsilon,k}\|_{L^\infty(U_k)} \le C N_\varepsilon^{-\frac{1}{p}+1}. \tag{5.4b}$$

Note that both S and $\tilde{S}_\varepsilon$ in Corollary 5.4 are monotonic, triangular bijections as they are the inverses of such maps.

6 Convergence of the Pushforward Measures

Theorem 5.2 established smallness of $\sum_{k\in\mathbb{N}} |\partial_k(T_k - \tilde{T}_k)|$. The relevance of this term stems from the formal calculation (cp. (3.6))

$$|\det dT - \det d\tilde{T}| = \left|\prod_{k\in\mathbb{N}} \partial_k T_k - \prod_{k\in\mathbb{N}} \partial_k \tilde{T}_k\right| \le \sum_{k\in\mathbb{N}} |\partial_k T_k - \partial_k \tilde{T}_k| \prod_{j<k} |\partial_j T_j| \prod_{i>k} |\partial_i \tilde{T}_i|.$$

Assuming that we can bound the last two products, the determinant $\det d\tilde{T}$ converges to $\det dT$ at the rate given in Theorem 5.2. This will allow us to bound the Hellinger distance (H), the total variation distance (TV), and the Kullback-Leibler divergence (KL) between $\tilde{T}_\sharp\rho$ and π, as we show in the following theorem. Recall that for two probability measures $\nu \ll \mu$, $\eta \ll \mu$ on U with densities $f_\nu = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, $f_\eta = \frac{\mathrm{d}\eta}{\mathrm{d}\mu}$,

$$\mathrm{H}(\nu,\eta) = \frac{1}{\sqrt{2}} \|\sqrt{f_\nu} - \sqrt{f_\eta}\|_{L^2(U,\mu)}, \quad \mathrm{TV}(\nu,\eta) = \frac{1}{2}\|f_\nu - f_\eta\|_{L^1(U,\mu)},$$
$$\mathrm{KL}(\nu,\eta) = \int_U \log\left(\frac{f_\nu}{f_\eta}\right) \mathrm{d}\nu.$$

Theorem 6.1 *Let* f_ρ, f_π *satisfy Assumption 2.2 for some* $p \in (0,1)$, *and let* $\tilde{T}_\varepsilon : U \to U$ *be the approximate transport from Theorem 5.2.*

Then there exists $C > 0$ *such that for* $\mathrm{dist} \in \{\mathrm{H}, \mathrm{TV}, \mathrm{KL}\}$ *and every* $\varepsilon > 0$

$$\mathrm{dist}((\tilde{T}_\varepsilon)_\sharp\mu, \pi) \le C N_\varepsilon^{-\frac{1}{p}+1}. \tag{6.1}$$

Next we treat the Wasserstein distance. Recall that for a Polish space (M, d) (i.e., M is separable and complete with the metric d on M) and for $q \in [1,\infty)$, the q-Wasserstein distance between two probability measures ν and η on M (equipped with the Borel σ-algebra) is defined as [37,Def. 6.1]

$$W_q(\nu,\eta) := \inf_{\gamma\in\Gamma}\left(\int_{M\times M} d(x,y)^q \,\mathrm{d}\gamma(x,y)\right)^{1/q}, \tag{6.2}$$

where Γ stands for the couplings between η and ν, i.e., the set of probability measures on $M\times M$ with marginals ν and η, cp. [37,Def. 1.1].

To bound the Wasserstein distance, we employ the following proposition. It has been similarly stated in [32,Theorem 2], but for measures on $\mathbb{R}^d$. To fit our setting, we extend the result to compact metric spaces,[1] but emphasize that the proof closely follows that of [32,Theorem 2], and the argument is very similar. As pointed out in [32], the bound in the proposition is sharp.

Proposition 6.2 *Let (M_1,d_1) be a compact metric space, and (M_2,d_2) a Polish space, both equipped with the Borel σ-algebra. Let $T: M_1\to M_2$ and $\tilde T: M_1\to M_2$ be two continuous functions and let ν be a probability measure on M_1. Then for every $q\in[1,\infty)$*

$$W_q(T_\sharp\nu,\tilde T_\sharp\nu)\le \sup_{x\in M_1} d_2(T(x),\tilde T(x))<\infty.$$

To apply Proposition 6.2 we first have to equip U with a metric. For a sequence $(c_j)_{j\in\mathbb{N}}\in\ell^1(\mathbb{N})$ of positive numbers set

$$d(\boldsymbol{x},\boldsymbol{y}) := \sum_{j\in\mathbb{N}} c_j|x_j-y_j| \qquad \forall\, \boldsymbol{x},\boldsymbol{y}\in U. \tag{6.3}$$

By Lemma A.1, d defines a metric that induces the product topology on U. Since U with the product topology is a compact space by Tychonoff's theorem [26,Theorem 37.3], (U,d) is a compact Polish space. Moreover:

Lemma 6.3 *Let f_ρ, f_π satisfy Assumption 2.2 and consider metric (6.3) on U. Then $T: U\to U$ and the approximation $\tilde T_\varepsilon: U\to U$ from Theorem 5.2 are continuous with respect to d. Moreover, if there exists $C>0$ such that with*

$$b_j := \max\{\|\psi_{\rho,j}\|_X, \|\psi_{\pi,j}\|_Y\} \tag{6.4}$$

it holds that $b_j\le Cc_j$ for all $j\in\mathbb{N}$ (cp. Assumption 2.2), then T and $\tilde T_\varepsilon$ are Lipschitz continuous.

With $d: U\times U\to\mathbb{R}$ as in (6.3), (U,d) is a compact Polish space and T and $\tilde T_\varepsilon$ are continuous, so that we can apply Proposition 6.2. Using Theorem 5.2 and $\sup_j c_j\in(0,\infty)$,

[1] The author of [37] mentions that such a result is already known, but without providing a reference. For completeness we have added the proof.

$$W_q(T_\sharp\mu, (\tilde{T}_\varepsilon)_\sharp\mu) \leq \sup_{\boldsymbol{y}\in U} d(T(\boldsymbol{y}), \tilde{T}_\varepsilon(\boldsymbol{y})) \leq \sum_{k\in\mathbb{N}} \|T_k - \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} c_k \leq C N_\varepsilon^{-\frac{1}{p}+1}. \tag{6.5}$$

Next let us discuss why $c_j := b_j$ as in (6.4) is a natural choice in our setting. Let $\Phi : U \to Y$ be the map

$$\Phi(\boldsymbol{y}) := \sum_{j\in\mathbb{N}} y_j \psi_{\pi,j} \in Y.$$

In the inverse problem discussed in Example 2.6, we try to recover an element $\Phi(\boldsymbol{y}) \in Y$. For computational purposes, the problem is set up to recover instead the expansion coefficients $\boldsymbol{y} \in U$. Now suppose that π is the posterior measure on U. Then $\Phi_\sharp\pi = (\Phi \circ T)_\sharp\rho$ is the corresponding posterior measure on Y (the space we are actually interested in). The map $\Phi : U \to Y$ is Lipschitz continuous w.r.t. the metric d on U, since for $\boldsymbol{x}, \boldsymbol{y} \in U$ due to $\|\psi_{\pi,j}\|_Y \leq b_j$,

$$\|\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{y})\|_Y = \left\| \sum_{j\in\mathbb{N}} (x_j - y_j)\psi_{\pi,j} \right\|_Y \leq \sum_{j\in\mathbb{N}} |x_j - y_j| b_j = d(\boldsymbol{x}, \boldsymbol{y}). \tag{6.6}$$

Therefore, $\Phi \circ T : U \to Y$ and $\Phi \circ \tilde{T}_\varepsilon : U \to Y$ are Lipschitz continuous by Lemma 6.3. Moreover, compactness of U and continuity of $\Phi : U \to Y$ imply that $\Phi(T(U)) = \Phi(\tilde{T}_\varepsilon(U)) = \Phi(U) \subseteq Y$ is compact and thus separable. Hence we may apply Proposition 6.2 also to the maps $\Phi \circ T : U \to \Phi(U)$ and $\Phi \circ \tilde{T}_\varepsilon : U \to \Phi(U)$. This gives a bound on the distance between the pushforward measures on the Banach space Y. Specifically, since $\|\Phi(T(\boldsymbol{y})) - \Phi(\tilde{T}_\varepsilon(\boldsymbol{y}))\|_Y \leq d(T(\boldsymbol{y}), \tilde{T}_\varepsilon(\boldsymbol{y}))$, which can be bounded as in (6.5), we have shown:

Theorem 6.4 *Let f_ρ, f_π satisfy Assumption 2.2 for some $p \in (0, 1)$, let $\tilde{T}_\varepsilon : U \to U$ be the approximate transport and let $N_\varepsilon \in \mathbb{N}$ be the number of degrees of freedom as in Theorem 5.2.*

Then there exists $C > 0$ such that for every $q \in [1, \infty)$ and every $\varepsilon > 0$

$$W_q((\tilde{T}_\varepsilon)_\sharp\mu, \pi) \leq C N_\varepsilon^{-\frac{1}{p}+1},$$

and for the pushforward measures on the Banach space Y

$$W_q((\Phi \circ \tilde{T}_\varepsilon)_\sharp\mu, \Phi_\sharp\pi) \leq C N_\varepsilon^{-\frac{1}{p}+1}. \tag{6.7}$$

Finally let us discuss how to efficiently sample from the measure $\Phi_\sharp\pi$ on the Banach space Y. As explained in the introduction, for a sample $\boldsymbol{y} \sim \rho$ we have $T(\boldsymbol{y}) \sim \pi$ and $\Phi(T(\boldsymbol{y})) = \sum_{j\in\mathbb{N}} T_j(\boldsymbol{y}_{[j]})\psi_{\pi,j} \sim \Phi_\sharp\pi$. To truncate this series, introduce $\Phi_s(\boldsymbol{y}_{[s]}) := \sum_{j=1}^{s} y_j \psi_{\pi,j}$. As earlier, denote by ρ_s the marginal measure of ρ on U_s. For $\boldsymbol{y}_{[s]} \sim \rho_s$, the sample

$$\Phi_s(\tilde{T}_{\varepsilon,[s]}(\boldsymbol{y}_{[s]})) = \sum_{j=1}^{s} T_{\varepsilon,j}(\boldsymbol{y}_{[j]})\psi_{\pi,j}$$

follows the distribution of $(\Phi_s \circ \tilde{T}_{\varepsilon,[s]})_\sharp \rho_s$, where $\tilde{T}_{\varepsilon,[s]} := (\tilde{T}_{\varepsilon,k})_{k=1}^s : U_s \to U_s$. In the next corollary we bound the Wasserstein distance between $(\Phi_s \circ \tilde{T}_{\varepsilon,[s]})_\sharp \rho_s$ and $\Phi_\sharp \pi$. Note that the former is a measure on Y, and in contrast to the latter, is supported on an s-dimensional subspace. Thus in general neither of these two measures need to be absolutely continuous w.r.t. the other. This implies that the KL divergence, the total variation distance, and the Hellinger distance, in contrast with the Wasserstein distance, need not tend to 0 as $\varepsilon \to 0$ and $s \to \infty$.

The corollary shows that the convergence rate in (6.7) can be retained by choosing the truncation parameter s as N_ε (the number of degrees of freedom in Theorem 5.2); in fact, it even suffices to truncate after the maximal k such that $\Lambda_{\varepsilon,k} \neq \emptyset$, as explained in Remark 6.7.

Corollary 6.5 *Consider the setting of Theorem 6.4 and assume that $(b_j)_{j\in\mathbb{N}}$ in (6.4) is monotonically decreasing. Then there exists $C > 0$ such that for every $q \in [1,\infty)$ and $\varepsilon > 0$*

$$W_q((\Phi_{N_\varepsilon} \circ \tilde{T}_{\varepsilon,[N_\varepsilon]})_\sharp \rho_{N_\varepsilon}, \Phi_\sharp \pi) \le C N_\varepsilon^{-\frac{1}{p}+1}.$$

Remark 6.6 Convergence in W_q implies weak convergence [37,Theorem 6.9].

Remark 6.7 Checking the proof of Theorem 5.2, we have $N_\varepsilon \le C\varepsilon^{-p}$, cp. (C.21). Thus the maximal activated dimension (represented by the truncation parameter $s = N_\varepsilon$) increases only algebraically as $\varepsilon \to 0$. The approximation error also decreases algebraically like ε^{1-p} as $\varepsilon \to 0$, cp. (C.22). Moreover, the function $\Phi_{s_\varepsilon} \circ \tilde{T}_{\varepsilon,[s_\varepsilon]}$ with $s_\varepsilon := \max\{k \in \mathbb{N} : \Lambda_{\varepsilon,k} \neq \emptyset\}$ leads to the same convergence rate in Corollary 6.5. In other words, we only need to use the components $\tilde{T}_{\varepsilon,k}$ for which $\Lambda_{\varepsilon,k} \neq \emptyset$.

7 Conclusions

The use of transportation methods to sample from high-dimensional distributions is becoming increasingly popular to solve inference problems and perform other machine learning tasks. Therefore, questions of when and how these methods can be successful are of great importance, but thus far not well understood. In the present paper we analyze the approximation of the KR transport in the high- (or infinite-)dimensional regime and on the bounded domain $[-1,1]^{\mathbb{N}}$. Under the setting presented in Sect. 2, it is shown that the transport can be approximated without suffering from the curse of dimension. Our approximation is based on polynomial and rational functions, and we provide an explicit *a priori* construction of the ansatz space. Moreover, we show

how these results imply that it is possible to efficiently sample from certain high-dimensional distributions by transforming a lower-dimensional latent variable.

As we have discussed in the finite-dimensional case [41,Sect. 5], from an approximation viewpoint there is also a link to neural networks, which can be established via [36, 39] where it is proven that ReLU neural networks are efficient at emulating polynomials and rational functions. While we have not developed this aspect further in the present manuscript, we mention that neural networks are used in the form of normalizing flows [29, 30] to couple distributions in spaces of equal dimension, and for example in the form of generative adversarial networks [2, 14] and, more recently, injective flows [19, 21], to map lower-dimensional latent variables to samples from a high-dimensional distribution. In Sect. 6 we provided some insight (for the present setting, motivated by inverse problems in science and engineering) into how low-dimensional the latent variable can be, and how expressive the transport should be, to achieve a certain accuracy in the Wasserstein distance (see Corollary 6.5). Further examining this connection and generalizing our results to distributions on unbounded domains (such as $\mathbb{R}^{\mathbb{N}}$ instead of $[-1,1]^{\mathbb{N}}$) will be the topic of future research.

Funding Open Access funding enabled and organized by Projekt DEAL.

A Proofs of Sect. 3

A.1 Lemma 3.2

Lemma A.1 *Let $(c_j)_{j\in\mathbb{N}} \in \ell^1(\mathbb{N})$ be a sequence of positive numbers. Then $d(\boldsymbol{x},\boldsymbol{y}):=\sum_{j\in\mathbb{N}} c_j|x_j-y_j|$ defines a metric on U that induces the product topology.*

Proof Recall that the family of sets

$$\{\boldsymbol{x}\in U\,:\,|x_j-y_j|<\varepsilon\;\forall j\le N\}\qquad \boldsymbol{y}\in U,\;\varepsilon>0,\;N\in\mathbb{N},$$

forms a basis of the product topology on U. Fix $\boldsymbol{y}\in U$ and $\varepsilon>0$, and let $N_\varepsilon\in\mathbb{N}$ be so large that $\sum_{j>N_\varepsilon}2c_j<\frac{\varepsilon}{2}$. Let $C_0:=\sum_{j=1}^{N_\varepsilon}c_j$. Then if $\boldsymbol{x},\boldsymbol{y}\in U$ satisfy $|x_j-y_j|<\frac{\varepsilon}{2C_0}$ for all $j\le N_\varepsilon$, we have $d(\boldsymbol{x},\boldsymbol{y})=\sum_{j\in\mathbb{N}}c_j|x_j-y_j|<\frac{\varepsilon}{2}\frac{\sum_{j=1}^{N_\varepsilon}c_j}{C_0}+\sum_{j>N_\varepsilon}2c_j\le\varepsilon$, and thus

$$\left\{\boldsymbol{x}\in U\,:\,|x_j-y_j|<\frac{\varepsilon}{2C_0}\;\forall j\le N_\varepsilon\right\}\subseteq\{\boldsymbol{x}\in U\,:\,d(\boldsymbol{x},\boldsymbol{y})<\varepsilon\}.$$

On the other hand, if we fix $\boldsymbol{y} \in U$, $\varepsilon > 0$ and $N \in \mathbb{N}$, and set $C_0 := \min_{j=1,\dots,N} c_j > 0$, then

$$\{\boldsymbol{x} \in U : d(\boldsymbol{x}, \boldsymbol{y}) < \varepsilon C_0\} = \left\{\boldsymbol{x} \in U : \sum_{j\in\mathbb{N}} c_j |x_j - y_j| < \varepsilon C_0\right\} \subseteq \{\boldsymbol{x} \in U : |x_j - y_j| < \varepsilon \ \forall j \le N\}.$$

□

Proof of Lemma 3.2 By [3,Lemma 6.4.2 (ii)], the Borel σ-algebra on U (with the product topology) coincides with the product σ-algebra on U. Since $f : U \to \mathbb{R}$ is continuous, and because U and $\mathbb{R}$ are equipped with the Borel σ-algebras, $f : U \to \mathbb{R}$ is measurable. Since f is bounded it belongs to $L^2(U, \mu)$.

Fix $(c_j)_{j\in\mathbb{N}} \in \ell^1(\mathbb{N})$ with $c_j > 0$ for all $j \in \mathbb{N}$, and let d be the metric on U from Lemma A.1. Since $f \in C^0(U, \mathbb{R}_+)$ and U is compact by Tychonoff's theorem [26,Theorem 37.3], the Heine-Cantor theorem yields f to be uniformly continuous. Thus for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that for all $\boldsymbol{x}, \boldsymbol{y} \in U$ with $d(\boldsymbol{x}, \boldsymbol{y}) < \delta_\varepsilon$ it holds $|f(\boldsymbol{x}) - f(\boldsymbol{y})| < \varepsilon$. Now let $k \in \mathbb{N}$ and $\varepsilon > 0$ arbitrary. Then for all $\boldsymbol{x}_{[k]}$, $\boldsymbol{y}_{[k]} \in U_k$ such that $\sum_{j=1}^k c_j |x_j - y_j| < \delta_\varepsilon$, we get

$$|\hat{f}_k(\boldsymbol{x}_{[k]}) - \hat{f}_k(\boldsymbol{y}_{[k]})| = \left| \int_U f(\boldsymbol{x}_{[k]}, \boldsymbol{t}) - f(\boldsymbol{y}_{[k]}, \boldsymbol{t}) \, \mathrm{d}\mu(\boldsymbol{t}) \right| \le \int_U |f(\boldsymbol{x}_{[k]}, \boldsymbol{t}) - f(\boldsymbol{y}_{[k]}, \boldsymbol{t})| \, \mathrm{d}\mu(\boldsymbol{t}) \le \varepsilon,$$

which shows continuity of $\hat{f}_k : U_k \to \mathbb{R}$.

Next, using that $\inf_{\boldsymbol{y}\in U} f(\boldsymbol{y}) =: r > 0$ (due to compactness of U and continuity of f), for $k > 1$ we have $\hat{f}_{k-1}(\boldsymbol{x}_{[k-1]}) \ge \min\{r, 1\} > 0$ independent of $\boldsymbol{x}_{[k-1]} \in U_{k-1}$. This implies that also $\frac{\hat{f}_k}{\hat{f}_{k-1}} = f_k : U_k \to \mathbb{R}_+$ is continuous, where the case $k = 1$ is trivial since $\hat{f}_0 \equiv 1$.

Finally we show (3.5). Let again $\varepsilon > 0$ be arbitrary and $N_\varepsilon \in \mathbb{N}$ so large that $\sum_{j>N_\varepsilon} 2c_j < \delta_\varepsilon$. Then for every $\boldsymbol{x}, \boldsymbol{t} \in U$ and every $k > N_\varepsilon$ we have $d((\boldsymbol{x}_{[k]}, \boldsymbol{t}), \boldsymbol{x}) \le \sum_{j>N_\varepsilon} c_j |x_j - t_j| \le \sum_{j>N_\varepsilon} 2c_j < \delta_\varepsilon$, which implies $|f(\boldsymbol{x}_{[k]}, \boldsymbol{t}) - f(\boldsymbol{x})| < \varepsilon$. Thus for every $\boldsymbol{x} \in U$ and every $k > N_\varepsilon$

$$|\hat{f}_k(\boldsymbol{x}_{[k]}) - f(\boldsymbol{x})| = \left| \int_U f(\boldsymbol{x}_{[k]}, \boldsymbol{t}) \, \mathrm{d}\mu(\boldsymbol{t}) - f(\boldsymbol{x}) \right| \le \int_U |f(\boldsymbol{x}_{[k]}, \boldsymbol{t}) - f(\boldsymbol{x})| \, \mathrm{d}\mu(\boldsymbol{t}) < \varepsilon,$$

which concludes the proof. □

A.2 Theorem 3.3

With $F_{*;k}(\boldsymbol{x}_{[k-1]}, x_k) := \int_{-1}^{x_k} f_{*;k}(\boldsymbol{x}_{[k-1]}, t_k)\,\mathrm{d}t_k$, the construction of $T_k : U_k \to U_k$ described in Sect. 3 amounts to the explicit formula $T_1(x_1) := (F_{\pi;1})^{-1} \circ F_{\rho;1}(x_1)$ and inductively

$$T_k(\boldsymbol{x}_{[k-1]}, \cdot) := F_{\pi;k}(T_{[k-1]}(\boldsymbol{x}_{[k-1]}), \cdot)^{-1} \circ F_{\rho;k}(\boldsymbol{x}_{[k-1]}, \cdot), \tag{A.1}$$

where $F_{\pi;k}(T_{[k-1]}(\boldsymbol{x}_{[k-1]}), \cdot)$ denotes the inverse of $x_k \mapsto F_{\pi;k}(T_{[k-1]}(\boldsymbol{x}_{[k-1]}), x_k)$.

Remark A.2 If $f_{*;k} \in C^0(U_k, \mathbb{R}_+)$ for $* \in \{\rho, \pi\}$, then by (A.1) it holds $T_k, \partial_k T_k \in C^0(U_k)$.

Proof of Theorem 3.3 We start with (i). As a consequence of Remark A.2 and Lemma 3.2, $T_k \in C^0(U_k, U_1)$ for every $k \in \mathbb{N}$. So each $T_k : U_k \to U_1$ is measurable and thus also $T_{[n]} = (T_k)_{k=1}^n : U_n \to U_n$ is measurable for each $n \in \mathbb{N}$. Furthermore $T : U \to U$ is bijective by Lemma 3.1 and because for every $\boldsymbol{x} \in U$ and $k \in \mathbb{N}$ it holds that $T_k(\boldsymbol{x}_{[k-1]}, \cdot) : U_1 \to U_1$ is bijective.

The product σ-algebra on U is generated by the algebra (see [3,Def. 1.2.1]) $\mathcal{A}_0$ given as the union of the σ-algebras $\mathcal{A}_n := \{A_n \times [-1, 1]^{\mathbb{N}} : A_n \in \mathcal{B}(U_n)\}$, $n \in \mathbb{N}$, where $\mathcal{B}(U_n)$ denotes the Borel σ-algebra. For sets of the type $A := A_n \times [-1, 1]^{\mathbb{N}} \in \mathcal{A}_n$ with $A_n \in \mathcal{B}(U_n)$, due to $T_j(\boldsymbol{y}_{[j]}) \in U_1$ for all $\boldsymbol{y} \in U$ and $j > n$, we have

$$\begin{aligned} T^{-1}(A) = \{\boldsymbol{y} \in U : T(\boldsymbol{y}) \in A\} &= \{\boldsymbol{y} \in U : T_{[n]}(\boldsymbol{y}_{[n]}) \in A_n\} \\ &= (T_{[n]})^{-1}(A_n) \times [-1, 1]^{\mathbb{N}}, \end{aligned}$$

which belongs to $\mathcal{A}_n$ and thus to the product σ-algebra on U since $T_{[n]}$ is measurable. Hence $T : U \to U$ is measurable w.r.t. the product σ-algebra.

Denote now by π_n and ρ_n the marginals on U_n w.r.t. the first n variables, i.e., e.g., $\pi_n(A) := \pi(A \times [-1, 1]^{\mathbb{N}})$ for every $A \in \mathcal{B}(U_n)$. By (3.4) (see [33,Proposition 2.18]), $(T_{[n]})_\sharp \rho_n = \pi_n$. For sets of the type $A := A_n \times [-1, 1]^{\mathbb{N}} \in \mathcal{A}_n$ with $A_n \in \mathcal{B}(U_n)$,

$$\begin{aligned} T_\sharp \rho(A) = \rho(\{\boldsymbol{y} \in U : T(\boldsymbol{y}) \in A\}) &= \rho(\{\boldsymbol{y} \in U : T_{[n]}(\boldsymbol{y}_{[n]}) \in A_n\}) \\ &= \rho_n(\{\boldsymbol{y} \in U_n : T_{[n]}(\boldsymbol{y}) \in A_n\}) \\ &= \pi_n(A_n) \\ &= \pi(A). \end{aligned}$$

According to [3,Theorem 3.5.1], the extension of a non-negative σ-additive set function on the algebra $\mathcal{A}_0$ to the σ-algebra generated by $\mathcal{A}_0$ is unique. Since $T : U \to U$ is bijective and measurable, it holds that both π and $T_\sharp \rho$ are measures on U and therefore $\pi = T_\sharp \rho$.

Finally we show (ii). Let $\hat{f}_{\pi,n} \in C^0(U_n, \mathbb{R}_+)$ and $\hat{f}_{\rho,n} \in C^0(U_n, \mathbb{R}_+)$ be as in (3.2), i.e., these functions denote the densities of the marginals π_n, ρ_n. Since $(T_{[n]})_\sharp \rho_n = \pi_n$, by a change of variables (see, e.g., [4,Proposition 2.5]), for all $\boldsymbol{x} \in U$

$$\hat{f}_{\rho,n}(\boldsymbol{x}_{[n]}) = \hat{f}_{\pi,n}(T_{[n]}(\boldsymbol{x}_{[n]})) \det dT_{[n]}(\boldsymbol{x}_{[n]}) = \hat{f}_{\pi,n}(T_{[n]}(\boldsymbol{x}_{[n]})) \prod_{j=1}^{n} \partial_j T_j(\boldsymbol{x}_{[j]}).$$

Therefore

$$\prod_{j=1}^{n} \partial_j T_j(\boldsymbol{x}_{[j]}) = \frac{\hat{f}_{\rho,n}(\boldsymbol{x}_{[n]})}{\hat{f}_{\pi,n}(T_{[n]}(\boldsymbol{x}_{[n]}))}. \tag{A.2}$$

According to Lemma 3.2 we have uniform convergence

$$\lim_{n\to\infty} \hat{f}_{\rho,n}(\boldsymbol{x}_{[n]}) = f_\rho(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in U$$

and uniform convergence of

$$\lim_{n\to\infty} \hat{f}_{\pi,n}(\boldsymbol{y}_{[n]}) = f_\pi(\boldsymbol{y}) \qquad \forall \boldsymbol{y} \in U.$$

The latter implies with $\boldsymbol{y} = T(\boldsymbol{x})$ that

$$\lim_{n\to\infty} \hat{f}_{\pi,n}(T_{[n]}(\boldsymbol{x}_{[n]})) = f_\pi(T(\boldsymbol{x})) \qquad \forall \boldsymbol{x} \in U$$

converges uniformly. Since $f_\pi : U \to \mathbb{R}_+$ is continuous and U is compact, we can conclude that $\hat{f}_{\pi,n}(\boldsymbol{x}) \geq r$ (cp. (3.2)) for some $r > 0$ independent of $n \in \mathbb{N}$ and $\boldsymbol{x} \in U_n$. Thus the right-hand side of (A.2) converges uniformly, and

$$\det dT(\boldsymbol{x}) := \lim_{n\to\infty} \prod_{j=1}^{n} \partial_j T_j(\boldsymbol{x}_{[j]}) = \frac{f_\rho(\boldsymbol{x})}{f_\pi(T(\boldsymbol{x}))} \in C^0(U, \mathbb{R}_+)$$

converges uniformly. Moreover $\det dT(\boldsymbol{x}) f_\pi(T(\boldsymbol{x})) = f_\rho(\boldsymbol{x})$ for all $\boldsymbol{x} \in U$. □

B Proofs of Sect. 4

B.1 Proposition 4.2

The proposition is a consequence of the finite-dimensional result shown in [41]. For better readability, we recall the statement here together with its requirements; see [41,Assumption 3.5, Theorem 3.6]:

Assumption B.1 Let $0 < \hat{M} \leq \hat{L}$, $\hat{C}_1 > 0$, $k \in \mathbb{N}$ and $\boldsymbol{\delta} \in (0,\infty)^k$ be given. For $* \in \{\rho, \pi\}$:

(a) $\hat{f}_* : U_k \to \mathbb{R}_+$ is a probability density and $\hat{f}_* \in C^1(\mathcal{B}_{\boldsymbol{\delta}}(U_1), \mathbb{C})$,
(b) $\hat{M} \leq |\hat{f}_*(\boldsymbol{x})| \leq \hat{L}$ for $\boldsymbol{x} \in \mathcal{B}_{\boldsymbol{\delta}}(U_1)$,

(c) $\sup_{\boldsymbol{y}\in\mathcal{B}_{\boldsymbol{\delta}}} |\hat{f}_*(\boldsymbol{x}+\boldsymbol{y}) - \hat{f}_*(\boldsymbol{x})| \le \hat{C}_1$ for $\boldsymbol{x} \in U_k$,

(d) $\sup_{\boldsymbol{y}\in\mathcal{B}_{\boldsymbol{\delta}_{[j]}}\times\{0\}^{k-j}} |\hat{f}_*(\boldsymbol{x}+\boldsymbol{y}) - \hat{f}_*(\boldsymbol{x})| \le \hat{C}_1\delta_{k+1}$ for $\boldsymbol{x} \in U_k$ and $j \in \{1,\dots,k-1\}$.

Theorem B.2 *Let $0 < \hat{M} \le \hat{L} < \infty$, $k \in \mathbb{N}$ and $\boldsymbol{\delta} \in (0,\infty)^k$. There exist $\hat{C}_1$, $\hat{C}_2$ and $\hat{C}_3 > 0$ depending on $\hat{M}$ and $\hat{L}$ (but not on k or $\boldsymbol{\delta}$) such that if Assumption B.1 holds with $\hat{C}_1$, then:*

Let $H : U_k \to U_k$ be the KR-transport as in (3.3) such that H pushes forward the measure with density $\hat{f}_\rho$ to the one with density $\hat{f}_\pi$. Set $R_k := \partial_k H_k$. With $\boldsymbol{\zeta} = (\zeta_j)_{j=1}^k$ where $\zeta_j := \hat{C}_2\delta_j$, it holds for all $j \in \{1,\dots,k\}$:

(i) $R_j \in C^1(\mathcal{B}_{\boldsymbol{\zeta}_{[j]}}(U_1), \mathcal{B}_{\hat{C}_3}(1))$ and $\Re(R_j(\boldsymbol{x})) \ge \frac{1}{\hat{C}_3}$ for all $\boldsymbol{x} \in \mathcal{B}_{\boldsymbol{\zeta}_{[j]}}(U_1)$,

(ii) if $j \ge 2$, $R_j : \mathcal{B}_{\boldsymbol{\zeta}_{[j-1]}}(U_1) \times U_1 \to \mathcal{B}_{\frac{\hat{C}_3}{\max\{1,\delta_j\}}}(1)$.

Proof of Proposition 4.2 For $* \in \{\rho,\pi\}$ and $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}}(U_1) \subseteq \mathbb{C}^k$ let

$$\hat{f}_{*,k}(\boldsymbol{z}) := \int_U f_*(\boldsymbol{z},\boldsymbol{y})\, \mathrm{d}\mu(\boldsymbol{y})$$

be the extension of (3.2) to complex numbers. By Lemma 3.2, $\hat{f}_{*,k} \in C^0(U_k)$. Moreover with ρ_k and π_k being the marginal measures on U_k in the first k variables, by definition $\hat{f}_{\rho,k} = \frac{\mathrm{d}\rho_k}{\mathrm{d}\mu}$ and $\hat{f}_{\pi,k} = \frac{\mathrm{d}\pi_k}{\mathrm{d}\mu}$. In other words, these functions are the respective marginal densities in the first k variables.

Let $H : U_k \to U_k$ be the KR-transport satisfying $H_\sharp\rho_k = \pi_k$, and let $T : U \to U$ be the KR-transport satisfying $T_\sharp\rho = \pi$. By construction (cp. (3.3)) and uniqueness of the KR-transport, it holds $T_{[k]} = (T_j)_{j=1}^k = (H_j)_{j=1}^k = H$. In order to complete the proof, we will apply Theorem B.2 to H. To this end we need to check Assumption B.1 for the densities $\hat{f}_{\rho,k}, \hat{f}_{\pi,k} : U_k \to \mathbb{R}$. We will do so with the constants

$$\hat{M} := \frac{M}{2}, \qquad \hat{L} := L + \frac{M}{2}, \qquad C_1(M,L) := \min\left\{\frac{M}{2}, \hat{C}_1(\hat{M},\hat{L})\right\}, \qquad \text{(B.1)}$$

where $\hat{C}_1(\hat{M},\hat{L})$ is as in Theorem B.2. Assume for the moment that $\hat{f}_{\rho,k}, \hat{f}_{\pi,k} : U_k \to \mathbb{R}$ satisfy Assumption B.1 with $\hat{M}$ and $\hat{L}$. Then Theorem B.2 immediately implies the statement of Proposition 4.2 with $C_2(M,L) := \hat{C}_2(\hat{M},\hat{L})$ and $C_3(M,L) := \hat{C}_3(\hat{M},\hat{L})$, where $\hat{C}_2$ and $\hat{C}_3$ are as in Theorem B.2.

It remains to verify Assumption B.1. We do so item by item and fix $* \in \{\rho,\pi\}$:

(a) By Lemma 3.2, $\hat{f}_{*,k} \in C^0(U_k)$ and $\int_{U_k} \hat{f}_{*,k}(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \int_U f_*(\boldsymbol{y})\,\mathrm{d}\mu(\boldsymbol{y}) = 1$, so that $\hat{f}_{*,k}$ is a positive probability density on U_k.
Fix $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}}(U_1) \subseteq \mathbb{C}^k$ and $i \in \{1,\dots,k\}$. We want to show that $z_i \mapsto \hat{f}_{*,k}(\boldsymbol{z}) \in \mathbb{C}$ is complex differentiable for $z_i \in \mathcal{B}_{\delta_i}(U_1)$. It holds:

- By Assumption 4.1 (a), $\boldsymbol{y} \mapsto f_*(\boldsymbol{z},\boldsymbol{y}) : U \to \mathbb{C}$ is continuous and therefore measurable for all $z_i \in \mathcal{B}_{\delta_i}(U_1)$.
- By Assumption 4.1 (b), for every fixed $\boldsymbol{y} \in U$, $z_i \mapsto f_*(\boldsymbol{z},\boldsymbol{y}) : \mathcal{B}_{\delta_i}(U_1) \to \mathbb{C}$ is differentiable.

- By Assumption 4.1 (a) $f_* : \mathcal{B}_{\boldsymbol{\delta}}(U_1) \times U \to \mathbb{C}$ is continuous. Thus, compactness of $\bar{\mathcal{B}}_r(z_i) \times U$ (w.r.t. the product topology), implies that for every $z_i \in \mathcal{B}_{\delta_i}(U_1)$ with $r > 0$ s.t. $\bar{\mathcal{B}}_r(z_i) \subseteq \mathcal{B}_{\delta_i}(U_1)$, holds $\sup_{x\in\mathcal{B}_r(z_i)} \sup_{\boldsymbol{y}\in U} |f_*(\boldsymbol{z}_{[i-1]}, x, \boldsymbol{z}_{[i+1:k]}, \boldsymbol{y})| < \infty$. Hence

$$z_i \in \mathcal{B}_{\delta_i}(U_1) \Rightarrow \exists r > 0 : \sup_{x\in\mathcal{B}_r(z_i)} \int_{\boldsymbol{y}\in U} |f_*(\boldsymbol{z}_{[i-1]}, x, \boldsymbol{z}_{[i+1:k]}, \boldsymbol{y})| \,\mathrm{d}\mu(\boldsymbol{y}) < \infty.$$

According to the main theorem in [24], this implies $z_i \mapsto \hat{f}_{*,k}(\boldsymbol{z}) = \int_U f_*(\boldsymbol{z}, \boldsymbol{y}) \,\mathrm{d}\mu(\boldsymbol{y})$ to be differentiable on $\mathcal{B}_{\delta_i}(U_1)$. Since $i \in \{1, \ldots, k\}$ was arbitrary, Hartog's theorem, e.g., [20,Theorem 1.2.5], yields $\boldsymbol{z} \mapsto \hat{f}_{*,k}(\boldsymbol{z}) : \mathcal{B}_{\boldsymbol{\delta}}(U_1) \to \mathbb{C}$ to be differentiable.

(b) By Assumption 4.1 (c), and because $f_*(\boldsymbol{y}) \in \mathbb{R}_+$ for $\boldsymbol{y} \in U$, we have $M \le f_*(\boldsymbol{y}) \le L$ for all $\boldsymbol{y} \in U$. Thus $\hat{f}_{*,k}(\boldsymbol{x}) = \int_U f(\boldsymbol{x}, \boldsymbol{y}) \,\mathrm{d}\mu(\boldsymbol{y}) \ge M$ and also $\hat{f}_{*,k}(\boldsymbol{x}) \le L$ for all $\boldsymbol{x} \in U_k$. Furthermore, for $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}} \subseteq \mathbb{C}^k$ and $\boldsymbol{x} \in U_k$, by Assumption 4.1 (d) and (B.1)

$$|\hat{f}_{*,k}(\boldsymbol{x}+\boldsymbol{z}) - \hat{f}_{*,k}(\boldsymbol{x})| \le \int_U |f_*(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}) - f_*(\boldsymbol{x}, \boldsymbol{y})| \,\mathrm{d}\mu(\boldsymbol{y}) \le C_1 \le \frac{M}{2}.$$

Thus, with $\hat{M} = \frac{M}{2} > 0$ and $\hat{L} = L + \frac{\hat{M}}{2}$ we have $\hat{M} \le |\hat{f}_{*,k}(\boldsymbol{z})| \le \hat{L}$ for all $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}}(U_1)$.

(c) For $\boldsymbol{x} \in U_k$ by Assumption 4.1 (d) and (B.1)

$$\begin{aligned}
&\sup_{\boldsymbol{z}\in\mathcal{B}_{\boldsymbol{\delta}}} |\hat{f}_{*,k}(\boldsymbol{x}+\boldsymbol{z}) - \hat{f}_{*,k}(\boldsymbol{x})| \\
&\quad\le \sup_{\boldsymbol{y}\in\mathcal{B}_{\boldsymbol{\delta}}} \int_U |f_*(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}) - f_*(\boldsymbol{x}, \boldsymbol{y})| \,\mathrm{d}\mu(\boldsymbol{y}) \le C_1(M, L) \\
&\quad\le \hat{C}_1(\hat{M}, \hat{L}).
\end{aligned}$$

(d) For $\boldsymbol{x} \in U_k$ and $j \in \{1, \ldots, k-1\}$ by Assumption 4.1 (e) and (B.1)

$$\begin{aligned}
&\sup_{\boldsymbol{z}\in\mathcal{B}_{\boldsymbol{\delta}_{[j]}}\times\{0\}^{k-j}} |\hat{f}_{*,k}(\boldsymbol{x}+\boldsymbol{z}) - \hat{f}_{*,k}(\boldsymbol{x})| \\
&\quad\le \sup_{\boldsymbol{z}\in\mathcal{B}_{\boldsymbol{\delta}_{[j]}}\times\{0\}^{\mathbb{N}}} \int_U |f_*(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}) - f_*(\boldsymbol{x}, \boldsymbol{y})| \,\mathrm{d}\mu(\boldsymbol{y}) \\
&\quad\le C_1(M, L)\delta_{j+1} \le \hat{C}_1(\hat{M}, \hat{L})\delta_{j+1}.
\end{aligned}$$

□

B.2 Verifying Assumption 4.1

In this section we show that densities as in Assumption 2.1 satisfy Assumption 4.1.

Lemma B.3 *Let $f(\boldsymbol{y}) = \mathfrak{f}(\sum_{j\in\mathbb{N}} y_j\psi_j)$ satisfy Assumption 2.1 for some $p \in (0, 1)$ and $0 < M \le L < \infty$. Let $(b_j)_{j\in\mathbb{N}} \subset (0, \infty)$ be summable and such that $b_j \ge \|\psi_j\|_Z$ for all $j \in \mathbb{N}$. Let $C_1 = C_1(M, L) > 0$ be as in Proposition 4.2.*

There exists a monotonically increasing sequence $(\kappa_j)_{j\in\mathbb{N}} \in (0,\infty)^{\mathbb{N}}$ and $\tau > 0$ (depending on $(b_j)_{j\in\mathbb{N}}$, C_1 and $\mathfrak{f}$) such that for every fixed $J \in \mathbb{N}$, $k \in \mathbb{N}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^k$, with

$$\delta_j = \delta_j(J, \boldsymbol{\nu}) := \kappa_j + \begin{cases} 0 & j < J,\ j \ne k \\ \frac{\tau\nu_j}{(\sum_{i=J}^{k-1}\nu_i)b_j} & j \ge J,\ j \ne k \\ \frac{\tau}{b_j} & j = k \end{cases} \qquad \forall j \in \{1, \ldots, k\}, \tag{B.2}$$

f satisfies Assumption 4.1.

Lemma B.4 *Let $\boldsymbol{b} = (b_j)_{j\in\mathbb{N}} \in \ell^1(\mathbb{N})$ with $b_j \ge 0$ for all j, and let $\gamma > 0$. There exists $(\kappa_j)_{j\in\mathbb{N}} \subset (0,\infty)$ monotonically increasing and such that $\kappa_j \to \infty$ and $\sum_{j\in\mathbb{N}} b_j\kappa_j < \gamma$.*

Proof If there exists $d \in \mathbb{N}$ such that $b_j = 0$ for all $j > d$, then the statement is trivial. Otherwise, for $n \in \mathbb{N}$ set $j_n := \min\{j \in \mathbb{N} : \sum_{i\ge j} b_i \le 2^{-n}\}$. Since $\boldsymbol{b} \in \ell^1(\mathbb{N})$, $(j_n)_{n\in\mathbb{N}}$ is well-defined, monotonically increasing, and tends to infinity (it may have repeated entries). For $j \in \mathbb{N}$ let

$$\tilde{\kappa}_j := \begin{cases} 1 & \text{if } j < j_1 \\ n & \text{if } j \in \mathbb{N} \cap [j_n, j_{n+1}), \end{cases}$$

which is well-defined since $j_n \to \infty$ so that

$$\mathbb{N} = \{1, \ldots, j_1\} \cup \bigcup_{n\in\mathbb{N}} (\mathbb{N} \cap [j_n, j_{n+1}))$$

and those sets are disjoint, in particular if $j_n = j_{n+1}$ then $[j_n, j_{n+1}) \cap \mathbb{N} = \emptyset$. Then

$$\begin{aligned} \sum_{j\in\mathbb{N}} b_j\tilde{\kappa}_j &= \sum_{j=1}^{j_1-1} b_j + \sum_{j\ge j_1} b_j\tilde{\kappa}_j \\ &= \sum_{j=1}^{j_1-1} b_j + \sum_{n\in\mathbb{N}} \sum_{j=j_n}^{j_{n+1}-1} b_j\tilde{\kappa}_j \le \sum_{j=1}^{j_1-1} b_j + \sum_{n\in\mathbb{N}} n2^{-n} < \infty. \end{aligned}$$

Set $\kappa_j := \frac{\gamma\tilde{\kappa}_j}{\sum_{j\in\mathbb{N}} b_j\tilde{\kappa}_j}$. □

Proof of Lemma B.3 In Steps 1–2 we will construct $(\kappa_j)_{j\in\mathbb{N}} \subset (0,\infty)$ and $\tau > 0$ independent of $J \in \mathbb{N}$, $k \in \mathbb{N}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^k$. In Steps 3–4, we verify that $(\kappa_j)_{j\in\mathbb{N}}$ and τ have the desired properties.

Moreover, we will use that Z is a Banach space, $Z_{\mathbb{C}}$ its complexification as introduced in (and before) Assumption 2.1, and $\psi_j \in Z \subseteq Z_{\mathbb{C}}$ for all j.

Step 1. Set $K:=\{\sum_{j\in\mathbb{N}} y_j\psi_j \,:\, \boldsymbol{y} \in U\} \subseteq Z$. According to [40,Remark 2.1.3], $\boldsymbol{y} \mapsto \sum_{j\in\mathbb{N}} y_j\psi_j : U \to Z$ is continuous and $K \subseteq Z$ is compact (as the image of a compact set under a continuous map). Compactness of K and continuity of $\mathfrak{f}$ imply $\sup_{\psi\in K} |\mathfrak{f}(\psi)| < \infty$ and

$$\lim_{\varepsilon\to 0} \sup_{\|\psi\|_Z<\varepsilon} \sup_{\phi\in K} |\mathfrak{f}(\psi+\phi) - \mathfrak{f}(\phi)| = 0. \tag{B.3}$$

Hence there exists $r > 0$ such that with $O_Z \subseteq Z_{\mathbb{C}}$ from Assumption 2.1

$$\{\phi + \psi \,:\, \phi \in K,\ \|\psi\|_{Z_{\mathbb{C}}} < r\} \subseteq O_Z \tag{B.4}$$

and

$$C_{\mathfrak{f}} := \sup_{\|\psi\|_Z<r} \sup_{\phi\in K} |\mathfrak{f}(\phi+\psi)| < \infty \tag{B.5}$$

and

$$\sup_{\|\psi\|_Z<r} \sup_{\phi\in K} |\mathfrak{f}(\psi+\phi) - \mathfrak{f}(\phi)| < C_1. \tag{B.6}$$

Step 2. We show the existence of $\boldsymbol{\kappa} = (\kappa_j)_{j\in\mathbb{N}} \subset (0,\infty)$ monotonically increasing, and $\tau > 0$ such that with $r > 0$ from Step 1

$$\sum_{j\in\mathbb{N}} \kappa_j b_j + 2\tau < r, \tag{B.7}$$

and additionally for every $j \in \mathbb{N}$ with $K \subseteq Z$ from Step 1

$$\sup_{\boldsymbol{z}\in\mathcal{B}_{\boldsymbol{\kappa}_{[j]}}\times\{0\}^{\mathbb{N}}} \sup_{\|\psi\|_{Z_{\mathbb{C}}}<2\tau} \sup_{\phi\in K} \left| \mathfrak{f}\left(\phi + \psi + \sum_{j\in\mathbb{N}} z_j\psi_j\right) - \mathfrak{f}(\phi) \right| \le C_1\kappa_{j+1}. \tag{B.8}$$

Let $(\tilde{\kappa}_j)_{j\in\mathbb{N}} \to \infty$ be as in Lemma B.4 such that $\sum_{j\in\mathbb{N}} \tilde{\kappa}_j b_j < \frac{r}{3}$ and with $\tilde{\tau} := \frac{r}{3}$ it holds

$$\sum_{j\in\mathbb{N}} \tilde{\kappa}_j b_j + 2\tilde{\tau} < r.$$

Since $\tilde{\kappa}_j \to \infty$ as $j \to \infty$, there exists $d \in \mathbb{N}$ such that $C_1\tilde{\kappa}_{j+1} \ge 2C_{\mathfrak{f}}$ for all $j \ge d$ (with $C_{\mathfrak{f}}$ as in (B.5)). For all $\boldsymbol{z} \in \mathcal{B}_{\tilde{\boldsymbol{\kappa}}} \subseteq \mathbb{C}^{\mathbb{N}}$ using $\|\psi_j\|_Z \le b_j$

$$\sup_{\|\psi\|_{Z_{\mathbb{C}}}<2\tilde{\tau}} \left\| \psi + \sum_{j\in\mathbb{N}} z_j\psi_j \right\|_{Z_{\mathbb{C}}} \le 2\tilde{\tau} + \sum_{j\in\mathbb{N}} \tilde{\kappa}_j \|\psi_j\|_Z \le 2\tilde{\tau} + \sum_{j\in\mathbb{N}} \tilde{\kappa}_j b_j \le r$$

and thus by (B.5) for $\phi \in K$ and $\|\psi\|_{Z_{\mathbb{C}}} < 2\tilde{\tau}$

$$\left|\mathfrak{f}\left(\phi+\psi+\sum_{j\in\mathbb{N}} z_j\psi_j\right)-\mathfrak{f}(\phi)\right| \le 2C_{\mathfrak{f}} \le C_1\tilde{\kappa}_{j+1} \tag{B.9}$$

for all $j \ge d$. Hence (B.8) holds for $\tilde{\boldsymbol{\kappa}}$ for all $j \ge d$.

To finish the construction of $\boldsymbol{\kappa}$, first define $\kappa_j := \tilde{\kappa}_j$ for all $j \ge d$. For $k \in \{1, \dots, d-1\}$, inductively (starting with $k = d-1$ and going backward) let $\tilde{\tau}_k > 0$ and $\kappa_k \in (0, \kappa_{k+1})$ be so small that

$$\sup_{\substack{|z_j|\le\kappa_k \\ \forall j\le k}} \sup_{\|\psi\|_Z<2\tilde{\tau}_k} \sup_{\phi\in K} \left|\mathfrak{f}\left(\phi+\psi+\sum_{j=1}^{k} z_j\psi_j\right)-\mathfrak{f}(\phi)\right| \le C_1\kappa_{k+1} \tag{B.10}$$

which is possible due to (B.3) and because $C_1\kappa_{k+1} > 0$. Letting $\tau := \min\{\tilde{\tau}, \tilde{\tau}_1, \dots, \tilde{\tau}_d\}$, it now follows by (B.9) and (B.10) that (B.8) holds for all $j \in \mathbb{N}$.

Step 3. We verify Assumption 4.1 (a), (b) and (c). Fix $J \in \mathbb{N}, k \in \mathbb{N}$ and $\mathbf{0} \ne \boldsymbol{\nu} \in \mathbb{N}_0^k$. By definition of $\boldsymbol{\delta} \in \mathbb{R}^k$ in (B.2) and with $b_j \ge \|\psi_j\|_Z$

$$\begin{aligned}\sup_{z\in\mathcal{B}_{\boldsymbol{\delta}}\times\{0\}^{\mathbb{N}}} \sup_{\boldsymbol{y}\in U} \left\|\sum_{j\in\mathbb{N}}(y_j+z_j)\psi_j\right\|_Z &\le \sum_{j\in\mathbb{N}} b_j(1+\kappa_j) + \tau\frac{b_k}{b_k} + \sum_{j=J}^{k-1}\frac{\tau\nu_j b_j}{(\sum_{i=J}^{k-1}\nu_i)b_j}\\ &\le \sum_{j\in\mathbb{N}} b_j + \sum_{j\in\mathbb{N}}\kappa_j b_j + 2\tau < \infty\end{aligned}$$

and, similarly, by (B.7)

$$\sup_{z\in\mathcal{B}_{\boldsymbol{\delta}}\times\{0\}^{\mathbb{N}}} \sup_{\boldsymbol{y}\in U} \left\|\sum_{j\in\mathbb{N}}(y_j+z_j)\psi_j - \sum_{j\in\mathbb{N}} y_j\psi_j\right\|_Z \le \sum_{j\in\mathbb{N}}\kappa_j b_j + 2\tau < r. \tag{B.11}$$

Thus by (B.4), $f(\boldsymbol{y}+\boldsymbol{z}) = \mathfrak{f}(\sum_{j\in\mathbb{N}}(y_j+z_j)\psi_j)$ is well-defined for all $\boldsymbol{y} \in U$, $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}} \times \{0\}^{\mathbb{N}}$, since then $\sum_{j\in\mathbb{N}}(y_j+z_j)\psi_j \in O_Z$, where $O_Z \subseteq Z_{\mathbb{C}}$ is the domain of definition of $\mathfrak{f}$. Summability of $(\|\psi_j\|_Z)_{j\in\mathbb{N}}$ implies continuity of $f : \mathcal{B}_{\boldsymbol{\delta}}(U_1) \times U \to O_Z$ w.r.t. the product topology on $\mathcal{B}_{\boldsymbol{\delta}}(U_1) \times U \subseteq \mathbb{C}^{\mathbb{N}}$ (see, e.g., [40,Remark 2.1.3]). Continuity of $\mathfrak{f} : O_Z \to \mathbb{C}$ thus implies $f \in C^0(\mathcal{B}_{\boldsymbol{\delta}} \times \{0\}^{\mathbb{N}}, \mathbb{C})$, i.e., Assumption 4.1 (a) holds.

Differentiability of $\mathfrak{f}$ implies that $f(\boldsymbol{z}) = \mathfrak{f}(\sum_{j\in\mathbb{N}} z_j\psi_j)$ is differentiable in each z_j for $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}}(U_1) \times U$, proving Assumption 4.1 (b). Finally Assumption 4.1 (c) is a consequence of Assumption 2.1 (d).

Step 4. We show Assumption 4.1 (d) and (e). Fix $k \in \mathbb{N}$, $\mathbf{0} \neq \boldsymbol{\nu} \in \mathbb{N}_0^k$ and $J \in \mathbb{N}$. Then for any $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}} \subseteq \mathbb{C}^k$, by (B.2) we can write $z_i = z_{i,1} + z_{i,2}$ with

$$|z_{i,1}| \leq \kappa_i, \qquad |z_{i,2}| \leq \begin{cases} 0 & i < J,\ i \neq k \\ \frac{\tau \nu_i}{(\sum_{r=J}^{k-1} \nu_r) b_i} & i \geq J,\ i \neq k \\ \frac{\tau}{b_k} & \text{if } i = k \end{cases} \qquad \forall i \leq k.$$

Thus for any $\boldsymbol{y} \in U$, $j \in \{1, \ldots, k-1\}$ and $\boldsymbol{z} \in \mathcal{B}_{\boldsymbol{\delta}_{[j]}} \times \{0\}^{\mathbb{N}}$

$$\sum_{i \in \mathbb{N}} (y_i + z_i) \psi_i = \underbrace{\left(\sum_{i \in \mathbb{N}} y_i \psi_i \right)}_{=:\phi} + \underbrace{\left(\sum_{i=1}^{j} z_{i,2} \psi_i \right)}_{=:\psi} + \left(\sum_{i=1}^{j} z_{i,1} \psi_i \right).$$

With $K \subseteq Z$ from Step 1, $\|\psi\|_{Z_{\mathbb{C}}} \leq \sum_{i \in \mathbb{N}} \tau \frac{\nu_i b_i}{(\sum_{r=J}^{k-1} \nu_r) b_i} + \tau \frac{b_k}{b_k} \leq 2\tau$ and $\phi \in K$. Thus (B.8) implies Assumption 4.1 (e). Finally, Assumption 4.1 (d) is a consequence of (B.6) and (B.11). □

C Proofs of Sect. 5

C.1 Theorem 5.2

First we show two summability results, similar to [7,Lemma 7.1] and [7,Theorem 7.2]. In the following we write $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. For an index set $I \subseteq \mathbb{N}$, $\boldsymbol{\nu} = (\nu_j)_{j \in I} \in \mathbb{N}_0^I$ and $\boldsymbol{\varrho} = (\varrho_j)_{j \in I} \in [0, \infty)^I$ we use the notation

$$\operatorname{supp} \boldsymbol{\nu} = \{j \in I \,:\, \nu_j \neq 0\}, \qquad |\boldsymbol{\nu}| := \sum_{j \in \operatorname{supp} \boldsymbol{\nu}} \nu_j, \qquad \boldsymbol{\varrho}^{\boldsymbol{\nu}} := \prod_{j \in \operatorname{supp} \boldsymbol{\nu}} \varrho_j^{\nu_j},$$

where empty sums equal 0 and empty products equal 1.

Lemma C.1 *Let $\tau > 0$ and let $\boldsymbol{\varrho} \in (1, \infty)^{\mathbb{N}}$ be such that $(\varrho_j^{-1}) \in \ell^p(\mathbb{N})$ for some $p \in (0, 1]$ and additionally $\sup_{j \in \mathbb{N}} \varrho_j^{-1} < 1$. Then with $\gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})$ as in (4.6),*

$$\sum_{k \in \mathbb{N}} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^k} \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})^p \prod_{j=1}^{k} (1 + 2\nu_j)^{\tau} < \infty.$$

Proof The assumptions on $\boldsymbol{\varrho}$ imply

$$C_0 := \sum_{\{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} \,:\, |\boldsymbol{\nu}| < \infty\}} \boldsymbol{\varrho}^{-p\boldsymbol{\nu}} \prod_{j \in \mathbb{N}} (1 + 2\nu_j)^{\tau} < \infty$$

see for example [42,Lemma 3.10]. Let $\mathbf{0}:=(0)_{j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$. For any $\mathbf{0} \neq \boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}}$ with $|\boldsymbol{\nu}| < \infty$, we have $\boldsymbol{\nu} = (\boldsymbol{\eta}, \mathbf{0})$ with $\boldsymbol{\eta} \in \mathbb{N}_0^{k-1} \times \mathbb{N}$ and $k:=\max_j \nu_j \neq 0$. Thus with the convention $\mathbb{N}_0^0 \times \mathbb{N}:=\mathbb{N}$,

$$\{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} : |\boldsymbol{\nu}| < \infty\} = \{\mathbf{0}\} \cup \bigcup_{k\in\mathbb{N}} \{(\boldsymbol{\nu}, \mathbf{0}) : \boldsymbol{\nu} \in \mathbb{N}_0^{k-1} \times \mathbb{N}\}.$$

Hence

$$1 + \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{k-1}\times\mathbb{N}} \boldsymbol{\varrho}_{[k]}^{-p\boldsymbol{\nu}} \prod_{j=1}^{k} (1 + 2\nu_j)^{\tau} = C_0.$$

Using the convention $\boldsymbol{\varrho}_{[0]}^{-\boldsymbol{\nu}_{[0]}} = 1$, by definition

$$\gamma(\boldsymbol{\varrho}, \boldsymbol{\nu}) = \begin{cases} \varrho_k^{-1} \boldsymbol{\varrho}_{[k-1]}^{-\boldsymbol{\nu}_{[k-1]}} & \text{if } \nu_k = 0 \\ \boldsymbol{\varrho}_{[k]}^{-\boldsymbol{\nu}} & \text{if } \nu_k > 0 \end{cases} \qquad \forall \boldsymbol{\nu} \in \mathbb{N}_0^k.$$

Partitioning $\mathbb{N}_0^k = (\mathbb{N}_0^{k-1} \times \{0\}) \cup (\mathbb{N}_0^{k-1} \times \mathbb{N})$ we get

$$\begin{aligned} &\sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})^p \prod_{j=1}^{k} (1 + 2\nu_j)^{\tau} \\ &\quad = \sum_{k\in\mathbb{N}} \varrho_k^{-p} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{k-1}\times\{0\}} \boldsymbol{\varrho}_{[k-1]}^{-p\boldsymbol{\nu}_{[k-1]}} \prod_{j=1}^{k-1} (1 + 2\nu_j)^{\tau} \\ &\qquad + \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{k-1}\times\mathbb{N}} \boldsymbol{\varrho}_{[k]}^{-p\boldsymbol{\nu}} \prod_{j=1}^{k} (1 + 2\nu_j)^{\tau} \\ &\quad \leq \sum_{k\in\mathbb{N}} \varrho_k^{-p} C_0 + C_0 < \infty, \end{aligned}$$

since $\sum_{k\in\mathbb{N}} \varrho_k^{-p} < \infty$. □

Lemma C.2 *Let $\tau > 0$ and let $\boldsymbol{\varrho} \in (1, \infty)^{\mathbb{N}}$ be such that $(\varrho_j^{-1}) \in \ell^p(\mathbb{N})$ for some $p \in (0, 1]$ and additionally $\sum_{j\in\mathbb{N}} \varrho_j^{-1} < 1$. Then with $\gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})$ as in (4.6)*

$$\sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} \left(\frac{|\boldsymbol{\nu}|^{|\boldsymbol{\nu}|}}{\boldsymbol{\nu}^{\boldsymbol{\nu}}} \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu}) \right)^p \prod_{j=1}^{k} (1 + 2\nu_j)^{\tau} < \infty.$$

Proof By [42,Lemma 3.11], the assumptions on $\boldsymbol{\varrho}$ imply with $w_{\boldsymbol{\nu}} = \prod_j (1+2\nu_j)^\tau$

$$1 + \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{k-1}\times\mathbb{N}} \left(\frac{|\boldsymbol{\nu}|^{|\boldsymbol{\nu}|}}{\boldsymbol{\nu}^{\boldsymbol{\nu}}}\boldsymbol{\varrho}_{[k]}^{-\boldsymbol{\nu}}\right)^p w_{\boldsymbol{\nu}} = \sum_{\{\boldsymbol{\nu}\in\mathbb{N}_0^{\mathbb{N}}\,:\,|\boldsymbol{\nu}|<\infty\}} \left(\frac{|\boldsymbol{\nu}|^{|\boldsymbol{\nu}|}}{\boldsymbol{\nu}^{\boldsymbol{\nu}}}\boldsymbol{\varrho}^{-\boldsymbol{\nu}}\right)^p w_{\boldsymbol{\nu}} =: C_0 < \infty.$$

Hence, similar as in the proof of Lemma C.1

$$\begin{aligned}
&\sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} \left(\frac{|\boldsymbol{\nu}|^{|\boldsymbol{\nu}|}}{\boldsymbol{\nu}^{\boldsymbol{\nu}}}\gamma(\boldsymbol{\varrho},\boldsymbol{\nu})\right)^p w_{\boldsymbol{\nu}} \\
&\quad = \sum_{k\in\mathbb{N}} \varrho_k^{-p} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{k-1}\times\{0\}} \left(\frac{|\boldsymbol{\nu}|^{|\boldsymbol{\nu}|}}{\boldsymbol{\nu}^{\boldsymbol{\nu}}}\boldsymbol{\varrho}_{[k-1]}^{-\boldsymbol{\nu}_{[k-1]}}\right)^p w_{\boldsymbol{\nu}} + \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{k-1}\times\mathbb{N}} \left(\frac{|\boldsymbol{\nu}|^{|\boldsymbol{\nu}|}}{\boldsymbol{\nu}^{\boldsymbol{\nu}}}\boldsymbol{\varrho}_{[k]}^{-\boldsymbol{\nu}}\right)^p w_{\boldsymbol{\nu}} \\
&\quad \le \sum_{k\in\mathbb{N}} \varrho_k^{-p} C_0 + C_0.
\end{aligned}$$

□

In the following we denote by $L_n : U_1 \to \mathbb{R}$ for $n \in \mathbb{N}_0$ the n-th Legendre polynomial normalized in $L^2(U_1,\mu)$. Then $(L_n)_{n\in\mathbb{N}_0}$ forms an orthonormal basis of this space. More generally, setting $L_{\boldsymbol{\nu}}(\boldsymbol{x}) := \prod_{j=1}^k L_{\nu_j}(x_j)$ with $\boldsymbol{\nu}\in\mathbb{N}_0^k$ for $\boldsymbol{x}\in U_k$, the family $(L_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathbb{N}_0^k}$ forms an orthonormal basis of $L^2(U_k,\mu)$, and any function f in this space allows the representation $f(\boldsymbol{x}) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} L_{\boldsymbol{\nu}}(\boldsymbol{x}) l_{\boldsymbol{\nu}}$ with the coefficients $l_{\boldsymbol{\nu}} = \int_{U_k} L_{\boldsymbol{\nu}}(\boldsymbol{x}) f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \in \mathbb{R}$. We have [28,§18.2(iii) and §18.3]

$$\|L_{\boldsymbol{\nu}}\|_{L^\infty(U_k)} \le \prod_{j=1}^k (1+2\nu_j)^{\frac{1}{2}}. \tag{C.1}$$

To prove Theorem 5.2 we will bound the Legendre coefficients of $\sqrt{\partial_k T_k} - 1$. To this end we will use the next lemma, which we have also used in the analysis of the finite-dimensional case; see [41,Lemma 4.1]. For a proof in the one-dimensional case we refer to Chapter 12 in [10]; see the calculation in equations (12.4.24)–(12.4.26). The multidimensional case follows by applying the result in each variable separately, e.g., [5] or [40,Corollary B.2.7].

Lemma C.3 *Let* $\boldsymbol{\zeta} \in (0,\infty)^k$. *Let* $f : \mathcal{B}_{\boldsymbol{\zeta}}(U_1) \to \mathcal{B}_{r_1}$ *be differentiable. Then*

(i) for all $\boldsymbol{\nu}\in\mathbb{N}_0^k$

$$\begin{aligned}
&\left|\int_{U_d} f(\boldsymbol{y}) L_{\boldsymbol{\nu}}(\boldsymbol{y})\,\mathrm{d}\mu(\boldsymbol{y})\right| \\
&\quad \le r_1 \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \left(\frac{2(\zeta_j+1)}{\zeta_j}(1+2\nu_j)^{3/2}\right) \prod_{j=1}^k (1+\zeta_j)^{-\nu_j},
\end{aligned} \tag{C.2}$$

(ii) if $f : \mathcal{B}_{\boldsymbol{\zeta}_{[k-1]}}(U_1) \times [-1,1] \to \mathcal{B}_{r_2}$ *then for all* $\boldsymbol{\nu} \in \mathbb{N}_0^{k-1} \times \{0\}$

$$\left|\int_{U_d} f(\boldsymbol{y}) L_{\boldsymbol{\nu}}(\boldsymbol{y})\, \mathrm{d}\mu(\boldsymbol{y})\right| \le r_2 \prod_{j \in \operatorname{supp} \boldsymbol{\nu}} \left(\frac{2(\zeta_j+1)}{\zeta_j}(1+2\nu_j)^{3/2}\right) \prod_{j=1}^{k} (1+\zeta_j)^{-\nu_j}. \tag{C.3}$$

Proof of Theorem 5.2 We first define some constants used throughout the proof. Afterward the proof proceeds in 5 steps.

Let $M \le |\mathfrak{f}_\rho(\psi)|, |\mathfrak{f}_\pi(\psi)| \le L$ as stated in Assumption 2.2. Let $C_1, C_2, C_3 > 0$ be the constants from Proposition 4.2 depending on M and L. Let $(\kappa_j)_{j\in\mathbb{N}} \subset (0,\infty)$ (monotonically increasing) and $\tau > 0$ be as in Lemma B.3 (depending on $(b_j)_{j\in\mathbb{N}}$, C_1 and $\mathfrak{f}_\rho, \mathfrak{f}_\pi$). Then $\kappa_{\min} := \min_{j\in\mathbb{N}} \kappa_j > 0$. Fix $J \in \mathbb{N}$ so large and $\alpha > 0$ so small that

$$\sum_{j\ge J}\left(\frac{b_j}{C_2\tau}\right)^p < 1 \quad \text{and} \quad 1 + \frac{\alpha}{b_j} < \begin{cases} 1 + C_2\kappa_{\min} & j < J \\ \frac{C_2\tau}{b_j} & j \ge J \end{cases} \quad \forall j \in \mathbb{N}. \tag{C.4}$$

This is possible because $\boldsymbol{b} \in \ell^p(\mathbb{N})$, since $b_j = \max\{\|\psi_{j,\rho}\|_{Z_\rho}, \|\psi_{j,\pi}\|_{Z_\pi}\}$ (cp. Assumption 2.2). Then by Lemma B.3, $f_\rho(\boldsymbol{y}) = \mathfrak{f}_\rho(\sum_{j\in\mathbb{N}} y_j\psi_{\rho,j})$ and $f_\pi(\boldsymbol{y}) = \mathfrak{f}_\pi(\sum_{j\in\mathbb{N}} y_j\psi_{\pi,j})$ satisfy Assumption 4.1 with $(\delta_j)_{j\in\mathbb{N}}$ as in (B.2) (and with our above choice of $J \in \mathbb{N}$).

Step 1. We provide bounds on the Legendre coefficient

$$l_{k,\boldsymbol{\nu}} := \int_{U_k} (\sqrt{R_k(\boldsymbol{x})} - 1) L_{\boldsymbol{\nu}}(\boldsymbol{x})\, \mathrm{d}\mu(\boldsymbol{x}) \tag{C.5}$$

with $R_k = \partial_k T_k$ and $\boldsymbol{\nu} \in \mathbb{N}_0^k$ for $k \in \mathbb{N}$.

Fix $k \in \mathbb{N}$ and $\boldsymbol{0} \ne \boldsymbol{\nu} \in \mathbb{N}_0^k$, and let $\delta_j = \delta_j(J, \boldsymbol{\nu})$ be as in (B.2). According to Proposition 4.2 (applied with $j = k$)

(i) $R_k \in C^1(\mathcal{B}_{\boldsymbol{\zeta}_{[k]}}(U_1), \mathcal{B}_{C_3}(1))$ and $\Re(R_k(\boldsymbol{x})) \ge \frac{1}{C_3}$ for all $\boldsymbol{x} \in \mathcal{B}_{\boldsymbol{\zeta}_{[k]}}(U_1)$,
(ii) if $k \ge 2$, $R_k : \mathcal{B}_{\boldsymbol{\zeta}_{[k-1]}}(U_1) \times U_1 \to \mathcal{B}_{\frac{C_3}{\delta_k}}(1)$,

where $\zeta_j = C_2\delta_j$, $j \in \{1,\dots,k\}$, and the constants C_2 and C_3 solely depend on M and L but not on k or $\boldsymbol{\nu}$. In particular for

$$Q_k := \sqrt{R_k} - 1 = \sqrt{\partial_k T_k} - 1 \tag{C.6}$$

we get with $C_4 := \sqrt{1 + C_3} + 1$

$$Q_k : \mathcal{B}_{\boldsymbol{\zeta}_{[k]}}(U_1) \to \mathcal{B}_{C_4}, \tag{C.7}$$

which follows by (i) and $|\sqrt{R_k(\boldsymbol{x})}-1| \le |\sqrt{R_k(\boldsymbol{x})}|+1 \le \sqrt{C_3+1}+1$ for all $\boldsymbol{x} \in \mathcal{B}_{\boldsymbol{\zeta}_{[k]}}(U_1)$. We claim that with $r:=2C_3C_4 \ge C_3$ if $k \ge 2$

$$Q_k : \mathcal{B}_{\boldsymbol{\zeta}_{[k-1]}}(U_1) \times U_1 \to \mathcal{B}_{\frac{r}{\delta_k}}. \tag{C.8}$$

To show it fix $\boldsymbol{x} \in \mathcal{B}_{\boldsymbol{\zeta}_{[k-1]}}(U_1) \times U_1$. We distinguish between $\frac{C_3}{\delta_k} \le \frac{1}{2}$ and $\frac{C_3}{\delta_k} > \frac{1}{2}$. For any $q \in \mathbb{C}$ with $|q| \le \frac{1}{2}$ we have with $g(q):=\sqrt{1+q}-1$ that $g(0)=0$ and $|g'(q)| \le \sqrt{\frac{1}{2}}$. Thus $|\sqrt{1+q}-1| \le |q|$ for all $|q| \le \frac{1}{2}$. Therefore if $\frac{C_3}{\delta_k} \le \frac{1}{2}$ then by (ii) $|R_k(\boldsymbol{x})-1| \le \frac{C_3}{\delta_k} \le \frac{1}{2}$ and thus

$$|Q_k(\boldsymbol{x})| = |\sqrt{(R_k(\boldsymbol{x})-1)+1}-1| \le |R_k(\boldsymbol{x})-1| \le \frac{C_3}{\delta_k} \le \frac{r}{\delta_k}.$$

For the second case $\frac{C_3}{\delta_k} > \frac{1}{2}$, by (ii) we have $|\sqrt{R_k(\boldsymbol{x})}-1| \le 1+|\sqrt{R_k(\boldsymbol{x})}| \le 1+\sqrt{C_3+1} = C_4$. Since $\frac{C_3}{\delta_k} > \frac{1}{2}$ and thus $\delta_k \le 2C_3$, we can bound C_4 by $C_4 = \frac{r}{2C_3} \le \frac{r}{\delta_k}$, which concludes the proof of (C.8).

The fact that R_k has non-negative real part implies that its composition with the square root, i.e., the map $\boldsymbol{x} \mapsto \sqrt{R_k(\boldsymbol{x})}$, is well-defined and differentiable on $\mathcal{B}_{\boldsymbol{\zeta}_{[k]}}(U_1)$. With $\kappa_{\min} = \min_{j\in\mathbb{N}} \kappa_j > 0$ set $\zeta_{\min}:=C_2\kappa_{\min} > 0$ and observe that $\zeta_j = C_2\delta_j \ge \zeta_{\min}$ for all $j \in \mathbb{N}$ (cp. (B.2)). Let

$$w_{\boldsymbol{\nu}} = \prod_{j=1}^{k}(1+2\nu_j)^{\theta}, \tag{C.9}$$

with $\theta = \frac{3}{2} + \log_3(\frac{2(1+\zeta_{\min})}{\zeta_{\min}})$. Then

$$\begin{aligned}\prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \frac{2(\zeta_j+1)}{\zeta_j}(1+2\nu_j)^{3/2} &= \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} 3^{\log_3(\frac{2(\zeta_j+1)}{\zeta_j})}(1+2\nu_j)^{3/2}\\ &\le \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} (1+2\nu_j)^{\theta} = w_{\boldsymbol{\nu}}.\end{aligned}$$

With Lemma C.3 (i) and (C.7) we obtain for the Legendre coefficients of Q_k in (C.5)

$$|l_{k,\boldsymbol{\nu}}| \le w_{\boldsymbol{\nu}} C_4 \prod_{j=1}^{k}(1+\zeta_j)^{-\nu_j} \qquad \forall \boldsymbol{\nu} \in \mathbb{N}_0^k. \tag{C.10}$$

Moreover with Lemma C.3 (ii) and (C.8)

$$|l_{k,\boldsymbol{\nu}}| \le w_{\boldsymbol{\nu}} \frac{r}{\delta_k} \prod_{j=1}^{k}(1+\zeta_j)^{-\nu_j} \qquad \forall \boldsymbol{\nu} \in \mathbb{N}_0^{k-1} \times \{0\}. \tag{C.11}$$

Step 2. We provide a bound on $|l_{k,\boldsymbol{\nu}}|$ in terms of $\gamma(\tilde{\boldsymbol{\varrho}}, \boldsymbol{\nu})$ for some $\tilde{\boldsymbol{\varrho}}$.

Fix again $k \in \mathbb{N}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^k$. For $j \in \mathbb{N}$ by definition of $\zeta_j = C_2\delta_j$ and $\delta_j = \delta_j(J, \boldsymbol{\nu})$ in (B.2)

$$\zeta_j = C_2\delta_j = C_2 \begin{cases} \kappa_j + 0 & j \neq k,\ j < J \\ \kappa_j + \frac{\tau\nu_j}{(\sum_{i=J}^{k-1}\nu_i)b_j} & j \neq k,\ j \geq J \\ \kappa_k + \frac{\tau}{b_k} & j = k. \end{cases}$$

Since $C_2 \in (0, 1]$ (see Proposition 4.2) and $\kappa_{\min} \leq \kappa_j$, it holds with $|\boldsymbol{\nu}_{[J:k]}| = \sum_{j=J}^{k}\nu_j \geq \sum_{j=J}^{k-1}\nu_j$

$$\zeta_j \geq \begin{cases} C_2\kappa_{\min} & j < J \\ \frac{C_2\tau\nu_j}{|\boldsymbol{\nu}_{[J:k]}|b_j} & j \geq J \end{cases} \qquad \forall j \in \{1, \ldots, k\}$$

and additionally

$$\frac{r}{\delta_k} = \frac{r}{\kappa_k + \tau/b_k} \leq \frac{b_k r}{\tau}.$$

Thus by (C.10) and (C.11) for $k \in \mathbb{N}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^k$

$$|l_{k,\boldsymbol{\nu}}| \leq C_4 w_{\boldsymbol{\nu}} \prod_{j=1}^{J-1}(1 + C_2\kappa_{\min})^{-\nu_j} \prod_{i=J}^{k} \frac{|\boldsymbol{\nu}_{[J:k]}|^{\nu_i}}{\nu_i^{\nu_i}} \prod_{i=J}^{k}\left(\frac{b_i}{C_2\tau}\right)^{\nu_i} \cdot \begin{cases} 1 & k \in \operatorname{supp}\boldsymbol{\nu} \\ \frac{b_k r}{\tau} & k \notin \operatorname{supp}\boldsymbol{\nu} \end{cases} \tag{C.12}$$

with empty products equal to 1 by convention.

Defining

$$\tilde{\varrho}_j := \begin{cases} 1 + C_2\kappa_{\min} & j < J \\ \frac{C_2\tau}{b_j} & j \geq J \end{cases} \qquad \forall j \in \mathbb{N} \tag{C.13}$$

bound (C.12) becomes with $\gamma(\tilde{\boldsymbol{\varrho}}, \boldsymbol{\nu}) = \tilde{\varrho}_k^{-\max\{1,\nu_k\}} \prod_{j=1}^{k-1} \tilde{\varrho}_j^{-\nu_j}$

$$\begin{aligned} |l_{k,\boldsymbol{\nu}}| &\leq C_4 w_{\boldsymbol{\nu}} \prod_{i=J}^{k} \frac{|\boldsymbol{\nu}_{[J:k]}|^{\nu_i}}{\nu_i^{\nu_i}} \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \tilde{\varrho}_j^{-\nu_j} \cdot \begin{cases} 1 & k \in \operatorname{supp}\boldsymbol{\nu} \\ \frac{b_k r}{\tau} & k \notin \operatorname{supp}\boldsymbol{\nu} \end{cases} \\ &= C_4 w_{\boldsymbol{\nu}} \prod_{i=J}^{k} \frac{|\boldsymbol{\nu}_{[J:k]}|^{\nu_i}}{\nu_i^{\nu_i}} \gamma(\tilde{\boldsymbol{\varrho}}, \boldsymbol{\nu}) \cdot \begin{cases} 1 & k \in \operatorname{supp}\boldsymbol{\nu} \\ \tilde{\varrho}_k \frac{b_k r}{\tau} & k \notin \operatorname{supp}\boldsymbol{\nu} \end{cases} \\ &\leq C_5 w_{\boldsymbol{\nu}} \gamma(\tilde{\boldsymbol{\varrho}}, \boldsymbol{\nu}) \prod_{i=J}^{k} \frac{|\boldsymbol{\nu}_{[J:k]}|^{\nu_j}}{\nu_i^{\nu_i}} \end{aligned} \tag{C.14}$$

where

$$C_5:=C_4 \sup_{k\in\mathbb{N}} \tilde{\varrho}_k \frac{b_k r}{\tau} < \infty$$

is finite by definition of $\tilde{\varrho}_j$ in (C.13).

With α in (C.4) introduce

$$\varrho_j:=1+\frac{\alpha}{b_j} < \tilde{\varrho}_j \qquad \forall j \in \mathbb{N}. \tag{C.15}$$

Then

$$\gamma(\tilde{\boldsymbol{\varrho}}, \boldsymbol{\nu}) \leq \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu}). \tag{C.16}$$

Step 3. We show a summability result for the Legendre coefficients.

For notational convenience we introduce the shortcuts

$$\boldsymbol{\nu}_E:=\boldsymbol{\nu}_{[J-1]}, \quad \boldsymbol{\nu}_F:=\boldsymbol{\nu}_{[J:k]}, \quad \tilde{\boldsymbol{\varrho}}_E:=\tilde{\boldsymbol{\varrho}}_{[J-1]}, \quad \tilde{\boldsymbol{\varrho}}_F:=\tilde{\boldsymbol{\varrho}}_{[J:k]}.$$

Hence $\tilde{\boldsymbol{\varrho}}_E^{-\boldsymbol{\nu}_E} = \prod_{j=1}^{J-1} \varrho_j^{-\nu_j}$, $\boldsymbol{\nu}_F^{\boldsymbol{\nu}_F} = \prod_{j\geq J} \nu_j^{\nu_j}$, $\gamma(\boldsymbol{\varrho}_F, \boldsymbol{\nu}_F) = \varrho_k^{-\max\{1,\nu_k\}} \prod_{j=J}^{k-1} \varrho_j^{-\nu_j}$ in case $k \geq J$ etc. For $k \geq J$ and $\boldsymbol{\nu} \in \mathbb{N}_0^k$

$$\begin{aligned}\gamma(\tilde{\boldsymbol{\varrho}}, \boldsymbol{\nu}) &= \tilde{\boldsymbol{\varrho}}_{[k]}^{-\boldsymbol{\nu}} \cdot \begin{cases} \tilde{\varrho}_k^{-1} & k \notin \operatorname{supp} \boldsymbol{\nu} \\ 1 & k \in \operatorname{supp} \boldsymbol{\nu} \end{cases} = \tilde{\boldsymbol{\varrho}}_E^{-\boldsymbol{\nu}_E} \tilde{\boldsymbol{\varrho}}_F^{-\boldsymbol{\nu}_F} \\ &\quad \cdot \begin{cases} \tilde{\varrho}_k^{-1} & k \notin \operatorname{supp} \boldsymbol{\nu} \\ 1 & k \in \operatorname{supp} \boldsymbol{\nu} \end{cases} = \tilde{\boldsymbol{\varrho}}_E^{-\boldsymbol{\nu}_E} \gamma(\tilde{\boldsymbol{\varrho}}_F, \boldsymbol{\nu}_F).\end{aligned}$$

By (C.16) and because $p < 1$ it holds $\gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})^{p-1} \leq \gamma(\tilde{\boldsymbol{\varrho}}, \boldsymbol{\nu})^{p-1}$. Thus by (C.14) and (C.15)

$$\begin{aligned}&\sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}| \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})^{p-1} \\ &\quad \leq C_5 \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}}^2 \gamma(\tilde{\boldsymbol{\varrho}}_{[k]}, \boldsymbol{\nu}) \frac{|\boldsymbol{\nu}_F|^{|\boldsymbol{\nu}_F|}}{\boldsymbol{\nu}_F^{\boldsymbol{\nu}_F}} \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})^{p-1} \\ &\quad \leq C_5 \sum_{k=1}^{J-1} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}}^2 \gamma(\tilde{\boldsymbol{\varrho}}_{[k]}, \boldsymbol{\nu})^p \\ &\qquad + C_5 \sum_{k\geq J} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}}^2 \gamma(\tilde{\boldsymbol{\varrho}}_{[k]}, \boldsymbol{\nu})^p \frac{|\boldsymbol{\nu}_F|^{|\boldsymbol{\nu}_F|}}{\boldsymbol{\nu}_F^{\boldsymbol{\nu}_F}}\end{aligned}$$

$$\leq C_5 \sum_{k=1}^{J-1} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}}^2 \gamma(\tilde{\boldsymbol{\varrho}}_{[k]}, \boldsymbol{\nu})^p$$

$$+C_5 \sum_{k\geq J} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}}^2 \tilde{\boldsymbol{\varrho}}_E^{-p\boldsymbol{\nu}_E} \gamma(\tilde{\boldsymbol{\varrho}}_F, \boldsymbol{\nu}_F)^p \frac{|\boldsymbol{\nu}_F|^{|\boldsymbol{\nu}_F|}}{\boldsymbol{\nu}_F^{\boldsymbol{\nu}_F}}. \tag{C.17}$$

By Lemma C.1 (here we use that $\sup_{j\in\{1,\dots,J-1\}} \tilde{\varrho}_j^{-1} < 1$, see (C.4) and (C.13)), the first sum is bounded. For the second sum in (C.17)

$$\sum_{k\geq J} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}} \tilde{\boldsymbol{\varrho}}_E^{-p\boldsymbol{\nu}_E} \gamma(\tilde{\boldsymbol{\varrho}}_F, \boldsymbol{\nu}_F)^p \frac{|\boldsymbol{\nu}_F|^{|\boldsymbol{\nu}_F|}}{\boldsymbol{\nu}_F^{\boldsymbol{\nu}_F}}$$

$$= \sum_{k\geq J} \left(\sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{J-1}} w_{\boldsymbol{\nu}} \tilde{\boldsymbol{\varrho}}_E^{-p\boldsymbol{\nu}} \right) \left(\sum_{\boldsymbol{\mu}\in\mathbb{N}_0^{k-J+1}} \frac{|\boldsymbol{\mu}|^{|\boldsymbol{\mu}|}}{\boldsymbol{\mu}^{\boldsymbol{\mu}}} \gamma(\tilde{\boldsymbol{\varrho}}_{[J:k]}, \boldsymbol{\mu})^p \right). \tag{C.18}$$

E.g., by [42,Lemma 3.10] (again due to $\sup_{j\in\{1,\dots,J-1\}} \tilde{\varrho}_j^{-1} < 1$)

$$\sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{J-1}} w_{\boldsymbol{\nu}} \tilde{\boldsymbol{\varrho}}_E^{-p\boldsymbol{\nu}} =: C_0 < \infty,$$

and thus (C.18) is bounded by

$$C_0 \sum_{k\geq J} \sum_{\boldsymbol{\mu}\in\mathbb{N}_0^{k-J+1}} \frac{|\boldsymbol{\mu}|^{|\boldsymbol{\mu}|}}{\boldsymbol{\mu}^{\boldsymbol{\mu}}} \gamma(\tilde{\boldsymbol{\varrho}}_{[J:k]}, \boldsymbol{\mu})^p = C_0 \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\mu}\in\mathbb{N}_0^k} \frac{|\boldsymbol{\mu}|^{|\boldsymbol{\mu}|}}{\boldsymbol{\mu}^{\boldsymbol{\mu}}} \gamma(\tilde{\boldsymbol{\varrho}}_{[J:J+k]}, \boldsymbol{\mu})^p < \infty \tag{C.19}$$

by Lemma C.2 and because $\sum_{j\geq J} (\tilde{\varrho}_j^p)^{-1} < 1$ by (C.4) and (C.13). In all

$$\sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}| \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})^{p-1} =: C_6 < \infty. \tag{C.20}$$

Step 4. As before, by Lemma C.1 and because $\sup_{j\in\mathbb{N}} \varrho_j^{-1} < 1$ and $(\varrho_j^{-1})_{j\in\mathbb{N}} \in \ell^p(\mathbb{N})$ (cp. (C.15)),

$$\sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu})^p =: C_7 < \infty.$$

For $k \in \mathbb{N}$ and $\varepsilon > 0$ set

$$\Lambda_{\varepsilon,k} = \{\boldsymbol{\nu} \in \mathbb{N}_0^k \, : \, \gamma(\boldsymbol{\varrho}, \boldsymbol{\nu}) \geq \varepsilon\} \quad \text{and} \quad N_\varepsilon := \sum_{k\in\mathbb{N}} |\Lambda_{\varepsilon,k}|.$$

Then

$$N_\varepsilon = \sum_{\{(k,\boldsymbol{\nu})\,:\,\gamma(\boldsymbol{\varrho},\boldsymbol{\nu})\geq\varepsilon\}} \gamma(\boldsymbol{\varrho},\boldsymbol{\nu})^p \gamma(\boldsymbol{\varrho},\boldsymbol{\nu})^{-p} \leq \varepsilon^{-p} \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} \gamma(\boldsymbol{\varrho},\boldsymbol{\nu})^p = C_7 \varepsilon^{-p}$$

and thus

$$\varepsilon \leq \left(\frac{N_\varepsilon}{C_7}\right)^{-\frac{1}{p}} \qquad \forall \varepsilon > 0. \tag{C.21}$$

On the other hand, assuming $\varepsilon > 0$ to be so small that $N_\varepsilon > 0$, by (C.20)

$$\begin{aligned}\sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\setminus\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}| &= \sum_{\{(k,\boldsymbol{\nu})\,:\,\boldsymbol{\nu}\in\mathbb{N}_0^k,\ \gamma(\boldsymbol{\varrho},\boldsymbol{\nu})<\varepsilon\}} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}| \\ &= \sum_{\{(k,\boldsymbol{\nu})\,:\,\boldsymbol{\nu}\in\mathbb{N}_0^k,\ \gamma(\boldsymbol{\varrho},\boldsymbol{\nu})<\varepsilon\}} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}| \gamma(\boldsymbol{\varrho},\boldsymbol{\nu})^{p-1} \gamma(\boldsymbol{\varrho},\boldsymbol{\nu})^{1-p} \\ &\leq C_6 \varepsilon^{1-p} \leq (C_6 C_7^{\frac{1}{p}-1}) N_\varepsilon^{-\frac{1}{p}+1}.\end{aligned} \tag{C.22}$$

Step 5. We finish the proof and verify (5.3).

For $k \in \mathbb{N}$ and $\varepsilon > 0$ define $p_{\varepsilon,k} := \sum_{\boldsymbol{\nu}\in\Lambda_{\varepsilon,k}} l_{k,\boldsymbol{\nu}} L_{\boldsymbol{\nu}} \in \mathbb{P}_{\Lambda_{\varepsilon,k}}$. We have $\sqrt{\partial_k T_k} - 1 = Q_k = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} l_{k,\boldsymbol{\nu}} L_{\boldsymbol{\nu}}$. Since $\|L_{\boldsymbol{\nu}}\|_{L^\infty(U_k)} \leq w_{\boldsymbol{\nu}}$ by (C.1) and (C.9), by (C.20) and because $\gamma(\boldsymbol{\varrho},\boldsymbol{\nu})^{p-1} \geq 1$

$$\sup_{k\in\mathbb{N}} \|Q_k\|_{L^\infty(U_k)} \leq \sup_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}| \leq \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}| \leq C_6 < \infty. \tag{C.23}$$

Similarly

$$\|Q_k - p_{\varepsilon,k}\|_{L^\infty(U_k)} \leq \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\setminus\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}} |l_{k,\boldsymbol{\nu}}|. \tag{C.24}$$

In [41,Lemma C.4] we showed that there exists $K \in (0, 1]$ and $C_K > 0$ (both independent of k) such that

$$\|Q_k - p_{\varepsilon,k}\|_{L^\infty(U_k)} < \frac{K}{1 + \|Q_k\|_{L^\infty(U_k)}} \tag{C.25}$$

implies

$$\|T_k - \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \leq C_K (1 + \|Q_k\|_{L^\infty(U_k)})^3 \|Q_k - p_{\varepsilon,k}\|_{L^\infty(U_k)} \tag{C.26}$$

and

$$\|\partial_k T_k - \partial_k \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \leq C_K (1 + \|Q_k\|_{L^\infty(U_k)})^3 \|Q_k - p_{\varepsilon,k}\|_{L^\infty(U_k)}. \tag{C.27}$$

We distinguish between two cases, first assuming

$$\sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\backslash\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}}|l_{k,\boldsymbol{\nu}}| < \frac{K}{1+\|Q_k\|_{L^\infty(U_k)}}. \tag{C.28}$$

By (C.24) and (C.28), (C.25) holds. Now, (C.23), (C.24) and (C.26) imply

$$\|T_k - \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \le C_K(1+C_6)^3 \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\backslash\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}}|l_{k,\boldsymbol{\nu}}|, \tag{C.29}$$

and by (C.27)

$$\|\partial_k T_k - \partial_k\tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \le C_K(1+C_6)^3 \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\backslash\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}}|l_{k,\boldsymbol{\nu}}|.$$

In the second case where

$$\sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\backslash\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}}|l_{k,\boldsymbol{\nu}}| > \frac{K}{1+\|Q_k\|_{L^\infty(U_k)}}, \tag{C.30}$$

we redefine $p_{\varepsilon,k}:=0$, so that $\tilde{T}_{\varepsilon,k}(\boldsymbol{x}) = x_k$ (cp. Remark 5.3). Since $T_k : U_k \to U_1$ and $\tilde{T}_{\varepsilon,k} : U_k \to U_1$, we get $\|T_k - \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \le 2$, and therefore by (C.23)

$$\begin{aligned}\|T_k - \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} &\le \frac{2}{\frac{K}{1+\|Q_k\|_{L^\infty(U_k)}}} \frac{K}{1+\|Q_k\|_{L^\infty(U_k)}}\\ &\le \frac{2(1+C_6)}{K} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\backslash\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}}|l_{k,\boldsymbol{\nu}}|.\end{aligned} \tag{C.31}$$

Next, using $Q_k = \sqrt{\partial_k T_k} - 1$, by (C.23) it holds $\|\sqrt{\partial_k T_k}\|_{L^\infty(U_k)} \le 1 + C_6$ as well as $\|\partial_k T_k\|_{L^\infty(U_k)} \le (1+C_6)^2$. Similarly $\|\sqrt{\partial_k \tilde{T}_{\varepsilon,k}}\|_{L^\infty(U_k)} = \|p_{\varepsilon,k}\|_{L^\infty(U_k)} \le 1 + C_6$ and $\|\partial_k\tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \le (1+C_6)^2$. Still assuming (C.30), we get analogous to (C.31)

$$\begin{aligned}\|\partial_k T_k - \partial_k\tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} &\le \|\partial_k T_k\|_{L^\infty(U_k)} + \|\partial_k\tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)}\\ &\le 2(1+C_6)^2 \le \frac{2(1+C_6)^3}{K} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\backslash\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}}|l_{k,\boldsymbol{\nu}}|.\end{aligned}$$

In total, by (C.29), (C.31), and (C.22)

$$\sum_{k\in\mathbb{N}} \|T_k - \tilde{T}_{\varepsilon,k}\|_{L^\infty(U_k)} \le C \sum_{k\in\mathbb{N}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^k\backslash\Lambda_{\varepsilon,k}} w_{\boldsymbol{\nu}}|l_{k,\boldsymbol{\nu}}| \le C N_\varepsilon^{-\frac{1}{p}+1},$$

for some $C > 0$ independent of $\varepsilon > 0$. An analogous estimate is obtained for $\sum_{k\in\mathbb{N}} \|\partial_k T_k - \partial_k \tilde{T}_k\|_{L^\infty(U_k)}$. □

C.2 Corollary 5.4

For the proof we will need the following two lemmata. The first one is classical, e.g., [41,Lemma 3.1].

Lemma C.4 *Let $\zeta > 0$. Assume that $f \in C^1(\mathcal{B}_\zeta(U_1), \mathbb{C})$ such that $\sup_{x\in\mathcal{B}_\zeta(U_1)} |f(x)| \le L$. Then $\sup_{x\in U_1} |f'(x)| \le \frac{L}{\zeta}$ and $f : U_1 \to \mathbb{C}$ is Lipschitz continuous with Lipschitz constant $\frac{L}{\zeta}$.*

For a function g denote by $\mathrm{Lip}[g] \in [0, \infty]$ its Lipschitz constant.

Lemma C.5 *Let f_ρ, f_π satisfy Assumption 2.2. Then there exists $K > 0$ such that for all $k \in \mathbb{N}$, all $j < k$ and all $\boldsymbol{x} \in U$ with $b_j := \max\{\|\psi_{\rho,j}\|_Z, \|\psi_{\pi,j}\|_Z\}$*

$$\mathrm{Lip}\big[U_1 \ni x_j \mapsto \partial_k T_k(\boldsymbol{x}_{[k]})\big] \le K b_k b_j, \qquad \mathrm{Lip}\big[U_1 \ni x_k \mapsto \partial_k T_k(\boldsymbol{x}_{[k]})\big] \le K b_k \tag{C.32}$$

and

$$\mathrm{Lip}\big[U_1 \ni x_j \mapsto T_k(\boldsymbol{x}_{[k]})\big] \le 2K b_k b_j, \qquad \mathrm{Lip}\big[U_1 \ni x_k \mapsto T_k(\boldsymbol{x}_{[k]})\big] \le 1 + K. \tag{C.33}$$

Proof Fix $k > 1$ and $j \in \{1, \dots, k-1\}$. First applying Lemma B.3 with $J := 1$ and the multiindex $\boldsymbol{\nu} \in \mathbb{N}_0^k$ with $\nu_i = 0$ if $i \neq j$ and $\nu_j = 1$, and then applying Proposition 4.2 (ii) it holds for some $\boldsymbol{\zeta} \in (0, \infty)^k$ where in particular

$$\zeta_j = \frac{C_2 \tau}{b_j}$$

that

$$\partial_k T_k - 1 : \mathcal{B}_{\boldsymbol{\zeta}_{[k-1]}}(U_1) \times U_1 \to \mathcal{B}_{\frac{C_3 b_k}{\tau}}.$$

Moreover this function is complex differentiable in $x_j \in \mathcal{B}_{\zeta_j}(U_1)$. Here the constants C_2, C_3 and τ solely depend on ρ and π, and we point out that we used the trivial lower bounded $\kappa_j \ge 0$ for κ_j in Lemma B.3. By Lemma C.4

$$\mathrm{Lip}\big[U_1 \ni x_j \mapsto \partial_k T_k(\boldsymbol{x}_{[k]})\big] = \mathrm{Lip}\big[U_1 \ni x_j \mapsto \partial_k T_k(\boldsymbol{x}_{[k]}) - 1\big] \le \frac{\frac{C_3 b_k}{\tau}}{\zeta_j} \le K b_k b_j,$$

with $K := \max\{\frac{C_3}{C_2}, \frac{C_3}{C_2 \tau}, C_3\} \ge \frac{C_3}{C_2}$. This shows the first inequality in (C.32).

Fix $k \in \mathbb{N}$. Similar as above, choosing $\boldsymbol{\nu} \in \mathbb{N}_0^k$ such that $\nu_i = 0$ if $i \neq k$ and $\nu_k = 1$ in Lemma B.3, we find with Proposition 4.2 (i) that

$$\partial_k T_k - 1 : \mathcal{B}_{\boldsymbol{\zeta}_{[k]}}(U_1) \to \mathcal{B}_{C_3}, \tag{C.34}$$

where now $\zeta_k = C_2 \frac{\tau}{b_k}$. Again by Lemma C.4

$$\mathrm{Lip}\big[U_1 \ni x_k \mapsto \partial_k T_k(\boldsymbol{x}_{[k]})\big] = \mathrm{Lip}\big[U_1 \ni x_k \mapsto \partial_k T_k(\boldsymbol{x}_{[k]}) - 1\big] \leq \frac{C_3}{\frac{C_2 \tau}{b_k}} \leq K b_k,$$

which shows the second inequality in (C.32).

Next we show the first inequality in (C.33) and fix $j < k$. For $\boldsymbol{y} \in U_k$ and with $\tilde{\boldsymbol{y}} := (\boldsymbol{y}_{[j-1]}, \tilde{y}_j, \boldsymbol{y}_{[j+1:k]})$

$$\begin{aligned} |T_k(\boldsymbol{y}) - T_k(\tilde{\boldsymbol{y}})| &\leq \int_{-1}^{y_k} |\partial_k T_k(\boldsymbol{y}_{[k-1]}, t) - \partial_k T_k(\tilde{\boldsymbol{y}}_{[k-1]}, t)| \,\mathrm{d}t \\ &\leq \int_{-1}^{1} K b_k b_j |y_j - \tilde{y}_j| \,\mathrm{d}t \leq 2 K b_k b_j |y_j - \tilde{y}_j|. \end{aligned}$$

For the second inequality in (C.33) let $\boldsymbol{y} \in U_k$ and $\tilde{\boldsymbol{y}} = (\boldsymbol{y}_{[k-1]}, \tilde{y}_k)$. Then

$$\begin{aligned} |T_k(\boldsymbol{y}) - T_k(\tilde{\boldsymbol{y}})| &\leq \int_{\tilde{y}_k}^{y_k} |\partial_k T_k(\boldsymbol{y}_{[k-1]}, t)| \,\mathrm{d}t \\ &\leq |y_k - \tilde{y}_k| \sup_{t \in U_1} |\partial_k T_k(\boldsymbol{y}, t)| \\ &\leq |y_k - \tilde{y}_k|(1 + C_3) \leq |y_k - \tilde{y}_k|(1 + K), \end{aligned}$$

where we used (C.34) to bound $\sup_{t \in U_1} |\partial_k T_k(\boldsymbol{y}, t)| \leq 1 + C_3$. □

Proof of Cor. 5.4 For notational convenience we drop the index ε and write $\tilde{T}_k$ instead of $\tilde{T}_{\varepsilon,k}$ etc.

Step 1. We show (5.4a). Since the assumptions on ρ and π are the same (see Assumption 2.2), switching the roles of the measures, (C.33) implies for the inverse transport $S = (S_k)_{k \in \mathbb{N}}$ (the KR transport satisfying $S_\sharp \pi = \rho$)

$$\mathrm{Lip}\big[U_1 \ni x_j \mapsto S_k(\boldsymbol{x}_{[k]})\big] \leq 2 K b_k b_j, \qquad \mathrm{Lip}\big[U_1 \ni x_k \mapsto S_k(\boldsymbol{x}_{[k]})\big] \leq 1 + K.$$

Recall the notation $T_{[k]} = (T_i)_{i=1}^k : U_k \to U_k$ and $T_{[j:k]} = (T_i)_{i=j}^k : U_k \to U_{k-j+1}$ for the components of the transport map. Then for any $k \in \mathbb{N}$ it holds on U_k

$$\begin{aligned} |S_k(\tilde{T}_{[k]}) - S_k(T_{[k]})| &\leq \textstyle\sum_{j=1}^k |S_k(\tilde{T}_{[j]}, T_{[j+1:k]}) - S_k(\tilde{T}_{[j-1]}, T_{[j:k]})| \\ &\leq (1 + K)|\tilde{T}_k - T_k| + \textstyle\sum_{j=1}^{k-1} 2 K b_j b_k |\tilde{T}_j - T_j|. \end{aligned} \tag{C.35}$$

Since $\tilde{T}: U \to U$ is a bijection (in particular $\tilde{T}_{[k]}: U_k \to U_k$ is bijective) we get

$$\begin{aligned}
&\sum_{k\in\mathbb{N}} \|S_k - \tilde{S}_k\|_{L^\infty(U_k)} \\
&\quad = \sum_{k\in\mathbb{N}} \|S_k \circ \tilde{T}_{[k]} - \tilde{S}_k \circ \tilde{T}_{[k]}\|_{L^\infty(U_k)} \\
&\quad = \sum_{k\in\mathbb{N}} \|S_k \circ \tilde{T}_{[k]} - S_k \circ T_{[k]}\|_{L^\infty(U_k)} \\
&\quad \le 2K \sum_{k\in\mathbb{N}} \sum_{j=1}^{k-1} b_j b_k \|\tilde{T}_j - T_j\|_{L^\infty(U_k)} + (1+K) \sum_{k\in\mathbb{N}} \|\tilde{T}_k - T_k\|_{L^\infty(U_k)} \\
&\quad = 2K \sum_{j\in\mathbb{N}} b_j \|\tilde{T}_j - T_j\|_{L^\infty(U_k)} \sum_{k>j} b_k + (1+K) \sum_{k\in\mathbb{N}} \|\tilde{T}_k - T_k\|_{L^\infty(U_k)} \\
&\quad \le \left(1 + K + 2K \max_{j\in\mathbb{N}} b_j \sum_{i\in\mathbb{N}} b_i\right) \sum_{k\in\mathbb{N}} \|\tilde{T}_k - T_k\|_{L^\infty(U_k)}. \qquad \text{(C.36)}
\end{aligned}$$

Since $\sum_{i\in\mathbb{N}} b_i < \infty$ this together with (5.3a) shows (5.4a).

Step 2. We show (5.4b). For $\boldsymbol{x} \in U_k$ holds $S_{[k]} \circ T_{[k]}(\boldsymbol{x}) = \boldsymbol{x}$. Thus $S_k(T_{[k]}(\boldsymbol{x})) = x_k$ and therefore $\partial_k S_k(T_{[k]}(\boldsymbol{x}))\partial_k T_k(\boldsymbol{x}) = 1$, where we used that $T_{[j]}$ with $j < k$ only depends on $\boldsymbol{x}_{[j]}$. After applying $T_{[k]}^{-1} = S_{[k]}$ this reads

$$\partial_k S_k(\boldsymbol{x}) = \frac{1}{\partial_k T_k(S_{[k]}(\boldsymbol{x}))}.$$

The second inequality in (C.33) gives $|\partial_k T_k| \le 1 + K$ and thus $\partial_k S_k(\boldsymbol{x}) \ge \frac{1}{1+K}$ for all $k \in \mathbb{N}$ and all $\boldsymbol{x} \in U_k$. Similarly $\partial_k \tilde{S}_k(\boldsymbol{x}) = \frac{1}{\partial_k \tilde{T}_k(\tilde{S}_{[k]}(\boldsymbol{x}))}$. By (5.3b) (as long as $N_\varepsilon \ge 1$) we have for $\boldsymbol{x} \in U_k$

$$|\partial_k \tilde{T}_k(\boldsymbol{x})| \le |\partial_k T_k(\boldsymbol{x})| + |\partial_k T_k(\boldsymbol{x}) - \partial_k \tilde{T}_k(\boldsymbol{x})| \le 1 + K + C$$

with the constant C from (5.3b). Thus $\partial_k \tilde{S}_k(\boldsymbol{x}) \ge \frac{1}{1+K+C}$ for $\boldsymbol{x} \in U_k$. Since $x \mapsto \frac{1}{x}$: $[\frac{1}{1+K+C}, \infty) \to \mathbb{R}$ has Lipschitz constant $(1+K+C)^2$, we get

$$\begin{aligned}
&\sum_{k\in\mathbb{N}} \|\partial_k S_{[k]} - \partial_k \tilde{S}_{[k]}\|_{L^\infty(U_k)} = \sum_{k\in\mathbb{N}} \left\| \frac{1}{\partial_k T_k \circ S_{[k]}} - \frac{1}{\partial_k \tilde{T}_k \circ \tilde{S}_{[k]}} \right\|_{L^\infty(U_k)} \\
&\quad \le (1+K+C)^2 \sum_{k\in\mathbb{N}} \|\partial_k T_k \circ S_{[k]} - \partial_k \tilde{T}_k \circ \tilde{S}_{[k]}\|_{L^\infty(U_k)} \\
&\quad \le (1+K+C)^2 \sum_{k\in\mathbb{N}} \Big(\|\partial_k T_k \circ S_{[k]} - \partial_k T_k \circ \tilde{S}_{[k]}\|_{L^\infty(U_k)} \\
&\qquad + \|\partial_k T_k \circ \tilde{S}_{[k]} - \partial_k \tilde{T}_k \circ \tilde{S}_{[k]}\|_{L^\infty(U_k)} \Big). \qquad \text{(C.37)}
\end{aligned}$$

Using (C.32) the same calculation as in (C.35) yields

$$|\partial_k T_k(\tilde{S}_{[k]}) - \partial_k T_k(S_{[k]})| \leq \sum_{j=1}^{k} |\partial_k T_k(\tilde{S}_{[j]}, S_{[j+1:k]}) - \partial_k T_k(\tilde{S}_{[j-1]}, S_{[j:k]})|$$
$$\leq K b_k |\tilde{S}_k - S_k| + \sum_{j=1}^{k-1} K b_j b_k |\tilde{S}_j - S_j|.$$

Thus by (C.37) (similar as in (C.36))

$$\sum_{k\in\mathbb{N}} \|\partial_k S_{[k]} - \partial_k \tilde{S}_{[k]}\|_{L^\infty(U_k)}$$
$$\leq (1+K+C)^2 \sum_{k\in\mathbb{N}} \left(K b_k \|\tilde{S}_k - S_k\|_{L^\infty(U_k)} + \sum_{j=1}^{k-1} K b_j b_k \|\tilde{S}_k - S_k\|_{L^\infty(U_k)} \right)$$
$$+ (1+K+C)^2 \sum_{k\in\mathbb{N}} \|\partial_k T_k - \partial_k \tilde{T}_k\|_{L^\infty(U_k)}$$
$$\leq (1+K+C)^2 K \max_{j\in\mathbb{N}} b_j \left(1 + \sum_{i\in\mathbb{N}} b_i\right) \sum_{k\in\mathbb{N}} \|\tilde{S}_k - S_k\|_{L^\infty(U_k)}$$
$$+ (1+K+C)^2 \sum_{k\in\mathbb{N}} \|\partial_k T_k - \partial_k \tilde{T}_k\|_{L^\infty(U_k)}.$$

Applying (5.4a) and (5.3b) shows (5.4b) and concludes the proof. □

D Proofs of Sect. 6

D.1 Theorem 6.1

Lemma D.1 *Let* $(a_j)_{j\in\mathbb{N}}, (b_j)_{j\in\mathbb{N}} \subseteq (0,\infty)$ *be such that* $\lim_{n\to\infty} \sum_{j=1}^n \log(a_j) \in \mathbb{R}$ *exists and* $\sum_{j\in\mathbb{N}} |a_j - b_j| < \infty$. *Then with* $a_{\min} = \min_{j\in\mathbb{N}} a_j > 0$, $b_{\min} = \min_{j\in\mathbb{N}} b_j > 0$ *and*

$$C := \frac{\exp\left(\sum_{j\in\mathbb{N}} \frac{|a_j - b_j|}{a_{\min}}\right) \lim_{n\to\infty} \prod_{j=1}^n a_j}{\min\{a_{\min}, b_{\min}\}} < \infty$$

the limit $\lim_{n\to\infty} \prod_{j=1}^n b_j \in \mathbb{R}$ *exists and it holds*

$$\left| \lim_{n\to\infty} \prod_{j=1}^n a_j - \lim_{n\to\infty} \prod_{j=1}^n b_j \right| \leq C \sum_{j\in\mathbb{N}} |a_j - b_j|. \tag{D.1}$$

Proof For $a > 0$, $\log : [a, \infty) \to \mathbb{R}$ has Lipschitz constant $\frac{1}{a}$. Thus

$$|\log(a_j) - \log(b_j)| \le \frac{|a_j - b_j|}{\min\{a_j, b_j\}} \qquad \forall j \in \mathbb{N}. \tag{D.2}$$

For $a > 0$, $\exp : (-\infty, a] \to \mathbb{R}$ has Lipschitz constant $\exp(a)$. Thus, since $a_j + |a_j - b_j| \ge \max\{a_j, b_j\}$,

$$\begin{aligned} \left|\prod_{j=1}^n a_j - \prod_{j=1}^n b_j\right| &= \exp\left(\sum_{j=1}^n \log(a_j)\right) - \exp\left(\sum_{j=1}^n \log(b_j)\right) \\ &\le \exp\left(\sum_{j=1}^n \log(a_j + |a_j - b_j|)\right) \sum_{j=1}^n \frac{|a_j - b_j|}{\min\{a_j, b_j\}}. \end{aligned} \tag{D.3}$$

Since $\lim_{n\to\infty} \sum_{j=1}^n \log(a_j) \in \mathbb{R}$, it must hold $\log(a_j) \to 0$ and $a_j \to 1$ as $j \to \infty$. Hence $a_{\min} := \min_{j\in\mathbb{N}} a_j > 0$. Using $\log(1 + x) \le x$ for $x \ge 0$ so that

$$\log(a_j + |a_j - b_j|) = \log\left(a_j\left(1 + \frac{|a_j - b_j|}{a_j}\right)\right) \le \log(a_j) + \frac{|a_j - b_j|}{a_{\min}}$$

we get

$$\begin{aligned} \lim_{n\to\infty} \exp\left(\sum_{j=1}^n \log(a_j + |a_j - b_j|)\right) &\le \lim_{n\to\infty} \exp\left(\sum_{j=1}^n \left(\log(a_j) + \frac{|a_j - b_j|}{a_{\min}}\right)\right) \\ &= \exp\left(\sum_{j\in\mathbb{N}} \frac{|a_j - b_j|}{a_{\min}}\right) \lim_{n\to\infty} \prod_{j=1}^n a_j < \infty. \end{aligned}$$

Equation (D.1) follows by taking the limit $n \to \infty$ in (D.3). □

Lemma D.2 *Let $T, \tilde{T}_\varepsilon : U \to U$, $N_\varepsilon \in \mathbb{N}_0$ and $p \in (0, 1)$ be as in Theorem 5.2 and let $S := T^{-1}$ and $\tilde{S}_\varepsilon := \tilde{T}_\varepsilon^{-1}$. Then there exists C such that for all $\varepsilon > 0$*

$$\sup_{\boldsymbol{y}\in U} \left|\lim_{n\to\infty} \prod_{j=1}^n \partial_j S_j(\boldsymbol{y}_{[j]}) - \lim_{n\to\infty} \prod_{j=1}^n \partial_j \tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]})\right| \le C N_\varepsilon^{-\frac{1}{p}+1}. \tag{D.4}$$

Proof If $N_\varepsilon = 0$ then (D.4) is trivial. As in Step 2 of the proof of Corollary 5.4, one shows that for any $k \in \mathbb{N}$ and $\varepsilon > 0$ so small that $N_\varepsilon \ge 1$ we have

$$\inf_{\boldsymbol{y}\in U} \min\left\{\partial_k S_k(\boldsymbol{y}_{[k]}), \partial_k \tilde{S}_{\varepsilon,k}(\boldsymbol{y}_{[k]})\right\} \ge \frac{1}{\bar{C}} \tag{D.5}$$

for a constant $\bar{C} < \infty$ independent of k and ε.

By Lemma D.1

$$\sup_{\boldsymbol{y}\in U}\left|\lim_{n\to\infty}\prod_{j=1}^{n}\partial_j S_j(\boldsymbol{y}_{[j]})-\lim_{n\to\infty}\prod_{j=1}^{n}\partial_j \tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]})\right|\le C_\varepsilon\sum_{j\in\mathbb{N}}\|\partial_j S_j-\partial_j\tilde{S}_{\varepsilon,j}\|_{L^\infty(U_j)}$$

with

$$C_\varepsilon=\bar{C}\exp\left(\bar{C}\sum_{j\in\mathbb{N}}\|\partial_j S_j-\partial_j\tilde{S}_{\varepsilon,j}\|_{L^\infty(U_j)}\right)\sup_{\boldsymbol{y}\in U}\lim_{n\to\infty}\partial_j S_j(\boldsymbol{y}).$$

By (D.5) and using Corollary 5.4, we conclude that C_ε is uniformly bounded for all $\varepsilon>0$ so small that $N_\varepsilon\ge 1$. Thus it holds (D.4). □

Proof of Theorem 6.1 Throughout we denote $\tilde{S}_\varepsilon=(\tilde{S}_{\varepsilon,j})_{j\in\mathbb{N}}:=\tilde{T}_\varepsilon^{-1}:U\to U$.

Step 1. By Theorem 3.3 (cp. Remark 3.4), $\det dS(\boldsymbol{y}):=\lim_{n\to\infty}\prod_{j=1}^{n}\partial_j S_j(\boldsymbol{y}_{[j]})\in C^0(U,\mathbb{R})$ exists and (cp. Assumption 2.2)

$$\frac{\mathrm{d}\pi}{\mathrm{d}\mu}(\boldsymbol{y})=f_\pi(\boldsymbol{y})=\det dS(\boldsymbol{y})f_\rho(S(\boldsymbol{y}))$$
$$=\det dS(\boldsymbol{y})\mathfrak{f}_\rho\left(\sum_{j\in\mathbb{N}}S_j(\boldsymbol{y}_{[j]})\psi_{\rho,j}\right)\qquad\forall\boldsymbol{y}\in U.\tag{D.6}$$

Next we claim

$$\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}(\boldsymbol{y})=\det d\tilde{S}_\varepsilon(\boldsymbol{y})f_\rho(\tilde{S}_\varepsilon(\boldsymbol{y}))$$
$$=\det d\tilde{S}_\varepsilon(\boldsymbol{y})\mathfrak{f}_\rho\left(\sum_{j\in\mathbb{N}}\tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]})\psi_{\rho,j}\right)\qquad\forall\boldsymbol{y}\in U.\tag{D.7}$$

By Remark 5.3, there exists $k_0\in\mathbb{N}$ such that $\tilde{T}_{\varepsilon,k}(\boldsymbol{y}_{[k]})=x_k$ for all $k\ge k_0$, and thus

$$\tilde{S}_{\varepsilon,k}(\boldsymbol{y}_{[k]})=x_k\qquad\forall k\ge k_0.\tag{D.8}$$

Fix $n_0\ge k_0$ and let $A\subseteq U$ be measurable and of the type $A=\times_{j=1}^{n_0}A_j\times U$ with $A_j\subseteq U_1$. To show (D.7), e.g., by [3,Theorem 3.5.1], it suffices to show

$$(\tilde{T}_\varepsilon)_\sharp\rho(A)=\int_A\det d\tilde{S}_\varepsilon(\boldsymbol{y})f_\rho(\tilde{S}_\varepsilon(\boldsymbol{y}))\,\mathrm{d}\mu(\boldsymbol{y}),\tag{D.9}$$

since these sets form an algebra that generate the σ-algebra on U. For any such A

$$(\tilde{T}_\varepsilon)_\sharp\rho(A) = \rho(\{\boldsymbol{y} \in U \,:\, \tilde{T}_\varepsilon(\boldsymbol{y}) \in A\}) = \rho(\tilde{S}_\varepsilon(A)) = \int_{\tilde{S}_\varepsilon(A)} f_\rho(\boldsymbol{y}) \,\mathrm{d}\mu(\boldsymbol{y}).$$

By (D.8) we have $\tilde{S}_\varepsilon(A) = \tilde{S}_{\varepsilon,[n_0]}(\times_{j=1}^{n_0} A_j) \times U$ (here $\tilde{S}_{\varepsilon,[n_0]} = (\tilde{S}_{\varepsilon,j})_{j=1}^{n_0} : U_j \to U_j$) and thus with $\boldsymbol{y}_{[n_0+1:]} := (y_j)_{j>n_0}$

$$\begin{aligned}(\tilde{T}_\varepsilon)_\sharp\rho(A) &= \int_U \int_{\tilde{S}_{\varepsilon,[n_0]}(\times_{j=1}^{n_0} A_j)} f_\rho(\boldsymbol{y}_{[n_0]}, \boldsymbol{y}_{[n_0+1:]}) \,\mathrm{d}\mu(\boldsymbol{y}_{[n_0]}) \,\mathrm{d}\mu(\boldsymbol{y}_{[n_0+1:]}) \\ &= \int_U \int_{\times_{j=1}^{n_0} A_j} f_\rho(\tilde{S}_{\varepsilon,[n_0]}(\boldsymbol{y}_{[n_0]}), \boldsymbol{y}_{[n_0+1:]}) \\ &\quad \det d\tilde{S}_{\varepsilon,[n_0]}(\boldsymbol{y}_{[n_0]}) \,\mathrm{d}\mu(\boldsymbol{y}_{[n_0]}) \,\mathrm{d}\mu(\boldsymbol{y}_{[n_0+1:]}). \end{aligned} \tag{D.10}$$

Again by (D.8) we have

$$\begin{aligned}\det d\tilde{S}_\varepsilon(\boldsymbol{y}) := \lim_{m\to\infty} \prod_{j=1}^{m} \partial_j \tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]}) &= \prod_{j=1}^{k_0} \partial_j \tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]}) \\ &= \prod_{j=1}^{n_0} \partial_j \tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]}) = \det d\tilde{S}_{\varepsilon,[n_0]}(\boldsymbol{y}_{[n_0]}).\end{aligned}$$

Since $(\tilde{S}_{\varepsilon,[n_0]}(\boldsymbol{y}_{[n_0]}), \boldsymbol{y}_{[n_0+1:]}) = \tilde{S}_\varepsilon(\boldsymbol{y})$, (D.10) shows (D.9).

Step 2. By Lemma D.2

$$\begin{aligned}\sup_{\boldsymbol{y}\in U} |\det dS(\boldsymbol{y}) - \det d\tilde{S}_\varepsilon(\boldsymbol{y})| &= \sup_{\boldsymbol{y}\in U} \left| \lim_{n\to\infty} \prod_{j=1}^{n} \partial_j S_j(\boldsymbol{y}_{[j]}) - \lim_{n\to\infty} \prod_{j=1}^{n} \partial_j \tilde{S}_j(\boldsymbol{y}_{[j]}) \right| \\ &\le C N_\varepsilon^{-\frac{1}{p}+1}. \end{aligned} \tag{D.11}$$

Using that the differentiable function $\mathfrak{f}_\rho : O_X \to \mathbb{C}$ has some Lipschitz constant $r < \infty$ on the compact set $\{\sum_{j\in\mathbb{N}} y_j \psi_{\rho,j} \,:\, \boldsymbol{y} \in U\} \subseteq O_X \subseteq X_\mathbb{C}$ (cp. Assumption 2.1), we have for all $\boldsymbol{y} \in U$ with $b_j := \|\psi_{\rho,j}\|_X$

$$\begin{aligned}&\left| \mathfrak{f}_\rho\left(\sum_{j\in\mathbb{N}} S_j(\boldsymbol{y}_{[j]})\psi_{\rho,j}\right) - \mathfrak{f}_\rho\left(\sum_{j\in\mathbb{N}} \tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]})\psi_{\rho,j}\right)\right| \\ &\quad \le r \sum_{j\in\mathbb{N}} |S_j(\boldsymbol{y}_{[j]}) - \tilde{S}_{\varepsilon,j}(\boldsymbol{y}_{[j]})| b_j \le C N_\varepsilon^{-\frac{1}{p}+1}\end{aligned} \tag{D.12}$$

by Corollary 5.4, and for some C depending on r and $\sup_{j\in\mathbb{N}} b_j < \infty$.

Therefore, using (D.6), (D.7), (D.11), (D.12) and the triangle inequality we find

$$\sup_{y\in U}\left|f_\pi(\boldsymbol{y})-\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}(\boldsymbol{y})\right| = \sup_{y\in U}\left|\frac{\mathrm{d}\pi}{\mathrm{d}\mu}(\boldsymbol{y})-\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}(\boldsymbol{y})\right| \le CN_\varepsilon^{-\frac{1}{p}+1} \quad \text{(D.13)}$$

for some suitable constant $C < \infty$ and all $\varepsilon > 0$.

Equation (D.13) yields (6.1) for the total variation distance. Moreover

$$\sup_{y\in U}\left|\sqrt{f_\pi(\boldsymbol{y})}-\sqrt{\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}(\boldsymbol{y})}\right| \le \sup_{y\in U}\frac{\left|f_\pi(\boldsymbol{y})-\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}(\boldsymbol{y})\right|}{|\sqrt{f_\pi(\boldsymbol{y})}|} \le \frac{CN_\varepsilon^{-\frac{1}{p}+1}}{\inf_{\boldsymbol{y}\in U}\sqrt{f_\pi(\boldsymbol{y})}},$$

which gives (6.1) for the Hellinger distance since $\inf_{\boldsymbol{y}\in U} f_\pi(\boldsymbol{y}) \ge M > 0$ by Assumption 2.1.

Finally, for the KL divergence, using that $a|\log(a)-\log(b)| \le (1+\frac{|a-b|}{b})|a-b|$ for all $a, b > 0$ (see [41,Lemma E.2]), by (D.13)

$$\begin{aligned}\mathrm{KL}((\tilde{T}_\varepsilon)_\sharp\rho||\pi) &\le \int_U \left|\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}(\boldsymbol{y})\right|\left|\log\left(\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}(\boldsymbol{y})\right)-\log(f_\pi(\boldsymbol{y}))\right|\,\mathrm{d}\mu(\boldsymbol{y})\\ &\le \left(1+\frac{\|f_\pi-\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}\|_{L^\infty(U)}}{\inf_{\boldsymbol{y}\in U} f_\pi(\boldsymbol{y})}\right)\left\|f_\pi-\frac{\mathrm{d}(\tilde{T}_\varepsilon)_\sharp\rho}{\mathrm{d}\mu}\right\|_{L^\infty(U)}\\ &\le CN_\varepsilon^{-\frac{1}{p}+1}.\end{aligned}$$

□

D.2 Proposition 6.2

Proof of Proposition 6.2 Compactness of M_1 and continuity of $T : M_1 \to M_2$ and $\tilde{T} : M_1 \to M_2$ imply that $K:=T(M_1)$ and $\tilde{K}:=\tilde{T}(M_1)$ are compact subsets of M_2. As compact subsets of a metric space, they are closed and thus measurable. Note that $\mathrm{supp}(T_\sharp\nu) \subseteq K$ and $\mathrm{supp}(\tilde{T}_\sharp\nu) \subseteq \tilde{K}$. Since the continuous function $(x, y) \mapsto d_2(x, y)$ is bounded on the compact set $K \times \tilde{K} \subseteq M_2 \times M_2$, the Wasserstein distance $W_q(T_\sharp\nu, \tilde{T}_\sharp\nu)$ in (6.2) is well-defined and finite.

Fix $\varepsilon > 0$ and let $(B_\varepsilon(x_i))_{i=1}^n$ with $B_\varepsilon(x_i):=\{x \in M_1 : d_1(x, x_i) < \varepsilon\}$ be a finite cover of M_1. Such a cover exists by compactness of M. Define $I_1:=B_\varepsilon(x_1)$ and inductively set $I_j:=B_\varepsilon(x_j)\backslash\bigcup_{i=1}^{j-1} I_i$, so that $(I_j)_{j=1}^n$ is a (measurable) partition of M_1.

Denote by μ_j the measure $\mu_j(A):=\nu(T^{-1}(A)\cap I_j)$ on M_2, and by $\tilde{\mu}_j$ the measure $\tilde{\mu}_j(A):=\nu(\tilde{T}^{-1}(A)\cap I_j)$ on M_2. Then $\sum_{j=1}^n \mu_j(A) = \nu(T^{-1}(A))$ for all measurable $A \subseteq M_2$, i.e., $T_\sharp\nu = \sum_{j=1}^n \mu_j$. Similarly $\tilde{T}_\sharp\nu = \sum_{j=1}^n \tilde{\mu}_j$. Note that

$$\mu_j(M_2) = \nu(T^{-1}(M_2)\cap I_j) = \nu(I_j) = \nu(\tilde{T}^{-1}(M_2)\cap I_j) = \tilde{\mu}_j(M_2).$$

Wlog $\mu_j(M_2) = \nu(I_j) = \tilde{\mu}_j(M_2) > 0$ for all $j \in \{1, \ldots, n\}$ (otherwise we can omit $\mu_j, \tilde{\mu}_j$). Let Γ_j be the couplings between μ_j and $\tilde{\mu}_j$ (the measures on $M_2 \times M_2$ with marginals μ_j and $\tilde{\mu}_j$). Note that Γ_j is not empty since $\frac{1}{\nu(I_j)}\mu_j \otimes \tilde{\mu}_j \in \Gamma_j$. Let Γ be the couplings between $T_\sharp\nu$ and $\tilde{T}_\sharp\nu$. Then $\{\sum_{j=1}^n \gamma_j \,:\, \gamma_j \in \Gamma_j\} \subseteq \Gamma$. Thus

$$\begin{aligned} W_q(T_\sharp\nu, \tilde{T}_\sharp\nu)^q &= \inf_{\gamma\in\Gamma} \int_{M_2\times M_2} d_2(x, y)^q \,\mathrm{d}\gamma(x, y) \\ &\le \inf_{\gamma_j\in\Gamma_j} \sum_{j=1}^n \int_{M_2\times M_2} d_2(x, y)^q \,\mathrm{d}\gamma_j(x, y). \end{aligned}$$

If $A \subseteq M_2$ has empty intersection with $T(I_j) = \{T(x) \,:\, x \in I_j\} \subseteq M_2$, then $T^{-1}(A) = \{x \in M_1 \,:\, T(x) \in A\}$ has empty intersection with I_j, which implies $\mu_j(A) = \nu(T^{-1}(A) \cap I_j) = 0$. Thus $\mathrm{supp}(\mu_j) \subseteq T(I_j)$ and similarly $\mathrm{supp}(\tilde{\mu}_j) \subseteq \tilde{T}(I_j)$. Hence

$$\begin{aligned} \int_{M_2\times M_2} d_2(x, y)^q \,\mathrm{d}\gamma_j(x, y) &= \int_{T(I_j)\times\tilde{T}(I_j)} d_2(x, y)^q \,\mathrm{d}\gamma_j(x, y) \\ &\le \mu_j(T(I_j)) \sup_{(x,y)\in T(I_j)\times\tilde{T}(I_j)} d_2(x, y)^q. \end{aligned}$$

Here we used that γ_j has marginal μ_j in the first argument. In total, using $\mu_j(T(I_j)) = \nu(I_j)$,

$$W_q(T_\sharp\nu, T_\sharp\mu)^q \le \nu(M_1) \max_{j=1,\ldots,n} \sup_{(x,y)\in T(I_j)\times\tilde{T}(I_j)} d_2(x, y)^q. \tag{D.14}$$

To bound the supremum we first note that continuity of $T : (M_1, d_1) \to (M_2, d_2)$ and compactness of M_1 imply uniform continuity by the Heine–Cantor theorem, i.e.,

$$\lim_{\delta\to 0} \sup_{x\in M_1} \sup_{y\in B_\delta(x)} d_2(T(x), T(y)) = 0 \tag{D.15}$$

and the same holds for $\tilde{T}$.

Then, with $I_j \subseteq B_\varepsilon(x_j)$ as constructed in the beginning, using (D.15) for T and $\tilde{T}$,

$$\begin{aligned} &\sup_{(x,y)\in T(I_j)\times\tilde{T}(I_j)} d_2(x, y) = \sup_{x,y\in I_j} d_2(T(x), \tilde{T}(y)) \\ &\le \sup_{x,y\in I_j} \Big(d_2(T(x), T(x_j)) + d_2(T(x_j), \tilde{T}(x_j)) + d_2(\tilde{T}(x_j), \tilde{T}(y)) \Big) \\ &\le \sup_{x\in M_1} \sup_{y\in B_\varepsilon(x)} d_2(T(x), T(y)) + \sup_{x\in M_1} d_2(T(x), \tilde{T}(x)) \\ &\quad + \sup_{x\in M_1} \sup_{y\in B_\varepsilon(x)} d_2(\tilde{T}(x), \tilde{T}(y)) = \sup_{x\in M_1} d_2(T(x), \tilde{T}(x)) + o(1) \qquad \text{as } \varepsilon \to 0. \end{aligned}$$

Together with (D.14), and because $\nu(M_1) = 1$ since ν is a probability measure, this concludes the proof. □

D.3 Lemma 6.3

Proof of Lemma 6.3 By Lemma A.1 d induces the product topology on U (independent of the choice of positive and summable sequence $(c_j)_{j\in\mathbb{N}}$ in (6.3)). Thus, to prove the lemma, it suffices to check Lipschitz continuity in case $b_j \le Cc_j$.

We begin with $\tilde{T}$. By Remark 5.3 there exists $k_0 \in \mathbb{N}$ such that $T_k(\boldsymbol{y}) = y_k$ for all $\boldsymbol{y} \in U_k$ and all $k \ge k_0$. By construction each $\tilde{T}_k : U_k \to U_1$ is a rational function with positive denominator (in particular C^∞) and thus there exists $L > 0$ such that L is a Lipschitz constant of $\tilde{T}_k : U_k \to U_1$ for all $k \in \{1, \dots, k_0 - 1\}$ w.r.t. the Euclidean norm $\|\cdot\|$. Thus for all $\boldsymbol{x}, \boldsymbol{y} \in U$

$$
\begin{aligned}
d(\tilde{T}(\boldsymbol{x}), \tilde{T}(\boldsymbol{y})) &= \sum_{k<k_0} c_k |\tilde{T}_k(\boldsymbol{x}_{[k]}) - \tilde{T}_k(\boldsymbol{y}_{[k]})| + \sum_{k\ge k_0} c_k |\tilde{T}_k(\boldsymbol{x}_{[k]}) - \tilde{T}_k(\boldsymbol{y}_{[k]})| \\
&\le \sum_{k<k_0} c_k L \|\boldsymbol{x}_{[k]} - \boldsymbol{y}_{[k]}\| + \sum_{k\ge k_0} c_k |x_k - y_k| \\
&\le \sum_{k<k_0} c_k L \sum_{j=1}^{k} |x_j - y_j| + \sum_{k\ge k_0} c_k |x_k - y_k| \\
&\le C_0 \sum_{k\in\mathbb{N}} c_k |x_k - y_k|,
\end{aligned}
$$

where $C_0 := 1 + L \sum_{k=1}^{k_0-1} c_k$.

The argument for T is similar as in the proof of Corollary 5.4. By (C.33) for all $\boldsymbol{x}$, $\boldsymbol{y} \in U$ and all $k \in \mathbb{N}$

$$
\begin{aligned}
|T_k(\boldsymbol{x}_{[k]}) - T_k(\boldsymbol{y}_{[k]})| &\le \sum_{j=1}^{k} |T_k(\boldsymbol{x}_{[j]}, \boldsymbol{y}_{[j+1:k]}) - T_k(\boldsymbol{x}_{[j-1]}, \boldsymbol{y}_{[j:k]})| \\
&\le (1 + K)|x_k - y_k| + \sum_{j=1}^{k-1} 2K b_k b_j |x_j - y_j|
\end{aligned}
$$

and thus since $b_j \le Cc_j$

$$
\begin{aligned}
d(T(\boldsymbol{x}), T(\boldsymbol{y})) &= \sum_{k\in\mathbb{N}} c_k |T_k(\boldsymbol{x}_{[k]}) - T_k(\boldsymbol{y}_{[k]})| \\
&\le \sum_{k\in\mathbb{N}} (1 + K) c_k |x_k - y_k| + \sum_{k\in\mathbb{N}} \sum_{j=1}^{k-1} 2K b_k b_j c_k |x_j - y_j| \\
&= (1 + K) d(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j\in\mathbb{N}} b_j |x_j - y_j| \sum_{k>j} 2K b_k c_k
\end{aligned}
$$

$$\leq (1+K)d(\boldsymbol{x},\boldsymbol{y}) + C\sum_{j\in\mathbb{N}} c_j|x_j-y_j|\sum_{k\in\mathbb{N}} 2Kb_kc_k$$
$$= \left(1+K+2CK\sum_{k\in\mathbb{N}} b_kc_k\right)d(\boldsymbol{x},\boldsymbol{y}).$$

□

D.4 Corollary 6.5

Proof of Corollary 6.5 Fix $\varepsilon > 0$. Set $H : U \to U$ via $H_j := \tilde{T}_{\varepsilon,j}$ if $j \leq N_\varepsilon$ and $H_j(\boldsymbol{y}) := 0$ for $j > N_\varepsilon$. Then $\Phi(H(\boldsymbol{y})) = \Phi_{N_\varepsilon}(\tilde{T}_{\varepsilon,[N_\varepsilon]}(\boldsymbol{y}_{[N_\varepsilon]}))$ for all $\boldsymbol{y} \in U$. Thus $(\Phi\circ H)_\sharp\rho = (\Phi_{N_\varepsilon}\circ\tilde{T}_{\varepsilon,[N_\varepsilon]})_\sharp\rho_{N_\varepsilon}$, and it suffices to bound the difference between $(\Phi\circ H)_\sharp\rho$ and $(\Phi\circ T)_\sharp\rho = \Phi_\sharp\pi$. To this end we compute similar as in (6.6) for all $\boldsymbol{y}\in U$ with b_j in (6.4)

$$\begin{aligned}
&\|\Phi(T(\boldsymbol{y})) - \Phi(H(\boldsymbol{y}))\|_Y \\
&= \left\| \sum_{j=1}^{N_\varepsilon} (T_j(\boldsymbol{y}_{[j]}) - \tilde{T}_{\varepsilon,j}(\boldsymbol{y}_{[j]}))\psi_{\pi,j} + \sum_{i>N_\varepsilon} T_j(\boldsymbol{y}_{[j]})\psi_{\pi,j} \right\|_Y \\
&\leq \sum_{j=1}^{N_\varepsilon} b_j\|T_j - \tilde{T}_{\varepsilon,j}\|_{L^\infty(U_j)} + \sum_{i>N_\varepsilon} b_j.
\end{aligned} \tag{D.16}$$

In case the $(b_j)_{j\in\mathbb{N}}$ are monotonically decreasing, Stechkin's lemma, which is easily checked, states that $\sum_{i>N_\varepsilon} b_j \leq \|(b_j)_{j\in\mathbb{N}}\|_{\ell^p(\mathbb{N})} N_\varepsilon^{-\frac{1}{p}+1}$. The ℓ^p-norm is finite by Assumption 2.2. Thus by Theorem 5.2 the last term in (D.16) is bounded by $C(N_\varepsilon^{-\frac{1}{p}+1} + N_\varepsilon^{-\frac{1}{p}+1})$. An application of Proposition 6.2 yields the same bound for the Wasserstein distance. □

References

1. Alexanderian, A.: A brief note on the Karhunen–Loève expansion (2015)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In D. Precup and Y. W. Teh (eds.) Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, pp. 214–223. PMLR, 06–11 Aug (2017)
3. Bogachev, V.I.: Measure Theory, vol. I–II. Springer, Berlin (2007)
4. Bogachev, V.I., Kolesnikov, A.V., Medvedev, K.V.: Triangular transformations of measures. Mat. Sb. **196**(3), 3–30 (2005)
5. Chkifa, A.: Sparse polynomial methods in high dimension: application to parametric PDE, Ph.D. thesis, UPMC, Université Paris 06, Paris, France (2014)
6. Cohen, A., Chkifa, A., Schwab, C.: Breaking the curse of dimensionality in sparse polynomial approximation of parametric pdes. J. Math. Pures et Appliquees **103**(2), 400–428 (2015)
7. Cohen, A., DeVore, R., Schwab, C.: Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. Found. Comput. Math. **10**(6), 615–646 (2010)
8. Cohen, A., Devore, R., Schwab, C.: Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. Anal. Appl. (Singap.) **9**(1), 11–47 (2011)
9. Dashti, M., Stuart, A.M.: The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification. Vol. 1, 2, 3, pp. 311–428. Springer, Cham (2017)

10. Davis, P.: Interpolation and Approximation. Dover Books on Mathematics, Dover Publications (1975)
11. De Philippis, G., Figalli, A.: Partial regularity for optimal transport maps. Publ. Math. Inst. Hautes Études Sci. **121**, 81–112 (2015)
12. Doersch, C.: Tutorial on variational autoencoders. *arXiv preprint* arXiv:1606.05908 (2016)
13. El Moselhy, T.A., Marzouk, Y.M.: Bayesian inference with optimal maps. J. Comput. Phys. **231**(23), 7815–7850 (2012)
14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Ed.) Advances in Neural Information Processing Systems, volume 27. Curran Associates, Inc., (2014)
15. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. *arXiv preprint* arXiv:1406.2661 (2014)
16. Jaini, P., Selby, K.A., Yu, Y.: Sum-of-squares polynomial flow. ICML, (2019)
17. Jerez-Hanckes, C., Schwab, C., Zech, J.: Electromagnetic wave scattering by random surfaces: shape holomorphy. Math. Mod. Meth. Appl. Sci. **27**(12), 2229–2259 (2017)
18. Kirwan, P.: Complexifications of multilinear and polynomial mappings, 1997. Ph.D. thesis, National University of Ireland, Galway
19. Kothari, K., Khorashadizadeh, A., de Hoop, M., Dokmanić, I.: Trumpets: injective flows for inference and inverse problems (2021)
20. Krantz, S. G.: Function theory of several complex variables. AMS Chelsea Publishing, Providence (Reprint of the 1992 edition 2001)
21. Kumar, A., Poole, B., Murphy, K.: Regularized autoencoders via relaxed injective probability flow (2020)
22. Lord, G.J., Powell, C.E., Shardlow, T.: An Introduction to Computational Stochastic PDEs. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York (2014)
23. Marzouk, Y., Moselhy, T., Parno, M., Spantini, A.: Sampling via measure transport: an introduction. In: Handbook of Uncertainty Quantification. Vol. 1, 2, 3, pp. 785–825. Springer, Cham (2017)
24. Mattner, L.: Complex differentiation under the integral. Nieuw Arch. Wiskd. (5), **2**(1), 32–35 (2001)
25. McLaughlin, D., Townley, L.R.: A reassessment of the groundwater inverse problem. Water Resour. Res. **32**(5), 1131–1161 (1996)
26. Munkres, J.R.: Topology. Prentice Hall Inc, Upper Saddle River (2000)
27. Muñoz, G.A., Sarantopoulos, Y., Tonge, A.: Complexifications of real Banach spaces, polynomials and multilinear maps. Studia Math. **134**(1), 1–33 (1999)
28. Olver, F.W.J., Lozier, D.W., Boisvert, R.F., Clark, C.W. editors. NIST handbook of mathematical functions. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge (2010)
29. Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S., Lakshminarayanan, B.: Normalizing flows for probabilistic modeling and inference. J. Mach. Learn. Res. **22**, 1–64 (2021)
30. Rezende, D., Mohamed, S.: Variational inference with normalizing flows. In: F. Bach and D. Blei (Eds.) Proceedings of the 32nd International Conference on Machine Learning, volume 37 of Proceedings of Machine Learning Research, pp. 1530–1538, Lille, France, 07–09 Jul 2015. PMLR
31. Robert, C.P., Casella, G.: Monte Carlo Statistical Methods (Springer Texts in Statistics). Springer, Berlin, Heidelberg (2005)
32. Sagiv, A.: The Wasserstein distances between pushed-forward measures with applications to uncertainty quantification. Commun. Math. Sci. **18**(3), 707–724 (2020)
33. Santambrogio, F.: Optimal Transport for Applied Mathematicians, Volume 87 of Progress in Nonlinear Differential Equations and their Applications. Birkhäuser/Springer, Cham, Calculus of variations, PDEs, and modeling (2015)
34. Schwab, C., Stuart, A.M.: Sparse deterministic approximation of Bayesian inverse problems. Inverse Problems, **28**(4), 045003 (2012)
35. Spantini, A., Bigoni, D., Marzouk, Y.: Inference via low-dimensional couplings. J. Mach. Learn. Res. **19**(1), 2639–2709 (2018)
36. Telgarsky, M.: Neural networks and rational functions. In: D. Precup and Y. W. Teh (Eds.) Proceedings of the 34th International Conference on Machine Learning, Volume 70 of Proceedings of Machine Learning Research, pp. 3387–3393. PMLR, 06–11 Aug (2017)
37. Villani, C.: Optimal Transport, Volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Berlin, Old and new (2009)

38. Wehenkel, A., Louppe, G.: Unconstrained monotonic neural networks. arXiv preprint arXiv:1908.05164 (2019)
39. Yarotsky, D.: Error bounds for approximations with deep ReLU networks. Neural Netw. **94**, 103–114 (2017)
40. Zech, J.: Sparse-Grid Approximation of High-Dimensional Parametric PDEs, Dissertation 25683, ETH Zürich. https://doi.org/10.3929/ethz-b-000340651 (2018)
41. Zech, J., Marzouk, Y.: Sparse approximation of triangular transports. Part I: the finite dimensional case. Constr. Approx. https://doi.org/10.1007/s00365-022-09569-2 (2022)
42. Zech, J., Schwab, C.: Convergence rates of high dimensional Smolyak quadrature. ESAIM Math. Model. Numer. Anal. **54**(4), 1259–1307 (2020)