

APPROXIMATION IN METRIC SOBOLEV SPACES: A GENERAL FRAMEWORK

MASSIMO FORNASIER AND GIACOMO ENRICO SODINI

ABSTRACT. In our recent work [17], we introduced a numerical framework for approximating Sobolev functions on Wasserstein spaces from finite samples, leveraging structural properties established in [18]. The present paper demonstrates that this methodology extends far beyond that specific setting. We identify a general class of metric measure spaces—including weighted Riemannian manifolds and spaces of measures equipped with the Hellinger–Kantorovich distance—for which the key hypotheses of Hilbertianity and the existence of a computable algebra of Lipschitz functions hold. Within this abstract framework, we recover and generalize the core approximation results of [17] for recovering functions from random point evaluations. Our main contribution is to show that the combination of theoretical foundations from [18] and algorithmic strategies from [17] is robust enough to apply to a wide variety of infinite-dimensional spaces of current interest.

CONTENTS

1	Introduction	1
2	The relaxation approach to Metric Sobolev spaces	3
2.1	Infinitesimal Hilbertianity	4
3	Regularized least squares approximation of bounded functions	7
3.1	Direct pre-Cheeger approach	7
3.2	Approximated Cheeger energy approach	14
	Acknowledgments	17
	References	18

1. INTRODUCTION

The primary goal of this work is to advance the program of bringing abstract functional analysis and, more specifically, the theory of metric measure spaces into the realm of practical computability. This is motivated by the growing need for robust mathematical foundations for data analysis and machine learning in settings where data is intrinsically infinite-dimensional and may come with their own intrinsic metric (e.g., distances between images or shapes). In our previous work [17], we showed the feasibility and power of this approach by developing a comprehensive framework for the numerical approximation of Sobolev-smooth functions defined on the Wasserstein space of probability measures. A central and successful application was the efficient approximation of the Wasserstein distance function itself, a task of significant importance in fields such as image comparison, retrieval, and recognition, where our method offered substantial computational speedups after training.

2020 *Mathematics Subject Classification.* Primary: 46E36, 49Q22; Secondary: 28A33, 49M25, 65D15.

Key words and phrases. Metric measure spaces, Sobolev spaces, Cheeger energy, optimal transport, Wasserstein distance, Hellinger–Kantorovich distance, empirical risk minimization, approximation theory, Tikhonov regularization.

The success of the constructive methods in [17] was critically based on the deep structural properties of the underlying metric measure space, which were established in our preceding theoretical work [18]. A key finding was that, for any finite reference measure $\mathbf{m}$ on the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, the associated Cheeger energy is a quadratic form, making the Sobolev space $H^{1,2}$ a Hilbert space. At this introductory level, let us just mention that the Cheeger energy serves as a surrogate of the integral norm of the gradient in the context of metric spaces, see (2.4) for the precise definition.

This Hilbertianity, combined with the explicit computability of the pre-Cheeger energy (see (2.2)) on a sufficiently large and practically implementable algebra of Lipschitz functions (cylinder functions), provided the essential tool for the approximation algorithms in [17]. A crucial step in those algorithms involved going from a continuous (or diffused) probability measure $\mathbf{p}$ on the Wasserstein space to its discrete (or empirical) approximations $\mathbf{p}_N$, drawn from random samples. This type of “discrete-to-continuous” transition is a well-established theme in approximation theory and numerical analysis, often explored to develop methods that are both theoretically sound and computationally viable. Notably, the work of [1] shares a similar spirit, investigating approximation problems in metric measure spaces using a discretization of the underlying space. However, a cornerstone of their analysis is the metric doubling property of the underlying space. Under such condition one can define partitions of the space with finitely many neighboring sets, whose number is uniform with respect to the resolution of the partition. Then a discrete Cheeger energy can be defined in terms of finite differences of integral evaluations over neighboring sets. As the resolution of the partition increases, [1, Theorem 7.4] proves that such discrete Cheeger energy tends – in the sense of Γ -convergence – to become equivalent, up to multiplicative constants, to the Cheeger energy. While the doubling condition is powerful, it also means that the metric space is intrinsically finite-dimensional, corresponding to the finitely many neighboring sets of partitions. Therefore the approximation proposed [1] cannot be applied in a genuinely infinite-dimensional metric space, such as the Wasserstein space, that was our focus in [17, 18], nor in many other infinite-dimensional metric spaces of interest in modern application areas such as machine learning.

This observation naturally leads to the central question of the present paper: can the methodology pioneered in [17] be abstracted and generalized beyond the specific case of the Wasserstein space to a broader class of metric measure spaces? The answer, as we will show, is affirmative.

We begin by formally introducing a class of metric measure spaces for which the key hypotheses enabling our earlier work hold. These hypotheses, which we clarify and discuss in detail, include the Hilbertianity of the pre-Cheeger energy and the existence of a sufficiently rich algebra of Lipschitz functions for which the pre-Cheeger energy is computable. We then present a series of relevant and non-trivial examples that fit this framework. Beyond the already-treated Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, these include weighted Riemannian manifolds and spaces of measures endowed with the Hellinger–Kantorovich distance. This latter example shows the applicability of our framework to other optimal transport geometries.

Within this general setting, we then show how to recover and extend some of the core approximation results from [17]. Specifically, we focus on the problem of approximating an unknown function from its values known only at a finite set of random points, distributed according to the underlying reference measure. We tackle this problem in two ways. First, we approach the approximation by working at the level of the pre-Cheeger energy. This is advantageous in concrete applications, as this form is often more amenable to direct computation and numerical implementation, for instance, via neural networks as demonstrated in [17]. Second, by introducing an additional approximation parameter to handle the squared field representing the Cheeger energy, we demonstrate that one can work directly at the level of the Cheeger energy itself.

This paper is primarily intended as a review and synthesis, aiming to demonstrate the robustness and generality of the techniques developed in [17]. While the results are not entirely surprising

given the foundational work in [18] and the algorithmic framework in [17], the primary contribution lies in showing that this powerful computational-analytical framework can be coherently and meaningfully extended to encompass a much wider class of infinite-dimensional metric measure spaces, including several of significant interest in modern applications.

Finally, this work is situated within the broader, and still partly unexplored, context of analysis in infinite-dimensional spaces, and in particular in spaces of measures. By demonstrating the robustness and applicability of a computational-analytical framework beyond the specific case of the Wasserstein space, we hope to highlight the richness of this field and encourage further investigation. Many fundamental questions, such as the existence of compact Sobolev embeddings, the validity of Poincaré inequalities, and the analysis of intrinsic distances induced by Dirichlet forms and their associated stochastic processes, remain interesting and open questions for future research in this infinite-dimensional setting.

2. THE RELAXATION APPROACH TO METRIC SOBOLEV SPACES

Several approaches to defining Sobolev spaces on metric-measure spaces have been studied in recent years. Here, a metric-measure space is a complete and separable metric space $(\mathcal{S}, d)$ endowed with a non-negative, finite Borel measure $\mathbf{m}$ on $(\mathcal{S}, d)$. This family of spaces includes, among others, weighted Banach spaces, Riemannian and Finsler manifolds, RCD spaces, Carnot groups, and Alexandrov spaces. At least four different approaches exist for defining a Sobolev space on such spaces:

- (i) The Newtonian approach, due to Shanmugalingam [29] after Koskela-MacManus [22];
- (ii) The Beppo-Levi approach, due to Ambrosio, Gigli, and Savaré [3, 4], later revisited by Savaré in [27];
- (iii) The derivation approach, due to Di Marino [15];
- (iv) The relaxation approach, originally proposed by Cheeger [8], and later revisited by Ambrosio, Gigli, and Savaré in [3, 4].

Under suitable assumptions, the above approaches are equivalent and give rise to the same object: the metric Sobolev space $H^{1,2}(\mathcal{S}, d, \mathbf{m})$, a Banach space of L^2 functions that are differentiable in a suitable sense. We cite the two monographs [20, 7] for an overview on the matter and, in particular, refer to [5, 6] for a comparison of the mentioned approaches. We also note that our exposition is limited to the exponent $p = 2$ in the Sobolev space; analogous constructions could be carried out for $p \neq 2$, as in the classical theory of Sobolev spaces on Euclidean and Riemannian settings.

Here we focus on the fourth approach, which is perhaps the easiest to present from scratch and has already been employed in the same spirit as the present paper in the work [17], which we take inspiration from for most of our constructions.

We briefly introduce a few relevant objects: for a d -Lipschitz and bounded function $f \in \text{Lip}_b(\mathcal{S}, d)$ we define its *d-Lipschitz asymptotic constant* as

$$\text{lip}_d h(x) := \lim_{r \downarrow 0} \text{Lip}(f, d, B_r(x)), \quad x \in \mathcal{S}, \quad (2.1)$$

where $B_r(x) := \{y \in \mathcal{S} : d(x, y) < r\}$ is the d -open ball of radius $r > 0$ and, for any given $A \subset \mathcal{S}$, the quantity $\text{Lip}(f, d, A)$ is the d -Lipschitz constant of f in A defined as

$$\text{Lip}(f, d, A) := \sup \left\{ \frac{|f(y) - f(z)|}{d(y, z)} : y, z \in A, y \neq z \right\}.$$

It is clear that $\text{lip}_d f$ serves as a surrogate for the norm of the gradient: in fact, for a smooth function $f \in C^\infty(\mathbb{R}^d)$ it holds

$$\text{lip}_{d_e} f(x) = |\nabla f(x)|,$$

where d_e is the distance induced by the Euclidean norm $|\cdot|$ in $\mathbb{R}^d$. Clearly lip_d is a pointwise and purely metric object: its integrated counterpart is the so-called *pre-Cheeger energy*:

$$\text{pCE}_{2,\mathbf{m}}(f) := \int_{\mathcal{S}} (\text{lip}_d f)^2 d\mathbf{m}, \quad f \in \text{Lip}_b(\mathcal{S}, d). \quad (2.2)$$

Note that pCE also depends on the underlying distance d , but we are not going to stress this dependence, which will always be clear from the context. At this point, pCE is defined only for d -Lipschitz and bounded functions, and there is no reason for it to be lower semicontinuous w.r.t. the L^2 topology. Therefore, we introduce its L^2 relaxation, the *Cheeger energy*, given by

$$\text{CE}_{2,\mathbf{m}}(f) = \inf \left\{ \liminf_{n \rightarrow \infty} \text{pCE}_{2,\mathbf{m}}(f_n) : (f_n)_n \subset \text{Lip}_b(\mathcal{S}, d), f_n \rightarrow f \text{ in } L^2(\mathcal{S}, \mathbf{m}) \right\}. \quad (2.3)$$

The resulting Sobolev space $H^{1,2}(\mathcal{S}, d, \mathbf{m})$ is then the vector space of functions $f \in L^2(\mathcal{S}, \mathbf{m})$ with finite Cheeger energy endowed with the norm

$$\|f\|_{H^{1,2}(\mathcal{S}, d, \mathbf{m})}^2 := \int_{\mathcal{S}} |f|^2 d\mathbf{m} + \text{CE}_{2,\mathbf{m}}(f), \quad (2.4)$$

which makes it a Banach space (cf. [19, Theorem 2.1.17]). Note that, for every $f \in \text{Lip}_b(\mathcal{S}, d)$, we have that $\text{lip}_d f \leq \text{Lip}(f, d, \mathcal{S})$ so that $\text{Lip}_b(\mathcal{S}, d) \subset H^{1,2}(\mathcal{S}, d, \mathbf{m})$.

2.1. Infinitesimal Hilbertianity

In general, $H^{1,2}(\mathcal{S}, d, \mathbf{m})$ is not a Hilbert space. For instance, for $\mathcal{S} = \mathbb{R}^d$, $d(x, y) := |x - y|_\infty$, $x, y \in \mathbb{R}^d$, is the distance induced by the infinity norm, and $\mathbf{m}$ is a Gaussian measure on $\mathbb{R}^d$, this is indeed not a Hilbert space. When the Hilbertian property is satisfied, we say that the metric-measure space $(\mathcal{S}, d, \mathbf{m})$ is *infinitesimally Hilbertian*, a crucial property which is the starting point of the theory of RCD spaces. When a (complete and separable) metric space $(\mathcal{S}, d)$ is such that $(\mathcal{S}, d, \mathbf{m})$ is infinitesimally Hilbertian for every non-negative, finite, Borel measure $\mathbf{m}$ on $(\mathcal{S}, d)$, then we say that $(\mathcal{S}, d)$ is *universally infinitesimally Hilbertian*. Weighted Riemannian manifolds, weighted Banach spaces, and other non-trivial spaces fall into this category, as we will see in the following.

It has been shown in [18] that a sufficient condition for a metric-measure space $(\mathcal{S}, d, \mathbf{m})$ to be infinitesimally Hilbertian is given by the existence of a unital subalgebra $\mathcal{A} \subset \text{Lip}_b(\mathcal{S}, d)$ (that is, $\mathcal{A}$ is a subalgebra of $\text{Lip}_b(\mathcal{S}, d)$ and contains constant functions) with the two following properties:

- (1) $\mathcal{A}$ is strongly dense in the metric Sobolev space: for every $f \in H^{1,2}(\mathcal{S}, d, \mathbf{m})$ there exists a sequence $(f_n)_n \subset \mathcal{A}$ such that $f_n \rightarrow f$ in $H^{1,2}(\mathcal{S}, d, \mathbf{m})$;
- (2) $(\text{pCE}_{2,\mathbf{m}}, \mathcal{A})$ is quadratic, equivalently:

$$\text{pCE}_{2,\mathbf{m}}(f + h) + \text{pCE}_{2,\mathbf{m}}(f - h) = 2\text{pCE}_{2,\mathbf{m}}(f) + 2\text{pCE}_{2,\mathbf{m}}(h) \quad \text{for every } f, h \in \mathcal{A}.$$

Note that condition (2) implies, by polarization, that pCE is a quadratic form. On the other hand, condition (1) also implies that we can approximate the Cheeger energy of f via the pre-Cheeger energy of the sequence $(f_n)_n$, that is,

$$\lim_{n \rightarrow \infty} \text{pCE}_{2,\mathbf{m}}(f_n) = \text{CE}_{2,\mathbf{m}}(f).$$

In general, the Cheeger energy can be difficult to compute or is defined only implicitly. In contrast, the pre-Cheeger energy often takes a more explicit form, for instance, when the following condition (which is stronger than (2)) holds:

- (2') there exists a symmetric bilinear map $\Gamma : \mathcal{A} \times \mathcal{A} \rightarrow L^1(\mathcal{S}, \mathbf{m})$ such that

$$(\text{lip}_d f)^2 = \Gamma(f, f) \quad \text{for every } f \in \mathcal{A}.$$

We now list several examples where conditions (1) and (2') (and consequently condition (2)) are satisfied. In all the following examples, the metric space under consideration is universally infinitesimally Hilbertian.

Example 2.1 (Weighted Euclidean spaces). In the Euclidean space $\mathbb{R}^d$, we consider the distance d_e induced by the Euclidean norm, and a non-negative, finite Borel measure $\mathbf{m}$ on $(\mathbb{R}^d, d_e)$. Then one can take $\mathcal{A} = C_c^\infty(\mathbb{R}^d)$, the space of smooth and compactly supported functions on $\mathbb{R}^d$, together with the map

$$\Gamma(f, h) := \langle \nabla f, \nabla h \rangle, \quad f, h \in C_c^\infty(\mathbb{R}^d),$$

where ∇ is the Euclidean gradient. Both properties (1) and (2) can be found in e.g. [27, Example 5.3.8]. The Hilbertianity of $H^{1,2}(\mathbb{R}^d, d_e, \mathbf{m})$ can be found in e.g. [16, Theorem 11].

Example 2.2 (Weighted Riemannian manifolds). Let (M, g) be a (smooth, complete) Riemannian manifold with induced Riemannian metric d_g , and let $\mathbf{m}$ be a non-negative, finite Borel measure on (M, d_g) . As in the Euclidean case, we can take $\mathcal{A} = C_c^\infty(M)$, the space of smooth and compactly supported functions on M , together with the map

$$\Gamma(f, h) := g(\nabla^g f, \nabla^g h), \quad f, h \in C_c^\infty(M),$$

where ∇^g is the gradient induced by g . Both properties (1) and (2) can be found in e.g. [27, Example 5.3.8]. The Hilbertianity of $H^{1,2}(M, d_g, \mathbf{m})$ can be found in [25, Theorem 3.11].

Example 2.3 (Weighted Hilbert spaces). If $(H, \langle \cdot, \cdot \rangle)$ is a separable Hilbert space with induced distance d_H , for any non-negative, finite Borel measure $\mathbf{m}$ on (H, d_H) , we can take as $\mathcal{A}$ the space $C_b^1(H)$ of continuously Fréchet-differentiable functions with bounded Fréchet differential on H , with the map

$$\Gamma(f, h) := \langle Df, Dh \rangle, \quad f, h \in C_b^1(H),$$

where D is the Fréchet differential. Properties (1) and (2) are contained in [27, Example 5.3.9] and the Hilbertianity of $H^{1,2}(H, d_H, \mathbf{m})$ in [27, Corollary 5.3.11].

Before introducing the next three examples, we define the algebra that will play the role of $\mathcal{A}$ in all of them: the space of so-called *cylinder functions*. Let $\mathbf{E}$ be a Euclidean space, a Riemannian manifold, or a separable Hilbert space, and let $\mathcal{M}_+(\mathbf{E})$ be the space of non-negative, finite Borel measures on $(\mathbf{E}, d_{\mathbf{E}})$, where $d_{\mathbf{E}}$ is the Euclidean, the Riemannian, or the Hilbertian distance, respectively. For a bounded Borel function $f : \mathbf{E} \rightarrow \mathbb{R}$, we define the map $f^* : \mathcal{M}_+(\mathbf{E}) \rightarrow \mathbb{R}$ by

$$f^*(\mu) := \int_{\mathbf{E}} f \, d\mu, \quad \mu \in \mathcal{M}_+(\mathbf{E}).$$

A function $u : \mathcal{M}_+(\mathbf{E}) \rightarrow \mathbb{R}$ is called a cylinder function if

$$u = F \circ (f_0^*, f_1^*, \dots, f_k^*), \tag{2.5}$$

for some $F \in C_c^\infty(\mathbb{R}^{k+1})$, $k \in \mathbb{N}_0$, smooth (in the appropriate sense, as in the examples above) functions f_i on $\mathbf{E}$ for $i = 1, \dots, k$, and f_0 equal to the constant function 1. The resulting algebra of functions is denoted by $\text{Cyl}(\mathbf{E})$. We mention that the following three examples belong to the larger framework of *Unbalanced Optimal Transport*, see [28].

Example 2.4 (L^2 -Hellinger spaces). Let He_2 be the L^2 -Hellinger distance on $\mathcal{M}_+(\mathbf{E})$ given by

$$\text{He}_2(\mu, \nu)^2 := \int_{\mathbf{E}} \left| \sqrt{\frac{d\mu}{d\eta}} - \sqrt{\frac{d\nu}{d\eta}} \right|^2 d\eta, \quad \mu, \nu \in \mathcal{M}_+(\mathbf{E}),$$

where $\eta \in \mathcal{M}_+(\mathbf{E})$ is any measure such that $\mu, \nu \ll \eta$, and $d\mu/d\eta$ (resp. $d\nu/d\eta$) denotes the Radon–Nikodym derivative of μ (resp. ν) with respect to η . We stress that, by 2-homogeneity, the above definition does not depend on the choice of η , so one can always take e.g. $\eta = \mu + \nu$. It can be shown that $(\mathcal{M}_+(\mathbf{E}), \text{He}_2)$ is a complete and separable metric space whose topology is

that of total variation; see e.g. [24]. Let $\mathbf{m}$ be any non-negative, finite Borel measure on $\mathcal{M}_+(\mathbf{E})$. For $\mathcal{A} = \text{Cyl}(\mathbf{E})$, the map $\mathbf{\Gamma}$ is given by

$$\mathbf{\Gamma}^{\text{ver}}(u, v)(\mu) := \int_{\mathbf{E}} (\nabla^{\text{ver}} u)_{\mu}(x) (\nabla^{\text{ver}} v)_{\mu}(x) d\mu(x), \quad \mu \in \mathcal{M}_+(\mathbf{E}), u, v \in \text{Cyl}(\mathbf{E}), \quad (2.6)$$

where ∇^{ver} is the *vertical gradient* given, for a cylinder function u as in (2.5), by

$$(\nabla^{\text{ver}} u)_{\mu}(x) := \sum_{i=0}^k \partial_i F((f_0^*, f_1^*, \dots, f_k^*)(\mu)) f_i(x), \quad x \in \mathbf{E}, \mu \in \mathcal{M}_+(\mathbf{E}).$$

The density property (1) and the Hilbertianity of $H^{1,2}(\mathcal{M}_+(\mathbf{E}), \mathbf{He}_2, \mathbf{m})$ are proven in [14, Corollary 4.20], while property (2) is shown in [14, Corollary 4.8].

Example 2.5 (L^2 -Kantorovich–Rubinstein–Wasserstein spaces). Let $\mathcal{P}_2(\mathbf{E})$ be the space of Borel probability measures μ on $\mathbf{E}$ with finite 2-moment, i.e.,

$$\int_{\mathbf{E}} d_{\mathbf{E}}(x, x_o)^2 d\mu(x) < \infty \quad \text{for some (hence for any) } x_o \in \mathbf{E}.$$

On $\mathcal{P}_2(\mathbf{E})$ we consider the L^2 -Kantorovich–Rubinstein–Wasserstein distance W_2 given by

$$W_2(\mu, \nu)^2 := \inf \left\{ \int_{\mathbf{E} \times \mathbf{E}} d_{\mathbf{E}}(x, y)^2 d\pi(x, y) : \pi \in \Pi(\mu, \nu) \right\}, \quad \mu, \nu \in \mathcal{P}_2(\mathbf{E}),$$

where $\Pi(\mu, \nu)$ is the set of Borel probability measures on $\mathbf{E} \times \mathbf{E}$ having μ and ν as first and second marginal, respectively. It is well known that $(\mathcal{P}_2(\mathbf{E}), W_2)$ is a complete and separable metric space whose topology is stronger than the weak (or narrow) one; see e.g. [2]. Let $\mathbf{m}$ be any non-negative, finite Borel measure on $\mathcal{P}_2(\mathbf{E})$. Since $\mathcal{P}_2(\mathbf{E}) \subset \mathcal{M}_+(\mathbf{E})$, we can still consider the algebra $\mathcal{A} = \text{Cyl}(\mathbf{E})$, with the map $\mathbf{\Gamma}$ given by

$$\mathbf{\Gamma}^{\text{hor}}(u, v)(\mu) := \int_{\mathbf{E}} \langle (\nabla^{\text{hor}} u)_{\mu}(x), (\nabla^{\text{hor}} v)_{\mu}(x) \rangle_{T_x \mathbf{E}} d\mu(x), \quad \mu \in \mathcal{P}_2(\mathbf{E}), u, v \in \text{Cyl}(\mathbf{E}), \quad (2.7)$$

where ∇^{hor} is the *horizontal gradient* given, for a cylinder function u as in (2.5), by

$$(\nabla^{\text{hor}} u)_{\mu}(x) := \sum_{i=0}^k \partial_i F((f_0^*, f_1^*, \dots, f_k^*)(\mu)) \nabla^{\mathbf{E}} f_i(x), \quad x \in \mathbf{E}, \mu \in \mathcal{P}_2(\mathbf{E}),$$

where $\nabla^{\mathbf{E}}$ is the appropriate differential or gradient operator on $\mathbf{E}$. The density property (1) and the Hilbertianity of $H^{1,2}(\mathcal{P}_2(\mathbf{E}), W_2, \mathbf{m})$ are contained in [18, Theorem 4.10, Corollary 4.11], while property (2) is [18, Proposition 4.9]. These results are further extended in [30, 14].

Example 2.6 (Hellinger–Kantorovich space). On the space $\mathcal{M}_+(\mathbf{E})$ we consider a different distance, namely the so-called *Hellinger–Kantorovich* (also known as Wasserstein–Fisher–Rao) distance given by

$$\mathbf{HK}(\mu, \nu)^2 := \inf \left\{ \int_0^1 \int_{\mathbf{E}} \left[|v_t|^2 + \frac{1}{4} |w_t|^2 \right] d\eta_t(x) dt : \partial_t \eta + \text{div}(v_t \eta_t) = w_t \eta_t, \eta_0 = \mu, \eta_1 = \nu \right\},$$

where the infimum is taken over narrowly continuous curves $\eta : [0, 1] \rightarrow \mathcal{M}_+(\mathbf{E})$ satisfying the PDE above in the distributional sense on $[0, 1] \times \mathbf{E}$. The space $(\mathcal{M}_+(\mathbf{E}), \mathbf{HK})$ is a complete and separable metric space; see the works [21, 9, 23] where it was introduced, and [12] for the connection between Wassertein, Hellinger, and Hellinger–Kantorovich distances. Considering again the algebra $\mathcal{A}$ of cylinder functions, the map $\mathbf{\Gamma}$ in this case is given by

$$\mathbf{\Gamma}(u, v) := \mathbf{\Gamma}^{\text{hor}}(u, v) + 4\mathbf{\Gamma}^{\text{ver}}(u, v), \quad u, v \in \text{Cyl}(\mathbf{E}),$$

where Γ^{hor} is as in (2.7) and Γ^{ver} is as in (2.6). Property (1) and the Hilbertianity of $(\mathcal{M}_+(\mathbf{E}), \mathbf{H}, \mathbf{m})$ for any non-negative, finite Borel measure $\mathbf{m}$ on $(\mathcal{M}_+(\mathbf{E}), \mathbf{H})$ are established in [14, Corollary 4.17]. Property (2) is proven in [14, Proposition 4.7].

Remark 2.7 (Extension to the Hilbertian case). We remark that, while the case of $\mathbf{E}$ being equal to a separable Hilbert case has been explicitly treated for the L^2 -Kantorovich–Rubinstein–Wasserstein setting in Example 2.5 in [18, Section 6.3], this has not been done explicitly for the L^2 -Hellinger and Hellinger–Kantorovich cases in Examples 2.4 and 2.6, respectively. However, exactly the same argument of [18, Theorem 6.4] can be used simply replacing W_2 with He_2 or $\mathbf{H}$.

We briefly comment on the general ideas used to establish properties (1) and (2) in Examples 2.1–2.6.

Regarding property (2), while its validity is straightforward in the Euclidean, Riemannian, and Hilbertian settings of Examples 2.1–2.3, it requires a more delicate approach in the less concrete settings of Examples 2.4–2.6. In the latter cases, the property is obtained by exploiting the fine structure of geodesics in the underlying metric space.

Property (1) can be derived in all examples from the general criterion proven in [18, Theorem 2.13] and later extended in [14, Theorem 2.12]. This criterion shows that density is achieved by approximating—in the Sobolev sense—functions of the form $\mathcal{S} \ni x \mapsto d(x, y)$ as the base point $y \in \mathcal{S}$ varies.

Finally, as noted above, Hilbertianity follows directly from properties (1) and (2) (see [18, Theorem 2.17]).

We conclude this section mentioning the paper [26] for the use of subalgebras of Lipschitz functions in metric BV spaces.

3. REGULARIZED LEAST SQUARES APPROXIMATION OF BOUNDED FUNCTIONS

We show in this section how to recover some results of [17] originally obtained for the metric-measure space $(\mathcal{P}_2(\mathbf{E}), W_2, \mathbf{m})$ as in Example 2.5, within the more general setting of arbitrary metric-measure spaces. In Subsection 3.1 we will work at the level of the pre-Cheeger energy as in (2.2) as it is more common, see the Examples 2.1–2.6, that one can compute it explicitly on relevant subalgebras $\mathcal{A}$ of functions.

We point out that, in case the pre-Cheeger energy is *closable* on $\mathcal{A}$ (that is, the $L^2(\mathcal{S}, \mathbf{m})$ closure of the graph of $\text{pCE}_{2, \mathbf{m}}$ restricted to $\mathcal{A}$ is still a graph) and $\mathcal{A}$ is strongly dense as in (1), then the Cheeger energy, when restricted to $\mathcal{A}$, coincides with the pre-Cheeger energy. Therefore, in the closable case, the procedure of Section 3.1 can be carried out with $\text{CE}_{2, \mathbf{m}}$ instead of $\text{pCE}_{2, \mathbf{m}}$. This is indeed one of the few cases in which the Cheeger energy is explicitly computable, at least for the functions in $\mathcal{A}$. Closability obviously depends on the choice of the measure $\mathbf{m}$ (besides, of course, on $(\mathcal{S}, d)$): it holds, for example, in Example 2.1 when $\mathbf{m}$ is the Lebesgue measure, in Example 2.2 when $\mathbf{m}$ is the Riemannian volume measure, and also in Examples 2.5 and 2.6 for specific choices of $\mathbf{m}$, see the works [13, 14] for possible choices of such measures.

However, in general, one cannot hope for the closability property or the possibility to explicitly compute the Cheeger energy. This is however crucial to carry out the procedure of Section 3.1 for the Cheeger energy. We can however, work also at the level of the Cheeger energy, at the cost of introducing an approximation parameter, via a Lusin type procedure. This is the content of Section 3.2.

3.1. Direct pre-Cheeger approach

We consider a metric-measure space $(\mathcal{S}, d, \mathbf{p})$, with $\mathbf{p}$ a probability, together with a unital subalgebra $\mathcal{A} \subset \text{Lip}_b(\mathcal{S}, d)$ satisfying (2') and additionally

- (3) the map Γ in (2) takes values in $C_b(\mathcal{S})$, the space of continuous and bounded functions on $(\mathcal{S}, d)$.

Note that (3) is satisfied in the Examples 2.1-2.6.

We fix natural numbers $n, N \in \mathbb{N}$, a finite dimensional subspace $\mathcal{V}_n \subset \mathcal{A} \subset H^{1,2}(\mathcal{S}, \mathbf{d}, \mathbf{p})$ with $\dim(\mathcal{V}_n) = n$ and N i.i.d. points $x_1, \dots, x_N$ distributed on $\mathcal{S}$ according to $\mathbf{p}$; this means that we are considering an underlying probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Accordingly we define the (random) empirical measure $\mathbf{p}_N := \frac{1}{N} \sum_{j=1}^N \delta_{x_j}$ and we fix $\lambda \geq 0$.

Our aim is to estimate the $L^2(\mathcal{S}, \mathbf{p})$ distance between a bounded Borel function f which is known only on the support of $\mathbf{p}_N$ and a suitable regularization of f itself. We start with a definition.

Definition 3.1. Given a function $f \in L^2(\mathcal{S}, \mathbf{p}_N)$, we define the functionals

$$\begin{aligned} \mathcal{J}_{N,f}^\lambda(g) &:= \|f - g\|_{L^2(\mathcal{S}, \mathbf{p}_N)}^2 + \lambda \mathbf{pCE}_{2,\mathbf{p}_N}(g), & g \in \text{Lip}_b(\mathcal{S}, \mathbf{d}), \\ \mathcal{N}_{N,f}(g) &:= \|f - g\|_{L^2(\mathcal{S}, \mathbf{p}_N)}^2, & g \in L^2(\mathcal{S}, \mathbf{p}_N). \end{aligned}$$

Both functionals admit a unique minimizer when restricted to $\mathcal{V}_n$ so that we can define

$$S_N^{\lambda,n}(f) := \operatorname{argmin}_{g \in \mathcal{V}_n} \mathcal{J}_{N,f}^\lambda(g), \quad P_N^n(f) := \operatorname{argmin}_{g \in \mathcal{V}_n} \mathcal{N}_{N,f}(g).$$

In order to make the objects above computable, we fix a suitable basis of $\mathcal{V}_n$.

Lemma 3.2. *There exists a finite family of functions $\mathcal{B}_n := \{f_1, \dots, f_n\} \subset \mathcal{V}_n$ such that*

- (i) $\mathcal{B}_n$ is a linear basis of $\mathcal{V}_n$;
- (ii) $\int_{\mathcal{S}} f_i f_j \, \mathbf{d}\mathbf{p} = 0$ for every $i, k \in \{1, \dots, n\}$, $i \neq k$;
- (iii) $\int_{\mathcal{S}} |f_i|^2 \, \mathbf{d}\mathbf{p} = 1$ for every $i = 1, \dots, n$;
- (iv) $\mathbf{pCE}_{2,\mathbf{p}}(f_i, f_k) = 0$ for every $i, k \in \{1, \dots, n\}$, $i \neq k$.

Proof. This follows by the simultaneous diagonalizations of the two bilinear forms

$$(f, g) \mapsto \int_{\mathcal{S}} f g \, \mathbf{d}\mathbf{p}, \quad (f, g) \mapsto \mathbf{pCE}_{2,\mathbf{p}}(f, g)$$

in the finite dimensional vector space $\mathcal{V}_n$. Note we are only using the bilinearity given by (2). $\square$

In the next Lemma we make the form of $S_N^{\lambda,n}(f)$ as in Definition 3.1 more explicit. To do that we first need to define two matrices.

Definition 3.3. Let $\{f_1, \dots, f_n\}$ be as in Lemma 3.2. We define the matrices $L_{N,n} \in \mathbb{R}^{N \times n}$ and $D_n \in \mathbb{R}^{n \times n}$ as

$$(L_{N,n})_{ji} := \frac{1}{\sqrt{N}} f_i(x_j), \quad (D_n)_{ik} = \mathbf{pCE}_{2,\mathbf{p}_N}(f_i, f_k).$$

The expected value of the matrices $L_{N,n}^\top L_{N,n}$ and D_n is easily computed:

$$\mathbb{E}[L_{N,n}^\top L_{N,n}] = I_n \quad \text{and} \quad \mathbb{E}[D_n] = \Gamma_n, \tag{3.1}$$

where I_n is the identity matrix of order n and $\Gamma_n \in \mathbb{R}^{n \times n}$ is a diagonal matrix with entries

$$(\Gamma_n)_{ii} = \mathbf{pCE}_{2,\mathbf{p}}(f_i), \quad i = 1, \dots, n.$$

In the next Lemma, we show how to render the a priori abstract quantity $S_N^{\lambda,n}(f)$ computable. This constitutes the first step toward utilizing such quantities in real-world computer simulations, as demonstrated in [17].

Lemma 3.4. *Let $\{f_1, \dots, f_n\}$ be as in Lemma 3.2 and let $f \in L^2(\mathcal{S}, \mathbf{p}_N)$. Then $g = S_N^{\lambda,n}(f)$ (cf. Definition 3.1) if and only if*

$$g = \sum_{i=1}^n f_i w_i^*,$$

where $w^\star \in \mathbb{R}^n$ is the unique minimizer in $\mathbb{R}^n$ of the functional $\mathcal{J}_{N,z^f}^{\lambda,n} : \mathbb{R}^n \rightarrow \mathbb{R}$ defined as

$$\mathcal{J}_{N,z^f}^{\lambda,n}(w) := |L_{N,n}(w - z^f)|_N^2 + \lambda w^\top D_n w, \quad w \in \mathbb{R}^n, \quad (3.2)$$

where $|\cdot|_N$ denotes the Euclidean norm in $\mathbb{R}^N$ and $z^f \in \mathbb{R}^n$ is such that

$$P_N^n f = \sum_{i=1}^n f_i z_i^f.$$

Equivalently $w^\star$ solves

$$(L_{N,n}^\top L_{N,n} + \lambda D_n) w^\star = L_{N,n}^\top L_{N,n} z^f. \quad (3.3)$$

In particular, $S_N^{\lambda,n}$ is a linear operator.

Proof. For every $g \in \mathcal{V}_n$, we can rewrite

$$\begin{aligned} \mathcal{J}_{N,f}^{\lambda,n}(g) &= \frac{1}{N} \sum_{j=1}^N |g(x_j) - f(x_j)|^2 + \lambda \mathbf{pCE}_{2,\mathbf{p}_N}(g) \\ &= \frac{1}{N} \sum_{j=1}^N |g(x_j) - P_N^n f(x_j)|^2 + \lambda \mathbf{pCE}_{2,\mathbf{p}_N}(g) + C(n, N, f) \end{aligned}$$

where $C(n, N, F)$ is a constant depending only on f , n and N . Furthermore, upon defining

$$\tilde{\mathcal{J}}_{N,f}^{\lambda,n}(g) := \frac{1}{N} \sum_{j=1}^N |g(x_j) - P_N^n f(x_j)|^2 + \lambda \mathbf{pCE}_{2,\mathbf{p}_N}(g)$$

we obtain that

$$\mathcal{J}_{N,f}^{\lambda,n}(G) = \tilde{\mathcal{J}}_{N,f}^{\lambda,n}(g) + C(n, N, f).$$

Since $\tilde{\mathcal{J}}_{N,f}^{\lambda,n}$ and $\mathcal{J}_{N,f}^{\lambda,n}$ differ only by a constant, they have the same minimizer. For a given $g \in \mathcal{V}_n$ there exists a unique $w^g \in \mathbb{R}^n$ such that

$$g = \sum_{i=1}^n f_i w_i^g$$

and it is immediate to check that $\tilde{\mathcal{J}}_{N,f}^{\lambda,n}(g) = \mathcal{J}_{N,z^f}^{\lambda,n}(w^g)$. This concludes the proof. $\square$

Remark 3.5. We note that (3.3) can equivalently be stated as

$$(L_{N,n}^\top L_{N,n} + \lambda D_n) w^\star = y^f,$$

where $y^f \in \mathbb{R}^n$ with

$$y_i^f = \frac{1}{N} \sum_{j=1}^N f(x_j) f_i(x_j), \quad i = 1, \dots, n.$$

In particular, we may obtain the solution $w^\star$ by solving a linear equation.

Next result is an auxiliary large deviation bound, which allows us to control the spectral behavior of the matrices introduced in Definition 3.3.

Proposition 3.6. *Let $L_{N,n}$ and D_n be defined as in Definition 3.3 and let I_n and Γ_n be as in (3.1). Then*

$$\mathbb{P}(\|L_{N,n}^\top L_{N,n} - I_n\| > \delta) \leq 2n \exp \left\{ -N \frac{c_\delta}{K_\lambda(n)} \right\} \quad \text{for every } 0 < \delta < 1 \quad (3.4)$$

and, for every $0 < \delta < 1$,

$$\mathbb{P} \left(\|L_{N,n}^\top L_{N,n} + \lambda D_n - (I_n + \lambda \Gamma_n)\| > \delta \sigma_{\max,n}^\lambda \right) \leq 2n \exp \left\{ -N \sigma_{\min,n}^\lambda \frac{c_\delta}{K_\lambda(n)} \right\}, \quad (3.5)$$

where

$$\begin{aligned} K_\lambda(n) &:= \sup_{x \in \mathcal{S}} \sum_{i=1}^n (|f_i(x)|^2 + \lambda \Gamma(f_i, f_i)(x)), \\ \sigma_{\min,n}^\lambda &:= 1 + \lambda \min_{i=1,\dots,n} \mathbf{pCE}_{2,\mathbf{p}}(f_i), \quad \sigma_{\max,n}^\lambda := 1 + \lambda \max_{i=1,\dots,n} \mathbf{pCE}_{2,\mathbf{p}}(f_i), \\ c_\delta &:= (1 + \delta) \log(1 + \delta) - \delta, \end{aligned}$$

and $\|\cdot\|$ denotes the spectral norm.

Proof. The first bound follows precisely as in the proof of [10, 11, Theorem 1]. The second one analogously: $L_{N,n}^\top L_{N,n} + \lambda D_n$ is the sum of N symmetric and positive semi-definite matrices X^j , $j = 1, \dots, N$, which are i.i.d. copies of the matrix X with entries

$$X_{ik} := X_{ik}^1 + X_{ik}^2 := \frac{1}{N} f_i(x) f_k(x) + \frac{\lambda}{N} \Gamma(f_i, f_k)(x), \quad i, k \in \{1, \dots, n\},$$

with x distributed according to $\mathbf{p}$.

Indeed, since both X^1 and X^2 are symmetric and positive semi-definite matrices, so is X and thus its spectral norm can be bounded by its trace; i.e., we have that

$$\|X\| \leq \frac{1}{N} \sum_{i=1}^n (|f_i(x)|^2 + \lambda \Gamma(f_i, f_i)(x)) \quad \text{for } \mathbf{p}\text{-a.e. } x \in \mathcal{S}.$$

Invoking the definition of $K_\lambda(n)$, this immediately yields the uniform bound

$$\|X\| \leq \frac{K_\lambda(n)}{N}.$$

In turn, for every $0 < \delta < 1$, the matrix Chernoff inequality, cf. [31, Theorem 1.1], yields that

$$\begin{aligned} \mathbb{P} \left(\lambda_{\min}(L_{N,n}^\top L_{N,n} + \lambda D_n) - \sigma_{\min,n}^\lambda \leq -\delta \sigma_{\min,n}^\lambda \right) &= \mathbb{P} \left(\lambda_{\min}(L_{N,n}^\top L_{N,n} + \lambda D_n) \leq (1 - \delta) \sigma_{\min,n}^\lambda \right) \\ &\leq n \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^{\frac{N \sigma_{\min,n}^\lambda}{K_\lambda(n)}} \\ &\leq n \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\frac{N \sigma_{\min,n}^\lambda}{K_\lambda(n)}} \end{aligned}$$

and

$$\begin{aligned} \mathbb{P} \left(\lambda_{\max}(L_{N,n}^\top L_{N,n} + \lambda D_n) - \sigma_{\max,n}^\lambda \geq \delta \sigma_{\max,n}^\lambda \right) &= \mathbb{P} \left(\lambda_{\max}(L_{N,n}^\top L_{N,n} + \lambda D_n) \geq (1 + \delta) \sigma_{\max,n}^\lambda \right) \\ &\leq n \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\frac{N \sigma_{\max,n}^\lambda}{K_\lambda(n)}} \\ &\leq n \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\frac{N \sigma_{\min,n}^\lambda}{K_\lambda(n)}} \end{aligned}$$

so that

$$\mathbb{P} \left(\|L_{N,n}^\top L_{N,n} + \lambda D_n - (I_n + \lambda \Gamma_n)\| > \delta \sigma_{\max,n}^\lambda \right) \leq 2n \exp \left\{ -N \sigma_{\min,n}^\lambda \frac{c_\delta}{K_\lambda(n)} \right\}.$$

□

The condition (3.6) below ensures that the probabilities in Proposition 3.6 are small: we think to fix the dimension n of the subspace $\mathcal{V}_n$ and an order $r > 0$. Then, up to choosing enough point evaluations N , we obtain that the above probability is smaller than N^{-r} . The required condition reads as

$$\frac{N}{\log(N)} \geq \frac{(1+r)K_\lambda(n)}{\sigma_{\min,n}^\lambda c_{\frac{1}{2\sigma_{\max,n}^\lambda}}}, \quad N \geq 2, \quad (3.6)$$

which further implies that

$$\frac{N}{\log(N)} \geq \frac{(1+r)K_\lambda(n)}{c_{1/2}}.$$

Indeed, given (3.6) and choosing $\delta = 1/2$ in (3.4) as well as $\delta = (2\sigma_{\max,n}^\lambda)^{-1}$ in (3.5) leads to

$$\begin{aligned} & \mathbb{P} \left(\|(L_{N,n}^\top L_{N,n} + \lambda D_n) - (I_n + \lambda \Gamma_n)\| > \frac{1}{2} \right) + \mathbb{P} \left(\|L_{N,n}^\top L_{N,n} - I_n\| > \frac{1}{2} \right) \\ & \leq 4 \frac{n}{N} N^{-r} \leq 4 \frac{K_\lambda(n)}{N} N^{-r} \leq \frac{4c_{1/2}}{\log(N)(1+r)} N^{-r} \leq \frac{4c_{1/2}}{\log(2)} N^{-r} \leq N^{-r}, \end{aligned} \quad (3.7)$$

where we have used that $K_\lambda(n) \geq n$ and that $N \geq 2$.

With the bound (3.7) at our disposal we can reproduce two of the main results of [17].

In the first result the data are noiseless samples of the function f we intend to recover.

Theorem 3.7. *Let $M > 0$ and let $f : \mathcal{S} \rightarrow [-M, M]$ be a measurable function and let $f^\star$ be defined as*

$$f^\star := -M \vee S_N^{\lambda,n}(f) \wedge M,$$

where $S_N^{\lambda,n}$ is as in Definition 3.1; let $r > 0$ be given and assume that $n, N \in \mathbb{N}$ satisfy (3.6). Then

$$\begin{aligned} & \mathbb{E} \left[\|f - f^\star\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right] \\ & \leq 2e(f, n) \left(1 + \frac{c_{1/2}}{\log(N)(1+r)(1/2 + \lambda\mu_{\min,n})^2} \right) + 4\lambda \mathbf{pCE}_{2,\mathbf{p},N}(P_n f) + 2M^2 N^{-r}, \end{aligned}$$

where $P_n f$ is the $L^2(\mathcal{S}, \mathbf{p})$ -orthogonal projection of f onto $\mathcal{V}_n$ and

$$\begin{aligned} e(f, n) &:= \|P_n f - f\|_{L^2(\mathcal{S}, \mathbf{p})}^2, \\ \mu_{\min,n} &:= \min_{i=1,\dots,n} \mathbf{pCE}_{2,\mathbf{p}}(f_i). \end{aligned}$$

Proof. Recall that we have fixed an underlying probability space $(\Omega, \mathcal{E}, \mathbb{P})$; we consider a partition of Ω into the two sets

$$\Omega_+ := \left\{ \omega \in \Omega : \|(L_{N,n}^\top L_{N,n} + \lambda D_n) - (I_n + \lambda \Gamma_n)\| \leq \frac{1}{2} \wedge \|L_{N,n}^\top L_{N,n} - I_n\| \leq \frac{1}{2} \right\}$$

and its complement $\Omega_- := \Omega \setminus \Omega_+$. Notice that (3.7) gives $\mathbb{P}(\Omega_-) \leq N^{-r}$. We have then

$$\mathbb{E} \left[\|f - f^\star\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right] \leq \int_{\Omega_+} \|f - f^\star\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} + 2M^2 N^{-r}.$$

We are left to estimate

$$\int_{\Omega_+} \|f - f^\star\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} \leq \int_{\Omega_+} \|f - S_N^{\lambda,n}(f)\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P}.$$

Since by Lemma 3.4 the operator $S_N^{\lambda,n}$ is linear, we have

$$f - S_N^{\lambda,n}(f) = (I - S_N^{\lambda,n})h + P_n f - S_N^{\lambda,n}(P_n f),$$

where $h = f - P_n f$.

Observe that for $\xi \in \mathcal{V}_n$ and a draw in Ω_+ it holds

$$\int_{\mathcal{S}} |\xi|^2 d\mathbf{p} \leq 2 \int_{\mathcal{S}} |\xi|^2 d\mathbf{p}_N. \quad (3.8)$$

By the very definition of $S_N^{\lambda,n}$ we further find that

$$\begin{aligned} \|S_N^{\lambda,n}(G) - G\|_{L^2(\mathcal{S}, \mathbf{p}_N)}^2 &\leq \|S_N^{\lambda,n}(G) - G\|_{L^2(\mathcal{S}, \mathbf{p}_N)}^2 + \lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(S_N^{\lambda,n}(G)) \\ &\leq \|G - G\|_{L^2(\mathcal{S}, \mathbf{p}_N)}^2 + \lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(G) \\ &= \lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(G) \end{aligned}$$

for any $G \in \mathcal{V}_n$. Combining this inequality with (3.8) yields

$$\begin{aligned} \int_{\Omega_+} \|P_n f - S_N^{\lambda,n}(P_n f)\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} &\leq 2 \int_{\Omega_+} \|P_n f - S_N^{\lambda,n}(P_n f)\|_{L^2(\mathcal{S}, \mathbf{p}_N)}^2 d\mathbb{P} \\ &\leq 2\lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(P_n f). \end{aligned}$$

It only remains to estimate

$$\int_{\Omega_+} \|(I - S_N^{\lambda,n})h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P}.$$

Since h is $L^2(\mathcal{S}, \mathbf{p})$ -orthogonal to $\mathcal{V}_n$ and $S_N^{\lambda,n}h$ belongs to $\mathcal{V}_n$, we have that

$$\|(I - S_N^{\lambda,n})h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 = \|h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 + \|S_N^{\lambda,n}h\|_{L^2(\mathcal{S}, \mathbf{p})}^2.$$

Precisely as in [10, 11] one then shows that

$$\int_{\Omega_+} \|S_N^{\lambda,n}h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} \leq \alpha_{\lambda,n}^2 \frac{K_{\lambda}(n)}{N} \|h\|_{L^2(\mathcal{S}, \mathbf{p})}^2,$$

where

$$\alpha_{\lambda,n} := \frac{1}{\frac{1}{2} + \lambda \min_{i=1, \dots, n} \mathbf{p} \mathbf{CE}_{2, \mathbf{p}}(f_i)}. \quad (3.9)$$

Putting everything together we get

$$\begin{aligned} \mathbb{E} \left[\|f - f^*\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right] &\leq \int_{\Omega_+} \|f - f^*\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} + 2M^2 N^{-r} \\ &\leq 2 \int_{\Omega_+} \|(I - S_N^{\lambda,n})h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} + 4\lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(P_n f) + 2M^2 N^{-r} \\ &\leq 2\|h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 + 2\alpha_{\lambda,n}^2 \frac{K_{\lambda}(n)}{N} \|h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 + 4\lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(P_n f) + 2M^2 N^{-r}. \end{aligned}$$

Finally, recalling the definition of h and employing (3.6) yield that

$$\begin{aligned} \mathbb{E} \left[\|f - f^*\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right] &\leq \|P_n f - f\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \left(2 + \frac{2\alpha_{\lambda,n}^2 c_{1/2}}{\log(N)(1+r)} \right) \\ &\quad + 4\lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(P_n f) + 2M^2 N^{-r}. \end{aligned}$$

□

We assume now to know the target function f at the (random) points $x_1, \dots, x_N$ only up to some noise. In particular, we assume that we have at our disposal the noisy values

$$\tilde{f}(x_j) := f(x_j) + \eta_j, \quad j = 1, \dots, N, \quad (3.10)$$

where $(\eta_j)_{j=1}^N$ are i.i.d. random variables with variance bounded by $\sigma^2 > 0$. Notice that the (random) function $\tilde{f}$ belongs to $L^2(\mathcal{S}, \mathbf{p}_N)$.

Theorem 3.8. *Let $M > 0$, let $f : \mathcal{S} \rightarrow [-M, M]$ be a measurable function and let $\tilde{f}^*$ be defined as*

$$\tilde{f}^* := -M \vee S_N^{\lambda, n}(\tilde{f}) \wedge M,$$

where $S_N^{\lambda, n}$ is as in Definition 3.1 and $\tilde{f}$ is as in (3.10) based on the noisy data $(y_j)_{j=1}^N$ whose i.i.d. errors $(\eta_j)_{j=1}^N$ have variance bounded by $\sigma^2 > 0$. Let $r > 0$ be given and assume that $n, N \in \mathbb{N}$ satisfy (3.6). Then

$$\begin{aligned} \mathbb{E} \left[\|f - \tilde{f}^*\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right] &\leq 2e(f, n) \left(1 + \frac{c_{1/2}}{\log(N)(1+r)(1/2 + \lambda\mu_{\min, n})^2} \right) + 8\lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(P_n f) + \\ &\quad + 4 \frac{\sigma^2}{(1 + \lambda\mu_{\min, n})^2} \frac{n}{N} + 2M^2 N^{-r}, \end{aligned}$$

where $e(f, n)$ and $\mu_{\min, n}$ are as in Theorem 3.7.

Proof. The proof follows the one of Theorem 3.7 and it is also inspired by [10, 11, Theorem 3]. We split again Ω into the sets Ω_+ and Ω_- . Performing the same calculations as before, we find that

$$\mathbb{E} \left[\|f - \tilde{f}^*\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right] \leq \int_{\Omega_+} \|f - S_N^{\lambda, n}(\tilde{f})\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} + 2M^2 N^{-r}.$$

We can write

$$f - S_N^{\lambda, n}(\tilde{f}) = (I - S_N^{\lambda, n})(h + P_n f) - S_N^{\lambda, n}(\eta)$$

where $h = f - P_n f$, P_n is the orthogonal $L^2(\mathcal{S}, \mathbf{p})$ projection on $\mathcal{V}_n$ and $\eta \in L^2(\mathcal{S}, \mathbf{p}_N)$ is the (random) function taking values $\eta(x_j) := \eta_j$, $j = 1, \dots, N$.

Following precisely the same steps as in the previous proof, we obtain the bound

$$\begin{aligned} \int_{\Omega_+} \|f - S_N^{\lambda, n}(\tilde{f})\|_{L^2(\mathcal{S}, \mathbf{p})}^2 d\mathbb{P} &\leq 2\|h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 + 8\lambda \mathbf{p} \mathbf{CE}_{2, \mathbf{p}_N}(P_n f) \\ &\quad + \int_{\Omega_+} \left(2\|S_N^{\lambda, n}(h)\|_{L^2(\mathcal{S}, \mathbf{p})}^2 + 4\|S_N^{\lambda, n}(\eta)\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right) d\mathbb{P}. \end{aligned}$$

Repeating the computations of [10, 11, Theorem 3] then leads to

$$\int_{\Omega_+} \left(2\|S_N^{\lambda, n}(h)\|_{L^2(\mathcal{S}, \mathbf{p})}^2 + 4\|S_N^{\lambda, n}(\eta)\|_{L^2(\mathcal{S}, \mathbf{p})}^2 \right) d\mathbb{P} \leq 2\alpha_{\lambda, n}^2 \frac{K_\lambda(n)}{N} \|h\|_{L^2(\mathcal{S}, \mathbf{p})}^2 + 4\alpha_{\lambda, n}^2 \sigma^2 \frac{n}{N},$$

where $\alpha_{\lambda, n}$ is as in (3.9). Combining all the estimates and doing some simple manipulations finally lead to the claimed upper bound. $\square$

Remark 3.9. In practical applications, it is not always straightforward to determine the approximation subspace $\mathcal{V}_n$ that optimizes $e(f, n)$. A possible heuristic remedy, explored in [17, Section 6.4], is the employment of neural networks.

3.2. Approximated Cheeger energy approach

In this subsection, we want to reproduce the results of Section 3.1 but working directly at the level of the Cheeger energy. In general, we can not assume that condition (3) holds with $\mathbf{pCE}$ being replaced by $\mathbf{CE}$ (unless, for example, we are in the special closable case discussed at the beginning of Section 3).

Still, in general, we have that the Cheeger energy can be represented in terms of a bilinear map, in the following sense; we consider a metric-measure space $(\mathcal{S}, \mathbf{d}, \mathbf{p})$ with $\mathbf{p}$ a probability, together with a unital subalgebra $\mathcal{A} \subset \text{Lip}_b(\mathcal{S}, \mathbf{d})$. Unlike in Section 3.1, we assume $(\mathcal{S}, \mathbf{d}, \mathbf{m})$ to be infinitesimally Hilbertian. In particular, there exists a symmetric bilinear map $\mathbf{\Gamma} : H^{1,2}(\mathcal{S}, \mathbf{d}, \mathbf{m}) \times H^{1,2}(\mathcal{S}, \mathbf{d}, \mathbf{m}) \rightarrow L^1(\mathcal{S}, \mathbf{m})$ such that

$$\mathbf{CE}_{2,\mathbf{p}}(f, h) = \int_{\mathcal{S}} \mathbf{\Gamma}(f, h) d\mathbf{p}, \quad \mathbf{\Gamma}(f, f) \geq 0 \text{ } \mathbf{p}\text{-a.e.} \quad \text{for every } f, h \in H^{1,2}(\mathcal{S}, \mathbf{d}, \mathbf{m}).$$

Here, we are denoting with $\mathbf{CE}_{2,\mathbf{p}}(\cdot, \cdot)$ the quadratic form associated by polarization to $\mathbf{CE}_{2,\mathbf{p}}$ which satisfies the parallelogram identity because of the Hilbertianity of $H^{1,2}(\mathcal{S}, \mathbf{d}, \mathbf{p})$. The existence of the map $\mathbf{\Gamma}$ as above can be deduced, for example, from [19, Theorem 4.3.3].

As in Section 3.1, we fix natural numbers $N \in \mathbb{N}$ and N i.i.d. points $x_1, \dots, x_N$ distributed on $\mathcal{S}$ according to $\mathbf{p}$; this means that we are considering an underlying probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Accordingly, we define the (random) empirical measure $\mathbf{p}_N := \frac{1}{N} \sum_{j=1}^N \delta_{x_j}$ and we fix $\lambda \geq 0$.

We consider a family of n -dimensional subspaces $(\mathcal{V}_n)_n \subset \text{Lip}_b(\mathcal{S}, \mathbf{d}) \subset H^{1,2}(\mathcal{S}, \mathbf{d}, \mathbf{p})$; arguing as in Lemma 3.2, for any $n \in \mathbb{N}$, we can find a family of functions $\mathcal{B}_n := \{f_1, \dots, f_n\}$ such that

- (a) $\mathcal{B}_n$ is a linear basis of $\mathcal{V}_n$;
- (b) $\int_{\mathcal{S}} f_i f_k d\mathbf{p} = 0$ for every $i, k \in \{1, \dots, n\}$, $i \neq k$;
- (c) $\int_{\mathcal{S}} |f_i|^2 d\mathbf{p} = 1$ for every $i = 1, \dots, n$;
- (d) $\mathbf{CE}_{2,\mathbf{p}}(f_i, f_k) = 0$ for every $i, k \in \{1, \dots, n\}$, $i \neq k$.

We denote

$$\mathcal{B} := \bigcup_n \mathcal{B}_n \subset \mathcal{A}.$$

Instead of assuming directly that $\mathbf{\Gamma}$ takes values in the space of continuous functions, we can produce an approximation of it with this property, at the price of introducing an additional error parameter.

Lemma 3.10. *For every $\varepsilon \in (0, 1)$ there exist a Borel subset $A_\varepsilon \subset \mathcal{S}$ with $\mathbf{p}(A_\varepsilon) \leq \varepsilon$ and a symmetric map $\mathbf{\Gamma}_\varepsilon : \mathcal{B} \times \mathcal{B} \rightarrow C_b(\mathcal{S})$ such that*

$$\mathbf{\Gamma}_\varepsilon(f, h) = \mathbf{\Gamma}(f, h) \text{ in } \mathcal{S} \setminus A_\varepsilon \text{ and } \int_{\mathcal{S}} |\mathbf{\Gamma}_\varepsilon(f, h) - \mathbf{\Gamma}(f, h)| d\mathbf{p} < \varepsilon \quad \text{for every } f, h \in \mathcal{B}.$$

Proof. Fix $(f, h) \in \mathcal{B} \times \mathcal{B}$ and $\delta > 0$. For simplicity, we set $u := \mathbf{\Gamma}(f, h)$; since $u \in L^1(\mathcal{S}, \mathbf{p})$, we can find $M_\delta > 1$ such that

$$\int_{\{|u| > M_\delta\}} |u| d\mathbf{p} < \delta/6.$$

Therefore, by Markov's inequality, we have

$$M_\delta \mathbf{p}(\{|u| > M_\delta\}) \leq \delta/6.$$

Set $u_\delta := -M_\delta \vee u \wedge M_\delta$. By Lusin's theorem applied to the finite Borel measure

$$\mathbf{m} := (|u| + 1)\mathbf{p}$$

and the bounded function u_δ , there exists a closed set $E_\delta \subset \mathcal{S}$ such that u_δ is continuous on E_δ and $\mathbf{m}(\mathcal{S} \setminus E_\delta) < \delta/(3M_\delta)$. Let us set $B_\delta := (\mathcal{S} \setminus E_\delta) \cup \{|u| > M_\delta\}$. We compute

$$\mathbf{p}(B_\delta) \leq \mathbf{p}(\mathcal{S} \setminus E_\delta) + \mathbf{p}(\{|u| > M_\delta\}) \leq \mathbf{m}(\mathcal{S} \setminus E_\delta) + \delta/(6M_\delta) < \delta.$$

Since E_δ is closed, by Tietze's extension theorem, there exists a continuous function v_δ such that $v_\delta = u_\delta$ on E_δ and $\|v_\delta\|_\infty \leq M_\delta$. On $\mathcal{S} \setminus B_\delta = E_\delta \setminus \{|u| > M_\delta\}$, we have $v_\delta = u_\delta$ and $u_\delta = u$ so that $v_\delta = u$ on $\mathcal{S} \setminus B_\delta$. We compute

$$\begin{aligned} \int_{\mathcal{S}} |u - v_\delta| \, d\mathbf{p} &= \int_{B_\delta} |u - v_\delta| \, d\mathbf{p} \\ &\leq \int_{\mathcal{S} \setminus E_\delta} |u| \, d\mathbf{p} + M_\delta \mathbf{p}(\mathcal{S} \setminus E_\delta) + \int_{\{|u| > M_\delta\}} |u| \, d\mathbf{p} + M_\delta \mathbf{p}(\{|u| > M_\delta\}) \\ &\leq \mathbf{m}(\mathcal{S} \setminus E_\delta)(1 + M_\delta) + \delta/6 + \delta/6 \\ &\leq \delta/(3M_\delta)(1 + M_\delta) + \delta/3 \\ &\leq \delta. \end{aligned}$$

We have shown that for every $\delta > 0$ and every $(f, h) \in \mathcal{B} \times \mathcal{B}$ there exists $v_\delta^{f,h} \in C_b(\mathcal{S})$ and a Borel subset $B_\delta^{f,h} \subset \mathcal{S}$ with $\mathbf{p}(\mathcal{S} \setminus B_\delta^{f,h}) < \delta$ such that

$$v_\delta^{f,h} = \Gamma(f, h) \text{ in } B_\delta^{f,h}, \quad \int_{\mathcal{S}} |v_\delta^{f,h} - \Gamma(f, h)| \, d\mathbf{p} < \delta.$$

Let $\varepsilon > 0$ be fixed and let $(f_l)_l$ be an enumeration of $\mathcal{B}$; for every $l, m \in \mathbb{N}$ with $m \leq l$ set $\delta_{l,m} := \varepsilon/2^{l+m}$ and

$$\Gamma_\varepsilon(f_l, f_m) := v_{\delta_{l,m}}^{f_l, f_m}, \quad A_\varepsilon := \bigcup_{l,m} B_{\delta_{l,m}}^{f_l, f_m}.$$

Then $\mathbf{p}(A_\varepsilon) \leq \varepsilon$, $\Gamma_\varepsilon(f_l, f_m) \in C_b(\mathcal{S})$ and, for every $l, m \in \mathbb{N}$ with $m \leq l$, on $\mathcal{S} \setminus A_\varepsilon \subset \mathcal{S} \setminus B_{\delta_{l,m}}^{f_l, f_m}$, we have $\Gamma_\varepsilon(f_l, f_m) = v_{\delta_{l,m}}^{f_l, f_m} = \Gamma(f_l, f_m)$ and

$$\int_{\mathcal{S}} |\Gamma_\varepsilon(f_l, f_m) - \Gamma(f_l, f_m)| \, d\mathbf{p} = \int_{\mathcal{S}} |v_{\delta_{l,m}}^{f_l, f_m} - \Gamma(f_l, f_m)| \, d\mathbf{p} < \delta_{l,m} \leq \varepsilon.$$

When $l, m \in \mathbb{N}$ with $m > l$, we simply set

$$\Gamma_\varepsilon(f_l, f_m) := \Gamma_\varepsilon(f_m, f_l).$$

By symmetry of Γ , we conclude the proof. $\square$

For every $\varepsilon \in (0, 1)$, and every non-negative, finite Borel measure $\mathbf{m}$ on $(\mathcal{S}, \mathbf{d})$, we define a map $\mathbf{CE}_{2,\mathbf{m}}^\varepsilon : \mathcal{B} \times \mathcal{B} \rightarrow [0, \infty)$ as

$$\mathbf{CE}_{2,\mathbf{m}}^\varepsilon(f, h) := \int_{\mathcal{S}} \Gamma_\varepsilon(f, h) \, d\mathbf{m} \quad f, h \in \mathcal{B} \times \mathcal{B},$$

where Γ_ε is as in Lemma 3.10.

We now fix $n \in \mathbb{N}$ (corresponding to the subspace $\mathcal{V}_n$) and a tolerance $\gamma \in (0, 1)$ for the rest of the subsection: the latter parameter will quantify the error we make in approximating the Cheeger energy $\mathbf{CE}_{2,\mathbf{p}}$ with the above approximation $\mathbf{CE}_{2,\mathbf{p}}^\varepsilon$. We choose $\varepsilon \in (0, \gamma)$ (clearly depending also on γ and n) such that

$$\left(\sum_{i=1}^n \sum_{k=1}^n |\mathbf{CE}_{2,\mathbf{p}}^\varepsilon(f_i, f_k) - \mathbf{CE}_{2,\mathbf{p}}(f_i, f_k)|^2 \right)^{1/2} < \gamma. \quad (3.11)$$

The possibility of doing so follow of course from Lemma 3.10.

We define the analogous matrices as in Definition 3.3, additionally depending on the parameter ε as above.

Definition 3.11. Let $\{f_1, \dots, f_n\}$ be as in (a)-(d). We define the matrices $L_{N,n} \in \mathbb{R}^{N \times n}$ and $I_n, D_{N,n}^\varepsilon, \Gamma_n^\varepsilon, \Gamma_n \in \mathbb{R}^{n \times n}$ as

$$(L_{N,n})_{ji} := \frac{1}{\sqrt{N}} f_i(x_j), \quad (D_{N,n}^\varepsilon)_{ik} = \int_{\mathcal{S}} \Gamma_\varepsilon(f_i, f_k) d\mathbf{p}_N, \quad (\Gamma_n^\varepsilon)_{ik} := \mathbf{CE}_{2,\mathbf{p}}^\varepsilon(f_i, f_k)$$

with I_n the identity matrix in $\mathbb{R}^{n \times n}$ and Γ_n the diagonal matrix with entries

$$(\Gamma_n)_{ii} = \mathbf{CE}_{2,\mathbf{p}}(f_i, f_i), \quad i = 1, \dots, n.$$

The expected value of the matrices $L_{N,n}^\top L_{N,n}$ and $D_{N,n}^\varepsilon$ is easily computed:

$$\mathbb{E}[L_{N,n}^\top L_{N,n}] = I_n \quad \text{and} \quad \mathbb{E}[D_{N,n}^\varepsilon] = \Gamma_n^\varepsilon; \quad (3.12)$$

By (3.11), we see that $\|\Gamma_n^\varepsilon - \Gamma_n\| < \gamma$ so that

$$\|D_{N,n}^\varepsilon - \Gamma_n\| \leq \|D_{N,n} - \Gamma_n^\varepsilon\| + \gamma.$$

Proposition 3.12. Let $I_n, L_{N,n}, D_{N,n}^\varepsilon$ and Γ_n^ε be defined as in Definition 3.11. Then

$$\mathbb{P}(\|L_{N,n}^\top L_{N,n} - I_n\| > \delta) \leq 2n \exp \left\{ -N \frac{c_\delta}{K_\lambda^\varepsilon(n)} \right\} \quad \text{for every } 0 < \delta < 1 \quad (3.13)$$

and, for every $0 < \delta < 1$, we have

$$\mathbb{P} \left(\|L_{N,n}^\top L_{N,n} + \lambda D_{N,n}^\varepsilon - (I_n + \lambda \Gamma_n^\varepsilon)\| > \delta \sigma_{\max,n}^{\lambda,\varepsilon} \right) \leq 2n \exp \left\{ -N \sigma_{\min,n}^{\lambda,\varepsilon} \frac{c_\delta}{K_\lambda^\varepsilon(n)} \right\} + N\varepsilon, \quad (3.14)$$

where

$$\begin{aligned} K_\lambda^\varepsilon(n) &:= \sup_{x \in \mathcal{S} \setminus A_\varepsilon} \sum_{i=1}^n (|f_i(x)|^2 + \lambda \Gamma_\varepsilon(f_i, f_i)(x)), \\ \sigma_{\min,n}^{\lambda,\varepsilon} &:= 1 + \lambda \min_{i=1,\dots,n} \mathbf{CE}_{2,\mathbf{p}}^\varepsilon(f_i, f_i), \quad \sigma_{\max,n}^\lambda := 1 + \lambda \max_{i=1,\dots,n} \mathbf{CE}_{2,\mathbf{p}}^\varepsilon(f_i, f_i), \\ c_\delta &:= (1 + \delta) \log(1 + \delta) - \delta, \end{aligned}$$

and $\|\cdot\|$ denotes the spectral norm.

Proof. The first bound follows precisely as in the proof of [10, 11, Theorem 1]. For the second one, we proceed as follows: we split Ω in two parts Ω_+ and Ω_- defined as

$$\Omega_+ := \{\omega \in \Omega : x_j(\omega) \in \mathcal{S} \setminus A_\varepsilon \text{ for every } j = 1, \dots, N\}, \quad \Omega_- := \Omega \setminus \Omega_+.$$

Consider the new probability space $(\Omega_+, \mathcal{E}_+, \mathbb{P}_+)$, where $\mathcal{E}_+$ is the trace of $\mathcal{E}$ on Ω_+ and $\mathbb{P}_+$ is simply the normalized restriction of $\mathbb{P}$ to Ω_+ . For every $E \in \mathcal{E}$, we clearly have

$$\mathbb{P}(E) \leq \mathbb{P}_+(E \cap \Omega_+) \mathbb{P}(\Omega_+) + \mathbb{P}(\Omega_-) \leq \mathbb{P}_+(E \cap \Omega_+) + N\varepsilon.$$

On Ω_+ , exactly the same estimates of Proposition 3.12 can be applied noticing that, there, $L_{N,n}^\top L_{N,n} + \lambda D_{N,n}^\varepsilon$ is the sum of N independent symmetric and positive semi-definite matrices X^j , $j = 1, \dots, N$, with entries

$$X_{ik}^j := \frac{1}{N} f_i(x_j) f_k(x_j) + \frac{\lambda}{N} \Gamma_\varepsilon(f_i, f_k)(x_j), \quad i, k \in \{1, \dots, n\}.$$

Indeed on Ω_+ , we have that $\Gamma_\varepsilon(f_i, f_k)(x_j) = \Gamma(f_i, f_k)(x_j)$ and the latter matrix is symmetric and positive semi-definite. $\square$

Once that we have the bounds (3.13), (3.14) at our disposal, we can derive the analogous results as in Theorems 3.7 and 3.8, with the estimates containing the error introduced with the approximation parameter $\varepsilon \in (0, 1)$.

Theorem 3.13. *Let $M > 0$ and let $f : \mathcal{S} \rightarrow [-M, M]$ be a measurable function and let $f^\star$ be defined as*

$$f^\star := -M \vee S_N^{\lambda, n}(f) \wedge M,$$

where $S_N^{\lambda, n}$ is as in Definition 3.1 with $\text{CE}_{2, \mathfrak{p}_N}^\varepsilon$ in place of $\text{pCE}_{2, \mathfrak{p}_N}$; let $r > 0$ be given and assume that $n, N \in \mathbb{N}$ satisfy (3.6). Then

$$\begin{aligned} \mathbb{E} \left[\|f - f^\star\|_{L^2(\mathcal{S}, \mathfrak{p})}^2 \right] \\ \leq 2e(f, n) \left(1 + \frac{c_{1/2}}{\log(N)(1+r)(1/2 + \lambda\mu_{\min, n}^\varepsilon)^2} \right) + 4\lambda\text{CE}_{2, \mathfrak{p}_N}^\varepsilon(P_n f) + 2M^2(N^{-r} + N\varepsilon), \end{aligned}$$

where $P_n f$ is the $L^2(\mathcal{S}, \mathfrak{p})$ -orthogonal projection of f onto $\mathcal{V}_n$ and

$$\begin{aligned} e(f, n) &:= \|P_n f - f\|_{L^2(\mathcal{S}, \mathfrak{p})}^2, \\ \mu_{\min, n}^\varepsilon &:= \min_{i=1, \dots, n} \text{CE}_{2, \mathfrak{p}}^\varepsilon(f_i). \end{aligned}$$

In case of noisy data, we can do the procedure as in Section 3.1: assume to know the target function f at the (random) points $x_1, \dots, x_N$ only up to some noise. In particular, we assume that we have at our disposal the noisy values

$$\tilde{f}(x_j) := f(x_j) + \eta_j, \quad j = 1, \dots, N, \quad (3.15)$$

where $(\eta_j)_{j=1}^N$ are i.i.d. random variables with variance bounded by $\sigma^2 > 0$. Notice that the (random) function $\tilde{f}$ belongs to $L^2(\mathcal{S}, \mathfrak{p}_N)$.

Theorem 3.14. *Let $M > 0$, let $f : \mathcal{S} \rightarrow [-M, M]$ be a measurable function and let $\tilde{f}^\star$ be defined as*

$$\tilde{f}^\star := -M \vee S_N^{\lambda, n}(\tilde{f}) \wedge M,$$

where $S_N^{\lambda, n}$ is as in Definition 3.1 with $\text{CE}_{2, \mathfrak{p}_N}^\varepsilon$ in place of $\text{pCE}_{2, \mathfrak{p}_N}$, and $\tilde{f}$ is as in (3.15) based on the noisy data $(y_j)_{j=1}^N$ whose i.i.d. errors $(\eta_j)_{j=1}^N$ have variance bounded by $\sigma^2 > 0$. Let $r > 0$ be given and assume that $n, N \in \mathbb{N}$ satisfy (3.6). Then

$$\begin{aligned} \mathbb{E} \left[\|f - \tilde{f}^\star\|_{L^2(\mathcal{S}, \mathfrak{p})}^2 \right] &\leq 2e(f, n) \left(1 + \frac{c_{1/2}}{\log(N)(1+r)(1/2 + \lambda\mu_{\min, n}^\varepsilon)^2} \right) + 8\lambda\text{CE}_{2, \mathfrak{p}_N}^\varepsilon(P_n f) + \\ &\quad + 4 \frac{\sigma^2}{(1 + \lambda\mu_{\min, n}^\varepsilon)^2} \frac{n}{N} + 2M^2(N^{-r} + N\varepsilon), \end{aligned}$$

where $e(f, n)$ and $\mu_{\min, n}^\varepsilon$ are as in Theorem 3.13.

ACKNOWLEDGMENTS

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This project has received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (grant agreement No. 101198055, project acronym NEITALG).

REFERENCES

- [1] L. Ambrosio, M. Colombo, and S. Di Marino. Sobolev spaces in metric measure spaces: reflexivity and lower semicontinuity of slope. *Adv. Stud. Pure Math.*, 67:1–58, 2015.
- [2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [3] L. Ambrosio, N. Gigli, and G. Savaré. Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces. *Rev. Mat. Iberoam.*, 29(3):969–996, 2013.
- [4] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 195(2):289–391, 2014.
- [5] L. Ambrosio, T. Ikonen, D. Lučić, and E. Pasqualetto. Metric Sobolev spaces I: Equivalence of definitions. *Milan J. Math.*, 92(2):255–347, 2024.
- [6] L. Ambrosio, T. Ikonen, D. Lučić, and E. Pasqualetto. Correction to: Metric Sobolev Spaces I: Equivalence of Definitions. *Milan J. Math.*, 93(2):551–555, 2025.
- [7] A. Björn and J. Björn. *Nonlinear potential theory on metric spaces*, volume 17 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2011.
- [8] J. Cheeger. Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.*, 9(3):428–517, 1999.
- [9] L. Chizat, G. Peyré, B. Schmitzer, and F. c.-X. Vialard. An interpolating distance between optimal transport and Fisher-Rao metrics. *Found. Comput. Math.*, 18(1):1–44, 2018.
- [10] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13(5):819–834, 2013.
- [11] A. Cohen, M. A. Davenport, and D. Leviatan. Correction to: On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 19(1):239, 2019.
- [12] N. De Ponti, G. E. Sodini, and L. Tamanini. The infimal convolution structure of the Hellinger-Kantorovich distance. 2025. arXiv 2503.12939.
- [13] L. Dello Schiavo. Massive Particle Systems, Wasserstein Brownian Motions, and the Dean-Kawasaki Equation. 2025. arXiv 2411.14936.
- [14] L. Dello Schiavo and G. E. Sodini. The Hellinger-Kantorovich metric measure geometry on spaces of measures. 2025. arXiv 2503.07802.
- [15] S. Di Marino. Sobolev and BV spaces on metric measure spaces via derivations and integration by parts, 2014. arXiv 1409.5620.
- [16] S. Di Marino, D. Lučić, and E. Pasqualetto. A short proof of the infinitesimal Hilbertianity of the weighted Euclidean space. *C. R. Math. Acad. Sci. Paris*, 358(7):817–825, 2020.
- [17] M. Fornasier, P. Heid, and G. E. Sodini. Approximation theory, computing, and deep learning on the Wasserstein space. *Math. Models Methods Appl. Sci.*, 35(4):825–903, 2025.
- [18] M. Fornasier, G. Savaré, and G. E. Sodini. Density of subalgebras of Lipschitz functions in metric Sobolev spaces and applications to Wasserstein Sobolev spaces. *J. Funct. Anal.*, 285(11):Paper No. 110153, 76, 2023.
- [19] N. Gigli and E. Pasqualetto. *Lectures on nonsmooth differential geometry*, volume 2 of *SISSA Springer Series*. Springer, Cham, 2020.
- [20] J. Heinonen, P. Koskela, N. Shanmugalingam, and J. T. Tyson. *Sobolev spaces on metric measure spaces*, volume 27 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2015. An approach based on upper gradients.
- [21] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11-12):1117–1164, 2016.
- [22] P. Koskela and P. MacManus. Quasiconformal mappings and Sobolev spaces. *Studia Math.*, 131(1):1–17, 1998.
- [23] M. Liero, A. Mielke, and G. Savaré. Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *Invent. Math.*, 211(3):969–1117, 2018.
- [24] G. Luise and G. Savaré. Contraction and regularizing properties of heat flows in metric measure spaces. *Discrete Contin. Dyn. Syst. Ser. S*, 14(1):273–297, 2021.
- [25] D. Lučić and E. Pasqualetto. Infinitesimal Hilbertianity of weighted Riemannian manifolds. *Canad. Math. Bull.*, 63(1):118–140, 2020.
- [26] E. Pasqualetto and G. E. Sodini. Functions of bounded variation and Lipschitz algebras in metric measure spaces. *ESAIM:COCV*, 32, 2026.
- [27] G. Savaré. Sobolev spaces in extended metric-measure spaces. In *New trends on analysis and geometry in metric spaces*, volume 2296 of *Lecture Notes in Math.*, pages 117–276. Springer, Cham, 2022.
- [28] G. Savaré and G. E. Sodini. A relaxation viewpoint to unbalanced optimal transport: duality, optimality and Monge formulation. *J. Math. Pures Appl.*, 188:114–178, 2024.

- [29] N. Shanmugalingam. Newtonian spaces: an extension of Sobolev spaces to metric measure spaces. *Rev. Mat. Iberoamericana*, 16(2):243–279, 2000.
- [30] G. E. Sodini. The general class of Wasserstein Sobolev spaces: density of cylinder functions, reflexivity, uniform convexity and Clarkson’s inequalities. *Calc. Var. Partial Differential Equations*, 62(7):Paper No. 212, 41, 2023.
- [31] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.

MASSIMO FORNASIER: TUM SCHOOL OF COMPUTATION, INFORMATION, AND TECHNOLOGY - DEPARTMENT OF MATHEMATICS, BOLTZMANNSTRASSE 3, 85748 GARCHING BEI MÜNCHEN (GERMANY)
& TUM INSTITUTE FOR ADVANCED STUDIES
& MUNICH DATA SCIENCE INSTITUTE
& MUNICH CENTER FOR MACHINE LEARNING
Email address: `massimo.fornasier@cit.tum.de`

GIACOMO ENRICO SODINI: INSTITUTE OF ANALYSIS AND SCIENTIFIC COMPUTING, TU WIEN, WIEDNER HAUPTSTRASSE 8-10, A-1040 VIENNA, AUSTRIA
Email address: `giacomo.sodini@tuwien.ac.at`